

Klasifikasi Tingkat Keparahan Pasien Penyakit Diabetes Mellitus di Puskesmas Pembantu Kota Binjai menggunakan Algoritma K-Nearest Neighbor (KNN)

Severity Classification of Diabetes Mellitus Patients at the Puskesmas Pembantu of Binjai City using the K-Nearest Neighbor (KNN) Algorithm

¹Nadya Putri Dwinta*, ²Asrianda, ³Rini Meiyanti

^{1,2,3}Universitas Malikussaleh

^{1,2,3}Jl. Batam, Bukit Indah -Lhokseumawe, Aceh, Indonesia

*e-mail: nadya.210170196@mhs.unimal.ac.id, asrianda@unimal.ac.id, rinimeiyanti@unimal.ac.id

(received: 4 August 2026, revised: 17 August 2026, accepted: 18 August 2026)

Abstrak

Diabetes Mellitus (DM) merupakan salah satu penyakit kronis yang jumlah penderitanya terus meningkat, termasuk di Kota Binjai. Penentuan tingkat keparahan pasien DM secara manual memerlukan ketelitian tinggi karena melibatkan banyak indikator medis, sehingga diperlukan metode klasifikasi yang mampu membantu tenaga kesehatan bekerja lebih cepat dan akurat. Penelitian ini bertujuan merancang sistem klasifikasi tingkat keparahan pasien DM tipe 1 dan tipe 2 di Puskesmas Pembantu Kota Binjai menggunakan algoritma K-Nearest Neighbor (KNN) serta mengukur tingkat akurasi. Data yang digunakan berjumlah 594 rekam medis pasien periode 2024–2025 dengan tiga belas atribut, yaitu usia, tinggi badan, berat badan, indeks massa tubuh, lingkar pinggang, tensi sistolik, tensi diastolik, kadar gula darah, serta atribut hasil encoding. Data melalui tahap praproses berupa transformasi atribut kategorikal dan normalisasi Z-Score sebelum dihitung kedekatannya menggunakan Manhattan Distance. Pengujian dilakukan pada nilai $k=1$ hingga $k=15$ untuk memperoleh nilai k optimum. Hasil pengujian menunjukkan bahwa nilai $k=1$ memberikan akurasi tertinggi sebesar 98,86%, dengan nilai precision 97,71%, recall 100%, dan F1-Score 0,99. Algoritma KNN selanjutnya diimplementasikan pada sistem berbasis website yang memungkinkan tenaga kesehatan menginputkan data pasien dan memperoleh hasil klasifikasi secara langsung. Hasil penelitian ini menunjukkan bahwa metode KNN dengan Manhattan Distance efektif digunakan dalam mengklasifikasikan tingkat keparahan pasien DM dan dapat menjadi alat bantu pendukung keputusan bagi tenaga kesehatan Puskesmas Pembantu Kota Binjai.

Kata kunci: klasifikasi, data mining, diabetes mellitus, K-Nearest Neighbor, manhattan distance

Abstract

Diabetes mellitus (DM) is a chronic disease with an increasing prevalence, including in Binjai City. Manually determining the severity level of DM patients requires a high degree of accuracy and careful assessment because it involves multiple medical indicators. Therefore, a classification method is needed to assist healthcare professionals in making faster and more accurate assessments. This study aims to develop a classification system for determining the severity of Type 1 and Type 2 diabetes mellitus in patients at Puskesmas Pembantu in Binjai City using the K-Nearest Neighbors (KNN) algorithm and to evaluate its classification performance. The dataset consisted of 594 patient medical records collected from 2024–2025, comprising 13 attributes, including age, height, weight, body mass index, waist circumference, systolic blood pressure, diastolic blood pressure, blood glucose level, and encoded attributes. The data preprocessing stage involved transforming categorical attributes and applying Z-score normalization before calculating similarity using the Manhattan distance metric. The model was evaluated using values of k ranging from 1 to 15 to determine the optimal k . The results show that $k = 1$ achieved the highest accuracy of 98.86%, with a precision of 97.71%, recall of 100%, and an F1-score of 0.99. The KNN algorithm was subsequently implemented in a web-based system that enables healthcare professionals to enter patient data and obtain classification results directly. These findings demonstrate that the KNN algorithm using the

<http://sistemasi.ftik.unisi.ac.id>

Manhattan distance metric is effective for classifying DM severity levels and can serve as a decision-support tool for healthcare professionals at Puskesmas Pembantu in Binjai City.

Keywords: *classification, data mining, diabetes mellitus, K-Nearest Neighbor, manhattan distance.*

1 Pendahuluan

Diabetes Mellitus (DM) merupakan salah satu penyakit kronis paling serius di era modern karena dapat menimbulkan komplikasi, kecacatan, biaya pengobatan yang tinggi, hingga penurunan harapan hidup penderitanya. Data IDF Diabetes Atlas mencatat prevalensi diabetes secara global pada kelompok usia 20–79 tahun mencapai 10,5% (sekitar 536,6 juta jiwa) pada tahun 2021 dan diproyeksikan meningkat menjadi 12,2% (sekitar 783,2 juta jiwa) pada tahun 2045 [1]. Pada tingkat nasional, prevalensi diabetes di Indonesia turut tergolong tinggi dengan angka mencapai 10,6% dari total populasi [9], sementara pada tingkat lokal, Badan Pusat Statistik Kota Binjai mencatat sebanyak 8.717 penderita DM pada tahun 2023 [2], sehingga penanganan penyakit ini menjadi perhatian penting bagi fasilitas pelayanan kesehatan tingkat pertama, khususnya di Provinsi Sumatera Utara.

