

# **PERBANDINGAN ALGORITMA MACHINE LEARNING DALAM KLASIFIKASI PENYAKIT DIABETES**

**Disusun guna memenuhi tugas mata kuliah  
Pembelajaran Mesin**

Dosen Pengampu:  
**Arif Rifai Dwiyanto, ST., MTI**



Oleh:

Adhwa Ardista Khoirunisa	202310715146
Sumih	202310715145
Putri Adelia Azizah	202310715084

**F5A2**

**UNIVERSITAS BHAYANGKARA JAKARTA RAYA  
FAKULTAS ILMU KOMPUTER  
PROGRAM STUDI INFORMATIKA  
JAKARTA**

**2026**

## KATA PENGANTAR

Kami panjatkan puji dan syukur ke hadirat Tuhan Yang Maha Esa atas rahmat dan karunia-Nya, sehingga kami dapat menyelesaikan laporan tugas yang berjudul “Perbandingan Algoritma Machine Learning dalam Klasifikasi Penyakit Diabetes” dengan baik.

Adapun tujuan dari penulisan laporan ini adalah untuk memenuhi tugas pada mata kuliah yang bersangkutan. Selain itu, laporan ini bertujuan untuk menambah wawasan mengenai penerapan algoritma *machine learning* dalam klasifikasi penyakit diabetes.

Kami mengucapkan terima kasih kepada Bapak Arif Rifai Dwiyanto, ST., MTI, selaku dosen pengampu mata kuliah, yang telah memberikan bimbingan dan tugas sehingga laporan ini dapat tersusun dengan baik serta menambah pemahaman mengenai penerapan machine learning dalam pengolahan data.

Kami menyadari bahwa laporan yang kami susun ini masih memiliki keterbatasan dan jauh dari kata sempurna. Oleh karena itu, kami mengharapkan kritik dan saran yang membangun demi perbaikan dan penyempurnaan laporan ini di masa mendatang. Kami mohon maaf apabila terdapat kesalahan maupun kekurangan dalam penyusunan laporan ini.

Jakarta, 17 Januari 2026

Penyusun

## DAFTAR ISI

<b>KATA PENGANTAR .....</b>	<b>i</b>
<b>DAFTAR ISI .....</b>	<b>ii</b>

### **BAB I PENDAHULUAN**

1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	1
1.3 Tujuan Penelitian .....	2

### **BAB II DASAR TEORI**

2.1 Diabetes Melitus .....	3
2.2 Machine Learning .....	3
2.3 Algoritma yang Digunakan.....	4
2.3.1 Naïve Bayes Classifier .....	4
2.3.2 K-Nearest Neighbor (KNN).....	5
2.3.3 Decision Tree .....	5

### **BAB III METODOLOGI PENELITIAN**

3.1 Dataset .....	6
3.2 Tahapan Penelitian.....	6
3.3 Preprocessing Data .....	7
3.4 Confusion Matrix.....	7

### **BAB IV HASIL DAN PEMBAHASAN**

4.1 Evaluasi Model Naïve Bayes.....	9
4.2 Evaluasi Model Nearest Neighbor (KNN) .....	10
4.3 Evaluasi Model Decision Tree.....	12

### **BAB V PENUTUP**

5.1 Kesimpulan.....	14
5.2 Saran .....	14

### **DAFTAR PUSTAKA**

# **BAB I**

## **PENDAHULUAN**

### **1.1 Latar Belakang**

Permasalahan kesehatan saat ini tidak hanya berkaitan dengan pengobatan, tetapi juga pada ketepatan dan kecepatan dalam proses diagnosis. Data kesehatan yang terus bertambah setiap tahunnya perlu dimanfaatkan secara optimal agar dapat memberikan informasi yang berguna dalam pengambilan keputusan medis.

Pemanfaatan teknik data mining dan machine learning menjadi salah satu solusi untuk mengolah data kesehatan dalam jumlah besar. Melalui proses klasifikasi, data pasien dapat dianalisis untuk memprediksi kondisi kesehatan berdasarkan atribut tertentu. Pendekatan ini diharapkan mampu membantu proses deteksi penyakit secara lebih sistematis dan berbasis data.

Dalam penelitian ini, dilakukan penerapan beberapa algoritma klasifikasi, yaitu Naive Bayes, K-Nearest Neighbors (KNN), dan Decision Tree, untuk mengklasifikasikan data penyakit diabetes. Ketiga algoritma tersebut dipilih karena memiliki karakteristik dan cara kerja yang berbeda, sehingga dapat dibandingkan performanya berdasarkan hasil evaluasi model. Hasil dari penelitian ini diharapkan dapat memberikan gambaran mengenai metode klasifikasi yang paling efektif untuk digunakan pada data penyakit diabetes.

