ISSN: 3110-0864 (Media Online)
JURNAL ILMU KOMPUTER DAN TEKNIK INFROMATIKA
Volume 2 No 1 Januari 2026 | pp: 119-130
https://journals.raskhamedia.or.id/index.php/juikti
DOI: https://doi.org/10.64803/juikti.v2i1.116

Sistem Klasifikasi Diabetes Mellitus Menggunakan Algoritma K-Nearest Neighbor (KNN) Berbasis Web

Muhammad Randy Fachrezi[1*], Alwi Syahputra[2], Hafiz Aryanda[3], Rusma Riansyah[4]

[1,2,3,4] Fakultas Sains dan Teknologi, Ilmu Komputer, Universitas Islam Negeri Sumatera Utara, Medan, Indonesia
Email: [1*]randyfachrezy6@gmail.com, [2]alwiisyahputra0@gmail.com, [3]hafizaryandaaa@gmail.com, [4]riansyahrusma@gmail.com
(* Email Corresponding Author: randyfachrezy6@gmail.com)
Dikirim: 28 Oktober 2025 | Diterima: 5 Januari 2026 | Dipublish: 27 Januari 2026

Abstrak

Diabetes mellitus merupakan penyakit metabolik kronis yang ditandai dengan peningkatan kadar glukosa darah akibat gangguan sekresi atau kerja insulin dan prevalensinya terus meningkat, termasuk di Indonesia yang mencapai 11,3% dari total populasi berdasarkan data International Diabetes Federation tahun 2024, sementara permasalahan utama dalam penanganannya adalah keterlambatan diagnosis yang dapat menyebabkan berbagai komplikasi serius seperti gangguan ginjal, penyakit kardiovaskular, dan kerusakan saraf, sehingga diperlukan suatu sistem pendukung keputusan yang mampu membantu proses deteksi dini secara cepat, akurat, dan mudah diakses; oleh karena itu, penelitian ini bertujuan untuk mengembangkan sistem klasifikasi diabetes mellitus berbasis web menggunakan algoritma K-Nearest Neighbor (KNN) sebagai metode klasifikasi utama dengan memanfaatkan dataset Pima Indians Diabetes Database yang terdiri dari 768 data dan delapan atribut klinis, termasuk usia, indeks massa tubuh, kadar glukosa, dan tekanan darah, dengan tahapan prapemrosesan data yang meliputi imputasi median untuk menangani nilai anomali, penanganan outlier menggunakan metode Interquartile Range (IQR), penyeimbangan distribusi kelas menggunakan teknik Synthetic Minority Over-sampling Technique (SMOTE), serta normalisasi Min–Max untuk menyamakan skala fitur; penentuan nilai parameter K optimal dilakukan menggunakan metode cross-validation dan menghasilkan nilai terbaik pada K = 8, kemudian model dievaluasi menggunakan data pengujian dengan hasil akurasi sebesar 73,68%, precision 71,15%, recall 78,72%, F1-score 74,75%, dan nilai ROC-AUC sebesar 0,8132 yang menunjukkan bahwa model memiliki kemampuan yang cukup baik dalam membedakan pasien diabetes dan non-diabetes, khususnya dalam mendeteksi kasus positif diabetes; sistem web dikembangkan menggunakan arsitektur client–server dengan React.js sebagai frontend dan Python Flask sebagai backend, serta dilengkapi dengan fitur edukasi kesehatan, formulir input data klinis pengguna, dan halaman hasil klasifikasi yang informatif disertai rekomendasi tindak lanjut, sehingga berdasarkan hasil pengujian dapat disimpulkan bahwa sistem yang dikembangkan mampu berfungsi dengan baik dan berpotensi digunakan sebagai alat bantu skrining awal diabetes mellitus sebelum dilakukan pemeriksaan medis lanjutan.

Kata Kunci: Diabetes Mellitus, K-Nearest Neighbor, Klasifikasi, Pembelajaran Mesin, Sistem Berbasis Web

Abstract

Diabetes mellitus is a chronic metabolic disease characterized by elevated blood glucose levels caused by impaired insulin secretion or function, and its prevalence continues to increase, including in Indonesia where it reached 11.3% of the total population according to the International Diabetes Federation in 2024, while delayed diagnosis remains a major challenge that often leads to serious complications such as kidney failure, cardiovascular diseases, and nerve damage; therefore, an accessible and reliable decision support system is required to facilitate early detection, and this study aims to develop a web-based diabetes mellitus classification system using the K-Nearest Neighbor (KNN) algorithm as the primary classification method by utilizing the Pima Indians Diabetes Database consisting of 768 instances with eight clinical attributes including age, body mass index, glucose level, and blood pressure, with data preprocessing stages involving median imputation to handle anomalous values, outlier treatment using the Interquartile Range (IQR) method, class distribution balancing through the Synthetic Minority Over-sampling Technique (SMOTE), and feature scaling using Min–Max normalization; the optimal K parameter was determined through cross-validation, resulting in K = 8 as the best configuration, and the developed model achieved a testing accuracy of 73.68%, with a precision of 71.15%, recall of 78.72%, F1-score of 74.75%, and an ROC-AUC value of 0.8132, indicating that the model has satisfactory capability in distinguishing diabetic and non-diabetic patients, particularly in identifying positive diabetes cases; the web-based system was implemented using a client–server architecture with React.js for the frontend interface and Python Flask for backend processing, and it provides health education features, a clinical data input form, and an informative classification result page accompanied by follow-up recommendations, demonstrating that the proposed system performs adequately and has strong potential to be used as an initial screening tool to support early detection of diabetes mellitus prior to comprehensive medical examination.

Keywords*: Diabetes Mellitus, K-Nearest Neighbor, Classification, Machine Learning, Web-Based System*

1. PENDAHULUAN

Diabetes mellitus merupakan salah satu penyakit metabolik kronis yang ditandai dengan meningkatnya kadar glukosa darah akibat gangguan produksi maupun fungsi insulin. Menurut laporan *World Health Organization* (WHO), jumlah penderita diabetes di dunia terus meningkat setiap tahunnya, dengan estimasi mencapai lebih dari 830 juta orang pada tahun 2022 [1]. Di Indonesia, prevalensi diabetes mencapai

ISSN: 3110-0864 (Media Online)
JURNAL ILMU KOMPUTER DAN TEKNIK INFROMATIKA
Volume 2 No 1 Januari 2026 | pp: 119-130
https://journals.raskhamedia.or.id/index.php/juikti
DOI: https://doi.org/10.64803/juikti.v2i1.116

11,3% dari total populasi berdasarkan data International Diabetes Federation (IDF) tahun 2024 [2]. Tingginya angka prevalensi ini menunjukkan bahwa diabetes merupakan masalah kesehatan masyarakat yang serius dan membutuhkan strategi deteksi dini yang efektif.

