

Perbandingan Klasifikasi antara *Naives Bayes* dan *Decision Tree* dalam Prediksi Penyakit Diabetes Tahap Awal

Akbar Wibowo Putra¹, Kevin Kosumo², Ayu Sitho Resmy Ratu³, Radik Rosyadi Mujayanto⁴, Muhammad Rafly⁵, Melati Mahandani Mintarum⁶, Vivine Nurcahyawati⁷

Program Studi S1 Sistem Informasi,
Fakultas Teknologi dan Informatika, Universitas Dinamika
Jl. Raya Kedung Baruk No.98, Kedung Baruk, Kec. Rungkut, Kota Surabaya, Jawa Timur
120410100057@dinamika.ac.id; 220410100071@dinamika.ac.id;
320410100051@dinamika.ac.id; 420410100074@dinamika.ac.id;
520410100078@dinamika.ac.id; 620410100079@dinamika.ac.id
7vivine@dinamika.ac.id(Corresponding author)

Abstract

Diabetes is a health condition characterized by an elevated blood glucose level. There are two types, namely diabetes type 1 and diabetes type 2. Diabetes type 1 is caused by a lack of insulin production by the pancreas. Symptoms of diabetes include excessive thirst, frequent urination, and constant hunger. Classification is a process that helps us group data or information into categories based on similar characteristics. In the context of diabetes, classification methods can be used to group individuals based on their risk levels of developing diabetes. By using classification methods, doctors can determine an individual's risk of diabetes and design an appropriate treatment plan. This study involves a comparison between the Naïve Bayes and Decision Tree methods. The results of this research indicate that the algorithm generated is the best among the two algorithms in identifying diabetes patients. An accuracy of 66.67% was obtained for Naïve Bayes, while an accuracy of 91.67% was obtained for Decision Tree. In this study, it was found that the Decision Tree method has a higher accuracy rate than the Naïve Bayes method in the case study and data testing.

Keywords: *Diabetes, Classification, Naïve Bayes, Decision Tree, ID3, Data Mining*

1. Pendahuluan

Diabetes adalah penyakit kronis yang ditandai oleh tingkat kadar gula darah (glukosa) dalam tubuh yang tinggi. Hal ini terjadi dikarenakan ketidakmampuan tubuh untuk menggunakan atau menghasilkan insulin yang cukup. Insulin merupakan hormon yang diperlukan untuk mengubah glukosa dalam makanan menjadi energi yang dapat digunakan oleh tubuh. Terdapat beberapa gejala umum diabetes, antara lain sering merasa haus, lapar, sering buang air kecil, kehilangan berat badan secara tiba-tiba, luka yang sulit sembuh, rasa kantuk berlebihan, serta kesulitan melihat. Namun, gejala diabetes dapat berbeda-beda dari satu individu ke individu lainnya, dan ada kemungkinan bahwa beberapa orang tidak mengalami gejala apa pun. Diabetes Mellitus telah menjadi sumber kekhawatiran yang signifikan bagi masyarakat. Penyakit ini terkait dengan berbagai komplikasi serius, termasuk risiko penyakit jantung koroner, gagal ginjal, stroke, kebutaan, dan bahkan kematian [1]. Data yang berasal dari Kementerian Kesehatan melalui *Sample Registration Survey* 2014 menunjukkan bahwa diabetes merupakan penyebab kematian tertinggi ketiga di Indonesia, mencapai 6,7% dari total kematian, berada di belakang stroke dengan 21,1% dan penyakit jantung koroner dengan 12,9%. Data yang berasal dari *International Diabetes Federation* (IDF) pada tahun 2017, Indonesia menempati peringkat ke-6 di dunia dengan jumlah penduduk yang menderita diabetes sebanyak 10,3 juta jiwa. Namun, jika tidak ditangani dengan baik, *World Health Organization* (WHO) bahkan memproyeksikan bahwa prevalensi diabetes di Indonesia dapat meningkat drastis menjadi 21,3 juta jiwa pada tahun 2030. Berdasarkan data dari LITBANG KEMENKES, tercatat peningkatan jumlah kasus baru penyakit

diabetes sebesar 56,82%. Pada tahun 2018, jumlah total penderita penyakit diabetes mencapai 230.871 orang, dengan jumlah kasus terbanyak terdapat di Pulau Jawa, yaitu sebanyak 93.478 kasus [2]. Dengan banyaknya jumlah kasus diabetes di Indonesia dan perkiraan dari WHO yang mana pasien diabetes akan melonjak drastis pada tahun 2030, oleh karena itu diperlukan langkah-langkah awal yang efektif untuk melakukan prediksi dan deteksi dini penyakit diabetes. Diabetes dapat diprediksi dengan menggunakan data pasien diabetes yang terdokumentasi dalam sebuah database. Data tersebut dapat diolah dengan algoritma tertentu untuk menghasilkan informasi yang dapat digunakan untuk diagnosis awal penyakit diabetes [3]. Teknik analisa secara manual yang digunakan tidaklah lagi efektif untuk mendiagnosa penyakit diabetes. Seiring dengan perkembangan waktu, semakin pentingnya penggunaan sistem pengetahuan berbasis komputer sebagai alat analisis dalam proses diagnosis penyakit adalah sebuah tuntutan yang mendesak.