### **1.2 Rumusan Masalah**

Berdasarkan latar belakang yang telah dijelaskan, maka rumusan masalah dalam penelitian ini adalah sebagai berikut:

1. Bagaimana proses klasifikasi penyakit diabetes menggunakan algoritma Naive Bayes, KNN, dan Decision Tree?
2. Bagaimana performa masing-masing algoritma dalam mengklasifikasikan data penyakit diabetes berdasarkan nilai akurasi dan metrik evaluasi lainnya?
3. Algoritma klasifikasi manakah yang memberikan hasil terbaik dalam penelitian ini?

### **1.3 Tujuan Penelitian**

Tujuan dari penelitian ini adalah sebagai berikut:

1. Menerapkan algoritma Naive Bayes, KNN, dan Decision Tree untuk klasifikasi penyakit diabetes.
2. Mengetahui dan membandingkan performa dari masing-masing algoritma klasifikasi.
3. Menentukan algoritma terbaik berdasarkan hasil evaluasi model.

## **BAB II**

### **DASAR TEORI**

#### **2.1 Diabetes Melitus**

Diabetes Melitus merupakan salah satu masalah kesehatan yang masih banyak terjadi dan jumlah penderitanya terus meningkat, baik di Indonesia maupun di dunia. Penyakit ini termasuk penyakit metabolik kronis yang disebabkan oleh gangguan pada hormon insulin, yaitu ketika pankreas tidak mampu memproduksi insulin dalam jumlah yang cukup atau ketika tubuh tidak dapat menggunakan insulin secara optimal. Padahal, insulin berperan penting dalam mengatur kadar gula darah di dalam tubuh.

Akibat gangguan tersebut, kadar gula darah dapat meningkat dan menyebabkan kondisi yang disebut hiperglikemia. Jika hiperglikemia terjadi secara terus-menerus, maka dapat mengganggu proses metabolisme karbohidrat, lemak, dan protein. Kondisi ini umumnya disebabkan oleh kerusakan sel beta pada pankreas atau menurunnya sensitivitas sel tubuh terhadap insulin. Apabila tidak ditangani dengan baik, diabetes melitus dapat menimbulkan berbagai komplikasi jangka panjang, terutama pada saraf dan pembuluh darah.

Dalam beberapa tahun terakhir, peningkatan kasus diabetes melitus menjadi perhatian serius karena penyakit ini tidak hanya menyerang orang dewasa, tetapi juga mulai banyak ditemukan pada anak-anak dan remaja. Selain berdampak pada penurunan kualitas hidup penderitanya, diabetes melitus juga dapat meningkatkan risiko kematian akibat komplikasi yang ditimbulkan. Oleh karena itu, diperlukan upaya deteksi dini dan pengelolaan penyakit yang tepat agar risiko komplikasi dapat diminimalkan.

#### **2.2 Machine Learning**

Machine Learning (ML) adalah bagian dari Artificial Intelligence (AI) yang memungkinkan sistem komputer untuk belajar dari data secara mandiri tanpa perlu diprogram secara eksplisit untuk setiap tugasnya. Prinsip kerjanya mengadopsi kemampuan belajar manusia, yaitu mengenali pola dan mengambil keputusan

berdasarkan pengalaman atau data historis. Agar dapat bekerja optimal, algoritma ML memerlukan data training sebagai bahan pembelajaran sebelum mampu menghasilkan prediksi atau keputusan.

Dalam implementasinya, ML sering digunakan untuk tugas klasifikasi dan prediksi. Klasifikasi berfungsi mengelompokkan data ke dalam kategori tertentu berdasarkan karakteristiknya, sedangkan prediksi (regresi) bertujuan memperkirakan nilai output berdasarkan pola yang dipelajari. Efektivitas model ini sangat bergantung pada kualitas dan volume data yang digunakan. Berbagai algoritma seperti Support Vector Machine (SVM) hingga Neural Network kini banyak diterapkan di berbagai sektor, termasuk kesehatan, untuk membantu analisis data berskala besar dengan akurasi tinggi.

## **2.3 Algoritma yang Digunakan**

### **2.3.1 Naïve Bayes Classifier**

Naïve Bayes merupakan algoritma klasifikasi yang berbasis pada Teorema Bayes dan menggunakan pendekatan probabilistik untuk menentukan kelas suatu data berdasarkan probabilitas posterior yang dihitung dari data historis. Algoritma ini bekerja dengan menghitung probabilitas setiap kelas terhadap fitur yang diberikan, kemudian menetapkan label kelas dengan nilai probabilitas tertinggi (Maximum A Posteriori).

