ISSN: 2301-5373
E-ISSN: 2654-5101
Volume 11 • Number 3 • February 2023

JELIKU

Jurnal Elektronik Ilmu Komputer Udayana

Informatics Study Program
Faculty of Mathematics and Natural Sciences
Udayana University

Table of Contents

SUSUNAN DEWAN REDAKSI
JURNAL ELEKTRONIK ILMU KOMPUTER UDAYANA (JELIKU)

This page is intentionally left blank.

Jurnal Elektronik Ilmu Komputer Udayana
Volume 11, No 3. February 2023

p-ISSN: 2301-5373
e-ISSN: 2654-5101

Diagnosis Penyakit Retinopati Diabetes Menggunakan SVM dengan Optimasi Algoritma Genetika

Ida Bagus Weda Baskara Adi Putra[a1], Luh Gede Astuti[a2], AAIN Eka Karyawati[a3] I Wayan Santiyasa[a4], Cokorda Rai Adi Pramartha[a5], I Gede Santi Astawa[a6]

[a]Program Studi Informatika, Universitas Udayana
Kuta Selatan, Badung, Bali, Indonesia
[1]wedabaskara219@gmail.com
[2]lg.astuti@unud.ac.id
[3]eka.karyawati@unud.ac.id
[4]santiyasa@unud.ac.id
[5]cokorda@unud.ac.id
[6]santi.astawa@unud.ac.id

Abstract

Diabetes mellitus is a non-communicable disease that is widely infected in the community. People with diabetes mellitus often do not realize that they are infected and are only known after complications occur. Diabetic retinopathy is one of the complications of diabetes mellitus. Diabetic retinopathy occurs when the vessels in the retina are damaged, causing visual disturbances. Diabetic retinopathy is divided into 2 stages, namely Non-Proliferative Diabetic Retinopathy (NPDR) and Proliferative Diabetic Retinopathy (PDR).This study diagnoses diabetic retinopathy using classification methods, namely SVM and GLCM as feature extraction methods. In addition, this study adds an optimization method, namely genetic algorithms to determine the optimal parameters for SVM. The amount of data used is 885 images with 3 labels, namely Normal, NPDR, and PDR. The test results in this study, SVM with genetic algorithms get better results than SVM without optimization. In SVM without F1 optimization the highest score was obtained at 0.7372 while in SVM with F1 optimization the highest score was obtained at 0.7578 with an increase in the percentage of F1 Score by 2.06%.
Keywords: *Diabetic Retinopathy, Diagnosis, SVM, GLCM, Genetic Algorithm, Parameter*

1. Pendahuluan

Penyakit diabetes terus meningkat prevelansinya diseluruh dunia. Berdasarkan data *International Diabetes Federation* (IDF) penyandang diabetes didunia sebanyak 463 juta orang dewasa pada awal tahun 2020 dengan prevalensi global mencapai 9,3 persen. Di Indonesia angka penderita diabetes mencapai 10,8 juta per tahun 2020 dan menempati peringkat ke-7 dari 10 negara pendertia diabetes terbanyak berdasarkan data *International Diabetes Federation* [1]. Penyakit diabetes melitus terjadi akibat kurangnya insulin di dalam tubuh sehingga kadar gula akan meningkat [2].

Penyandang diabetes melitus sering tidak menyadari bahwa ia terjangkit dan baru diketahui setelah terjadinya komplikasi[3]. Komplikasi dari penyakit diabetes melitus salah satunya adalah retinopati diabetes. Retinopati diabetes adalah gangguan pada mata akibat pembuluh darah pada retina rusak. Retinopati diabetes berpotensi menyebabkan kebutaan jika tidak diagnosis sedini mungkin.

Diagnosis retinopati diabetes dapat dilakukan dengan menggunakan metode klasifikasi. Klasifikasi adalah metode pengelompokan data berdasarkan label/kelas yang ada. Penelitian mengenai klasifikasi penyakit retinopati diabetes sebelumnya sudah pernah dilakukan pada tahun 2018. Dimana dalam penelitian tersebut dapat mengindenfikasi pasien pengidap retinopati diabetes melalui citra fundus retina. Metode klasifikasi yang digunakan terdapat 4, dimana metode *random forest* memperoleh hasil terbaik dan metode ekstraksi fitur yang digunakan adalah GLCM [4]. Selain itu penelitian lainnya dengan menggunakan *naïve bayes* dan GLCM juga dapat melakukan klasifikasi retinopati diabetes [5].

Pada penelitian ini menggunakan *Support Vector Machine* (SVM) pada metode klasifikasi dan GLCM pada metode ekstraksi fitur. Pada metode SVM terdapat parameter yang harus ditentukan seperti parameter C dan parameter lainnya. Penentuan nilai parameter yang tepat akan berpengaruh terhadap model SVM yang dihasilkan [6].

Penentuan parameter pada SVM biasanya dilakukan manual dengan cara mencoba setiap kombinasi nilai dan akan membutuhkan waktu lebih lama. Dari permasalahan tersebut diperlukan suatu teknik yang dapat mengoptimasi pemilihan nilai parameter tersebut dengan cara otomatis yaitu dengan menggunakan algoritma genetika.

Penelitian mengenai SVM dengan algoritma genetika sebagai optimasi sebelumnya sudah pernah dilakukan pada tahun 2019. Dimana pada penelitian tersebut membandingkan SVM dengan optimasi dan tanpa optimasi dimana SVM dengan optimasi memperoleh akurasi lebih baik dibandingkan dengan metode SVM tanpa optimasi [7]. Selain itu terdapat juga penelitian lainnya yang menggunakan algoritma genetika sebagai penentuan nilai parameter optimal pada SVM pada tahun 2017. Pada penelitian tersebut penentuan nilai parameter SVM dengan algoritma genetika memperoleh hasil lebih baik [8].

Dari penjabaran diatas maka dalam penelitian ini akan menggunakan algoritma genetika dalam menentukan parameter pada SVM untuk diagnosis penyakit retinopati diabetes berdasarkan ciri tekstur pada citra fundus retina dengan GLCM.

2. Metode Penelitian

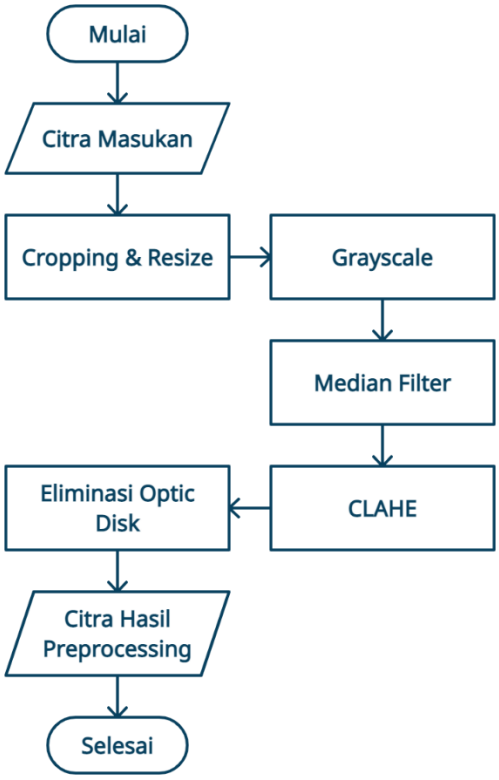
2.1. Pengumpulan Data

Data yang digunakan pada penelitian ini adalah data sekunder dari APTOS 2019 *Blindness Detection resized version* yang didapat dari kaagle.com. Jumlah dataset yang digunakan adalah 800 data citra fundus retina. Dimana dataset yang didapatkan akan dibagi menjadi 3 label yaitu Normal, *Non Proliferative Diabetic Retinopathy* (NPDR) , dan *Proliferative Diabetic Retinopathy* (PDR). NPDR adalah fase awal retinopati diabetik yang ditandai dengan adanya microaneurisma, pendarahan, dan eksudat pada retina. PDR adalah fase selanjutnya yang ditandai dengan adanya neovaskularisasi dan pendarahan pada vitreous [9].

2.2. *Preprocessing*

Seperti terlihat pada Gambar 1, tahap *preprocessing* diawali dengan melakukan *cropping* untuk menghilangkan area hitam. Setelah melalui tahap *cropping*, setiap data citra diresize kedalam dimensi 200x200 piksel. Kemudian setiap citra yang telah diresize dikonversi ke *greyscale*. Setelah itu setiap citra difilter menggunakan *median filter* yang bertujuan untuk menghilangkan *noise*. Setelah *noise* dihilangkan citra ditingkatkan menggunakan metode CLAHE. *Contrast Limited Adaptive Histogram Equalization* (CLAHE) digunakan untuk memperoleh kontras yang lebih baik sehingga fitur eksudat, mikroaneurisma dan pembuluh darah lebih terlihat jelas. Dilanjutkan tahap eliminasi *optic disk*. Eliminasi optic disk dilakukan dengan cara mencari nilai intesitas maksimum pada citra karena *optic disk* pada retina biasanya memiliki nilai intesitas paling tinggi. Kemudian nilai dari intesitas tersebut dapat digunakan untuk mencari letak titik dari nilai intesitas tersebut. Pada titik tersebut dapat menggunakan *masking* dengan objek lingkaran untuk menghilangkan *optic disk*.

p-ISSN: 2301-5373
e-ISSN: 2654-5101



Gambar 1. *Preprocessing* Citra

2.3. Ekstraksi Fitur dengan GLCM

Setelah tahap *preprocessing* citra dilanjutkan ke tahap ekstraksi fitur menggunakan GLCM, seperti terlihat pada Gambar 2.

Gambar 2. Tahap Ekstraksi Fitur

Gray Level Co-Occurance Matrix (GLCM) adalah sebuah metode yang dapat digunakan dalam menganalisis dan mengekstrasi fitur pada sebuah citra. GLCM bekerja dengan membuat sebuah matriks yang digunakan untuk menggambarkan hubungan antara dua piksel pada jarak dan arah yang sudah ditentukan dalam suatu gambar [10]. Ekstrasi fitur pada GLCM dilakukan dalam 4 arah sudut yaitu 0, 45, 90, 135.

Adapun beberapa fitur yang dapat diekstraksi oleh GLCM, yaitu

a. *Contrast*

Contrast adalah fitur yang digunakan untuk mengukur tingkat perbedaan intensitas pada citra.

$$Contrast = \sum_{i=1}^{k}\sum_{j=1}^{k}(i-j)^2 P_{ij} \quad (1)$$

b. *Correlation*

Correlation adalah fitur untuk mengukur korelasi antar pixel pada citra.

$$Correlation = \sum_{i=1}^{k}\sum_{j=1}^{k}\frac{(i-m_4)(j-m_c)pij}{\sigma_r\sigma_c} \quad (2)$$

c. *Energy*

Energy adalah fitur untuk mengukur keseragaman intensitas pada citra.

$$Energy = \sum_{i=1}^{k}\sum_{j=1}^{k} p^2 ij \quad (3)$$

d. *Entropy*

Entropy adalah fitur untuk mengukur ketidakaturan distribusi intensitas suatu citra.

$$Entropy = \sum_{i=1}^{k}\sum_{j=1}^{k} p_{ij} log_2 P_{ij} \quad (4)$$

e. *Homogeneity*

Homogeneity adalah untuk mengukur kehomogenan variasi intensitas suatu citra.

$$Homogeneity = \sum_{i=1}^{k}\sum_{j=1}^{k}\frac{Pij}{1+|i-j|} \quad (5)$$

2.4. *Min – Max Normalization*

Selanjutnya nilai hasil ekstraksi fitur GLCM ditranformasi ke skala 0 sampai 1 menggunakan metode *min – max*. Berikut rumus dari metode *min - max* :

$$\text{v'} = \frac{v-min_A}{max_A - min_A}\ (\text{new\_maxA} - \text{new\_minA}) + \text{new\_minA} \quad (6)$$

Keterangan :

v′ : Nilai data yang telah dinormalisasi

v : Nilai data yang belum dinormalisasi

maxA : Nilai tertinggi dari data

minxA : Nilai terkecil dari data

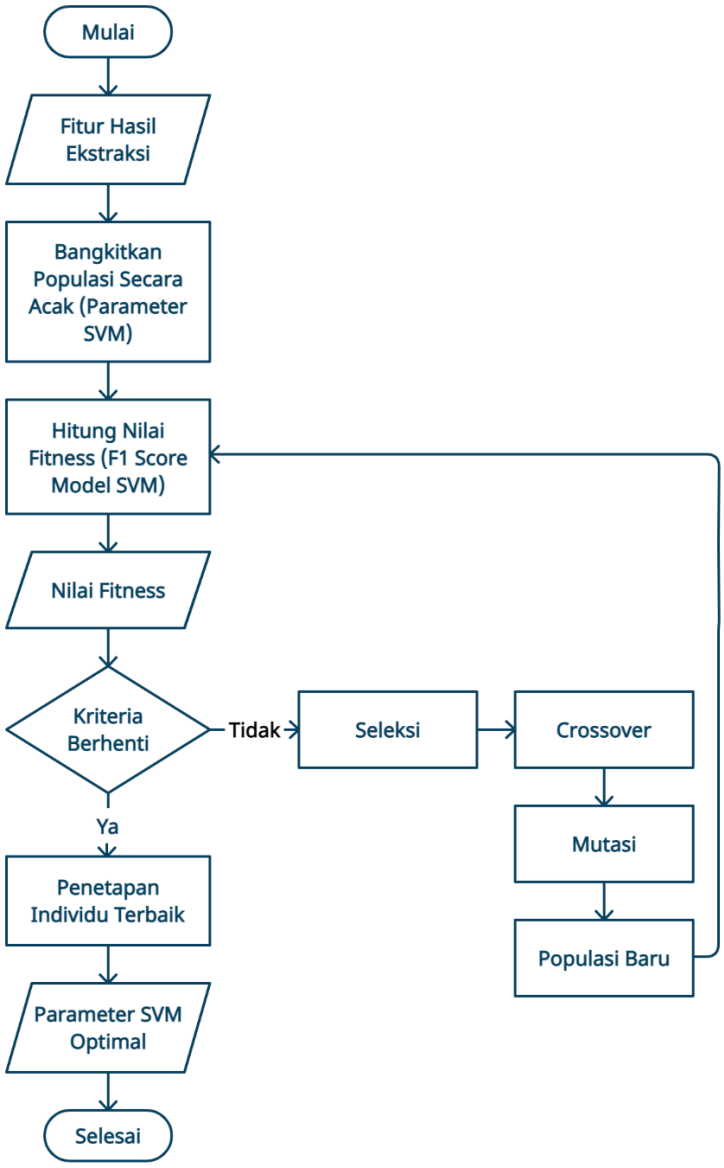
new_maxA : Nilai tertinggi data baru

new_minA : Nilai terkecil data baru

p-ISSN: 2301-5373
e-ISSN: 2654-5101

2.5. Penentuan Parameter Optimal pada SVM dengan Algoritma Genetika

Proses penentuan parameter optimal pada SVM dilakukan dengan menggunakan algoritma genetika. *Flowchart* dari proses penentuan nilai parameter SVM dengan algoritma genetika dapat dilihat pada Gambar 3.



Gambar 3. *Flowchart* penentuan parameter SVM

2.5.1. Algoritma Genetika

Algoritma genetika adalah algoritma pencarian yang berlandaskan pada prinsip seleksi alam. Algoritma genetika dapat digunakan sebagai solusi untuk permasalahan optimasi pada pencarian.

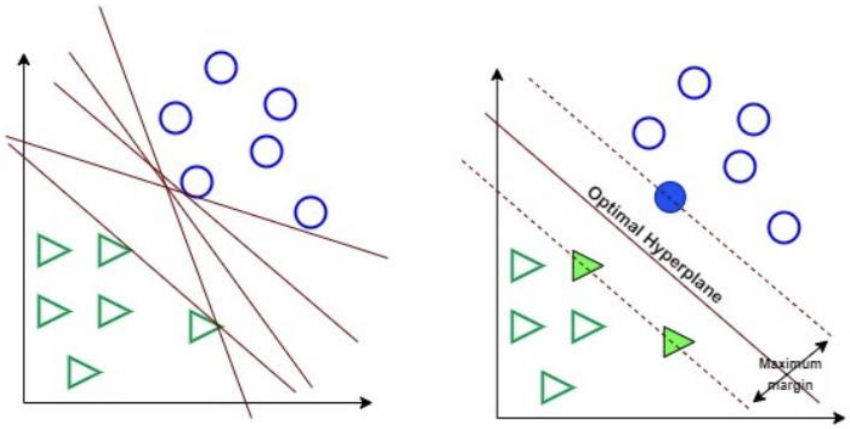
Terdapat struktur secara umum dari algoritma genetika, sebagai berikut [11].

1. Membangkitkan populasi awal
   Tahapan awal dalam algoritma genetika dimulai dengan pembangkitkan populasi yang dibentuk dari sekumpulan N individu secara *random*. Setiap individu akan memiliki sebuah kromosom. Individu - individu ini mempresentasikan sekumpulan solusi yang diinginkan. Dalam penelitian ini jumlah kromosom setiap individu adalah sama dengan jumlah parameter pada SVM yang ingin dioptimalkan

2. Evaluasi *Fitness*
   Setiap individu pada populasi akan dilakukan evaluasi dengan menghitung nilai fitnessnya. Nilai *fitness* dalam penelitian ini adalah F1 *Score* dari model klasifikasi SVM. Nilai F1 *Score* masing – masing individu akan dievaluasi sampai terpenuhi kriteria berhenti, jika tidak terpenuhi populasi baru akan dibentuk.
3. Seleksi
   Proses ini menyeleksi individu yang akan digunakan pada proses crossover nanti sebagai orang tua. *Roulette – Wheel* merupakan salah satu metode yang dapat digunakan untuk melakukan proses seleksi. Langkah pertama untuk menggunakan seleksi *Roulette – Wheel* adalah membuat interval nilai kumulatif dari rank nilai fitnes masing – masing individu dibagi total rank semua individu. Langkah selanjutnya membangkitkan bilangan *random*. Jika bilangan *random* berada dalam interval nilai kumulatif suatu individu maka individu tersebut akan dipilih.
4. *Crossover*
   Orang tua yang telah dipilih pada proses seleksi akan dilakukan *crossover* (kawin silang) untuk menghasilkan individu baru. Tetapi sebelum itu ditentukan terlebih dahulu apakah *crossover* akan dilakukan atau tidak dengan membangkitkan bilangan *random*. Jika bilangan *random* yang dibangkitkan lebih kecil dari probabilitas *crossover* (pc) yang ditentukan maka *crossover* dilakukan.
5. Mutasi
   Mutasi merupakan proses mengubah nilai gen dalam suatu individu. Proses mutasi dilakukan jika kondisi berikut terpenuhi yaitu bilangan *random* yang dibangkitkan suatu individu lebih kecil dari probabilitas mutasi (pm) yang telah ditentukan.

2.5.2. *Support Vector Machine* (SVM)

SVM adalah metode klasifikasi yang bekerja dengan cara membuat sebuah garis pemisah untuk memisahkan kelas yang berbeda yaitu +1 dan -1. Garis pemisah ini disebut dengan *hyperplane*.



Gambar 4. *Hyperplane* optimal SVM (Sumber : www.dicoding.com)

Untuk mencari *hyperplane* yang optimal dapat ditemukan dengan mengukur atau mencari margin dari *hyperplane* tersebut yang maksimal. Jarak antara *hyperplane* dengan *support vector* pada masing – masing kelas disebut dengan margin [12]. Pada Gambar 4, segita berwarna hijau dan lingkaran berwarna biru merupakan sebuah *support vector*.
SVM dapat digunakan untuk melakukan klasifikasi pada kasus *linear* dan *non linear*. Pada kasus *non linear* SVM menggunakan konsep kernel pada ruang berdimensi tinggi.

Berikut merupakan fungsi kernel pada SVM *non linear* :
1. Kernel *Linear*
   $$K(x, y) = x.y \quad (7)$$
   Keterangan :
   x, y = fitur data

2. Kernel *Polynomial*

$$K(x, y) = (\gamma(x. y) + r)^d \quad (8)$$

Keterangan :
x, y = fitur data
r = koefisien *polynomial*
d = derajat *polynomial*
γ = *control* variabel

3. Kernel *Radial Basis Function* (RBF)

$$K(x, y) = exp\ (-\gamma||x-y||^2) \quad (9)$$

Keterangan :
x, y = fitur data
γ = *control* variabel

Untuk melakukan klasifikasi menggunakan SVM dapat menggunakan persamaan berikut :

$$u = \sum_{i=1}^{N} y_i \alpha_i K(x_i, x) - b \quad (10)$$

Keterangan :
u = hasil klasifikasi
y_i = kelas ke – i (+1 atau -1)
α_i = *support vector* dari data ke – i
$K(x_i,x)$ = fungsi kernel
b = nilai bias

Nilai *support vector* dan bias dapat ditemukan pada proses pembelajaran SVM. Metode pembelajaran SVM yang dapat digunakan adalah metode *Sequential Minimal Optimization* (SMO).

3. Hasil dan Pembahasan

Pengujian dilakukan dengan menghitung nilai rata – rata F1 *Score* menggunakan k *fold cross validation* pada proses klasifikasi SVM dengan algoritma genetika. Kemudian dibandingkan dengan SVM tanpa optimasi. Nilai k yang digunakan pada k *fold cross validation* adalah 5. Dan kernel SVM yang akan diuji adalah kernel *polynomial*. Terdapat beberapa parameter pada kernel *polynomial* yakni C, *gamma*, r dan *degree*.

3.1. Pengujian Tanpa Optimasi

Pada pengujian ini dilakukan pengujian model SVM kernel *polynomial* dengan menggunakan beberapa kombinasi nilai parameter yang telah ditentukan.

Tabel 1 menyajikan hasil k *fold cross validation* untuk model SVM dengan kernel *polynomial* menggunakan 16 kombinasi nilai parameter. Setiap kombinasi nilai parameter memperoleh nilai rata - rata F1 Score yang berbeda – beda. Dimana pada tabel tersebut terlihat nilai rata – rata F1 *Score* terbaik diperoleh sebesar 0.7372 dengan mengunakan kombinasi nilai parameter *degree* = 4, r = 1, *gamma* = 1 dan C = 0.5.

Tabel 1. Hasil pengujian SVM tanpa optimasi pada kernel *polynomial*

Degree	r	gamma	C	F1 Score
2	1	0.50	0.50	0.6432
2	1	0.50	1.00	0.6924
2	1	1.00	0.50	0.7222
2	1	1.00	1.00	0.6992
3	1	0.50	0.50	0.6740
3	1	0.50	1.00	0.6725
3	1	1.00	0.50	0.6970
3	1	1.00	1.00	0.7092
4	1	0.50	0.50	0.6547
4	1	0.50	1.00	0.7276
4	1	1.00	0.50	0.7372
4	1	1.00	1.00	0.7283
5	1	0.50	0.50	0.7362
5	1	0.50	1.00	0.7016
5	1	1.00	0.50	0.7117
5	1	1.00	1.00	0.7210

3.2. Pengujian SVM dengan Optimasi

Pada pengujian SVM dengan optimasi terdapat 3 pengujian yaitu pengujian nilai probabilitas mutasi, nilai probabilitas *crossover* dan jumlah populasi.

3.2.1. Pengujian Probabilitas Mutasi

Pengujian probabilitas mutasi dilakukan 4 kali dimana nilai yang diuji 0.1, 0.2, 0.3, 0.4 dengan menggunakan nilai probabilitas *crossover* 0.6 dan jumlah populasi 10.

Tabel 2 menyajikan nilai parameter yang dihasilkan dari algoritma genetika serta hasil k *fold cross validation* untuk model SVM dengan kernel *polynomial*. Pada tabel tersebut terlihat nilai rata - rata F1 *score* terbaik adalah 0.7578 dengan nilai parameter *degree* = 5, r = 0.40586, *gamma* = 0.59903 dan C = 0.43708. Nilai parameter ini dihasilkan dari algoritma genetika ketika menggunakan nilai probabilitas mutasi 0.2.

Tabel 2. Hasil pengujian nilai mutasi algoritma genetika pada SVM kernel *polynomial*

Pm	d	r	Gamma	C	F1 Score
0.1	5	0.29757	0.55123	0.68195	0.7455
0.2	5	0.40586	0.59903	0.43708	0.7578
0.3	4	0.69073	0.64586	0.69073	0.7411
0.4	5	0.57659	0.76878	0.13953	0.7437

p-ISSN: 2301-5373
e-ISSN: 2654-5101

3.2.2. Pengujian Probabilitas *Crossover*

Pengujian probabilitas *crossover* dilakukan 4 kali dimana nilai yang diuji 0.6, 0.7, 0.8, 0.9 dengan menggunakan nilai probabilitas mutasi 0.1 dan jumlah populasi 10.

Tabel 3 menyajikan nilai parameter yang dihasilkan dari algoritma genetika serta hasil k *fold cross validation* untuk model SVM dengan kernel *polynomial*. Pada tabel tersebut terlihat nilai rata - rata F1 *score* terbaik adalah 0.7577 dengan nilai parameter *degree* = 5, r = 0.12197, *gamma* = 0.76390 dan C = 0.17465. Nilai parameter ini dihasilkan dari algoritma genetika ketika menggunakan nilai probabilitas *crossover* 0.7.

Tabel 3. Hasil pengujian nilai *crossover* algoritma genetika pada SVM kernel *polynomial*

Pc	d	r	Gamma	C	F1 Score
0.6	5	0.29757	0.55123	0.68195	0.7455
0.7	5	0.12197	0.76390	0.17465	0.7577
0.8	5	0.40489	0.81756	0.17953	0.7460
0.9	5	0.69854	0.89756	0.39708	0.7415

3.2.3. Pengujian Jumlah Populasi

Jumlah populasi yang diuji adalah 10, 20, 30, 40 dengan menggunakan probabilitas *crossover* 0.6 dan probabilitas mutasi 0.1.

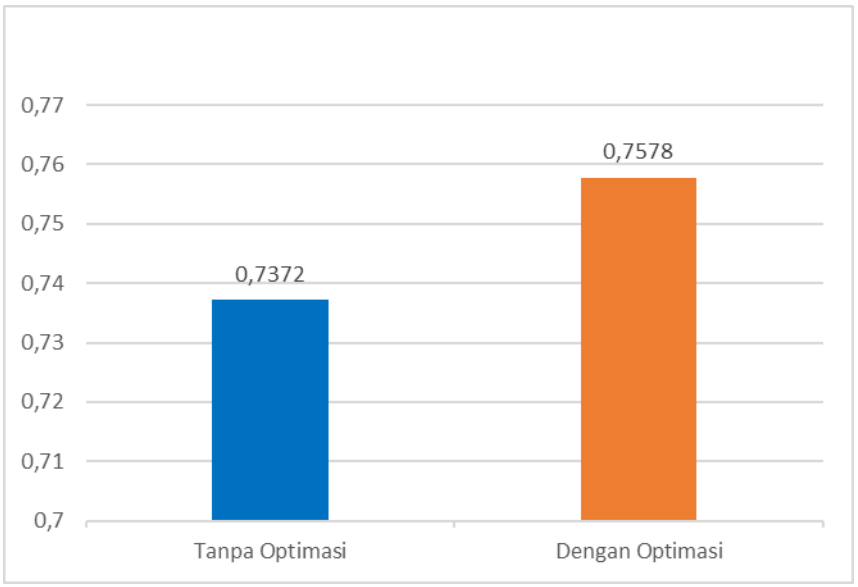
Tabel 4 menyajikan nilai parameter yang dihasilkan dari algoritma genetika serta hasil k *fold cross validation* untuk model SVM dengan kernel *polynomial*. Pada tabel tersebut terlihat nilai rata - rata F1 *score* terbaik adalah 0.7524 dengan nilai parameter *degree* = 4, r = 0.90829, *gamma* = 0.95804 dan C = 0.81366. Nilai parameter ini dihasilkan dari algoritma genetika ketika menggunakan jumlah populasi 40.

Tabel 4. Hasil pengujian jumlah populasi algoritma genetika pada SVM kernel *polynomial*

Pop	d	r	Gamma	C	F1 Score
10	5	0.29757	0.55123	0.68195	0.7455
20	4	0.40099	0.65561	0.92878	0.7404
30	4	0.76976	0.93170	0.32001	0.7405
40	4	0.90829	0.95804	0.81366	0.7524

3.3. Perbandingan Tanpa Optimasi dan Dengan Optimasi

Berikut perbandingan nilai F1 Score tertinggi yang diperoleh dari masing - masing pengujian SVM tanpa optimasi dan dengan optimasi.



Gambar 5. Perbandingan F1 Score tanpa optimasi dan dengan optimasi

Berdasarkan pada Gambar 5, SVM dengan optimasi memperoleh hasil lebih baik dibandingkan tanpa optimasi dalam diagnosis retinopati diabetik berdasarkan ciri tekstur pada citra fundus retina pada penelitian ini. SVM tanpa optimasi memperoleh F1 *Score* sebesar 0.7372 sedangan dengan optimasi F1 *Score* yang diperoleh sebesar 0.7578. Dengan menggunakan optimasi, persentase F1 *Score* yang diperoleh meningkat sebesar 2,06 %.

4. Kesimpulan

Pada pengujian algoritma genetika, perubahan parameter probabilitas *crossover*, mutasi dan jumlah populasi berpengaruh terhadap nilai F1 *Score* yang diperoleh. Jika nilai probabilitas *crossover* dan mutasi semakin besar, F1 *Score* yang diperoleh cenderung menurun. Sedangkan pada jumlah populasi, jika nilainya semakin besar, F1 *Score* yang diperoleh cenderung meningkat.

Model klasifikasi SVM dengan optimasi menghasilkan nilai F1 *Score* yang lebih tinggi dibandingkan dengan tanpa optimasi. Pada SVM tanpa optimasi F1 *score* tertinggi diperoleh sebesar 0.7372 sedangkan pada SVM dengan optimasi F1 *score* tertinggi diperoleh sebesar 0.7578 dengan kenaikan persentase F1 *Score* sebesar 2,06 %.

Daftar Pustaka

[1] IDF, IDF Diabetes ATLAS, 9th ed, Belgia: International Diabetes Federation, 2019.
[2] D. Hardianto, “Telaah Komprehensif Diabetes Melitus: Klasifikasi, Gejala, Diagnosis, Pencegahan, Dan Pengobatan” *Bioteknologi & Biosains Indonesia*, vol. 7, no. 2, p. 304 – 317, 2020.
[3] Kemenkes RI, “Profil Kesehatan Indonesia 2014”, Jakarta: Kemenkes RI, 2014.
[4] A. Anitha dan T. Sridevi, “Classifying Diabetic Retinopathy In Retinal Images Utilizing GLCM And Evolutionary PSO Features” *International Journal of Computer Engineering and Applications*, vol 12, no. 3, p. 168 – 178, 2018.
[5] Erwin, A. L. Nurjanah, S. D. Noviyanti dan Yurika, “Klasifikasi Penyakit Diabetik Retinopathy dengan Metode Naïve Bayes pada Citra Retina” *Annual Research Seminar*, vol. 4, no. 1, p. 126 – 131, 2018.

[6] L. C. Huang dan J. C. Wang, “A Ga-based Feature Selection and Parameters Optimization for Support Vector Machines” *Expert Systems with Application*, vol. 31, no. 2, p. 231 – 240, 2006.
[7] H. Harfani dan A. Maulana, “Penerapan Algoritma Genetika pada Support Vector Machine Sebagai Pengoptimasi Parameter untuk Memprediksi Kesuburan” *Jurnal Teknik Informatika STMIK Antar Bangsa*, vol. 5, no. 1, p. 51-59, 2019.
[8] J. Manurung, H. Mawengkang dan E. Zamzani, “Optimizing Support Vector Machine Parameters with Genetic Algorithm for Credit Risk Assessment” *Journal of Physics*, vol. 930, no. 1, p. 1 – 5, 2017.
[9] A. S. Soelistijo, Pengelolaan dan Pencegahan Diabetes Melitus Tipe 2 Dewasa di Indonesia, Jakarta: PB PARKENI, 2019.
[10] Rizal, "Analisis Gray Level Co-Occurrence Matrix (GLCM) Dalam Mengenali Citra Ekspresi Wajah" *Jurnal Mantix*, vol. 3, no. 2, pp. 31-38, 2019.
[11] P.G.L.N Suwirmayanti, M. I. Sudarsana dan S. Darmayasa, “Penerapan Algoritma Genetika Untuk Penjadwalan Mata Pelajaran” *Journal of Applied Intelligent System*, vol. 1, no. 3, p. 220-233, 2016.
[12] M. I. Rosadi, C. B. Sanjaya dan L. Hakim, “Klasifikasi Diabetic Retinopathy Menggunakan Seleksi Fitur Dan Support Vector Machine” *Jurnal Rekayasa Sistem Komputer System*, vol. 1, no. 2, p. 109 - 117, 2018.

This page is intentionally left blank.

Jurnal Elektronik Ilmu Komputer Udayana
Volume 11, No 3. February 2023

p-ISSN: 2301-5373
e-ISSN: 2654-5101

Pemodelan Topik Artikel Berita Menggunakan *Structural Topic Model* dan *Latent Dirichlet Allocation*

Ayu Kadek Nadya Oktaviana[a1], Ngurah Agus Sanjaya ER[a2], Ida Bagus Made Mahendra[a3], I Gede Santi Astawa[a4], I Gede Arta Wibawa[a5], I Komang Ari Mogi[a6]

[a]Program Studi Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam,
Universitas Udayana
Bali, Indonesia
[1]oktaviananadya5@gmail.com
[2]agus_sanjaya@unud.ac.id
[3]ibm.mahendra@unud.ac.id
[4]santi.astawa@unud.ac.id
[5]gede.arta@unud.ac.id
[6]arimogi@unud.ac.id

Abstract

An online news portal is one of the technologies in the form of online media that provides information services in the form of news articles. The number of news articles on online news portals continues to grow over time and more news article data will be available. A large amount of data is a challenge in itself to be processed into a more useful form, namely by conducting topic modeling based on news article data so that the data can be categorized based on the topics discussed in it. Topic modeling groups text data into a specific set of topics based on their similarities. In this study, the dataset used was 44,425 news articles from November 2021 to March 2022 which were taken from the online news portal detik.com. News exploration was carried out by topic modeling using two methods, Latent Dirichlet Allocation (LDA) and Structural Topic Model (STM). The LDA method produces 8 topics derived from the calculation of the highest probabilistic coherence value. The STM method produces 11 topics based on the highest semantic coherence and exclusivity values.

Keywords: *topic modelling, latent dirichlet allocation, structural topic model, news articles, text mining*

1. Pendahuluan

Berita merupakan sarana yang digunakan oleh masyarakat untuk memperoleh informasi mengenai suatu peristiwa atau kejadian faktual secara tertulis. Teknologi yang semakin berkembang menjadikan berita lebih mudah untuk dijangkau oleh masyarakat, salah satunya dengan adanya portal berita *online*. Banyaknya jumlah portal berita *online* yang ada maka artikel berita yang beredar juga semakin banyak. Data berita terus bertambah seiring berjalannya waktu sehingga menjadi tantangan tersendiri untuk mengolah data berita menjadi bentuk yang lebih bermanfaat yaitu dengan menemukan topik tersembunyi dari sekumpulan berita melalui pemodelan topik. Menemukan topik dari sekumpulan berita, dapat menghemat lebih banyak waktu daripada membaca semuanya untuk mengetahui peristiwa yang terjadi dalam jangka waktu tertentu.

Pemodelan topik adalah teknik pembelajaran mesin tanpa pengawasan untuk mengeksplorasi dan mengungkapkan struktur semantik dari sekumpulan dokumen berbentuk teks [1]. Pemodelan topik menggunakan pendekatan *bag-of-words* dimana makna kalimat tidak dievaluasi. Sebaliknya, pemodelan topik mengevaluasi frekuensi kata-kata. Oleh karena itu diasumsikan bahwa kata-kata yang paling sering muncul dalam suatu topik akan menunjukkan tentang topik tersebut [2]. Pemodelan topik telah membuktikan dirinya sebagai alat untuk analisis eksplorasi

sekumpulan dokumen, terutama untuk menemukan topik. Salah satunya adalah penelitian menemukan distribusi topik kemudian klasterisasi dokumen dari cerita berbahasa Bali menggunakan *Latent Dirichlet Allocation* (LDA) [3]. Penelitian lain yang telah dilakukan yaitu menemukan topik dari sekumpulan judul berita menggunakan *Latent Dirichlet Allocation* (LDA) berdasarkan hasil analisis sentimen yang menghasilkan lima topik dari masing-masing sentimen yaitu positif, negatif, dan netral [4]. Selanjutnya penelitian menemukan topik dari portal berita *online* selama masa Pembatasan Sosial Berskala Besar (PSBB) menggunakan metode *Latent Dirichlet Allocation* (LDA) yang menghasilkan empat kelompok besar topik pemberitaan [5].

Penelitian yang disebutkan sebelumnya memiliki kesamaan yaitu menggunakan metode *Latent Dirichlet Allocation* (LDA) untuk menemukan suatu topik dari sekumpulan dokumen. Namun, pemodelan topik menggunakan LDA sering mengabaikan metadata dokumen lain yang bisa mempengaruhi penemuan topik, seperti tanggal publikasi berita, lokasi, penulis, kategori, dan lain-lain. Oleh karena itu pada penelitian ini, penulis melakukan pemodelan topik artikel berita berbahasa Indonesia menggunakan metode lain yaitu *Structural Topic Model* (STM). STM memungkinkan untuk mengeksplorasi korelasi topik dan memahami bagaimana metadata dokumen berhubungan dengan distribusi topik [5]. Penelitian *topic modelling* menggunakan STM telah banyak dilakukan oleh peneliti lain seperti penelitian deteksi topik laten dan tren pada artikel jurnal teknologi pendidikan menggunakan metadata dokumen berupa wilayah negara atau region dan institusi sebagai penentu hasil topik [6]. Penelitian lain menerapkan STM pada 4636 makalah penelitian S2ORC yang berhasil mengungkapkan dua belas topik penelitian, mengetahui korelasi antar topik, dan mengevaluasi topik dari waktu ke waktu [7].

Berdasarkan penelitian-penelitian terdahulu, hasil pada penelitian ini dapat digunakan untuk mengetahui perbandingan topik yang dihasilkan dari metode LDA dan STM dari kumpulan artikel berita berbahasa Indonesia dan mengetahui isu yang terjadi dalam kurun waktu tertentu di masyarakat.

2. Metode Penelitian

2.1. Deskripsi Data

Data yang digunakan pada penelitian ini berasal dari portal berita *online* detik (www.detik.com). Selama lima bulan dari 1 November 2021 sampai 31 Maret 2022, data diambil menggunakan teknik *web scraping* dengan bantuan *framework* Scrapy. Hasil akhir pengumpulan data disimpan dalam format CSV. Tabel 1 menunjukkan deskripsi dari data yang telah dikumpulkan.

Tabel 1. Deskripsi Dataset

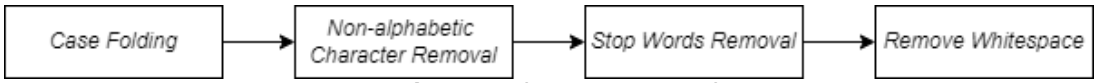
Tanggal	November 2021 – Maret 2022
Sumber	www.detik.com
Tipe Dokumen	Artikel Berita
Fitur	Title, Year, Month, Day, Auhtor, Location, Content
Total	44.425

2.2. *Data Preprocessing*

Preprocessing adalah proses mengolah data yang tidak terstruktur menjadi lebih struktur. Terdiri dari penyetaraan teks, menghilangkan angka, tanda baca, karakter khusus, mengekstrak *stopwords*, dan menghapus spasi kosong. Data teks dibersihkan dengan menghilangkan atau mengubah kata-kata yang tidak bernilai. Semua kata diubah menjadi huruf kecil, dan tanda baca serta spasi dihapus. Karakter khusus dan *stopwords* dihapus karena sering kali tidak berkontribusi pada identifikasi topik. Contoh *stopwords* adalah “dan”, “yang”, dan “ia”. Kata-kata

p-ISSN: 2301-5373
e-ISSN: 2654-5101

ini tidak menambah nilai tentang topik. Gambar 1 menunjukkan tahap *preprocessing* yang dikerjakan pada penelitian ini.



Gambar 1. Tahap *Preprocessing*

2.3. *Document Term Matrix* (DTM)

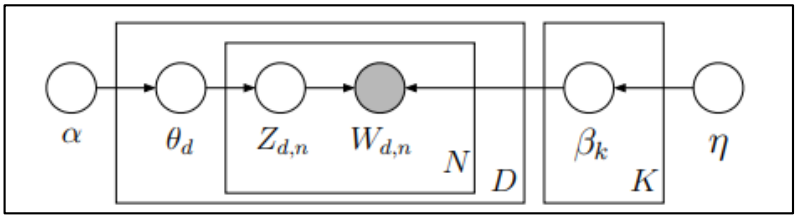
Topic modelling menganggap kata sebagai unit dasar dari sebuah dokumen. Kumpulan kata dalam *corpus* atau sekumpulan dokumen disebut kosa kata (*vocabulary*). *Topic modelling* bergantung pada asumsi *bag-of-words* yang berarti bahwa kata-kata dalam dokumen dapat ditukar sehingga mengabaikan urutan kata dan menghitung frekuensi atau kemunculan kata dari dokumen [8]. Hal ini mengarah pada representasi koleksi dokumen sebagai matriks istilah dokumen atau *document-term-matrix* (DTM) di mana frekuensi kata-kata dalam dokumen dihitung [9]. DTM terdiri dari baris yang mewakili dokumen asli, kolom yang mewakili setiap kata dalam korpus, dan setiap sel berisi frekuensi kemunculan kata tertentu [10].

2.4. *Latent Dirichlet Allocation* (LDA)

Latent Dirichlet Allocation (LDA) mengasumsikan dokumen dan kata-kata sebagai campuran acak atas topik tersembunyi (laten) dimana setiap topik dianggap sebagai distribusi kata yang berasal dari model probabilistik generatif [8]. Model probabilistik generatif bekerja dengan mengamati data, kemudian menghasilkan data yang mirip dengannya untuk memahami data yang diamati. LDA sebagai model probalistik generatif [11]:

(1) Untuk setiap topik,
   (a) Pilih distribusi di atas kata-kata $\beta_k \sim \text{Dir}V(\eta)$.

(2) Untuk setiap dokumen,
   (a) Pilih vektor proporsi topik $\theta_d \sim$ Dir(α).
   (b) Untuk setiap kata dalam dokumen,
      (i) Pilih penugasan topik $Z_{d,n} \sim$ Multinomial(θ_d), $Z_{d,n} \epsilon \{1,..,K\}$.
      (ii) Pilih kata $W_{d,n} \sim$ Multinomial($\beta_{z,d,n}$), $W_{d,n} \epsilon \{1,\ldots,K\}$.

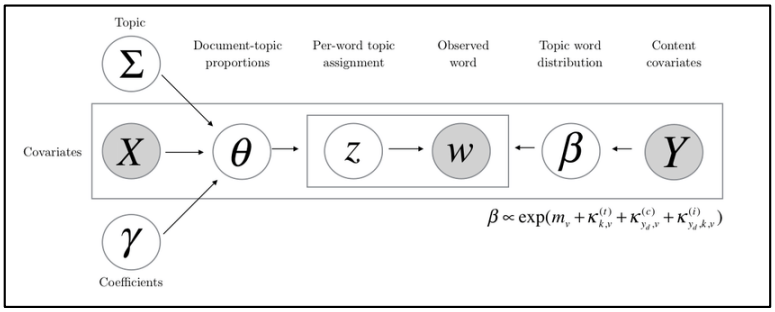


Gambar 2. Model Grafik *Latent Dirichlet Allocation* (LDA)

2.5. *Structural Topic Model* (STM)

Structural Topic Model (STM) merupakan perluasan dari LDA yang mengakomodasi struktur korpus atau metadata dokumen ke dalam model topik [12]. Pendekatan STM dapat digunakan untuk mengeksplorasi korelasi topik dan memahami bagaimana metadata dokumen (kovariat) berhubungan dengan distribusi topik di atas dokumen (prevalensi topik) dan distribusi kata di atas topik (konten topik) [13]. Prevalensi topik adalah tingkat topik yang ada di setiap dokumen. Konten topik adalah frekuensi kata yang digunakan dalam topik.

Pada Gambar 2 yaitu model grafik dari LDA, distribusi topik (θ) bergantung pada parameter *dirichlet* (α) yang menghasilkan distribusi multinomial. Dengan demikian, distribusi topik pada semua dokumen dipengaruhi oleh prior tunggal tersebut. Demikian juga pada distribusi kata (β) terhadap topik juga dikendalikan oleh *dirichlet* sebelumnya (η). Akibatnya, tidak ada kovariat tingkat dokumen yang mempengaruhi distribusi topik (prevalensi topik) dan distribusi kata (konten topik).



Gambar 3. Model Grafik *Structural Topic Modelling* (STM)

Sebaliknya, pada STM prevalensi topik dapat dipengaruhi oleh satu atau lebih kovariat tingkat dokumen (X), yang menghasilkan distribusi unik atas topik untuk setiap dokumen (Gambar 3). Hal ini dilakukan untuk menerapkan *logistic normal linier prior* [$\theta \sim \text{LogisticNormal}(X)$] daripada menerapkan *dirichlet prior* sebelumnya [$\theta \sim \text{Dir}(\alpha)$] ke prevalensi topik dimana *logistic normal linier prior* adalah fungsi dari tingkat dokumen kovariat seperti tanggal terbit artikel, penulis, dan lokasi. Selain itu, STM memungkinkan penggunaan kata dalam topik (konten topik) lebih bervariasi menurut kovariat kategorikal (Y) menggunakan fungsi *multinomial logit* [14].

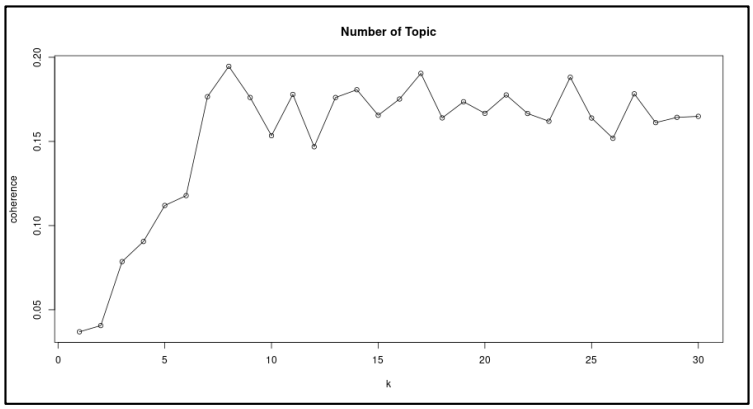
3. Hasil dan Pembahasan

3.1. Menentukan Jumlah Topik Optimal

Salah satu tantangan utama pemodelan topik adalah mengidentifikasi jumlah topik optimal yang laten dalam korpus. Jumlah topik ini digunakan sebagai masukan dalam menjalankan *topic modelling.*

a. *Latent Dirichlet Allocation* (LDA)

Jumlah topik yang optimal untuk metode LDA ditentukan berdasarkan nilai *probabilistic coherence* dengan menguji 30 kandidat topik. Berdasarkan Tabel 2 dan Gambar 4 di bawah ini, jumlah topik yang optimal adalah 8 topik dengan nilai *coherence* terbesar yaitu 0.19387522.



Gambar 4. Hasil Evaluasi Jumlah Topik LDA

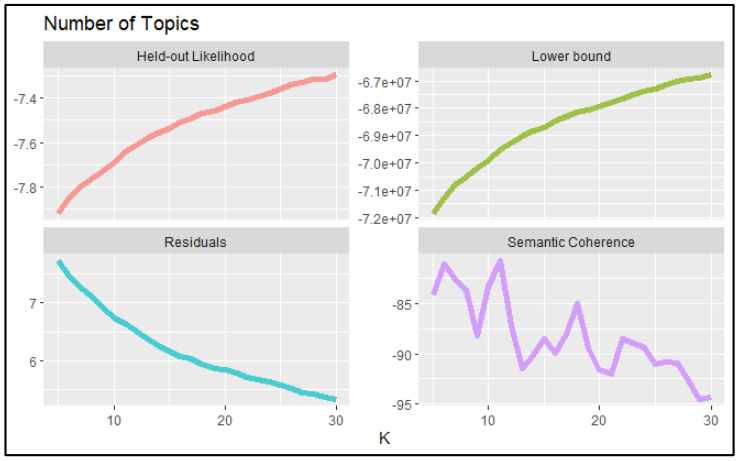
b. *Structural Topic Model* (STM)

Jumlah topik dalam metode STM ditentukan berdasarkan dua kriteria yaitu *semantic coherence* dan *exclusivity* yang keduanya harus setinggi mungkin. Pemilihan kedua kriteria ini dikarenakan hasil dari perhitungan *semantic coherence* akan menghasilkan beberapa topik di mana kata-kata yang sangat umum mendominasi. Oleh karena itu, perlu mempertimbangkan eksklusivitas dari topik untuk memberikan perbandingan. *Semantic coherence* menunjukkan seberapa konsisten suatu topik atau seberapa sering istilah yang menggambarkan suatu topik terjadi bersamaan dan muncul bersama dalam sebuah dokumen. *Exclusivity* adalah seberapa eksklusif topik yang muncul dengan probabilitas

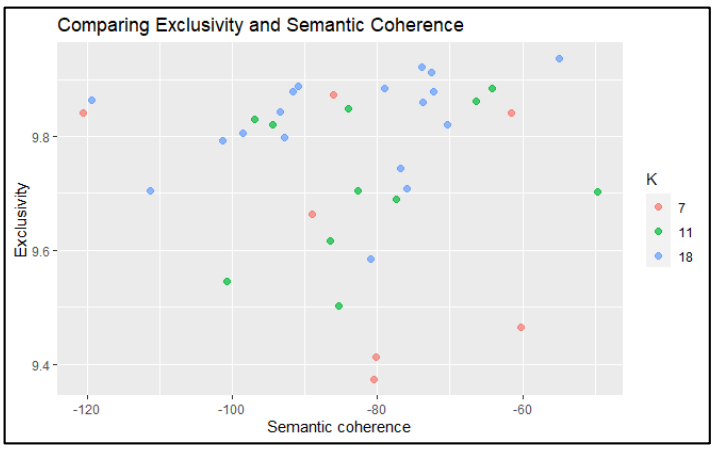
p-ISSN: 2301-5373
e-ISSN: 2654-5101

tinggi untuk suatu topik atau seberapa berbeda topik satu sama lain, dan topik tersebut menggambarkan hal yang berbeda. Semakin tinggi nilai koherensi semantik dan eksklusivitas kata-kata dalam suatu topik, semakin "baik" model topik tersebut.

Gambar 5 menunjukkan hasil evaluasi model STM. Dari Gambar 5 terlihat bahwa jumlah topik tertinggi berdasarkan *semantic coherence* pada kisaran topik 7, 11, dan 18. Untuk memilih topik yang optimal diantara ketiga jumlah topik ini, pertimbangkan juga nilai eksklusivitas dari ketiga topik tersbut yang ditunjukkan pada Gambar 6. Dari Gambar 6, jumlah topik yang optimal adalahh 11 topik dengan nilai koherensi dan nilai eksklusivitas yang tinggi.



Gambar 5. Hasil Evaluasi Jumlah Topik STM



Gambar 6. Perbandingan Nilai *Exclusivity* dan *Semantic Coherence*

3.2. Identifikasi Topik

a. *Latent Dirichlet Allocation* (LDA)

Tabel 2 menunjukkan hasil pemodelan topik LDA berdasarkan jumlah topik yaitu K=8. Nilai koherensi untuk kedelapan topik tersebut memiliki nilai yang beragam. Dari topik dengan nilai tertinggi yang mudah diinterpretasikan, hingga topik dengan nilai koherensi terendah yang sulit untuk ditafsirkan secara sekilas.

Topik 1 membahas mengenai kasus korupsi yang diidentifkasi dengan kata-kata seperti "kpk", "uang", dan "pasal". Hal ini menunjukkan bahwa Topik 1 membahas kasus korupsi. Topik 2 berkaitan dengan kegiatan olahraga, khususnya sepak bola, yang digambarkan dengan kata-kata "gol", "pemain", "laga", dan lainnya. Topik 3 merupakan salah satu topik yang memiliki nilai koheren paling rendah diantara topik lainnya. Dari kata-kata yang dihasilkan, secara ambigu bahwa topik ini membahas mengenai peristiwa yang terjadi di jalanan diidentifikasi dari kata "jalan", "mobil", dan kejadian di daerah yang digambarkan oleh kata "kecamatan", "desa", "air". Topik 4 adalah topik dengan nilai koherensi paling

rendah sehingga kata-kata yang dihasilkan lebih sulit untuk diinterpretasikan dibandingkan topik lainnya. Topik 4 mencakup berita yang terjadi di ibukota Jakarta termasuk pembangunan, program, pendidikan, dan lain-lain. Topik 5 membahas isu COVID19 dimulai dengan vaksinasi, varian terbaru Corona yaitu Omicron, dan menjelaskan cara penanganan PPKM karena terdapat kata "level" menandakan masih ada kabar kenaikan atau penurunan level PPKM di berbagai daerah. Topik 6 berkaitan dengan harga minyak goreng di Pasaran. Topik ini memberitahu bahwa dalam kurun waktu lima bulan yaitu November 2021 sampai Maret 2022 terdapat berita mengenai permasalahan harga minyak goreng di Indonesia. Topik 7 membahas mengnai pemerintah Indonesia yang diidentifikasi dengan kata-kata seperti "presiden", "dpr", dan "partai". Topik 8 terkait dengan negara Rusia dan Ukraina dimana diketahui bahwa kedua negara ini memiliki konflik yang dapat diindentifikasi dari kata "militer", "pesawat" dan "wilayah".

Tabel 2. Hasil *Topic Modelling* dengan LDA

Topik	Coherence	Prevalence	Top Terms
1	0.191	13.243	pasal, pidana, kpk, uang, bukti, laporan, jaksa, dugaan, perkara, jakarta
2	0.401	10.512	pemain, tim, gol, laga, pertandingan, menit, poin, bola, liga, kali
3	0.082	21.569	jalan, rumah, lokasi, desa, mobil, kecamatan, kabupaten, kota, air, kejadian
4	0.061	10.993	kota, jakarta, kerja, dki, daerah, pembangunan, program, pendidikan, kepala, nomor
5	0.111	9.787	covid, kesehatan, kota, vaksinasi, kabupaten, vaksin, corona, omicron, varian, level
6	0.169	13.585	harga, minyak, pt, perusahaan, pasar, juta, goreng, ekonomi, produk, bank
7	0.218	10.521	ketua, jokowi, dpr, presiden, anggota, tni, partai, ri, komisi, politik
8	0.318	9.789	rusia, ukraina, as, dunia, militer, china, presiden, pesawat, wilayah, amerika

b. *Structural Topic Model* (STM)

Hasil pemodelan topik menggunakan STM ditunjukkan oleh Tabel 3. Kata-kata yang tercantum pada Tabel 3 berasal dari keluaran STM dengan jumlah topik K = 11. Tabel tersebut memiliki empat jenis pembobotan kata untuk setiap topik yaitu *Highest Prob*, FREX, *Lift*, dan *Score* [15]. *Highest Prob* adalah kata-kata dengan probabilitas tertinggi atau kata yang paling umum untuk topik tertentu, namun tidak eksklusif dalam arti kata-kata tersebut dapat dikaitkan dengan sejumlah topik. FREX (*frequency-exclusivity*) memberi bobot pada kata-kata berdasarkan seberapa eksklusif kata-kata tersebut terhadap suatu topik pada frekuensi keseluruhannya. *Lift* membobot kata-kata dengan membaginya dengan frekuensinya di topik lain sehingga memberi bobot lebih pada kata-kata yang lebih jarang di topik lain. *Score* mirip dengan *lift*, membagi frekuensi log kata dalam topik dengan frekuensi log kata dalam topik lain.

Tabel 3. Hasil *Topic Modelling* dengan STM

Topic	Proportion	Top Terms			
		Highest Prob	FREX	Lift	Score
1	0.098	video, akun, media, agama, keluarga, suara, foto, sosial, acara, maaf	mui, wayang, yahya, nu, pbnu, gus, yaqut, lagu, ustaz, munarman	bimasislam, dikaruniai, pdri, tarian, mempelai, aurel, slavina, antiteror, najah, akhyar	nu, mui, munarman, wayang, gus, agama, pbnu, yaqut, masjid, arteria
2	0.058	harga, minyak, goreng, juta,	minyak, goreng, liter, harga, bbm,	bersubsidi, kedelai, menstabilkan,	minyak, goreng, harga, liter,

p-ISSN: 2301-5373
e-ISSN: 2654-5101

Topic	Proportion	Top Terms			
		Highest Prob	**FREX**	**Lift**	**Score**
		pasar, ribu, barang, uang, pedagang, kenaikan	kemasan, kg, kedelai, tempe, kripto	bakunya, rti, dow, nurwan, jualnya, pengecer, crude	pasar, kedelai, tempe, kripto, pedagang, curah
3	0.082	covid, kesehatan, vaksinasi, vaksin, corona, omicron, varian, level, kota, persen	vaksinasi, vaksin, omicron, ppkm, dosis, booster, pcr, corona, pasien, prokes	antigen, spesimen, faskes, adisasmito, immunity, herd, penyebarannya, terjangkit, penularannya, eswatini	omicron, vaksin, vaksinasi, corona, covid, varian, ppkm, pasien, dosis, booster
4	0.080	jokowi, presiden, ketua, dpr, partai, ri, anggota, pendidikan, kota, komisi	ikn, pemilu, puan, demokrat, seleksi, fadli, ruu, beasiswa, jokowi, gerindra	pileg, bawaslu, imin, parpol, maharani, ahy, pilpres, fadli, civitas, kelulusan	partai, jokowi, pemilu, ikn, dpr, politik, puan, snmptn, ruu, gerindra
5	0.083	pasal, pidana, kpk, jaksa, hakim, perkara, uang, tindak, dugaan, terdakwa	kpk, jaksa, terdakwa, pn, persidangan, pidana, suap, kejaksaan, korupsi, penjara	dakwaannya, disidangkan, ezer, stepanus, narkotika, narapidana, novia, bnn, lpsk, terpidana	kpk, terdakwa, pidana, jaksa, pasal, penyidik, penjara, korupsi, perkara, hakim
6	0.127	jalan, desa, kota, kabupaten, air, lokasi, hujan, rumah, kecamatan, banjir	sampah, bpbd, longsor, banjir, pohon, hujan, ruas, petir, jembatan, genangan	jpo, bandang, bpbd, majenang, ambarawa, menggenang, bpjt, lengkong, candipuro, lor	banjir, hujan, desa, tol, kecamatan, sungai, kabupaten, longsor, bmkg, bpbd
7	0.056	rusia, ukraina, as, wilayah, militer, china, barat, pesawat, presiden, kapal	rusia, ukraina, putin, gempa, invasi, rudal, kapal, pesawat, biden, israel	kharkiv, belarusia, associated, pemberontakan, artileri, istanbul, kyiv, yahudi, kuleba, pentagon	rusia, ukraina, putin, gempa, invasi, rudal, militer, as, nato, kiev
8	0.068	jakarta, dki, tni, nomor, gubernur, kerja, anies, jenderal, peraturan, sesuai	buruh, ump, upah, mk, formula, tni, bpjs, cipta, ketenagakerjaan, minimum	kartiko, ciptaker, kis, menaker, pangkostrad, feo, jkp, kspi, onlyfans, permenaker	anies, ump, tni, dki, upah, buruh, mk, formula, uu, dudung
9	0.129	mobil, kejadian, rumah, motor, pria, diduga, luka, jalan, ditemukan, lokasi	tkp, mayat, pengemudi, reskrim, polsek, sopir, pengeroyokan, anggiat, handi, kompol	sabetan, diautopsi, inafis, dagu, rakitan, spion, bejatnya, memaki, tusuk, jatanras	mobil, luka, motor, tewas, zulpan, polsek, reskrim, kombes, pengeroyokan, arteria
10	0.109	pemain, tim, gol, laga, pertandingan, menit, poin, bola, liga, kali	pemain, gol, laga, pertandingan, bola, liga, mandalika, juara, motogp, pelatih	newcastle, persik, napoli, ducati, pss, suporter, arsenal, league, juventus, persebaya	gol, laga, pertandingan, pemain, liga, gawang, piala, pelatih, motogp, kemenangan
11	0.110	perusahaan, pt, ekonomi, program, bank, digital, produk, kerja, triliun, keuangan	fitur, nasabah, ekosistem, triliun, digital, baterai, kredit, umkm, oppo, asuransi	multiplier, ekuitas, consumer, baterainya, warjiyo, payment, property, lifestyle, app, emission	triliun, umkm, fitur, digital, produk, listrik, bank, nasabah, saham, pt

Dari Tabel 3, Topik 1 dapat diinterpretasikan dengan lebih mudah diinterpretasikan berdasarkan kata-kata FREX: ‘mui’, ‘wayang’, ‘nu’, ‘pbnu’, ‘gus’, ‘lagu’, ‘ustaz’. Jika dihubungkan dengan kata-kata Highest Prob Topik 1, topik ini terkait dengan agama Islam yang diidentifikasi dari kata ‘mui’, ‘nu’.

Topik 2 memiliki kata-kata FREX dan Highest Prob yang hampir mirip. Oleh karena itu, topik ini mengangkat isu harga minyak goreng. Topik ini juga muncul pada hasil pemodelan topik dengan LDA. Topik 2 memiliki kata ‘harga’, ‘minyak’, ‘kenaikan’. Dari hal ini, dapat disimpulkan bahwa terjadi kejadian kenaikan harga minyak goreng di Indonesia.

Topik 3 yang dihasilkan dengan metode STM juga berhasil muncul di Topik 5 dari hasil pemodelan topik dengan LDA, yaitu membahas mengenai isu kesehatan terutama COVID-19. Keunggulan STM sendiri adalah menghasilkan berbagai kata yang tidak ditemukan pada hasil LDA, seperti kata-kata ‘antigen’, ‘spesimen’, ‘faskes’, ‘immunity’, ‘herd’, ‘penyebarannya’, ‘terjangkit’, dan ‘penularannya’.

Topik 4 berisi pembahasan yang sama dengan Topik 7 dari hasil pemodelan topik dengan LDA, yaitu membahas mengenai Pemerintah Indonesia. Berita ini bisa berupa berita tentang pemerintahan, pemilu, hingga ibu kota baru, yang ditunjukkan dengan kata ‘ikn’.

Topik 5 menghasilkan topik yang mirip dengan Topik 1 pada hasil LDA yaitu Korupsi. Secara sekilas dengan meilhat kata-kata FREX, Lift, hingga Score dapat dipahami bahwa topik ini adalah berita tentang kasus korupsi di Indonesia.

Topik 6 membahas mengenai bencana alam yang terjadi di Indonesia. Diidentifikasi dengan kata-kata seperti ‘banjir’, ‘bpbd’, dan ‘longsor’. Hasil pemodelan topik menggunakan LDA, topik yang berhubungan dengan bencana alam secara ambigu dihasilkan dari satu topik yaitu Topik 3. Topik 3 tersebut membahas dua kejadian dalam satu topik, tetapi pemodelan topik dengan STM, berhasil memisahkan dua kejadian tersebut menjadi dua topik sehingga topik lebih mudah dimengerti.

Topik 7 memiliki kesamaan dengan Topik 8 dari hasil pemodelan topik dengan LDA yaitu isu yang terjadi antara negara Rusia dan Ukraina. Hasil FREX mencantumkan kata ‘invansi’ dan ‘rudal’, yang menunjukkan bahwa kedua negara ini sedang berperang.

Topik 8 membahas mengenai berita yang terjadi di ibu kota Jakarta, ditandai dengan kata ‘jakarta’, ‘anies’ dan lain-lain. Topik ini juga berhasil dihasilkan dari pemodelan topik menggunakan LDA pada topik 4.

Topik 9 adalah pengembangan dari Topik 3 hasil pemodelan topik dengan LDA yaitu kejadian yang terjadi di jalan. Dari hasil Topik 9 terlihat bahwa sering terdapat berita mengenai kejadian di jalan yang diidentifikasi dari kata ‘jalan’, ‘mobil’, ‘motor’, ‘luka’ yang menunjukkan bahwa kecelakaan lalu lintas merupakan hal yang umum diberitakan dalam kurun waktu 5 bulan.

Topik 10 memiliki kesamaan dengan Topik 2 dari hasil pemodelan topik dengan LDA. Topik ini menjelaskan mengenai olahraga khususnya sepak bola yang dijelaskan dengan kata-kata seperti ‘bola’, ‘pemain’, ‘gol’. Kata eksklusif untuk topik ini juga menjelaskan olahraga lain seperti ‘motogp’ dan ‘mandalika’. Seperti yang diketahui bahwa pertandingan MotoGP diadakan pada bulan Maret 2022 di sirkuit Mandalika Lombok.

Topik 11 membahas mengenai ekonomi jika dilihat dari keempat bobot yaitu Highest Prob, FREX, Lift, dan Score. Topik ini diwakili oleh kata-kata seperti ‘bank’, ‘ekonomi’, ‘perusahaan’, ‘digital’, dan ‘keuangan’.

Dari 11 Topik yang dihasilkan dari metode STM, jika dilihat secara sekilas STM menghasilkan topik yang mirip dengan LDA. Topik 3 dan 4 yang dihasilkan dari LDA memiliki nilai koherensi rendah sehingga sulit untuk diinterpretasi secara sekilas. Topik-topik tersebut juga muncul dari hasil STM dan hasil topik dari STM ini lebih mudah dipahami dengan melihat kata-kata teratas dan eksklusif (FREX) kata dari topik tersebut.

p-ISSN: 2301-5373
e-ISSN: 2654-5101

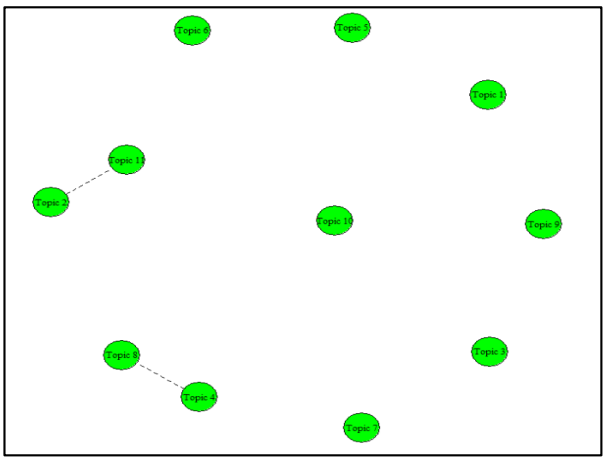
Seperti Topik 3 dari hasil LDA yang membahas mengenai peristiwa yang terjadi di jalanan diidentifikasi dari kata “jalan”, “mobil”, dan kejadian di daerah yang digambarkan oleh kata “kecamatan”, “desa”, “air”. Pada STM, topik ini berhasil dipisah menjadi dua topik yaitu Topik 6 mengenai bencana alam dan 9 mengenai kecelakaan lalu lintas. Kemudian Topik 4 hasil LDA mengenai berita yang terjadi di ibukota Jakarta, topik ini muncul sebagai Topik 8 dari hasil STM.

3.3. Korelasi Antar Topik

Topik yang berkorelasi dapat diidentifikasi menggunakan STM dengan menganalisis seberapa sering topik muncul bersamaan dalam dokumen yang sama. Topik model adalah model campuran yang setiap topiknya memiliki “nilai”, sehingga topik mungkin saja tidak terkait dan berkorelasi. Korelasi positif menunjukkan bahwa dokumen yang memiliki konten tentang topik A cenderung juga menyajikan beberapa konten topik B. Korelasi negatif menunjukkan bahwa dokumen yang menyajikan topik A secara sistematis tidak menyajikan topik B. Semakin banyak topik yang ekslusif atau tidak tumpang tindih, maka korelasi akan menuju nilai -1.

Untuk memudahkan melihat korelasi antar topik dapat dengan melihat Gambar 7. Pada Gambar 7 tersebut, menunjukkan dua node yang terhubung artinya topik tersebut memiliki nilai korelasi yang tinggi. Terdapat empat topik yang saling terhubung yaitu korelasi antara Topik 11 dan Topik 2, dan korelasi antara Topik 8 dan Topik 4.

Topik 11 dan Topik 2 memiliki nilai korelasi 0,06. Jika dilihat dari kata-kata yang dihasilkan dari kedua topik ini memiliki kesamaan mengenai masalah ekonomi dimana Topik 2 membahas mengenai permasalahan harga minyak goreng dan Topik 11 yang membahas mengenai ekonomi digital. Topik 8 dan 4 memiliki nilai korelasi 0,04 dan memiliki kesamaan mengenai Pemerintahan. Topik 8 berkaitan dengan Pemerintahan Indonesia yang diidentifikasi dari kata-kata seperti ‘dpr’ dan ‘jokowi’. Sedangkan Topik 4 membahas mengenai berita yang terjadi di Ibukota Jakarta yang ditunjukkan oleh kata-kata seperti ‘anies’ dan ‘jakarta’. Kedua topik ini rata-rata terjadi di Jakarta.



Gambar 7. Korelasi Antar Topik

4. Kesimpulan

Kesimpulan penelitian ini didasarkan pada dataset yang digunakan yaitu 44.425 artikel berita dari bulan November 2021 hingga Maret 2022 diambil dari portal berita online detik.com dengan melakukan *topic modelling* menggunakan Latent Dirichlet Allocation (LDA) dan Structural Topic Model (STM). Metode LDA menghasilkan 8 Topik yang berasal dari perhitungan *probabilistic coherence* nilai tertinggi. Metode STM menghasilkan 11 Topik berdasarkan nilai *semantic coherence* dan *exclusivity* tertinggi. Dari kedua hasil topik masing-masing metode, LDA memiliki 2 dari 8 Topik yang tidak mudah diinterpretasikan. Sedangkan STM berhasil memudahkan interpretasi topik tersebut dengan melihat eksklusivitas topiknya. Pada metode STM, terdapat 2 pasang topik yang saling berkorelasi yaitu Topik 11 dan Topik 2 memiliki kesamaan mengenai masalah ekonomi. Topik 8 dan 4 memiliki kesamaan mengenai pemerintahan.

Referensi

[1] H. Jelodar *et al.*, "Latent Dirichlet allocation (LDA) and topic modeling: models, applications, a survey," *Multimed. Tools Appl.*, vol. 78, no. 11, pp. 15169–15211, 2019, doi: 10.1007/s11042-018-6894-4.

[2] C. B. Asmussen and C. Møller, "Smart literature review: a practical topic modelling approach to exploratory literature review," *J. Big Data*, vol. 6, no. 1, Dec. 2019, doi: 10.1186/s40537-019-0255-7.

[3] N. A. Sanjaya ER, "Implementasi Latent Dirichlet Allocation (LDA) untuk Klasterisasi Cerita Berbahasa Bali," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 8, no. 1, p. 127, 2021, doi: 10.25126/jtiik.0813556.

[4] C. Naury, D. H. Fudholi, and A. F. Hidayatullah, "Topic Modelling pada Sentimen Terhadap Headline Berita Online Berbahasa Indonesia Menggunakan LDA dan LSTM," *J. Media Inform. Budidarma*, vol. 5, no. 1, p. 24, 2021, doi: 10.30865/mib.v5i1.2556.

[5] W. Wahyudin, "APLIKASI TOPIC MODELING PADA PEMBERITAAN PORTAL BERITA ONLINE SELAMA MASA PSBB PERTAMA," *Semin. Nas. Off. Stat.*, vol. 2020, no. 1, pp. 309–318, 2021, doi: 10.34123/semnasoffstat.v2020i1.579.

[6] X. Chen, D. Zou, G. Cheng, and H. Xie, "Detecting latent topics and trends in educational technologies over four decades using structural topic modeling: A retrospective of all volumes of Computers & Education," *Comput. Educ.*, vol. 151, no. September 2019, p. 103855, 2020, doi: 10.1016/j.compedu.2020.103855.

[7] R. B. Mishra and H. Jiang, "Management and organizational research: Structural topic modeling for a better understanding of theory application," *Sustain.*, vol. 14, no. 1, 2022, doi: 10.3390/su14010159.

[8] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent Dirichlet allocation," *J. Mach. Learn. Res.*, vol. 3, no. 4–5, pp. 993–1022, 2003, doi: 10.1016/b978-0-12-411519-4.00006-9.

[9] M. Ponweiser, "Latent Dirichlet Allocation in R," no. May, pp. 2–21, 2012, [Online]. Available: http://epub.wu.ac.at/3558/.

[10] A. K. Johnson, R. Bhaumik, D. Nandi, A. Roy, and S. D. Mehta, "' Is this Herpes or Syphilis ?': Latent Dirichlet Allocation Analysis of Sexually Transmitted Disease-Related Reddit Posts During the COVID-19 Pandemic," 2022, doi: https://doi.org/10.1101/2022.02.13.22270890.

[11] D. M. Blei and J. D. Lafferty, "TOPIC MODELS," 2009. doi: 10.1007/978-981-16-0100-2_7.

[12] M. E. Roberts *et al.*, "Structural topic models for open-ended survey responses," *Am. J. Pol. Sci.*, vol. 58, no. 4, pp. 1064–1082, 2014, doi: 10.1111/ajps.12103.

[13] J. Bohr and R. E. Dunlap, "Key Topics in environmental sociology, 1990–2014: results from a computational text analysis," *Environ. Sociol.*, vol. 4, no. 2, pp. 181–195, 2018, doi: 10.1080/23251042.2017.1393863.

[14] E. (Olivia) Park, B. (Kevin) Chae, and J. Kwon, "The structural topic model for online review analysis: Comparison between green and non-green restaurants," *J. Hosp. Tour. Technol.*, vol. 11, no. 1, pp. 1–17, 2020, doi: 10.1108/JHTT-08-2017-0075.

[15] M. E. Roberts, B. M. Stewart, and D. Tingley, "Stm: An R package for structural topic models," *J. Stat. Softw.*, vol. 91, no. 2, 2019, doi: 10.18637/jss.v091.i02.

Jurnal Elektronik Ilmu Komputer Udayana
Volume 11, No 3. February 2023

p-ISSN: 2301-5373
e-ISSN: 2654-5101

Fuzzy Inference System as a Method of Determining Drug Demand in Pharmacy Systems

Ni Made Rai Nirmala Santhi[a1], Ida Bagus Made Mahendra [a2], Ngurah Agus Sanjaya ER [a3], I Putu Gede Hendra Suputra [a4], I Komang Ari Mogi [a5], Agus Muliantara[a6]

[a]Program Studi Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Udayana
Badung, Bali, Indonesia
[1]ranirma.santhi@gmail.com
[2] ibm.mahendra@unud.ac.id
[3] agus_sanjaya@unud.ac.id
[4] hendra.suputra@unud.ac.id
[5] arimogi@unud.ac.id
[6] muliantara@unud.ac.id

Abstract

In providing health services, puskesmas cannot be separated from the need for pharmaceutical preparations. Given the large influence of pharmaceutical preparations on the smoothness and performance of services at the puskesmas, it is necessary to pay special attention to managing them. However, prediction of demand is often difficult for officers, because it cannot be done by just anyone. Therefore, this study was conducted to provide consideration for the demand for pharmaceutical preparations in order to minimize the lack of pharmaceutical preparations. In the process, a demand analysis is carried out based on the number of uses and remaining inventory from the previous time period, to produce a prediction of the number of requests for pharmaceutical preparations. The method that will be used in this research is the Mamdani model fuzzy inference system, with interval determination based on the quartile method and the decile method. Which results obtained that the evaluation of predictive competence using both methods is sufficient. Namely, the Mean Absolute Percentage Error (MAPE) of the two methods is 26.43% for the decile method and 28.77% for the quartile method.

Keywords: *Fuzzy Inference System, Prediction, Demand, Pharmaceutical inventory, Mamdani Method*

1. Pendahuluan

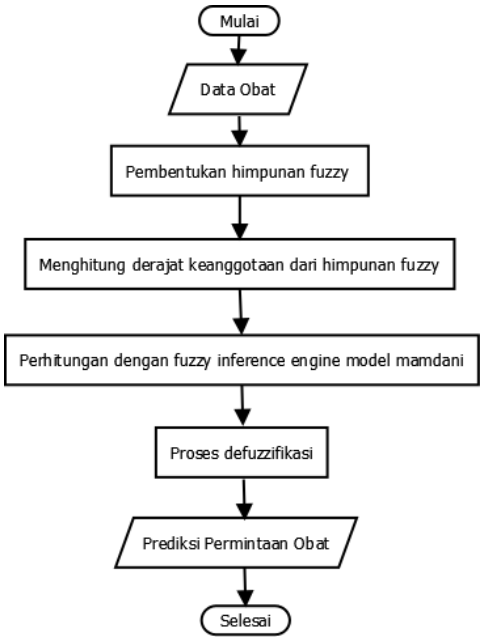
Pusat kesehatan masyarakat yang seterusnya disebut puskesmas merupakan pelaksana fasilitas pelayanan dasar dalam upaya mewujudkan derajat kesehatan yang optimal bagi masyarakat [1]. Dalam memberikan pelayanan kesehatan, puskesmas tidak terlepas dari komidi utamanya yakni sediaan farmasi yang berupa obat, bahan obat, maupun kosmetika. Mengingat besarnya pengaruh sediaan farmasi untuk kelancaran pelayanan di puskesmas, maka perlu adanya perhatian khusus untuk mengelolanya. Pengelolaan sediaan farmasi, sangat penting dilakukan untuk menjaga ketersediaannya dalam membantu peningkatan kualitas pelayanan fasilitas kesehatan kepada masyarakat [2].

Kegiatan pengelolaan sediaan farmasi berupa perencanaan memiliki peran penting untuk menjaga persediaan farmasi. Perencanaan tersebut meliputi proses peramalan jumlah permintaan yang diperlukan dalam pemenuhan persediaan puskesmas. Prediksi jumlah permintaan sering menjadi kesulitan bagi petugas akibat tidak dapat dilakukan oleh sembarang orang karena perlunya pengalaman yang cukup lama untuk memberikan peramalan yang tepat agar tidak terjadinya kekurangan persediaan. Persediaan dilakukan untuk menjamin adanya kepastian bahwa pada saat dibutuhkan produk-produk tersebut tersedia. Untuk itu perlu adanya perencanaan persediaan yang berdasar dalam menyikapi pola permintaan yang pada dasarnya tidak beraturan [3].

Maka dari itu, untuk memberikan pertimbangan permintaan sediaan farmasi guna meminimalisir kurangnya, diperlukan analisis permintaan yang berdasar pada banyaknya penggunaan dan sisa persediaan dari periode waktu sebelumnya[4]. Dari data tersebut kemudian menghasilkan prediksi jumlah permintaan sediaan farmasi. Adapun metode yang akan digunakan dalam penelitian ini yaitu *fuzzy inference system* model mamdani, dengan penentuan interval dari himpunan *fuzzy* menggunakan metode kuartil maupun metode desil. Dengan dilakukannya prediksi permintaan sediaan farmasi menggunakan metode *fuzzy inference system* model mamdani, diharapkan dapat dijadikan bahan pertimbangan yang berdasar guna memberikan perancanaan yang baik dalam pengelolaan sediaan farmasi.

2. Metode Penelitian

Terdapat beberapa tahapan dalam penerapan *fuzzy inference sistem* model mamdani, yaitu dengan alur sebagaimana pada gambar 1.



Gambar 1. Alur Umum Penelitian

Sebelum dilakukannya proses penerapan *fuzzy inference system*, terlebih dahulu dilakukan proses mempersiapkan data. Adapun proses mempersiapkan data meliputi pengambilan data, pembersihan data yang sesuai, serta melakukan proses input ke dalam system. Kemudian baru dapat dilakukan pengiolahan data sebagaimana alur penerapan *fuzzy inference system*. Adapun alur tersebut lebih jelasnya seperti berikut.

1. Pembentukan himpunan *fuzzy* menggunakan metode kuartile dan metode desil.
2. Menghitung fungsi keanggotaan dari himpunan *fuzzy*. Adapun fungsi keanggotaan yang digunakan yaitu triangular membership function.
3. Perhitungan dengan *fuzzy inference engine* model mamdani.
4. Proses konversi hasil dari *inference engine* yang diekspresikan dalam bentuk *fuzzy set* ke suatu bilangan riil, proses ini disebut juga proses defuzzifikasi.

2.1 Pembentukan Himpunan Fuzzy

Pembentukan himpunan *fuzzy* yakni menentukan semua variabel yang terkait seperti jumlah sisa persediaan dan penggunaan sebagai fungsi *input*, yang menghasilkan permintaan sediaan farmasi sebagai fungsi *output*[5]. Adapun dalam penelitian ini dalam menentukan interval setiap himpunan menggunakan metode kuartil dan metode desil, dengan hasil sepert pada tabel 1.

p-ISSN: 2301-5373
e-ISSN: 2654-5101

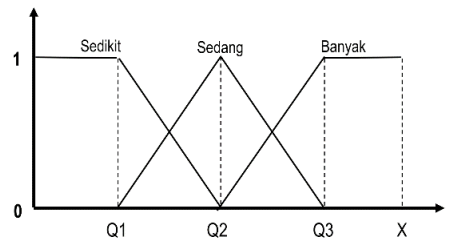
Tabel 1. Interval dari Himpunan Fuzzy

Nama Variabel	Semesta Pembicara	Himpunan *Fuzzy*	Interval Metode Kuartil	Interval Metode Desil
Jumlah Sisa Persediaan	[0 Xs]	Sedikit	[0 Q2]	[0 D5]
		Sedang	[Q1 Q3]	[D1 D9]
		Banyak	[Q2 Xs]	[D5 Xs]
Jumlah Penggunaan	[0 Xp]	Sedikit	[0 Q2]	[0 D5]
		Sedang	[Q1 Q3]	[D1 D9]
		Banyak	[Q2 Xp]	[D5 Xp]
Jumlah Permintaan	[0 Xr]	Sedikit	[0 Q2]	[0 D5]
		Sedang	[Q1 Q3]	[D1 D9]
		Banyak	[Q2 Xr]	[D5 Xr]

Keterangan :
- Xs : Nilai terbesar sisa persediaan.
- Xp : Nilai terbesar sisa penggunaan.
- Xr : Nilai terbesar sisa permintaan.
- Dn : Desil ke-n dari variabel.
- Qn : Kuartil ke-n dari variabel.

2.2 Menghitung Derajat Keanggotaan

Fungsi keanggotaan divisualisasikan sebagai kurva yang memetakan nilai derajat dari keanggotaan *input* yang diberikan. Derajat keanggotaan tersebut merupakan representasi dari titik input dengan nilai keanggotaannya. Adapun interval nilai derajat keanggotaan suatu *input* antara 0 sampai dengan 1[5]. Adapun fungsi keanggotaan yang digunakan pada penelitian ini yaitu *linear-triangular membership function*, dengan batas bawah dari membership function ini yang akan ditentukan menggunakan metode kuartil dan metode desil. Yang selanjutnya masing – masing hasil perhitungan *membership function* akan melalui tahap selanjutnya untuk menghasilkan hasil prediksi dari kedua metode tersebut. Adapun contoh visualisasi *membership function* seperti berikut.



Gambar 2. Visualisasi *Membership Function*

Dari visualisasi tersebut, maka untuk perhitungan derajat keanggotaan seperti persamaan berikut.

$$\mu\, Sedikit = \begin{cases} 1 & ; \quad x_i \le Q_1 \\ \frac{Q_2 - Xi}{Q_2 - Q_1} & ; \quad Q_1 \le x_i \le Q_2 \\ 0 & ; \quad x_i \ge Q_2 \end{cases} \tag{1}$$

$$\mu\, Sedang = \begin{cases} 0 & ; Xi \le Q_1 \quad atau\, x_i \ge Q_3 \\ \frac{x_i - Q_1}{Q_2 - Q_1} & ; \quad Q_1 \le x_i \le Q_2 \\ \frac{Q_3 - x_i}{Q_3 - Q_2} & ; Q_2 \le x_i \le Q_3 \end{cases} \tag{2}$$

$$\mu\, Banyak = \begin{cases} 0 & ; \quad x_i \le Q_2 \\ \frac{x_i - Q_2}{Q_3 - Q_2} & ; \quad Q_2 \le x_i \le Q_3 \\ 1 & ; \quad x_i \ge Q_3 \end{cases} \tag{3}$$

2.3 *Fuzzy Inference Engine* Model Mamdani

Fuzzy inference engine Model Mamdani atau sering disebut juga dengan metode min-max. Pada tahap ini diawali dengan menetapkan aturan dengan betuk fungsi implikasi. Adapun aturan tersebut yaitu sabagaimana tabel 2. Setelah ditetapkan aturan implikasi, kemudian dilakukan proses implementasi fungsi implikasi dengan menggunakan min function atau disebut juga fungsi minimum berdasarkan aturan implikasi yang telah dibuat, untuk mendapatkan α-predikat setiap aturan[6]. Adapun fungsi minimum diekspresikan sebagai persamaan berikut.

$$\mu_{\tilde{A}\cap\tilde{B}} = \min\left(\mu_{\tilde{A}}(S), \mu_{\tilde{B}}(P)\right) \quad (4)$$

Keterangan:

$\mu_{\tilde{A}\cap\tilde{B}}$: Nilai hasil operator AND derajat keanggotaan atau $\alpha_{-predikat}$ aturan ke-i.
$\mu_{\tilde{A}}(S)$: Derajat keanggotaan dari variabel sisa persediaan.
$\mu_{\tilde{B}}(P)$: Derajat keanggotaan dari variabel pemakaian.

Selanjutnya yaitu mencari solusi *fuzzy set* dengan proses menghitung nilai maksimum tiap daerah himpunan *fuzzy*. Yang mana proses ini direpresentasikan dengan persamaan berikut.

$$\mu(x_i) = max\left(\mu_{af}(x_i), \mu_{bf}(x_i), \mu_{cf}(x_i)\right) \quad (5)$$

Keterangan:

$\mu(x_i)$: Solusi fuzzy set himpunan fuzzy ke-i.
$\mu_{af}(x_i)$: $\alpha_{-predikat}$ aturan ke-a anggota himpunan *fuzzy* ke-i.
$\mu_{bf}(x_i)$: $\alpha_{-predikat}$ aturan ke-b anggota himpunan *fuzzy* ke-i.
$\mu_{cf}(x_i)$: $\alpha_{-predikat}$ aturan ke-c anggota himpunan *fuzzy* ke-i.

Tabel 2. Fungsi Implikasi

Jumlah Sisa Persediaan	Jumlah Penggunaan	Jumlah Permintaan
Sedikit	Sedikit	Sedikit
	Sedang	Sedang
	Banyak	Banyak
Sedang	Sedikit	Sedikit
	Sedang	Sedang
	Banyak	Sedang
Banyak	Sedikit	Sedikit
	Sedang	Sedikit
	Banyak	Sedang

2.4 Defuzzifikasi

Defuzzifikasi merupakan proses mengkonversi setiap hasil dari inference engine yang diekspresikan dalam bentuk fuzzy set ke suatu bilangan real. Adapun disini menggunakan metode centroid atau disebut *juga center of area (center of gravity)* yang mana mencari hasil prediksi dengan melakukan perhitungan untuk mencari titik pusat daerah *fuzzy*[5]. Dalam menghitung titik tengah dilakukan dengan mengitung nilai moments dan luas setiap daerah.

$$\mathrm{M}_i = \int_{\mathrm{b}}^{\mathrm{a}} \mu\, z_i \quad d(z_i) \quad (6)$$

$$A_i = (a - b) \times \mu \quad (7)$$

$$A_i = \frac{(a - b) \times (\mu_a + \mu_b)}{2} \quad (8)$$

$$Z = \frac{\sum_{i=1}^{n} \mathrm{M}_i}{\sum_{i=1}^{n} \mathrm{A}_i} \quad (9)$$

Keterangan:

M_i : Moment daerah ke-i.
A_i : Luas daerah ke-i.
$z_i \quad d(z_i)$: Diintegralkan terhadap nilai daerah ke-i.
a : Batas atas/akhir (daerah solusi).

p-ISSN: 2301-5373
e-ISSN: 2654-5101

b : Batas bawah/awal (daerah solusi).
Z : Nilai titik tengah (*output*)

2.5 *Mean Absolute Percentage Error* (MAPE)

Mean Absolute Percentage Error (MAPE) merupakan nilai rata – rata perbedaan absolut diantara nilai prediksi dan nilai realisasi yang dikalikan dengan seratus persen[7]. MAPE disebut juga sebagai hasil persenan dari nilai realisasi. Penggunaan MAPE yaitu untuk evaluasi hasil peramalan berupa tingkat akurasi angka peramalan terhadap angka realisasi[8]. Dengan kemudian hasil perhitungan MAPE kemudian dievaluasi berdasarkan pada tabel 3.

$$MAPE = \frac{1}{n}\sum_{t=1}^{n} \frac{|Qt-Pt|}{Qt} \times 100\% \quad (10)$$

keterangan:

Qt : nilai actual
Pt : nilai prediksi
n : jumlah data

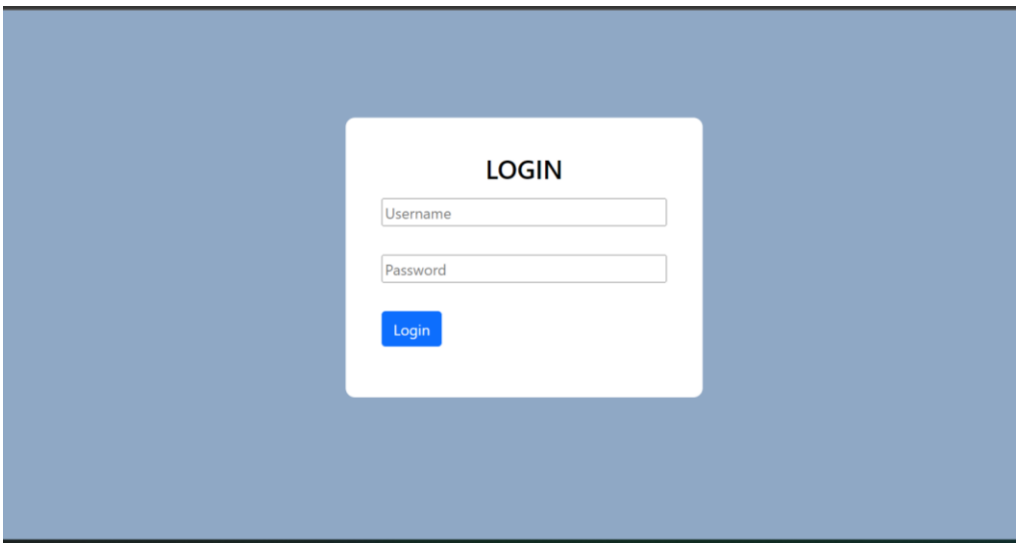
Tabel 3. Indikator evaluasi

MAPE Range	Keterangan
< 10%	Kompetensi model prediksi sangat baik.
10% - 20%	Kompetensi Model Prediksi Baik.
21% - 50%	Kompetensi model prediksi cukup.
>50%	Kompetensi model prediksi buruk.

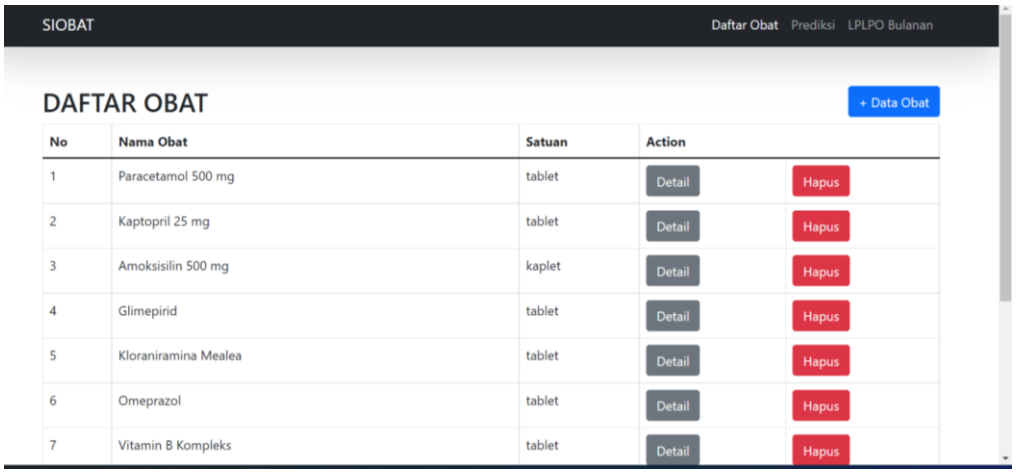
3. Hasil dan Pembahasan

3.1. Tampilan Antarmuka Sistem

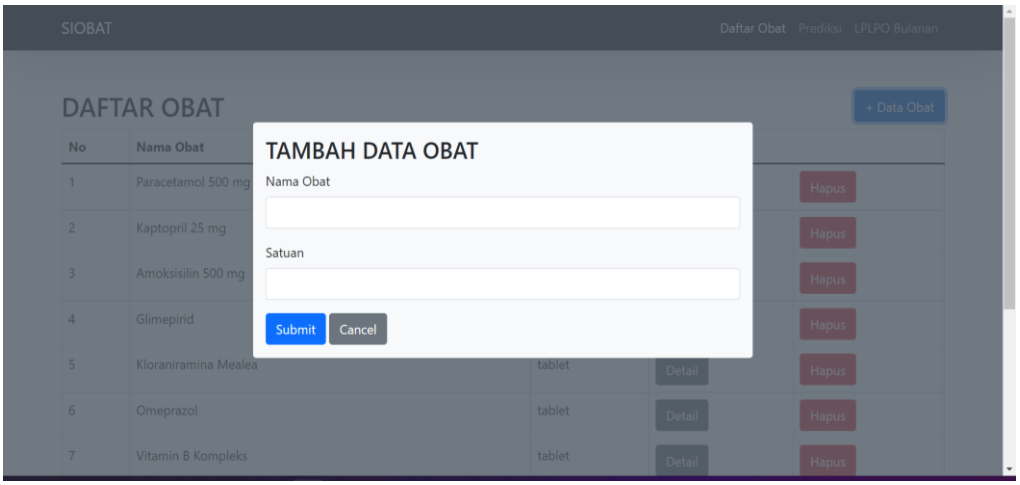
Adapun implementasi prediksi menggunakan *fuzzy inference system* model mamdani ke dalam visualisasi website dapat dilihat seperti pada gambar berikut.



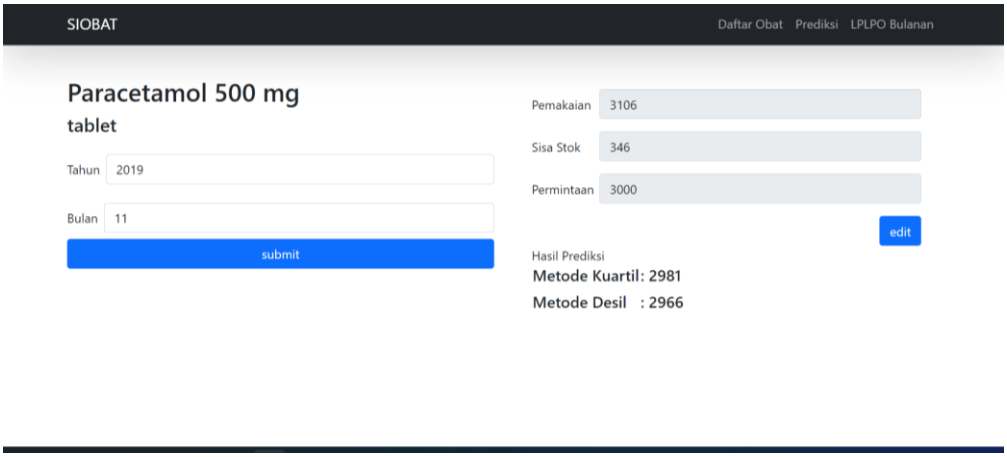
Gambar 3. Halaman Login



Gambar 4. Halaman Daftar Obat



Gambar 5. Halaman Tambah Data Obat



Gambar 6. Halaman Detail Obat

Adapun website akan digunakan oleh petugas kefarmasian. Dengan petugas terlebih dahulu login menggunakan username dan password pada kolom yang disediakan, seperti pada gambar 3 alaman login website. Selanjutnya petugas akan diarahkan pada tampilan daftar obat seperti pada gambar 4.

p-ISSN: 2301-5373
e-ISSN: 2654-5101

Petugas dapat melakukan penambahan data dengan tombol tambah data obat melakukan update data serta melihat prediksi dengan tombol detail, dan menghapus data dengan tombol hapus. Adapun tampilan untuk menambah obat seperti pada gambar 5. Pada gambar 6 menampilkan detail dari suatu obat, dengan detail tampilan terdapat *input* data dan menampilkan hasil prediksi. Petugas terlebih dahulu melakukan *input* tahun dan bulan dari satu obat. Kemudian dilakukan proses *input* dari pemakaian dan sisa stok, yang selanjutnya menghasilkan nilai prediksi permintaan obat. Selanjutnya petugas dapat memasukan nilai dari permintaan yang sebenarnya. Jika data sebelumnya sudah ada, maka akan menampilkan data sebelumnya yang telah tersimpan.

3.2. Hasil dan Evaluasi

Pada penelitian ini dilakukan menggunakan data sisa persediaan dan penggunaan obat untuk menentukan jumlah permintaan obat yang dibutuhkan untuk memenuhi kebutuhan bulan selanjutnya. Adapun prediksi permintaan obat dilakukan menggunakan *fuzzy inference system* model mamdani, dengan penentuan interval himpunan *fuzzy* dilakukan dengan metode kuartil serta metode desil. Berikut pada tabel 3 merupakan contoh dari penentuan himpunan *fuzzy.*

Tabel 4. Penentuan Himpunan Fuzzy

Variable	Semesta Pembicara	Himpunan Fuzzy	Interval Metode Kuartil		Interval Metode Desil	
Jumlah Sisa Persediaan	[0 1304]	Sedikit	[0	652]	[0	652]
		Sedang	[326	978]	[131	1174]
		Banyak	[652	1304]	[652	1304]
Jumlah Penggunaan	[0 2375]	Sedikit	[0	1188]	[0	1188]
		Sedang	[594	1782]	[238	2138]
		Banyak	[1188	2375]	[1188	2375]
Jumlah Permintaan	[0 3000]	Sedikit	[0	1500]	[0	1500]
		Sedang	[750	2250]	[300	2700]
		Banyak	[1500	3000]	[1500	3000]

Setelah ditentukannya interval dari setiap himpunan, selanjutnya dilakukan proses menghitung derajat keanggotaan. Adapun nilai dari derajat keanggotaan berkisar antara 0 hingga 1. Pada tabel 4 berikut merupakan contoh menghitung derajat keanggotaan dari variabel jumlah sisa persediaan dengan interval menggunakan metode kuartil.

