

Perbandingan Implementasi Machine Learning Menggunakan Metode KNN, Naive Bayes, Dan Logistik Regression Untuk Mengklasifikasi Penyakit Diabetes

^{1*}Dewi Nasien, Ricalvin Darwin,² Alexander Cia³, Andrian Leo Winata⁴, Jerry Go⁵, Richard M.C⁶, Ryan Charles Wijaya⁷, Kevin Charles Lo⁸

¹⁻⁸ Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Institut Bisnis dan Teknologi Pelita Indonesia
e-mail: ¹[*¹dewinasien@lecturer.pelitaindonesia.ac.id](mailto:dewinasien@lecturer.pelitaindonesia.ac.id), ²ricalvin.darwin@student.pelitaindonesia.ac.id,

³alexander.cia@student.pelitaindonesia.ac.id, ⁴andrian.leo@student.pelitaindonesia.ac.id,

⁵jerry.go@student.pelitaindonesia.ac.id, ⁶richard@student.pelitaindonesia.ac.id, ⁷ryan.charles@student.pelitaindonesia.ac.id,

⁸kevin.charles@student.pelitaindonesia.ac.id

Abstract – Diabetes is in the spotlight because of its chronic nature, with the main symptom being an increase in blood sugar levels above normal limits. Diabetes occurs when the body cannot efficiently take glucose into cells to be converted into energy, causing extra sugar to accumulate in the bloodstream. This research uses feature extraction with Principal Component Analysis (PCA) with a threshold of 80%, producing 5 main features. These features are then used as input for three classifiers, namely K-Nearest Neighbors (KNN), Naive Bayes, and Logistic Regression. The data used comes from Kaggle, with a data split of 70:30 and 80:20. The research results show that the Naive Bayes method provides the best accuracy, reaching 79% at an 80:20 data split. Therefore, it can be concluded that the Naive Bayes algorithm is the best choice for diabetes data classification in this study.

Keywords: Feature_extraction, pca, diabetes, knn, bayes, logistic_regression

Abstrak – Penyakit diabetes menjadi sorotan karena sifatnya yang kronis, dengan gejala utama berupa peningkatan kadar gula darah di atas batas normal. Diabetes terjadi ketika tubuh tidak dapat efisien mengambil glukosa ke dalam sel untuk diubah menjadi energi, menyebabkan penumpukan gula ekstra dalam aliran darah. Penelitian ini menggunakan ekstraksi fitur dengan Analisis Komponen Utama (Principal Component Analysis - PCA) dengan threshold 80%, menghasilkan 5 fitur utama. Fitur-fitur ini kemudian digunakan sebagai input untuk tiga classifier, yaitu K-Nearest Neighbors (KNN), Naive Bayes, dan Regresi Logistik. Data yang digunakan berasal dari Kaggle, dengan pembagian data 70:30 dan 80:20. Hasil penelitian menunjukkan bahwa metode Naive Bayes memberikan akurasi terbaik, mencapai 79% pada pembagian data 80:20. Oleh karena itu, dapat disimpulkan bahwa algoritma Naive Bayes adalah pilihan terbaik untuk klasifikasi data diabetes dalam penelitian ini.

Kata Kunci – Ekstraksi_fitur, pca, diabetes, knn, bayes ,regresi_logistik.

I. PENDAHULUAN

Diabetes merupakan penyakit yang ditandai dengan kenaikan kadar gula dalam darah. Penyakit ini bisa disebabkan melalui keturunan, ataupun bisa dari racun makanan dan infeksi. Penyakit ini menyebabkan gangguan yang terjadi pada mata, ginjal, syaraf, dan pembuluh darah [1]

Algoritma klasifikasi menggunakan dataset pelatihan untuk membentuk suatu model. Model yang telah dibuat ini selanjutnya digunakan untuk meramalkan label kelas dari data baru yang belum diketahui. Klasifikasi merupakan fungsi dalam data mining yang menciptakan suatu model untuk meramalkan kelas atau kategori dari objek-objek dalam basis data. Proses klasifikasi melibatkan dua tahap, yaitu pembelajaran dan pengklasifikasian [2]

Pengelompokan data merupakan tahap dimana data diatur dan dikelompokkan ke dalam variasi jenis, bentuk, atau kelas yang berbeda. Proses ini memungkinkan pemisahan dan kategorisasi data sesuai dengan kebutuhan koleksi data untuk berbagai keperluan, baik bisnis maupun pribadi. Secara khusus, ini mencerminkan langkah dalam manajemen data [3]

Ekstraksi informasi bisa diterapkan menggunakan teknik machine learning, dan banyak algoritma machine learning telah sukses diaplikasikan untuk mendeteksi kasus diabetes (Karo & Hendriyana, 2022).

