Jurnal Infortech
Volume 7 No. 2 Desember 2025
E-ISSN: 2715-8160

Pendekatan Algoritma Naive Bayes Dalam Memprediksi Penyakit Diabetes

Qudsiah Nur Azizah[1], Diah Puspitasari[2], Sulaeman Hadi Sukmana[3], Erma Delima Sikumbang[4], Kresna Ramanda[5]

[1,2,3,4,5]Universitas Bina sarana Informatika
e-mail: [1]qudsiah.qna@bsi.ac.id, [2]diah.puspitasari@bsi.ac.id, [3]sulaeman.sdu@bsi.ac.id, [4]erma@bsi.ac.id, [5] kresna.kra@bsi.ac.id

Diterima	Direvisi	Disetujui
30-07-2025	18-09-2025	22-12-2025

Abstrak - Diabetes melitus merupakan penyakit metabolik yang bersifat kronis dan multifaktorial. Penyakit ini menunjukkan gejala peningkatan kadar gula darah (hiperglikemia) akibat proses metabolisme karbohidrat yang tidak normal, lemak, dan protein yang tidak normal. Hingga saat ini, Melampaui angka 150 juta orang yang tercatat secara global mengidap penyakit ini dan perkembangan penyakit yang terus meningkat dapat menyebabkan komplikasi yang fatal. Faktanya, sebagian besar masyarakat mengabaikan tanda-tanda awal ini: rasa lapar yang berlebihan, rasa lelah yang tidak wajar, dan luka yang lambat sembuh. Penelitian ini dilakukan untuk analisis algoritma Naïve Bayes pada klasifikasi penyakit diabetes untuk mendapatkan hasil optimal dengan akurasi yang ditawarkan dengan cara ini. Dataset diambil dari website Kaggle yang berjumlah 10.000 data dengan jumlah 2 kelas yaitu Diabetes dan Non Diabetes, kelas *Diabetes* mencakup sebanyak 8.500 data, sementara kelas *Non-Diabetes* mencakup 91.500 data. Metodologi yang diterapkan yakni Algoritma Naïve Bayes. Evaluasi model membuktikan bahwa Naïve Bayes mampu mencapai skor akurasi sebesar 90,66% yang artinya Algoritma Naïve Bayes merupakan pendekatan yang reliabel dan efektif dalam mengklasifikasikan penyakit diabetes.

Kata Kunci: Naive Bayes, Diabetes, Machine Learning, Klasifikasi.

Abstract *– Diabetes mellitus is a chronic and multifactorial metabolic disorder characterized by hyperglycemia resulting from abnormalities in carbohydrate, lipid, and protein metabolism. Currently, more than 150 million people worldwide are diagnosed with this condition, and its increasing prevalence poses a risk of fatal complications. In practice, early clinical manifestations—such as polyphagia, abnormal fatigue, and delayed wound healing—are frequently overlooked by the general population. This study was conducted to analyze the Naïve Bayes algorithm in the classification of diabetes to achieve optimal accuracy. The dataset, sourced from Kaggle, comprises 100,000 instances categorized into two classes: 8,500 instances for the 'Diabetes' class and 91,500 instances for the 'Non-Diabetes' class. The research methodology employs the Naïve Bayes algorithm. Experimental results demonstrate that Naïve Bayes achieves an accuracy rate of 90.66%, indicating that this algorithm is an effective and appropriate method for the classification of diabetes.*

Keywords: *Naïve Bayes, Diabetes, Machine Learning, Classification*.

