

## KOMPARASI ALGORITMA DECISION TREE DAN NAIVE BAYES DALAM KLASIFIKASI PENYAKIT DIABETES

Rizky Maulana, Riri Narasati, Ruli Herdiana, Ryan Hamonangan, Saeful Anwar.

Teknik Informatika, STMIK IKMI Cirebon

Jalan Perjuangan No 10 B Karyamulya Kesambi Kota Cirebon

rizkym71338@gmail.com

### ABSTRAK

Rata-rata nilai akurasi klasifikasi penyakit diabetes adalah sebesar 82%. Meskipun demikian, nilai tersebut masih kurang baik. Karena nilai akurasi tersebut masih dibawah rata-rata yaitu 85%. Masalah utama muncul karena tingkat akurasi yang kurang baik dapat menimbulkan risiko kesalahan yang signifikan dalam mengidentifikasi penyakit diabetes. Kecenderungan nilai rata-rata yang rendah ini menunjukkan adanya hambatan dalam keandalan sistem klasifikasi penyakit diabetes. Penelitian ini menerapkan metode *Knowledge Discovery in Databases* (KDD). Adalah sebuah teknik yang dipakai untuk mengeksplorasi dan memperoleh pemahaman yang mendalam pada suatu *database*. Tujuan utama pada tahapan KDD adalah menghasilkan informasi yang dapat digunakan sebagai landasan untuk pengambilan keputusan. Hasil penelitian menandakan bahwa *Naive Bayes* mencapai tingkat keakuratan yang lebih baik sebesar 91.56%. Sedangkan, *Decision Tree* memiliki tingkat akurasi 87.01%. Dari evaluasi ini, dapat disimpulkan bahwa *Naive Bayes* menghasilkan kinerja yang lebih unggul dalam klasifikasi penyakit diabetes pada dataset yang digunakan dibandingkan dengan *Decision Tree*.

**Kata kunci:** *Diabetes, Data Mining, Decision Tree, Naive Bayes.*

### 1. PENDAHULUAN

Diabetes adalah penyakit kronis yang menyebabkan peningkatan kadar gula darah dan dapat disebabkan oleh gaya hidup tidak sehat, termasuk kegemukan, pola makan tinggi lemak dan kalori, serta kurangnya aktivitas fisik. Menurut *International Diabetes Federation* (2021), di tahun 2021, terdapat 531 juta manusia dewasa di umur 20–73 tahun yang menderita diabetes, dengan perkiraan peningkatan menjadi 643 juta pada 2030 dan 783 juta pada 2045. Di Asia Tenggara, terdapat 90 juta penderita diabetes. Diabetes juga menjadi penyebab 6.7 juta kematian pada tahun 2021 setiap 5 detik dan menyebabkan pengeluaran kesehatan mencapai setidaknya 966 miliar dolar Amerika (USD), dengan peningkatan sebesar 316% dalam 15 tahun terakhir [1].

Meskipun telah dilakukan banyak penelitian menggunakan *machine learning* untuk mengklasifikasikan diabetes, rata-rata akurasi masih kurang baik, yaitu sekitar 82%. Ini menciptakan risiko kesalahan yang signifikan dalam identifikasi penyakit. Oleh karena itu, penelitian ini memfokuskan pada perbandingan di antara metode *Decision Tree* dan *Naive Bayes* dalam meningkatkan akurasi pengklasifikasi diabetes.

Beberapa penelitian sebelumnya telah mengeksplorasi penerapan algoritma *machine learning* dalam klasifikasi penyakit diabetes. Penelitian oleh Afif tahun 2020 berfokus pada klasifikasi diabetes mellitus menggunakan dataset dari Rumah Sakit Aisyah dengan 17 atribut [2]. Penelitian oleh Wahyu tahun 2023 berfokus pada klasifikasi diabetes dengan menerapkan algoritma *Decision Tree* [3]. Hasil penelitian Habibah tahun

2023 menunjukkan akurasi sebesar 76,67%, presisi 72%, dan *recall* 41,67% dalam klasifikasi penyakit diabetes [4].