Metode klasifikasi dalam supervised learning pada machine learning ada banyak sekali, diantaranya yaitu K-Nearest Neighbor (KNN), Naive Bayes Classifier (NBC), Support Vector Machine (SVM), Neural Network (NN), Random Forest Classifier (RFC), Ada Boost Classifier (ABC), serta Quadratic Discriminant Analysis (QDA) [4]

Naïve Bayes adalah metode klasifikasi dengan perhitungan peluang. Naïve Bayes adalah metode supervised learning. Naïve Bayes diketahui mempunyai tingkat akurasi yang baik dengan perhitungan sederhana [5]

Naïve Bayes Classifier (NBC) adalah bentuk yang sederhana dari jaringan Bayesian dalam bidang data mining. Kesederhanaan metode ini didasarkan pada asumsi bahwa semua fitur bersifat independen satu sama lain. NBC menonjol karena akurasi tinggi dan tingkat kesalahan klasifikasi yang rendah [6]

K-Nearest Neighbor berfungsi untuk mengklasifikasikan subjek baru berdasarkan data sampel pelatihan dan atribut. Proses ini melibatkan pengelompokan hasil sampel uji baru sesuai dengan mayoritas kategori yang terdapat dalam K-NN [7]

Algoritma ini bekerja berdasarkan jarak terdekat dari data latih ke data uji untuk menetapkan K-Nearest Neighbor. Sesudah mengumpulkan K- Nearest Neighbors, lalu diambil sebagian besar K- Nearest Neighbors untuk dibuat prediksi dari sampel uji [5]

Logistic Regression adalah metode yang digunakan untuk menganalisis data yang menggambarkan antara satu variabel respon (dependen) atau lebih variabel predictor. Logistik Regression digunakan pada saat variabel predictor (y) memiliki skala kategorik atau nominal yang terdiri dari dua atau lebih kategori. Oleh karena itu, metode ini dikembangkan dengan tujuan memastikan bahwa, tak peduli apapun perkiraannya, selalu berada dalam rentang antara 0 hingga 1 [8]. Metode regresi logistik digunakan untuk menemukan korelasi antara variabel respons yang memiliki sifat dikotomis (ordinal atau nominal) atau polikotomis (memiliki skala nominal atau ordinal dengan lebih dari 2 kategori) [9]

Penelitian ini akan mendalami tentang perbandingan kinerja tiga metode klasifikasi, yaitu K-Nearest Neighbor (KNN), Naive Bayes, dan Logistic Regression, khususnya dalam konteks klasifikasi data diabetes. Fokus utama penelitian ini adalah pada evaluasi ketiga metode tersebut dengan menggunakan metode ekstraksi fitur Principal Component Analysis (PCA) dengan threshold 80%. Output dari ekstraksi fitur ini akan dijadikan sebagai input untuk KNN, Naive Bayes, dan Logistic Regression. Data yang digunakan akan dibagi menjadi dua skenario, masing-masing 70-30 dan 80-20, mencerminkan proporsi data latih dan uji yang umum digunakan dalam eksperimen klasifikasi.

Evaluasi kinerja akan menggunakan confusion matrix pada kedua skenario pembagian data, memberikan gambaran yang jelas tentang kemampuan ketiga metode dalam mengklasifikasikan kasus diabetes. Analisis akurasi prediksi akan menyoroti ketepatan dan keakuratan hasil klasifikasi dari masing-masing metode. Dengan pendekatan ini, penelitian ini bertujuan untuk memberikan pemahaman mendalam tentang potensi dan kelemahan metode klasifikasi tertentu dalam menangani permasalahan klasifikasi data diabetes melalui pemilihan metode ekstraksi fitur yang tepat, pembagian data yang representatif, dan evaluasi kinerja yang cermat.

Tujuan penelitian ini terfokus pada menganalisis kinerja tiga metode klasifikasi utama, yaitu K-Nearest Neighbor (KNN), Naive Bayes, dan Logistic Regression, dalam konteks pengklasifikasian data diabetes. Pertama, penelitian ini akan mengevaluasi dan menganalisis kinerja KNN dengan mempertimbangkan pembagian data 70-30 dan 80-20, serta hasilnya akan disajikan dalam bentuk confusion matrix. Selanjutnya, Naive Bayes akan diuji dan dievaluasi dalam dua skenario pembagian data yang sama, dengan hasil juga disajikan dalam bentuk confusion matrix. Selanjutnya, penelitian akan menilai efektivitas Logistic Regression dalam mengidentifikasi kasus diabetes, juga dengan pembagian data 70-30 dan 80-20, dan hasilnya akan direpresentasikan melalui confusion matrix. [10], [11], [12], [13], [14], [15], [16]