Pada era saat ini, teknologi komputer telah membuktikan kemampuannya dalam membantu pendeteksian penyakit secara akurat, yang juga dapat menghemat waktu dalam proses diagnosis. *Data Mining* sebagai bidang penting dalam ilmu komputer, memiliki peran krusial dalam memprediksi berbagai fenomena. Ini melibatkan proses penemuan informasi baru dari data yang sudah ada melalui analisis data yang cermat [4]. Salah satu cara pengelompokan sesuatu berdasarkan model yang dimiliki yaitu dengan proses Klasifikasi. Klasifikasi membangun model pada *data training* dan digunakan untuk mengklasifikasikan data baru [5]. Algoritma klasifikasi memiliki beberapa macam yakni *Decision Tree*, *Naïve Bayes*, dan *K-means*. Pada penelitian lain juga menggunakan algoritma lainnya seperti *Naïve Bayes*, Regresi Logistik dan *Random Forest*, klasifikasi melalui Regresi, *Bayes Net* [7]. Namun, hingga saat ini belum ada kepastian mengenai algoritma mana yang paling akurat dalam memprediksi penyakit diabetes.

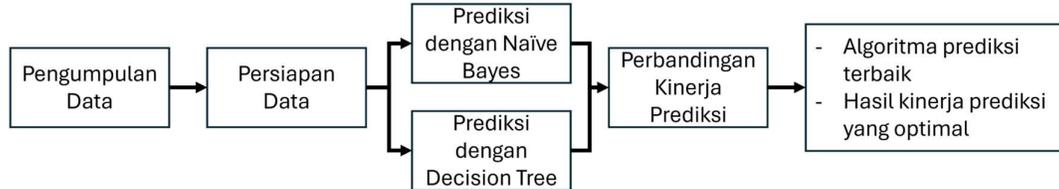
Decision Tree memang merupakan salah satu teknik klasifikasi yang sangat populer. Pada *Decision Tree*, terdapat simpul akar yang diberi nama atribut, dan dari simpul akar tersebut terdapat cabang-cabang yang menggambarkan nilai-nilai atribut. Setiap simpul daun dalam pohon tersebut mewakili berbagai kelas yang berbeda, sehingga memungkinkan pengelompokan data ke dalam kelas yang sesuai berdasarkan kondisi yang ada pada cabang-cabang pohon [6]. Pohon keputusan dibangun secara manual atau dapat tumbuh secara otomatis dengan penerapan *Decision Tree* untuk pemodelan yang belum diklasifikasi pada himpunan data. *Iterative Dichotomiser 3* atau yang biasa dikenal dengan ID3 merupakan algoritma pada *data mining* dengan tujuan klasifikasi yang mendasar. Pada algoritma ini melakukan pencarian komprehensif pada semua potensi struktur pohon keputusan. Algoritma ID3 menggunakan konsep dari entropi informasi [8].

Selanjutnya yang menjadi algoritma yang cukup populer dalam klasifikasi *data mining* yaitu algoritma *Naïve Bayes*. *Naïve Bayes* juga termasuk ke dalam statistik klasifikasi yang dapat memprediksi probabilitas kelas keanggotaan. *Naïve Bayes* berdasarkan prinsip teori Bayes. Pada *Naïve Bayes*, digunakan teknik prediksi berbasis probabilitas sederhana yang didasarkan pada penerapan teorema Bayes dengan asumsi independensi yang kuat [9]. Pengklasifikasi *Naïve Bayes* adalah pengklasifikasi probabilistik sederhana yang mengaplikasikan teorema Bayesian dengan asumsi independensi yang kuat [10]. Penggunaan algoritma *Naïve Bayes* dan *Decision Tree* dalam *data mining* sebagai algoritma alternatif untuk mendiagnosa penyakit diabetes bisa menjadi pilihan yang tepat. Oleh karena itu, dalam penelitian ini, akan dilakukan perbandingan antara *Naïve Bayes* dan *Decision Tree* untuk menentukan model dengan tingkat akurasi terbaik. Hal ini diharapkan akan meningkatkan akurasi hasil prediksi secara signifikan.

