



Comparison of Naive Bayes and K-Nearest Neighbors Algorithms for Metabolic Syndrome Classification

Perbandingan Algoritma Naive Bayes dan K-Nearest Neighbors untuk Klasifikasi Metabolik Sindrom

Fitriana Sholekhah¹, Adinda Dwi Putri², Rahmaddeni^{3*}, Lusiana Efrizoni⁴

^{1,2,3,4}Jurusan Teknik Informatika, STMIK Amik Riau, Indonesia

E-Mail: ¹2110031802099@sar.ac.id, ²2110031802112@sar.ac.id,
³rahmaddeni@sar.ac.id, ⁴lusiana@stmik-amik-riau.ac.id

Received Nov 28th 2023; Revised Jan 17th 2024; Accepted Feb 20th 2024
Corresponding Author: Rahmaddeni

Abstract

A medical condition known as metabolic syndrome potentially increases the chances of coronary heart disease, stroke, heart attack and type 2 diabetes. Metabolic syndrome can also cause high blood sugar, low cholesterol levels, concurrent obesity and excess fat in the waist area. If a combination of these three conditions occurs, it can be said that this disease is metabolic syndrome. In addition, metabolic syndrome is also associated with insulin resistance, meaning that the body's cells do not respond well to the effects of insulin which causes high blood sugar levels because sugar is not absorbed into the cells properly. Metabolic syndrome is growing as obesity increases in Asia, with prevalence estimates continuing to rise. This has the potential to increase cases of cardiovascular disease and mortality risk. Therefore, it is necessary to develop a model to diagnose metabolic syndrome. This study aims to compare the performance of the main classification algorithms, namely Naive Bayes (NB) and K-Nearest Neighbors (KNN) in detecting metabolic syndrome. The results of this study show that the use of the Naive Bayes algorithm produces an accuracy of 79%, while the highest accuracy of the K-Nearest Neighbors (KNN) algorithm is 82%. In conclusion, the results of this study show that the K-NN algorithm with 50:50 data division is more effective in predicting and classifying metabolic syndrome.

Keyword: Classification, Comparison, K-Nearest Neighbors, Metabolic Syndrome, Naive Bayes

Abstrak

Kondisi medis yang dikenal sebagai sindrom metabolik berpotensi meningkatkan kemungkinan penyakit jantung koroner, stroke, serangan jantung dan diabetes tipe 2. Sindrom metabolik juga dapat menyebabkan gula darah tinggi, kadar kolesterol rendah, obesitas secara bersamaan dan kelebihan lemak di daerah pinggang. Jika kombinasi dari ketiga kondisi ini terjadi maka dapat dikatakan penyakit ini sebagai sindrom metabolik. Selain itu, sindrom metabolik juga dikaitkan dengan resistensi insulin, artinya dimana sel-sel tubuh tidak merespon baik terhadap efek insulin yang menyebabkan kadar gula darah tinggi karena gula tidak terserap ke dalam sel dengan baik. Sindrom metabolik tumbuh seiring meningkatnya obesitas di Asia, dengan perkiraan prevalensi yang terus naik. Ini berpotensi meningkatkan kasus penyakit kardiovaskular dan risiko kematian. Oleh karena itu, perlu dikembangkan model untuk mendiagnosis sindrom metabolik. Penelitian ini bertujuan untuk membandingkan kinerja algoritma klasifikasi utama, yaitu Naive Bayes (NB) dan K-Nearest Neighbors (KNN) dalam mendeteksi sindrom metabolik. Hasil dari penelitian ini menunjukkan bahwa penggunaan algoritma Naive Bayes menghasilkan akurasi sebesar 79%, sedangkan akurasi tertinggi dari algoritma K-Nearest Neighbors (KNN) adalah 82%. Kesimpulannya, dari hasil penelitian ini menunjukkan bahwa algoritma K-NN dengan pembagian data 50:50 lebih efektif dalam memprediksi dan mengklasifikasikan sindrom metabolik.

Kata Kunci: Klasifikasi, K-Nearest Neighbors, Naive Bayes, Perbandingan, Sindrom Metabolik

1. PENDAHULUAN

Metabolic syndrome atau sindrom metabolik merupakan faktor risiko yang meningkatkan risiko penyakit kardiovaskular, diabetes, stroke, dan masalah kesehatan lainnya [1]. Jika mengalami beberapa gejala dari komponen bisa dikatakan menderita sindrom metabolik seperti *hipertensi*, *colestrol*, *obesitas*. Obesitas sentral, kadar trigliserida (TG) yang tinggi, hipertensi (HI), kadar lipoprotein densitas tinggi (HDL) yang rendah, dan tingginya kadar glukosa darah puasa (GDP) adalah semua gejala dari sindrom metabolik (SM) [2].