DM muncul ketika pankreas tidak mampu memproduksi insulin dalam jumlah memadai atau ketika tubuh tidak dapat memanfaatkan insulin secara efisien, sehingga terjadi hiperglikemia yang dapat merusak sistem saraf dan pembuluh darah apabila tidak dikendalikan [3]. Kondisi ini juga berdampak pada kualitas hidup penderita, baik dari sisi kesehatan mental, kemampuan fisik, maupun peran sosial, sehingga kualitas hidup menjadi salah satu indikator penting dalam menilai keberhasilan intervensi pelayanan kesehatan [4]. Berdasarkan karakteristik genetik dan patofisiologinya, DM dibedakan menjadi beberapa tipe, di antaranya DM tipe 1 yang disebabkan kerusakan autoimun sel beta pankreas, DM tipe 2 yang disebabkan resistensi insulin dan penurunan fungsi sel beta tanpa proses autoimun, serta diabetes gestasional yang muncul selama masa kehamilan [5].

Penentuan tingkat keparahan pasien DM di fasilitas kesehatan tingkat pertama seperti Puskesmas Pembantu masih dilakukan secara manual dengan mempertimbangkan berbagai indikator medis, sehingga memerlukan waktu dan ketelitian yang tidak sedikit. Data mining dapat dimanfaatkan untuk membantu proses ini melalui teknik klasifikasi, yaitu proses menemukan model yang mampu membedakan kelas data untuk memprediksi kategori suatu objek yang labelnya belum diketahui. Salah satu algoritma klasifikasi yang banyak digunakan karena kesederhanaan dan efektivitasnya adalah K-Nearest Neighbor (KNN), yang telah terbukti mampu memberikan hasil klasifikasi yang baik pada berbagai studi kasus penyakit, termasuk klasifikasi penderita diabetes melitus [6] dan klasifikasi tingkat keparahan penyakit diare pada warga desa [7]. Penelitian lain juga menunjukkan bahwa perhitungan jarak Manhattan pada algoritma KNN dapat digunakan secara efektif untuk keperluan klasifikasi data sosial berskala besar [8]. Meskipun demikian, penelitian klasifikasi penderita diabetes melitus yang telah ada [6] umumnya belum menggunakan data rekam medis riil dari fasilitas kesehatan tingkat pertama seperti Puskesmas Pembantu, serta belum melibatkan atribut spesifik seperti lingkaran pinggang dan hasil encoding kategori sebagaimana digunakan pada penelitian ini. Pemilihan Puskesmas Pembantu Kota Binjai sebagai lokasi penelitian didasarkan pada tingginya jumlah kasus DM di kota tersebut serta belum tersedianya sistem klasifikasi otomatis yang dapat membantu tenaga kesehatan pada fasilitas tersebut dalam menentukan tingkat keparahan pasien secara cepat.

Berdasarkan permasalahan dan temuan penelitian terdahulu tersebut, penelitian ini menerapkan algoritma K-Nearest Neighbor dengan perhitungan jarak Manhattan Distance untuk mengklasifikasikan tingkat keparahan pasien penyakit Diabetes Mellitus tipe 1 dan tipe 2 di Puskesmas Pembantu Kota Binjai berdasarkan data rekam medis pasien periode 2024–2025. Penelitian ini bertujuan untuk (1) merancang sistem klasifikasi tingkat keparahan pasien penyakit DM menggunakan metode KNN, (2) mengukur tingkat akurasi metode tersebut, dan (3) mengimplementasikan sistem klasifikasi tersebut ke dalam sebuah aplikasi berbasis website yang dapat digunakan secara langsung oleh tenaga kesehatan Puskesmas Pembantu Kota Binjai dalam menginputkan data pasien dan memperoleh hasil klasifikasi. Hasil penelitian ini diharapkan dapat menjadi alat bantu pendukung keputusan bagi tenaga kesehatan dalam mengidentifikasi tingkat keparahan penyakit DM secara lebih cepat dan akurat, serta membantu manajemen Puskesmas Pembantu Kota Binjai dalam perencanaan dan pengelolaan sumber daya kesehatan.

2 Tinjauan Literatur

2.1 Data Mining dan Klasifikasi

Data mining adalah proses yang mengombinasikan matematika, statistika, kecerdasan buatan, dan pembelajaran mesin untuk mengekstraksi serta mengidentifikasi informasi penting dari kumpulan data berukuran besar [10]. Salah satu peran utama data mining adalah klasifikasi (classification), yaitu proses menemukan model atau pola yang dapat menggambarkan atau membedakan kategori data untuk memprediksi kelas dari suatu objek ketika labelnya belum diketahui. Klasifikasi banyak diterapkan pada berbagai studi kasus, mulai dari klasifikasi citra hewan [11], klasifikasi gangguan kejiwaan berbasis web [12], klasifikasi opini masyarakat [13], klasifikasi buah [14], hingga analisis sentimen ulasan pengguna aplikasi [15].

