

Penerapan Algoritma K-Nearest Neighbors (KNN) Dalam Klasifikasi Penyakit Asma

John Franz Cristian E D L¹, Indra Kelana Jaya², Tamado Simon Sagala³

¹²³Fakultas Ilmu Komputer, Universitas Methodist Indonesia

Info Artikel

Histori Artikel:

Received, 10 Agustus 2026

Revised, 3 September 2026

Accepted, 25 September 2026

Keywords:

K-Nearest Neighbors, KNN, Klasifikasi, Penyakit Asma, Machine Learning

ABSTRAK

Penelitian ini bertujuan untuk mengimplementasikan algoritma K-Nearest Neighbors (KNN) dalam melakukan klasifikasi penyakit asma. Penyakit asma merupakan salah satu gangguan pernapasan kronis yang memerlukan diagnosis cepat dan tepat agar penanganannya dapat dilakukan secara efektif. Oleh karena itu, dibutuhkan metode klasifikasi berbasis algoritma yang mampu memberikan hasil akurasi yang baik dalam mengidentifikasi kondisi pasien. Dalam penelitian ini, data pasien yang mengandung atribut terkait kondisi kesehatan digunakan sebagai input untuk model KNN. Proses klasifikasi dilakukan dengan menentukan parameter nilai k terbaik guna memperoleh performa optimal. Evaluasi kinerja model dilakukan dengan menggunakan confusion matrix dan classification report yang mencakup metrik akurasi, precision, recall, dan f1-score. Hasil penelitian menunjukkan bahwa algoritma KNN mampu memberikan tingkat akurasi sebesar 85% dalam mengklasifikasikan penyakit asma. Selain itu, evaluasi classification report menghasilkan nilai precision, recall, dan f1-score pada kelas positif (1) masing-masing sebesar 0,88, 0,90, dan 0,89, sementara pada kelas negatif (0) masing-masing sebesar 0,77, 0,73, dan 0,75. Dengan hasil tersebut, dapat disimpulkan bahwa algoritma KNN dapat digunakan secara efektif sebagai metode klasifikasi penyakit asma dan berpotensi menjadi dasar pengembangan sistem pendukung keputusan dalam bidang medis.

This is an open access article under the [CC BY-SA](#) license



Penulis Koresponden:

John Franz Cristian E D L
Fakultas Ilmu Komputer,
Universitas Methodist Indonesia, Medan,
Jl. Hang Tuah No.8, Medan - Sumatera Utara.
Email: jhonfranz1010@gmail.com

1. PENDAHULUAN

Perkembangan teknologi kecerdasan buatan (Artificial Intelligence/AI) dan pembelajaran mesin (Machine Learning) dalam beberapa tahun terakhir menunjukkan kemajuan yang sangat pesat. Teknologi ini banyak dimanfaatkan dalam berbagai bidang untuk mendukung pengolahan data, analisis informasi, serta pengambilan keputusan berbasis data secara lebih akurat dan efisien. Dalam bidang informatika, machine learning berperan penting dalam pengembangan sistem cerdas yang mampu mempelajari pola dari data historis dan menghasilkan proses klasifikasi maupun prediksi secara otomatis. Peningkatan ketersediaan data dalam jumlah besar mendorong kebutuhan akan metode analisis yang andal dan sistematis agar informasi yang dihasilkan dapat dimanfaatkan secara optimal.

Machine learning terdiri dari berbagai pendekatan analisis data, di antaranya teknik klasifikasi, regresi, pengelompokan data (clustering), serta pengembangan model berbasis deep learning. Masing-

masing pendekatan memiliki karakteristik dan fungsi yang berbeda dalam menyelesaikan permasalahan nyata [2]. Pada penelitian ini, fokus diarahkan pada teknik klasifikasi, yaitu proses pengelompokan data ke dalam kelas tertentu berdasarkan karakteristik yang dimiliki oleh data tersebut. Salah satu algoritma klasifikasi yang banyak digunakan adalah K-Nearest Neighbors (KNN), yang dikenal memiliki konsep sederhana serta mampu memberikan hasil yang cukup baik pada data dengan pola yang kompleks dan non-linear[3].

Dalam bidang kesehatan, penerapan machine learning semakin berkembang dan memberikan kontribusi besar dalam mendukung proses diagnosis penyakit, pemantauan kondisi pasien, serta pengambilan keputusan klinis[4]. Salah satu penyakit yang memerlukan perhatian khusus dalam proses diagnosis adalah asma. Asma merupakan salah satu penyakit kronis pada sistem pernapasan yang ditandai oleh adanya peradangan dan penyempitan saluran napas. Kondisi ini dapat menyebabkan berbagai gejala, seperti sesak napas, mengi, batuk, serta rasa tidak nyaman pada dada. Gejala-gejala tersebut sering kali menyerupai gangguan pernapasan lainnya, sehingga proses identifikasi penyakit asma pada tahap awal dapat menjadi cukup sulit apabila hanya mengandalkan pemeriksaan klinis konvensional.

Dalam praktik klinis, diagnosis asma umumnya dilakukan oleh tenaga medis dengan menganalisis gejala yang dialami pasien, disertai dengan pertimbangan berbagai faktor pendukung, seperti usia, jenis kelamin, riwayat merokok, hasil pemeriksaan fungsi paru, serta faktor risiko lainnya. Proses ini sangat bergantung pada pengalaman dan ketelitian dokter, sehingga berpotensi menimbulkan subjektivitas dan risiko kesalahan diagnosis. Oleh karena itu, diperlukan suatu pendekatan berbasis data yang mampu membantu tenaga medis dalam melakukan klasifikasi penyakit secara lebih objektif, sistematis, dan konsisten.

