

Komparasi Algoritma Naïve Bayes dan Neural Network untuk Menentukan Ketepatan Masa Studi Mahasiswa (Studi Kasus: Program Studi Teknologi Informasi Universitas Udayana)

Ni Putu Erika Sari Bintari^{a1}, I Ketut Gede Darma Putra^{a2}, I Made Sunia Raharja^{a3}

^aProgram Studi Teknologi Informasi, Fakultas Teknik, Universitas Udayana, Bali

e-mail: 1erikasari@gmail.com, 2ikgdarmaputra@unud.ac.id, 3sunia.raharja@unud.ac.id

Abstrak

Perguruan tinggi memiliki akreditasi sebagai acuan untuk penilaian terhadap kualitas kampus yang ditentukan oleh Badan Akreditasi Nasional Perguruan Tinggi (BAN-PT). Tingkat akreditasi perguruan tinggi salah satunya dinilai dari kualitas lulusan mahasiswa yaitu lama studi mahasiswa. Lama studi mahasiswa dapat diprediksi menggunakan metode data mining. Metode data mining yang digunakan untuk memprediksi masa studi mahasiswa adalah algoritma naïve bayes dan neural network. Data mahasiswa berupa NIM, jenis kelamin, jalur masuk, IP Semester 1, IP Semester 2, IP Semester 3, IP Semester 4, dan kategori. Data sampel merupakan alumni mahasiswa angkatan 2008-2013 sebanyak 363 mahasiswa. Metode data mining akan dilakukan sebanyak empat kali pengujian berdasarkan jumlah data training dan data testing. Hidden layer yang digunakan sebanyak 1 dan 2 dengan dua neuron, training cycles menggunakan angka kisaran 25 hingga 2000, momentum sebesar 0,9 dan learning rate sebesar 0.5. Nilai performa algoritma naïve bayes akan menampilkan confusion matrix berupa akurasi sedangkan nilai performa algoritma neural network akan menampilkan confusion matrix berupa akurasi dan nilai RMSE.

Kata kunci: Akreditasi, data mining, masa studi, naïve bayes, neural network

Abstract

Universities have accreditation as a reference for the assessment of campus quality determined by the National Accreditation Board for Higher Education (BAN-PT). One of the levels of university accreditation is assessed from the quality of student graduates, namely the length of student study. The length of student study can be predicted using data mining methods. Data mining methods used to predict student study period are naïve Bayes algorithm and neural network. Student data in the form of NIM, gender, entry path, IP Semester 1, IP Semester 2, IP Semester 3, IP Semester 4, and categories. The sample data are alumni of the 2008-2013 class of students as many as 363 students. The data mining method will be tested four times based on the amount of training data and testing data. The hidden layers used are 1 and 2 with two neurons, training cycles using numbers ranging from 25 to 2000, momentum of 0.9 and learning rate of 0.5. The performance value of the naïve bayes algorithm will display a confusion matrix in the form of accuracy, while the performance value of the neural network algorithm will display a confusion matrix in the form of accuracy and RMSE value.

Keywords: Accreditation, data mining, study period, naïve bayes, neural network

1. Pendahuluan

Perguruan tinggi saat ini saling bersaing untuk memiliki keunggulan memberikan kualitas yang baik. Universitas Udayana, khususnya Program Studi Teknologi Informasi yang merupakan bagian dari Fakultas Teknik pun ingin memberikan kualitas yang baik bagi mahasiswanya. Salah satu hal untuk meningkatkan kualitas yang baik adalah dalam hal meningkatkan nilai kelulusan mahasiswa dengan tepat waktu. Tingkat kelulusan mahasiswa suatu kampus berpengaruh dalam pemberian nilai akreditasi oleh BAN-PT. Program Studi Teknologi Informasi Universitas Udayana memiliki akreditasi 'B' yang diberikan oleh BAN-PT

pada tahun 2018 dan berlaku sampai tahun 2023. Nilai akreditasi dapat ditingkatkan dengan beberapa kriteria yang sudah ditetapkan oleh BAN-PT, salah satunya adalah kualitas mahasiswa yang dapat dinilai dari lulusan mahasiswa tepat waktu.