2.2 K-Nearest Neighbor (KNN)

K-Nearest Neighbor (KNN) merupakan metode klasifikasi yang mengelompokkan data baru berdasarkan kemiripannya dengan sejumlah k data pelatihan yang letaknya paling dekat. Proses kerja algoritma KNN secara umum terdiri atas empat tahap, yaitu (1) menentukan nilai parameter k , (2) menghitung jarak antara data uji dengan seluruh data latih, (3) mengurutkan hasil perhitungan jarak dari yang terkecil, dan (4) menetapkan kelas hasil klasifikasi berdasarkan kelas mayoritas dari k tetangga terdekat [16]. Algoritma ini tergolong lazy learner karena tidak memerlukan proses pelatihan model secara eksplisit, sehingga seluruh perhitungan dilakukan langsung pada saat proses prediksi berlangsung [17].

Sebelum proses klasifikasi dilakukan, data terlebih dahulu melalui tahap normalisasi agar seluruh atribut memiliki rentang nilai yang sebanding sehingga tidak ada atribut yang mendominasi perhitungan jarak. Salah satu metode normalisasi yang umum digunakan adalah Z-Score Normalization, yang dirumuskan pada Persamaan (1).

$$Z = (X - \mu) / \sigma \quad (1)$$

Keterangan: Z adalah nilai standar (Z-Score), X adalah nilai data yang akan dihitung, μ adalah nilai rata-rata (mean) dari seluruh data dalam distribusi, dan σ adalah standar deviasi atribut yang mengukur sebaran data.

Perhitungan jarak antara data uji dan data latih dalam algoritma KNN dapat menggunakan beberapa metode, di antaranya Euclidean Distance, Manhattan Distance, Minkowski Distance, dan Hamming Distance. Pemilihan metode jarak bergantung pada karakteristik data dan kompleksitas permasalahan yang dihadapi. Penelitian ini menggunakan Manhattan Distance karena perhitungannya yang sederhana dan telah terbukti efektif digunakan pada klasifikasi data sosial berskala besar [8]. Rumus Manhattan Distance ditunjukkan pada Persamaan (2).

$$d(x, y) = \sum_{i=1}^n |x_i - y_i| \quad (2)$$

Keterangan: $d(x,y)$ adalah jarak antara data x dan y , x_i adalah nilai atribut ke- i pada data uji, y_i adalah nilai atribut ke- i pada data latih, dan n adalah jumlah atribut yang digunakan. Semakin kecil nilai jarak yang diperoleh, semakin tinggi tingkat kemiripan antara data uji dan data latih.

2.3 Confusion Matrix

Confusion matrix digunakan untuk mengevaluasi kinerja model klasifikasi dengan membandingkan kelas hasil prediksi dan kelas aktual data. Beberapa metrik evaluasi yang diturunkan dari confusion matrix meliputi akurasi, precision, recall, dan F1-Score, sebagaimana ditunjukkan pada Persamaan (3) hingga Persamaan (6) [18].

$$\text{Akurasi} = (TP + TN) / (TP + TN + FP + FN) \quad (3)$$

$$\text{Precision} = TP / (TP + FP) \quad (4)$$

$$\text{Recall} = TP / (TP + FN) \quad (5)$$

$$\text{F1-Score} = 2 \times (\text{Precision} \times \text{Recall}) / (\text{Precision} + \text{Recall}) \quad (6)$$

Nilai True Positive (TP) dan True Negative (TN) menunjukkan jumlah data yang diklasifikasikan dengan benar, sedangkan False Positive (FP) dan False Negative (FN) menunjukkan jumlah data yang

diklasifikasikan keliru. Akurasi mengukur proporsi keseluruhan prediksi yang benar, precision mengukur ketepatan prediksi kelas positif, recall mengukur kemampuan model dalam menemukan seluruh data kelas positif, dan F1-Score merupakan rata-rata harmonik antara precision dan recall.

2.4 Penelitian Terdahulu

Beberapa penelitian terdahulu telah menerapkan algoritma KNN pada berbagai studi kasus. Penelitian klasifikasi jenis kucing menggunakan kombinasi Principal Component Analysis (PCA) dan KNN memperoleh akurasi 80% [11]. Penelitian klasifikasi gangguan jiwa berbasis web menggunakan KNN dengan $k=10$ memperoleh akurasi hingga 100% pada data uji [12]. Penelitian klasifikasi opini masyarakat terhadap naturalisasi pemain sepak bola dengan KNN dan SMOTE memperoleh akurasi 77,06% [13]. Penelitian klasifikasi buah menggunakan KNN dan PCA memperoleh akurasi 93% [14], sedangkan penelitian analisis sentimen ulasan pengguna e-commerce menggunakan KNN memperoleh akurasi 82% [15]. Penelitian klasifikasi data bunga iris menggunakan KNN memperoleh akurasi rata-rata antara 68% hingga 100% bergantung pada nilai k yang digunakan [16], dan penelitian prediksi harga bahan pangan menggunakan KNN Regressor memperoleh nilai MAE dan RMSE terbaik pada $k=2$ [19]. Penerapan KNN pada bidang kesehatan juga terbukti efektif pada klasifikasi penderita diabetes melitus [6] serta pada penentuan tingkat keparahan penyakit diare [7].