Penelitian ini bertujuan meningkatkan akurasi klasifikasi penyakit diabetes dengan membandingkan algoritma *Decision Tree* dan *Naive Bayes*. Tujuan perbandingan adalah untuk menentukan algoritma yang paling sesuai dalam meningkatkan akurasi diagnosis diabetes. Penelitian ini akan mengevaluasi efektivitas kedua algoritma tersebut, dengan fokus utama pada pengukuran dan perbandingan tingkat akurasi.

Hasilnya ditujukan kepada dokter spesialis endokrin, apoteker, dan perawat. Dokter dapat meningkatkan diagnosis dan manajemen penyakit dengan memanfaatkan temuan penelitian. Apoteker dapat memberikan dukungan terkait obat-obatan kepada pasien berdasarkan informasi hasil penelitian. Perawat dapat menggunakan hasil penelitian untuk mendukung perawatan dan edukasi pasien tentang manajemen pola hidup dan pengobatan diabetes. Selain itu, diharapkan bahwa peningkatan nilai akurasi hasil penelitian akan memberikan dampak positif dalam upaya pencegahan dan manajemen penyakit diabetes.

### 2. TINJAUAN PUSTAKA

Pada tahap ini, dijelaskan kajian literatur yang sesuai dengan penelitian yang tengah dilakukan saat ini.

#### 2.1. Data Mining

*Data mining* merupakan langkah kompleks yang melibatkan penghimpunan, pengolahan, dan pemanfaatan dataset berukuran luas dengan tujuan

menemukan aturan atau pola yang dapat disimpan dalam basis data atau media penyimpanan lainnya. Proses ini mencakup langkah-langkah analisis mendalam menggunakan metode-metode seperti *machine learning* untuk mengidentifikasi hubungan-hubungan yang mungkin tidak terlihat secara langsung. Hasil temuan dari data mining memberikan pemahaman yang lebih mendalam tentang informasi yang terkandung dalam data besar tersebut, dan dapat diterapkan dalam konteks pengambilan keputusan atau perencanaan strategis [5] [6].

**2.2. Klasifikasi**

Klasifikasi adalah suatu teknik dalam pengolahan data yang melibatkan pembagian objek yang digunakan ke dalam kelompok atau kelas-kelas, di mana jumlah kelas dapat disesuaikan dengan kebutuhan analisis. Tujuan utama dari proses klasifikasi ini adalah untuk menciptakan pola atau aturan yang memungkinkan pemisahan yang jelas antara setiap kelompok data. Dengan demikian, hasil klasifikasi dapat memberikan informasi tentang keanggotaan suatu objek dalam suatu kategori tertentu, berdasarkan karakteristik, perilaku, dan atribut yang telah ditentukan sebelumnya untuk setiap kelompok tersebut. Proses ini membantu dalam pemahaman dan interpretasi data, serta memfasilitasi pengambilan keputusan berdasarkan pengelompokan yang telah didefinisikan sebelumnya [7] [8].

**2.3. Decision Tree**

*Decision Tree* merupakan suatu metode dalam *machine learning* yang menggunakan sampel latihan, yakni sekumpulan data, untuk membangun struktur pohon keputusan. Proses ini melibatkan serangkaian keputusan berbasis aturan-aturan yang dibuat untuk mengklasifikasikan data ke dalam kelompok atau kategori tertentu. Setelah pohon keputusan dibangun, ia kemudian diuji untuk memastikan kebenarannya dalam melakukan prediksi atau klasifikasi terhadap data yang belum pernah terlihat sebelumnya. Metode tersebut sering digunakan dalam tugas-tugas klasifikasi dan prediksi, serta dapat memberikan pemahaman yang intuitif terhadap faktor-faktor yang mempengaruhi pengambilan keputusan dalam suatu model [9].

