



Department of Digital Business

Journal of Artificial Intelligence and Digital Business (RIGGS)

Homepage: <https://journal.ilmudata.co.id/index.php/RIGGS>

Vol. 5 No. 2 (2026) pp: 2812-2819

P-ISSN: 2963-9298, e-ISSN: 2963-914X

Analisis Akurasi Dua Metode Klasifikasi: K-Nearest Neighbor vs Naïve Bayes pada Data Diabetes

Fidalina Nirigi, Mochammad Triyanto, Mohammad Rezza Pahlevi, Athia Saelan, Fadhlanrashif Ibrahim Supriana

Universitas Indonesia Membangun

fidalinanirigi585@gmail.com, mochiyani300717@gmail.com, mohammadrezzapahlevi@inaba.ac.id,
athia.saelan@inaba.ac.id, fadhlanrashif.ibrahim@inaba.ac.id

Abstract

Diabetes merupakan kondisi metabolik yang ditandai oleh tingginya kadar glukosa darah dan telah menjadi masalah kesehatan global. Apabila tidak ditangani dengan tepat, diabetes dapat menyebabkan komplikasi serius seperti penyakit kardiovaskular, stroke, kerusakan ginjal, mata, dan sistem saraf. Perkembangan teknologi machine learning memberikan peluang dalam membantu proses klasifikasi dan prediksi penyakit diabetes secara lebih cepat dan akurat. Penelitian ini bertujuan untuk menganalisis tingkat akurasi dua metode klasifikasi, yaitu K-Nearest Neighbor (KNN) dan Naïve Bayes pada data diabetes. Dataset yang digunakan adalah Pima Indians Diabetes Database dengan pembagian data sebesar 80% untuk data latih dan 20% untuk data uji. Tahapan penelitian meliputi preprocessing data, pelatihan model, dan pengujian klasifikasi. Variabel yang digunakan meliputi kadar glukosa, usia, indeks massa tubuh (BMI), tekanan darah, serta riwayat diabetes. Hasil penelitian menunjukkan bahwa kedua algoritma mampu melakukan klasifikasi data diabetes dengan baik. Namun, algoritma K-Nearest Neighbor memperoleh tingkat akurasi lebih tinggi sebesar 81%, sedangkan Naïve Bayes memperoleh akurasi sebesar 77%. Berdasarkan hasil tersebut, metode K-Nearest Neighbor dinilai lebih efektif dalam proses prediksi penyakit diabetes dibandingkan metode Naïve Bayes. Penelitian ini diharapkan dapat menjadi referensi dalam pengembangan sistem pendukung keputusan berbasis machine learning di bidang kesehatan, khususnya untuk deteksi dini penyakit diabetes.

Keywords: Klasifikasi, Diabetes, Machine Learning

1. Introduction

Machine learning telah mendorong berbagai inovasi dalam diagnosis penyakit, termasuk diabetes—suatu kondisi kronis dengan dampak kesehatan global yang besar [1]. Pendekatan seperti K-Nearest Neighbors dan Naïve Bayes banyak diterapkan untuk mengatasi kompleksitas diagnosis. Menurut data Federasi Diabetes Internasional, terdapat 382 juta penderita diabetes di dunia, dan angka ini diperkirakan meningkat menjadi 592 juta pada tahun 2035 [2]. Menjaga keseimbangan gula darah menjadi sangat krusial, karena kadar gula yang tinggi dapat menyebabkan diabetes serta komplikasi seperti penyakit jantung, stroke, obesitas, serta kerusakan mata, ginjal, dan saraf. Pengelolaan gaya hidup dan pemantauan glukosa darah secara proaktif dapat menurunkan risiko komplikasi tersebut. Dengan dukungan pengetahuan dan teknologi, tantangan diabetes dapat diatasi melalui deteksi awal dan pengobatan yang tepat, sehingga meningkatkan kualitas hidup pasien serta memberikan kontribusi bagi kesehatan masyarakat secara luas [3]. Diabetes tipe 1 merupakan penyakit autoimun di mana tubuh menghancurkan sel-sel penghasil insulin yang diperlukan untuk menyerap gula menjadi energi. Jenis ini dapat menyebabkan obesitas dan umumnya muncul pada masa kanak-kanak atau remaja. Diabetes tipe 2 biasanya menyerang orang dewasa yang mengalami obesitas, di mana tubuh gagal memproduksi insulin secara memadai. Tipe ini sering terjadi pada kelompok usia paruh baya atau lanjut. Selain itu, faktor lain seperti infeksi bakteri atau virus, paparan racun atau zat kimia dalam makanan, reaksi autoimun, obesitas, pola makan buruk, perubahan gaya hidup, kebiasaan makan, serta polusi lingkungan juga dapat memicu diabetes. Diabetes selanjutnya dapat menimbulkan berbagai penyakit seperti komplikasi kardiovaskular, gangguan ginjal, retinopati, ulkus kaki, dan lain-lain. Algoritma machine learning umumnya digunakan untuk menemukan pola tersembunyi dalam kumpulan data besar serta menghasilkan prediksi yang diinginkan [1].