Data *mining* dapat digunakan untuk mencari pola kelulusan mahasiswa dalam suatu kampus. Data *mining* memiliki banyak algoritma yang dapat diterapkan untuk membantu mengolah data. Salah satunya adalah melakukan prediksi untuk kelulusan mahasiswa. Prediksi kelulusan mahasiswa menggunakan atribut NIM, semester, nilai IPS selama 4 semester, jenis kelamin, jalur masuk, dan sks 4 semester.

Penelitian terdahulu yang berkaitan dengan penelitian ini adalah penelitian dari Nurul Amini Abidin (2021). Penelitian tersebut menggunakan satu metode data mining, yaitu algoritma neural network dengan tiga kali pelatihan data. Hasil dari tingkat akurasi yaitu 90,19%, 81,82%, dan 45,45% menggunakan *software* MATLAB. [1]

Penelitian terdahulu yang berkaitan dengan penelitian ini adalah penelitian dari Eko Prasetyo Rohmawan (2018). Penelitian tersebut menggunakan dua metode data mining, yaitu decision tree dan artificial neural network. Hasil dari tingkat akurasi decision tree adalah 74,51% dan artificial neural network adalah 79,74%. [2]

Penelitian terdahulu yang berkaitan dengan penelitian ini adalah penelitian dari Ridwan (2020). Penelitian tersebut menggunakan satu metode data mining, yaitu algoritma neural network. Hasil dari tingkat akurasi algoritma neural network adalah 98,27%. [3]

Penelitian terdahulu yang berkaitan dengan penelitian ini adalah penelitian dari Neni Purwati (2020). Penelitian tersebut menggunakan satu metode data mining, yaitu algoritma neural network. Hasil dari tingkat akurasi algoritma neural network adalah 92,83%. [4]

Penelitian terdahulu yang berkaitan dengan penelitian ini adalah penelitian dari Muhammad Dedek Yalidhan (2018). Penelitian tersebut menggunakan 318 sampel data dengan satu metode data mining, yaitu algoritma neural network. Hasil dari tingkat akurasi algoritma neural network adalah 98,97%. [5]

Penelitian terdahulu yang berkaitan dengan penelitian ini adalah penelitian dari Ananda Fiqri Firdaus (2021). Penelitian tersebut menggunakan satu metode data mining, yaitu algoritma naïve bayes. Hasil dari tingkat akurasi algoritma naïve bayes adalah 90,78%. [6]

Penelitian terdahulu yang berkaitan dengan penelitian ini adalah penelitian dari Anak Agung Arimas Purnamaswari (2022). Penelitian tersebut menggunakan dua metode data mining, yaitu algoritma neural network backpropagation dan algoritma support vector machines. Hasil dari tingkat error Neural Network Backpropagation RMSE 0,048 dan Support Vector Machines RMSE 0,108. [7]

2. Metodologi Penelitian

Metodologi penelitian merupakan suatu cara ilmiah berupa tahapan penelitian untuk melakukan penelitian. Data yang digunakan dalam penelitian ini adalah sebagai berikut.