**2.4. Naïve Bayes**

*Naïve Bayes* merupakan metode dalam *machine learning* yang bukan mengikuti pedoman tertentu dan didasarkan pada bidang matematika, yaitu konsep probabilitas. Metode ini menggunakan pendekatan probabilitas untuk menentukan klasifikasi suatu data, memanfaatkan peluang terbesar dalam kemungkinan klasifikasi tersebut. Dalam prosesnya, *Naïve Bayes* menghitung frekuensi masing-masing klasifikasi dengan menggunakan data pelatihan. Pendekatan ini memungkinkan model untuk memperoleh pemahaman tentang kemungkinan hasil klasifikasi berdasarkan peluang munculnya atribut-atribut

tertentu dalam data latihan. *Naïve Bayes* sering digunakan dalam tugas klasifikasi teks, pengelompokan dokumen, dan masalah klasifikasi lainnya. [10].

**2.5. Confusion Matrix**

*Confusion Matrix* adalah teknik dalam bidang *data mining* yang secara umum dipakai untuk menilai kinerja suatu algoritma atau model pengklasifikasi. Metode ini memberikan informasi mengenai tingkat akurasi, presisi, dan tingkat kesalahan dari algoritma tersebut. Representasi umum dari *Confusion Matrix* adalah tabel yang menampilkan keempat hasil klasifikasi yang mungkin. Baris matriks mencerminkan jumlah data uji untuk kelas yang sebenarnya, sementara kolom mencerminkan kelas prediksi [11].

Tabel 1. *Confusion matrix*

		Kelas Sebenarnya	
		True	False
Kelas Prediksi	True	TP	FP
	False	FN	TN

Tabel 1 adalah *Confusion Matrix* yang digunakan sebagai sarana evaluasi hasil klasifikasi dari suatu model atau algoritma. Dalam matriks ini, empat hasil klasifikasi yang mungkin dapat diidentifikasi.

- a. *True Positive* (TP) mencerminkan jumlah data yang berhasil diidentifikasi secara tepat sebagai kelas positif.
- b. *True Negative* (TN) merujuk pada jumlah data yang berhasil diidentifikasi secara tepat sebagai kelas negatif.
- c. *False Positive* (FP) merujuk pada jumlah data yang seharusnya termasuk ke dalam kelas negatif, namun keliru diidentifikasi sebagai kelas positif.
- d. *False Negative* (FN) mencakup jumlah data yang seharusnya masuk ke dalam kelas positif, tetapi keliru diidentifikasi sebagai kelas negatif.

Pengujian yang melibatkan *Confusion Matrix* menghasilkan sejumlah perhitungan, antara lain yaitu:

- a. Akurasi adalah ukuran yang digunakan untuk mengevaluasi sejauh mana kemampuan suatu model dalam mengklasifikasikan data secara tepat secara keseluruhan. Berikut ini merupakan formula untuk menghitung nilai akurasi.

$$Akurasi = \frac{TP+TN}{TP+TN+FP+FN} \tag{1}$$

- b. Presisi merupakan metrik yang digunakan untuk mengevaluasi proporsi prediksi yang positif yang benar dibandingkan dengan keseluruhan prediksi yang positif, dengan tujuan untuk mengurangi kesalahan positif palsu. Berikut merupakan formula untuk mengkalkulasi nilai presisi.

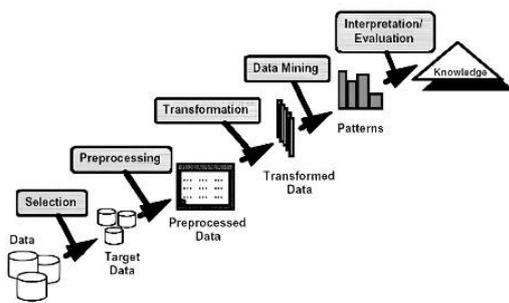
$$\text{Presisi} = \frac{TP}{TP + FP} \tag{2}$$

