 ISSN:2302-8149
e-ISSN:2540-9719

Klasifikasi Diabetes Melitus menggunakan Algoritma *K-Nearest Neighbor* (*KNN*) (Studi Kasus: Data Pasien RSUD Salatiga)

Classification of Diabetes Mellitus using the K-Nearest Neighbor (KNN) Algorithm: A Case Study of Patient Data at Salatiga Regional Hospital

[1]Gwen Theresia Grandis Aritonang*, [2]Magdalena A. Ineke Pakereng
[1,2]Jurusan Teknik Informatika, Fakultas Teknologi Informasi, Universitas Kristen Satya Wacana
[1,2]Jl. Diponegoro 52-60, Salatiga 50711, Indonesia
*e-mail: 672022250@student.uksw.edu

(***received*:** 11 April 2026, ***revised*:** 22 May 2026, ***accepted*:** 23 May 2026)

Abstrak

Salah satu penyakit metabolik yang paling umum di Indonesia, termasuk di RSUD Salatiga, adalah diabetes melitus. Diagnosis dini penyakit ini sangat penting untuk mencegah perkembangan komplikasi yang lebih serius, tetapi hal ini sering terkendala karena gejala awalnya sulit dikenali. Studi ini mengimplementasikan algoritma K-Nearest Neighbor (KNN) sebagai teknik klasifikasi risiko diabetes melitus berbasis data klinis pasien RSUD Salatiga. Dataset terdiri dari penelitian ini mencakup variabel-variabel, seperti jenis kelamin, usia, riwayat hipertensi, kadar glukosa, kadar HbA1c, status merokok, riwayat penyakit hati, dan indeks massa tubuh (BMI). Adapun tahapan penelitian terdiri dari pengumpulan data, pre-processing (pembersihan, normalisasi, dan *encoding* variabel), seleksi fitur, serta optimasi parameter k. Evaluasi dilakukan menggunakan *confuison matrix* dengan metrik akurasi, presisi, *recall, dan F1-score*. Hasil penelitian menunjukkan bahwa algoritma KNN mampu memberikan akurasi tertinggi sebesar 92,08% dengan penerapan seleksi fitur dan nilai k = 6. Peningkatan akurasi ini menunjukkan bahwa optimasi parameter k dan seleksi fitur memiliki dampak yang signifikan terhadap kinerja model. Oleh karena itu, KNN dapat digunakan sebagai alat prediksi dini yang efektif untuk membantu tenaga medis memeriksa pasien diabetes melitus yang memungkinkan intervensi lebih cepat dan tepat.

Kata kunci: klasifikasi, diabetes melitus, prediksi dini, seleksi fitur, k-nearest neighbor (KNN)

Abstract

One of the most common metabolic diseases in Indonesia, including at RSUD Salatiga, is Diabetes Mellitus. Early diagnosis of this disease is crucial to prevent the development of more severe complications; however, this process is often challenging because the initial symptoms are difficult to recognize. This study implements the K-Nearest Neighbor (KNN) algorithm as a classification technique for predicting diabetes mellitus risk based on clinical data from patients at RSUD Salatiga. The dataset used in this research included variables such as gender, age, hypertension history, glucose level, HbA1c level, smoking status, liver disease history, and Body Mass Index (BMI). The research stages consisted of data collection, preprocessing (data cleaning, normalization, and variable encoding), feature selection, and optimization of the k parameter. Model evaluation was conducted using a confusion matrix with performance metrics including accuracy, precision, recall, and F1-score. The results indicate that the KNN algorithm achieved the highest accuracy of 92.08% when feature selection was applied with k = 6. This improvement demonstrates that both k-parameter optimization and feature selection significantly affect model performance. Therefore, the KNN algorithm can serve as an effective early prediction tool to assist medical personnel in identifying diabetes mellitus patients, enabling faster and more accurate interventions.

Keywords: *classification, diabetes mellitus, early prediction, feature selection, k-nearest neighbor (KNN)*

 ISSN:2302-8149
e-ISSN:2540-9719

1 Pendahuluan

Diabetes melitus adalah salah satu penyakit yang sering dijumpai di kalangan orang dewasa dan lanjut usia. Berdasarkan data dari *World Health Organization* (WHO), sekitar 70% kematian di dunia disebabkan oleh penyakit tidak menular, dan diabetes menjadi salah satu penyebab kematian yang cukup tinggi di Indonesia [1]. Lalu, menurut daftar kasus di RSUD Salatiga, penderita diabetes melitus biasanya berusia antara 30 hingga 80 tahun. Peningkatan kasus ini sebagian besar ditimbulkan oleh perilaku hidup yang tidak sehat, contohnya melalui konsumsi makanan berlemak dan tidak berolahraga. Namun, penderita sering kali tidak menyadari kondisi tersebut karena gejala awal yang relatif ringan dan menyebabkan diagnosis serta penanganan yang tertunda. Akibatnya, diabetes melitus dapat berkembang menjadi kondisi yang lebih serius dan menyebabkan banyak komplikasi jika tidak ditangani dengan benar [2].