Tabel 5. Menghitung Derajat Keanggotaan

Himpunan Fuzzy	Interval	Perhitungan	Pemenuhan Kondisi	Derajat Keanggotaan
(phi) S sedikit	x1 <= 326	1	no	0,9447
	326 <= x1 <= 652	(652 - x1) : 326	yes	
	x1 >= 652	0	no	
(phi) S sedang	x1 <= 326 or X1 >= 978	0	no	
	326 <= x1 <= 652	(X1 - 326) : 326	yes	0,0552
	652 <= x1 <= 978	(978 - x1) : 326	no	
(phi) S banyak	x<=652	0	yes	
	652 <= x1 <= 978	(X1 - 652) : 326	no	0
	x1 >=978	1	no	

Selanjutnya dilakukan proses implementasi fungsi implikasi, dengan penerapan fungsi minimum dan maksimum. Pada tabel 5 merupakan contoh penerapan fungsi minimum, dan dilanjutkan dengan penerapan fungsi maksimum dengan hasil seperti tabel 6. Adapun hasil tersebut berdasar dari hasil

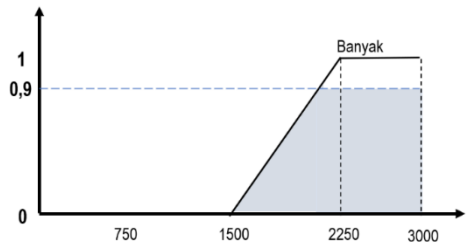
perhitungan derajat keanggotaan sebelumnya, yang kemudian menghasilkan daerah penyelesaian *fuzzy*. Adapun visualisasi daerah penyelesaian *fuzzy* untuk variabel jumlah sisa persediaan seperti pada gambar 7.

Tabel 6. Penerapan Fungsi Minimum

Jumlah Sisa Persediaan	Jumlah Penggunaan	Jumlah Permintaan	Himpunan Fungsi Rule
0,9447	0	0	Sedikit
0,9447	0	0	Sedang
0,9447	1	0,9447	Banyak
0,0552	0	0	Sedikit
0,0552	0	0	Sedang
0,0552	1	0,0552	Banyak
0	0	0	Sedikit
0	0	0	Sedang
0	1	0	Banyak

Tabel 7. Penerapan Fungsi Maksimum

Himpunan *Fuzzy*	*Alpha Predicate*
Sedikit	0
Sedang	0
Banyak	0,9447



Gambar 7. Visualisasi Daerah Penyelesaian *Fuzzy*

Setelah dilakukan perhitungan dari setiap variabel sehingga menghasilkan daerah penyelesaian *fuzzy*, kemudian dilakukan proses defuzzifikasi untuk mengubah bentuk *fuzzy set* ke suatu bilangan riil. Adapun proses defuzzifikasi diawali dengan menghitung momen dan luas daerah penyelesaian *fuzzy* seperti pada persamaan (6), (7), dan (8). Lalu dilakukan proses menghitung titik tengah sebagaimana persamaan (9). Berikut merupakan contoh hasil prediksi permintaan obat.

Tabel 8. Penerapan Fungsi Maksimum

Nama Obat	Sisa Persediaan	Pemakaian	Prediksi Kuartil	Prediksi Desil	Permintaan Sebenarnya
Amoksisilin kaplet 500 mg	631	1990	2431	2193	1500
Glimepirid tablet 1mg	837	1747	1559	1363	1000
Kaptopril tablet 25 mg	416	3140	3665	3191	3000
Ctm tablet 4 mg	147	1279	1612	1673	1500
Omeprazol	0	600	385	308	300
Parasetamol tablet 500mg	313	3865	2981	3180	3700
Vitamin b kompleks tablet	683	5794	6318	5015	5500

Dari hasil prediksi permintaan obat menggunakan *fuzzy inference system* dengan penentuan interval dengan metode kuartil dan desil, dilakukan evaluasi berdasarkan pada *mean absolute percentage error* sebagaimana persamaan (10) menghasilkan hasil prediksi seperti pada tabel 8. Berdasarkan

p-ISSN: 2301-5373
e-ISSN: 2654-5101

hasil perolehan tersebut, prediksi menggunakan *fuzzy inference system* dengan penentuan interval dengan metode kuartil dan desil memperoleh evaluasi cukup.

Tabel 9. Hasil Prediksi

Nama Metode	MAPE
Metode Desil	26.43 %
Metode Kuartil	28.77 %

4. Kesimpulan

Prediksi permintaan dengan *fuzzy inference system* diawali dengan proses penentuan himpunan *fuzzy*, dan dilanjutkan dengan proses pembentukan interval untuk setiap himppunan *fuzzy*. Dalam penentuan interval dilakukan dengan metode kuartil serta metode desil. Setelah ditentukann interval, dilakukan proses perhitungan *fuzzy inference engine* model mamdani terhadap 84 data histori. Berdasarkan perhitungan tersebut diperoleh hasil prediksi dengan *mean absolute percentage error* sebesar 26.43 % untuk metode desil, serta 28.77 % untuk metode kuartil. Evaluasi hasil penerapan *fuzzy inference system* model mamdani dengan penentuan interval himpunan menggunakan metode kuartil maupun metode desil dalam prediksi permintaan sediaan farmasi di puskesmas diperoleh hasil kompetensi prediksi cukup. Yang mana berdasarkan hasil penelitian diketahui bahwa interval dari himpunan *fuzzy* memberikan pengaruh terhadap kompetensi prediksi yang dihasilkan. Untuk penelitian selajutnya diharapkan dapat meneliti pengaruh instrumen penelitian seperti bentuk representasi jumlah himpunan *fuzzy*, maupun pengaruh desil yang digunakan terhadap hasil prediksi yang diperoleh.

Daftar Pustaka

[1] Kementrian Kesehatan Republik Indonesia, *Standar Pelayanan Kefarmasian Di Puskesmas*, Peraturan. Jakarta: Berita Negara Republik Indonesia Tahun 2017 Nomor 206, 2017.

[2] N. Febriany, “Implementatiton of fuzzy time series methods,” *J. Math.*, pp. 29–49, 2016.

[3] A. Sukoco and R. Yuli Endra, “Penerapan Fuzzy Inference System Metode Mamdani Untuk Pemilihan Jurusan,” *J. Manaj. Sist. Inf. dan Teknol.*, pp. 89–99, 2016.

[4] C. P. P. Maibang and A. M. Husein, “Prediksi Jumlah Produksi Palm Oil Menggunakan Fuzzy Inference System Mamdani,” *J. Teknol. dan Ilmu Komput. Prima*, vol. 2, no. 2, p. 19, 2019, doi: 10.34012/jutikomp.v2i2.528.

[5] A. A. I. D. Fibriayora, G. K. Gandhiadi, N. K. T. Tastrawati, and E. N. Kencana, “APPLICATION OF MAMDANI FUZZY METHOD TO DETERMINE ROUND BREAD PRODUCTION AT PT VANESSA BAKERY,” vol. 8, no. 3, pp. 204–210, 2019.

[6] K. Harefa, “Penerapan Fuzzy Inference System untuk Menentukan Jumlah Pembelian Produk Berdasarkan Data Persediaan dan Penjualan,” *J. Inform. Univ. Pamulang*, vol. 2, no. 4, p. 205, 2017, doi: 10.32493/informatika.v2i4.1487.

[7] M. B. S. Junianto, “Fuzzy Inference System Mamdani dan the Mean Absolute Percentage Error (MAPE) untuk Prediksi Permintaan Dompet Pulsa pada XL Axiata Depok,” *J. Inform. Univ. Pamulang*, vol. 2, no. 2, p. 97, 2017, doi: 10.32493/informatika.v2i2.1511.

[8] J. G. Nayak, L. G. Patil, and V. K. Patki, “Groundwater for Sustainable Development Development of water quality index for Godavari River (India) based on fuzzy inference system,” *Groundw. Sustain. Dev.*, vol. 10, no. December 2019, p. 100350, 2020, doi: 10.1016/j.gsd.2020.100350.

This page is intentionally left blank.

Jurnal Elektronik Ilmu Komputer Udayana
Volume 11, No 3. February 2023

p-ISSN: 2301-5373
e-ISSN: 2654-5101

Pengembangan Model Ontologi Pada Sistem Informasi Bahasa Bali

I Made Yoga Mahendra[a1], Cokorda Pramartha[a2], I Putu Gede Hendra Suputra[a3], I Made Widiartha[a4], I Gede Arta Wibawa[a5], I Wayan Supriana[a6]

[a]Program Studi Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Udayana Bali, Indonesia
[1]yoga.mahendra2000@gmail.com
[2]cokorda@unud.ac.id
[3]hendra.suputra@unud.ac.id
[4]madewidiartha@unud.ac.id
[5]gede.arta@unud.ac.id
[6]wayan.supriana@unud.ac.id

Abstract

Balinese is the language of communication for the Balinese ethnic community. The Balinese language has a coarse to fine vocabulary called anggah-ungguhing basa Bali. The Balinese language is increasingly being abandoned by the younger generation of Bali. To help overcome these problems, the Balinese language needs to be built using the Balinese language system using the concept of a semantic ontology. The ontology development method used is Methontology and produces a consistent and valid ontology. The system built will be able to provide information about Balinese words. The system features semantic browsing, semantic search, and opposite words. To ensure system functionality, Black-Box testing was carried out. The result is that the system has good functionality.

Keywords: *Knowledge Management System, Balinese language, Semantic Web, Methontology, Ontology, Levenshtein Distance*

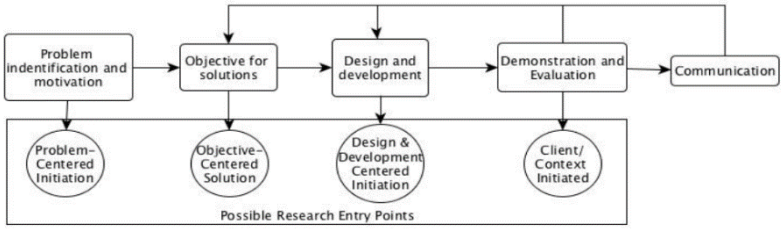
1. Introduction

Bali merupakan salah satu provinsi di Indonesia yang terkenal dengan banyak tempat wisata, seni budaya dan keunikan bahasanya, Penggunaan bahasa Bali sebagai bahasa utama memberikan kemudahan bagi masyarakat bali dalam berkomunikasi dalam suasana formal maupun informal. Keberadaan bahasa Bali sekarang tidak seperti dulu, dimana dalam pemakaian bahasa Bali tidak lagi sebagai bahasa utama dalam berkomunikasi bagi masyarakat Bali [1]. Generasi muda Bali yang seharusnya menjadi pelestari bahasa Bali malah enggan menggunakannya. Keberadaan bahasa Bali yang mulai bergeser, tidak mempengaruhi perhatian generasi muda terhadap keberadaan bahasa Bali. Maka Untuk membantu melestarikan Warisan Budaya Bali khususnya Bahasa Bali.Maka dari itu perlu dibangun sistem pengetahuan Bahasa Bali [2].

Dalam melestarikan warisan budaya ini tentu diperlukan sesuatu yang dapat mentransformasikan ke dalam bentuk digital dan eksplisit. Ontologi dapat membantu kemungkinan sebuah sistem manajemen pengetahuan untuk dapat membuka kemungkinan berpindah dari pandangan orientasi dokumen ke arah pengetahuan yang saling terikat, serta berkombinasi untuk dimaanfaatkan kembali secara fleksibel dan dinamis. Berdasarkan hal tersebut, Sistem yang dibangun yaitu Sistem Informasi Bahasa Bali Berbasis Web Semantik Ontologi. Pada penelitian ini, penulis mengembangkan ontologi yang kemudian akan diterapkan pada sistem Pencarian Bahasa Bali yang dibagi menjadi Bahasa Bali Alus (sor,singgih,mider), Bahasa Bali Andap dan Bahasa Bali Kasar. Pembangunan ontologi menggunakan metode Methontology.

2. Metode Penelitian

Metode penelitian yang digunakan dalam penelitian ini yaitu DSRM (Design Science Research Methodology). Metode DSRM memberikan pendekatan yang berguna untuk melakukan penelitian untuk membuat dan mengevaluasi desain untuk memecahkan masalah [9].



Gambar 1. Tahapan Design Science Research Methodology

Gambar 1 merupakan tahap dari metode DSRM, metode DSRM mempunyai 5 tahapan yaitu Identifikasi masalah dan inovasi, solusi objektif, desain dan pengembangan, pengujian dan evaluasi serta komunikasi.

2.1 Identifikasi Masalah dan Motivasi

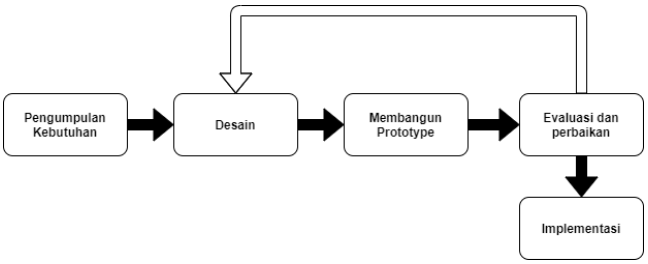
Tahapan ini merupakan tahapan untuk indentifikasi masalah yang diangkat pada penelitian ini. Dasar Permasalahan dalam penelitian ini adalah kurangnya atau minimnya sumber informasi yang diketahui masyarakat terhadap Pengetahuan Bahasa bali, baik itu Bahasa Bali Alus, Bahasa Bali Kasar dan Bahasa Bali Andap, hal ini diperparah dengan minimnya minat anak muda di Bali menggunakan Bahasa Bali dan kurang paham terkait tingkatan Bahasa Bali sehingga warisan budaya leluhur ini terancam punah.

2.2 Solusi Objektif

Tahap ini merupakan tahap penentuan solusi yang digunakan untuk mengatasi permasalahan pada indetifikasi dan motivasi yaitu dengan membuat aplikasi sistem yang dapat membantu pengguna mencari, mencari, dan tidak setuju dengan kata-kata dalam Bahasa Bali. Sistem yang akan dibangun menggunakan model ontologi, karena model ontologi dapat digunakan untuk menyajikan informasi dengan cara yang lebih semantik (bermakna) dan terorganisir, serta dapat memetakan kumpulan sumber informasi yang terstruktur dan sistematis. Sehingga ketika pengetahuan dari Bahasa Bali dikumpulkan secara ontologi, maka akan memberikan kemudahan dalam pengorganisasian dan manajemen data [5].

2.3 Desain dan Pengembangan

Metode yang digunakan dalam pengembangan sistem adalah metode prototyping [10]. Kelebihan dari metode ini yaitu umpan balik yang didapat dari pengguna dengan cepat [11].



Gambar 2. *Tahapan Pembangunan Sistem Dengan Metode Prototyping*

Dalam Gambar 2 terdapat 5 tahapan dalam implementasi dari metode prototyping yaitu pengumpulan kebutuhan, desain, membangun prototype, evaluasi dan perbaikan serta implementasi.

a. Analisis Kebutuhan

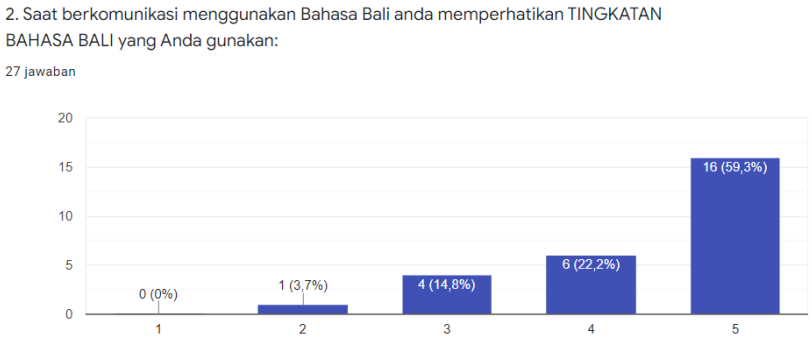
   Langkah-langkah yang dilakukan pengembangan sistem memerlukan penilaian komponen berupa kebutuhan awal dan analisa tentang ide maupun gagasan yang digunakan dalam membangun atau mengembangkan sebuah sistem. Analisis dilakukan untuk mengetahui seperangkat kebutuhan yang dibutuhkan ataupun hal yang digunakan dalam menunjang penelitian [4]. Pada tahapan ini terdiri dari 2 yaitu analisis kebutuhan fungsional meliputi

p-ISSN: 2301-5373
e-ISSN: 2654-5101

kegunaan sistem, dan analisis kebutuhan non-fungsional meliputi komponen eksternal yang digunakan dalam menunjang penelitian.

b. Data

Data kata bahasa bali yang digunakan dalam penelitian ini untuk membangun model ontologi dibagi menjadi 2 (tiga). Pertama yaitu melakukan wawancara dengan ahli Bahasa Bali atau narasumber yang dirasa memiliki pengetahuan tinggi dalam Bahasa Bali. Kedua yaitu melalui Buku sor singgih basa bali, kamus kruna alus mider bahasa bali, kamus anggah-ungguh kruna basa bali dan wawancara melalui ahli bahasa bali. Dalam pengumpulan data awal untuk penentuan kriteria dalam sistem, dilakukan survei secara online yang melibatkan beberapa pihak melalui kuesioner yang terdiri dari 27 responden. Dari hasil kuesioner yang ditampilkan pada gambar 3, didapatkan 6 kriteria yang digunakan dalam sistem yaitu Tingkatan Bahasa, Bentuk Kata, Kategori Kata, Imbuhan, Sisipan dan Akhiran.



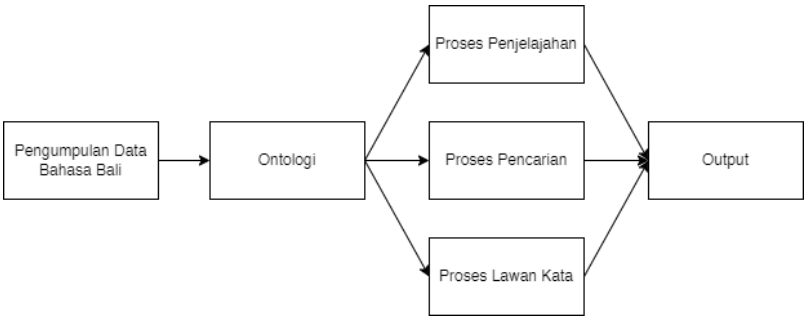
Gambar 3. Diagram Tingkatan Bahasa sebagai Kriteri Pencarian

c. Pembangunan Model Ontologi

Metode yang digunakan dalam pembangunan model ontologi untuk penelitian ini adalah methontology. Methontology ini merupakan salah satu metodologi untuk membangun model ontologi, dan memiliki keunggulan dalam memberikan detail dari setiap proses yang dilakukan. Selain itu, methontology juga memiliki kemampuan untuk menggunakan kembali ontologi yang dibuat untuk pengembangan sistem lebih lanjut [6].

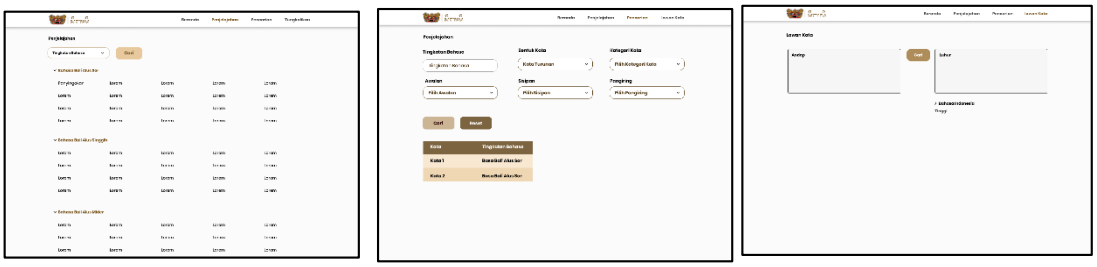
d. Desain

Dalam Tahapan ini, dilakukan pembangunan user interface dari fitur sistem yaitu penjelajahan, pencarian dan lawan kata serta desain umum sistem. Desain sistem yang akan dibuat akan melalui beberapa tahapan yang akan dilalui. Dimulai dari tahapan pengumpulan dan penyimpanan data, proses penjelajahan, dan proses pencarian hingga tahap evaluasi kinerja sistem yang dapat dilihat pada gambar 4 yang menunjukan rancangan desain umum sistem informasi bahasa bali.



Gambar 4. Desain Umum Sistem

Pada sistem yang akan dibangun ini, rancang user interface sistem hanya akan ditujukan untuk guest user. Berikut penjabaran dari rancang antarmuka fitur penjelajahan dan pencarian pada sistem dapat dilihat pada gambar 5.



Gambar 5. Rancangan *User Interface* penjelajahan, pencarian dan lawan kata pada sistem

2.4 Evaluasi Sistem

Tahapan evaluasi prototyping memiliki tujuan untuk memastikan prototipype yang dibangun sudah sesuai dengan kriteria yang dibutuhkan. Setelah dilakukan evaluasi, terdapat perbaikan dalam perancangan, sehingga sistem dinyatakan benar dan layak untuk dibangun [12].

2.5 Komunikasi

Tahapan terakhir dalam pengembangan aplikasi ini yaitu mendokumentasikan segala pengetahuan yang berkaitan dengan penelitian agar hasil dari penelitian dapat disimpan dalam bentuk arsip tulisan pada buku tugas akhir kemudian hasil tulisan ini dapat diterbitkan dalam jurnal ilmiah.

3. Hasil dan Pembahasan

Hasil dan pembahasan pada penelitian ini berkaitan dengan hasil pengembangan sistem informasi bahasa bali, serta hasil evaluasi sistem.

3.1 Desain dan Pengembangan Sistem

Dalam mengimplementasikan metode *Design Science Research Methodology* memerlukan tahapan desain dan pembangunan.

a. Pengembangan Model Ontologi

Pada bagian ini akan dijabarkan implementasi dari pengembangan ontologi sesuai dengan tahapan yang telah ditentukan. Berikut ini implementasi dari tahapan metode pembangunan ontologi dengan metode Methontology.

1. Tahapan Spesifikasi

   Dalam tahap ini, dihasilkan deskripsi dari ontologi Bahasa Bali yaitu Domain Bahasa Bali, tujuan untuk mengembangkan model ontologi yang memberikan kemudahan dalam pengklasifikasian Kata pada Bahasa Bali, Tingkat formalitas penelitian formal, ruang lingkup penelitian kata pada bahasa bali dan Sumber pengetahuan yang digunakan melalui buku serta wawancara

2. Tahap Akuisisi Pengetahuan

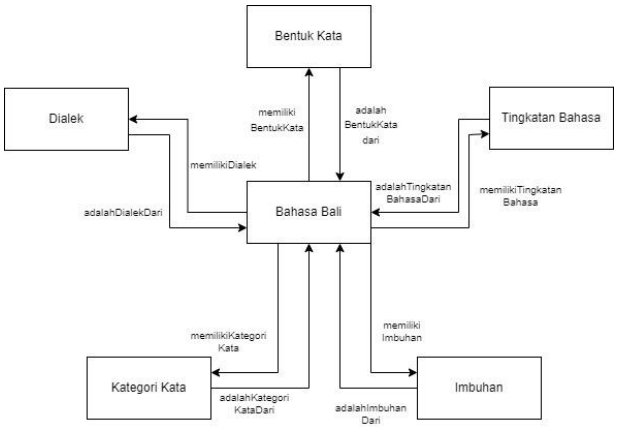
   Dalam tahap ini, teknik-teknik yang penulis gunakan untuk mengakuisisi pengetahuan ontologi Bahasa Bali yaitu Berdiskusi dengan dosen dan wawancara dengan salah satu Dosen Sastra Bali, disini peneliti melakukan wawancara untuk membangun pengetahuan yang ada pada kata bahasa bali. Melakukan analisis teks informal, untuk mempelajari alur konsep yang dijelaskan dalam buku dan studi pegangan, dan melakukan analisis teks formal. Melakukan analisis teks formal untuk memberikan pemahaman struktur yang akan ditemukan(definisi, penegasan, dan lain-lain) dan jenis pengetahuan yang dikontribusikan oleh masing-masing (konsep, atribut, nilai, dan hubungan).

3. Tahap Konseptualisasi

   Pada gambar 6 merupakan gambaran dari model formal ontologi dimana setiap relasinya dapat memiliki triplet (Subjek-Predikat-Objek), dimana salah satu contohnya adalah pada triplet daerah dari Bahasa Bali, individual “ngamaang.andap” menjadi subject,

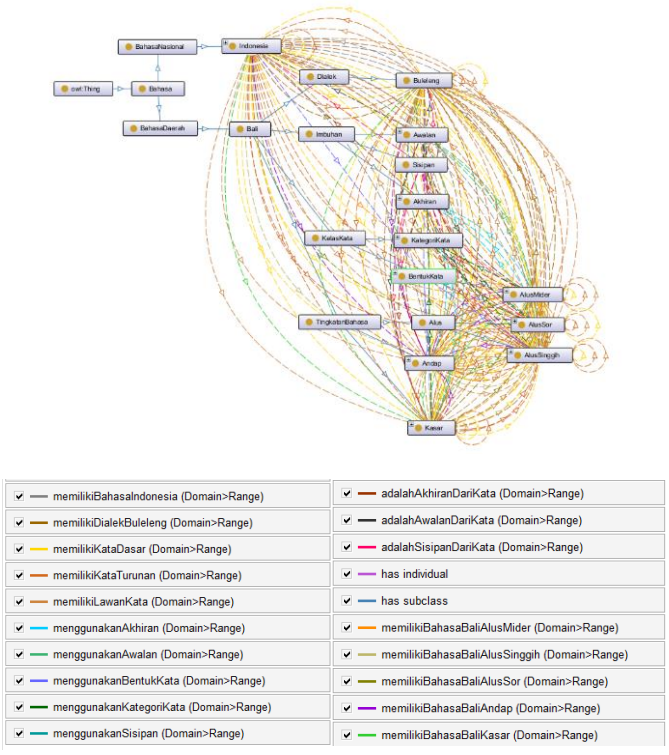
p-ISSN: 2301-5373
e-ISSN: 2654-5101

"memilikiKategoriKata" (object property) sebagai predicate, dan individual "kata_kerja" sebagai object nya.



Gambar 6. *Concept Taxonimies* pada ontologi Bahasa Bali

4. Tahap Integrasi
   Dalam tahap ini, penulis mengintegrasikan model ontologi yang dibuat dengan struktur pembentuk dari kata Bahasa Bali yang telah didiskusikan dengan ahli ontologi dan ahli Bahasa Bali.
5. Tahap Implementasi
   Perancangan konseptual ontologi telah diselesaikan menggunakan metode Methontology kemudian diimplementasikan menggunakan perangkat lunak Protégé 5.5.0. Pada perangkat lunak Protégé 5.5.0, setiap bagian dari ontologi diberikan makna sesuai dengan hasil dari alur tahapan tugas pada Methontology, concept didefinisikan sebagai class, ad-hoc binary relation didefinisikan sebagai object properties, dan instances didefinisikan sebagai individual. Pada gambar 7 terdapat hasil visualisasi dari ontologi Bahasa Bali berupa ontograf, Ontograf pada ontologi Bahasa Bali mendefinisikan 21 buah object properties, 222 individual ,1 data properties dan 20 class.



Gambar 7. Ontograf pada Ontologi Bahasa Bali

6. Tahap Evaluasi

Dalam tahap ini, model formal ontologi yang sudah melewati tahap pengembangan dilakukan inferensi menggunakan Pellet Reasoner untuk mengecek konsistensi dari ontologi. Dari proses reasoning yang dilakukan, ontologi Kata Bahasa Bali telah konsisten, yang dibuktikan dengan tidak munculnya pesan "Reasoner Error" yang terlihat pada gambar 8.

```
INFO  16:19:21  ------------------------------ Running Reasoner ------------------------------
INFO  16:19:22  Pre-computing inferences:
INFO  16:19:22      - class hierarchy
INFO  16:19:22      - object property hierarchy
INFO  16:19:22      - data property hierarchy
INFO  16:19:22      - class assertions
INFO  16:19:22      - object property assertions
INFO  16:19:22      - same individuals
INFO  16:19:23  Ontologies processed in 2024 ms by Pellet
INFO  16:19:23
INFO  17:36:15  ------------------------------ Running Reasoner ------------------------------
INFO  17:36:15  Pre-computing inferences:
INFO  17:36:15      - class hierarchy
INFO  17:36:15      - object property hierarchy
INFO  17:36:15      - data property hierarchy
INFO  17:36:15      - class assertions
INFO  17:36:15      - object property assertions
INFO  17:36:15      - same individuals
INFO  17:36:15  Ontologies processed in 74 ms by Pellet
INFO  17:36:15
```

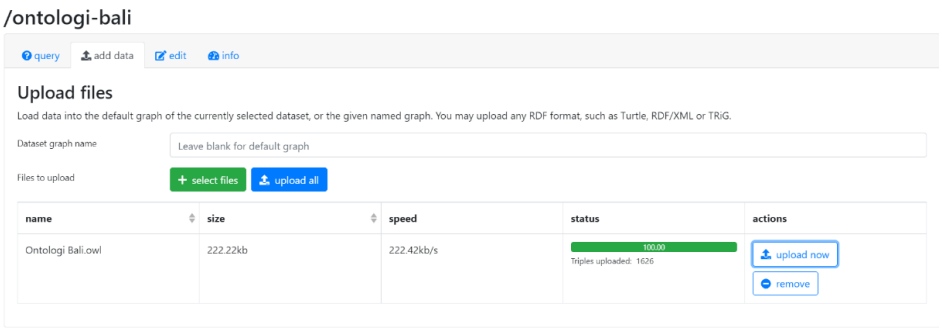
Gambar 8. Log Proses Reasoning Ontologi Bahasa Bali

b. Implementasi Sistem

Pada tahapan implementasi sistem dibagi menjadi dua yaitu proses menghubungkan ontology ke sistem dan mengimplementasikan fitur Penjelajahan, Pencarian dan lawan kata Bahasa Bali dengan hasil inputan pengguna.

1. Implementasi Ontologi kedalam sistem

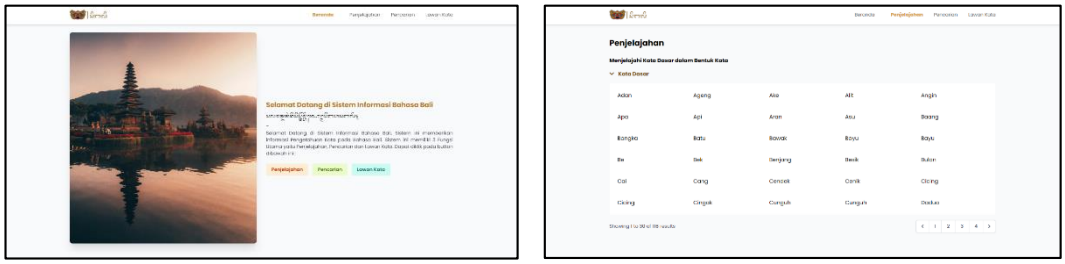
Dalam pembangunan sebuah website berbasis Ontology maka dibutuhkan sebuah server yang mampu mengelola data – data yang terdapat pada suatu Ontology. Server yang digunakan untuk mengelola sebuah Ontologi dinamakan server Apache Jena Fuseki. Pada gambar 9 menunjukan proses untuk mengunggah file Ontologi Bali.owl



Gambar 9. Upload OWL ke server Apache Jena Fuseki

2. Implementasi *User Interface* (Antarmuka) Sistem

User interface merupakan hal yang penting dalam sistem karena akan langsung berhubungan dengan pengguna. Berikut merupakan user interface dari fitur yang ada pada sistem

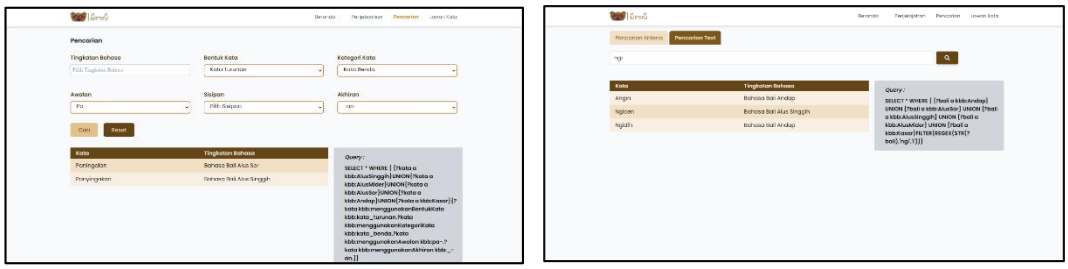


Gambar 10. *User Interface* Halaman Beranda dan Penjelajahan

Pada gambar 10 terdapat *user interface* Beranda dan Penjelajahan dari sistem. Pada menu beranda terdapat deskripsi singkat tentang sistem. Pada halaman penjelajahan pengguna

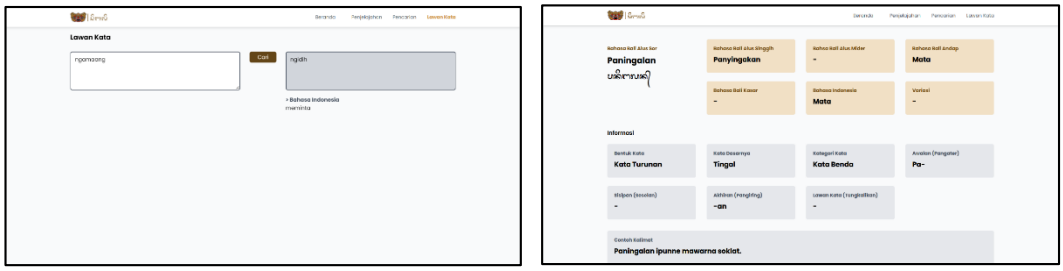
p-ISSN: 2301-5373
e-ISSN: 2654-5101

dapat melakukan penjelajahan berdasarkan kriteria yang diinginkan. Kriteria yang dimaksud adalah Tingkatan Bahasa, Bentuk Kata dan Kategori Kata. Setelah diklik akan tampil detail pengetahuan dari kata tersebut.



Gambar 11. *User Interface* Halaman Pencarian Kriteria dan Pencarian Text

Pada gambar 11 terdapat user interface Beranda dan Penjelajahan dari sistem. Pada halaman Pencarian pengguna dapat melakukan pencarian terhadap informasi pengetahuan Bahasa Bali dengan menentukan kriteria kata yang diinginkan dan dapat menggunakan kolom teks. Ketika permintaan sudah diinputkan maka query akan mencocokan dengan pemodelan ontologi menggunakan kata kunci yang dipilih. Data yang akan ditampilkan sesuai dengan permintaan yang diinginkan dan terdapat detail dari Kata tersebut.



Gambar 12. *User Interface* Halaman Lawan Kata dan Detail Kata

Pada gambar 12 terdapat user interface Beranda dan Penjelajahan dari sistem. Pada menu Lawan Kata terdapat sebuah inputan untuk mencari lawan kata dalam bahasa bali. Pada halaman detail kata, jika user memilih kata pada penjelajahan, pencarian maupun lawan kata akan diarahkan ke halaman detail. Pada halaman ini diberikan informasi dari pengetahuan yang ada pada kata bahasa bali.

3.2 Pengujian Sistem

Dalam pengujian sistem, dilakukan pengujian menggunakan Black-box. Pengujian Black-Box bertujuan untuk melakukan pengujian fungsional pada keseluruhan fitur yang ada pada sistem . Fitur yang diuji yaitu fitur pencarian, penjelajahan dan lawan kata. Skenario yang diuji yaitu halaman tampilan data, inputan pengguna, output pengguna.

a. Kode Pengujian

Tabel 1. Kode Pengujian Black Box

Nama Pengujian	Kode Pengujian	Pengguna	Hasil Diharapkan
Pencarian	P1	*Guest User*	Berhasil
Penjelajahan	P2	*Guest User*	Berhasil
Lawan Kata	P3	*Guset User*	Berhasil

Pada Table 1 terdapat keterangan terkait pengujian Black-Box terhadap masing-masing fitur utama dari sistem. Pada tabel tersebut terdapat kode P1 untuk fitur pencarian, P2 untuk fitur penjelajahan, dan P3 untuk fitur lawan kata dengan hasil yang diharapkan berhasil.

b. Pengujian Fitur Pencarian

Tabel 2. Pengujian Blackbox Fitur Pencarian

Nama Pengujian: Pencarian			Kode Pengujian: P1		
Pengguna: *Guest User*					
No	Kode	Nama Skenario	Hasil yang diharapkan	Hasil	Kesimpulan

1	P1-1	Halaman Pencarian	Sistem mampu menampilkan halaman pencarian	Sesuai	Berhasil
2	P1-2	Input Pencarian	Sistem mampu memasukan inputan kedalam fitur pencarian	Sesuai	Berhasil
3	P1-3	Output Pencarian	Sistem mampu menampilkan keluaran dari masukan pengguna	Sesuai	Berhasil
4	P1-4	Tampilan Halaman Pencarian	Tampilan dari sistem responsif, bisa diakses dalam mobile, tablet dan pc	Sesuai	Berhasil

Pada Tabel 2 merupakan hasil dari pengujian Black-Box pada fitur pencarian. Hasil yang diharapkan dan hasil dari pengujian sistem sesuai, maka dapat disimpulkan berhasil.

c. Pengujian Fitur Penjelajahan

Tabel 3. Pengujian Blackbox Fitur Penjelajahan

Nama Pengujian: Penjelajahan		Kode Pengujian: P2			
Pengguna: *Guest User*					
No	Kode	Nama Skenario	Hasil yang diharapkan	Hasil	Kesimpulan
1	P2-1	Halaman Penjelajahan	Sistem mampu menampilkan halaman penjelajahan	Sesuai	Berhasil
2	P2-2	Input Penjelajahan	Sistem mampu memasukan inputan kedalam fitur penjelajahan	Sesuai	Berhasil
3	P2-3	Output Penjelajahan	Sistem mampu menampilkan keluaran dari masukan pengguna	Sesuai	Berhasil
4	P2-4	Tampilan Halaman Penjelajahan	Tampilan dari sistem responsif, bisa diakses dalam mobile, tablet dan pc	Sesuai	Berhasil

Pada Tabel 3 merupakan hasil dari pengujian Black-Box pada fitur penjelajahan. Hasil yang diharapkan dan hasil dari pengujian sistem sesuai, maka dapat disimpulkan berhasil.

d. Pengujian Fitur Lawan Kata

Tabel 4. Pengujian Blackbox Fitur Lawan Kata

Nama Pengujian: Lawan Kata		Kode Pengujjian: P3			
Pengguna: *Guest User*					
No	Kode	Nama Skenario	Hasil yang diharapkan	Hasil	Kesimpulan
1	P3-1	Halaman Lawan Kata	Sistem mampu menampilkan halaman Lawan Kata	Sesuai	Berhasil
2	P3-2	Input Lawan Kata	Sistem mampu memasukan inputan kedalam fitur Lawan Kata	Sesuai	Berhasil
3	P3-3	Output Lawan Kata	Sistem mampu menampilkan keluaran dari masukan pengguna	Sesuai	Berhasil
4	P3-4	Tampilan Halaman Lawan Kata	Tampilan dari sistem responsif, bisa diakses dalam mobile, tablet dan pc	Sesuai	Berhasil

Pada Tabel 4 merupakan hasil dari pengujian Black-Box pada fitur lawan kata. Hasil yang diharapkan dan hasil dari pengujian sistem sesuai, maka dapat disimpulkan berhasil.

4. Kesimpulan.

Berdasarkan segala tahapan penelitian yang telah dilakukan dari awal hingga akhir, dapat menghasilkan kesimpulan yaitu Pengembangan Ontologi pada sistem informasi bahasa bali dapat dibangun dengan menggunakan metode *Methontology*. Ontologi yang dihasilkan sudah konsisten dan valid. Dari pengujian yang dilakukan terhadap sistem dapat disimpulkan bahwa hasil yang didapat sudah berjalan sesuai dengan apa yang diharapkan.

Referensi

[1] C. Cokorda, I. A. I. Iswara, and I. K. A. Mogi, "Digital Humanities: Community Participation in the Balinese Language Digital Dictionary," *J. Sist. Inf.*, vol. 16, no. 2, pp. 18–30, 2020, doi: 10.21609/jsi.v16i2.956.

[2] N. N. A. Suciartini, "Eksistensi Bahasa Bali Di Ranah Milenial (Studi Kasus Kemunculan Parodi Hai Puja)," *J. Ilmu Agama*, vol. 1, no. 2, pp. 134–149, 2018, [Online]. Available: http://marefateadyan.nashriyat.ir/node/150.

[3] I. made Wardana and C. Pramartha, "Development of Semantic Ontology Modeling in Knowledge Representation of Balinese Gamelan Instruments," *J. Elektron. Ilmu Komput. Udayana*, vol. 8, no. 2, p. 8, 2019.

[4] F. Abdussalaam and I. Oktaviani, "Perancangan Sistem Informasi Nilai Berbasis Web Menggunakan Metode Prototyping," *J. E-KOMTEK*, vol. 4, no. 1, pp. 16–29, 2020.

[5] A. Nguyen, L. Gardner, and D. Sheridan, "Towards Ontology-Based Design Science Research for Knowledge Accumulation and Evolution," *Proc. 52nd Hawaii Int. Conf. Syst. Sci.*, no. January, pp. 5755–5764, 2019, doi: 10.24251/hicss.2019.694.

[6] K. W. Triyoga, D. E. Cahyani, S. W. Sihwi, F. Matematika, and U. S. Maret, "Pembangunan Ontology Berbasis Metode Methontology Untuk Domain Tuberculosis," pp. 47–54, 2019.

[7] A. M. Sinaga, R. J. Sipahutar, and D. I. P. Hutasoit, "Penerapan Ontology Web Language pada Domain Ulos Batak Toba," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 5, no. 4, p. 493, 2018, doi: 10.25126/jtiik.201854903.

[8] I. L. Koten and C. Pramartha, "Semantic Representation of Balinese Traditional Dance. Jurnal Elektronik Ilmu Komputer Udayana," *J. Elektron. Ilmu Komput. Udayana*, vol. 8, no. 4, pp. 411–419, 2020.

[9] C. Pramartha, J. Davis, K. Kuan, and J. G. Davis, "Digital Preservation of Cultural Heritage: An Ontology-Based Approach," *Australas. Conf. Inf. Syst.*, no. December, pp. 1–12, 2017.

[10] C. R. A. Pramartha and N. P. S. H. Mimba, "Udayana University International Student Management: A Business Process Reengineering Approach," *ComTech Comput. Math. Eng. Appl.*, vol. 11, no. 2, pp. 57–64, 2020, doi: 10.21512/comtech.v11i2.6383.

[11] E. R. Subhiyakto and D. W. Utomo, "Analisis dan Perancangan Aplikasi Pemodelan Kebutuhan Perangkat Lunak Menggunakan Metode Prototyping," *Pros. Semin. Nas. Multi Disiplin Ilmu Call Pap. UNISBANK Ke-3*, pp. 57–62, 2017.

[12] C. Pramartha, P. Informatika, and U. Udayana, "Pengembangan sistem informasi penanganan penderita gangguan jiwa dengan pendekatan," vol. 5, no. 1, pp. 31–41, 2022.

This page is intentionally left blank.

Jurnal Elektronik Ilmu Komputer Udayana
Volume 11, No 3. February 2023

p-ISSN: 2301-5373
e-ISSN: 2654-5101

Perancangan Aplikasi Sistem Informasi Kain Tenun Endek Bali

Gde Deva Dimastawan Saputra[a1], Cokorda Pramartha [a2], I Gusti Agung Gede Arya Kadyanan[a3], Ida Bagus Gede Dwidasmara[a4], I Ketut Gede Suhartana[a5], Luh Arida Ayu Rahning Putri[a6]

[a]Program Studi Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Udayana
Bukit Jimbaran, Badung, Bali, Indonesia
[1]devadimastawan@gmail.com
[2]cokorda@unud.ac.id
[3]gungde@unud.ac.id
[4]dwidasmara@unud.ac.id
[5]ikg.suhartana@unud.ac.id
[6]rahningputri@unud.ac.id

Abstract

Bali memiliki banyak warisan budaya yang dapat menambah daya tarik wisatawan domestik maupun internasional. Salah satu warisan budaya tersebut adalah kain Endek Bali. Endek adalah kain tenun yang berasal dari Bali. Endek Bali umumnya dipakai untuk upacara adat, namun seiring berjalannya waktu kini banyak digunakan sebagai pakaian sehari-hari, seragam kantor dan seragam sekolah. Setiap Endek memiliki ciri khas berupa motif yang berbeda-beda. Umumnya ada yang memiliki motif fauna dan fauna, ukiran hingga wayang. Karena warisan budaya ini perlu dilestarikan agar tidak punah, maka diperlukan solusi untuk mewujudkannya dalam bentuk digital. Dalam mengatasi masalah tersebut, digunakan konsep ontologi semantik untuk merepresentasikan warisan budaya ini dalam bentuk digitalisasi. Pengembangan model ontologi ini nantinya dapat digunakan kembali untuk terus dikembangkan oleh penelitian selanjutnya. Ontologi Endek Bali menghasilkan 19 kelas, 16 properti objek, 2 properti data, dan 124 individu.

Kata kunci : *Endek, Ontologi, Semantik web, Methontology, SPARQL query.*

1. Pendahuluan

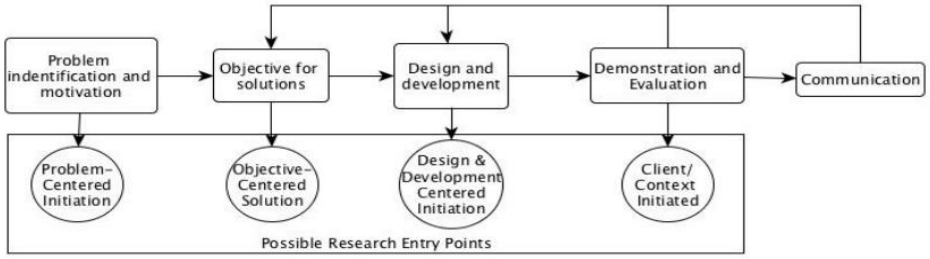
Endek adalah kain tradisional Bali yang termasuk kedalam kain tenun ikat, penggunaannya sudah tidak asing lagi bagi masyarakat Bali. Kain Endek adalah hasil dari seni karya dari seni rupa terapan, dalam hal ini kain endek dapat diterapkan atau digunakan dalam kehidupan sehari-hari seperti penerapan kain endek dalam pakaian adat, seragam sekolah maupun kantor. Industri kerajinan kain endek berasal dari wilayah Kabupaten Klungkung. Kain Endek merupakan warisan dari leluhur yang sekarang menjadikan kekayaan budaya bagi Bali. Berbagai jenis kain tenun geringsing, endek, cepuk, songket, dan yang lainnya. Kain Endek Bali juga dapat dikatakan warisan seni budaya Bali. Motif tersebut memberikan ciri khas tersendiri pada kain Endek dibandingkan dengan motif-motif kain pada umumnya [1]

Dalam melestarikan warisan budaya ini tentu diperlukan sesuatu yang dapat mentransformasikan ke dalam bentuk digital dan eksplisit. Ontologi dapat membantu kemungkinan sebuah sistem manajemen pengetahuan untuk dapat membuka kemungkinan berpindah dari pandangan orientasi dokumen ke arah pengetahuan yang saling terikat, serta berkombinasi untuk dimaanfaatkan kembali secara fleksibel dan dinamis [2] Ontologi juga memiliki keterkaitan dengan *web* semantik. Web semantik merupakan teknologi pada web yang dapat membantu sebuah komputer untuk memahami makna suatu kata atau kalimat yang diberikan oleh pengguna. Maka dengan web semantik komputer dapat lebih mudah memproses informasi serta mengerti informasi yang diinginkan oleh pengguna. Dengan dilakukan penelitian ini penulis ingin mendapatkan sebuah informasi terkait tentang Endek Bali untuk melestarikan budaya kita dan juga memberikan pemahaman kepada seluruh generasi tentang

warisan budaya ini. Data yang didapatkan dengan melakukan wawancara kepada narasumber ahli dari pengrajin kain Endek yang ada di Klungkung.

2. Metode Penelitian

Pada penelitian ini menggunakan alur metode *Design Science Research Methodology* (DSRM). DSRM ini merupakan metode yang berfokus untuk pada solusi permasalahan dan pengembangan sistem. Dimana dengan cara melakukan pendekatan yang dapat berguna untuk memberikan solusi dari permasalah yang diangkat[3]. Metodologi penelitian DSRM Ini akan menilai kemampuan kontribusi ahli dan non-ahli untuk melestarikan, dan bereksperimen dengan sistem prototipe yang akan dibuat. DSRM terdiri dari beberapa tahap: (1) Identifikasi masalah dan motivasi; (2) Tujuan solusi; (3) Desain dan pengembangan; (4) Demonstrasi dan Evaluasi; dan (5) Komunikasi[4]. Alur metode DSRM dapat dilihat pada gambar 1.



Gambar 1. Metode *Design Science Research Methontology*

2.1 Identifikasi Masalah dan Motivasi

Tahapan ini merupakan identifikasi masalah yang diangkat oleh peneliti yaitu permasalah yang di angkat oleh penulis adalah masih banyaknya masyarakat baik dari kalangan golongan muda maupun lansia masih kurang memahami tentang salah satu warisan budaya kita yaitu Kain Tenun Endek Bali karena kalau tidak dilestarikan warisan budaya kita ini nanti bisa saja generasi setelah nya tidak mengetahui tentang hal ini. Oleh karena itu penulis ingin membuat aplikasi sebagai wadah untuk masyarakat yang ingin mempelajari lebih dalam mengenai kain Endek dan juga dengan aplikasi ini diharapkan dapat membantu melestarikan warisan budaya Bali.

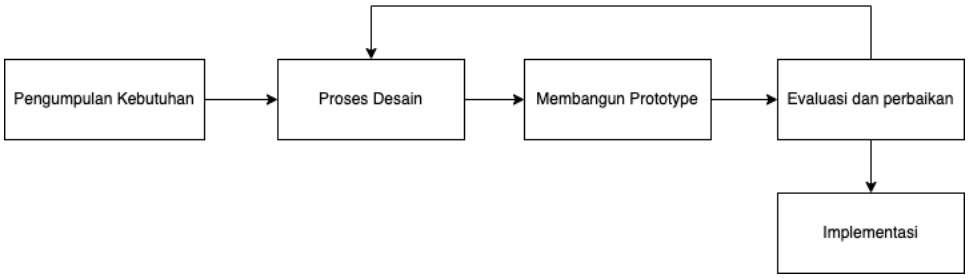
2.2 Solusi Objektif

Pada tahapan ini adalah menemukan solusi dari permasalahan yang ada yaitu dengan membuat sebuah aplikasi berbasis website. Solusi yang diciptakan oleh penulis nantinya diharapkan berguna sebagai wadah informasi tentang kain Endek. Saat ini kain Endek banyak digunakan oleh masyarakat saat ini tapi masih banyak yang belum mengetahui apa nama kain Endek tersebut, apa saja penyusun kain endek tersebut, apa saja motif yang ada pada kain endek dll. Diharapkan aplikasi ini juga nantinya memberikan nilai positif dan dapat diterima oleh masyarakat mengingat sekarang semua orang sudah banyak beralih menggunakan teknologi maka oleh sebab itu warisan budaya juga kita bisa lestarikan dan kembangkan dengan menggunakan teknologi yang ada saat ini.

2.3 Desain dan Pembangunan Sistem

Pada tahapan ini terdapat sebuah proses pengembangan sebuah sistem yang dibangun. Dimana dalam perancangan desain sistem peneliti menggunakan metode *prototyping*, penggunaan metode *prototyping* dalam perancangan sistem berguna untuk interaksi antara pengembang dan pengguna untuk saling berinteraksi selama proses pembuatan *prototype* pada sistem. Tahapan desain dan pengembangan.yaitu memberikan tampilan prototipe kepada calon pengguna. Klasifikasi dari kesetiaan prototipe dapat dibagi menjadi rendah, sedang, dan tinggi. Dalam penelitian termasuk kedalam penilaian medium-fidelity dimana prototyping dalam bentuk berbasis web aplikasi [5].Tahapan desain serta pengembangan sistem pada penelitian ini dilakukan mulai dari meliputi Analisis Kebutuhan, Pengumpulan data, Pembangunan Model, Pengembangan dan Desain Sistem, dan Perancangan Sistem. [6]

p-ISSN: 2301-5373
e-ISSN: 2654-5101



Gambar 2. Metode Prototyping

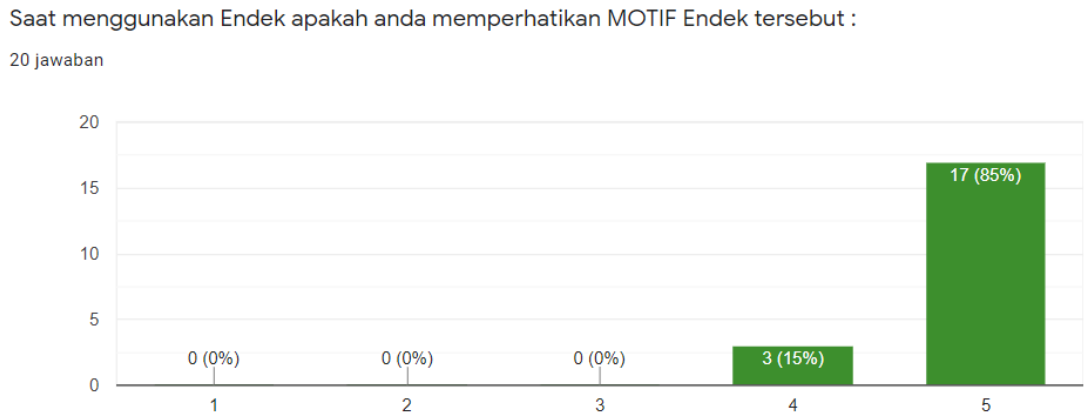
Pada gambar 2 merupakan alur dari metode *prototyping* pada sistem yang dimana pada yang (1) dimulai dengan pengumpulan kebutuhan dari kebutuhan sistem (2) merupakan tahapan desain dari sistem yang itu berupa desain mockup dari aplikasi (3) ketika sudah merancang desain maka dilanjutkan dengan membangun *prototype* dari sistem dengan menggunakan yang sudah dirancang pada tahapan sebelumnya (4) untuk tahapan ini merupakan tahapan yang dilakukan sebelum melakukan implementasi sistem yang dibangun seperti melakukan evaluasi atau perbaikan kepada data maupun desain pada sistem yang dirasa kurang (5) kemudian dilanjutkan dengan tahapan implementasi untuk menghasilkan sistem yang diinginkan.

a. Analisis Kebutuhan

Analisis kebutuhan adalah tahapan awal yang berguna untuk membangun struktur awal dalam pengembangan sistem seperti menganalisis mengenai ide atau inovasi dalam pembangunan sistem. Analisis juga dilakukan untuk mengetahui komponen apa saja yang sedang berjalan pada sistem baik *hardware* maupun software, jaringan dan pemakaian sistem sebagai level pengguna akhir pada sistem [7]. Dalam analisis kebutuhan ini terdapat dua hal yaitu analisis kebutuhan fungsional dan analisis kebutuhan nonfungsional yang difokuskan pada fungsi dari sistem tersebut .

b. Data

Tahapan pengumpulan data pada aplikasi SIEndek dibagi menjadi tiga tahapan yaitu pengumpulan data awal, pengumpulan data untuk pembangunan ontologi dan yang terakhir pengumpulan data pengujian. ada pengumpulan data awal peneliti melakukan penyebaran kuisioner kepada beberapa masyarakat yang berusia 17-24 tahun. Kuisioner yang diberikan bertujuan untuk mengetahui perbedaan apa saja yang menjadi hal diperhatikan oleh masyarakat ketika menggunakan endek seperti menggunakan endek memperhatikan bahan, motif dan warna.

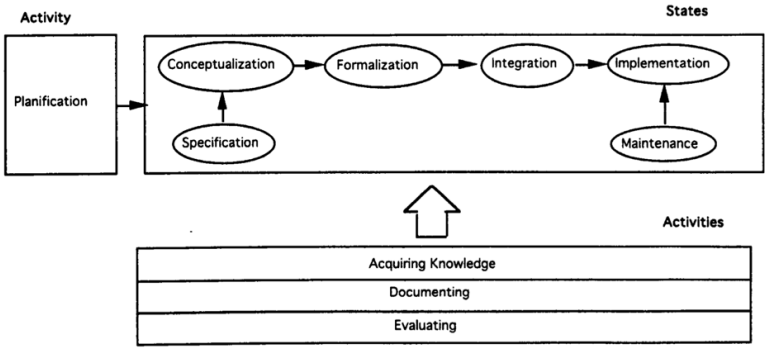


Gambar 3. Grafik Data Kuisioner Menggunakan Endek Memperhatikan Motif

Pada gambar 3 dapat dilihat hasil dari penyebaran kuisioner kepada masyarakat yang berjumlah 20 jawaban dimana diperoleh jawaban pertanyaan mengenai pemilihan Endek yang memperhatikan motif dari skala 1-5 terhadap 20 orang, dimana 85% responden sangat memerhatikan motif saat memilih endek.

c. Pembangunan Model

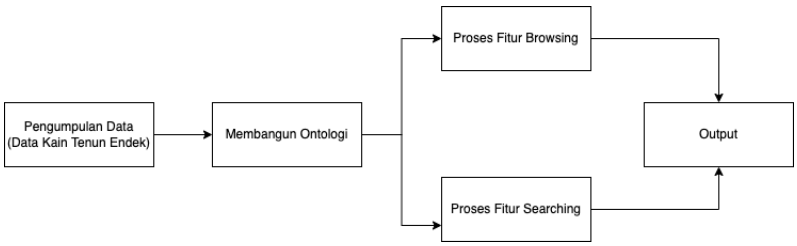
Dalam pembangunan model ontologi ini menggunakan metode *methontology*. *Methontology* merupakan sebuah metodologi pembangunan model ontologi yang memiliki keunggulan terkait dengan deskripsi setiap aktivitas yang harus dilakukan secara detail. Selain itu juga, dengan *methontology* ontologi yang kita bangun dapat digunakan kembali oleh pengembang sistem selanjutnya[8].



Gambar 4. Tahapan Methontology

d. Desain

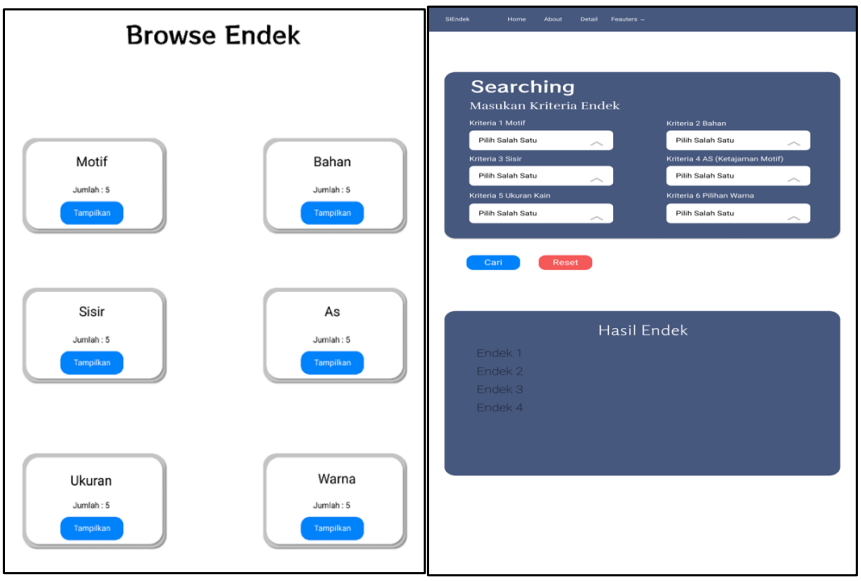
Dalam pembuatan desain sistem yang dibangun diperlukan beberapa tahapan yang harus dilalui yaitu, dimulai pengumpulan data ontologi kain tenun endek, selanjutnya membangun ontologi dari data yang dikumpulkan, selanjutnya proses fitur *searching,* fitur *browsing*, hingga tahapan evaluasi kinerja sistem.



Gambar 5. Desain umum sistem

Pada gambar 5 tahapan pertama yaitu pengumpulan data, dimana peneliti mengumpulkan data terkait informasi kain endek. Selanjutnya data yang sudah dikumpulkan tersebut akan diinput oleh peneliti pada model ontologi yang telah dibangun dan selanjutnya akan diimplementasikan pada sistem. Pada tahapan selanjutnya yaitu proses fitur *searching* dan *browsing,* merupakan tahapan dimana sistem akan mengeluarkan hasil atau output dari inputan yang dilakukan pengguna terhadap kain endek.

p-ISSN: 2301-5373
e-ISSN: 2654-5101



Gambar 6. Desain Fitur *Browsing dan Searching*

Pada halaman gambar 6 merupakan desain dari halaman *browsing* dan *searching*, pada *browsing* pengguna dapat memilih satu inputan kriteria Endek yang disediakan oleh sistem untuk menampilkan hasil endek yang terdapat pada sistem. Perbedaan dengan fitur searching dimana hanya dapat melakukan inputan dengan lebih banyak memilih kriteria, namun pada fitur browsing pengguna hanya dapat salah satu dari enam kriteria itu. Kriteria yaitu berupa motif, bahan, sisir, as, ukuran, dan warna. Jika pengguna memilih salah satu kriteria tersebut maka sistem akan menampilkan lagi bagian dari kriteria tersebut sebelum menampilkan endek yang sesuai dengan pilihan pengguna.

2.4 Pengujian dan evaluasi

Dalam tahap ini merupakan tahapan implementasi sistem untuk mendapatkan dan mengembangkan *software* maupun *hardware*, melakukan pelatihan serta perpindahan dan melakukan pengujian. pengujian *software* disini dilakukan untuk memastikan apakah sistem yang dibuat sesuai dengan desainnya dan semua fungsi dapat dipergunakan dengan baik tanpa adanya kesalahan.

3 Hasil dan Pembahasan

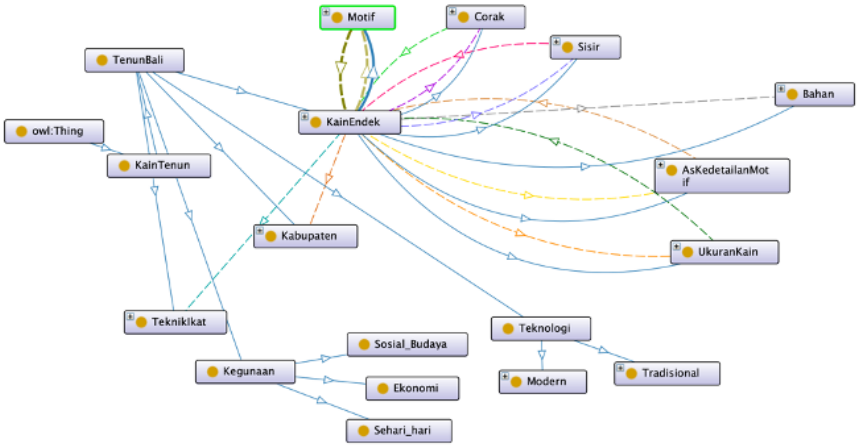
Dalam hasil dan pembahasan yang terdapat pada penelitian ini mengenai hasil dari pengembangan sistem informasi Kain Tenun Endek Bali yang telah dibuat. Dan juga menjelaskan hasil pengujian serta evaluasi sistem.

3.1. Desain dan Pengembangan Sistem

a. Pembangunan model Ontologi.

Perancangan konseptual ontologi yang dilakukan dalam menggunakan metode *methontology* dalam perancangan ontologi yang dibangun menggunakan aplikasi Protégé 5.5.0. Dalam pembangunan ontologi tersusun berdasarkan hirarki dari masing-masing class yang ada pada kain endek bali. Hirarki dari ontologi tersebut disusun berdasarkan dengan komponen yang ada pada kain endek merupakan hirarki kain endek terdapat 19 class yaitu kelas pertama merupakan class KainTenun dan memiliki subclass tenun bali kemudian subclass dari tenun terdapat 5 yaitu kabupaten, teknologi, kegunaan, teknik ikat dan domain dari penelitian yaitu KainEndek.

Dalam class teknologi dibagi menjadi 2 subclass yaitu teknologi modern dan teknologi tradisional. , sedangkan pada class kegunaan Kain Endek juga dibagi berdasarkan 3 subclass yaitu subclass sosial budaya, ekonomi, dan sehari-hari. Ontograf dari Kain Tenun Endek Bali dapat dilihat pada gambar 8 :

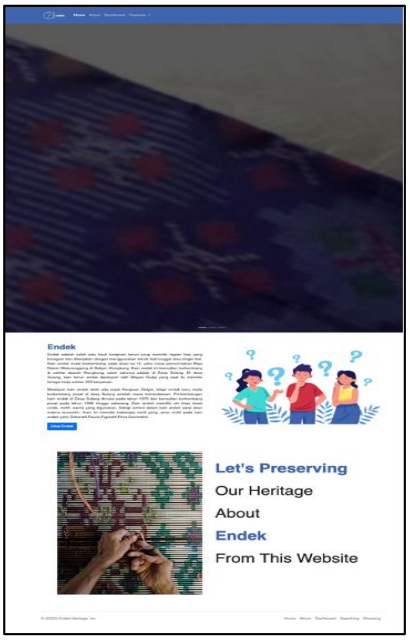


Gambar 8. Diagram Ontograf Kain Tenun Endek Bali

b. Implementasi Sistem

Dalam implementasi sistem ini dijelaskan sesuai dengan tahapan yang telah ditentukan. Dalam sistem ini hanya akan ada 1 jenis pengguna yang disebut dengan *guest user*, yang dimana *guest user* tersebut dapat melakukan pengoperasian fitur sistem seperti *searching, simple searching* dan *browsing.* Dalam pembangunan tampilan *user interface* pada sistem informasi Kain Tenun Endek yang berbasis web dengan menggunakan *framework* Laravel 9.2 dan *bootstrap-5* digunakan dalam membangun tampilan *front-end* halaman website. Dalam pengolahan data ontologi website digunakan sebuah server *Apache Jena Fuseki*. Berikut adalah contoh tampilan website dari Sistem Informasi Kain Tenun Endek Bali.

a)

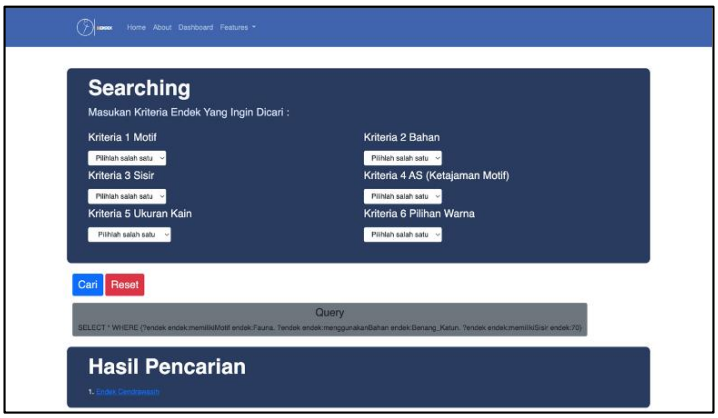


Gambar 9. Implementasi Antarmuka Halaman Landing Page

Pada gambar 9 merupakan halaman landing page merupakan tampilan awal dari aplikasi SIEndek yang dimana berisi gambar kain endek serta kata-kata ajakan untuk melestarikan endek. Dan pada halaman tersebut juga terdapat button untuk menuju ke halaman dashboard yang berisi data seluruh endek pada aplikasi SIEndek.

p-ISSN: 2301-5373
e-ISSN: 2654-5101

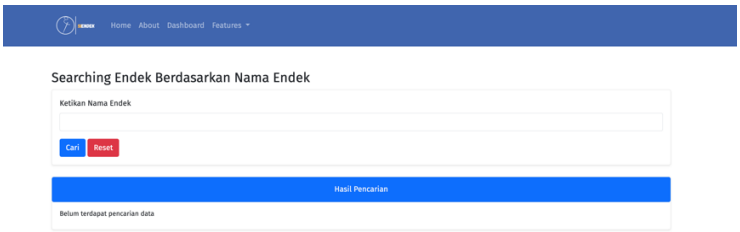
b)



Gambar 10. Implementasi Antarmuka Halaman *Searching*

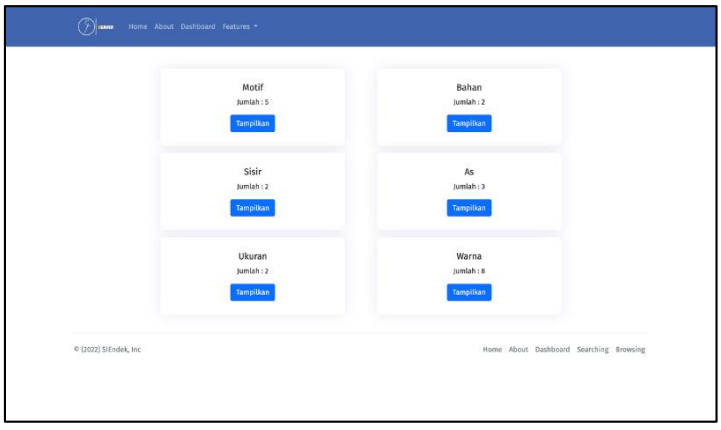
Gambar 10 ini merupakan tampilan dari salah satu fitur utama SIEndek yaitu searching. Dimana pada fitur ini pengguna dapat melakukan pencarian endek yang ada di aplikasi berdasarkan dengan kriteria umum kain endek yang ada pada aplikasi seperti motif, bahan, sisir, askedetailanmotif, ukuran kain dan warna. Pada fitur ini pengguna memilih 6 kriteria endek tersebut secara bersamaan sehingga hasil dari endek yang didapatkan dapat sesuai yang diinginkan berdasarkan dengan inputan kriteria.

c)



Gambar 11. Implementasi Antarmuka Halaman *Simple Searching*

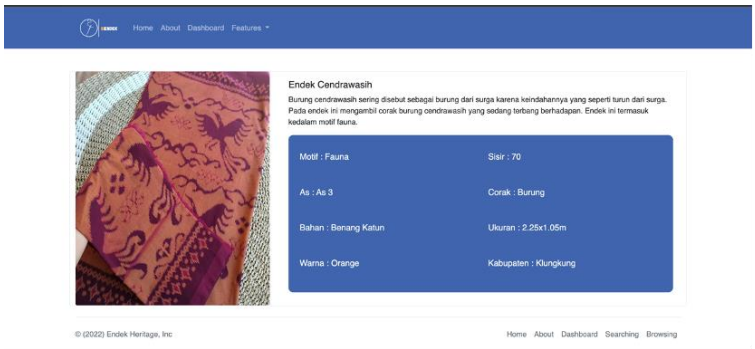
Gambar 11 adalah halaman *simple searching* pengguna dapat melakukan pencarian dari Kain Endek dengan cara mengetikan nama dari endek yang ingin dicari maka selanjutnya sistem akan menampilkan endek yang



Gambar 12. Implementasi Antarmuka Halaman *Browsing*

Pada halaman 12 adalah halaman *browsing*, pengguna dapat memilih satu inputan kriteria Endek yang disediakan oleh sistem untuk menampilkan hasil endek yang terdapat pada sistem. Kriteria yaitu berupa motif, bahan, sisir, as, ukuran, dan warna. Jika pengguna memilih salah satu kriteria tersebut maka sistem akan menampilkan lagi bagian dari kriteria tersebut sebelum menampilkan endek yang sesuai dengan pilihan pengguna.

d)



Gambar 12. Implementasi Antarmuka Halaman Detail

Halaman 12 merupakan halaman detail yang berisi spesifikasi atau informasi terkait tentang endek seperti deskripsi endek tersebut, komponen endek seperti motif, sisir, askedetailanmotif, ukuran kain, warna, dan bahan.

3.2. Pengujian dan evaluasi

Pengujian dengan *Black-Box Testing* berfungsi untuk menguji fungsionalitas dari fitur yang ada pada sistem dengan menggunakan metode Black-Box Testing sedangkan terdapat terdapat penugasan pemaanfaat sistem kepada calon pengguna dengan memberikan beberapa pertanyaan mengenai fitur-fitur pada sistem. Berikut merupakan alur dari tahapan pengujian :
Hasil pengujian *Black-Box Testing* dari fitur pencarian (*searching)* dan fitur penjelajahan (*browsing)* ditunjukkan oleh Tabel 2 dan Tabel 3. Berdasarkan hasil pengujian *Black-Box Testing* tersebut dapat dilihat bahwa hasil yang diberikan oleh sistem telah sesuai dan dapat dikatakan sistem telah memiliki fungsionalitas yang baik.

a. Blackbox-testing

Tabel 1. Pengujian *Blackbox Testing* Fitur *Searching*

Nama Pengujian : *Searching*		Kode Pengujjian: F1		
Kode	Nama Skenario	Hasil yang diharapkan	Hasil	Kesimpulan
F1-1	Halaman Searching	Sistem mampu menampilkan halaman searching	Sesuai	Normal
F1-2	Input Searching	Sistem mampu memasukan inputan kedalam fitur searching	Sesuai	Normal
F1-3	Output Searching	Sistem mampu menampilkan keluaran dari masukan pengguna	Sesuai	Normal

Tabel 2. Pengujian *Blackbox Testing* Fitur *Browsing*

Nama Pengujian: Penjelajahan			Kode Pengujjian: F2		
No	Kode	Nama Skenario	Hasil yang diharapkan	Hasil	Kesimpulan

1	F2-1	Halaman Browsing	Sistem mampu menampilkan halaman penjelajahan	Sesuai	Normal
2	F2-2	Input Browsing	Sistem mampu memasukan inputan kedalam fitur penjelajahan	Sesuai	Normal
3	F2-3	Output Browsing	Sistem mampu menampilkan keluaran dari masukan pengguna	Sesuai	Normal

4. Kesimpulan

Hasil pembahasan dalam perancangan sistem informasi kain endek yang dibangun menggunakan bahasa pemograman PHP dan dalam pengelola sumber daya informasi kain endek menggunakan ontologi untuk merepresentasikan pengetahuan pada sekumpulan konsep pada sebuah domain informasi yang dimana ontologi juga termasuk kedalam teknologi dari web semantik. Sedangkan untuk fitur-fitur *searching* dan *browsing* pada aplikasi sudah sesuai dengan apa yang diharapkan, dimana sudah dibuktikan dengan hasil pengujian *black box* pada pembahasan diatas yang menunjukan sistem mampu menampilkan halaman dari masing-masing fitur dan mengeluarkan input serta output yang diberikan dengan hasil yang sesuai dan memiliki kesimpulan normal. Diharapkan dengan adanya aplikasi ini dapat membantu masyarakat memberikan informasi terutama bagi yang memiliki terkaitan dengan kain endek.

Daftar Pustaka

[1] N. M. Ariani, “Pengembangan Kain Endek Sebagai Produk Penunjang Pariwisata Budaya Di Bali Endek Fabric Development As a Cultural,” vol. 9, no. 2, pp. 146–159, 2019.

[2] L. Mutawalli, I. F. Suhriani, and S. Supardianto, “Implementasi Sparql Dengan Framework Jena Fuseki Untuk Melakukan Pencarian Pengetahuan Pada Model Ontologi Jalur Klinis Tata Laksana Perawatan Penyakit Katarak,” *Jurnal Informatika dan Rekayasa Elektronik*, vol. 1, no. 2, p. 68, 2018, doi: 10.36595/jire.v1i2.66.

[3] C. Pramartha, “Assembly the Semantic Cultural Heritage Knowledge,” *Jurnal Ilmu Komputer*, vol. 11, no. 2, p. 83, 2018, doi: 10.24843/jik.2018.v11.i02.p03.

[4] C. Pramartha and J. G. Davis, “Digital preservation of cultural heritage: Balinese Kulkul artefact and practices,” *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 10058 LNCS, no. June 2018, pp. 491–500, 2016, doi: 10.1007/978-3-319-48496-9_38.

[5] C. R. A. Pramartha and N. P. S. H. Mimba, “Udayana University International Student Management: A Business Process Reengineering Approach,” *ComTech: Computer,*

Mathematics and Engineering Applications, vol. 11, no. 2, pp. 57–64, Dec. 2020, doi: 10.21512/comtech.v11i2.6383.

[6] Z. Zakaria *et al.*, “The Development of Personality Ontology Based on the Methontology Approach,” 2018.

[7] D. Purnomo, “Model Prototyping Pada Pengembangan Sistem Informasi,” *JIMP-Jurnal Informatika Merdeka Pasuruan*, vol. 2, no. 2, 2017.

[8] C. Pramartha, J. G. Davis, and K. K. Y. Kuan, “A Semantically-Enriched Digital Portal for the Digital Preservation of Cultural Heritage with Community Participation,” *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 11196 LNCS, no. October, pp. 560–571, 2018, doi: 10.1007/978-3-030-01762-0_49.

Jurnal Elektronik Ilmu Komputer Udayana
Volume 11, No 3. February 2023

p-ISSN: 2301-5373
e-ISSN: 2654-5101

Sistem Pencarian Produk Skincare Berbasis Ontologi

Made Rusdinda Hartani[a1], I Gusti Agung Gede Arya Kadyanan[a2], Cokorda Rai Adi Pramartha[a3], I Gusti Ngurah Anom Cahyadi Putra[a4], I Gede Arta Wibawa[a5], I Made Widiartha[a6]

[a]Program Studi Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Udayana
Bukit Jimbaran, Badung, Bali, Indonesia
[1]dindahartani24@gmail.com
[2]gungde@unud.ac.id
[3]cokorda@unud.ac.id
[4]anom.cp@unud.ac.id
[5]gede.arta@unud.ac.id
[6]madewidiartha@unud.ac.id

Abstrak

Perawatan kulit (skincare) adalah prosedur merawat kulit yang dilakukan dengan menggunakan produk kecantikan serta disesuaikan dengan jenis kulit wajah masing-masing individu. Solusi untuk membantu calon pembeli skincare dalam memilih produk skincare yang sesuai dengan masalah kulitnya adalah dengan membuat sebuah sistem pencarian produk skincare yang memiliki knowledge base. Salah satu pendekatan yang dapat dilakukan adalah dengan pendekatan ontologi dan diwujudkan dalam bentuk web semantik. Pembangunan ontologi produk skincare dibangun dengan metode Methontologi, dan dihasilkan model ontologi yang terbentuk dari 7 class utama, 12 object properties, dan 88 instances. Sedangkan, pembangunan sistem berbasis web dilakukan dengan mengimplementasikan metode Prototyping dan sistem yang dibangun memiliki 2 buah fitur, yaitu fitur pencarian (searching) dan fitur penjelajahan (browsing). Evaluasi sistem menggunakan Technology Acceptance Model (TAM) yang terdiri dua faktor yaitu persepsi kemudahan dan persepsi kegunaan. Pada evaluasi kemudahan dan evaluasi kegunaan sistem diperoleh nilai rerata sebesar 6.44 untuk kedua evaluasi, nilai tersebut menunjukkan persepsi "Setuju" dari peserta pengujian. Hasil tersebut menunjukkan bahwa rata-rata peserta pengujian setuju bahwa sistem pencarian produk skincare yang dibangun adalah sistem yang berguna dan mudah untuk digunakan.

Keywords: Skincare, Ontologi, Methontologi, Prototyping, TAM

1. Pendahuluan

Bagi wanita, kulit wajah yang sehat dan indah menjadi faktor dalam menunjang penampilan dan meningkatkan kepercayaan diri. Sebuah upaya yang dapat dilakukan untuk memelihara kesehatan kulit wajah yaitu dengan menggunakan produk perawatan kulit atau skincare. Perawatan kulit (skincare) adalah prosedur merawat kulit yang dilakukan dengan menggunakan produk kecantikan dengan kandungan bahan yang aman serta disesuaikan dengan jenis kulit wajah masing-masing individu [1].

Berdasarkan riset yang dilakukan oleh Zap Clinic dan MarkPlus. Inc [2] ditemukan bahwa demi menjaga penampilan 81,7% wanita Indonesia lebih memilih untuk menggunakan produk skincare yang tepat dibanding menggunakan make-up. Riset tersebut juga mengungkapkan fakta bahwa produk skincare yang wajib dimiliki oleh wanita Indonesia yaitu face wash, moisturizer, toner, dan serum. Saat ini telah beredar ratusan produk skincare di pasaran dengan kriteria dan jenis yang beragam. Semakin beragam produk yang tersedia mengakibatkan pembeli harus lebih teliti dalam memilih produk yang sesuai dengan permasalahan kulitnya. Dalam artikel limone.id, dr. Irma Tarida Listyawati, SpKK yang merupakan dokter spesialis kulit dan kelamin berpendapat bahwa pemilihan skincare harus disesuaikan dengan masalah kulit dan target yang ingin dicapai dalam perawatan tersebut [3].

Solusi yang dapat dilakukan untuk membantu calon pembeli skincare dalam memilih produk skincare yang sesuai dengan masalah kulitnya adalah dengan membuat sebuah sistem pencarian produk

skincare yang memiliki knowledge base. Salah satu pendekatan yang dapat dilakukan adalah dengan pendekatan ontologi dan diwujudkan dalam bentuk web semantik. Penggunaan ontologi memungkinkan pendeskripsian data menjadi lebih baik, sehingga dapat memiliki keterhubungan atau keterkaitan yang jelas antara satu data dengan data yang lainnya pada sebuah halaman website [4]. Sedangkan, pemanfaatan teknologi informasi berbasis internet akan memudahkan calon pembeli skincare untuk melakukan pencarian informasi mengenai produk skincare yang sesuai dengan masalah kulit yang dialami. Namun dalam pelaksanaannya, para calon pembeli skincare sering kali mendapatkan informasi lengkap yang tidak berada pada website yang sama. Sehingga memerlukan waktu dan tenaga lebih untuk menyusun informasi yang sesuai kebutuhan calon pembeli. Selain itu, calon pembeli harus memastikan bahwa informasi yang didapat telah relevan. Adopsi teknologi web semantik dapat mengatasi permasalahan tersebut.

Pada penelitian ini akan dibangun sebuah sistem pencarian produk skincare sesuai dengan masalah kulit wajah. Sistem pencarian ini akan dibuat dengan memanfaatkan teknologi web semantik dan data akan dipetakan ke dalam bentuk ontologi sebagai basis pengetahuan. Model ontologi dibangun menggunakan metode Methontologi, sedangkan pengembangan sistem menggunakan metode Prototyping. Penelitian ini diharapkan mampu memfasilitasi calon pembeli skincare dalam melakukan pemilihan produk sesuai masalah kulitnya.

2. Metode Penelitian

Alur dari penelitian ini meliputi beberapa tahapan yaitu identifikasi permasalahan, studi literatur, pengumpulan data, pembangunan ontologi, perancangan dan pengembangan sistem, serta tahap pengujian dan evaluasi.

Tabel 1. Contoh Data Produk Skincare

Merek Skincare	Nama Produk	Jenis Skincare	Tipe Kulit	Waktu Penggunaan	Masalah Kulit	Usia
SM	Supple Power Hyaluronic9+Onsen Moisture Bomb Gel	*Moisturizer*	Berminyak, kering, sensitif, dan kombinasi	Pagi dan Malam	Garis halus, kusam, kerutan, dan dehidrasi	Remaja
TS	3% Astaxanthin + Chlorelina Serum	Serum	Berminyak, kering, sensitif, dan kombinasi	Pagi dan malam	Dehidrasi dan kerutan	Remaja
TS	Mugwort Cica Essence Toner - Soothing & Hydrating	Toner	Berminyak, kering, sensitif, dan kombinasi	Pagi dan malam	Kulit kasar, dehidrasi, dan kusam	Remaja
WL	Acne Calming Serum	Serum	Berminyak, kering, sensitif, dan kombinasi	Pagi dan malam	Jerawat, kemerahan, dehidrasi, dan kusam	Remaja
AV	Miraculous Acne Solution Bubble-Serum Infused Cleanser	*Face wash*	Berminyak, kering, sensitif, dan kombinasi	Pagi dan malam	Jerawat, kusam, komedo hitam, komedo putih, dan pori-pori besar	Remaja

2.1. Identifikasi Permasalahan

Tahap ini sangat penting dalam proses penelitian karena jalannya proses penelitian terlaksana berdasarkan permasalahan yang terjadi. Permasalahan dalam penelitian ini adalah beredarnya produk skincare di pasaran dengan kriteria dan jenis yang beragam semakin meningkat. Semakin beragam produk yang tersedia mengakibatkan pembeli harus lebih teliti dalam memilih produk yang sesuai dengan permasalahan kulitnya. Pemilihan produk skincare yang tepat menjadi sebuah permasalahan untuk calon pembeli skincare.

2.2. Studi Literatur

Studi literatur dilakukan dengan peninjauan terhadap beberapa literatur berupa artikel-artikel ilmiah dan penelitian-penelitian yang didokumentasikan ke dalam bentuk jurnal, serta sumber internet yang berhubungan dengan domain skincare, ontologi dan web semantik.

2.3. Pengumpulan Data

Pada pengumpulan data, data yang digunakan dalam penelitian ini adalah data produk skincare dari beberapa merek skincare lokal. Pengumpulan data dibedakan menjadi dua, yaitu metode pengumpulan data untuk pembangunan model ontologi dan metode pengumpulan data untuk pengujian dan evaluasi sistem. Metode pengumpulan data untuk pembangunan model ontologi juga dibedakan menjadi dua, yaitu metode pengumpulan data untuk menentukan kriteria skincare dan metode pengumpulan data untuk produk skincare. Pengumpulan data kriteria skincare dengan melakukan wawancara kepada narasumber yang dirasa memiliki pengetahuan mengenai skincare dan pengumpulan data produk skincare dilakukan dengan studi literatur terkait melalui sumber internet terpercaya yaitu website resmi dari beberapa produsen skincare lokal. Contoh data produk skincare yang telah dikumpulkan dapat dilihat pada Tabel 1. Kemudian untuk metode pengumpulan data pengujian dan evaluasi sistem didapatkan melalui kuesioner yang dibagikan kepada peserta pengujian dan evaluasi sistem. Peserta pengujian dan evaluasi memenuhi kriteria usia 11-30 tahun.

2.4. Pembangunan Ontologi

Metode yang digunakan dalam membangun model ontologi pada penelitian ini adalah metode Methontologi. Methontologi adalah sebuah metodologi yang memungkinkan pembangunan ontologi pada level pengetahuan. Methontologi memiliki kemampuan untuk melakukan life cycle ontologi yang didasarkan pada pengembangan prototype yang mengijinkan untuk melakukan penambahan, perubahan, dan penghapusan terms pada tiap versi terbarunya [5].

a. Spesifikasi
   Tahap spesifikasi dilakukan untuk menghasilkan dokumen spesifikasi ontologi informal, semi formal atau formal yang ditulis dalam bahasa alami menggunakan representasi menengah atau pertanyaan kompetensi.
b. Akuisisi Pengetahuan
   Tahap akuisisi pengetahuan adalah kegiatan independen dalam proses pengembangan ontologi. Dalam tahap ini, teknik-teknik yang penulis gunakan untuk mengakuisisi pengetahuan ontologi Produk skincare.
c. Konseptualisasi
   Pada tahap konseptualisasi akan disusun pengetahuan domain dalam model konseptual yang menggambarkan masalah dan solusinya dalam hal kosa kata domain yang diidentifikasi dalam aktivitas spesifikasi ontologi. Hal yang dilakukan pada tahap ini yaitu membangun daftar istilah lengkap yang mencakup konsep, instance, kata kerja, dan properti. Dalam tahap ini, akan dihasilkan model konseptual dari ontologi produk skincare.
d. Integrasi
   Pada tahap integrasi ini mempertimbangkan penggunaan kembali definisi yang telah dibangun dalam ontologi karena ontologi memiliki sifat reusable. Dalam mempertimbangkan penggunaan kembali definisi yang sudah dibangun ke dalam ontologi, perlu dilakukan pemeriksaan meta-ontologi untuk memilih kesesuaian konsep yang akan dikembangkan pada domain ontologi yang akan dibangun. Hal ini bertujuan untuk menjamin bahwa set definisi baru dan yang digunakan kembali didasarkan pada set istilah dasar yang sama spesifikasi.
e. Implementasi

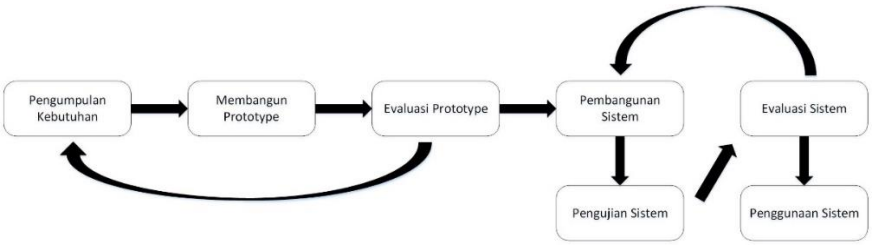
Tahap ini merupakan proses implementasi dari perancangan ontologi yang telah dibuat pada tahapan sebelumnya. Dalam tahap ini, dilakukan proses pendefinisian kembali dan proses implementasi dari perancangan ontologi menggunakan perangkat lunak Protégé.

f. Evaluasi
Tahap evaluasi merupakan tahap dengan melaksanakan penilaian teknis ontologi, lingkungan perangkat lunak, dan dokumentasinya sehubungan dengan kerangka acuan selama setiap tahapan dan antara tahapan dari siklus hidup mereka. Evaluasi merangkum istilah verifikasi dan validasi yang mengacu pada proses teknis yang menjamin kebenaran ontologi, lingkungan perangkat lunak terkait, dan dokumentasi sehubungan dengan kerangka acuan selama setiap tahapan dan antara tahapan dari siklus hidup mereka.

g. Dokumentasi
Pada tahap terakhir ini, dilakukan proses dokumentasi baik dalam kode ontologi, teks bahasa alami yang dilampirkan pada definisi formal, maupun makalah yang diterbitkan dalam proses konferensi dan jurnal yang mengatur pertanyaan-pertanyaan penting dari ontologi yang sudah dibangun.

2.5. Perancangan dan Pengembangan Sistem

Pada tahap perancangan dan pengembangan sistem akan menggunakan metode Prototyping. Metode Prototyping bertujuan untuk mengumpulkan informasi dari pengguna sehingga pengguna dapat berinteraksi dengan model prototype yang dikembangkan, sebab prototype menggambarkan versi awal dari sistem untuk kelanjutan sistem sesungguhnya yang lebih besar [6]. Tahapan-tahapan metode Prototyping dapat dilihat pada Gambar 1.



Gambar 1. Diagram Alur Prototyping

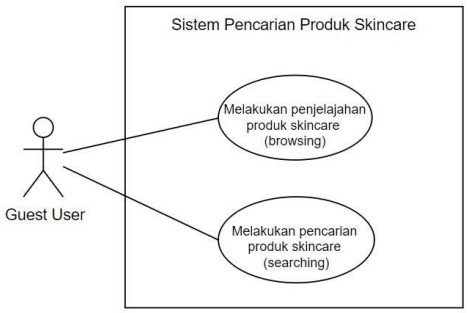
a. Pengumpulan Kebutuhan

1. Kebutuhan Fungsional
Sistem yang dirancang memungkinkan pengguna melakukan penjelajahan (browsing) dan pencarian (searching) informasi mengenai produk skincare dengan kriteria tertentu.
2. Kebutuhan Non-fungsional
Kebutuhan perangkat keras sistem yaitu laptop atau komputer yang akan digunakan dalam membangun sistem berbasis web. Sedangkan kebutuhan perangkat lunak sistem meliputi Protégé untuk pembangunan ontologi, SPARQL sebagai query dari model ontologi, Visual Studio Code sebagai text editor dalam pembangunan program berbasis web, Apache Jena Fuseki sebagai penghubung antara ontologi semantik dengan web.

b. Perancangan Prototype
Prototype yang akan dirancang merupakan sistem berbasis web yang memiliki fitur untuk penjelajahan semantik (semantic browsing), dan pencarian semantik (semantic searching).

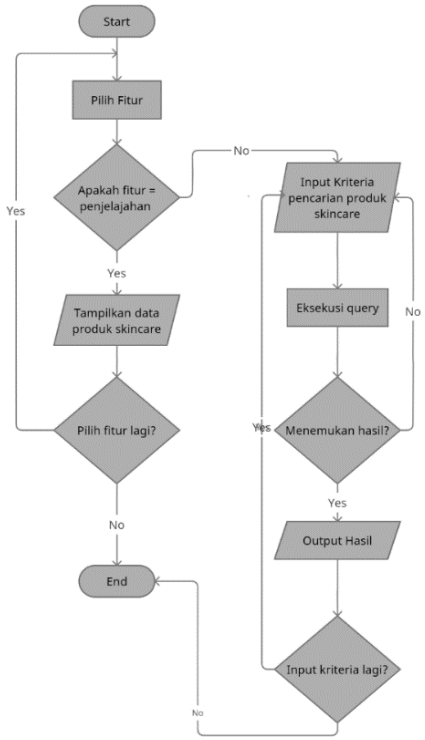
1. Use Case Diagram
Sistem Pencarian Produk Skincare memiliki sebuah aktor yaitu guest user. Guest User memiliki hak akses untuk melakukan 2 use case, yaitu melakukan penjelajahan (browsing) dan melakukan pencarian (searching) produk skincare. Use Case Diagram Sistem Pencarian Produk Skincare dapat dilihat pada Gambar 2.

p-ISSN: 2301-5373
e-ISSN: 2654-5101



Gambar 2. Use Case Diagram

2. Diagram Alir Sistem
   Prototype sistem akan dirancang berbasis web yang memiliki fitur untuk penjelajahan semantik (semantic browsing), dan pencarian semantik (semantic searching). Diagram alir sistem dari fitur-fitur prototype dapat dilihat pada Gambar 3.



Gambar 3. Diagram Alir Sistem

c. Evaluasi Prototype
   Pada tahap evaluasi prototype, dilakukan evaluasi terhadap prototype yang telah dibangun sebelumnya. Tahapan evaluasi bertujuan untuk mengetahui apakah prototype telah sesuai dengan keinginan. Tahap evaluasi ini juga mencakup kegiatan perbaikan perancangan dan berlangsung hingga sistem dinyatakan benar dan layak untuk dibuat.
d. Pembangunan Sistem
   Pada tahapan ini akan mulai dilakukan pembangunan sistem sesuai dengan rancangan atau prototyping yang telah dihasilkan tahapan sebelumnya. Adapun beberapa tahapan dalam pembangunan sistem ini, yaitu sebagai berikut:

   1. Penyiapan ontologi produk skincare yang sebelumnya telah dibangun menggunakan software Protégé 5.5.0 yang nantinya akan menjadi basis data dari sistem pencarian skincare.
   2. Menyiapkan *environment* sebagai tempat melakukan *deployment* sistem.

3. Tahap pengkodean dilakukan proses mengintegrasikan *library* EasyRDF ke dalam bahasa pemrograman PHP dan bahasa *query* SPARQL.
4. Menyiapkan *Virtual Private Server (VPS)* sistem sebagai tempat *running* sistem secara *online* agar sistem dapat diakses secara *online* oleh banyak orang.

e. Pengujian Sistem
Setelah tahap pembangunan sistem selesai, kemudian dilanjutkan dengan proses pengujian terhadap sistem. Tujuan pengujian sistem adalah agar sistem yang dibangun dapat berfungsi sesuai dengan yang diharapkan sebelumnya.
Pengujian Black-Box Testing merupakan kumpulan seri pengujian yang dilakukan pada user interface untuk menguji apakah hasil eksekusi telah sesuai dengan masukan yang diberikan dan memeriksa fungsional dari sistem. Pada tahap Black-Box Testing dilakukan pengujian terhadap fitur penjelajahan (browsing) dan pencarian (searching) sistem.

f. Evaluasi Sistem
Evaluasi sistem dilakukan dengan merekrut sejumlah peserta yang bersedia dan akan diberikan kuesioner yang terdiri dari beberapa pernyataan evaluasi sistem dan peserta harus memberikan skor persetujuan terhadap pernyataan tersebut. Evaluasi sistem menggunakan Technology Acceptance Model (TAM) yang terdiri dua faktor yaitu persepsi kemudahan (perceived ease of use) dan persepsi kegunaan (perceived usefulness).
Setelah jawaban peserta untuk pernyataan evaluasi terkumpulkan, maka peneliti akan melanjutkan dengan proses pengolahan dan analisis data evaluasi dengan melakukan analisis statistik berdasarkan skor yang telah ditetapkan pada skala Likert 7 poin (sangat setuju = 7, setuju = 6, agak setuju = 5, tidak setuju maupun tidak-setuju (netral) = 4, agak tidak setuju = 3, tidak setuju = 2, dan sangat tidak setuju = 1).

3. Hasil dan Pembahasan

Bagian hasil dan pembahasan menjelasakan mengenai implementasi ontologi, implementasi sistem, dan implementasi pengujian dan evaluasi sistem.

3.1. Implementasi Ontologi

a. Spesifikasi
Pada tahapan ini telah dihasilkan spesifikasi ontologi produk skincare dengan deskripsi sebagai berikut:
1. Domain: Produk skincare
2. Tanggal: 30 September 2021
3. Tingkat formalitas: Formal
4. Ruang lingkup: Produk skincare
5. Sumber pengetahuan: Wawancara dan studi literatur melalui website resmi produsen skincare lokal.

b. Akuisisi Pengetahuan
Pada tahap akuisisi pengetahuan ontologi produk skincare dilakukan dengan melalui beberapa teknik, yaitu sebagai berikut:
1. Berdiskusi dengan pembimbing terkait untuk membangun draf awal dokumen spesifikasi persyaratan.
2. Analisis teks informal, untuk mempelajari konsep-konsep utama yang diberikan dalam buku dan studi pegangan.
3. Analisis teks formal. Hal yang dilakukan adalah mengidentifikasi struktur yang akan dideteksi (definisi, penegasan, dan lain-lain) dan jenis pengetahuan yang dikontribusikan oleh masing-masing (konsep, atribut, nilai, dan hubungan).

Data yang digunakan untuk membangun model ontologi dalam penelitian ini adalah data produk skincare dari beberapa produsen skincare lokal. Data ini diperoleh melalui pengumpulan data yang bersumber dari internet yaitu web resmi dari beberapa produsen skincare lokal, dan pengambilan data langsung dengan wawancara kepada seorang narasumber yang dirasa memiliki pengetahuan mengenai skincare.