Penerapan algoritma K-Nearest Neighbors (KNN) dalam klasifikasi penyakit asma diharapkan dapat menjadi solusi pendukung dalam proses diagnosis awal. Dengan memanfaatkan data gejala dan faktor risiko pasien, algoritma KNN dapat mengklasifikasikan kondisi pasien ke dalam kategori penderita asma atau bukan berdasarkan kedekatan pola data dengan data pasien sebelumnya. Pendekatan ini tidak dimaksudkan untuk menggantikan peran dokter, melainkan sebagai second opinion berbasis sistem cerdas yang dapat membantu meningkatkan ketepatan diagnosis, mengurangi risiko kesalahan, serta mempercepat proses pengambilan keputusan medis.

Beberapa penelitian sebelumnya menunjukkan bahwa algoritma KNN memiliki performa yang cukup baik dalam klasifikasi penyakit. Salah satu penelitian menerapkan algoritma KNN untuk mendiagnosis diabetes gestasional (Gestational Diabetes Mellitus/GDM) menggunakan data klinis dari Pima Indians Diabetes Database. Melalui tahapan praproses data, penanganan missing values, normalisasi data, serta pemilihan nilai parameter k yang optimal, model KNN tersebut mampu mencapai tingkat akurasi sekitar 76%[5]. Hasil penelitian tersebut menunjukkan bahwa pemilihan fitur yang relevan dan pengolahan data yang tepat sangat berpengaruh terhadap performa algoritma KNN dalam klasifikasi penyakit.

Penelitian ini bertujuan untuk menerapkan algoritma K-Nearest Neighbors (KNN) dalam melakukan klasifikasi penyakit asma berdasarkan data gejala dan faktor risiko pasien. Diharapkan hasil penelitian ini dapat memberikan kontribusi dalam pengembangan sistem pendukung keputusan di bidang kesehatan, khususnya dalam membantu tenaga medis melakukan diagnosis awal penyakit asma secara lebih cepat, akurat, dan berbasis data.

2. METODE PENELITIAN

Framework penelitian dalam penelitian ini dirancang untuk menggambarkan secara sistematis seluruh tahapan yang dilakukan agar penelitian berjalan sesuai dengan tujuan yang telah ditetapkan. Kerangka kerja ini mencakup proses mulai dari studi literatur, pengumpulan data, penerapan metode, hingga evaluasi hasil. Kerangka kerja pada penelitian ini dapat dilihat pada Gambar 1.



Gambar 1. Framework Penelitian

Penelitian dimulai Peneliti melakukan tahap studi literatur. Tahap studi literatur merupakan kegiatan untuk mengumpulkan informasi dari berbagai sumber, dilanjutkan dengan pengumpulan data yang diperoleh dari website Kaggle. Dataset yang diunduh berjudul Penyakit Asma dengan variabel berjumlah 18 yang disusun berdasarkan data medis dan klinis terkait pasien yang didiagnosis dengan Penyakit Asma. Kemudian dilakukan analisis menggunakan metode K-Nearest Neighbors (KNN) hingga diperoleh hasil dan kesimpulan dari analisis yang dilakukan.

3. HASIL DAN PEMBAHASAN

3.1 Implementasi Algoritma K- Nearest Nighbor (KNN)

Implementasi Algoritma K- Nearest Nighbor (KNN) dilakukan dengan tahapan sebagai berikut[6].

3.1.1 Penentuan Nilai K

Proses perhitungan dilakukan dengan menggunakan nilai $K=3$ dan menggunakan perbandingan data yaitu 80:20. Pada perhitungan manual digunakan 20 data sebagai data sampel. Pemilihan nilai $K = 3$ dipilih karena pada algoritma K-Nearest Neighbors (KNN), nilai K yang terlalu kecil (misalnya $K = 1$ atau $K = 2$) dapat membuat model menjadi terlalu sensitif terhadap noise atau data yang menyimpang, sehingga hasil klasifikasi cenderung kurang stabil. Dengan menggunakan $K = 3$, prediksi yang dihasilkan lebih representatif karena keputusan klasifikasi didasarkan pada mayoritas dari tiga tetangga terdekat, bukan hanya satu atau dua tetangga, sehingga hasilnya lebih seimbang dan akurat[7]. Data sampel dapat dilihat pada Tabel 1

Tabel 1 Data Sampel

No	Age	Gender	BMI	Smoking	Physical Activity	...	Diagnosis
1	60	1	32,713	0	0,417		1
2	5	1	34,582	0	5,783		1
3	49	0	21,742	0	3,355		1
4	32	0	22,697	0	4,238		1
5	21	0	16,339	1	7,889		1
6	54	1	30,307	0	9,387		0
7	23	1	26,827	0	9,409		0
8	77	1	28,904	0	6,970		0
9	33	0	15,221	0	4,033		0
10	48	1	39,174	1	0,396		1
11	58	0	35,016	0	8,679		0
12	77	0	36,090	0	4,401		1
13	15	1	36,897	0	7,723		0
14	42	1	15,180	0	1,288		1
15	78	0	16,777	0	7,024		0
16	8	0	35,997	0	7,601		0
17	27	1	34,901	0	1,028		0

18	19	1	37,914	0	5,596	0
19	40	1	19,521	0	1,634	1
20	36	1	30,049	0	1,262	1

Data 1-16 digunakan sebagai data latih (total data latih 16 data) dan data 17-20 digunakan sebagai data uji (total data uji 4 data).

3.1.2 Menghitung Nilai Euclidean Distance

Perhitungan jarak setiap data uji terhadap masing-masing data latih dilakukan menggunakan rumus jarak Euclidean Distance hingga ditemukan nilai minimum terhadap setiap data latih. Berikut merupakan tahapan dari perhitungan menggunakan K-Nearest Neighbor (KNN) dengan mencari nilai Euclidean Distance[8].