Sindrom metabolik bukan penyakit tetapi itu adalah kumpulan faktor risiko metabolik yang terkait langsung dengan penyakit tidak menular, terutama arterosklerosis kardiovaskuler [3]. Seiring dengan perubahan gaya hidup manusia, peningkatan kejadian sindrom metabolik adalah salah satu masalah kesehatan baru. Sindrom metabolik terjadi ketika tubuh tidak seimbang dalam beberapa bagian penting dari kesehatan. Misalnya, tubuh mungkin kesulitan mengelola gula darah, tekanan darah naik, dan kadar kolesterol yang tidak sehat. Jika faktor-faktor ini digabungkan, ini dapat meningkatkan risiko penyakit serius seperti jantung dan diabetes.

Kondisi ini menyebabkan masalah kesehatan global yang semakin mendesak dan kompleks. Seseorang dengan sindrom metabolik memiliki kemungkinan lebih besar untuk mengalami berbagai masalah kesehatan yang serius. Sindrom metabolik berkembang seiring dengan prevalensi obesitas yang terjadi pada populasi Asia [4] Diperkirakan prevalensi sindrom metabolik terus meningkat, yang berdampak pada peningkatan jumlah penderita penyakit kardiovaskuler dan kelangsungan hidup [5]. Oleh karena itu, diperlukan adanya model yang dapat digunakan untuk menentukan diagnosis sindrom metabolik.

Berdasarkan hasil penelitian klasifikasi algoritma Naïve Bayes dan K-Nearest Neighbors pada penelitian terdahulu didapatkan bahwa metode K-Nearest Neighbors dengan $K=5$ memiliki nilai akurasi tertinggi sebesar 90%, sedangkan metode Naïve Bayes memiliki nilai akurasi sebesar 80% [6]. Penelitian selanjutnya didapatkan tingkat akurasi pada metode K-Nearest Neighbors sebesar 67% sedangkan metode *Naïve Bayes Classifier* tingkat akurasi yang didapatkan sebesar 58% [7]. Sementara itu dari penelitian klasifikasi kanker prostat dengan hasil akurasi dengan menggunakan algoritma Naïve Bayes adalah 80% dan K-NN sebesar 90% [8]. Kemudian penelitian lainnya mendapatkan hasil yang menunjukkan metode Naïve Bayes memiliki kinerja yang lebih baik dengan tingkat *accuracy* 70%, sedangkan metode K-Nearest Neighbors memiliki tingkat *accuracy* yang cukup rendah yaitu 40% [9].

Perkembangan teknologi informasi dalam memproses kumpulan data yang sangat besar selalu dikaitkan dengan kemajuan dalam data mining. Oleh karena itu, ada banyak algoritma yang digunakan untuk memproses data besar, salah satunya adalah algoritma klasifikasi [10]. Klasifikasi adalah metode analisis data yang mengekstrak model dan memasukkannya ke dalam kelas tertentu [11]. Salah satu algoritma yang banyak digunakan untuk melakukan klasifikasi adalah algoritma Naïve Bayes dan K-Nearest Neighbors. Karena, KNN merupakan algoritma yang sederhana untuk diimplementasikan tetapi menghasilkan akurasi yang baik [12]. Semakin banyak data maka waktu proses membandingkannya akan semakin lama, sehingga akan sangat sulit mendapatkan proses *real-time* untuk kasus klasifikasi [13]. *Naïve Bayes Classifier* adalah algoritma pembelajaran Bayesian yang paling cepat dan sederhana [14]. Sedangkan, metode K-Nearest Neighbors (K-NN) merupakan suatu algoritma dari metode *supervised* dimana hasil dari *query instance* dapat diklasifikasikan berdasarkan mayoritas dari label *class* pada K-NN [15].

Hasil dari penelitian ini adalah K-Nearest Neighbors lebih tinggi nilai keakuratannya dari pada naïve bayes [16]. Banyak peneliti telah membandingkan kedua metode tersebut. Oleh karena itu, penelitian ini membandingkan dua algoritma sekaligus, K-Nearest Neighbors dan *Naïve Bayes Classifier*. Tujuan dari penelitian ini adalah untuk menentukan algoritma model mana yang memiliki nilai akurasi, *pressi*, dan *recall* yang lebih tinggi. Penelitian ini juga melakukan mengevaluasi tingkat akurasi dari tiga *splitting data*. Dengan berkonsentrasi pada algoritma Naïve Bayes dan K-Nearest Neighbors. Diharapkan penelitian selanjutnya menggunakan algoritma lainnya dan dapat menggunakan penelitian ini sebagai landasan untuk penelitian sindrom metabolik.