Karakteristik utama metode Naïve Bayes adalah asumsi independensi yang kuat atau bersifat “naif”, di mana setiap fitur dianggap saling bebas dan tidak saling memengaruhi dalam menentukan probabilitas suatu kelas. Keunggulan utama algoritma ini terletak pada kesederhanaan, efisiensi, serta kecepatan komputasinya, bahkan pada dataset berdimensi tinggi. Meskipun secara umum lebih sesuai untuk atribut bertipe diskrit, Naïve Bayes juga dapat diterapkan pada data kontinu dengan asumsi distribusi tertentu, seperti pada Gaussian Naïve Bayes, sehingga algoritma ini tetap efektif digunakan dalam berbagai studi klasifikasi penyakit, termasuk diabetes.

### 2.3.2 K-Nearest Neighbor (KNN)

K-Nearest Neighbor (KNN) termasuk salah satu algoritma machine learning yang sering digunakan dalam tugas klasifikasi. Algoritma ini berada dalam kelompok *instance-based learning* dan dikenal sebagai *lazy learning*, karena tidak melakukan proses pembelajaran secara eksplisit pada tahap pelatihan, melainkan bekerja ketika proses klasifikasi data uji dilakukan. Pendekatan ini memanfaatkan data latih yang tersedia untuk menentukan label data baru berdasarkan tingkat kemiripan atau kedekatan jarak.

Prinsip kerja KNN berfokus pada pencarian sejumlah  $k$  tetangga terdekat dari suatu data uji di dalam ruang fitur. Penentuan kelas dilakukan melalui mekanisme voting, sehingga kelas dengan jumlah terbanyak di antara tetangga terdekat dipilih sebagai hasil klasifikasi. Untuk mengukur tingkat kedekatan antar data, metode penghitungan jarak seperti Euclidean distance sering digunakan.

Sebagai metode nonparametrik, KNN tidak memerlukan asumsi khusus terhadap distribusi data dan dapat diterapkan pada berbagai jenis dataset. Namun demikian, performa algoritma ini sangat dipengaruhi oleh pemilihan nilai  $k$  yang tepat serta metode pengukuran jarak yang digunakan.

### 2.3.3 Decision Tree

Algoritma ini memetakan sekumpulan data ke dalam struktur hierarki yang menyerupai pohon, sehingga mampu menyederhanakan basis data yang kompleks menjadi aturan-aturan keputusan yang logis. Secara struktural, pohon keputusan terdiri dari root node sebagai titik awal dataset, internal node yang berfungsi sebagai penguji atribut, serta leaf node yang merepresentasikan hasil akhir atau label kelas.

Dalam prosesnya, Decision Tree membagi dataset menjadi kelompok-kelompok yang lebih kecil berdasarkan atribut yang paling informatif hingga mencapai simpul daun. Setiap jalur yang terbentuk dari akar hingga ke daun berperan sebagai aturan keputusan untuk mengklasifikasikan data baru. Keunggulan utama metode ini terletak pada kemampuannya memvisualisasikan proses pengambilan keputusan secara sederhana, sehingga hasil analisisnya sangat mudah diinterpretasikan oleh pengguna dalam berbagai bidang.



## BAB III

### METODOLOGI PENELITIAN

#### 3.1 Dataset

Penelitian ini memanfaatkan dataset yang diperoleh dari platform Kaggle. Data awal terdiri dari 2.000 entri dengan 11 atribut yang didominasi oleh tipe data numerik (*integer* dan *float*). Dalam tahap awal pengolahan, dilakukan eliminasi terhadap dua atribut yang dinilai tidak relevan atau tidak memiliki nilai, sehingga menyisakan 9 variabel inti. Struktur akhir dataset ini mencakup 8 atribut input dan 1 atribut label (Outcome) sebagaimana dirinci pada Tabel 1.