Permasalahan utama dalam penanganan diabetes adalah keterlambatan diagnosis. Banyak penderita tidak menyadari kondisi mereka hingga penyakit berkembang ke tahap lanjut dengan komplikasi serius, seperti gangguan kardiovaskular, nefropati, neuropati, maupun retinopati [3]. Oleh karena itu, deteksi dini menjadi kunci dalam pencegahan komplikasi dan penanganan yang lebih efektif. Pemanfaatan teknologi informasi melalui pendekatan *data mining* dan *machine learning* dapat menjadi solusi alternatif dalam mendukung proses klasifikasi risiko diabetes secara lebih cepat dan akurat.

Salah satu algoritma *machine learning* yang banyak digunakan dalam klasifikasi adalah K-Nearest Neighbor (KNN). Algoritma ini bekerja dengan prinsip kedekatan (*similarity*) antara data uji dengan data latih berdasarkan perhitungan jarak [4]. Sejumlah penelitian terdahulu menunjukkan efektivitas KNN dalam klasifikasi diabetes dengan tingkat akurasi yang tinggi. Namun, penerapan KNN dalam bentuk sistem berbasis web yang dapat diakses secara *real-time* masih relatif terbatas.

Salah satu penelitian dilakukan oleh Gunawan & Fenriana yang mengembangkan aplikasi prediksi diabetes berbasis web menggunakan algoritma KNN dengan *dataset Early Stage Diabetes Risk Prediction*. Penelitian ini menggunakan Z-Score normalization dan jarak Euclidean, menghasilkan akurasi optimal sebesar 97,12%. Hasilnya menunjukkan efektivitas KNN dalam *screening diabetes* melalui aplikasi interaktif [5]. Penelitian lain oleh Ibanez, Wiriasto & Rosmaliati mengombinasikan PCA dengan algoritma K-Means untuk klasterisasi data stunting. Walaupun fokusnya bukan pada diabetes, penelitian ini relevan karena menekankan pentingnya *preprocessing* dan reduksi dimensi dalam meningkatkan kualitas klasifikasi data kesehatan [6].

Selanjutnya, Oktaviana dkk. menerapkan KNN untuk prediksi diabetes tipe 2 menggunakan data klinis lokal dari Puskesmas Mlati II, Sleman. Penelitian ini berfokus pada penerapan algoritma K-Nearest Neighbor (KNN) dalam memprediksi penyakit diabetes melitus tipe 2 dengan memanfaatkan data klinis dari Puskesmas Mlati II, Sleman. Dataset terdiri dari 200 sampel dengan lima atribut, yaitu usia, tekanan darah, BMI, kadar gula darah puasa, serta status diabetes. Tahapan penelitian mencakup preprocessing menggunakan MinMax Normalization, pembagian data dengan Stratified 5-fold Cross Validation, serta implementasi KNN menggunakan Manhattan Distance dengan nilai tetangga (k) = 13, ditentukan dari rumus √n. Hasil pengujian menunjukkan bahwa model KNN mampu memberikan performa yang cukup baik dengan akurasi 88%, precision 83%, recall 87%, dan f1-score 85%. Temuan ini memperlihatkan bahwa KNN efektif digunakan untuk klasifikasi diabetes, khususnya dengan data klinis lokal di Indonesia [7].

Kemudian, Pratama dkk membandingkan kinerja Random Forest, KNN, dan Logistic Regression dalam prediksi diabetes. Penelitian ini berfokus membandingkan kinerja tiga algoritma machine learning, yaitu Random Forest, K-Nearest Neighbor, dan Logistic Regression, dalam memprediksi risiko diabetes menggunakan dataset UCI yang tersedia di Kaggle. Dataset terdiri dari 768 data dengan delapan parameter. Penelitian ini menerapkan preprocessing data, normalisasi, serta reduksi dimensi dengan PCA, kemudian mengevaluasi performa model berdasarkan akurasi, confusion matrix, dan ROC-AUC. Dari hasil analisis, Logistic Regression memperoleh performa terbaik dengan akurasi 77% dan AUC 0,83, diikuti oleh KNN dengan akurasi 75% dan AUC 0,81, serta Random Forest dengan akurasi 74% dan AUC 0,81. Penelitian ini menekankan bahwa pemilihan algoritma yang tepat dan penerapan preprocessing yang baik sangat berpengaruh terhadap hasil prediksi[8].

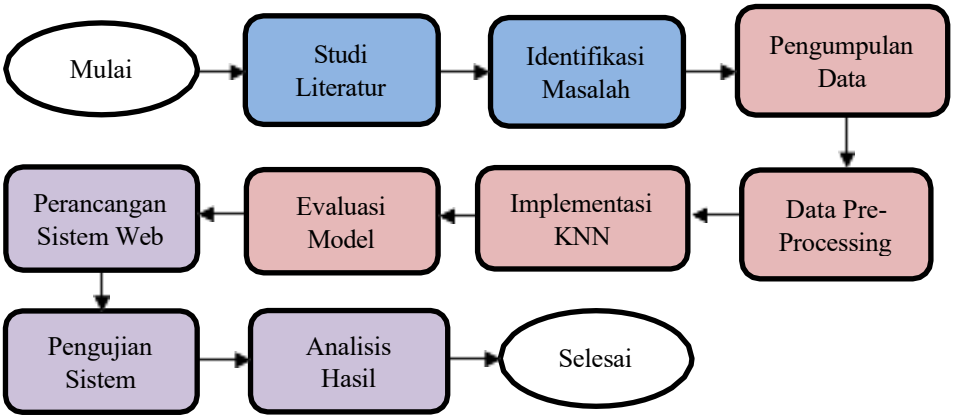
Terakhir, Amri dkk mengombinasikan KNN dengan teknik SMOTE-ENN untuk mengatasi ketidakseimbangan data pada *dataset Pima Indians*. Penelitian ini berfokus untuk optimasi prediksi diabetes pada wanita Pima Indian dengan mengombinasikan algoritma KNN dan teknik SMOTE-ENN untuk mengatasi ketidakseimbangan data. Dataset penelitian berasal dari Kaggle dan UCI Repository dengan total 768 data yang mencakup sembilan variabel, seperti usia, jumlah kehamilan, glukosa, tekanan darah, ketebalan kulit, insulin, BMI, diabetes pedigree function, dan outcome. Teknik SMOTE-ENN digunakan untuk memperbaiki representasi kelas minoritas sekaligus mengurangi noise. Proses pelatihan dan pengujian dilakukan dengan pembagian data 70% training dan 30% testing menggunakan Jupyter Notebook. Hasil penelitian menunjukkan peningkatan performa signifikan, dengan akurasi mencapai 96%, classification error 0,04, serta AUC 0,95. Hal ini membuktikan bahwa kombinasi KNN dengan SMOTE-ENN mampu memberikan prediksi yang sangat baik dalam kasus imbalanced dataset[9].