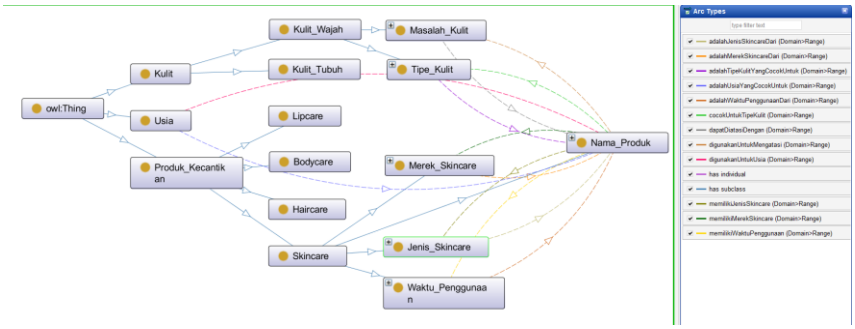
c. Konseptualisasi
Setelah model konseptual dibangun, maka metodologi akan mengubah model konseptual menjadi model formal, yang dimana model formal ini akan diimplementasikan dengan bahasa implementasi ontologi.

d. Integrasi

p-ISSN: 2301-5373
e-ISSN: 2654-5101

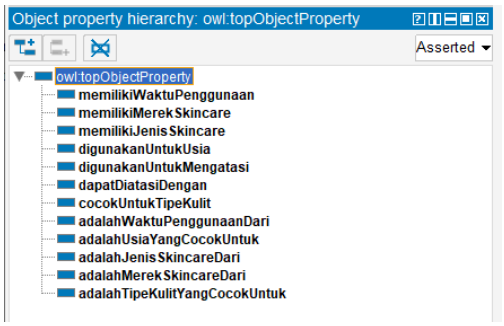
Pembangunan ontologi pada domain produk skincare pertama kali dilakukan dan belum ditemukan model ontologi pada domain yang berkaitan dengan produk skincare, maka pada penelitian ini tahap integrasi ontologi tidak dilakukan.

e. Implementasi
Model ontologi diimplementasikan dengan menggunakan perangkat lunak Protégé 5.5.0. Pada perangkat lunak Protégé 5.5.0, setiap bagian ontologi didefinisikan sesuai dengan hasil dari tiap tahapan tugas pada Methontologi.
Gambar 4 merupakan implementasi ontologi ini terdapat 7 (tujuh) class utama yaitu, Nama Produk, Jenis Skincare, Merek Skincare, Tipe Kulit, Waktu penggunaan, Masalah Kulit, dan Usia yang ditunjukkan melalui ontograf.



Gambar 4. Diagram Ontograf Produk Skincare

Gambar 5 adalah *object properties* yang terdapat pada ontologi produk skincare. Jumlah dari object properties yang digunakan adalah 12 *object properties*.



Gambar 5. Object Properties Ontologi Produk Skincare

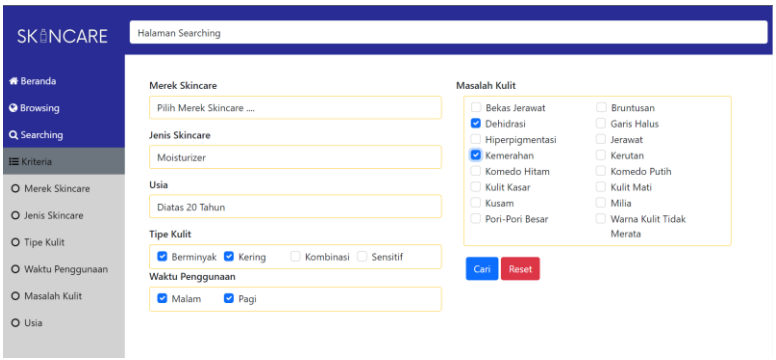
f. Evaluasi
Pada tahap evaluasi dilanjutkan dengan proses *reasoners* untuk melihat konsistensi ontologi yang dibangun. Proses *reasoners* dilakukan menggunakan ekstensi HermiT yang tedapat pada *tools* Protégé 5.5.0.

g. Dokumentasi
Dari ontologi produk skincare yang telah dibangun, dihasilkan *metric* ontologi yang tersusun untuk memberikan gambaran secara matematis komponen yang ada dalam rancangan ontologi tersebut. *Metrics* ontologi produk skincare yang telah dibangun menghasilkan 1358 axiom.

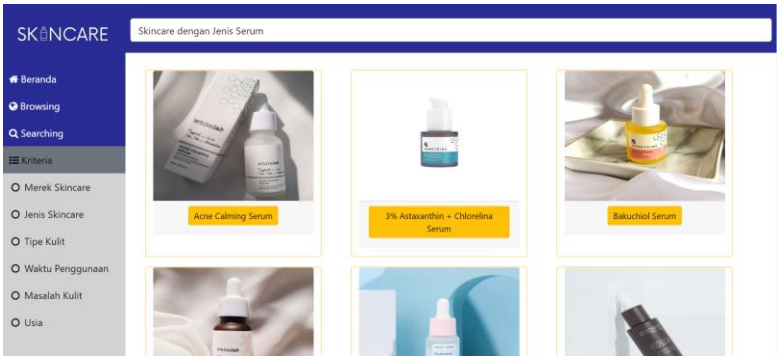
3.2. Implementasi Sistem

Pada tahap implementasi ontologi ke dalam sistem dilakukan dengan mengunggah ontologi produk skincare ke server Apache Jena Fuseki. Server Fuseki akan digunakan untuk mengelola dan menyimpan data ontologi yang bertujuan agar ontologi nantinya dapat digunakan oleh sistem. Ontologi yang diunggah merupakan sebuah file dengan format owl. Apabila proses pengunggahan telah selesai, maka dilanjutkan dengan proses koneksi ontologi dengan sistem menggunakan *library* EasyRDF. Pada implementasi antarmuka, rancangan antarmuka sebelumnya akan diimplementasikan menggunakan bahasa pemrograman HTML dengan *framework* Laravel 8.83.1 dan CSS dengan *framework* Bootstrap 5.1.3.

Gambar 7 adalah hasil implementasi antarmuka halaman pencarian (*searching*) sistem. Pada halaman ini, *guest user* dapat melakukan pencarian produk skincare dengan cara mengisi *form input* sesuai dengan filter kriteria yang diinginkan, lalu mengklik tombol "Cari". Hasil pencarian produk skincare akan ditampilkan beserta *query* SPARQL yang digunakan untuk melakukan pencarian. *Guest user* kemudian dapat mengakses tautan hasil yang diinginkan.
Gambar 8 adalah implementasi antarmuka dari halaman penjelajahan sistem setelah memilih salah satu tautan kriteria pada halaman sebelumnya. Pada halaman ini akan terdapat beberapa daftar produk skincare dari kriteria yang telah dipilih *guest user*.



Gambar 7. Implementasi Antarmuka Halaman Pencarian



Gambar 8. Implementasi Antarmuka Halaman Penjelajahan

3.3. Implementasi Pengujian Sistem

Hasil pengujian *Black-Box Testing* dari fitur pencarian (*searching)* dan fitur penjelajahan (*browsing)* ditunjukkan oleh Tabel 2 dan Tabel 3. Berdasarkan hasil pengujian *Black-Box Testing* tersebut dapat dilihat bahwa hasil yang diberikan oleh sistem telah sesuai dan dapat dikatakan sistem telah memiliki fungsionalitas yang baik.

Tabel 2. Hasil Pengujian Black-Box Testing Fitur Penjelajahan

Kasus: Penjelajahan *Guest User*

No.	Nama Skenario	Hasil yang Diharapkan	Hasil Pengujian	Kesimpulan
1.	Menampilkan halaman penjelajahan	Sistem menampilkan halaman *Browsing*	Sesuai harapan	Valid
2.	Melakukan pemilihan *hyperlink* pada fitur penjelajahan	a. Sistem menampilkan *hyperlink* kriteria b. Sistem berhasil melakukan *browsing* saat *guest user* melakukan pemilihan *hyperlink*	Sesuai harapan	
3.	Menampilkan hasil penjelajahan	Sistem menampilkan hasil penjelajahan sesuai hyperlink kriteria yang dipilih	Sesuai harapan	

p-ISSN: 2301-5373
e-ISSN: 2654-5101

	dengan benar			

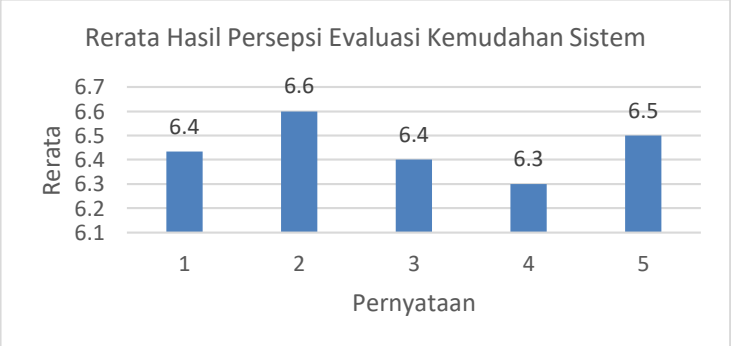
Tabel 3. Hasil Pengujian Black-Box Testing Fitur Pencarian

Kasus: Pencarian *Guest User*				
No.	Nama Skenario	Hasil yang Diharapkan	Hasil Pengujian	Kesimpulan
1.	Menampilkan halaman pencarian	Sistem menampilkan halaman *Searching*	Sesuai harapan	Valid
2.	Melakukan pemilihan filter kriteria pada fitur pencarian	a. Sistem menampilkan *form input* untuk *searching* b. Sistem berhasil melakukan *searching* saat *guest user* melakukan filter kriteria	Sesuai harapan	
3.	Menampilkan hasil pencarian dengan benar	Sistem menampilkan hasil pencarian dan *query* yang sesuai dengan filter kriteria	Sesuai harapan	

3.4. Implementasi Evaluasi Sistem

Jumlah peserta evaluasi sistem yang optimal yaitu berjumlah minimal 20 orang [8]. Pada penelitian ini, peneliti merekrut peserta yang berjumlah 30 orang. Pada tahap evaluasi sistem dilakukan pengolahan data terhadap penilaian sistem yang diberikan oleh peserta pengujian. Pengolahan data dilakukan dalam 2 aspek yaitu evaluasi kemudahan sistem dan evaluasi kegunaan sistem sesuai dengan konsep dari *Technology Acceptance Model* (TAM).

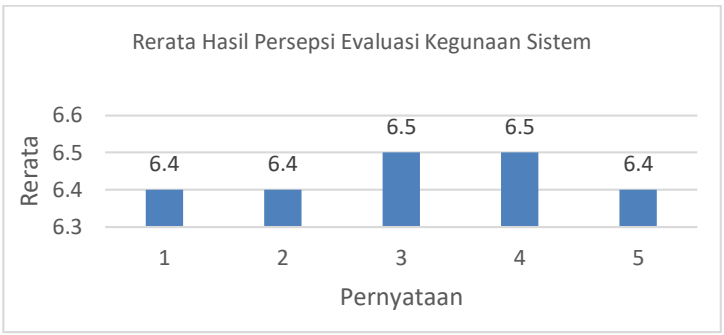
a. Evaluasi Kemudahan Sistem
Berdasarkan pengolahan data evaluasi kemudahan, diketahui bahwa nilai rata-rata (mean yang dimiliki peserta evaluasi pada keseluruhan pernyataan evaluasi kemudahan adalah 6.44 (dibulatkan 6) dan dikonversikan ke dalam skala Likert menjadi “Setuju”. Rerata tersebut menggambarkan skala jawaban yang diberikan peserta mengenai kemudahan penggunaan sistem adalah setuju.
Berdasarkan hasil analisis statistik tersebut, adapun grafik dari rerata yang dimiliki peserta dalam melakukan evaluasi kemudahan sistem pada masing-masing pernyataan dapat dilihat pada Gambar 10.



Gambar 10. Grafik Batang Rerata Hasil Evaluasi Persepsi Kemudahan Sistem

b. Evaluasi Kegunaan Sistem
Berdasarkan dengan pengolahan data evaluasi kegunaan, diketahui bahwa nilai rata-rata (mean) yang dimiliki peserta evaluasi pada keseluruhan pernyataan evaluasi kegunaan adalah 6.44 (dibulatkan 6) dan dikonversikan ke dalam skala Likert menjadi “Setuju”. Rerata tersebut menggambarkan skala jawaban yang diberikan peserta mengenai kegunaan penggunaan sistem adalah setuju.

Berdasarkan hasil analisis statistik tersebut, adapun grafik dari rerata yang dimiliki peserta dalam melakukan evaluasi kegunaan sistem pada masing-masing pernyataan dapat dilihat pada Gambar 11.



Gambar 11. Grafik Batang Rerata Hasil Evaluasi Persepsi Kegunaan Sistem

4. Kesimpulan

Adapun kesimpulan yang dapat ditarik dari penelitian yang telah dilakukan, adalah sebagai berikut.

a. Pembangunan ontologi produk skincare telah dibangun dengan mengimplementasikan metode Methontologi, yang memiliki beberapa tahapan, antara lain spesifikasi, akuisisi pengetahuan, konseptualisasi, integrasi, implementasi, evaluasi, dan dokumentasi. Pada pembangunan ontologi produk skincare, telah dihasilkan model ontologi yang terbentuk dari 7 class utama, 12 object properties, dan 88 instances.
b. Pembangunan sistem pencarian produk skincare berbasis ontologi telah dibangun dengan mengimplementasikan metode Prototyping yang memiliki beberapa tahapan, antara lain pengumpulan kebutuhan, perancangan prototype, evaluasi prototype, pembangunan sistem, serta pengujian dan evaluasi sistem. Pembangunan prototype sistem memiliki 2 buah fitur, yaitu fitur pencarian (searching) dan fitur penjelajahan (browsing).
c. Pada evaluasi kemudahan dan evaluasi kegunaan sistem diperoleh nilai rerata sebesar 6.44 untuk kedua evaluasi, dimana apabila dikonversi ke dalam skala Likert akan menghasilkan persepsi “Setuju”. Hasil rerata tersebut menunjukkan bahwa rata-rata peserta pengujian setuju bahwa sistem pencarian produk skincare yang dibangun adalah sistem yang berguna dan mudah untuk digunakan.

Referensi

[1] V. Maarif, H. M. Nur, dan T. A. Septianisa, “Sistem Pendukung Keputusan Pemilihan Skincare Yang Sesuai Dengan Jenis Kulit Wajah Menggunakan Logika Fuzzy,” *EVOLUSI J. Sains dan Manaj.*, vol. 7, no. 2, hal. 73–80, 2019.

[2] ZAP Clinic Index dan MarkpPlus, “ZAP Beauty Index 2020,” *Zap Clinic Index*, hal. 1–36, 2020.

[3] S. Day, “Limone”, 30 Juli 2020. [Online]. Available: https://www.limone.id/urutan-pemakaian-skincare/. [1 April 2021]

[4] Himawan, T. W. Harjanti, R. Supriati, dan H. Setiyani, “Evolusi Penggunaan Teknologi Web 3.0 : Semantic Web,” *J. Inf. Syst. Hosp. Technol.*, vol. 2, no. 02, hal. 54–60, 2020.

[5] K. D. P. Novianti dan R. A. N. Diaz, “Sistem Pencarian Program Studi Pada Perguruan Tinggi Di Bali Berbasis Semantik,” *JST (Jurnal Sains dan Teknol.*, vol. 6, no. 1, hal. 93–104, 2017.

[6] D. Purnomo, “Model Prototyping Pada Pengembangan Sistem Informasi,” *J I M P - J. Inform. Merdeka Pasuruan*, vol. 2, no. 2, hal. 54–61, 2017.

[7] C. R. A. Pramartha, “Assembly the Semantic Cultural Heritage Knowledge,” *J. Ilmu Komput.*, vol. 11, no. 2, hal. 83, 2018.

[8] T. A. Ghaffur dan Nurkhamid, “Analisis Kualitas Sistem Informasi Kegiatan Sekolah Berbasis Mobile Web Di Smk Negeri 2 Yogyakarta,” *Elinvo (Electronics, Informatics, Vocat. Educ.*, vol. 2, no. 1, hal. 94–101, 2017.

Jurnal Elektronik Ilmu Komputer Udayana
Volume 11, No 3. February 2023

p-ISSN: 2301-5373
e-ISSN: 2654-5101

Pengaruh Metode Reduced Error Pruning pada Algoritma C4.5 untuk Prediksi Penyakit Diabetes

Luh Putu Eka Nadya Wati[a1], Ida Bagus Made Mahendra[a2], Ngurah Agus Sanjaya ER [a3], I Gusti Ngurah Anom Cahyadi Putra[a4], Agus Muliantara [a5], Luh Arida Ayu Rahning Putri [a6]

[a]Program Studi Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam,
Universitas Udayana
Bali, Indonesia
[1]ekanadya96@gmail.com
[2]ibm.mahendra@unud.ac.id
[3]agus_sanjaya@unud.ac.id
[4]anom.cp@unud.ac.id
[5]muliantara@unud.ac.id
[6]rahningputri@unud.ac.id

Abstract

Degenerative disease is one of the conditions that can cause the performance of several organs in the human body to decrease and affect health conditions. The prevalence rate of diabetes is predicted to continue to increase in 2030 to reach 578 million and 700 million in 2045. In this research, a diabetes prediction system was formed using the C4.5 Algorithm with Reduced Error Pruning (REP). This research is focused on the application of the Reduced Error Pruning method on the C4.5 Algorithm and used two datasets containing several medical predictors of diabetes symptoms. Based on the research that has been done, the prediction process using the C4.5 Algorithm with Reduced Error Pruning based on the first dataset resulted in an average accuracy of 92,4% with an average accuracy before Reduced Error Pruning of 91,6%. In comparison, in the second dataset, average accuracy was obtained without Reduced Error Pruning by 81,2% and 83,4% for results with Reduced Error Pruning. Based on this percentage, the Reduced Error Pruning method does not have a big influence on the level of accuracy produced.

Keywords: *Diabetes, Data Mining, Sistem Prediksi, Algoritma C4.5, Reduced Error Pruning*

1. Pendahuluan

Penyakit degeneratif merupakan salah satu kondisi yang dapat menyebabkan kinerja dari beberapa organ dalam tubuh manusia mengalami penurunan sehingga dapat mempengaruhi kondisi kesehatan. Penyakit degeneratif dapat diderita oleh semua orang dan merupakan masalah mendesak dan kontroversial bagi beberapa negara termasuk negara Indonesia. Tanpa disadari, jutaan orang memiliki kebiasaan yang tidak sehat sehingga memicu terjadinya berbagai masalah kesehatan [1]. Diabetes merupakan salah satu dari sekian banyaknya penyakit degeneratif. Banyak masyarakat mengidap penyakit ini. Berdasarkan hasil statistik dari *International Diabetes Federation* (IDF) pada tahun 2019 menyatakan bahwa Indonesia masuk ke dalam daftar negara dengan jumlah pengidap diabetes terbanyak hingga mencapai 10,7 juta kasus dan menduduki urutan ke-7 dari 10 negara. Seiring dengan pertambahan jumlah penduduk diperkirakan tingkat prevalensi penyakit diabetes akan meningkat di kalangan masyarakat pada rentang usia 65-79 tahun sebesar 19,9% atau 111,2 juta orang. Peningkatan prevalensi penyakit diabetes diprediksi akan terus bertambah pada tahun 2030 mencapai 578 juta hingga 700 juta di tahun 2045.

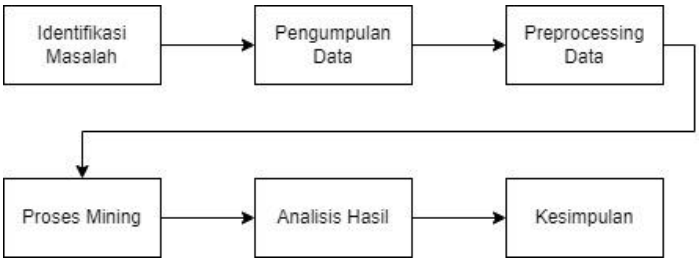
Adanya peningkatan kadar glukosa darah dalam tubuh merupakan salah satu tanda dari penyakit diabetes. Terdapat beberapa komplikasi yang bisa terjadi apabila penyakit ini tidak ditangani dengan serius. Komplikasi tersebut dapat menyebabkan adanya penurunan fungsi penglihatan yang dapat menyebabkan kebutaan, adanya penurunan fungsi ginjal, kerusakan pada beberapa saraf hingga dapat berujung pada kematian [2]. Sebanyak 25% pasien yang baru menyadari mengidap penyakit diabetes sudah mengalami komplikasi seperti neuropati sebanyak 9%, retinopati, dan nefropati sebanyak 8% berdasarkan data *United Kingdom Prospective Diabetes Study*. Pasien diabetes yang mengalami beberapa komplikasi diperkirakan telah mengidap penyakit diabetes 4-7 tahun

sebelumnya [3]. Penelitian ini akan membahas mengenai pembangunan sistem prediksi penyakit diabetes menggunakan Algoritma C4.5 dengan metode *Reduced Error Pruning*. Terdapat beberapa tahapan dalam penerapan Algoritma C4.5. Tahapan tersebut diantaranya adanya pemilihan atau pencarian atribut yang berperan sebagai *root*, pembentukan cabang untuk setiap nilai pada pohon keputusan yang dapat mendukung proses prediksi. Penerapan Algoritma C4.5 akan menghasilkan beberapa *rule* sebagai solusi permasalahan yang ingin diatasi dan mendukung proses prediksi [4]. *Reduced Error Pruning* (REP) berguna untuk meningkatkan akurasi dan menyederhanakan struktur *decision tree*. Apabila struktur *tree* yang dihasilkan sangat kompleks, hal ini dapat menyebabkan *rule* yang dihasilkan susah untuk diimplementasikan [5]. Dengan menerapkan *Reduced Error Pruning* pada Algoritma C4.5, maka akan diperoleh pengaruh REP pada tingkat akurasi sistem prediksi yang dihasilkan apabila dibandingkan dengan penerapan Algoritma C4.5 tanpa metode REP serta dapat melihat pengaruh metode REP dalam pembentukan pohon keputusan.

2. Metode Penelitian

2.1 Kerangka Penelitian

Kerangka penelitian berguna untuk mengidentifikasi tahap-tahap yang perlu dilakukan sehingga dapat memenuhi tujuan penelitian. Gambar 1 berikut ini merupakan diagram yang mencangkup beberapa tahapan kegiatan pada penelitian.



Gambar 1. Alur Penelitian

Pada kerangka penelitian diatas menjelaskan tentang tahap-tahap yang dilakukan pada penelitian ini. Identifikasi masalah merupakan tahapan awal yang dilakukan. Topik permasalahan yang diangkat pada penelitian ini yaitu perancangan sistem prediksi penyakit diabetes menggunakan Algoritma C4.5 dengan metode *Reduced Error Pruning*. Pada tahap pengumpulan data, penulis menggunakan data sekunder yang mengandung beberapa prediktor medis penyakit diabetes yang menjadi tolak ukur proses prediksi. Tahapan preprocessing data berguna untuk melihat kualitas data sebelum diproses. Tahap ini bertujuan untuk mengatasi adanya *missing value* yang terkandung pada data serta menyeleksi data-data yang ingin digunakan untuk tahap analisis sehingga dapat menghasilkan hasil yang optimal. Proses mining yang dilakukan menggunakan Algoritma C4.5 dengan *Reduced Error Pruning*. Penelitian ini membandingkan hasil proses mining Algoritma C4.5 tanpa *Reduced Error Pruning* dengan hasil proses mining Algoritma C4.5 yang ditambah dengan metode *Reduced Error Pruning*. Hasil tersebut akan dianalisis untuk mengetahui pengaruh metode *Reduced Error Pruning* terhadap Algoritma C4.5. Sehingga, pada penelitian ini dapat ditarik kesimpulan berdasarkan hasil pengujian yang dilakukan.

2.2 Penyakit Diabetes

Diabetes adalah suatu penyakit yang membuat penderitanya memiliki kadar glukosa yang sangat tinggi di dalam tubuhnya. Kurangnya produksi insulin dalam tubuh dapat menyebabkan kadar gula darah meningkat. Jika peningkatan kadar gula darah yang tinggi tidak ditangani dengan cepat maka akan berdampak pada munculnya masalah kesehatan yang lebih serius. Tingkat glukosa yang tinggi disebabkan oleh adanya kerusakan pada sel beta organ pankreas yang merupakan sumber penghasil insulin. Hal tersebut memicu ketidakefektifan hormon insulin dalam mengatur keseimbangan tingkat gula darah pada seseorang [6]. Tingkat glukosa darah yang bernilai 200 mg/dl atau lebih merupakan tanda bahwa glukosa darah tersebut tinggi. Penyakit ini dapat menyerang seluruh sistem pada tubuh seperti kulit, jantung dan hingga menimbulkan komplikasi serius. Terjadinya penyakit diabetes pada seseorang dapat dipengaruhi karena adanya resistensi insulin. Seseorang yang memiliki berat badan berlebih sangat rentan mengalami resistensi insulin. Resistensi insulin merupakan suatu keadaan yang membuat sel tubuh tidak bisa memanfaatkan hormon insulin. Insulin membantu sel tubuh menggunakan gula glukosa untuk energi. Gula darah akan menumpuk

apabila terjadi resisten terhadap insulin. Kadar gula darah akan meningkat sehingga dapat memicu hiperglikemia kronik jika produksi insulin tidak cukup kuat untuk mengkompensasi resistensi insulin. Penyebab peningkatan prevalensi diabetes adalah pertambahan populasi dan perilaku gaya hidup yang tidak sehat. Seseorang penderita penyakit ini dapat memiliki gejala antara lain adanya berkurangnya berat badan yang cukup drastis, sering merasa haus, dan memiliki rasa lapar berlebih. Selain itu, keluhan seperti badan lemah dan kurangnya energi, kesemutan di tangan atau kaki, gatal, penyembuhan luka yang lama, dan mata kabur merupakan gejala lain dari diabetes. Perubahan pola hidup menuju yang lebih sehat merupakan bentuk pencegahan dari ancaman penyakit diabetes.

2.3 Pengumpulan Dataset

Penelitian ini menggunakan dataset diabetes yang mengandung beberapa prediktor dengan dua jenis label atau kelas diabetes. Dataset yang digunakan terdiri dari dua dataset yang berbeda. Pengambilan dua dataset ini bermaksud untuk memperluas atau menambah beberapa prediktor yang menjadi tolak ukur seseorang dengan penyakit diabetes sehingga dapat meningkatkan kualitas prediksi. Dataset pertama merupakan data hasil kuesioner dari pasien rumah sakit di Sylhet, Bangladesh yang didapat dari web kaggle.com. Sedangkan untuk dataset kedua merupakan data yang didapat dari web datahub.io yang merupakan data dari Institut Nasional Diabetes dan Penyakit Pencernaan dan Ginjal. Total data prediktor medis yang digunakan terdiri dari 1.285 record data. Adapun prediktor yang akan digunakan berdasarkan dataset pertama yaitu gender, poliuria, polidipsia, penurunan berat badan secara tiba-tiba, cepat lelah, polifagia, pandangan kabur, penyembuhan tertunda, kegemukan, dan kelas. Prediktor yang digunakan pada dataset kedua yaitu kadar glukosa plasma, tekanan darah, ketebalan kulit trisep, kadar serum insulin, indeks massa tubuh, riwayat silsilah diabetes, umur serta kelas. Dataset yang digunakan termasuk ke dalam jenis data sekunder. Hal ini karena data yang digunakan diperoleh dari beberapa sumber yang tersedia di internet dan dapat diakses banyak orang. Terdapat dua kelas diabetes pada dataset yaitu positif untuk penderita penyakit diabetes dan negatif bagi yang tidak menderita penyakit diabetes.

2.4 Data Preprocessing

Data preprocessing merupakan langkah penting dalam proses data mining. Tahap ini berguna untuk meningkatkan kualitas data sebelum masuk ke tahap analisis penerapan metode [7]. Pada tahap ini dilakukan pembersihan data dengan melakukan pengisian nilai terhadap atribut yang memuat *missing value*. Data yang memiliki *missing value* akan dibersihkan dengan memberikan nilai yang frekuensinya dominan berdasarkan atribut tersebut yang nantinya akan diisi pada setiap masing-masing baris atribut yang mengandung *missing value* dengan mengacu pada atributnya. Pada dataset kedua ditemukan beberapa data *missing value* pada 5 atribut yaitu lipatan kulit trisep, insulin, glukosa plasma, tekanan darah diastolik, dan indeks massa tubuh. Pada dataset tertera bahwa terdapat nilai 0 untuk 5 atribut tersebut. Nilai 0 dianggap kurang relevan karena pada umumnya tidak mungkin seseorang memiliki indeks massa tubuh dengan nilai 0, sehingga nilai 0 pada beberapa atribut dianggap sebagai *missing value*. Data yang telah diproses kemudian akan digunakan untuk training dan testing dalam proses prediksi.

2.5 Implementasi Metode

Sistem yang telah dirancang akan diterjemahkan ke dalam bentuk kode pemrograman komputer. PHP merupakan bahasa pemrograman scripting yang digunakan pada penelitian ini untuk mengembangkan sistem prediksi dalam bentuk web. Selain itu terdapat penggunaan HTML, CSS, *framework bootstrap* untuk pembuatan tampilan website serta Javascript untuk beberapa proses validasi yang ada pada sistem. Pada penelitian ini juga memanfaatkan MySQL dalam proses pembangunan database.

2.6 Decision Tree

Decision tree adalah jenis diagram alur yang menunjukkan jalur yang jelas untuk menuju suatu keputusan dan sangat berguna untuk analisis data. Metode *decision tree* akan menghasilkan struktur pohon atau struktur hirarki keputusan yang merupakan representasi dari beberapa aturan (*rule*) sehingga mudah dipahami [8]. Klasifikasi dalam bentuk pohon keputusan dapat digunakan pada kumpulan data dengan domain medis [9]. *Decision tree* bekerja dengan mempartisi data secara rekursif berdasarkan nilai atribut input. Partisi data disebut cabang. Cabang *decision tree* akan mencakup semua parameter data yang digunakan. Selain itu, simpul daun pada *decision tree* merupakan representasi dari label suatu kelas.

2.7 Algoritma C4.5

Algoritma C4.5 digunakan dalam data mining sebagai teknik klasifikasi yang dapat digunakan untuk menghasilkan keputusan berdasarkan sampel data tertentu. Pohon keputusan yang terbentuk dapat mendukung proses pengambilan suatu keputusan. Algoritma ini memiliki beberapa kelebihan yaitu dapat menangani data numerik dan kategorik serta dapat mengolah dataset yang rumit dan besar [10]. Terdapat beberapa elemen yang harus dicari dalam pemodelan pohon keputusan menggunakan Algoritma C4.5, yaitu:

1. Entropy (S), Entropy merupakan parameter yang berguna untuk mengukur tingkat keragaman nilai atribut masing-masing terhadap suatu atribut keputusan.

   $$Entropy\ (S) = \Sigma_{i=n}^{n} - pi * log_2\ (pi) \quad (1)$$

   Keterangan:
   S = Jumlah sampel kasus (sampling)
   n = Banyaknya partisi untuk S
   pi = Perbandingan Si ke S
2. Gain (S, A) adalah nilai yang berguna sebagai dasar dalam pembuatan simpul atau akar dan cabang dari suatu pohon keputusan.

   $$Gain\ (S, A) = E(S) - \Sigma_{i=1}^{n} \frac{|Si|}{|S|} * E(Si) \quad (2)$$

   Keterangan:
   E = Entropy
   S = Jumlah sampel kasus (sampling)
   A = Atribut
   n = Banyaknya partisi untuk S
   | Si | = Banyaknya kasus pada i-partisi
   | S | = Banyaknya kasus di S

2.8 Reduced Error Pruning

Reduced Error Pruning (REP) merupakan salah satu bentuk pemangkasan dengan teknik postpruning. Postpruning yaitu proses untuk meminimalkan pohon keputusan yang terbentuk dengan cara memangkas beberapa cabang pada pohon keputusan yang sebelumnya sudah selesai dibangun. Metode REP ini merupakan metode yang bisa digunakan pada Algoritma C4.5 [5]. Cara kerja REP adalah mempertimbangkan setiap simpul atau cabang pohon untuk pemangkasan. Pemangkasan merupakan tahap menghapus subpohon dan menetapkan kelas yang paling dominan muncul di simpul itu. Sebuah simpul akan dihapus jika akurasi pohon yang dihasilkan tidak lebih buruk daripada hasil saat training serta dapat meminimalkan nilai estimasi error. Node atau cabang akan dihapus secara iteratif jika dapat meningkatkan akurasi pohon keputusan. Selain itu, proses pemangkasan dilakukan dari bagian bawah menuju bagian atas pohon. Nilai estimasi error merupakan penentu proses *pruning* perlu dilakukan atau tidak. Selama nilai estimasi error pada proses *pruning* lebih kecil atau sama dengan nilai estimasi error sebelumnya maka proses *pruning* dilakukan [11]. Proses tersebut terus dilakukan selama ditemukan nilai estimasi error yang lebih baru dan tidak mengurangi nilai akurasi yang dihasilkan dibandingkan nilai error lama. Berkaitan dengan hal ini maka terdapat proses update nilai estimasi error yang terbaru [12]. Perhitungan untuk mencari nilai estimasi error dapat dilihat sebagai berikut.

$$e = \frac{f + \frac{z^2}{2N} + z\sqrt{\frac{f}{N} + \frac{f^2}{N} + \frac{z^2}{4N^2}}}{1 + \frac{z^2}{N}} \quad (3)$$

Keterangan:

e = Estimasi Error
f = Jumlah data yang salah terklasifikasi yang dibagi dengan jumlah data sampel.
N = Total jumlah data sampel yang digunakan.
z = Nilai konstanta yang ditetapkan Algoritma C4.5 dengan nilai 0,69.

p-ISSN: 2301-5373
e-ISSN: 2654-5101

2.9 Desain Evaluasi Metode

Pengujian pada penelitian ini dilakukan untuk mengetahui tingkat akurasi serta nilai estimasi error berdasarkan proses prediksi. Pada umumnya untuk mengevaluasi kinerja algoritma klasifikasi menggunakan skenario pengujian *K-Fold Cross Validation*. Pengujian menggunakan *K-Fold Cross Validation* berguna untuk mengetahui hasil rata-rata akurasi model sebanyak k pengujian dengan memprediksi beberapa kumpulan data yang diinput secara acak. Pada *K-Fold Cross Validation* terdapat pembagian dataset dengan rasio yang sama sebanyak nilai k [13]. Masing-masing dari *k-fold* diberikan kesempatan untuk digunakan sebagai data testing, sementara semua *fold* lainnya secara kolektif digunakan sebagai data training. Pada penelitian ini menggunakan jumlah *fold* sebanyak 5 untuk menguji validitas dalam model. Penelitian ini menggunakan perbandingan data sebesar 80% data untuk proses training dan 20% data digunakan pada proses testing. Hal ini juga dapat memastikan bahwa sistem akan menghasilkan keluaran sesuai dengan yang diinginkan berdasarkan masukan tertentu.

Setelah dilakukannya penerapan Algoritma C4.5 dengan *Reduce Error Pruning* (REP), maka selanjutnya dilakukan proses untuk menghitung nilai akurasi berdasarkan pada data yang berhasil diprediksi secara benar, yaitu dengan menjumlahkan data yang diprediksi dengan benar oleh sistem yang kemudian dibagi dengan jumlah total data yang digunakan dan dikalikan dengan 100%. Berikut ini merupakan rumus perhitungan akurasi dari sistem prediksi pada penelitian ini [14], yaitu:

$$\text{Akurasi} = \frac{\sum klasifikasi\ benar}{\sum data\ uji} \times 100\% \quad (4)$$

Keterangan :

Akurasi = hasil tingkat akurasi
klasifikasi benar = banyaknya data yang berhasil prediksi secara benar
data uji = total keseluruhan data uji yang digunakan

3. Hasil dan Pembahasan

Berdasarkan pada keseluruhan alur kerangka penelitian yang telah dilakukan. Maka untuk hasil penelitian dapat dilihat pada bagian berikut ini.

3.1 Pengujian Akurasi

Percobaan dilakukan beberapa kali dengan skenario *K-Fold Cross Validation* dengan menentukan nilai k terlebih dahulu sebanyak 5. Pada penelitian ini menggunakan dua dataset sehingga terdapat dua tabel hasil akurasi menggunakan *Reduced Error Pruning* berdasarkan dataset. Jumlah rasio terkait data yang berguna untuk proses training dan data uji pada setiap percobaan yaitu sebesar 80% untuk data training dan 20% berguna sebagai data uji. Berdasarkan persentase tersebut maka untuk dataset pertama jumlah data training yang digunakan sebanyak 416 dan 104 sebagai data uji. Sedangkan untuk dataset kedua menggunakan data sebanyak 612 untuk data training dan 153 digunakan sebagai data uji.

3.1.1 Pengujian Dataset Pertama

Hasil pengujian yang sudah dilakukan pada dataset pertama dapat diamati pada Tabel 1 dan Tabel 2 berikut ini.

Tabel 1. Pengujian Dataset 1 tanpa *Reduced Error Pruning*

Pengujian ke-	Banyaknya Prediksi Benar	Banyaknya Prediksi Salah	Akurasi	Estimasi Error
1	92	12	88 %	12 %
2	97	7	93 %	7 %
3	96	8	92%	8 %

4	95	9	91 %	9 %
5	98	6	94 %	6 %
	Rata-rata		91,6%	8,4%

Tabel 2. Pengujian Dataset 1 dengan *Reduced Error Pruning*

Pengujian ke-	Banyaknya Prediksi Benar	Banyaknya Prediksi Salah	Akurasi	Estimasi Error
1	93	11	89 %	11 %
2	98	6	94 %	6 %
3	96	8	92%	8 %
4	95	9	91 %	9 %
5	100	4	96 %	4 %
	Rata-rata		92,4%	7,6%

Tabel 1 menunjukkan hasil dari pengujian sistem prediksi menggunakan Algoritma C4.5 tanpa metode *Reduced Error Pruning* berdasarkan prediktor medis pada dataset pertama. Pada pengujian tersebut yang terlihat pada Tabel 3 menghasilkan rata-rata akurasi sebesar 91,6 %. Berdasarkan pada Tabel 2 menunjukkan hasil dari pengujian sistem prediksi pada dataset pertama menggunakan Algoritma C4.5 dengan *Reduced Error Pruning*. Hasil rata-rata akurasi secara keseluruhan pada skenario pengujian menggunakan *K-Fold Cross Validation* dengan *fold* sebanyak 5 sebesar 92,4 %. Mengacu pada hasil tersebut maka perolehan nilai akurasi dengan penerapan *Reduced Error Pruning* mengalami peningkatan yang tidak banyak yaitu sebesar 0,8%.

3.1.2 Pengujian Dataset Kedua

Pada proses implementasi Algoritma C4.5 untuk dataset kedua dilakukan pemisahan nilai ambang batas dikarenakan seluruh atribut yang terkandung pada dataset kedua bersifat numerik. Pemisahan nilai numerik dilakukan dengan mencoba setiap nilai sebagai titik ambang dan menghitung perolehan akurasi untuk setiap nilai ambang batas [15]. Hal ini berguna untuk melihat nilai ambang batas yang dapat menghasilkan model prediksi terbaik. Berikut ini Tabel 3 merupakan penentuan nilai ambang batas yang digunakan untuk penanganan data numerik berdasarkan pada dataset kedua.

Tabel 3. Penentuan Nilai Ambang Batas

No.	Nama Field	Nilai Ambang Batas
1.	Kadar Glukosa Darah	<= 140 dan > 140
		<= 126 dan > 126
		<= 99 dan > 99
		<= 70 dan > 70
2.	Tekanan Darah	<= 110 dan > 110
		<= 90 dan > 90
		<= 80 dan > 80
3.	Ketebalan Lipatan Kulit Trisep	<= 35 dan > 35
4.	Kadar Serum Insulin	<= 750 dan > 750
		<= 450 dan > 450
		<= 150 dan > 150
5.	Indeks Massa Tubuh	<= 30,1 dan > 30,1
		<= 23,4 dan > 23,4
		<= 18,5 dan > 18,5

6.	Riwayat Silsilah Diabetes	<= 0,56 dan > 0,56 <= 0,39 dan > 0,39
7.	Umur	<= 61 dan > 61 <= 45 dan > 45 <= 28 dan > 28 <= 25 dan > 25

Berikut ini pada Tabel 4 dan Tabel 5 merupakan hasil pengujian berdasarkan nilai ambang batas yang digunakan untuk menangani data numerik pada dataset kedua, seperti yang tertera pada Tabel 3.

Tabel 4. Pengujian Kedua Dataset 2 tanpa *Reduced Error Pruning*

Pengujian tahap-	Banyaknya Prediksi Benar	Banyaknya Prediksi Salah	Akurasi	Estimasi Error
1	123	30	80 %	20 %
2	129	24	84 %	16 %
3	127	26	83 %	17 %
4	120	33	78 %	22 %
5	124	29	81 %	19 %
Rata-rata			81,2%	18,8%

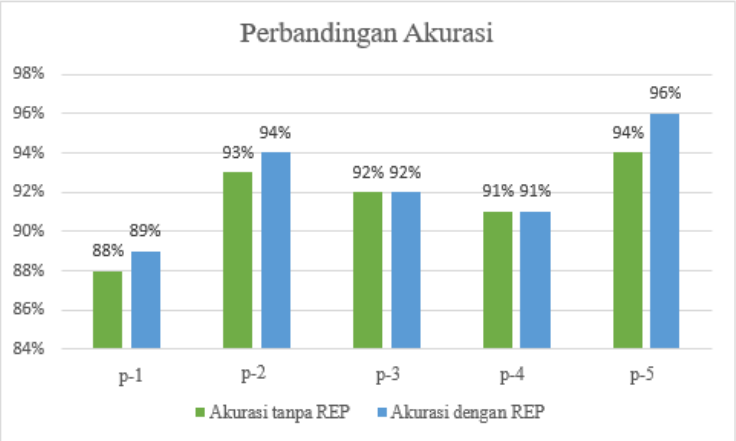
Tabel 5. Pengujian Kedua Dataset 2 dengan *Reduced Error Pruning*

Pengujian tahap-	Banyaknya Prediksi Benar	Banyaknya Prediksi Salah	Akurasi	Estimasi Error
1	125	28	82 %	18 %
2	131	22	86 %	14 %
3	130	23	85 %	15 %
4	124	29	81 %	19 %
5	127	26	83 %	17 %
Rata-rata			83,4%	16,6%

Tabel 5 berikut ini merupakan tabel hasil uji Algoritma C4.5 dengan *Reduced Error Pruning* pada dataset kedua. Dari hasil tersebut dapat dilihat bahwa persentase akurasi yang dihasilkan cukup baik jika dibandingkan dengan hasil pada Tabel 4. Berdasarkan hasil diatas maka didapatkan rata-rata akurasi dengan *Reduced Error Pruning* dari keseluruhan percobaan dengan beberapa nilai ambang batas pada dataset kedua sebesar 83,4 %. Nilai akurasi meningkat sebesar 2,2% menggunakan metode *Reduced Error Pruning*.

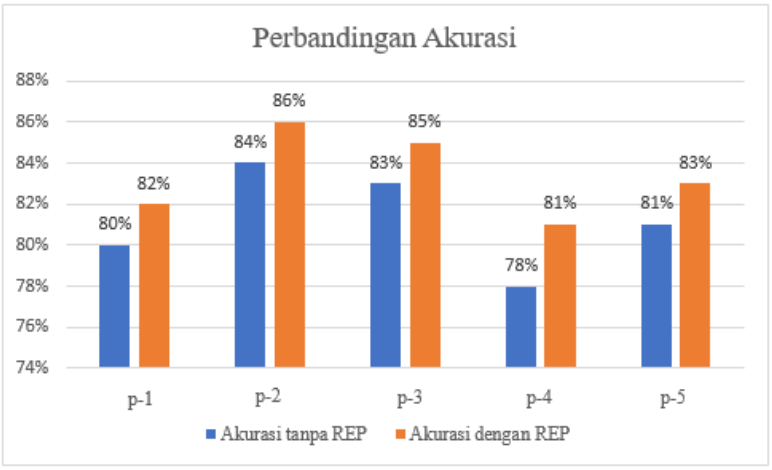
3.1 Pengaruh Metode Reduced Error Pruning

Pada penelitian ini, pengujian dilakukan dengan membandingkan hasil Algoritma C4.5 tanpa *Reduced Error Pruning* dan hasil dengan *Reduced Error Pruning* untuk mengetahui seberapa besar pengaruh penerapan *Reduced Error Pruning*. Proses eksperimen ini membandingkan akurasi serta jumlah *rule* yang dihasilkan. Pengaruh penerapan metode *Reduced Error Pruning* terhadap besar akurasi yang dihasilkan dapat dilihat pada grafik berikut yang tercantum pada Gambar 2 serta Gambar 3.



Gambar 2. Perbandingan Akurasi Dataset 1

Berdasarkan Gambar 2 tersebut menunjukkan bahwa metode *Reduced Error Pruning* tidak memberikan pengaruh terhadap peningkatan akurasi yang dihasilkan berdasarkan dataset pertama. Hal tersebut dapat dilihat dari hasil pengujian ke-3 dan ke-4 yang menghasilkan akurasi sama dengan hasil akurasi tanpa *Reduced Error Pruning*. Sementara itu, jika dilihat pada hasil pengujian ke-1, ke-2 dan ke-5 penerapan *Reduced Error Pruning* memiliki pengaruh terhadap akurasi yang dihasilkan hal tersebut dilihat dari adanya peningkatan nilai akurasi dari sebelumnya. Pada pengujian ke-1 dan ke-2 nilai akurasi naik sebesar 1% sedangkan pada pengujian ke-5 nilai akurasi naik sebesar 2% menggunakan *Reduced Error Pruning*.

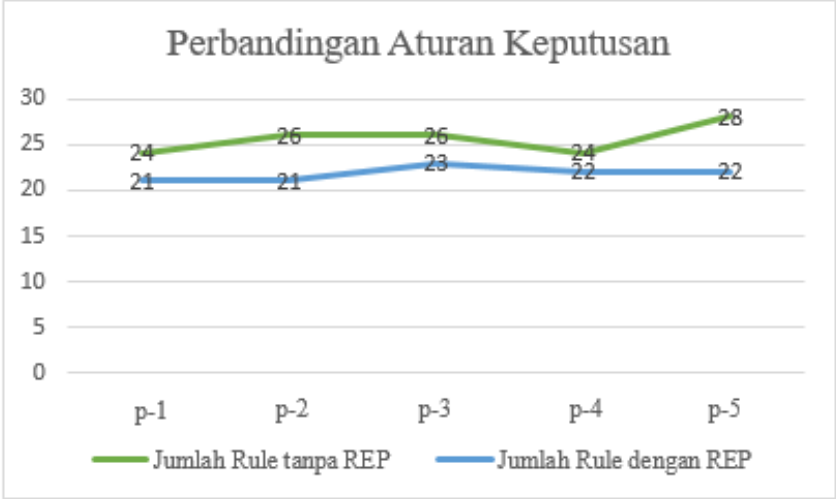


Gambar 3. Perbandingan Akurasi Dataset 2

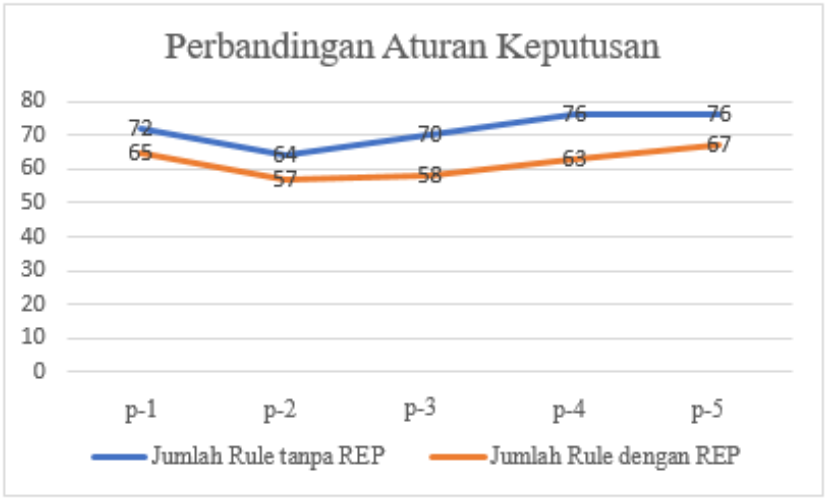
Berbeda pada hasil pengujian pada dataset pertama, perbandingan akurasi berdasarkan pada pengujian yang dilakukan pada dataset kedua menunjukkan bahwa metode *Reduced Error Pruning* dapat meningkatkan akurasi pada setiap percobaan yang dilakukan. Berdasarkan pada Gambar 3 terlihat bahwa peningkatan akurasi dengan metode *Reduced Error Pruning* pada setiap percobaan tidak menghasilkan nilai akurasi yang jauh berbeda dari sebelumnya jika dibandingkan dengan hasil tanpa *Reduced Error Pruning*. Adanya perbandingan akurasi antara percobaan pada dataset pertama dan kedua, maka dapat diketahui bahwa penerapan metode *Reduce Error Pruning* pada Algoritma C4.5 tidak selalu memberikan pengaruh terhadap nilai akurasi yang dihasilkan pada setiap percobaan.

Selain perbandingan hasil akurasi, penelitian ini juga membandingkan jumlah aturan keputusan yang berhasil dibentuk menggunakan Algoritma C4.5 tanpa *Reduced Error Pruning* maupun dengan *Reduced Error Pruning*.

p-ISSN: 2301-5373
e-ISSN: 2654-5101



Gambar 4. Perbandingan Aturan Keputusan Dataset 1



Gambar 5. Perbandingan Aturan Keputusan Dataset 2

Berdasarkan grafik yang tertera pada Gambar 4 dan Gambar 5, terlihat bahwa baik untuk dataset pertama maupun dataset kedua yang diuji dengan menggunakan metode *Reduced Error Pruning* menghasilkan aturan keputusan yang lebih sedikit jika dibandingkan dengan aturan keputusan yang dihasilkan tanpa adanya *Reduced Error Pruning*. Jika dilihat dari hasil percobaan tersebut maka dapat disebutkan bahwa jumlah *rule* yang berhasil terbentuk menggunakan Algoritma C4.5 dapat disederhanakan melalui penerapan metode *Reduced Error Pruning*. Hal ini dikarenakan jumlah aturan keputusan yang terbentuk mengalami penurunan sebesar 14,58% untuk pengujian dataset pertama dan 13,32% untuk dataset kedua berdasarkan rata-rata persentase penurunan jumlah aturan keputusan antara penerapan metode Algoritma C4.5 tanpa adanya *Reduced Error Pruning* dibandingkan dengan adanya *Reduced Error Pruning*.

4. Kesimpulan

Berdasarkan hasil uji coba proses prediksi penyakit diabetes menggunakan Algoritma C4.5 dengan *Reduced Error Pruning* berdasarkan pada dataset pertama dihasilkan rata-rata akurasi sebesar 92,4% dengan rata-rata akurasi sebelum *Reduced Error Pruning* sebesar 91,6% sedangkan pada dataset kedua dihasilkan rata-rata akurasi tanpa *Reduced Error Pruning* sebesar 81,2% dan 83,4% untuk hasil dengan *Reduced Error Pruning*. Tingkat akurasi menggunakan metode *Reduced Error Pruning* mengalami peningkatan sebesar 0,8% untuk dataset pertama dan 2,2% untuk dataset kedua. Pada penelitian ini metode *Reduced Error Pruning* memiliki pengaruh terhadap jumlah aturan keputusan yang terbentuk. Jumlah aturan keputusan yang dihasilkan dengan *Reduced Error Pruning* lebih sedikit dengan rata-rata penurunan sebesar 14,58% untuk pengujian dataset pertama dan

13,32% untuk dataset kedua dibandingkan dengan aturan keputusan yang terbentuk melalui Algoritma C4.5 tanpa *Reduced Error Pruning*.

References

[1] I. Istianah, S. Septiani, and G. K. Dewi, "Mengidentifikasi Faktor Gizi pada Pasien Diabetes Mellitus Tipe 2 di Kota Depok Tahun 2019," *J. Kesehat. Indones. (The Indones. J. Heal.*, vol. X, no. 2, pp. 72–78, 2020.

[2] I. Irma, L. O. Alifariki, and A. Kusnan, "Uji Sensitifitas dan Spesifisitas Keluhan Penderita Diabetes Melitus Berdasarkan Keluhan dan Hasil Pemeriksaan Gula Darah Sewaktu (GDS)," *J. Kedokt. dan Kesehat.*, vol. 16, no. 1, p. 25, 2020, doi: 10.24853/jkk.16.1.25-34.

[3] "Issn: 2089-9084 ism, vol. 6 no.1, mei - agustus," vol. 6, no. 1, 2015.

[4] E. Elisa, "Analisa dan Penerapan Algoritma C4.5 Dalam Data Mining Untuk Mengidentifikasi Faktor-Faktor Penyebab Kecelakaan Kerja Kontruksi PT.Arupadhatu Adisesanti," *J. Online Inform.*, vol. 2, no. 1, p. 36, 2017, doi: 10.15575/join.v2i1.71.

[5] R. K. Amin, Indwiarti, and Y. Sibaroni, "Implementasi Klasifikasi Decision Tree Dengan Algoritma C4 . 5 Dalam Pengambilan Keputusan Permohonan Kredit Oleh Debitur," *e-Proceeding Eng.*, vol. 2, no. 1, 2015.

[6] W. Wijanarto and R. Puspitasari, "Optimasi Algoritma Klasifikasi Biner dengan Tuning Parameter pada Penyakit Diabetes Mellitus," *Eksplora Inform.*, vol. 9, no. 1, pp. 50–59, 2019, doi: 10.30864/eksplora.v9i1.257.

[7] A. G. Lazuardy, H. S. S. Kom, and M. Eng, "Data Cleansing Pada Data Rumah Sakit," pp. 1–6, 2019.

[8] M. Yunus, H. Ramadhan, D. R. Aji, and A. Yulianto, "Penerapan Metode Data Mining C4.5 Untuk Pemilihan Penerima Kartu Indonesia Pintar (KIP)," *Paradig. - J. Komput. dan Inform.*, vol. 23, no. 2, 2021, doi: 10.31294/p.v23i2.11395.

[9] A. Kumar and B. K. Sarkar, "A hybrid predictive model integrating C4.5 and decision table classifiers for medical data sets," *J. Inf. Technol. Res.*, vol. 11, no. 2, pp. 150–167, 2018, doi: 10.4018/JITR.2018040109.

[10] D. M. Farid, L. Zhang, C. M. Rahman, M. A. Hossain, and R. Strachan, "Hybrid decision tree and naïve Bayes classifiers for multi-class classification tasks," *Expert Syst. Appl.*, vol. 41, no. 4 PART 2, pp. 1937–1946, 2014, doi: 10.1016/j.eswa.2013.08.089.

[11] Y. Kustiyahningsih and E. Rahmanita, "Aplikasi Sistem Pendukung Keputusan Menggunakan Algoritma C4.5. untuk Penjurusan SMA," *J. Semantec*, vol. 5, no. 2, pp. 101–108, 2016.

[12] I. Iskandar *et al.*, "Prediksi Kelulusan Mahasiswa Menggunakan Algoritma Decision Tree C4 . 5 Dengan Teknik Pruning," *J. Ilmu Komput. dan Sist. Inf.*, vol. 6, no. 1, pp. 64–68, 2018.

[13] H. Azis, P. Purnawansyah, F. Fattah, and I. P. Putri, "Performa Klasifikasi K-NN dan Cross Validation Pada Data Pasien Pengidap Penyakit Jantung," *Ilk. J. Ilm.*, vol. 12, no. 2, pp. 81–86, 2020, doi: 10.33096/ilkom.v12i2.507.81-86.

[14] A. Prasatya, R. R. A. Siregar, and R. Arianto, "Penerapan Metode K-Means Dan C4.5 Untuk Prediksi Penderita Diabetes," *Petir*, vol. 13, no. 1, pp. 86–100, 2020, doi: 10.33322/petir.v13i1.925.

[15] D. Putra and I. G. A. G. A. Kadnyanana, "Implementation of Feature Selection using Information Gain Algorithm and Discretization with NSL-KDD Intrusion Detection System," *JELIKU (Jurnal Elektron. Ilmu Komput. Udayana)*, vol. 9, no. 3, p. 359, 2021, doi: 10.24843/jlk.2021.v09.i03.p06.

Jurnal Elektronik Ilmu Komputer Udayana
Volume 11, No 3. February 2023

p-ISSN: 2301-5373
e-ISSN: 2654-5101

Article Classification Using Convolutional Neural Network (CNN) And Chi-Square Feature Selection

I Gede Laksmana Yudha[a1], Ngurah Agus Sanjaya ER[a2], Anak Agung Istri Ngurah Eka Karyawati[a3], Ida Bagus Gede Dwidasmara[a4], I Gusti Ngurah Anom Cahyadi Putra[a5], Ida Bagus Made Mahendra[a6]

[a]Program Studi Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam
Universitas Udayana
Bali, Indonesia
[1]laksmanayudha22@gmail.com
[2]agus_sanjaya@unud.ac.id
[3]eka.karyawati@unud.ac.id
[4]dwidasmara@unud.ac.id
[5]anom.cp@unud.ac.id
[6]ibm.mahendra@unud.ac.id

Abstract

News articles are reports or information about events that are actual, reliable, and based on facts or reality. The increase in internet users has resulted in the growth of the amount of available information increasing rapidly. Easy internet access makes many types of Indonesian news articles published digitally. With a very large number of news articles, it will be easier to find a news article if the news has been organized and has been grouped according to its respective categories. Text classification is a problem that aims to determine the topic or theme of a document. In achieving this goal, the classification process forms a model that can distinguish data into different classes based on certain rules. The method used to build the model is Convolutional Neural Network (CNN) with Chi-Square feature selection. News articles are divided into six classes, namely news, technology, football, health, lifestyle, and automotive. In this study, the best CNN model was obtained with the number of filters used was 200 and the feature selection being 40%. The test results on the test data provide an accuracy value of 96,074%, precision of 96,079%, recall of 96,074%, and an f-1 score of 96,070%.

Keywords: *Convolutional Neural Network, Text Classification, Chi-Square, K-Fold Cross Validation, Article*

1. Pendahuluan

Artikel berita adalah laporan atau informasi mengenai kejadian/peristiwa yang bersifat aktual, dapat dipercaya, dan berdasarkan fakta atau realita. Karena kemajuan teknologi dan internet, artikel berita semakin sedikit ditemui pada media cetak seperti koran dan banyak beralih pada media digital misalnya website. Peningkatan jumlah informasi yang pesat diakibatkan oleh meningkatnya pula penggunaan internet oleh masyarakat. Mudahnya akses internet menyebabkan beragam jenis artikel berita Bahasa Indonesia secara digital banyak dipublikasikan, hal ini dapat memberi manfaat dalam akses informasi berbahasa Indonesia secara mudah dan cepat. Namun, dengan meningkatnya jumlah artikel berita Bahasa Indonesia menimbulkan masalah lain yaitu pada pengkategorian topik pada setiap artikel berita yang dapat memakan waktu.

Terdapat beberapa permasalahan dalam pengelompokan dokumen, salah satunya yaitu klasifikasi. Selain klasifikasi, klasterisasi juga merupakan permasalahan dalam pengelompokan dokumen [1]. Namun pada kasus kali ini, klasifikasi menjadi permasalahan yang tepat, karena dokumen ditentukan berdasarkan kategori yang telah ditentukan. Klasifikasi teks merupakan permasalahan yang memiliki tujuan untuk menentukan topik atau tema suatu dokumen [2]. Untuk menentukan topik ke dalam kelasnya masing - masing, maka pada proses klasifikasi

akan dibentuk suatu model berdasarkan aturan tertentu sehingga dapat membedakan topik dari suatu dokumen.

Berbagai macam metode telah digunakan dalam penelitian - penelitian sebelumnya terkait dengan klasifikasi teks. Salah satunya penelitian yang dilakukan oleh Ramdhani yaitu klasifikasi berita indonesia menggunakan CNN. Hasil yang diperoleh dari pengujian adalah akurasi terbaik sebesar 90,74% dan nilai loss sebesar 29,05% [3].

Terdapat beberapa cara yang dilakukan untuk dapat meningkatkan performa algoritma klasifikasi. Salah satunya dengan menggunakan seleksi fitur. Pada jurnal Suharno melakukan klasifikasi menggunakan *K-Nearest Neighbors* dan *Chi-Square* sebagai seleksi fitur pada dokumen pengaduan Sambat. Dari hasil pengujian dengan seleksi fitur sebesar 25%, didapatkan nilai *precision*, *recall*, dan *f-measure* secara berturut - turut sebesar 90%, 78%, dan 78%. Dari hasil penelitian tersebut, nilai *f-measure* dapat ditingkatkan melalui penggunaan seleksi fitur dan Metode KNN pada klasifikasi dokumen pengaduan SAMBAT[4].

Kemudian, menurut Farid pada penelitiannya tentang deteksi hoax pada twitter dengan metode CNN dan seleksi fitur information gain. Dalam penelitian ini Farid et al (2020) menyimpulkan bahwa metode *Convolutional Neural Network* (CNN) dapat dengan baik mengklasifikasikan berita hoax. Penggunaan TF-IDF dan *Information Gain* untuk seleksi fitur juga sangat mempengaruhi hasil klasifikasi karena pada pengujian diperoleh nilai akurasi tertinggi sebesar 95,56% [5].

Sehingga dalam melakukan penelitian ini, penulis berlandaskan pada permasalahan dan penelitian terkait yang telah disebutkan. Sehingga, penulis bermaksud untuk melakukan implementasi algoritma *Convolutional Neural Network* (CNN) pada klasifikasi artikel berita bahasa Indonesia dengan seleksi fitur *Chi-Square*.

2. Metodologi Penelitian

Pada metode penelitian, secara umum terdapat alur penelitian mulai dari input data yang berupa teks berita, yang setelah itu akan dilakukan tahap prepocessing dan seleksi fitur *Chi-Square*. Kemudian, dilakuakn TF-IDF untuk mengubah bentuk data dokumen menjadi vektor sebelum dimasukan ke algoritma CNN. Selanjutnya akan masuk ke tahap klasifikasi *Convolutional Neural Network* (CNN) yang menghasilkan output berupa nama kelas dari probabilitas tertinggi hasil klasifikasi. Kemudian untuk pengukuran performa model digunakan *k-fold cross validation* pada tahap evaluasi.

2.1. Pengumpulan Data

Data yang digunakan adalah artikel berita berbahasa Indonesia yang bersumber dari situs Kompas.com, hal ini dikarenakan situs-situs tersebut merupakan situs portal berita terpercaya dan situs-situs tersebut memiliki beberapa kategori berita yang sama dan kedua website tersebut dirasa cukup untuk memenuhi kebutuhan jumlah data. Data diperoleh secara sekunder menggunakan web-scraping. Kelas yang digunakan adalah kategori yang telah tersedia pada situs berita sebanyak enam kategori yaitu news, teknologi, bola, health, lifestyle, dan otomotif. Kelas – kelas tersebut dipilih berdasarkan kategori yang paling sering muncul pada kebanyakan situs portal berita. Jumlah data yang akan digunakan adalah 13.500 data teks berita dengan 2.250 data pada setiap kelasnya.

2.2. Preprocessing

Untuk menyiapkan data menjadi siap diolah dan bersih maka memerlukan proses preprocessing. Urutan proses *preprocessing* secara berturut turut dimulai dari input data dokumen, *case folding*, *punctuation removal*, stemming, tokenisasi, stopword removal. Pada tabel 1 diperlihatkan contoh proses *preprocessing* pada teks.
Pada *case folding* dokumen diubah ke dalam bentuk huruf kecil, untuk menyeragamkan ukuran teks. Sehingga, kata yang sama dengan salah satu kata menggunakan huruf kapital akan dianggap berbeda. Kata ‘Sukses’ akan sama dengan ‘sukses’ pada *case folding*. Selanjutnya, untuk menghilangkan tanda baca dilakukan punctuation removal, agar mempermudah proses tokenisasi. Kemudian, kata - kata yang memiliki imbuhan akan diubah ke bentuk dasarnya melalui proses stemming. Selanjutnya, kata - kata dipecah menjadi potongan token pada

p-ISSN: 2301-5373
e-ISSN: 2654-5101

proses tokenisasi. Terakhir, dilakukan penghapusan kata - kata umum yang tidak memberikan perbedaan konten secara signifikan melalui proses stopword removal [6].

Tabel 1. *Text Preprocessing*

Preprocessing	Hasil
Data Awal	Setelah sukses dengan penjualan perdana terbatas pada model sebelumnya, Katalis kembali luncurkan skuter listrik yang unik. Motor ramah lingkungan ini disebut terinspirasi dari robot.
Case Folding	setelah sukses dengan penjualan perdana terbatas pada model sebelumnya, katalis kembali luncurkan skuter listrik yang unik. motor ramah lingkungan ini disebut terinspirasi dari robot.
Punctuation Removal	setelah sukses dengan penjualan perdana terbatas pada model sebelumnya katalis kembali luncurkan skuter listrik yang unik motor ramah lingkungan ini disebut terinspirasi dari robot
Stemming	telah sukses dengan jual perdana batas pada model belum katalis kembali luncur skuter listrik yang unik motor ramah lingkung ini sebut inspirasi dari robot
Tokenisasi	['telah', 'sukses', 'dengan', 'jual', 'perdana', 'batas', 'pada', 'model', 'belum', 'katalis', 'kembali', 'luncur', 'skuter', 'listrik', 'yang', 'unik', 'motor', 'ramah', 'lingkung', 'ini', 'sebut', 'inspirasi', 'dari', 'robot']
Stopword	['sukses', 'jual', 'perdana', 'batas', 'model', 'katalis', 'luncur', 'skuter', 'listrik', 'unik', 'motor', 'ramah', 'lingkung', 'inspirasi', 'robot']

2.3. Seleksi Fitur

Seleksi fitur juga diperlukan dalam klasifikasi, ini sangat penting dalam klasifikasi teks karena tingginya dimensi fitur teks dan keberadaan fitur yang tidak relevan (*noisy*) [7]. Chi-Square adalah salah satu teknik filter dalam seleksi fitur [8]. Hasil klasifikasi dipengaruhi oleh jumlah fitur yang digunakan, seleksi fitur membantu dalam mengurangi jumlah fitur yang dianggap tidak relevan pada kumpulan dokumen. Sehingga, kinerja metode klasifikasi dapat ditingkatkan efektifitas dan efisiensinya. Pengujian derajat kepentingan sebuah term terhadap kategorinya dilakukan menggunakan seleksi fitur *chi-square*. Berikut ini merupakan fungsi persamaan dari Chi-Square :

$$x^2(t,c) = \frac{N(AD-CB)^2}{(A+C)(B+D)(A+B)(C+D)} \quad (1)$$

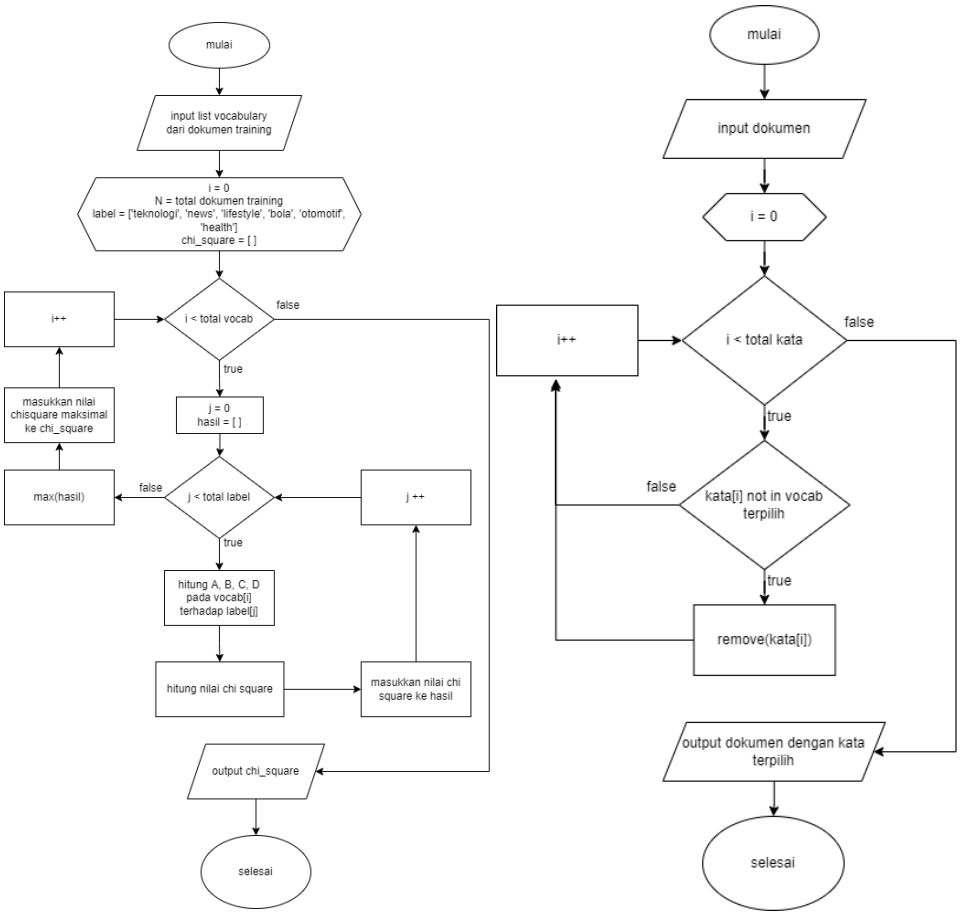
keterangan :
t = Term
c = Kelas/kategori topik dokumen
N = Total dokumen latih/*training*
A = Total dokumen pada kelas/kategori topik c yang mengandung t
B = Total dokumen bukan kelas/kategori topik c yang mengandung t
C = Total dokumen pada kelas/kategori topik c yang tidak mengandung t
D = Total dokumen bukan kelas/kategori topik c yang tidak mengandung t

Nilai Chi-Square ini dapat diglobalisasikan pada semua kategori dengan memilih nilai maksimal atau dengan menggunakan nilai rata – rata di antara semua kategori [4].

$$x^2{}_{avg}(t) = \sum_{i=1}^{m} P(c_i) \,.\, x^2(t,c_i) \quad (2)$$

$$x^2{}_{max}(t) = max_{i=1}^{m}\{x^2(t, c_i)\} \quad (3)$$

Gambar 1 adalah flowchart untuk menghitung nilai chi-square tunggal untuk setiap kata pada vocabulary. Suatu kata i pada satu kategori akan dihitung nilai A, B, C, D sesuai dengan persamaan (1). Kemudian, nilai – nilai tersebut digunakan untuk menghitung nilai *chi-square* menggunakan persamaan (1). Kemudian, perhitungan nilai *chi-square* dilanjutkan untuk semua kategori. Setelah mendapatkan nilai *chi-square* kata i dengan semua kategori, selanjutnya dipilih nilai chi-square maksimal yang akan digunakan sebagai nilai *chi-square* tunggal untuk kata i. Proses ini dilakukan terhadap semua kata yang ada pada *vocabulary*. Pada proses akhir, kata – kata pada *vocabulary* akan diurutkan berdasarkan nilai *chi-square* tertinggi ke terendah. Kemudian pada *flowchart*, kata – kata dalam dokumen yang tidak termasuk dalam kata – kata rasio seleksi fitur akan dihilangkan.



Gambar 1. Seleksi Fitur *Chi-Square*

2.4. Pembobotan TF-IDF

Setelah melalui tahap *preprocessing*, agar dapat diproses data perlu diubah ke dalam bentuk numerik. TF-IDF mentransformasi data hasil preprocessing ke dalam bentuk numerik yaitu vektor. Pembobotan ini menentukan tingkat hubungan kata terhadap dokumen dengan cara memberikan nilai bobot pada setiap kata. TF-IDF mengkombinasikan konsep tingkat kemunculan term atau kata pada sebuah dokumen (*term frequency*) dan tingkat kepentingan dari term atau kata pada dokumen (*inverse document frequency*) [9]. Persamaan dari TF-IDF dapat dilihat pada rumus (4), (5), dan (6).

p-ISSN: 2301-5373
e-ISSN: 2654-5101

a. Term Frequency

$$tf_{ik} = \frac{f_{ik}}{\sum_{j=1}^{t} f_{ij}} \quad (4)$$

yang mana, f_{ik} merupakan kemunculan kata atau term *k* pada dokumen ke-i.

b. Inverse Document Frequency

$$idf_k = log\frac{N}{n_k} \quad (5)$$

dimana *N* adalah jumlah total dokumen, n_k merupakan jumlah total dokumen dengan kemunculan kata atau term *k*.

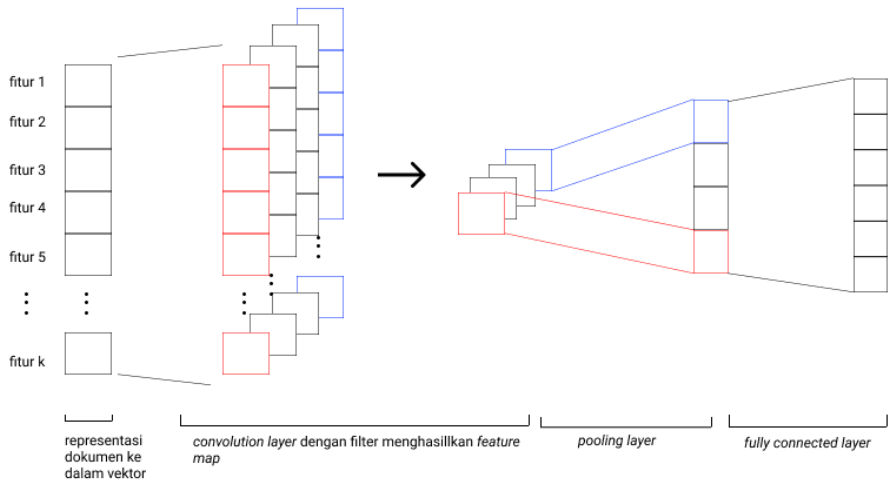
c. Term Frequency - Inverse Document Frequency

Kombinasi dari nilai TF dan IDF akan menghasilkan nilai TF-IDF.

$$W_{ik} = tf_{ik} \times idf_k \quad (6)$$

2.5. Convolutional Neural Network

Algoritma CNN merupakan metode pengembangan *multilayer perceptron* yang memiliki tingkat kedalaman jaringan yang tinggi sehingga tergolong ke salah satu jenis *Deep Neural Network* [10]. CNN umumnya terdiri dari beberapa lapisan, yaitu lapisan konvolusi, lapisan *pooling*, dan lapisan *fully connected*. Data yang menjadi masukkan pada CNN akan dipelajari fiturnya pada lapisan konvolusi dan *pooling*, yang selanjutnya akan dilakukan klasifikasi pada lapisan *fully connected*. Menurut Goldberg, ide utama di balik konvolusi dan *pooling* adalah untuk menerapkan filter pada setiap instansiasi dari h-word sliding window pada kalimat [11].



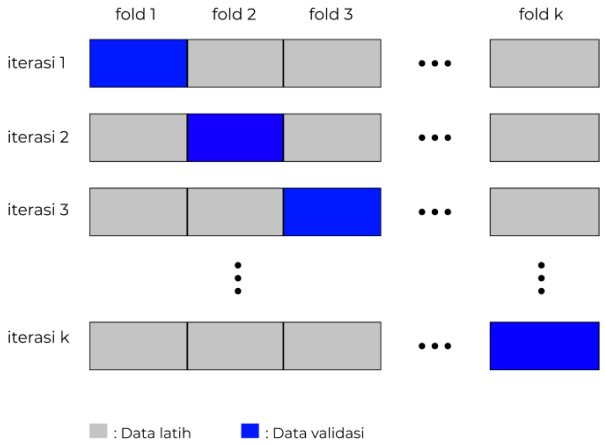
Gambar 2. Arsitektur CNN

Salah satu ciri utama dari CNN dengan algoritma lainnya adalah lapisan konvolusi. Lapisan ini menghasilkan *feature map* dengan menggeser sebuah filter yang melakukan proses konvolusi terhadap data input pada lapisan sebelumnya.

Setelah *feature map* dihasilkan, maka selanjutnya feature map tersebut menjadi masukan pada lapisan pooling. Lapisan pooling bertugas untuk meringkas/mereduksi dimensi dari feature map, sehingga ukuran data menjadi lebih kecil yang berpengaruh pada kecepatan komputasi. Max Pooling adalah salah satu contoh dari lapisan pooling yang memperoleh informasi penting dari feature map dengan mengambil nilai maksimal pada elemen-elemen yang berada pada lingkup window [12].

Kemudian, lapisan fully connected akan melakukan proses klasifikasi terhadap fitur - fitur yang dihasilkan dari proses sebelumnya. Lapisan ini terdiri atas lapisan input, lapisan tersembunya atau hidden layer, dan lapisan output yang masing - masing lapisan memiliki neuron – neuron yang saling terhubung. Proses pemetaan fitur ke dalam kategori akan dibantu dengan nilai bobot yang dimiliki oleh setiap neuron - neuron tersebut.

2.6. K-Fold Cross Validation



Gambar 3. *K-Fold Cross Validation*

K-Fold Cross Validation membagi keseluruhan dataset dalam k bagian, sehingga terdapat sebanyak k iterasi yang setiap iterasinya secara bergantian bagian – bagian data akan menjadi data latih dan data validasi [13]. Gambar 3 merupakan Ilustrasi dari *k-fold cross validation*.

2.7. Evaluasi

Pengukuran performa didasarkan pada jumlah pengujian yang diprediksi dengan jumlah benar dan salah oleh model yang ditabulasikan dalam tabel yang disebut sebagai confusion matrix [14]. Kemudian nilai benar dan salah tersebut dihitung kedalam TP, FN, FP, dan TN yang selanjutnya secara berturut – turut diukur melalui persamaan akurasi, *precision, recall,* dan *f-1 score* melalui rumus (7), (8), (9), dan (10). Tabel 2 adalah contoh dari *confusion matrix*.

Tabel 2. *Confusion Matrix*

		Kelas Prediksi	
		Positif	Negatif
Kelas Sebenarnya	Positif	TP	FN
	Negatif	FP	TN

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \times 100\% \quad (7)$$

$$Precision = \frac{TP}{TP + FP} \quad (8)$$

$$Recall = \frac{TP}{TP + FN} \quad (9)$$

$$F - 1\ Score = \frac{2 \times precision \times recall}{precision + recall} \quad (10)$$

p-ISSN: 2301-5373
e-ISSN: 2654-5101

True Positive (TP) terjadi ketika data positif berhasil dengan benar diklasifikasikan oleh model. *False Negative* (FN) terjadi ketika data positif diklasifikasikan salah sebagai data negatif oleh model. *False Positive* (FP) terjadi ketika data negatif diklasifikasikan salah sebagai data positif oleh model. *True Negative* (TN) terjadi ketika data negatif berhasil diklasifikasikan dengan benar oleh model.

3. Hasil dan Diskusi

Tabel 3. Hasil Uji Coba Data Validasi

Rasio	Filter	Akurasi	*Precision*	*Recall*	*F-1*
10 %	50	94,657%	94,664%	94,665%	94,642%
	100	94,926%	94,945%	94,937%	94,913%
	200	94,787%	94,805%	94,793%	94,774%
	400	94,889%	94,879%	94,902%	94,879%
20%	50	95,065%	95,058%	95,084%	95,061%
	100	94,870%	94,881%	94,876%	94,853%
	200	94,889%	94,885%	94,904%	94,876%
	400	95,065%	95,062%	95,080%	95,056%
30%	50	95,213%	95,213%	95,233%	95,204%
	100	95,315%	95,322%	95,327%	95,308%
	200	95,176%	95,188%	95,195%	95,164%
	400	95,232%	95,242%	95,242%	95,220%
40%	50	95,185%	95,181%	95,195%	95,177%
	100	95,278%	95,271%	95,297%	95,273%
	200	**95,315%**	**95,314%**	**95,336%**	**95,312%**
	400	95,157%	95,159%	95,178%	95,153%
50%	50	95,102%	95,101%	95,115%	95,100%
	100	95,120%	95,113%	95,134%	95,115%
	200	95,074%	95,080%	95,084%	95,067%
	400	95,065%	95,080%	95,088%	95,062%
60%	50	95,093%	95,093%	95,110%	95,095%
	100	95,037%	95,035%	95,051%	95,023%
	200	95,102%	95,114%	95,121%	95,098%
	400	95,204%	95,208%	95,222%	95,194%
70%	50	95,129%	95,122%	95,147%	95,125%
	100	95,009%	95,023%	95,013%	94,995%
	200	95,278%	95,283%	95,293%	95,269%
	400	95,000%	95,003%	95,019%	94,996%
80%	50	95,083%	95,079%	95,099%	95,076%
	100	95,102%	95,110%	95,117%	95,093%
	200	95,065%	95,057%	95,089%	95,062%
	400	95,074%	95,083%	95,090%	95,068%
90%	50	95,102%	95,100%	95,123%	95,096%
	100	95,111%	95,107%	95,124%	95,106%
	200	95,120%	95,121%	95,137%	95,115%
	400	95,167%	95,169%	95,192%	95,162%
100%	50	94,796%	94,801%	94,812%	94,789%
	100	94,870%	94,865%	94,885%	94,862%
	200	94,972%	94,975%	94,999%	94,972%
	400	94,778%	94,785%	94,786%	94,764%

Pada pengujian, kombinasi parameter yang digunakan adalah rasio seleksi fitur dan jumlah filter pada CNN. Rasio seleksi fitur adalah 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, dan 100% dari total fitur yang diperoleh. Kemudian Jumlah filter pada CNN adalah 50, 100, 200, dan 400. Pengujian menggunakan *k-fold cross validation* dengan nilai *k* yang digunakan adalah 5. Hasil *k-fold cross validation* selanjutnya akan dihitung rata – rata nilai evaluasi yaitu akurasi, *precision, recall*, dan *f-1 score* yang kemudian dipilih kombinasi parameter dengan nilai evaluasi terbaik.

Tabel 3 merupakan hasil dari pengujian dengan menggunakan *k-fold cross validation* dengan nilai akurasi, *precision, recall,* dan *f-1* score terhadap rasio seleksi fitur dan jumlah filter CNN yang digunakan. Berdasarkan hasil pengujian pada Tabel 3, didapat kombinasi parameter terbaik yaitu rasio seleksi fitur sebesar 40% dan jumlah filter sebanyak 200 dengan nilai rata – rata dari akurasi sebesar 95,315%, *precision* sebesar 95,314%, *recall* sebesar 95,336%, dan f-1 *score sebesar* 95,312%. Berdasarkan Tabel 3 didapat bahwa perubahan filter CNN dan rasio seleksi fitur cenderung tidak memberikan pengaruh yang signifikan terhadap akurasi model.

Tabel 4. Hasil Uji Coba Data Testing

Akurasi	***Precision***	***Recall***	***F-1***
96,074%	96,079%	96,074%	96,070%

Tabel 5. Pengaruh Filter CNN Terhadap Akurasi Model

Rasio Seleksi Fitur	**Filter CNN**	**Akurasi**
100%	50	94,796%
	100	94,870%
	200	94,972%
	400	94,778%

Tabel 6. Pengaruh Rasio Seleksi Fitur Terhadap Akurasi Model

Filter CNN	**Rasio Seleksi Fitur**	**Akurasi**
200	10%	94,787%
	20%	94,889%
	30%	95,176%
	40%	95,315%
	50%	95,074%
	60%	95,102%
	70%	95,278%
	80%	95,065%
	90%	95,120%
	100%	94,972%

p-ISSN: 2301-5373
e-ISSN: 2654-5101

Setelah mendapatkan model dengan kombinasi parameter terbaik, selanjutnya model tersebut akan diujikan pada data testing. Tabel 4 adalah hasil uji coba model terhadap data *testing*. Berdasarkan hasil uji coba model terhadap data *testing*, nilai evaluasi rata – rata yang diperoleh adalah akurasi sebesar 96,074%, precision sebesar 96,079%, recall sebesar 96,074%, dan f-1 score sebesar 96,070%. Kemudian, mengenai pengaruh yang diberikan melalui perubahan hyperparameter filter CNN dan rasio seleksi fitur terhadap akurasi model dapat dilihat pada Tabel 5 dan 6.

Berdasarkan Tabel 5 dan Tabel 6 didapat bahwa perubahan rasio seleksi fitur dan filter CNN cenderung tidak memberikan pengaruh yang signifikan terhadap akurasi model.

Tabel 7. *Running Time Training* Terhadap Filter CNN

Rasio Seleksi Fitur	Filter CNN	*Running Time* (menit)
100%	50	12,47
	100	30,01
	200	124,05
	400	224,44

Tabel 8. *Running Time Training* Terhadap Rasio Seleksi Fitur

Filter CNN	Rasio Seleksi Fitur	*Running Time* (menit)
200	10%	2,70
	20%	6,78
	30%	12,14
	40%	21,08
	50%	38,58
	60%	44,91
	70%	86,39
	80%	95,87
	90%	121,55
	100%	124,05

Berdasarkan Tabel 7 dan Tabel 8 didapat bahwa semakin tinggi rasio seleksi fitur dan semakin banyak filter pada CNN yang digunakan maka semakin lama waktu yang diperlukan oleh model untuk proses training.

4. Kesimpulan

Berdasarkan hasil pengujian yang dilakukan, maka dapat ditarik kesimpulan bahwa:

a. Dari hasil validasi model dengan *k-fold cross validation* diperoleh bahwa model klasifikasi artikel berita Bahasa Indonesia terbaik adalah model *Convolutional Neural Network* (CNN) dengan jumlah filter 200 dan seleksi fitur sebesar 40%, dimana akurasi, *precision*, *recall*, dan *f-1 score* dari validasi adalah secara berturut – turut sebesar 95,315%, 95,314%, 95,336%, dan 95,312%.

b. Hasil *k-fold cross validation* menunjukkan kecenderungan bahwa perubahan rasio seleksi fitur tidak memberikan pengaruh yang signifikan terhadap performa model klasifikasi menggunakan *Convolutional Neural Network* (CNN), namun peningkatan rasio seleksi fitur dapat mempengaruhi running time pada sistem. Semakin tinggi rasio seleksi fitur yang digunakan maka semakin lama waktu yang diperlukan oleh model untuk proses training. Sehingga, seleksi fitur dapat mengurangi fitur yang dianggap kurang relevan untuk proses klasifikasi agar proses training menjadi lebih cepat. Hasil terbaik diperoleh pada saat seleksi fitur sebesar 40% dengan nilai akurasi, *precision*, *recall*, dan *f-1 score* model klasifikasi terhadap data *testing* (*unseen* data) adalah berturut-turut sebesar 96,074%, 96,079%, 96,074%, dan 96,070%.

.

Daftar Pustaka

[1] N. A. S. ER, “Implementasi Latent Dirichlet Allocation (Lda) Untuk Implementation of Latent Dirichlet Allocation (Lda) for,” vol. 8, no. 1, pp. 127–134, 2021, doi: 10.25126/jtiik.202183556.

[2] C. Manning and H. Schutze, *Foundations of statistical natural language processing*. MIT press, 1999.

[3] M. A. Ramdhani, M. A. Ramdhani, D. S. adillah Maylawati, and T. Mantoro, “Indonesian news classification using convolutional neural network,” *Indones. J. Electr. Eng. Comput. Sci.*, vol. 19, no. 2, pp. 1000–1009, 2020, doi: 10.11591/ijeecs.v19.i2.pp1000-1009.

[4] C. F. Suharno, M. A. Fauzi, and R. S. Perdana, “Klasifikasi Teks Bahasa Indonesia Pada Dokumen Pengaduan Sambat Online Menggunakan Metode K-Nearest Neighbors Dan Chi-square,” *Syst. Inf. Syst. Informatics J.*, vol. 3, no. 1, pp. 25–32, 2017, doi: 10.29080/systemic.v3i1.191.

[5] H. K. Farid, E. B. Setiawan, and I. Kurniawan, “Selection for Hoax News Detection on Twitter using Convolutional Neural Network,” *Indones. J. Comput.*, vol. 5, no. December 2020, pp. 23–36, 2020, doi: 10.34818/indojc.2021.5.3.506.

[6] F. Taufiqurrahman, S. Al Faraby, and M. D. Purbolaksono, “Klasifikasi Teks Multi Label pada Hadis Terjemahan Bahasa Indonesia Menggunakan Chi Square dan SVM,” *e-Proceeding Eng.*, vol. 8, no. 5, pp. 10650–10659, 2021.

[7] C. C. Aggarwal and C. Zhai, *Mining text data*. Springer Science & Business Media, 2012.

[8] I. Listiowarni and N. Puspa Dewi, “Pemanfaatan Klasifikasi Soal Biologi Cognitive Domain Bloom’s Taxonomy Menggunakan KNN Chi-Square Sebagai Penyusunan Naskah Soal,” *Digit. Zo. J. Teknol. Inf. dan Komun.*, vol. 11, no. 2, pp. 186–197, 2020, doi: 10.31849/digitalzone.v11i2.4798.

[9] K. D. Yonatha Wijaya and A. A. I. N. E. Karyawati, “The Effects of Different Kernels in SVM Sentiment Analysis on Mass Social Distancing,” *JELIKU (Jurnal Elektron. Ilmu Komput. Udayana)*, vol. 9, no. 2, p. 161, 2020, doi: 10.24843/jlk.2020.v09.i02.p01.

[10] A. R. Maulana and N. Rochmawati, “Opinion Mining Terhadap Pemberitaan Corona di Instagram menggunakan Convolutional Neural Network,” *JINACS*, vol. 02, pp. 53–59, 2020.

[11] Y. Goldberg, *Neural network methods for natural language processing*, vol. 10, no. 1. Morgan & Claypool Publishers, 2017.

[12] Y. Kim, “Convolutional neural networks for sentence classification,” *EMNLP 2014 - 2014 Conf. Empir. Methods Nat. Lang. Process. Proc. Conf.*, pp. 1746–1751, 2014, doi: 10.3115/v1/d14-1181.

[13] H. Rhomadhona and J. Permadi, “Klasifikasi Berita Kriminal Menggunakan NaÃ¯ve Bayes Classifier (NBC) dengan Pengujian K-Fold Cross Validation,” *J. Sains dan Inform.*, vol. 5, no. 2, pp. 108–117, 2019, doi: 10.34128/jsi.v5i2.177.

[14] P.-N. Tan, M. Steinbach, and V. Kumar, *Introducing to Data Mining*. Boston: Pearson Education, 2006.

Jurnal Elektronik Ilmu Komputer Udayana
Volume 11, No 3. February 2023

p-ISSN: 2301-5373
e-ISSN: 2654-5101

Implementasi Metode Hybrid Particle Swarm Optimization dan Genetic Algorithm Pada Penjadwalan Job Shop Scheduling

Anak Agung Putra Adnyana [1], I Made Widiartha [2], Agus Muliantara [3], Luh Gede Astuti [4], Made Agung Raharja [5], I Dewa Made Bayu Atmaja Darmawan [6]

Program Studi Informatika Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Udayana
Jl. Raya Kampus Unud, Indonesia Bukit Jimbaran, Badung, Bali
Kode Pos: 80364 Indonesia
[1] putraadnyana@student.unud.ac.id, [2] madewidiartha@unud.ac.id, [3] muliantara@unud.ac.id, [4] lg.astuti@unud.ac.id, [5] made.agung@unud.ac.id, [6] dewabayu@unud.ac.id

Abstract

Job shop problem is one of the non-deterministic combinatorial optimization problems with polynomial time (NP-complete). Genetic Algorithm optimization will be applied to solve Job Shop problems. hybrid particle swarm optimization. In this study.This Study is an attempt to solve Job Shop Scheduling problem using hybrid particle swarm optimization and genetic algorithm method, to find minimum Makespan. 5 parameters, C1, C2, inertia weight, crossover rate and mutation rate, will be compared with a range from 0.1 to 1 with difference 0.2, the test will look for combination parameter that get the minimum Makespan, The results of the implementation of the hybrid particle swarm optimization method and genetic algorithm are makespan of 29 days is obtained with an objective function value of 0.0043, with optimal parameters (C1) = 0.7, (C2) = 0.3, (w) = 0.3, (Cr) = 0.5, and (Mr) = 0.7.

Keywords: *Job shop scheduling, Particle swarm optimization, Genetic algorithm, Makespan, Parameter*

1. Pendahuluan

Dalam dunia industri, proses produksi merupakan proses pembuatan barang atau jasa. Sebelum proses produksi dimulai atau konversi input menjadi produk perusahaan, perlu untuk mengembangkan jadwal produksi yang komprehensif. Diharapkan dengan dilakukannya penjadwalan produksi proses produksi dapat berjalan dengan lancar dan efisien [1]. Dengan waktu yang efisien tersebut diharapkan industri dapat mengurangi pengeluaran biaya produksi serta dapat memenuhi kebutuhaan konsumen tepat waktu, oleh karena itu pemahaman mengenai konsep penjadwalan sangat penting, sehingga para pekerja mengetahui kapan waktu harus memulai suatu pekerjaan dan kapan waktu mengakhirinya. Penjadwalan disusun dengan mempertimbangkan berbagai batasan yang ada. Penjadwalan yang baik akan memberikan dampak positif, yaitu rendahnya biaya operasi dan waktu pengiriman, yang akhirnya dapat meningkatkan kepuasan pelanggan.

Penjadwalan produksi merupakan bagian integral dari sistem perusahaan manufaktur. Permasalahan penjadwalan adalah menugaskan mesin produksi untuk melakukan serangkaian kegiatan melakukan pekerjaan dalam waktu tertentu sampai mencapai tujuan tertentu [2]. Konsep penjadwalan produksi merupakan salah satu contoh dari permasalahan dunia nyata dari Scheduling Problem (SP), SP adalah masalah yang berkaitan dengan pengurutan pemrosesan sejumlah pekerjaan pada sejumlah mesin. SP merupakan permasalahan yang memiliki karakteristik terdiri dari m mesin dan n pekerjaan. Setiap pekerjaan harus diproses pada setiap mesin. Masing-masing mesin dapat memroses paling banyak satu pekerjaan pada suatu waktu. Setiap pekerjaan harus diproses pada suatu mesin selarna suatu periode waktu tertentu tanpa interupsi. Setiap pekerjaan hanya dapat diproses oleh satu mesin dalam satu waktu. perusahaan manufaktur bekerja dengan berbagai sistem produksi, antara lain *flow shop* dan *job shop*. Sistem produksi job shop merupakan penjadwalan yg mempunyai hambatan urutan pemrosesan pekerjaan, dan setiap pekerjaan wajib melalui setiap mesin tepat satu kali. Penjadwalan *job shop* bisa dikelompokkan menurut waktu kedatangannya. *job shop* statis adalah jika seluruh pekerjaan diterima dalam waktu yg sama. *Job shop* dinamis yaitu jika ketika kedatangan pekerjaan diterima dalam waktu yg bervariasi. *Job shop* deterministik apabila variasi waktu kedatangannya diketahui sebelumnya. Tetapi apabila waktu kedatangan yang bervariasi tidak diketahui sebelumnya maka dianggap penjadwalan *job shop* non deterministik atau stokastik [3]. Dalam penelitian ini tujuan

yang akan dicapai adalah menghasilkan jadwal dengan *makespan* yang pendek,dimana *makespan* merupakan waktu penyelesaian paling akhir dari seluruh pekerjaan pada seluruh mesin [4].