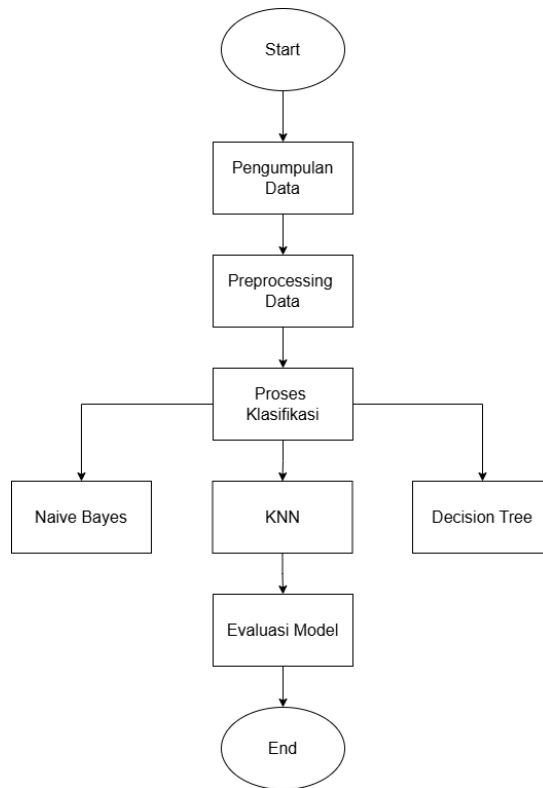
Tabel 1. Atribut Dataset

No	Nama Atribut	Tipe Data	Keterangan
1.	Pregnancies	Integer	Jumlah kehamilan
2.	Glucose	Integer	Konsentrasi glukosa plasma
3.	BloodPressure	Integer	Tekanan darah (mmHg)
4.	SkinThickness	Integer	Ketebalan lipatan kulit
5.	Insulin	Integer	Kadar insulin darah
6.	BMI	Float	Indeks massa tubuh
7.	DiabetesPedigreeFunction	Float	Skor faktor genetik diabetes
8.	Age	Integer	Usia pasien
9.	Outcome	Integer	Label kelas diabetes

#### 3.2 Tahapan Penelitian

Penelitian ini dilakukan secara sistematis, dimulai dari pengumpulan dan preprocessing dataset diabetes. Dataset kemudian dibagi menjadi data latih dan data uji. Selanjutnya dilakukan proses klasifikasi menggunakan algoritma Naïve Bayes, K-Nearest Neighbor (KNN), dan Decision Tree. Untuk meningkatkan keandalan model, pada tahap klasifikasi diterapkan teknik cross validation. Hasil klasifikasi

selanjutnya dievaluasi untuk mengukur kinerja masing-masing metode. Alur penelitian ditunjukkan pada Gambar 1.



Gambar 1. Tahapan Penelitian

### 3.3 Preprocessing Data

Data menjalani tahap preprocessing untuk memastikan kualitas dan konsistensi sebelum digunakan dalam analisis. Tahap ini meliputi penanganan nilai yang hilang, normalisasi, dan pemrosesan lain agar data siap digunakan secara optimal.

Selain itu, dataset dibagi menjadi 80% data training dan 20% data testing. Pembagian ini bertujuan agar model dapat dilatih secara efektif pada data training dan diuji pada data testing untuk mengevaluasi performa secara akurat.

### 3.4 Confusion Matrix

Confusion matrix, juga disebut error matrix, menampilkan kinerja model klasifikasi dengan membandingkan prediksi dan data aktual. Pada penelitian ini,

confusion matrix menggunakan dua kelas, yaitu kelas positif dan kelas negatif. Metode evaluasi dengan dua kelas tersebut ditunjukkan pada Tabel 2.

Tabel 2. Confusion Matrix

		Actual Values	
		Positive (P)	Negative (N)
Predicted Values	Positive (P)	True Positive (TP)	False Positive (FP)
	Negative (N)	False Negative (FN)	True Negative (TN)

**Keterangan:**

- › True Positive (TP) = Jumlah data positif yang diklasifikasikan sebagai positif
- › False Positive (FP) = Jumlah data negatif yang diklasifikasikan sebagai positif
- › False Negative (FN) = Jumlah data positif yang diklasifikasikan sebagai negatif
- › True Negative (TN) = Jumlah data negatif yang diklasifikasikan sebagai negatif

**Metrik Turunan dari Confusion Matrix**

1. Accuracy (Akurasi)

Persentase prediksi yang benar dari seluruh data

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

2. Precision (Presisi)

Proporsi prediksi positif yang benar-benar positif

$$Precision = \frac{TP}{TP + FP}$$

3. Recall / True Positive Rate (TPR)

Proporsi data positif yang berhasil dikenali model

$$Recall = \frac{TP}{TP + FN}$$

4. F1-Score

Menggabungkan presisi dan recall untuk menilai keseimbangan prediksi

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} = \frac{2TP}{2TP + FP + FN}$$