1. Menghitung jarak antara data uji 17 dengan setiap data latih menggunakan Euclidean Distance[9].

$$D1 = \sqrt{(D1_1 - D17_1)^2 + (D1 \dots - D17 \dots)^2 + (D1_{24} - D17_{24})^2}$$

$$D1 = \sqrt{(60 - 27)^2 + (\dots - \dots)^2 + (0 - 1)^2}$$

$$D1 = 34,15$$

$$D2 = \sqrt{(D2_1 - D17_1)^2 + (D2 \dots - D17 \dots)^2 + (D2_{24} - D17_{24})^2}$$

$$D2 = \sqrt{(5 - 27)^2 + (\dots - \dots)^2 + (0 - 1)^2}$$

$$D2 = 25,55$$

$$D3 = \sqrt{(D3_1 - D17_1)^2 + (D3 \dots - D17 \dots)^2 + (D3_{24} - D17_{24})^2}$$

$$D3 = \sqrt{(49 - 27)^2 + (\dots - \dots)^2 + (1 - 1)^2}$$

$$D3 = 26,79$$

.

.

.

$$D16 = \sqrt{(D16_1 - D17_1)^2 + (D16 \dots - D17 \dots)^2 + (D16_{24} - D17_{24})^2}$$

$$D16 = \sqrt{(8 - 27)^2 + (\dots - \dots)^2 + (1 - 1)^2}$$

$$D16 = 23,08$$

Perhitungan di atas dilakukan untuk menghitung euclidean distance antara data 17 dengan setiap data latih. Berdasarkan hasil perhitungan kemudian diurutkan jaraknya (Rank) antara data latih dengan data uji, dari yang terdekat hingga terjauh, hasil perhitungan dapat dilihat dilihat pada Tabel 2

Tabel 2 Hasil Euclidean Distance Data 17

Data	Label (Diagnosis)	Nilai Euclidean	Ranking Euclidean
D1	1	34,15	13
D2	1	25,55	8
D3	1	26,79	10
D4	1	16,61	2
D5	1	23,37	7
D6	0	30,66	11
D7	0	15,05	1
D8	0	51,98	15
D9	0	22,53	4
D10	1	22,94	5
D11	0	33,00	12
D12	1	50,93	14
D13	0	17,67	3
D14	1	26,52	9
D15	0	55,80	16
D16	0	23,08	6

Menggunakan nilai K tetangga terdekat (K = 3) dari hasil Euclidean Distance di dapatkan bahwa nilai euclidean distance 3 terkecil berurut yaitu D7=15,05, D4=16,61 dan D13=17,67. Selanjutnya

mencari mayoritas Label dari 3 nilai euclidean terkecil tersebut yaitu : 0, 1 dan 0. Berdasarkan K- NN dengan K = 3, K17 masuk kedalam Label “0”.

Jika digunakan K = 2, maka dua jarak terkecil yang diperoleh adalah D7 = 15,05 (Label 0) dan D4 = 16,61 (Label 1). Hasil ini menimbulkan kondisi seri antara label 0 dan 1 sehingga klasifikasi menjadi tidak efektif, karena KNN tidak dapat menentukan mayoritas kelas secara jelas. Oleh karena itu, pemilihan nilai K = 3 lebih tepat digunakan agar keputusan klasifikasi lebih stabil.

2. Menghitung jarak antara data uji 18 dengan setiap data latih menggunakan Euclidean Distance.

$$D1 = \sqrt{(D1_1 - D18_1)^2 + (D1 \dots - D18 \dots)^2 + (D1_{24} - D18_{24})^2}$$

$$D1 = \sqrt{(60 - 19)^2 + (\dots - \dots)^2 + (0 - 1)^2}$$

$$D1 = 42,34$$

$$D2 = \sqrt{(D2_1 - D18_1)^2 + (D2 \dots - D18 \dots)^2 + (D2_{24} - D18_{24})^2}$$

$$D2 = \sqrt{(5 - 19)^2 + (\dots - \dots)^2 + (0 - 1)^2}$$

$$D2 = 15,33$$

$$D3 = \sqrt{(D3_1 - D18_1)^2 + (D3 \dots - D18 \dots)^2 + (D3_{24} - D18_{24})^2}$$

$$D3 = \sqrt{(49 - 19)^2 + (\dots - \dots)^2 + (1 - 1)^2}$$

$$D3 = 35,19$$

.

.

.

$$D16 = \sqrt{(D16_1 - D18_1)^2 + (D16 \dots - D18 \dots)^2 + (D16_{24} - D18_{24})^2}$$

$$D16 = \sqrt{(8 - 19)^2 + (\dots - \dots)^2 + (1 - 1)^2}$$

$$D16 = 13,84$$

Perhitungan di atas dilakukan untuk menghitung euclidean distance antara data 18 dengan setiap data latih. Berdasarkan hasil perhitungan kemudian diurutkan jaraknya (Rank) antara data latih dengan data uji, dari yang terdekat hingga terjauh, hasil perhitungan dapat dilihat pada Tabel 3

Tabel 3 Hasil Euclidean Distance Data 18

Data	Label (Diagnosis)	Nilai Euclidean	Ranking Euclidean
D1	1	42,34	13
D2	1	15,33	3
D3	1	35,19	10
D4	1	20,85	5
D5	1	24,93	6
D6	0	37,64	11
D7	0	16,67	4
D8	0	58,88	15
D9	0	28,82	7
D10	1	31,47	8
D11	0	39,96	12
D12	1	58,62	14
D13	0	12,88	1
D14	1	33,19	9
D15	0	63,32	16
D16	0	13,84	2

Menggunakan nilai K tetangga terdekat (K = 3) dari hasil Euclidean Distance di dapatkan bahwa nilai euclidean distance 3 terkecil berurut yaitu D13=12,88, D16=13,84 dan D2=15,33. Selanjutnya mencari mayoritas Label dari 3 nilai euclidean terkecil tersebut yaitu : 0, 0 dan 1. Berdasarkan K- NN dengan K = 3, K18 masuk kedalam Label “0”.