Setelah itu, dilakukan komparasi dan analisis akurasi prediksi dari ketiga metode klasifikasi pada data diabetes, dengan mempertimbangkan kedua skenario pembagian data. Tujuan akhirnya adalah mengidentifikasi metode klasifikasi yang paling optimal berdasarkan hasil evaluasi dan komparasi, dengan fokus khusus pada metode yang memberikan akurasi tertinggi. Melalui pencapaian tujuan-tujuan ini, penelitian ini diharapkan

dapat memberikan wawasan mendalam tentang efektivitas metode klasifikasi dalam menangani tantangan klasifikasi data diabetes, dengan potensi aplikasi pada pemahaman dan penanganan penyakit ini secara lebih efisien.[2], [9], [17], [18], [19], [20], [21]

II. PENELITIAN YANG TERKAIT

Adapun penelitian terdahulu yang telah dilakukan dan dapat dijadikan referensi untuk penelitian ini, diantaranya adalah: Penelitian Latifah Uswatun Khasanah, Yuki Novia Nasution, Dan Fidia Deny Tisna Amijaya (2022) yang berjudul Klasifikasi Penyakit Diabetes Melitus Menggunakan Algoritma Naive Bayes Classifier. Penelitian ini membahas tentang bagaimana cara mengklasifikasikan penyakit diabetes dengan menggunakan algoritma Naive Bayes. Adapun hasil dari penelitian ini, yaitu pada proporsi data 60:40 akurasi adalah 92,31%, pada proporsi data 70:30 akurasi adalah 89,74%, pada proporsi data 80:20 akurasi adalah 92,31%, dan pada proporsi data 90:10 akurasi adalah 84,62% [22]

Penelitian Umikulsum Indah Lestari, Anis Yusrotun Nadhiroh, dan Cahyuni Novia (2021) yang berjudul Penerapan Metode K-Nearest Neighbor Untuk Sistem Pendukung Keputusan Identifikasi Penyakit Diabetes Melitus. Penelitian ini membahas bagaimana metode KNN ini diterapkan atau diimplementasikan untuk identifikasi penyakit diabetes. Adapun hasil dari penelitian ini adalah pengujian dengan nilai $k=23$ didapatkan akurasi sebesar 96% yang dimana dinilai cukup tinggi dan telah berhasil menerapkan metode KNN untuk sistem pendukung keputusan identifikasi penyakit diabetes melitus [23]

Penelitian Happy Andrian Dwi Fasnuri, Haris Yuana, dan M. Taofik Chulkamdi (2022) yang berjudul Penerapan Algoritma K-Nearest Neighbor (K-NN) Untuk Klasifikasi Penyakit Diabetes Melitus Studi Kasus: Warga Desa Jatitengah. Penelitian ini membahas tentang bagaimana algoritma KNN diterapkan untuk mengklasifikasikan penyakit diabetes pada studi kasus warga desa Jatitengah. Adapun hasil dari penelitian ini, yaitu dengan $k=9$ didapatkan hasil akurasi yang baik sebesar 93%, presisi 100%, recall 60%, dan F1-Score sebesar 75%, yang dimana dengan akurasi sebesar 93% maka penelitian ini dinilai berhasil dalam menerapkan atau mengimplementasikan metode KNN tersebut untuk mengklasifikasikan penyakit diabetes melitus [24]