Dengan uraian penelitian terdahulu tersebut, terlihat bahwa KNN telah terbukti efektif dalam klasifikasi diabetes, baik dengan data internasional maupun lokal. Namun, sebagian besar penelitian masih berfokus pada optimasi algoritma dan *preprocessing*, sementara penerapan KNN dalam bentuk sistem berbasis web yang dapat diakses secara *real-time* masih jarang dilakukan. Inilah yang menjadi GAP Analysis dari penelitian ini, yaitu mengintegrasikan KNN ke dalam sistem web interaktif untuk mendukung deteksi dini diabetes secara praktis dan mudah diakses.

ISSN: 3110-0864 (Media Online)
JURNAL ILMU KOMPUTER DAN TEKNIK INFROMATIKA
Volume 2 No 1 Januari 2026 | pp: 119-130
https://journals.raskhamedia.or.id/index.php/juikti
DOI: https://doi.org/10.64803/juikti.v2i1.116

2. METODOLOGI PENELITIAN

Dengan menggunakan algoritma K-NN untuk melakukan klasifikasi penyakit diabetes mellitus, setiap tahapan penelitian ini disajikan dalam diagram alir untuk memastikan bahwa penelitian berjalan sesuai dengan tujuan yang telah ditetapkan.Tahapan penelitian disajikan pada gambar 1.



Gambar 1. Tahapan dalam pembuatan model dan sistem web untuk deteksi diabetes.

2.1. Mulai

Tahap ini merupakan titik awal dimulainya seluruh rangkaian proses penelitian, yang diawali dengan penetapan topik dan judul penelitian.

2.2. Studi Literatur

Pada tahap ini, peneliti melakukan pengumpulan landasan teori dan tinjauan pustaka yang relevan dengan penelitian. Kegiatan ini mencakup studi terhadap penelitian-penelitian sebelumnya (*penelitian terdahulu*) terkait klasifikasi diabetes, penerapan algoritma K-Nearest Neighbor (KNN), serta perancangan sistem pendukung keputusan berbasis web. Tujuannya adalah untuk mendapatkan pemahaman mendalam mengenai topik dan mengidentifikasi celah penelitian.

2.3. Identifikasi Masalah

Setelah melakukan studi literatur, peneliti mengidentifikasi masalah yang akan diselesaikan. Berdasarkan temuan, masalah dirumuskan, misalnya, "Bagaimana merancang sebuah sistem berbasis web yang mampu mengklasifikasikan risiko diabetes mellitus menggunakan algoritma KNN secara efektif dan mudah diakses oleh pengguna?"

2.4. Pengumpulan Data

Tahap ini adalah proses pengumpulan *dataset* yang akan digunakan untuk melatih dan menguji model. Sesuai dengan batasan penelitian, data yang digunakan adalah *dataset* sekunder yang tersedia secara publik, yaitu **Pima Indians Diabetes Database**. *Dataset* ini berisi atribut-atribut relevan seperti kadar glukosa, tekanan darah, indeks massa tubuh (BMI), umur, dan variabel lainnya, beserta outcome (label kelas) apakah seorang pasien terdiagnosis diabetes atau tidak. Dataset ini berasal dari **National Institute of Diabetes and Digestive and Kidney Diseases** dengan tujuan utama untuk memprediksi status diabetes secara diagnostik. Seluruh data diambil dari pasien perempuan berusia minimal 21 tahun dengan latar belakang etnis Pima Indian, sehingga memberikan karakteristik khusus pada populasi penelitian. Variabel prediktor yang tersedia mencakup jumlah kehamilan, kadar insulin, BMI, usia, serta faktor klinis lainnya, sedangkan variabel target adalah *Outcome* yang menunjukkan apakah pasien terdiagnosis diabetes atau tidak [10].

Tabel 1. Informasi tentang *dataset*

Parameter Data	Keterangan
Jumlah Data	768
Atribut/Fitur	8 (Pregnancies, Glucose, BloodPressure, SkinThickness, Insulin, BMI, DiabetesPedigreeFunction, Age)
Target/Label	1 (Outcome)
Tipe data	Int64 (7), float64 (2)

ISSN: 3110-0864 (Media Online)
JURNAL ILMU KOMPUTER DAN TEKNIK INFROMATIKA
Volume 2 No 1 Januari 2026 | pp: 119-130
https://journals.raskhamedia.or.id/index.php/juikti
DOI: https://doi.org/10.64803/juikti.v2i1.116

Anomali/Kecacatan	Memiliki banyak nilai dengan angka 0 (anomali), Memiliki nilai Outliers Distribusi label tidak merata

2.5. Data Pre-Processing

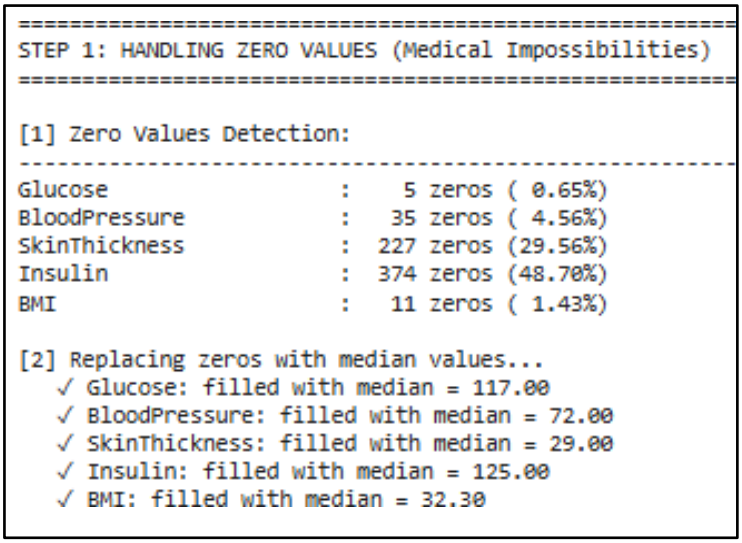
Data mentah yang telah dikumpulkan tidak dapat langsung digunakan karena seringkali mengandung ketidaksempurnaan seperti nilai kosong, anomali, maupun distribusi yang tidak seimbang. Oleh karena itu, tahap *pre-processing* (pra-pemrosesan) data menjadi langkah krusial untuk memastikan kualitas *dataset* sebelum digunakan dalam proses klasifikasi [11]. Proses ini bertujuan untuk membersihkan, menormalkan, dan menyiapkan data agar sesuai dengan kebutuhan algoritma K-Nearest Neighbor (KNN) yang berbasis perhitungan jarak. Proses ini mencakup:

2.5.1 Pembersihan Data:

a. Menangani data yang hilang (*missing values*) atau data anomali.