Pada penelitian ini metode *hybrid particle swarm optimization* dan *genetic algorithm* digunakan untuk mencari solusi optimal dari masalah penjadwalan *job shop*. Tujuan dari penelitian ini adalah untuk meminimalkan total waktu produksi untuk menyelesaikan semua pekerjaan. Untuk mencapai rencana produksi yang optimal (biaya produksi minimal)

2. Metodelogi Penelitian

2.1 *Job Shop Scheduling*

Job Shop Scheduling Problem (JSSP) merupakan masalah penjadwalan yang banyak digunakan pada industri dengan tipikal pesanan adalah *made to order*, barang Pesanan akan diproduksi apabila hanya terdapat pesanan dari pelanggan. Penjadwalan JSSP adalah jenis penjadwalan yang rumit, salah satu penyebabnya adalah jenis produk atau variasi produk yang ditangani sangat bervariasi. Banyaknya variasi berdasarkan pesanan mengakibatkan timbul banyak jenis pekerjaan dan kebutuhan penggunaan alat yang berbeda. JSSP dapat dijelaskan sebagai berikut:

Terdapat n pekerjaan yang harus diselesaikan selama rentang waktu [T1, T2]. Pekerjaan ini akan diproses pada m mesin dengan prosedur pemesinan yang diberikan. Setiap pekerjaan dapat diproses pada satu dan hanya satu mesin pada satu waktu dan setiap mesin dapat memproses satu dan hanya satu pekerjaan pada satu waktu. Waktu pemrosesan setiap pekerjaan pada setiap mesin adalah tetap dan diasumsikan telah diketahui sebelumnya [5]. Adapun komponen-komponen pembentuk permasalah JSSP adalah sebagai berikut:

1. Rentang waktu pengerjaan [T1, T2], biasanya dalam satuan hari, minggu, atau bulan. (T1 adalah waktu pekerjaan di terima, T2 adalah batas waktu pengerjaan)
2. Pekerjaan $J = \{J_1, J_2, J_3, \ldots, J_n\}$, n adalah jumlah dari seluruh pekerjaan.
3. Mesin $M = \{M_1, M_2, M_3, \ldots, M_m\}$, m adalah dari seluruh mesin.
4. Matriks urutan pengerjaan SJM untuk setiap pekerjaan J, dapat digambarkan sebagai berikut:

$$S_{JM} = \begin{bmatrix} S_{11} & S_{12} \ldots . & S_{ik} \\ S_{21} & S_{22} \ldots . & S_{ik} \end{bmatrix}$$

   Dimana S_{ik} adalah urutan pengerjan dari job ke-i di mesin ke-k
5. Matriks waktu pengerjaan TJM untuk setiap pekerjaan J, dapat digambarkabn sebagai berikut:

$$T_{JM} = \begin{bmatrix} T_{11} & T_{12} \ldots . & T_{ik} \\ T_{21} & T_{22} \ldots . & T_{ik} \end{bmatrix}$$

   Dimana T_{ik} adalah urutan pengerjan dari job ke-i di mesin ke-k

2.2 Metode *Hybrid Particle Swarm Optimization* dan *Genetic Algorithm*

Seperti Namanya metode yang diusulkan dalam penelitian ini menggabungkan dua buah algoritma yaitu Genetic Algorithm yang diusulkan oleh Golberg sebagai metode pencarian *heuristic* yang menyalin seleksi alam [6]. Dan Particle Swarm Optimization atau yang biasa disingkat PSO dimana merupakan salah satu metode *heuristic* yang biasa digunakan dalam optimasi. Particle Swarm Optimization terinspirasi oleh pola berkelompok (*swarm*) burung atau ikan [7]. Proses penerapan metode kedalam permasalahan secara tidak langsung di bagi menjadi dua tahap yaitu tahap GA dan juga Tahap PSO, yang nantinya akan di gabungkan menjadi metode hybrid. Berikut merupakan langkah langkahnya:

1. Menginputkan data yang akan di proses dalam algoritma, yaitu:
   a. Seq_machine adalah matriks urutan pengerjaan.
   b. pcs_time adalah matriks waktu pengerjaan.
2. Menginputkan parameter yang akan digunakan yaitu:
   a. W adalah faktor inertia.
   b. c_1 adalah koefisien percepatan partikel.
   c. c_2 adalah koefisien percepatan populasi.
   d. Pe_{num} adalah ukuran populasi.
   e. r_{max} adalah penghitung iterasi maks.
   f. P_c adalah probabilitas crossover partikel.
   g. P_m adalah probabilitas mutasi.

h. V_{ij} adalah kecepatan partikel
i. $X_{ij}(t)$ adalah Partikel saat ini
j. R_1 & R_2 adalah bilangan random (0-1)

3. Inisialisasi populasi partikel dengan posisi dan kecepatan acak pada dimensi-D, setiap partikel mewakili posisi kandidat pekerjaan. Sebuah partikel dianggap sebagai titik dalam ruang dimensi-D.
4. Hitung nilai fungsi objektif untuk setiap partikel populasi pada iterasi r.
5. Perulangan iterasi terjadi selama r < r_{max} dan nilai selisih G_{best} [r] – G_{best} [r -1] < 0.00001, selisih terjadi lebih dari 20 kali, jika tidak terpenuhi, hentikan algoritma dan G_{best} terakhir menjadi solusi optimal.
6. *Elitism*. Masukan partikel dari populasi dengan nilai fitness terbaik ke generasi berikutnya.
7. Operasi *Crossover*. Dua partikel dalam populasi dipilih dengan probabilitas P_c sebagai partikel induk untuk operasi crossover. Silangkan induk dengan partikel dalam populasi.
8. Operasi *Mutation*. Hasilkan partikel baru dengan menukar gen dari partikel terbaik. Operator ini dapat secara signifikan meningkatkan keragaman populasi dan menghindari masalah lokal optima.
9. Operasi *Repair*. Operator ini digunakan untuk memeriksa kelayakan partikel di populasi. Operator *Repair* digunakan untuk memperbaiki gen yang tidak layak dari partikel menjadi layak. Ide tentang operator ini adalah untuk memisahkan gen yang layak dalam suatu partikel dari yang tidak layak, dan mengembangkan bersama gen yang tidak layak sampai mereka menjadi layak. Cara spesifik untuk melakukan langkah ini adalah menemukan gen yang hilang dari gen partikel, dan kemudian menempelkan gen yang hilang di akhir gen yang baru dihasilkan.
10. Update nilai *fittnes* dari setiap partikel. P_{best} adalah nilai terbaik untuk partikel i selama proses iterasi, dan P_r^i Adalah fitness partikel i saat ini. Jika P_r^i < P_{best} ditetapkan, lalu set P_{best} = P_r^i, jika tidak, P_{best} tetap.
11. Update nilai *fittnes* dari populasi. Petakan posisi setiap partikel ke dalam ruang solusi dan mengevaluasi nilai kesesuaiannya sesuai dengan fungsi objektif. Tentukan Gbest sebagai nilai terbaik dari populasi partikel, G_{best} = min (P_{best}), (nilai i = 1, ..., Pe_{num}).
12. Update kecepatan dan posisi, hal tersebut diperbarui sesuai dengan persamaan (1) dan persamaan (2). Kembali ke Langkah 3 setelah memperbarui kecepatan dan posisi dan memulai iterasi baru.

$$V_{ij}(t+1) = W \times V_{ij}(t) + c_1 \times R_1 \times \left(Pbest - X_{ij}(t)\right) + c_2 \times R_2 \times \left(Gbest - X_{ij}(t)\right) \quad (1)$$

$$X'_{ij}(t+1) = X_{ij}(t) + V_{ij}(t+1) \quad (2)$$

2.3 Data

Data yang digunakan untuk penelitian ini adalah data mesin, data order/pekerjaan, data urutan pengerjaan, data waktu pengerjaan. Jenis data yang dipakai pada penelitian ini merupakan data sekunder. Data sekunder yang di gunakan dalam penelitian kali ini bersumber dari penelitian Purwaningsih dan Fitriana (2016) [8]. Jumlah keseluruhan data yang digunakan terdiri dari 27 buah pesanan barang yang berbentuk furniture.

3. Hasil dan pengujian

Fungsi objektif merupakan tujuan dari penjadwalan yang dilakukan. Dalam permasalahan penelitian ini fungsi objektif merupakan meminimasi makespan.

$$makespan = PT_{X(ms)} +_{\ldots..} PT_{X(ms)+n} \quad (3)$$

Lantaran tujuan penjadwalan merupakan meminimalkan sedangkan prinsip algoritma genetika merupakan memaksimasia, maka nilai fungsi fitness dalam masalah ini menjadi:

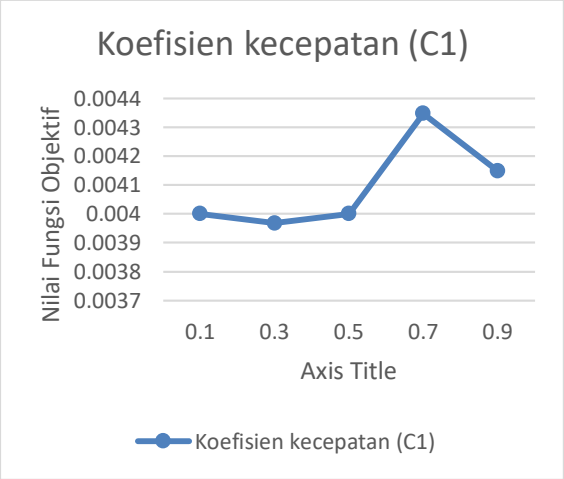
$$F(x) = \frac{1}{makespan} \quad (4)$$

3.1. Pengujian

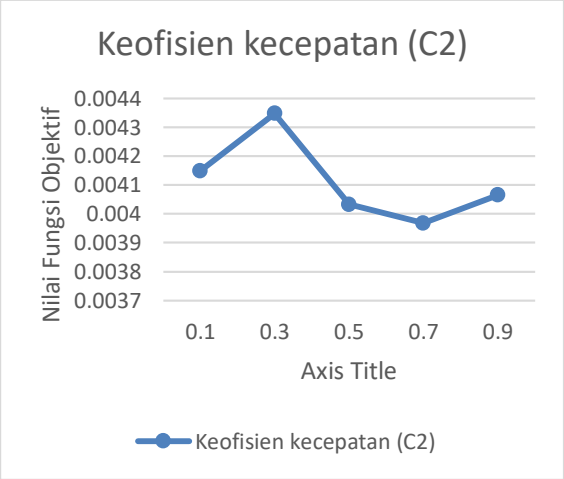
Pada penelitian ini akan dilakukan pengujian menentukan solusi optimal dengan beberapa parameter antara lain koefisien percepatan partikel (C1), koefisien percepatan populasi (C2), Inertia weight (w), probabilitas crossover, probabilitas mutation. Data akan diuji parameternya untuk melihat pengaruh nilai parameter terhadap makespan atau waktu pengerjaan dari jadwal yang dihasilkan. Keseluruhan Parameter yang akan diuji memiliki rentang uji dari 0.1 sampai 1 dengan selisih 0,2, dan dengan nilai parameter tetap seperti jumlah populasi sama dengan 100 jumlah iterasi sama dengan 1000.

3.2. Hasil Penelitian

Berikut merupakan hasil terbaik dari pengujian parameter yang direncanakan sebelumnya, nantinya hasil ini akan dibandingkan dengan hasil dari penelitian sebelumnya [8].

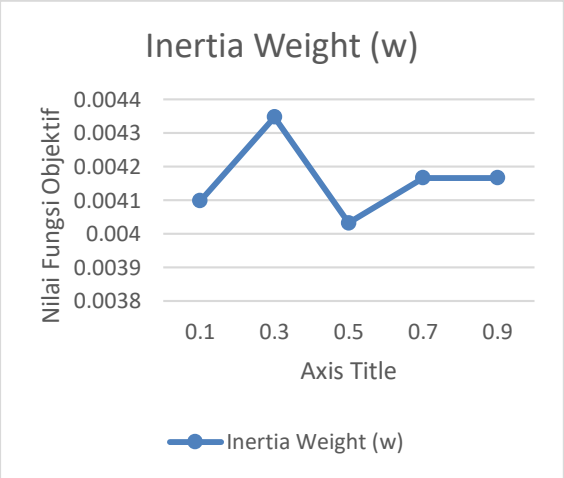


Gambar 1. Hasil pengujian parameter koefisien kecepatan (C1) PSO-GA

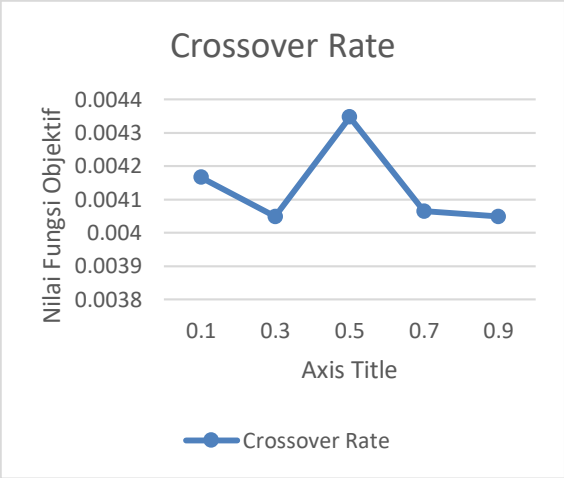


Gambar 2. Hasil pengujian parameter koefisien kecepatan (C2) PSO-GA

Pada Gambar 1 menunjukan perubahan dari parameter koefisien C1 dimana fungsi objektif memiliki puncak/ nilai terbaik pada nilai 0.7 yang mana pada nilai parameter sebelumnya memiliki nilai fungsi objektif yang hampir sama, ini menyebabkan nilai 0.7 menjadi nilai parameter optimal untuk koefisien percepatan C1. Sedangkan pada gambar 2 menunjukan perubahan dari parameter koefisien C2 dimana fungsi objektif memiliki puncak/ nilai terbaik pada nilai 0.3 dan nilai setelahnya mengalami penurunan, ini menyebabkan nilai 0.3 menjadi nilai parameter optimal untuk koefisien percepatan C2.



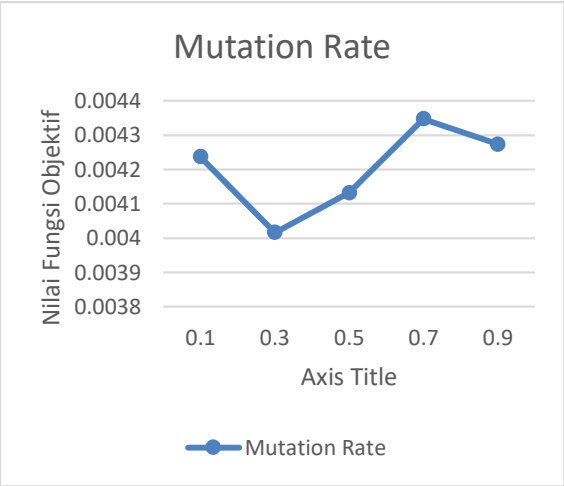
Gambar 3. Hasil pengujian parameter inertia weight (w) PSO-GA



Gambar 4. Hasil pengujian parameter crossover rate (Cr) PSO-GA

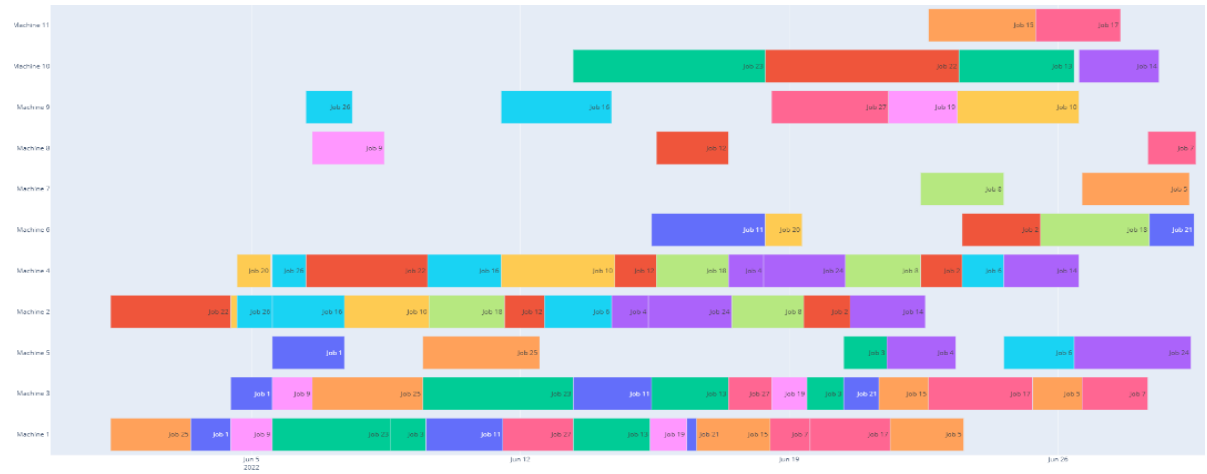
Pada Gambar 3 menunjukan perubahan dari parameter inertia weight (w) dimana fungsi objektif memiliki puncak/ nilai terbaik pada nilai 0.3 yang mana pada nilai parameter selanjutnya mengalami penurunan lalu mendatar, ini menyebabkan nilai 0.3 menjadi nilai parameter optimal untuk inertia weight (w). Pada Gambar 4 menunjukan perubahan dari parameter crossover rate (Cr) dimana fungsi objektif memiliki puncak/ nilai terbaik pada nilai 0.5 dimana nilai sebelumnya mengalami penurunan dan setelahnya mendatar, ini menyebabkan nilai 0.5 menjadi nilai parameter optimal untuk crossover rate (Cr).

p-ISSN: 2301-5373
e-ISSN: 2654-5101



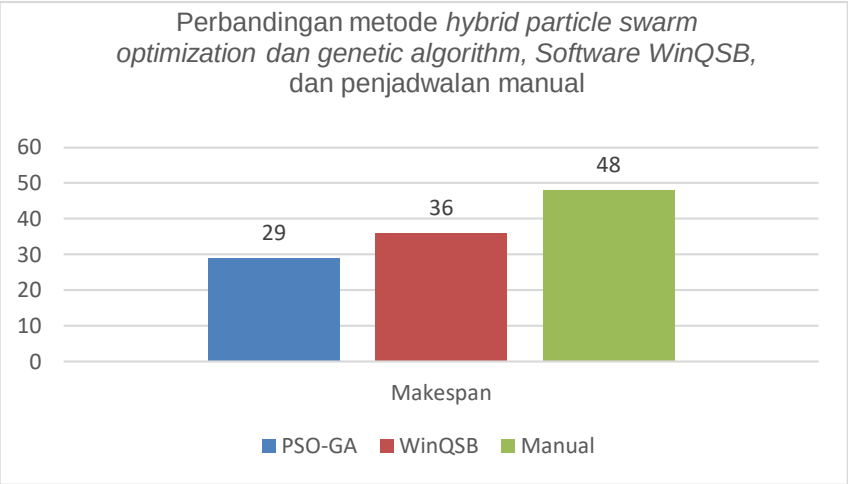
Gambar 5. Hasil pengujian parameter mutation rate (Mr) PSO-GA

Pada Gambar 5 menunjukan perubahan dari parameter mutation rate (Mr) dimana fungsi objektif memiliki puncak/ nilai terbaik pada nilai 0.7 dimana nilai sebelumnya mengalami penurunan lalu perlahan naik ke titik optimal dan setelahnya nilai menurun, ini menyebabkan nilai 0.7 menjadi nilai parameter optimal untuk mutation rate (Mr).



Gambar 6. Gant chart solusi optimal

Dapat dilihat seperti gambar 6 setiap warna dalam jadwal merepresentasikan sebuah pekerjaan yang dikerjakan, dan setiap baris merepresentasikan mesin yang berbeda. Panjang dari setiap bar merepresentasikan waktu dari pengerjaan sebuah pekerjaan, dan dari pengujian sebelumnya didapatkan hasil nilai fungsi objektif optimal sebesar 3.7×10^{-3}, dan gantt chart dari jadwal yang di generate dengan parameter optimal dapat dilihat di gambar 6, *makespan* dari hasil yang didapat adalah sebesar 29 hari.



Gambar 7. Hasil Perbandingan metode *hybrid, Software WinQSB,* dan penjadwalan manual

Perbandingan hasil dari metode hybrid dengan penelitian sebelumnya dapat dilihat pada grafik yang terdapat pada gambar 7, jika dibandingkan dengan hasil dari penelitian sebelumnya [8] metode metode hybrid particle swarm optimization dan genetic algorithm mampu mengungguli dari hasil software WinQSB dan penjadwalan manual pada penelitian sebelumnya. Dari perbandingan tersebut dapat dikatakan metode *hybrid particle swarm optimization* dan *genetic algorithm* dapat menyelesaikan permasalahan Job shop scheduling dengan baik.

4. Kesimpulan

Penelitian ini sudah berhasil mengimplementasikan metode *hybrid particle swarm optimization* dan *genetic algorithm* untuk mendapatkan jadwal pekerjaan optimal. adapun nilai parameter yang optimal untuk mendapatkan nilai fungsi objektif terbaik adalah parameter (C1) = 0.7, (C2) = 0.3, (w) = 0.3, (Cr) = 0.5, dan (Mr) = 0.7, dan variabel tetap jumlah populasi = 100, jumlah generasi = 1000 yang memiliki nilai fungsi objektif 0.0043, waktu pengerjaan atau *makespan* 29 hari dan biaya operasional sebesar Rp 21.085.198.

References

[1] I. Raharja, “Analisa Penjadwalan Proyek Dengan Metode Pert Di PT. Hasana Damai Putra YOGYAKARTA Pada Proyek Perumahan Tirta Sani,” *Bentang*, vol. 2, no. 1, 2014.

[2] Y. Muharni dan D. A. Utami, “Menggunakan Metode Nawaz Enscore Ham Dan Genetic Algorithm,” vol. V, no. 2, hal. 29–39, 2019.

[3] M. Saidah, Nafiuna Hidayatus, “Implementasi algoritma optimasi bee colony untuk penjadwalan job shop,” *J. Tek. ITS Vol. 1*, no. October 2016, 2013.

[4] N. Azmi, I. Jamaran, Y. Arkeman, dan D. Mangunwidjaja, “Penjadwalan Pesanan Menggunakan Algoritma Genetika untuk Tipe Produksi Hybrid and Flexible Flowshop pada Industri Kemasan Karton,” *J. Tek. Ind.*, hal. 176–188.

[5] L. L. Liu, R. S. Hu, X. P. Hu, G. P. Zhao, dan S. Wang, “A hybrid PSO-GA algorithm for job shop scheduling in machine tool production,” *Int. J. Prod. Res.*, vol. 53, no. 19, hal. 5755–5781, 2015, doi: 10.1080/00207543.2014.994714.

[6] M. O. Okwu dan L. K. Tartibu, “Genetic Algorithm,” *Stud. Comput. Intell.*, vol. 927, hal. 125–132, 2021, doi: 10.1007/978-3-030-61111-8_13.

[7] T. Mahardhika, “Hybrid Algorithm as alternative method for optimization, a combination Genetic Algorithm and Particle Swarm Optimization,” *J. Phys. Conf. Ser.*, vol. 1764, no. 1, 2021, doi: 10.1088/1742-6596/1764/1/012040.

[8] R. Purwaningsih dan I. C. Fitriana, “Analisis Penjadwalan Produk PT Eksotika Logam Bali (DECO BALI) dengan Minimasi Makespan,” hal. 124–131, 2016.

Jurnal Elektronik Ilmu Komputer Udayana
Volume 11, No 3. February 2023

p-ISSN: 2301-5373
e-ISSN: 2654-5101

Klasifikasi Motif Kain Tradisional Cepuk Menggunakan GLCM dan KNN

Wayan Kiki Oktalao[a1], I Dewa Made Bayu Atmaja Darmawan[a2], I Wayan Santiyasa[b3], I Putu Gede Hendra Suputra [b4], I Gusti Ngurah Anom Cahyadi Putra [b5]

[a]Department of Math and Science, Udayana University
South Kuta, Badung, Bali, Indonesia
[1]kiki.oktalao@cs.unud.ac.id
[2]dewabayu@unud.ac.id
[3]made.agung@unud.ac.id
[4]santiyasa@unud.ac.id
[5]hendra.suputra@unud.ac.id
[6]anom.cp@unud.ac.id

Abstract

Cepuk weaving is one of the typical woven fabrics from the Balinese area, precisely in Tanglad Village, Nusa Penida District, Klungkung Regency, which is usually used by the people of Nusa Penida for ceremonial/ritual needs, such as cutting teeth, cremation, melukat to daily clothing needs. For generations, Tanglad has six types of cepuk, namely Mekawis, Amethyst, Lingking Paku, Tangi Gede, Sudamala, and Kurung. Each type has a different motif with a distinctive color. Therefore, this study was conducted to determine whether the selection of features affects the accuracy value resulting from the system testing carried out, as well as introducing cepuk fabrics to the general public using AI that can classify types of cepuk woven fabrics using the (K-Nearest Neighbor) KNN method, after performing the feature extraction using (Grey Level Co-occurrence Matrix) GLCM method. The features taken are Contrast, Energy, Entropy, Homogeneity, Dissimilarity, ASM (Angular Second Moment), and IDM (Inverse Differential Moment), with variations in the angle of 0º, 45º, 90º, 135º. Based on research conducted on 44 testing data with 11 data for each class, the results obtained are 91.7% accuracy using the parameter value of k = 3.

Keywords: *Cepuk, Classification, Feature Extraction, GLCM, KNN*

1. Pendahuluan

Indonesia adalah negara memiliki beragam warisan budaya seperti kesenian, musik, tarian, masakan, sastra, Bahasa daerah serta pakaian. Dari segi pakaian, Indonesia juga memiliki pakaian tradisional daerah diantaranya ada kain batik, kain songket, kebaya, koteka ataupun kain tenun yang merupakan pakaian tradisional khas Indonesia yang paling terkenal. Tenun merupakan kerajinan berbentuk kain dan terbuat dari benang dengan metode menggabungkan benang secara vertikal dan horizontal. Tenun tradisional khas Indonesia berasal dari berbagai daerah antara lain Bali, Sumatera, NTT, Jawa, dan NTB. Bali memiliki aturan untuk mengenakan Endek pada hari yang telah ditetapkan di daerah Bali (Surat Edaran Nomor 04 Tahun 2021), namun selain kain Endek, ada beberapa jenis tekstil atau kain tenun seperti kain Gringsing dan kain Cepuk khas Nusa Penida.

Kain cepuk merupakan salah satu kerajinan khas dari Tanglad, Kec. Nusa Penida, Kab. Klungkung, Bali, yang diwariskan melalui nenek moyang dari generasi ke generasi. Sejarah dari nama cepuk sendiri berasal dari bahasa Sanskerta, yaitu cepuk artinya kayu canging. Kayu canging adalah jenis tumbuhan yang digunakan sebagai bahan dasar membuat kain tenun. Selain itu cepuk juga asal katanya yaitu tepuk yang berarti bertemu, tiap motif pada kain cepuk selalu saling bertemu, diantaranya ada yang membentuk geometris belah ketupat. Disisi lain warna yang dipakai untuk benang kain cepuk memiliki simbol-simbol penjuru mata angin menurut kepercayaan masyarakat bali. Warna kuning

terletak di barat yang melambangkan Dewa Mahadewa, Merah terletak di selatan yang melambangkan Dewa Brahma, putih terletak di timur lambang dari Dewa Iswara, hitam terletak di utara yang melambangkan Dewa Wisnu, dan campuran keseluruhan warna tersebut yang melambangkan Dewa Siwa yang terletak di tengah. Karenanya kain cepuk biasanya dipakai Ketika mengadakan ritual keagamaan, Contohnya digunakan sebagai kain (pakaian yang dililitkan), selendang, tempat persembahyangan, serta ornamen bangunan upacara [1].

Dari generasi ke generasi daerah Tanglad mempunyai enam jenis kain cepuk, yaitu Mekawis, Kecubung, Lingking Paku, Tangi Gede, Sudamala, dan Kurung. Setiap jenis memiliki motif yang berbeda dengan warna yang khas. Adapun penelitian yang mengangkat kain cepuk sebagai topik penelitian yaitu, tentang ekspansi tenun cepuk sebagai penunjang daya tarik wisata kebubudayaaan di Nusa Penida [2], perancangan film dokumenter nilai makna serta peranan kain tenun endek dan cepuk di bali [3], namun penelitian tersebut hanya membahas tentang pengembangan daya tarik wisata terhadap kain cepuk dan pembuatan film dokumenter tentang pengenalan kain cepuk.

Oleh karena itu dikerjakannya penelitian ini untuk mengetahui apakah pemilihan fitur mempengaruhi hasil akurasi pada sistem yang dibangun dan juga untuk memperkenalkan kain cepuk dengan pemanfaatan AI. Adapun penelitian yang membahas tentang klasifikasi kerajinan kain dilakukan oleh [4] yang membahas mengenai klasifikasi kain tenun berdasarkan tekstur dan warna menggunakan metode KNN (K-Nearest Neighbour), Pengenalan pola motif batik Sleman dengan metode GLCM (Grey Level Co-occurrence Matrix) [5], ekstraksi ciri GLCM atas KNN pada saat melakukan klasifikasi motif batik [6], serta ekstraksi ciri GLCM atas KNN dalam mengklasifikasi motif batik [7].

Pada penelitian ini akan dibuat sistem yang dapat mengklasifikasikan jenis-jenis dari kain tenun cepuk dengan metode KNN (K-Nearest Neighbour), untuk melakukan klasifikasi sebelumnya akan dilakukan ekstraksi fitur dengan metode GLCM (Grey Level Co-occurrence Matrix. Metode ini digunakan dalam pengenalan tekstur, segmentasi citra, analisis warna pada citra, klasifikasi citra, dan pengenalan objek [5]. Setelah proses klasifikasi maka diperoleh hasil klasifikasi jenis kain tenun cepuk tersebut. Dengan adanya penelitian ini diharapkan dapat membantu dalam mengklasifikasikan jenis kain tenun cepuk serta sistem yang dibangun dapat memberikan informasi untuk membantu dalam melestarikan dan mengenalkan budaya tentang kain cepuk.

2. Metode Penelitian

Penelitian dilakukan melalui beberapa langkah, diantaranya studi literatur, pengumpulan data, pengolahan data, implementasi, serta pengujian. Studi literatur dilakukan untuk mengumpulkan summber-sumber tulisan, menyusun permasalahan pada tulisan, serta menjadi acuan penelitian ini. Selanjutnya dilakukan pengumpulan data baik secara primer ataupun sekunder, pada penelitian ini, data yang dipakai merupakan data primer yang diambil langsung ke pengrajin di Desa Tanglad, Nusa Penida. Setelah mendapatkan data berupa dataset citra kain Cepuk, kemudian dilakukan preprocessing data dengan cara cropping, resize, dan conversi citra. Setelah tahap preprocessing, kemudian data diimplementasikan dengan metode ekstraksi ciri/fitur dan metode klasifikasi. Pada tahap akhir akan dilakukan pengujian pada data yang telah diimplementasikan untuk memperoleh akurasi.

Gambar 1. Denah Alur Metode Penelitian

2.1. Studi Literatur

Studi literatur/referensi merupakan runtutan aktivitas berkenaan dengan metode pengumpulan data, melalui membaca, mencatat, pustaka, mengolah subjek penelitian, dan studi literatur. Bertujuan untuk mencari referensi tertulis, seperti buku, majalah, artikel/jurnal, majalah, maupun dokumen/arsip yang relevan terhadap kajian masalah. Sehingga data yang diperoleh melalui studi literatur bisa dijadikan sitasi/rujukan dalam memperkuat tulisan yang telah dibuat.

2.2. Pengumpulan Data

Data untuk penelitian ini adalah data primer, yang didapat melalui pengambilan gambar kain cepuk di Desa Tanglad, Nusa Penida dengan masing-masing jenis kain. Data yang diambil kemudian

p-ISSN: 2301-5373
e-ISSN: 2654-5101

dibagi dua yaitu, data training/latih dan data testing/uji dengan jumlah perbandingan pada data set 75% data training serta 25% data testing atau 3:1. Kain yang akan diambil gambarnya dibentangkan dengan arah sudut yang sama pada setiap jenis kain, kemudian ditentukan jarak yang tepat dan diambil gambarnya dengan memperhatikan tekstur (pola garis, titik maupun bentuk gambar pada kain) agar tekstur pada kain didapatkan dengan tepat, masing-masing akan diambil gambar untuk 4 jenis kain dan untuk masing-masing gambar akan dilakukan preprocessing yaitu tahap cropping untuk mendapatkan data dengan pola yang tepat untuk mewakili setiap jenis kainnya. Pada penelitian ini data latih akan difokuskan untuk klasifikasi 4 jenis kain cepuk pada gambar 2, yaitu kain cepuk kecubung (1), liking paku (2), mekawis (3) dan kain cepuk kurung atau biasa (4) [8].

Gambar 2. Contoh Kain Cepuk Khas Nusa penida

2.3. Pengolahan Data

Tahap *preprocessing* dilakukan bertujuan untuk mempermudah dalam mengolah dan mendapatkan citra final yang akan digunakan untuk mendapatkan fitur/ciri dari citra. Adapun tahapan yang akan dilakukan, yaitu:

2.3.1. *Cropping*

Pada tahap ini gambar yang telah diambil di hari dan dengan tiga device yang berbeda selanjutnya akan dipotong/cropping sebagai tahap awal *preprocessing* [9], untuk diambil bagian yang akan difokuskan untuk dijadikan sebagai data latih ataupun data uji, hal ini dilakukan untuk mereduksi volume data citra agar mempermudah pemrosesan, pemotongan citra akan dilakukan secara manual dengan memperhatikan pola tekstur pada setiap kain.

2.3.2. *Resize*

Pada tahap ini Resize dilakukan sebelum melakukan cropping yang bertujuan untuk mengubah *size* citra arah horizontal atau vertikal menjadikan *size* yang telah ditentukan yaitu 500x500 piksel [10]. Hal ini bertujuan menyeragamkan ukuran citra/gambar kain yang akan dipakai saat melakukan proses pelatihan dan pengujian.

2.3.3. Conversi Citra

Pada tahap ini citra RGB akan dikonversi menjadi *grayscale* yang nanti akan berguna dalam proses ekstraksi ciri/fitur dengan metode GLCM (*Gray Level Co-occurrence Matrix*) [11]. Tahapan ini dilakukan sebelum tahap ekstraksi ciri/fitur dengan Bahasa pemrograman python.

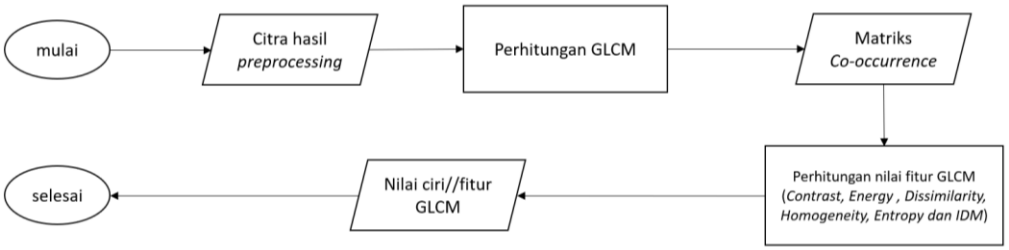
2.4. Implementasi Data

Tahap implementasi data dilakukan setelah data melalui tahap *preprocessing* yang menghasilkan dataset yang siap digunakan untuk tahap implementasi sistem [12]. Adapun proses yang dilalui data pada tahap ini yaitu:

2.4.1. Ekstraksi Ciri/Fitur

Tahap ini dilakukan ekstraksi ciri dari citra yang sudah melewati tahap *preprocessing*. Tujuan dilakukan tahap ini dengan metode GLCM (*Gray Level Co-occurrence Matrix*) yaitu untuk menganalisis serta membedakan tekstur pola atau bentuk pada motif kain cepuk, yang akan dibedakan dengan fitur-fitur yang ada pada GLCM seperti *Contrast*, *Energy*, *Entropy*, *Homogeneity*, *Dissimilarity, Angular Second Moment* (ASM) dan *Inverse Differential Moment* (DM).

Tekstur merupakan kesesuaian pola-pola tertentu pada susunan piksel-piksel dalam citra digital. Salah satu bagian penting dalam analisis tekstur yaitu dengan matriks pasangan intensitas (Gray Level Co-occurence Matrix) adalah matriks keterkaitan dua dimensi[13]. Tekstur juga bisa dikatakan sebagai karakteristik intrinsik suatu citra yang terkait pada tingkat kekerasan (roughness), keteraturan (regularity), dan granulitas (granulation) susunan struktural pada piksel [14].



Gambar 3. Alur Proses Ekstraksi Ciri/Fitur

2.4.2. Klasifikasi

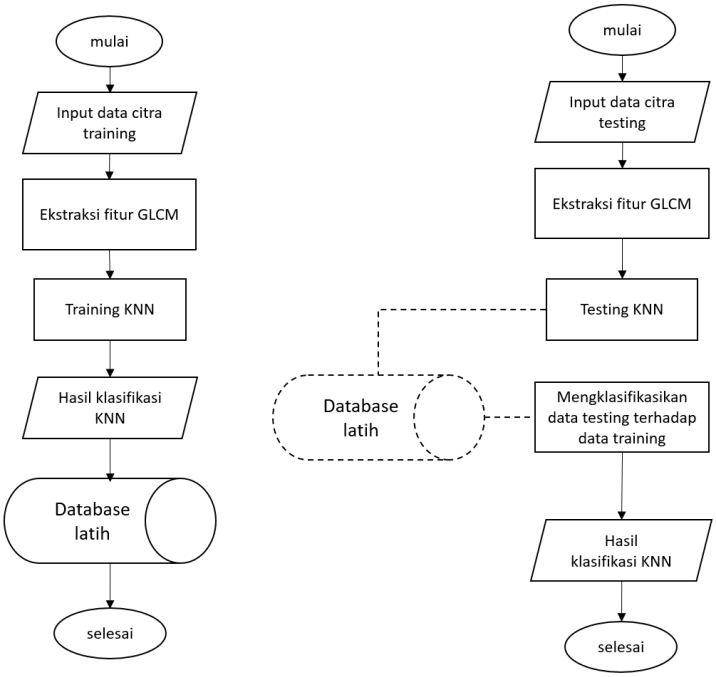
Penelitian ini menggunakan metode pengukuran kemiripan atau klasifikasi yang sederhana yaitu metode K-Nearest Neighbour (KNN). Untuk menghitung jarak ketetanggaan antara dua titik dari data training dan data testing menggunakan rumus Euclidean [15]. Kelas jenis kain yang digunakan pada penelitian ini yaitu kain cepuk kecubung, liking paku, mekawis dan kain cepuk kurung atau biasa, deskripsi pada setiap kelas yang terlihat pada tabel 1.

Tabel 1. Jenis Kain

Jenis Kain	Deskripsi
Cepuk Kecubung	Digunakan anak perempuan saat upacara potong gigi (satu paket motif terdiri dari bunga gede-kurung-bunga gede ditambah apit gunung, panggeh taji, pacit genggong, dan mata titiran).
Cepuk Liking Paku	Digunakan anak laki-laki ketika upacara potong gigi (motif garis mata titiran diganti dengan pancit genggong).
Cepuk Mekawis	Kain ini digunakan sebagai pembungkus tulang pada upacara pengabenan/kematian (motif bebas (kecubung atau kurung)).
Cepuk Kurung	Bebas digunakan oleh masyarakat umum dan bisa dimodifikasi (motif hampir sama dengan kecubung, namun hanya kurung, tidak ada bunga gede).
Cepuk Sudamala	Digunakan ketika melukat (membersihkan diri), hampir sama dengan kecubung berwarna hitam-putih.
Cepuk Tangi Gede	Berfungsi untuk sanan empeg, yang digunakan ketika upacara anak ke-2 dari tiga bersaudara jika kakak pertama dan adik ketiganya meninggal (motif hampir sama seperti kecubung, tetapi pada bagian kurung berwarna hitam).

Untuk flowchart alur proses klasifikasi menggunakan metode KNN (*K-Nearest Neighbour*) yang terlihat pada gambar 4. Pada gambar tersebut terdapat dua flowchart yang hampir serupa, karena flowchart yang kiri merupakan alur klasifikasi untuk data latih sedangkan yang kanan adalah flowchart klasifikasi data uji, yang telah memiliki perbandingan atau model system dari data training sebelumnya.

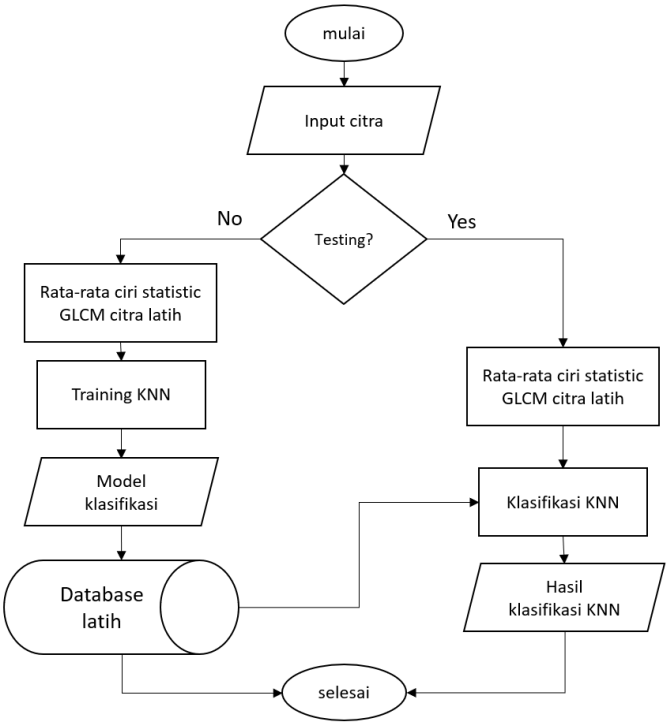
p-ISSN: 2301-5373
e-ISSN: 2654-5101



Gambar 4. Alur Proses Klasifikasi

2.5. Pengujian

Pengujian dilakukan untuk mengukur kinerja dari metode klasifikasi yaitu metode KNN. Pada pengujian ini akan diuji akurasi dari sistem yang telah dibuat dengan parameter uji yaitu jumlah data yang diklasifikasikan sesuai dengan kelas yang ditentukan dan dibagi dengan jumlah data yang diuji dikali dengan 100, dengan melakukan pengujian untuk mengetahui bagaimana pengaruh fitur GLCM terhadap hasil dengan mengidentifikasi pola dan diklasifikasikan dengan KNN. Pada gambar 5 menampilkan alur pengujian system, berfungsi untuk memprediksi jenis kain yang telah diinput oleh user baik data latih maupun data uji.



Gambar 5. Alur Proses Klasifikasi

Tabel 2 merupakan tabel skenario uji yang menampilkan hasil pengujian fitur-fitur terhadap nilai akurasi yang dihasilkan.

Tabel 2. Skenario Uji Akurasi Fitur

Fitur yang Dipakai	Fitur Uji 1	Fitur Uji 2	Fitur Uji 3
Akurasi	Nilai Akurasi	Nilai Akurasi	Nilai Akurasi

3. Hasil dan Pembahasan

Bagian hasil akan membahas mengenai pengujian sistem yang telah dilakukan sebelumnya. Pembahasan ini memuat tentang pengujian akurasi terhadap pemilihan fitur pada metode GLCM. Pengujian akurasi dipakai untuk mengukur akurasi dari system, yaitu dengan cara mencocokkan hasil yang didapat melalui perhitungan sistem. Berdasarkan dari perankingan data yang dilakukan melalui proses perhitungan pada sistem, diperoleh prediksi jenis-jenis kain yang diinputkan kemudian dilakukan perbandingan untuk mengetahui tingkat keakuratan saat menentukan prediksi jenis kain [16].

3.1. Hasil

Setelah dilakukan pengujian untuk memperoleh akurasi dengan metode klasifikasi KNN dan metode ekstraksi fitur GLCM dengan melakukan tiga kali pengujian yang masing-masing pengujian menggunakan fitur yang berbeda-beda, maka dihasilkan akurasi seperti pada tabel 3. Fitur yang diuji adalah fitur contras, yang merupakan fitur yang nilainya dipengaruhi oleh intensitas cahaya, sehingga jika fitur tersebut digunakan maka akan terlihat apakah intensitas cahaya mempengaruhi hasil dari akurasi sistem atau tidak.

Tabel 3. Hasil Pengujian Akurasi Fitur

Fitur yang Dipakai	Akurasi
contras, dissimilarity, homogeneity, IDM, entropy, labels	k1 : 0.6477272727272727 k3 : 0.5568181818181818 k5 : 0.5681818181818182 k7 : 0.5625 k9 : 0.5511363636363636 k11 : 0.5568181818181818 k13 : 0.5909090909090909 k15 : 0.625 k17 : 0.6477272727272727 k19 : 0.625
contras, energy, dissimilarity, homogeneity, IDM, entropy, ASM, labels	k1 : 0.6477272727272727 k3 : 0.5568181818181818 k5 : 0.5681818181818182 k7 : 0.5625 k9 : 0.5511363636363636 k11 : 0.5568181818181818 k13 : 0.5909090909090909 k15 : 0.625 k17 : 0.6477272727272727 k19 : 0.625
energy, dissimilarity, homogeneity, IDM, entropy, ASM, labels	k1 : 1.0 k3 : 0.9166666666666666 k5 : 0.8276515151515151 k7 : 0.8049242424242424 k9 : 0.7897727272727273 k11 : 0.7481060606060606 k13 : 0.6988636363636364 k15 : 0.6666666666666666 k17 : 0.6553030303030303 k19 : 0.6534090909090909

Tabel 3 memperlihatkan hasil pengujian yang telah dilakukan, yaitu setiap fitur yang digunakan menghasilkan akurasi yang berbeda terutama pada fitur *contrast* yang membuat akurasi dari sistem menurun, dikarenakan data yang digunakan diambil pada waktu yang berbeda dan membuat intensitas cahaya dari tiap data berbeda pula, sehingga nilai yang dihasilkan fitur contrast berbeda-beda pada jenis kain yang sama.

3.2. Pembahasan

Setelah dilakukan pengujian, diperoleh hasil berupa nilai fitur, dan nilai akurasi yang berbeda. Seperti pada tabel 3, yaitu untuk fitur yang menggunakan contras sebagai salah satu fiturnya memperoleh nilai akurasi terbaik 64,8% untuk nilai k = 1 dan k = 17 serta untuk nilai akurasi yang diperoleh dari hasil pengujian tanpa fitur contras dan ditambah dengan fitur energy yaitu 100% untuk nilai k = 1 dan 91,7% untuk nilai k = 3. Namun pada saat fitur energy dan contras digunakan bersamaan menghasilkan nilai akurasi yang sama dengan pengujian hanya menggunakan fitur contras.

4. Kesimpulan

Berdasarkan hasil penelitian yang dilakukan menggunakan metode ektraksi fitur GLCM dan metode klasifikasi KNN, maka diperoleh hasil bahwa pemilihan fitur berpengaruh terhadap nilai akurasi pada sistem yang telah dibangun dibuktikan dengan adanya perubahan pada nilai akurasi yang diperoleh ketika menggunakan jenis fitur yang berbeda, dengan adanya penurunan akurasi saat menggunakan fitur *contrast*. Untuk akurasi terbaik yang dapat dihasilkan oleh sistem yang dibangun yaitu 100% untuk k = 1 dan 91,7% untuk nilai k = 3 dengan menggunakan fitur *Energy*, *Entropy*, *Homogeneity*, *Dissimilarity, Angular Second Moment* (ASM) dan *Inverse Differential Moment* (IDM).

References

[1] N. L. W. Sayang Telagawathi, “Pelatihan dan Pendampingan Manajemen Usaha Kelompok Perajin Tenun Endek di Desa Sulang Klungkung,” *Proceeding TEAM*, vol. 2, p. 687, 2017, doi: 10.23887/team.vol2.2017.208.

[2] F. L. Amir, “Pengembangan Kain Tenun Cepuk Sebagai Pendukung Daya Tarik Wisata Budaya Di Nusa Penida,” *J. Master Pariwisata*, vol. 4, pp. 327–339, 2018, doi: 10.24843/JUMPA.2018.v04.i02.p12.

[3] J. I. Tedja, D. T. Ardianto, and P. B. Setyawan, “KAIN TENUN ENDEK DAN CEPUK DI BALI Abstrak Pendahuluan Metode Penelitian Metode Analisis Data Rumusan Masalah Landasan Teori Tujuan.”

[4] Kevin, J. Hendryli, and D. E. Herwindiati, “Klasifikasi kain tenun berdasarkan tekstur & warna dengan metode K-NN,” *J. Comput. Sci. Inf. Syst.*, vol. 3, no. 2, pp. 85–95, 2019.

[5] M. Metode and G. Level, “Ekstraksi citra fitur pada pengenalan pola motif batik sleman menggunakan metode gray level co-occurrence matrix,” vol. 5, pp. 3–6, 2019.

[6] C. Jatmoko and D. Sinaga, “Ektraksi Fitur Glcm Pada K-Nn Dalam Mengklasifikasi Motif Batik,” pp. 978–979, 2019.

[7] R. Dani, A. Sugiharto, and G. A. Winara, “Aplikasi Pengolahan Citra Dalam Pengenalan Pola Huruf Ngalagena Menggunakan MATLAB,” *Konf. Nas. Sist. Inform.*, pp. 772–777, 2015.

[8] A. Kurnianingsih and widayanti Arioka, *Menenun Waktu (Kisah Tradisi Tenun di Tenganan Pagringsingan, Sidemen, dan Tanglad Nusa Penida)*. bali: wisnu press, 2019.

[9] A. W. Bawono, I. B. Hidayat, S. Nugroho, and S. Si, “Deteksi Area Hutan Berbasis Citra Google Earth Menggunakan Metode Grey Level Co-Occurrence Matrix (GLCM) dan Support Vector Machine (SVM),” vol. 6, no. 1, pp. 524–530, 2019.

[10] Y. Kusumawati, A. Susanto, I. Utomo, W. Mulyono, and D. P. Prabowo, “KLASIFIKASI BATIK KUDUS BERDASARKAN POLA MENGGUNAKAN K-NN DAN,” pp. 509–514, 2020.

[11] I. Amalia, “Ekstraksi Fitur Citra Songket Berdasarkan Tekstur Menggunakan Metode Gray Level Co-occurrence Matrix (GLCM),” *J. Infomedia*, vol. 3, no. 2, pp. 64–68, 2018, doi: 10.30811/jim.v3i2.715.

[12] A. J. T, D. Yanosma, and K. Anggriani, “Implementasi Metode K-Nearest Neighbor (Knn) Dan Simple Additive Weighting (Saw) Dalam Pengambilan Keputusan Seleksi Penerimaan Anggota Paskibraka,” *Pseudocode*, vol. 3, no. 2, pp. 98–112, 2017, doi: 10.33369/pseudocode.3.2.98-112.

[13] K. A. Nugraha, W. Hapsari, and N. A. Haryono, “Analisis Tekstur Pada Citra Motif Batik Untuk Klasifikasi K-NN,” *Informatika*, vol. 10, no. 2, pp. 135–140, 2014.

[14] W. P. (Universitas S. D. Adnyana, “Pengenalan Tekstur Dengan Statistical Texture Descriptor,” 2018.

[15] Y. I. N, A. Ana, and D. Permatasari, “Pengenalan Pembicara untuk Menentukan Gender Menggunakan Metode MFCC dan VQ,” *MIND J.*, vol. 2, no. 1, pp. 34–47, 2018, doi: 10.26760/mindjournal.v2i1.34-47.

[16] N. N. Dzikrulloh and B. D. Setiawan, “Penerapan Metode K – Nearest Neighbor (KNN) dan Metode Weighted Product (WP) Dalam Penerimaan Calon Guru Dan Karyawan Tata Usaha Baru Berwawasan Teknologi (Studi Kasus : Sekolah Menengah Kejuruan Muhammadiyah 2 Kediri),” *Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 1, no. 5, pp. 378–385, 2017.

This page is intentionally left blank.

Jurnal Elektronik Ilmu Komputer Udayana
Volume 11, No 3. February 2023

p-ISSN: 2301-5373
e-ISSN: 2654-5101

Klasifikasi Penyakit Paru-Paru Berdasarkan Suara Paru-Paru Menggunakan Metode MFCC dan SVM

Ni Putu Subhasini Dewi Sukma[a1], I Gede Arta Wibawa[a2], I Gusti Ngurah Anom Cahyadi Putra[a3], I Gusti Agung Gede Arya Kadyanan[a4], I Gede Santi Astawa[a5], Agus Muliantara[a6]

Program Studi Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Udayana
Bukit Jimbaran, Badung, Bali, Indonesia
[1]subhadewi59@gmail.com
[2]gede.arta@unud.ac.id
[3]anom.cp@unud.ac.id
[4]gungde@unud.ac.id
[5]santi.astawa@unud.ac.id
[6]muliantara@unud.ac.id

Abstract

The lungs are vital organs that function as a place for oxygen exchange. The lungs are one of the organs that have an important role for humans to survive. The importance of maintaining lung health to avoid dangerous diseases. If the health of the lungs is not taken care of properly, it will make the lungs unable to function normally and can cause disturbances in the respiratory system. The types of lung diseases that are suffered by many people in the world such as cough, asthma, bronchitis, tuberculosis, and so on. Therefore, this study will classify lung diseases based on lung sounds using the MFCC method which will be used at the feature extraction stage and the SVM method used at the data classification stage. The testing phase used is K-Fold Cross Validation and also Confusion Matrix to determine the amount of data that is identified correctly and to determine the level of accuracy generated. After going through several processes to classify the lungs based on the sound of the lungs, the accuracy rate is quite good, which is 80%.

Keywords: *Classification, Lung Sound, Lung Disease, MFCC, SVM, Accuracy*

1. Pendahuluan

Pada sistem pernafasan manusia terdapat beberapa macam organ penting yang bekerja dengan fungsinya masing-masing dalam proses pernafasan. Paru-paru merupakan bagian dari sistem pernafasan manusia yang bekerja sebagai tempat pertukaran oksigen. Bagi manusia kesehatan paru-paru sangat penting untuk diperhatikan dengan baik. Setiap orang perlu untuk mengetahui gejala penyakit paru-paru guna menghindari munculnya penyakit yang dapat membahayakan kesehatan manusia terutama kesehatan paru-paru sehingga dapat berakibat fatal bahkan hingga menyebabkan kematian.

Penyakit paru-paru adalah suatu kondisi yang menyebabkan paru-paru tidak dapat bekerja dengan baik atau normal. Gejala penyakit yang dianggap biasa saja ketika seseorang sedang mengalami kondisi seperti batuk berdahak, bau mulut, demam, berat badan turun, dan lainnya yang ada kemungkinan beberapa gejala yang disebutkan adalah tanda-tanda dari penyakit paru-paru [8]. Di kehidupan sehari-hari terdapat banyak yang menderita penyakit paru-paru seperti bronkitis, TBC, asma, hingga batuk serta demam. Adapun penyebab dari penyakit paru-paru seperti polusi udara, virus, bakteri, dan sebagainya [5].

Beberapa penyakit paru-paru memiliki indikasi yang sama, maka agar dapat mengetahui macam-macam penyakit yang dialami penderita, diperlukan seorang dokter yang ahli dalam ilmu kesehatan paru-paru. Seorang dokter dalam melakukan pemeriksaan pada seorang pasien biasanya menggunakan alat berupa stetoskop untuk mendengarkan suara pernafasan paru-paru. Suara paru-paru merupakan suatu informasi penting yang dibutuhkan oleh seorang dokter dalam menentukan tingkat kesehatan pernafasan seseorang [4].

Beberapa peneliti sebelumnya telah melakukan penelitian dengan mengembangkan berbagai metode dalam mendiagnosis penyakit paru-paru melalui suara paru-paru. Peneliti sebelumnya melakukan penelitian dengan menggunakan *Autoencoder* yang merupakan salah satu arsitektur DNN dan diuji oleh SVM untuk mengklasifikasi suara paru-paru yang normal dan abnormal menghasilkan tingkat

akurasi sejumlah 82,38% [1]. Terdapat penelitian yang melakukan pengelompokan suara paru-paru menggunakan CNN, di mana penelitian ini menyinggung data suara paru-paru yang berupa spektrogram suara paru-paru yang kemudian diklasifikasi dengan CNN yang memperoleh tingkat keakuratan yang didapat sejumlah 74% [7]. Sebuah penelitian menggunakan *Probabilistic Neural Network* dan SVM *One Against All*, SVM *One Against One* dalam membandingkan suara paru-paru yang normal dan abnormal dengan memperoleh tingkat keakuratan rata-rata SVM *One against all* sejumlah 47,55%, dengan tingkat ketepatan sejumlah 70%. Akurasi rata-rata SVM *One Against One* sejumlah 50.92%, dengan tingkat ketepatan sejumlah 75%. Akurasi rata-rata *Probabilistic Neural Network* sejumlah 70%, dengan tingkat ketepatan sejumlah 70% [3].

Pada penelitian ini berfokus untuk meneliti dan mencoba metode MFCC dan SVM untuk klasifikasi penyakit paru-paru berdasarkan suara paru-paru. Beberapa tahapan penelitian ini yaitu tahap akuisisi data, selanjutnya akan dilakukan tahap ekstraksi ciri menggunakan metode MFCC, kemudian mengklasifikasi data menggunakan metode SVM, dan dilakukan tahap pengujian dengan menggunakan *K-Fold Cross Validation* dan Matriks konfusi. Terakhir akan mendapatkan hasil berupa tingkat akurasi.

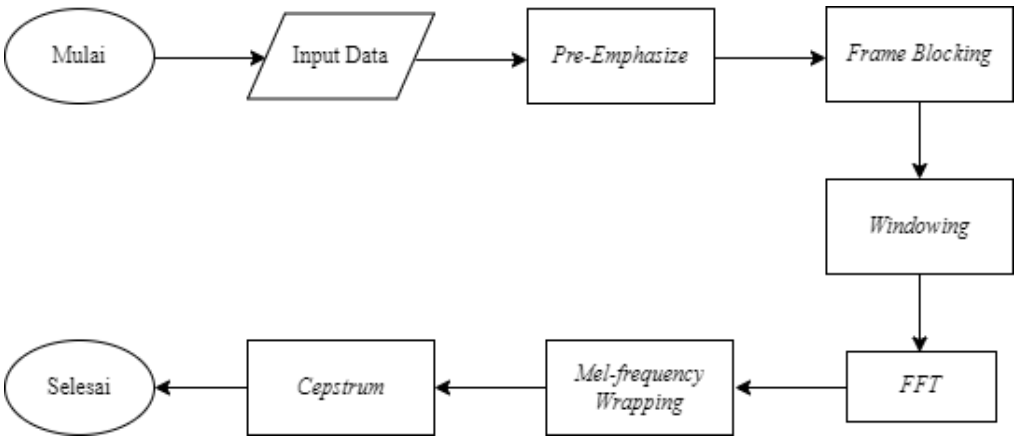
2. Metode Penelitian

2.1 Akuisi Data

Dataset yang digunakan yaitu data sekunder yang didapat dari website ICBHI 2017. Dataset ini berisi file rekaman suara paru-paru dengan format .wav. Dataset tersebut diambil dari 126 subjek yang berisi rekaman suara paru-paru sebanyak 6898 siklus pernafasan, di mana dataset akan dibagi menjadi 2 yaitu data *training* dan data *testing* dengan proporsi data adalah 80:20.

2.2 Ekstraksi Ciri

Ekstraksi ciri merupakan proses yang digunakan untuk mengekstraksi ciri atau fitur dari suatu objek yang dapat membuat karakteristik dari objek tersebut. Pada tahap ekstraksi ciri menggunakan metode MFCC. MFCC merupakan metode yang digunakan untuk representasi fitur sinyal suara dalam bentuk angka [10].



Gambar 1. Tahap Ekstraksi Ciri

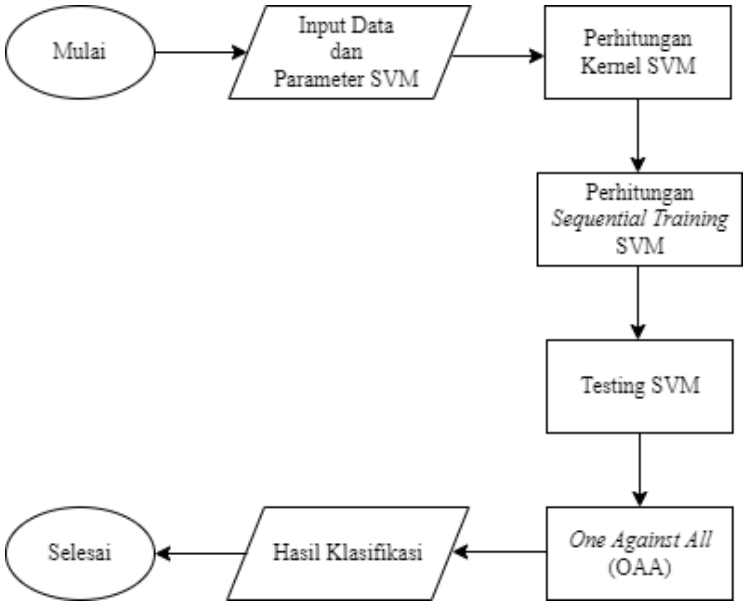
Gambar 1 merupakan tahap ekstraksi ciri MFCC, di mana akan dilakukan beberapa tahap. Diawali dengan tahap *pre-emphasize* dilakukan untuk meningkatkan frekuensi tinggi pada spektrum. Selanjutnya tahap *frame blocking* dilakukan untuk memotong sinyal audio menjadi beberapa *frame*. Kemudian berikutnya tahap *windowing* dilakukan untuk meminimalisasi terjadinya efek diskontinuitas dari potongan-potongan sinyal. Dilanjutkan ke tahap FFT, di mana tahap ini dilakukan guna mengubah domain waktu sinyal ke domain frekuensi. Selanjutnya tahap *mel-frequency wrapping* dilakukan untuk penyaringan atau filter dari spectrum pada setiap frame. Tahap terakhir adalah tahap *cepstrum* dilakukan untuk mendapatkan koefisien MFCC dengan mengubah *mel spectrum* menjadi domain waktu.

2.3 Klasifikasi

Klasifikasi merupakan proses mendeteksi model yang mendeskripsikan dan memilah suatu kelas data yang bertujuan supaya model tersebut dapat berfungsi untuk memprediksi kelas objek yang label pada kelasnya tidak diketahui [7]. Pada tahap klasifikasi menggunakan metode SVM yang bertujuan untuk

p-ISSN: 2301-5373
e-ISSN: 2654-5101

mendapatkan hyperplane pemisah klaster data yang memaksimalkan margin antar support vector [2]. Data yang telah diperoleh fiturnya pada saat tahap ekstraksi ciri maka akan dilanjut ke tahap klasifikasi.



Gambar 2. Tahap Klasifikasi

Gambar 2 merupakan tahap klasifikasi SVM, di mana diawali dengan tahap menginput data dan menentukan parameter yang akan digunakan, pada penelitian ini menggunakan parameter *Complexity C* dan γ (gamma). Kemudian dilanjutkan dengan menghitung kernel SVM, di mana pada penelitian ini menggunakan kernel *Radial Basis Function* (RBF), *Linear*, dan *Polynomial*. Berikutnya dilakukan perhitungan *Sequential Training* SVM untuk proses *training* dan selanjutnya dilakukan proses *testing* SVM pada data uji. Kemudian dilakukan tahap *one against all* guna mengatasi permasalahan *multi class*. Terakhir akan mengeluarkan hasil klasifikasi.

2.4 Pengujian

Tahap pengujian menggunakan metode *K-Fold Cross Validation* yang bertujuan untuk mengukur ketepatan hasil analisis generalisasi pada dataset *independent* [6] dan Matriks konfusi merupakan metode yang diperlukan untuk mengukur akurasi dalam konsep data *mining* [9]. Pada tahap *K-Fold Cross Validation* akan dilakukan pembagian terhadap seluruh data menjadi beberapa bagian untuk *testing* dan *training*. Di mana pengujian akan dilakukan percobaan sebanyak 5 tahapan (5 *Fold CV)*. Kemudian hasil dari 5 *Fold Cross Validation* ini akan dicatat nilai evaluasi performa dari model tersebut menggunakan Matriks Konfusi untuk dapat mengetahui jumlah data yang teridentifikasi benar serta mengetahui akurasi sistem tersebut.

Tabel 1. Matriks Konfusi

Matriks Konfusi		Kelas Hasil Prediksi	
		Positif	Negatif
Kelas Asli	Positif	TP	FN
	Negatif	FP	TN

Adapun 4 *output* yang dihasilkan dari matriks ini yaitu *recall*, *precission*, *f1-score* dan *accuracy*. Rumus untuk menghitung 4 *output* tersebut dapat dilihat pada persamaan berikut.

$$Recall = \frac{TP}{TP+FN} \times 100\% \quad (1)$$

$$Precission = \frac{TP}{TP+FP} \times 100\% \quad (2)$$

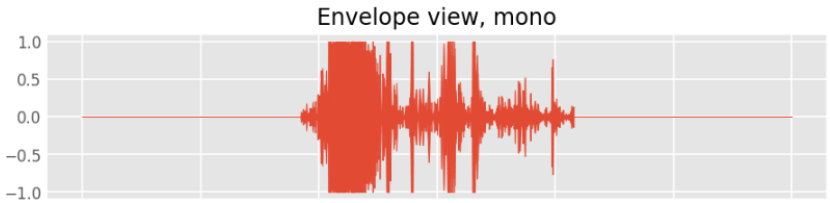
$$F1 - score = \frac{1}{F1} = \frac{1}{2}\left(\frac{1}{precision} + \frac{1}{recall}\right) \times 100\% \quad (3)$$

$$Accuracy = \frac{TP+TN}{TP+FN+FP+TN} \times 100\% \quad (4)$$

Di mana,
TP (*True Positive*) = jumlah data positif yang tergolong benar
TN (*True Negative*) = jumlah data negatif yang tergolong salah
FP (*False Positive*) = jumlah data positif yang tergolong benar
FN (*False Negative*) = jumlah data negatif yang tergolong salah

3. Hasil dan Pembahasan

Penelitian ini menggunakan dataset rekaman suara paru-paru yang didapat dari website ICBHI 2017. Rekaman suara paru-paru disimpan dalam format .wav.

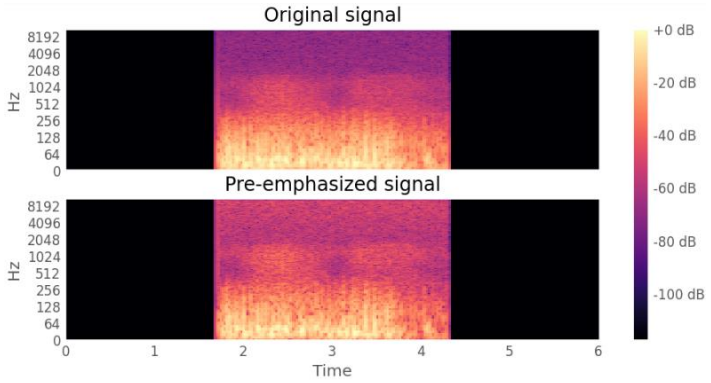


Gambar 3. Grafik Sinyal Suara Paru-Paru

Gambar 3 merupakan salah satu grafik sinyal suara paru-paru dari rekaman audio pada dataset. Data rekaman suara ini diperoleh dari 126 subjek yang berisi 6898 siklus pernafasan paru-paru. Di mana suara tersebut terbagi menjadi 1864 suara *crackle*, 886 suara *wheeze*, dan 506 gabungan dari suara *crackle* dan *wheeze*. Data akan dibagi menjadi data *train* dan data *test* dengan proporsi data *train* sejumlah 80% dan data *test* sejumlah 20%.

Kemudian dilakukan tahap ekstraksi ciri menggunakan *Mel Frequency Cepstral Coefficient* (MFCC) dengan hasil akhir berupa koefisien MFCC yang selanjutnya akan digunakan pada tahap klasifikasi. Tahap pertama dalam ekstraksi ciri MFCC ini dilakukan tahap *pre-emphasize. Output* dari tahap *pre-emphasize* dapat dilihat pada gambar 4.

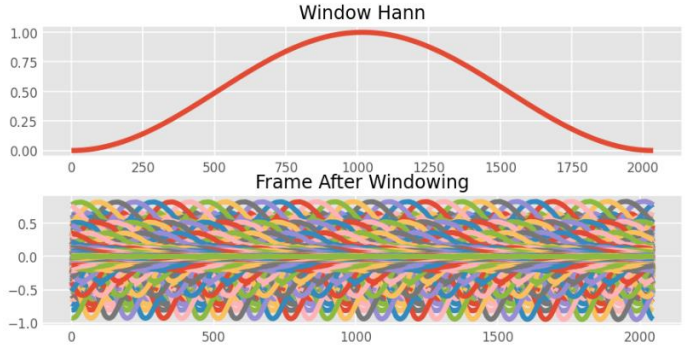
p-ISSN: 2301-5373
e-ISSN: 2654-5101



Gambar 4. Output Tahap Pre-Emphasize

Gambar 4 merupakan *output* dari tahap *pre-emphasize* guna meningkatkan energi sinyal pada frekuensi tinggi.

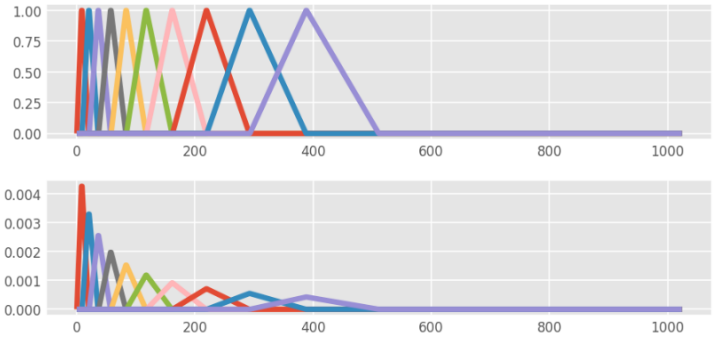
Tahap berikutnya yaitu tahap *frame blocking*, di mana sinyal suara akan disegmentasi mejadi beberapa *frame* yang kemudian akan dilanjutkan pada tahap *windowing*. *Output* dari tahap *windowing* dapat dilihat pada gambar 5.



Gambar 5. Output Tahap Windowing

Gambar 5 merupakan *output* dari tahap *windowing* guna mengurangi atau meminimalisasi efek diskontinuitas dari potongan-potongan sinyal tersebut.

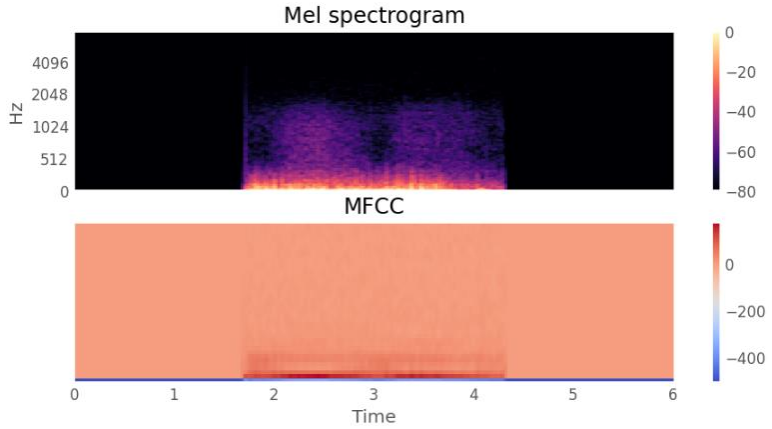
Selanjutnya adalah tahap FFT, di mana sinyal akan diubah dari domain waktu menjadi domain frekuensi. Setelah sinyal diubah ke dalam domain frekuensi, berikutnya akan masuk pada tahap *mel-frequency wrapping*. *Output* dari tahap *mel-frequency wrapping* dapat dilihat pada gambar 6.



Gambar 6. Output Tahap Mel-Frequency Wrapping

Gambar 6 merupakan *output* dari tahap *mel-frequency wrapping*, di mana tahap ini pada umumnya menggunakan *filterbank* untuk penyaringan atau filter yang bertujuan untuk mengetahui energi dari *frequency band*.

Setelah melalui tahap *mel-frequency wrapping*, selanjutnya tahap terakhir yaitu tahap *cepstrum*. *Output* dari tahap *cepstrum* dapat dilihat pada gambar 7.



Gambar 7. Output Tahap Cepstrum

Gambar 7 merupakan *output* dari tahap *cepstrum* yang bertujuan untuk mengubah *mel spectrum* menjadi domain waktu dengan tahap ini akan diperoleh koefisien MFCC.

Hasil akhir dari ekstraksi ciri MFCC tersebut selanjutnya akan diaplikasikan pada tahap klasifikasi. Selanjutnya mengklasifikasi data menggunakan metode SVM dan pengujian dari data menggunakan parameter yang telah ditentukan yaitu parameter *Complexity C* dan γ (gamma), serta terdapat 3 kernel yang digunalan yaitu RBF, *Linear*, dan *Polynomial*. Hasil *training* setiap parameter SVM dapat dilihat pada tabel 2.

Tabel 2. Hasil Training Parameter SVM

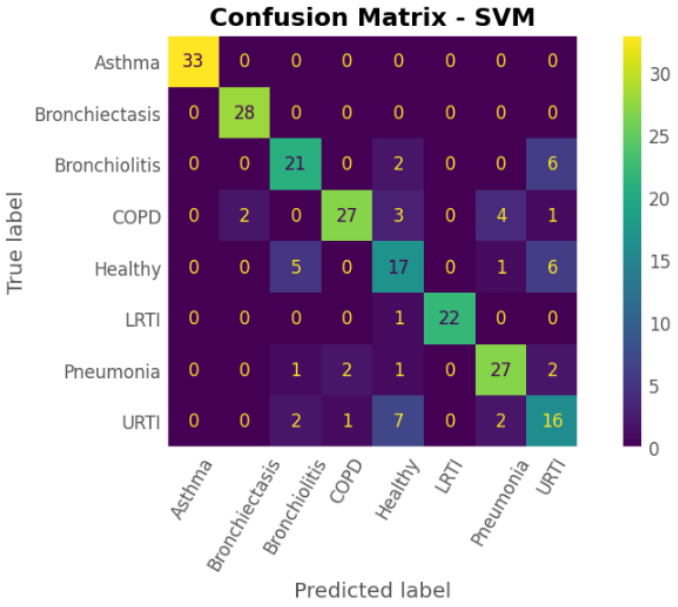
params	split0_test_score	split1_test_score	split2_test_score	split3_test_score	split4_test_score	mean_test_score	std_test_score	rank_test_score
{'SVM__C': 0.1, 'SVM__gamma': 1, 'SVM__kernel': 'linear'}	0.770833	0.822917	0.791667	0.791667	0.807292	0.796875	0.01743	1
{'SVM__C': 10, 'SVM__gamma': 0.01, 'SVM__kernel': 'linear'}	0.770833	0.822917	0.791667	0.791667	0.807292	0.796875	0.01743	1

p-ISSN: 2301-5373
e-ISSN: 2654-5101

{'SVM__C': 10, 'SVM__gamma': 0.1, 'SVM__kernel': 'linear'}	0.770833	0.822917	0.791667	0.791667	0.807292	0.796875	0.01743	1
{'SVM__C': 10, 'SVM__gamma': 1, 'SVM__kernel': 'linear'}	0.770833	0.822917	0.791667	0.791667	0.807292	0.796875	0.01743	1
{'SVM__C': 1, 'SVM__gamma': 0.001, 'SVM__kernel': 'linear'}	0.770833	0.822917	0.791667	0.791667	0.807292	0.796875	0.01743	1
{'SVM__C': 100, 'SVM__gamma': 1, 'SVM__kernel': 'linear'}	0.770833	0.822917	0.791667	0.791667	0.807292	0.796875	0.01743	1
{'SVM__C': 1, 'SVM__gamma': 0.01, 'SVM__kernel': 'linear'}	0.770833	0.822917	0.791667	0.791667	0.807292	0.796875	0.01743	1
{'SVM__C': 1, 'SVM__gamma': 0.1, 'SVM__kernel': 'linear'}	0.770833	0.822917	0.791667	0.791667	0.807292	0.796875	0.01743	1

Tabel 2 merupakan hasil *training* setiap parameter SVM. Pada tabel terdapat kolom *params* yang di mana berisi parameter yang digunakan, dengan nilai yang berbeda-beda dari setiap parameter. Kemudian pada kolom berikutnya terdapat *split test score* yang di mana berisi nilai hasil pengujian yang dilakukan menggunakan *K-Fold Cross Validation* dengan 5 kali iterasi. Pada kolom selanjutnya terdapat *mean test score* yang berisi nilai rata-rata dari *split test score*. Berikutnya terdapat kolom *std test score* merupakan nilai hasil standar deviasi dari model. Terakhir terdapat kolom *rank test score* yang merupakan hasil peringkat dari setiap parameter yang digunakan. Hasil *training* tersebut menunjukkan

parameter terbaik untuk SVM model yaitu parameter SVM__C : 0,1 , SVM__gamma : 1, serta SVM__kernel : linear dengan mendapatkan *score* terbaik sebesar 0.7968749999999999.



Gambar 8. Hasil Matriks Konfusi

Gambar 8 merupakan hasil tahap pengujian dengan Matriks Konfusi. Di mana dari hasil tersebut terlihat bahwa hasil yang terbanyak terdapat 33 data yang teridentifikasi benar dengan diagnosis asma dan beberapa diagnosis penyakit paru masih ada yang teridentifikasi salah dikarenakan hasil identifikasi dari beberapa data asli dengan hasil prediksi diagnosis masih terdapat ketidaksesuaian ciri-ciri dari beberapa penyakit yang diprediksi.

	precision	recall	f1-score	support
Asthma	1.00	1.00	1.00	33
Bronchiectasis	0.93	1.00	0.97	28
Bronchiolitis	0.72	0.72	0.72	29
COPD	0.90	0.73	0.81	37
Healthy	0.55	0.59	0.57	29
LRTI	1.00	0.96	0.98	23
Pneumonia	0.79	0.82	0.81	33
URTI	0.52	0.57	0.54	28
accuracy			0.80	240
macro avg	0.80	0.80	0.80	240
weighted avg	0.80	0.80	0.80	240

Gambar 9. Hasil Akurasi

Gambar 9 merupakan hasil akurasi dari klasifikasi paru-paru berdasarkan suara paru-paru menggunakan metode MFCC dan SVM dengan memperoleh tingkat akurasi sebesar 80%. Pada masing-masing diagnosis penyakit paru-paru memperoleh hasil yang berbeda-beda, di mana diagnosis penyakit asma memperoleh hasil tertinggi sebesar 100%.

4. Kesimpulan

Pada penelitian yang telah dilakukan dapat disimpulkan bahwa penelitian dengan menggunakan metode MFCC dan SVM untuk mengklasifikasi penyakit paru-paru berdasarkan suara paru-paru memperoleh hasil yang cukup baik setelah melalui beberapa proses dengan tingkat akurasi yang dihasilkan yaitu sebesar 80%.

Daftar Pustaka

[1] A. H. Falah dan Jondri, “Klasifikasi Suara Paru Normal Dan Abnormal Menggunakan Deep Neural Network dan Support Vector Machine”, *e-Proceeding of Engineering*, vol. 6, no.1, hlm. 2451 – 2459, April 2019.
[2] A. R. Isnain, A. I. Sakti, D. Alita dan N. S. Marga, “SENTIMEN ANALISIS PUBLIK TERHADAP KEBIJAKAN LOCKDOWN PEMERINTAH JAKARTA MENGGUNAKAN ALGORITMA SVM”, *JDMSI*, vo. 2, no. 1, hlm. 31 – 37, 2021.
[3] C. Aridela, A. Rizal, dan Y. S. Hariyani, “PERBANDINGAN SUARA PARU NORMAL DAN ABNORMAL MENGGUNAKAN PROBABILISTIC NEURAL NETWORK DAN SUPPORT VECTOR MACHINE”, *e-Proceeding of Engineering*, vol. 4, no. 1, hlm. 165 – 172, April 2017.
[4] D. S. Wiradikusuma, J. Jondri dan A. Rizal, “Klasifikasi Suara Paru-Paru menggunakan Jaringan Syaraf Tiruan dan Wavelet Transform”, *e-Proceeding of Engineering*, vol. 8, no. 2, hlm. 3224 – 3231, April 2021.
[5] E. R. Ritonga dan M. D. Irawan, “SISTEM PAKAR DIAGNOSA PENYAKIT PARU-PARU PADA ANAK DENGAN METODE DEMPSTER-SHAFER”, *CESS (Journal Of Computer Engineering, System And Science)*, vol. 2, no. 1, hlm. 39 – 47, Januari 2017.
[6] F. Rabani, Jondri dan A. Rizal, “Klasifikasi Suara Paru Normal dan Abnormal Menggunakan Ekstraksi Fitur Discrete Wavelet Transform dengan Klasifikasi Menggunakan Jaringan Saraf Tiruan yang Dioptimasi dengan Algoritma Genetika”, *Jurnal ELEMENTER*, vol. 7, no. 1, hlm. 20 – 34, Mei 2021.
[7] I. H. Wafii, Drs. Jondri dan Dr. A. Rizal, “Klasifikasi Suara Paru-Paru Menggunakan Convolutional Neural Network (CNN)”, *e-Proceeding of Engineering*, vol. 8, no. 2, hlm. 3218 – 3223, April 2021.
[8] Karimah, Z. I. Nikmah, S. K. Aditya dan E. G. Wahyuni, “Aplikasi Web Untuk Pendeteksi Penyakit Paru – Paru Menggunakan Metode Certainty Factor”, *Seminar Nasional Informatika Medis (SNIMed)*, hlm. 86 – 91, 2019.
[9] M. F. Rahman, M. I. Darmawidjadja dan D. Alamsah, “KLASIFIKASI UNTUK DIAGNOSA DIABETES MENGGUNAKAN METODE BAYESIAN REGULARIZATION NEURAL NETWORK (RBNN)”, *JURNAL INFORMATIKA*, vol. 11, no. 1, hlm. 36 – 45, Januari 2017.
[10] . A. Sadewa, T. A. Budi W dan S. Sa’adah, “IMPLEMENTASISPEAKER RECOGNITIONUNTUKOTENTIKASI MENGGUNAKAN MODIFIED MFCC–VECTOR QUANTIZATIONALGORITMA LBG”, *e-Proceeding of Engineering*, vol. 2, no. 1, hlm. 1453 – 1463, April 2015.

This page is intentionally left blank.

Jurnal Elektronik Ilmu Komputer Udayana
Volume 11, No 3. February 2023

p-ISSN: 2301-5373
e-ISSN: 2654-5101

Implementation of the Weighted Product Method for Recommendations Halal Dinning Places in Bali

Sandi[a1], Ida Bagus Made Mahendra[a2], Agus Muliantara [a3], I Ketut Gede Suhartana [a4], I Wayan Santiyasa [a5], Luh Arida Ayu Rahning Putri [a6]

[a]Program Studi Informatika, Universitas Udayana
Denpasar, Indonesia
[1]sandiesawara@gmail.com
[2]ibm.mahendra@unud.ac.id
[3]muliantara@unud.ac.id
[4]ikg.suhartana@unud.ac.id
[5]rahningputri@unud.ac.id
[6]santiyasa@unud.ac.id

Abstract

The growing tourism industry in Bali must be responded to by improving service facilities for tourists, so that Bali remains popular with many people. Including providing facilities to show the location of halal dinning places according to the criteria required by the user. Also considering that current technological developments can be easily accessed by many people, for some of these reasons the authors in this study intend to create an information system by implementing the weighted product method in order to provide recommendations for halal places to eat according to the criteria required by users. The results of this study are a recommendation system for halal places to eat in Bali on website and mobile platforms, with a recommendation accuracy of 90% based on the results of accuracy testing and error values.

Keywords: *Information System, Halal Dinning Places, Weighted Product, Recommendation System*

1. Introduction

Awareness of the importance of creating comfort and convenience for tourists and local residents in Bali, especially in choosing culinary products. It must be realized by innovation and creation in the culinary field, so that the culinary field can become one of the fields that encourage many tourist visits, because culinary has its own charm to be enjoyed. The number of restaurants, angkringan, stalls and shops that provide food for customers in the Bali area makes buyers have many choices and their own goals. However, it is often found that Muslim tourists visiting Bali are still confused and wondering whether the places to eat really provide halal-guaranteed culinary products.

On the one hand, the level of awareness of culinary business actors on the importance of certifying culinary products using halal labels is not yet so high, whereas on the other hand the government is actively campaigning for the huge potential of halal tourism in tourist destinations throughout Indonesia, including in Bali. The government realizes the importance of improving the performance of tourism services in various fields, including in the field of culinary tourism. The government's attention to the development of halal tourism requires culinary business actors to increase product penetration by following standardization rules through product certification. So through this system, it can be used to educate the public to have an awareness of the importance of halal certification and culinary business actors.

So that this system is expected to be a means of promotion for halal food businesses in Bali, as well as responding to public concerns regarding the concept of halal tourism development, so this system can be an easy alternative to be accepted and understood by the public. So with some of the reasons above, the author decided to build a system that can provide recommendations and can help consumers choose a place to provide halal culinary products according to the required criteria. It is hoped that this system will be integrated with the recommendation system and also implement other relevant technologies to be applied in overcoming this problem.

2. Reseach Methods

In this study, the authors chose to use the weighted product method as the main method in providing recommendations for halal dining places in Bali. The weighted product method was chosen because it can make good recommendations on multiple criteria in accordance with the needs of the system to be made, as evidenced by the results of previous studies that served as a guide in this study. The implementation of this method is done by using dart programming on a mobile application.
The recommended steps for determining halal dinning places using the WP method are:

a. The assessment criteria used as a reference are:
   1. Price
   2. Distance
   3. Operation Time

b. Determination of initial value, normalization with: $W_j = \frac{w_j}{\sum w_j}$

   Information:
   W : weight criteria
   j : criteria

c. Then the vector calculation process *s* is calculated based on the equation:

$$S_{i} = \prod_{j}^{n} = x_{ij}^{wj}$$

   Information:
   S : Alternative preferences are analogous to *vektor S*
   X : Criteria value
   W : weight of criteria /sub-criteria
   i : Alternative
   j : criteria
   n : many criteria

d. After the *vector s* value is obtained, the next step is to add up all *vector s* to calculate *vector v*. the calculation is as follows:

$$n: VA = \frac{\prod_{j=1}^{n} x_{ij}^{wj}}{\prod_{j=1}^{n} (x_j^*)}$$

   Information:
   V: Alternative preferences are analogous to *vector V*
   X: Criteria value
   W: weight of criteria/sub-criteria
   i : Alternative
   j : Criteria
   n : many criteria
   * : the number of citeria that have been assessed on the *vector S*

p-ISSN: 2301-5373
e-ISSN: 2654-5101

3. Result and Discussion

3.1. Database schema

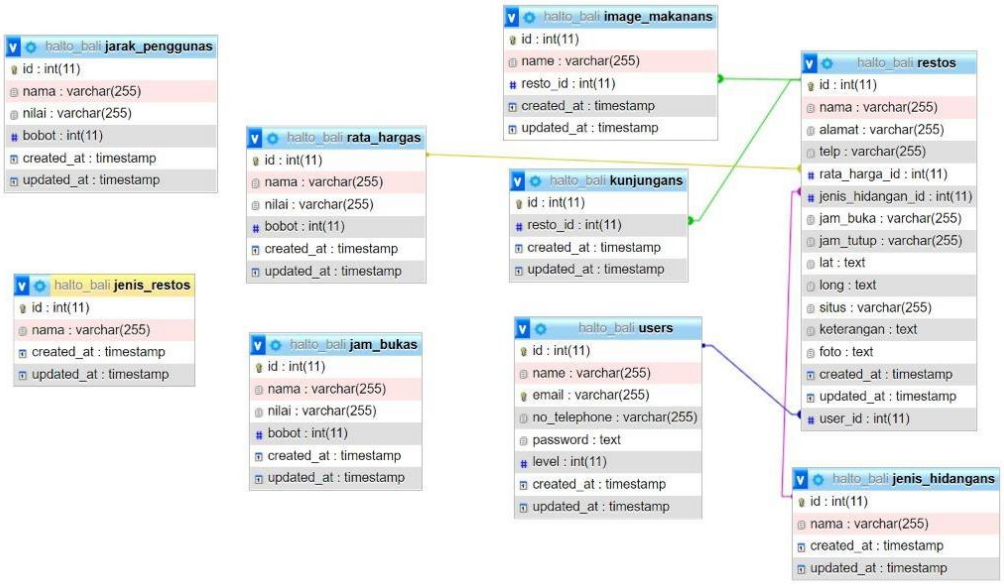


Figure 1. Database schema

The implementation of the database is carried out in a SQL database environment, the SQL database has tables, column data types, and relations between tables. This system consists of 9 tables, namely:

a. jarak_users, to store location update data for each user.
b. jenis_restos, to store the type data of the place to eat.
c. rata_hargas, stores the average menu price data at the restaurant.
d. jam_bukas, to store information about the opening time of a place to eat.
e. image_makanans, to store photo data of places to eat and food menus.
f. kunjungans, to save the details of a visit to a place to eat
g. users, to store user data.
h. restos, to store detailed data on where to eat.
i. jenis_hidangans, to store data on the types of dishes available at the dining area.

3.2. Interface Design

a. Admin Webpage
The web application is provided to perform actions as a user, in this case the admin and food service provider.

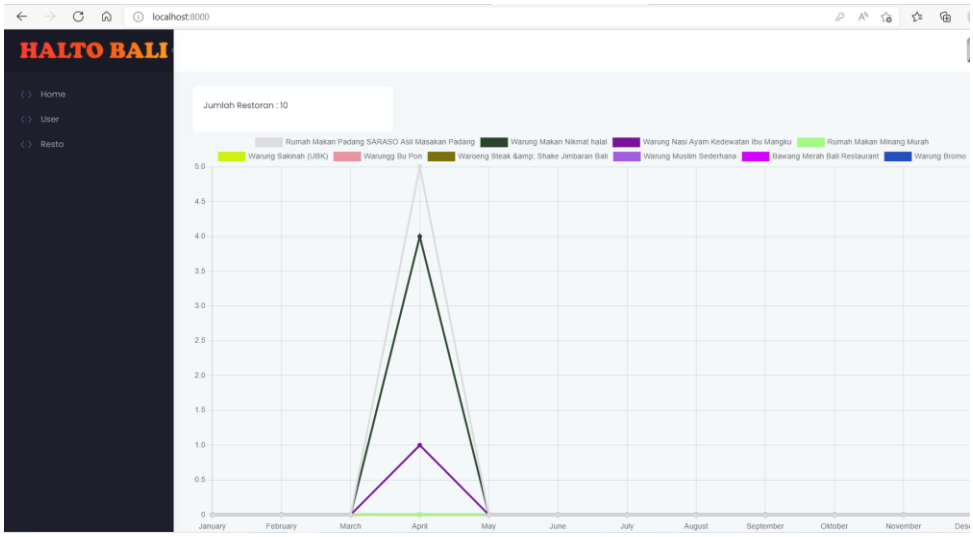


Figure 2. Dashboard

The main page after the admin and provider logs in, besides the main features seen on this page, a data card on the number of providers is presented and a graph of visits to a place to eat.

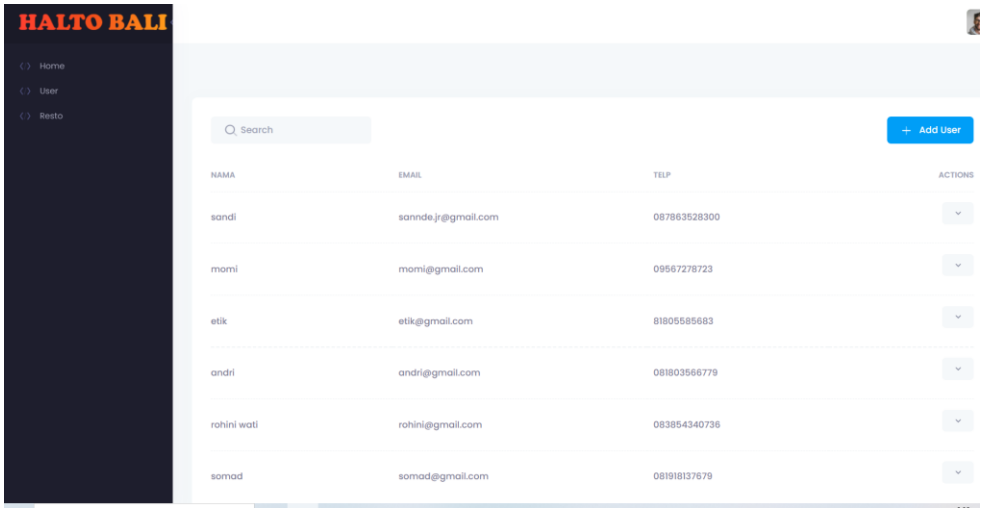


Figure 3. User Page

Is a page to add new place provider data. Administrators can enter data through the available fields.

p-ISSN: 2301-5373
e-ISSN: 2654-5101

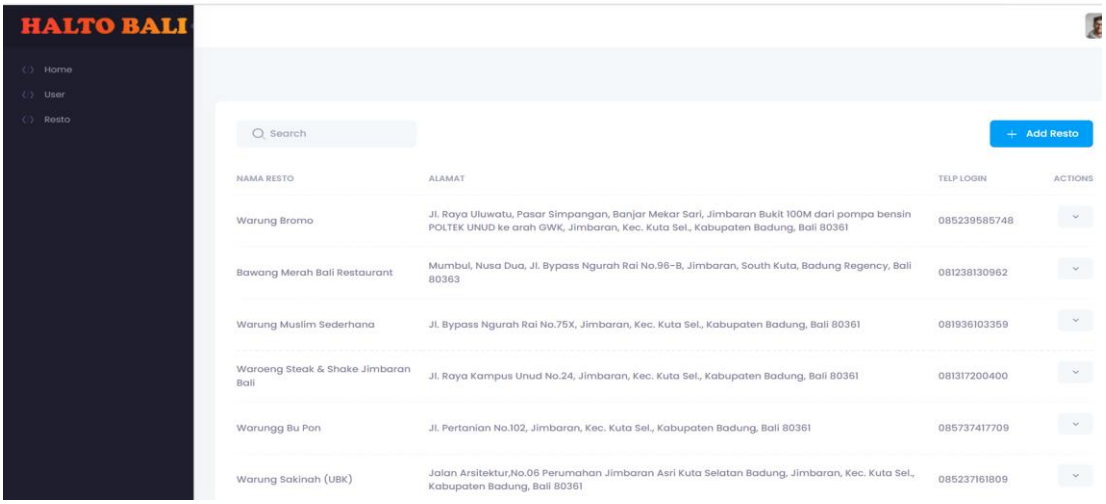


Figure 4. Restos Page

This is a page to add data for a new place to eat. Administrators can enter data through the available fields.

b. Mobile User Page
The mobile page is provided specifically for users, in this case seekers of places to eat.

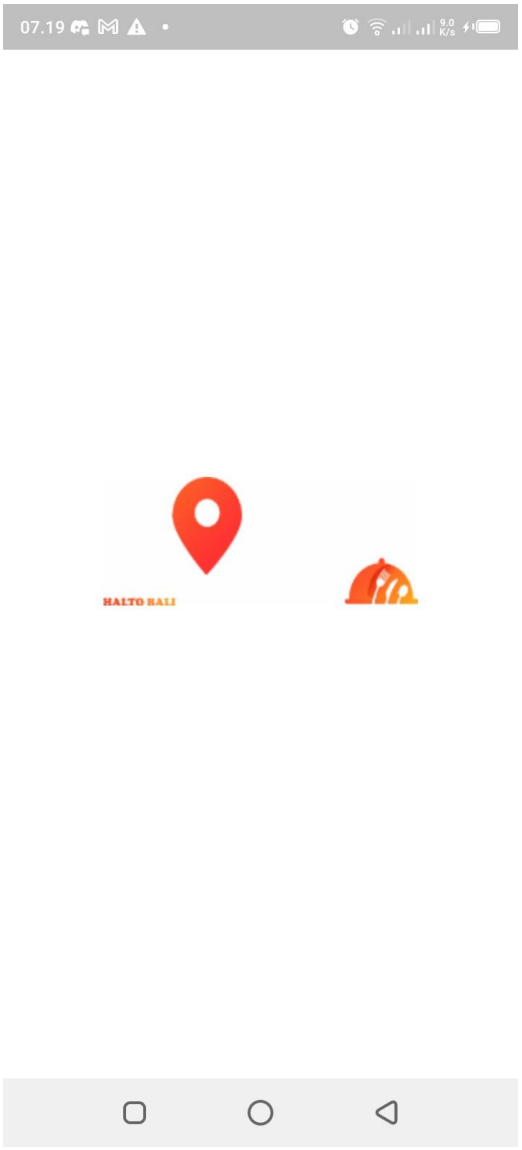


Figure 5. Splash Screen

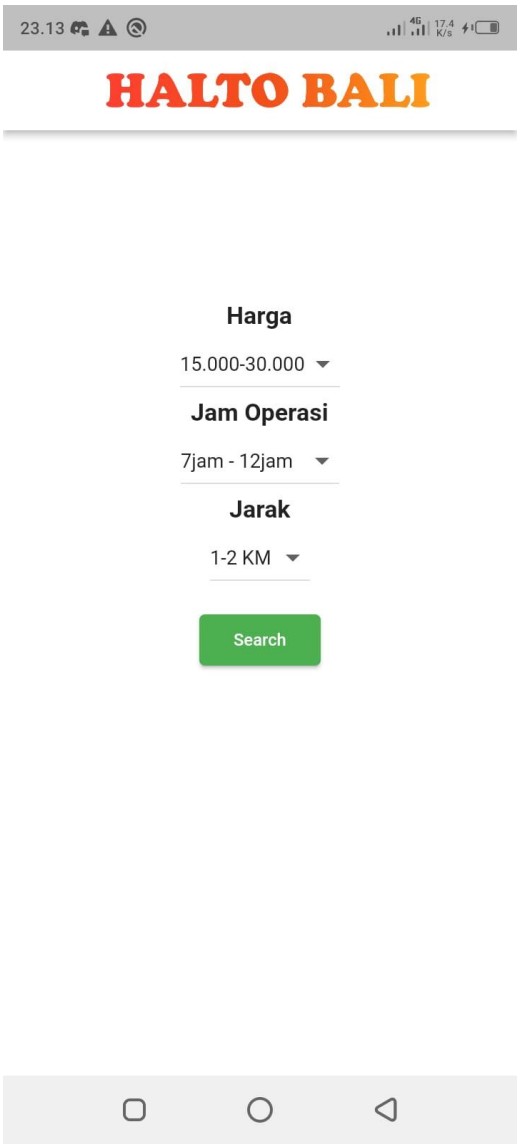


Figure 6. Halaman Input Nilai Referensi

On the Splash Screen page there are 2 components, namely Image, Title Text. Image displays the logo of the application system, Title Text displays information on the name of the application, namely Halto Bali.

While on the input page the reference value is where the user enters the desired criteria to become the user's preference value. The preference value is sent to the server as a weight value in ranking using the weighted product method.

p-ISSN: 2301-5373
e-ISSN: 2654-5101

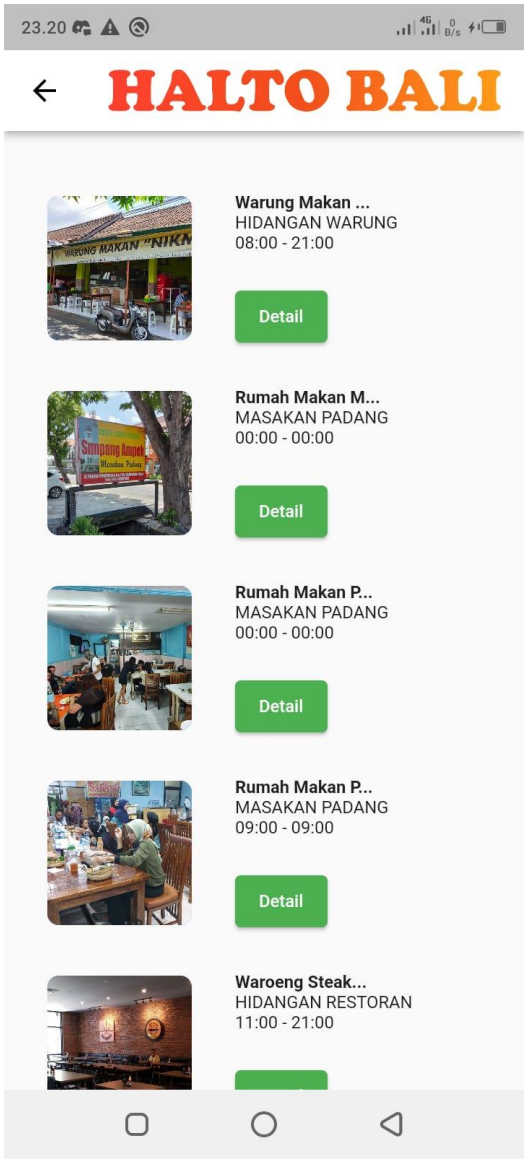


Figure 7. Recommendations list page

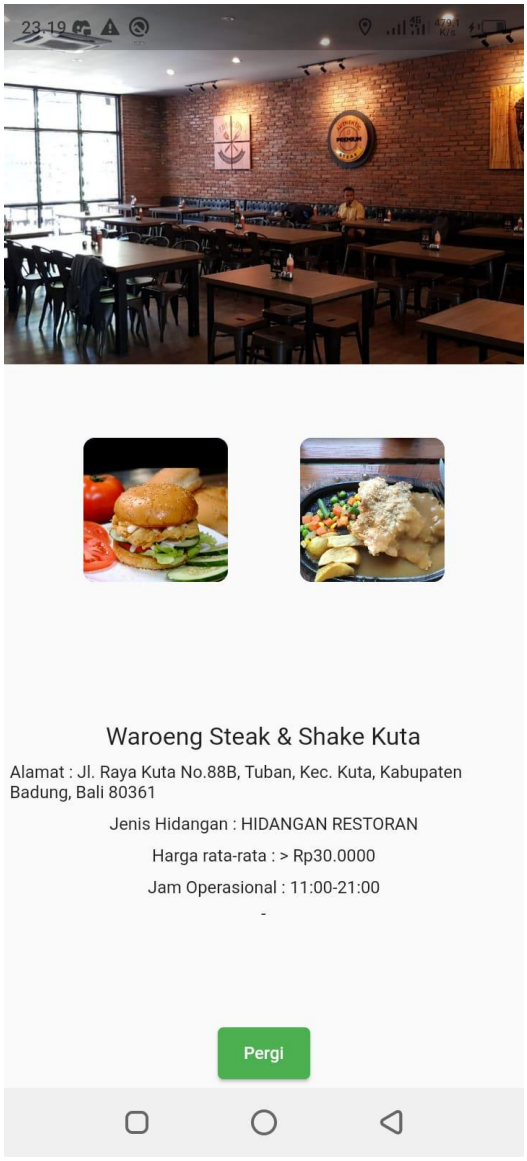


Figure 8. Details Page

The recommendation list page is a page to display the ranking results of alternative solutions that are processed from the weighted product method. Then the dining details page is a page that displays complete details of where to eat, such as the name of the place to eat, the type of place to eat, and complete information on the food menu.