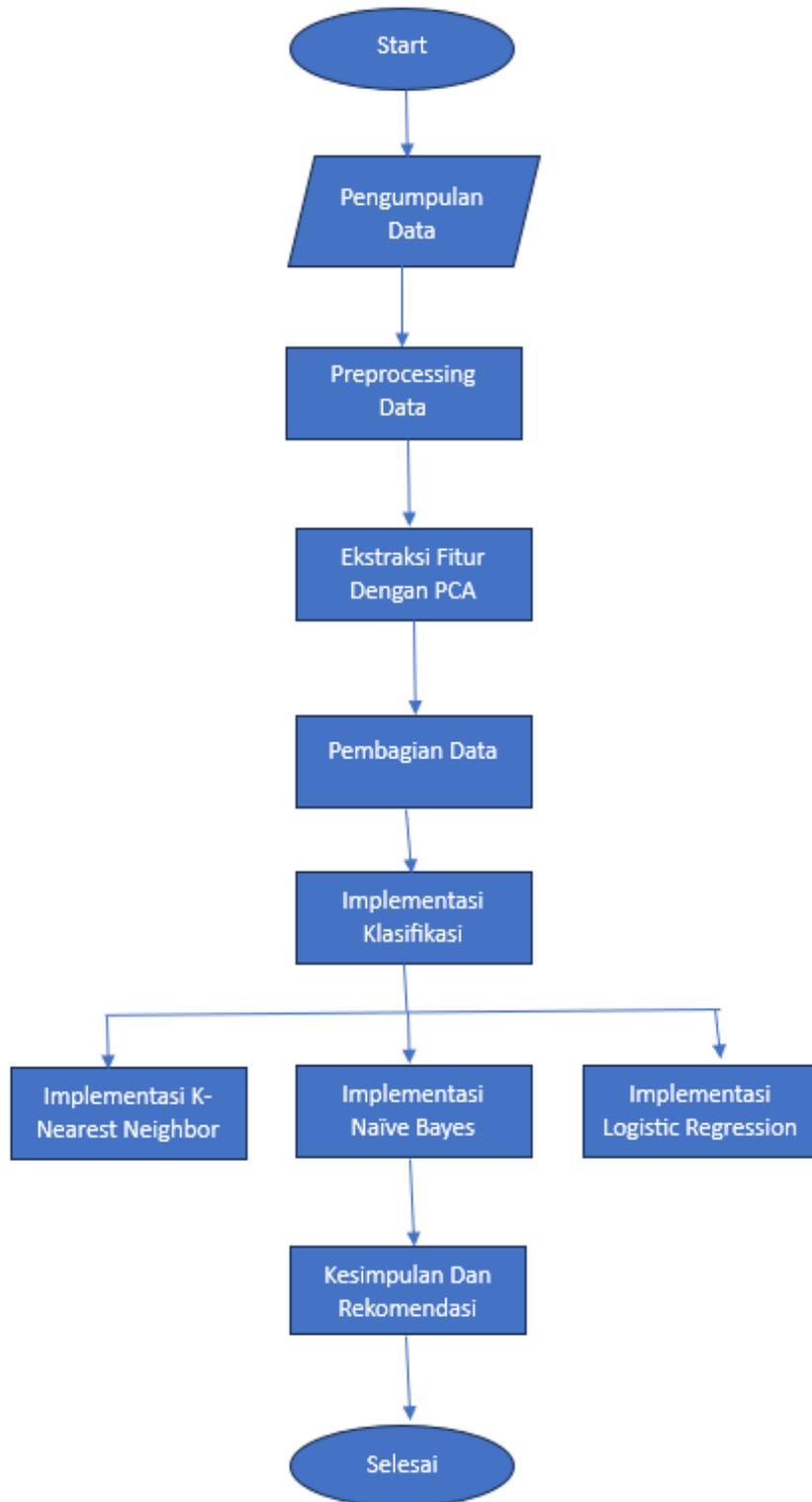
Penelitian Nabilla Yolanda Paramitha, Aang Nuryaman, Ahmad Faisol, Eri Setiawan, Dan Dina Eka [25] yang berjudul Klasifikasi Penyakit Stroke Menggunakan Metode Naive Bayes. Penelitian ini membahas tentang bagaimana cara mengklasifikasikan penyakit stroke dengan menggunakan metode naive bayes. Adapun hasil dari penelitian ini, yaitu pada proporsi data 80:20 mendapatkan akurasi sebesar 80% yang juga merupakan akurasi yang paling baik diantara proporsi data lain yang diuji cobakan [26].

Dalam konteks ini, penggunaan teknik-teknik analisis data dan machine learning dapat memberikan kontribusi besar dalam identifikasi dan klasifikasi kasus diabetes. Metode klasifikasi, seperti K-Nearest Neighbor (KNN), Naive Bayes, dan Logistic Regression, dapat digunakan untuk memproses data dan memberikan prediksi yang berguna dalam pengelolaan penyakit ini [1], [2], [3], [4], [5], [6], [7], [8], [9], [10], [11], [12], [13], [14], [15], [16], [17], [18], [19], [20], [21], [22], [23], [24], [26], [27]

Penelitian ini dilatarbelakangi oleh kebutuhan untuk memahami dan membandingkan efektivitas KNN, Naive Bayes, dan Logistic Regression dalam mengklasifikasikan data diabetes. Dengan melibatkan ekstraksi fitur menggunakan Principal Component Analysis (PCA), penelitian ini bertujuan untuk mengidentifikasi metode klasifikasi mana yang paling optimal dan dapat memberikan kontribusi signifikan dalam mendukung diagnosis dini serta pengelolaan penyakit diabetes.