Evaluasi dan validasi hasil klasifikasi akan menggunakan matriks *confusion matrix*. Penggunaan *confusion matrix* dalam evaluasi dapat menghasilkan nilai *precision*, *recall*, dan *accuracy* yang dapat membantu dalam menentukan hasil dari perbandingan algoritma. *Precision* merupakan pengukur proporsi kasus yang diprediksi sebagai positif yang juga benar-benar positif pada data yang sebenarnya. *Recall* merupakan penunjuk beberapa persen data kategori positif yang diklasifikasikan secara benar. *Accuracy* merupakan pengukur persentase ketepatan *data training* yang diklasifikasikan secara benar setelah dilakukan pengujian pada hasil klasifikasi [11].

2. Metode Penelitian

Pada penelitian kali ini menggunakan beberapa tahap yaitu: pengumpulan data, persiapan data, prediksi dengan Naive Bayes dan Decision Tree, perbandingan kinerja prediksi, dan evaluasi hasil kinerja prediksi yang optimal seperti terlihat pada Gambar 1.



Gambar 1. Tahapan Penelitian

Data yang digunakan dalam penelitian ini diperoleh dari repositori yang dapat diakses melalui *Kaggle*. Data ini dikumpulkan melalui pengisian kuesioner langsung oleh pasien dari Rumah Sakit Diabetes Sylhet di Sylhet, Bangladesh, dan pengumpulan data ini telah disetujui oleh dokter [12]. Dataset yang digunakan merupakan dataset primer terdiri atas 520 data pasien dan 17 atribut yang terdiri dari:

a. *Age* → umur

Penderita diabetes tidak mengenal umur, baik itu masih anak-anak maupun lansia dapat mengalami diabetes bila pola hidup yang tidak sehat, terutama pengonsumsian gula, karbohidrat, dan lain-lain. yang berlebihan dan tak diatur dengan baik. Umur merupakan salah satu faktor paling umum dalam diabetes [13].

b. *Gender* → jenis kelamin

Seperti halnya dengan umur, diabetes tidak memandang jenis kelamin, sehingga baik laki-laki maupun perempuan dapat mengalami penyakit diabetes.

c. *Polyuria* → sering buang air kecil

Sel-sel dalam tubuh penderita diabetes mengalami kesulitan dalam menyerap glukosa, sehingga ginjal mencoba untuk mengeluarkan sebanyak mungkin glukosa dari tubuh. Glukosa juga memiliki kemampuan untuk menarik cairan ke dalam air kemih, yang dapat menyebabkan volume air kemih menjadi berlebihan dan mengakibatkan penderita sering buang air kecil [14]. Ditandai dengan kondisi urin yang mengandung glukosa ± 180 mg/dL, serta *osmolaritas filtrat glomerulus* yang meningkat dan reabsorpsi air yang terhambat oleh *tubulus* ginjal [15].

d. *Polydipsia* → sering haus

Dengan kehilangan banyak cairan tubuh karena sering buang air kecil, penderita akan sering merasa haus dan membutuhkan banyak minum. Kehausan yang terus-menerus dapat mendorong penderita untuk mengonsumsi lebih banyak cairan [14].

e. *Sudden weight loss* → penurunan berat badan

Kadar gula darah yang tinggi juga dapat menyebabkan penurunan berat badan yang signifikan. Penderita diabetes dapat mengeluarkan sekitar 500 gram glukosa dalam urin per 24 jam yang setara dengan kehilangan sekitar 2000 kalori per hari [16].

f. *Weakness* → melemah

Kadar gula darah tinggi menyebabkan sering terjadi rasa tidak enak badan. Sering buang air di malam hari merupakan hal yang melelahkan untuk tubuh. Akibatnya orang cenderung mudah tersinggung.

g. *Polyphagia* → sering makan/mudah lapar

Rasa lapar yang berlebihan juga merupakan tanda lain dari diabetes. Ketika kadar gula darah turun, tubuh dapat mengira bahwa belum diberi makan dan akhirnya menginginkan glukosa tambahan untuk memenuhi kebutuhan sel-selnya. Hal ini juga ditandai dengan

merasa kurang bertenaga. Permasalahan pada insulin pada penderita diabetes sehingga gula yang masuk ke dalam sel tubuh kurang dan kurang terbentuknya energi [16].