Diabetes menjadi perhatian penting bagi praktisi medis karena adanya ketergantungan kompleks antar berbagai faktor. Penyakit ini mempengaruhi organ seperti ginjal, mata, jantung, saraf, dan kaki [2]. Banyak metode klasifikasi yang telah diterapkan untuk mendiagnosis diabetes, misalnya *support vector machine* (SVM) untuk klasifikasi berdasarkan data medis, teorema Bayes untuk memprediksi akurasi, serta pohon keputusan dan Naïve Bayes untuk menganalisis pola diagnosis diabetes [4]. Dalam penelitian ini, kami menggunakan metode klasifikasi K-Nearest Neighbor dan Naïve Bayes untuk mengukur akurasi dalam diagnosis diabetes.

Dalam penelitian mengenai klasifikasi penyakit diabetes, pemilihan metode klasifikasi yang tepat menjadi faktor penting untuk memperoleh tingkat akurasi prediksi yang optimal. Algoritma klasifikasi seperti K-Nearest Neighbor (K-NN) dan Naïve Bayes banyak digunakan dalam penelitian data mining karena memiliki kemampuan dalam mengolah data kesehatan dan menghasilkan prediksi yang cukup akurat. Diabetes mellitus merupakan penyakit kronis yang jumlah penderitanya terus meningkat setiap tahun, sehingga diperlukan sistem pendeteksian dini berbasis teknologi untuk membantu proses klasifikasi data pasien diabetes.[1].

Diabetes melitus merupakan salah satu penyakit kronis yang terus mengalami peningkatan jumlah penderita setiap tahun, baik di tingkat nasional maupun global. Penyakit ini terjadi akibat gangguan metabolisme yang ditandai dengan meningkatnya kadar glukosa dalam darah karena tubuh tidak mampu memproduksi atau menggunakan insulin secara optimal. Kondisi tersebut menyebabkan diabetes menjadi salah satu penyakit yang memerlukan penanganan serius karena dapat menimbulkan berbagai komplikasi, seperti gangguan ginjal, penyakit jantung, hingga kerusakan saraf.[2]

Berdasarkan penelitian Ginting et al., [10], berbagai metode machine learning mampu memberikan hasil prediksi yang baik terhadap penyakit diabetes melitus tipe 2. Penelitian tersebut menjelaskan bahwa faktor usia, obesitas, riwayat keluarga, dan kurangnya aktivitas fisik menjadi faktor risiko utama penyebab diabetes. Selain itu, penerapan machine learning dinilai mampu membantu proses pencegahan serta mempercepat penanganan penyakit diabetes melalui prediksi dini.

Penelitian yang dilakukan oleh Puji Dwi Rinanda dkk. dengan judul *Perbandingan Klasifikasi Antara Naive Bayes dan K-Nearest Neighbor Terhadap Risiko Diabetes pada Ibu Hamil* membahas penerapan teknik data mining untuk mengklasifikasikan risiko diabetes pada ibu hamil menggunakan algoritma Naive Bayes dan K-Nearest Neighbor (KNN). Penelitian ini dilatarbelakangi oleh tingginya risiko diabetes gestasional yang dapat membahayakan kesehatan ibu maupun janin apabila tidak dideteksi sejak dini. [12]

Dalam penelitian tersebut, data yang digunakan diperoleh dari dataset diabetes ibu hamil yang kemudian diproses menggunakan tahapan data preprocessing, pembagian data latih dan data uji, serta pengujian menggunakan metode K-Fold Cross Validation. Algoritma Naive Bayes digunakan karena mampu melakukan klasifikasi berdasarkan probabilitas, sedangkan algoritma K-Nearest Neighbor bekerja dengan menghitung kedekatan jarak antar data untuk menentukan hasil klasifikasi.