Pada beberapa atribut dalam dataset, ditemukan nilai dengan angka 0 yang tidak logis, misalnya pada variabel kadar glukosa, tekanan darah, atau ketebalan kulit. Nilai nol pada atribut-atribut tersebut dapat dianggap sebagai anomali atau bentuk missing values terselubung, karena secara medis tidak mungkin seseorang memiliki kadar glukosa atau tekanan darah bernilai nol. Jika tidak ditangani, kondisi ini dapat menurunkan kualitas data dan memengaruhi akurasi model klasifikasi [12].

Untuk mengatasi masalah tersebut, dilakukan proses imputasi median, yaitu mengganti nilai anomali atau missing values dengan nilai tengah (median) dari distribusi data pada atribut yang sama. Pemilihan median lebih tepat dibandingkan mean karena median lebih tahan terhadap pengaruh *outlier* dan mampu merepresentasikan nilai yang lebih realistis dari populasi data. Dengan cara ini, distribusi data menjadi lebih konsisten dan model KNN dapat melakukan perhitungan jarak dengan lebih akurat.



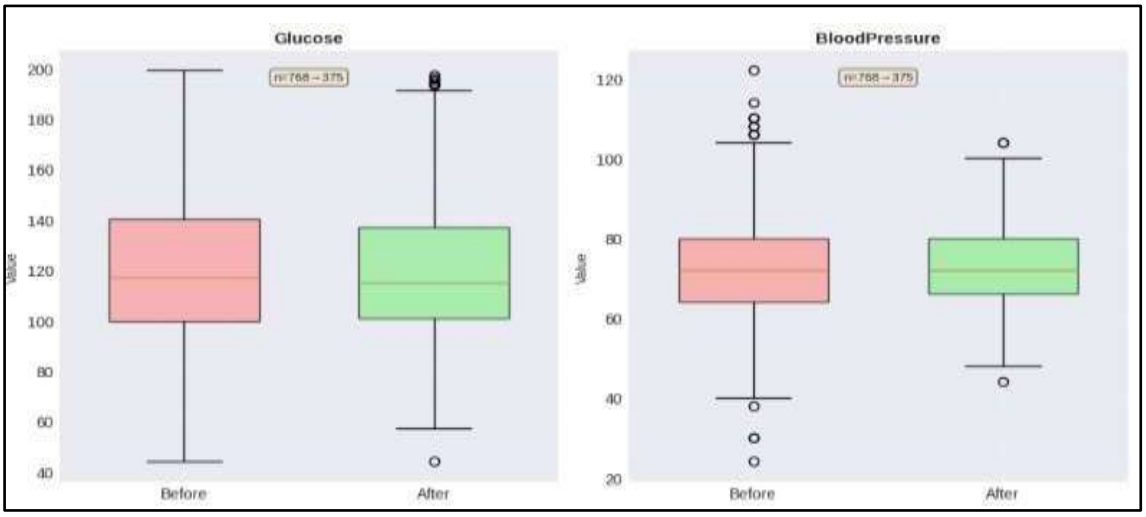
```
============================================================
STEP 1: HANDLING ZERO VALUES (Medical Impossibilities)
============================================================

[1] Zero Values Detection:
------------------------------------------------------------
Glucose                   :    5 zeros ( 0.65%)
BloodPressure             :   35 zeros ( 4.56%)
SkinThickness             :  227 zeros (29.56%)
Insulin                   :  374 zeros (48.70%)
BMI                       :   11 zeros ( 1.43%)

[2] Replacing zeros with median values...
   ✓ Glucose: filled with median = 117.00
   ✓ BloodPressure: filled with median = 72.00
   ✓ SkinThickness: filled with median = 29.00
   ✓ Insulin: filled with median = 125.00
   ✓ BMI: filled with median = 32.30
```

Gambar 2. Menangani data anomaly dengan imputasi median

b. Menangani nilai Outliers.

Pada beberapa atribut, terdapat data *outlier* yang merupakan nilai-nilai ekstrem, baik terlalu rendah maupun terlalu tinggi yang menyimpang jauh dari distribusi utama data lainnya, sehingga dapat mendistorsi analisis statistik dan performa model *machine learning* seperti KNN. Visualisasi boxplot efektif mendeteksi *outlier* melalui "whiskers" (batas Q1 - 1.5×IQR hingga Q3 + 1.5×IQR), di mana titik di luar rentang tersebut dianggap anomali. Penanganan dilakukan dengan IQR (*Interquartile Range*), yaitu selisih kuartil ketiga (Q3, persentil 75%) dan kuartil pertama (Q1, persentil 25%), memungkinkan *capping* (batasi ke batas whiskers) atau removal untuk pertahankan ukuran *dataset* sambil kurangi bias [13].

ISSN: 3110-0864 (Media Online)
JURNAL ILMU KOMPUTER DAN TEKNIK INFROMATIKA
Volume 2 No 1 Januari 2026 | pp: 119-130
https://journals.raskhamedia.or.id/index.php/juikti
DOI: https://doi.org/10.64803/juikti.v2i1.116

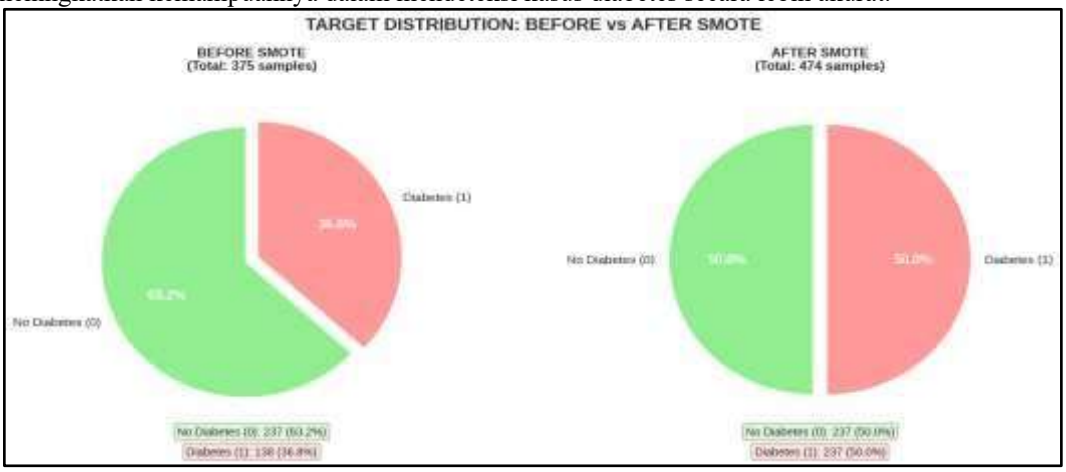


Gambar 3. Distribusi outlier sebelum dan setelah penanagan dengan IQR

c. Meratakan data label yang tidak seimbang

Dataset yang digunakan dalam penelitian ini memiliki masalah ketidakseimbangan jumlah data antar kelas, di mana jumlah data pasien non-diabetes jauh lebih banyak dibandingkan data pasien diabetes. Kondisi ini dapat menyebabkan model KNN lebih cenderung "belajar" dari kelas yang jumlahnya besar, sehingga sering mengabaikan atau salah mengenali kasus diabetes yang jumlahnya lebih sedikit. Akibatnya, meskipun hasil akurasi terlihat cukup tinggi, kemampuan model dalam mendeteksi pasien yang benar-benar menderita diabetes bisa menjadi kurang optimal. Oleh karena itu, penyeimbangan data menjadi langkah penting agar model tidak hanya fokus pada kelas mayoritas, tetapi juga mampu mengenali pola pada kelas minoritas dengan lebih baik dan menghasilkan prediksi yang lebih adil dan stabil.