Diabetes melitus merupakan gangguan proses metabolisme yang bercirikan dengan kondisi hiperglikemia, yaitu meningkatnya konsentrasi glukosa dalam darah yang disebabkan oleh gangggguan produksi insulin dan penurunan efektivitas kerja insulin [2]. Perubahan pola hidup masyarakat modern, baik pada kalangan remaja maupun dewasa, turut berkontribusi terhadap peningkatan kasus diabetes. Konsumsi makanan cepat saji dan produk pangan tinggi gula yang berlebihan menjadi salah satu faktor risiko utama yang memicu perkembangan penyakit ini [3]. Solusi untuk masalah ini adalah dengan cara mendeteksi dini berbasis teknologi [4].

Penelitian ini bertujuan untuk mendeteksi potensi seseorang yang mengidap diabetes melitus dengan menggunakan pendekatan algoritma *K-Nearest Neighbor* (KNN). Metode KNN bekerja dengan membandingkan data baru terhadap kedekatan jarak dengan data latih [5]. Data yang dianalisis dipetakan ke dalam ruang berdimensi banyak, yang setiap dimensinya merepresentasikan karakteristik tertentu. Ruang tersebut kemudian dibagi berdasarkan kelompok atau kelas data yang berbeda [6].

Penelitian ini memilih metode KNN karena metode ini relatif mudah untuk diimplementasikan dan mampu memberikan hasil yang akurat, terutama jika data telah melalui proses pembersihan *(cleaning)* dan normalisasi [7]. KNN dikenal sebagai algoritma klasifikasi yang sederhana namun efektif dalam berbagai kasus, termasuk pada permasalahan klasifikasi di bidang kesehatan. Beberapa penelitian menunjukkan bahwa KNN mampu memberikan kinerja yang kompetitif dibandingkan algoritma klasifikasi lainnya, khususnya pada dataset dengan distribusi data yang bersifat non-linear. Selain itu, penelitian lainnya juga menunjukkan bahwa algoritma KNN menghasilkan tingkat akurasi yang lebih tinggi dibandingkan *Naïve Bayes* dalam memprediksi penyakit diabetes [8].

Meskipun demikian, KNN memiliki keterbatasan, salah satunya adalah sensitivitas terhadap data berdimensi tinggi, sehingga diperlukan tahapan *pre-processing* data terlebih dahulu sebelum penerapan metode ini untuk meningkatkan keakuratan hasil klasifikasi [7]. Dengan mempertimbangkan karakteristik data serta hasil penelitian terdahulu, penerapan algoritma KNN yang dikombinasikan dengan tahapan *pre-processing* diharapkan dapat menjadi pendekatan yang efektif dalam memprediksi risiko diabetes melitus berdasarkan data klinis pasien. Selain itu, berbagai penelitian sebelumnya menunjukkan bahwa kinerja KNN sangat dipengaruhi oleh pemilihan parameter k dan kualitas fitur yang digunakan. Di sisi lain, meskipun banyak penelitian telah menunjukkan bahwa KNN dapat memberikan kinerja yang baik dalam klasifikasi diabetes, sebagian besar penelitian ini masih bergantung pada dataset publik seperti *Pima Indian Diabetes Database* [3]. Ketergantungan ini menghasilkan keterbatasan dalam secara akurat merepresentasikan kondisi nyata pasien di tingkat lokal, terutama di rumah sakit regional. Selain itu, beberapa penelitian terus berkonsentrasi pada metrik akurasi tanpa mempertimbangkan metrik evaluasi lain yang lebih komprehensif, seperti presisi, *recall*, dan skor F1 [9].

Penelitian ini bertujuan untuk mengklasifikasikan risiko diabetes melitus menggunakan algoritma KNN berdasarkan data klinis dari pasien di RSUD Salatiga, dengan menangani masalah yang telah diidentifikasi. Penelitian ini mengisi celah penelitian dengan memanfaatkan data klinis nyata dan melaksanakan proses seleksi fitur serta optimasi parameter k untuk meningkatkan kinerja model. Kontribusi utama dari penelitian ini adalah pengembangan model klasifikasi yang lebih relevan secara praktis dan menawarkan evaluasi yang lebih komprehensif untuk mendukung pengambilan keputusan di sektor kesehatan.

 ISSN:2302-8149
e-ISSN:2540-9719

Berdasarkan latar belakang tersebut, maka rumusan masalah dalam penelitian ini adalah sebagai berikut (1) bagaimana penerapan algoritma *K-Nearest Neighbor* (KNN) dalam mengklasifikasikan risiko diabetes melitus berdasarkan data pasien RSUD Salatiga? (2) berapa tingkat performa terbaik yang dapat dihasilkan oleh algoritma KNN berdasarkan optimasi parameter k dan seleksi fitur?