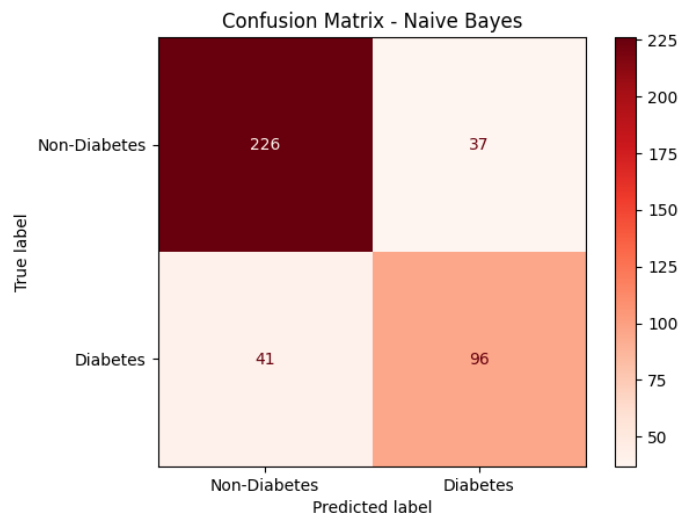
## BAB IV

### HASIL DAN PEMBAHASAN

#### 4.1 Evaluasi Model Naïve Bayes

Dataset yang digunakan dalam penelitian ini berjumlah 2.000 data dengan 9 atribut, terdiri dari 8 variabel input dan 1 variabel kelas. Data dibagi menjadi 80% data latih dan 20% data uji, sehingga diperoleh 400 data uji. Model klasifikasi dibangun menggunakan algoritma Naïve Bayes dan diimplementasikan menggunakan Google Colab.

Evaluasi performa model dilakukan menggunakan confusion matrix berdasarkan 400 data uji. Hasil pengujian menunjukkan nilai True Negative (TN) sebesar 226, yang berarti model berhasil mengklasifikasikan pasien non-diabetes dengan benar. Nilai False Positive (FP) sebesar 37 menunjukkan adanya pasien non-diabetes yang diprediksi sebagai diabetes. Nilai False Negative (FN) sebesar 41 menunjukkan pasien diabetes yang gagal terdeteksi oleh model, sedangkan nilai True Positive (TP) sebesar 96 menunjukkan jumlah pasien diabetes yang berhasil diprediksi dengan benar.



Gambar 2. Confusion Matrix Model Naïve Bayes

Berdasarkan hasil evaluasi, model Naïve Bayes memperoleh akurasi sebesar 81%, yang menunjukkan kemampuan klasifikasi yang cukup baik. Pada kelas 0, diperoleh nilai precision sebesar 0,85, recall 0,86, dan F1-score 0,85, sedangkan

pada kelas 1 diperoleh precision 0,72, recall 0,70, dan F1-score 0,71, yang menunjukkan bahwa performa model pada kelas minoritas masih lebih rendah dibandingkan kelas mayoritas.

Nilai macro average sebesar 0,78 dan weighted average sebesar 0,80 menunjukkan bahwa model memiliki performa yang cukup stabil meskipun terdapat ketidakseimbangan distribusi data. Hasil evaluasi model Naïve Bayes ditunjukkan pada Gambar 3.

	precision	recall	f1-score	support
0	0.85	0.86	0.85	263
1	0.72	0.70	0.71	137
accuracy			0.81	400
macro avg	0.78	0.78	0.78	400
weighted avg	0.80	0.81	0.80	400

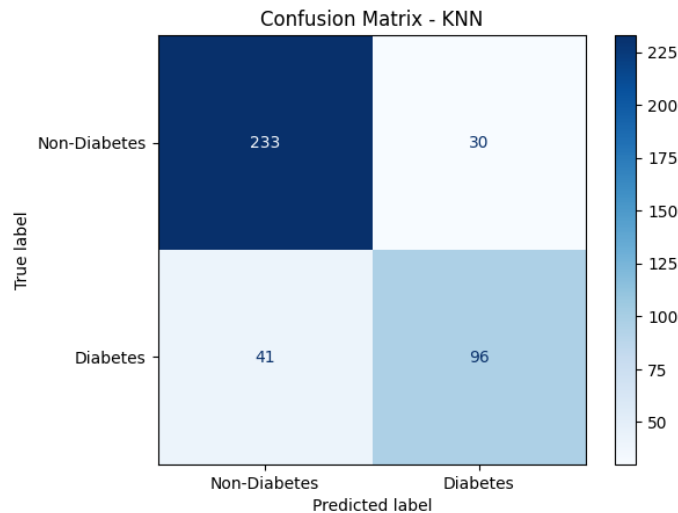
Gambar 3. Hasil Evaluasi Model Naïve Bayes

## 4.2 Evaluasi Model K-Nearest Neighbor (KNN)

Dataset yang digunakan dalam penelitian ini berjumlah 2.000 data dengan 9 atribut, terdiri dari 8 variabel input dan 1 variabel kelas. Data dibagi menjadi 80% data latih dan 20% data uji, sehingga diperoleh 400 data uji. Model klasifikasi dibangun menggunakan algoritma K-Nearest Neighbors (KNN) dan diimplementasikan menggunakan Google Colab.