Untuk mengatasi masalah tersebut, digunakan teknik SMOTE (Synthetic Minority Over-sampling Technique). Teknik ini bekerja dengan cara menambah data baru pada kelas minoritas secara buatan, bukan dengan menyalin data lama, tetapi dengan membuat data baru yang mirip berdasarkan karakteristik data yang sudah ada [14]. SMOTE membentuk data sintetis dengan melihat beberapa tetangga terdekat dari kelas minoritas, kemudian membuat contoh baru di antara data-data tersebut. Dengan cara ini, jumlah data antar kelas menjadi lebih seimbang [15], sehingga model KNN dapat belajar dari kedua kelas secara proporsional dan meningkatkan kemampuannya dalam mendeteksi kasus diabetes secara lebih akurat.



Gambar 4. Distribusi label sebelum dan setelah penanagan dengan SMOTE

2.5.2 Transformasi Data

Mengubah data ke format yang sesuai**.** Data mentah sering kali berbentuk tidak seragam, ukuran gambar bervariasi, tipe data campur aduk (teks, angka, tanggal), atau skala nilai berbeda antar kolom. Sehingga model *machine learning* seperti KNN tidak dapat memprosesnya secara efektif, menyebabkan error atau performa buruk. Transformasi format data merupakan tahap pra-pemrosesan krusial yang menstandardisasi struktur agar sesuai *input* model [16].

ISSN: 3110-0864 (Media Online)
JURNAL ILMU KOMPUTER DAN TEKNIK INFROMATIKA
Volume 2 No 1 Januari 2026 | pp: 119-130
https://journals.raskhamedia.or.id/index.php/juikti
DOI: https://doi.org/10.64803/juikti.v2i1.116

2.5.3 Normalisasi Data

Menyamakan rentang nilai dari setiap atribut (fitur) menggunakan metode seperti Normalisasi Min-Max. Langkah ini sangat penting untuk algoritma KNN, karena KNN berbasis jarak. Tanpa normalisasi, fitur dengan rentang nilai besar (misalnya, glukosa) akan mendominasi perhitungan jarak dibandingkan fitur dengan rentang nilai kecil (misalnya, BMI).

2.5.4 Split Data

Pembagian *dataset* menjadi data *training* (80%) dan *testing* (20%) merupakan praktik standar dalam pemodelan *machine learning*, termasuk algoritma KNN, untuk memastikan model belajar pola dari sebagian besar data sambil dievaluasi pada data tak terlihat. Proporsi ini memberi model cukup contoh pelatihan agar akurat, sekaligus cukup data uji untuk ukur performa nyata tanpa *overfitting* [17].

2.6. Implementasi KNN

Ini adalah tahap inti dari penelitian. Algoritma K-Nearest Neighbor diimplementasikan menggunakan bahasa pemrograman (dalam hal ini Python). Logika algoritma yang dibangun mencakup:

a. Fungsi untuk menghitung jarak antara data uji (input pengguna) dengan seluruh data latih di dataset, umumnya menggunakan metrik Jarak Euclidean**.**

b. Proses pengurutan data latih berdasarkan jarak terdekat**.**

c. Pengambilan 'K' tetangga terdekat (misalnya, K=5).

d. Proses *voting* (pemungutan suara mayoritas) dari 'K' tetangga tersebut untuk menentukan kelas prediksi (Risiko Tinggi atau Risiko Rendah).

2.7. Evaluasi Model

Setelah implementasi model KNN, tahap evaluasi dilakukan untuk mengukur tingkat keandalan dan generalisasi model terhadap data baru. Model kemudian dijalankan pada data uji untuk menghasilkan prediksi, yang dibandingkan dengan label aktual melalui *Confusion Matrix* guna menghitung metrik kinerja utama: akurasi (proporsi prediksi benar), presisi (akurasi prediksi positif), dan *recall* (kemampuan deteksi kasus positif), sehingga memberikan gambaran komprehensif tentang efektivitas model KNN dalam menangani pola data yang telah melalui pra-pemrosesan.

2.8. Perancangan Sistem Web

Model KNN yang telah terverifikasi kemudian diintegrasikan ke dalam aplikasi web melalui perancangan arsitektur *client-server*, dengan *frontend* berbasis React untuk antarmuka responsif dan *backend* Python Flask untuk eksekusi logika prediksi. Desain UI/UX difokuskan pada user-friendliness, mencakup formulir input gejala interaktif, tombol toggle mode gelap/terang, serta halaman hasil analisis yang menampilkan probabilitas diagnosis beserta rekomendasi pencegahan. Fungsionalitas sistem dirancang mengikuti user journey intuitif: halaman informasi edukasi (Gejala, Pencegahan, FAQ), proses analisis real-time via API call ke model KNN, dan navigasi seamless antar modul untuk mendukung pengambilan keputusan medis awal yang akurat bagi pengguna awam

2.9. Pengujian Sistem

Pengujian sistem web dilakukan secara komprehensif untuk memverifikasi integritas keseluruhan aplikasi, berbeda dari evaluasi model murni yang fokus pada performa algoritma. Verifikasi fungsional mencakup pengujian formulir input (validasi field wajib, handling input kosong), eksekusi prediksi (konsistensi hasil KNN dengan evaluasi offline), dan kontrol UI. Selain itu, pengujian error handling memastikan sistem tangguh terhadap *edge cases* seperti input ekstrem atau koneksi gagal, dengan *logging backend* untuk debugging, sehingga menjamin reliabilitas deployment produksi dan user experience optimal tanpa crash atau hasil *misleading.*

2.10. Analisis Hasil

Analisis hasil dilakukan secara biner terhadap performa model dan fungsionalitas sistem untuk mengevaluasi pencapaian tujuan penelitian. Analisis performa model menginterpretasikan metrik evaluasi KNN. Analisis fungsionalitas sistem menilai kesesuaian implementasi web terhadap rumusan masalah: apakah UI memenuhi usability heuristic (Nielsen), API response time <2 detik, dan coverage fitur 100% (*input-output*-prediksi), sehingga menyimpulkan bahwa sistem berhasil mengoperasionalkan model ML untuk aplikasi praktis dengan tingkat keberhasilan optimal.