2. METODOLOGI PENELITIAN

Penelitian ini dilakukan untuk membangun model klasifikasi gangguan metabolik dengan membandingkan hasil dari algoritma KNN dan Naïve Bayes. Alur penelitian dimulai dari: (1) Pengumpulan Data; (2) Preprocessing; (3) Splitting Data; (4) Pemodelan. Alur penelitian digambarkan pada gambar 1.

2.1. Pengumpulan Data

Data utama penelitian ini berasal dari dataset publik *Metabolic Syndrome* yang tersedia di platform *Kaggle* dan terdiri dari 2041 entri data, masing-masing dengan 15 fitur yang menjelaskan berbagai aspek yang terkait dengan sindrom metabolik. Dataset ini terakhir diperbarui pada tahun 2023, menunjukkan bahwa itu lengkap dan akurat. Secara keseluruhan, kumpulan data ini memberikan landasan yang kokoh untuk melakukan analisis menyeluruh tentang sindrom metabolik yang terkait dengan penelitian ini.

2.2. Preprocessing

Preprocessing data adalah proses menyiapkan dan membersihkan data sebelum lakukan analisis [17]. Pada tahap ini data yang akan diolah bertujuan untuk menghindarkan dari data yang mengganggu (*noise*) atau data yang tidak konsisten [18]. Proses *preprocessing* mencakup tiga tahapan yaitu *cleaning data*, *tokenizing* dan *transform data* [19].

1. *Data Cleaning*

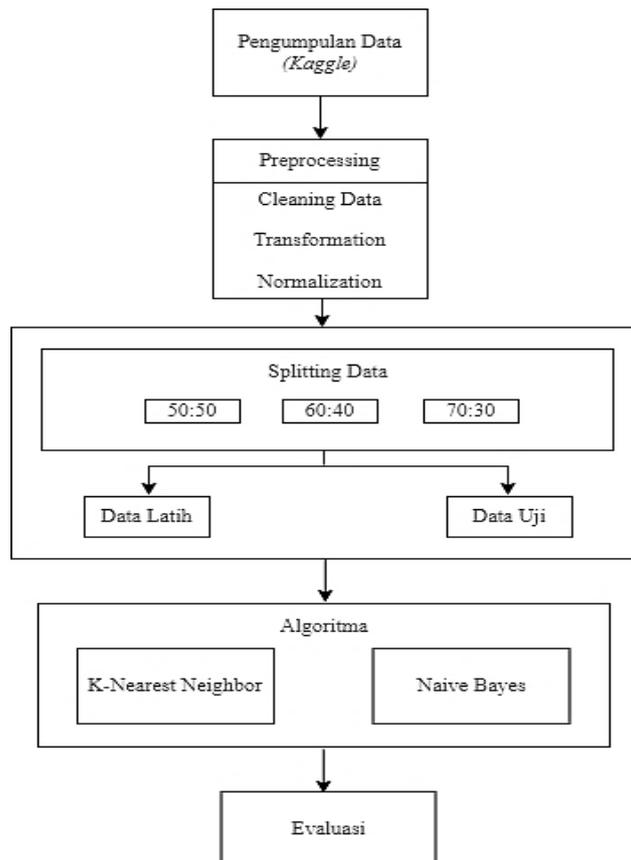
Data cleaning adalah menghapus atau mengisi nilai kosong untuk seluruh kumpulan data dengan nilai rata-rata dari tiap kolom pada nilai yang kosong [20].

2. *Data Transformation*

Data Transformation adalah mengubah data untuk mendapatkan data yang lebih baik [21] dilakukan untuk mengubah data yang akan dianalisis sehingga dapat menghasilkan nilai yang lebih relevan[22].

3. *Normalization*

Normalisasi Data merupakan proses transformasi skala atribut numerik ke dalam bentuk lebih sederhana seperti 0 hingga 1 [23].



Gambar 1. Alur Penelitian

2.3. *Splitting Data*

Splitting data berfungsi untuk membagi dataset menjadi dua bagian, yaitu data pelatihan dan data uji dengan proporsi tertentu [24]. Tahap berikutnya adalah pembagian data. Proses ini membagi data menjadi dua bagian, yaitu data latih dan data uji. Tujuan dari pembagian data adalah untuk memastikan bahwa model yang dibangun dapat diterapkan dengan baik pada data yang belum pernah dilihat sebelumnya. Studi ini dilakukan dengan membagi data tiga kali, masing-masing 50:50, 60:40, dan 70:30. Algoritma K-Nearest Neighbors dan Naïve Bayes akan digunakan untuk menguji data yang telah dibagi.

1. 50:50, 50 dari data digunakan untuk pelatihan model, dan 50 digunakan untuk menguji model.
2. 60:40, 60 dari data digunakan untuk pelatihan model, dan 40 digunakan untuk menguji model.
3. 70:30, 70 dari data digunakan untuk pelatihan model, dan 30 digunakan untuk menguji model.