Penelitian ini diharapkan dapat berkontribusi pada kemajuan sistem prediksi dini diabetes mellitus yang lebih akurat dan dapat diterapkan pada pasien RSUD Salatiga , sehingga membantu tenaga medis dalam melakukan proses skrining awal dengan lebih cepat dan tepat, serta berpotensi membantu menekan peningkatan jumlah kasus diabetes melitus di wilayah Salatiga dan sekitarnya.

2 Tinjauan Literatur

Berbagai penelitian terbaru menunjukkan bahwa pendekatan *machine learning* telah banyak digunakan dalam deteksi dan prediksi diabetes melitus, baik dengan memanfaatkan data klinis dari rumah sakit maupun dataset publik seperti *Pima Indian Diabetes Database* (PIDD) [3]. Secara umum, penelitian-penelitian tersebut berfokus pada perbandingan kinerja algoritma klasifikasi, optimasi parameter, penerapan teknik penyeimbangan data, serta seleksi fitur guna meningkatkan akurasi model yang dihasilkan [5].

Algoritma yang sering digunakan dalam klasifikasi diabetes melitus meliputi *K-Nearest Neighbor* (KNN), *Naïve Bayes*, *Support Vector Machine* (SVM), *Random Forest* (RF), serta *Logistic Regression* (LR) [7]. Di antara berbagai metode tersebut, KNN dikenal sebagai algoritma yang sederhana namun memiliki performa yang kompetitif dalam berbagai studi kasus klasifikasi. Beberapa penelitian menunjukkan bahwa KNN mampu menghasilkan tingkat akurasi yang cukup tinggi, bahkan dalam beberapa kasus lebih unggul dibandingkan metode lainnya. Sebagai contoh, penelitian yang membandingkan KNN dengan *Naïve Bayes* menunjukkan bahwa KNN dapat memberikan hasil akurasi yang lebih baik pada prediksi diabetes [8]. Selain itu, penelitian lain juga menunjukkan bahwa KNN dengan penerapan seleksi fitur seperti *Information Gain* mampu meningkatkan performa model, meskipun peningkatannya tidak selalu signifikan [6].

Namun demikian, hasil dari berbagai penelitian menunjukkan adanya variasi performa yang cukup signifikan, dengan rentang akurasi KNN berkisar antara sekitar 70% hingga di atas 90%, tergantung pada karakteristik dataset dan metode evaluasi yang digunakan [10]. Variasi ini mengindikasikan bahwa kinerja KNN sangat dipengaruhi oleh distribusi data, pemilihan parameter k, serta teknik pra-pemrosesan yang diterapkan. Dalam konteks penanganan ketidakseimbangan data, beberapa penelitian mengintegrasikan teknik *Synthetic Minority Oversampling Technique* (SMOTE) dan melaporkan adanya peningkatan performa model, baik dari segi akurasi maupun metrik evaluasi lainnya seperti presisi, *recall*, dan *F1-score*. Selain itu, optimasi parameter menggunakan *cross-validation* juga terbukti dapat meningkatkan stabilitas dan generalisasi model [7].

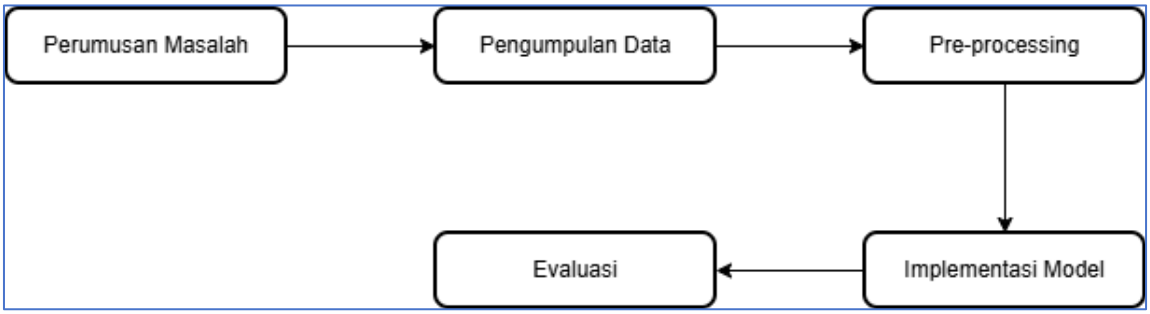
Di sisi lain, beberapa penelitian menunjukkan bahwa algoritma lain seperti *Random Forest* dan *Logistic Regression* dapat memberikan performa yang lebih baik pada dataset tertentu [11]. Hal ini menunjukkan bahwa belum terdapat satu algoritma yang secara konsisten unggul dalam semua kondisi, sehingga pemilihan metode harus disesuaikan dengan karakteristik data yang digunakan. Selain itu, sebagian besar penelitian masih berfokus pada dataset publik seperti PIDD, sementara penelitian berbasis data klinis lokal, khususnya di tingkat rumah sakit daerah, masih relatif terbatas [4].