ISSN: 3110-0864 (Media Online)
JURNAL ILMU KOMPUTER DAN TEKNIK INFROMATIKA
Volume 2 No 1 Januari 2026 | pp: 119-130
https://journals.raskhamedia.or.id/index.php/juikti
DOI: https://doi.org/10.64803/juikti.v2i1.116

2.11. Selesai

Ini Adalah tahap akhir penelitian, di mana sistem aplikasi web klasifikasi diabetes telah selesai dibangun, diuji, dan dianalisis. Peneliti kemudian menyusun laporan akhir atau skripsi berdasarkan seluruh tahapan yang telah dilalui.

3. HASIL DAN PEMBAHASAN

3.1 Implementasi Sistem

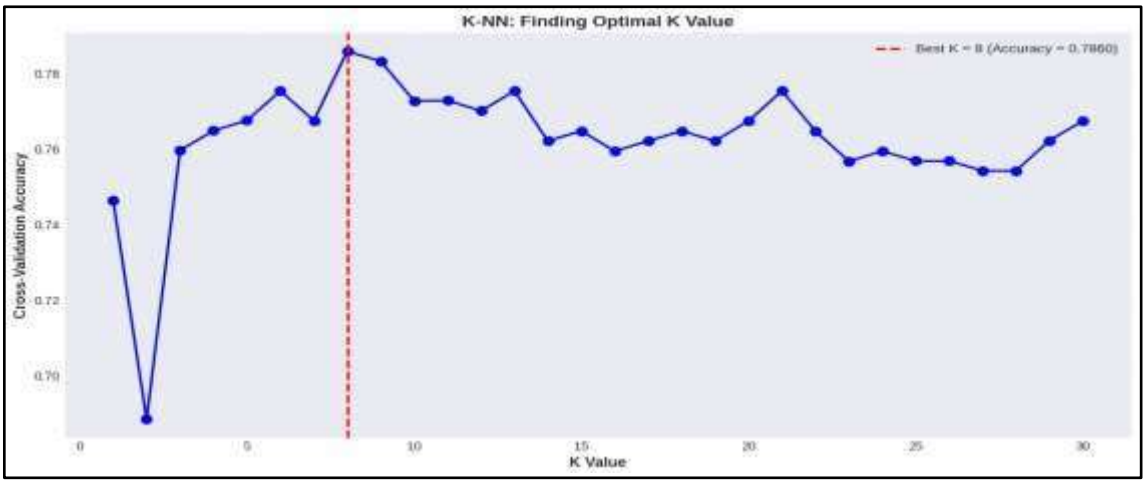
3.1.1 Arsitektur Sistem

Sistem klasifikasi risiko diabetes yang dibangun menggunakan arsitektur client–server berbasis web, di mana bagian client ditangani oleh React.js dan bagian server ditangani oleh Python dengan framework Flask yang mengintegrasikan model K-Nearest Neighbor (KNN). Pada sisi klien, React.js berfungsi sebagai antarmuka pengguna (UI) yang memungkinkan pengguna memasukkan data kesehatan melalui form interaktif dan melihat hasil prediksi secara real-time, sedangkan sisi server Flask menerima permintaan dari klien, menjalankan proses klasifikasi menggunakan model KNN, kemudian mengembalikan hasil ke frontend dalam format JSON.

3.1.2 Akurasi dan Evaluasi Pemodelan

Hasil evaluasi menunjukkan bahwa model K-Nearest Neighbor (KNN) mampu mencapai akurasi sebesar 82,06% pada data training dan 73,68% pada data testing. Perbedaan nilai akurasi ini mengindikasikan adanya overfitting tingkat sedang dengan selisih sebesar 0,0837, namun masih dalam batas wajar untuk model berbasis jarak seperti KNN. Akurasi pada data testing yang mendekati 74% menunjukkan bahwa model memiliki kemampuan generalisasi yang cukup baik dalam memprediksi status diabetes pada data yang belum pernah dilihat sebelumnya.

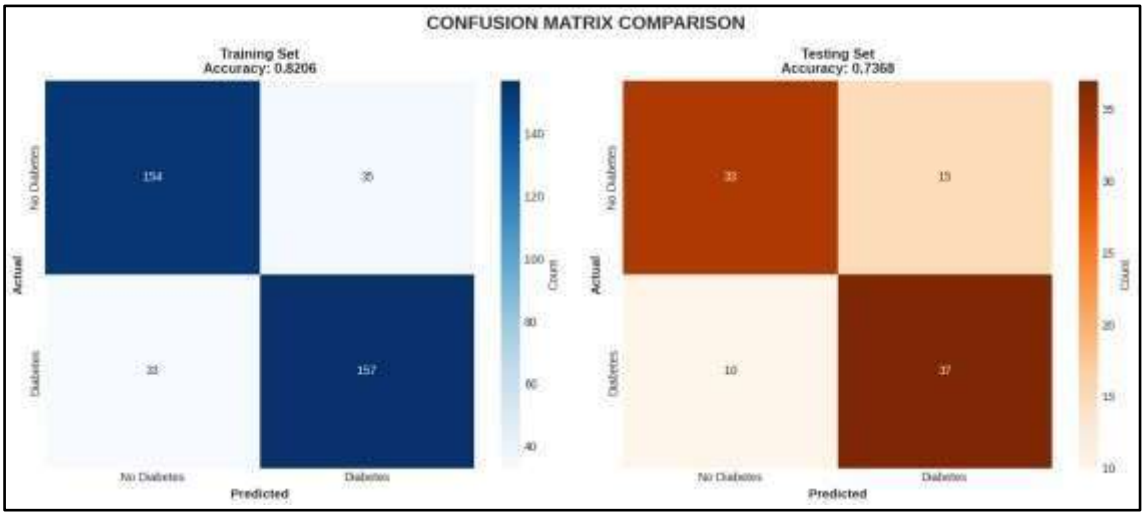
Pemilihan nilai K dilakukan menggunakan metode cross-validation dengan menguji nilai K dari 1 hingga 30. Berdasarkan grafik akurasi validasi silang, nilai akurasi tertinggi diperoleh pada K = 8 dengan skor sebesar 0,7860. Hal ini menunjukkan bahwa penggunaan 8 tetangga terdekat memberikan keseimbangan terbaik antara bias dan varians, sehingga mampu meningkatkan stabilitas dan performa model. Oleh karena itu, nilai K = 8 dipilih sebagai parameter optimal untuk proses pelatihan model KNN.