Jika digunakan K = 2, maka dua jarak Euclidean terkecil adalah D13 = 12,88 (Label 0) dan D16 = 13,84 (Label 0). Karena kedua tetangga terdekat memiliki label yang sama, yaitu 0, maka berdasarkan

algoritma K-Nearest Neighbors (KNN) dengan $K = 2$, data K18 tetap diklasifikasikan ke dalam Label 0. Namun demikian, penggunaan $K = 2$ kurang direkomendasikan karena berpotensi menimbulkan kondisi seri pada kasus lain, sehingga klasifikasi bisa menjadi tidak efektif dibandingkan dengan penggunaan nilai $K = 3$.

3. Menghitung jarak antara data uji 19 dengan setiap data latih menggunakan Euclidean Distance.

$$D1 = \sqrt{(D1_1 - D19_1)^2 + (D1 \dots - D19 \dots)^2 + (D1_{24} - D19_{24})^2}$$

$$D1 = \sqrt{(60 - 40)^2 + (\dots - \dots)^2 + (0 - 1)^2}$$

$$D1 = 26,05$$

$$D2 = \sqrt{(D2_1 - D19_1)^2 + (D2 \dots - D19 \dots)^2 + (D2_{24} - D19_{24})^2}$$

$$D2 = \sqrt{(5 - 40)^2 + (\dots - \dots)^2 + (0 - 1)^2}$$

$$D2 = 39,58$$

$$D3 = \sqrt{(D3_1 - D19_1)^2 + (D3 \dots - D19 \dots)^2 + (D3_{24} - D19_{24})^2}$$

$$D3 = \sqrt{(49 - 40)^2 + (\dots - \dots)^2 + (1 - 1)^2}$$

$$D3 = 12,39$$

.

.

.

$$D16 = \sqrt{(D16_1 - D19_1)^2 + (D16 \dots - D19 \dots)^2 + (D16_{24} - D19_{24})^2}$$

$$D16 = \sqrt{(8 - 40)^2 + (\dots - \dots)^2 + (1 - 1)^2}$$

$$D16 = 37,61$$

Perhitungan di atas dilakukan untuk menghitung euclidean distance antara data 19 dengan setiap data latih. Berdasarkan hasil perhitungan kemudian diurutkan jaraknya (Rank) antara data latih dengan data uji, dari yang terdekat hingga terjauh, hasil perhitungan dapat dilihat pada Tabel 4

Tabel 4 Hasil Euclidean Distance Data 19

Data	Label (Diagnosis)	Nilai Euclidean	Ranking Euclidean
D1	1	26,05	9
D2	1	39,58	13
D3	1	12,39	3
D4	1	13,44	4
D5	1	24,27	8
D6	0	21,70	5
D7	0	21,88	6
D8	0	39,98	15
D9	0	11,00	2
D10	1	23,51	7
D11	0	26,08	10
D12	1	41,78	16
D13	0	33,89	11
D14	1	10,14	1
D15	0	39,79	14
D16	0	37,61	12

Menggunakan nilai K tetangga terdekat ($K = 3$) dari hasil Euclidean Distance di dapatkan bahwa nilai euclidean distance 3 terkecil berurut yaitu $D14=10,14$, $D9=11,00$ dan $D3=12,39$. Selanjutnya mencari mayoritas Label dari 3 nilai euclidean terkecil tersebut yaitu : 1, 0 dan 1. Berdasarkan K- NN dengan $K = 3$, K18 masuk kedalam Label “1”.

Jika digunakan $K = 2$, maka dua jarak Euclidean terkecil adalah $D14 = 10,14$ (Label 1) dan $D9 = 11,00$ (Label 0). Hasil ini menimbulkan kondisi seri antara label 1 dan 0, sehingga klasifikasi dengan $K = 2$ menjadi tidak efektif karena tidak ada mayoritas kelas yang jelas.

4. Menghitung jarak antara data uji 20 dengan setiap data latih menggunakan Euclidean Distance.

$$D1 = \sqrt{(D1_1 - D20_1)^2 + (D1 \dots - D20 \dots)^2 + (D1_{24} - D20_{24})^2}$$

$$D1 = \sqrt{(60 - 36)^2 + (\dots - \dots)^2 + (0 - 1)^2}$$

$$D1 = 26,48$$

$$D2 = \sqrt{(D2_1 - D20_1)^2 + (D2 \dots - D20 \dots)^2 + (D2_{24} - D20_{24})^2}$$

$$D2 = \sqrt{(5 - 36)^2 + (\dots - \dots)^2 + (0 - 1)^2}$$

$$D2 = 34,62$$

$$D3 = \sqrt{(D3_1 - D20_1)^2 + (D3 \dots - D20 \dots)^2 + (D3_{24} - D20_{24})^2}$$

$$D3 = \sqrt{(49 - 36)^2 + (\dots - \dots)^2 + (1 - 1)^2}$$

$$D3 = 16,44$$

.

.