2.4. *Algoritma*

Dua algoritma klasifikasi yang digunakan dalam penelitian ini adalah *K-Nearest Neighbors* (KNN) dan *Naive Bayes* (NB). KNN menggunakan data latih terdekat untuk mengklasifikasikan data uji, sementara NB menggunakan prinsip probabilitas untuk menentukan kelas instance.

2.4.1. K-Nearest Neighbors

Algoritma *K-Nearest Neighbors* adalah metode klasifikasi objek berdasarkan data pelatihan yang jaraknya paling dekat[25]. Rinsip sederhana dari *K-Nearest Neighbors* adalah bekerja berdasarkan jarak terdekat dari sampel uji ke sampel latih[26]. Untuk memprediksi kelas baru, metode k-NN menggunakan prinsip ketetanggaan[27]. Tahapan proses KNN meliputi:

1. Menentukan parameter K
2. Menghitung jarak antara data training dan data testing
Perhitungan jarak *Euclidean* adalah perhitungan jarak yang paling sering digunakan dalam algoritma KNN. Sebagai contoh, rumusnya adalah sebagai berikut:

$$euc = \sqrt{(\sum_{i=1}^n (p_i - q_i)^2)} \quad (1)$$

dimana:

- p_i = sample data/data training
- q_i = data uji/ data testing
- i = variable data
- n = dimensi data

3. Mengurutkan jarak yang terbentuk
4. Tentukan jarak terdekat sampai urutan K
5. Menyusun kelas yang sesuai
6. Cari jumlah kelas dari tetangga terdekat Anda dan tetapkan kelas tersebut sebagai kelas data untuk dievaluasi[28].

2.4.2. Naïve Bayes

Klasifikasi Bayes adalah pengelompokan statistik yang memperkirakan kemungkinan keanggotaan kelas misalnya, kemungkinan bahwa sebuah tuple masuk ke dalam kelas tertentu[29]. *Teorema Bayes* adalah algoritma Naïve Bayes yang menggunakan pengalaman masa lalu untuk memprediksi apa yang akan terjadi di masa depan[30]. Tahapan proses Naïve Bayes meliputi:

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)} \quad (2)$$

Keterangan dari persamaan (2) diatas yaitu:

- $P(A|B)$ = probabilitas A dan B terjadi bersama-sama
- $P(B|A)$ = probabilitas B dan A terjadi bersama-sama
- $P(A)$ = probabilitas kejadian
- $A \cdot P(B)$ = probabilitas kejadian B[31].

2.5. Evaluasi

Pada tahap ini, evaluasi metode prediksi dilakukan untuk mengukur kinerja algoritma *K-Nearest Neighbors* (KNN) dan *Naïve Bayes* (NB). Evaluasi ini akan membandingkan nilai akurasi yang dihasilkan oleh kedua algoritma dengan menggunakan dataset yang sama dan menerapkan tahapan *splitting data* yang identik, yaitu 50:50, 60:40, dan 70:30.

3. HASIL DAN PEMBAHASAN

3.1. Pengumpulan Data

Tahap pertama yaitu pengumpulan data dimana data diambil dari situs kaggle. Hasil analisis menunjukkan bahwa informasi terdiri dari 505 data, Tipe data untuk variabel dependen, sindrom metabolik, memiliki nilai 0 yang menunjukkan bahwa sindrom metabolik tidak ada, dan nilai 1 menunjukkan bahwa sindrom metabolik ada. Dari semua data dalam dataset tersebut, perbandingan antara ada sindrom metabolik dan tidak ada sindrom metabolik pada Tabel 1.

Tabel 1. Data Metabolik Sindrom

No	seqn	Age	Sex	Marital	Income	...	MetabolicSyndrome
0	62161	22	Male	Single	8200	...	0
1	62164	44	Female	Married	4500	...	0
2	62169	21	Male	Single	800	...	0
...
504	64998	54	Male	Single	5400	...	1

Tabel 1 menunjukkan dataset diabetes yang mencakup berbagai fitur, seperti *seqn*, *Age*, *Sex*, *Marital*, *Income*, *Race*, *WaistCirc*, *BMI*, *Albuminuria*, *UrAlbCr*, *UricAcid*, *BloodGlucose*, *HDL*, *Triglycerides*, dan *MetabolicSyndrome*. Di bawah ini adalah penjelasan singkat tentang masing-masing fitur:

1. *seqn* adalah identifikasi unik untuk setiap subjek dalam dataset.
2. *Age*: usia subjek dalam tahun.
3. *Sex*: jenis kelamin subjek (pria atau wanita).
4. *Marital*: status pernikahan.
5. *Income*: tingkat pendapatan.
6. *Race*: kategori ras.
7. *WaistCirc*: lingkar pinggang.
8. *BMI*: Indeks Massa Tubuh subjek (BMI).
9. *Albuminuria*: tingkat albuminuria dalam urin subjek.
10. *UrAlbCr*: rasio albumin terhadap kreatin dalam tubuh subjek.
11. *UricAcid* adalah tingkat asam urat yang ditemukan dalam darah subjek.
12. *BloodGlucose*: Nilai glukosa darah dari subjek.
13. *HDL* adalah subjek kolesterol lipoprotein tinggi densitas.
14. *Triglycerides*: Jumlah trigliserida yang ditemukan dalam darah subjek.
15. *MetabolicSyndrome*: Kondisi subjek sindrom metabolik.