- h. *Genital thrush* → terjaid sariawan/infeksi jamur pada area kelamin
Pada daerah genital memiliki glukosa yang tinggi dalam urin membuat daerah genital terjadi sariawan atau infeksi jamur, dan berakibat pembengkakan dan gatal.
- i. *Visual blurring* → pengelihatan memburam
Kadar gula darah yang tinggi dapat mengakibatkan penglihatan kabur atau sesekali melihat kilatan cahaya. Jika gula darah tidak terkendali dalam jangka waktu yang lama, dapat menyebabkan kerusakan mata yang permanen, bahkan berisiko mengakibatkan kebutaan.
- j. *Itching* → gatal-gatal pada kulit
Kulit kering seringkali dapat menjadi tanda peringatan adanya penyakit diabetes, seperti juga kondisi kulit lainnya dan perubahan warna kulit yang menggelap di sekitar leher atau ketiak.
- k. *Irritability* → iritabilitas
Penderita diabetes juga memiliki risiko yang lebih tinggi terhadap berbagai jenis infeksi, walaupun yang paling umum adalah infeksi jamur seperti candida dan jenis infeksi jamur lainnya. Kelebihan gula dalam tubuh dapat menciptakan lingkungan yang mendukung pertumbuhan jamur dan bakteri.
- l. *Delayed healing* → penyembuhan yang melambat
Salah satu gejala yang umum pada diabetes yaitu lambannya proses penyembuhan pada infeksi, luka, dan memar. Kerusakan pada pembuluh darah dan arteri dapat terjadi karena adanya kadar glukosa yang berlebihan dalam tubuh yang mengelilingi pembuluh darah tersebut.
- m. *Partial paresis* → melemahnya gerak badan atau terjadi gangguan gerakan pada anggota tubuh
Kondisi sering buang air kecil dapat terjadi baik pada pagi maupun malam hari, penderita diabetes harus sering bergerak ke kamar mandi, dan hal tersebut yang menyebabkan melemahnya tubuh apalagi sudah terjadi komplikasi pada tubuh.
- n. *Muscle stiffness* → otot kaku
Kesemutan, mati rasa pada otot tangan dan kaki, serta rasa sakit yang membakar atau bengkak merupakan tanda-tanda bahwa saraf dapat mengalami kerusakan akibat diabetes. Jika kadar gula darah tidak terkendali dalam jangka waktu yang lama, kerusakan saraf ini dapat menjadi permanen.
- o. *Alopecia* → kerontokan rambut
Penderita diabetes juga dapat kehilangan/kerontokan rambut lebih dari biasanya. Rambut rontok tidak hanya terjadi di kepala, namun juga di lengan, kaki dan bagian tubuh lainnya. Biasanya proses pertumbuhan rambut akan lebih lambat daripada biasanya.
- p. *Obesity* → berat badan berlebih/obesitas
Karena penderita diabetes mudah merasakan lapar, bila pola makan yang tidak diatur maka penderita diabetes akan lebih rentan dengan obesitas karena konsumsi makanan yang berlebih. Terutama pengonsumsi glukosa, gula, karbohidrat yang berlebih akan menyebabkan diabetes menjadi lebih parah. Obesitas merupakan tanda utama yang menunjukkan ke dalam keadaan pradiabetes [13].
- q. *Class* → hasil akhir apakah positif diabetes atau tidak.

Berikut ini dalam Tabel 1 adalah merupakan contoh data dari dataset yang digunakan dalam penelitian ini. Dalam tabel tersebut mendaftar data dari 520 penderita diabetes berdasarkan 17 atribut yang telah dijelaskan sebelumnya. Isi data dalam data mentah ini belum keseluruhan dapat langsung diproses dalam klasifikasi, sehingga dibutuhkan pembersihan data.

Tabel 1. Dataset

No	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	40	Male	No	Yes	No	Yes	No	No	No	Yes	No	Yes	No	Yes	Yes	Yes	Positive
2	58	Male	No	No	No	Yes	No	No	Yes	No	No	No	Yes	No	Yes	No	Positive
3	41	Male	Yes	No	No	Yes	Yes	No	No	Yes	No	Yes	No	Yes	Yes	No	Positive
4	45	Male	No	No	Yes	Yes	Yes	Yes	No	Yes	No	Yes	No	No	No	No	Positive
5	60	Male	Yes	Yes	Yes	Yes	Yes	No	Yes	Positive							
6	55	Male	Yes	Yes	No	Yes	Yes	No	Yes	Yes	No	Yes	No	Yes	Yes	Yes	Positive
7	57	Male	Yes	Yes	No	Yes	Yes	Yes	No	No	No	Yes	Yes	No	No	No	Positive
8	66	Male	Yes	Yes	Yes	Yes	No	No	Yes	Yes	Yes	No	Yes	Yes	No	No	Positive
9	67	Male	Yes	Yes	No	Yes	Yes	Yes	No	Yes	Yes	No	Yes	Yes	No	Yes	Positive
10	70	Male	No	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	No	No	No	Yes	No	Positive
11	44	Male	Yes	Yes	No	Yes	No	Yes	No	No	Yes	Yes	No	Yes	Yes	No	Positive
12	38	Male	Yes	Yes	No	No	Yes	Yes	No	Yes	No	Yes	No	Yes	No	No	Positive
13	35	Male	Yes	No	No	No	Yes	Yes	No	No	Yes	Yes	No	No	Yes	No	Positive
14	61	Male	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	No	No	No	Yes	Yes	Positive
...
52	0	42	Male	No	Negative												