.

$$D16 = \sqrt{(D16_1 - D20_1)^2 + (D16 \dots - D20 \dots)^2 + (D16_{24} - D20_{24})^2}$$

$$D16 = \sqrt{(8 - 36)^2 + (\dots - \dots)^2 + (1 - 1)^2}$$

$$D16 = 31,18$$

Perhitungan di atas dilakukan untuk menghitung euclidean distance antara data 20 dengan setiap data latih. Berdasarkan hasil perhitungan kemudian diurutkan jaraknya (Rank) antara data latih dengan data uji, dari yang terdekat hingga terjauh, hasil perhitungan dapat dilihat pada Tabel 5

Tabel 4 Hasil Euclidean Distance Data 20

Data	Label (Diagnosis)	Nilai Euclidean	Ranking Euclidean
D1	1	26,48	11
D2	1	34,62	13
D3	1	16,44	2
D4	1	14,26	1
D5	1	23,85	8
D6	0	22,68	7
D7	0	17,19	3
D8	0	43,43	15
D9	0	19,08	4
D10	1	19,21	5
D11	0	24,71	9
D12	1	42,42	14
D13	0	25,04	10
D14	1	20,19	6
D15	0	45,95	16
D16	0	31,18	12

Menggunakan nilai K tetangga terdekat ($K = 3$) dari hasil Euclidean Distance di dapatkan bahwa nilai euclidean distance 3 terkecil berurut yaitu $D4=14,26$, $D3=16,44$ dan $D7=17,19$. Selanjutnya mencari mayoritas Label dari 3 nilai euclidean terkecil tersebut yaitu : 1, 1 dan 0. Berdasarkan K- NN dengan $K = 3$, K18 masuk kedalam Label “1”. Jika digunakan $K = 2$, maka dua jarak Euclidean terkecil adalah $D4 = 14,26$ (Label 1) dan $D3 = 16,44$ (Label 1). Karena kedua tetangga terdekat memiliki label yang sama, yaitu 1, maka berdasarkan algoritma K-Nearest Neighbors (KNN) dengan $K = 2$, data K18 juga diklasifikasikan ke dalam Label 1. Namun demikian, penggunaan nilai $K = 2$ tetap kurang direkomendasikan karena pada kasus lain dapat menimbulkan kondisi seri, sedangkan dengan $K = 3$ hasil klasifikasi lebih stabil.

3.2 Aplikasi Klasifikasi Penyakit Asma

Aplikasi Klasifikasi Penyakit Asma dirancang menggunakan bahasa pemrograman Python dan framework Streamlit untuk membangun antarmuka pengguna berbasis web. Program ini dirancang untuk melakukan analisis data pasien dengan algoritma K- Nearest Nighbor (KNN).

1. Tampilan Antar Muka Awal

Tampilan awal dari sebuah aplikasi berbasis web yang dikembangkan menggunakan Streamlit, dengan judul "Klasifikasi Penyakit Asma dengan K-Nearest Neighbors (KNN)". Aplikasi ini dirancang untuk melakukan klasifikasi terhadap data pasien guna menentukan apakah seseorang menderita penyakit asma atau tidak, dengan memanfaatkan algoritma KNN sebagai metode klasifikasinya. Pada bagian awal antarmuka, pengguna diarahkan untuk mengunggah file Excel (.xlsx) yang berisi data pasien dan hasil diagnosis sebelumnya. Terdapat komponen unggah file dengan tampilan drag and drop atau pilihan untuk memilih file secara manual melalui tombol Browse files, serta informasi bahwa format file yang diterima adalah Excel dengan ukuran maksimal 200MB. Selain itu, ditampilkan juga notifikasi yang berbunyi "Silakan upload file Excel terlebih dahulu" sebagai pengingat bahwa proses klasifikasi tidak dapat dilanjutkan sebelum data dimasukkan. Secara keseluruhan, tampilan awal aplikasi ini sudah cukup informatif dan user-friendly, memudahkan pengguna dalam memahami langkah awal yang perlu dilakukan sebelum melanjutkan ke proses klasifikasi penyakit asma. Tampilan antar muka awal dapat dilihat pada Gambar 1



Gambar 1 Tampilan Antar Muka Awal

2. Menampilkan Data

Tampilan bagian dari dataset asli yang telah diunggah oleh pengguna ke dalam aplikasi. Dataset ini merupakan hasil pembacaan dari file Excel bernama Data Asma.xlsx, yang berisi informasi berbagai variabel terkait pasien, yang akan digunakan dalam proses klasifikasi penyakit asma menggunakan algoritma K-Nearest Neighbors (KNN). Tabel data memperlihatkan sejumlah kolom seperti: Age, Gender, BMI, Smoking, PhysicalActivity, DietQuality, SleepQuality, PollutionExposure, PollenExposure, DustExposure, PetAllergy, FamilyHistoryAsthma, HistoryOfAllergies, Eczema, HayFever, GastroesophagealReflux, LungFunctionFEV1, LungFunctionFVC, Wheezing, ShortnessOfBreath, ChestTightness, Coughing, NighttimeSymptoms, ExerciseInduced dan Diagnosis.

Setiap baris pada tabel mewakili satu entri pasien, dengan nilai-nilai numerik yang mencerminkan kondisi atau faktor risiko tertentu yang dapat mempengaruhi diagnosis penyakit asma. Data seperti ini sangat cocok untuk digunakan dalam klasifikasi berbasis pembelajaran mesin, khususnya KNN, karena formatnya bersih, terstruktur, dan mayoritas terdiri dari variabel numerik yang dapat dihitung jaraknya. Tampilan visualisasi data ini juga memberikan transparansi kepada pengguna untuk melihat apakah data yang diunggah telah dimuat dengan benar sebelum dilakukan proses pelatihan dan prediksi model. Tampilan data yang telah diunggah ke aplikasi dapat dilihat pada Gambar 2

Data Asma.xlsx 342.340

Data Asli

No	Age	Gender	BMI	Smoking	PhysicalActivity	DietQuality	SleepQuality	PollutionExposure	PollenExposure	DustExposure	PetAllergy	FamilyHistory
0	1	40	1	32.713	0	0.4167	0.0267	4.7511	3.1404	7.6463	3.5024	0
1	2	5	1	34.5619	0	5.7834	2.7961	5.7963	0.4163	8.6984	1.2329	0
2	3	49	0	21.7418	0	3.355	5.9676	6.3934	6.4625	2.1582	1.249	0
3	4	32	0	22.6971	0	4.2383	5.8025	7.0502	1.3682	7.4043	0.6863	1
4	5	21	0	16.3392	1	7.8893	7.9412	9.6043	2.2627	0.7771	6.8411	0
5	6	54	1	30.3068	0	9.3871	1.7557	6.7012	0.458	2.5412	8.4088	0
6	7	23	1	26.8266	0	9.4091	4.8963	4.0422	9.9744	2.9562	5.7306	1
7	8	77	1	28.904	0	6.9697	4.4966	6.6122	0.001	7.9751	1.6457	0
8	9	33	0	15.2214	0	4.0328	2.4541	6.931	8.7396	6.3137	7.3066	0