a. Data training

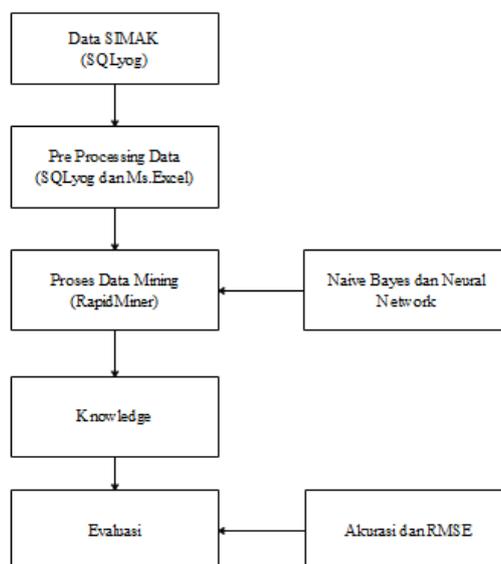
Data training yang digunakan akan dibagi menjadi empat kali pengujian dengan menggunakan atribut NIM, kategori, jenis kelamin, jalur masuk, SKS, IPS semester 1, IPS semester 2, IPS semester 3, dan IPS semester 4. Pengujian pertama menggunakan data alumni mahasiswa tahun 2008-2009 sebanyak 108 mahasiswa. Pengujian kedua menggunakan data alumni mahasiswa tahun 2008-2010 sebanyak 162 mahasiswa. Pengujian ketiga menggunakan data alumni mahasiswa tahun 2008-2011 sebanyak 240 mahasiswa. Pengujian keempat menggunakan data alumni mahasiswa tahun 2008-2012 sebanyak 309 mahasiswa.

b. Data testing

Data testing yang digunakan akan dibagi menjadi empat kali pengujian dengan menggunakan atribut yang sama dengan data training. Pengujian pertama menggunakan data alumni mahasiswa tahun 2010-2013 sebanyak 255 mahasiswa. Pengujian kedua menggunakan data alumni mahasiswa tahun 2011-2013 sebanyak 201 mahasiswa. Pengujian ketiga menggunakan data alumni mahasiswa tahun 2012-2013 sebanyak 123 mahasiswa. Pengujian keempat menggunakan data alumni mahasiswa tahun 2013 sebanyak 54 mahasiswa.

2.1 Gambaran Umum

Gambaran umum merupakan tahapan atau alur proses penelitian dari pengumpulan data hingga data selesai diolah. Berikut merupakan gambaran umum penelitian.



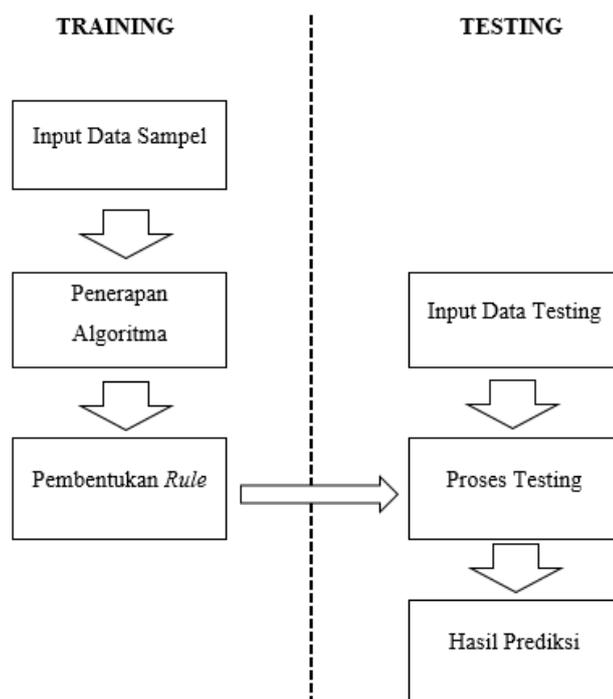
Gambar 1. Gambaran Umum

Gambar 1 merupakan tampilan dari gambaran umum sistem yang dibuat untuk melakukan data mining dalam prediksi kelulusan mahasiswa di perguruan tinggi menggunakan algoritma *naive bayes* dan *neural network*. Data mahasiswa akan disimpan ke dalam *database* menggunakan aplikasi MySQL dengan menghubungkan ke dalam *database server* XAMPP. Data yang sudah disimpan ke dalam *database* akan dilakukan seleksi data dan penyaringan data sehingga hanya data yang dibutuhkan saja yang akan diolah. Atribut yang digunakan adalah NIM, jenis kelamin, jalur masuk, sks yang ditempuh selama 4 semester, IP semester 1, IPS semester 2, IP semester 3, IP semester 4, dan kategori kelulusan. Data yang digunakan adalah data alumni mahasiswa program studi Teknologi Informasi, Universitas Udayana angkatan 2008 – 2013.