Berdasarkan celah penelitian tersebut, penelitian ini berfokus pada penerapan dan evaluasi algoritma *K-Nearest Neighbor* (KNN) untuk mendeteksi potensi diabetes melitus dengan menggunakan data klinis pasien dewasa di RSUD Salatiga. Penelitian ini menekankan pada optimasi parameter k melalui pendekatan *cross-validation*, penerapan tahapan pra-pemrosesan data secara menyeluruh, serta evaluasi kinerja model menggunakan berbagai metrik klasifikasi. Dengan demikian, penelitian ini diharapkan tidak hanya menguji kembali efektivitas KNN, tetapi juga memberikan kontribusi dalam pengembangan model prediksi berbasis data klinis lokal yang lebih relevan dan aplikatif sebagai sistem pendukung keputusan di lingkungan layanan kesehatan.

 ISSN:2302-8149
e-ISSN:2540-9719

3 Metode Penelitian

Penelitian akan dilakukan melalui beberapa tahap yaitu perumusan masalah, pengumpulan data, *pre-processing*, implementasi model, dan evaluasi hasil penelitian yang diilustrasikan dalam Gambar 1.



Gambar 1 Tahapan penelitian

Tahap pertama adalah perumusan masalah. Penelitian ini bertujuan untuk mengembangkan sistem prediksi dengan cepat dan akurat dalam mengidentifikasi kemungkinan diabetes melitus di RSUD Salatiga. Penelitian ini menganalisis bagaimana algoritma *K-Nearst Neighbors* (KNN) secara efektif menghasilkan pengklasifikasikan resiko diabetes melitus berdasarkan data medis [5]. Hal ini dilakukan melalui studi literatur, pengumpulan data kesehatan, dan analisis kebutuhan sistem.

Tahap kedua yaitu pengumpulan data. Data yang digunakan dalam penelitian ini merupakan data sekunder yang diperoleh dari RSUD Salatiga sebanyak 502 data pasien. *Dataset* terdiri dari beberapa variabel, yaitu jenis kelamin, usia, riwayat hipertensi, kadar glukosa, kadar HbA1c, status merokok, riwayat penyakit hati, dan indeks massa tubuh (BMI), dengan satu variabel target yaitu status diabetes. Selanjutnya, dilakukan proses pengumpulan dan pemeriksaan data yang akan digunakan dalam penelitian. Data ini kemudian dikompilasi dan disesuaikan dengan kebutuhan algoritma KNN yang membutuhkan data numerik dan bersih [9].

Tahap ketiga dalam penelitian ini adalah *pre-processing*, yang dijalankan dengan tujuan menjaga kualitas data sebelum tahap pemodelan dilakukan. Tahap *pre-processing* data dilakukan untuk memastikan kualitas data sebelum digunakan dalam proses pemodelan, karena algoritma *K-Nearest Neighbor* (KNN) sangat sensitif terhadap skala data dan kualitas input. Proses ini dimulai dengan tahap pembersihan data yang mencakup pemeriksaan nilai yang hilang (*missing values*), duplikasi data, serta validitas nilai pada setiap atribut. Selanjutnya, dilakukan transformasi data melalui *encoding* untuk variabel kategorikal, seperti jenis kelamin, dengan metode label *encoding* agar data bisa direpresentasikan dalam bentuk numerik dan dapat diproses oleh algoritma KNN. Tahap berikutnya adalah normalisasi data menggunakan metode *StandardScaler* untuk menyamakan skala antar fitur, sehingga setiap variabel memiliki distribusi dengan rata-rata 0 dan standar deviasi 1. Normalisasi ini sangat penting untuk mencegah dominasi fitur tertentu dalam perhitungan jarak *Euclidean* yang digunakan oleh KNN. Setelah semua proses tersebut selesai, *dataset* dibagi menjadi dua bagian, yaitu data latih *(training data)* sebesar 80% dan data uji *(testing data)* sebesar 20% dengan metode pengambilan acak, sehingga distribusi data tetap representatif. Tahapan *pre-processing* ini sangat penting karena kualitas data yang baik akan langsung mempengaruhi kinerja model dalam menghasilkan klasifikasi yang akurat [4].