Gambar 1 Tampilan Data

3. Tampilan Data Setelah Dibersihkan

Gambar tersebut menunjukkan tabel “Data Setelah Dibersihkan” yang ditampilkan dalam aplikasi klasifikasi penyakit asma. Tabel ini berisi data pasien yang telah melalui proses pembersihan, yaitu penghapusan nilai-nilai kosong (missing values), sehingga hanya menyisakan baris data yang lengkap dan siap untuk dianalisis lebih lanjut. Setiap baris mewakili satu pasien, sementara kolom mencakup berbagai fitur atau variabel yang relevan terhadap risiko asma. Beberapa fitur tersebut antara lain: Age, Gender, BMI, Smoking, PhysicalActivity, DietQuality, SleepQuality, PollutionExposure, PollenExposure, DustExposure, PetAllergy, FamilyHistoryAsthma, HistoryOfAllergies, Eczema, HayFever, GastroesophagealReflux, LungFunctionFEV1, LungFunctionFVC, Wheezing, ShortnessOfBreath, ChestTightness, Coughing, NighttimeSymptoms, ExerciseInduced dan Diagnosis.

Data ini sangat penting karena akan digunakan sebagai input dalam model pembelajaran mesin berbasis K-Nearest Neighbors (KNN) untuk mengklasifikasikan kemungkinan pasien mengalami penyakit asma. Visualisasi ini memberikan gambaran jelas kepada pengguna mengenai bentuk dan kualitas data yang digunakan sebelum dimasukkan ke dalam algoritma. Penyajian data yang rapi dan bersih juga memastikan bahwa hasil prediksi yang dihasilkan lebih akurat dan dapat dipercaya. Tampilan hasil data yang telah dibersihkan dapat dilihat pada Gambar 3

Data Setelah Dibersihkan

No	Age	Gender	BMI	Smoking	PhysicalActivity	DietQuality	SleepQuality	PollutionExposure	PollenExposure	DustExposure	PetAllergy	FamilyHistory
0	1	40	1	32.713	0	0.4167	0.0267	4.7511	3.1404	7.6463	3.5024	0
1	2	5	1	34.5619	0	5.7834	2.7961	5.7963	0.4163	8.6984	1.2329	0
2	3	49	0	21.7418	0	3.355	5.9676	6.3934	6.4625	2.1582	1.249	0
3	4	32	0	22.6971	0	4.2383	5.8025	7.0502	1.3682	7.4043	0.6863	1
4	5	21	0	16.3392	1	7.8893	7.9412	9.6043	2.2627	0.7771	6.8411	0
5	6	54	1	30.3068	0	9.3871	1.7557	6.7012	0.458	2.5412	8.4088	0
6	7	23	1	26.8266	0	9.4091	4.8963	4.0422	9.9744	2.9562	5.7306	1
7	8	77	1	28.904	0	6.9697	4.4966	6.6122	0.001	7.9751	1.6457	0
8	9	33	0	15.2214	0	4.0328	2.4541	6.931	8.7396	6.3137	7.3066	0
9	10	48	1	35.1739	1	0.3064	9.5108	5.6728	5.3953	8.9068	6.2256	0

Gambar 2 Tampilan Data Setelah Dibersihkan

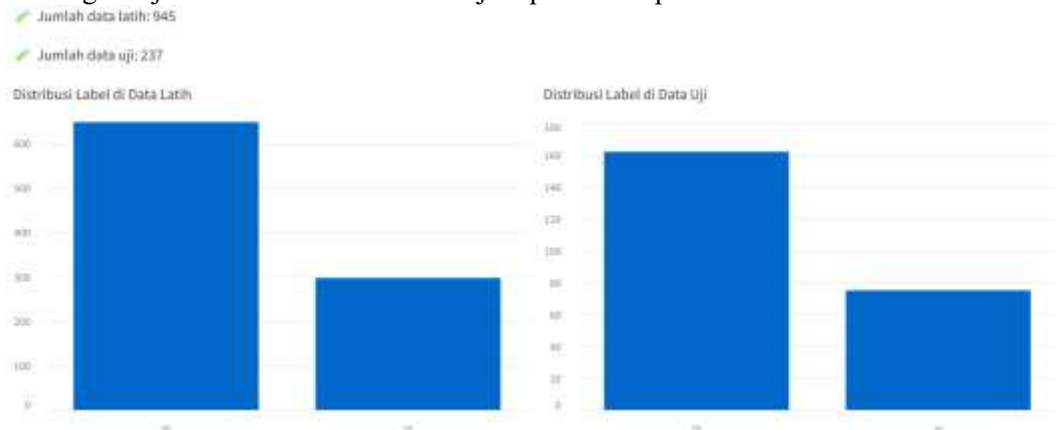
4. Tampilan Pembagian Data Latih dan Data Uji

Tampilan informasi mengenai jumlah data latih dan data uji yang digunakan dalam proses klasifikasi penyakit asma menggunakan algoritma K-Nearest Neighbors (KNN). Jumlah data latih yang digunakan adalah 945, sedangkan jumlah data uji adalah 237. Data ini telah dibagi secara proporsional agar model dapat dilatih secara optimal dan dievaluasi dengan baik.