- c. *Recall* adalah metrik yang digunakan untuk mengukur seberapa banyak prediksi positif yang benar dibandingkan dengan total data aktual yang positif, dengan tujuan untuk mengurangi kesalahan negatif palsu. Berikut ini merupakan formula untuk menghitung nilai *recall*.

$$\text{Recall} = \frac{TP}{TP + FN} \tag{3}$$

### 3. METODE PENELITIAN

Penelitian ini akan menerapkan metode *Knowledge Discovery in Databases (KDD)*. Merupakan suatu metode yang dimanfaatkan untuk memperoleh informasi berharga dari suatu basis data. Tujuan akhir dari metode ini adalah untuk menyajikan materi yang dapat dipertimbangkan dalam pengambilan keputusan [12]. Berikut adalah tahapan-tahapan dalam penggunaan metode *KDD*.



Gambar 1. Langkah penelitian [13]

Gambar 1 menggambarkan langkah-langkah penelitian meliputi Pemilihan, Pra-pemrosesan, Transformasi, *Data Mining*, dan Interpretasi / Evaluasi.

#### 3.1. Selection

Pada langkah ini, dilakukan seleksi data yang akan diterapkan dalam studi ini. Di samping itu, dilakukan pemilihan variabel yang akan dijadikan label dengan dukungan dari seorang pakar di bidang penyakit diabetes.

#### 3.2. Preprocessing

Pada langkah ini, dilakukan serangkaian langkah untuk memastikan kualitas dataset. Proses ini melibatkan penghapusan data yang tidak lengkap dan identifikasi serta eliminasi data ganda guna menjamin akurasi hasil analisis. Reduksi fitur juga diterapkan dengan menghapus atribut atau kolom yang dianggap tidak relevan untuk mencapai tujuan analisis, sehingga dataset menjadi lebih sederhana tanpa kehilangan informasi kunci. Selain itu, penanganan data yang hilang menjadi fokus utama, di mana upaya

dilakukan untuk mengisi nilai yang kosong atau menghapus entri yang tidak lengkap. Dengan melakukan pembersihan data ini secara cermat, diharapkan dapat menghasilkan dataset yang bersih, lengkap, dan siap digunakan dalam proses analisis dengan kedua metode tersebut.

#### 3.3. Transformation

Pada langkah ini, data diubah menjadi format yang optimal guna mengikuti proses, menggunakan kedua metode tersebut. Perubahan ini melibatkan penyesuaian struktur data agar sesuai dengan persyaratan dan karakteristik kedua algoritma tersebut. Perubahan ini bertujuan untuk memastikan kualitas data yang memadai sehingga algoritma dapat memberikan hasil yang akurat dan informatif dalam tahap selanjutnya. Dengan merinci format data yang sesuai, proses analisis dengan *Decision Tree* dan *Naïve Bayes* dapat dilakukan secara efisien, memberikan dukungan untuk membuat keputusan yang lebih baik dan memperoleh pemahaman yang lebih mendalam terhadap informasi yang tersimpan dalam dataset.

#### 3.4. Data Mining

Pada langkah ini, terjadi proses penting dalam penerapan algoritma *Decision Tree* dan *Naïve Bayes*, yaitu pemisahan dataset menjadi dua komponen utama, yaitu dataset untuk pelatihan dan pengujian. Pemisahan ini bertujuan untuk melakukan pelatihan pada model dengan memanfaatkan data pelatihan dan menguji kinerja model dengan menggunakan data pengujian. Proses pelatihan model melibatkan penggunaan algoritma *Decision Tree* dan *Naïve Bayes* untuk memahami pola dan aturan yang terkandung dalam data pelatihan. Setelah model dilatih, langkah selanjutnya adalah menguji performa model menggunakan data pengujian untuk mengevaluasi sejauh mana model dapat mengklasifikasikan data yang belum pernah dikenali sebelumnya dengan tingkat ketepatan yang tinggi. Proses ini menjadi langkah kritis dalam memastikan keandalan model dan mengidentifikasi apakah algoritma yang digunakan dapat memberikan hasil yang memuaskan dalam konteks klasifikasi dataset yang diberikan.