Tahap keempat mulai mengimplementasi model. Tahap ini melibatkan penerapan *data mining* dalam mengklasifikasikan objek baru dan membandingkannya dengan data uji yang paling dekat jaraknya. Proses klasifikasi dengan algoritma KNN mencakup dua langkah, yaitu: (1) menentukan nilai k, serta (2) menghitung jarak dengan metode *Euclidean Distance* [5], dengan menggunakan rumus (1) sebagai berikut [15].

$$f(a,b) = \sqrt{\sum_{n=1}^{p} \quad (X_{ag} - X_{bg})^2} \quad (1)$$

 ISSN:2302-8149
e-ISSN:2540-9719

dengan:
X_{ag} = Nilai atribut ke – k pada data latih (objek a)
X_{bg} = Nilai atribut ke – k pada data uji (objek b)
p = Banyaknya atribut (fitur) yang dibandingkan

Setelah mengimplemntasi model juga dilakukan metode *Sequential Forward Selection* (SFS), *Sequential Forward Selection* digunakan untuk menentukan ukuran subset fitur yang optimal merujuk pada proses memilih sebagian dari semua atribut dalam sebuah *dataset*, yang bertujuan untuk meningkatkan kinerja model atau analisis data [12]. Tujuan utama dari pemilihan fitur adalah untuk mengurangi dimensi data, meningkatkan interpretabilitas, mencegah *overfitting*, dan mempercepat proses pelatihan model. Dengan memilih hanya atribut yang paling relevan dan informatif, seleksi fitur dapat meningkatkan efisiensi algoritma dan kualitas hasil analisis [13]. Tahap akhir adalah melakukan penilaian kinerja algoritma KNN untuk mengetahui tingkat akurasi klasifikasi penyakit diabetes melitus pada data pasien RSUD Salatiga. Proses pengujian dilakukan dengan menggunakan *confusion matrix* yang dianalisis melalui bahasa pemrograman Python [11].

4 Hasil dan Pembahasan

Analisis Data

Data sekunder yang diperoleh dari rekam medis manual RSUD Salatiga berfungsi sebagai sumber data untuk penelitian ini. Struktur dataset yang digunakan dalam penelitian ditunjukkan pada Tabel 1. Menurut gambar tersebut, terdapat beberapa fitur atau atribut yang relevan dengan diagnosis diabetes, termasuk jenis kelamin, usia, riwayat hipertensi, riwayat penyakit hati, status merokok, indeks massa tubuh (BMI), kadar HbA1c, dan kadar glukosa. Sementara itu, kolom diabetes berfungsi sebagai variabel target (label). Secara keseluruhan, dataset ini terdiri dari 502 sampel. [10].

Tabel 1 Tampilan awal dataset penelitian

No	Gender	Umur	Hipertensi	Penyakit Hati	Merokok	BMI	HbA1c level	Kadar glukosa	Diabetes
1	L	47	1	0	1	34.6	8.8	174	0
2	L	55	1	0	0	38.3	5.5	365	1
3	L	42	1	0	1	36.1	13.8	161	0
4	P	53	1	0	0	33.1	13.7	143	0
5	P	66	1	0	0	31.9	13.6	653	1
6	P	73	1	0	0	37.4	4.7	341	1
7	L	46	1	0	1	22.2	13.9	490	1
8	L	47	0	1	1	38.2	13.1	311	1
9	L	47	0	0	1	24.3	12.6	211	1

Pengolahan Data

Tujuan dari tahap pengolahan data penelitian untuk menghasilkan informasi yang berperan dalam pengambilan keputusan dan penyusunan laporan. Data tersebut dikumpulkan dari rekam medis manual RSUD Salatiga. Analisis mencakup persiapan data, *pre-processing*, pembentukan model, pelatihan, dan pengujian untuk mendapatkan nilai akurasi dan prediksi yang optimal. Metode ini menggunakan algoritma *K-Nearest Neighbor* (KNN) dan *Naïve Bayes* yang diimplementasikan pada platform *Google Colab* menggunakan bahasa pemrograman *Python* [5].

Pre-processing

Sebelum melanjutkan analisis lebih lanjut, *data cleaning* dilakukan untuk memeriksa keberadaan data kosong atau data yang hilang. Hasil pemeriksaan nilai yang hilang pada *dataset* disajikan pada Gambar 2. Menurut Gambar 2, terlihat jelas bahwa sebagian besar atribut berisi data lengkap (dengan nilai 0); namun, terdapat nilai yang hilang pada atribut jenis kelamin sebanyak 1 data dan pada atribut hipertensi sebanyak 1 data . Identifikasi nilai-nilai yang hilang ini berfungsi sebagai dasar untuk

 ISSN:2302-8149
e-ISSN:2540-9719

menerapkan langkah-langkah penanganan data (seperti imputasi atau penghapusan baris) sebelum tahap pengkodean variabel kategorikal.

```
Missing value :
gender            1
umur              0
hipertensi        1
penyakit hati     0
merokok           0
BMI               0
HbA1c_level       0
kadar_glukosa     0
diabetes          0
dtype: int64
```