Pada tahap awal, untuk memastikan kualitas data yang baik akan dilakukan proses pembersihan data melalui beberapa langkah, yang mencakup:

1. Menghapus data yang dianggap tidak lengkap (*missing value*).
2. Membuat kamus data untuk melakukan konversi pada kolom "Age" guna untuk pengelompokan data dan mempermudah dalam penerapan pada algoritma yang digunakan.
3. Merapikan kembali dataset yang digunakan guna untuk mempermudah dalam penerapan algoritma dan membaca dataset.

Dalam penelitian ini, kami akan membandingkan dua algoritma *data mining* dengan tujuan klasifikasi, yaitu *Decision Tree* dan *Naïve Bayes*. Selanjutnya dilakukan evaluasi menggunakan *confusion matrix* untuk menentukan algoritma yang paling baik dalam memprediksi penyakit diabetes dan mengukur tingkat akurasi kedua algoritma yang digunakan. Penggunaan alat bantuan berupa *RapidMiner* untuk mengeksekusi algoritma *Naïve Bayes* dan *Decision Tree*, serta untuk menganalisis hasil antara lain nilai *precision*, *recall*, dan *accuracy*.

Algoritma *Naïve Bayes* adalah mengklasifikasi statistik yang didasarkan pada teorema Bayes. Algoritma ini berfungsi sebagai alat untuk mengambil keputusan berdasarkan informasi yang tersedia. Teorema Bayes dapat dituliskan dalam rumus 1 sebagai berikut [3]:

$$P(H|E) = \frac{P(E|H) \times P(H)}{P(E)} \tag{1}$$

Algoritma *Decision Tree* yang digunakan yaitu ID3. *Iterative Dichotomiser 3* atau yang biasa dikenal dengan ID3 merupakan algoritma pada *data mining* dengan tujuan klasifikasi yang mendasar. Pada algoritma ini melakukan pencarian komprehensif pada semua potensi struktur pohon keputusan. Algoritma ID3 menggunakan konsep dari entropi informasi [8]. Langkah-langkah dalam penerapan algoritma ID3 sebagai berikut [17]:

1. Menghitung *Information Gain* dari setiap atribut dilakukan dengan menggunakan rumus 2 dan 3 berikut:

$$Gain(S, A) = Entropy(S) - \sum_{v \in Nilai(A)} \frac{|S_v|}{|S|} Entropy(S_v) \tag{2}$$

Di mana:

$$Entropy(S) = -P_+ \log_2 P_+ - P_- \log_2 P_- \tag{3}$$

2. Memilih atribut yang memiliki nilai *Information Gain* terbesar.
3. Membentuk simpul yang berisi atribut tersebut.
4. Proses perhitungan *Information Gain* akan diulang terus menerus sampai semua data telah termasuk dalam kelas yang sama. Atribut yang telah dipilih sebelumnya tidak akan lagi diikuti dalam perhitungan nilai *Information Gain*.

3. Hasil dan Pembahasan

Berikut adalah hasil dan pembahasan yang diperoleh dari proses pengolahan data menggunakan algoritma *Naïve Bayes* dan *Decision Tree*. Adapun *data testing* yang digunakan adalah *data testing* yang berjumlah 12 orang seperti pada Tabel 2, yaitu:

Tabel 2. Data Testing

No	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	Dewasa	Male	No	Yes	No	Yes	No	No	No	Yes	No	Yes	No	Yes	Yes	Yes	Positive
2	Dewasa	Female	Yes	No	Yes	Yes	Yes	No	No	No	No	Yes	No	No	No	No	Positive
3	Lansia	Male	Yes	Yes	Yes	Yes	No	No	Yes	Yes	Yes	No	Yes	Yes	No	No	Positive
4	Lansia	Female	Yes	No	Yes	Yes	Yes	No	No	Positive							
5	Muda	Female	No	No	No	No	No	No	Yes	No	No	No	Yes	Yes	No	No	Positive
6	Muda	Female	No	No	No	No	No	No	Yes	No	No	No	Yes	Yes	No	No	Positive
7	Dewasa	Male	No	No	No	Yes	No	No	No	No	No	No	Yes	Yes	No	No	Negative
8	Dewasa	Female	No	Negative													
9	Lansia	Male	Yes	No	No	No	Yes	No	Yes	Yes	No	Yes	Yes	Yes	Yes	No	Negative
10	Lansia	Female	No	No	No	Yes	No	No	No	Yes	No	Yes	No	No	Yes	No	Negative
11	Muda	Male	No	Negative													
12	Muda	Female	No	No	No	No	Yes	No	No	No	No	No	Yes	No	No	No	Negative

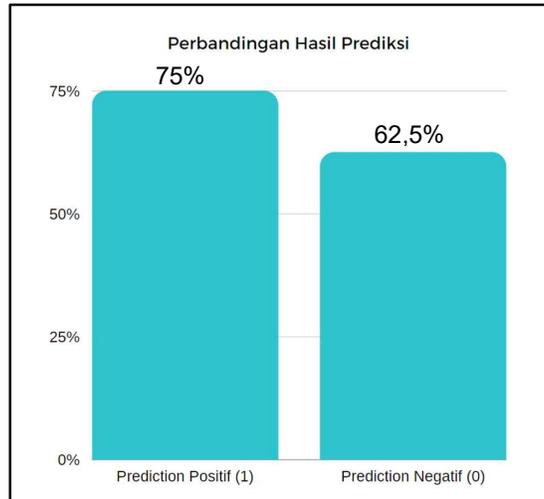
3.1. Hasil Perhitungan Algoritma Naïve Bayes

Berikut ini adalah hasil kinerja prediksi menggunakan algoritma *Naïve Bayes* dengan menggunakan aplikasi *RapidMiner* dengan *data testing* berjumlah 12 dan dataset berjumlah 520 data yang sudah dilakukan *cleansing* dan pengelompokan.

Tabel 3. Hasil Perhitungan Naïve Bayes

Accuracy: 66,67%			
	true Positive	True Negative	class precision
pred. Positive	3	1	75.00%
pred. Negative	3	5	62.50%
class recall	50.00%	83.33%	

Pada Tabel 3 menunjukkan hasil perhitungan menggunakan algoritma *Naïve Bayes* dengan tingkat *accuracy* sebesar 66,67%. Adapun *class precision* dari *Prediction Positive* (1) adalah 75% dan *precision* dari *Prediction Negative* (0) adalah sebesar 62,5% seperti terlihat pada Gambar 2 berikut.



Gambar 2. Perbandingan Hasil Prediksi Naïve Bayes

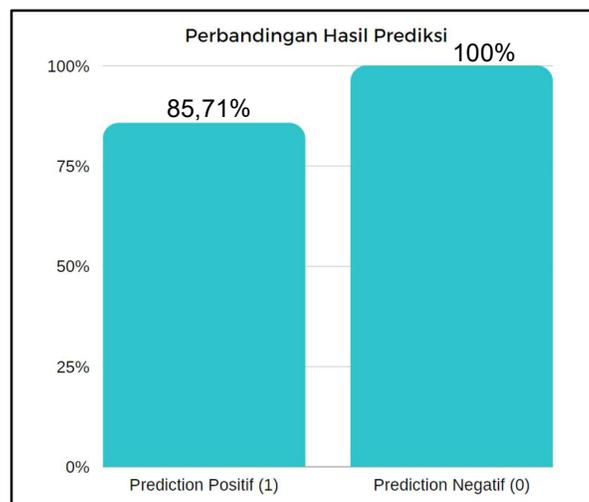
3.2. Hasil Perhitungan Algoritma *Decision Tree*

Berikut ini adalah hasil kinerja prediksi menggunakan algoritma *Decision Tree* dengan algoritma ID3, dan menggunakan aplikasi *RapidMiner* dengan *data testing* berjumlah 12 dan dataset berjumlah 520 data yang sudah dilakukan *cleansing* dan pengelompokan.