Gambar 5. Menemukan Nilai K yang paling tinggi yaitu terletak pada K8

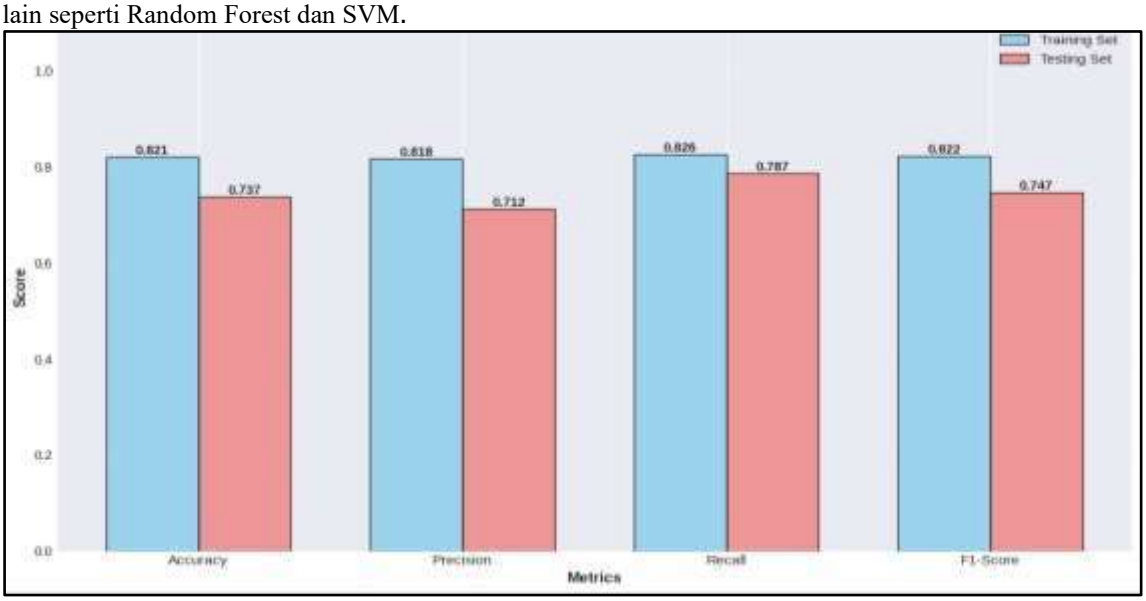
Confusion matrix pada data training menunjukkan bahwa model berhasil mengklasifikasikan 154 data non-diabetes dan 157 data diabetes secara benar, dengan kesalahan klasifikasi masing-masing sebanyak 35 dan 33 data. Sementara itu, pada data testing, model mengklasifikasikan dengan benar 33 data non-diabetes dan 37 data diabetes, dengan kesalahan prediksi sebanyak 15 data non-diabetes yang salah diprediksi sebagai diabetes dan 10 data diabetes yang salah diprediksi sebagai non-diabetes. Hasil ini menunjukkan bahwa model cenderung sedikit lebih sensitif dalam mendeteksi kasus diabetes dibandingkan dengan mendeteksi non-diabetes.

ISSN: 3110-0864 (Media Online)
JURNAL ILMU KOMPUTER DAN TEKNIK INFROMATIKA
Volume 2 No 1 Januari 2026 | pp: 119-130
https://journals.raskhamedia.or.id/index.php/juikti
DOI: https://doi.org/10.64803/juikti.v2i1.116



Gambar 6. Perbandingan Confusion Matrix untuk training dan untuk testing

Pada data testing, model menghasilkan precision sebesar 71,15%, recall sebesar 78,72%, dan F1-score sebesar 74,75%, serta nilai ROC-AUC sebesar 0,8132. Nilai recall yang relatif tinggi menunjukkan bahwa model cukup baik dalam mendeteksi pasien yang benar-benar menderita diabetes, sehingga cocok digunakan sebagai alat bantu skrining awal. Secara keseluruhan, performa model dikategorikan cukup baik (fair performance) dengan kemampuan klasifikasi yang seimbang antara kedua kelas, meskipun masih terdapat ruang untuk peningkatan melalui rekayasa fitur, tuning hiperparameter lanjutan, atau penggunaan algoritma lain seperti Random Forest dan SVM.



Gambar7. Klasifikasi peforma model dengan algoritma KNN

3.2. Antarmuka Web

3.2.1 Halaman Landing (Beranda)

Halaman landing merupakan halaman pertama yang dilihat pengguna saat mengakses sistem. Di bawah judul terdapat deskripsi singkat yang menjelaskan bahwa sistem menggunakan teknologi machine learning untuk menganalisis risiko diabetes dan memberikan rekomendasi pencegahan sejak dini. Pada bagian tengah halaman terdapat tombol "Mulai Analisis" berwarna cyan yang menonjol, memudahkan pengguna untuk langsung memulai proses klasifikasi risiko diabetes. Di bagian atas halaman terdapat navigation bar yang terdiri dari Gejala, Pencegahan, Evaluasi Model, FAQ, dan tombol "Cek Risiko" untuk akses cepat ke menu tertentu.

ISSN: 3110-0864 (Media Online)
JURNAL ILMU KOMPUTER DAN TEKNIK INFROMATIKA
Volume 2 No 1 Januari 2026 | pp: 119-130
https://journals.raskhamedia.or.id/index.php/juikti
DOI: https://doi.org/10.64803/juikti.v2i1.116

Gambar 8. Tampilan beranda Web

3.2.2 Halaman Pencegahan

Halaman ini berfungsi sebagai sumber edukasi yang memberikan penjelasan detail tentang gejala-gejala umum diabetes. Dengan judul "Gejala Umum Diabetes", halaman ini menyajikan enam kartu informasi yang disusun secara grid. Setiap kartu menjelaskan satu gejala spesifik beserta deskripsi singkat: Sering Haus & Lapar (polidipsia dan polifagia), Sering Buang Air Kecil (poliuria). Halaman ini membantu pengguna memahami kondisi kesehatan mereka sebelum menggunakan fitur analisis.

3.2.3 Halaman FaQ

Halaman FAQ menyediakan jawaban untuk pertanyaan umum seputar diabetes. Menggunakan desain accordion, pengguna dapat membuka dan menutup setiap pertanyaan sesuai kebutuhan. Beberapa pertanyaan yang tersedia meliputi:

a. Apakah diabetes bisa disembuhkan?
b. Apa perbedaan antara prediabetes dan diabetes?
c. Apakah penderita diabetes tidak boleh makan nasi sama sekali?
d. Apakah stres dapat mempengaruhi kadar gula darah?