III. METODE PENELITIAN

Penelitian ini akan membandingkan akurasi atau hasil dari 3 algoritma, yaitu KNN, Naive Bayes, dan juga Logistic Regression. Berikut merupakan langkah-langkah yang dilakukan dalam penelitian ini, yaitu:



Gambar 1. Langkah penelitian perhitungan

Pada gambar 1 dijelaskan kerangka penelitian yang terdiri dari pengumpulan data, preprocessing data, ekstraksi fitur dengan PCA, pembagian data, implementasi klasifikasi yang dibagi menjadi 3, yaitu KNN, Naive Bayes, dan juga Logistic Regression, dan yang terakhir merupakan kesimpulan dan rekomendasinya. Adapun penjelasan dari masing-masingnya, yaitu:

1. Pengumpulan Data

Data yang digunakan dalam penelitian ini berasal dari Kaggle, dengan fokus pada dataset diabetes. Data tersebut mencakup variabel-variabel yang relevan untuk klasifikasi penyakit diabetes.

2. Preprocessing Data

Data akan mengalami tahap preprocessing, termasuk penanganan nilai-nilai yang hilang, normalisasi, dan pemrosesan lainnya agar dapat digunakan secara optimal dalam proses klasifikasi.

3. Ekstraksi Fitur dengan PCA

Metode ekstraksi fitur yang digunakan adalah Principal Component Analysis (PCA) dengan threshold 80%. Tujuan dari PCA adalah untuk mengurangi dimensi fitur dan mempertahankan informasi yang signifikan untuk digunakan sebagai input pada metode klasifikasi.

4. Pembagian Data

Data akan dibagi menjadi dua skenario, yaitu 70-30 dan 80-20. Pembagian ini akan memberikan data latih dan data uji yang diperlukan untuk mengevaluasi kinerja metode klasifikasi.

5. Implementasi Klasifikasi K-Nearest Neighbor (KNN)

Implementasi KNN akan menggunakan bahasa pemrograman Python dengan memanfaatkan library scikit-learn.

Hasil klasifikasi akan ditampilkan dalam bentuk confusion matrix.

6. Implementasi Klasifikasi Naive Bayes

Implementasi metode Naive Bayes juga akan dilakukan dengan menggunakan Python, dengan fokus pada library scikit-learn.

Evaluasi kinerja akan didasarkan pada confusion matrix.

7. Implementasi Klasifikasi Naive Bayes

Implementasi Logistic Regression akan mengikuti pendekatan yang serupa dengan KNN dan Naive Bayes.

Hasil klasifikasi akan dievaluasi menggunakan confusion matrix.

8. Evaluasi Kinerja

Setiap metode klasifikasi akan dievaluasi berdasarkan confusion matrix yang mencakup parameter-parameter seperti akurasi, presisi, recall, dan F1-score.

Analisis akan dilakukan terhadap hasil evaluasi untuk memahami kekuatan dan kelemahan masing-masing metode.

9. Komparasi Akurasi

Akurasi prediksi dari ketiga metode pada kedua skenario pembagian data akan dikomparasikan untuk menentukan metode yang paling optimal dalam klasifikasi data diabetes.

10. Kesimpulan dan Rekomendasi

Penelitian akan diakhiri dengan menyusun kesimpulan berdasarkan hasil evaluasi dan komparasi. Rekomendasi akan diberikan terkait dengan pilihan metode klasifikasi yang paling cocok untuk aplikasi pada data diabetes. Dengan mengikuti langkah-langkah metodologi ini, diharapkan penelitian dapat memberikan wawasan yang mendalam tentang efektivitas metode klasifikasi tertentu dalam mengatasi tantangan klasifikasi data diabetes.

IV. HASIL DAN PEMBAHASAN

3.1. K-Nearest Neighbor

Pengujian yang dilakukan dengan KNN ini menggunakan bahasa Python pada Google Colaboratory. Hasil pengujian dari metode ini akan ditampilkan dalam confusion matrix, yang dimana split datanya terdiri dari 70-30 dan 80-20 yang tertera pada gambar sebagai berikut.

Laporan Klasifikasi:				
	precision	recall	f1-score	support
Negative	0.72	0.89	0.80	151
Positive	0.63	0.34	0.44	80
accuracy			0.70	231
macro avg	0.67	0.62	0.62	231
weighted avg	0.69	0.70	0.67	231

Gambar 2. Gambar Confusion Matrix KNN (70-30)

Laporan Klasifikasi:				
	precision	recall	f1-score	support
Negative	0.72	0.89	0.79	99
Positive	0.65	0.36	0.47	55
accuracy			0.70	154
macro avg	0.68	0.63	0.63	154
weighted avg	0.69	0.70	0.68	154

Gambar 3. Gambar Confusion Matrix KNN (80-20)

3.2. Naive Bayes

Pengujian yang dilakukan dengan metode Naive Bayes menggunakan bahasa python pada Google Colaboratory. Hasil pengujiannya ditampilkan dalam confusion matrix dengan split data 70-30 dan 80-20 yang tertera pada gambar sebagai berikut.