PENDAHULUAN

Peningkatan kadar glukosa darah merupakan indikator adanya gangguan metabolisme yang dikenal sebagai diabetes melitus. Secara global, prevalensi penyakit ini menunjukkan bahwa sekitar satu dari sebelas orang dewasa terdampak oleh kondisi tersebut diabetes melitus tipe 2, dengan sekitar 75% penderitanya berada di negara-negara berkembang. Tipe 2 dari penyakit ini terjadi akibat Defisiensi produksi insulin karena kerusakan fungsi sel pankreas serta adanya resistensi terhadap insulin. Faktor-faktor risiko penyakit ini terdiri dari yang bisa diubah dan yang tidak dapat diubah (Widiasari et al., 2021). Diabetes Manifestasi gangguan klinis jangka panjang yang serius dan terus mengalami peningkatan jumlah kasus, serta menjadi salah satu penyebab utama kematian. Kondisi ini terjadi karena defisiensi fungsi metabolik yang menghambat proses asimilasi nutrien secara optimal oleh sistemik insulin dengan baik atau defisiensi sintesis insulin pada organ pankreas. (Nasution et al., 2021).

Untuk mendapatkan perbedaan antara objek, klasifikasi harus dilaksanakan berdasarkan prosedur yang terstandardisasi demi mencapai luaran yang paling akurat. Machine learning adalah sebuah cabang teknologi mutakhir yang berkembang dengan sangat cepat di era modern, di mana sistem komputer diprogram menggunakan algoritma khusus sehingga mampu melakukan pembelajaran secara otomatis dari data tanpa intervensi manusia secara langsung (Suwitono & Kaunang, 2022). Data mining adalah suatu metode untuk mengekstraksi informasi baru

dengan cara menganalisis pola-pola tertentu dari data berukuran besar. Proses ini terdiri dari beberapa tahap yang bertujuan untuk menemukan wawasan atau pengetahuan tersembunyi yang tidak bisa diidentifikasi secara langsung atau manual. Istilah ini secara terminologis diidentifikasi sebagai *Knowledge Discovery in Databases* (KDD) (Muntiari & Hanif, 2022). Naive Bayes adalah algoritma klasifikasi yang didasarkan pada pendekatan probabilistik yang bekerja dengan menganalisis frekuensi kemunculan serta hubungan nilai-nilai dalam data untuk memperkirakan kemungkinan suatu kategori (Putra et al., 2024).

Pada Penelitian (Rizki et al., 2022) mengeksplorasi faktor risiko dan algoritma Kernel K-Nearest Neighbor untuk memprediksi kemungkinan pengidap diabetes memprediksi potensi pengidap diabetes berdasarkan factor risiko dengan menggunakan kernel linier dan kernel polynomial degree 1, menghasilkan rata-rata akurasi sebesar 93%. Pada Penelitian Lumbanraja (R Lumbanraja et al., 2022) penelitian tentang penerapan metode *Support Vector Machine* (SVM) digunakan dalam proses pengklasifikasian penderita diabetes melitus menunjukkan bahwa penggunaan kernel Gaussian menghasilkan tingkat akurasi yang tinggi, rata-rata 82,76%, dalam klasifikasi SVM.

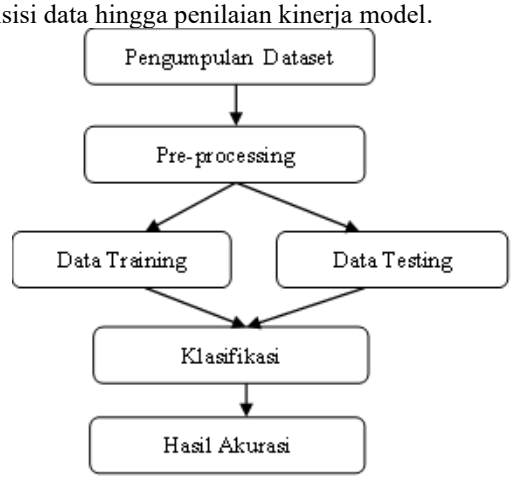
Penelitian terdahulu yang berjudul "Pendekatan Algoritma Naïve Bayes Dalam Memprediksi Penyakit Diabetes" mengungkapkan bahwa Algoritma Naive Bayes memiliki kemampuan untuk menghasilkan akurasi yang tinggi dalam proses klasifikasi terhadap penyakit diabetes. Angka tersebut mencerminkan performa yang cukup baik dalam konteks klasifikasi berbasis pendekatan probabilistik. Algoritma Naive Bayes menunjukkan kemampuan untuk kinerja yang lebih baik dalam penelitian ini dengan capaian tingkat akurasi sebesar 90,66%, yang mengindikasikan bahwa algoritma ini memiliki kapabilitas yang sangat baik dalam mengidentifikasi dan mengelompokkan data penderita diabetes secara lebih akurat. Perbedaan hasil akurasi antara kedua penelitian tersebut dapat disebabkan oleh beragam faktor, antara lain adanya perbedaan pada karakteristik dataset yang digunakan, teknik prapemrosesan data, parameter yang diterapkan, Termasuk pula rasio distribusi data antara himpunan latih (training set) dan himpunan uji (testing set).