3.2. Preprocessing

Proses ini melibatkan tahap preprocessing data, di mana outliers pada kumpulan data ditangani melalui normalisasi atau standarisasi nilai. Fokus penanganan outliers adalah fitur *seqn*, *sex*, *marital*, *income*, *race*, *waistcirc*, *albuminuria*, *urabc*, *uricacid*, *hdl*. Hasil *preprocessing outliers* berhasil pada Tabel 2.

Tabel 2. Hasil *Preprocessing Outliers*

No	Age	BMI	Blood Glucose	Triglycerides	Metabolic Syndrome
0	22	23.3	92	84	0
1	44	23.2	82	56	0
2	21	20.1	107	78	0
...
500	22	26.6	84	57	0

3.3. Splitting Data

Penelitian ini membagi data dalam tiga tahap dengan rasio 50:50, 60:40, dan 70:30. Pada tahap 60:40, 60% dari total data digunakan sebagai data latih, sementara 40% digunakan sebagai data uji. Pada tahap berikutnya, 70% dari total data digunakan sebagai data latih, dan 30% digunakan sebagai data uji. Metode ini memungkinkan eksplorasi kinerja model dalam berbagai skenario pembagian data latih dan uji. Data yang telah diikuti ketiga tahap ini kemudian diuji menggunakan algoritma *K-Nearest Neighbors* dan *Naïve Bayes* untuk mengevaluasi kinerja model pada setiap skenario pembagian data.

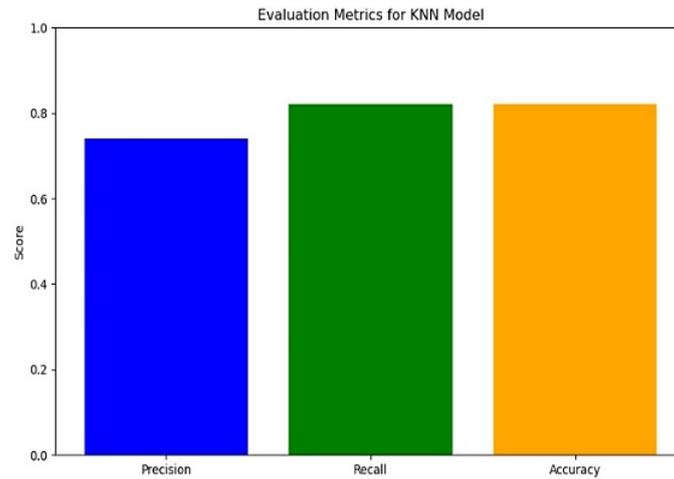
3.4. Algoritma

3.4.1. K-Nearest Neighbors

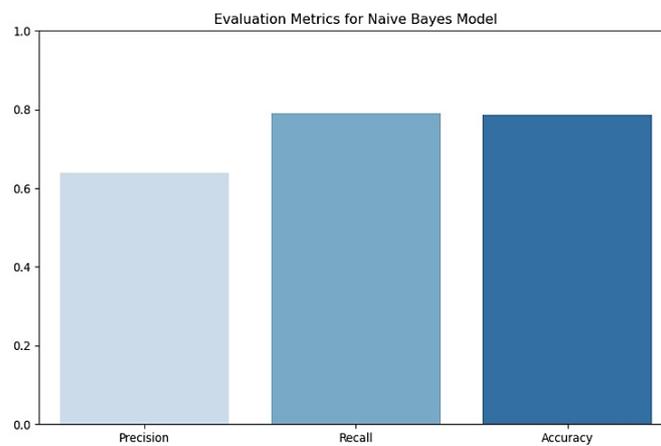
Gambar 2 menunjukkan visualisasi hasil akurasi K-NN untuk tiga skema pembagian data, yaitu 50:50, 60:40, dan 70:30. Hasil eksperimen menunjukkan bahwa akurasi K-NN mencapai 82% pada skema 50:50, 80% pada skema 60:40, dan 79% pada skema 70:30. 60:40 menunjukkan tren bahwa kinerja K-NN meningkat seiring dengan peningkatan proporsi data pada set pelatihan. Skema 60:40 memberikan hasil akurasi terbaik.