Hasil penelitian menunjukkan bahwa algoritma Naive Bayes memiliki performa yang lebih baik dibandingkan K-Nearest Neighbor dalam klasifikasi risiko diabetes pada ibu hamil. Pengujian menggunakan K-Fold Cross Validation dengan nilai K=10 menghasilkan tingkat akurasi sebesar 75,78% pada algoritma Naive Bayes, sedangkan algoritma KNN memperoleh akurasi sebesar 74,48% dengan nilai K=25. Berdasarkan hasil tersebut, algoritma Naive Bayes dinilai lebih efektif dalam melakukan klasifikasi risiko diabetes pada ibu hamil.

Penelitian ini menunjukkan bahwa penerapan data mining dan machine learning dapat membantu tenaga medis dalam melakukan deteksi dini terhadap risiko diabetes secara lebih cepat dan efisien. Selain itu, penggunaan algoritma klasifikasi seperti Naive Bayes dan KNN dapat dijadikan sebagai sistem pendukung keputusan dalam bidang kesehatan, khususnya untuk analisis risiko penyakit diabetes pada ibu hamil.

Penelitian yang dilakukan oleh M. Iqbal, R. Saputra, dan N. Hidayat dengan judul *Implementasi Algoritma Naive Bayes Untuk Prediksi Penyakit Diabetes Mellitus* membahas penerapan algoritma Naive Bayes dalam memprediksi penyakit diabetes mellitus menggunakan teknik data mining. Penelitian ini bertujuan untuk membantu proses diagnosis dini terhadap penyakit diabetes sehingga penanganan dapat dilakukan lebih cepat dan tepat.[13]

Dalam penelitian tersebut, algoritma Naïve Bayes digunakan karena memiliki kemampuan klasifikasi berbasis probabilitas yang sederhana, cepat, dan efektif dalam mengolah data medis. Data pasien diproses melalui beberapa tahapan, mulai dari preprocessing data, pembagian data latih dan data uji, hingga proses klasifikasi untuk menentukan apakah pasien berpotensi menderita diabetes mellitus atau tidak. Variabel yang digunakan meliputi kadar glukosa, tekanan darah, usia, indeks massa tubuh (BMI), dan faktor kesehatan lainnya yang berkaitan dengan diabetes.

Hasil penelitian menunjukkan bahwa algoritma Naïve Bayes mampu memberikan performa yang baik dalam proses prediksi penyakit diabetes mellitus dengan tingkat akurasi yang cukup tinggi. Penelitian ini membuktikan bahwa metode data mining dapat dimanfaatkan sebagai sistem pendukung keputusan dalam bidang kesehatan, khususnya untuk membantu tenaga medis dalam melakukan deteksi dini penyakit diabetes secara lebih efektif dan efisien. Selain itu, penggunaan algoritma Naïve Bayes dinilai mampu mempercepat proses analisis data pasien dibandingkan metode diagnosis secara manual.

Penelitian yang dilakukan oleh A. Prasetyo dan R. Nugroho dengan judul *Perbandingan Metode KNN dan Naïve Bayes Pada Klasifikasi Penyakit Diabetes* membahas perbandingan performa dua algoritma data mining, yaitu K-Nearest Neighbor (KNN) dan Naïve Bayes, dalam mengklasifikasikan penyakit diabetes. Penelitian ini dilakukan untuk mengetahui metode yang memiliki tingkat akurasi terbaik dalam proses prediksi penyakit diabetes berdasarkan data pasien.[14]

Dalam penelitian tersebut, data pasien diproses menggunakan teknik klasifikasi untuk menentukan apakah seseorang termasuk kategori penderita diabetes atau tidak. Algoritma KNN bekerja dengan menghitung kedekatan jarak antar data sehingga klasifikasi dilakukan berdasarkan tetangga terdekat, sedangkan algoritma Naïve Bayes menggunakan pendekatan probabilitas untuk menentukan hasil klasifikasi. Kedua metode diuji menggunakan data medis yang meliputi kadar gula darah, tekanan darah, usia, indeks massa tubuh (BMI), dan faktor kesehatan lainnya.