Implementasi algoritma Naive Bayes dalam penelitian ini difokuskan pada proses klasifikasi jenis penyakit diabetes dengan memanfaatkan teknik pengolahan data secara sistematis, dimulai dari tahapan prapemrosesan, pelatihan model, hingga evaluasi performa klasifikasi. Dengan mempertimbangkan hasil yang diperoleh, dapat disimpulkan bahwa Naive Bayes adalah salah satu teknik terbaik untuk membuat sistem klasifikasi berbasis data mining untuk mendeteksi penyakit diabetes. Penelitian ini juga memberikan kontribusi terhadap penguatan bukti empiris mengenai efektivitas Naive Bayes dalam domain medis, khususnya untuk diagnosis awal penyakit kronis seperti diabetes melitus.

METODE PENELITIAN

Fokus utama dari riset ini adalah untuk mengkaji serta mengklasifikasikan jenis diabetes. pada individu dengan memanfaatkan pendekatan pengolahan data dan algoritma *Naive Bayes*. Proses penelitian diawali dengan tahapan akuisisi dataset sebagai basis analisis. Dataset yang telah diperoleh selanjutnya dipisahkan secara terstruktur ke dalam dua klasifikasi primer, yang masing-masing direpresentasikan sebagai himpunan latih (training set) dan himpunan uji (testing set), guna mendukung proses dan meningkatkan performa model pada tahap klasifikasi.

Metodologi dalam penelitian klasifikasi penyakit diabetes ini dirancang secara metodis dan diorganisasikan ke dalam format visualisasi untuk memberikan pemahaman yang komprehensif mengenai alur dan tahapan penelitian yang dilakukan. Pengembangan model prediksi dilakukan dengan menerapkan pendekatan Naive Bayes, yang dibangun berdasarkan data yang telah melalui serangkaian proses prapemrosesan, seperti pembersihan data, penanganan nilai hilang, dan pembagian dataset. Setelah model dikembangkan, kinerjanya dievaluasi secara kuantitatif berdasarkan parameter evaluasi yang mencakup tingkat akurasi, presisi, recall, serta F1-score, serta dibandingkan dengan model referensi atau pendekatan lain yang telah ada dalam studi terdahulu. Rangkaian tahapan kerangka metodologis yang diajukan dalam studi ini divisualisasikan secara jelas pada Gambar 1, guna memperjelas struktur logis dan proses kerja dalam penelitian ini, mencakup seluruh tahapan komprehensif, mulai dari fase akuisisi data hingga penilaian kinerja model.



Sumber : Penelitian (2025)

Gambar 1.Metode Yang Diusulkan

1. Dataset

Dataset yang diimplementasikan dalam penelitian ini mencakup data patologi diabetes yang bersumber dari situs *open-source* www.kaggle.com. Dataset tersebut terdiri atas 100.000 entri yang terklasifikasi ke dalam dua kategori fundamental, yakni kelas Diabetes sebanyak 8.500 data dan kelas Non-Diabetes sebanyak 91.500 data. Pemilihan dataset ini didasarkan pada kelengkapan dan relevansi data terhadap tujuan penelitian, yaitu untuk mendukung proses analisis klasifikasi dalam mendeteksi kondisi diabetes secara lebih akurat melalui penerapan metode algoritma yang diusulkan.