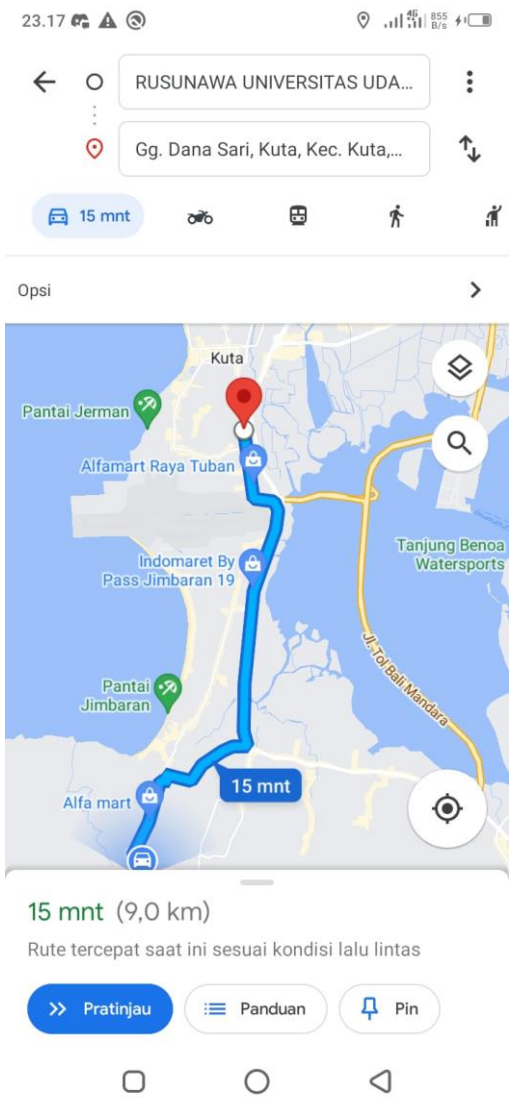


Figure 9. Dinning Place Route

This page is a page for displaying a map and route location from a place to eat. This page is equipped with a map as a geographical visualization of where to eat.

3.3. Testing

The tests carried out are in the form of validity testing using blackbox testing and accuracy testing.

a. Blackbox testing
Blackbox testing is carried out to check the functionality of the system, testing is carried out on web applications that are accessed by administrators and providers and mobile applications that are accessed by place seekers. The following is a blackbox test table on the system.

Table 1 Blackbox test results

Platform test	Functional	Identification	Test Results
Mobile	Registration	UM1	Valid
	Login	UM2	Valid
	Logout	UM3	Valid

p-ISSN: 2301-5373
e-ISSN: 2654-5101

	input criteria	UM4	Valid
	input dinning place data	UM5	Valid
	Input alternative data	UM6	Valid
	Search for dinning place	UM7	Valid
	Detile for dinning place	UM8	Valid
	Recommended the dinning place	UM9	Valid
	The location of the dinning place	UM10	Unvalid

b. Accuracy Testing

System accuracy testing is done by comparing the ranking results performed by the system with the ranking results with real data or existing data. The test was carried out at the location of Rusunawa, Udayana Campus, Street Udayana Campus Jimbaran, Kec. Kuta Sel., Badung Regency, Bali 80361, Indonesia with latitude: -8.7981624 and longitude: 115.1617221. With detailed criteria as follows:

1. Distance
   a. <=1 km = near
   b. 1-2 km = medium
   c. > 2 km = far
2. Price
   a. <15.000 = cheap
   b. 15.000-30.000 = medium
   c. >30.0000 = expensive
3. Opration Time
   a. 12jam -24 jam = very long
   b. 7 jam-11 jam = long
   c. <7 jam = short

Tested on 50 data. Based on the trials that have been carried out and making comparisons, of course there are some differences. The difference or error will be calculated the error value. To calculate the accuracy as follows:

Appropriate amount of data = 45
Incorrect amount of data = 5

$$Accuracy = \frac{appropriate\ amount\ of\ data}{amount\ of\ test\ data} \times 100\%$$

$$= \frac{45}{50} \times 100\% = 90\%$$

$$Error = \frac{incorrect\ amount\ of\ data}{amount\ of\ test\ data} \times 100\%$$

$$= \frac{5}{50} \times 100\% = 10\%$$

Calculation of the accuracy test obtained the appropriate results as many as 45 alternatives and 5 alternatives that were not suitable. Thus the level of accuracy obtained from the WP.

4. Conclusion

From the results of the trials and discussions conducted, the following conclusions can be drawn:

a. The system created can apply calculations with the WP method. The application of the WP method on an android-based application produces accurate calculations after being tested with system calculations and real data calculations.
b. The system calculation using the WP method shows an accuracy rate of 90%, so the results are in the good category.

References

[1] Khairina, D.M., Ivando, D., Maharani, S, “Implementasi Metode Weighted Product Untuk Aplikasi Pemilihan Smartphone Android”, *Jurnal Infotel,* Vol. 8. 2016.
[2] McGinty, L., Smyth, B. “Adaptive selection: analysis ofcritiquing and preference based feed back in conversation on recommender system”, *International J Electron Commerce.* 2006
[3] Sucipto, “Analisa hasil rekomendasi pembimbing menggunakan multi-attribute Dengan metode weighted product”, *Fountain of Informatics Journal*, Vol.2. 2017
[4] Darmaja, “Rancang Bangun Sistem Rekomendasi Warung Makanan Khas Bali Menggunakan Metode Collaborative Filtering Berbasis Mobile”. Jimbaran: Ilmu Komputer, UNUD, 2016.
[5] Pandean, S.S., Hansun, S, “Aplikasi Web Untuk Rekomendasi Restoran Menggunakan Weighted Product”, *Jurnal Teknologi Informasi dan Ilmu Komputer,* Vol.5. 2018
[6] Connolly, T., & Begg, C. Database Systems: a Practical Approach to Design, Implementation, and Management. 5th Edition. America: Pearson Education. 2010.

Jurnal Elektronik Ilmu Komputer Udayana
Volume 11, No 3. February 2023

p-ISSN: 2301-5373
e-ISSN: 2654-5101

Pengamanan Audio Menggunakan Metode RSA dan Steganografi *Spread spectrum* Berbasis Android

I Ketut Kusuma Merdana[a1], I Komang Ari Mogi[a2], I Gede Arta Wibawa[a3], Agus Muliantara[a4], I Ketut Gede Suhartana[a5], I Putu Gde Hendra Suputra[a6]

[a]Program Studi Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Udayana
Badung, Bali, Indonesia
[1]ketutkusuma0910@gmail.com
[2]arimogi@unud.ac.id
[3]gede.arta@unud.ac.id
[4]muliantara@unud.ac.id
[5]ikg.suhartana@unud.ac.id
[6]hendra.suputra@unud.ac.id

Abstract

Audio is a multimedia element whose function is the same as text, video or images, but for the recipient it can only be heard through human hearing, which is 20 to 20,000 Hz. Audio can also be in the form of a song that is often listened to and has a high selling power, therefore audio is also often traded or rented. In this study, the application was developed using the Spread spectrum method, Rivest Shamir Adleman and assisted also by the Standard Advanced Encryption method to secure audio to protect against data theft by unauthorized parties. This application is based on Android which can be further developed to run more flexibly and make it easier for users to use it. From the test results the application can process input and output as expected. For testing the results of image quality, which is an audio password insertion container, in this test, the PSNR value with an average of 56,851 is obtained, and the MSE value with an average value of 0.124 which is a good result for testing image quality.

Keywords: *Steganography, RSA, Spread spectrum, PSNR*

1. Pendahuluan

Saat ini, meningkatnya penggunaan teknologi komputasi dan telekomunikasi mengubah pandangan masyarakat tentang komunikasi. Kemajuannya di bidang jaringan komunikasi seluler dengan konsep sistem terbuka yang memudahkan seseorang untuk membobol jaringan adalah data yang tidak aman karena dapat digunakan oleh pihak yang tidak bertanggung jawab untuk mengambil data penting yang berujung pada proses pengiriman informasi [1]. Melindungi informasi sangat penting sehingga ada dua cara untuk melindunginya: enkripsi dan stenografi. Enkripsi pada umumnya adalah ilmu penyandian data, menggunakan kunci enkripsi untuk mengacak data [2]. Stegnografi adalah ilmu menulis pesan tersembunyi atau tersembunyi sehingga hanya pengirim dan penerima yang dapat mengenalinya [3]. Kriptografi memiliki banyak algoritma untuk melindungi informasi. Salah satunya adalah algoritma RSA yang merupakan algoritma enkripsi asimetris. Keuntungan utama dari algoritma RSA adalah menggabungkan teknik keamanan informasi, atau stenografi, untuk meningkatkan keamanan. Stegnografi *spread spectrum* membutuhkan proses penyisipan melalui proses propagasi dan modulasi, yang mengacak pesan propagasi *spread spectrum* di seluruh media penampung. Untuk pesan dan informasi yang dikirim secara acak.

Pada penelitian sebelumnya [2] tentang kombinasi enkripsi RSA dan stenografi *spread spectrum* untuk melindungi data teks, penelitian ini menggunakan enkripsi RSA dan stenografi *spread spectrum* sebagai file teks pada sebuah gambar, saya berhasil menggabungkannya dengan. Penelitian selanjutnya tentang implementasi stenografi SMS pada audio menggunakan metode *spread spectrum* [4] menguji PSNR dan MSE penelitian ini dan membandingkan file asli dengan file hasil stagnografi.

Pengujian 3x menghasilkan rata-rata MSE sebesar 2.2692E-06 dengan nilai PSNR rata-rata sebesar 111.9842 dB. Jumlah karakter atau ukuran file yang disisipkan berpengaruh signifikan terhadap nilai MSE dan PSNR. Berdasarkan referensi di atas, sebuah penelitian yang berjudul "Implementasi *Spread spectrum* RSA Algorithms and Stegnography for Installing Text Messages in Speech" telah diusulkan. Penelitian ini mengimplementasikan transaksi atau aplikasi keamanan audio yang dapat digunakan untuk bertransaksi. Data audio dalam aplikasi ini menggunakan audio terenkripsi dan audio berformat .wav terenkripsi RSA. Format WAV dipilih karena dapat menyimpan file audio terkompresi dan tidak terkompresi. Pembeli yang memutar audio harus menggunakan aplikasi keamanan audio ini untuk memutar audio setelah proses dekripsi dan ekstraksi.

2. Metode Penelitian

2.1 *Riverst Shamir Adleman*

Rivest Shamir Adleman (RSA) adalah algoritma enkripsi asimetris yang menggunakan dua angka acak sebagai kunci untuk menghasilkan dua kunci untuk enkripsi pesan. Proses dekripsi plaintext (P) dan ciphertext (C) dari algoritma RSA menggunakan persamaan berikut [5]:

$$C = P^e \ mod \ n \quad (1)$$

$$P = C^d \ mod \ n \quad (2)$$

Proses pencarian kunci publik (e,n) dan kunci privat (d,n) tidak gratis. Untuk menentukan kunci publik dan kunci privat: (Rudiyanto,.

1. Pilih 2 buah bilangan prima p dan q yang mana nilainya harus berbeda.
2. Menghitung n = p*q
3. Mencari nilai dari persamaan m = (p-1)*(q-1)
4. Menda[atkan nilai e, yang mana berasal dari nilai e yang merupakan bilangan relative prima dari m
5. Mendapatkan nilai d, yang berasal dari persamaan e.d mod m = 1.

Untuk mendapatkan nilai d yang memenuhi persamaan pada langkah 5 digunakan algoritma extended Euclidean pada langkah selanjutnya.

1. Mengubah persamaan e.d dalam mode m = 1 menjadi (m.x) + (e.y) = 1 yang merupakan persamaan Euclidean
2. Mentukan nilai a dan b yang memenuhi persmaan m = (a.e) + b
3. Menyimpan persamaan yang ditemukan dan mengganti nilai m dan nilai e dengan nilai e dan nilai b
4. Mengulangi hingga nilai b menjadi 1 untuk langkah 2 dan 3
5. Mengubah menjadi b = m – (a.3) untuk semua persamaan yang tersimpan
6. Pada setiap persamaan yang nilainya b = e, ubah nilai e
7. Mengubah agar menjadi mirip dengan persamaan yang dibuat pada langkah 1 untuk setiap persamaan
8. Pada setiap persamaan yang telah diubah tersebut, variable y merupakan kunci privat yang dapat digunakan untuk proses dekripsi. Kunci merupakan bilangan positif karenanya jika y bernilai negatif maka kunci dekripsi akan menggunakan persamaan m – y.

2.2 *Spread spectrum*

Metode *spread spectrum* adalah teknologi transmisi yang menggunakan kode semu yang tidak bergantung pada data informasi sebagai modulator bentuk gelombang dan menyebarkan energi sinyal pada jangkauan yang lebih luas dari jalur komunikasi asli (bandwidth). Melalui penerima, sinyal dikumpulkan kembali dalam replika *encoder pseudo-noise* yang disinkronkan. Metode *spread spectrum* memerlukan cakupan sebagai *noise* atau sebagai upaya untuk menambahkan pseudo-noise ke cakupan. Penyisipan menggunakan metode ini memiliki tiga proses: difusi pesan, modulasi, dan penyisipan amplop, dan ekstraksi memiliki tiga proses: akuisisi pesan, demodulasi, dan *despreading* [6].

Penyisipan *Spread Spectrum* memiliki langkah-langkah utama sebagai berikut.

1. Gunakan kata kunci untuk menghasilkan kebisingan pseudo-acak.
2. Proses penyebaran untuk setiap pesan yang akan disisipkan.
3. Modulasi antar pesan hasil penyebaran dengan pseudo-acak yang telah menghasilkan kebisingan.

p-ISSN: 2301-5373
e-ISSN: 2654-5101

4. Menyisipkan kebisingan ke dalam setiap LSB pada penampung yang mana pesan rahasia akan disisipkan.

Untuk langkah ekstrasi yang merupakan langkah untuk mengambil data yang disisipkan terdiri dari langkah-langkah sebagai berikut [7].

1. Mengambil data pada setiap LSB data gambar.
2. Hasilkan kebisingan pseudo-acak dengan menggunakan kata kunci yang sama seperti proses penyisipan.
3. Melakukan proses de-modulasi antara bit pesan yang telah diekstrak dengan deret bilangan pseudo-acak.
4. Proses penyebaran balik yang bertujuan untuk mendapatkan pesan rahasia yang sudah disisipkan.

2.3 *Advanced Encryption Standar (AES)*

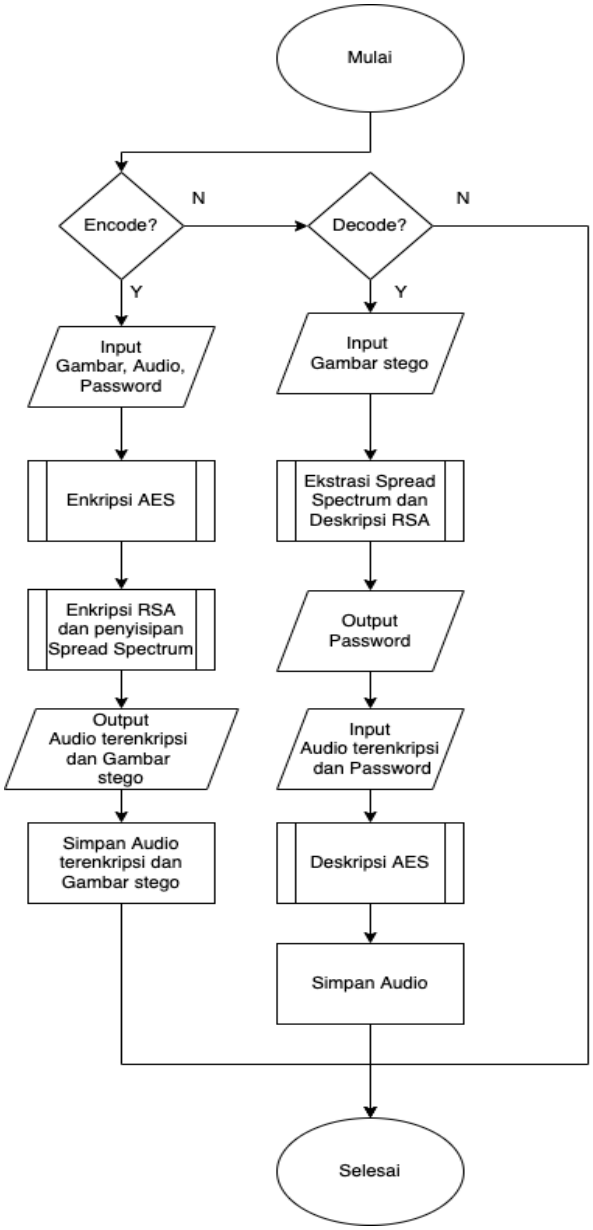
Algoritma AES (*Advanced Encryption Standard)* merupakan algoritma enkripsi yang dapat digunakan untuk melindungi data atau informasi. Algortima AES adalah sebuah *block ciphertext* simetris yang dapat mengenkripsi (encrypt) dan mendekripsi (decrypt) informasi. Kunci enkripsi AES terdiri dari kunci panjang 128-bit, 192-bit atau 256-bit. Perbedaan panjang kunci mempengaruhi jumlah putaran yang diimplementasikan oleh algoritma AES. Pada daarnya, operasi AES dilakukan pada *array byte* dua dimensi yang disebut status. Ukuran bagian adalah NRowXNCol, pada awal enkripsi, data input, disalin ke *array* keadaan dalam format in0, in2, in3, in4, in5, in6, in7, in8, in9, in10, in11, in12, in13, in14, in15. Status ini akan dilakukan kemudian oleh operasi enkripsi/dekripsi. Kemudian *output* dilangkahkan ke dalam *array output*[8].

2.4 Analisis Masalah

Dalam penelitian ini mencoba mengembangkan aplikasi yang dapat mengenkripsi dan mendekripsi audio menggunakan metode RSA, tetapi setelah membaca dan menjalankan eksperimen itu sendiri, peneliti tidak dapat melakukannya. Hal ini juga diperkuat dengan penelitian sebelumnya yang gagal mendekripsikan audio yang terenkripsi RSA. Masalah utama dengan dekripsi adalah bahwa RSA tidak cocok untuk mengenkripsi data dalam jumlah besar. Semakin besar data yang akan dienkripsi, semakin panjang kunci yang dihasilkan oleh RSA harus mengikuti panjang data. Untuk itu peneliti menggunakan metode enkripsi yang lebih sederhana yaitu metode AES. Ini unggul dalam mengenkripsi dan mendekripsi data dalam jumlah besar. Penggunaan metode atau algoritma akan dilanjutkan pada penelitian ini dengan menggabungkan *spread spectrum steganography*. Oleh karena itu, dienkripsi menggunakan metode RSA dan dimasukkan ke dalam wadah sebelum Anda memasukkan kunci atau kata sandi untuk membuka audio.

2.5 Desain Perancangan Sistem

Desain perancangan sistem berisikan langkah pada rancangan pembuatan perangkat lunak, yang bertujuan untuk mempermudah untuk pengembangan sistem aplikasi yang dibangun. Sesuai dengan Gambar 1, sistem memiliki dua fungsi utama, enkoding dan dekoding, dimana fungsi enkoding memasukkan gambar, audio dan password. Gambar-gambar ini dienkripsi menggunakan metode AES audio dan enkripsi RSA, dan ketika dimasukkan dengan *Spread spectrum*, audio dihasilkan. Gambar terenkripsi dan gambar yang disisipkan. Fungsi dekoding pertama memasukkan gambar yang disisipkan, pertama-tama mengekstraknya dengan *Spread spectrum*, dan kemudian mendekodekannya dengan RSA. Setelah mendapatkan kata sandi, itu akan dimasukkan dengan audio terenkripsi yang didekripsi menggunakan AES.



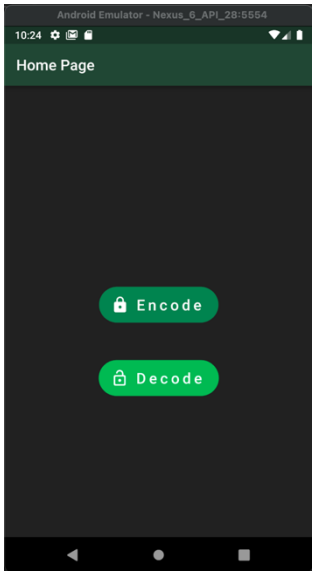
Gambar 1. Alur umum sistem

3. Hasil dan Pembahasan

3.1 Tampilan Antarmuka Sistem Pengamanan Audio

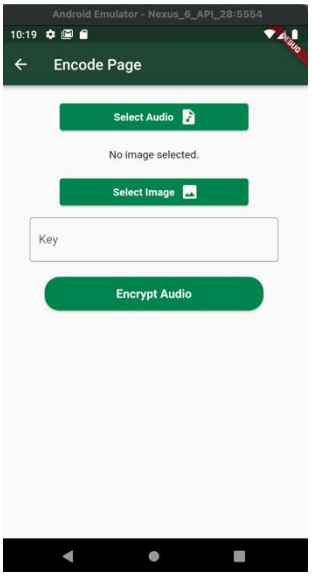
Tampilan antarmuka sistem terdapat tiga halaman yang difokuskan, yaitu halaman utama, halaman enkode, dan halaman dekode. Pada saat aplikasi dibuka maka akan ditujukan kepada halaman utama yang mana berisikan dua fitur seperti yang ditujukan pada Gambar 2.

p-ISSN: 2301-5373
e-ISSN: 2654-5101



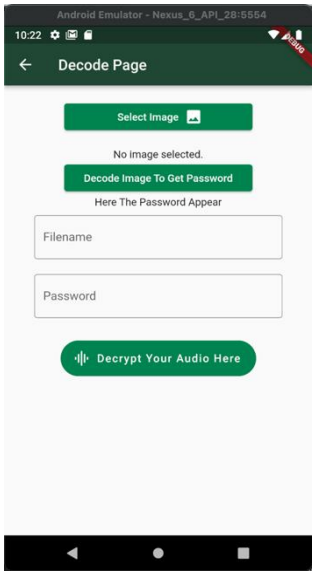
Gambar 2. Halaman utama

Gambar 2 menunjukan halaman utama yang mana merupakan halaman pertama yang ditampilkan kepada pengguna yang di dalamnya terdapat fitur encode dan decode. Fitur-fitur tersebut merupakan fitur utama yang ada pada aplikasi. Oleh karenanya halaman utama sangat penting dan harus ada.



Gambar 3. Halaman Enkode

Gambar 3 menunjukkan halaman Enkode yang digunakan untuk enkoding. Pertama, ada tombol *"Select Audio"* yang digunakan pengguna untuk memilih audio. Dengan fungsi pemilihan audio ini, *. Hanya file audio format wav yang dapat dibaca. Di bawahnya terdapat tombol *Select Image*, yang digunakan untuk memilih gambar yang akan digunakan sebagai pengganti password terenkripsi. Lalu ada area entri teks untuk memasukkan kata sandi. Ini harus berisi 8 karakter.



Gambar 4. Halaman Dekode

Selanjutnya halaman decode yang merupakan halaman dimana fungsi dekripsi dan ekstrasi berlangsung pada sistem. Tampilan halaman dekode bisa dilihat pada Gambar 4 yang mana dilihat bahwa pada halaman dekode akan menerima masukan berupa gambar yang telah disisipkan yang akan mengambil kunci untuk membuka audio. Setelah mendapatkan kunci maka kunci akan digunakan untuk membuka audio terenkripsi yang sebelumnya sudah dipilih oleh pengguna.

3.2 Pengujian Kualitas Gambar Tersisipkan

Pada penelitian ini, untuk menguji apakah gambar yang sudah disisipkan memiliki kualitas yang bagus agar tidak dapat diketahui oleh pengguna asing bahwa gambar tersebut sudah memiliki data rahasia. Maka dilakukan pengujian PSNR dan MSE yang merupakan pengujian terhadap kualitas gambar yang disisipkan kunci audio. Hasil pengujian tersebut dapat dilihat pada Tabel 1.

Tabel 1. Hasil Pengujian Kualitas Gambar Dengan PSNR dan MSE

No.	Berkas Gambar	Ukuran Gambar (Steganografi)	MSE	PSNR
1.	Gambar 1	11,6 MB	0.11664017220786	57.462322087 dB
2.	Gambar 2	7,2 MB	0.15389515817901	56.258554045 dB
3.	Gambar 3	5,11 MB	0.0880213131703	58.6849251745 dB
4.	Gambar 4	3,5 MB	0.12594370659722	57.1290389017 dB
5.	Gambar 5	9,4 MB	0.13714898003472	56.758877787 dB
6.	Gambar 6	4,9 MB	0.17089168495418	55.8035942902 dB
7.	Gambar 7	6,7 MB	0.06488634478788	60 dB
8.	Gambar 8	9,6 MB	0.16076400945216	56.068915319 dB

Pada Tabel 1 dapat dibandingkan dan diuji dengan MSE dan PSNR untuk menentukan kualitas gambar yang baik. Semakin kecil nilai MSE, semakin baik kualitas gambarnya. Untuk PSNR, gambar berkualitas tinggi adalah gambar dengan PSNR besar. Pada pengujian PSNR dan MSE rata-rata PSNR untuk pengujian ini adalah 56,851 dB dan nilai MSE rata-rata 0,124 dB. Pengujian ini dilakukan untuk menguji bagaimana kualitas gambar setelah proses *embedding* atau penyisipan dibandingkan dengan rata-rata skor PSNR 56,851 dan rata-rata skor MSE 0,124. Hal ini menunjukkan bahwa proses steganografi menggunakan *Spread spectrum* dapat berhasil menyisipkan data ke dalam citra.

3.3 Pengujian Pengubahan Gambar

Pengujian ini menjalankan pengujian untuk menganalisis bagaimana sebuah gambar dengan tombol atau pesan mengubah atributnya berupa ukuran gambar, warna gambar, panjang dan lebar gambar.

p-ISSN: 2301-5373
e-ISSN: 2654-5101

Hal ini dilakukan untuk menguji hasil ekstraksi apakah hasilnya cocok dengan kunci yang dimasukkan sebelumnya setelah atribut diubah. Setelah dilakukan pengujian dan mendapatkan hasil yang ditunjukkan pada Tabel 2, dapat disimpulkan bahwa citra steganografi dengan atribut yang dimodifikasi tidak dapat diekstraksi.

Tabel 2. Hasil pengujian Pengubah Gambar

No	Pengujian	Hasil
1.	Ubah warna gambar	Tidak berhasil ekstrasi
2.	Ubah ukuran gambar	Tidak berhasil ekstrasi
3.	Ubah panjang dan lebar gambar	Tidak berhasil ekstrasi

4. Kesimpulan

Aplikasi yang dibangun menggunakan penyebaran spektrum dan enkripsi dengan RSA berhasil menyisipkan dan mengekstraksi teks cipher ke dalam gambar yang disiapkan dengan baik. Kualitas gambar yang telah disisipkan menerima rata-rata 56.185 dB untuk nilai PNSR dan untuk nilai MSE rata-rata dengan nilai 0,113431 yang dapat disimpulkan bahwa kualitas gambar sulit untuk dibedakan dengan yang aslinya. Namun sistem ini memiliki kelemahan berupa panjang minimum wadah format gambar yang digunakan untuk penyisipan harus sepanjang 1504 piksel, ini dapat mempersulit sistem untuk pengambilan data gambar.

Daftar Pustaka

[1] Yusfrizal, “RANCANG BANGUN APLIKASI KRIPTOGRAFI PADA TEKS MENGGUNAKAN METODE REVERSE CHIPER DAN RSA BERBASIS ANDROID Yusfrizal 1),” *Jurnal Teknik Informatika Kaputama (JTIK)*, vol. 3, no. 2, 2019.

[2] Rajamah Limbong, “KOMBINASI KRIPTOGRAFI RSA DAN STEGANOGRAFI *SPREAD SPECTRUM*,” vol. 7, no. 1, pp. 97–100, 2019.

[3] A. Septayuda, I. Bambang Hidayat, and H. Hudan Nuha, “ANALISIS STEGANOGRAFI CITRA DIGITAL MENGGUNAKAN METODE *SPREAD SPECTRUM* BERBASISANDROID ANALYSIS OF DIGITAL IMAGE STEGANOGRAPHY USING *SPREAD SPECTRUM* METHOD BASED ON ANDROID,” Bandung, Dec. 2014.

[4] M. M. Assyahid, R. Rihartanto, and D. S. B. Utomo, “Implementasi Steganografi Pesan Text ke Dalam Audio Dengan Metode *Spread spectrum*,” *Juristek*, vol. 3, no. 2, pp. 27–34, 2018.

[5] A. Hidayat and A. Faizin, “PERBANDINGAN KRIPTOGRAFI MENGGUNAKAN ALGORITMA DATA ENCRYPTION STANDART (DES) DAN ALGORITMA RIVEST SHAMIR ADLEMAN (RSA) UNTUK KEAMANAN DATA,” *JASIEK*, vol. 1, no. 2, 2019, doi: 10.12928/JASIEK.v13i2.xxxx.

[6] B. Oktavianto, T. W. Purboyo, and R. E. Saputra, “A Proposed Method for Secure Steganography on PNG Image Using *Spread spectrum* Method and Modified Encryption,” 2017. [Online]. Available: http://www.ripublication.com

[7] M. Iqbal, T. Zebua, and R. D. Sianturi, “Implementasi Algoritma Spritz Dan *Spread spectrum* Untuk Menyembuyikan Pesan Enkripsi Kedalam File Audio Mp3,” *KOMIK (Konferensi Nasional Teknologi Informasi dan Komputer)*, vol. 3, no. 1, pp. 486–492, 2019, doi: 10.30865/komik.v3i1.1631.

[8] A. Fauzi, “Analisa Kombinasi Pesan Teks Ke Dalam File Audio Memanfaatkan Algoritma Data Encryption Standard Dan Metode End of File,” *Jurnal Teknik Informatika Kaputama (JTIK)*, vol. 3, no. 1, pp. 1–8, 2019.

This page is intentionally left blank.

Jurnal Elektronik Ilmu Komputer Udayana
Volume 11, No 3. February 2023

p-ISSN: 2301-5373
e-ISSN: 2654-5101

Klasifikasi Aksara Bali Berbasis Suara Dengan Metode KNN dan FastDTW

I Putu Rama Anadya[a1], I Gusti Agung Gede Arya Kadyanan[a2], I Dewa Made Bayu Atmaja Darmawan[b3], Cokorda Rai Adi Pramartha[b4], Anak Agung Istri Ngurah Eka Karyawati[b5], Ngurah Agus Sanjaya ER[b6]

[a]Program Studi Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Udayana
Badung, Bali, Indonesia
[1]gusrama18@gmail.com
[2]gungde@unud.ac.id
[3]dewabayu@unud.ac.id
[4]cokorda@unud.ac.id
[5]eka.karyawati@unud.ac.id
[6]agus_sanjaya@unud.ac.id

Abstrak

Bahasa Bali merupakan bahasa yang digunakan masyarakat Bali untuk berkomunikasi. Untuk menulis bahasa Bali, mereka dapat menggunakan Aksara Bali. Pada era globalisasi ini, aksara Bali mulai di tinggalkan oleh masyarakat atau mulai menuju kepunahan. Saat ini aksara Bali membutuhkan pengembangan di dunia digital sehingga mampu menyesuaikan dengan perkembangan jaman, khususnya pada media pembelajaran seperti sebuah aplikasi. sebelum membangun aplikasi perlu dibuatkan sebuah sistem yang dapat mendukung aplikasi tersebut. Penulis membuat sistem klasifikasi dengan metode KNN dan FastDTW untuk mengenali suara aksara Bali khususnya aksara Wreasta. Hasil penelitian yang telah dilakukan mendapatkan akurasi yang baik yaitu 100% pada model klasifikasi dengan nilai k-3, k-5, k-7 dan k-9 serta akurasi terendah pada k-19 dengan akurasi 94,44%

Kata Kunci: Aksara Bali, pengenalan suara, KNN, FastDTW, suara

1. Pendahuluan

Bahasa Bali merupakan bahasa yang digunakan masyarakat Bali untuk berkomunikasi. Untuk menulis bahasa Bali, mereka dapat menggunakan Aksara Bali. Pada era globalisasi ini, aksara Bali mulai di tinggalkan oleh masyarakat atau mulai menuju kepunahan[2].

Untuk meningkatkan minat masyarakat dalam menggunakan aksara Bali, pemerintah mengeluarkan Perda No. 1 tahun 2018 Tentang Bahasa, Aksara, dan Sastra Bali juga Peraturan Gubernur No. 80 Tahun 2018 tentang Perlindungan dan Penggunaan Bahasa, Aksara dan Sastra Bali serta Penyelengaraan Bulan Bahasa Bali. Saat ini aksara Bali membutuhkan pengembangan di dunia digital sehingga mampu menyesuaikan dengan perkembangan jaman, khususnya pada media pembelajaran.

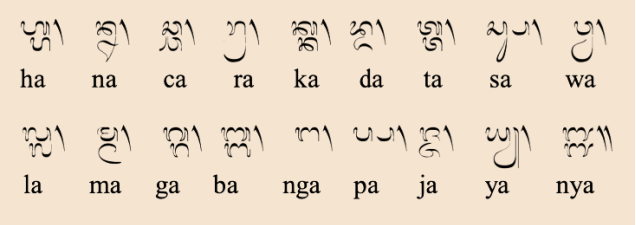
Salah satu pengembangan digital untuk media pembelajaran aksara Bali yaitu dengan membuat sebuah aplikasi yang dapat digunakan sebagai media pembelajaran baru. Dengan adanya aplikasi akan membuat masyarakat lebih tertarik untuk belajar aksara Bali. Akan tetapi sebelum membuat aplikasi, tentu perlu dikembangkan sebuah sistem untuk mendukung sebuah aplikasi tersebut. pada penelitian ini, penulis akan mengembangkan sebuah sistem klasifikasi aksara Bali berbasis suara dengan menggunakan metode KNN sebagai metode klasifikasi dan metode FastDTW sebagai metode untuk mengukur kedekatan antara dua data. Penulis akan menggunakan data berupa suara pengucapan aksara Bali khususnya aksara *Wreasta* sebanyak 18 karakter dan mengukur kemampuan metode KNN dan FastDTW dalam mengenali aksara Bali tersebut.

2. Metodologi Penelitian

2.1. Aksara Bali

Aksara Bali merupakan sekumpulan karakter, lambang atau tanda yang digunakan oleh masyarakat Bali untuk menulis aksara Bali [4]. aksara Bali dapat dibagi menjadi empat bagian yaitu aksara *Wreasta*, aksara *Swalalita*, Aksara *Wicaksara*, dan Aksara *Modre*. Akan tetapi penelitian ini hanya akan berfokus pada aksara Wreasta.

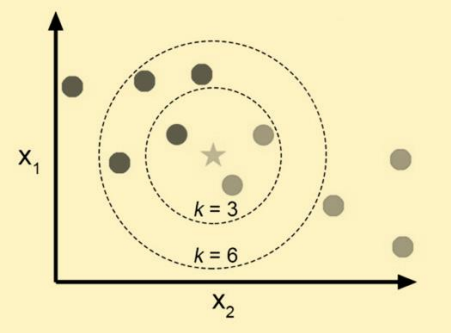
Aksara *Wreasta* merupakan aksara yang digunakan yang paling umum digunakan oleh masyarakan untuk menulis bahasa Bali[4]. Aksara *Wreasta* terdiri dari 18 karakter yang dapat dilihat pada Gambar 1.



Gambar 1. 18 Karakter dari Aksara *Wreasta*[5]

2.2. K-Nearest Neighbor

Dikutip dari penelitian yang dilakukan oleh Wu pada tahun 2009, metode *K-Nearest Neighbor* (KNN) merupakan suatu metode yang digunakan dalam proses klasifikasi yang mana dilakukan dengan mencari kelompok k objek dalam data latih yang paling mirip dengan data uji[3]. Klasifikasi ini didasarkan pada pembelajaran dengan analogi, yaitu mencari kelas terdekat dari objek yang dimiliki dengan *dataset* yang sudah ada dengan mengukur jarak dari setiap objek dengan dataset sebanyak jumlah *dataset* yang dimiliki. Untuk ilustrasi dari algoritma KNN dapat dilihat pada Gambar 2.



Gambar 2. Ilustrasi metode KNN

Adapun algoritm dari KNN secara umum adalah sebagai berikut.

1. Menentukan jumlah tetangga atau K yang digunakan untuk pertimbangan menentukan kelas dari data uji.
2. Hitung jarak dari fitur data uji dengan masing fitur data dalam dataset. Dalam penelitian ini, metode pengukuran jarak yang digunakan adalah metode fastDTW.
3. Ambil sejumlah k data terdekat
4. Pilih kelas yang paling banyak dari k data yang diambil

2.3. Fast Dynamic Time Warping

Fast Dynamic Time Warping (FastDTW) adalah versi terbaru dari *Dynamic Time Warping* (DTW) yang dikembangkan oleh Salvador dan Chan pada tahun 2007 untuk mengatasi masalah kompleksitas kuadratik yang terjadi pada DTW[1]. Menurut [1] FastDTW menggunakan pendekatan multilevel dengan tiga operasi kunci yaitu *coarsening, projection* dan *refinement*

2.3.1 *Coarsening*

FastDTW melakukan pemangkasan atau mengubah ukuran data tetapi data masih dapat merepresentasikan kurva yang sama seakurat mungkin dengan data yang lebih sedikit.

2.3.2 *Projection*

FastDTW mencari *warp path* dengan jarak minimum pada resolusi yang lebih rendah, dan gunakan jalur lusi itu sebagai tebakan awal untuk jalur lusi jarak minimum resolusi yang lebih tinggi.

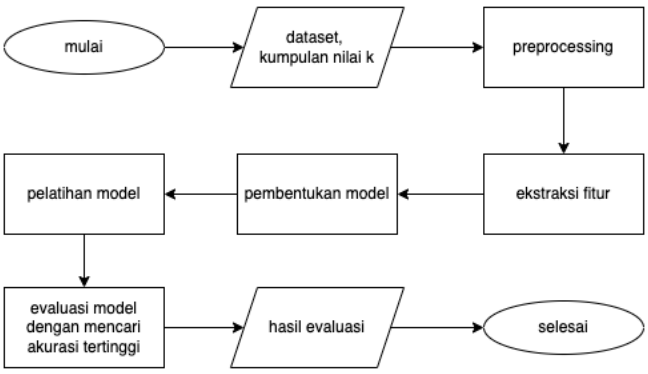
2.3.3 *Refinement*

FastDTW memperbaiki jalur lusi yang diproyeksikan dari resolusi yang lebih rendah melalui penyesuaian *warp path*.

3. Hasil dan Pembahasan

3.1. Alur Penelitian

Berdasarkan alur penelitian pada Gambar 3, proses pertama yaitu menginputkan dataset dan nilai ka yang akan digunakan sebagai parameter pengujian. dataset yang digunakan merupakan data primer berupa *file* suara berformat *waveform* (.wav). kemudian setiap data akan melalui *preprocessing* dan ekstraksi fitur untuk mendapatkan fitur dari setiap audio. Setelah itu data akan dibagi menjadi data latih dan data uji. Kemudian membuat model klasifikasi dengan algoritma KNN dengan metode pengukur jarak fastDTW. Model akan dilatih dengan data latih kemudian akan dievaluasi dengan data uji. Kelas yang didapat akan dibandingkan dengan kelas sebenarnya dari data uji sehingga didapatkan akurasi. Hasil evaluasi akan menjadi luaran dari penelitian yang dilakukan.



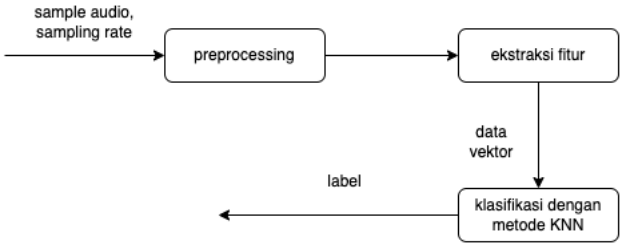
Gambar 3. Alur Penelitian

3.2. Data Penelitian

Data yang digunakan merupakan data primer yang berupa file suara dengan format *waveform* (.wav) yang diperoleh dengan menggunakan *voice recorder* bawaan smartphone. setiap orang akan direkam 27 kali untuk setiap labelnya. Label terdiri dari 18 karakter yang merupakan bagian dari aksara Wreasta. Total data yang dimiliki yaitu 486 data yang akan dibagi menjadi 70% data latih dan 30% data uji.

3.3. Desain Sistem

Pada penelitian ini penulis menggunakan metode KNN dan FastDTW untuk melakukan klasifikasi pada suara. sebelum suara di klasifikasi, diperlukan pra proses seperti penghilangan *noise* dan ekstraksi fitur untuk mengubah suara ke bentuk vektor agar suara dapat di klasifikasi. adapun alur dari program dapat dilihat pada Gambar 4.



Gambar 4. Alur klasifikasi dengan metode KNN

Dilihat pada Gambar 4, program menerima inputan berupa *sample audio* dan *sampling rate* dari data suara. Kemudian masuk ke pra proses untuk menghilangkan *noise* dari suara. setelah pra proses, *sample audio* tersebut akan di ekstraksi menggunakan metode ekstraksi fitur. Dalam penelitian ini penulis menggunakan metode MFCC dengan koefisien 13. Setelah didapatkan fitur dari audio, lanjut ke tahap klasifikasi dimana fitur tersebut dicocokkan dengan fitur data yang ada pada sistem. sesuai dengan algoritma KNN, sebanyak k data dengan jarak terdekat akan diambil dan label dari data yang paling banyak muncul akan digunakan sebagai label dari data suara yang baru.

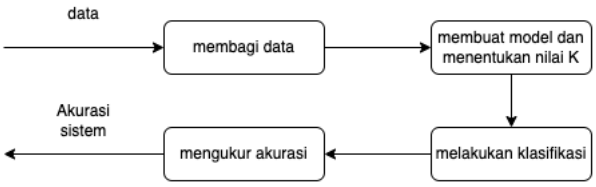
Sebelum sistem dapat melakukan klasifikasi, perlu dilakukan pelatihan terlebih dahulu untuk menghasilkan model yang paling optimal untuk melakukan klasifikasi. Pelatihan yang akan dilakukan pada penelitian ini yaitu dengan mencari nilai k terbaik dari model klasifikasi. Adapun alur dari pelatihan yang akan dilakukan dapat dilihat pada Gambar 5.

Dapat dilihat pada Gambar 5. Sistem menerima inputan berupa kumpulan data. Kemudian data tersebut akan dibagi menjadi data uji dan data latih dengan perbandingan 3:7. Kemudian sistem membuat model klasifikasi dengan metode KNN dan menentukan nilai K. setelah itu sistem akan melakukan klasifikasi dengan mengukur setiap data uji dengan data latih yang dimiliki. Kemudian sistem akan mengukur akurasi dari model dengan membandingkan label data uji dari hasil klasifikasi dengan label data uji sebenarnya. Dari hal tersebut akan didapatkan nilai akurasi dari siste.

Proses ini akan diulangi sebanyak 9 kali dengan nilai k yaitu k-3, k-5, k-7, k-9, k-11, k-13, k-15, k-17 dan k-19. Nilai K yang memiliki akurasi tertinggi akan digunakan sebagai model terbaik

3.4. Pengujian Sistem

Sistem yang dibangun oleh penulis menggunakan bahasa pemrograman *Python* pada aplikasi *Jupyter Notebook*. Adapun hasil pengujian dari penelitian ini dapat dilihat pada Tabel 1.



Gambar 5. Alur Pelatihan Sistem

Dari hasil pengujian yang telah dilakukan didapat hasil bahwa model terbaik memiliki nilai k-3,k-5,k-7 dan k-9 dengan akurasi 100% dan model dengan akurasi terendah ada pada model dengan nilai k-19 dengan akurasi 94,44%.

4. Kesimpulan

Berdasarkan penelitian yang telah dilakukan dihasilkan sebuah sistem klasifikasi dengan menggunakan metode KNN sebagai metode klasifikasi yang didalamnya terdapat metode FastDTW untuk mengukur kedekatan antara kedua data. melalui pengujian yang telah dilakukan didapat hasil bahwa model terbaik memiliki nilai K-3,K-5,K-7 dan K-9 dengan akurasi 100% dan model dengan akurasi terendah ada pada model dengan nilai k-19 dengan akurasi 94,44%. Dapat disimpulkan bahwa

penggunaan metode KNN dan FastDTW dapat digunakan dalam pengenalan aksara Bali khususnya aksara Wreasta berbasis suara.

Tabel 1. Hasil Pengujian

No	Nilai K	Akurasi	Waktu Komputasi
1	K-3	100%	6,2 detik
2	K-5	100%	6,5 detik
3	K-7	100%	6,5 detik
4	K-9	100%	6,5 detik
5	K-11	98,15%	6,4 detik
6	K-13	98,15%	6,5 detik
7	K-15	98,15%	6,4 detik
8	K-17	96,30%	6,5 detik
9	K-19	94,44%	6,4 detik

References

[1] Haviana, S.F.C. 2015. Sistem Gesture Accelerometer dengan Metode Fast Dynamic Time Warping (FastDTW). *Jurnal Sistem Informasi Bisnis*. 5, 10-21456.

[2] Indrawan, G. Paramarta, I.K., Agustini, K., Sariyasa. 2018. Latin-to-Balinese Script Transliteration Method on Mobile Application: A Comparison. *Indonesian Journal of Electrical Engineering and Computer Science*. No. 3 (Vol. 10). Hal 1331-1342

[3] Leidiyana, H. 2013. Penerapan Algoritma K-Nearest Neighbor Untuk Penentuan Resiko Kredit Kepemilikan Kendaraan Bemotor. *Jurnal Penelitian Ilmu Komputer, System Embedded & Logic.* No. 1 (Vol. 1). Hal 65-76

[4] Suasta, I.B.M, Mayun, I.B., Rupa, W. 1996. *Modernisasi dan Pelestarian : Perkembangan Metode dan Teknik Penulisan Aksara Bali*. Jakarta: Proyek Pengkajian dan Pembinaan Nilai-nilai Budaya Direktorat Sejarah dan Nilai Tradisional Direktorat Jenderal Kebudayaan Departemen Pendidikan dan Kebudayaan.

[5] Suwija, I.N. 2015. *Menulis Bahasa Bali 2*. Diktat. Denpasar: Fakultas Pendidikan Bahasa dan Seni IKIP PGRI Bali

This page is intentionally left blank.

Jurnal Elektronik Ilmu Komputer Udayana p-ISSN: 2301-5373
Volume 11, No 3. February 2023 e-ISSN: 2654-5101

Sistem Rekomendasi Film Dengan *Item-Based Collaborative Filtering* Menggunakan Flask Framework

I Made Nusa Yudiskara[a1], I Gede Santi Astawa[a2], Luh Gede Astuti[a3], Made Agung Raharja[a4], I Made Widiartha[a5], I Wayan Supriana[a6]

[a]Program Studi Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Udayana
Badung, Bali, Indonesia
[1]yudiskara@gmail.com
[2]santi.astawa@cs.unud.ac.id
[3]lg.astuti@unud.ac.id
[4]made.agung@unud.ac.id
[5]madewidiartha@unud.ac.id
[6]wayan.supriana@unud.ac.id

Abstrak

Jumlah informasi yang tersedia seiring dengan berkembangnya teknologi akan menyulitkan pengguna internet untuk mencari dan menyaring informasi di internet. Sistem rekomendasi dirancang untuk membantu pengguna menemukan data dengan cepat yang mungkin sesuai dengan data pribadi mereka dan mengurangi informasi yang berlebihan. Item atau film dalam hal ini dapat disarankan berdasarkan rating film yang diberikan oleh pengguna atau berdasarkan pengguna lain dengan kebiasaan serupa. Dalam penelitian ini penulis mengimplementasikan item-based collaborative filtering dengan membuat sistem berupa aplikasi web menggunakan python dan flask framework dimana aplikasi ini dapat memberikan user rekomendasi film berdasarkan rating yang diberikan sebelumnya oleh user. Dataset pada penelitian ini diperoleh dari website grouplens.org dengan total jumlah data rating 100.836 dari 610 user berbeda terhadap 9742 judul film. Hasil pengujian MAE dan RMSE, dengan 20 top-N similarity untuk menghitung prediksi rating, mendapatkan skor yang terbaik dengan nilai MAE 0,513536 dan nilai RMSE 0,659217674. Dengan melihat nilai RMSE dan MAE, variance error dari implementasi sistem tidak terlalu besar. Untuk pengujian pemberian ranking menggunakan Spearman Rank Correlation, IBCF dengan 20 top-N similarity dan 15 top-N similarity mendapatkan skor yang paling tinggi yaitu 0,91515 yang berarti peringkat yang diberikan sistem untuk rating prediksi sudah hampir mirip dengan peringkat rating aktual.

Kata Kunci: *Rekomedasi, Recommendation System, Collaborative Filtering, Item-based CF, Aplikasi Web*

1. Pendahuluan

Sistem rekomendasi dirancang untuk membantu pengguna menemukan data dengan cepat yang mungkin sesuai dengan data pribadi mereka dan mengurangi informasi yang berlebihan. Sistem rekomendasi yang efektif tidak hanya dapat memfasilitasi proses pencarian informasi pengguna [1], tetapi juga dapat menciptakan loyalitas pelanggan dan meningkatkan keuntungan bagi perusahaan. Dengan pentingnya peran sistem rekomendasi dalam sistem informasi online, sistem rekomendasi telah menjadi topik penelitian yang aktif dan semakin menarik perhatian dalam bidang sistem temu kembali informasi dan komunitas data mining [2].

Amazon pertama kali menciptakan dan menggunakan Item-based colaborative filtering pada tahun 1998. Daripada mencocokkan pengguna dengan pelanggan serupa, Item colaborative based filtering mencocokkan setiap produk yang dibeli dan dirating oleh pelanggan dengan item serupa, kemudian mengumpulkan item-item serupa, yang nantinya akan dihitung prediksi nilai rating dimana item dengan prediksi rating tertinggi akan dimasukkan ke dalam daftar item rekomendasi [3]. Penelitian dengan topik item based collaborative filtering sudah banyak dilakukan. Contohnya penelitian oleh Lisniati Dzumiroh dan Ristu Saptono (2016). dengan judul "Penerapan Metode Collaborative Filtering Menggunakan Rating Implisit pada Sistem Perekomendasi Pemilihan Film di Rental VCD", berdasarkan hasil penelitian, item-based collaborative filtering menghasilkan kualitas rekomendasi

yang lebih baik dibandingkan user-based collaborative filtering [4]. Nilai akurasi yang diperoleh dari item-based collaborative filtering yang diukur dengan menggunakan metrik akurasi MAE mendapatkan nilai yang lebih rendah dibandingkan user-based collaborative filtering. Peneliti juga mengkombinasikan item-based collaborative filtering dengan fitur konten, dimana peningkatan nilai akurasi tidak signifikan, dan hasil rekomendasi tidak memiliki karakteristik rekomendasi oleh collaborative filtering.

2. Metode Penelitian

2.1. Pengumpulan Data

Data pada penelitian ini menggunakan data sekunder dimana data-data tersebut berupa data film, dan rating yang di berikan pada film dari tanggal 29 Maret 1996 sampai 24 September, 2018. User dipilih secara acak, semua user setidaknya merating 20 film, tidak ada informasi demografis yang disertakan. Setiap user diwakili oleh “Id” yang diberikan, tidak ada informasi lain yang disediakan.

movieId	title	genres
1	Toy Story (1995)	Adventure|Animation|Children|Comedy|Fantasy
2	Jumanji (1995)	Adventure|Children|Fantasy
3	Grumpier Old Men (1995)	Comedy|Romance
4	Waiting to Exhale (1995)	Comedy|Drama|Romance
5	Father of the Bride Part II (1995)	Comedy
6	Heat (1995)	Action|Crime|Thriller
7	Sabrina (1995)	Comedy|Romance
8	Tom and Huck (1995)	Adventure|Children
9	Sudden Death (1995)	Action
10	GoldenEye (1995)	Action|Adventure|Thriller
11	American President, The (1995)	Comedy|Drama|Romance
12	Dracula: Dead and Loving It (1995)	Comedy|Horror
13	Balto (1995)	Adventure|Animation|Children
14	Nixon (1995)	Drama

Gambar 1. Dataset Movies

Gambar berikut merupakan data movies, yang digunakan dalam penelitian ini, movieId merupakan Id film, title merupakan judul film, dan genre yang merupakan genre film. Total data pada dataset ini adalah 9.742 data

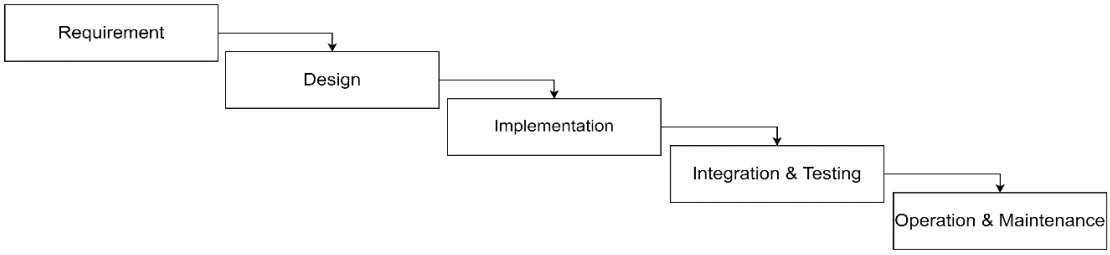
userId	movieId	rating	timestamp
1	31	2.5	1260759144
1	1029	3	1260759179
1	1061	3	1260759182
1	1129	2	1260759185
1	1172	4	1260759205
1	1263	2	1260759151
1	1287	2	1260759187
1	1293	2	1260759148
1	1339	3.5	1260759125
1	1343	2	1260759131
1	1371	2.5	1260759135
1	1405	1	1260759203
1	1953	4	1260759191

Gambar 2. Dataset Ratings

Gambar berikut merupakan data ratings, yang digunakan dalam penelitian ini, userId merupakan Id user, movieId merupakan Id film, timestamp merupakan waktu pemberian rating oleh user format yang digunakan adalah timestamp with milliseconds, data rating memperlihatkan nilai yang diberikan oleh user kepada film yang telah mereka tonton. Jumlah user pada data ini adalah 610, dan total data pada dataset ini adalah 100.836 data

2.2. Metode Pengembangan Sistem

Pada proses pengembangan sistem digunakan model *Waterfall*. Dalam metode Waterfall proses pengembangan sistem dibagi kedalam beberapa tahap, setiap tahap harus diselesaikan sebelum tahap selanjutnya dan tidak boleh ada tahap yang dilakukan secara bersamaan. Berikut merupakan ilustrasi model *Waterfall* [5].



Gambar 3. Tahapan model waterfall

Berikut merupakan tahapan dalam model waterfall :

1. Requirements
   Semua kebutuhan yang mungkin diperlukan dalam sistem yang akan dikembangkan dicatat dan didokumentasikan. Pada tahapan ini pengembang akan mendapatkan informasi kebutuhan sistem baik itu kebutuhan fungsional ataupun non-fungsional
2. Design
   Sesuai dengan kebutuhan sistem yang didapatkan pada tahap pertama. Tahap selanjutnya adalah membuat desain sistem, mencatat kebutuhan harware dan software, serta mendokumentasikan hasil desain sistem
3. Implementation
   Berdasarkan inputan yang didapatkan dari tahap *system design,* sistem pertama dikembangkan kedalam kumpulan program yang disebut *units,* yang mana akan diintegrasikan pada tahap selanjutnya. Setiap unit akan dikembangkan dan diuji kegunaannya
4. Integration and Testing
   Semua unit yang dikembangkan pada tahap implementation akan diintegrasikan pada sistem setelah dilakukan pengujian pada setiap unit. Setelah dilakukan integrasi, sistem akan diuji untuk mencari kesalahan yang mungkin terjadi dan melakukan perbaikan atas kesalahan tersebut.
5. Operation and Maintenance.
   Setelah sistem diuji baik itu fungsional maupun non-fungsional, Sistem dapat dijalankan atau dioperasikan oleh pengguna. Jika terdapat masalah yang terjadi pada saat penggunaan akan dilakukan maintenance untuk memperbaiki masalah tersebut.

2.3. Item-based Collaborative Filtering

Amazon pertama kali menciptakan dan menggunakan Item-based colaborative filtering pada tahun 1998. Daripada mencocokkan pengguna dengan pelanggan serupa, Item colaborative based filtering mencocokkan setiap produk yang dibeli dan dirating oleh pelanggan dengan item serupa, kemudian mengumpulkan item-item serupa, yang nantinya akan dihitung prediksi nilai rating dimana item dengan prediksi rating tertinggi akan dimasukkan ke dalam daftar item rekomendasi [6]. Tahap pertama dalam IBCF adalah mencari similarity antar item, kemudian berdasarkan similarity yang didapatkan item yang telah dirating akan dicari dan user akan direkomendasikan berdasarkan item yang telah dirating [7].

2.4. Pearson Correlation Coefficient

Kalkulasi similarity atau nilai kemiripan antar item merupakan langkah penting dalam *Item-Based Collaborative Filtering* [8]. Setiap pasangan item akan dihitung tingkat kemiripan similarity antar item menggunakan persamaan Pearson Correlation Coefficient, dimana nilai yang dihasilkan oleh persamaan *Pearson Correlation Coefficient* memiliki rentang nilai antara +1,0 dan -1,0 [9]. Berikut merupakan keterangan dari nilai yang dihasilkan:

- Nilai similarity 0: pasangan item tidak berkorelasi (independen)
- Nilai similarity mendekati +1,0: pasangan item memiliki korelasi tinggi
- Nilai similarity mendekati -1,0: pasangan item bertolak belakang.

Persamaan Pearson Correlation Coefficient adalah sebagai berikut:

$$r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}} \quad (1)$$

Keterangan:
r = Nilai kemiripan antara x dan y
xi = Nilai sebuah item x
yi = Nilai sebuah item y
$\overline{x}$ = Nilai rata-rata semua item x
$\overline{y}$ = Nilai rata-rata semua item y

2.5. Weighted Sum

Setelah menghitung nilai similarity dari semua pasangan item, langkah selanjutnya adalah menghitung prediksi rating yang akan diberikan pada item yang belum dirating oleh pengguna. Hasil perhitungan prediksi kemudian akan diurutkan sesuai dari nilai yang paling tinggi, kemudian sistem akan merekomendasikan item-item yang diprediksi akan dirating dengan nilai tinggi oleh user. Persamaan yang digunakan untuk menghitung nilai prediksi adalah weighted sum [10]. Untuk menghitung prediksi nilai rating yang akan diberikan user, untuk semua item yang telah dirating user,

Jumlah nilai similarity dari N-item dengan nilai similarity teratas (*Top-N Similarity*) terhadap item prediksi dikalikan dengan nilai rating yang telah dirating user pada setiap item tersebut, dibagi dengan jumlah dari semua N nilai similarity teratas.

$$P(u,i) = \frac{\sum semua\ item\ similar(S_{i,N} * R_{u,i})}{\sum semua\ item\ similar(|S_{i,N}|)} \quad (2)$$

Keterangan ;
P(u,i) = Nilai prediksi rating untuk user u pada item i.
R(u,i) = Rating yang diberikan user u pada item i.
S(i,N) = Nilai similarity untuk item i dan Top-N item similar

2.6. Pengujian Akurasi

Uji akurasi akan dilakukan menggunakan 3 metrik, yaitu MAE, RMSE, dan Spearman Correlation Coefficient. RMSE dan MAE dapat digunakan bersama untuk mengetahui variance error pada kumpulan nilai prediksi, nilai RMSE akan selalu lebih besar atau sama dengan MAE, semakin besar perbedaan antara keduanya maka semakin besar pula variance error untuk setiap individu pada sampel. Spearman Rank Correlation digunakan untuk menghitung seberapa mirip peringkat item aktual yang diberikan user dengan peringkat prediksi hasil rekomendasi sistem.
Berikut merupakan skenario pengujian akurasi sistem:

1. Persiapkan data yang berisi user serta rating yang telah diberikan pada berbagai judul film.
2. Dari data awal, sejumlah rating yang telah diberikan oleh beberapa user, akan sengaja di-hilangkan.
3. Sistem akan memprediksi nilai rating yang telah dihilangkan dan dibandingkan dengan nilai rating sebenarnya yang diberikan user untuk memperoleh nilai akurasi.
4. Penulis menghitung akurasi menggunakan 3 metrik yaitu MAE, RMSE, dan Spearman Correlation Coefficient
5. Setelah dilakukan perhitungan akurasi dilakukan, laporan hasil pengujian akan dilampirkan dalam bentuk tabel.

2.6.1 MAE

Dalam statistika, MAE (Mean Absolute Error) adalah cara untuk mengukur tingkat kesalahan antara dua buah nilai untuk kasus yang serupa. Contohnya pada perbandingan X dan Y dimana X adalah nilai prediksi dan Y adalah nilai aktual. Semakin rendah nilai MAE maka semakin baik hasil prediksi, begitu pula sebaliknya. Secara formal didefinisikan sebagai berikut:

$$MAE = \frac{\sum_{i=1}^{N}|x_i - \hat{x_i}|}{N} \quad (3)$$

i = *variable i*.
N = Jumlah data.
x_i = Nilai data aktual.
$\hat{x}_i$ = Nilai hasil prediksi.
Pengujian dilakukan pada 5 user dimana 20 rating yang sudah diberikan akan sengaja dihilangkan untuk digunakan sebagai data uji. Data tersebut kemudian dibandingkan dengan nilai sebenarnya untuk mendapatkan skor MAE

2.6.2 RMSE

RMSE (Root Mean Square Error) adalah cara yang sering digunakan untuk mengukur perbedaan nilai prediksi dan nilai aktual pada sebuah hasil prediksi, nilai RMSE selalu bernilai positif atau 0 (Hampir tidak pernah terjadi dalam praktik) yang akan menandakan nilai prediksi yang dihasilkan sempurna dan tidak memiliki kesalahan. Pada umumnya semakin kecil nilai RMSE maka semakin baik hasil prediksi. Secara formal didefinisikan sebagai berikut:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(x_i - \hat{x_i})^2}{N}} \quad (4)$$

N = Jumlah data.
x_i = Nilai data aktual.
$\hat{x}_i$ = Nilai hasil prediksi

p-ISSN: 2301-5373
e-ISSN: 2654-5101

2.6.3 Spearman Correlation Coefficient

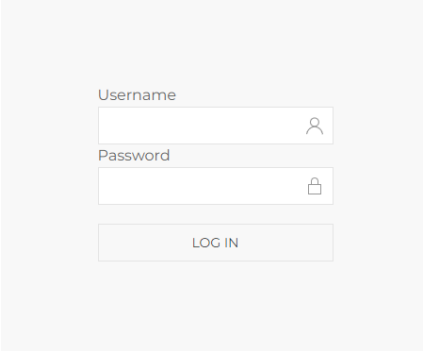
Uji Spearman merupakan salah satu uji statistik non paramateris yang digunakan untuk mengukur korelasi peringkat. Pengujian spearman akan menghitung skor seberapa baik sistem memberikan ranking dibandingkan dengan bagaimana urutan ranking yang seharusnya diberikan. Secara formal didefinisikan sebagai berikut:

$$r_s = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)} \quad (5)$$

Keterangan:
$d_i = R(X_i) – R(Y_i)$ adalah perbedaan peringkat prediksi dan peringkat sebenarnya.
n = jumlah data
Spearman rank correlation memiliki rentang nilai antara 1 dan -1. Jika nilai mendekati 1 maka urutan ranking item semakin sesuai, jika nilai mendekati -1 maka urutan ranking item berbanding terbalik dengan urutan peringkat sebenarnya, dan nilai mendekati 0 maka peringkat prediksi dan peringkat sebenarnya tidak memiliki korelasi.

3. Hasil dan Pembahasan

3.1. Implementasi antarmuka sistem



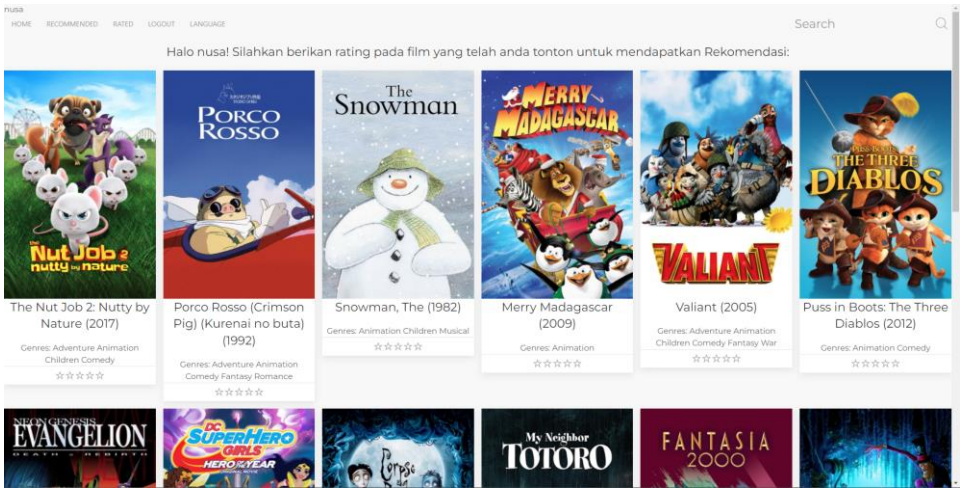
Gambar 4. Tampilan Halaman Login dan Register

Gambar 4 merupakan tampilan halaman login dan register. Halaman ini digunakan user untuk masuk ke dalam sistem tujuan dari fitur ini adalah agar user dapat melakukan peratingan dan mendapatkan rekomendasi berdasarkan rating yang diberikan oleh film-film yang sebelumnya telah ditonton oleh user. Untuk melakukan login pada sistem user harus menginput username dan password yang sebelumnya sudah didaftarkan. Apabila user belum mendaftarkan username dan password, maka user harus menuju ke halaman register untuk mendaftar

Gambar 5. Tampilan Halaman Home

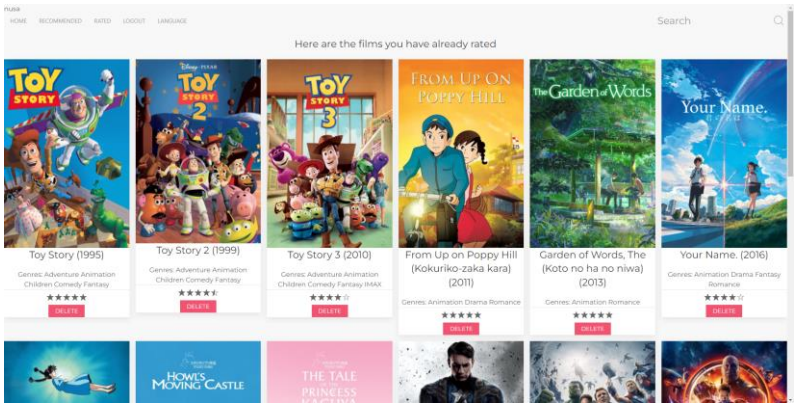
Gambar 5 merupakan tampilan halaman home. Halaman ini merupakan halaman utama pada sistem, user dapat mengakses halaman ini sebelum login ataupun register username dan password,

namun user tidak dapat melakukan peratingan dan tidak akan bisa mengakses halaman rekomendasi, karena belum memiliki akun. Pada halaman ini user dapat menjelajahi film-film yang sudah terdaftar pada database. Dalam satu halaman akan ditampilkan 18 film yang dapat dirating oleh user setelah melakukan login, untuk menampilkan halaman selanjutnya pada bagian paling bawah terdapat tombol Next dan Previous untuk menuju halaman sebelumnya. Selain itu user dapat melakukan pencarian dengan menginputkan judul atau memfilter genre film yang ingin ditampilkan



Gambar 6. Tampilan Halaman Recommendation

Gambar 6 merupakan tampilan halaman recommendation. Pada halaman ini user akan ditampilan dengan film-film yang belum dirating oleh user. setiap item pada halaman ini diurutkan berdasarkan prediksi rating paling besar yang akan diberikan user setelah menonton film tersebut



Gambar 7. Tampilan Halaman Rating

Gambar 7 merupakan tampilan halaman rating. Pada halaman ini user dapat mengubah ataupun menghapus rating yang telah diberikan pada sebuah film dengan menghapus tombol delete. Mengubah ataupun menghapus item pada halaman ini akan mempengaruhi fim-film yang akan ditampilkan pada halaman Rekomendasi

3.2. Pengujian Sistem

Tujuan dari pengujian adalah untuk memastikan semua fitur yang diimplementasikan pada sistem berjalan sesuai dengan ekspektasi dan seberapa akurat prediksi rating pada item yang belum dirating oleh pengguna, yang mana nantinya akan digunakan untuk memberikan rekomendasi film kepada pengguna. Pada penelitian ini dilakukan dua jenis pengujian yaitu Black Box Test, untuk menguji fitur yang ada pada aplikasi, dan pengujian akurasi untuk menghitung seberapa akurat rating prediksi, pengujian akurasi pada penelitian ini menggunakan 3 metrik yaitu MAE, RMSE, dan Spearman Correlation Coefficient.

p-ISSN: 2301-5373
e-ISSN: 2654-5101

3.3. Pengujian Black Box

Tabel 1. Black Box Test Skenario Pendaftaran

Ekspektasi dan Hasil Observasi (Data Benar)			
Input	**Ekspektasi**	**Hasil Observasi**	**Kesimpulan**
Username dan Password	Setelah menekan tombol register user akan diarahkan ke halaman login	User diarahkan ke halaman login	VALID
Ekspektasi dan Hasil Observasi (Data Salah)			
Input	**Ekspektasi**	**Hasil Observasi**	**Kesimpulan**
Menginput username terdaftar	Setelah menekan tombol register akan ditampilkan pesan error	User ditampilkan pesan error	VALID
Tidak menginput username dan/atau password	Setelah menekan tombol register akan ditampilkan pesan error	User ditampilkan pesan error	VALID

Tabel 2. Black Box Test Skenario Login

Ekspektasi dan Hasil Observasi (Data Benar)			
Input	**Ekspektasi**	**Hasil Observasi**	**Kesimpulan**
Username dan Password	Setelah menekan tombol login akan diarahkan ke halaman home	User diarahkan ke halaman home	VALID
Ekspektasi dan Hasil Observasi (Data Salah)			
Input	**Ekspektasi**	**Hasil Observasi**	**Kesimpulan**
Menginput username terdaftar	Setelah menekan tombol login akan ditampilkan pesan error	User ditampilkan pesan error	VALID
Tidak menginput username dan/atau password	Setelah menekan tombol login akan ditampilkan pesan error	User ditampilkan pesan error	VALID

Tabel 3. Black Box Test Skenario Logout

Ekspektasi dan Hasil Observasi			
Input	**Ekspektasi**	**Hasil Observasi**	**Kesimpulan**
Klik logout pada menu navigasi	Setelah menekan tombol logout sesi user akan selesai dan akan dikembalikan ke halaman index	Sesi user selesai dan dikembalikan ke halaman index	VALID

Tabel 4. Black Box Test Skenario Pencarian

Ekspektasi dan Hasil Observasi			
Input	**Ekspektasi**	**Hasil Observasi**	**Kesimpulan**
Kata kunci judul film	Setelah menginputkan kata kunci dan menekan enter akan ditampilkan film yang terkait dengan kata kunci	User ditampilkan film yang terkait dengan kata kunci	VALID

Tabel 5. Black Box Test Skenario Filter

Ekspektasi dan Hasil Observasi			
Input	**Ekspektasi**	**Hasil Observasi**	**Kesimpulan**
Kategori film	Setelah menginputkan kategori film dengan mencentang salah satu atau beberapa kategori film pada menu drop down pencarian kemudian tekan “Enter” pada keyboard, user ditampilkan film dengan kategori inputan kategori	User ditampilkan film dengan kategori sesuai inputan kategori	VALID

Tabel 6. Black Box Test Skenario Pencarian Dengan Filter

Ekspektasi dan Hasil Observasi			
Input	**Ekspektasi**	**Hasil Observasi**	**Kesimpulan**
Kategori film	Setelah menginputkan kata kunci dan kategori film dengan mencentang salah satu atau beberapa kategori film pada menu drop down pencarian, user ditampilkan film dengan inputan kata kunci dan kategori.	User ditampilkan film dengan kategori sesuai inputan kata kunci dan kategori	VALID

Tabel 7. Black Box Test Pemberian Rating

Ekspektasi dan Hasil Observasi			
Input	**Ekspektasi**	**Hasil Observasi**	**Kesimpulan**
Rating Film	Setelah menginputkan nilai rating dengan cara memilih jumlah bintang yang akan diberikan kepada film yang telah ditonton, data rating akan tersimpan dan ditampilkan pada halaman rated	Data rating tersimpan dan ditampilkan pada halaman rated	VALID

Tabel 8. Black Box Test Pemberian Rekomendasi

Ekspektasi dan Hasil Observasi (Data Benar)			
Input	**Ekspektasi**	**Hasil Observasi**	**Kesimpulan**
≥15 inputan rating film	Setelah memilih recommendation pada menu navigasi user akan ditampilkan sejumlah film rekomendasi dari sistem	User ditampilkan sejumlah film rek-omendasi dari sistem	VALID
Ekspektasi dan Hasil Observasi (Data Salah)			
Input	**Ekspektasi**	**Hasil Observasi**	**Kesimpulan**
≤15 inputan rating film	Setelah memilih recommendation pada menu navigasi user akan ditampilkan pesan error dan diarahkan ke halaman home untuk merating paling tidak 15 film	User ditampilkan pe-san error dan diarahkan ke halaman home untuk merating paling tidak 15 film	VALID

3.4. Hasil Pengujian Akurasi

3.4.1 MAE

Dari hasil pengujian akurasi sistem dengan metrik MAE, mengikuti skenario pengujian yang dijelaskan sebelumnya pada 2.6. dilakukan pengujian untuk top-N similarity 5, 10, 15, 20. Dengan meningkatkan jumlah nilai similarity yang digunakan untuk menghitung prediksi rating, nilai akurasi prediksi rating juga semakin baik. Top-20 Similarity mendapatkan akurasi terbaik dengan nilai MAE 0,513536 nilai tersebut 4,48% lebih baik dari Top-5 Similarity

Tabel 9. Hasil pengujian akurasi sistem dengan metrik MAE

Top-N Similarity	MAE
5	0,537652
10	0,531548
15	0,523733
20	0,513536

3.4.2 RMSE

Dari hasil pengujian akurasi sistem dengan metrik RMSE, mengikuti skenario pengujian yang dijelaskan sebelumnya pada 2.6. dilakukan pengujian untuk top-N similarity 5, 10, 15, 20. Dengan meningkatkan jumlah nilai similarity yang digunakan untuk menghitung prediksi rating, nilai akurasi prediksi rating juga semakin baik. Top-20 Similarity mendapatkan akurasi terbaik dengan nilai RMSE 0,659217674 nilai tersebut 6,79% lebih baik dari Top-5 Similarity

Tabel 10. Hasil pengujian akurasi sistem dengan metrik RMSE

Top-N Similarity	RMSE
5	0,707284891
10	0.675281266
15	0,671004491
20	0,659217674

3.4.3 Spearman Rank Correlation

Dari hasil pengujian akurasi sistem dengan menggunakan Spearman Rank Correlation, mengikuti skenario pengujian yang dijelaskan sebelumnya pada 2.6. dilakukan pengujian untuk top-N similarity 5, 10, 15, 20. Dengan meningkatkan jumlah nilai similarity yang digunakan untuk menghitung prediksi

rating, korelasi urutan peringkat tabel rekomendasi dan urutan peringkat nilai rating yang diberikan oleh pengguna semakin baik. Top-20 Similarity dan Top-15 Similarity mendapatkan nilai terbaik yaitu 0,91515 yang berarti urutan peringkat dari tabel rekomendasi pengguna memiliki korelasi tinggi dengan urutan peringkat nilai rating yang diberikan oleh pengguna memiliki korelasi tinggi.

Tabel 11. Hasil pengujian akurasi sistem dengan Spearman Rank Correlation

.**Top-N Similarity**	**Spearman Rank Correlation**
5	0,00606
10	0,491
15	0,91515
20	0,91515

4. Kesimpulan

Sistem Rekomendasi Film Dengan Algoritma Item-Based Collaborative Filtering yang telah dibangun berupa aplikasi web, telah berhasil menerapkan Algoritma Item-Based Collaborative Filtering dengan menggunakan bahasa python, web-framework flask, dan menggunakan sqlite sebagai database. Dari hasil pengujian MAE dan RMSE, yang menggunakan 20 nilai similarity tertinggi untuk menghitung prediksi rating, mendapatkan skor yang cukup baik yaitu nilai MAE sebesar 0,513536 dan RMSE sebesar 0,659217674. Selain itu RMSE dan MAE dapat digunakan bersama untuk mengetahui variance pada kumpulan prediksi, nilai RMSE akan selalu lebih besar atau sama dengan MAE, semakin besar perbedaan antara keduanya maka semakin besar pula variance error untuk setiap individu pada sampel, dari hasil RMSE dan MAE variance error dari implementasi sistem tidak terlalu besar. Untuk pengujian pemberian ranking menggunakan Spearman Rank Correlation pada top-n similarity 15 dan 20 mendapatkan skor sebesar 0,91515 yang berarti peringkat yang diberikan sistem untuk rating prediksi sudah hampir mirip dengan peringkat rating Aktual.

Daftar Pustaka

[1] I. Dwicahya, P. H. Rosa, and R. Nugroho, “Movie Recommender System Comparison of User-based and Item-based Collaborative Filtering Systems,” 2019.

[2] J. P. Verma, B. Patel, and A. Patel, “Big data analysis: Recommendation system with hadoop framework,” 2015.

[3] M. K. Kharita, A. Kumar, and P. Singh, “Item-Based Collaborative Filtering in Movie Recommendation in Real time,” 2018.

[4] L. Dzumiroh and R. Saptono, “Penerapan Metode Collaborative Filtering Menggunakan Rating Implisit pada Sistem Perekomendasi Pemilihan Film di Rental VCD,” J. Teknol. Inf. ITSmart, vol. 1, no. 2, 2016.

[5] P. W. W. Andrew Hans Ritdrix, “Sistem rekomendasi buku menggunakan metode item-based collaborative filtering skripsi,” J. Masy. Inform., vol. 9, 2018.

[6] S. Sari and D. Tri Hendra, “Aplikasi Rekomendasi Film menggunakan Pendekatan Collaborative Filtering dan Euclidean Distance sebagai Ukuran kemiripan Rating,” Semin. Nas. Teknol. Inf. dan Komun. Terap., vol. 1, no. 1, 2015.

[7] V. L. Jaja, B. Susanto, and L. R. Sasongko, “Penerapan Metode Item-Based Collaborative Filtering Untuk Sistem Rekomendasi Data MovieLens,” d’CARTESIAN, vol. 9, no. 2, 2020.

[8] M. M. Dewi and I. Artikel, “Optimasi Pearson Correlation untuk Sistem Rekomendasi menggunakan Algoritma Firefly,” J. Inform., vol. 9, no. 1, pp. 1–5, 2022,

[9] B. Prasetyo, H. Haryanto, S. Astuti, E. Z. Astuti, and Y. Rahayu, “Implementasi Metode Item-Based Collaborative Filtering dalam Pemberian Rekomendasi Calon Pembeli Aksesoris Smartphone,” Eksplora Inform., vol. 9, no. 1, 2019.

[10] A. Rosita, N. Puspitasari, and V. Z. Kamila, “REKOMENDASI BUKU PERPUSTAKAAN KAMPUS DENGAN METODE ITEM-BASED COLLABORATIVE FILTERING,” Sebatik, vol. 26, no. 1, 2022.

Jurnal Elektronik Ilmu Komputer Udayana
Volume 11, No 3. February 2023

p-ISSN: 2301-5373
e-ISSN: 2654-5101

Apllication of Naïve Bayes Algorithm in Education Games Learn to Write Aksara Bali

I Made Satya Vyasa[1)], I Gede Arta Wibawa [2)], I Gusti Ngurah Anom Cahyadi Putra [3)], I Gusti Agung Gede

Arya Kadyanan [4),] Ngurah Agus Sanjaya ER[5),] I Putu Gede Hendra Suputra [6)]

[a)]Informatics Department, Faculty of Mathematics and Natural Sciences, Udayana University
South Kuta, Badung, Bali, Indonesia
[1] satyavyasa77@gmail.com
[2] gede.arta@unud.ac.id
[3] anomcp@unud.ac.id
[4] gungde@unud.ac.id
[5] agus_sanjaya@unud.ac.id
[6] hendra.suputra[@]unud.ac.id

Abstract

Balinese script is a script from the Balinese area that was commonly used by people in ancient times to describe a word, there are many ways to keep this script sustainable and not extinct. One of them is by making a Balinese character recognition game. Which in this application the user will provide input in the form of Balinese characters written on the application, which then the input will go through a preprocessing process and then proceed with diagonal feature extraction, then the results of the feature extraction will go through a classification process using the Naïve Bayes method. The result is a web-based application that can recognize Balinese script writing using the Naïve Bayes classification method with an accuracy rate of 70.3% and get a good response from every respondent who has tested the application.

Keyword : *Balinese Scription, Diagonal Feature Extraction, Naïve Bayes.*

1. Introduction

Balinese script itself is divided into 4 types when viewed in terms of its function, namely, Wreastra script which is a Balinese script commonly used by Balinese people, both for writing place names, object names, and the like. The next is the Swalalita script. This script is used by the Balinese people in writing Kekawin (poems in Balinese culture) and also seloka derived from Kawi and Sanskrit languages. After that there is the Wijaksana script, this script is a script that is sacred by Hindus in Bali because this script is used to write religious mantras.

By utilizing the media technology that is currently developing, one of which is a smartphone. Almost everyone around the world must have this device, including the people on the island of Bali. By utilizing this technology, the learning process will run more pleasantly, Walat. W also emphasized that the use of media in the learning process is one of the efforts to create more meaningful and quality learning [7]. One of the features of smartphones that can be applied as a learning tool is games. Using games as learning media can give a relaxed and fun impression to attract students' interest in learning.

In the character recognition process, feature extraction and methods are needed to classify the data so that the data held can be separated into certain data classes. One of the classification methods is the Naive Bayes method. Based on research conducted by Arif Muhamad, this nave Bayes classification has a reasonably high accuracy rate of 85.10%.

2. Literature

2.1 Aksara Bali

Aksara comes from Sanskrit, which by type is classified as a noun neutral or sissy which means letters, syllables, or words [1]. One of the scripts used by people in Indonesia to date is the Balinese

script. Balinese script is a sign or symbol used by Balinese people to write words or sentences using Balinese, but in its development apart from being written in Balinese script, Balinese is also often written in Latin script.

Based on its function, the Balinese script can be divided into four, namely the Wreastra script, Swalelita script, Wijaksana script, and Modre script. The most common script used by Balinese people is the Wreastra script, where this script is used to write street names, building names, and the name of an agency. This Wreastra script consists of 18 consonant scripts, namely ha, na, ca, ra, ka, da, ta, sa, wa, la, ma, ga, ba, nga, pa, ja, yes, nya [1].

2.2 Color Model in Image

In applications that use the concept of digital image processing, the digital image is known into several types, namely the RGB color model which generally in this color image each pixel has a certain color, namely red (red), green (green), and blue (blue). This RGB color space can be visualized as a cube, which on the 3 axes will represent the red (red) R, green (green) G, and blue (blue) B color components. Then the grayscale or gray color model is an image with black and white color gradations that produce a gray color effect. Each pixel value of the image represents the degree of gray or white intensity. and the last is the binary color model which is a color space that only has 2 possible values for each pixel, namely black (0) and white (1). Binary images are also often referred to as B & W images (black and white) or monochrome images [5]. The process of converting RGB color to binary is needed to make it easier at the preprocessing stage, feature extraction, and data classification process.

2.3 Diagonal Feature Extraction

Feature extraction is a process that is needed to get the characteristics of an image which will later describe how the character of the object. There are many algorithms in feature extraction including diagonal. A diagonal algorithm in feature extraction is an algorithm used to get character traits from handwriting. This diagonal algorithm is done by dividing an image into several equal zones. In each zone, the point characteristics are calculated on each diagonal [3].

2.4 Naïve Bayes

The Naïve Bayes classification method is an information retrieval method that uses a probabilistic approach in its inference process, which is based on Bayes' theorem in general. The application that uses this method the most is text classification. This algorithm assumes that an attribute in an object is independent. This method assumes that the absence of a certain feature will not affect the presence of other features.

3. Research Method

3.1 Research Data

In this study, the data used is primary data which is data obtained from interviews or data collection through sources or respondents. The primary data in this study was in the form of Balinese script characters obtained through collecting data from several volunteers who wrote Balinese script characters manually or by hand. In the data collection process itself, one respondent will be asked to write 10 times per character [2].

3.1.1 Training Data

Training data is data used to train the architecture of the Naïve Bayes method. The training data used in this study were primary. Data were obtained from 10 respondents who wrote each character on the paper provided 10 times so that the total training data collected was 100 pieces of writing on each Balinese script character . The details of the characters collected are the characters in the Wreastra script, totaling 18 characters, and the voice character cast, totaling 5 characters so that the total characters are 23 characters so that the total of all training data obtained is 2300 training data. An example of the training data can be seen in figure 1.

p-ISSN: 2301-5373
e-ISSN: 2654-5101

Figure 1 Training Data

3.1.2 Testing Data

The test data is the same as the training data, namely in the form of images of Balinese script characters. Taking the image of the Balinese script character by writing the Balinese script character on the application that was built directly.

3.2 Design Method

3.2.1 Preprocessing

The initial data processing process, is divided into 2 processes, namely the process of processing test data and training data, where these two processes have the same stages but use different data. This process consists of binarization, normalization, and thining .

3.2.2 Diagonal Feature Extraction

The preprocessing results are broken down into a size of 10x10 pixels so that the image has 54 areas or features. Next, the value for each feature will be searched. For each of these features, a diagonal line will be drawn according to the size of the feature, where with a size of each feature of 10x10 pixels it will produce a total of 19 diagonal lines, of which 19 lines are sub-features. Next, the total value of all foreground pixels in the feature will be calculated. These values will be averaged to form a single value on the feature it occupies [4]. This process will be repeated until all features are scored. The entire value of this feature will be the weighted value that will be used in the classification process. The diagonal feature extraction process can be seen in figure 2.

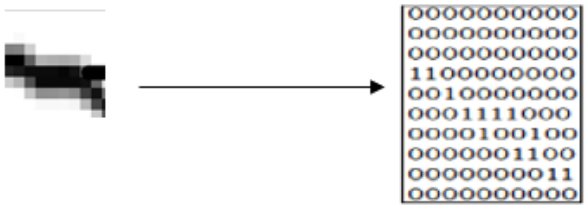


Image 2 Diagonal Feature Extraction

For example, in one of the features x with a size, the value will be searched based on the binary color it has. Number 1 shows the foreground part of the image which has black color, as can be seen in figure 3.

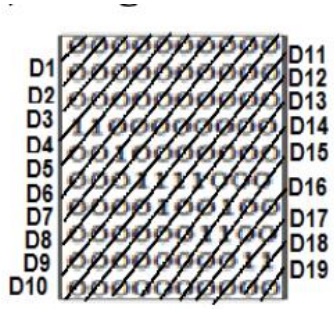


Image 3 Feature Diagonal Line

The search for the value of this feature is done by moving diagonally. Where this movement produces 19 diagonal lines or equal to 19 sub-features. The values of these sub-features will be averaged to form the value for the x feature. For example, in Figure 4 it has a feature value of 0.6842. This is

obtained from all the foreground values in the image which is worth 13, then the average feature will be searched so that.

$$Feature\ Value\ x = \frac{13}{19} = 0{,}6842$$

After this feature x value is found, it will proceed to the next feature until all feature values are obtained. After all 54 features are obtained, as many as 15 additional features are obtained from the 54 feature areas obtained previously [8]. The number of additional features is obtained through 9 features from the row average and 6 features from the column average. The number of features that are obtained after the entire process is complete is 69 feature areas.

3.2.2 Data Classification

The first step in this process is to input data from a character dataset where the data set will be used as training data in this application. Furthermore, after inputting the data, the process of calculating the amount of data in each class will be carried out, whereas in this study there were 23 classes. With each class having 90 data each, then the prior probability calculation process will be carried out.

After getting the prior probability in each class, the next process is to calculate the number of cases per class or the probability of each feature where the probability X is based on the conditions in the hypothesis H or P(X|H). After the prior probabilities and cases per class are obtained, then the probability value is multiplied by all class variables. The last step is to calculate the value and compare the results per class using equation (2.3) (posterior probability).

3.3 Application Needs Analysis

This section will explain the functional and non-functional requirements of the application to be made. Functional Requirements are application requirements that are seen from their functions. This requirement usually describes the functionality that the author expects to have in the application and to run properly. Meanwhile, non-functional requirements are application requirements for software (software) and hardware (hardware), both in making applications and in implementing the application itself. The two needs can be seen in table 1 and tables 2 and 3.

Table 1 Application Functional Requirements

No.	Functional Needs
1.	The application can display information about how to use the application.
2.	Applications can display information about the application developer.
3.	The application can be a canvas on which the user writes Balinese script.
4.	The application can apply the Bayesian naive classification algorithm when performing the classification process for the Balinese script entered by the user.

Table 2 Developer Non-functional Requirements

Type	Non-Functional Needs
Hardware	Prosessor : ryzen 5
	Kartu grafik : Intel Iris Plus
	RAM : 8GB
	Storage : 128GB
	OS : windows 10
Software	IDE :Visual Studio Code
	Image Processing Software : Adobe Illustrator
	Programming Language : Python 3.9.6, HTML, CSS, JS
	Framework : Flask, Boostrap
	Library : OpenCV, numpy, os, glob, sklearn, pickle

p-ISSN: 2301-5373
e-ISSN: 2654-5101

Table 3 User Non-functional Requirements

Type	Non-Functional Needs
Hardware	A laptop that can open a browser
Software	Web Browser : Google Chrome, Safari, Opera, Mozilla Firefox

4. Result and Discussion

4.1 App Workflow

The application flow design is used as a description of the application application from function to function. With regard to the game request, the user is faced with 3 choices, namely, how to start, start and get started. If you choose to play, the game application offers information about the playback playback. If you want the user to be confronted with information about the developer, the user will be returned to the main menu using the function by using the function, and the last selection starts the user presenting the game in the game. This start menu contains 2 more options, namely learning and training. If the user wants to learn, the user will enter the learning mode of the written Balinese script, and when it is finished, it returns to the main menu. If you, if you select the practice menu, the user is faced with the challenge mode of writing Balinese script, and when finished, returns to the main menu.

4.2 Interface Implementation

Interface ImplementationImplementation of the interface in this system will be divided into 8 pages. Can be seen in figures 4 and 5, namely the main menu page (1), how to play page (2), application page (3), game start page (4), study page (5), exercise page (6) , true page (7), and false page (8).

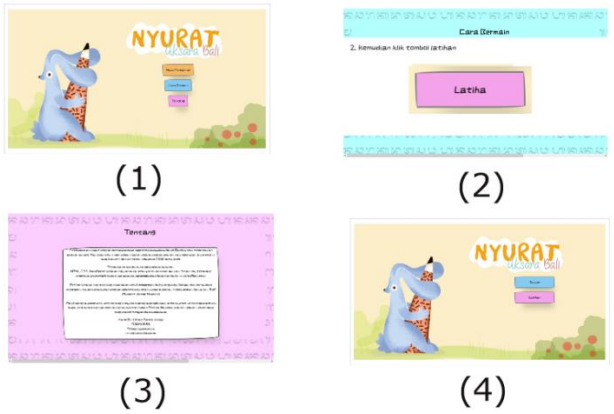


Image 4 Interface Implementation 1

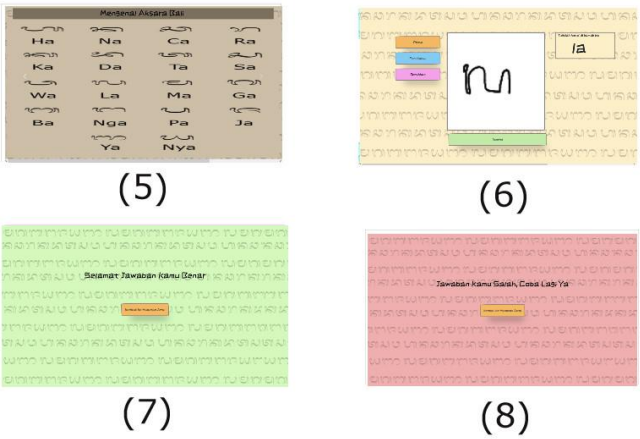


Image 5 Interface Implementation2

4.3 Testing

4.3.1 Black Box Testing

In this test, the functionality of the application will be tested whether it is following the expected expectations or not. The results of this test are as follows:

Table 3 Black Box Test Results

No	Testing	Result	Description
1.	Starting the app	Success	Can be opened and display the main page.
2.	Pressing the “Start” button	Success	The button can direct the application to the game mode select page.
3.	Pressing the “How to Play” button”	Success	The button can direct the application to the how-to page.
4.	Pressing the “About” button	Success	The button can direct the app to the about page.
5.	Pressing the “Learn” button	Success	The button can direct the application to the study page.
6.	Pressing the “Practice” button	Success	The button can direct the app to the workout page.
7.	Writing on canvas	Success	User can write on canvas.
8.	Pressing the “Eraser” button	Success	Users can change from pencil mode to eraser mode, and can erase doodles on the canvas well.
9.	Pressing the “Pencil” button	Success	Users can change the mode from eraser back to pencil mode just fine.
10.	Hit the “Clean” button	Success	Users can clean the entire canvas content automatically.
11.	Pressing the “Gather” button	Success	This button will stop writing activities and then send data to enter into the classification process.
12.	Question	Success	Questions can appear and display questions on the practice menu.
14.	Pressing the “Back to Start Page” button	Success	This button will direct the user to the start page.

p-ISSN: 2301-5373
e-ISSN: 2654-5101

4.3.2 UAT Testing (User Acceptnance Test)

This test is carried out to measure the user's perception of the application that is built. In this test, 20 respondents will assess the application built. This test is carried out by asking several questions to users who have tried this application based on a predetermined rating scale. This application is run locally and is not included in the URL link. Here are the questions and the percentage of answers from respondents who have tried this application:

Table 4 UAT Test Results (User Acceptance Test)

(User Acceptnance Test)	(User Acceptnance Test)	(User Acceptnance Test)					(User Acceptnance Test)				
		(User Acceptnance Test)	(User Acceptnance Test)	**C**	**D**	**E**	**A**	**B**	**C**	**D**	**E**
(User Acceptnance Test)	(User Acceptnance Test)	(User Acceptnance Test)	(User Acceptnance Test)	1	0	0	70%	25%	5%	0%	0%
(User Acceptnance Test)	(User Acceptnance Test)	(User Acceptnance Test)	(User Acceptnance Test)	0	0	0	50%	50%	0%	0%	0%
(User Acceptnance Test)	(User Acceptnance Test)	(User Acceptnance Test)	(User Acceptnance Test)	3	2	0	35%	40%	15%	10%	0%
(User Acceptnance Test)	(User Acceptnance Test)	(User Acceptnance Test)	(User Acceptnance Test)	0	0	0	60%	40%	0%	0%	0%
(User Acceptnance Test)	(User Acceptnance Test)	(User Acceptnance Test)	(User Acceptnance Test)	3	0	0	75%	10%	15%	0%	0%
(User Acceptnance Test)	(User Acceptnance Test)	(User Acceptnance Test)	(User Acceptnance Test)	9	1	0	30%	20%	45%	5%	0%
(User Acceptnance Test)	(User Acceptnance Test)	(User Acceptnance Test)	(User Acceptnance Test)	2	0	0	65%	25%	10%	0%	0%
(User Acceptnance Test)	(User Acceptnance Test)	(User Acceptnance Test)	(User Acceptnance Test)	1	0	0	60%	35%	5%	0%	0%

From the table above, it can be found that, as many as 70% of respondents strongly agree that the appearance of the application is attractive. Then as many as 50% of respondents strongly agree that the menus of the application are easy to understand. As many as 35% of respondents agree that the application is easy to use. As many as 60% of respondents strongly agree that the application provides examples of Balinese script and can help respondents to know the shape of the Balinese script. Then as many as 75% of respondents strongly agree that the exercise menu can help respondents to know their ability to write Balinese script. As many as 30% of respondents strongly agree that the application can improve writing skills in Balinese script. As many as 65% of respondents strongly agree that with the application that is built the experience of learning to write Balinese script becomes fun. Finally, 60% of respondents strongly agree that the application built can be a medium for learning Balinese script well.

4.3.3 Accuracy Test

This accuracy test is done by calculating the average of the results of the tests carried out. In this accuracy test, the Balinese script characters are written 20 times per character so the tests carried out are 460 times. This test is carried out directly on the application that has been built. accuracy test results can be seen in Figure 6.