#### 3.5. Interpretation / Evaluation

Pada tahap ini, hasil dari penerapan algoritma *Decision Tree* dan *Naïve Bayes* dianalisis secara mendalam. Proses ini melibatkan penilaian kinerja kedua algoritma berdasarkan metrik-metrik evaluasi yang relevan, seperti akurasi, presisi dan *recall*. Evaluasi ini memberikan gambaran tentang seberapa baik kedua algoritma dapat mengklasifikasikan data, serta kecenderungan masing-masing dalam mengatasi *false positive* dan *false negative*. Selain itu, dilakukan perbandingan performa antara *Decision Tree* dan *Naïve Bayes* untuk menentukan metode yang lebih efektif dalam konteks dataset yang digunakan. Hasil

evaluasi ini menjadi dasar untuk mengidentifikasi kelebihan dan kekurangan masing-masing algoritma, serta menyusun rekomendasi terkait pemilihan algoritma yang paling sesuai untuk tujuan klasifikasi dataset yang spesifik. Evaluasi yang cermat pada tahap ini sangat penting untuk memastikan keberlanjutan dan kehandalan model pengklasifikasi.

**4. HASIL DAN PEMBAHASAN**

**4.1. Selection**

Dalam langkah ini, menggunakan dataset sekunder yang diperoleh dari *open source* yaitu dataset penyakit diabetes yang berasal dari situs <https://www.kaggle.com/datasets/akshaydattatraykha/re/diabetes-dataset> dengan jumlah 768 baris data dan 9 variabel, dimana dataset tersebut dalam bentuk format *Comma Separated Values (CSV)*. *Outcome* menjadi variabel yang berfungsi sebagai label dalam studi ini.

**4.2. Preprocessing**

Pada langkah ini, dilakukan proses *cleaning data* yang mencakup eliminasi beberapa variabel yang tidak relevan untuk keperluan studi ini. Variabel yang dihilangkan adalah “*Pregnancies*” dan “*Skin Thickness*”. Terlebih lagi, dataset yang dimanfaatkan dalam penelitian ini sudah lengkap dan tidak mengandung kekosongan data.

Tabel 2. *Preprocessing* penyakit diabetes

V1	V2	V3	V4	V5	V6	V7
148	72	0	33.6	0.627	50	1
85	66	0	26.6	0.351	31	0
183	64	0	23.3	0.672	32	1
89	66	94	28.1	0.167	21	0
137	40	168	43.1	2.288	33	1
116	74	0	25.6	0.201	30	0
78	50	88	31	0.248	26	0
115	0	0	35.3	0.134	29	0
197	70	543	30.5	0.158	53	1
125	96	0	0	0.232	54	0
110	92	0	37.6	0.191	30	0
168	74	0	38	0.537	34	1
139	80	0	27.1	1.441	57	1
189	60	846	30.1	0.398	59	1
166	72	175	25.8	0.587	51	1

Tabel 2 merupakan data penyakit diabetes setelah dilakukan *preprocessing* data yaitu pengurangan atribut “*Pregnancies*” dan “*Skin Thickness*”.

Keterangan:

- a. V1 = *Glucose*
- b. V2 = *Blood Pressure*
- c. V3 = *Insulin*
- d. V4 = *BMI*
- e. V5 = *Diabetes Pedigree Function*
- f. V6 = *Age*
- g. V7 = *Outcome*

**4.3. Transformation**

Pada langkah ini, dilakukan perubahan pada data atribut “*Outcome*” dengan mengubah representasinya dari bentuk angka 1 dan 0 menjadi “*yes*” dan “*no*”. Proses ini bertujuan untuk memberikan interpretasi yang lebih intuitif terhadap nilai atribut, mempermudah pemahaman hasil analisis, dan meningkatkan kejelasan makna dari atribut tersebut. Dengan mengganti representasi angka menjadi kata-kata, hasil analisis menjadi lebih mudah diinterpretasikan oleh pemangku kepentingan dan memfasilitasi komunikasi antara pengguna data dan model. Perubahan ini juga dapat memberikan nilai tambah pada pemahaman konteks khusus dari atribut “*Outcome*” dalam konteks dataset yang digunakan.