Evaluasi performa model dilakukan menggunakan confusion matrix berdasarkan 400 data uji. Hasil pengujian menunjukkan nilai True Negative (TN) sebesar 233, yang berarti model KNN berhasil memprediksi pasien non-diabetes dengan benar. Nilai False Positive (FP) sebesar 30 menunjukkan adanya pasien non-diabetes yang diprediksi sebagai diabetes oleh model.

Selanjutnya, nilai False Negative (FN) sebesar 41 menunjukkan bahwa model KNN gagal mendeteksi pasien yang sebenarnya mengalami diabetes, sedangkan nilai True Positive (TP) sebesar 96 menunjukkan jumlah pasien diabetes yang berhasil diprediksi dengan benar. Confusion matrix ini merepresentasikan kinerja model KNN dan menjadi dasar dalam perhitungan metrik evaluasi.



Gambar 4. Confusion Matrix Model K-Nearest Neighbor (KNN)

Berdasarkan hasil evaluasi model KNN, diperoleh nilai akurasi sebesar 82%, yang menunjukkan bahwa model mampu mengklasifikasikan data uji dengan cukup baik. Pada kelas 0, model menghasilkan nilai precision sebesar 0,85, recall 0,89, dan F1-score 0,87, yang menunjukkan performa klasifikasi yang baik pada kelas mayoritas. Sementara itu, pada kelas 1, diperoleh nilai precision sebesar 0,76, recall 0,70, dan F1-score 0,73, yang menunjukkan bahwa performa model pada kelas minoritas masih lebih rendah dibandingkan kelas mayoritas.

Nilai macro average sebesar 0,80 menunjukkan performa rata-rata model pada kedua kelas, sedangkan weighted average sebesar 0,82 menandakan bahwa model memiliki performa yang cukup stabil meskipun distribusi data tidak seimbang. Hasil evaluasi model K-Nearest Neighbor (KNN) ditunjukkan pada Gambar 5.

	precision	recall	f1-score	support
0	0.85	0.89	0.87	263
1	0.76	0.70	0.73	137
accuracy			0.82	400
macro avg	0.81	0.79	0.80	400
weighted avg	0.82	0.82	0.82	400

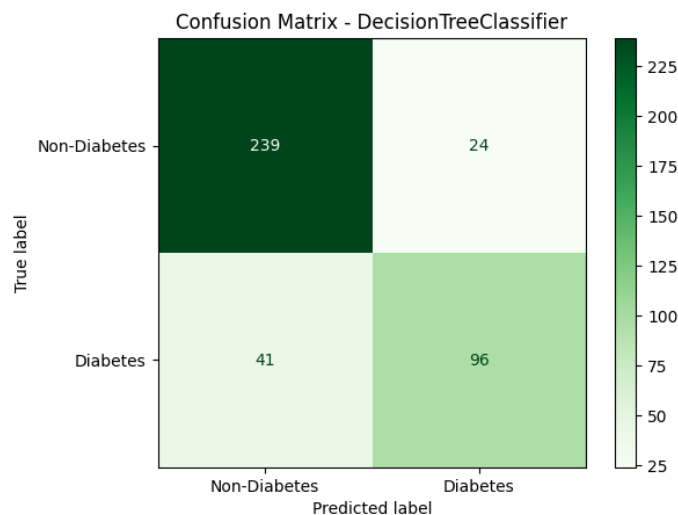
Gambar 5. Hasil Evaluasi Model K-Nearest Neighbor (KNN)

### 4.3 Evaluasi Model Decision Tree

Dataset yang digunakan dalam penelitian ini berjumlah 2.000 data dengan 9 atribut, terdiri dari 8 variabel input dan 1 variabel kelas. Data dibagi menjadi 80% data latih dan 20% data uji, sehingga diperoleh 400 data uji. Model klasifikasi dibangun menggunakan algoritma Decision Tree dan diimplementasikan menggunakan Google Colab.

Evaluasi performa model dilakukan menggunakan confusion matrix berdasarkan 400 data uji. Hasil pengujian menunjukkan nilai True Negative (TN) sebesar 239, yang berarti model Decision Tree berhasil memprediksi pasien non-diabetes dengan benar. Nilai False Positive (FP) sebesar 24 menunjukkan adanya pasien non-diabetes yang diprediksi sebagai diabetes oleh model.

Selanjutnya, nilai False Negative (FN) sebesar 41 menunjukkan bahwa model Decision Tree gagal mendeteksi pasien yang sebenarnya mengalami diabetes, sedangkan nilai True Positive (TP) sebesar 96 menunjukkan jumlah pasien diabetes yang berhasil diprediksi dengan benar. Confusion matrix ini merepresentasikan kinerja model Decision Tree dan menjadi dasar perhitungan metrik evaluasi.