Gambar 2 Tahapan missing value

Pada tahap pra-pemrosesan data ini, proses pengkodean diterapkan pada variabel kategorikal jenis kelamin. Hasil transformasi data ini diilustrasikan pada Gambar 2. Menurut Gambar 4, nilai asli variabel jenis kelamin, yang awalnya dalam format teks (seperti laki-laki dan perempuan), telah dikonversi menjadi numerik menggunakan tipe data bilangan bulat biner. Terlihat jelas dari lima baris pertama sampel data bahwa kategori jenis kelamin telah berhasil direpresentasikan oleh angka 0 dan 1, memastikan bahwa strukturnya kompatibel dan dapat diproses oleh algoritma pembelajaran mesin.

```
0    0
1    0
2    0
3    1
4    1
Name: gender, dtype: int64
```

Gambar 3 Tahapan encoding

Setelah memisahkan fitur prediktor dari variabel target (diabetes), fase normalisasi data dilakukan menggunakan metode *MinMaxScaler*. Visualisasi data yang dinormalisasi dapat dilihat pada Gambar 4. Gambar 4 mengilustrasikan representasi numerik baru dari semua fitur seperti jenis kelamin, usia, hipertensi, penyakit hati, merokok, BMI, kadar HbA1c, dan kadar glukosa, yang sekarang semua nilainya diskalakan secara seragam dalam rentang 0 hingga 1, sehingga siap digunakan dalam pemodelan diagnosis diabetes.

	gender	umur	hipertensi	penyakit hati	merokok	BMI	HbA1c_level	kadar_glukosa
0	0.0	0.364706	1.0	0.0	1.0	0.752336	0.48	0.143143
1	0.0	0.458824	1.0	0.0	0.0	0.925234	0.15	0.334334
2	0.0	0.305882	1.0	0.0	1.0	0.822430	0.98	0.130130
3	0.5	0.435294	1.0	0.0	0.0	0.682243	0.97	0.112112
4	0.5	0.588235	1.0	0.0	0.0	0.626168	0.96	0.622623

Gambar 4 Tahapan normalisasi

Proses pra-pemrosesan menghasilkan total 502 titik data valid yang siap untuk pemodelan. Karakteristik distribusi kelas data ini dapat diamati pada Gambar 5. Gambar 5 mengilustrasikan komposisi variabel target, di mana jumlah titik data pada kategori 0 (non-diabetes) ada 244, sedangkan pada kategori 1 (positif diabetes) berjumlah 258. [14].

 ISSN:2302-8149
e-ISSN:2540-9719

```
Jumlah data 0 (tidak diabetes): 244
Jumlah data 1 (positif diabetes): 258
```

Gambar 5 Jumlah data berdasarkan kelas diabetes

Seleksi Fitur

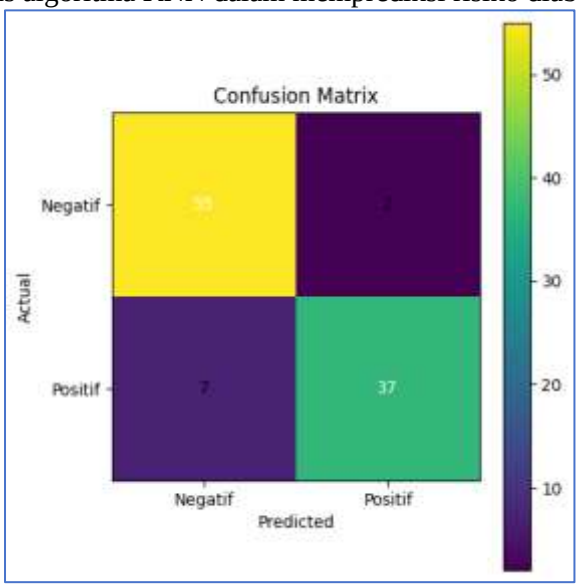
Hasil pengujian pemilihan fitur ini dapat diamati pada Gambar 6. Gambar 6 mengilustrasikan bahwa akurasi model telah meningkat secara signifikan, dari 78,22% sebelum pemilihan fitur menjadi 95,05% setelah proses pemilihan. Melalui proses eliminasi SFS, algoritma berhasil menyaring dataset, mempertahankan empat fitur yang paling berpengaruh dalam mendiagnosis diabetes, yang meliputi variabel hipertensi, penyakit hati, merokok, dan kadar glukosa. Pengurangan atribut yang kurang relevan ini secara efektif mengoptimalkan perhitungan jarak untuk tetangga terdekat dalam model KNN. [13].

```
Akurasi sebelum seleksi fitur: 0.7822
Fitur terpilih: ['hipertensi', 'penyakit hati', 'merokok', 'kadar_glukosa']
Akurasi sesudah seleksi fitur: 0.9505
```

Gambar 6 Akurasi sebelum dan sesudah seleksi fitur

Confusion Matrix

Gambar 7 menggambarkan hasil *confusion matrix*, yang memberikan analisis mendalam tentang kinerja model. Menurut data, model secara akurat mengklasifikasikan 55 pasien negatif dan 37 pasien diabetes positif. Meskipun terdapat 9 kesalahan klasifikasi (2 positif palsu dan 7 negatif palsu), hasil ini menunjukkan efektivitas algoritma KNN dalam memprediksi risiko diabetes di RSUD Salatiga.