Laporan Klasifikasi:				
	precision	recall	f1-score	support
Negative	0.80	0.83	0.81	151
Positive	0.65	0.61	0.63	80
accuracy			0.75	231
macro avg	0.73	0.72	0.72	231
weighted avg	0.75	0.75	0.75	231

Gambar 4. Gambar Confusion Matrix Naive Bayes (70-30)

Laporan Klasifikasi:				
	precision	recall	f1-score	support
Negative	0.82	0.86	0.84	99
Positive	0.72	0.65	0.69	55
accuracy			0.79	154
macro avg	0.77	0.76	0.76	154
weighted avg	0.78	0.79	0.78	154

Gambar 5. Gambar Confusion Matrix Naive Bayes (80-20)

3.3. Logistic Regression

Pengujian yang dilakukan dengan metode Logistic Regression menggunakan bahasa python pada Google Colaboratory. Hasil pengujiannya ditampilkan dalam confusion matrix dengan split data 70-30 dan 80-20 yang tertera pada gambar sebagai berikut.

Laporan Klasifikasi:				
	precision	recall	f1-score	support
Negative	0.76	0.82	0.79	151
Positive	0.60	0.50	0.54	80
accuracy			0.71	231
macro avg	0.68	0.66	0.67	231
weighted avg	0.70	0.71	0.70	231

Gambar 6. Gambar Confusion Matrix Regresi Logistik (70-30)

Laporan Klasifikasi:				
	precision	recall	f1-score	support
Negative	0.76	0.84	0.80	99
Positive	0.64	0.53	0.58	55
accuracy			0.73	154
macro avg	0.70	0.68	0.69	154
weighted avg	0.72	0.73	0.72	154

Gambar 7. Gambar Confusion Matrix Regresi Logistik(80-20)

3.4. Pengujian Sistem

Pengujian dan komparasi dari metode-metode ini akan dilakukan dengan melihat akurasi prediksi pada data diabetes yang digunakan, untuk melihat algoritma atau metode yang paling cocok untuk digunakan dalam data tersebut, yang dimana komparasinya akan ditunjukkan pada tabel berikut ini.

Tabel 1. Tabel Komparasi Tiga Metode

Split Data	Akurasi		
	KNN	Naive Bayes	Logistic Regression
70-30	70%	75%	71%
80-20	70%	79%	73%

Dari hasil tabel diatas, dapat dilihat bahwa algoritma Naive Bayes memiliki akurasi yang paling tinggi, yaitu sekitar 79% dengan pembagian data 80-20. Oleh karena itu, algoritma Naive Bayes merupakan algoritma yang paling baik untuk digunakan dalam klasifikasi data ini.

V. KESIMPULAN

Berdasarkan hasil analisis dan pembahasan serta pengujian menggunakan tiga metode klasifikasi, yaitu K-Nearest Neighbor (KNN), Naive Bayes, dan Logistic Regression, terdapat beberapa temuan signifikan. Pada split data 70-30, KNN memberikan akurasi sebesar 70%, sedangkan pada split 80-20, akurasinya meningkat menjadi 71%. Meskipun KNN memberikan hasil yang cukup baik, tetapi terdapat peningkatan yang terbatas pada akurasi dengan peningkatan ukuran data. Metode Naive Bayes menunjukkan kinerja yang cukup baik, dengan akurasi 75% pada pembagian data 70-30 dan meningkat menjadi 79% pada pembagian data 80-20. Naive Bayes memiliki performa yang konsisten dan memberikan akurasi tertinggi dibandingkan dengan KNN dan Logistic Regression pada kedua pembagian data. Logistic Regression menunjukkan akurasi sebesar 71% pada split data 70-30 dan 73% pada split data 80-20. Meskipun hasilnya bersaing, Logistic Regression tidak sebaik Naive Bayes pada split data 80-20. Dari hasil tabel komparasi akurasi, dapat disimpulkan bahwa Naive Bayes memiliki akurasi paling tinggi (79%) pada pembagian data 80-20. Dengan demikian, algoritma Naive Bayes dapat dianggap sebagai pilihan terbaik untuk klasifikasi data diabetes dalam konteks penelitian ini. Secara keseluruhan, Naive Bayes merupakan metode klasifikasi yang paling optimal untuk dataset diabetes yang digunakan dalam penelitian ini, dengan memberikan akurasi prediksi yang tinggi pada kedua pembagian data.