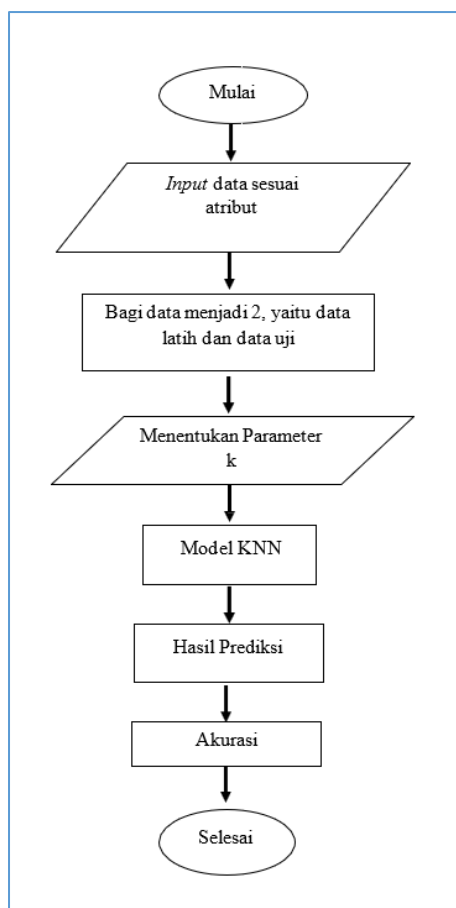
Berdasarkan tinjauan tersebut, algoritma KNN terbukti fleksibel diterapkan pada berbagai domain data, termasuk data kesehatan. Namun demikian, sebagian besar penelitian terdahulu belum secara khusus menerapkan KNN dengan perhitungan Manhattan Distance untuk mengklasifikasikan tingkat keparahan pasien Diabetes Mellitus berdasarkan data rekam medis riil pada fasilitas kesehatan tingkat pertama, sekaligus mengimplementasikannya ke dalam sistem berbasis website yang dapat langsung digunakan oleh tenaga kesehatan. Kebaruan penelitian ini terletak pada penerapan tersebut, khususnya pada studi kasus Puskesmas Pembantu Kota Binjai.

3 Metode Penelitian

Penelitian ini dilaksanakan di Puskesmas Pembantu Kota Binjai, tempat penulis melaksanakan studi lapangan, dengan pelaksanaan penelitian dimulai pada bulan Juni 2025 hingga selesai. Penelitian dilakukan secara luring dengan pengambilan data langsung dari sistem informasi manajemen Puskesmas Pembantu Kota Binjai.

Data yang digunakan merupakan data sekunder berupa rekam medis pasien penyakit Diabetes Mellitus tipe 1 dan tipe 2 periode Januari 2024 hingga Desember 2025, dengan total 594 baris data. Variabel yang digunakan mencakup usia (X_1), tinggi badan (X_2), berat badan (X_3), indeks massa tubuh/IMT (X_4), lingkar pinggang (X_5), tensi sistolik (X_6), tensi diastolik (X_7), kadar gula darah (X_8), serta atribut hasil encoding untuk jenis kelamin, IMT, tensi, lingkar pinggang, dan kadar gula (X_9 – X_{13}). Data disimpan dalam format digital (Microsoft Excel) untuk mempermudah proses analisis lebih lanjut menggunakan bahasa pemrograman Python. Dari total 594 data tersebut, sebanyak 331 data (55,7%) digunakan sebagai data latih dan 263 data (44,3%) digunakan sebagai data uji. Pembagian data dilakukan secara acak (random split) dengan tetap menjaga proporsi kelas Diabetes Mellitus Tipe 1 dan Tipe 2 pada kedua subset data guna menjaga validitas proses pelatihan dan pengujian model.

Tahapan penelitian meliputi input data pasien sesuai atribut yang telah ditentukan, pembagian data menjadi data latih dan data uji, penentuan parameter k , pembentukan model KNN, penghasilan prediksi klasifikasi, hingga pengukuran akurasi. Skema sistem klasifikasi yang diterapkan pada penelitian ini ditunjukkan pada Gambar 1.



Gambar 1 Skema sistem klasifikasi KNN

Berdasarkan Gambar 1, proses dimulai dengan admin memasukkan data pasien DM tipe 1 dan tipe 2 ke dalam sistem. Data tersebut kemudian dibagi menjadi data latih (training) dan data uji (testing), dilanjutkan dengan penentuan parameter k yang akan digunakan pada perhitungan KNN. Sistem selanjutnya menghitung jarak Manhattan Distance antara nilai atribut data uji dengan setiap atribut data latih, kemudian menentukan hasil klasifikasi berdasarkan kelas mayoritas dari k tetangga terdekat. Hasil klasifikasi ditampilkan dalam bentuk tabel dan grafik akurasi, sehingga admin dapat melihat tingkat keparahan penyakit pasien DM.