Gambar 7 *Confusion matrix* evaluasi model KNN

Nilai k

Hasil pengujian kinerja algoritma KNN terkait berbagai nilai k disajikan dalam Tabel 2. Menentukan nilai k yang optimal sangat penting untuk mencapai hasil prediksi yang akurat sekaligus mempertahankan keseimbangan antara kondisi *overfitting* dan *underfitting*. Berdasarkan data dalam tabel, jelas bahwa variasi nilai k memiliki dampak langsung pada kinerja algoritma, di mana pemilihan k yang terlalu kecil atau terlalu besar dapat menurunkan kualitas klasifikasi. Temuan penelitian menunjukkan bahwa tingkat akurasi tertinggi dicapai pada k = 6, yaitu 92,08%.

 ISSN:2302-8149
e-ISSN:2540-9719

Tabel 2 Hasil Pengujian akurasi algoritma KNN berdasarkan variasi nilai K.

Nilai K	Akurasi
1	0,8416
2	0,8713
3	0,8713
4	0,9109
5	0,9010
6	0,9208

Evaluasi

Berdasarkan serangkaian metode yang telah diterapkan, hasil evaluasi kinerja model disajikan dalam Tabel 3. Data dalam tabel ini menunjukkan bahwa model KNN menunjukkan kinerja yang kuat dengan tingkat akurasi sebesar 92,08%, bersama dengan nilai presisi, *recall*, dan *F1-score* yang tinggi. Nilai presisi yang tinggi menunjukkan efektivitas model dalam meminimalkan kesalahan klasifikasi pada pasien non-diabetes (positif palsu). Sementara itu, nilai recall yang baik mencerminkan kemampuan model untuk secara akurat mengidentifikasi pasien yang benar-benar menderita diabetes. Hal ini sangat penting dalam bidang medis untuk mencegah keterlambatan dalam perawatan pasien. Dengan *F1-score* yang seimbang, model KNN yang diusulkan terbukti stabil dan dapat diandalkan sebagai alat deteksi dini diabetes mellitus.

Tabel 3 Hasil evaluasi model KNN

Akurasi Evaluasi Model KNN	Presisi		*Recall*		*F1-Score*	
0,87	0 (tidak diabetes)	1 (diabetes)	0 (tidak diabetes)	1 (diabetes)	0 (tidak diabetes)	1 (diabetes)
	0,91	0,83	0,86	0,89	0,88	0,86

5 Kesimpulan

Penelitian ini menunjukkan bahwa algoritma *K-Nearest Neighbor* (KNN) dapat digunakan secara efektif untuk mengklasifikasikan risiko diabetes melitus berdasarkan data klinis dari pasien di RSUD Salatiga. Penerapan seleksi fitur menggunakan metode *Sequential Forward Selection* (SFS) dan optimasi parameter k terbukti meningkatkan kinerja model, dengan akurasi meningkat dari 87,13% menjadi 92,08% pada k = 6. Selain itu, model ini menunjukkan kinerja yang baik menurut metrik presisi, *recall*, dan *F1-score*. Hasil seleksi fitur menunjukkan bahwa riwayat hipertensi, penyakit hati, status merokok, dan tingkat glukosa merupakan fitur yang paling berpengaruh dalam klasifikasi diabetes. Model yang dihasilkan memiliki potensi untuk berfungsi sebagai sistem pendukung keputusan untuk membantu tenaga kesehatan dalam deteksi dini diabetes mellitus. Penelitian di masa depan disarankan untuk memanfaatkan dataset yang lebih besar dan menerapkan metode validasi yang lebih komprehensif, seperti *k-fold*, *cross-validation*, serta membandingkan kinerja KNN dengan algoritma machine learning lainnya.

Referensi

[1] I. A. K. Adi and W. A. E. Prabowo, "Klasifikasi Penyakit Diabetes menggunakan Pendekatan Pembelajaran Mesin dengan Model Non-linier," *JURIKOM (Jurnal Ris. Komputer)*, Vol. 12, No. 3, pp. 262–268, 2025, DOI: 10.30865/jurikom.v12i3.8586.

[2] Lestari, Zulkarnain, Sijid, and S. Aisyah, "Diabetes Melitus: Review Etiologi, Patofisiologi, Gejala, Penyebab, Cara Pemeriksaan, Cara Pengobatan dan Cara Pencegahan," *UIN Alauddin Makassar*, Vol. 1, No. 2, pp. 237–241, 2021, [Online]. Available: http://journal.uin-alauddin.ac.id/index.php/psb

[3] S. S. Bhat, V. Selvam, and G. A. Ansari, "*Predicting Life Style of Early Diabetes Mellitus using Machine Learning Technique*," *Int. J. Comput.*, Vol. 22, No. 3, pp. 345–351, 2023, DOI:

 ISSN:2302-8149
e-ISSN:2540-9719

10.47839/ijc.22.3.3230.