Gambar 6. Confusion Matrix Model Decision Tree

Berdasarkan hasil evaluasi model Decision Tree, diperoleh akurasi sebesar 84%, yang menunjukkan bahwa model mampu mengklasifikasikan data uji dengan sangat baik. Pada kelas 0, model menghasilkan nilai precision sebesar 0,85, recall

0,91, dan F1-score 0,88, yang menunjukkan performa klasifikasi yang sangat baik pada kelas mayoritas.

Sementara itu, pada kelas 1, diperoleh nilai precision sebesar 0,80, recall 0,70, dan F1-score 0,75, yang menunjukkan bahwa performa model pada kelas minoritas masih lebih rendah dibandingkan kelas mayoritas.

Nilai macro average sebesar 0,81 menunjukkan performa rata-rata model pada kedua kelas, sedangkan weighted average sebesar 0,83 menandakan bahwa model Decision Tree memiliki performa paling stabil secara keseluruhan meskipun distribusi data tidak seimbang. Hasil evaluasi model Decision Tree ditunjukkan pada Gambar 7.

	precision	recall	f1-score	support
0	0.85	0.91	0.88	263
1	0.80	0.70	0.75	137
accuracy			0.84	400
macro avg	0.83	0.80	0.81	400
weighted avg	0.84	0.84	0.83	400

Gambar 7. Hasil Evaluasi Model Decision Tree



## **BAB V**

### **KESIMPULAN DAN SARAN**

#### **5.1 Kesimpulan**

Berdasarkan hasil penelitian yang telah dilakukan terhadap dataset penyakit diabetes menggunakan metode Naive Bayes, K-Nearest Neighbors (KNN), dan Decision Tree, dapat disimpulkan bahwa:

1. Hasil analisis data menunjukkan bahwa kadar gula darah merupakan variabel yang paling berpengaruh dalam menentukan penyakit diabetes.
2. Dataset yang digunakan memiliki ketidakseimbangan jumlah data antara kelas diabetes dan non-diabetes. Kondisi ini memengaruhi hasil klasifikasi, terutama pada kelas diabetes sebagai kelas minoritas. Pada penelitian ini digunakan data asli tanpa proses penyeimbangan agar hasil yang diperoleh tetap mencerminkan kondisi sebenarnya.
3. Berdasarkan hasil evaluasi model, Decision Tree menghasilkan performa terbaik dengan nilai akurasi tertinggi dibandingkan dengan Naive Bayes dan KNN, sehingga dipilih sebagai model terbaik dalam penelitian ini.

#### **5.2 Saran**

Berdasarkan hasil penelitian, beberapa saran yang dapat diberikan adalah sebagai berikut:

1. Penelitian selanjutnya disarankan untuk menerapkan teknik penyeimbangan data agar hasil klasifikasi pada kelas minoritas dapat ditingkatkan.
2. Dapat dilakukan pengujian menggunakan metode klasifikasi lain untuk membandingkan hasil dan memperoleh performa yang lebih baik.
3. Pengaturan parameter model dan penggunaan data yang lebih beragam dapat dipertimbangkan agar hasil penelitian menjadi lebih optimal.

## DAFTAR PUSTAKA

- Ahmad, A. (2017). Mengenal artificial intelligence, machine learning, neural network, dan deep learning. [https://www.academia.edu/download/54674088/Perbedaan\\_Deep\\_learn.pdf](https://www.academia.edu/download/54674088/Perbedaan_Deep_learn.pdf)
- Aktavera, B., & Wijaya, H. O. L. (2023). Klasifikasi produk menggunakan algoritma decision tree. <https://pdfs.semanticscholar.org/6155/d13a2fa66a7d36f544e7b842d24cff6e9126.pdf>
- Alfarizi, M. R. S., Al-Farish, M. Z., Taufiqurrahman, M., Ardiansah, G., & Elga, M. (2023). Penggunaan Python sebagai bahasa pemrograman untuk *machine learning* dan *deep learning*. *Karimah Tauhid*, 2(1), Article 7518. <https://ojs.unida.info/karimahtauhid/article/view/7518>
- Arda, D., Aminuddin, A., Sima, Y., Izza, N. C., & Lalla, N. S. N. (2023). Edukasi kesehatan tentang penyakit diabetes melitus bagi masyarakat. *Abdimas Polsaka*, 2(1), 7–12. <https://doi.org/10.35816/abdimaspolsaka.v2i1.25>
- Cahyantia, D., Rahmayania, A., & Husniara, S. A. (2020). Analisis performa metode KNN pada dataset pasien pengidap kanker payudara. *IJODAS*, 1(1), Article 13. <https://jurnal.yoctobrain.org/index.php/ijodas/article/view/13/10>
- Chazar, C., & Widhiaputra, B. E. (2020). Machine learning diagnosis kanker payudara menggunakan algoritma support vector machine. *Informasi*, 5(2), 48. <https://doi.org/10.37424/informasi.v12i1.48>
- Devi, E. S., Wahono, B. B., & Wibowo, G. N. (2022). Komparasi algoritma klasifikasi Naïve Bayes, decision tree (C4.5) dan support vector machine (SVM) dalam diagnosis penyakit diabetes. *Jurnal Teknik Informatika (J-TInFo)*, 5(1), Article 1268. <https://journal.unisnu.ac.id/JTINFO/article/view/1268>
- Fitria, R., Yulisda, D., & Ula, M. (2021). Data mining classification algorithms for diabetes dataset using WEKA tool. *Jurnal Informatika dan Sistem Informasi*, 5(2), 156–165. <https://doi.org/10.29103/sisfo.v5i2.6236>