Sebelum data diproses menggunakan algoritma KNN, dilakukan tahap praproses berupa pembersihan data, transformasi atribut kategorikal menjadi numerik, dan normalisasi seluruh atribut numerik menggunakan metode Z-Score sebagaimana Persamaan (1). Transformasi atribut kategorikal dilakukan menggunakan metode label encoding, yaitu mengubah setiap kategori pada atribut IMT, tensi, lingkaran pinggang, kadar gula, dan jenis kelamin menjadi nilai numerik ordinal sesuai urutan tingkat keparahannya (misalnya kategori IMT diberi label 1 untuk kurus, 2 untuk normal, 3 untuk gemuk, dan 4 untuk obesitas), sehingga nilai tersebut dapat diikutsertakan dalam perhitungan jarak Manhattan Distance. Normalisasi ini bertujuan agar tidak ada atribut yang mendominasi perhitungan jarak akibat perbedaan satuan dan rentang nilai antaratribut. Setelah dinormalisasi, jarak antara data uji dan setiap data latih dihitung menggunakan Manhattan Distance sebagaimana Persamaan (2), kemudian hasil jarak diurutkan dari nilai terkecil untuk menentukan k tetangga terdekat. Kelas hasil klasifikasi ditentukan berdasarkan kelas mayoritas dari tetangga terdekat tersebut, yaitu kelas Diabetes Mellitus Tipe 1 atau Tipe 2. Pengujian nilai k dilakukan pada rentang $k=1$ hingga $k=15$ untuk memperoleh nilai k yang memberikan akurasi klasifikasi terbaik, yang kemudian dievaluasi menggunakan confusion matrix sebagaimana Persamaan (3) hingga Persamaan (6).

Perancangan dan implementasi sistem dilakukan menggunakan bahasa pemrograman Python dengan dukungan Visual Studio Code sebagai editor kode dan Microsoft Excel sebagai media penyimpanan data awal. Sistem dibangun berbasis website agar dapat diakses dengan mudah oleh

tenaga kesehatan Puskesmas Pembantu Kota Binjai dalam melakukan input data pasien dan memperoleh hasil klasifikasi tingkat keparahan penyakit Diabetes Mellitus secara langsung.

4 Hasil dan Pembahasan

Data yang digunakan dalam penelitian ini berjumlah 594 rekam medis pasien Diabetes Mellitus dari Puskesmas Pembantu Kota Binjai, yang terdiri atas pasien DM Tipe 1 dan Tipe 2. Contoh data kategori sebelum dilakukan praproses ditunjukkan pada Tabel 1.

Tabel 1 Contoh data kategori pasien diabetes mellitus

Usia	TB/BB	Tensi	L. Pinggang	IMT	Kadar Gula	Tipe
48	154/62	130/80	93	26,1	200	DM2
46	155/63	130/80	90	26,2	200	DM2
37	153/60	130/80	90	25,6	200	DM2
33	155/62	120/80	65	16,2	200	DM1
50	154/60	130/80	90	25,3	200	DM2

Sebelum dilakukan perhitungan jarak, seluruh atribut numerik dinormalisasi menggunakan metode Z-Score sesuai Persamaan (1). Nilai rata-rata (mean) dan standar deviasi setiap atribut yang diperoleh dari 594 data ditunjukkan pada Tabel 2.

Tabel 2 Parameter normalisasi z-score

Atribut	Mean (μ)	Standar Deviasi (σ)
Usia	51,6143	8,0041
Tinggi Badan	165,5148	101,5142
Berat Badan	57,5061	5,8238
IMT	23,8336	2,1302
Lingkar Pinggang	87,7017	23,9254
Tensi Sistolik	126,0638	9,4436
Tensi Diastolik	80,9043	4,3577
Kadar Gula	205,0673	14,9371

Sebagai contoh, hasil normalisasi Z-Score pada baris pertama data latih menghasilkan nilai atribut usia sebesar $-1,1147$, tinggi badan $-0,0928$, berat badan $-1,0931$, IMT $0,4286$, lingkar pinggang $-0,3032$, tensi sistolik $-0,6769$, tensi diastolik $-0,2294$, dan kadar gula $-0,0680$. Proses yang sama diterapkan pada seluruh atribut dan seluruh baris data hingga diperoleh data yang telah ternormalisasi secara menyeluruh.

4.1 Perhitungan Manhattan Distance

Setelah data dinormalisasi, tahap selanjutnya adalah menghitung jarak antara data uji dan data latih menggunakan Manhattan Distance sesuai Persamaan (2). Sebagai ilustrasi, perhitungan jarak antara data uji nomor urut 1 dan data latih nomor urut 1 menghasilkan rincian selisih Z-Score pada setiap atribut sebagaimana ditunjukkan pada Tabel 3.

Tabel 3 Hasil perhitungan jarak manhattan distance

Atribut	Selisih Z-Score
Usia	1,7344
Tinggi Badan	0,0239
Berat Badan	1,2043
IMT	1,1513
Lingkar Pinggang	0
Tensi Sistolik	0
Tensi Diastolik	0
Kadar Gula	0,4228
Jenis Kelamin (encoded)	2,0236
IMT (encoded)	0

Tensi (encoded)	0
Lingkar Pinggang (encoded)	2,1798
Kadar Gula (encoded)	0
Total Jarak Manhattan	8,7396

Nilai total jarak Manhattan yang diperoleh sebesar 8,7396 menunjukkan tingkat kedekatan antara data uji dan data latih tersebut; semakin kecil nilai jarak, semakin tinggi tingkat kemiripan kedua data. Perhitungan yang sama dilakukan terhadap seluruh data latih, kemudian hasil jarak diurutkan dari nilai terkecil hingga terbesar untuk menentukan tetangga terdekat sesuai nilai k yang digunakan.