3.4.2. Naïve Bayes

Gambar 3 menunjukkan visualisasi hasil akurasi *Naïve Bayes* untuk tiga skema pembagian data, yaitu 50:50, 60:40, dan 70:30. Hasil eksperimen menunjukkan bahwa akurasi K-NN mencapai 79% pada skema 50:50, 77% pada skema 60:40, dan 77% pada skema 70:30. 60:40 dan 70:30 menunjukkan tren bahwa kinerja *Naïve Bayes* meningkat seiring dengan peningkatan proporsi data pada set pelatihan. Skema 60:40 dan 70:30 memberikan hasil akurasi terbaik.



Gambar 2. Visualisasi KNN



Gambar 3. Visualisasi *Naïve Bayes*

3.5 Evaluasi

Dalam penelitian ini, metode K-NN dan *Naïve Bayes* digunakan, dengan tiga *splitting* data (50:50, 60:40 dan 70:30). Hasil akurasi K-NN dibandingkan dengan K = 5 dan *Naïve Bayes*, yang dapat dilihat pada Tabel 3.

Tabel 3. Hasil Perbandingan KNN dan NB

Splitting data	K-Nearest Neighbors (K=5)			Naïve Bayes		
	Akurasi	Presisi	Recall	Akurasi	Presisi	Recall
50:50:00	82%	85%	87%	79%	79%	92%
60:40:00	80%	85%	85%	77%	77%	92%
70:30:00	79%	83%	85%	77%	77%	91%

Berdasarkan Tabel 2, hasil akurasi tertinggi diperoleh pada *splitting* data 50:50 baik untuk K-NN maupun *Naïve Bayes*. Model K-NN dengan akurasi 82% dan *Naïve Bayes* 79%. Sementara 60:40, KNN 80% sedangkan NB 77% dan 70:30, KNN 79% sedangkan NB 77%.

4. KESIMPULAN

Penelitian yang dilakukan pada dataset *Metabolic Syndrome* yang diambil dari *Kaggle*, yang terdiri dari 505 data, menemukan bahwa penggunaan dua algoritma, K-NN dan *Naïve Bayes*, memberikan hasil yang paling optimal pada skema pembagian data 50:50. Dengan parameter K = 5, *accuracy* mencapai 82%, *precision* mencapai 87%, dan *recall* mencapai 85% pada KNN. *Naïve Bayes* menunjukkan hasil dengan nilai *Accuracy* sebesar 79%, *Precision* sebesar 77%, dan *Recall* sebesar 77%.

Berdasarkan hasil penelitian, dapat disimpulkan bahwa algoritma K-NN dan *Naïve Bayes* paling baik bekerja ketika data dibagi dengan rasio 50:50. Dalam skema pembagian data ini, algoritma K-NN mencapai tingkat akurasi tertinggi, yaitu sebesar 82%. Hasil ini menunjukkan bahwa dalam situasi di mana dataset

Metabolic Syndrome digunakan, K-NN dengan pembagian data 50:50 lebih efektif dalam memprediksi dan mengklasifikasikan hasil.

Untuk meningkatkan performa model ke tingkat yang sangat baik, penelitian yang akan datang harus melihat penambahan satu algoritma klasifikasi. Dengan memasukkan algoritma klasifikasi tambahan, penelitian dapat memberikan gambaran yang lebih komprehensif tentang berbagai teknik yang dapat diterapkan pada dataset *Metabolic Syndrome*. Dengan memasukkan algoritma klasifikasi tambahan, hasil yang lebih kuat dan generalisasi model dapat dicapai.