2. Pre-processing

Tahapan ini merupakan langkah krusial sebelum proses klasifikasi, yang mencakup pembersihan data dari *noise*, penanganan terhadap nilai yang hilang maupun tidak konsisten, serta Proses pengolahan data dilakukan dengan memisahkan dataset ke dalam data latih sebagai bahan pembelajaran model dan data uji untuk tahap evaluasi. Pada fase partisi data, dialokasikan proporsi sebesar 90% untuk tahap pelatihan model dan 10% untuk keperluan evaluasi performa.

Prosedur ini diarahkan guna mengoptimalkan kualitas data secara keseluruhan agar model klasifikasi yang dikembangkan mampu menunjukkan performansi yang optimal. Selain itu, tahapan ini juga bertujuan untuk menjamin bahwa data yang diimplementasikan dalam proses penambangan data (data mining) memiliki derajat konsistensi yang memadai dan kesiapan yang tinggi untuk diolah lebih lanjut.

3. Klasifikasi

Pada fase ini, dilaksanakan prosedur pembelajaran (*training*) menggunakan data latih (*training data*) yang telah dilengkapi dengan label kelas, dengan tujuan untuk membentuk suatu model klasifikasi yang mampu mengenali pola-pola tertentu dalam data dan menggunakannya untuk Melakukan klasifikasi data uji (testing data) ke dalam kategori atau label kelas yang telah ditetapkan sebelumnya.

Proses pembelajaran ini merupakan inti dari metode *supervised learning*, di mana algoritma secara iteratif menyesuaikan parameter-parameter internalnya untuk meminimalkan kesalahan prediksi. Dengan demikian, diharapkan model yang dihasilkan memiliki kapabilitas generalisasi yang mumpuni serta mampu memberikan prediksi presisi terhadap data prematur (unseen data) yang belum pernah terlibat dalam proses pelatihan.

Proses penyusunan model atau fungsi yang bertujuan untuk mengenali serta Prosedur pemisahan data ke dalam kategori yang spesifik didefinisikan sebagai klasifikasi. Mekanisme ini selanjutnya diimplementasikan guna mengidentifikasi. Proses ini kemudian digunakan untuk menentukan kelas data baru yang belum diketahui kategorinya. (Azizah, 2023)

4. Naive Bayes

Naive Bayes merepresentasikan salah satu instrumen klasifikasi statistik yang diimplementasikan guna memperkirakan probabilitas suatu data termasuk ke dalam kelas tertentu. Pendekatan kategorisasi data menggunakan teknik ini. Dengan fase penerapan algoritma untuk meneliti dataset dalam sistem implementasi Naive Bayes (Gusrialni Fitri et al., 2023). Persamaan Teorema Bayes diberikan dalam rumus sebagai berikut :

$$P(x|y) = \frac{p(x)=p\,(x)}{p(y)} \quad \text{.....(1)}$$

Keterangan :

y : Data dengan kelas yang tidak terklasifikasi.
x : Hipotesis data y merupakan suatu kelas spesifik.
P(x|y) : Kemungkinan berdasarkan kondisi x dan' y (posteriosi probobality).
P(x) :Kemungkinan premis x (prior probability).
P(y|x) : Probabilitas y mengingat keadaan dalam hipotesis x.
P(y) : Kemungkinan y.

Algoritma Naive Bayes adalah salah satu metode klasifikasi dalam bidang data mining yang berlandaskan berdasarkan Teorema Bayes, metodologi ini diakui memiliki proses komputasi yang cepat serta tingkat akurasi yang baik, khususnya ketika diterapkan pada kumpulan data berukuran besar (Rinanda et al., 2022). Naive Bayes merepresentasikan metode klasifikasi yang bersifat fundamental namun memiliki efektivitas tinggi dalam mengatribusikan data ke dalam kelas yang relevan dengan derajat akurasi yang relatif unggul.