- Hadi, A. P., & Santoso, J. (2024). Analisis perbandingan kinerja model machine learning dalam prediksi diabetes: Studi kasus dataset Kaggle 2022. *Jurnal Ilmiah Komputer*, 10(2), 123–135.  
<https://ejurnal.esaunggul.ac.id/index.php/JIK/article/view/8784/4378>
- Haq, O. M., Ridwan, A., & Pratama, T. G. (2025). Analisis perbandingan kinerja algoritma Naïve Bayes dan KNN untuk memprediksi penyakit diabetes. *Progresif*, 5(1), Article 2424.  
<http://dx.doi.org/10.35889/progresif.v2i1i1.2424>
- Khasanah, L. U., Nasution, Y. N., & Amijaya, F. D. T. (2022). Klasifikasi penyakit diabetes melitus menggunakan algoritma Naïve Bayes classifier. *Basis*, 9(1), Article 918.  
<https://jurnal.fmipa.unmul.ac.id/index.php/Basis/article/view/918/437>
- Kustiyahningsih, Y., & Syafa'ah, N. (2015). Sistem pendukung keputusan untuk menentukan jurusan pada siswa SMA menggunakan metode KNN dan SMART. <https://www.researchgate.net/profile/Yeni-Kustiyahningsih-2/publication/345942709>
- Nawawi, I., & Fatah, Z. (2024). Penerapan decision trees dalam mendeteksi pola tidur sehat berdasarkan kebiasaan gaya hidup. <https://pdfs.semanticscholar.org/f1a4/97d3bdce22e1c496652f04ca0dfd8838d70c.pdf>
- Puspita, R., & Widodo, A. (2021). Perbandingan metode KNN, decision tree, dan Naïve Bayes terhadap analisis sentimen pengguna layanan BPJS. <https://pdfs.semanticscholar.org/58ad/ceecf8f3b1e315f36625e2691181fd4505d7.pdf>
- Ridwan, A. (2020). Penerapan algoritma Naïve Bayes untuk klasifikasi penyakit diabetes mellitus. *Siskom-KB*, 1(2), Article 169.  
<https://jurnal.tau.ac.id/index.php/siskom-kb/article/view/169/146>
- Saleh, A., Sari, R. E., Ramadani, Fujiati, & Lestari, R. (2025). Analisis komparatif algoritma klasifikasi machine learning untuk memprediksi diabetes. *Algoritma*, 9(1), Article 23794.  
<https://doi.org/10.30829/algoritma.v9i1.23794>

- Septhya, D., Rahayu, K., Rabbani, S., Fitria, V., Rahmadden, Irawan, Y., & Hayami, R. (2023). Implementasi algoritma decision tree dan support vector machine untuk klasifikasi penyakit kanker paru. *Malcom*, 3(1), Article 591. <https://doi.org/10.57152/malcom.v3i1.591>
- Tasya Syamsudin, T. Handhayani, & M. I. Syaifudin. (2024). Perbandingan klasifikasi penyakit diabetes menggunakan metode machine learning. *Jurnal Ilmu Komputer dan Informasi*, 17(3), 789–802. <https://journal.untar.ac.id/index.php/jiksi/article/view/28197/16763>
- Watratana, A. F., Puspita, A. B., & Moeis, D. (2020). Implementasi algoritma Naïve Bayes untuk memprediksi tingkat penyebaran COVID-19 di Indonesia. *JACOST*, 1(1), Article 9. <https://doi.org/10.52158/jacost.v1i1.9>