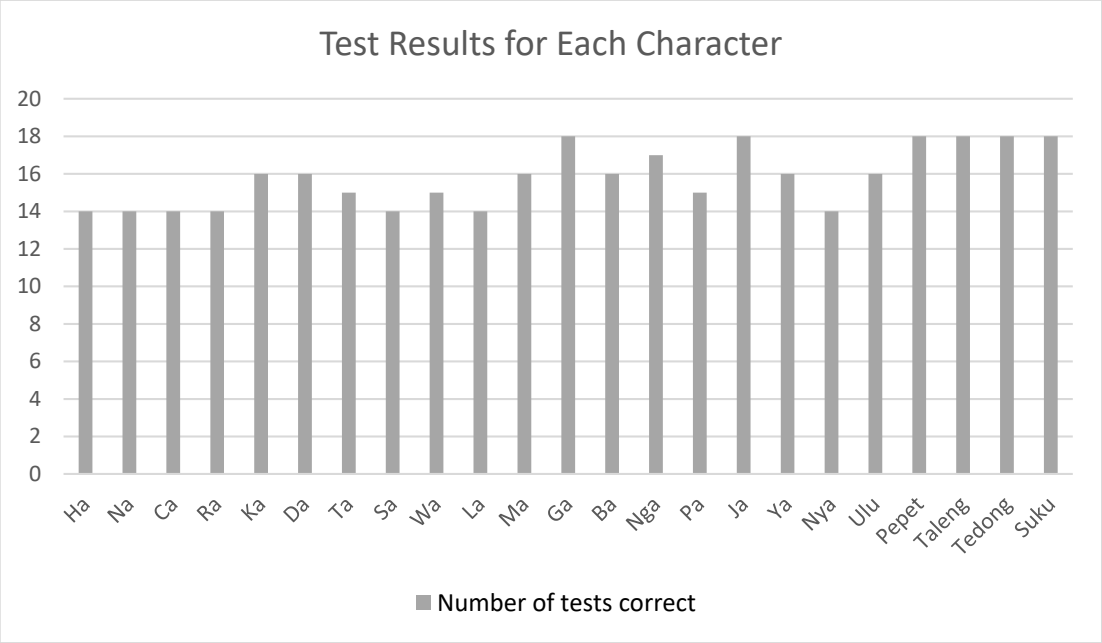


Figure 6 Accuracy Test Results

From the tests carried out, where each character was tested 20 times with a total of 364 trials which were true. It was found that the accuracy of the classification of Balinese characters using the Naïve Bayes method with diagonal feature extraction was 79.1%.

5. Conclusion

obtained are in the form of a Balinese script recognition game application that is intended for the community, both students and the general public. In testing, this research applies 3 types of testing, namely black box testing, where the results of black box testing in this study indicate that all the functions and features in this application can run well as expected. Then in the UAT test (user acceptance test), this application received a positive response from 20 respondents who had tested this application based on the questions and assessment criteria that had been determined. Furthermore, this study also tested the accuracy of the classification method used, in this study the classification method used was the Naïve Bayes method with diagonal feature extraction in classifying Balinese characters written by the user. The result of this accuracy test is the percentage of success of the Naïve Bayes algorithm in classifying Balinese characters, which is 79.1% in a total of 460 trials.

References

[1] Astra, I. S. (1982). *Prasasti Sibang Kaja di Kabupaten Badung.* Badung: Fakultas Sastra, Universitas Udayana.

[2] Cahyadhi, I. P., Sunarya, I. M., & Wirawan, I. M. (2016). Pengembangan Game Edukasi "Aksara Bali" Berbasis Android. *Kumpulan Artikel Mahasiswa Pendidikan Teknik Informatika (KARMAPATI)*.

[3] Firmansyah, M. A., Ramadhani, K. N., & Arifianto, A. (2018). *Pengenalan Angka Tulis Tangan Menggunakan Diagonal Feature Extraction dan Klasifikasi Artificial Neural Network Multiayer Perceptron*, 10.

[4] Parker, J. (2010). *Algorithms for Image Processing and Computer Vision, Second Edition.* Alberta: Wiley Publishing, Inc.

[5] Putra, D. (2010). *Pengolahan Citra Digital.* Yogyakarta: C.V ANDI OFFSET.

[6] Zang, T. Y., & Suen, C. Y. (1984). A Fast Parallel Algorithm for Thinning Digital Patterns. *Communications of the ACM*, 239.

[7] Walat, W. (2010). Conception of Media Education. *Journal of Technology and Information Education vol. 2*, 10.
[8] Win , T., Dr. , E. H., & Dr. , S. Y. (2019). *License Plate Detection and Recognition using OCR based on Morphological Operation*, 5.

This page is intentionally left blank.

Jurnal Elektronik Ilmu Komputer Udayana
Volume 11, No 3. February 2023

p-ISSN: 2301-5373
e-ISSN: 2654-5101

Sistem Klasifikasi Kelayakan Debitur Lembaga Perkreditan Desa Menggunakan Algoritma C4.5 dan Bagging

Ni Putu Novia Ardiyanti[a1], Made Agung Raharja[a2], Luh Gede Astuti[a3], I Komang Ari Mogi[a4], Luh Arida Ayu Rahning Putri[a5], I Dewa Made Bayu Atmaja Darmawan[a6]

[a]Program Studi Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Udayana
Badung, Bali, Indonesia
[1]putunovia546@gmail.com
[2]made.agung@unud.ac.id
[3]lg.astuti@unud.ac.id
[4]arimogi@unud.ac.id
[5]rahningputri@unud.ac.id
[6]dewabayu@unud.ac.id

Abstract

LPD Penarukan is one of the financial institutions that provides credit loans to people in the Penarukan traditional village area, Bali Province. This study aims to assist credit officers in classifying loans submitted by prospective debtors based on analysis of character, capacity, capital, collateral, and economic conditions by developing a credit classification system with a case study on LPD Penarukan using the C4.5 algorithm with bagging technique. Based on the results of research and testing that have been carried out, the system built is able to carry out the process of classifying credit data indicated by the results of usability testing giving the learnability component has an average value of 78.2%, the memorability component has an average value of 74.07%. The efficiency component gets an average value of 83.70%, the error component gets an average of 82.22%, the satisfaction component gets an average value of 83.89%. Tests were also carried out by calculating the accuracy and F1 score using the C4.5 algorithm with the bagging technique which resulted in an accuracy value of 81.87% and an F1 score of 89.62%.

Keywords: *Village Credit Institutions, algorithm C4.5, bagging techniques, classification, 5C analysis, usability*

1. Pendahuluan

Lembaga Perkreditan Desa (LPD) merupakan lembaga pengelola keuangan yang berada di lingkungan desa pakraman yang memberikan layanan berupa pinjaman kredit. LPD di Provinsi Bali mulai berkembang pada tahun 1985 dengan tujuan pemberdayaan desa pakraman dan memberikan pelayanan berupa pinjaman kredit kepada masyarakat [1]. Kredit menjadi salah satu pelayanan yang memberikan profit kepada LPD, namun juga beresiko akan terjadinya masalah. Berbeda dengan koperasi yang mengutamakan asas kekeluargaan anggota pada saat memberikan pinjaman kredit. Dalam memberikan kredit, LPD memperhatikan kriteria-kriteria analisis kredit yang disebut kriteria 5C yaitu *character, capital, capacity, collateral,* dan *condition of economy* [2]. LPD Desa Pakraman Penarukan merupakan lembaga keuangan milik Desa Adat Penarukan yang memberikan produk jasa seperti deposito, tabungan, pembayaran tagihan dan kredit dengan ruang lingkup yang lebih spesifik untuk masyarakat desa. Dalam kegiatan operasionalnya, LPD Desa Pakraman Penarukan mengalami permasalahan di mana untuk menganalisis data calon debitur dan menentukan kelayakan pemberian kredit masih dilakukan secara manual atau belum terdapat sistem analisis kredit dan klasifikasi nasabah seperti yang dimiliki oleh bank pada umumnya. Sehingga LPD membutuhkan waktu yang cukup lama dalam memberikan keputusan pemberian kredit kepada calon debitur.

Berdasarkan permasalahan yang dihadapi oleh LPD Desa Pakraman Penarukan, maka diperlukan suatu sistem yang dapat membantu dan mempercepat pekerjaan petugas kredit. Pada penelitian ini akan dibangun sistem berbasis *website* yang dapat mendukung dan membantu petugas kredit dalam

menganalisis data calon debitur berdasarkan kriteria penilaian 5C. Berdasarkan penelitian sebelumnya oleh Khasanah (2017) yang melakukan klasifikasi pada nasabah perusahaan keuangan dengan kelas lancar dan bermasalah menggunakan analisa kriteria 5C mendapatkan nilai akurasi sebesar 88,52% [3]. Penelitian oleh Jayanthi dkk (2017) mengenai prediksi kesehatan keuangan perusahaan memperoleh nilai rata-rata tertinggi akurasi sebesar 82,9% dan *f1-score* yaitu 82,4% pada algoritma C4.5 [4].

Berdasakan analisis pada penelitian sebelumnya, sistem yang dibangun pada penelitian ini akan menerapkan algoritma C4.5 dalam penentuan kelayakan kredit calon debitur. Algoritma C4.5 memiliki kelemahan pada data dengan ketidakseimbangan kelas di mana algoritma C4.5 secara signifikan memiliki kinerja lebih buruk [5]. Untuk mengatasi kelemahan tersebut dapat diterapkan metode *ensemble* salah satunya teknik *bagging*. Pada penelitian ini akan diterapkan pendekatan dengan teknik *bagging* pada algoritma C4.5. Diterapkannya teknik *bagging* karena tidak menimbulkan *overfitting* pada model klasifikasi, data penelitian ini terdiri atas dua kelas (layak dan tidak layak) dan meningkatkan keberagaman data latih. Data yang akan digunakan berupa data calon debitur LPD pada tahun 2018-2020 sebanyak 400 data. *Dataset* ini didominasi oleh jumlah debitur dengan kelas layak sehingga *dataset* memiliki kelas yang tidak seimbang. Teknik *bagging* melakukan resampling terhadap data latih sehingga terbentuk beberapa pohon keputusan. Dari pohon keputusan yang terbentuk, akan dilakukan pemilihan pada kelas yang paling banyak terbentuk dan menjadi pohon keputusan akhir. Dengan melakukan pemilihan pada kelas yang banyak muncul pada teknik *bagging*, dapat memberikan hasil keputusan yang lebih akurat dibandingkan hanya menggunakan satu buah pohon keputusan.

2. Metode Penelitian

Pengembangan sistem klasifikasi kredit pada penelitian ini menggunakan metode waterfall. Metode waterfall merupakan model pengembangan sistem yang memiliki fase-fase berurutan dan sistematis dengan tahapan *requirement analysis and definition, system and software design, implementation and unit testing, integration and unit testing* dan *operation and maintenance*.

2.1 Dataset

Penelitian ini menggunakan jenis data primer yang diperoleh dari di LPD Desa Pakraman Penarukan untuk mengklasifikasi kredit yang diajukan oleh calon debitur. Data penelitian yang digunakan adalah data calon debitur dari 2018-2020 yang berjumlah 400 data debitur yang di dalamnya terdiri dari informasi identitas nasabah dan penjamin, data kredit yang diajukan serta 2 label kelas yaitu layak dan tidak layak. Data debitur selanjutnya dibagi menjadi data latih dengan persentase 70% dan data uji sebesar 30% sehingga masing-masing data berjumlah 280 data dan 120 data.

2.2 Penilaian Kredit

Lembaga keuangan harus memperhatikan asas-asas perkreditan dalam memberikan kredit kepada debitur karena beresiko bagi kesehatan lembaga tersebut. Sehingga untuk mengurangi resiko, pembiayaan kredit oleh kreditur didasarkan pada kemampuan dan kesanggupan debitur. Terdapat penilaian kredit di antaranya watak, kemampuan, modal, angunan dan prospek usaha yang dimiliki debitur [2]. Penilaian tersebut kemudian berkembang dan dikenal sebagai analisis 5C dan yang dijelaskan sebagai berikut beserta inputan yang digunakan pada sistem [6].

a. *Character* (watak) yaitu sifat yang dimiliki oleh debitur, dengan inputan berupa karakter dari debitur
b. *Capacity* (kemampuan) yaitu kemampuan debitur dalam menjalankan kewajiban kredit, dengan inputan berupa pekerjaan dan kemampuan bayar debitur.
c. *Capital* yaitu sumber modal yang dimiliki oleh debitur, inputan yang digunakan yaitu kondisi aset yang dimiliki debitur.
d. *Collateral* (jaminan) adalah aset berharga yang digunakan debitur sebagai jaminan hutang atau pelunasan, inputan yang digunakan adalah perbandingan nilai CEV taksiran jaminan terhadap plafon dan jangka waktu pinjaman.
e. *Condition of economy* adalah kondisi ekonomi dari debitur.

2.3 Analisis Kebutuhan Sistem

Analisis kebutuhan dilakukan setelah mengenali masalah melalui wawancara dan pengamatan langsung terhadap sistem yang ada. Tabel 1 menjelaskan kebutuhan fungsional sistem yang akan digunakan dalam merancang dan mengembangkan sistem klasifikasi kelayakan debitur.

p-ISSN: 2301-5373
e-ISSN: 2654-5101

Tabel 1. Kebutuhan Fungsional Sistem

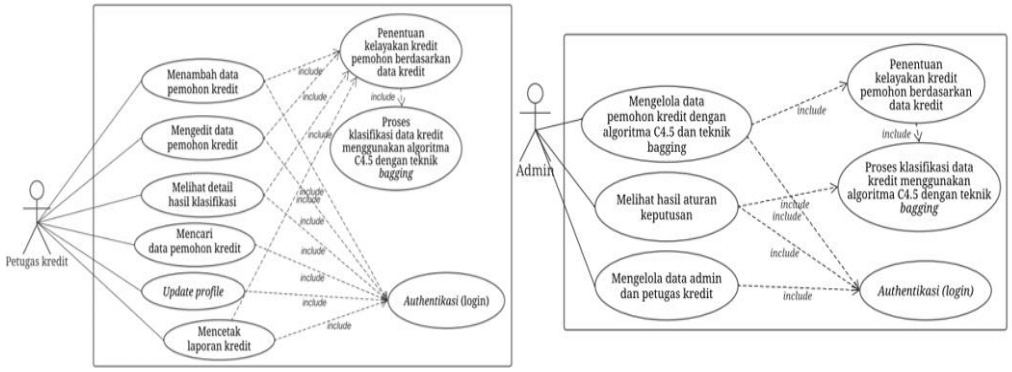
Kode	Deskripsi Kebutuhan	Target Pengguna
KF01	Sistem dapat menyediakan pengguna dengan kemampuan login dan logout, tergantung pada hak akses yang dimiliki.	Admin dan pengguna (petugas kredit)
KF02	Sistem menyediakan fitur *view* dan *update* profil.	Admin dan pengguna (petugas kredit)
KF03	Sistem dapat melakukan proses pada manajemen pengguna sistem (petugas kredit) dan data pemohon kredit.	Admin
KF04	Sistem dapat mengelola data pemohon kredit antara lain: a. Menambah pemohon kredit b. Mengedit data pemohon kredit c. Melihat detail hasil pemohon kredit	Pengguna (petugas kredit)
KF05	Sistem dapat melakukan pencarian pada data pemohon kredit berdasarkan kata kunci.	Pengguna (petugas kredit)
KF06	Sistem dapat melakukan proses klasifikasi data pemohon kredit berdasarkan data yang dimasukkan pada sistem.	Admin dan pengguna (petugas kredit)
KF07	Sistem dapat menampilkan dan mencetak laporan data pemohon kredit.	Pengguna (petugas kredit)

2.4 Perancangan Sistem

Tahapan perancangan sistem menerjemahkan analisa kebutuhan ke dalam desain perangkat lunak yang dapat dievaluasi sebelum pengkodean. Sistem dirancang menggunakan *Unified Modeling Language* (UML).

2.4.1 Use Case Diagram

Setelah memperoleh kebutuhan fungsional sistem, dilakukan pemodelan sistem menggunakan *use case diagram.* Model dari sistem dijabarkan dengan menggambarkan hubungan antara *use case*, aktor, dan sistem secara terstruktur [7]. Berikut merupakan *use case diagram* admin dan petugas kredit pada gambar 1.



Gambar 1*. Use Case Diagram*

Gambar 1 merupakan *use case diagram* untuk pengguna (petugas kredit) dan admin. Pada *use case diagram* pengguna (petugas kredit) terdapat beberapa fungsi antara lain menambah, mengedit, melihat hasil klasifikasi data pemohon kredit, mencari data pemohon kredit, *update profile*, dan mencetak laporan data pemohon kredit. Untuk fungsi menambah, mengedit, melihat detail data dan mencetak laporan kredit melibatkan proses persiapan data dan klasifikasi algoritma C4.5 dengan teknik bagging. Kemudian untuk *use case diagram* admin memiliki dua fungsi dalam mengelola data pemohon kredit, yaitu menggunakan algoritma C4.5 yang melibatkan proses klasifikasi dengan algoritma C4.5 dan mengelola data kredit yang

melibatkan proses klasifikasi algoritma C4.5 dengan teknik bagging untuk menentukan kelayakan pada pemohon kredit. Selain itu, admin juga memiliki fungsi mengelola data pengguna sistem.

2.4.2 Class Diagram

Sistem ini merancang *class diagram* dengan berbasis *model view controller* (MVC) dan *service class* yang dapat dipanggil di dalam *controller.* Pembuatan *class diagram* dengan metode tersebut disesuaikan dengan pembangunan sistem menggunakn *framework* laravel. Terdapat *class controller* yang terdiri dari RegisterController, LoginController, UserController, CustomerController, dan BaggingC45Controller. Kemudian untuk *service class* terdiri dari AlgorithmC45Service, BaggingAlgorithmsService, TreeNodeService, dan AlgorithmC45Controller. Terdapat juga *class model* antara lain User, Role, Credit, Customer, TempCredit, C45Tree, TestingResults dan TrainingData.

2.5 Algoritma C4.5

Algoritma C4.5 adalah perbaikan daripada algoritma ID3 yang memiliki kelemahan pengelolaan data dengan missing value dan mengatasi data kontinyu. Algoritma C4.5 melakukan penugasan atribut ke dalam kelas. Ini dapat diterapkan pada klasifikasi baru untuk menghitung probabilitas setiap *record* untuk suatu kategori atau mengelompokkan *record* ke dalam kelas dan mengklasifikasikannya [8]. Adapun proses dari algoritma C4.5 sebagai berikut [9].

1. Masukan *dataset*.
2. Lakukan perhitungan nilai *entropy* pada setiap atribut menggunakan persamaan 1.
3. Setelah nilai *entropy* telah ditemukan maka dapat dilakukan perhitungan untuk nilai *information gain* menggunakan persamaan 2. Selanjutnya menghitung nilai *split information* dengan persamaan 3 dan nilai *gain ratio* menggunakan persamaan 4.
4. Atribut data dengan nilai *gain ratio* tertinggi akan menjadi node akar.
5. Setelah itu lakukan kembali proses 2 dan 3. Atribut dengan *gain ratio* tertinggi akan menjadi simpul pohon. Jika sampel memiliki kelas yang sama, maka simpul tersebut akan digunakan sebagai simpul daun sesuai dengan kelas data.
6. Lanjutkan proses yang sama (secara rekursif) untuk membangun pohon keputusan untuk setiap sampel.
7. Rekursi akan berhenti jika data berada dalam kelas yang sama dan tidak ditemukan atribut lain yang dapat digunakan untuk partisi data selanjutnya.

Terdapat beberapa persamaan yang digunakan dalam proses pembentukan keputusan sebagai berikut [10].

a. Entropy

$$Entropy\ (S) = \sum_{i=1}^{n} -pi * log2pi \quad \text{.....1}$$

b. Gain

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^{n} \frac{|Si|}{|S|} \times Entropy(Si) \quad \text{.....2}$$

c. Split Information

$$SplitInfo(A) = -\sum_{i=1}^{n} \frac{|Si|}{|S|} \times log_2 \left(\frac{|Si|}{|S|}\right) \quad \text{.....3}$$

d. Gain Ratio

$$Gain\ ratio\ (A) = \frac{Gain\ (S,A)}{SplitInfo(A)} \quad \text{.....4}$$

Keterangan:

a. pi = Rasio Si terhadap S
b. S = Himpunan kasus.
c. A = Fitur
d. n = Jumlah partisi atribut A.
e. |S*i*| = Jumlah kasus pada partisi ke-i.
f. |S| = Jumlah kasus sampel S.

p-ISSN: 2301-5373
e-ISSN: 2654-5101

2.6 Teknik Bagging

Bagging (*Bootstrap Agregating*) merupakan salah satu metode *ensemble* yang bekerja dengan menciptakan beberapa sampel data baru dari data pelatihan asli. Sampel data dipilih secara acak menurut distribusi yang seragam dengan beberapa set data pelatihan yang diambil untuk sampel data dapat dipanggil kembali untuk sampel data tersebut [11]. Teknik *bagging* terdiri dari dua tahapan yaitu tahapan *bootstrap* dan tahapan *aggregating*. Proses *bootstrap* dilakukan mengambil sampel dari data pelatihan (resampling) dan proses *aggregating* yaitu menggabungkan banyak prediksi menjadi satu prediksi. Salah satu cara untuk mendapatkan nilai prediksi adalah dengan menggunakan suara terbanyak.[9].

2.7 Usability

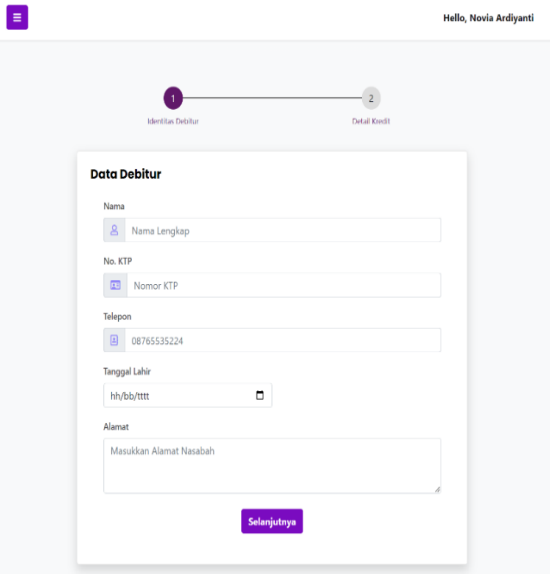
Usability adalah kunci utama HCI (*Human Computer Interaction*) yang melakukan kajian pada interaksi antara user dan sistem. Kriteria yang terdapat pada *usability* yang mencakup ruang lingkup HCI (*Human Computer Interaction*) sebagai berikut [12].

a. *Learnability* yaitu kemudahan pengguna pada saat pertama kali menggunakan sistem.
b. *Efficiency* yaitu kecepatan pengguna dalam menyelesaikan tugasnya dan kesulitan yang dihadapi pengguna.
c. *Memorability* yaitu kemungkinan pengguna dalam mengingat interaksi yang pernah dilakukan pada sistem untuk mengulangi kebenaran dan mencegah terjadinya kesalahan.
d. *Errors* merupakan kesalahan yang terdapat di dalam sistem atau kesalahan pengguna.
e. *Satisfaction* adalah tingkat kepuasan pengguna dalam sistem dan efisiensi serta efektifitas yang dirasakan pengguna saat menggunakan sistem.

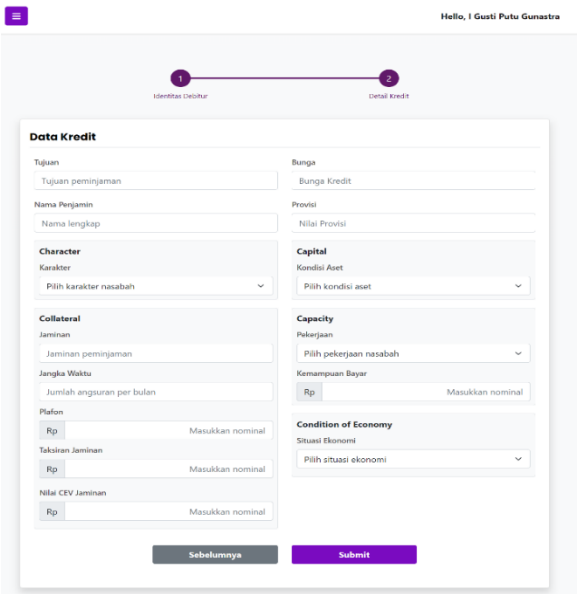
3. Hasil dan Diskusi

3.1. Implementasi Antarmuka Website

Sistem klasifikasi kredit pada penelitian memiliki tampilan antarmuka yang disesuaikan dengan rancangan yang ditampilkan pada *use case diagram*. Pada hasil implementasi, hanya ditampilkan beberapa antarmuka website yang menjadi fokus penelitian sebagai berikut.



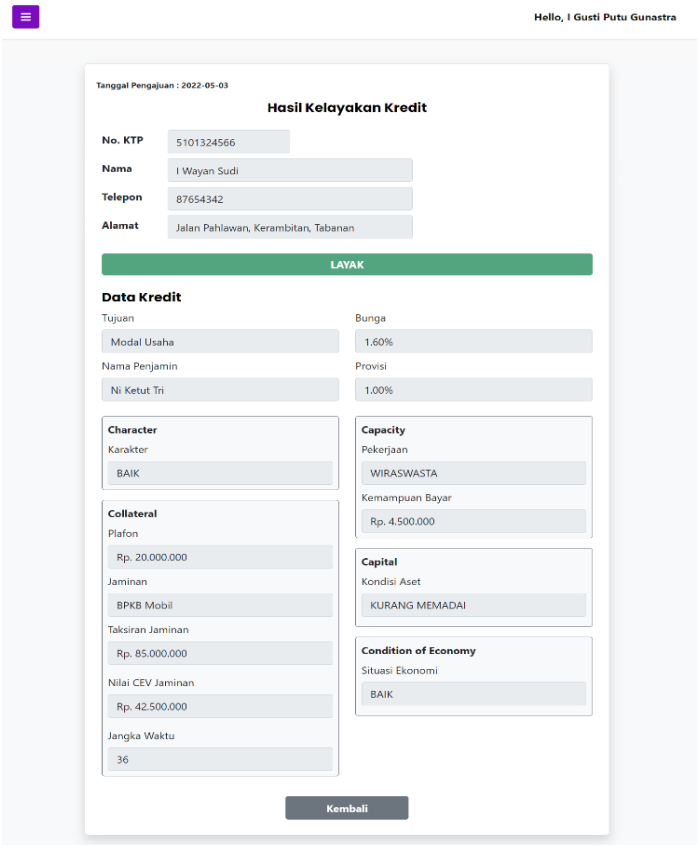
Gambar 2. Tambah Pemohon Bagian Identitas



Gambar 3. Tambah Pemohon Bagian Kredit

Gambar 2 dan 3 merupakan tampilan tambah nasabah pada bagian data debitur dan kredit. Bagian data debitur pengguna dapat memasukkan identitas lengkap debitur. Jika sudah selesai, pengguna dapat melajutkan pengisian data kredit dengan mengklik tombol selanjutnya. Kemudian, untuk bagian kredit pengguna dapat memasukkan data kredit dan mengisi sub bagian analisis 5C yang terdiri dari *character*

dengan inputan karakter, *collateral* dengan inputan plafon, taksiran jaminan dan nilai CEV jaminan serta jaminan untuk mengetahui nama jaminan yang digunakan, *capital* dengan inputan berupa kondisi aset calon debitur, *capacity* dengan inputan kemampuan bayar dan pekerjaan dari debitur serta *condition of economy* dengan inputan kondisi ekonomi calon debitur saat mengajukan kredit.



Gambar 4. Halaman Hasil Klasifikasi Kredit

Gambar 4 merupakan tampilan hasil kelayakan klasifikasi kredit setelah melakukan penambahan nasabah maupun mengedit data nasabah dan dapat diakses melalui halaman daftar nasabah pada tombol detail. Pada halaman ini terdapat data identitas debitur seperti nomor KTP, nama dan nomor telepon. Selain itu, juga terdapat data kredit yang telah diisi dan dianalisis untuk mengetahui hasil klasifikasi kredit.

3.2. Pengujian dan Evaluasi

Pengujian dari sistem klasifikasi kredit dilakukan untuk menguji ketergunaan dan ketercapaian tujuan dari sistem yang dibuat. Selain itu, tahapan ini juga dilakukan untuk menentukan ketepatan metode yang digunakan dalam mengklasifikasikan data dan memenuhi kebutuhan sistem. Pada sistem klasifikasi kredit sebagai dasar penerima kredit menggunakan algoritma C4.5 dengan teknik *bagging* dilakukan pengujian sistem berupa *usability testing*. Adapun pernyataan yang digunakan pada kuisioner sebagai berikut.

3.2.1 Pengujian Akurasi

Dalam sistem, untuk mengetahui hasil dan akurasi dari algoritma C4.5 dengan teknik *bagging* dilakukan dengan metode *confusion matrix* yang menghitung nilai akurasi dan F1 *score*. Pengujian dilakukan pada data uji sebanyak 120 data calon debitur didapatkan jumlah kelas layak yang dinilai benar sebanyak 95 dan kelas tidak layak ang dinilai benar adalah 3 data. Pengujian akurasi pada algoritma C4.5 dengan teknik *bagging* dilakukan pada pohon keputusan dari proses pelatihan data. Tabel 2 adalah hasil pengujian akurasi dan f1 *score* algoritma C4.5 dengan teknik *bagging*.

p-ISSN: 2301-5373
e-ISSN: 2654-5101

Tabel 2. Hasil Pengujian Akurasi

Pengujian	Akurasi	*F1 Score*
Algoritma C4.5 dengan teknik *bagging*	81,67%	89,62%

3.2.2 Pengujian Usability

Responden pada penelitian ini merupakan pengguna yang terdiri dari kepala LPD Desa Pakraman Penarukan, pegawai bagian kredit, pelaksana keliling, pegawai administrasi kredit dan kepala bagian dana. Dalam melakukan identifikasi dan pengujian terhadap masalah *usability* diperlukan minimal 5 sampai dengan 7 orang responden untuk hasil yang optimal pada project kecil dan 15 orang responden untuk project besar [13]. Pada penelitian ini menggunakan responden sebanyak 9 orang dikarenakan sistem yang dibangun hanya digunakan dalam lingkup LPD Desa Pakraman Penarukan dan jumlah tersebut memenuhi jumlah minimal responden untuk project dalam skala kecil. Kuesioner usability pada penelitian ini ditujunkan pada tabel 3.

Tabel 3. Kuesioner Usability

Kode	Pernyataan
Learnability	
L1	Sistem klasifikasi kredit berbasis website ini dapat dipelajari dengan mudah
L2	Saya dapat memahami tampilan tata letak sistem dan konten informasi yang disajikan dengan mudah
L3	Saya dengan mudah dan cepat menerima informasi secara detail dan juga spesifik pada sistem klasifikasi kredit
L4	Saya mampu dan mudah memahami dan mengerti alur navigasi, label dan simbol yang ada di dalam sistem
L5	Tanpa instruksi tertulis atau *manual book* saya mampu mempelajari penggunaan sistem klasifikasi kredit
Memorability	
M6	Saya dapat dengan mudah mengingat langkah-langkah penggunaan sistem
M7	Saya mudah mengingat alur navigasi, label dan simbol yang ada di dalam sistem
M8	Saya merasa mudah kapanpun menggunakan menggunakan sistem
Efficiency	
EF9	Saya mampu melakukan akses menu sistem dengan mudah
EF10	Saya dengan mudahnya dapat melakukan pencarian informasi yang ada terkait dengan sistem
EF11	Saya mampu langsung menemukan informasi yang saya ingin cari dari awal membuka sistem
Errors	
ER12	Saya tidak menemukan menu yang error atau tidak sesuai dengan fungsinya
ER13	Saya tidak menemukan eror apapun saat menggunakan website
ER14	Saya dapat menemukan fitur dan menu yang saya cari pada sistem klasifikasi kredit
Satisfaction	
S15	Saya senang dengan design antarmuka yang ada pada sistem secara keseluruhan
S16	Saya merasa nyaman dalam menggunakan sistem
S17	Panduan warna dan tata letak konten nyaman untuk dilihat
S18	Sistem klasifikasi kredit sesuai dengan ekspektasi saya ketika melihat judul yang terdapat pada halaman sistem

(Sumber: Sukmasetya dkk (2020) [14])

Dari 9 orang responden, diperoleh bahwa responden dengan usia 33 sampai dengan 41 tahun berjumlah empat responden (44,44%), usia responden 42 sampai dengan 50 tahun berjumlah 2 responden (22,22%)

dan rentang usia 51 – 59 tahun berjumlah tiga responden (33,33%). Kemudian, berdasarkan jabatan responden di LPD Penarukan, diperoleh responden dengan jabatan yaitu Kepala LPD Desa Pakraman Penarukan berjumlah 1 orang (11%), kepala bagian kredit berjumlah 1 orang (11%), pelaksana keliling/kolektor tabungan sebanyak 3 orang (33%), petugas kredit sebanyak 2 orang (22%), administrasi kredit/*customer service* berjumlah 1 orang (11%) dan kepala bagian dana berjumlah 1 orang (11%).

a. Pengujian Validitas dan Reliabilitas

Untuk mengukur kelayakan dari kuisioner, setiap pernyataan akan diuji validitas dan reliabilitasnya dengan aplikasi *IBM SPSS Statistics 25*. Analisis dan pengujian dilakukan untuk membuktikan bahwa kuesioner memenuhi persyaratan alat ukur yang baik serta intrumen penelitian yang valid dan reliabel [15].

Penelitian ini menggunkan nilai r_{tabel} sebesar 0,666 yang diperoleh dari nilai r_{tabel} dengan taraf signifikansi untuk pengujian dua arah dengan nilai 0,05 dan derajat bebas N-2 (N = 9 - 2 = 7). Berdasarkan hasil uji validitas yang dilakukan pada 18 item pernyataan, dapat disimpulan bahwa 18 item pernyataan yang terdapat pada komponen *learnability, memorability, efficiency, errors* dan *satisfaction* adalah valid dimana nilai r_{hitung} setiap pernyataan berada pada rentang 0.686-0.908, sehingga disimpulkan nilai r_{hitung} lebih besar dari r_{tabel}. Kemudian hasil dari uji reliabilitas yang telah dilakukan, diperoleh bahwa nilai *cronbach's alpha* sebesar 0,970 > 0.666, sehingga dapat disimpulkan bahwa 18 item pernyataan yang terdapat di dalam kuesioner pada lampiran 3 dapat diterima dan konsisten.

b. Analisis Pengujian Usability

Pengujian usability dilakukan dengan membagikan kuesioner kepada 9 responden penelitian yang dilaksanakan selama 3 hari. Setelah itu, dilakukan dengan perhitungan indeks persentase dari masing-masing pernyataan menggunakan penilaian yang diberikan oleh pengguna pada kuesioner. Hasil analisis pengujian usability ditunjukkan oleh tabel 4.

Tabel 4. Hasil Pengujian Usability

No.	Komponen	Kode	Skala Likert					Total	Index	Rata-rata
			1	2	3	4	5			
			STS	TS	KS	S	SS			
1	*Learnability*	L1	0	0	0	28	10	38	84.44%	78.22%
2		L2	0	0	0	28	10	38	84.44%	
3		L3	0	0	0	20	5	25	55.56%	
4		L4	0	0	0	32	5	37	82.22%	
5		L5	0	0	0	28	10	38	84.44%	
6	*Memorability*	M6	0	0	15	16	0	31	68.89%	74.07%
7		M7	0	0	0	28	10	38	84.44%	
8		M8	0	0	15	16	0	31	68.89%	
9	*Efficiency*	EF9	0	0	0	32	5	37	82.22%	83.70%
10		EF10	0	0	0	28	10	38	84.44%	
11		EF11	0	0	0	28	10	38	84.44%	
12	*Errors*	ER12	0	0	0	32	5	37	82.22%	82.22%
13		ER13	0	0	0	32	5	37	82.22%	
14		ER14	0	0	3	24	10	37	82.22%	
15	*Satisfaction*	S15	0	0	0	28	10	38	84.44%	
16		S16	0	0	0	32	5	37	82.22%	83.89 %
17		S17	0	0	0	28	10	38	84.44%	
18		S18	0	0	0	28	10	38	84.44%	

Berdasarkan hasil pengujian usability yang terlihat pada tabel 3 dari 9 responden dengan 18 pernyataan diperoleh rata-rata indeks persentase setiap komponen sebagai berikut.

1. Komponen learnability memiliki nilai rata-rata sebesar 78.22% yang berada pada kategori responden setuju terhadap penggunaan sistem klasifikasi kredit yang mudah dipelajari dan tampilan tata letak sistem dan konten informasi yang disajikan mudah dimengerti.
2. Komponen memorability memiliki nilai rata-rata sebesar 74.07 % yang masuk ke dalam kategori responden setuju dengan kemudahan mengingat alur navigasi, label dan simbol yang ada di dalam sistem.
3. Komponen efficiency memperoleh nilai rata-rata sebesar 83.70% yaitu pada kategori responden sangat setuju dengan kemudahan dalam melakukan pencarian dan menemukan informasi yang terkait dengan sistem.
4. Komponen errors memperoleh rata-rata sebesar 82.22% yang berada pada kategori responden sangat setuju bahwa tidak terdapat error pada sistem.
5. Komponen satisfaction mendapatkan nilai rata-rata sebesar 83.89 % yang berada pada kategori responden sangat setuju dan puas terhadap desain antarmuka yang ditampilkan, pemilihan warna dan tata letak serta kesesuaian antara judul halaman dan informasi yang diberikan.

4. Kesimpulan

Penerapan algoritma C4.5 dengan teknik *bagging* pada sistem klasifikasi nasabah LPD Desa Pakraman Penarukan dapat memberikan klasifikasi nasabah berdasarkan analisis 5C (*character, capital, capacity, collateral* dan *condition of economy*). Pengujian dilakukan pada 120 data uji algoritma C4.5 dengan teknik *bagging* menghasilkan akurasi sebesar 81,87% dan F1 *score* sebesar 89,62%. Pengujian usability dengan 9 responden penelitian yang berada di lingkungan LPD Desa Pakraman Penarukan menghasilkan komponen *learnability* memiliki nilai rata-rata sebesar 78,2%, komponen *memorability* memiliki nilai rata-rata sebesar 74,07 %. Komponen *efficiency* memperoleh nilai rata-rata sebesar 83,70%, komponen *errors* memperoleh rata-rata sebesar 82,22%, komponen *satisfaction* mendapatkan nilai rata-rata sebesar 83,89 %. Hasil pengujian *usability* secara keseluruhan menunjukkan bahwa seluruh komponen memiliki nilai rata-rata indeks setiap komponen berada di atas nilai 70%, sehingga dapat dikatakan bahwa sistem klasifikasi calon debitur yang telah dibuat memiliki nilai dalam aspek *usability* dan sangat mudah dipelajari dan dipahami oleh pengguna.

References

[1] Keputusan Gubernur Bali, *Surat Keputusan Kepala Daerah Tingkat I Bali Nomor 972*. 1984, pp. 0–6.

[2] Presiden Republik Indonesia, *Undang-Undang Republik Indonesia Nomor 10 Tahun 1998*. Indonesia, 2014.

[3] S. N. Khasanah, “Penerapan Algortima C4.5 Untuk Penentuan Kelayakan Kredit,” *J. Techno Nusa Mandiri*, vol. 14, no. 1, pp. 9–14, 2017.

[4] J. Jayanthi, G. Kaur, and K. S. Joseph, “Financial Forecasting Using Decision Tree (REPTree & C4.5) and Neural Networks (K*) For Handling The Missing Values,” *ICTACT J. Soft Comput.*, vol. 7, no. 3, pp. 1473–1477, 2017, doi: 10.21917/ijsc.2017.0204.

[5] I. Brown and C. Mues, “An experimental comparison of classification algorithms for imbalanced credit scoring data sets,” *Expert Syst. Appl.*, vol. 39, no. 3, pp. 3446–3453, 2012, doi: 10.1016/j.eswa.2011.09.033.

[6] I. W. Supriana, M. A. Raharja, and P. W. Gunawan, “Sistem Informasi Prediksi Penilaian Kredit Perbankan Menggunakan Algoritma K-Nearest Neighbor Classification,” *JST (Jurnal Sains dan Teknol.*, vol. 8, no. 1, pp. 44–54, 2019, doi: 10.23887/jstundiksha.v8i1.16470.

[7] F. Rohman and D. Kurniawan, “Pengukuran Kualitas Website Badan Nasional Penanggulangan Bencana Menggunakan Metode Webqual,” *J. ILMU Pengetah. DAN Teknol. Komput.*, vol. 3, no. 1, pp. 31–38, 2017.

[8] A. Ridwan and A. T. Khoiriyah, “Penerapan Teknik Bagging Pada Algoritma Naive Bayes dan Algoritma C4.5 Untuk Mengatasi Ketidakseimbangan Kelas,” *J. BISNIS Digit. DAN Sist. Inf.*, vol. 1, no. 1, pp. 41–48, 2020, [Online]. Available: https://ejr.stikesmuhkudus.ac.id/index.php/BIDISFO/article/view/914.

[9] E. Prasetyo and B. Prasetiyo, “Increased Classification Accuracy C4 . 5 Algorithm Using Bagging Techniques in Diagnosing Heart Disease,” *J. Teknol. Inf. dan Ilmu Komput.*, vol. 7, no. 5, pp.

1035–1040, 2020, doi: 10.25126/jtiik.202072379.

[10] M. Mirqotussa'adah, M. A. Muslim, E. Sugiharti, B. Prasetiyo, and S. Alimah, "Penerapan Dizcretization dan Teknik Bagging Untuk Meningkatkan Akurasi Klasifikasi Berbasis Ensemble pada Algoritma C4.5 dalam Mendiagnosa Diabetes," *Lontar Komput. J. Ilm. Teknol. Inf.*, vol. 8, no. August, p. 135, 2017, doi: 10.24843/lkjiti.2017.v08.i02.p07.

[11] M. U. Nuha, I. Arieshanti, and Y. Purwananto, "Pengembangan Perangkat Lunak Prediktor Kebangkrutan Menggunakan Metode Bagging Nearest Neighbor Support Vector Machine," *J. Tek. Pomits*, vol. 1, no. 1, pp. 1–6, 2012, [Online]. Available: https://pdfslide.tips/documents/pengembangan-perangkat-lunak-prediktor-kebangkrutan-dapat-mencegah-atau.html.

[12] M. A. Raharja, S. Purnawati, I. P. G. Adiatmika, I. N. Adiputra, and I. B. A. Swamardika, "Usability Analysis of Tembang Sekar Alit Learning (SekARAI) Applications Using The Human Computer Interaction (HCI) Model In Bali Students," *Proc. Second Asia Pacific Int. Conf. Ind. Eng. Oper. Manag.*, pp. 2870–2879, 2021.

[13] J. Lazar, J. H. Feng, and H. Hochheiser, *Research Methods in Human-Computer Interaction, 2nd Edition*, Second Edi. Cambrigde: Morgan Kaufmann, 2017.

[14] P. Sukmasetya, A. Setiawan, and E. R. Arumi, "Penggunaan Usability Testing Sebagai Metode Evaluasi Website Krs Online Pada Perguruan Tinggi," *JST (Jurnal Sains dan Teknol.*, vol. 9, no. 1, pp. 58–67, 2020, doi: 10.23887/jst-undiksha.v9i1.24691.

[15] B. O. Lubis, A. Salim, and J. Jefi, "Evaluasi Usability Sistem Aplikasi Mobile JKN Menggunakan Use Questionnaire," *J. SAINTEKOM*, vol. 10, no. 1, p. 65, 2020, doi: 10.33020/saintekom.v10i1.131.

Jurnal Elektronik Ilmu Komputer Udayana
Volume 11, No 3. February 2023

p-ISSN: 2301-5373
e-ISSN: 2654-5101

Pengembangan Sistem Pengenalan Karakter Aksara Suku Simalungun Berbasis Android

Theresia Seftiani Girsang[a1], I Dewa Made Bayu Atmaja Darmawan[a2], Ngurah Agus Sanjaya ER[a3], AAIN Eka Karyawati[a4], I Putu Gede Hendra Suputra[a5], Cokorda Rai Adi Pramartha[a6]

[a]Program Studi Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Udayana
Badung, Bali, Indonesia
[1]theresiagirsang89@gmail.com
[2]dewabayu@unud.ac.id
[3]agus_sanjaya@unud.ac.id
[4]eka.karyawati@unud.ac.id
[5]hendra.suputra@unud.ac.id
[6]cokorda@unud.ac.id

Abstract

Simalungun script is the script used by the Simalungun people to communicate with each other in their time. But over time, this character is rarely used. Therefore, the Simalungun script needs special attention because it is already threatened with extinction due to limited data and information. To overcome this, it is necessary to use the role of information technology. In this study, a system was built that can classify and introduce the Simalungun tribal characters using the Android-based Convolutional Neural Network (CNN) method. In addition, in this study, CNN MobileNetV2 and TensorFlow Lite architectures were used for deploying android needs. From the results of the training using the MobileNetV2 architecture by testing 29 characters, the accuracy results are 81% and the test results are 82%. In testing the feasibility of the application, the author uses the concept of usability testing by involving 15 respondents and giving a percentage of 78%.

Keywords: *Simalungun Script, Image Classification, Convolutional Neural Network, Android, Usability Testing*

1. Pendahuluan

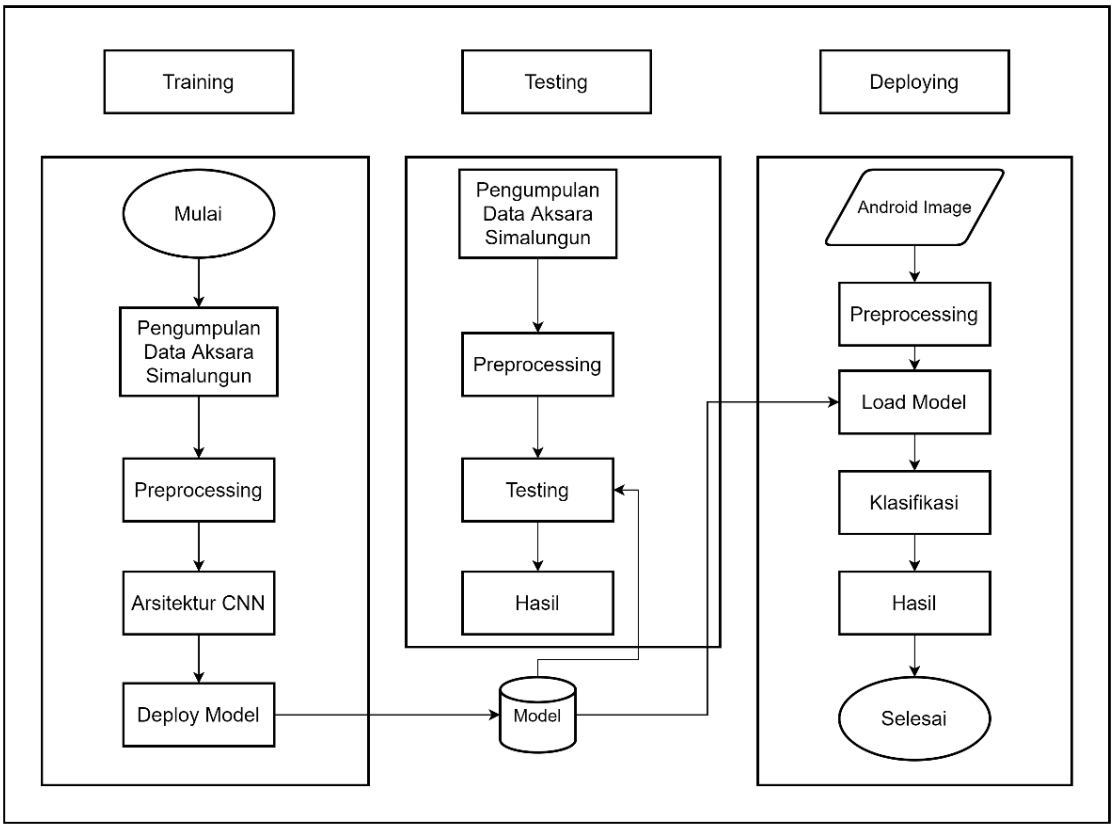
Aksara Simalungun merupakan aksara yang digunakan oleh masyarakat Simalungun untuk saling berkomunikasi. Aksara Simalungun terdiri dari Induk Surat dan Anak Surat. Induk Surat merupakan sembilan belas huruf aksara simalungun. Anak Surat merupakan tanda-tanda atau diakritik yang dapat mengubah nilai atau bunyi dari Induk Surat. Aksara Simalungun ini adalah salah satu rumpun dari Aksara Batak yang memerlukan perhatian khusus [1]. Hal ini dikarenakan perkembangan teknologi informasi yang sangat pesat, sehingga perlahan-lahan aksara ini mulai ditinggalkan. Aksara Simalungun merupakan warisan budaya yang pantas untuk tetap dijaga dan dilestarikan. Untuk mengatasi permasalahan tersebut maka perlu adanya peran teknologi informasi yang digunakan. Machine Learning dan Deep Learning merupakan teknologi kecerdasan buatan yang popular saat ini dan dapat digunakan untuk mengatasi permasalahan tersebut [2]. Convolutional Neural Network merupakan salah satu metode Deep Learning dari pengembangan Multi-Layer Perceptron yang didesain untuk mengolah data dua dimensi, yaitu gambar dan suara.

Penelitian lainnya dilakukan oleh Susilo, dkk pada tahun 2017, dimana pada penelitian tersebut menggunakan metode CNN dalam pengenalan pola karakter bahasa jepang hiragana. Pada penelitian tersebut arsitektur jaringan CNN menghasilkan akurasi yang baik mencapai 96,2 %. Metode *Convolutional Neural Network* dianggap baik dan mampu dalam mengenali berbagai bentuk citra. Penelitian diatas menjadi dasar usulan peneliti untuk membuat suatu teknologi yang dapat digunakan dalam pengenalan karakter Aksara Simalungun [3]. Berdasarkan latar belakang permasalahan di atas, maka penelitian kali ini akan membangun sebuah sistem yang dapat mendeteksi dan mengenalkan aksara suku Simalungun dengan lebih baik. Penelitian ini akan dilakukan dengan menggunakan metode *Convolutional Neural Network* (CNN). Supaya dapat membantu melestarikan warisan budaya

supaya tidak punah. Peneliti menganggap hal ini penting untuk dilakukan. Oleh karena itu peneliti akan membuat suatu sistem berbasis *android* sederhana sebagai aplikasi pengenalan karakter aksara Simalungun.

2. Metode Penelitian

Pada Gambar 1. terdapat tiga proses yang akan dilakukan dalam penelitian ini. Proses pertama yaitu *training* (pelatihan), proses kedua *testing* (pengujian) dan proses yang terakhir *deploying* (penyematan).



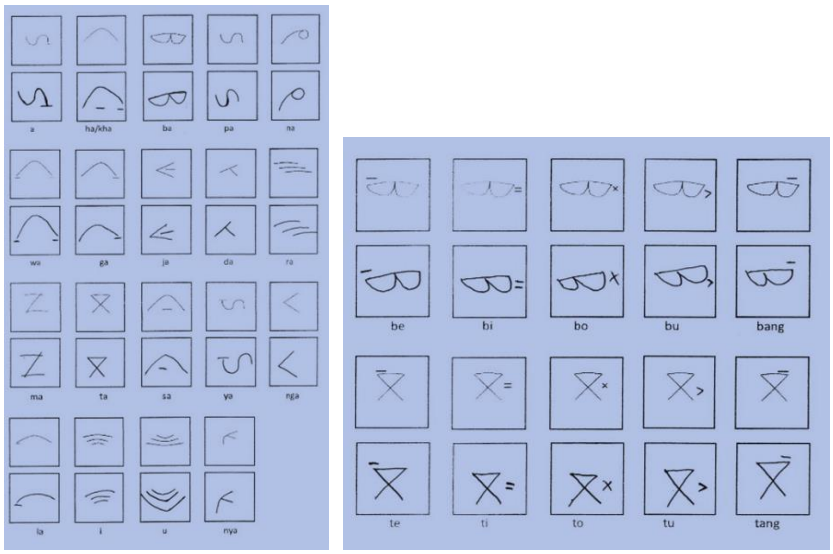
Gambar 1. Alur Sistem Pengenalan Karakter Aksara Suku Simalungun

Pada tahap training (pelatihan) terdiri dari pengumpulan data Aksara Simalungun yang diambil langsung dari beberapa responden terkait, proses preprocessing data yang siap diolah ke proses klasifikasi, proses klasifikasi menggunakan arsitektur CNN dan menghasilkan model yang siap untuk di testing. Kemudian pada tahap testing (pengujian) akan dilakukan pengujian terhadap model yang sudah dilatih pada tahap training. Jika model yang dihasilkan sudah memberikan akurasi yang cukup baik pada saat testing, maka akan dilanjutkan dengan proses deploying (penyematan). Tahap deploying (penyematan) adalah tahap yang berisi proses konversi klasifikasi CNN pada aplikasi android. Dimana pada tahap ini akan menghasilkan aplikasi yang dapat memberikan prediksi deskripsi karakter aksara.

2.1 Pengumpulan Data

Data primer adalah data yang diperoleh langsung dengan cara mengambil sendiri tanpa adanya perantara. Sedangkan data sekunder adalah data yang sebelumnya sudah dikumpulkan oleh orang lain, sehingga peneliti boleh meminta data yang sudah ada tersebut kepada instansi atau organisasi [4]. Data yang digunakan dalam penelitian ini adalah data primer. Pengumpulan data akan dilakukan dengan memberikan sampel data kepada 33 orang responden. Kemudian responden akan menulis sesuai dengan sampel data yang diberikan. Responden yang dituju adalah siswa/siswi tingkat sekolah dasar kelas VI (enam). Data tersebut merupakan scan citra karakter Aksara Simalungun berjumlah 29 karakter, yang mana terdiri dari 19 karakter induk huruf dan dengan tambahan 10 karakter anak huruf. Berikut dapat dilihat 29 karakter Aksara Simalungun. Contoh data dapat dilihat pada Gambar 2.

p-ISSN: 2301-5373
e-ISSN: 2654-5101



Gambar 2. (a) Induk Huruf, (b) Anak Huruf

Gambar 2. menunjukkan data yang akan dikumpulkan dari seluruh responden. Data yang berhasil dikumpulkan terdiri dari 33 variasi jenis tulisan tangan yang berbeda. Setiap responden menuliskan sebanyak lima kali penulisan aksara sesuai sampel yang diberikan. Oleh karena itu, maka total keseluruhan data aksara yang terkumpul adalah sebanyak 4.640 karakter. Kemudian untuk validasi daripada tulisan karakter Aksara Simalungun ini secara detail diperiksa oleh guru aksara yang bersangkutan.

2.2 Pra Proses Citra Input

Praproses citra input ini dilakukan pada saat implementasi dalam penulisan kode program (*coding*) dengan bahasa pemrograman Python. Tujuan praproses input adalah untuk menyesuaikan format citra supaya pada saat citra memasuki arsitektur CNN dapat terbaca dengan baik. Tahapannya sebagai berikut:

1. Mengubah data citra (.png) ke dalam bentuk array.
2. Mendapatkan fitur (X) dan label (y) citra, hasilnya menjadi *variable* berupa *X_train, y_train, X_test, y_test*.
3. Normalisasi (*feature scaling*) untuk *X_train* dan *X_test*, yaitu mengubah rentang nilai 0-225 menjadi 0-1.
4. *One-hot-encoding* untuk *y_train dan y_test*, yaitu mengubah setiap nilai di dalam kolom menjadi kolom baru dan mengisinya dengan nilai biner yaitu 0 dan 1.
5. Hasil akhir berupa *X_train_norm* (*X_train* yang telah di normalisasi), *y_train_encode* (*y_train* yang telah di *encoding*), *X_test_norm* (*X_test* yang telah di normalisasi), *y_test_encode* (*y_test* yang telah di *encoding*).
6. Selanjutnya variabel yang merepresentasikan citra tersebut sudah siap diolah kedalam algoritma CNN untuk dilakukan proses *training* kemudian *testing*.

2.3 Arsitektur CNN: MobileNetV2

Dalam penelitian ini, peneliti memilih arsitektur CNN MobileNetV2. Dengan alasan selain memang score akurasinya cukup tinggi, juga yang menjadi keunggulan utamanya adalah jumlah training parameters yang kecil dibandingkan dengan arsitektur CNN lainnya. Sehingga kebutuhan akan komputasinya lebih ringan. Oleh karena itu, jika model tersebut akan di deploy ke dalam sebuah real app, misalnya aplikasi android ataupun aplikasi berbasis website akan ringan dan berukuran kecil [5].
Pada penelitian ini akan digunakan tipe arsitektur CNN *MobileNetV2* dengan tambahan *dropout*. Arsitektur CNN akan dilatih (*training* dan *validation*) terhadap dataset yang sudah tersedia. Selain itu, juga akan diterapkan *hyperparameter* berupa *learning rate* dengan nilai 0.0003 yang akan menghasilkan model CNN. Model tersebut akan diujikan (*testing*) terhadap data test untuk kemudian dievaluasi hasilnya, jika hasilnya sudah cukup baik maka yang terakhir adalah men-*deploy*nya ke aplikasi siap pakai.

2.4 Perancangan Sistem

Dalam penelitian ini sistem yang akan dikerjakan merupakan sistem berbasis android dengan menggunakan bahasa pemrograman Kotlin. Sistem dibangun menggunakan model tensorflow lite, dimana model ini sudah dikonversi dari bahasa pemrograman Python 3.9.0 dengan tambahan library yang sudah tersedia pada Google Collaboratory. Tujuan utama pembuatan sistem ini adalah untuk

memudahkan pengenalan karakter Aksara Simalungun menggunakan metode Convolutional Neural Network. Hasil dari proses ini adalah berupa deskripsi karakter aksara.

Sistem dirancang untuk mengklasifikasi masing-masing karakter dari Aksara Simalungun berdasarkan citra input menggunakan teknologi Deep Learning metode Convolutional Neural Network. Karakter Aksara Simalungun yang akan diklasifikasikan terdiri dari 19 induk huruf dan 10 anak huruf. Secara keseluruhan, proses klasifikasi citra Aksara Simalungun ini melalui beberapa tahapan. Pertama, dilakukan pengumpulan dataset citra Aksara Simalungun secara langsung. Kedua, akan dilakukan praproses dataset agar data citra sesuai dengan kriteria standar saat memasuki arsitektur CNN.

Praproses dataset ini dilakukan dengan harapan menghasilkan model CNN yang baik. Tahap ketiga, dataset akan dipecah menjadi data train dan data test, dimana data train nantinya akan digunakan untuk melatih model dan data test nantinya akan digunakan untuk menguji model. Untuk mengevaluasi model digunakan K-Fold Cross Validation. Kemudian untuk mengevaluasi model klasifikasi digunakan Confusion Matrix. Setelah model selesai dievaluasi, maka akan dikonversi menggunakan tensorflow lite dan akan menghasilkan model.tflite yang selanjutnya dilakukan deploy ke Android Studio. Proses yang selanjutnya berjalan adalah implementasi aplikasi menggunakan model.tflite pada Android Studio.

2.5 Evaluasi Sistem

Untuk evaluasi performansi hasil klasifikasi dalam penelitian ini, maka digunakan teknik confusion matrix dan classification report. Dalam confusion matrix akan diperlihatkan tabel yang berisi jumlah prediksi benar dan salah dari model klasifikasi. Dalam confusion matrix hasil yang ditampilkan sudah cukup detail, namun jika ditambah dengan classification report akan semakin menambah pemahaman hasil klasifikasi dan dapat mengukur kinerja model.

Pengujian menggunakan *usability testing* yang merupakan tingkat suatu produk yang digunakan oleh user untuk mencapai target yang ditetapkan, seperti efektifitas, efisiensi dan mencapai kepuasan pengguna pada suatu aplikasi [6]. Kriteria yang dapat diukur dalam pengujian *usability* ini yaitu sebagai berikut:

a. Kemudahan (l*earnability*), mendefinisikan seberapa cepat user dalam menggunakan sistem dan kemudahan dalam penggunaan aplikasi.
b. Efisiensi (*efficiency*), mendefiniskan sebagai sumber daya yang dikeluarkan guna mencapai ketepatan dan kelengkapan tujuan.
c. Mudah diingat (*memorability*), mendefinisikan bagaimana kemampuan user dapat mempertahankan pengetahuannya setelah jangka waktu tertentu.
d. Kepuasan (*satisfaction*), mendefinisikan sikap positif user terhadap aplikasi.

Dari keempat aspek *usability* tersebut, diambil 15 responden dengan kriteria responden seperti berikut:

a. Responden berasal dari daerah Simalungun.
b. Responden berusia 18 – 30 tahun.
c. Responden dengan jenis kelamin laki-laki dan perempuan.
d. Responden yang mengerti dan memahami tentang Aksara Simalungun.
e. Responden yang memahami teknologi dan informasi.

Ketika seluruh kriteria responden terpenuhi, maka 15 responden tersebut akan diarahkan untuk mengikuti sesi latihan dalam menggunakan aplikasi. Peneliti akan memberikan panduan dalam menggunakan aplikasi kepada seluruh responden. Hal ini diberikan supaya responden dapat lebih mengerti dalam menggunakan aplikasi dan dapat mengisi kuesioner yang telah disediakan dengan penilaian yang valid.

3. Hasil dan Pembahasan

3.1 Implementasi CNN

Proses pelatihan (training) dilakukan dengan menggunakan dataset yang sudah diolah pada tahap pra proses sebelumnya. Menggunakan metode CNN dengan arsitektur MobileNetV2 + Dropout. Selain itu, proses training dilakukan menggunakan teknik K-Fold Cross Validation dengan nilai K=3, dimana dataset yang digunakan 80% data train (3712) dan 20% data testing (928). Digunakan juga beberapa hyperparameter sebagai berikut:

- Input shape citra : 128x128x3
- Batch size : 32
- Epoch : 20
- Optimizer : Adam
- Learning rate : 0.0003

Berikut adalah hasil *accuracy* dan *loss* dari Fold ke-1, Fold ke-2 dan Fold ke-3.

Tabel 1. Hasil Training dengan K-Fold

No.	*Accuracy*	*Loss*	*Accuracy*	*Loss*	*Accuracy*	*Loss*

	Fold 1	*Fold* 1	*Fold* 2	*Fold* 2	*Fold* 3	*Fold* 3
1.	81.6	0.538	80.8	0.522	82.4	0.506

Berdasarkan hasil *training* pada Tabel 1. yang dilakukan maka didapatkan total rata-rata akurasi sebesar 81.6%. Dapat dikatakan bahwa akurasi ini sudah cukup baik dan dapat dilanjutkan ke tahap *testing*.

3.2 Implementasi Android

a. Splash Screen Aplikasi

Halaman ini merupakan halaman splashscreen aplikasi. *Splash Screen* muncul pada saat aplikasi mulai dijalankan dengan memperlihatkan logo aplikasi. Saat aplikasi pertama kali dibuka, maka akan ditampilkan logo aplikasi sebagai pengantar ke halaman selanjutnya. Halaman *splash screen* ini akan berdurasi 3 detik, setelah 3 detik berlalu maka akan ditampilkan halaman berikutnya, yaitu halaman utama. Berikut dapat dilihat *splash screen* aplikasi.

Gambar 3. Splash Screen Aplikasi

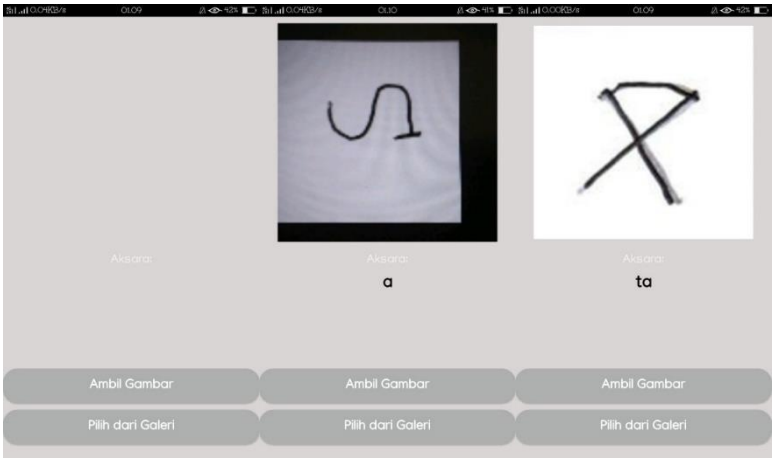
b. Halaman Utama Aplikasi

Halaman utama aplikasi, dimana terdapat beberapa menu yang dapat dijalankan. Pada halaman ini juga terdapat tulisan yang merupakan sambutan saat masuk ke dalam halaman utama aplikasi. Terdapat logo aplikasi juga pada halaman ini. Halaman utama tentunya dilengkapi dengan empat menu, yaitu menu klasifikasi, menu induk surat, menu anak surat dan menu informasi. Setiap menu dapat dijalankan dengan menekan *button* masing-masing menu, yang selanjutnya akan diarahkan ke halaman menu yang dipilih.

Gambar 4. Halaman Utama Aplikasi

c. Halaman Klasifikasi

Halaman klasifikasi adalah halaman yang digunakan untuk memproses pengambilan gambar melalui kamera dan galeri, yang selanjutnya akan di prediksi hasil dari masing-masing karakter yang diinput. Pada halaman ini terdapat 2 (dua) pilihan, pengguna dapat memilih button ambil gambar untuk mengambil gambar aksara secara real time sebagai inputan untuk diprediksi. Kemudian terdapat pilihan pilih dari galeri, dimana pengguna dapat memasukkan inputan dari kumpulan foto aksara yang sudah ada atau sudah diambil sebelumnya. Inputan dari kedua pilihan tersebut kemudian akan diprediksi oleh aplikasi, aplikasi akan memberikan output berupa deskripsi aksara.



Gambar 5. Halaman Klasifikasi Aplikasi

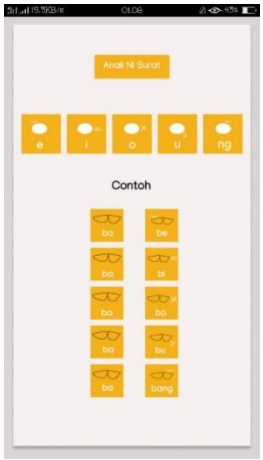
d. Halaman Induk Surat

Halaman ini merupakan halaman induk surat pada aplikasi, yang didalamnya berisi seluruh karakter induk surat Aksara Simalungun. Terdapat image dan label ke-23 induk surat Aksara Simalungun pada halaman ini. Halaman ini bertujuan untuk memperlihatkan image dan label induk surat aksara yang benar kepada pengguna.

Gambar 6. Halaman Induk Surat Aplikasi

e. Halaman Anak Surat

Halaman ini merupakan halaman anak surat pada aplikasi, yang didalamnya berisi lima pola/tanda dalam anak surat. Halaman ini bertujuan untuk memperlihatkan image dan label anak surat aksara yang benar kepada pengguna. Dalam hal ini, anak surat bertujuan sebagai tanda/pola yang dapat digunakan untuk mengubah bunyi suatu aksara.



Gambar 7. Halaman Anak Surat Aplikasi

f. Halaman Informasi

Halaman ini merupakan halaman informasi pada aplikasi yang berisi informasi singkat mengenai Aksara Simalungun. Dalam halaman ini juga ditampilkan satu image yang merupakan sejarah Aksara

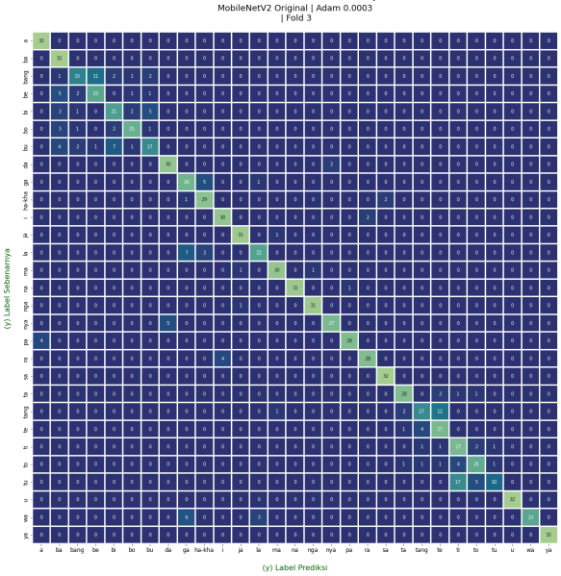
p-ISSN: 2301-5373
e-ISSN: 2654-5101

Simalungun, yaitu aksara yang ditulis pada media bambu. Pengguna dapat membaca secara singkat sejarah dari Aksara Simalungun pada halaman informasi.

Gambar 8. Halaman Informasi Aplikasi

3.3 Evaluasi Confusion Matrix dan Classification Report

Dari hasil pelatihan (*training*) dan pengujian (*testing*), dilakukan analisis terhadap model dengan menggunakan *Confusion Matrix* dan *Classification Report*. Berikut ditampilkan untuk detail hasil dari *Confusion Matrix* dan *Classification Report*.



Accuracy : 81.897%
Loss : 0.503

Classification Report :

	precision	recall	f1-score	support
a	0.89	1.00	0.94	32
ba	0.67	1.00	0.80	32
bang	0.71	0.47	0.57	32
be	0.66	0.72	0.69	32
bi	0.66	0.66	0.66	32
bo	0.83	0.78	0.81	32
bu	0.65	0.53	0.59	32
da	0.86	0.94	0.90	32
ga	0.65	0.81	0.72	32
ha-kha	0.78	0.91	0.84	32
i	0.88	0.94	0.91	32
ja	0.94	0.97	0.95	32
la	0.85	0.69	0.76	32
ma	0.94	0.94	0.94	32
na	1.00	0.97	0.98	32
nga	0.97	0.97	0.97	32
nya	0.93	0.84	0.89	32
pa	0.97	0.88	0.92	32
ra	0.93	0.88	0.90	32
sa	0.94	1.00	0.97	32
ta	0.88	0.88	0.88	32
tang	0.74	0.53	0.62	32
te	0.63	0.84	0.72	32
ti	0.55	0.84	0.67	32
to	0.75	0.75	0.75	32
tu	0.83	0.31	0.45	32
u	1.00	1.00	1.00	32
wa	1.00	0.72	0.84	32
ya	1.00	1.00	1.00	32
accuracy			0.82	928
macro avg	0.83	0.82	0.81	928
weighted avg	0.83	0.82	0.81	928

Gambar 9. (a) Confusion Matrix, (b) Classification Report

Berdasarkan informasi pada Gambar 9. *Confusion Matrix* dan *Classification Report* diatas, beberapa hal yang dapat disimpulkan yaitu:

1. Hasil *testing* dengan data *test* menghasilkan paling banyak salah prediksi yaitu pada aksara be, bi, bu, bang, te, ti, tu dan tang. Jika diperhatikan secara seksama kesalahan prediksi beberapa aksara ini disebabkan oleh kemiripan bentuk huruf dan beberapa bentuk huruf yang sama, namun hanya dibedakan oleh simbol atau tanda tertentu.
2. Model *MobileNetV2* + *dropout* (Adam 0.0003) dengan total data *test*, secara keseluruhan menghasilkan *score* akurasi dan *f1-score* yang cukup baik.
3. Pengujian dengan *confusion matrix* dan *classifaction report* menunjukkan akurasi yang didapatkan sebesar 82% dan dengan akurasi seperti ini dapat dikatakan sudah cukup baik.

3.4 Analisis Kepuasan Pengguna

Analisis kepuasaan pengguna bertujuan untuk menilai tanggapan dari pengguna sistem, yaitu siapa saja yang penasaran dan mau belajar tentang Aksara Simalungun. Untuk mencapai target yang ditetapkan dengan efektifitas, efisiensi dan mencapai kepuasan pengguna pada suatu aplikasi, maka dibutuhkan pengujian usability yang digunakan oleh pengguna. Evaluasi kepuasaan pengguna dilakukan dengan menggunakan data kuesioner yang berasal dari 15 responden yang sudah memenuhi kriteria. Berdasarkan hasil kuesioner yang dikumpulkan maka didapatkan hasil sebagai berikut:

Skor maksimum = 15 (responden) x 12 (pertanyaan) x 5 (score) = 900

Skor maksimum yang didapat yaitu sebesar 900, maka hasil kuesioner usability ini sudah bisa ditentukan persentasenya sebagai berikut:

$$\frac{702}{900}\ x\ 100\% = 78\%$$

Dapat disimpulkan bahwa hasil dari persentase menunjukkan bahwa keempat aspek usability yang didapatkan dari 15 responden yang langsung menjadi pengguna aplikasi Aksara Simalungun, memberikan penilaian aplikasi sebesar 78%.

4. Kesimpulan

Penelitian ini menghasilkan sebuah sistem pengenalan karakter Aksara Simalungun berdasarkan citra input yang menggunakan teknologi *deep learning* dengan metode *Convolutional Neural Network* (CNN) berbasis android. Berikut beberapa kesimpulan yang dapat diambil dari penelitian ini yaitu:

1. Berdasarkan metode *Convolutional Neural Network* menggunakan arsitektur *MobileNetV2* + *Dropout* dengan *hyperparameter* berupa *Adam Optimizer* dan *Learning Rate* 0.0003 di *fold* ke-3. Memberikan *score* akurasi *training* sebesar 81% dan *score* akurasi *testing* sebesar 82%. Dengan *score* akurasi seperti berikut, arsitektur *MobileNetV2* teruji menghasilkan *score* akurasi *testing* yang cukup tinggi.
2. Paremeter pelatihan yang digunakan pada penelitian ini seperti *Adam Optimizer*, *Learning Rate* (0.0003) dan *epoch* (20) memberikan pengaruh yang cukup signifikan terhadap *score* akurasi yang dihasilkan oleh metode *Convolutional Neural Network*. Dengan adanya parameter pelatihan ini dapat mengoptimalisasi kinerja metode, mempercepat waktu komputasi dan menghasilkan score akurasi yang tinggi.
3. Dari hasil pengujian dengan analisis kepuasan pengguna menggunakan skala *likert* pada pengujian *usability* yang berguna untuk mengetahui tingkat kelayakan aplikasi. Peneliti melibatkan 15 orang responden dan pengujian ini dilakukan dengan cara menyebar kuesioner yang berisi pertanyaan seputar aplikasi. Hasil dari kuesioner tersebut menentukan kelayakan aplikasi dari 4 aspek *usability* dengan hasil persentase sebesar 78%.

Daftar Pustaka

[1] Kozok, U. Warisan Leluhur: sastra lama dan aksara Batak (Vol. 17). Kepustakaan Populer Gramedia. 1999.
[2] Nurhikmat, Triakno. "Implementasi deep learning untuk image classification menggunakan algoritma Convolutional Neural Network (CNN) pada citra wayang golek." (2018).
[3] Susilo, M. M., Wonohadidjojo. D. M., & Sugianto, N. (2017). Pengenalan Pola Karakter Bahasa Jepang Hiragana Menggunakan 2D Convolutional Neural Network. *J. Inform. dan Sist. Inf. Univ. Ciputra, 3(02), 28-36.*
[4] Khasanah. Dqlab.id. 07 Juli 2021. https://dqlab.id/perbedaan-data-primer-dan-data-sekunder. Diakses pada tanggal 12 Juli 2022.
[5] Afif. Medium.id. 28 April 2020. https://medium.com/@hafizhan.aliady/membuat-klasifikasi-gambar-images-menggunakan-keras-tensorflow-tf-keras-dan-python-53f7ae953cea. Diakses pada 12 Juli 2022.
[6] Rahadi, Dedi Rianto. "Pengukuran usability sistem menggunakan use questionnaire pada aplikasi android." *JSI: Jurnal Sistem Informasi (E-Journal)* 6.1 (2014).

Jurnal Elektronik Ilmu Komputer Udayana
Volume 11, No 3. February 2023

p-ISSN: 2301-5373
e-ISSN: 2654-5101

Analisis Sentimen Berbasis Aspek Ulasan Pelanggan Hotel di Bali Menggunakan Metode *Decision Tree*

Ni Putu Ambalika Dewi[a1], Ngurah Agus Sanjaya ER[a2], AAIN Eka Karyawati[a3], Ida Bagus Made Mahendra[a4], Ida Bagus Gede Dwidasmara[a5], I Gede Arta Wibawa[a6]

[a]Program Studi Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam,
Universitas Udayana
Bali, Indonesia
[1]ambalikaambalikadewi@gmail.com
[2]agus_sanjaya@unud.ac.id
[3]eka.karyawati@unud.ac.id
[4]ibm.mahendra@unud.ac.id
[5]dwidasmara@unud.ac.id
[6]gede.arta@unud.ac.id

Abstract

The main means of tourism is the accommodation industry. Therefore, accommodation cannot be separated from the tourism industry because they both need each other. One of the accommodations that is most closely related to tourism is hospitality accommodation. With the increasing number of hotels in Bali, the hotel certainly needs the right marketing strategy. So, it is necessary to process customer reviews automatically to determine sentiment analysis based on customer reviews based on certain aspects. In this study, the author builds a model for aspect-based sentiment analysis using the Decision Tree method. The data used in this study is hotel customer review data in Indonesian language. Evaluation is done by measuring the performance of the Decision Tree model. The Decision Tree model for aspects produces performance, accuracy, precision, recall, and F1-Score, respectively 82,5%, 80%, 90,9%, and 85,1%, the Decision Tree model for service aspect sentiment produces accuracy, precision, recall performance , and F1-Score, respectively, which are 75%, 72,7%, 80%, and 76,2%, while the Decision Tree model for the sentiment of cleanliness aspect produces performance of accuracy, precision, recall, and F1-Score, respectively, which is 81,8%, 87,5%, 77,8%, and 82,4%.

Keywords: *Decision Tree*, *Confusion Matrix*, TF-IDF, *Aspect-Based Sentiment Analysis*, *Review*

1. Pendahuluan

Sektor pariwisata di Bali tidak dapat diragukan kemajuan dan popularitasnya. Pulau Bali mampu memikat jutaan wisatawan mancanegara ataupun domestik setiap tahunnya. Menurut data yang didapatkan dari https://kompas.com, pada tahun 2015, Bali menempati posisi kedua sebagai pulau dengan destinasi tujuan wisata terbaik setelah kepulauan Galapagos Ekuador. Sarana pokok kepariwisataan adalah industri akomodasi. Oleh karena itu, akomodasi sulit dipisahkan dengan industri pariwisata karena keduanya saling berkaitan. Salah satu akomodasi yang sangat erat kaitannya dengan pariwisata adalah akomodasi perhotelan. Pembangunan hotel-hotel di Bali semakin meningkat seiring dengan meningkatnya jumlah wisatawan yang berwisata ke Bali setiap tahunnya. Reservasi hotel juga semakin berkembang dengan adanya media-media digital berbasis internet yang memudahkan wisatawan untuk memesan kamar hotel tanpa perlu datang langsung ke hotel yang diinginkan. Beragamnya hotel yang ada di Bali membuat wisatawan cenderung melihat ulasan yang telah ditinggalkan oleh wisatawan sebelumnya untuk menentukan pilihan hotel yang mereka inginkan. Selain itu, harga juga menjadi faktor lain wisatawan dalam menentukan akomodasi hotel. Semakin banyaknya hotel yang ada di Bali, pihak hotel tentu membutuhkan strategi pemasaran yang tepat. Sehingga, diperlukan pengolahan ulasan pelanggan secara otomatis untuk menentukan analisis sentimen berdasarkan ulasan pelanggan berdasarkan aspek-aspek tertentu.

Sumber data yang umum digunakan untuk analisis sentimen adalah jaringan sosial yang menyimpan dan menyimpan banyak informasi. Media sosial amatlah penting sebagai sumber data yang

digunakan untuk analisis sentimen. Beberapa penelitian selanjutnya telah berkembang, dengan fokus pada pengembangan model terbaik untuk memperluas aplikasi analisis sentimen [8]. Pada penelitian sebelumnya, beberapa peneliti telah melakukan penelitian untuk memprediksi harga saham menggunakan metode *Decision Tree* dengan pembobotan TF-IDF dan TF-RF. Hasilnya menunjukkan bahwa penelitian yang dilakukan dengan TF-RF mendapatkan nilai F-Measure yang lebih besar dibandingkan dengan TF-IDF [1]. Pada penelitian lainnya, penulis melakukan penelitian untuk melihat karakteristik dari mahasiswa Universitas Cokroaminoto Palopo. Dalam penelitiannya, penulis menggunakan metode naive bayes dan metode pohon keputusan. Penelitian ini menggunakan variabel bebas. Dari penelitian yang dilakukan penulis, hasil analisis yang ditemukan oleh penulis merupakan hasil ketepatan dalam mengklasifikasi karakteristik pendaftar Universitas Cokroaminoto Palopo menggunakan metode *Naïve Bayes* sebesar 98,18% dan menggunakan metode *Decision Tree* sebesar 97,82% [2]. Dalam penelitian lain tentang klasifikasi berbasis *machine learning* menggunakan metode *decision tree*, peneliti mengeksplorasi lebih lanjut tentang metode *decision tree*. Dari studi yang dilakukan, terlihat bahwa penggunaan dataset yang berbeda mempengaruhi hasil klasifikasi menggunakan metode *decision tree* [3].

Untuk itu pada penelitian kali ini, penulis ingin mengetahui bahwa metode yang digunakan menghasilkan nilai akurasi, *precision, recall* dan *f-1 score* yang baik. Berdasarkan pemasalahan dan penelitian-penelitian sebelumnya yang menjadi dasar untuk penelitian ini, maka penulis bermaksud untuk melakukan penelitian terhadap performa dari metode *Decision Tree* dalam analisis sentiment berbasis aspek pada ulasan pelanggaan hotel di Bali dan diharapkan penelitian ini menunjukan hasil klasifikasi yang baik.

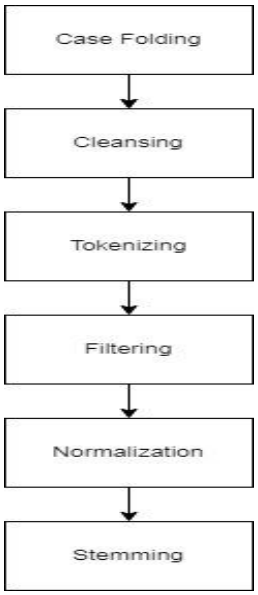
2. Metode Penelitian

2.1 Dataset

Data yang digunakan pada penelitian ini merupakan data dalam bentuk ulasan pelanggan hotel di Bali berbahasa Indonesia. Data berjumlah 800 dengan format *file* *.xlsx yang meliputi 200 data ulasan untuk aspek pelayanan bernilai positif, 200 data ulasan untuk aspek pelayanan bernilai negatif, 200 data ulasan untuk aspek kebersihan bernilai positif dan 200 data ulasan untuk aspek kebersihan bernilai negatif.. Seluruh data sudah dilabeli ahli. Data berita kemudian dibagi menjadi dua yaitu data latih dan data uji, dengan sebanyak 80% data latih dan 20% data uji. Data latih tersebut kemudian dibagi lagi menjadi data latih dan data validasi untuk digunakan dalam proses pelatihan model dengan menggunakan *K-fold cross validation* dengan K=10.

2.2 *Preprocessing*

Preprocessing adalah proses pengolahan data yang digunakan untuk membuat format yang lebih baik. Tahapan-tahapan *preprocessing* yang dilakukan dalam penelitian ini adalah sebagai pada gambar berikut [4].



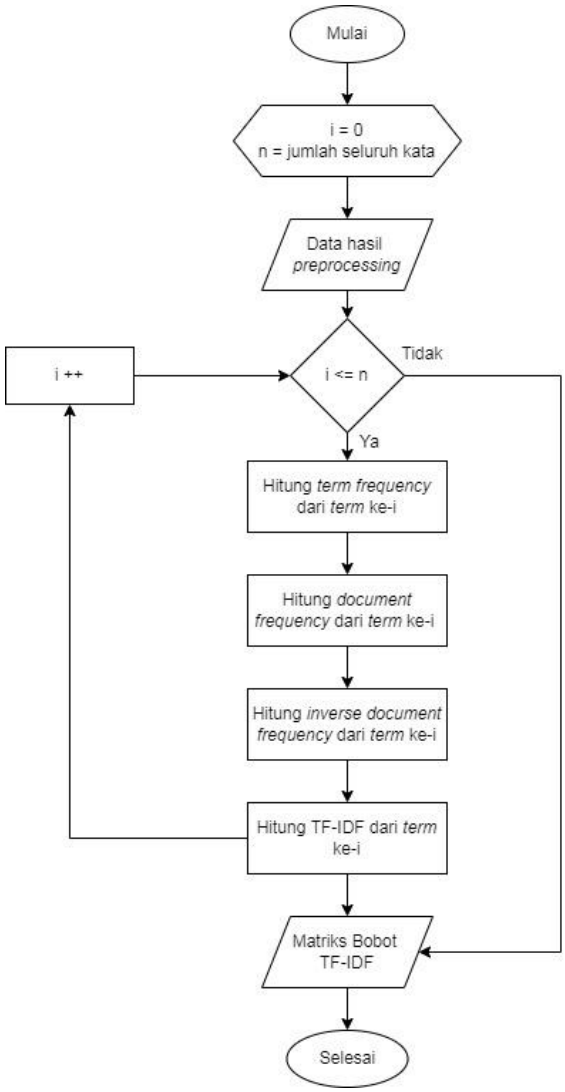
Gambar 1. Alur Preprocessing

p-ISSN: 2301-5373
e-ISSN: 2654-5101

Proses pertama pada *preprocessing* adalah *case folding* yang digunakan untuk mengubah semua huruf menjadi huruf kecil [5]. Setelah itu dilanjutkan dengan *cleansing,* pada proses ini akan dihapus karakter yang tidak memiliki kaitan terhadap analisis sentimen berbasis aspek, termasuk penghapusan tanda baca. Selanjutnya adalah *tokenizing*, yang digunakan untuk memisahkan kata dalam suatu paragraf atau kalimat menjadi token-token tertentu [5]. Selanjutnya *filtering* atau *stopwords removal*, yaitu*u*penghapusan kata-kata yang tidak berpengaruh terhadap analisis sentimen berbasis aspek pada ulasan pelanggan [5]. Kemudian proses *normalization* yang berfungsi untuk mengubah dan mengembalikan bentuk penulisan tidak baku ke bentuk penulisan yang sesuai dengan KBBI. Proses terakhir merupakan *stemming,* yaitu proses ekstraksi kata-kata berimbuhan untuk mendapatkan kata dasar [5].

2.3 *Term Frequency Inverse Document Frequency* (TF-IDF)

Setelah melewati tahapan *preprocessing*, selanjutnya dilanjutkan dengan melakukan pembobotan dengan metode *Term Frequency Inverse Document Frequency*. Pembobotan TF-IDF dimulai dengan memasukkan data hasil *preprocessing* yang dikerjakan berulang hingga jumlah total seluruh kata. Proses selanjutnya dilakukan perhitungan *term frequency* yaitu menghitung frekuensi munculnya kata pada suatu dokumen [6]. Kemudian dilanjutkan dengan perhitungan *document frequency* untuk menghitung jumlah dokumen yang mengandung suatu kata tertentu. Selanjutnya diilanjutkan dengan perhitungan bobot *inverse document frequency* untuk menghitung distribusi kata pada koleksi dokumen dengan menggunakan persamaan (2) [6]. Terakhir, proses menghitung bobot TF-IDF dilakukan dengan mengalikan nilai *term frequency* dan nilai *inverse document frequency* dengan menggunakan persamaan (3) [6]. Gambar berikut merupakan tahapan TF-IDF.



Gambar 2. TF-IDF

a. Menghitung *term frequency* dengan persamaan
$tf = 0{,}5 + 0{,}5 \times \frac{tf}{\max(tf)}$ (1)

b. Menghitung *inverse document frequency* dengan persamaan
$idf_t = \log\left(\frac{D}{dft}\right)$ (2)

c. Menghitung bobot TF-IDF dengan persamaan
$W_{n,x} = tf_{n,x} \times idf_{n,x}$ (3)
Keterangan:
tf : banyaknya kata yang sering muncul pada dokumen.
max(tf) : jumlah kata atau *term* pada data yang sama yang sering bermunculan.
Nilai D : jumlah seluruh data yang digunakan
df_x : banyaknya data yang mengandung kata x.
idf : banyaknya data *inverse*.
n : data ke-n.
x : kata ke-x.
$W_{d,t}$: bobot TF-IDF dari data ke-n dan kata ke-x.

2.4 *Decision Tree*

Decision tree atau bisa disebut analisis pohon keputusan adalah salah satu metode klasifikasi yang memanfaatkan teori graf dalam membagi kelompok data menjadi himpunan data. *Decision tree* secara visual mirip dengan sebuah pohon yang bercabang dengan ranting-rantingnya [7]. Data yang sudah melalui tahapan *preprocessing* dan pembobotan TF-IDF kemudian akan melalui tahap klasifikasi menggunakan metode *Decision Tree* untuk melakukan analisis sentiment berbasis aspek. Klasifikasi dilakukan dengan mengelompokkan data berupa ulasangan pelanggan yang telah di olah sebelumnya ke kelompok-kelompok aspek yang telah ditentukan. Tahap pertama yang dilakukan adalah menghitung nilai *entropy* yang akan digunakan untuk menghitung nilai *gain*. Hitung nilai *gain*, dan pilih atribut sebagai *root* berdasarkan nilai *gain* yang tertinggi. Buat cabang untuk masing-masing nilai *gain* kemudian temukan atribut terbaik dan split terbaik pada atribut. Selanjutnya lakukan pembagian data berdasarkan split, pohon akan bertumbuh dan kembali pada proses pembuatan cabang. Apabila sudah tidak terdapat cabang yang bisa dibuat maka proses telah selesai. Gambar 3 merupakan alur proses *Decision Tree*.

Tahapan-tahapan dari metode decision tree adalah sebagai berikut [1]:

a. Hitung nilai *entropy* menggunakan rumus berikut.
$\text{Entropy}(\text{S}) = \sum_{i=l}^{n} - p_i \ log_2 p_i$ (4)

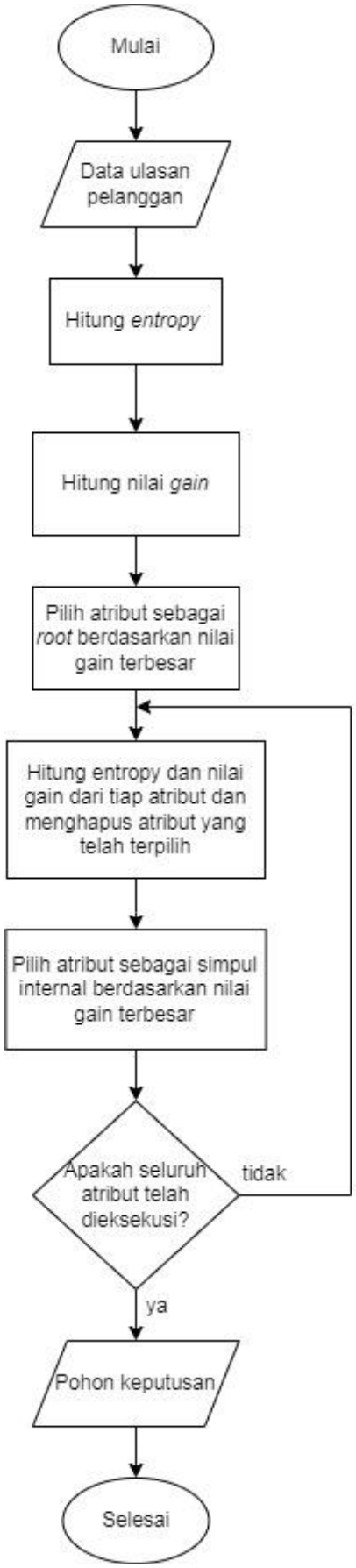
Keterangan:
S : Sekelompok data latih
n : Proporsi partisi pada S
p_i : Jumlah sampel pada kelas i

b. Hitung nilai *gain* menggunakan rumus berikut.
$\text{Gain}(\text{S}, \text{A}) = Entropy(S) - \sum_{i=l}^{n} \frac{S_i}{S} \ x \ Entropy(S_i)$ (5)

Keterangan:
S : Sekelompok data latih
A : Atribut
n : Pembagian total untuk seluruh atribut A
S_i : Pembagian total ke-i

c. Tentukan atribut sebagai *root*
d. Proses partisi akan berhenti jika kondisi berikut terpenuhi, yaitu:
   - Seluruh data yang ada pada atribut *A* mendapatkan kategori yang sama
   - Atribut yang ada dalam data sudah habis terbagi
   - Tidak ada data yang tersisa pada cabang yang kosong

p-ISSN: 2301-5373
e-ISSN: 2654-5101



Gambar 3. Decision Tree

2.5 Evaluasi

Evaluasi yang digunakan untuk menyelesaikan penelitian ini adalah *confusion matrix* yang dimaksudkan untuk menghitung akurasi, *precision*, *recall* dan *F1-score*. Evaluasi ditujukan untuk mengukur performa model terbaik yang sudah dipilih dari proses validasi. Ukuran evaluasi yang digunakan adalah akurasi, *precision, recall,* dan *F1-Score,* dan akurasi. Akurasi digunakan untuk mengevaluasi banyakya label prediksi yang sesuai dengan label aktual, dihitung menggunakan persamaan (6). *Precision* mengukur presentase dokumen bernilai positif benar di antara seluruh dokumen yang diidentifikasi positif, dihitung menggunakan persamaan (7). *Recall* mengukur presentase dokumen bernilai positif benar yang dapat diidentifikasikan di antara seluruh dokumen yang relevan, dihitung menggunakan persamaan (8). *F1-Score* adalah kombinasi hasil dari *precision* dan *recall*, dihitung menggunakan persamaan (9).

Tabel 1. *Confusion Matrix*

Data	Nilai Sesungguhnya	
	Relevan	Tidak Relevan
Retrived	TP	FP
Not Retrived	FN	TN

Keterangan:

TP (*True Positive*) : Proyeksi data yang benar sepenuhnya benar

FN (*False Negative*) : Terdapat data yang salah pada proyeksi data yang benar

FP (*False Positive*) : Terdapat data yang benar pada proyeksi data yang salah

TN (*True Negative*) : Proyeksi data yang salah sepenuhnya salah

Rumus berikut digunakan untuk menghitung akurasi, *precision, recall* dan *F-1 score* [1].

$$\text{Akurasi} = \frac{TP+TN}{(TP+FP+TN+FN)} \quad (6)$$

$$\textit{Precision} = \frac{TP}{(TP+FP)} \quad (7)$$

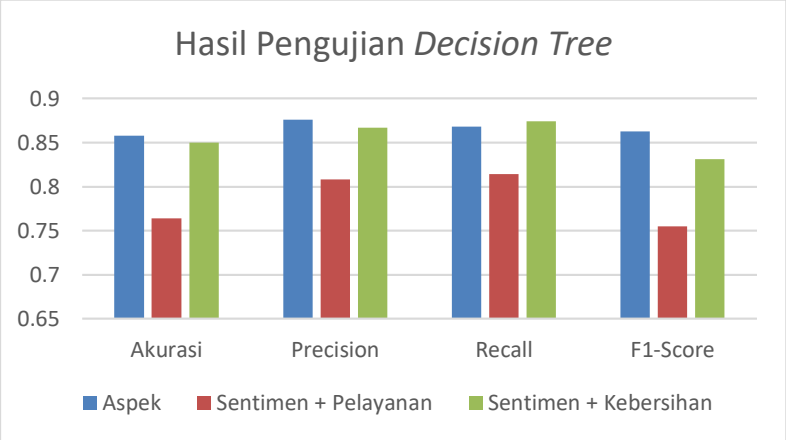
$$\textit{R}\text{ecall} = \frac{TP}{(TP+FN)} \quad (8)$$

$$\textit{F-1 Score} = \frac{2\ x\ Precison\ x\ Recall}{(Precision+Recall)} \quad (9)$$

3. Hasil dan Pembahasan

Pada pengujian penelitian ini, metode *Decision Tree* diuji untuk dapat menghasilkan nilai akurasi, *precision, recall* dan *F1-Score*. Proses pelatihan dan validasi menggunakan metode *K-Fold Cross Validation* untuk dapat menghasilkan performa akurasi terbaik. Pengujian dilakukan untuk membuat tiga model yaitu model aspek, model sentimen aspek pelayanan dan model sentimen aspek kebersihan. Setelah melakukan pengujian dengan menggunakan 10 *fold* untuk membuat model aspek, didapatkan nilai akurasi rata-rata sebesar 85,8%, nilai *precision* rata-rata sebesar 87,6%, nilai *recall* rata-rata sebersar 86,8%, dan nilai F1-Score rata-rata sebesar 86,2%. Setelah melakukan pengujian dengan menggunakan 10 *fold* untuk membuat model sentimen aspek pelayanan, didapatkan nilai akurasi rata-rata sebesar 76,4%, nilai *precision* rata-rata sebesar 80,8%, nilai *recall* rata-rata sebersar 81,4%, dan nilai F1-Score rata-rata sebesar 75,5%. Setelah melakukan pengujian dengan menggunakan 10 *fold* untuk membuat model sentimen aspek kebersihan, didapatkan nilai akurasi rata-rata sebesar 85%, nilai *precision* rata-rata sebesar 86,7%, nilai *recall* rata-rata sebersar 87,4%, dan nilai F1-Score rata-rata sebesar 83,1%. Setelah didapatkan model terbaik, selanjutnya model akan digunakan pada proses pengujian data uji dengan menggunakan data baru yang belum digunakan pada saat proses pelatihan dan validasi. Gambar berikut merupakan rata-rata hasil pengujian *Decision Tree* pada ketiga model.

p-ISSN: 2301-5373
e-ISSN: 2654-5101



Gambar 4. Hasil Pengujian Decision Tree

Setelah melakukan pengujian tehadap ketiga model dengan menggunakan data baru, hasil perbandingan performa model pada saat *training* dan *testing* dapat dilihat pada tabel .