Terdapat pula dua diagram batang yang menggambarkan distribusi label pada masing-masing set data. Diagram pertama menunjukkan distribusi label pada data latih, di mana label 0 (tidak terkena asma)

jauh lebih banyak dibanding label 1 (terkena asma). Hal serupa juga terlihat pada data uji, yang memperlihatkan ketidakseimbangan antara jumlah sampel dari dua kelas tersebut.

Ketidakseimbangan distribusi label ini penting untuk diperhatikan karena dapat mempengaruhi kinerja model KNN. Model cenderung bias terhadap kelas mayoritas (label 0), sehingga langkah-langkah seperti penyeimbangan data, penggunaan metrik evaluasi yang tepat, atau penyesuaian parameter model mungkin diperlukan untuk memastikan hasil klasifikasi yang adil dan akurat. Grafik ini membantu pengguna memahami karakteristik data sebelum proses pelatihan dan pengujian model dilakukan. Tampilan informasi mengenai jumlah data latih dan data uji dapat dilihat pada Gambar 4



Gambar 3 Tampilan informasi mengenai jumlah data latih dan data uji

5. Tampilan Penentuan Nilai K

Tampilan penentuan nilai k menunjukkan antarmuka pengguna berupa slider untuk memilih nilai jumlah tetangga terdekat (k) dalam algoritma K-Nearest Neighbors (KNN). Nilai k saat ini dipilih sebesar 3, yang berarti bahwa untuk menentukan kelas dari suatu data uji, sistem akan melihat 3 tetangga terdekat dari data tersebut berdasarkan jarak (Euclidean). Tampilan penentuan nilai k dapat dilihat pada Gambar 5



Gambar 5 Tampilan Penentuan Nilai K

6. Tampilan Hasil Klasifikasi

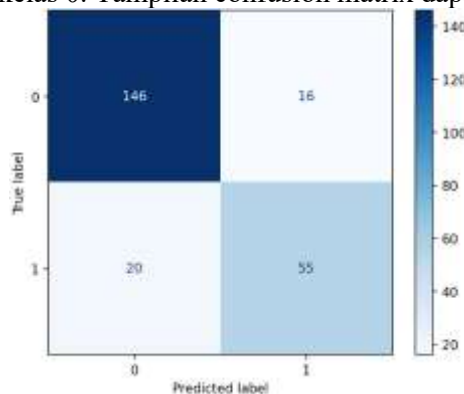
Tampilan hasil klasifikasi menampilkan sebuah tabel yang berisi hasil evaluasi prediksi model klasifikasi menggunakan algoritma K-Nearest Neighbors (KNN). Tabel ini membandingkan antara label aktual dan label prediksi dari data uji yang digunakan untuk mengukur akurasi model. Kolom pertama menunjukkan nomor atau indeks dari masing-masing data uji. Kolom kedua berisi label aktual, yaitu nilai sebenarnya dari data yang sudah diketahui sebelumnya, sementara kolom ketiga menunjukkan label prediksi yang dihasilkan oleh model KNN. Dari hasil yang ditampilkan, dapat diamati bahwa sebagian besar prediksi sudah sesuai dengan label aktual, namun masih terdapat beberapa data yang salah klasifikasi. Misalnya, pada salah satu baris, label aktual bernilai 1 tetapi diprediksi menjadi 0. Kesalahan seperti ini penting untuk dianalisis lebih lanjut guna mengetahui kelemahan model. Tabel ini menjadi acuan awal untuk melihat performa klasifikasi model KNN dan dapat dikembangkan lebih lanjut untuk menampilkan metrik evaluasi seperti akurasi, presisi, recall, dan confusion matrix.

Hasil Prediksi			
Number Data (y)	Label Aktual	Label Prediksi	
0	258	1	1
1	961	0	8
2	896	1	8
3	636	0	8
4	44	0	8
5	250	1	1
6	665	0	8
7	533	0	8
8	140	0	8
9	528	0	8

Gambar 6 Tampilan Hasil Klasifikasi

3.2.1 Confusion Matrix

Tampilan confusion matrix yang digunakan untuk mengevaluasi performa model klasifikasi. Pada sumbu horizontal ditampilkan label hasil prediksi model, sedangkan pada sumbu vertikal ditampilkan label sebenarnya (true label)[10]. Matriks ini terbagi menjadi empat bagian, yaitu: True Negative (TN) sebanyak 146, False Positive (FP) sebanyak 16, False Negative (FN) sebanyak 20, dan True Positive (TP) sebanyak 55. Nilai TN menunjukkan jumlah data yang benar diprediksi sebagai kelas 0, sedangkan TP menunjukkan data yang benar diprediksi sebagai kelas 1. Sementara itu, FP adalah jumlah data yang sebenarnya termasuk kelas 0 tetapi salah diprediksi sebagai kelas 1, dan FN adalah jumlah data yang sebenarnya termasuk kelas 1 tetapi salah diprediksi sebagai kelas 0. Tampilan confusion matrix dapat dilihat pada Tabel 7



Gambar 4 Tampilan confusion matrix

Dari matriks diatas dapat disimpulkan bahwa model memiliki performa yang cukup baik dengan jumlah prediksi yang benar ($TN + TP = 201$) lebih besar dibanding jumlah prediksi yang salah ($FP + FN = 36$). Visualisasi ini juga memperkuat evaluasi sebelumnya bahwa model masih menghasilkan kesalahan, khususnya dalam mengklasifikasikan kelas 1, yang terlihat dari nilai FN (20) yang relatif lebih tinggi. Oleh karena itu, model masih bisa dioptimalkan, misalnya dengan menyesuaikan parameter atau menyeimbangkan distribusi label dalam data latih.