Rumus Naive Bayes :

$$P\,(Ci|X) = \frac{P(X|Ci).P(Ci)}{P(X)} \quad \text{.....(2)}$$

X : Data tanpa informasi kelas
Ci : Prediksi bahwa data X berada pada kelas tertentu
P(Ci|X) : Prediksi Peluang Ci sesuai keadaan X
P(Ci) : Prediksi Peluang Ci

P(X|Ci) : Peluang X sesuai keadaan pada Prediksi Ci
P(X) : Peluang dari X

5. Machine Learning

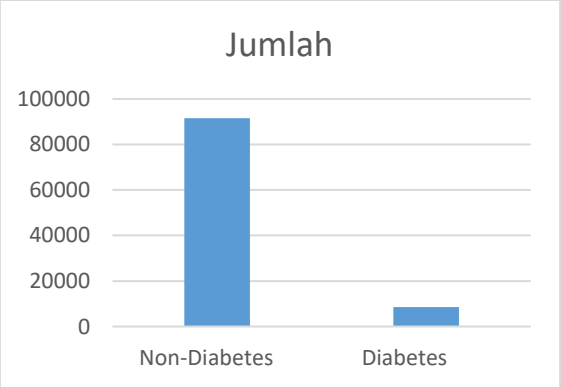
Machine learning merupakan salah satu subdisiplin dari bidang kecerdasan buatan (artificial intelligence) yang dikembangkan untuk meniru proses berpikir manusia. Teknologi ini memungkinkan sistem untuk melakukan ekstraksi pengetahuan dan pembelajaran secara mandiri berbasis data, serta mengambil keputusan secara otonom tanpa memerlukan pemrograman ulang secara eksplisit. Dalam machine learning, hasil yang akurat dapat dicapai melalui penerapan teknik-teknik cerdas.

Pengajaran yang diawasi, pengajaran yang tidak diawasi, pengajaran semi-pengawasan, dan penguatan pengajaran adalah beberapa metode pengajaran mesin (Valerian et al., 2025).

HASIL DAN PEMBAHASAN

1. Dataset

Dataset penderita diabetes mellitus yang diimplementasikan dalam penelitian ini diekstraksi dari laman resmi penyedia dataset terbuka *Kaggle*, dengan total sebanyak 100.000 record. Dataset tersebut terdiri atas dua kelas utama, yaitu Diabetes dan Non-Diabetes, yang masing-masing merepresentasikan kondisi pasien berdasarkan diagnosis medis. Data yang digunakan mencakup atribut-atribut medis dan demografis, seperti usia, jenis kelamin, kadar glukosa, tekanan darah, indeks massa tubuh (*body mass index*), serta beberapa parameter klinis lainnya yang relevan dengan risiko diabetes. Informasi ini disusun secara sistematis dan divisualisasikan pada Gambar 2, guna memberikan gambaran awal mengenai distribusi data serta proporsi antar kelas yang menjadi basis fundamental dalam prosedur klasifikasi serta tahapan analisis komprehensif pada penelitian ini.



Sumber : Penelitian (2025)

Gambar 2. Visualisasi Jumlah Dataset

Pada Gambar 2, divisualisasikan distribusi data dari dataset yang digunakan dalam penelitian ini, di mana kelas Non-Diabetes terdiri atas jumlah entitas data yang jauh lebih besar, yaitu sebanyak 91.500 data, dibandingkan dengan kelas Diabetes yang hanya terdiri atas 8.500 data.

2. Pre-processing

Tahap prapemrosesan data diawali dengan proses pembersihan data (data cleaning). Hasil pemeriksaan menunjukkan bahwa dataset tidak mengandung nilai hilang (missing value). Dengan demikian, analisis dapat dilanjutkan ke tahap pembagian data (data splitting) menggunakan rasio 90:10, Proporsi sebesar 90% dari total dataset diimplementasikan sebagai himpunan latih (training data), sedangkan 10% residunya difungsikan sebagai himpunan uji (testing data).