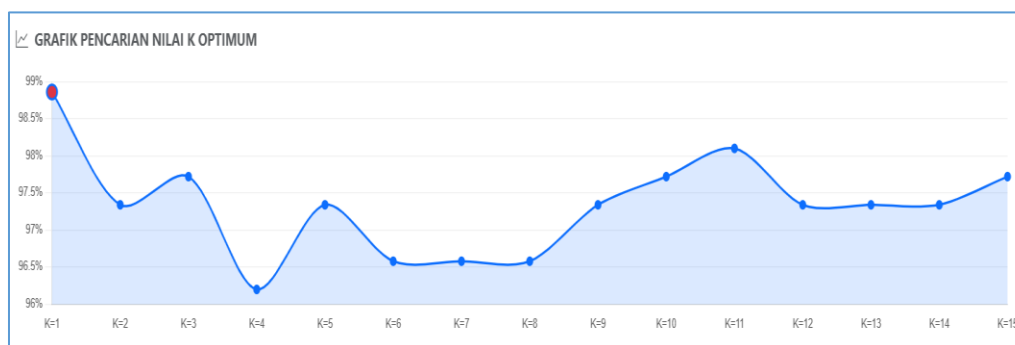
4.2 Hasil Klasifikasi dan Penentuan Nilai K Optimum

Berdasarkan hasil perhitungan Manhattan Distance, diperoleh peringkat data latih dengan jarak terdekat terhadap data uji nomor urut 1, sebagaimana ditunjukkan pada Tabel 4. Kelima belas tetangga terdekat pada data uji tersebut seluruhnya berlabel Diabetes Tipe 1, sehingga untuk nilai k=1 hingga k=15 hasil voting mayoritas tetap memprediksi kelas Tipe 1, sesuai dengan label aktualnya.

Tabel 4 Contoh hasil klasifikasi berdasarkan tetangga terdekat

No.	Data Latih	Jarak Manhattan	Peringkat	Label
1	435	0,0815	1	Tipe 1
2	200	0,4009	2	Tipe 1
3	351	0,4684	3	Tipe 1
4	36	0,4941	4	Tipe 1
5	328	0,4973	5	Tipe 1

Pengujian nilai k dilakukan pada rentang k=1 hingga k=15 untuk memperoleh nilai k yang memberikan akurasi klasifikasi tertinggi, sebagaimana ditunjukkan pada Gambar 2.



Gambar 2 Grafik pencarian nilai k optimum

Berdasarkan Gambar 2, akurasi tertinggi diperoleh pada k=1 dengan nilai 98,86%. Akurasi kemudian mengalami penurunan pada rentang k=2 hingga k=4, dengan nilai terendah pada k=4 sebesar 96,20%. Akurasi meningkat kembali pada rentang k=5 hingga k=11 dan mencapai nilai tertinggi kedua pada k=11 sebesar 98,10%. Setelah k=11, akurasi cenderung stabil pada kisaran 97,3%–97,7% hingga k=15. Berdasarkan hasil tersebut, nilai k=1 dipilih sebagai nilai k optimum karena memberikan akurasi tertinggi dibandingkan nilai k lainnya. Perlu dicatat bahwa penggunaan nilai k=1 memiliki potensi overfitting karena klasifikasi hanya mengacu pada satu tetangga terdekat, sehingga model menjadi sangat sensitif terhadap noise atau data pencilan (outlier) pada data latih dan berisiko kurang mampu melakukan generalisasi terhadap data baru. Meskipun demikian, selisih akurasi antara k=1 (98,86%) dan k=11 (98,10%) yang relatif kecil, disertai nilai precision, recall, dan F1-Score yang tetap tinggi pada rentang k tersebut, mengindikasikan bahwa model cukup stabil. Untuk mengurangi risiko overfitting, nilai k=11 dapat menjadi alternatif yang lebih konservatif, dan penelitian lanjutan disarankan untuk memvalidasi kestabilan model menggunakan teknik cross-validation.

4.3 Evaluasi Model Menggunakan Confusion Matrix

Evaluasi lebih lanjut terhadap model dengan $k=1$ dilakukan menggunakan confusion matrix yang membandingkan hasil prediksi terhadap 263 data uji. Hasil confusion matrix ditunjukkan pada Tabel 5.

Tabel 5 Confusion matrix hasil klasifikasi KNN manhattan distance

	Prediksi DM Tipe 1	Prediksi DM Tipe 2
Aktual DM Tipe 1	132	3
Aktual DM Tipe 2	0	128

Berdasarkan Tabel 5, dengan kelas Diabetes Tipe 2 sebagai kelas positif, diperoleh nilai True Positive (TP) sebanyak 128 data, True Negative (TN) sebanyak 132 data, False Positive (FP) sebanyak 3 data (data aktual Tipe 1 yang diprediksi sebagai Tipe 2), dan False Negative (FN) sebanyak 0 data (data aktual Tipe 2 yang diprediksi sebagai Tipe 1). Berdasarkan nilai tersebut, perhitungan metrik evaluasi menggunakan Persamaan (3) hingga Persamaan (6) menghasilkan akurasi sebesar $(128+132)/(128+132+3+0) \times 100\% = 98,86\%$, precision sebesar $128/(128+3) \times 100\% = 97,71\%$, dan recall sebesar $128/(128+0) \times 100\% = 100\%$. Secara rinci, hasil evaluasi per kelas ditunjukkan pada Tabel 6.