Berdasarkan hasil evaluasi model yang ditampilkan pada gambar 4.8 berikut, diperoleh akurasi sebesar 0.85 atau 85%. Nilai ini menunjukkan bahwa model mampu mengklasifikasikan data dengan cukup baik. Dari tabel klasifikasi, kelas 0 memiliki nilai precision sebesar 0.88, recall 0.90, dan f1-score 0.89, yang mengindikasikan bahwa model sangat baik dalam mengidentifikasi data yang termasuk dalam kelas 0. Sementara itu, untuk kelas 1, nilai precision adalah 0.77, recall 0.73, dan f1-score 0.75, yang menunjukkan bahwa kinerja model pada kelas ini masih cukup baik, meskipun tidak sekuat pada kelas 0. Nilai macro average untuk precision, recall, dan f1-score masing-masing adalah 0.83, 0.82, dan 0.82, yang menunjukkan rata-rata kinerja model tanpa mempertimbangkan proporsi jumlah data tiap kelas. Sedangkan weighted average, yang mempertimbangkan proporsi data tiap kelas, menunjukkan nilai 0.85 untuk precision, recall, dan f1-score, sejalan dengan nilai akurasi keseluruhan. Secara keseluruhan, model menunjukkan performa yang stabil dan dapat diandalkan dalam melakukan klasifikasi dua kelas dapat dilihat pada Gambar 8

Classification Report (Precision, Recall, F1-Score, Support):				
	precision	recall	f1-score	support
0	0.88	0.90	0.89	162
1	0.77	0.73	0.75	75
accuracy			0.85	237
macro avg	0.83	0.82	0.82	237
weighted avg	0.85	0.85	0.85	237

Gambar 5 Hasil Klasifikasi

4. KESIMPULAN & SARAN

4.1 Kesimpulan

Dari hasil penelitian yang telah dilakukan dengan metode K-Nearest Neighbor (KNN) untuk klasifikasi penyakit Asma maka dapat disimpulkan bahwa metode K-Nearest Neighbor (KNN) dapat digunakan untuk klasifikasi penyakit. Tingkat akurasi yang didapatkan oleh metode K-Nearest Neighbor (KNN) dalam klasifikasi penyakit asma yaitu sebesar 85%. Hasil evaluasi lebih lanjut melalui classification report menunjukkan bahwa precision, recall, dan f1-score untuk kelas 1 masing-masing sebesar 0,88, 0,90, dan 0,89, sementara untuk kelas 0 masing-masing sebesar 0,77, 0,73, dan 0,75.

4.2 Saran

Untuk penelitian yang lebih baik di masa yang akan datang, peneliti menyarankan agar penelitian selanjutnya menggunakan dataset yang lebih besar dan lebih beragam guna meningkatkan kemampuan generalisasi model. Selain itu, pada penelitian berikutnya disarankan untuk menentukan nilai K yang optimal melalui teknik validasi silang (cross-validation) agar hasil klasifikasi yang diperoleh menjadi lebih akurat dan andal.

REFERENSI

- [1] A. R. Angga, "Implementasi Mechine Learning Pada Hybrid Intelligence Sistem Menggunakan Metode," vol. 4307, no. 1, pp. 685–690, 2025.
- [2] M. A. Kiram, E. Darnila, and I. Sahputra, "MACHINE LEARNING KLASIFIKASI PENYAKIT JIWA MENGGUNAKAN METODE K-NEAREST NEIGHBOR BERBASIS WEB," vol. 9, pp. 2445–2456, 2025.
- [3] H. Susanto, A. Setyanto, and A. H. Muhammad, "Analisis Sentimen Berita terhadap Bitcoin dengan Metode Klasifikasi K- Nearest Neighbor," vol. 4, no. 2, 2024.
- [4] S. A. Sabita and Y. Yahfizham, "Penerapan algoritma klasifikasi nearest neighbor dalam mendeteksi penyakit diabetes," *J. Bintang Pendidik. dan Bhs.*, vol. 2, no. 1, pp. 149–158, 2024, [Online]. Available: <https://doi.org/10.59024/bhinneka.v2i1.645>
- [5] S. Mansouri, S. Boulares, and S. Chabchoub, "Machine Learning for Early Diabetes Detection and Diagnosis," *J. Wirel. Mob. Networks, Ubiquitous Comput. Dependable Appl.*, vol. 15, no. 1, pp. 216–230, 2024, doi: 10.58346/JOWUA.2024.II.015.
- [6] A. P. Silalahi and H. G. Simanullang, "Supervised Learning Metode K-Nearest Neighbor Untuk Prediksi Diabetes Pada Wanita," *METHOMIKA J. Manaj. Inform. dan Komputerisasi Akunt.*, vol. 7, no. 1, pp. 144–149, 2023, doi: 10.46880/jmika.vol7no1.pp144-149.
- [7] F. M. N. Akbar, "Metode KNN (K-Nearest Neighbor) untuk Menentukan Kualitas Air," *J. Tekno Kompak*, vol. 18, no. 1, p. 28, 2024, doi: 10.33365/jtk.v18i1.3241.
- [8] S. Qoris, F. Yasin, and A. W. Widodo, "Klasifikasi Tingkat Obesitas Berdasarkan Pola Hidup dan Kebiasaan Konsumsi Makanan Menggunakan Metode K-Nearest Neighbor," vol. 9, no. 3, pp. 1–6, 2025.
- [9] G. H. Martono and N. Sulistianingsih, "Perbandingan Matriks Jarak Pada Algoritma K-NN Untuk Prediksi Penyakit Diabetes," *JoMI J. Millenn. Informatics*, vol. 2, no. 1, pp. 1–6, 2024.
- [10] H. N. Zuhdi and B. Prasetyo, "Penerapan Algoritma K-Nearest Neighbors dalam Evaluasi Kelayakan Kredit: Studi Kasus pada Bank ABC," *IJRSE Indones. J. Inform. Res. Softw. Eng.*, vol. 4, no. 1, pp. 40–46, 2024.