Tabel 1. Distribusi Himpunan Data Training dan Data Testing

Dataset	Presentase	Jumlah Data
Data Training	90%	90.000
Data Testing	10%	10.000
Total		100.000

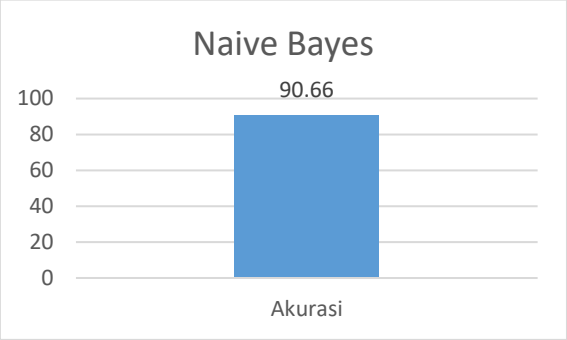
Sumber : Penelitian (2025)

Tabel 1. Merepresentasikan partisi data ke dalam dua sub-himpunan, yakni data latih (training) dan data uji (testing), yang masing-masing diimplementasikan pada fase pembelajaran sistem dan evaluasi performa model klasifikasi.. Dari keseluruhan 100.000 data, sebanyak 90.000 data (90%) dialokasikan sebagai data training, Data tersebut digunakan dalam fase pelatihan untuk mengonstruksi kapabilitas model dalam mengenali karakteristik dataset. Sebaliknya, 10.000 sampel sisanya (10%) dialokasikan sebagai instrumen pengujian untuk mengevaluasi efektivitas model pada domain data baru, sehingga dapat diketahui kemampuan model dalam mengeneralisasi pada data baru.

3. Klasifikasi

Setelah melalui serangkaian tahap prapemrosesan data (*data pre-processing*), yang mencakup pembersihan data, penanganan nilai hilang, normalisasi atribut, serta segmentasi basis data menjadi kategori training data dan testing data, proses klasifikasi kemudian dilakukan dengan mengimplementasikan algoritma Naive Bayes. Algoritma ini dipilih atas dasar kapabilitasnya dalam

menangani data berskala besar dengan efisiensi komputasi yang tinggi serta pendekatan probabilistik yang sederhana namun efektif. Dengan karakteristik tersebut, Naive Bayes dinilai sesuai dan layak digunakan dalam penelitian ini untuk mendukung proses identifikasi Diabetes Mellitus secara lebih optimal.



Sumber : Penelitian (2025)

Gambar 3. Visualisasi Hasil Akurasi

Berdasarkan hasil evaluasi yang diperoleh melalui fase pengujian terhadap data uji (testing data), model klasifikasi menunjukkan performansi akurasi yang signifikan sebesar 90,66%, yang mengindikasikan bahwa algoritma Naive Bayes memiliki performansi yang representatif dalam mengidentifikasi dan memprediksi pola-pola data yang terkait dengan kondisi diabetes mellitus. Capaian nilai akurasi tersebut mengonfirmasi bahwa model yang dikembangkan memiliki kapabilitas untuk mengeksekusi prosedur klasifikasi dengan derajat ketepatan yang tinggi dalam lingkup penelitian ini, serta berpotensi untuk diimplementasikan lebih lanjut dalam sistem pendukung keputusan medis guna membantu proses deteksi dini dan diagnosis penyakit diabetes secara lebih efektif dan efisien.

Tabel 2. Confusion Matrix

	True No Diabetes	True Diabetes	Class Precision
No Diabetes	8506	290	96,70%
Diabetes	644	560	46,51%
Class Recall	92,96%	65,88%	

Sumber : Penelitian (2025)

Pada Tabel 2, menyajikan hasil evaluasi kinerja model klasifikasi yang dikonstruksi berbasis algoritma Naive Bayes, dengan dua kategori kelas yaitu No Diabetes dan Diabetes. Evaluasi dilakukan menggunakan dua metrik utama dalam pengukuran kinerja model klasifikasi, yakni precision dan recall. Precision adalah ukuran seberapa akurat model memiliki kapabilitas untuk menghasilkan prediksi yang presisi terhadap kelas target tertentu, atau dengan kata lain, sejauh mana relevansi prediksi positif yang dihasilkan oleh model terhadap data aktual.