DAFTAR PUSTAKA

- [1] K. M. Amien, I. L. F., Astuti, W., & Lhaksamana, "Perbandingan Metode Naive Bayes dan KNN (K-Nearest Neighbor) dalam Klasifikasi Penyakit Diabetes. E-Proceeding of Engineering, 10(2), 1911–1920." 2023.
- [2] K. Susana, H., Suarna, N., Fathurrohman, F., & Kaslani, "Penerapan Model Klasifikasi Metode Naive Bayes Terhadap Penggunaan Akses Internet. Jurnal Riset Sistem Informasi Dan Teknologi Informasi (JURSISTEKNI), 4(1), 1–8. <https://doi.org/10.52005/jursistekni.v4i1.96>," 2022.
- [3] S. Putra, P., Pardede, A. M. H., & Syahputra, "ANALISIS METODE K-NEAREST NEIGHBOUR (KNN) DALAM KLASIFIKASI DATA IRIS BUNGA. 6(1), 297–305." 2022.
- [4] A. M. (2020). Argina, "Penerapan Metode Klasifikasi K-Nearest Neighbor pada Dataset Penderita Penyakit Diabetes. Indonesian Journal of Data and Science, 1(2), 29–33. <https://doi.org/10.33096/ijodas.v1i2.11>," 2020.
- [5] B. Ikhromr, F. N., Sugiyarto, I., Faddillah, U., & Sudarsono, "Implementasi Data Mining Untuk Memprediksi Penyakit Diabetes Menggunakan Algoritma Naives Bayes dan K-Nearest Neighbor. INTECOMS: Journal of Information Technology and Computer Science, 6(1), 416–428. <https://doi.org/10.31539/intecom.v6i1.5916>," 2023.
- [6] I. R. Padiku, "Penerapan Metode Naive Bayes Classifier (Nbc) Untuk Klasifikasi Kondisi Internal Program Studi. Jurnal Teknik, 19(1), 65–74. <https://doi.org/10.37031/jt.v19i1.118>," 2021.
- [7] M. T. (2019). Maliha, D. Z., Santoso, E., & Furqon, ("). Penerapan Metode Neighbor Weighted K-Nearest Neighbor Dalam Klasifikasi Diabetes Mellitus. Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer, 3(3), 2910–2915. <https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/4817>," 2019.