Hasil penelitian menunjukkan bahwa metode KNN dan Naïve Bayes sama-sama mampu digunakan dalam klasifikasi penyakit diabetes dengan tingkat akurasi yang cukup baik. Namun, terdapat perbedaan performa antara kedua algoritma berdasarkan hasil pengujian yang dilakukan. Penelitian ini membuktikan bahwa penerapan data mining dalam bidang kesehatan dapat membantu proses diagnosis penyakit diabetes secara lebih cepat, efektif, dan efisien. Selain itu, hasil penelitian dapat dijadikan sebagai referensi dalam pengembangan sistem pendukung keputusan untuk deteksi dini penyakit diabetes.

Penelitian yang dilakukan oleh Fitriyani dengan judul *Prediksi Diabetes Menggunakan Algoritma Naïve Bayes dan Greedy Forward Selection* membahas penerapan teknik data mining untuk memprediksi penyakit diabetes menggunakan algoritma Naïve Bayes yang dikombinasikan dengan metode seleksi fitur Greedy Forward Selection. Penelitian ini bertujuan untuk meningkatkan akurasi prediksi diabetes dengan memilih atribut atau fitur yang paling relevan sehingga performa model klasifikasi menjadi lebih optimal. [15]

Dalam penelitian tersebut digunakan dataset publik diabetes yang kemudian diproses melalui tahapan preprocessing, seleksi fitur, dan klasifikasi data. Algoritma Naïve Bayes digunakan karena memiliki proses perhitungan yang sederhana dan cepat dalam melakukan klasifikasi berbasis probabilitas, sedangkan metode Greedy Forward Selection digunakan untuk menyeleksi fitur-fitur penting yang berpengaruh terhadap hasil prediksi. Dengan menghilangkan atribut yang tidak relevan, performa model dapat meningkat secara signifikan.

Hasil penelitian menunjukkan bahwa kombinasi algoritma Naïve Bayes dan Greedy Forward Selection mampu menghasilkan tingkat akurasi sebesar 91,73%, sedangkan algoritma Naïve Bayes tanpa seleksi fitur hanya memperoleh akurasi sebesar 87,69%. Hasil tersebut membuktikan bahwa penggunaan metode seleksi fitur dapat meningkatkan performa klasifikasi dalam prediksi penyakit diabetes. Penelitian ini juga menunjukkan bahwa penerapan machine learning dapat membantu tenaga medis dalam melakukan deteksi dini penyakit diabetes secara lebih cepat, efektif, dan efisien.

2. Research Methods

Penelitian yang dilakukan sebelumnya dengan berbagai metode yang berbeda. Dalam penelitian [5] memprediksi dengan algoritma Naïve Bayes total pengumpulan data sampel adalah 62,5% dan efisiensi prediksi Dukungan

mesin vektor untuk kumpulan data sampel adalah 82%. menyimpulkan bahwa algoritma mesin vektor Dukungan adalah lebih efisien daripada algoritma Naive Bayes dalam memprediksi probabilitas penderita diabetes atas sampel dataset. [4] melakukan penelitian dengan menggunakan berbagai jenis algoritma KNN. Hasil simulasi menunjukkan bahwa KNN jenis Fine, Medium, Cubic, dan Weighted memiliki kinerja yang unggul. Model pengklasifikasi dari semua jenis KNN dibutuhkan kurang dari 0,7 detik untuk memprediksi target. Selain itu, dapat dikatakan bahwa algoritma Fine KNN cocok untuk menyelesaikan masalah klasifikasi diabetes dengan akurasi yang lebih tinggi dari pada tipe yang lain. Pada penelian [6] Algoritma klasifikasikan berdasarkan indikasi apakah seseorang menderita diabetes atau tidak. Model ini mencapai akurasi sekitar 81%. Sistem yang diusulkan juga mendukung streaming langsung input data, di mana hasilnya diperoleh secara real-time untuk pasien yang masuk.