Di sisi lain, parameter recall berfungsi untuk mengevaluasi efektivitas model dalam mengidentifikasi keseluruhan entitas positif pada kelas tersebut; suatu aspek yang krusial terutama dalam domain deteksi patologi, seperti diabetes, di mana kesalahan dalam mengidentifikasi kasus positif dapat berimplikasi serius terhadap penanganan pasien. Dengan menggunakan kedua metrik ini, diperoleh representasi yang lebih komprehensif terkait efektivitas model dalam melakukan klasifikasi terhadap masing-masing kelas, serta mengidentifikasi potensi kekuatan dan kelemahan dari metodologi klasifikasi yang diimplementasikan dalam penelitian ini.

Berdasarkan data yang dipaparkan pada Tabel 2, tercatat nilai True Positive (TP) sebanyak 560 kasus, yang merepresentasikan jumlah pasien diabetes yang terklasifikasi secara akurat. Nilai True Negative (TN) sebesar 8.506 menunjukkan jumlah kasus non-diabetes yang berhasil diidentifikasi dengan benar oleh model. Sebaliknya, terdapat nilai False Positive (FP) sebanyak 644 kasus, di mana subjek non-diabetes secara keliru dikategorikan sebagai penderita diabetes. Terakhir, nilai False Negative (FN) sebesar 290 kasus mengindikasikan jumlah penderita diabetes yang salah terdeteksi sebagai non-diabetes. Oleh karena itu, Algoritma Naive Bayes sangat efektif dalam mendeteksi pasien diabetes.

KESIMPULAN

Berdasarkan hasil evaluasi terhadap model klasifikasi yang diterapkan menggunakan algoritma Naive Bayes, diperoleh kinerja yang menunjukkan tingkat ketepatan yang tinggi dalam mengklasifikasikan penyakit diabetes. Hal ini ditunjukkan oleh nilai akurasi sebesar 90,66% . Saran untuk penelitian selanjutnya dilakukan *data balancing* atau penerapan metode klasifikasi alternatif untuk meningkatkan deteksi pada kelas minoritas (diabetes), terutama dalam konteks penerapan medis yang menuntut sensitivitas tinggi terhadap kasus penyakit.

Untuk penelitian selanjutnya, disarankan agar dilakukan pengusulan dan eksplorasi terhadap algoritma klasifikasi lainnya, atau dilakukan perbandingan kinerja antar beberapa metode klasifikasi yang berbeda, Guna menghasilkan model dengan derajat akurasi yang lebih superior serta performansi yang lebih optimal. sehingga dapat dievaluasi kelebihan dan kekurangannya secara komprehensif dalam konteks klasifikasi penyakit diabetes. Selain itu, perlu dilakukan studi lanjutan yang berfokus pada efisiensi komputasi, seperti pengurangan kompleksitas algoritma atau penerapan teknik optimasi, untuk meminimalisir proses

komputasi yang berulang (*redundant computations*) yang berpotensi memperpanjang waktu pemrosesan dan menghambat efisiensi sistem secara keseluruhan. Dengan demikian, penelitian pada tahap selanjutnya diharapkan tidak hanya berfokus pada peningkatan tingkat akurasi model, tetapi juga mampu mengembangkan pendekatan klasifikasi yang lebih cepat, efisien, dan mudah diterapkan. Langkah tersebut diorientasikan untuk mendatangkan implikasi positif bagi penguatan sistem pengambilan keputusan berbasis data di sektor kesehatan, sehingga hasil penelitian dapat dimanfaatkan secara lebih optimal dalam konteks analisis dan pengambilan keputusan klinis.