REFERENSI

- [1] U. Q. Ayunin, E. Retnowati, and J. H. Prayitno, "Prevalensi Komponen Sindrom Metabolik Pada Pegawai Bumn PT Wijaya Karya Divisi IV Surabaya," *Ghidza J. Gizi dan Kesehat.*, vol. 3, no. 1, p. 7, 2019, doi: 10.22487/j26227622.2019.v3.i1.12117.
- [2] D. P. R. Rahma Listyandini, Fenti Dewi Pertiwi, "Asupan Makan, Stress, dan Aktivitas Fisik dengan Sindrom Metabolik pada Pekerja Di Jakarta," *J. Kaji. dan Pengemb. Kesehat. Masy.*, vol. 01, no. 01, pp. 19–32, 2020, [Online]. Available: <https://jurnal.umj.ac.id/index.php/AN-NUR>
- [3] R. Christijani, "Penentuan Diagnosis Sindrom Metabolik Berdasarkan Penilaian Skor Sindrom Metabolik Dan Ncep Atp-Iii Pada Remaja [Penelitian Di Beberapa Sma Di Kota Bogor]," *Penelit. Gizi dan Makanan (The J. Nutr. Food Res.)*, vol. 42, no. 1, pp. 21–28, 2019, doi: 10.22435/pgm.v42i1.2418.
- [4] Septiyanti, N. Jafar, and Hendrayati, "Hubungan Pola Konsumsi Sayur dan Buah dengan Kejadian Sindrom Metabolik pada Pasien Rawat Jalan di RSUD Labuang Baji Kota Makassar," *Wind. Public Heal. J.*, vol. 1, no. 1, pp. 38–47, 2020, doi: 10.33096/woph.vi.40.
- [5] A. T. Abeng, J. P. Jasmani, and R. Fakultas, "Faktor Risiko Aktifitas Fisik dan Pola Makan Pada Sindrom Metabolik Di Kecamatan Jekan Raya Kota Palangka Raya," *J. Kesehat.*, pp. 1–5, 2019.
- [6] I. L. F. Amien, W. Astuti, and K. M. Lhaksamana, "Perbandingan Metode Naïve Bayes dan KNN (K-Nearest Neighbors) dalam Klasifikasi Penyakit Diabetes," *e-Proceeding Eng.*, vol. 10, no. 2, pp. 1911–1920, 2023.
- [7] S. Sahar, "Analisis Perbandingan Metode K-Nearest Neighbors dan Naïve Bayes Clasiffier Pada Dataset Penyakit Jantung," *Indones. J. Data Sci.*, vol. 1, no. 3, pp. 79–86, 2020, doi: 10.33096/ijodas.v1i3.20.
- [8] A. Muzakir, A. Desiani, and A. Amran, "Klasifikasi Penyakit Kanker Prostat Menggunakan Algoritma Naïve Bayes dan K-Nearest Neighbors," *Komputika J. Sist. Komput.*, vol. 12, no. 1, pp. 73–79, 2023, doi: 10.34010/komputika.v12i1.9629.
- [9] Ericha Apriliyani and Y. Salim, "Analisis performa metode klasifikasi Naïve Bayes Classifier pada Unbalanced Dataset," *Indones. J. Data Sci.*, vol. 3, no. 2, pp. 47–54, 2022, doi: 10.56705/ijodas.v3i2.45.
- [10] A. Putri *et al.*, "Komparasi Algoritma K-NN, Naive Bayes dan SVM untuk Prediksi Kelulusan Mahasiswa Tingkat Akhir," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 3, no. 1, pp. 20–26, 2023, doi: 10.57152/malcom.v3i1.610.
- [11] G. Abdurrahman, "Klasifikasi Penyakit Diabetes Melitus Menggunakan Adaboost Classifier," *JUSTINDO (Jurnal Sist. dan Teknol. Inf. Indones.)*, vol. 7, no. 1, pp. 59–66, 2022, doi: 10.32528/justindo.v7i1.4949.
- [12] W. Ramdhani, D. Bona, R. B. Musyaffa, and C. Rozikin, "Klasifikasi Penyakit Kanker Payudara Menggunakan Algoritma K-Nearest Neighbors," *J. Ilm. Wahana Pendidik.*, vol. 8, no. 12, pp. 445–452, 2022.
- [13] wahyono, "Improving the Speed of K-NN Algorithm for Vehicle Classification," *Tehno.Com*, vol. 19, no. 2, pp. 1–7, 2020.
- [14] T. Arifin and D. Ariesta, "Prediksi Penyakit Ginjal Kronis Menggunakan Algoritma Naive Bayes Classifier Berbasis Particle Swarm Optimization," *J. Tekno Insentif*, vol. 13, no. 1, pp. 26–30, 2019, doi: 10.36787/jti.v13i1.97.
- [15] K. Kartarina, N. K. Sriwinarti, and N. luh P. Juniarti, "Analisis Metode K-Nearest Neighbors (K-NN) Dan Naive Bayes Dalam Memprediksi Kelulusan Mahasiswa," *JTIM J. Teknol. Inf. dan Multimed.*, vol. 3, no. 2, pp. 107–113, 2021, doi: 10.35746/jtim.v3i2.159.
- [16] A. Noviriandini, P. Handayani, and Syahriani, "Prediksi Penyakit Liver Dengan Menggunakan Metode," *Pros. TAU SNAR-TEK Semin. Nas. Rekayasa dan Teknol.*, no. November, pp. 75–80, 2019.
- [17] M. Habibi and P. W. Cahyo, "Journal Classification Based on Abstract Using Cosine Similarity and Support Vector Machine," *JISKA (Jurnal Inform. Sunan Kalijaga)*, vol. 4, no. 3, p. 48, 2020, doi: 10.14421/jiska.2020.43-06.
- [18] F. Alghifari and D. Juardi, "Penerapan Data Mining Pada Penjualan Makanan Dan Minuman Menggunakan Metode Algoritma Naïve Bayes," *J. Ilm. Inform.*, vol. 9, no. 02, pp. 75–81, 2021, doi: 10.33884/jif.v9i02.3755.
- [19] Adhitya Karel Maulaya and Junadhi, "Analisis Sentimen Menggunakan Support Vector Machine