Data yang sudah di seleksi selanjutnya akan dilakukan proses *data mining* menggunakan aplikasi *open source* RapidMiner 9.4. Data yang sudah di seleksi kemudian di *import* dan dibagi menjadi dua jenis data, yaitu *data training* dan *data testing*. Algoritma yang digunakan dalam penelitian ini adalah algoritma *naive bayes* dan *neural network*. Uji masing-masing algoritma dibagi menjadi 4 (empat) pengujian berdasarkan jumlah *data training* dan *data testing*. Hasil data mining akan memperlihatkan hasil nilai akurasi keberhasilan masing-masing algoritma dalam menentukan kategori ketepatan masa studi mahasiswa. Kedua algoritma akan dilakukan komparasi untuk menentukan algoritma terbaik yang bisa digunakan untuk melakukan prediksi ketepatan masa studi mahasiswa di masa mendatang.

2.2 Deskripsi Proses

Alur proses algoritma yang akan berjalan akan dijelaskan pada blok diagram sebagai berikut. Blok diagram adalah alur kerja yang menyatakan urutan dari proses *data mining* pada RapidMiner. Terdapat dua proses yaitu, proses pengolahan data training dan proses data testing



Gambar 2. Deskripsi Proses

Gambar 2 merupakan tampilan dari diagram blok untuk menunjukkan alur proses kerja pengolahan *data mining*. Terdapat dua proses utama dalam alur, yaitu proses *training* dan proses *testing*. Proses pengolahan *data training* dilakukan dengan menginputkan data sampel yaitu untuk dilakukan proses pengolahan data dengan algoritma *naïve bayes* dan *neural network*. Kemudian akan didapatkan *rule* untuk menjadi dasar pengolahan *data testing*.

Proses pengolahan *data testing* dilakukan dengan menginputkan data uji yaitu mahasiswa dengan ketentuan jumlah dan nama atribut harus sama dengan *data training*. Selanjutnya akan dilakukan perbandingan pada *data testing* terhadap data training yang telah menghasilkan *rule*. Hasil prediksi merupakan kategori kelulusan mahasiswa, yaitu tepat dan terlambat.

2.3 Populasi dan Sampel

Populasi dalam penelitian ini adalah seluruh mahasiswa Program Studi Teknologi Informasi, Fakultas Teknik, Universitas Udayana angkatan 2008-2013 sebanyak 537 mahasiswa. Sampel adalah mahasiswa dari program studi teknologi informasi yang sudah lulus angkatan 2008-2013 sebanyak 363 mahasiswa.

3. Kajian Pustaka

Kajian pustaka merupakan tinjauan pustaka yang berisi landasan teori secara umum yang dijadikan acuan dan pendukung dari pengerjaan laporan ini.

3.1 Data Mining

Data mining merupakan suatu proses atau cara dalam menemukan pola tersembunyi sehingga ditemukan suatu pengetahuan (*knowledge*) baru yang belum pernah diketahui sebelumnya dari proses penggalan informasi dari tumpukan data yang sangat banyak.

3.2 Metode Naïve Bayes

Algoritma *naïve bayes* digunakan untuk klasifikasi dengan menggunakan probabilitas untuk memprediksi peluang di masa depan berdasarkan pengalaman yang ada sebelumnya.