Tabel 6 Laporan evaluasi klasifikasi per kelas

Target Kelas	Precision	Recall	F1-Score
Diabetes Tipe 1	1,00	0,98	0,99
Diabetes Tipe 2	0,98	1,00	0,99

Hasil evaluasi menunjukkan bahwa metode KNN dengan Manhattan Distance memperoleh nilai akurasi sebesar 98,86%, serta nilai precision, recall, dan F1-Score yang berkisar antara 0,98 hingga 1,00. Hasil tersebut menunjukkan bahwa algoritma mampu memberikan hasil klasifikasi yang sangat baik terhadap data pasien Diabetes Mellitus yang digunakan dalam penelitian ini, sehingga layak dijadikan dasar bagi tenaga kesehatan dalam mengidentifikasi tingkat keparahan penyakit pasien secara lebih cepat dan akurat.

4.4 Implementasi Sistem

Algoritma KNN dengan Manhattan Distance yang telah diuji selanjutnya diimplementasikan ke dalam sistem berbasis website. Sistem tersebut terdiri atas beberapa halaman utama, yaitu halaman login sebagai akses awal yang membatasi pengguna sistem hanya kepada admin yang memiliki otoritas, halaman dashboard sebagai pusat navigasi yang menampilkan ringkasan data pasien serta informasi model KNN, halaman dataset untuk mengelola data rekam medis pasien dan melakukan pelatihan ulang model, halaman form prediksi untuk memasukkan data pasien baru yang akan diklasifikasikan, serta halaman hasil prediksi yang menampilkan tipe Diabetes Mellitus hasil klasifikasi beserta riwayat prediksi pasien. Melalui sistem ini, tenaga kesehatan Puskesmas Pembantu Kota Binjai dapat memperoleh hasil klasifikasi tingkat keparahan pasien Diabetes Mellitus secara cepat sebagai alat bantu pendukung keputusan, dengan tetap mempertimbangkan bahwa keputusan medis akhir tetap berada pada tenaga kesehatan yang berwenang.

5 Kesimpulan

Berdasarkan hasil penelitian mengenai klasifikasi tingkat keparahan pasien penyakit Diabetes Mellitus di Puskesmas Pembantu Kota Binjai menggunakan algoritma K-Nearest Neighbor, dapat disimpulkan bahwa algoritma KNN berhasil diterapkan dalam mengklasifikasikan 594 data rekam medis pasien ke dalam kategori Diabetes Mellitus Tipe 1 dan Tipe 2 melalui tahapan transformasi data kategorikal, normalisasi Z-Score, dan perhitungan jarak menggunakan Manhattan Distance. Hasil pengujian nilai k pada rentang $k=1$ hingga $k=15$ menunjukkan bahwa nilai $k=1$ memberikan akurasi klasifikasi tertinggi sebesar 98,86%, dengan nilai precision 97,71%, recall 100%, dan F1-Score 0,99,

yang menunjukkan bahwa model memiliki kemampuan sangat baik dalam mengenali pola data berdasarkan atribut usia, indeks massa tubuh, lingkaran pinggang, tensi sistolik, tensi diastolik, kadar gula darah, dan jenis kelamin pasien. Algoritma KNN dengan Manhattan Distance tersebut selanjutnya berhasil diimplementasikan ke dalam sistem berbasis website yang menyediakan halaman login, dashboard, pengelolaan dataset, form prediksi pasien, dan halaman hasil prediksi, sehingga tenaga kesehatan dapat memasukkan data kondisi pasien dan memperoleh hasil klasifikasi Diabetes Mellitus secara lebih terstruktur dan efisien.

Hasil klasifikasi dari sistem ini diharapkan dapat digunakan sebagai alat bantu pendukung keputusan bagi tenaga kesehatan Puskesmas Pembantu Kota Binjai, dan tidak dijadikan sebagai satu-satunya dasar penentuan diagnosis medis, karena penentuan kondisi dan penanganan pasien tetap memerlukan pemeriksaan serta pertimbangan dari tenaga kesehatan yang berwenang. Penelitian selanjutnya disarankan untuk menggunakan jumlah dan variasi data yang lebih besar, membandingkan algoritma KNN dengan metode klasifikasi lain seperti Random Forest atau Support Vector Machine, menguji metode perhitungan jarak lain sebagai pembanding terhadap Manhattan Distance, serta menerapkan teknik cross-validation untuk memvalidasi kestabilan model dan mengurangi risiko overfitting guna memperoleh kombinasi parameter yang lebih optimal.