Tabel 2. Hasil Pengujian Model

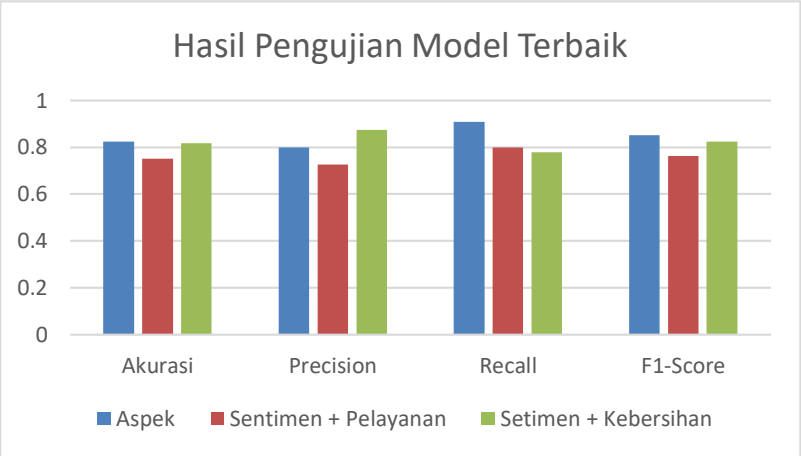
Model	Ukuran Evaluasi	Pengujian	
		Training Validasi	Testing Data Baru
Aspek	Akurasi	0.858	0.825
	Precision	0.876	0.8
	Recall	0.868	0.909
	F1-Score	0.862	0.851
Sentimen + Pelayanan	Akurasi	0.764	0.75
	Precision	0.808	0.727
	Recall	0.814	0.8
	F1-Score	0.755	0.762
Sentimen + Kebersihan	Akurasi	0.85	0.818
	Precision	0.867	0.875
	Recall	0.874	0.778
	F1-Score	0.831	0.824

Pada Gambar dapat dilihat bahwa implementasi metode *Decision Tree* dalam analisis sentimen berbasis aspek menghasilkan nilai akurasi sebesar 82,5% pada model aspek, 75% pada model sentimen aspek pelayanan dan 81,8% pada model sentimen aspek kebersihan. Akurasi yang tinggi menunjukkan bahwa model dapat dengan baik mengklasifikasikan ulasan pelanggan ke dalam aspek pelayanan dan kebersihan serta mengklasifikasikan ulasan pelanggan ke sentimen positif dan negatif. Selanjutnya, perlu adanya ukuran evaluasi lain yang memperhitungkan kesalahan prediksi, yaitu *precision, recall,* dan *F1-Score*.

Nilai *precision* yang dihasilkan sebesar 80% pada model aspek, 72,7% pada model sentimen aspek pelayanan dan 87,5% pada model sentimen aspek kebersihan. Nilai *precision* yang tinggi menunjukkan bahwa model dapat dengan baik mengklasifikasikan aspek pelayanan ke dalam aspek pelayanan dan sentimen positif ke dalam sentimen positif, serta sedikit kesalahan prediksi aspek pelayanan ke dalam aspek kebersihan dan sentimen positif ke dalam sentimen negatif.

Nilai *recall* yang dihasilkan sebesar 90,9% pada model aspek, 80% pada model sentimen aspek pelayanan dan 77,8% pada model sentimen aspek kebersihan. Nilai *recall* yang tinggi menunjukkan bahwa model dapat dengan baik mengklasifikasikan mengklasifikasikan aspek pelayanan ke dalam aspek pelayanan dan sentimen positif ke dalam sentimen positif, serta sedikit kesalahan prediksi aspek kebersihan ke dalam aspek pelayanan dan sentimen negatif ke dalam sentimen positif.

Nilai *F1-Score* yang dihasilkan sebesar 85,1% pada model aspek, 76,2% pada model sentimen aspek pelayanan dan 82,4% pada model sentimen aspek kebersihan. Nilai *F1-Score* digunakan untuk melihat keseimbangan antara *precision* dan *recall*, sehingga nilai ini bisa digunakan juga sebagai ukuran evaluasi sebuah model selain menggunakan akurasi.



Gambar 5. Hasil Pengujian Model Terbaik

4. Kesimpulan

Berdasarkan penelitian yang telah dilakukan, ditarik kesimpulan bahwa implementasi metode *Decision Tree* dalam analisis sentimen berbasis aspek pada ulasan pelanggan dengan menggunakan *K-Fold Cross Validation* menghasilkan tiga model terbaik. Pada pengujian data baru, model *Decision Tree* untuk aspek menghasilkan performa akurasi, *precision*, *recall,* dan *F1-Score* secara berturut-turut yaitu 82,5%, 80%, 90,9%, dan 85,1%. Pada pengujian model *Decision Tree* untuk sentimen aspek pelayanan menghasilkan performa akurasi, *precision*, *recall,* dan *F1-Score* secara berturut-turut yaitu 75%, 72,7%, 80%, dan 76,2%. Pada pengujian model *Decision Tree* untuk sentimen aspek kebersihan menghasilkan performa akurasi, *precision*, *recall,* dan *F1-Score* secara berturut-turut yaitu 81,8%, 87,5%, 77,8%, dan 82,4%.

Daftar Pustaka

[1] M. G. Tambunan and E. B. Setiawan, "Prediksi Kepribadian DISC pada Twitter Menggunakan Metode Decision Tree C4.5 dengan Pembobotan TF-IDF dan TF-RF," *e-Proceeding of Engineering*. Vol. 7, No. 1, page 2725-2738, 2020

[2] Y. Hastuti, "Klasifikasi Karakteristik Mahasiswa Universitas Cokroaminoto Palopo Menggunakan Metode *Naïve Bayes* dan *Decision Tree,*" *Jurnal Dinamika*. Vol. 7, No.2, page 34-41, 2016

[3] B. T. Jijo and A.M. Abdulazeez, "Classification Based on Decision Tree Algorithm for Machine Learning," *Journal of Applied Science and Technology Trends*, vol. 2, no. 1, page 20–28, 2021

[4] E. Supriyanti and M. Iqbal, "Pengukuran *Similarity* Tema pada Juz 30 Al Qur'an Menggunakan Teks Klasifikasi," *Jurnal SIMETRIS*, Vol. 9, No. 1, page. 361–370, 2018

[5] K. D. Y. Wijaya and A.A.I.N.E. Karyawati, "The Effects of Different Kernels in SVM Sentiment Analysis on Mass Social Distancing," *JELIKU*, vol. 9, no. 2, page 161-168, Nov. 2020
[6] I. W. Santiyasa, G. P. A. Brahmantha, I. W. Supriana, I. G. G. A. Kadyanan, I. K. G. Suhartana, and I. B. M. Mahendra, "Identification of Hoax Based on Text Mining Using K-Nearest Neighbor Method," *JELIKU (Jurnal Elektron. Ilmu Komput. Udayana)*, vol. 10, no. 2, page 217–226, 2021
[7] C. Z. Janikow, "Fuzzy decision trees: issues and methods, " *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*. Vol. 28, No. 1, page 1–14, 1998
[8] N. C. Dang, M. N. Moreno-García, and F. De la Prieta, "Sentiment Analysis Based on Deep Learning: A Comparative Study," *Electronics*, vol. 9, no. 3, page 483, Mar. 2020.

This page is intentionally left blank.

Jurnal Elektronik Ilmu Komputer Udayana
Volume 11, No 3. February 2023

p-ISSN: 2301-5373
e-ISSN: 2654-5101

Hybrid Implementation Fuzzy K-Nearest Neighbor (FK-NN) And Particle Swarm Optimization (PSO) Methods for Classification of Liver Disease

I Gede Bagus Semara Wijaya [a1], Luh Gede Astuti[a2], I Putu Gede Hendra Suputra[a3], I Dewa Made Bayu Atmaja Darmawan [a4], I Wayan Santiyasa[a5], I Gede Santi Astawa[a6]

[a]Departement of Informatics, Faculty of Mathematics and Natural Science, Udayana University
Jl. Raya Kampus Universitas Udayana Bukit Jimbaran, Badung, Bali
[1] gedebagus097@gmail.com
[2] lg.astuti@unud.ac.id
[3]hendra.suputra@unud.ac.id
[4]dewabayu@unud.ac.id
[5]santiyasa67@gmail.com
[6]santi.astawa@cs.unud.ac.id

Abstract

Liver disease is a disease that attacks the liver or liver where this disease is caused by viral infections, toxic materials and bacteria, causing inflammation of the liver and causing the liver to not function properly. Therefore, the author will conduct research to make a Liver Disease Classification program. This research will use Fuzzy K-Nearest Neighbor (FK-NN) and Particle Swarm Optimization (PSO) methods. Fuzzy K-Nearest Neighbor is a classification method that combines fuzzy and k-nearest neighbor algorithms. Particle Swarm Optimization is a simple optimization technique to apply and modify several parameters. This research will implement the application design to the lines of web-based program code using the Python language and the Django framework. This study resulted in the value of the accuracy range obtained by the PSO-FKNN hybrid method is 66 to 74 (in percent) compared to the range of accuracy values of FKNN without the hybrid method is 64.90% to 68.29% (in percent), the difference in the accuracy values obtained by PSO-FKNN FKNN is affected by changes in the position of the training and testing data in each test.

Keywords: *Liver Disease, Fuzzy K-Nearest Neighbor (FK-NN), Particle Swarm Optimization (PSO)*

1. Introduction

Liver disease is a disease that attacks the liver or liver where this disease is caused by viral infections, toxic materials, and bacteria, causing inflammation of the liver and causing the liver to not function properly. There are several types of liver disease including hepatitis, liver cirrhosis, and liver cancer. People with the liver disease generally experience fatigue, loss of appetite, sudden weight loss, and abdominal pain. If left unchecked, these symptoms can cause several disorders in the liver to cause diseases such as hepatitis, liver cirrhosis, and liver cancer which can end in death. Many researchers research to classify whether a person suffers from liver disease or not. Research on liver disease was previously conducted by Rahmawati in 2015 with the title "Comparative Analysis of the Naïve Bayes Algorithm and C4.5 for Predicting Liver Disease" where this study used the Naïve Bayes algorithm and C4.5 to predict liver disease.

This study tested the Naïve Bayes and C4.5 algorithms using the cross-validation and split presentation methods and then measured using a confusion matrix. The results obtained an

accuracy value of 69.828% on the C4.5 algorithm and 63.362% on the Naïve Bayes algorithm [1]. One method that can also be used for classification is Fuzzy K-Nearest Neighbor (FK-NN).

Fuzzy K-Nearest Neighbor is a classification method that combines fuzzy and k-nearest neighbor algorithms. Research on Fuzzy K-Nearest Neighbor has been conducted previously by Shanti, Hidayat, and Wihandika in 2018 entitled "Implementation of the F-KNN (Fuzzy K-Nearest Neighbor) Method for Dog Disease Diagnosis". This study resulted in the highest accuracy obtained from the test results when K = 5 with a value of 98.67% [2]. The advantage of Fuzzy K-Nearest Neighbor compared to K-Nearest Neighbor is that the relationship between data and class is not rigid because each data has a certain membership value from each class. But Fuzzy K-Nearest Neighbor also has a weakness where the parameters k (neighborhood value) and m (fuzzy power of weight) are fundamental factors that affect the classification results. In determining these parameters it is often difficult to control because it is not certain what the exact values of k and m are so other methods are needed to improve accuracy [3]. The way that can be done is by using a hybrid method that uses 2 methods on a problem. One method that can be used to implement the hybrid method is Particle Swarm Optimization (PSO).

Particle Swarm Optimization is a simple optimization technique to apply and modify several parameters. PSO is also widely used for weight optimization and feature selection. Research on the hybrid method with Particle Swarm Optimization (PSO) has been carried out previously by Muhammad Ridwan Lubis in 2017 entitled Hybrid Particle Swarm Optimization - Neural Network Backpropagation Method for Predicting the results of Football Matches that produce better results on each test with a percentage 0.03% [4].

Next is the research conducted by Siringoringo and Perangin-angin in 2017 as well as their research, namely Hybridization of the Fuzzy K-Nearest Neighbor Method with the Modified Particle Swarm Optimization Method on Classification of Soybean Plant Diseases. This study proves that the application of MPSO to FK-NN can improve performance with the resulting accuracy of 84% for MPSO and FK-NN and 81% for FK-NN only [3].

From the results of the implementation that has been carried out in previous research regarding the hybrid method, this study will conduct research on the Implementation of the Hybrid Fuzzy K-Nearest Neighbor (FK-NN) and Particle Swarm Optimization (PSO) Method for Classification of Liver Disease. PSO here is used to find the optimal k and m parameter values which will then be used in the Fuzzy K-Nearest Neighbor method. This research will conduct a test to get the best combination of parameters produced by PSO and FK-NN. The next test is to compare the accuracy results between the hybrid Fuzzy K-Nearest Neighbor (FK-NN) and Particle Swarm Optimization (PSO) methods without using the hybrid method.

2. Research Methods

2.1. Classification

Classification is a technique for designing functions based on observations of the data and attributes of the data so that it can be done mapping of data that does not have a class into data that has been classified according to the rules given. There are many algorithms used to classify data, including fuzzy, artificial neural network, support vector machine and K-nearest neighbor. Classification using these algorithms is generally carried out based on 3 stages:

1. Model Design
   The process of building solutions to solve problems based on data that has been classified (training data).
2. Model Implementation
   The process of determining the class for test data is based on the function model and data parameters that have been determined at the design stage.
3. Model Evaluation
   The process aims to evaluate the results of the implementation of the function model in classifying test data based on predetermined parameters [6].

p-ISSN: 2301-5373
e-ISSN: 2654-5101

2.2. Fuzzy K-Nearest Neighbor (F-KNN)

Fuzzy K-NN is a classification method that combines fuzzy and K-NN techniques. This method differs from other methods in that it explicitly predicts which class the test data follows based on the closest K comparison. The basis of the FK-NN algorithm is the assignment of membership values, which is a function of the K-NN's distance vector and the membership of their neighbors to possible classes.

This approach plays an important role in disambiguating classification. Furthermore, the instances in each class have a certain membership value, so it gives the instance more strength or confidence to be a class [7]. This procedure is performed using Equation (2.2) before computing the membership value in the fuzzy K-NN.

$$u_{ij} = \begin{cases} 0.51 + \left(\frac{n_j}{n}\right) * 0.49, if\ j = 1 \\ \left(\frac{n_j}{n}\right) * 0.49, if\ j \neq 1 \end{cases} \quad \text{.....(1)}$$

Descriptions:
nj = The number of members of class j in a training data n
n = Amount of training data used
j = Data Class
Then calculate the membership value of each class with the formula (2).

$$U_i(x) = \frac{\sum_{j=1}^{k} U_{ij}(\|x-xj\|^{-2/(m-1)})}{\sum_{j=1}^{k}(\|x-xj\|^{-2/(m-1)})} \text{.....(2)}$$

Descriptions:
uij = fuzzy membership value in the test sample (x, xj)
k = nearest neighbor value
j = test data membership data variables
m = power to the power of magnitude m > 1

2.3. Particle Swarm Optimization (PSO)

Particle Swarm Optimization (PSO for short) is based on the behavior of a swarm of insects such as ants, termites, bees or birds. The PSO algorithm mimics the social behavior of these creatures. Social behavior includes the behavior of individuals and the influence of others in the group. The word "particle" refers to an individual, such as a bird in a flock. Each individual or particle uses its intelligence (intelligence) to behave in an interrelated manner and is also influenced by the behavior of the collective group. So if a particle or a bird finds a short path to a food source, other members of the group will be able to follow that path quickly, even if they are far away in the group.

2.4. Research Data

This study uses secondary data obtained from the UCI Machine Learning Repository. This data consists of 583 data records, with a percentage of 71.3% (416 records) for the class with liver disease and 28.7% (167 records) for the class without liver disease. The following is an attribute table for liver disease data:

Table 1 Liver Disease Dataset Features

Attribute	Domain
Age	4 – 90
Gender	0, 1
Total bilirubin	0 – 75
Direct Bilirubin	0– 13.07
Total Proteins	63 - 2110
Albumin	10 – 2000
Albumin and Globulin Ratio	11 – 4929
Alamine aminotransferase	2.07 – 9.06

Aspartate aminotransferase	0 – 5.05
Alkaline Phospotase	0 – 2.08
Is_Patients	1 *for liver disease*, 2 *for not liver disease*

2.5. Method Design

2.5.1. Data Normalization

The data used in this study has several types of data values that differ in each of its features, such as the values of tens, hundreds, and decimals, for this reason, normalization will be carried out at the preprocessing stage of the data before use. The normalization technique used in this study is min-max normalization[8]. The following is an example of unnormalized data:

Table 2 Example of unnormalized data (part 1)

age	*gender*	*tot_bilirubin*	*direct_bilirubin*	*Tot_proteins*	*albumin*	*ag_ratio*
65	0	0,7	0,1	187	16	18
62	1	10,9	5,5	699	64	100
62	1	7,3	4,1	490	60	68
58	1	1	0,4	182	14	20
72	1	3,9	2	195	27	59

Table 3 Example of unnormalized data (part 2)

sgpt	*sgot*	*alkphos*	*is_patients*
6,8	3,3	0,9	1
7,5	3,2	0,74	1
7	3,3	0,89	1
6,8	3,4	1	1
7,3	2,4	0,4	1

Then after the normalization process, the data will look like the following:

Table 2 Example of normalized data (part 1)

age	*Gender*	*tot_bilirubin*	*direct_bilirubin*	*Tot_proteins*	*albumin*	*ag_ratio*
0,709302	0	0,004021	0,007692	0,060576	0,003015	0,001423
0,674419	1	0,140751	0,423077	0,310699	0,027136	0,018097
0,674419	1	0,092493	0,315385	0,208598	0,025126	0,01159
0,627907	1	0,008043	0,030769	0,058134	0,00201	0,00183
0,790698	1	0,046917	0,153846	0,064485	0,008543	0,00976

Table 3 Example of normalized data (part 2)

sgpt	*sgot*	*alkphos*	*is_patients*
0,685714	0,585366	0,352941	1
0,785714	0,560976	0,258824	1
0,714286	0,585366	0,347059	1
0,685714	0,609756	0,411765	1
0,757143	0,365854	0,058824	1

p-ISSN: 2301-5373
e-ISSN: 2654-5101

2.5.2. Optimization of F-KNN Parameters with PSO

The PSO process will go through the initial stages, namely the initialization of parameters and training data, and data testing. This PSO process will produce optimal values of the parameters k and m used in the F-KNN process. This optimal parameter value will be used in the process of testing the F-KNN method to obtain classification results.

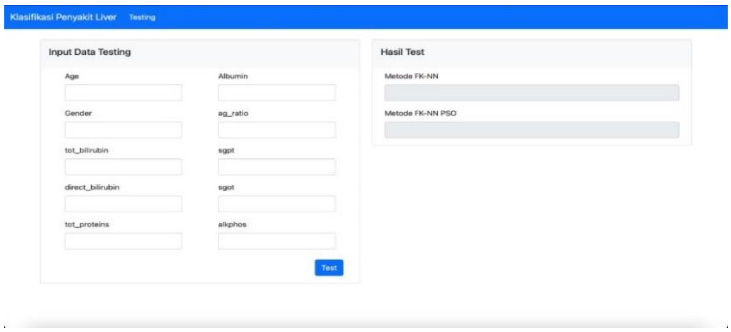
2.6. Testing Fuzzy K-Nearest Neighbor

In the F-KNN process, the optimal k and m parameter values are obtained from the PSO process. The parameters obtained are the optimal combination to get the best accuracy results which will later be used as the fitness value in the PSO process. Each combination of k and m parameters will be tested on the F-KNN method until the stopping condition has been met, namely the maximum iteration or convergent value. The results of this test will be used to obtain the results of the classification in the F-KNN method.

3. Result and Discussion

3.1. Application Interface Design

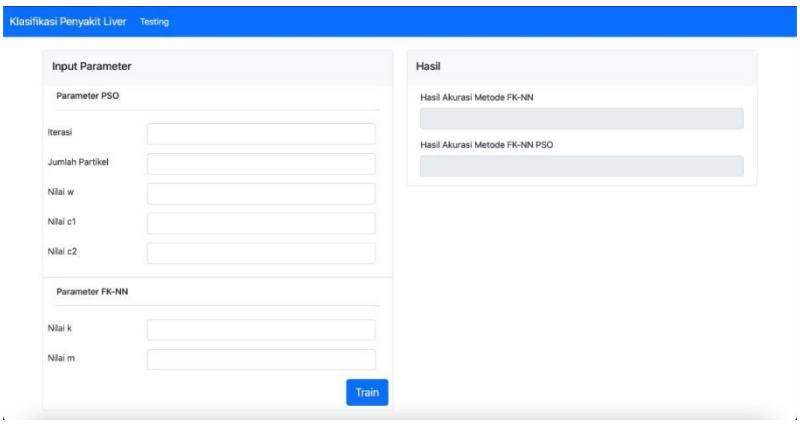
This research will implement the application design to the lines of web-based program code using the Python language and the Django framework. Furthermore, this program will be run on a PC device with an interface that has input from the researcher as the user. The following is the system interface design which can be seen in Picture 1, and Picture 2 below.



Picture 1 Testing interface display

Picture 3 is the page used to test the classification of liver disease. On this page, the user (researcher) is asked to input parameters and 10 features of the data to be tested. Furthermore, the system will provide feedback in the form of classification results obtained from each method.

Picture 4 is the main page of the classification system. This page consists of 2 parts, namely on the left side is a form that functions to input the parameters needed from the FKNN and PS methods for the training process. On the right side is the result of the accuracy of the classification of the implementation of different methods, namely the hybrid FKNN and PSO methods compared



Picture 2 Training interface display

to the FKNN method without the hybrid method. The train button is a button to bring up a table of test results for all training data presented in full.

3.2. Testing

3.2.1. Effect of FKNN Parameters

The values of k and m are the parameters used in the FKNN method to determine the effect of the parameter w, then the experiment is carried out 10 times with the values of k and m using different combinations and the value of w is 0.5, C1 = 0.7 and C2 = 1.3 with 100 iterations.

Table 6 FKNN Parameter Test Results

Number	K	M	FKNN Accuracy
1	18	20	65%
2	14	10	70%
3	14	11	70.10%
4	16	4	71.92%
5	17	6	69.29%
6	18	2	72.80%
7	13	20	68.42%
8	10	10	66.66%
9	6	19	64.91%
10	4	20	63.15%

Table 6 shows that the smaller the M value used, the greater the accuracy is inversely proportional to the K value, whereas the larger the K value used, the greater the accuracy value. The best accuracy results are obtained when K has a value of 18 and M has a value of 2 with an accuracy of 72.80%.

3.2.2. Effect of Parameter W

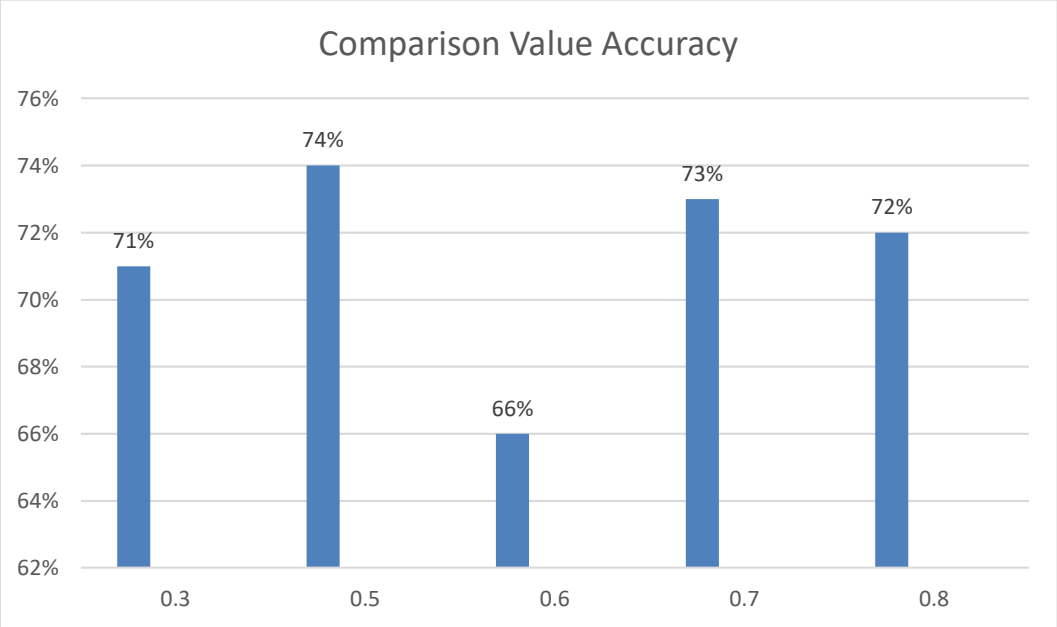
The value of w is a parameter used to control the velocity given by the particle. To determine the effect of the w parameter, the w value will be used as an independent parameter, where the w value to be tested has a range between 0.3 to 0.8. The tests carried out were 10 trials with 100 iterations, based on the results of experiments that have been carried out, the average accuracy and iteration can be seen in Table 7 below.

Table 7 Parameter Test Results W

W	0.5	0.3	0.6	0.7	0.8
Mean Accuracy	74%	71%	66%	73%	72%

p-ISSN: 2301-5373
e-ISSN: 2654-5101

Table 7 shows the value of w 0.5 has the highest average accuracy value in this experiment with a value of 74%. This is due to changes in the position of the testing and training data during the experiment. The difference in the average value of accuracy can be seen in Picture 3 below.



Picture 3 W Accuracy Value Comparison

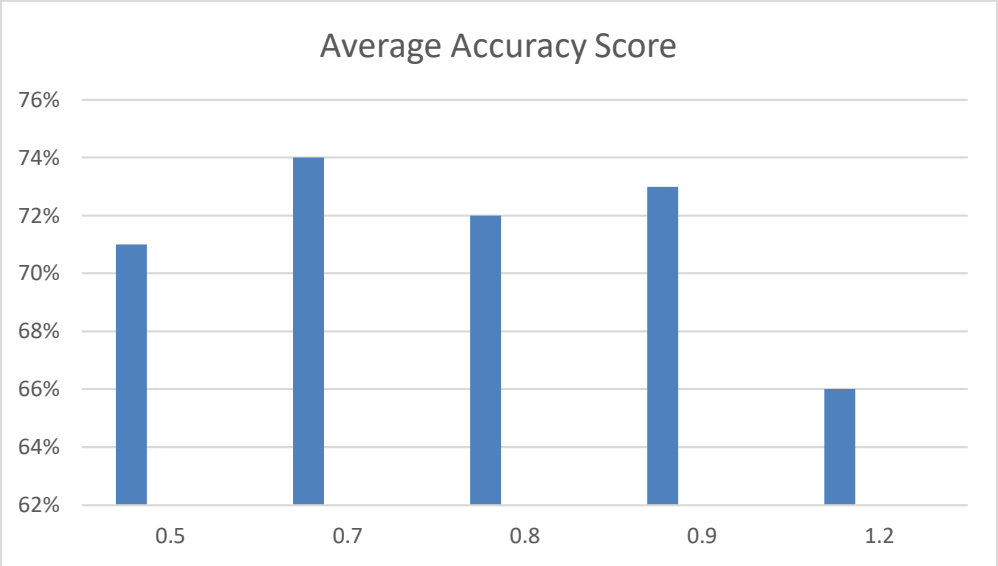
3.2.3. Effect of Parameter C1

Testing the c1 parameter, in this test the c1 parameter is an independent parameter whose test value is in the range of 0.1 to 1.5. The results of the average accuracy and iteration can be seen in table 8 below.

Table 8 Parameter Test Results C1

Nilai C1	0.5	0.7	0.8	0.9	1.2
Mean Akurasi	71%	74%	66%	73%	72%

Based on table 8 shows the value of C1 0.7 has the highest average accuracy value in this experiment with a value of 74%. This is also caused by the displacement of the testing and training data positions during the experiment. The difference in the average value of accuracy can be seen in Picture 4 below.



Picture 4 Accuracy Value Comparison C1

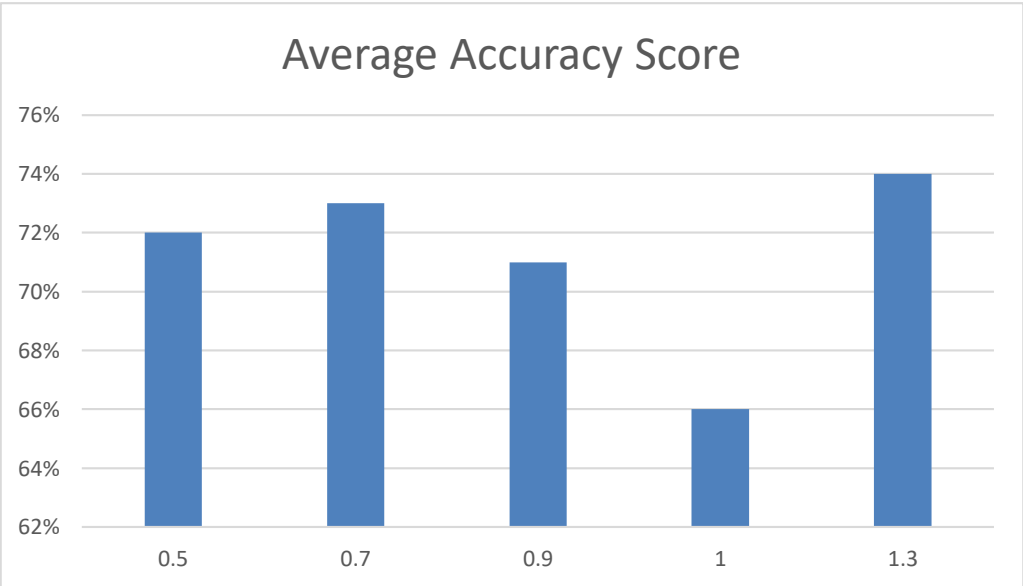
3.2.4. Effect of Parameter C2

Testing the c2 parameter, in this test the c2 parameter is an independent parameter whose test value is in the range of 0.1 to 1.5. The results of the average accuracy and iteration can be seen in table 9 below.

Table 9 Parameter Test Results C2

C1	0.5	0.7	0.9	1	1.3
Mean Accuracy	72%	73%	71%	66%	74%

Based on table 9 shows the value of C1 1.3 has the highest average accuracy value in this experiment with a value of 74%. This is also affected by the displacement of the testing and training data positions during the experiment. The difference in the average value of accuracy can be seen in Picture 5 below.



Picture 5 Accuracy Value Comparison C2

p-ISSN: 2301-5373
e-ISSN: 2654-5101

3.3. Comparison of PSO-FKNN and FKNN Hybrid Methods

The results of testing the best parameters for PSO and FKNN will be used as parameter values for PSO-FKNN to obtain optimal parameter values for liver disease classification so as to get the best accuracy results for classifying liver disease. Based on the test results obtained, the best value of the parameter w is 0.5, c1 is 0.7 and the value of c2 is 1.3. Furthermore, the best combination of parameters for FKNN is when k = 18 and m = 2. The test to compare the accuracy values of the PSO-FKNN and FKNN hybrid methods is carried out with 5 test scenarios where each test scenario is carried out 10 times with 100 iterations, training data and testing on each test will change its position. The results of the comparison of the average accuracy of the PSO-FKNN hybrid method with FKNN can be seen in table 10 below.

Table 10 Comparison of Accuracy Results of PSO-FKNN and FKNN Hybrid Methods

NO.	Used Methods	
	FKNN	PSO-FKNN
1.	68.29%	74%
2.	66.74%	71%
3.	66.57%	66%
4.	67.53%	73%
5.	64.90%	72%
Average	66.80%	71.2%

Based on table 10, shows that the average results obtained from the overall PSO-FKNN hybrid method experiment were 4.4% higher which got an average accuracy value of 71.2% compared to FKNN which only got 66.80% accuracy. The highest accuracy value obtained by the PSO-FKNN hybrid method is 74% and the lowest accuracy value is 66%. Changes in accuracy are also obtained due to changes in the position of the training and testing data on each test while the highest accuracy value for FKNN is 68.29% and the lowest accuracy is 64.90 % this is also due to the displacement of the training and testing data positions in each test.

4. Conclusion

In the research that has been done, the conclusions obtained are as follows.

1. The effect of the PSO parameter on the results of liver disease classification is when the value of w 0.5 has the highest average accuracy value, the value of C1 0.7 has the highest average accuracy value, indicating that the value of C2 1.3 has the highest average accuracy value, The effect of the FKNN parameter is also obtained when the K value is 18 and the M value is 2 in this experiment resulting in the best accuracy.
2. The PSO-FKNN hybrid method for weight optimization of FKNN can increase the accuracy of the FKNN classification process. This is shown in the comparison of the accuracy of FKNN compared to the PSO-FKNN hybrid method, which increased by 4.4 (in percent) for the average accuracy of the PSO-FKNN hybrid method. By using the PSO-FKNN hybrid method, there is an increase in accuracy for the classification of liver disease when compared to the classification of liver disease using the FKNN method.
3. The accuracy range value obtained by the PSO-FKNN hybrid method is 66 to 74 (in percent) compared to the FKNN accuracy value range is 64.90 to 68.29 (in percent), the difference in the accuracy value obtained by PSO-FKNN is influenced by changes in data-position training and testing in each test. The final result of the accumulated average value of the PSO-FKNN hybrid method is 71.2% and for FKNN is 66.80%, so the increase obtained is 4.4% after using PSO.

References

[1] Rahmawati, E. (2015). Analisa Komparasi Algoritma Naïve Bayes dan C4.5 Untuk Prediksi Penyakit Liver. *Jurnal Techno Nusa Mandiri*.

[2] Satria Dwi Nugraha, R. R. (2017). Penerapan Fuzzy K-Nearest Neighbor (FK-NN) Dalam Menentukan Status Gizi Balita. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer, hlm. 925-932*, 8.

[3] Rimbun Siringoringo, R. P.-a. (2017). Hibridisasi Metode Fuzzy K-Nearest Neighbor Dengan Metode Modified Particle Swarm Optimization Pada Pengklasifikasian Penyakit Tanaman Kedelai. *Jurnal & Penelitian Teknik Informatika*.

[4] Wildan Gita Akbari, N. H. (2019). Diagnosis Penyakit Cabai Menggunakan Metode Fuzzy K-Nearest Neighbor (FKNN). *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer, hlm. 1070-1074*.

[5] Wafiyah, F., Hidayat, N., & Perdana, R. S. (2017). Implementasi Algoritma Modified K-Nearest Neighbor (MKNN) untuk Klasifikasi Penyakit Demam. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 1210-1219.

[6] Dizka Maryam Febri Shanti, N. H. (2018). Implementasi Metode F-KNN (Fuzzy K-Nearest Neighbor) Untuk Diagnosis Penyakit Anjing. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer, hlm. 7401-7407*.

[7] Yunus, W. (2018). Algoritma K-Nearest Neighbor Berbasis Particle Swarm Optimization Untuk Prediksi Penyakit Ginjal Kronik. *Jurnal Teknik Elektro CosPhi.*

[8] Darnisa Azzahra Nasution, H. H. (2019). Perbandingan Normalisasi Data Untuk Klasifikasi Wine Menggunakan Algoritma K-NN. *CESS (Journal of Computer Engineering System and Science)*.

Jurnal Elektronik Ilmu Komputer Udayana
Volume 11, No 3. February 2023

p-ISSN: 2301-5373
e-ISSN: 2654-5101

Solar Eclipse Augmented Reality Application Using Natural Feature Tracking Method

Diky Rizky Awan[a1], Agus Muliantara[a2], I Gede Arta Wibawa[a3], Cokorda Rai Adi Pramartha[a4], Ida Bagus Gede Dwidasmara[a5], I Putu Gede Hendra Suputra[a6]

[a]Informatics Department, Faculty of Mathematics and Natural Sciences, Udayana University
South Kuta, Badung, Bali, Indonesia
[1]dikyrizkyawan@gmail.com
[2]muliantara@unud.ac.id
[3]gede.arta@unud.ac.id
[4]cokorda@unud.ac.id
[5]dwidasmara@unud.ac.id
[6]hendra.suputra@unud.ac.id

Abstract

An eclipse is a phenomenon that occurs when a celestial body moves into the shadow of another celestial body, be it a solar eclipse or a lunar eclipse. In the process of understanding this phenomenon, people do not understand what is meant by the eclipse phenomenon. Because of this, we need a teaching aid to help us understand the material.

The application will apply Augmented Reality technology using the Natural Feature Tracking (NFT) method. The application in this research will be made in the form of a native website. With python 3.7.0 64-bit programming language using the IDE (integrated development engineer), namely Visual Studio Code. This application will take advantage of the flask frameworks. This website will be native. Application implementation using native website templates with HTML (Hypertext Markup Language), CSS (Cascading Style Sheets), and Javascript. This application will be intended for a platform website resolution of 16:9 with a breakdown of 1920 x 1080 pixels. However, this ratio will be able to adjust because it applies the bootstrap 4 framework. From some tests conducted, the percentage of accuracy of the percentage of 3D object appearance is 52% on objects performing in the correct position, 34% on objects performing in the wrong position, and 10.67% on objects not appearing. In terms of functions, it has been running and meets the requirements with black box testing with 100% accuracy which is carried out based on 9 functions in the use case diagram.

Keywords: *Solar Eclipse, Augmented Reality, Natural Feature Tracking*

1. Introduction

An eclipse is a phenomenon that occurs when a space object moves into the shadow of another space object, including a solar eclipse or lunar eclipse [1]. A solar eclipse is a phenomenon when the moon is right between the sun and the earth so that the moon blocks the sun's rays from reaching the earth causing the sun not visible from the earth because it is covered by the moon's shadow [2].

In the process of understanding this phenomenon, people do not understand what is meant by the eclipse phenomenon because the eclipse is a fairly rare phenomenon because the frequency of eclipses with identical configurations is quite long repeatedly using the Saros cycle period which will occur every 18 years and 10 days [3]. And in his explanation in the book, only contains theories and images that can only be seen in two dimensions. Because of this, people tend to feel bored quickly and less interested so they have difficulty in understanding and reduce interest in learning about the eclipse phenomenon. Because of this, we need a prop to help in understanding the material.

Augmented Reality (AR) is the combination of real and virtual objects or two-dimensional or three-dimensional virtual in a real environment that runs interactively in real-time, where the virtual objects are integrated into the real world [4]. Natural Feature Tracking (NFT) is one of the image tracking methods to detect and track features naturally in images from angles, lines, or blobs [5].

NFT itself has many algorithms in its application such as the SIFT algorithm, SURF, ORB, BRISK, and others. This study uses the ORB algorithm because the ORB algorithm is a binary descriptor algorithm that is fast and has high resistance in feature detection so it is suitable for applying to AR applications [6].

2. Research Method

2.1 Augmented Reality

The merging of real and virtual objects is possible with the appropriate display technology, interactivity is possible through certain input devices, and good integration requires effective tracking. AR allows users to see the real world equipped with virtual objects that are combined with the real world. AR plays a role in complementing reality rather than completely replacing it. This helps users bring up the desired objects as if the virtual and real objects coexist in the same space [4].

2.2 Natural Feature Tracking

The initial process of NFTs is to rely on a Point of Interest (POI), where objects visible in the environment can be applied directly to detect the feature [7]. After going through the detection process, a feature description will be carried out which gets results in the form of floating-point or binary-based vectors, then continued by matching features that match the image in the database which aims to extract every six degrees of freedom (6DoF) pose which is used as a reference in the freedom of movement of an object. Then, a feature matching process is carried out where the two descriptors on the image will be matched. The matching of each camera frame must be calculated by the descriptor. If a sufficient connection is found between the features in the database and the camera frame, then the pose of the 6DoF camera can be calculated as well as determine the position of the visual object in the real environment.

2.3 Feature Detection

The detection feature is used to search for and identify objects in the image [8]. Feature detection must meet several requirements including fast computational time, stability of changing lighting conditions, and image defocusing, stability to observations from different points of view and invariant scale [7]. The ORB algorithm for feature detection uses an application based on the FAST corner detection algorithm with the initial image that has been converted to grayscale. Here is the flow of the feature detection process in the FAST algorithm. It first determines the p point on the image with the starting position (x,y) and the threshold value [9]. Then determine the radius of 3 pixels from point p so that 16 pixel points are obtained. Next, determine the location of 4 points of 16 pixels. The first point on the coordinates (xp, yp+3), the second point on the coordinates (xp+3, yp), the third point on the coordinates (xp, yp−3), and the fourth point on the coordinates (xp−3, yp). It then compares the intensity of the center point p with the reference of the previous four points for up to 16-pixels. The center point p is the angular point when there are at least 3 points without normal intensity. These categories include as in equation (1).

$$Sp \rightarrow x = \begin{cases} Dark, & lp \rightarrow \mathrm{x} \leq \mathrm{lp} - \mathrm{t} \\ Normal, & lp - t < lp \rightarrow \mathrm{x} < \mathrm{lp} + \mathrm{t} \\ Light, & \mathrm{lp} + \mathrm{t} \leq lp \rightarrow \mathrm{x} \end{cases} \quad \text{.....(1)}$$

Where :

$Sp\rightarrow$: Intensity of the center point (p point)

$Ip\rightarrow$: Pixel intensity x (neighbor intensity point)

t : threshold

2.4 Feature Descriptor

The descriptor aims to describe a region around a key point to generate a description vector because the descriptor describes each detected feature as a vector of a fixed length [6]. The ORB algorithm uses the basis of the improved BRIEF algorithm to calculate the descriptor for each point and the descriptor vector consists of the numbers 0 and 1 because the algorithm is of binary type. To increase the resistance to digital noise, gaussian filtering was first carried out on the image. The binary test can be formulated as in equation (2).

p-ISSN: 2301-5373
e-ISSN: 2654-5101

$$\tau(;,) = \begin{cases} 1, \mathrm{p(x)} < \mathrm{p(y)} \\ 0, \mathrm{p(x)} \geq \mathrm{p(y)} \end{cases} \quad \ldots\ldots(2)$$

Where (x) is the gray value in the x plane around the image feature point, and (y) is the gray value in the y plane around the image feature point. Then, determine a patch with a size of S × S and randomly select N which usually amounts to 256. Then, the brightness value of each pair of points is compared according to the equation and binary assignment. Obtained an N-dimensional vector consisting of a binary string N as in equation (3).

$$f_{\mathrm{N}}(p) = \sum_{1 \leq i \leq N} 2^{i-1} \tau\,(p; \mathrm{x_i}, y_i) \quad \ldots\ldots(3)$$

Where fN (p) is an N-dimensional vector that stores binary values from patches that are 2i-1 in size. The ORB algorithm uses the Steer BRIEF algorithm to calculate the main direction of each feature point because the BRIEF descriptor has no invariants to the rotation, causing easy data loss when the image is rotated.

2.5 Feature Matching

The feature point-based matching method is the main method for image matching because of its simple, fast, high-accuracy calculations, and has invariants to grayscale scales, lighting, and graphic distortions [10]. Feature matching is performed to determine which characteristics are represented in the descriptor of the two images according to the criteria that can be matched using a brute-force matcher.

2.6 Data

The data used in this study is divided into two types, namely image data and virtual object data. There are two kinds of reference imagery data, namely those obtained from camera photos and camera frames in real-time. Virtual object data is component data from an eclipse that is used as a reference in making 3D modeling objects. This data is obtained through books and library sources. That is solar eclipse data which is when the eclipse has a sun-moon-earth configuration on one straight line and a total solar eclipse is when the solar eclipse and the entire sun are covered by the moon.

2.7 Flowchart

First, the image on the camera and the reference image is inputted, followed by the feature detection process for both images. Then after the image is detected the feature is continued with the feature description process. Then, the two features are matched and when the two features match, it is continued by estimating the image pose. Then proceed with the process of displaying virtual objects in a real environment. If the two features do not match then the process will be repeated on the camera image. The process flow can be seen in figure 1.

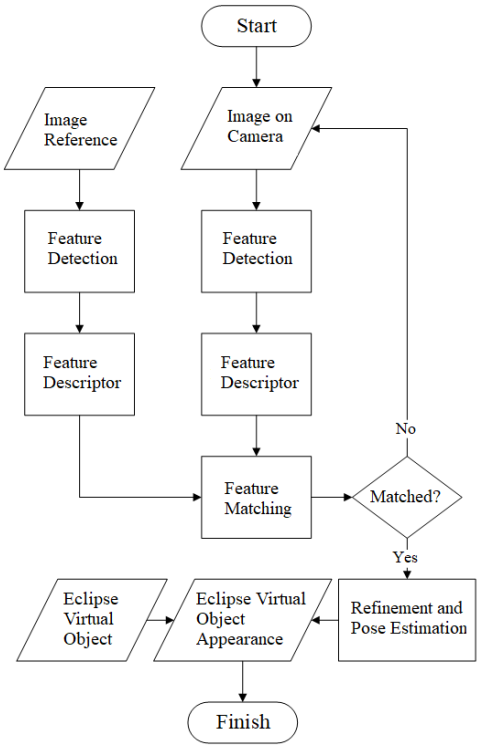


Figure 1 Flowchart

2.8 System Design

This interface design was created using Adobe Illustrator CC 2017 as a design maker or application design framework and implemented in the form of a native website. This design implementation uses a native website template with HTML (Hypertext Markup Language), CSS (Cascading Style Sheets), and Javascript. This design is intended for website platforms with a resolution of 16:9 with a resolution of 1920 x 1080 pixels. This interface design consists of 5 pages, namely the home page that can be seen in figure 2, how to use that can be seen in figure 3, about that can be seen in figure 4, new marker that can be seen in figure 5, and AR that can be seen in figure 6.

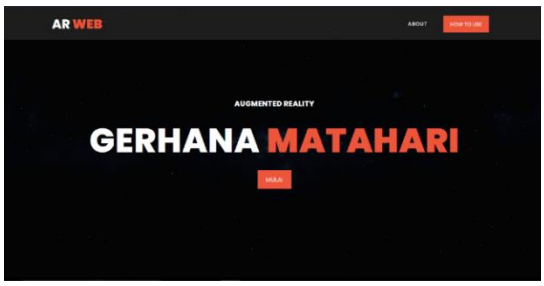


Figure 3 *Home Page*

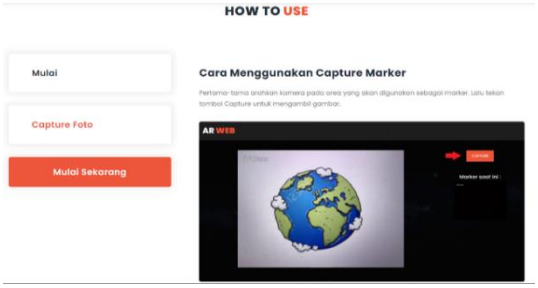


Figure 2 *How to Use*

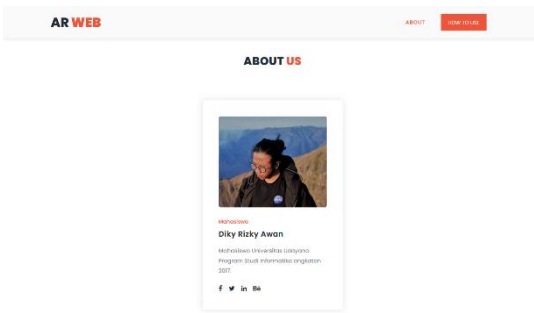


Figure 5 *About*

Figure 4 *New Marker*

p-ISSN: 2301-5373
e-ISSN: 2654-5101

Figure 6 *Augmented Reality*

3. Result and Discussion

3.1. Virtual Object Testing

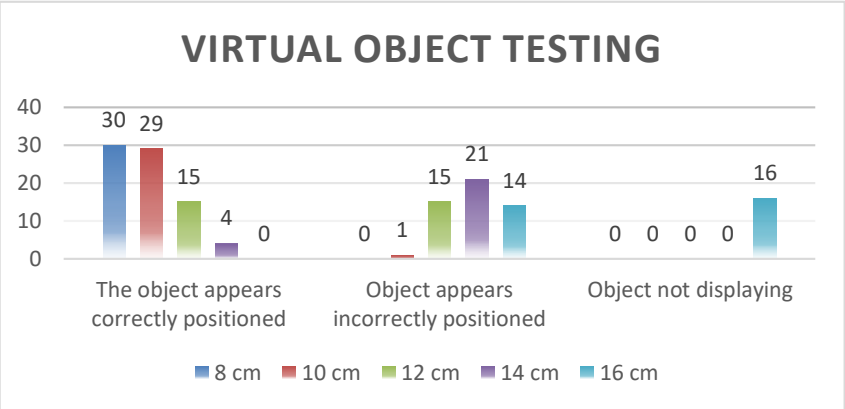
This test, it was carried out at a distance of 10 cm, 12 cm, 14 cm, and 16 cm. Tests are performed with a range of 30 frames on each test which in each frame will be counted as points. Then the points of the number of successful tests will be totaled and divided by the total number of tests and the result will be multiplied by 100%.

Table 1 Distance Aspect Virtual Object Testing Results

No	Result	Distance					Total Point	Percentage
		8 cm	10 cm	12 cm	14 cm	16 cm		
1	The object appears correctly positioned	30	29	15	4	0	78	52%
2	The object appears incorrectly positioned	0	1	15	21	14	51	34%
3	Object not displaying	0	0	0	0	16	16	10.67%

In table 1, it can be seen that the farther the distance, the more likely the accuracy of the object decreases. Testing the objects of this system produced quite good results, judging from the 3D objects that continue to appear at close range, although at long distances they get quite poor results. All points are summed up and divided by the overall test amount and multiplied by 100%. Testing for objects to perform correctly positioned earned a percentage of 52%. Meanwhile, testing for objects to appear incorrectly positioned got a percentage of 34%. The last test for non-performing objects got a percentage of 10.67%.

Figure 7 Virtual Object Testing Graph



3.2. Black Box Testing

This test will be carried out by testing the 9 functions already described in the use case diagram. This test will be performed on several people and each test will be recorded Test Results obtained from each function tried, whether it has been successful or not. Black box testing of this system can be seen in table 2.

Table 2 Black Box Testing

No	Use Case Function	Description	Result
1	Running NFTs with the ORB algorithm	It is the main process of the system which is to apply the ORB method In the NFT process series. This process is used to apply Augmented Reality technology	True
2	View the Home page	It is a process of displaying the home that will appear when the application first runs which has 3 main buttons, namely How to Use, About, Start	True
3	Display the How To use page	It is a process of displaying a How to use page that contains the steps to use the application so that users do not have difficulties.	True
4	Display the About page	It is a process of displaying an about page containing an application description and author biodata	True
5	Displaying 3D objects	Is the process of displaying 3D objects above a predetermined target as an application of Augmented Reality	True
6	Displays marker turnover	It is the process of displaying the marker substitution page that was used when the first start process appeared	True

7	View an AR page	It is the process of displaying a page that is used to turn on the camera and display 3D objects	True
8	Turning on the Camera	It is the process of turning on the camera so that it can be used to see the target and make 3D objects appear in Real-Time	True
9	Retrieving and storing markers	Is a process to retrieve marker data entered by the user. Then the marker is saved as a new marker by the system	True

The results of the black box testing of this system are very satisfactory. This can be seen from each function that the 9 functions in the use case diagram can fit all. It can be concluded that this black box test was successful with 100% accuracy.

4. Conclusion

From the research that has been carried out, several conclusions can be drawn as follows:

1. From virtual object testing, it can be seen that the farther the distance, the more likely the accuracy of the object decreases. Testing the objects of this system produced quite good results, judging from the 3D objects that continue to appear at close range, although at long distances they get quite poor results. Testing for objects to perform correctly positioned earned a percentage of 52%. Meanwhile, testing for objects to appear incorrectly positioned got a percentage of 34%. The last test for non-performing objects got a percentage of 10.67%.
2. From the black box test, it can be concluded that it gets 100% accuracy because the overall functional features of the application tested through the application can run as expected. Features tested according to the features described in the 9-feature Use Case diagram.

References

[1] Sulaiman, R. (2017). Gerhana dan Keharusan Kosmologis Manusia: Tinjauan Filsafat Wujud. Jurnal Kependidikan Dan Sosial Keagamaan.
[2] Khazin, M. (2004). Ilmu Falak Dalam Teori dan Praktek. Yogyakarta: Buana Pustaka.
[3] Siregar, S. (2017). Fisika Tata Surya.
[4] Azuma, R. T. (1997). A Survey of Augmented Reality. Malibu: Hughes Research Laboratories.
[5] Ćuković, S., Gattullo, M., Pankratz, F., & Devedžić, G. (2015). Marker Based vs. Natural Feature Tracking Augmented Reality Visualization of the 3D Foot Phantom.
[6] Hamidia, M., Zenati-Henda, N., Belghit, H., & Bellarbi, A. (2015). Object Recognition Based on ORB Descriptor for Markerless Augmented Reality. 9ème Conférence sur le Génie Electrique.
[7] Carmigniani, J., & Furht, B. (2011). Handbook Of Augmented Reality. Florida: Springer-Verlag New York.
[8] Jakubovic, A., & Velagic, J. (2018). Image Feature Matching and Object Detection Using Brute-Force Matchers. Conference: 2018 International Symposium ELMAR. Zadar: Institute of Electrical and Electronics Engineers (IEEE).
[9] Wahyudi, N., Harianto, R. A., & Endang Setyati, D. I. (2019). Augmented Reality Marker Based Tracking Visualisasi Drawing 2D ke dalam Bentuk 3D dengan Metode FAST Corner Detection. Journal of Intelligent Systems and Computation.
[10] Luo, C., Yang, W., Huang, P., & Zhou, J. (2019). Overview of Image Matching Based on ORB Algorithm. ICSP 2019. Weihai: IOP Publishing

This page is intentionally left blank.

Jurnal Elektronik Ilmu Komputer Udayana
Volume 11, No 3. February 2023

p-ISSN: 2301-5373
e-ISSN: 2654-5101

Pendeteksian Pelanggaran Rambu Larangan Berhenti Berbasis Mobile dengan Pendekatan Spatial Map Matching

Nusan Bagus Wibisana[a1], I Komang Ari Mogi[a2], Ida Bagus Gede Dwidasmara[a3], Cokorda Rai Adi Pramartha[a4], I Gusti Agung Gede Arya Kadyanan[a5], I Wayan Supriana[a6]

[a]Program Studi Informatika, Universitas Udayana
Jimbaran, Badung, Bali, Indonesia
[1]bagusnusan25@gmail.com
[2]arimogi@unud.ac.id
[3]dwidasmara@unud.ac.id
[4]cokorda@unud.ac.id
[5]gungde@unud.ac.id
[6]wayan.supriyana@unud.ac.id

Abstrak

Kemacetan merupakan salah satu hal yang tidak disukai oleh semua orang, kemacetan juga dapat meningkatkan risiko kecelakaan berkendara. Adapun salah satu penyebab kemacetan ialah karena pengendara berhenti pada pinggir jalan yang mengurangi tingkat efektifitas jalan, oleh karena itu dibutuhkan penempatan rambu larangan berhenti yang berguna untuk mempertegas bahwa pada zona yang ditentukan benar-benar dilarang untuk berhenti sehingga dapat mengurangi kemacetan yang ada. Namun penempatan rambu ini terkadang masih memiliki kekurangan dikarenakan kurangnya instrumen pendeteksian untuk mendeteksi pelanggaran terhadap rambu yang ada. Oleh karena itu dibutuhkan sebuah sistem pendeteksi pelanggaran rambu larangan berhenti yang dimana dengan menggunakan bantuan GPS dapat diketahui posisi pengguna dan dilakukan perhitungan menggunakan pendekatan metode spatial map matching untuk memutuskan apakah pengguna sedang berhenti pada zona larangan berhenti atau tidak. Adapun berdasarkan hasil pengujian yang dilakukan menggunakan 100 data didapat tingkat akurasi dari sistem yang dibangun mencapai 97%

Kata Kunci : Deteksi Pelanggaran lalu lintas, GPS, Spatial Map Matching, Rambu Larangan Berhenti, Aplikasi Mobile.

1. Pendahuluan

Kemacetan merupakan hal yang tidak disenangi oleh semua orang. Kemacetan juga merupakan salah satu faktor yang meningkatkan risiko kecelakaan dalam berkendara. Adapun salah satu penyebab kemacetan yang terjadi ialah terdapat pengendara yang berhenti pada pinggir jalan yang sebagaimana tertuang dalam pasal 118 Undang-Undang Nomor 22 Tahun 2009 tentang Lalu Lintas dan Angkatan Jalan (LLAJ) ayat 15 dijelaskan dimana setiap kendaraan bermotor dapat berhenti dijalan raya terkecuali terdapat rambu atau marka larangan berhenti atau tempat-tempat yang sekiranya dapat membahayakan keselamatan dan mengganggu lalu lintas serta pengendara. [6]

Pada kondisi jalan raya sesungguhnya meskipun sudah terpasang rambu larangan berhenti terkadang pengendara tidak menyadari akan keberadaan rambu tersebut dan tetap melanggar sehingga dibutuhkan sebuah instrumen yang dapat mendeteksi dan menberitahu bahwa pengendara sedang berhenti pada zona larangan berhenti sehingga pengendara dapat mengetahui bahwa ia telah melakukan pelanggaran dan segera meninggalkan zona larangan berhenti.

GPS merupakan salah satu instrumen yang dapat mengetahui posisi pengguna menggunakan bantuan satelit yang dimana nilai dari GPS akan berupa garis lintang dan garis bujur [1][2][7][8]. Nilai dari gps ini tidak dapat serta merta menentukan posisi pengendara pada jalan sesungguhnya sehingga dibutuhkan pemetaan dan penentuan apakah kondisi pengendara sedang melakukan pelanggaran atau tidak.

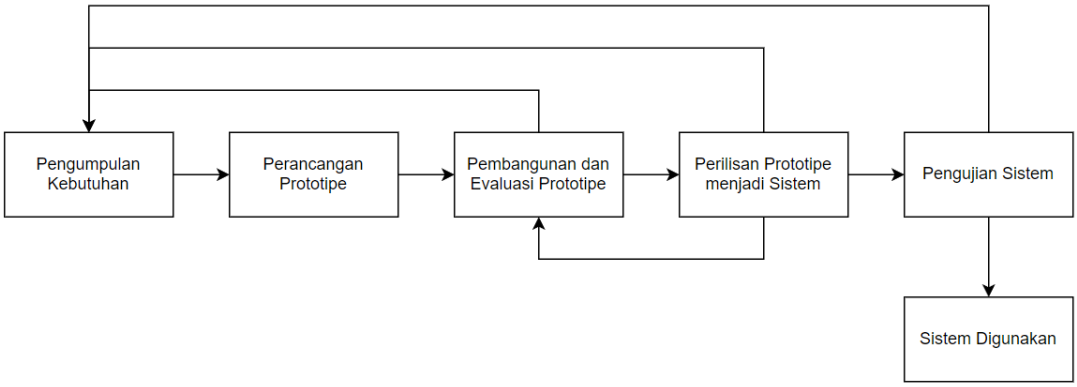
Salah satu metode yang dapat digunakan dalam mencocokan antara lokasi pengguna dengan rambu ialah dengan menggunakan pendekatan metode spatial map matching yang dimana metode ini akan mencocokan antara lokasi titik pengguna dengan lokasi titik rambu larangan berhenti. Sehingga dapat diputuskan apakah pengguna melakukan pelanggaran atau tidak [3][4].

Perangkat mobile smartphone merupakan salah satu perangkat yang hampir dimiliki oleh semua orang. Perangkat mobile smartphone jaman sekarang umumnya sudah terintegrasi sensor GPS sehingga tidak diperlukan menanam atau membawa sensor GPS tambahan[5].

Adapun hasil dari penelitian ini diharapkan diketahui tingkat efektivitas dari metode yang diterapkan dengan cara mengukur tingkat akurasi dari sistem yang dibangun sehingga sistem yang dibangun dapat digunakan sebagaimana mestinya.

2. Metodologi Penelitian

Adapun metodologi penelitian yang dilakukan lebih berfokus kepada pengembangan perangkat lunak yang dimana pengembangan perangkat lunak ini menggunakan pendekatan kerangka kerja prototipe. Gambaran alur pengembangan sistem dapat dilihat pada gambar 1



Gambar 1. Pengembangan Sistem

Adapun penjelasan dari tiap-tiap bagian pada gambar dapat dijelaskan sebagai berikut:

2.1. Pengumpulan Kebutuhan

Pengumpulan kebutuhan dilakukan untuk mengetahui kebutuhan yang sekiranya diperlukan oleh sistem sehingga sistem dapat berjalan dengan sebagaimana mestinya. Adapun pengumpulan kebutuhan terdiri dari 3 yaitu:

a. Kebutuhan masukan sistem

adapun masukan sistem yang dibutuhkan antara lain:

1. Data lokasi pengguna

   Adapun lokasi pengguna yang akan menjadi inputan dapat dilihat pada tabel 1.

Table 1. Posisi Pengguna

id	Id_perjalanan	latitude	longitude	Waktu
1	3	-8.675829	115.2593835	09/05/2022 10:42:18
2	3	-8.6758314	115.2593835	09/05/2022 10:42:23
3	3	-8.6758345	115.2593694	09/05/2022 10:42:28

p-ISSN: 2301-5373
e-ISSN: 2654-5101

2. Titik rambu zona larangan berhenti

Table 2. Lokasi titik rambu

Id	Status	latitude	longitude
205	1	-8.675415687617011	115.25950938460483
207	1	-8.675672844784271	115.25942891833438
209	0	-8.675928318837846	115.25934576985493

b. Kebutuhan pengolahan data

pengolahan data dilakukan melakukan proses dari inputan data yang telah diterima oleh sistem. Adapun pengolahan data pada sistem dilakukan dengan menggunakan pendekatan spasial map matching yang dimana akan dijelaskan sebagai berikut:

rumus utama untuk melakukan map matching kali ini yaitu dengan menghitung tingkat probabilitas observasi antara titik pengguna dengan titik kandidat dengan radius tertentu. Adapun pemilihan titik kandidat yaitu menggunakan jarak radius sejauh 30 meter dari titik pengguna. Rumus untuk probabilitas observasi ialah [4]:

$$N\left(C_i^j\right) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}$$

Adapun beberapa penjelasan-penjelasan dari penggunaan rumus diatas antara lain:

1. Perhitungan jarak

Perhitungan jarak digunakan untuk mengetahui jarak antara titik pengguna dengan titik kandidat kemudian diambil hanya titik dengan jarak kurang atau sama dengan 30 meter, perhitungan jarak menggunakan pendekatan perhitungan Euclidean yaitu dengan menghitung jarak titik vektor (X1,Y1) dengan titik vektor (X2,Y2) yang dimana X merupakan nilai latitude dan Y merupakan Nilai Longitude. Kemudian hasil perhitungan dikonversi menjadi satuan meter dengan melakukan perkaliaan dengan 11.1 dan dibagi dengan 0.0001 yang dimana tiap jarak 0.0001 pada peta setara dengan 11.1 meter pada dunia nyata. Rumus perhitungan jarak dapat dilihat sebagai berikut.

$$jarak = \sqrt{(X_1 - X_2)^2 + (Y_1 - Y_2)^2} * \frac{11.1}{0.0001}$$

2. Perhitungan jarak rata-rata (Mean)

Setelah kita mengetahui jarak masing-masing antara titik pengguna dan titik kandidat maka dilakukan perhitungan rata-rata antara titik pengguna dengan seluruh titik kandidat yang ada. Adapun rumus mencari nilai rata-rata dapat dilihat sebagai berikut.

$$\mu\,(rata - rata) = \frac{1}{N}\sum(X^i)$$

3. Perhitungan standar deviasi

Setelah mengetahui nilai rata-rata maka dapat dilanjutkan dengan perhitungan standart deviasi yang dimana rumus untuk mencari nilai standar deviasi dapat dilihat sebagai berikut.

$$\sigma^2(deviasi) = \frac{1}{N}\sum(X_i - \mu)^2$$

c. Kebutuhan keluaran sistem

Kebutuhan keluaran sistem merupakan kebutuhan yang berisikan hasil keluaran dari sistem yang dimana pada penelitian ini keluaran dari sistem ialah hasil penentuan yang menunjukan apakah pengguna berhenti pada zona larangan berhenti atau tidak. Adapun kondisi keluaran sistem dapat dilihat pada tabel 3.

Table 3. Keluaran Sistem

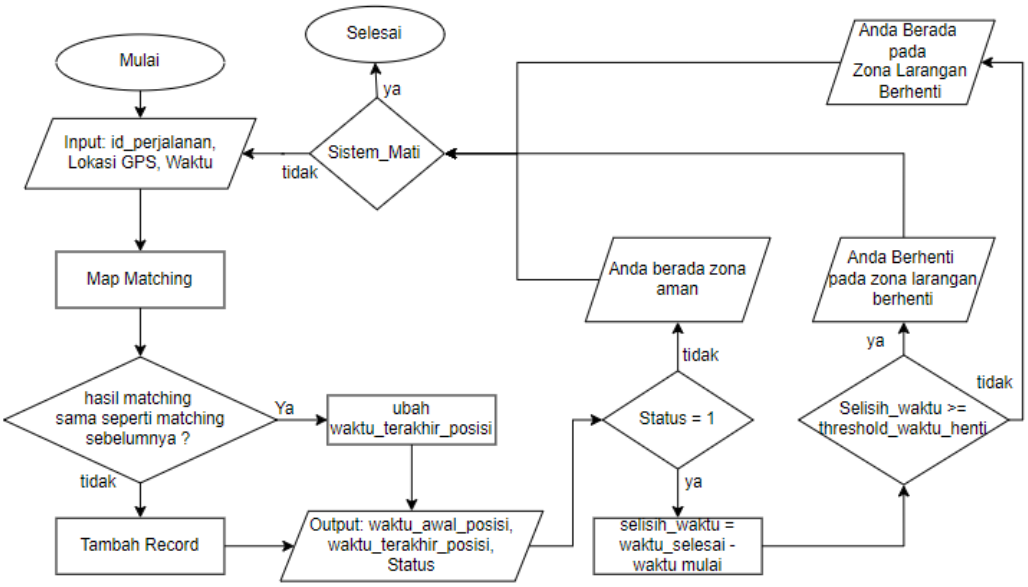
Kondisi Pengguna	Waktu	Keluaran yang diharapkan
Berada pada Rambu Larangan	Waktu pengguna > waktu threshold	Pengguna berhenti pada zona larangan berhenti
Berada pada Rambu Larangan	Waktu pengguna < waktu threshold	Pengguna berada pada zona larangan berhenti
Diluar Zona larangan	-	Zona Aman
Hanya melintasi zona larangan	-	Pengguna tidak ditampilkan sedang berhenti pada zona larangan berhenti

2.2. Perancangan Prototipe

Tahapan perancangan prototipe dilakukan untuk mengetahui apa saja yang diperlukan sehingga kebutuhan-kebutuhan pada pengumpulan kebutuhan dapat terpenuhi dan terlaksana. Adapun perancangan prototipe dari sistem antara lain:

a. Perancangan FlowChart

Flowchart atau diagram alur berisikan alur kerja dari sistem yang dibangun. Adapun alur kerja dari sistem ini dapat dilihat pada gambar 2.



Gambar 2. Flowchart Sistem

b. Penentuan fitur sistem

setelah flowchart terbentuk maka kita dapat mengetahui fitur fitur yang dibutuhkan pada sistem, adapun fitur-fitur tersebut antara lain:

1. Fitur Penambahan titik rambu larangan

   Fitur penambahan titik rambu larangan merupakan salah satu fitur yang sangat penting dikarenakan melalui fitur ini akan ditentukan dilokasi mana saja rambu larangan berhenti berada.

p-ISSN: 2301-5373
e-ISSN: 2654-5101

2. Fitur perekaman posisi pengguna
Fitur perekaman berfungsi untuk melakukan perekaman pada posisi pengguna sehingga dapat diketahui dimana lokasi pengguna yang sedang menggunakan sistem.
3. Fitur perhitungan pemrosesan data
Fitur perhitungan pemrosesan data bertujuan untuk melakukan perhitungan terhadap lokasi pengguna sehingga dapat diputuskan apakah pengguna sedang berhenti atau tidak pada rambu larangan
4. Fitur menampilkan hasil perhitungan
Fitur menampilkan hasil perhitungan digunakan untuk mengetahui kondisi pengguna saat ini apakah sedang melanggar terhadap zona larangan atau tidak.

2.3. Pengembangan dan Evaluasi Prototipe
Dari rancangan prototipe yang telah dibuat maka dilanjutkan dengan pengembangan terhadap prototipe dan evaluasi terhadap prototipe yang ada. Adapun prototipe yang dibangun berbentuk pecahan-pecahan dari fitur yang belum disatukan menjadi sebuah sistem utuh. dan evaluasi dilakukan agar mengetahui apakah fitur yang dibuat telah dapat bekerja menggunakan data buatan atau tidak.

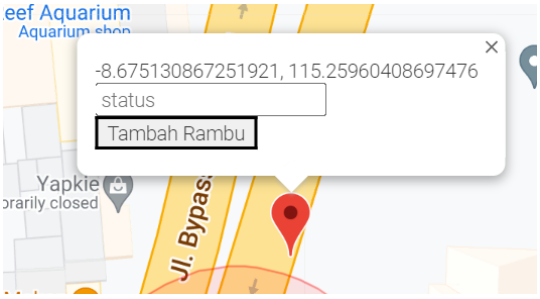
2.4. Pengembangan prototipe menjadi sistem utuh
Pengembangan prototipe pada tahap ini dilakukan bertujuan untuk menggabungkan antara beberapa prototipe lain yang sudah dapat berjalan menjadi sebuah sistem yang utuh sehingga aplikasi dapat berjalan sebagaimana mestinya'

2.5. Pengujian dan Evaluasi Sistem
Pengujian sistem dilakukan dengan cara menguji kesesuai antara masukan pengguna dengan keluaran sistem yang diharapkan dalam mendeteksi pelanggaran rambu larangan berhenti. Hasil dari pengujian kemudian dihitung tingkat ketepatan masukan dengan keluarannya untuk mengetahui tingkat akurasi dari sistem yang dibangun. Adapun skenario pengujian dari sistem dapat dilihat pada tabel 1.

Tabel 1. Tabel skenario pengujian

No	Kondisi pengujian	Hasil keluaran	Nilai
1	Pengguna berhenti pada zona larangan	Pengguna berhenti pada zona larangan	True Positive
2	Pengguna tidak berada pada zona larangan	Pengguna berhenti pada zona larangan	False Positive
3	Pengguna tidak berhenti pada zona larangan	Pengguna tidak berhenti pada zona larangan / Zona Aman	True Negative
4	Pengguna berhenti pada zona larangan	Pengguna tidak berhenti pada zona larangan / Zona Aman	False Negative

3. Hasil dan Pembahasan
3.1. Penentuan titik zona larangan berhenti
Untuk menentukan titik zona larangan berhenti admin pertama perlu menambahkan titik zona larangan berhenti pada peta, kemudian menambahkan arah berhenti zona larangan. Adapun penambahan dapat dilihat pada gambar 3.



Gambar 3. Gambar penentuan titik rambu larangan berhenti

Pada gambar 3 Dapat dijelaskan saat admin melakukan klik pada peta maka akan menampilkan sebuah pop-up yang berisi titik koordinat dan pengguna harus memasukan status yang dimana 1 untuk rambu larangan berhenti dan 0 untuk titik henti zona larangan.

Untuk melakukan validasi kesesuaian titik pada lokasi sebenarnya pengguna dapat melakukan klik animasi orang yang dapat dilihat pada gambar 4

Gambar 4. Ikon Animasi Validasi

Setelah dilakukan klik maka akan muncul tampilan lokasi titik pada jalan yang dapat dilihat pada gambar 5

Gambar 5. Validasi lokasi titik

Pada gambar 5 lokasi titik rambu rambu larangan dan titik berhenti rambu dilingkari dengan warna merah.

p-ISSN: 2301-5373
e-ISSN: 2654-5101

3.2. Perhitungan map matching.

Setelah admin menambahkan titik pengguna selanjutnya dilakukan perhitungan secara manual dengan mengambil titik sampel koordinat pengguna dan dilakukan perhitungan yang dimana nilai hasil perhitungan dapat dilihat pada tabel 2.

Tabel 2. Perhitungan Map Matching

Titik Pengguna	Id_titik kandidat	Titik Kandidat	Jarak	Mean (μ)	Dev (σ^2)	N
-8.675829, 115.2593835	207	-8.675415687617011, 115.25950938460483	18.05151	14.9223	3.12921	0.07735
	209	-8.675672844784271, 115.25942891833438	11.79309		3.12921	0.07735

Dari hasil perhitungan didapat bahwa nilai kandidat memiliki nilai N atau nilai probabilitas yang sama, oleh karena itu sistem akan memilih titik dengan id titik kandidat yang lebih kecil yaitu titik 207. Kemudian dilakukan penambahan pada sistem terhadap waktu durasi pada titik yang dimana dapat dilihat pada tabel 3

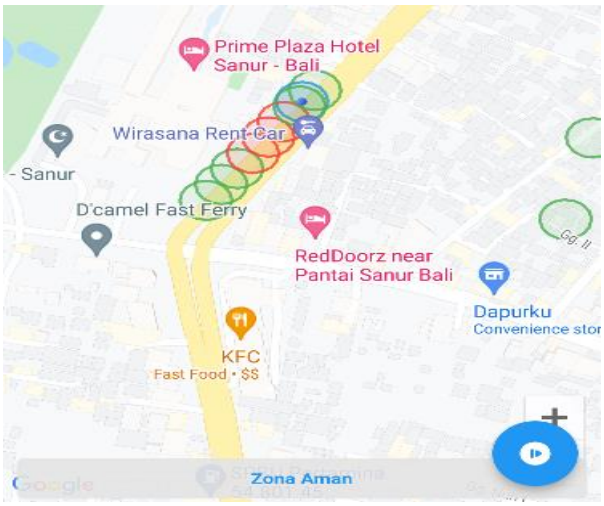
Tabel 3. Kondisi pengguna

Id_perjalanan	Titik_matching	Status	Waktu mulai	Waktu selesai
3	207	1	09/05/2022 10:42:18	09/05/2022 10:43:08

Pada tabel 3 dapat dijelaskan id perjalanan merupakan id yang membedakan tiap kendaraan beserta dengan waktu saat kendaraan melakukan perjalanan. Titik matching merupakan titik posisi pengguna berdasarkan titik kandidat hasil perhitungan. Status 1 merupakan status dimana lokasi tersebut merupakan zona larangan berhenti. Waktu mulai dan waktu selesai merupakan waktu untuk mengetahui berapa lama pengguna berada pada titik tersebut.

3.3. Pembangunan fitur perekaman data

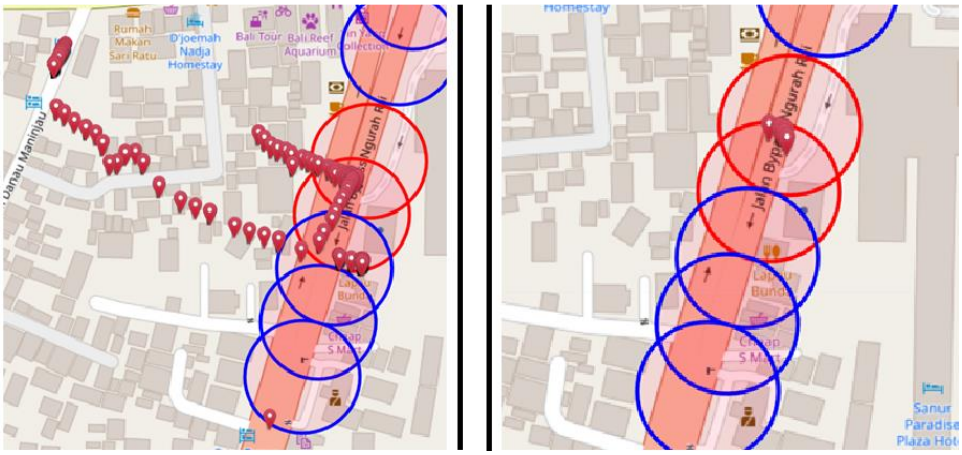
Fitur ini dilakukan untuk mengetahui posisi pengguna secara otomatis, sehingga nantinya dapat dilakukan perhitungan posisi secara cepat. Adapun hasil dari perekaman data dapat dilihat pada gambar 6.



Gambar 6. Fitur Perekaman data

3.4. Sebaran data titik perekaman

Setelah perekaman data secara otomatis dilakukan maka selanjutnya dilakukan pemeriksaan terhadap hasil pengambilan data guna melakukan validasi bahwa nilai masukan dari sistem sudah sesuai dengan hasil yang diinginkan. Adapun sebaran data dapat dilihat pada gambar 7.



Gambar 7. Hasil sebaran data

Pada gambar 7 dapat dijelaskan, pada awal perekaman sistem menggunakan waktu interval pengambilan data 1 detik. namun setelah melihat hasil perekaman sistem ternyata memiliki kejanggalan dimana pengambilan data yang terlihat seperti bergerak seperti pada gambar 7 bagian kiri namun pada kondisi sebenarnya kondisi perekaman sedang dalam kondisi berhenti.

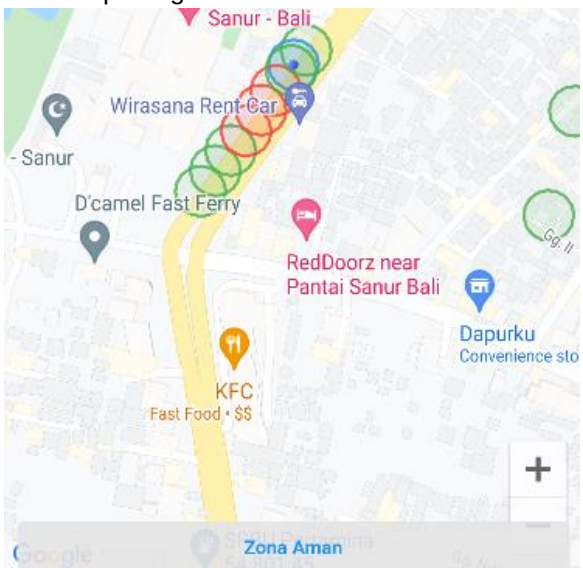
Melihat hasil rekaman yang seperti ini maka dilakukan perubahan interval dengan menggunakan beberapa waktu berbeda sehingga didapatkan waktu optimal. Setelah dilakukan beberapa kali percobaan maka diputuskan proses pengambilan data lokasi pengguna dilakukan dengan interval sistem akan meminta data baru apabila request permintaan data sebelumnya telah selesai atau menggunakan fungsi location.onlocationchanged pada pemrograman selesai di proses sehingga sebaran dapat jauh lebih mengumpul yang dapat dilihat pada gambar 7 bagian kanan

3.5. Output hasil perhitungan

Sesuai dengan perencanaan pembanguann sistem, keluaran dari sistem dibagi menjadi 3 yaitu peringatan tidak berada pada zona larangan, pengguna memasukin zona larangan dan pengguna berada pada zona larangan berhenti yang dimana akan dijelaskan sebagai berikut:

1. Tampilan Pengguna tidak berada pada zona larangan berhenti

Apabila pengguna tidak berada pada zona larangan berhenti maka sistem akan menampilkan Zona aman yang dapat dilihat pada gambar 8.



Gambar 8. Tampilan berada pada Zona Aman

p-ISSN: 2301-5373
e-ISSN: 2654-5101

2. Tampilan saat pengguna memasuki zona larangan berhenti

Apabila pengguna memasuki zona larangan berhenti tetapi belum melewati batas waktu threshold maka akan ditampilkan peringatan pengguna berada pada zona larangan berhenti yang dapat dilihat pada gambar 9.

Gambar 9. Tampilan pengguna memasuki zona larangan berhenti

3. Tampilan pengguna berhenti / melanggar pada zona Larangan Berhenti

Apabila pengguna telah melewati batas waktu yang ditetapkan maka akan muncul tampilan pengguna berhenti pada zona larangan berhenti yang dapat dilihat pada gambar 10.

Gambar 10. Tampilan peringatan berhenti

3.6. Pengujian Sistem

Penjelasan penilaian skenario pengujian yang dilakukan dapat dilihat pada tabel 4.

Tabel 4. Penilaian skenario pengujian

Skenario Pengujian	Hasil Keluaran Sistem	Nilai
Pengguna Berhenti pada zona larangan	Pengguna berhenti pada zona larangan	True Positif
Pengguna berhenti pada zona larangan	Pengguna tidak berhenti pada zona larangan	False Negatif
Pengguna tidak berhenti pada zona larangan	Pengguna tidak berhenti pada zona larnagan	True Negatif
Pengguna hanya melintasi Zona larangan	Pengguna tidak berhenti pada zona larangan	True Negatif

Adapun hasil nilai pengujian dari sistem menggunakan 100 buah data dapat dilihat pada tabel 5

Tabel 5. Rangkuman nilai hasil pengujian

Hasil Prediksi	Jumlah Data
True Positif (TP)	37
False Positif (FP)	0
True Negatif (TN)	60
False Negatif (FN)	3

Setelah diketahui jumlah hasil prediksi maka dilakukan perhitungan akurasi sehingga didapat angka sebagai berikut.

$$\text{Akurasi} = \frac{TP+TN}{TP+FP+TN+FN} * 100\ \% = \frac{37+60}{37+0+60+3} * 100\% = 97\%$$

4. Kesimpulan

Berdasarkan hasil dari pengembangan sistem yang dilakukan dapat diketahui dalam menentukan titik larangan berhenti pertama admin harus menempatkan rambu larangan berhenti sesuai dengan lokasi yang diinginkan, kemudian admin dapat memastikan lokasi titik rambu dengan fitur melihat titik pada peta. Dari hasil pengujian yang dilakukan didapatkan tingkat akurasi dari sistem ialah 97% yang dimana 37 data bernilai true positif, 3 data false negatif dan 60 data bernilai true negatif.

Referensi

[1] Abidin, H.Z. 2007. Penentuan Posisi dengan GPS dan Aplikasinya , PT Pradnya Paramita, Jakarta

[2] Andi. 2009. Global Positioning System, Penerbit Andi, Yogyakarta

[3] Danil, Ode, dkk. 2021. Implementasi Global Navigation Satellite Sistem (GNSS) Pada Sistem Presensi Menggunakan Metode Spatial Map Matching. semanTIK. 7(1). Kendari

[4] Maftukhin, M.R. 2018. Implemetasi Algoritma Spatial Map Matching Untuk Mengetahui Lokasi kendaraan Melalui Aplikasi GPS Tracker.

[5] Qorib, Widiartha. dkk. 2018. Rancang Bangun Sistem Deteksi Posisi Dengan Memanfaatkan GPS Pada Smartphone Berbasis Google Maps API Studi Kasus Pemantauan Pada Anak dan Remaja

[6] Undang-Undang no 22 tahun 2009 tentang lalu lintas dan angkutan jalan. 2009. http://jdih.dephub.go.id/assets/uudocs/uu/uu_no.22_tahun_2009.pdf dilihat 5 April 2021

[7] Winardi. 2006. Penentuan Posisi Dengan GPS Untuk Survei Terumbu Karang, Puslit Oseanografi – Lipi, Jakarta.

[8] https://www.gps.gov/technical/ps/2020-SPS-performance-standard.pdf dilihat 18 Juni 2021

Jurnal Elektronik Ilmu Komputer Udayana
Volume 11, No 3. February 2023

p-ISSN: 2301-5373
e-ISSN: 2654-5101

Pengenalan Pola Karakter Tulisan Tangan Aksara Bali Menggunakan Fitur *Zoning*, *Direction,* dan *Backpropagation*

I Kadek Agus Chandra Pradika[a1], Luh Arida Ayu Rahning Putri[a2], I Gede Santi Astawa[a3], Ida Bagus Gede Dwidasmara[a4], I Gede Arta Wibawa[a5], Made Agung Raharja[a6]

[a]Program Studi Informatika, Universitas Udayana
Jl. Kampus Bukit Jimbaran, Gedung BF Jimbaran, Badung, Bali 80361, Indonesia
[1]chandra.pradika@student.unud.ac.id
[2]rahningputri@unud.ac.id
[3]santi.astawa@unud.ac.id
[4]dwidasmara@unud.ac.id
[5]gede.arta@unud.ac.id
[6]made.agung@unud.ac.id

Abstract

Balinese script has a character with high similarity. Identifying classes between characters requires an optimal pattern recognition model by maximizing the use of feature extraction methods. The use of feature extraction methods in the dataset aims to obtain the characteristic value of each character, in the case of Balinese script data, a method with detailed feature retrieval is used using the zoning method. This study also added the additional factor of the direction feature. The learning method uses a neural network with a backpropagation algorithm. Tests on character data get the highest accuracy in the combination of 16x16 zone ICZ + ZCZ zoning features with the addition of direction features that are 91.18%, ZCZ zone 16x16 zoning and direction with 86.82% accuracy, and ICZ zoning 16x16 and direction zones with 82.43% accuracy. The highest increase in accuracy is found in the ZCZ feature with a difference of addition of 4.36%. The implementation of the model in word testing has an accuracy of 66.2 % and the results of segmentation testing are 97.33 %.

Keywords: *Balinese Script, Zoning, Direction, Neural Network, Backpropagation*

1. Pendahuluan

Aksara Bali merupakan simbol visual dari bahasa Bali. Aksara Bali dapat dikelompokkan berdasarkan bentuk dan fungsinya, yaitu aksara biasa dan aksara suci. Aksara biasa terdiri dari aksara wreastra dan swalalita. Disebut aksara biasa karena telah terbiasa dipergunakan oleh masyarakat Bali di dalam tulis-menulis untuk memenuhi kebutuhan kehidupan sehari-hari dalam berhubungan dengan sesama melalui aksara [1]. Aksara wreastra dipergunakan sebagai media komunikasi nonverbal, seperti menulis kesusasteraan, ilmu pengetahuan, seni budaya, hukum (awig-awig), dan lain-lain.

Perkembangan globalisasi mempengaruhi pergeseran fungsi dari aksara Bali oleh karena itu, diperlukan sistem untuk mengenali karakter atau huruf aksara Bali sebagai media pembelajaran masyarakat. Saat ini di Bali telah dilakukan berbagai upaya pelestarian aksara Bali, salah satu implementasi teknologi yang digunakan adalah pengenalan pola. Karakter yang dihasilkan dari tulisan tangan memiliki variasi berbeda-beda tergantung pada penulisnya. Satu orang bahkan dapat memiliki pola penulisan yang berbeda pada karakter yang sama. Variasi pola penulisan yang tidak konsisten menyebabkan kesulitan dalam mengenali pola karakter antar kelas. Cara mendapatkan ciri dari karakter, salah satunya dapat dilakukan dengan membagi karakter menjadi beberapa wilayah, sehingga menghasilkan set fitur yang memiliki ciri unik pada satu karakter. Pendekatan metode ekstraksi fitur yang menerapkan pembagian wilayah adalah metode *Zoning.*

Secara umum, dengan metode ekstraksi fitur *zoning*, citra akan dibagi menjadi beberapa zona yang berukuran sama dan diambil cirinya [2]. Keunggulan *zoning* dibandingkan metode lainnya, antara lain merupakan metode pencirian yang sederhana, kompleksitas yang rendah dan memiliki perhitungan

yang cepat dalam mengekstraksi ciri suatu karakter [3]. Variasi ekstraksi fitur zoning yang digunakan adalah *Image Centroid and Zone* (ICZ), *Zone Centroid and Zone* (ZCZ), dan gabungan ICZ dan ZCZ. Pengenalan menggunakan metode *zoning* gabungan ICZ dan ZCZ pada metode *Feed Forward Backpropagation Neural Network.*

Penerapan metode *zoning* pada aksara Bali sudah dilakukan sebelumnya oleh Wiguna dan Muliantara pada tahun 2019, dari hasil pengenalan mendapatkan akurasi sebesar 72,31%. Rendahnya nilai akurasi yang diperoleh dalam penelitian ini disebabkan oleh penggunaan metode zonasi yang terkadang memberikan nilai yang sama untuk karakter yang berbeda [4]. Pada aksara Bali terdapat beberapa karakter yang memiliki tingkat kemiripan yang tinggi. Analisa terhadap kemiripan 18 aksara wianjana mendapatkan beberapa aksara Bali yang memiliki tingkat kemiripan yang tinggi [5]. Beberapa karakter aksara Bali dengan tingkat kemiripan yang tinggi sulit untuk dikenali. Oleh karena itu, dibutuhkan sebuah metode yang sesuai untuk membedakan antar karakter, sehingga dalam proses pengenalan pola lebih akurat.

Metode ektraksi fitur *zoning* memerlukan tambahan metode untuk mengekstraksi ciri dari karakter aksara Bali yang memiliki tingkat kemiripan yang tinggi. Ekstraksi fitur yang mengekstraksi ciri berbasis arah garis, salah satunya adalah metode *Direction*. Metode ekstraksi fitur *direction* menghasilkan nilai label arah pada piksel *foreground* citra [6]. Penerapan metode *direction* digunakan untuk mengekstraksi fitur guratan garis pada aksara Bali, sehingga dapat mengenali aksara dengan tingkat kemiripan yang tinggi.

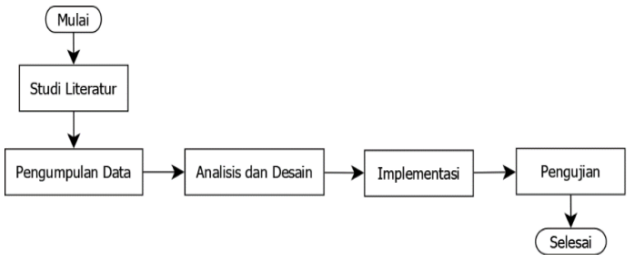
Berdasarkan pemaparan diatas, penulis mengajukan sebuah penelitian tentang pengenalan pola tulisan tangan aksara Bali, menggunakan metode ekstraksi fitur *Zoning* dengan variasi *algoritma Image Centroid and Zone* (ICZ), *Zone Centroid and Zone* (ZCZ), dan gabungan ICZ dan ZCZ dan penambahan metode ekstraksi fitur *direction* untuk mengekstraksi ciri karakter dengan tingkat kemiripan yang tinggi. Metode klasifikasi menggunakan *Backpropagation*. Penggunaan metode tersebut diharapkan dapat memberikan tingkat akurasi yang tinggi pada pengenalan pola karakter tulisan tangan aksara Bali.

2. Metode Penelitian

2.1. Desain Penelitian

Secara umum desain penelitian dapat diuraikan menjadi beberapa tahapan, sebagai berikut:
1. Studi Literatur: Mencari sumber-sumber penelitian yang terkait dengan penelitian.
2. Pengumpulan Data: Mengumpulkan data yang digunakan dalam sistem.
3. Analisis dan Desain: Menganalisa metode penelitian dan melakukan perancangan implementasi.
4. Implementasi: Mengimplementasikan metode yang dikaji ke dalam rancangan sistem.
5. Pengujian: Pengujian akurasi dari metode yang digunakan dalam sistem.

Alur penelitian pada pengenalan pola karakter tulisan tangan aksara Bali ditunjukkan pada gambar 1.



Gambar 1. Diagram Alir Alur Penelitian

2.2. Pengumpulan Data

Jenis data yang digunakan dalam penelitian ini adalah data primer. Terdapat dua tipe masukan data yang digunakan. Pertama adalah citra digital tulisan tangan huruf aksara Bali, meliputi aksara wianjana, gantungan dan gempelan aksara wianjana, pengangge suara, pengangge tengenan, dan angka Bali, data ini digunakan sebagai data latih dan pengujian untuk mengetahui akurasi setiap karakter terutama dengan kemiripan yang tinggi. Kedua adalah citra digital tulisan tangan kata aksara Bali, yang

p-ISSN: 2301-5373
e-ISSN: 2654-5101

digunakan dalam pengujian untuk mengetahui hasil pengenalan karakter yang sama pada struktur penulisan.

Target sebanyak 40 responden. Setiap responden memberikan data sampel sebanyak dua kali penulisan, sehingga total dataset adalah 4400 karakter tulisan tangan aksara Bali. Total dataset aksara didapatkan dengan mengalikan 80 data sampel dari responden dengan 55 kelas karakter yaitu 4400. Sedangkan total dataset kata yaitu 100. Jumlah dataset yang didapatkan disesuaikan dengan kebutuhan model *backpropagation* untuk data latih dan data uji.

2.3. Prepocessing

Tahap *preprocessing* berguna untuk mendapatkan piksel pembentuk citra dan mereduksi *noise*. Pertama citra digital dikonversi menjadi citra biner melalui proses *greyscaling* dan *thresholding* menggunakan metode *local adaptive thresholding*. Proses menghilangkan *noise* pada citra biner mengggunakan filter *mean*. Berikutnya, citra hasil filter melalui proses *thinning* menggunakan metode *zhang-zuen* untuk mendapatkan piksel pembentuk karakter dan mempercepat komputasi pengenalan pola karena jumlah piksel menjadi lebih kecil.

2.4. Segmentasi

Pada tahap segmentasi, memisahkan piksel *foreground* dengan background pada citra untuk memudahkan proses ekstraksi fitur. Proses segmentasi digunakan pada data kata aksara Bali, untuk mendapatkan karakter atau huruf pada kata tersebut. Metode segmentasi yang digunakan adalah proyeksi dan CCL. Metode proyeksi digunakan karena struktur dari aksara Bali memiliki pola penulisan vertikal dan horizontal. Untuk memisahkan beberapa karakter dalam satu dimensi yang sama, menggunakan metode CLL.

2.5. Ekstraksi Fitur

Pada penelitian ini menggunakan ekstraksi fitur *zoning* dan *direction*. Citra hasil segmentasi akan diekstraksi menggunakan kedua metode untuk mengetahui perbandingan akurasi dengan penambahan metode *direction*. Terdapat tiga algoritma *zoning* dengan mengambil ciri metrik dari objek yaitu ICZ, ZCZ, dan gabungan ICZ dan ZCZ. Hasil dari ekstaksi fitur adalah vektor baris atau kolom satu dimensi. Ciri yang dihasilkan dari metode *zoning* adalah nilai rata-rata perhitungan jarak piksel dengan *image centroid* atau *zone centroid* pada setiap zona. Sedangkan, ciri yang dihasilkan dari fitur *direction* adalah nilai arah piksel *foreground* setiap zona. Data fitur yang dihasilkan akan menjadi data latih untuk sistem dan sebagai data masukan untuk proses klasifikasi pada *backpropagation*.

2.6. Klasifikasi *Backpropagation*

Proses klasifikasi untuk mengenali karakter aksara Bali menggunakan metode *backpropagation.* Klasifikasi menggunakan *backpropagation* membutuhkan data masukan yaitu hasil ekstraksi fitur *zoning* dan *direction*. Data hasil ekstraksi fitur dibagi menjadi data latih dan data uji. Keluaran dari proses klasifikasi yaitu hasil pengenalan berupa nama dari 56 karakter aksara Bali. Tahap klasifikasi dibagi menjadi proses pelatihan dan pengujian.

3. Hasil dan Pembahasan

3.1. Pengumpulan Data

Dataset adalah karakter tulisan tangan aksara bali. Metode pengumpulan data menggunakan kuisioner yang diisi oleh 40 orang, sehingga dataset yang didapatkan berjumlah 4400 data karakter. Data kata aksara bali yang terkumpul berjumlah 100 kata, sehingga total dataset yang terkumpul berjumlah 4500 buah. Data kata didapatkan dari tulisan tangan satu orang pakar basa bali, dengan tujuan memvalidasi kebenaran aksara yang dibutuhkan oleh sistem. Pakar beratas nama Ni Kadek Witari, S.S. seorang guru Bahasa Bali SMP PGRI 8 Denpasar. Pakar memvalidasi kata yang diajukan oleh penulis sebagai kata yang memiliki kerakter dengan kemiripan yang tinggi. Kata yang membutuhkan validasi seperti sueca, dimana kata tersebut memiliki dua karakter yang mirip yaitu SA dan CA mengacu pada penelitian (Wibawa, 2019) tentang kemiripan akrasa bali.

3.2. Implementasi Sistem

Preprocessing berfungsi untuk mempersiapkan dataset berupa citra ke dalam bentuk yang lebih optimal untuk diolah atau pada tahap selanjutnya yaitu ekstraksi fitur. Pada penelitian ini proses *preprocessing* yang dilakukan yaitu *cropping, resize, grayscaling*, binerisasi, dan *thinning*. Citra diolah sehingga memiliki dimensi yang optimal dan pengikisan piksel atau *thinning* sehingga ekstraksi berfokus pada kerangka utama aksara.

3.3. Implementasi Segmentasi

Proses segmentasi dalam dua proses untuk memisahkan berdasarkan baris dan karakter kata. Hitung *contour* citra untuk menentukan garis batas yang kontinu, diimplementasikan pada baris kode 2-3. Inisialisasi variabel untuk menentukan *Region of Interes*t (ROI) dari citra, maka akan tercetak citra tersegmentasi berdasarkan dimensi baris, karakter dan batas piksel. Selanjutnya simpan roi sebagai gambar terpisah pada folder terpisah, diimplementasikan pada baris kode 4-5. Karakter diurutkan berdasarkan baris.

3.4. Implementasi Ekstraksi Fitur

Terdapat empat fitur yang diambil dari dataset. Fitur *centroid* dari pusat citra, *centroid* dari pusat zona, gabungan keduanya, dan fitur *direction*. Semua fitur akan dikombinasikan untuk mendapatkan sistem pengenalan pola paling optimal.

3.5. Pengujian Segmentasi

Pengujian ini bertujuan untuk mengetahui akurasi metode segmantasi yang digunakan. Berdasarkan hasil penelitian menggunakan 100 kata Aksara Bali, akurasi segmentasi sebesar 87.36 %. Hasil Pengujian Kata Aksara Bali. Hasil ini mengindikasikan bahwa medote segmentasi proyeksi dan *Connect Component Labeling* (CCL) sesuai diimplementasikan pada Aksara Bali. Berdasarkan penelitian, Aksara Bali dengan gantungan yang menyatu dengan aksaranya tidak dapat disegmentasi sehingga pada pengujian kata terbaca aksara lain oleh sistem.

3.6. Pengujian Optimasi Parameter *Neural Network*

3.6.1 Uji Pengaruh *Hidden Layer*

Pengujian ini bertujuan untuk mengetahui pengaruh jumlah hidden layer terhadap performa model *neural network* pada *hidden layer* satu sampai lima. Untuk parameter jumlah *neuron, learning rate,* dan *epoch* menggunakan nilai terendah dari setiap parameter. Hasil pengujian *hidden layer* ditunjukkan pada Tabel 1.

Tabel 1. Hasil Pengujian *Hidden Layer*

Jumlah *Hidden Layer*	Akurasi (%)
1	43.39
2	**45.16**
3	38.57
4	35.41
5	29.84

Berdasarkan hasil pengujian, jumlah *hidden layer* 2 memiliki performa yang lebih baik daripada *hidden layer* 1 atau *hidden layer* 3. Model dengan arsitektur *hidden layer* 2 dapat memberikan akurasi tertinggi yaitu sebesar 45.16 %. Oleh karena itu, *hidden layer* 2 dipilih untuk diujikan dengan parameter jumlah *neuron*.

3.6.2 Uji Pengaruh *Neuron Hidden Layer*

Pengujian ini bertujuan untuk mengetahui pengaruh jumlah *neuron hidden layer* terhadap performa model *neural network* pada neuron 32, 64, 128, 256, dan 512. Parameter lainnya yang digunakan yaitu

p-ISSN: 2301-5373
e-ISSN: 2654-5101

hidden layer 2, sedangkan untuk *learning rate* dan *epoch* menggunakan nilai terendah dari setiap parameter. Hasil pengujian *neuron hidden layer* ditunjukkan pada Tabel 2.