Penelitian yang dilakukan dalam [7] yaitu menggunakan model baru berdasarkan teknik penambahan data untuk memprediksi diabetes mellitus tipe 2 (T2DM). Masalah utama yang coba pecahkan adalah meningkatkan akurasi model prediksi, dan membuatnya model yang adaptif terhadap lebih dari satu himpunan data. Berdasarkan serangkaian prosedur prapemrosesan, modelnya adalah terdiri dari dua bagian, algoritma K-means yang ditingkatkan dan algoritma regresi logistik. Hasil yang didapatkan K-means adalah 77% dan algoritma regresi logistic adalah 72%. Penelitian yang dilakukan oleh [8] mengeksplorasi prediksi awal diabetes menggunakan berbagai teknik data mining. Dataset yang digunakan mengambil 768 instans dari PIMA Himpunan Data India untuk menentukan keakuratan penambahan data teknik dalam prediksi. Analisis membuktikan bahwa Modified J48 Pengklasifikasi memberikan akurasi tertinggi daripada teknik lainnya. Hasil dari Naive bayes adalah 76,30%.

Penelitian sebelumnya yang dilakukan oleh [9] menunjukkan bahwa algoritma Naive Bayes memperoleh nilai akurasi sebesar 84,07%, sedangkan algoritma K-Nearest Neighbor memperoleh nilai akurasi sebesar 85,99% pada dataset diabetes sebelum dilakukan optimasi. Setelah diterapkan metode optimasi Particle Swarm Optimization (PSO), akurasi Naive Bayes meningkat menjadi 89,01% dan K-Nearest Neighbor meningkat menjadi 94,51%. Hasil tersebut menunjukkan bahwa algoritma K-Nearest Neighbor memiliki performa yang lebih baik dibandingkan Naive Bayes dalam proses klasifikasi data diabetes.

Selain itu, penelitian [11] juga menjelaskan bahwa algoritma K-Nearest Neighbor sangat efektif digunakan dalam klasifikasi data kesehatan, khususnya penyakit diabetes, karena mampu mengidentifikasi kedekatan antar data berdasarkan perhitungan jarak tertentu. Penelitian tersebut membandingkan beberapa metode pengukuran jarak pada algoritma KNN, yaitu Euclidean, Manhattan, dan Chebyshev. Hasil penelitian menunjukkan bahwa metode Manhattan Distance menghasilkan tingkat akurasi tertinggi sebesar 98,17%, dibandingkan Euclidean sebesar 96,33% dan Chebyshev sebesar 93,58%. Temuan ini membuktikan bahwa performa algoritma KNN sangat dipengaruhi oleh metode pengukuran jarak yang digunakan dalam proses klasifikasi data.

2.1. Metodologi Yang Diusulkan

Dalam penelitian ini, terdapat 4 tahapan penelitian yang dimulai dari pengumpulan data, preprocessing, processing, dan evaluasi. Rancangan model pada penelitian ini seperti pada Fig.1

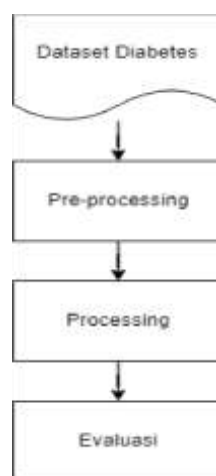


Fig. 1 Workflow Penelitian

2.2. Pengumpulan Dataset

Pada penelitian menggunakan dataset Pima Indians Diabetes Database yang digunakan penelitian [7] dan [8] untuk melakukan proses diagnosis penyakit diabetes. Total sampel dataset ini adalah 768 dan memiliki 9 atribut. Dataset yang digunakan untuk melakukan proses diagnosis penyakit diabetes berasal dari website Kaggle.

TABLE. 1 DATASET ATRIBUT

BMI	Index massa tubuh
DiabetesPedigreeFuction	Indikator riwayat diabetes dalam keluarga
Age	Umur
Outcome	Efek jangka panjang

2.3. Preprocessing Data

Preprocessing data termasuk peran terpenting dalam menemukan data. Langkah -langkah yang penting dalam preprocessing data adalah membersihkan data, integrasi data, reduksi data dan transformasi data pada fig.2.