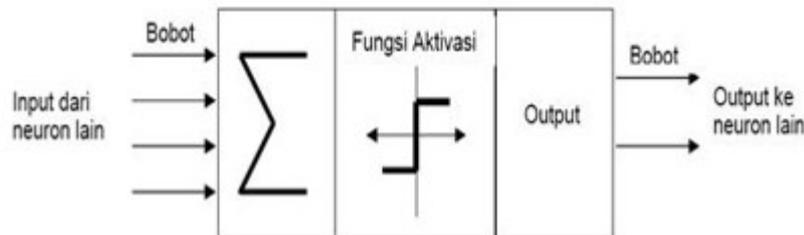
$$P(C|X) = \frac{P(x|c)P(c)}{P(x)} \quad (1)$$

Keterangan :

- X : kejadian x
- C : kelas yang tersedia
- $P(c|x)$: probabilitas kemunculan kelas c dengan kondisi x (*posterior probability*)
- $P(c)$: probabilitas kemunculan kelas c (*prior probability*)
- $P(x|c)$: probabilitas kemunculan kejadian x dengan kondisi kelas c (*likelihood probability*)
- $P(x)$: probabilitas kemunculan kejadian x (*evidence probability*)

3.3 Metode Neural Network

Artificial Neural Network atau disebut dengan Jaringan Syaraf Tiruan adalah sebuah teknik pengolahan informasi yang diambil dari cara kerja otak manusia untuk memproses informasi.



Gambar 3 Tahapan Proses Neural Network [8]

Tahapan proses algoritma neural network dimulai dari *layer input* yang berfungsi untuk menerima *input* yang berisi bobot. *Input* akan masuk ke dalam hidden layer yang berisi neuron untuk melakukan penjumlahan yang disimbolkan dengan sigma (Σ) seperti pada gambar diatas. Selanjutnya hasil penjumlahan akan dibandingkan dengan nilai ambang (*threshold*) pada fungsi aktivasi, apabila hasil penjumlahan lebih kecil dari *threshold* maka aktivasi neuron akan diaktifkan begitu pula sebaliknya. Neuron yang aktif akan mengirimkan nilai bobotnya ke *layer* selanjutnya sampai mendapatkan nilai pada *layer output*.

3.4 Root Mean Square Error (RMSE)

Root Mean Square Error merupakan sebuah metode untuk mengukur *error* dari hasil prediksi. Nilai dari RMSE adalah 0 hingga tak terbatas, semakin mendekati angka 0 suatu nilai *error* maka hasil prediksi semakin akurat.

Persamaan untuk menghitung nilai RMSE adalah sebagai berikut.

$$RMSE = \sqrt{\frac{\sum (y - y')^2}{n}} \quad (2)$$

Keterangan :

- Σ : penjumlahan dari hasil pengurangan
- y : nilai aktual
- y' : nilai prediksi
- n : banyaknya data

4. Hasil dan Pembahasan

Pembahasan dan hasil dari implementasi metode data mining menggunakan algoritma *naïve bayes* dan *neural network* dibagi menjadi 4 kali pengujian. Berikut merupakan hasil pengujian dari implementasi *data mining*.

4.1. Pengujian Algoritma Naïve Bayes

Evaluasi hasil uji algoritma naïve bayes digunakan untuk mencari akurasi terbaik dari pengujian metode data mining. Hasil uji pertama hingga uji keempat akan dibandingkan nilai akurasinya.

Tabel 1. Hasil Uji Algoritma Naïve Bayes

| Hasil Uji | Data Training | | Data Testing | | Akurasi |
|-------------|---------------|----------------|--------------|----------------|---------|
| | Tahun | Jumlah Records | Tahun | Jumlah Records | |
| Uji Pertama | 2008-2009 | 108 | 2010-2013 | 255 | 54,51% |
| Uji Kedua | 2008-2010 | 162 | 2011-2013 | 201 | 55,22% |
| Uji Ketiga | 2008-2011 | 240 | 2012-2013 | 123 | 59,35% |
| Uji Keempat | 2008-2012 | 309 | 2013 | 54 | 55,56% |

Hasil uji algoritma *naïve bayes* dengan akurasi terbaik yaitu pada uji ketiga menggunakan *data training* tahun 2008-2011 dan *data testing* tahun 2012-2013 sebesar 59,35%. Akurasi terendah pengujian yaitu uji pertama dengan menggunakan *data training* tahun 2008-2009 dan *data testing* tahun 2010-2013 sebesar 54,51%.