[4] S. S. Putro *et al.*, "*Classification of Diabetes Mellitus Disease at Rato Ebuh Hospital-Indonesia using the K-Nearest Neighbors Method based on Missing Value*," *BIO Web Conf.*, Vol. 146, 2024, DOI: 10.1051/bioconf/202414601081.

[5] M. Saputra, J. P. Sidabuke, R. P. Sinulingga, and R. B. Tamba, "Analisis Metode Algoritma *K-Nearest Neighbor* (*KNN*) dan *Naive Bayes* untuk Klasifikasi Diabetes Mellitus," *J. TEKINKOM*, Vol. 6, No. 2, p. 2023, 2023, DOI: 10.37600/tekinkom.v6i2.942.

[6] D. Devian, P. Nurul Sabrina, and A. Komarudin, "Prediksi Penyakit Diabetes dengan Metode *K-Nearest Neighbor* (*KNN*) dan Seleksi Fitur Information Gain," *JATI (Jurnal Mhs. Tek. Inform.*, Vol. 8, No. 6, pp. 11320–11326, 2024, DOI: 10.36040/jati.v8i6.11364.

[7] A. Mulyani, S. Khoerunisa, and D. Kurniadi, "*Comparison of KNN and SVM Algorithms Performance using SMOTE to Classify Diabetes," J. Nas. Tek. Elektro dan Teknol. Inf.*, Vol. 14, No. 1, pp. 25–34, 2025.

[8] Dewi Nasien *et al.*, "Perbandingan Implementasi *Machine Learning* menggunakan Metode K*NN, Naive Bayes*, dan Logistik Regression untuk mengklasifikasi Penyakit Diabetes," *JEKIN - J. Tek. Inform.*, Vol. 4, No. 1, pp. 10–17, 2024, DOI: 10.58794/jekin.v4i1.640.

[9] D. Fabiyanto and Y. Rianto, "*Performance Evaluation of Multiple Machine Learning Models for Wine Quality Prediction Evaluasi Kinerja Multiple Model Machine Learning untuk Prediksi Kualitas Wine*," *J. Inform. dan Teknol. Inf.*, Vol. 21, No. 2, pp. 209–223, 2024, DOI: 10.31515/telematika.v21i2.

[10] M. D. Nurmalasari, K. Kusrini, and S. Sudarmawan, "Komparasi Algoritma *Naive Bayes* dan *K-Nearest Neighbor* untuk membangun Pengetahuan Diagnosa Penyakit Diabetes," *J. Komtika (Komputasi dan Inform.*, Vol. 5, No. 1, pp. 52–59, 2021, DOI: 10.31603/komtika.v5i1.5140.

[11] N. Trisna and A. Mahessya, "Prediksi Risiko Diabetes Tahap Awal menggunakan *Machine Learning* dengan Algoritma *K-Nearest Neighbor*," Vol. 5, No. 2, pp. 481–487, 2025, [Online]. Available: https://jurnal.pustakagalerimandiri.co.id/index.php/pustakadataDOI:https://doi.org/10.55382/jurnalpustakadata.v5i2.1550

[12] D. Kuswandani and R. Umar, "Perbandingan Metode Seleksi Fitur Filter dan Wrapper pada Klasifikasi Risiko Penyakit Jantung menggunakan k - NN *Comparison of Filter and Wrapper Feature Selection Methods for Heart Disease Risk Classification using K - Nearest Neighbors (k - NN)*," Vol. 15, pp. 871–885, 2026.

[13] R. Hadi, N. L. G. P. Suwirmayanti, I. G. N. A. Kusuma, I. G. A. D. Saryanti, and P. D. Novayanti, "Implementasi Metode Normalisasi dan Seleksi Fitur dalam Optimasi Algoritma *K-Nearest Neighbor (KNN)* untuk Klasifikasi Data Bank," *J. Nas. Komputasi dan Teknol. Inf.*, Vol. 7, No. 5, pp. 1064–1071, 2024, DOI: 10.32672/jnkti.v7i5.7989.

[14] U. I. Lestari, "Penerapan Metode *K-Nearest Neighbor* untuk Sistem Pendukung Keputusan Identifikasi Penyakit Diabetes Melitus," *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, Vol. 8, No. 4, pp. 2071–2082, 2021, DOI: 10.35957/jatisi.v8i4.1235.

[15] E. Kavlakoglu, "Apa Algoritma K-Tetangga Terdekat (*KNN*)?," IBM Research. [Online]. Available: https://www.ibm.com/id-id/think/topics/knn