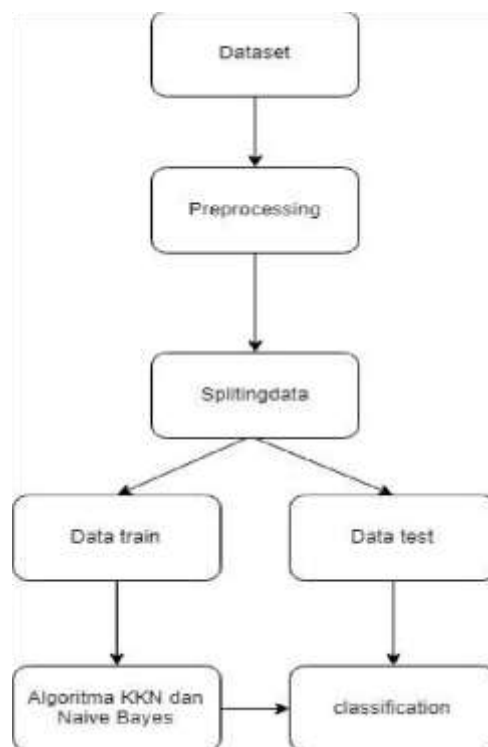


Fig2. Model KNN dan Naïve Bayes

2.4. Klasifikasi

Pada tahapan klasifikasi menggunakan algoritma KNN dan Naïve bayes untuk mencari tingkat akurasi tertinggi dari tiap tiap algoritma dalam diagnosa penyakit diabetes menggunakan 9 atribut pada dataset yang digunakan. Untuk pengukuran tingkat akurasi dari masing-masing algoritma menggunakan Confusion matrix untuk pengukuran performa untuk klasifikasi pada Table 1. Berdasarkan hasil pengujian, diperoleh bahwa algoritma K-Nearest Neighbor memiliki tingkat akurasi lebih tinggi dibandingkan algoritma Naïve Bayes, yaitu sebesar 81%,

sedangkan Naïve Bayes memperoleh akurasi sebesar 77%. Hasil tersebut menunjukkan bahwa algoritma KNN lebih efektif digunakan dalam proses klasifikasi data diabetes pada penelitian ini.

Keunggulan algoritma KNN terletak pada kemampuannya dalam mengenali pola kedekatan antar data melalui perhitungan jarak tertentu. Pada dataset diabetes, atribut seperti kadar glukosa, usia, tekanan darah, dan indeks massa tubuh (*Body Mass Index / BMI*) memiliki hubungan yang cukup kuat, sehingga pendekatan berbasis jarak pada algoritma KNN mampu menghasilkan klasifikasi yang lebih optimal. Sementara itu, algoritma Naïve Bayes menggunakan pendekatan probabilitas dengan asumsi bahwa setiap atribut bersifat independen, sehingga pada beberapa kondisi data kesehatan hasil klasifikasinya menjadi kurang maksimal dibandingkan algoritma KNN.

		True Class	
		Positive	Negative
Predicted Class	Positive	TP	FP
	Negative	FN	TN

Fig 3. Confusion Matrix

Evaluasi model dilakukan dengan menghitung nilai akurasi, *recall*, *precision*, dan *F1-score*. Dari *confusion matrix* diperoleh empat kategori: True Positive (TP), True Negative (TN), False Positive (FP), dan False Negative (FN).

3. Results and Discussions

3.1. Hasil Pengujian KNN

Dalam penerapan dua algoritma seperti KNN dan Naïve Bayes hasil akurasi untuk mendeteksi penyakit diabetes berbeda-beda. Penggunaan data training dan data testing untuk mencari skor akurasi terhadap algoritma tersebut sebesar 80% untuk data training dan 20%. Setelah melakukan proses klasifikasi menggunakan model algoritma KNN .