- Masyarakat Indonesia Di Twitter Terkait Bjorka,” *J. CoSciTech (Computer Sci. Inf. Technol.*, vol. 3, no. 3, pp. 495–500, 2022, doi: 10.37859/coscitech.v3i3.4358.
- [20] Gde Agung Brahmana Suryanegara, Adiwijaya, and Mahendra Dwifabri Purbolaksono, “Peningkatan Hasil Klasifikasi pada Algoritma Random Forest untuk Deteksi Pasien Penderita Diabetes Menggunakan Metode Normalisasi,” *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 5, no. 1, pp. 114–122, 2021, doi: 10.29207/resti.v5i1.2880.
- [21] B. Prasajo and E. Haryatmi, “Analisa Prediksi Kelayakan Pemberian Kredit Pinjaman dengan Metode Random Forest,” *J. Nas. Teknol. dan Sist. Inf.*, vol. 7, no. 2, pp. 79–89, 2021, doi: 10.25077/teknosi.v7i2.2021.79-89.
- [22] A. E. Satriatama *et al.*, “Analisis Klaster Data Pasien Diabetes untuk Identifikasi Pola dan Karakteristik Pasien,” *J. Teknol. Dan Sist. Inf. Bisnis*, vol. 5, no. 3, pp. 172–182, 2023, doi: 10.47233/jteksis.v5i3.828.
- [23] M. D. Purbolaksono, M. Irvan Tantowi, A. Imam Hidayat, and A. Adiwijaya, “Perbandingan Support Vector Machine dan Modified Balanced Random Forest dalam Deteksi Pasien Penyakit Diabetes,” *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 5, no. 2, pp. 393–399, 2021, doi: 10.29207/resti.v5i2.3008.
- [24] V. Ariyani, P. Putri, A. B. Prasetijo, and D. Eridani, “Perbandingan Kinerja Algoritme Naïve Bayes Dan K-Nearest Neighbors (Knn) Untuk Prediksi Harga Rumah,” *J. Ilm. Tek. Elektro*, no. 4, 2022, [Online]. Available: <https://ejournal.undip.ac.id/index.php/transmisi>
- [25] M. M. Baharuddin, H. Azis, and T. Hasanuddin, “Analisis Performa Metode K-Nearest Neighbors Untuk Identifikasi Jenis Kaca,” *Ilk. J. Ilm.*, vol. 11, no. 3, pp. 269–274, 2019, doi: 10.33096/ilkom.v11i3.489.269-274.
- [26] A. D. Adhi Putra, “Analisis Sentimen pada Ulasan pengguna Aplikasi Bibit Dan Bareksa dengan Algoritma KNN,” *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 8, no. 2, pp. 636–646, 2021, doi: 10.35957/jatisi.v8i2.962.
- [27] R. Siringoringo, “Klasifikasi Data Tidak Seimbang Menggunakan Algoritma SMOTE dan k-Nearest Neighbors,” *J. ISD*, vol. 3, no. 1, pp. 44–49, 2018.
- [28] D. Cahyanti, A. Rahmayani, and S. A. Husniar, “Analisis performa metode Knn pada Dataset pasien pengidap Kanker Payudara,” *Indones. J. Data Sci.*, vol. 1, no. 2, pp. 39–43, 2020, doi: 10.33096/ijodas.v1i2.13.
- [29] S. Cumel, David Zamri, Rahmaddeni, “Perbandingan Metode Data Mining untuk Prediksi Banjir Dengan Algoritma Naïve Bayes dan KNN,” *SENTIMAS Semin. Nas. Penelit. dan ...*, pp. 40–48, 2022, [Online]. Available: <https://journal.irpi.or.id/index.php/sentimas/article/view/353%0Ahttps://journal.irpi.or.id/index.php/sentimas/article/download/353/132>
- [30] D. Duei Putri, G. F. Nama, and W. E. Sulistiono, “Analisis Sentimen Kinerja Dewan Perwakilan Rakyat (DPR) Pada Twitter Menggunakan Metode Naive Bayes Classifier,” *J. Inform. dan Tek. Elektro Terap.*, vol. 10, no. 1, pp. 34–40, 2022, doi: 10.23960/jitet.v10i1.2262.
- [31] N. Indriyani, E. Ali, U. Rio, and R. Rahmaddeni, “Menentukan Kualitas Pelayanan Maskapai Penerbangan Domestik Dengan Menggunakan Metode Naïve Bayes,” *SATIN - Sains dan Teknol. Inf.*, vol. 6, no. 1, pp. 37–44, 2020, doi: 10.33372/stn.v6i1.605.