Tabel 4. Hasil Perhitungan *Decision Tree*

Accuracy: 91,67%			
	true Positive	True Negative	class precision
pred. Positive	6	1	85.71%
pred. Negative	0	5	100%
class recall	100%	83.33%	

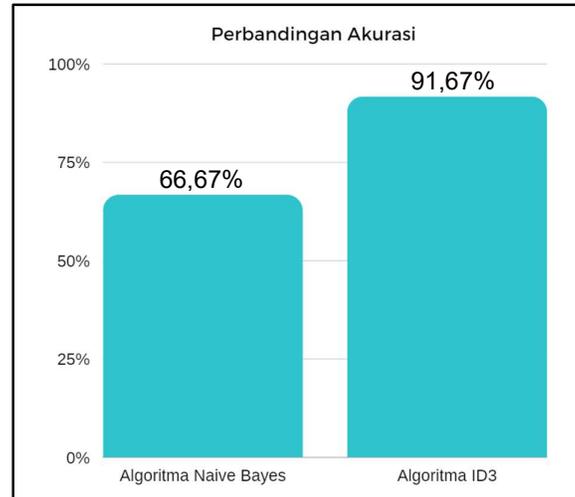
Pada Tabel 4 menunjukkan hasil perhitungan menggunakan algoritma *Decision Tree* dengan algoritma ID3 dengan tingkat *accuracy* sebesar 91,67%. Adapun *class precision* dari *Prediction Positive* (1) adalah 85,71% dan *precision* dari *Prediction Negative* (0) adalah sebesar 100% seperti terlihat pada Gambar 3 berikut.



Gambar 3. Perbandingan Hasil Prediksi *Decision Tree*

3.3. Hasil Perbandingan Akurasi Algoritma *Naïve Bayes* dan *Decision Tree*

Setelah selesai mengolah data dengan kedua algoritma, hasil dari akurasi kedua algoritma tersebut di bandingkan dengan tujuan menemukan hasil yang lebih tinggi tingkat akurasinya. Hasil perbandingan kedua algoritma tersebut dapat dilihat pada Gambar 4 di bawah ini.



Gambar 4. Perbandingan Akurasi Algoritma *Naïve Bayes* dan ID3

Adapun dari Gambar 4 dapat diketahui bahwa perhitungan dengan algoritma *Naïve Bayes* memiliki akurasi yang lebih rendah yaitu 66,67% jika dibandingkan dengan perhitungan algoritma ID3 yaitu dengan akurasi sebesar 91,67%. Hasil tersebut menunjukkan bahwa algoritma ID3 lebih unggul jika digunakan pada data yang telah digunakan dalam penelitian ini.

4. Kesimpulan

Pada penelitian ini telah melakukan klasifikasi penderita diabetes dengan algoritma *Naïve Bayes* dan *Decision Tree*. Data yang digunakan adalah sebanyak 520 dengan 12 data uji. Setelah mengimplementasikan algoritma *Naïve Bayes* dan *Decision Tree* untuk mengklasifikasi penderita Diabetes dengan mempertimbangkan hasilnya, dapat disimpulkan bahwa algoritma *Decision Tree* memiliki tingkat akurasi yang lebih tinggi. Hal ini dapat ditemukan melalui hasil pengolahan data dari kedua algoritma, yaitu algoritma *Naïve Bayes* memiliki tingkat akurasi sebesar 66,67% dan algoritma ID3 sebesar 91,67%. Faktor-faktor yang dapat memengaruhi tingkat akurasi tersebut meliputi jumlah dataset yang digunakan dan jumlah data testing yang diterapkan.

Daftar Pustaka

- [1] I. Alfarobi, "Sistem Pakar Deteksi Dini Gejala Awal Diabetes Mellitus," (*Indonesian J. Comput. Inf. Technol.*, vol. 4, no. 1, pp. 67–72, 2019, [Online]. Available: https://ejournal.bsi.ac.id/ejurnal/index.php/ijcit/article/download/5307/pdf_1.
- [2] R. Dwi, Y. Prakoso, and D. R. Sari, "Perancangan Sistem Klasifikasi Penyakit Diabetes Menggunakan Algoritma *Naïve Bayes*," *Perwira J. Sci. Eng.*, vol. 02, no. 01, pp. 24–31, 2022.
- [3] B. A. C. Permana and I. K. Dewi, "Komparasi Metode Klasifikasi Data Mining Decision Tree dan *Naïve Bayes* Untuk Prediksi Penyakit Diabetes," *Infotek J. Inform. dan Teknol.*, vol. 4, no. 1, pp. 63–69, 2021, doi: 10.29408/jit.v4i1.2994.
- [4] A. Ridwan, "Penerapan Algoritma *Naïve Bayes* Untuk Klasifikasi Penyakit Diabetes Mellitus," *J. SISKOM-KB (Sistem Komput. dan Kecerdasan Buatan)*, vol. 4, no. 1, pp. 15–21, 2020, doi: 10.47970/siskom-kb.v4i1.169.
- [5] B. Delvika, S. Nurhidayarnis, and P. D. Rinada, "Perbandingan Klasifikasi Antara *Naive Bayes* dan K-Nearest Neighbor Terhadap Resiko Diabetes Pada Ibu Hamil," *Indones. J. Mach. Learn. Comput. Sci.*, vol. 2, no. October, pp. 68–75, 2022.