Tabel 2. Hasil precision recall dan f1 score

Precision	0.72
Recall	0.57
F1 score	0.64

Melihat hasil pada Fig. 2, selanjutnya dilakukan proses perhitungan untuk mengetahui tingkat performa model algoritma K-Nearest Neighbor (KNN) pada data pengujian. Tabel 2 menunjukkan hasil evaluasi performa model berdasarkan nilai *precision*, *recall*, *F1-score*, dan akurasi. Hasil pengujian menunjukkan bahwa algoritma KNN memperoleh nilai *precision* sebesar 0,72, *recall* sebesar 0,57, dan *F1-score* sebesar 0,64. Selain itu, tingkat akurasi yang dihasilkan mencapai 81%. Hasil tersebut menunjukkan bahwa metode KNN mampu melakukan klasifikasi data diabetes dengan cukup baik karena bekerja berdasarkan kedekatan jarak antar data, sehingga mampu mengenali pola data pasien diabetes secara lebih optimal.

Tabel 3. Hasil Akurasi KNN

<i>Model</i>	Testing Accuracy
KNN	81%

3.2. Hasil Pengujian Naïve bayes

Hasil Pengujian algoritma K-nearest neighbors menghitung tingkat keakuratan algoritma K-nearest neighbors pada data adalah sebesar 81% pada tabel 3. Proses klasifikasi menggunakan model algoritma Naïve bayes menghasilkan precision, recall dan f1 score pada tabel 4.

Tabel 4. Hasil precision, recall dan f1 score

Precision	0.65
Recall	0.55
F1 score	0.59

Berdasarkan hasil pada Tabel 4, selanjutnya dilakukan perhitungan untuk mengetahui tingkat performa model algoritma Naïve Bayes pada data pengujian. Tabel 5 menunjukkan hasil evaluasi akurasi model berdasarkan nilai *precision*, *recall*, *F1-score*, dan akurasi. Hasil pengujian menunjukkan bahwa algoritma Naïve Bayes memperoleh nilai *precision* sebesar 0,65, *recall* sebesar 0,55, serta *F1-score* sebesar 0,59. Selain itu, tingkat akurasi yang dihasilkan mencapai 77%. Berdasarkan hasil tersebut, algoritma Naïve Bayes dinilai mampu melakukan klasifikasi data diabetes dengan cukup baik. Namun, tingkat akurasi yang diperoleh masih lebih rendah dibandingkan algoritma K-Nearest Neighbor (KNN) dalam proses klasifikasi data diabetes.

Tabel 5. Hasil Akurasi Naïve bayes

<i>Model</i>	Testing Accuracy
Naïve bayes	77%

Hasil Pengujian algoritma Naïve bayes menghitung tingkat keakuratan algoritma K-nearest neighbors pada data adalah sebesar 77% pada tabel 5. Perbandingan dari penelitian terdahulu yang menggunakan model KNN dan Naïve Bayes dapat dilihat pada tabel 6.

Tabel 6. Hasil penelitian terdahulu

Source	Model	Akurasi
[7]	KNN	77%
[8]	Naïve Bayes	76%

4. Conclusion

Tujuan dari penelitian ini untuk membandingkan skor akurasi dengan algoritma K-nearest neighbors dan Naïve bayes untuk memprediksi penyakit diabetes dengan dataset Classification with Diabetes. Dalam penelitian ini algoritma K-nearest neighbors memiliki skor tinggi dibandingkan menggunakan algoritma Naïve bayes dalam memprediksi penyakit diabetes dengan hasil skor akurasi algoritma Knearest neighbors 81% dan algoritma Naïve

bayes 77%. Dalam penelitian terdahulu yang mendapatkan skor akurasi dengan model KNN dengan hasil 77% dan Naïve Bayes 76 %