Tabel 2. Hasil Pengujian *Neuron Hidden Layer*

Neuron Hidden Layer	Akurasi (%)
32	32 45.16
64	46.09
128	51.16
256	58.18
512	**62.05**

Berdasarkan hasil pengujian, jumlah *neuron hidden layer* 512 memiliki performa yang lebih baik daripada *neuron* 32, 64, 128, dan 256. Model dengan arsitektur *neuron hidden layer* 512 dapat memberikan akurasi tertinggi yaitu sebesar 62.05 %. Oleh karena itu, *neuron hidden layer* 512 dipilih untuk diujikan dengan parameter *learning rate*. Berdasarkan hasil pengujian dapat disimpulkan bahwa semakin besar *neuron* maka semakin besar perolehan akurasi.

3.6.3 Uji Pengaruh *Learning Rate*

Learning rate merupakan salah satu parameter untuk menghitung nilai koreksi bobot pada proses pelatihan model. Pengujian ini bertujuan untuk mengetahui pengaruh nilai *learning rate* terhadap performa model *neural network*. Nilai *learning rate* yang digunakan yaitu 0.001, 0.002, 0.003, 0.004, dan 0.005. Parameter lainnya yang digunakan yaitu *hidden layer* 2, *neuron* 512, dan *epoch* menggunakan nilai terendah dari parameter. Hasil pengujian *learning rate* ditunjukkan pada Tabel 3.

Tabel 3. Hasil Pengujian *Learning Rate*

Learning Rate	Akurasi (%)
0.001	62.05
0.002	**62.25**
0.003	59.82
0.004	55.52
0.005	43.09

Berdasarkan hasil pengujian, *learning rate* 0.002 memiliki performa yang lebih baik daripada 0.001, 0.003, 0.004, dan 0.005. Model dengan arsitektur *learning rate* 0.002 dapat memberikan akurasi tertinggi yaitu sebesar 62.25 %. Oleh karena itu, *learning rate* 0.002 dipilih untuk diujikan dengan parameter *epoch*.

3.6.4 Uji Pengaruh *Epoch*

Pengujian ini bertujuan untuk mengetahui pengaruh jumlah *epoch* terhadap performa model *neural network*. Nilai *epoch* yang digunakan yaitu 100, 200, 400, 800, dan 1600. Parameter lainnya yang digunakan yaitu *hidden layer* 2, *neuron* 512, *learning rate* 0.002. Hasil pengujian *epoch* ditunjukkan pada Tabel 4.

Tabel 4. Hasil Pengujian *Epoch*

Epoch	Akurasi (%)
100	62.56
200	63.7
400	67.13
800	68.28
1600	**70.68**

Berdasarkan hasil pengujian, *epoch* 1600 memiliki performa yang lebih baik daripada 100, 200, 400, dan 800. Model dengan arsitektur *epoch* 1600 dapat memberikan akurasi tertinggi yaitu sebesar 70.68 %. Oleh karena itu, *epoch* 1600 dipilih untuk pengujian jenis fitur. Parameter yang terpilih yaitu jumlah *hidden layer* 2, *neuron* 512, *learning rate* 0.002, dan *epoch* 1600.

3.7. Klasifikasi Menggunakan Fitur *Zoning*

Pengujian metode *zoning* terdapat sembilan skenario yaitu kombinasi pengujian pada jenis dan zona. Variasi ekstraksi fitur *zoning* yang digunakan yaitu ICZ, ZCZ, dan gabungan ICZ dan ZCZ, sedangkan untuk pembagian zona yaitu 4x4, 8x8, dan 16x16. Fitur berekstensi *microsoft excel comma separated values (csv).* Hasil pengujian metode *zoning* dengan variasi jumlah zona ditunjukkan pada Tabel 5.

Tabel 5. Hasil Pengujian Metode *Zoning* dengan Jumlah Zona

Jumlah zona	ICZ (%)	ZCZ (%)	ICZ dan ZCZ (%)
4 x 4	70.68	71.85	75.94
8 x 8	80.39	85.76	86.97
16 x 16	**82.43**	**86.82**	**91.18**
Rata – rata	77.83	81.47	84.69

Berdasarkan hasil pengujian didapatkan kombinasi terbaik dengan akurasi pelatihan tertinggi adalah ICZ dan ZCZ zona 16x16 dengan akurasi mencapai 91.18 %. Diikuti oleh ZCZ zona 16x16 dengan akurasi 86.82 %. Berdasarkan akurasi di setiap variasi *zoning*, kelompok ICZ dan ZCZ memiliki rata-rata akurasi tertinggi yaitu 84.69 %. Kelompok ZCZ mendapatkan rata-rata akurasi sebesar 81.47 % dengan akurasi terbaik pada zona 16x16 yaitu sebesar 86.82 %. Kelompok ICZ mendapatkan rata-rata akurasi terkecil, sebesar 77.83 % dengan akurasi terbaik pada zona 16x16 yaitu 82.43 %.

3.8. Klasifikasi Menggunakan Fitur *Zoning* dan Fitur *Direction*

Pengujian kedua adalah kombinasi ekstraksi fitur *zoning* dan *direction*. Setiap fitur *zoning* yang terpilih ditambahkan fitur *direction*. Tujuan pengujian ini adalah menambahkan kemungkinan peningkatan akurasi fitur untuk pengujian model *neural network*. Hasil perbandingan dari pengujian penambahan metode *direction* ditunjukkan pada Tabel 6.

Tabel 6. Perbandingan Akurasi Pengenalan Pola Aksara Bali

Algoritma *Zoning*	Akurasi Metode *Zoning* (%)	Akurasi Metode *Zoning* dengan Penambahan *Direction* (%)	Peningkatan Akurasi (%)
ICZ	82.43	88.85	6.42
ZCZ	86.82	96.03	9.21
ICZ dan ZCZ	**91.18**	**97.61**	**6.43**

Berdasarkan hasil pelatihan *neural network* kombinasi fitur ICZ dan ZCZ zona 16x16 dan *direction* mendapatkan hasil akurasi tertinggi yaitu 91.18 %. Diikuti fitur ZCZ zona 16x16 dan *direction* mendapatkan akurasi sebesar 86.82 %. Akurasi terendah didapatkan oleh fitur ICZ zona 16x16 dan *direction* dengan akurasi sebesar 82.43 %. Fitur *direction* ditujukan untuk menambahkan keberagaman ciri dari citra, dengan tujuan untuk meningkatkan akurasi dari metode *zoning*. Akurasi fitur *direction* sendiri sebesar 82.8 %.

3.9. Identifikasi Kata Aksara Bali

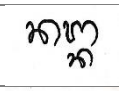
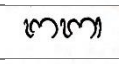
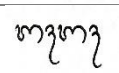
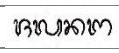
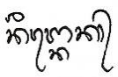
Pengujian keempat adalah pengujian terhadap data kata dimana model *train neural network* menggunakan kombinasi fitur ICZ dan ZCZ zona 16x16 dan *direction*. Kata yang diuji berjumlah 100 kata. Pertama import kata dan lakukan proses *preprocessing*, segmentasi, ekstraksi fitur, pengenalan pola, dan prediksi per karakter dengan keluaran label karakter.

Pengujian pada dataset kata Aksara Bali mencapai tingkat akurasi 66.2 %. Dan pada akurasi segmentasi mencapai 97.33 %. Keseluruhan hasil pengujian terlampir pada Lampiran 3 Akurasi

p-ISSN: 2301-5373
e-ISSN: 2654-5101

Pengujian Kata Aksara Bali. Hasil ini dipengaruhi karena terdapat persamaan bentuk karakter namun berbeda kelas data dan tidak didukung penggunaan *rule base* untuk membedakan kedua karakter dalam satu susunan kata tersebut. Contoh pada kasus ini adalah karakter aksara NA dengan gantungan KA yang memiliki karakter yang sama namun di kelas yang berbeda. Dengan menggunakan sistem yang dirancang hanya dapat membedakan karakter berdasarkan bentuk, tidak berdasarkan letak karakter pada susunan kata. Oleh karena itu, akurasi pada pengenalan karakter tersebut cenderung rendah. Hasil pengujian pada pengenalan kata ditunjukkan pada Tabel 7.

Tabel 7. Perbandingan Akurasi Pengenalan Pola Aksara Bali

No	Data Kata	Citra	Total Karakter	Karakter Segmentasi Benar	Jumlah Karakter Benar	Akurasi Segmentasi (%)	Akurasi (%)
1	Nangka		3	3	3	100	100
2	Ngaba		2	2	2	100	100
3	Ngahngah		4	4	4	100	100
4	Jalananga		4	4	4	100	100
5	Nengkek		7	7	7	100	100
		Rata – rata Akurasi				**87.36**	**66.2**

Faktor lainnya yang menentukan akurasi adalah hasil segmentasi karakter. Segmentasi memiliki kekurangan pada pemisahan karakter yang menyatu. Karakter tersebut dibaca menjadi satu karakter, sehingga kesulitan dalam mengidentifikasi kelas karakter. Contohnya pada kasus ini adalah karakter yang memiliki gantungan karena cenderung responden menuliskan gantungan menyatu dengan aksara dasarnya.

4. Kesimpulan

Pada penelitian ini mendapatkan hasil pengujian dari dataset karakter dan kata Aksara Bali menggunakan empat jenis fitur dengan tiga kombinasi fitur yaitu *zoning image centroid zone* (ICZ), *zone centroid zone* (ZCZ), gabungan keduanya (ICZ+ZCZ), dan *direction*, serta metode pembelajaran *neural network* dengan *backpropagation*. Berdasarkan hasil penelitian, didapatkan kesimpulan sebagai berikut.

1. Berdasarkan pengujian individual terhadap keempat jenis ekstraksi fitur, akurasi tertinggi didapatkan dari metode *zoning* ICZ dan ZCZ dengan zona 16x16 yaitu sebesar 91.18 %. Diikuti oleh *zoning* ZCZ zona 16x16 dengan akurasi 86.82 %, *zoning* ICZ zona 16x16 dengan akurasi 82.43 %, dan *direction* dengan akurasi 82.8 %.

2. Berdasarkan pengujian kombinasi akurasi *zoning* sebelumnya dengan penambahan *direction*, akurasi tertinggi didapatkan dari metode *zoning* ICZ dan ZCZ dengan akurasi mencapai 97.61 %. Diikuti oleh *zoning* ZCZ dengan akurasi 96.03 %, dan *zoning* ICZ dengan akurasi 88.85 %. Peningkatan akurasi tertinggi setelah penambahan *direction* terdapat pada fitur ZCZ dengan kenaikan sebesar 9.21 %. Hal tersebut membuktikan bahwa penambahan fitur *direction* dapat memberikan peningkatan performa model pengenalan pola khususnya pada data tulisan tangan.

3. Berdasarkan pengujian terhadap 100 kata mendapatkan rata-rata akurasi sebesar 66.2 %. Hasil ini disebabkan karena kesamaan bentuk karakter namun berbeda kelas dan kekurangan segmentasi pada pemisahan karakter yang menyatu.

Daftar Pustaka

[1] BW, T. A. Hermanto, I. G. R and D, R. N, "Pengenalan Huruf Bali Menggunakan Metode Modified Direction Feature (MDF) Dan Learning Vector Quantization (LVQ)", Konferensi Nasional Sistem dan Informatika 2009, 7–12, 2009.

[2] I. D. A. M. Sartini, M. W. A. Kesiman and I. G. M. Darmawiguna, “Pengembangan Text to Digital Image Converter untuk Dokumen Aksara Bali”, Jurnal Nasional Pendidikan Teknik Informatika (JANAPATI), 2(1), 2013.

[3] I. G. A. Wibawa, “Analisa Kesamaan Aksara Bali Menggunakan Template Matching”, Jurnal Elektronik Ilmu Komputer Udayana, 8(2), 2019.

[4] I. K. A. G. Wiguna and A. Muliantara, “Introduction of Balinese Script Handwriting Using Zoning and Multilayer Perceptron”, International Journal of Application Computer Science and Informatic Engineering (ACSIE), 1(1), 1–10, 2019.

[5] I. Mulia, “Pengenalan Akasara Sunda Menggunakan Ekstraksi Ciri Zoning dan Klasifikasi Support Vector Machine. Skripsi”, Skripsi. Departemen Ilmu Komputer Fakultas Matematika Dan Ilmu Pengetahuan Alam, Institut Pertanian Bogor, Bogor, 2012.

[6] R. Aristantya, I. Santoso and A. A. Zahra, ”Identifikasi Tanda Tangan Menggunakan Metode Zoning dan SVM (Support Vector Machine)”, Transient, 7(1), 174–178, 2018.

Jurnal Elektronik Ilmu Komputer Udayana
Volume 11, No 3. February 2023

p-ISSN: 2301-5373
e-ISSN: 2654-5101

Prediksi Hasil Panen Padi Di Kabupaten Jembrana Dengan Metode Linear Regression

I. B. M. Swarbawa[a1], I. G. Arta Wibawa[a2], I. K. G. Suhartana[a3]

[a]Informatics Engineering, Faculty of Math and Science, University of Udayana
South Kuta, Badung, Bali, Indonesia
[1]ibmswarbawa@gmail.com
[2]gede.arta@unud.ac.id
[3]ikg.suhartana@unud.ac.id

Abstract

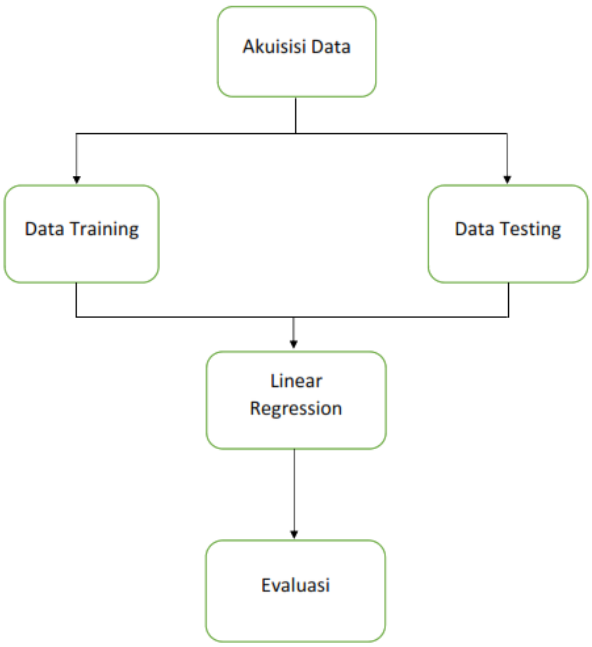
Prediction of agricultural yields is very much needed in terms of planning and decision making as well as in national food security policies. One of the strategic commodities that requires special attention is rice. This study aims to predict rice yields, especially in Jembrana Regency by using Linear Regression. Prediction is an important tool in effective and efficient planning. These results indicate that the rice yield prediction model can be used for forecasting purposes in the future. It is hoped that with this rice yield prediction system, it can help predict rice yields in the coming year and overcome the soaring increase in rice prices.

Keywords *: Yields, Prediction, Linear Regression*

1. Pendahuluan

Jembrana merupakan salah satu kabupaten penyangga pangan di provinsi Bali. Ada berbagai macam tanaman pangan yang dihasilkan di Kabupaten Jembrana setiap tahunnya salah satunya adalah padi. Padi merupakan salah satu tanaman pangan terpenting, karena padi menjadi sumber karbohidrat yang utama [1]. Kebutuhan bahan pangan terutama beras akan terus meningkat sejalan dengan pertambahan jumlah penduduk dan peningkatan konsumsi perkapita akibat peningkatan pendapatan. Hasil panen padi di kabupaten Jembrana mengalami naik turun. Sebagai upaya untuk mengatasi kenaikan harga beras yang melambung tinggi perlu adanya ramalan yang tepat tentang hasil panen padi di Kabupaten Jembrana yaitu sistem prediksi.

Prediksi merupakan proses memperkirakan sesuatu yang akan terjadi di masa depan secara sistematis berdasarkan informasi yang dimiliki di masa lalu dan sekarang [2]. Dalam penelitian ini akan digunakan metode Linear Regression untuk mendapatkan prediksi dari hasil panen padi di Kabupaten Jembrana. Algoritma linear regression adalah metode statistik yang digunakan untuk mengetahui pengaruh antara satu atau beberapa variabel terhadap satu buah variabel. Variabel merupakan besaran yang berubah-ubah nilainya. Variabel yang memengaruhi disebut dengan variabel bebas, variabel independen, atau variabel penjelas. Variabel yang dipengaruhi disebut dengan variabel terikat atau variabel dependen [3].



Gambar 1. Tahapan penelitian

Tahapan-tahapan yang dilakukan dalam penelitian adalah pengumpulan data, kemudian membagi data training dan data testing, kemudian mencari prediksi dengan regresi linier dan terakhir melakukan pengujian atau evaluasi.

2. Metode Penelitian

Dalam penelitian ini, tahap-tahap yang akan dilakukan adalah pengumpulan data, membagi data training dan data testing, penerapan algoritma linear regression, dan validasi hasil.

2.1. Pengumpulan Data

Dalam penelitian ini data yang digunakan adalah data yang berasal website Jembrana Satu Data Dari Desa atau disingkat JSDDD yang mana websitenya bernama https://survey.jembranakab.go.id/. Terdapat lima atribut yang digunakan dalam data yaitu nama kecamatan, luas tanam, luas panen, hasil produksi padi dan rata-rata hasil produksi padi.

2.2. Pre-Processing

Pre-processing merupakan langkah awal yang dilakukan dalam pengolahan data untuk membantu metode yang digunakan agar dapat berjalan dengan baik dan nantinya menghasilkan faktor kesalahan Root Mean Squared Error (RMSE) yang rendah. Pada tahap pre-processing, dilakukan proses cleaning data yang digunakan untuk menghilangkan informasi yang tidak diperlukan dalam proses pre-processing, yaitu menghilangkan atribut Kecamatan

2.3. Linear Regression

Metode Prediksi yang akan digunakan pada penelitian ini adalah Linear Regression. Linear Regression digunakan untuk membangun model untuk memprediksi hubungan antara dua variabel dengan menerapkan persamaan linier pada data yang diamati.

Metode utama untuk melakukan prediksi yakni membangun model regresi dengan mencari hubungan antara satu atau lebih variabel independen atau prediktor (X) dengan variabel dependen atau respons (Y). Linear regression memodelkan hubungan antara variabel skalar dan satu atau lebih variabel penjelas.

Metode linear regression tersusun atas dasar pola hubungan data yang relevan di masa lalu [4]. Algoritma linear regression dibagi menjadi dua jenis, yaitu simple linear regression dan multiple linear

regression. Simple linear regression merupakan hubungan antara satu variabel dependen dengan satu variabel independen, sedangkan multiple linear regression merupakan hubungan antara satu variabel dependen dengan dua atau lebih variabel independen.

Dalam penelitian ini digunakan multiple linear regression karena terdapat lebih dari satu variabel independen. Dataset penelitian terdiri dari satu variabel dependen (Y) dan tiga variabel independen (X). Variabel dependen tersebut adalah atribut Hasil Produksi Padi, sedangkan variabel independen adalah atribut Luas Tanam, Luas Panen, dan Rata-Rata Produksi Padi.

Perhitungan yang digunakan untuk multiple linear regression dinyatakan dalam persamaan 1

$$Y = a + a_1X_1 + b_2X_2 + ... + b_nX_n \quad (1)$$

dimana

Y = variabel dependen (nilai yang akan diprediksi)
$X_1, X_2,....X_n$ = variabel independen atau variabel bebas
a = konstanta
$b_1, b_2...b_n$ = koefisien regresi.
Dengan mengacu pada persamaan 1, untuk menghitung hasil panen tanaman padi dengan algoritme *linear regression*, digunakan persamaan 2

$$Y = a + b_1X_1 + b_2X_2 + b_3X_3 \quad (2)$$

dimana

Y = hasil produksi
a = konstanta
b_1 = koefisien variabel Luas Tanam
X_1 = Luas Tanam
b_2 = koefisien variabel Luas Panen
X_2 = Luas Panen
Langkah awal yang dilakukan adalah pembentukan model, dengan mencari nilai a, b_1, b_2,…, b_n menggunakan kuadrat terkecil dengan persamaan umum metode kuadrat terkecil

$$\begin{aligned} a_n + b1\sum X1 + b2\sum X2 &= \sum \quad Y \\ a\sum X1 + b1\sum (X1)^2 + b2\sum (X1X2) &= \sum X1Y \\ a\sum X2 + b1\sum (X1X2) + b2\sum (X2)^2 &= \sum X2Y \end{aligned} \quad (3)$$

Setelah hasil inversi diketahui, selanjutnya dilakukan perkalian matriks determinan dengan $\sum Y$, $\sum X_1Y$, $\sum X_2Y$. Untuk menghitung determinan matriks A, A_0, A_1, dan A_2, digunakan persamaan 4

$$\begin{bmatrix} \mathrm{N} & \sum \mathrm{X_1} & \sum \mathrm{X_2} \\ \sum \mathrm{X_1} & \sum \mathrm{X_1X_1} & \sum \mathrm{X_1X_2} \\ \sum \mathrm{X_2} & \sum \mathrm{X_2X_1} & \sum \mathrm{X_2X_2} \end{bmatrix} \mathrm{x} \begin{bmatrix} a \\ b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} \sum \mathrm{Y} \\ \sum X_1Y \\ \sum X_2Y \end{bmatrix} \quad (4)$$

Selanjutnya, mencari nilai a, b_1, dan b_2 dengan persamaan 5

$$\begin{aligned} \mathrm{a} &= \frac{Det\ \mathrm{A0}}{Det\ A} \\ \mathrm{b_1} &= \frac{Det\ \mathrm{A1}}{Det\ A} \\ \mathrm{b_2} &= \frac{Det\ \mathrm{A2}}{Det\ A} \end{aligned} \quad (5)$$

Selanjutnya, dilakukan perhitungan uji korelasi parsial untuk mengetahui tingkat keterkaitan masing-masing variabel independen terhadap variabel dependen. Perhitungan korelasi didapatkan dengan persamaan 6

$$\sum \mathrm{X_1Y} = \sum \mathrm{X_1Y} - \frac{(\sum X_1).(\sum Y)}{n} \quad (6)$$

$$\sum X_2Y = \sum X_2Y - \frac{(\sum X_2).(\sum Y)}{n}$$
$$\sum X_1X_2 = \sum X_1X_2 - \frac{(\sum X_1).(\sum X_2)}{n}$$

Selanjutnya, dilakukan perhitungan uji koefisien determinasi untuk mengetahui besar pengaruh antara variabel independen terhadap variabel dependen, sehingga diketahui kecocokan model linear regression. Koefisien determinasi berkisar antara 0 sampai 1. Jika R2 = 0, maka tidak ada pengaruh antara variabel independen dengan variabel dependen, dan jika R2 semakin mendekati nilai 1, maka semakin kuat pengaruh variabel independen terhadap variabel dependen. Perhitungan untuk mencari R^2 dapat dilihat pada persamaan 7

$$R^2 = 1 - \frac{ss_{error}}{ss_{total}} = 1 - \frac{y_i - y_i}{y_i - y} \quad (7)$$

dimana

y_i = observasi respons ke-i
y = rata-rata
y_i = ramalan respons ke-i
Koefisien determinasi dapat dihitung dengan persamaan 8

$$R = \frac{b_1\sum X_1Y + b_2\sum X_2Y}{\sum Y^2} \quad (8)$$

dimana

R = jumlah koefisien determinasi

2.4. Evaluasi

Evaluasi dilakukan untuk menguji dan melihat tingkat akurasi algoritme linear regression terhadap pemrosesan data. Dalam pengujian metode digunakan cross validation untuk memvalidasi keakuratan sebuah model yang dibangun berdasarkan dataset tertentu. Cross validation adalah cara untuk menemukan parameter terbaik dari suatu model dengan menguji besarnya error pada data uji [5].

K-Fold Cross Validation adalah metode yang digunakan untuk mengetahui rata-rata keberhasilan dari suatu sistem dengan melakukan perulangan dengan mengacak atribut masukan untuk menguji atribut input yang dimasukkan. K-Fold cross validation dilakukan dengan membagi data sejumlah n-fold.

Percobaan 1	Test	Train	Train	Train	Train
Percobaan 2	Train	Test	Train	Train	Train
Percobaan 3	Train	Train	Test	Train	Train
Percobaan 4	Train	Train	Train	Test	Train
Percobaan 5	Train	Train	Train	Train	Test

Gambar 2. K-Fold Cross Validation

Pada penelitian ini akan dilakukan perhitungan K-Fold hingga 5-fold. Pada gambar 2 diperlihatkan 5-fold cross validation dengan data testing ditampilkan berawarna kuning dan data latih ditampilkan berwarna putih.

Pengujian akurasi dilakukan dengan melihat perkembangan nilai RMSE. RMSE merupakan metode untuk mengevaluasi hasil teknik peramalan yang digunakan dengan mengukur tingkat akurasi dari hasil prakiraan suatu model [6].

$$RMSE = \sqrt{\frac{\sum_i^n (y_{i-} \widehat{y_i})^2}{n}} \quad (9)$$

p-ISSN: 2301-5373
e-ISSN: 2654-5101

dimana
y_i = data awal
y_i = data akhir
n = jumlah data

RMSE merupakan nilai rata-rata dari jumlah kuadrat kesalahan yang dapat menyatakan ukuran besarnya kesalahan yang dihasilkan oleh suatu model prakiraan. Nilai RMSE yang rendah menunjukkan bahwa variasi nilai yang dihasilkan oleh suatu model prakiraan mendekati variasi nilai aslinya [7]. Sedangkan apabila nilai RMSE semakin besar, maka keakuratan suatu nilai yang dihasilkan semakin sedikit atau tidak akurat.

3. Hasil dan Pembahasan

Terdapat 5 kecamatan di kabupaten jembrana. Tabel II menunjukan data hasil panen padi yang berisi atribut Kecamatan, Luas Tanam, Luas Panen, Hasil Produksi Padi dan Rata-Rata Hasil Produksi Padi. Pada tahap Pre-Processing dilakukan proses data cleaning. Proses data cleaning digunakan untuk menghilangkan atribut data yang tidak dibutuhkan dalam proses pre-processing, yaitu dengan menghilangkan atau membuang informasi atribut Kecamatan yang tidak diperlukan dalam pengolahan data

Tabel I
Dataset Padi

Kecamatan	Luas Tanam	Luas Panen	Hasil Produksi Padi
Pekutatan	7	6	3
Mendoyo	40	40	19
Jembrana	11	12	7
Negara	40	32	17
Melaya	20	18	9

Tabel II
Data Cleaning

Luas Tanam	Luas Panen	Hasil Produksi Padi
7	6	3
40	40	19
11	12	7
40	32	17
20	18	9

Hasil perhitungan algoritma linear regression me nyatakan hubungan antara Luas Tanam, Luas Panen dan Hasil Produksi Padi.

$$Y = a + b_1(\text{Luas Tanam}) + b_2(\text{Luas Panen}) \quad (10)$$

Untuk mendapatkan persamaan linear regression berganda dilakukan model *multiple linear regression.*

3.1. Multiple Linear Regression

Nilai a, b1, dan b2 merupakan nilai konstanta dan koefisien regresi yang dapat diperoleh dengan menggunakan perhitungan matriks determinan. Pada penelitian ini, terdapat persamaan variabel yang tidak diketahui nilainya, yaitu a, b1, dan b2. Persamaannya adalah

$$A = \begin{bmatrix} N & \sum X_1 & \sum X_2 \\ \sum X_1 & \sum X_1 X_1 & \sum X_1 X_2 \\ \sum X_2 & \sum X_2 X_1 & \sum X_2 X_2 \end{bmatrix} \quad (11)$$

A merupakan persamaan matriks yang digunakan untuk mencari nilai Det(A), dengan N adalah banyaknya data, $\sum X_1$ adalah jumlah dari variabel Luas Tanam dan $\sum X_2$ adalah jumlah dari variabel Luas Panen.

$$A_0 = \begin{bmatrix} \sum Y & \sum X_1 & \sum X_2 \\ \sum X_1 Y & \sum X_1 X_1 & \sum X_1 X_2 \\ \sum X_2 Y & \sum X_2 X_1 & \sum X_2 X_2 \end{bmatrix} \quad (12)$$

A_0 merupakan persamaan matriks yang digunakan untuk mencari nilai Det(A_0), dengan $\sum Y$ adalah jumlah dari variabel Hasil Produksi, $\sum X_1$ adalah jumlah dari variabel Luas Tanam, $\sum X_2$ adalah jumlah dari variabel Luas Panen

$$A_1 = \begin{bmatrix} N & \sum Y & \sum X_2 \\ \sum X_1 & \sum X_1 Y & \sum X_1 X_2 \\ \sum X_2 & \sum X_2 Y & \sum X_2 X_2 \end{bmatrix} \quad (13)$$

A_1 merupakan persamaan matriks yang digunakan untuk mencari nilai Det(A_1), dengan N adalah banyaknya data, $\sum Y$ adalah jumlah dari variabel Hasil Produksi, $\sum X_2$ adalah jumlah dari variabel Luas Panen.

$$A_2 = \begin{bmatrix} N & \sum X_1 & \sum Y \\ \sum X_1 & \sum X_1 X_1 & \sum X_1 Y \\ \sum X_2 & \sum X_2 X_1 & \sum X_2 Y \end{bmatrix} \quad (14)$$

A_2 merupakan persamaan matriks yang digunakan untuk mencari nilai Det(A_2), dengan N adalah banyaknya data, $\sum X_1$ adalah jumlah dari variabel Luas Tanam, $\sum Y$ adalah jumlah dari variabel Hasil Produksi. Langkah selanjutnya adalah menentukan determinasi matriks A, A_0, A_1, A_2

$$\text{Det (A)} = \{N.\sum X_1 X_1.\sum X_2 X_2\} + \{\sum X_1.\sum X_1 X_2.\sum X_2\} + \{\sum X_2.\sum X_1.\sum X_2 X_1\} - \{\sum X_2.\sum X_1 X_1.\sum X_2\} - \{N.\sum X_1 X_2.\sum X_2 X_1\} - \{\sum X_1.\sum X_1.\sum X_2 X_2\}$$

$$\text{Det } (A_0) = \{\sum Y\ .\sum X_1 X_1.\sum X_2 X_2\} + \{\sum X_1.\sum X_1 X_2.\sum X_2 Y\} + \{\sum X_2.\sum X_1 Y\ .\sum X_2 X_1\} - \{\sum X_2.\sum X_1 X_1.\sum X_2 Y\ \} - \{\sum Y\ .\sum X_1 X_2.\sum X_2 X_1\} - \{\sum X_1.\sum X_1 Y.\sum X_2 X_2\}$$

$$\text{Det } (A_1) = \{N.\sum X_1 Y.\sum X_2 X_2\} + \{\sum Y.\sum X_1 X_2.\sum X_2\} + \{\sum X_2.\sum X_1.\sum X_2 Y\} - \{\sum X_2.\sum X_1 Y.\sum X_2\} - \{N.\sum X_1 X_2.\sum X_2 Y\} - \{\sum Y.\sum X_1.\sum X_2 X_2\}$$

$$\text{Det } (A_2) = \{N.\sum X_1 X_1.\sum X_2 Y\} + \{\sum X_1.\sum X_1 Y.\sum X_2\} + \{\sum Y.\sum X_1.\sum X_2 X_1\} - \{\sum Y.\sum X_1 X_1.\sum X_2\} - \{N.\sum X_1 Y.\sum X_2 X_1\} - \{\sum X_1.\sum X_1.\sum X_2 Y\}$$

Berdasarkan perhitungan determinasi matriks, didapatlah hasil nilai a, b1, dan b2, seperti ditunjukkan pada Tabel III.

Tabel III.
Hasil Perhitungan Konstanta dan Koefisien

a	b_1	b_2
-0,79822	0,14939	0,31532

Berdasarkan hasil yang diperoleh untuk koefisien a, b1, dan b2 tersebut, hasil model multiple linear regression menjadi sebagai berikut.

$\sum Y = -0{,}79822 + 0{,}14939\sum X_1 + 0{,}31532\sum X_2$

p-ISSN: 2301-5373
e-ISSN: 2654-5101

Tabel IV
Data aktual dan hasil prediksi

Data Aktual	Hasil Prediksi
3	2,1
19	18
7	5
17	16
9	8

$$RMSE = \sqrt{\frac{\sum_i^n (y_{i-} \widehat{y_i})^2}{n}}$$
$$= \sqrt{\frac{0{,}9^2 + 1^2 + 2^2 + 1^2 + 1^2}{5}}$$
$$= 1{,}249$$

3.2. Uji Korelasi

Perhitungan uji korelasi parsial digunakan untuk mengetahui besar keterkaitan masing-masing variabel independen terhadap variabel dependen.

Perhitungan rX_1Y:

$$\sum X_1Y = \sum X_1Y - \frac{(\sum X_1).(\sum Y)}{n}$$
$$= 1718 - \frac{(118).(55)}{5}$$
$$= 420$$

Perhitungan rX_2Y:

$$\sum X_2Y = \sum X_2Y - \frac{(\sum X_2).(\sum Y)}{n}$$
$$= 1568 - \frac{(108).(55)}{5}$$
$$= 380$$

Perhitungan rX_1X_2:

$$\sum X_1X_2 = \sum X_1X_2 - \frac{(\sum X_1).(\sum X_2)}{n}$$
$$= 3414 - \frac{(118).(108)}{5}$$
$$= 865{,}2$$

Tabel V
Hasil Uji Korelasi

Keterkaitan Variabel	Nilai Korelasi
rX_1Y	420
rX_2Y	380
rX_1X_2	865,2

3.3. Uji Koefisien Determinasi

Perhitungan uji koefisien determinasi digunakan untuk mengetahui besar pengaruh antara variabel bebas terhadap variabel tak bebas, sehingga dapat diketahui kecocokan model Linear Regression.

$$R = \frac{b_1 \sum X_1 Y + b_2 \sum X_2 Y}{\sum Y^2}$$
$$= \frac{0{,}149 \,.1718 + 0{,}315 \,.1568}{789}$$
$$= 0{,}950446$$

Hasil koefisien determinasi sebesar 0,950446. Artinya tingkat kecocokan model multiple linear regression memiliki tingkat keandalan sebesar 95,04%. Sebanyak 95,04% variasi nilai Hasil Produksi

bergantung pada variabel independen yang diukur, yaitu Luas Tanam dan Luas Panen Sedangkan sisanya, yaitu sebesar 4,96%, dipengaruhi oleh variabel lain yang tidak diukur pada penelitian ini.

4. Kesimpulan

Berdasarkan analisis data dan pembahasan yang telah dipaparkan, dapat ditarik kesimpulan bahwa algoritma linear regression dapat diterapkan untuk prediksi hasil panen padi dengan beberapa variabel yang memengaruhi yaitu Luas Tanam, Luas Panen, dan Hasil Produksi Padi

Dari jumlah data sebanyak 5 Kecamatan, dihasilkan tingkat kecocokan model multiple linear regression sebesar 95,04%. Artinya sebanyak 95,04% variasi nilai hasil panen bergantung pada variabel independen yang diukur, yaitu Luas Tanam dan Luas Panen. Sedangkan sisanya sebesar 4,96%, dipengaruhi oleh variabel lain yang tidak diukur pada penelitian ini.

Referensi

[1] C.V. Donggulo, I.M. Lapanjang, dan U. Made, "Pertumbuhan dan Hasil Tanaman Padi (Oryza sativa L) pada Berbagai Pola Jajar Legowo dan Jarak Tanam," *J. Agrol,* vol. Vol. 24, p. hal. 27–35, 2017.

[2] Z.A. Matondang, "Sistem Pendukung Keputusan Forecasting Harga Emas Lelang pada Pegadaian dengan Metode Single Moving Average," *J. Tek. Inform. Unika St. Thomas,* Vols. Vol. 3, No. 1, p. hal. 72–77, 2018.

[3] F. Anis dan Suprayogi, "Estimasi Luas Panen Padi di Kabupaten Rembang Menggunakan Algoritma Linear Regression," *Skripsi, Universitas Dian Nuswantoro, Semarang, Indonesia,* 2015.

[4] M.F. Saputri dan S. Slamet, "Analisa Data Penjualan Menggunakan Metode Regresi Linier untuk Prediksi Persediaan Barang pada TB.Kawankita," *Skripsi, Universitas Dian Nuswantoro, Semarang, Indonesia,* 2016.

[5] A. Fikri, "Penerapan Data Mining untuk Mengetahui Tingkat Kekuatan Beton yang Dihasilkan dengan Metode Estimasi Menggunakan Linear Regression," *Skripsi, Universitas Dian Nuswantoro, Semarang, Indonesia,* 2009.

[6] T. Chai dan R.R. Draxler, "Root Mean Square Error (RMSE) or Mean Absolute Error (MAE)? - Arguments Against Avoiding RMSE in the Literature," *Geosci. Model Dev,* Vols. Vol. 7, No. 3, p. hal. 1247–1250, 2014.

[7] P. Choirunisa dan Kariyam, "Perbandingan Metode Triple Exponential Smoothing dan Metode Seasonal ARIMA untuk Peramalan Inflasi di Kota Tamjung Pandan," *Prosiding Sendika,* Vols. Vol. 5, No. 2, p. hal. 76–83, 2019.

Jurnal Elektronik Ilmu Komputer Udayana
Volume 11, No 3. February 2023

p-ISSN: 2301-5373
e-ISSN: 2654-5101

Classification of Bird Sounds Using the Mel-Frequency Cepstral Coefficient (MFCC) and K-Nearest Neighbor (KNN) Methods

Anak Agung Gde Ramananda Kartikeya Pattraksha[a1], I Wayan Supriana[a2], I Made Widiartha[a3], Luh Arida Ayu Rahning Putri[a4], Ida Bagus Gede Dwidasmara[a5], I Dewa Made Bayu Atmaja Darmawan[a6]

[a]Informatics Departement, Udayana University
Jimbaran, Bali, Indonesia
[1]kartikeyaramananda@email.com
[2]wayan.supriana@unud.ac.id
[3]madewidiartha@unud.ac.id
[4]rahningputri@unud.ac.id
[5]dwidasmara@unud.ac.id
[6]dewabayu@unud.ac.id

Abstract

With the rapid development of this technology, of course, it can provide benefits to the wider community, one of which is to provide convenience in meeting the needs of an agency or group, one of which is forest rangers and animal supervisors who can assist in classifying the types of birds or poultry, because birds have voices. which differ by type. For example, in the West Bali National Park itself, there are several types of bird species that are there, for example, the Java Dederuk bird or birds belonging to the Columbidae species. In this case technology plays an important role as a medium for information exchange regarding information about birds and can easily determine the type of bird. So, to fulfill this, we need a system that can classify bird species based on their sound. In this study, the system built uses the MFCC method for feature extraction. This method was chosen because the success rate in speech recognition using MFCC feature extraction is higher than feature extraction using FFT. And for the classification stage using the KNN method. This method was chosen because KNN is easy to represent compared to other methods. From the results of the Classification of Bird Sounds Using the Mel-Frequency Cepstral Coefficient (MFCC) and K-Nearest Neighbor (KNN) methods, it can be concluded that the resulting K-Nearest Neighbor model is able to classify bird species based on their sound properly because after a single data classification with data outside of the training data, the system is able to classify it very accurately, the accuracy percentage is 80%. And in the test results using K-fold Cross Validation, the best K value is obtained in the K-Nearest Neighbor model, namely K = 1 with 77% accuracy.

Keywords: *Mel Frequency Cepstral Coefficient, K-Nearest Neighbor, Birds*

1. Pendahuluan

Dengan pesatnya perkembangan teknologi ini, tentunya dapat memberikan manfaat bagi masyarakat luas, salah satunya mempermudah pemenuhan kebutuhan suatu instansi atau kelompok, salah satunya jagawana dan pengawas satwa yang dapat membantu dalam mengklasifikasikan jenis burung atau unggas, karena burung memiliki suara. yang berbeda menurut jenisnya. Misalnya saja di Taman Nasional Bali Barat sendiri terdapat beberapa jenis jenis burung yang ada disana, misalnya burung Dederuk Jawa atau burung yang termasuk dalam jenis Columbidae. Dalam hal ini, teknologi berperan penting sebagai sarana pertukaran informasi terkait informasi burung dan dapat dengan mudah menentukan jenis burung tersebut. Untuk mencapai hal tersebut diperlukan suatu sistem yang dapat mengklasifikasikan jenis burung berdasarkan suaranya.

Oleh karena itu pada penelitian ini sistem yang dibangun menggunakan metode MFCC untuk ekstraksi fitur. Metode ini dipilih karena tingkat keberhasilan pengenalan suara menggunakan ekstraksi fitur MFCC lebih tinggi dibandingkan dengan ekstraksi fitur menggunakan FFT [1]. Dan untuk tahap klasifikasi menggunakan metode KNN. Metode ini dipilih karena KNN mudah direpresentasikan dibandingkan dengan metode lainnya [2].

2. Metode Penelitian

2.1. Data dan Metode Pengumpulan Data

Data untuk penelitian ini diambil dari burung-burung di Bali. Jumlah data yang digunakan dalam penelitian ini adalah 50 data suara burung, dimana terdapat 5 kelas dan untuk setiap kelas terdapat 10 data suara. Famili burung yang dipilih untuk klasifikasi ini adalah perkutut jawa, gagak hutan, cerek besar, dan dara laut batu. Data suara burung diperoleh dari database burung dunia avibase.bsc-eoc.org. Format data yang digunakan adalah data audio dalam format .wav. Jenis burung yang akan digunakan dalam penelitian ini dapat dilihat pada Tabel 1.

Tabel 1. Bird Data

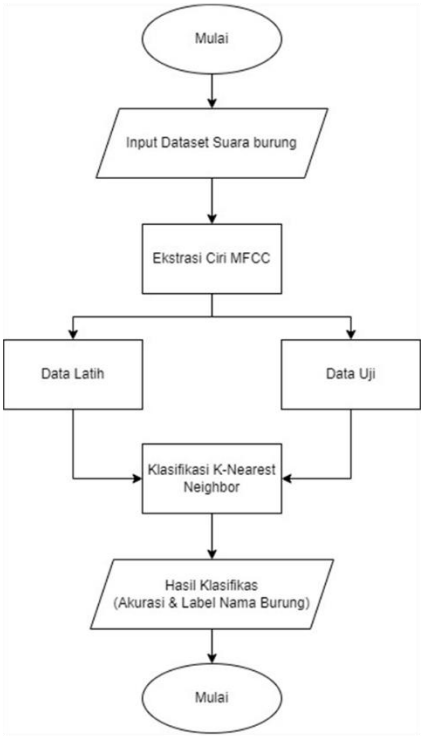
Nama ilmiah	Nama Burung	Famili	Label (.wav)
Geopelia striata	perkutut jawa	columbidae	748014
Geopelia striata	perkutut jawa	columbidae	738793
Geopelia striata	perkutut jawa	columbidae	618483
Geopelia striata	perkutut jawa	columbidae	616403
Geopelia striata	perkutut jawa	columbidae	598171
Geopelia striata	perkutut jawa	columbidae	588080
Geopelia striata	perkutut jawa	columbidae	576417
Geopelia striata	perkutut jawa	columbidae	764718
Geopelia striata	perkutut jawa	columbidae	756428
Geopelia striata	perkutut jawa	columbidae	713711
Corvus enca	gagak hutan	corvidae	657295
Corvus enca	gagak hutan	corvidae	628125
Corvus enca	gagak hutan	corvidae	616145
Corvus enca	gagak hutan	corvidae	616142
Corvus enca	gagak hutan	corvidae	616141
Corvus enca	gagak hutan	corvidae	614886
Corvus enca	gagak hutan	corvidae	614885
Corvus enca	gagak hutan	corvidae	614882
Corvus enca	gagak hutan	corvidae	614881
Corvus enca	gagak hutan	corvidae	614880
Pluvialis squatarola	cerek besar	charadriidae	775072
Pluvialis squatarola	cerek besar	charadriidae	775071
Pluvialis squatarola	cerek besar	charadriidae	775069
Pluvialis squatarola	cerek besar	charadriidae	775068
Pluvialis squatarola	cerek besar	charadriidae	774421
Pluvialis squatarola	cerek besar	charadriidae	774420
Pluvialis squatarola	cerek besar	charadriidae	774419
Pluvialis squatarola	cerek besar	charadriidae	774418
Pluvialis squatarola	cerek besar	charadriidae	771426

p-ISSN: 2301-5373
e-ISSN: 2654-5101

Pluvialis squatarola	cerek besar	charadriidae	770143
Onychoprion anaethetus	dara laut batu	Laridae	612658
Onychoprion anaethetus	dara laut batu	Laridae	589066
Onychoprion anaethetus	dara laut batu	Laridae	740587
Onychoprion anaethetus	dara laut batu	Laridae	735967
Onychoprion anaethetus	dara laut batu	Laridae	589062
Onychoprion anaethetus	dara laut batu	Laridae	735240
Onychoprion anaethetus	dara laut batu	Laridae	612657
Onychoprion anaethetus	dara laut batu	Laridae	589067
Onychoprion anaethetus	dara laut batu	Laridae	539164
Onychoprion anaethetus	dara laut batu	Laridae	539136
Haliastur indus	Elang Bondol	Accipitridae	743675
Haliastur indus	Elang Bondol	Accipitridae	736313
Haliastur indus	Elang Bondol	Accipitridae	702713
Haliastur indus	Elang Bondol	Accipitridae	685003
Haliastur indus	Elang Bondol	Accipitridae	663577
Haliastur indus	Elang Bondol	Accipitridae	663576
Haliastur indus	Elang Bondol	Accipitridae	663575
Haliastur indus	Elang Bondol	Accipitridae	663573
Haliastur indus	Elang Bondol	Accipitridae	644542
Haliastur indus	Elang Bondol	Accipitridae	578793

2.2. Desain Penelitian

Jalannya penelitian tentang klasifikasi suara dengan ekstraksi fitur MFCC menggunakan KNN diamati dengan flowchart. Flowchart digunakan untuk memudahkan penggambaran aliran informasi dan aliran penelitian. Alur desain penelitian ini dapat dilihat pada Gambar 1.



Gambar 1. Flowchart Research Design

Pada tahap klasifikasi dengan K-Nearest Neighbor, data dengan fitur terpilih akan dilatih terlebih dahulu. Kemudian data yang telah dilatih akan digunakan untuk memprediksi data uji masukan. Kemudian hasil prediksi ditampilkan dengan menggunakan metode K-Nearest Neighbor.

Setelah melakukan prediksi terhadap data uji, selanjutnya dilakukan pengujian terhadap program. Untuk pengujiannya sendiri digunakan metode K-Fold Cross Validation. Setelah itu akan muncul berapa akurasi yang dihasilkan oleh program tersebut. Kemudian hasil perhitungan KNN akan mendeteksi apakah suara burung termasuk kelas perkutut jawa, gagak hutan, cerek besar, dan dara laut batu.

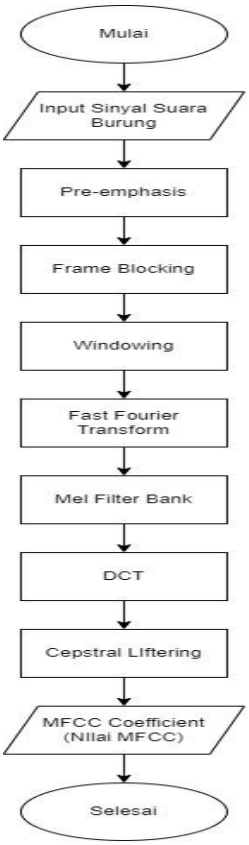
2.3. Pre-Processing

Sebelum dilakukan ekstraksi fitur menggunakan Mel Frequency Cepstral Coefficient, file suara burung dalam format .wav akan diproses terlebih dahulu yaitu pembacaan data menggunakan tools SciPy dimana file suara berupa sinyal analog akan diubah menjadi bentuk digital.

2.4. Feature Ekstraction Using Mel Frequency Cepstral Coefficient

Setelah melalui proses preprocessing, langkah selanjutnya adalah ekstraksi fitur menggunakan MFCC. Ekstraksi fitur MFCC merupakan adaptasi terhadap sistem pendengaran manusia, di mana sinyal audio difilter secara linear untuk frekuensi di bawah 1000 Hz dan secara logis untuk frekuensi di bawah 1000 Hz [3]. Tahapan MFCC dapat dijelaskan pada Gambar 2.

p-ISSN: 2301-5373
e-ISSN: 2654-5101



Gambar 2. Flowchart MFCC

Penjelasan tahapan ekstraksi ciri dengan MFCC pada penelitian ini adalah sebagai berikut:

a. Pre-Emphasis

Pre-Emphasis adalah langkah pertama dalam ekstraksi fitur MFCC. Proses ini akan mempertahankan frekuensi tinggi dalam spektrum yang biasanya hilang atau dihilangkan dalam proses produksi suara. Pre-ephasis bertujuan untuk mengurangi rasio noise terhadap sinyal sehingga dapat meningkatkan kualitas suatu sinyal dan menyeimbangkan spektrum suara yang disuarakan [3].

b. Frame Blocking

Frame Blocking adalah tahap dimana sinyal suara disegmentasi menjadi beberapa frame. Secara umum, dalam proses pemblokiran frame, setiap frame berdurasi 20-25 milidetik dengan porsi tumpang tindih (M) antara satu frame dengan frame lainnya sebesar 30-50% dari panjang frame. Panjang frame (N) yang digunakan dalam proses ini sangat berpengaruh. untuk sukses dalam analisis spektral [3]. Semakin panjang ukuran frame yang digunakan, semakin banyak resolusi frekuensi yang akan dituju, namun hal ini akan mempengaruhi resolusi waktu yang dihasilkan.

c. Windowing

Pada proses Windowing dilakukan tahap pembobotan data untuk setiap frame yang dibentuk pada proses sebelumnya dengan menggunakan fungsi window [4]. Windowing ini dilakukan untuk mengurangi gap setelah proses frame blocking.

d. Fast Fourier Transform

Fast Fourier Transform (FFT) merupakan pengembangan dari Discrete Fourier Transform (DFT) yang bertujuan untuk mengubah sinyal dari domain waktu ke domain frekuensi [5]. Pada langkah FFT ini, masing-masing N sampel dikonversi dari domain waktu ke domain frekuensi.

Setelah mendapatkan nilai FFT, nilai ini digunakan untuk menghitung densitas spektral energi. Densitas spektral energi ini akan digunakan untuk memetakan nilai FFT ke bank filter yang akan digunakan pada tahap selanjutnya.

e. Filter Bank

Selanjutnya dilakukan perhitungan bank filter terhadap sinyal data yang diperoleh dari proses FFT sebelumnya. Untuk menyederhanakan nilai ini diubah menjadi satuan dB, sehingga akan dihasilkan nilai bank filter signal.

f. Discrete Cosinus Transform

Pada tahap ini, nilai spektrum mel pada domain frekuensi akan diubah menjadi domain waktu untuk mendapatkan nilai koefisien.

g. Cepstral Liftering

Pada proses Cepstral lifting ini merupakan teknik yang digunakan untuk memperkecil sensitivitas koefisien cepstral yang dihasilkan dari tahapan utama dalam ekstraksi fitur menggunakan MFCC. Proses ini berfungsi untuk meningkatkan kualitas pengenalan suara.

Setelah melakukan perhitungan yang sama untuk setiap nilai koefisien cepstral, maka akan didapatkan nilai koefisien cepstral terfilter. Nilai ini nantinya akan digunakan sebagai dataset pada tahap klasifikasi.

2.5. Pembagian Data

Proses selanjutnya adalah memisahkan dataset menjadi data latih dan data uji. Data pelatihan adalah data yang diperiksa model, sedangkan data uji adalah data yang digunakan untuk mengetahui seberapa baik kinerja model pada data yang tidak terlihat. Scikit-learn memiliki fungsi yang disebut "train_test_split" yang memudahkan pemisahan dataset menjadi data pelatihan dan data pengujian.

2.6. Klasifikasi Menggunakan K-Nearest Neighbor Method

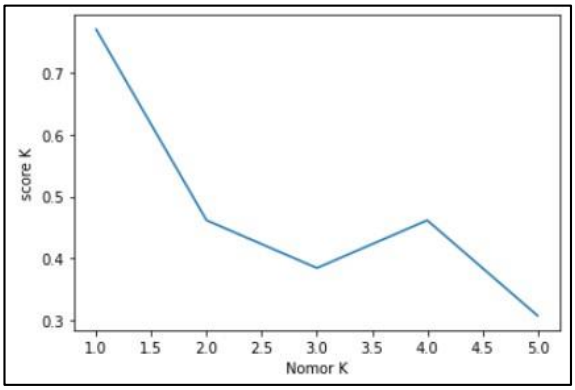
Metode K-nearest neighbor menggunakan semua data uji pada data, setelah itu sistem menentukan data latih dan uji. Setelah Anda menentukan data pelatihan dan pengujian, Anda dapat menemukan jarak minimum dan maksimum dalam perhitungan data pelatihan. Kemudian cari jarak minimum dari data uji ke data pelatihan. Setelah diperoleh jarak antara data latih dan data uji, ditentukan kelas sesuai dengan jarak minimum interval yang telah ditentukan. Setelah itu, hasil klasifikasi ditampilkan dalam bentuk akurasi. Alur metode KNN dapat dilihat pada Gambar 3.

Gambar 3. Flowchart KNN

p-ISSN: 2301-5373
e-ISSN: 2654-5101

3. Hasil dan Pembahasan

Pada Gambar 4 merupakan grafik yang diperoleh dari nilai K yang telah diuji dengan validasi K-Fold cross. Dari hasil pengujian dapat disimpulkan bahwa nilai K berpengaruh terhadap akurasi sistem. Dari hasil pengujian diperoleh nilai K terbaik yaitu K = 1 dengan akurasi 77%.



Gambar 4. The results of the K-Nearest Neighbor test with K-Fold Cross Validation

Setelah mendapatkan nilai K terbaik yaitu K = 1, dilakukan laporan Klasifikasi dan didapatkan hasil precision, recall dan f1-score. Dapat dilihat pada Gambar 5.

	precision	recall	f1-score	support
cerek besar	1.00	1.00	1.00	2
dara laut batu	0.67	0.67	0.67	3
elang bondol	0.67	1.00	0.80	2
gagak hutan	0.75	1.00	0.86	3
perkutut jawa	1.00	0.33	0.50	3
accuracy			0.77	13
macro avg	0.82	0.80	0.76	13
weighted avg	0.81	0.77	0.74	13

Gambar 5. Classification Report

3.1. Pengujian Sistem dengan Inputan Audio Baru Burung

Pada tahap ini dilakukan pengujian terhadap masukan data audio burung yang baru. Data baru yang akan diuji adalah 5 kelas dengan masing-masing kelas berisi 5 data audio burung baru. Setelah dilakukan pengujian terhadap 25 data audio tunggal untuk burung baru, diperoleh hasil klasifikasi seperti pada Tabel 2.

Tabel 2. Classification Results New bird audio input

Nama ilmiah	Famili	Nama File (.wav)	Nama Burung	Hasil Klasifikasi (Nama Burung)
Geopelia striata	columbidae	197147	perkutut jawa	gagak hutan
Geopelia striata	columbidae	665873	perkutut jawa	dara laut batu
Geopelia striata	columbidae	769104	perkutut jawa	perkutut jawa
Geopelia striata	columbidae	767816	perkutut jawa	perkutut jawa
Geopelia striata	columbidae	576417	perkutut jawa	gagak hutan
Corvus enca	corvidae	614879	gagak hutan	gagak hutan
Corvus enca	corvidae	614878	gagak hutan	gagak hutan
Corvus enca	corvidae	105944	gagak hutan	gagak hutan
Corvus enca	corvidae	769106	gagak hutan	gagak hutan
Corvus enca	corvidae	614858	gagak hutan	gagak hutan
Pluvialis squatarola	charadriidae	593996	cerek besar	cerek besar

Pluvialis squatarola	charadriidae	770143	cerek besar	cerek besar
Pluvialis squatarola	charadriidae	774418	cerek besar	cerek besar
Pluvialis squatarola	charadriidae	774420	cerek besar	cerek besar
Pluvialis squatarola	charadriidae	775068	cerek besar	cerek besar
Onychoprion anaethetus	Laridae	769107	dara laut batu	dara laut batu
Onychoprion anaethetus	Laridae	767680	dara laut batu	dara laut batu
Onychoprion anaethetus	Laridae	759450	dara laut batu	elang bondol
Onychoprion anaethetus	Laridae	37739	dara laut batu	dara laut batu
Onychoprion anaethetus	Laridae	251964	dara laut batu	elang bondol
Haliastur indus	Accipitridae	743675	elang bondol	elang bondol
Haliastur indus	Accipitridae	736313	elang bondol	elang bondol
Haliastur indus	Accipitridae	702713	elang bondol	elang bondol
Haliastur indus	Accipitridae	685003	elang bondol	elang bondol
Haliastur indus	Accipitridae	663577	elang bondol	elang bondol

Berdasarkan Tabel 2. Apakah dapat diperoleh hasil pada pengujian data tunggal dimana pada data perkutut Jawa terdapat 2 data yang berhasil diklasifikasikan dengan benar dari 5 data perkutut Jawa. Pada data gagak hutan terdapat 5 data yang berhasil diklasifikasikan dengan benar dari 5 data gagak hutan. pada data burung besar terdapat 5 data yang sudah terklasifikasi dengan benar dari 5 data burung besar. pada data merpati karang terdapat 3 data yang berhasil diklasifikasikan dengan benar dari 5 data merpati karang. pada data elang botak terdapat 5 data yang berhasil diklasifikasikan dengan benar dari 5 data elang botak. aDengan demikian diperoleh hasil dari total 20 data burung yang diuji, 20 data burung berhasil diklasifikasikan dengan benar. Yang mana jika dihitung akurasinya sebagai berikut:

$$accuracy = \frac{20}{25} \times 100\% = 80\%$$

Jadi akurasi yang didapat pada uji input data single audio baru diluar data training yang didapatkan adalah akurasi sebesar 80%.

4. Kesimpulan

Dari hasil penelitian yang dilakukan dengan menggunakan metode Mel-Frequency Cepstral Coefficient (MFCC) dan K-Nearest Neighbor (KNN) klasifikasi suara burung, dapat disimpulkan sebagai berikut:

- Model K-nearest neighbor yang dihasilkan cukup baik dalam mengklasifikasikan jenis burung berdasarkan suara karena setelah mengklasifikasikan data individual dengan data di luar data training, sistem dapat mengklasifikasikannya dengan akurasi sebesar 80%
- Pada hasil pengujian yang diperoleh dengan K-fold cross-validation didapatkan nilai K terbaik pada model K-nearest neighbor yaitu H. K = 1 dengan akurasi 77%.

References

[1] F. Budiman, M. A. Nursyeha, M. Rivai, and Suwito, “Pengenalan Suara Burung Menggunakan Mel Frequency Cepstrum Coefficient dan Jaringan Syaraf Tiruan Pada Sistem Pengusir Hama Burung” *Jurnal Nasional Teknik Elektro*, vol. 5. No.1, 64-72, 2016

[2] M. I. Sikki,” Pengenalan wajah menggunakan k-nearest neighbour dengan praproses transformasi wavelet” *paradigma*, Vol.10. No. 2, 159-172, 2009

[3] M. A. Hilmi, ”Metoda Mel Frequency Cepstrum Coefficients (MFCC) untuk Mengenali Ucapan pada Bahasa Indonesia” *Jurnal sains dan teknologi informasi*, Vol. 1. No.1. 22-31, 2012

[4] T. Chamidy, ”Metoda Mel Frequency Cepstrum Coefficients (MFCC) pada Klasifikasi Hidden Markov Model (HMM) untuk Kata Arabic pada penutur Indonesia” *Jurnal MATICS*, Vol. 8. No.1. 36-39, 2016

[5] H. Heriyanto, S. Hartati, and A. E. Putra, “Ekstraksi Ciri Mel Frequency Cepstral Coefficient (Mfcc) Dan Rerata Coefficient Untuk Pengecekan Bacaan Al-Qur’an” *Telematika: Jurnal Informatika dan Teknologi Informasi*, Vol. 15 No. 2, 99-108, 2018

This page is intentionally left blank.

Jurnal Elektronik Ilmu Komputer Udayana
Volume 11, No 3. February 2023

p-ISSN: 2301-5373
e-ISSN: 2654-5101

Steganografi Gambar Menggunakan *Least Significant Bit Random Placement* Untuk Perlindungan Hak Cipta Manuskrip Lontar Bali Berbasis Android

Muhammad Akbar Hamid[a1], I Gede Arta Wibawa[a2], Agus Muliantara[a3], I Wayan Santiyasa[a4], Anak Agung Istri Ngurah Eka Karyawati[a5], I Made Widiartha[a6]

[a]Program Studi Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Udayana
Kuta Selatan, Badung, Bali, Indonesia
[1]muhakbarhamid21@gmail.com
[2]gede.arta@unud.ac.id
[3]muliantara@unud.ac.id
[4]santiyasa@unud.ac.id
[5]eka.karyawati@unud.ac.id
[6]madewidiartha@unud.ac.id

Abstract

The Balinese lontar manuscripts are a valuable ancestral cultural heritage of the Balinese people because they contain ancient knowledge and historical records. Along with the development of ejection age, the material from ejection manuscripts is very vulnerable to damage. This damage can result in loss of information on the ejection, but this can be minimized by digitizing it. At this time if there are many types of lontar manuscripts that are digitized, it will be difficult to find proof of ownership. So, the authors conducted research on creating a copyright protection system for Balinese lontar manuscripts by inserting proof of ownership. The method used for insertion is the LSB method with ChaCha20 as additional security and random key generation. The research was successfully carried out based on the results of testing the computational process on the imperceptibility criterion that there was no difference between the stego image and the original image, then on fidelity got the MSE value of 0.00143 and PSNR 76.56 dB, the results of the recovery criteria stated that the inserted embedded image could be extracted with an average the average time is 20.629 seconds, and on the robustness criterion, the results with the embedded image cannot be extracted because it is not resistant to the processing of the stego image. Another measure of success based on software testing using black box testing is to find that the functional application works well and there are no problematic features and the UAT results get very good ratings from respondents with an average weighting value of 89.2%.

Keywords: *LSB, ChaCha20, Android, Copyright Protection, Lontar Bali*

1. Pendahuluan

Manuskrip lontar Bali merupakan naskah kuno warisan budaya leluhur masyarakat Bali yang berharga dan penting karena pada manuskrip terdapat ilmu pengetahuan dan catatan sejarah zaman dahulu. Manuskrip lontar Bali merupakan sarana menulis pada zaman dahulu dengan menggunakan daun pohon lontar sebelum adanya kertas. Pada saat ini manuskrip yang ditulis menggunakan daun lontar masih menjadi bagian aktif dari budaya literasi masyarakat Bali modern.

Seiring dengan perkembangan usia lontar, material dari manuskrip lontar sangat rentan terhadap kerusakan seperti pelapukan. Adanya kerusakan dapat berdampak hilangnya informasi yang terdapat pada manuskrip lontar. Maka dari itu perlunya tindakan pelestarian seperti digitalisasi. Digitalisasi memiliki manfaat untuk menjaga keamanan informasi dan memberikan akses pengetahuan yang terdapat pada manuskrip lontar kepada masyarakat umum.

Sementara itu, kebutuhan keamanan dan kerahasiaan data atau informasi pada saat ini semakin meningkat, terlebih jika banyak jenis manuskrip lontar yang didigitalisasikan. Akan sulit menemukan

kepemilikan jika terdapat banyak jenis manuskrip lontar yang didigitalisasikan sehingga apresiasi untuk pemilik lontar akan hilang. Maka dari itu perlu dilakukan cara untuk melindungi kepemilikan hak cipta atas manuskrip lontar digital dengan cara menyisipkan identitas kepemilikan ke dalam manuskrip lontar Bali. Salah satu metode yang digunakan untuk melakukan penyisipan informasi pada suatu media digital adalah steganografi.

Steganografi adalah proses penyisipan informasi rahasia atau pribadi di dalam suatu media digital dengan cara tidak terdeteksi dan tidak tampak secara langsung sehingga tidak mengganggu tampilan dari media penampung. Objek penyisipan (*embedded object*) dan media penyisipan (*cover object*) yang digunakan memiliki tipe data bervariasi seperti teks, gambar, audio, dan video [1]. Salah satu metode steganografi yang digunakan untuk penyisipan informasi rahasia adalah metode *Least Significant Bit* (LSB).

Metode LSB memiliki kelebihan dari cara mengamankan *embedded object* yaitu perubahan yang dihasilkan pada *cover object* sangat kecil, sehingga sulit untuk diketahui oleh manusia. Dikarenakan penyisipan yang dilakukan oleh LSB disisipkan pada bit terakhir dari nilai biner RGB (*red*, *green*, dan *blue*) *cover object* [2]. Dilihat berdasarkan dari metode keamanan data yang terbaru seperti *blockchain*, LSB memiliki keunggulan dari segi biaya dan kecepatan dikarenakan pada hal tersebut *blockchain* membutuhkan biaya yang mahal dari perancangan sampai perawatan dan kecepatan yang diberikan memerlukan waktu yang lama untuk prosesnya. Sedangkan LSB memberikan biaya yang murah dikarenakan implementasi yang tidak terlalu rumit dan waktu proses yang relatif singkat. Dari segi autentikasi keaslian datanya, blockchain dan LSB sama-sama dapat dibuktikan keaslian datanya [3].

Pada implementasi pengembangan metode LSB, menurut Hernandes et al. (2019) mengatakan bahwa metode LSB yang digunakan dikombinasikan dengan *Pseudorandom Number Generator* (PRNG) untuk membangkitkan bilangan acak dari posisi piksel yang disisipkan pesan rahasia dapat menghasilkan nilai *Peak Signal to Noise Ratio* (PSNR) mencapai 51 dB. Pada penelitian dengan *cover* dan *embedded object* berupa gambar, Das et al. (2018) mengatakan bahwa kombinasi LSB dan PRNG dapat menghasilkan nilai PSNR sebesar 68 dB. Hasil dari penelitian-penelitian tersebut dapat dikatakan cukup baik, akan tetapi dapat dikembangkan lagi untuk mendapatkan hasil yang lebih baik.

Berdasarkan permasalahan yang ada dan penelitian yang sudah dilakukan, penulis ingin membangun sistem untuk melestarikan serta menjaga kepemilikan dari manuskrip lontar Bali dengan cara menyisipkan bukti kepemilikan dengan tanda tangan digital menggunakan *Quick Response* (QR) *code*. Tanda-tangan digital QR *code* dipilih dikarenakan umum digunakan untuk mengidentifikasi keaslian dari tanda tangan atau bukti kepemilikan. Untuk metode yang digunakan untuk penyisipan dan ekstraksi adalah menggunakan metode LSB dan ChaCha20. Metode ChaCha20 digunakan sebagai C*ryptographically Secure Pseudorandom Number Generator* (CSPRNG) atau keamanan ketika proses penyisipan dan ekstraksi serta memetakan persebaran piksel penyisipan dan ekstraksi secara acak.

2. Metode Penelitian

Metode penelitian yang digunakan menggunakan metode pengembangan perangkat lunak *Rapid Application Development* (RAD). Metode RAD yang digunakan memiliki beberapa tahapan, yaitu analisis kebutuhan sistem, pemodelan sistem, implementasi sistem, dan pengujian sistem.

2.1. Analisis Kebutuhan Sistem

2.1.1. Analisis Kebutuhan Data Penelitian

Pada analisis kebutuhan data penelitian, data manuskrip lontar Bali yang digunakan sebagai *cover image* sedangkan data tanda tangan digital QR *code* digunakan sebagai *embedded image*.

a. Sumber Data

Sumber data manuskrip lontar Bali berasal dari Palm Leaf Wiki yang menyediakan arsip digital berbasis website dengan koleksi digital mengenai manuskrip lontar, salah satunya adalah koleksi digital manuskrip lontar Bali. Sedangkan untuk sumber data tanda tangan digital QR *code* berasal dari aplikasi generator QR *code*.

b. Jenis Data

Jenis data yang digunakan pada manuskrip lontar Bali dan tanda tangan digital QR *code* merupakan citra digital dengan ekstensi *portable network graphics* atau *.png.

c. Jumlah Data

Jumlah data manuskrip lontar Bali yang digunakan berjumlah sepuluh. Sedangkan untuk jumlah data tanda tangan digital QR *code* berjumlah dua.

2.1.2. Analisis Kebutuhan Fungsional

Analisis kebutuhan fungsional meliputi apa saja fitur-fitur yang harus ada dan merepresentasikan jalannya suatu sistem. Berikut adalah analisis kebutuhan fungsional dari sistem yang dibangun pada penelitian ini:

a. Sistem harus dapat mengirimkan data *input* pengguna ke server.
b. Server harus dapat mengolah masukan data pengguna dan mengembalikan nilai hasil proses komputasi.
c. Sistem harus dapat menangkap dan menampilkan hasil nilai *return* dari server.

2.1.3. Analisis Kebutuhan Non Fungsional

Analisis kebutuhan non fungsional berperan untuk membantu fungsionalitas dari sistem. Berikut merupakan analisis kebutuhan non fungsional untuk aplikasi yang dibangun pada penelitian ini:

a. Kegunaan (*usability*)

Agar dapat mudah dipahami oleh pengguna, aplikasi harus dilengkapi dengan yang sederhana, ikon yang mudah dipahami, serta ukuran teks dan tombol yang dapat dibaca dengan baik.

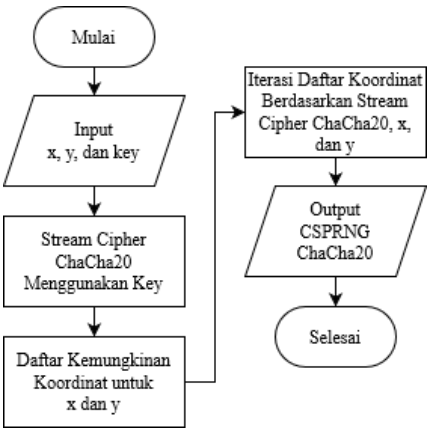
b. Portabilitas (*portability*)

Agar sistem dapat berjalan dengan baik pada lintas versi android *device*, maka sistem yang dibangun harus memerhatikan spesifikasi minimal dari *device* yang dipakai oleh pengguna.

2.2. Pemodelan Sistem

2.2.1. Rancangan Komputasi Aplikasi

a. Proses Pembangkitan Kunci

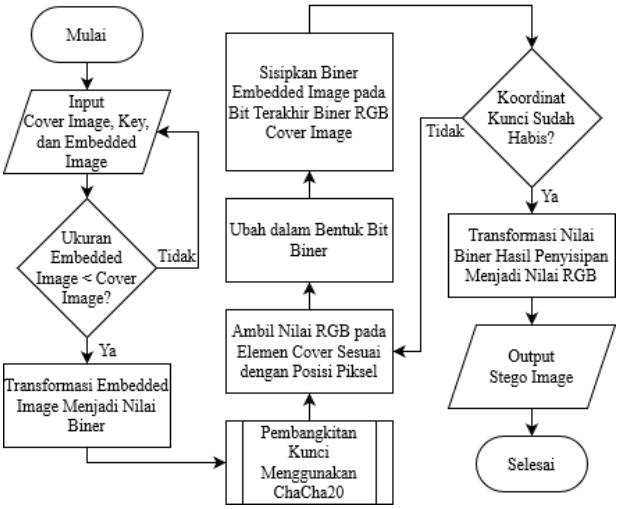


Gambar 1. *Flowchart* Proses Pembangkitan Kunci

Proses pembangkitan kunci menggunakan metode ChaCha20. Metode ChaCha20 digunakan sebagai *Cryptographically Secure Pseudorandom Number Generator* (CSPRNG) untuk pembangkitan kunci secara acak serta menentukan koordinat penyisipan *embedded image* secara acak pada *cover image* dan mendapatkan atau menentukan lokasi acak penyisipan pada *stego image* untuk mengekstrak *embedded image*. Proses *flowchart* proses pembangkitan kunci terdapat pada gambar 1.

b. Proses Penyisipan

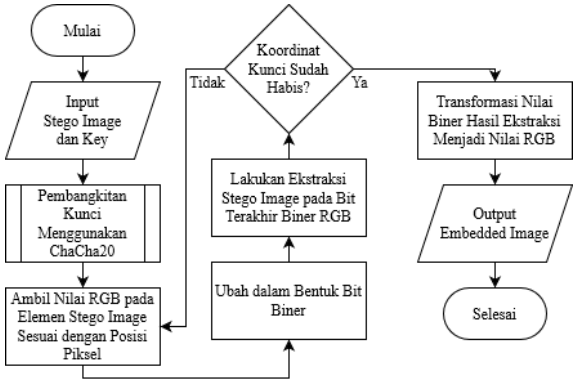
Proses penyisipan akan menggunakan metode LSB. Metode LSB berfungsi sebagai metode utama untuk melakukan penyisipan *embedded image* pada *cover image*. *Flowchart* proses penyisipan terdapat gambar 2.



Gambar 2. *Flowchart* Proses Komputasi Penyisipan

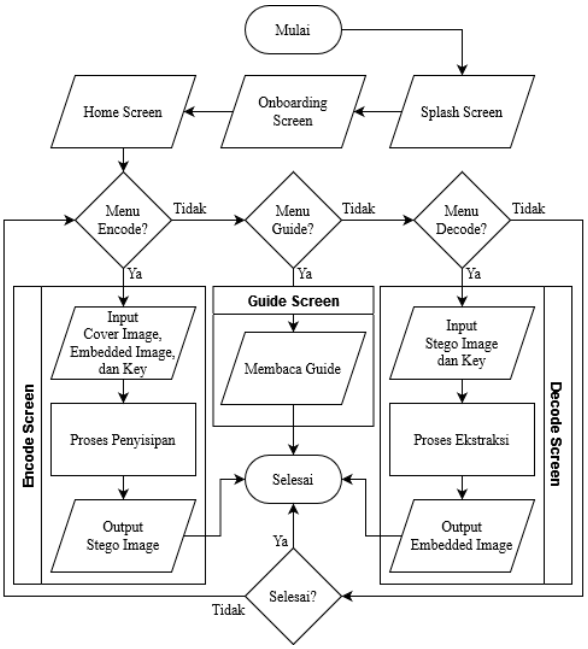
c. Proses Ekstraksi

Proses ekstraksi akan menggunakan metode LSB. Metode LSB berfungsi sebagai metode utama untuk melakukan ekstraksi *embedded image*. *Flowchart* proses ekstraksi terdapat pada gambar 3.



Gambar 3. *Flowchart* Proses Komputasi Ekstraksi

2.2.2. Rancangan Alur Aplikasi



Gambar 4. *Flowchart* Rancangan Alur Aplikasi

Rancangan alur aplikasi digunakan sebagai gambaran jalannya aplikasi dari satu fitur ke fitur lainnya. Untuk *flowchart* rancangan alur aplikasi terdapat pada gambar 4.

2.3. Rancangan Implementasi Sistem

Implementasi hasil rancangan komputasi ke dalam bentuk program menggunakan bahasa pemrograman python sebagai *backend* dan menggunakan *framework* flask pada python untuk API. Untuk rancangan implementasi perangkat lunak menggunakan bahasa pemrograman dart dengan *framework* flutter sebagai *frontend*.

2.4. Rancangan Pengujian Sistem

2.4.1. Rancangan Pengujian Komputasi

Pengujian komputasi dilakukan bertujuan untuk mengukur performa yang dimiliki oleh sistem. Pengujian komputasi dibagi berdasarkan kriteria dari steganografi, yaitu *imperceptibility*, *fidelity*, *recovery*, dan *robustness*.

a. Imperceptibility

Pengujian *imperceptibility* adalah pengujian untuk menguji kualitas *stego image* apakah *stego image* mengalami perubahan yang mencurigakan atau tidak berdasarkan pandangan mata manusia secara langsung dengan perbandingan antara gambar asli *stego image* sebelum dilakukan penyisipan atau *original image*. Pengujian ini dikatakan berhasil apabila *stego image* yang dihasilkan tidak mengalami distorsi dibandingkan dengan *original image* berdasarkan pengamatan secara langsung.

b. Fidelity

Pengujian *fidelity* adalah pengujian untuk mengetahui kualitas *stego image* setelah ditambahkan *embedded image*. Untuk mengetahui nilai kualitas tersebut, pengujian yang dilakukan pada penelitian ini adalah menggunakan perhitungan nilai *Mean Square Error* (MSE) dan *Peak Signal-to-Noise Ratio* (PSNR).

c. Recovery

Pengujian *recovery* merupakan pengujian untuk mengetahui keberhasilan dari *embedded image* yang telah disisipkan dalam *stego image* untuk dapat diekstrak kembali atau tidak. Untuk dapat diekstraksi, maka syaratnya adalah *key* atau kunci yang digunakan harus sesuai dengan kunci yang digunakan saat proses penyisipan gambar. Ukuran keberhasilan dari pengujian *recovery* akan diacu berdasarkan keberhasilan ekstraksi pesan dan kecepatan waktu ekstraksi.

d. Robustness

Pengujian *robustness* merupakan pengujian untuk membuktikan bahwa *embedded image* tahan terhadap berbagai operasi manipulasi atau *editing* yang dilakukan pada *stego image* dan untuk mengetahui pesan yang disembunyikan apakah berhasil diekstrak atau tidak. Adapun operasi manipulasi atau *editing* yang digunakan adalah pemotongan (*cropping*), mengubah resolusi (*resizing*), rotasi (*rotation*), *gaussian blur*, dan *grayscale*.

2.4.2. Rancangan Pengujian Perangkat Lunak

a. *Black Box Testing*

Pengujian *black box* dilakukan untuk menguji setiap fungsi pada aplikasi berjalan sesuai dengan rancangan atau tidak. Pengujian dilakukan dengan mendefinisikan kumpulan kondisi *input* dan *output* serta mendemonstrasikan spesifikasi fungsional pada aplikasi. Pengujian akan menghasilkan validasi dari *input* dan *output* aplikasi sesuai dengan rancangan.

b. *User Acceptance Testing* (UAT)

Pengujian aplikasi dengan UAT dilakukan berdasarkan tiga aspek, yaitu aspek aplikasi, pengguna, dan interaksi. Masing-masing dari aspek tersebut akan berisikan pertanyaan-pertanyaan yang nantinya akan diberikan kepada responden atau pengguna (*user*) untuk dinilai berdasarkan aplikasi. Hasil jawaban responden akan dihitung menggunakan skala pembobotan untuk mendapatkan hasil nilai pembobotan berdasarkan tabel 1.

Tabel 1. Skala Pembobotan UAT

Pernyataan	Nilai	Bobot
Sangat Setuju	A	5
Setuju	B	4

Netral	C	3
Tidak Setuju	D	2
Sangat Tidak Setuju	E	1

Berikut merupakan pertanyaan-pertanyaan berdasarkan aspek yang akan diberikan kepada responden:

- Aspek Aplikasi
  1) Apakah tampilan aplikasi menarik?
  2) Apakah tampilan warna dan *interface* pada aplikasi enak dilihat dan tidak membosankan?
  3) Apakah tampilan tata aplikasi jelas dan mudah dipahami?
- Aspek Pengguna
  4) Apakah aplikasi mempunyai menu-menu yang mudah dipahami?
  5) Apakah aplikasi mudah dioperasikan?
  6) Apakah aplikasi ini dapat digunakan untuk melakukan penyisipan dan ekstraksi?
- Aspek Interaksi
  7) Apakah hasil dari penyisipan dapat terlihat?
  8) Apakah hasil dari ekstraksi dapat terlihat?
  9) Apakah fitur guide atau panduan penggunaan mudah dipahami?

3. Hasil dan Pembahasan

3.1. Data Penelitian

a. *Cover Image*

Data gambar manuskrip lontar Bali yang digunakan sebagai *cover image* berjumlah sepuluh dengan memiliki kepemilikan pada masing-masing manuskrip lontar Bali. Gambar manuskrip lontar Bali yang dimiliki oleh Gedong Kirtya Singaraja berjudul Awig-awig Desa Silangjana, Carcan Kucing, Gaguritan Nengah Jimbaran, dan Atlas Bhumi (Kakawin). Sedangkan gambar manuskrip lontar Bali yang dimiliki oleh Pusat Dokumentasi Dinas Kebudayaan Provinsi Bali berjudul Babad Gumi, Geguritan Rusak Banjar, Gaguritan Cangak, Keputusan Sri Aji Malayu, Kidung Rajapala, dan Usada Buda Kecapi.

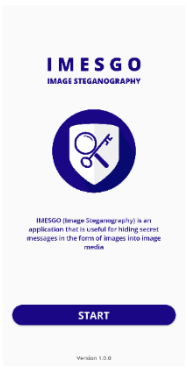
b. *Embedded Image*

Data gambar tanda tangan digital QR *code* yang digunakan sebagai embedded image berjumlah dua dengan kepemilikan data gambar manuskrip lontar Bali, yaitu Gedong Kirtya Singaraja dan Pusat Dokumentasi Dinas Kebudayaan Provinsi Bali.

3.2. Implementasi Sistem

Implementasi sistem merupakan tahapan untuk merealisasikan analisis dan rancangan yang telah dibuat. Implementasi yang dilakukan menggunakan *frontend* dengan *framework* flutter pada bahasa pemrograman dart dan *backend* menggunakan python. Untuk menghubungkan antara *frontend* dan *backend* menggunakan flask sebagai API. Hasil dari implementasi sistem terdapat pada gambar 5, gambar 6, gambar 7, gambar 8, gambar 9, gambar 10, gambar 11, dan gambar 12.

Gambar 5. *Splash Screen*

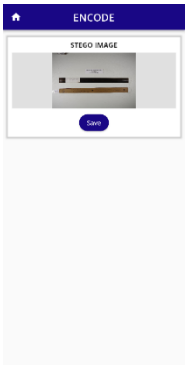


Gambar 6. *Onboarding Screen*

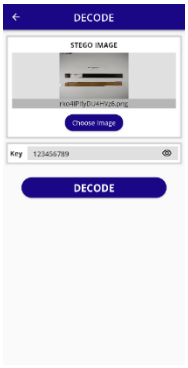
Gambar 7. *Home Screen*

Gambar 8. *Encode Screen*

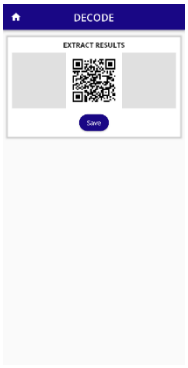
p-ISSN: 2301-5373
e-ISSN: 2654-5101



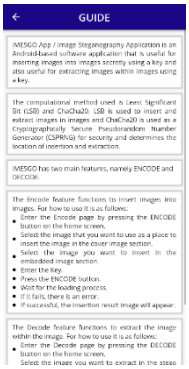
Gambar 9. *Encode Result Screen*



Gambar 10. *Decode Screen*



Gambar 11. *Decode Result Screen*



Gambar 12. *Guide Screen*

3.3. Pengujian Sistem

3.3.1. Pengujian Komputasi

a. *Imperceptibility*

Pengujian *imperceptibility* menggunakan kuesioner untuk dibagikan kepada responden dan membandingkan perbedaan antara gambar *original* dan gambar *stego*. Hasil jawaban kuesioner telah diisi sebanyak 20 orang responden dan jawaban kuesioner dirangkum pada tabel 2.

Tabel 2. Jawaban Kuesioner Pengujian *Imperceptibility*

No	Judul Manuskrip Lontar Bali	Jawaban		Persentase	
		Berbeda	Tidak Berbeda	Berbeda	Tidak Berbeda
1	Awig-awig Desa Silangjana	2	18	10	90
2	Babad Gumi	3	17	15	85
3	Carcan Kucing	1	19	5	95
4	Gaguritan Nengah Jimbaran	1	19	5	95
5	Geguritan Rusak Banjar	3	17	15	85
6	Gaguritan Cangak	3	17	15	85
7	Atlas Bhumi (Kakawin)	1	19	5	95
8	Keputusan Sri Aji Malayu	1	19	5	95
9	Kidung Rajapala	1	19	5	95
10	Usada Buda Kecapi	2	18	10	90

Berdasarkan tabel 3, responden memilih jawaban tidak berbeda dengan rata-rata 91% dan jawaban berbeda dengan rata-rata 9%. Dapat dikatakan hasil yang diberikan sangat baik dikarenakan responden menjawab bahwa pada perbandingan perbedaan antara gambar *original* dan gambar *stego* tidak memiliki perbedaan yang cukup terlihat atau tidak memiliki distorsi pada gambar *stego*.

b. *Fidelity*

Hasil dari pengujian *fidelity* pada tabel 3 memiliki nilai rata-rata MSE sebesar 0,00143 dan nilai rata-rata PSNR sebesar 76,56 dB. Hasil tersebut dapat dikatakan baik dan memiliki tingkat kemiripan yang tinggi dikarenakan pada nilai MSE nilainya mendekati nol dan pada nilai PSNR nilainya lebih dari 30 dB.

Tabel 3. Hasil Pengujian *Fidelity*

No	Judul Manuskrip Lontar Bali	*Key*	MSE	PSNR
1	Awig-awig Desa Silangjana	4o7AzH1Tzg	0,00136	76,79 dB
2	Babad Gumi	hROpPBpHOP	0,00150	76,36 dB
3	Carcan Kucing	dKyjgvoYIR	0,00135	76,81 dB
4	Gaguritan Nengah Jimbaran	MzlfYTo9xA	0,00134	76,83 dB
5	Geguritan Rusak Banjar	Q5f0RdDSsd	0,00149	76,39 dB
6	Gaguritan Cangak	V5iUqNDPR6	0,00146	76,46 dB
7	Atlas Bhumi (Kakawin)	F2GZ9rblwR	0,00136	76,79 dB
8	Keputusan Sri Aji Malayu	ZKGP2XVdDC	0,00150	76,36 dB
9	Kidung Rajapala	vkajpnThy6	0,00148	76,40 dB
10	Usada Buda Kecapi	tmD1lLML2l	0,00149	76,38 dB

c. *Recovery*

Pada pengujian *recovery*, *stego image* berhasil diekstrak dengan kunci yang sama dengan ketika proses penyisipan dan mendapatkan *embedded image* dengan rata-rata waktu proses ekstraksi 20,629 detik. Hasil dari pengujian *recovery* dapat dilihat pada tabel 4.

Tabel 4. Hasil Pengujian *Recovery*

No	Judul Manuskrip Lontar Bali	*Key*	Waktu Ekstraksi (detik)	Hasil Ekstraksi
1	Awig-awig Desa Silangjana	4o7AzH1Tzg	19,93	Berhasil
2	Babad Gumi	hROpPBpHOP	20,42	Berhasil
3	Carcan Kucing	dKyjgvoYlR	19,87	Berhasil
4	Gaguritan Nengah Jimbaran	Q5f0RdDSsd	20,98	Berhasil
5	Geguritan Rusak Banjar	MzlfYTo9xA	22,05	Berhasil
6	Gaguritan Cangak	V5iUqNDPR6	21,24	Berhasil
7	Atlas Bhumi (Kakawin)	F2GZ9rblwR	18,99	Berhasil
8	Keputusan Sri Aji Malayu	ZKGP2XVdDC	21,16	Berhasil
9	Kidung Rajapala	vkajpnThy6	21,35	Berhasil
10	Usada Buda Kecapi	tmD1lLML2l	20,30	Berhasil

d. *Robustness*

Pengujian *robustness* dilakukan manipulasi atau *editing* pada *stego image* untuk mengetahui ketahanan dan keberhasilan dari *embedded image* ketika diekstrak. Kunci yang digunakan atau yang terdapat pada *stego image* adalah acak dengan panjang satu sampai sepuluh karakter berdasarkan masing-masing *stego image*. Untuk parameter manipulasi atau *editing* menyesuaikan dengan kemungkinan yang terjadi pada proses manipulasi atau *editing.* Untuk hasil pengujian pemotongan (*cropping*) terdapat pada tabel 5, pengujian mengubah resolusi (*resizing*) terdapat pada tabel 6, pengujian rotasi (*rotation*) terdapat pada tabel 7, pengujian *gaussian blur* terdapat pada tabel 8, dan pengujian *grayscale* terdapat pada tabel 9.

Tabel 5. Hasil Pengujian Pemotongan (*cropping*)

No	*Stego Image*	*Key*	Ukuran Pemotongan	Lokasi Pemotongan	Hasil Ekstraksi
1	Awig-awig Desa Silangjana	dh	50%	Pojok Kiri Atas	Gagal
2	Babad Gumi	HM1ps	50%	Pojok Kanan Atas	Gagal
3	Carcan Kucing	nBf9wS	50%	Tengah	Gagal
4	Gaguritan Nengah Jimbaran	NUL8KFJ40	50%	Pojok Kiri Bawah	Gagal
5	Geguritan Rusak Banjar	o	50%	Pojok Kanan Bawah	Gagal
6	Gaguritan Cangak	OUIEKT6d	75%	Pojok Kiri Atas	Gagal
7	Atlas Bhumi (Kakawin)	P19z	75%	Pojok Kanan Atas	Gagal
8	Keputusan Sri Aji Malayu	qpybsnvayf	75%	Tengah	Gagal
9	Kidung Rajapala	qSzQUU6	75%	Pojok Kiri Bawah	Gagal
10	Usada Buda Kecapi	uvO	75%	Pojok Kanan Bawah	Gagal

Tabel 6. Hasil Pengujian Mengubah Resolusi (*resizing*)

No	*Stego Image*	*Key*	Resolusi		Perubahan Ukuran	Hasil Ekstraksi
			Sebelum	Sesudah		
1	Awig-awig Desa Silangjana	b	2808 x 1872	5616 x 3744	+ 100%	Gagal
2	Babad Gumi	tkyBM1	2808 x 1872	4914 x 3276	+ 75%	Gagal
3	Carcan Kucing	C	2808 x 1872	4212 x 2808	+ 50%	Gagal
4	Gaguritan Nengah Jimbaran	lYaO	2808 x 1872	3510 x 2340	+ 25%	Gagal
5	Geguritan Rusak Banjar	o5	2808 x 1872	2948 x 1965	+ 5%	Gagal
6	Gaguritan Cangak	wTq	2808 x 1872	2668 x 1779	- 5%	Gagal
7	Atlas Bhumi (Kakawin)	9AfoRyfzK	2808 x 1872	2106 x 1404	- 25%	Gagal
8	Keputusan Sri Aji Malayu	1x9gNn1	2808 x 1872	1404 x 936	- 50%	Gagal
9	Kidung Rajapala	akeXCcJb	2808 x 1872	702 x 468	- 75%	Gagal
10	Usada Buda Kecapi	O6dn6	2808 x 1872	141 x 94	- 95%	Gagal

Tabel 7. Hasil Pengujian Rotasi (*rotation*)

No	*Stego Image*	*Key*	Perubahan Rotasi	Hasil Ekstraksi
1	Awig-awig Desa Silangjana	pLNU06NAi9	-150°	Gagal
2	Babad Gumi	V	-120°	Gagal

p-ISSN: 2301-5373
e-ISSN: 2654-5101

3	Carcan Kucing	vlvEmQq	-90°	Gagal
4	Gaguritan Nengah Jimbaran	kA	-60°	Gagal
5	Geguritan Rusak Banjar	mRTJ4v	-30°	Gagal
6	Gaguritan Cangak	SmueBe0ba	30°	Gagal
7	Atlas Bhumi (Kakawin)	6xB	60°	Gagal
8	Keputusan Sri Aji Malayu	guYP	90°	Gagal
9	Kidung Rajapala	Y77dS	120°	Gagal
10	Usada Buda Kecapi	v0HmQDfU	150°	Gagal

Tabel 8. Hasil Pengujian *Gaussian Blur*

No	*Stego Image*	*Key*	Radius (piksel)	Hasil Ekstraksi
1	Awig-awig Desa Silangjana	GRE	1,0	Gagal
2	Babad Gumi	nRDEZ	2,0	Gagal
3	Carcan Kucing	s7se8vnRuV	3,0	Gagal
4	Gaguritan Nengah Jimbaran	N71tQqV	4,0	Gagal
5	Geguritan Rusak Banjar	M	5,0	Gagal
6	Gaguritan Cangak	FA	6,0	Gagal
7	Atlas Bhumi (Kakawin)	0Xgmd0J	7,0	Gagal
8	Keputusan Sri Aji Malayu	3yZx	8,0	Gagal
9	Kidung Rajapala	qwUGHEys	9,0	Gagal
10	Usada Buda Kecapi	yWK1PWoAA	10,0	Gagal

Tabel 9. Hasil Pengujian *Grayscale*

No	*Stego Image*	*Key*	Hasil Ekstraksi
1	Awig-awig Desa Silangjana	JAl2WAN2e	Gagal
2	Babad Gumi	yW7Fe0h	Gagal
3	Carcan Kucing	v13bSFYJ	Gagal
4	Gaguritan Nengah Jimbaran	i	Gagal
5	Geguritan Rusak Banjar	OkOqhhr3JA	Gagal
6	Gaguritan Cangak	3hbz	Gagal
7	Atlas Bhumi (Kakawin)	wZ	Gagal
8	Keputusan Sri Aji Malayu	4CNBU	Gagal
9	Kidung Rajapala	3RXAvV	Gagal
10	Usada Buda Kecapi	7fQ	Gagal

3.3.2. Pengujian Perangkat Lunak

a. *Black Box Testing*

Pengujian *black box* dilakukan untuk memastikan apakah masukan serta keluaran pada aplikasi telah berjalan sesuai dengan rancangan. Hasil dari pengujian *black box* terdapat pada tabel 10.

Tabel 10. Hasil Pengujian *Black Box*

No	Pengujian	Hasil
1	Memencet tombol *start* pada *onboarding screen* untuk menampilkan *home screen*	Valid
2	Memencet tombol *encode* pada *home screen* untuk menampilkan *encode screen*	Valid
3	Memencet tombol *decode* pada *home screen* untuk menampilkan *decode screen*	Valid
4	Memencet tombol *guide* pada *home screen* untuk menampilkan *guide screen*	Valid
5	Memasukkan *cover image* dan *embedded image* lalu menampilkan gambar dan namanya pada *encode screen*	Valid
6	Memasukkan *key* di dalam *text field* pada *encode screen*	Valid
7	Memencet tombol *encode* pada *encode screen* untuk melakukan proses *encode*	Valid
8	Melakukan proses *encode*	Valid
9	Menampilkan *stego image* atau hasil proses *encode*	Valid
10	Memasukkan *stego image* lalu menampilkan gambar dan namanya pada *decode screen*	Valid
11	Memasukkan *key* di dalam *text field* pada *decode screen*	Valid
12	Memencet tombol *decode* pada *decode screen* untuk melakukan proses *decode*	Valid
13	Melakukan proses *decode*	Valid
14	Menampilkan hasil ekstrak atau hasil proses *decode*	Valid

b. *User Acceptance Testing* (UAT)

Hasil dari pengujian UAT didapatkan 20 orang responden. Kuesioner yang telah diisi oleh responden diolah berdasarkan bobotnya pada tabel 1. Didapatkan hasil nilai pembobotan seperti pada tabel 11.

Tabel 11. Hasil Nilai Pembobotan Pengujian UAT

No	Pertanyaan		Bobot					Jumlah Bobot	Nilai (%)
	Aspek	Soal	A	B	C	D	E		
1	Aspek Aplikasi	Apakah tampilan aplikasi menarik?	35	44	3	2	0	84	84
2		Apakah tampilan warna dan *interface* pada aplikasi enak dilihat dan tidak membosankan?	45	28	9	2	0	84	84
3		Apakah tampilan tata aplikasi jelas dan mudah dipahami?	70	24	0	0	0	94	94
4	Aspek Pengguna	Apakah aplikasi mempunyai menu-menu yang mudah dipahami?	65	28	0	0	0	93	93
5		Apakah aplikasi mudah dioperasikan?	70	24	0	0	0	94	94
6		Apakah aplikasi ini dapat digunakan untuk melakukan penyisipan dan ekstraksi?	50	24	12	0	0	86	86
7	Aspek Interaksi	Apakah hasil dari penyisipan dapat terlihat?	50	24	9	2	0	85	85
8		Apakah hasil dari ekstraksi dapat terlihat?	55	24	9	0	0	88	88
9		Apakah fitur *guide* atau panduan penggunaan mudah dipahami?	45	32	6	2	0	85	85

Berdasarkan tabel 12 didapatkan hasil nilai pembobotan dengan rata-rata pada aspek aplikasi mendapatkan nilai 87,3%, aspek pengguna mendapatkan nilai 94,3%, dan aspek interaksi mendapatkan nilai 86%. Jadi dapat diketahui bahwa ketiga aspek tersebut telah memenuhi kebutuhan dari pengguna (*user*).

4. Kesimpulan

Penelitian berhasil dilakukan berdasarkan hasil pengujian proses komputasi pada kriteria *imperceptibility* dengan tidak adanya perbedaan antara *stego image* dan *original image*, lalu pada *fidelity* mendapatkan nilai MSE 0,00143 dan PSNR 76,56 dB, hasil kriteria *recovery* menyatakan *embedded image* yang disisipkan dapat diekstrak dengan rata-rata waktu 20,629 detik, dan pada kriteria *robustness* mendapatkan hasil dengan *embedded image* tidak dapat diekstrak karena tidak tahan terhadap proses manipulasi pada *stego image*. Ukuran keberhasilan lainnya berdasarkan pengujian perangkat lunak menggunakan *black box testing* mendapatkan bahwa fungsional aplikasi bekerja dengan baik serta tidak ada fitur yang bermasalah dan pada hasil UAT mendapatkan penilaian sangat baik dari responden dengan rata-rata hasil nilai pembobotan sebesar 89,2%.

Referensi

[1] I. J. Kadhim, P. Premaratne, P. J. Vial, and B. Halloran, “Comprehensive survey of image steganography: Techniques, Evaluations, and trends in future research,” *Neurocomputing*, vol. 335, pp. 299–326, 2019.

[2] Alvin, A. Wicaksana, and M. I. Prasetiyowati, “Digital Watermarking for Color Image Using DHWT and LSB,” in *2019 5th International Conference on New Media Studies (CONMEDIA)*, 2019, pp. 94–99. doi: 10.1109/CONMEDIA46929.2019.8981835.

[3] B. S. Riza, “Blockchain Dalam Pendidikan: Lapisan Logis di Bawahnya,” *ADI Bisnis Digital Interdisiplin Jurnal*, vol. 1, no. 1 Juni, pp. 41–47, 2020.

[4] A. Hernandes, H. Hartini, and D. Sartika, “Steganografi Citra Menggunakan Metode Least Significant Bit (LSB) Dan Linear Congruential Generator (LCG),” *JATISI (Jurnal Teknik Informatika dan Sistem Informasi)*, vol. 5, no. 2, pp. 137–146, 2019.

[5] K. Das, D. Choudhury, and S. K. Bandyopadhyay, “An Ameliorate Image Steganography Method using LSB Technique & Pseudo Random Numbers,” *Journal for Research| Volume*, vol. 4, no. 09, 2018.

ISSN

E-ISSN