- [6] A. H. Nasrullah, "Implementasi Algoritma Decision Tree Untuk Klasifikasi Data Peserta Didik," *J. Pilar Nusa Mandiri*, vol. 7, no. 2, p. 217, 2021.
- [7] R. Annisa, "Analisis Komparasi Algoritma Klasifikasi Data Mining Untuk Prediksi Penderita Penyakit Jantung," *J. Tek. Inform. Kaputama*, vol. 3, no. 1, pp. 22–28, 2019, [Online]. Available: <https://jurnal.kaputama.ac.id/index.php/JTIK/article/view/141/156>.
- [8] N. Nurdiana and A. Algifari, "Studi Komparasi Algoritma Id3 Dan Algoritma Naive Bayes Untuk Klasifikasi Penyakit Diabetes Mellitus," *INFOTECHjournal*, vol. 6, no. 2, pp. 18–23, 2020.
- [9] H. D. Wijaya and S. Dwiasnati, "Implementasi Data Mining dengan Algoritma Naïve Bayes pada Penjualan Obat," *J. Inform.*, vol. 7, no. 1, pp. 1–7, 2020, doi: 10.31311/ji.v7i1.6203.
- [10] F. Fitriyani, "Prediksi Diabetes Menggunakan Algoritma Naive Bayes dan Greedy Forward Selection," *J. Nas. Teknol. dan Sist. Inf.*, vol. 7, no. 2, pp. 61–69, 2021, doi: 10.25077/teknosi.v7i2.2021.61-69.
- [11] B. P. Pratiwi, A. S. Handayani, and S. Sarjana, "Pengukuran Kinerja Sistem Kualitas Udara Dengan Teknologi Wsn Menggunakan Confusion Matrix," *J. Inform. Upgris*, vol. 6, no. 2, pp. 66–75, 2021, doi: 10.26877/jiu.v6i2.6552.
- [12] I. Dutta, "Early Stage Diabetes Risk Prediction Dataset," *kaggle*, 2020. <https://www.kaggle.com/datasets/ishandutta/early-stage-diabetes-risk-prediction-dataset> (accessed Dec. 22, 2022).
- [13] I. Rahmasari and E. S. Wahyuni, "Efektivitas Memordoca Carantia (Pare) terhadap Penurunan Kadar Glukosa Darah," *Infokes*, vol. 9, no. 1, pp. 57–64, 2019.
- [14] D. A. Suntara, "PENGARUH KULIT SALAK (SALACA ZALACCA) TERHADAP GLUKOSA DARAH," *J. Ilm. Hosp.*, vol. 12, no. 1, 2023, [Online]. Available: <https://repositorio.ufsc.br/xmlui/bitstream/handle/123456789/167638/341506.pdf?sequence=1&isAllowed=y%0Ahttps://repositorio.ufsm.br/bitstream/handle/1/8314/LOEBLEIN%2C%20LUCINEIA%20CARLA.pdf?sequence=1&isAllowed=y%0Ahttps://antigo.mdr.gov.br/saneamento/proees>.
- [15] D. Hardianto, "TELAH KOMPREHENSIF DIABETES MELITUS KLASIFIKASI, GEJALA, DIAGNOSIS, PENCEGAHAN, DAN PENGOBATAN," *J. Bioteknol. Biosains Indones.*, vol. 7, no. 2, pp. 304–317, 2020, [Online]. Available: <http://ejurnal.bppt.go.id/index.php/JBBI>.
- [16] Lestari, Zulkarnain, and S. A. Sijid, "Diabetes Melitus: Review Etiologi, Patofisiologi, Gejala, Penyebab, Cara Pemeriksaan, Cara Pengobatan dan Cara Pencegahan," *UIN Alauddin Makassar*, no. November, pp. 237–241, 2021, [Online]. Available: <http://journal.uin-alauddin.ac.id/index.php/psb>.
- [17] D. Sismai Siahaan and A. Nuariana Sebayang, "Penetapan Mahasiswa Penerima Beasiswa Menggunakan Decision Tree Pada Strmik Logika," *J. Merdeka Inform.*, vol. 1, no. 1, pp. 7–15, 2023.