Tabel 3. *Transformation* penyakit diabetes

V1	V2	V3	V4	V5	V6	V7
148	72	0	33.6	0.627	50	<i>yes</i>
85	66	0	26.6	0.351	31	<i>no</i>
183	64	0	23.3	0.672	32	<i>yes</i>
89	66	94	28.1	0.167	21	<i>no</i>
137	40	168	43.1	2.288	33	<i>yes</i>
116	74	0	25.6	0.201	30	<i>no</i>
78	50	88	31	0.248	26	<i>no</i>
115	0	0	35.3	0.134	29	<i>no</i>
197	70	543	30.5	0.158	53	<i>yes</i>
125	96	0	0	0.232	54	<i>no</i>
110	92	0	37.6	0.191	30	<i>no</i>
168	74	0	38	0.537	34	<i>yes</i>
139	80	0	27.1	1.441	57	<i>yes</i>
189	60	846	30.1	0.398	59	<i>yes</i>
166	72	175	25.8	0.587	51	<i>yes</i>

Tabel 3 merupakan data penyakit diabetes setelah dilakukan perubahan data.

Keterangan:

- a. V1 = *Glucose*
- b. V2 = *Blood Pressure*
- c. V3 = *Insulin*
- d. V4 = *BMI*
- e. V5 = *Diabetes Pedigree Function*
- f. V6 = *Age*
- g. V7 = *Outcome*

**4.4. Data Mining**

Pada langkah pembuatan model, diterapkan dua algoritma klasifikasi utama, yaitu algoritma *Decision Tree* dan *Naïve Bayes*. Proses ini melibatkan penggunaan dataset yang telah melalui tahap pembersihan dan perubahan sebelumnya untuk melatih kedua model. Algoritma *Decision Tree* akan membangun pohon keputusan berdasarkan pola dan aturan yang terdapat dalam data latihan, sementara *Naïve Bayes* akan memanfaatkan teori probabilitas untuk menghasilkan model klasifikasi. Pemilihan kedua algoritma ini memungkinkan perbandingan performa keduanya dalam konteks dataset khusus yang digunakan, sehingga dapat dipilih algoritma

yang paling sesuai untuk tujuan klasifikasi penyakit diabetes pada penelitian ini.

**4.5. Interpretation / Evaluation**

Pada langkah evaluasi model, dilakukan analisis kinerja terhadap metode *Decision Tree* dan *Naive Bayes*. Evaluasi mencakup perhitungan akurasi, presisi, dan *recall* untuk setiap metode guna menilai seberapa baik keduanya dapat mengklasifikasikan data penyakit diabetes. Akurasi memberikan gambaran keseluruhan sejauh mana model dapat memberikan klasifikasi yang benar, sedangkan presisi dan *recall* memberikan *insight* terkait ketepatan dalam mengidentifikasi *instance* positif. Hasil evaluasi ini memungkinkan perbandingan performa kedua algoritma, sehingga dapat dipilih algoritma yang paling sesuai untuk meningkatkan akurasi dan ketepatan diagnosis penyakit diabetes pada penelitian ini.

Setelah menyelesaikan proses *data mining* dan memperoleh hasil klasifikasi melalui algoritma *Decision Tree* dan *Naive Bayes*, langkah berikutnya adalah mengevaluasi kinerja model dengan menghitung metrik-metrik seperti akurasi, presisi, dan *recall* menggunakan *Confusion Matrix*.