Tabel 2. *Confusion Matrix* Uji Terbaik Naïve Bayes

| Akurasi: 59,35% | | |
|-----------------|----------------|------------|
| | True Terlambat | True Tepat |
| Pred. Terlambat | 13 | 5 |
| Pred. Tepat | 45 | 60 |

Tabel 2 merupakan tabel *confusion matrix* yang menunjukkan bahwa kelas *True Positive* berjumlah 60, kelas *True Negative* berjumlah 13, kelas *False Positive* berjumlah 45, dan kelas *False Negative* berjumlah 5.

4.2. Pengujian Algoritma Neural Network

Evaluasi hasil uji algoritma neural network digunakan untuk mencari akurasi terbaik dari pengujian metode data mining. Hasil uji pertama hingga uji keempat akan dibandingkan nilai akurasi ke dalam bentuk tabel.

Tabel 3. Hasil Uji Algoritma Neural Network

| Hasil Uji | Data Training | | Data Testing | | Epoch | Hidden Layer | Akurasi | RMSE |
|-------------|---------------|----------------|--------------|----------------|-------|--------------|---------|-------|
| | Tahun | Jumlah Records | Tahun | Jumlah Records | | | | |
| Uji Pertama | 2008-2009 | 108 | 2010-2013 | 255 | 800 | 1 | 48,63% | 0,697 |
| Uji Kedua | 2008-2010 | 162 | 2011-2013 | 201 | 1800 | 1 | 57,71% | 0,612 |
| Uji Ketiga | 2008-2011 | 240 | 2012-2013 | 123 | 200 | 2 | 60,16% | 0,531 |
| Uji Keempat | 2008-2012 | 309 | 2013 | 54 | 1500 | 2 | 59,26% | 0,510 |

Hasil uji algoritma *neural network* dengan akurasi terbaik yaitu pada uji ketiga menggunakan *data training* tahun 2008-2011 dan *data testing* tahun 2012-2013 sebesar

60,16%. Akurasi terendah pengujian yaitu uji pertama dengan menggunakan *data training* tahun 2008-2009 dan *data testing* tahun 2010-2013 sebesar 48,63%.

Tabel 4. *Confusion Matrix* Uji Terbaik Neural Network

| Akurasi: 60,16% | | |
|-----------------|----------------|------------|
| | True Terlambat | True Tepat |
| Pred. Terlambat | 20 | 11 |
| Pred. Tepat | 38 | 54 |

Tabel 4 merupakan tabel *confusion matrix* yang menunjukkan bahwa kelas *True Positive* berjumlah 54, kelas *True Negative* berjumlah 20, kelas *False Positive* berjumlah 38, dan kelas *False Negative* berjumlah 11.

4.3 Hasil Analisa Kedua Algoritma

Hasil pengujian algoritma *naïve bayes* serta algoritma *neural network* menggunakan hasil akurasi terbaik akan dilakukan komparasi ke dalam bentuk tabel. Komparasi kedua algoritma untuk mencari algoritma terbaik berdasarkan hasil uji pertama hingga hasil uji keempat. Hasil komparasi kemudian divisualisasikan ke dalam bentuk grafik untuk memudahkan dalam memperoleh *knowledge*.