- [8] G. I. (2021). Reviantika, F., Azhar, Y., & Marthasari, "Analisis Klasifikasi SMS Spam Menggunakan Logistic Regression. *Jurnal Sistem Cerdas*, 4(3), 155–160. <https://doi.org/10.37396/jsc.v4i3.166>," 2021.
- [9] N. Susetyoko, R., Yuwono, W., Purwantini, E., & Ramadijanti, "Perbandingan Metode Random Forest, Regresi Logistik, Naïve Bayes, dan Multilayer Perceptron Pada Klasifikasi Uang Kuliah Tunggal (UKT). *Jurnal Infomedia*, 7(1), 8. <https://doi.org/10.30811/jim.v7i1.2916>," 2022.
- [10] D. Setiawan, D. Arisandi, and L. Trisnawati, "Aplikasi Prediksi Penyakit Sirosis Hati Menggunakan Algoritma Genetika," *J. SANTI (Sistem Inf. dan Teknol. Informasi)*, vol. 2, no. 1, pp. 31–40, 2022.
- [11] I. Puspita Sari and D. Setiawan, "Aplikasi D'Laundry Berbasis Android Menggunakan Model Design Thinking Android-Based D'Laundry Application Using Model Design Thinking," *Juli*, vol. 2, no. 3, pp. 136–142, 2022.
- [12] D. Setiawan, R. N. Putri, and R. Suryanita, "Implementasi Algoritma Genetika Untuk Prediksi Penyakit Autoimun," *Rabit J. Teknol. dan Sist. Inf. Univrab*, vol. 4, no. 1, pp. 8–16, 2019, doi: 10.36341/rabit.v4i1.595.
- [13] I. P. Sari, D. Setiawan, and D. W. Marwan, "Sosialisasi Pelaksanaan Posyandu Melalui Transformasi Digital Kesehatan Menggunakan Aplikasi Mobile Posyandu Q," *J. Abdidas*, vol. 3, no. 5, pp. 870–876, 2022, doi: 10.31004/abdidas.v3i5.691.
- [14] M. P. Efendy and D. Setiawan, "PERANCANGAN APLIKASI MAKANAN EMPAT SEHAT LIMA SEMPURNA," vol. 5, no. 1, pp. 13–19, 2021.
- [15] P. Addie, D. Pengembangan, K. Pada, L. Elvitaria, D. Setiawan, and L. Susanti, "PENDAHULUAN Perkembangan ilmu pengetahuan dan teknologi yang sangat pesat saat ini menunjang dan mempermudah semua kegiatan di berbagai bidang dalam kehidupan manusia , salah satunya di Pendidikan . Dengan pemanfaatan teknologi media pembelajaran akan leb," vol. 8, no. 2, pp. 231–246, 2023.
- [16] I. Engineering, D. Setiawan, M. C. Pratama, D. Arisandi, T. Informatika, and U. Abdurrah, "IMPLEMENTASI SISTEM KEAMANAN JARINGAN MENGGUNAKAN RULE-," vol. 7, no. 2, pp. 381–389, 2023.
- [17] D. Setiawan, R. N. Putri, and R. Suryanita, "Perbandingan Algoritma Genetika dan Backpropagation pada Aplikasi Prediksi Penyakit Autoimun," *Khazanah Inform. J. Ilmu Komput. dan Inform.*, vol. 5, no. 1, pp. 21–27, 2019, doi: 10.23917/khif.v5i1.7173.
- [18] Liza Trisnawati, D. Setiawan, and S. Ryanto, "ANTIDERM Dermatitis Consultation Application User Interface Design Model Using Design Sprint Method," *J. Int. Multidiscip.*, vol. 1, no. 1, pp. 1–17, 2023, doi: 10.58794/jim.v1i1.193.
- [19] N. Maryam, D. Setiawan, and R. Noratama Putri, "Aplikasi Deteksi Dini Penyakit Tidak Menular Pada Lansia Dengan menggunakan metode Dynamic System Development Method (DSDM)," *JEKIN - J. Tek. Inform.*, vol. 3, no. 1, pp. 1–8, 2022, doi: 10.58794/jekin.v3i1.219.
- [20] Rudiansyah, D. Setiawan, and Shucy Ramawati Fadila, "Identifikasi Faktor Risiko Stunting pada Anak-anak dengan Metode K means Clustering menggunakan Dataset Kaggle," *JEKIN - J. Tek. Inform.*, vol. 3, no. 1, pp. 10–16, 2023, doi: 10.58794/jekin.v3i1.443.
- [21] F. Yunita, "Sistem Klasifikasi Penyakit Diabetes Mellitus Menggunakan Metode," 2016.
- [22] F. D. T. Khasanah, L. U., Nasution, Y. N., & Amijaya, "Klasifikasi Penyakit Diabetes Melitus Menggunakan Naive Bayes Classifier. *BASIS*, 1(1), 41–50.," 2022.
- [23] C. Lestari, U. I., Nadhiroh, A. Y., & Novia, "Penerapan Metode K-Nearest Neighbor Untuk Sistem Pendukung Keputusan Identifikasi Penyakit Diabetes Melitus. *JATISI (Jurnal Teknik Informatika Dan Sistem Informasi)*, 8(4), 2071–2082. <https://doi.org/10.35957/jatisi.v8i4.1235>," 2021.
- [24] M. T. Fasnuari, H. A. D., Yuana, H., & Chulkamdi, "Penerapan Algoritma K-Nearest Neighbor Untuk Klasifikasi Penyakit Diabetes Melitus. *Antivirus: Jurnal Ilmiah Teknik Informatika*, 16(2), 133–142. <https://doi.org/10.35457/antivirus.v16i2.2445>," 2022.
- [25] D. Widyastutik, Ernawati, E. N. Pratiwi, and R. Wulandari, "UPAYA PENINGKATAN PERILAKU IBU POSTPARTUM MELALUI EDUKASI FAMILY CENTERED MATERNITY CARE (FCMC) TENTANG PERAWATAN MASA POSTPARTUM DI WILAYAH KERJA PUSKESMAS GAMBIRSARI SURAKARTA Program Studi Kebidanan Program Sarjana dan Pendidikan Profesi Bidan Progra," *J. Salam Sehat Masy.*, vol. 2, no. 2, pp. 43–50, 2021.
- [26] D. E. Paramitha, Y. N., Nuryaman, A., Faisol, A., Setiawan, E., & Nurvazly, "Klasifikasi Penyakit Stroke Menggunakan Metode Naïve Bayes. *Jurnal Siger Matematika*, 04(01), 11–16. <https://www.kaggle.com/datasets/zzettrkalkpakbal/full-filled->," 2023.
- [27] H. Karo, I. M. K., & Hendriyana, "Klasifikasi Penderita Diabetes menggunakan Algoritma Machine Learning dan Z-Score. *Jurnal Teknologi Terpadu*, 8(2), 94–99," 2022.