Tabel 4. *Confusion matrix decision tree*

	<i>True. Yes</i>	<i>True. No</i>
<i>Pred. Yes</i>	38	4
<i>Pred. No</i>	16	96

Berdasarkan hasil *Confusion Matrix* pada tabel 4 terlihat bahwa sejumlah 38 data diklasifikasikan sebagai "Yes" dengan benar, sementara 96 data diklasifikasikan sebagai "No" dengan benar.

Tabel 5. *Confusion matrix naïve bayes*

	<i>True. Yes</i>	<i>True. No</i>
<i>Pred. Yes</i>	46	5
<i>Pred. No</i>	8	95

Berdasarkan hasil *Confusion Matrix* pada tabel 5 terlihat bahwa sejumlah 46 data diklasifikasikan sebagai "Yes" dengan benar, sementara 95 data diklasifikasikan sebagai "No" dengan benar. Setelah mengetahui *Confusion Matrix* dari kedua algoritma, langkah berikutnya melibatkan evaluasi data dengan mengukur akurasi, presisi, dan *recall* dari metode *Decision Tree* dan *Naive Bayes*.

Tabel 6. Komparasi dua algoritma

Metode	Akurasi	Presisi	Recall
<i>Decision Tree</i>	87.01%	85.71%	96.00%
<i>Naive Bayes</i>	91.56%	92.23%	95.00%

Tabel 6 menunjukkan hasil uji perbandingan di antara dua metode, yakni *Decision Tree* dan *Naive Bayes*, dalam klasifikasi penyakit diabetes. Hasil uji

tersebut mengonfirmasi bahwa penggunaan algoritma *Naive Bayes* menghasilkan tingkat ketepatan yang lebih tinggi daripada metode *Decision Tree*, dengan akurasi mencapai 91.56%. Angka akurasi ini melebihi nilai rata-rata klasifikasi penyakit diabetes sebesar 82%. Hal ini mengindikasikan bahwa penerapan metode *Naive Bayes* dalam penelitian ini memberikan kinerja yang sangat baik dalam klasifikasi penyakit diabetes, mencapai tingkat akurasi yang signifikan.

**5. KESIMPULAN DAN SARAN**

Berdasarkan hasil penelitian komparasi metode *Decision Tree* dan *Naive Bayes* dalam mengklasifikasi penyakit diabetes, dua algoritma klasifikasi dievaluasi untuk memahami kinerja dalam klasifikasi penyakit diabetes. Dari hasil penelitian, terlihat bahwa *Naive Bayes* berhasil mencapai tingkat akurasi yang lebih tinggi sebesar 91,56%, sedangkan *Decision Tree* memiliki tingkat akurasi sebesar 87,01%. Berdasarkan evaluasi ini, dapat disimpulkan bahwa *Naive Bayes* memberikan kinerja yang lebih unggul dalam mengklasifikasikan penyakit diabetes pada dataset yang digunakan jika dibandingkan dengan *Decision Tree*. Tingginya akurasi pada *Naive Bayes* menunjukkan keahlian model ini dalam membedakan di antara kelas positif dan negatif dengan lebih baik daripada *Decision Tree*. Adapun saran-saran yang disampaikan berdasarkan hasil pengujian dalam penelitian ini adalah penelitian selanjutnya diharapkan melakukan eksplorasi model klasifikasi lainnya untuk lebih meningkatkan tingkat akurasinya dan penelitian selanjutnya diharapkan melakukan optimasi *hyperparameter* pada algoritma *Naive Bayes* untuk meningkatkan tingkat akurasinya.