Tabel 5. Komparasi Kedua Algoritma

| Hasil Uji | Data Training | | Data Testing | | Akurasi | |
|-------------|---------------|----------------|--------------|----------------|-------------|----------------|
| | Tahun | Jumlah Records | Tahun | Jumlah Records | Naïve Bayes | Neural Network |
| Uji Pertama | 2008-2009 | 108 | 2010-2013 | 255 | 54,51% | 48,63% |
| Uji Kedua | 2008-2010 | 162 | 2011-2013 | 201 | 55,22% | 57,71% |
| Uji Ketiga | 2008-2011 | 240 | 2012-2013 | 123 | 59,35% | 60,16% |
| Uji Keempat | 2008-2012 | 309 | 2013 | 54 | 55,56% | 59,26% |

Tabel 5 merupakan tampilan dari hasil uji kedua algoritma dari uji pertama hingga keempat. Tampilan hasil akurasi algoritma *neural network* menggunakan tingkat akurasi terbaik dari hasil uji pertama hingga keempat. Perbandingan hasil uji pertama pada kedua algoritma dengan akurasi terbaik adalah algoritma *naïve bayes* sebesar 54,51%. Perbandingan hasil uji kedua pada kedua algoritma dengan akurasi terbaik adalah algoritma *neural network* sebesar 57,71%. Perbandingan hasil uji ketiga pada kedua algoritma dengan akurasi terbaik adalah algoritma *neural network* sebesar 60,16%. Perbandingan hasil uji keempat pada kedua algoritma dengan akurasi terbaik adalah algoritma *neural network* sebesar 59,26%.

5. Kesimpulan

Komparasi dari metode *data mining* yang digunakan menunjukkan bahwa algoritma *neural network* memiliki tingkat akurasi lebih besar dari algoritma *naïve bayes* yaitu sebesar 60,16% dengan RMSE 0,531 pada pengujian ketiga. Berdasarkan hasil analisa pencarian atribut optimal, jumlah SKS merupakan atribut yang paling berpengaruh berdasarkan perhitungan *weight by information gain*, *information gain ratio*, dan *weight by correlation* IPS.

Referensi

- [1] N. A. Abidin, M. Assidiq and A. Qaslim, "Sistem Prediksi Kelulusan Mahasiswa dengan Metode Backpropagation Neural Network," *Jurnal Ilmiah Maju*, vol. 4, no. 2, 2021.
 - [2] E. P. Rohmawan, "Prediksi Kelulusan Mahasiswa Tepat Waktu Menggunakan Metode Decision Tree dan Artificial Neural Network," *Jurnal Ilmiah Matrik*, vol. 20, no. 1, 2018.
 - [3] R. H. Lubis and P. Kustanto, "Implementasi Algoritma Neural Network dalam Memprediksi Tingkat Kelulusan Mahasiswa," *Jurnal Media Informatika Budidarma*, vol. 4, no. 2, 2020.
 - [4] N. Purwati, R. Nurlistiani and O. Devinsen, "Data Mining dengan Algoritma Neural Network dan Visualisasi Data untuk Prediksi Kelulusan Mahasiswa," *Jurnal Informatika*, vol. 20, no. 2, 2020.
 - [5] M. D. Yalidhan and M. F. Amin, "Implementasi Algoritma Backpropagation untuk Memprediksi Kelulusan Mahasiswa," *Kumpulan Jurnal Ilmu Komputer (KLIK)*, vol. 5, no. 2, 2018.
 - [6] A. F. Firdaus, R. Saedudin and R. Andeswari, "Implementasi Metode Klasifikasi Naive Bayes dalam Memprediksi Kelulusan Mahasiswa," *e-Proceeding of Engineering*, vol. 8, no. 5, p. 9274, 2021.
 - [7] A. A. A. Purnamaswari, I. K. G. D. Putra and I. M. S. Putra, "Komparasi Metode Neural Network Backpropagation dan Support Vector Machines dalam Prediksi Volume Sampah TPA Suwung," *Jitter*, vol. 3, no. 1, 2022.
 - [8] S. M. Derwin Suhartono, " Binus Universitas School of Computer Science," Binus Universitas, 26 July 2012. [Online]. Available: <https://socs.binus.ac.id/2012/07/26/konsep-neural-network/>. [Accessed 1 November 2019].
-