REFERENSI

Azizah, Q. N. (2023). Klasifikasi Penyakit Daun Jagung Menggunakan Metode Convolutional Neural Network AlexNet. *Sudo Jurnal Teknik Informatika*, *2*(1), 28–33. https://doi.org/10.56211/sudo.v2i1.227

Gusrialni Fitri, N., Adilya, S., & Azizi, F. (2023). Comparison of the Naive Bayes Classification System and C4.5 for the Diagnosis of Stroke Perbandingan Sistem Klasifikasi Naive Bayes dan C4.5 Untuk Diagnosa Penyakit stoke. *SENTIMAS: Seminar Nasional Penelitian Dan Pengabdian Masyarakat*, 49–55.

Muntiari, N. R., & Hanif, K. H. (2022). Klasifikasi Penyakit Kanker Payudara Menggunakan Perbandingan Algoritma Machine Learning. *Jurnal Ilmu Komputer Dan Teknologi*, *3*(1), 1–6. https://doi.org/10.35960/ikomti.v3i1.766

Nasution, M. K., Saedudin, R. R., & Widartha, V. P. (2021). Perbandingan Akurasi Algoritma Naïve Bayes Dan Algoritma Xgboost Pada Klasifikasi Penyakit Diabetes. *E-Proceeding of Engineering*, *8*(5), 9765–9772. https://journal.ubpkarawang.ac.id/mahasiswa/index.php/ssj/article/view/424/338%0Ahttps://openlibrarypublications.telkomuniversity.ac.id/index.php/engineering/article/view/15759

Putra, F. K., Purnomo, A. S., Mercu, U., Yogyakarta, B., Informatika, J., Informasi, F. T., Mercu, U., Yogyakarta, B., Media, K., & Gigi, P. (2024). *Diagnosa Penyakit Gigi Di RSUD Solok Selatan Menggunakan Sistem Pakar Naïve Bayes*. *11*(3), 234–247.

R Lumbanraja, F., Lufiana, F., Heningtyas, Y., & Muludi, K. (2022). Implementasi Support Vector Machine (Svm) Untuk Klasifikasi Pederita Diabetes Mellitus. Jurnal Komputasi, 10(1), 75–83. https://doi.org/10.23960/komputasi.v10i1.2940

Rinanda, P. D., Delvika, B., Nurhidayarnis, S., Abror, N., & Hidayat, A. (2022). Perbandingan Klasifikasi Antara Naive Bayes dan K-Nearest Neighbor Terhadap Resiko Diabetes pada Ibu Hamil. MALCOM: Indonesian Journal of Machine Learning and Computer Science, 2(2), 68–75. https://doi.org/10.57152/malcom.v2i2.432

Rizki, R., Athallah, R., Cholissodin, I., & Adikara, P. P. (2022). Prediksi Potensi Pengidap Penyakit Diabetes berdasarkan Faktor Risiko Menggunakan Algoritme Kernel K-Nearest Neighbor. Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer, 6(8), 3777–3785. https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/11439

Suwitono, Y. A., & Kaunang, F. J. (2022). Implementasi Algoritma Convolutional Neural Network (CNN) Untuk Klasifikasi Daun Dengan Metode Data Mining SEMMA Menggunakan Keras. Jurnal Komtika (Komputasi Dan Informatika), 6(2), 109–121. https://doi.org/10.31603/komtika.v6i2.8054

Valerian, F. R., Syarief, M., Fatah, D. A., Informasi, S., Madura, U. T., Kamal, K., & Timur, J. (2025). Klasifikasi tingkat obesitas menggunakan metode gbm dan confusion matrix. 9(2), 2242–2249.

Widiasari, K. R., Wijaya, I. M. K., & Suputra, P. A. (2021). Diabetes Melitus Tipe 2: Faktor Risiko, Diagnosis, Dan Tatalaksana. Ganesha Medicine, 1(2), 114. https://doi.org/10.23887/gm.v1i2.40006