**DAFTAR PUSTAKA**

- [1] M Ivan Mahdi, "Penderita Diabetes Indonesia Terbesar Kelima di Dunia," Data Indonesia. [Online]. Available: <https://dataindonesia.id/kesehatan/detail/penderita-diabetes-indonesia-terbesar-kelima-di-dunia>
- [2] A. Afif, "Penerapan Algoritma Naive Bayes Untuk Klasifikasi Penyakit Diabetes Mellitus di Rumah Sakit Aisyiah," *J. Ilmu Komput. dan Mat.*, vol. 1, no. 2, pp. 15–21, 2020, doi: 10.47970/siskom-kb.v4i1.169.
- [3] B. A. R. P. Wahyu, A. F. Farozzi, C. P. Mahendra, and R. K. Hapsari, "Klasifikasi Penderita Penyakit Diabetes Berdasarkan Decision Tree Menggunakan Algoritma C4.5," *INTEGER J. Inf. Technol.*, vol. 8, no. 1, pp. 80–89, 2023.
- [4] N. N. Habibah, A. Nazir, I. Iskandar, and ..., "Pemodelan Klasifikasi Untuk Menentukan Penyakit Diabetes dengan Faktor Penyebab Menggunakan Decision Tree C4. 5 Pada Wanita," *J. Sist. ...*, 2023, [Online]. Available: <http://ejurnal.stmik-budidarma.ac.id/index.php/JSON/article/view/6202>

- [5] R. R. Khaerullah, N. Suarna, and O. Nurdiawan, "Analisa Pengelompokan Dataset Komputer Menggunakan Algoritma X-Means," *J. Inform. dan Teknol. Inf.*, vol. 1, no. 2, pp. 125–132, 2023, doi: 10.56854/jt.v1i2.135.
- [6] M. R. Sulistio, N. Suarna, and O. Nurdiawan, "Analisa Penerapan Metode Clustering X-Means Dalam Pengelompokan Penjualan Barang," *J. Teknol. Ilmu Komput.*, vol. 1, no. 2, pp. 37–42, 2023, doi: 10.56854/jtik.v1i2.49.
- [7] R. P. Fadhillah, R. Rahma, and ..., "Klasifikasi Penyakit Diabetes Mellitus Berdasarkan Faktor-Faktor Penyebab Diabetes menggunakan Algoritma C4. 5," *JUPI (Jurnal ...)*, 2022, [Online]. Available: <https://jurnal.stkipppgritulungagung.ac.id/index.php/jupi/article/view/3248>
- [8] O. N. Washilaturrizqi, "Implementasi Algoritma C4. 5 Untuk Menentukan Penerima," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 7, no. 1, pp. 373–377, 2023.
- [9] 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.
- [10] N. Nurahman and P. Prihandoko, "Perbandingan Hasil Analisis Teknik Data Mining 'Metode Decision Tree, Naive Bayes, Smo Dan Part' Untuk Mendiagnosa Penyakit Diabetes Mellitus," *J. Inf.* [pdfs.semanticscholar.org](https://pdfs.semanticscholar.org/2019/cb36/1875687463178d3c0bada1f496572a37201c.pdf), 2019. [Online]. Available: <https://pdfs.semanticscholar.org/cb36/1875687463178d3c0bada1f496572a37201c.pdf>
- [11] N. F. Arminda *et al.*, "IMPLEMENTASI ALGORITMA MULTINOMIAL NAIVE BAYES PADA ANALISIS SENTIMEN TERHADAP ULASAN PENGGUNA APLIKASI BRIMO," vol. 7, no. 3, pp. 1817–1822, 2023.
- [12] T. Prasetya, I. Ali, C. L. Rohmat, and O. Nurdiawan, "Klasifikasi Status Stunting Balita Di Desa Slangit Menggunakan Metode K-Nearest Neighbor," *INFORMATICS Educ. Prof. J. Informatics*, vol. 5, no. 1, p. 93, 2020, doi: 10.51211/itbi.v5i1.1431.
- [13] R. K. Purnomo, H. A. Saufi, and R. Hs, "Implementasi Data Mining Untuk Memprediksi Kelulusan Siswa SMK Al Huda Kedungwungu Dengan Menggunakan Algoritma Naïve Bayes Classifier Putriyana?1," *Exp. Student Exp.*, vol. 1, no. 1, pp. 1–7, 2023.