Reference

- [1] S. K. Dey, A. Hossain, dan M. M. Rahman, "Implementation of a Web Application to Predict Diabetes Disease: An Approach Using Machine Learning Algorithm," *2018 21st Int. Conf. Comput. Inf. Technol. ICCIT 2018*, hal. 1–5, 2019, doi: 10.1109/ICCITECHN.2018.8631968.
- [2] M. Komi, J. Li, Y. Zhai, dan Z. Xianguo, "Application of data mining methods in diabetes prediction," *2017 2nd Int. Conf. Image, Vis. Comput. ICIVC 2017*, no. S IX, hal. 1006–1010, 2017, doi: 10.1109/ICIVC.2017.7984706.
- [3] A. M. Argina, "Penerapan Metode Klasifikasi K-Nearest Neighbor pada Dataset Penderita Penyakit Diabetes," *Indones. J. Data Sci.*, vol. 1, no. 2, hal. 29–33, 2020, doi: 10.33096/ijodas.v1i2.11. [4] A. Ali, M. Alrubei, L. F. M. Hassan, M. Al-Ja'afari, dan S. Abdulwahed, "Diabetes classification based on KNN," *IJUM Eng. J.*, vol. 21, no. 1, hal. 175–181, 2020, doi: 10.31436/iiumej.v21i1.1206.
- [5] R. S. Raj, D. S. Sanjay, M. Kusuma, dan S. Sampath, "Comparison of Support Vector Machine and Naïve Bayes Classifiers for Predicting Diabetes," *1st Int. Conf. Adv. Technol. Intell. Control. Environ. Comput. Commun. Eng. ICATIECE 2019*, hal. 41–45, 2019, doi: 10.1109/ICATIECE45860.2019.9063792.
- [6] K. Shah, R. Punjabi, P. Shah, dan M. Rao, "Real Time Diabetes Prediction using Naïve Bayes Classifier on Big Data of Healthcare," *Int. Res. J. Eng. Technol.*, vol. 07, no. 05, hal. 102–107, 2020.
- [7] H. Wu, S. Yang, Z. Huang, J. He, dan X. Wang, "Type 2 diabetes mellitus prediction model based on data mining," *Informatics Med. Unlocked*, vol. 10, no. August 2017, hal. 100–107, 2018, doi: 10.1016/j.imu.2017.12.006.
- [8] M. R. ASengamuthu, M. R. Birami, dan ..., "Various Data Mining Techniques Analysis to Predict Diabetes Mellitus," *Int Res J Eng Technol ...*, 2018, [Daring]. Tersedia pada: https://www.academia.edu/download/57013446/IRJ_ETV515134.pdf.
- [9] A. Fauzi and A. H. Yunial, "Optimasi Algoritma Klasifikasi Naive Bayes, Decision Tree, K – Nearest Neighbor, dan Random Forest menggunakan Algoritma Particle Swarm Optimization pada Diabetes Dataset," *J. Edukasi dan Penelit. Inform.*, vol. 8, no. 3, p. 470, 2022, doi: 10.26418/jp.v8i3.56656.
- [10] R. G. Ginting, E. Girsang, J. B. Ginting, and H. Hartono, "Analisis Determinan Dan Prediksi Penyakit Diabetes Melitus Tipe 2 Menggunakan Metode Machine Learning: Scoping Review," *J. Matern. Kebidanan*, vol. 7, no. 1, pp. 58–72, 2022, doi: 10.34012/jumkep.v7i1.2538.
- [11] S. Anggraini, M. A. Hamzah, and F. Ariyanto, "Analisis Perbandingan Jarak Euclidean, Manhattan, Dan Chebyshev Distance Pada Algoritma KNN Berdasarkan Data Diabetes Di Puskesmas Larangan," vol. 11, no. 1, pp. 617–624, 2025.
- [12] 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: Comparison of Classification Between Naive Bayes and K-Nearest Neighbor on Diabetes Risk in Pregnant Women. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 2(2), 68-75. <https://doi.org/10.57152/malcom.v2i2.432>
- [13] M. Iqbal, R. Saputra, dan N. Hidayat, "Implementasi Algoritma Naïve Bayes Untuk Prediksi Penyakit Diabetes Mellitus," *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 8, no. 4, hal. 455–462, 2021.
- [14] A. Prasetyo dan R. Nugroho, "Perbandingan Metode KNN dan Naïve Bayes Pada Klasifikasi Penyakit Diabetes," *Jurnal Sistem Informasi dan Teknologi*, vol. 10, no. 1, hal. 33–40, 2022.
- [15] Fitriyani Fitriyani, " **Prediksi Diabetes Menggunakan Algoritma Naive Bayes dan Greedy Forward Selection,**" <https://doi.org/10.25077/TEKNOSI.v7i2.2021.61-69>