Referensi

- [1] H. Sun et al., "Erratum to 'IDF Diabetes Atlas: Global, Regional and Country-Level Diabetes Prevalence Estimates for 2021 and Projections for 2045' [Diabetes Res. Clin. Pract. 183 (2022) 109119]," *Diabetes Research and Clinical Practice*, vol. 204, 2023.
- [2] Badan Pusat Statistik Kota Binjai, "Jumlah Kasus 10 Penyakit Terbanyak di Kota Binjai," [binjaikota.bps.go.id](https://binjaikota.bps.go.id/id/statistic-table/1/NzY1IzE=/jumlah-kasus-10-penyakit-terbanyak-di-kota-binjai-2024.html), 2024. [Online]. Available: <https://binjaikota.bps.go.id/id/statistic-table/1/NzY1IzE=/jumlah-kasus-10-penyakit-terbanyak-di-kota-binjai-2024.html>.
- [3] A. Aminuddin, Y. Sima, et al., "Edukasi Kesehatan tentang Penyakit Diabetes Mellitus bagi Masyarakat," *Abdimas Polsaka: Jurnal Pengabdian Kepada Masyarakat*, Vol. 2, No. 1, 2023.
- [4] Nina, et al., "Determinan Risiko dan Pencegahan terhadap Kejadian Penyakit Diabetes Melitus Tipe 2 pada Usia Produktif di Wilayah DKI Jakarta," *Journal of Public Health Education*, Vol. 2, No. 4, 2023.
- [5] American Diabetes Association Professional Practice Committee, "Diagnosis and Classification of Diabetes: Standards of Care in Diabetes-2024," *Diabetes Care*, Vol. 47, Supplement_1, pp. S20-S42, 2024.
- [6] A. M. Argina, "Penerapan Metode Klasifikasi K-Nearest Neighbor pada Dataset Penderita Penyakit Diabetes," *Indonesian Journal of Data and Science*, Vol. 1, No. 2, 2020.
- [7] H. A. D. Fasnuari, et al., "Penerapan Algoritma K-Nearest Neighbor (KNN) untuk Klasifikasi Penyakit Diabetes Melitus (Studi Kasus: Warga Desa Jati Tengah)," *Antivirus: Jurnal Ilmiah Teknik Informatika*, Vol. 16, No. 2, 2022.
- [8] A. D. J. Saputro, A. Darmawan, and B. N. Sari, "Klasifikasi Persentase Kemiskinan di Jawa Barat menggunakan Data Mining Algoritma K-Nearest Neighbor (KNN)," *JATI (Jurnal Mahasiswa Teknik Informatika)*, Vol. 7, No. 1, 2023.
- [9] F. Pramonodjati, "Skrining Diabetes Melitus melalui Pemeriksaan Gula Darah Masyarakat Kota Surakarta pada Acara Car Free Day," *JIPMASLAB (Jurnal Inovasi dan Pemberdayaan Masyarakat Laboratorium Kesehatan)*, Vol. 1, No. 1, 2025.
- [10] M. H. Elison, R. Asrianto, and Aryanto, "Prediksi Penjualan Papan Bunga menggunakan Metode Double Exponential Smoothing," *JURSISTEKNI: Jurnal Riset Sistem Informasi dan Teknologi Informasi*, Vol. 2, No. 3, 2020.
- [11] M. Dadang Iskandar and S. Nurhaliza, "Klasifikasi Jenis Kucing menggunakan Algoritma K-Nearest Neighbor," *JTIK: Jurnal Teknologi Informasi dan Komunikasi*, Vol. 9, No. 3, 2025.
- [12] M. A. Kiram, E. Darnila, and I. Sahputra, "Machine Learning Klasifikasi Penyakit Jiwa menggunakan Metode K-Nearest Neighbor berbasis Web," *Jurnal Ners*, Vol. 9, No. 2, 2025.
- [13] M. Graciela, Rikky, and H. Irsyad, "Klasifikasi Opini Masyarakat terhadap Naturalisasi Pemain Sepak Bola menggunakan KNN dan SMOTE," in *Proc. AICOMS*, Vol. 3, No. 1, 2024.

- [14] I. Rosa and A. Ramadhanu, "Penerapan Algoritma *K-Nearest Neighbor (KNN)* dan PCA untuk Klasifikasi Apel Hijau, Apel Fuji dan Jeruk," JATI (Jurnal Mahasiswa Teknik Informatika), Vol. 9, No. 2, 2025.
- [15] I. H. Kusuma and N. Cahyono, "Analisis Sentimen Masyarakat terhadap Penggunaan *E-Commerce* menggunakan Algoritma *K-Nearest Neighbor*," JPIT: Jurnal Pengembangan IT, Vol. 8, No. 3, 2023.
- [16] P. Putra, A. M. H. Pardede, and S. Syahputra, "Analisis Metode *K-Nearest Neighbor (KNN)* dalam Klasifikasi Data Iris Bunga," JTIK: Jurnal Teknik Informatika Kaputama, Vol. 6, No. 1, 2022.
- [17] Nunsina, Tulus, and Z. Situmorang, "*Analysis Optimization K-Nearest Neighbor Algorithm with Certain Factor in Determining Student Career*," in Proc. MECnIT: Mechanical, Electronics, Computer and Industrial Technology, 2020.
- [18] M. Rahmadiyah and P. Suparman, "Penerapan Metode *K-Nearest Neighbor* untuk Sistem Penentuan Peminjaman Modal Nasabah Bank Syariah Indonesia Cabang Cikarang berbasis *Website*," JIK: Jurnal Informasi dan Komputer, Vol. 10, No. 2, 2022.
- [19] R. Rahmadini, et al., "Penerapan Data Mining untuk memprediksi Harga Bahan Pangan di Indonesia menggunakan Algoritma *K-Nearest Neighbor*," JMAS: Jurnal Mahasiswa Akuntansi Samudra, Vol. 6, No. 4, 2023.
- [20] Undang-Undang Republik Indonesia Nomor 44 Tahun 2009 Tentang Rumah Sakit, 2009.