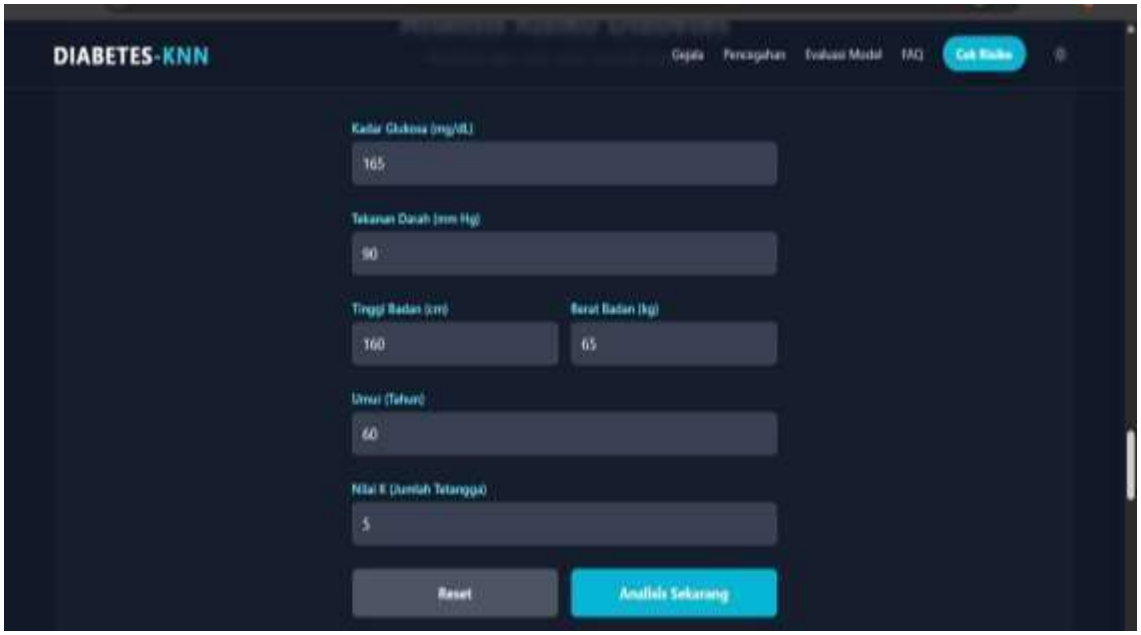
Fitur ini membantu pengguna mendapatkan informasi cepat tanpa harus mencari dari sumber lain.

3.2.4 Halaman Form Input Data

Halaman ini merupakan fitur utama sistem di mana pengguna memasukkan data kesehatan untuk mendapatkan hasil klasifikasi. Form dirancang dengan field input yang terstruktur dan mudah dipahami. Field input yang harus diisi pengguna meliputi:

a. Kadar Glukosa (mg/dL): Kadar glukosa darah
b. Tekanan Darah (mm Hg): Tekanan darah diastolik
c. Tinggi Badan (cm) dan Berat Badan (kg): Untuk perhitungan BMI
d. Umur (Tahun): Usia pengguna
e. Nilai K (Jumlah Tetangga): Parameter algoritma KNN dengan nilai default 5

ISSN: 3110-0864 (Media Online)
JURNAL ILMU KOMPUTER DAN TEKNIK INFROMATIKA
Volume 2 No 1 Januari 2026 | pp: 119-130
https://journals.raskhamedia.or.id/index.php/juikti
DOI: https://doi.org/10.64803/juikti.v2i1.116



Gambar 9. Tampilan menu *input*

3.2.5 Halaman Hasil Klasifikasi

Setelah data diproses, sistem menampilkan hasil klasifikasi dalam format visual yang informatif berupa circular progress bar disertai persentase keyakinan model dan label kategori risiko. Untuk kasus risiko tinggi diabetes, sistem menampilkan persentase keyakinan tinggi seperti 100% dengan label Berisiko Tinggi Diabetes. Penjelasan tambahan menyatakan bahwa model telah mengklasifikasikan data yang dimasukkan pengguna ke dalam kategori risiko tinggi, disertai informasi jumlah tetangga terdekat yang memiliki riwayat diabetes, misalnya 5 dari 5 tetangga terdekat memiliki riwayat diabetes. Informasi ini membantu pengguna memahami logika kerja algoritma KNN secara sederhana

Gambar 10. Tampilan hasil klasifikasi pada menu *input*

4. KESIMPULAN

Penelitian ini berhasil menjawab permasalahan utama terkait keterlambatan diagnosis diabetes mellitus dan keterbatasan akses masyarakat terhadap alat bantu deteksi dini yang praktis, dengan mengembangkan sistem klasifikasi berbasis web menggunakan algoritma K-Nearest Neighbor (KNN). Melalui penerapan tahapan prapemrosesan data yang sistematis, seperti penanganan nilai anomali, outlier, penyeimbangan distribusi kelas menggunakan SMOTE, serta normalisasi data, model yang dibangun mampu mempelajari pola data pasien secara lebih representatif dan tidak bias terhadap kelas mayoritas. Pemilihan nilai parameter K optimal melalui cross-validation juga terbukti meningkatkan stabilitas dan performa model dalam proses klasifikasi. Hasil pengujian menunjukkan bahwa sistem memiliki tingkat akurasi, kemampuan deteksi

ISSN: 3110-0864 (Media Online)
JURNAL ILMU KOMPUTER DAN TEKNIK INFROMATIKA
Volume 2 No 1 Januari 2026 | pp: 119-130
https://journals.raskhamedia.or.id/index.php/juikti
DOI: https://doi.org/10.64803/juikti.v2i1.116

kasus positif diabetes, serta kemampuan pemisahan kelas yang cukup baik, sehingga dapat diandalkan sebagai alat bantu skrining awal sebelum dilakukan pemeriksaan medis lanjutan. Integrasi model ke dalam sistem berbasis web dengan antarmuka yang informatif dan mudah digunakan turut menjawab permasalahan keterjangkauan teknologi bagi pengguna umum, karena sistem dapat diakses secara fleksibel tanpa memerlukan perangkat atau keahlian khusus. Dengan demikian, penelitian ini tidak hanya membuktikan bahwa algoritma KNN efektif diterapkan dalam klasifikasi diabetes mellitus, tetapi juga menunjukkan bahwa pemanfaatan teknologi pembelajaran mesin dalam bentuk sistem berbasis web memiliki potensi besar untuk mendukung upaya pencegahan dan pengendalian penyakit secara lebih dini, efisien, dan terjangkau. Meskipun masih terdapat ruang untuk peningkatan performa melalui pengembangan fitur, penambahan data, atau penggunaan metode klasifikasi lain, sistem yang dihasilkan telah mampu memberikan solusi awal yang relevan terhadap permasalahan diagnosis dini diabetes mellitus di masyarakat.

REFERENCES

[1] WHO, "Diabetes," World Healt Organization (WHO). Accessed: Jan. 12, 2026. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/diabetes

[2] IDF, "Indonesia," Federasi Diabetes Internasional. International Diabetes Federation. Accessed: Jan. 12, 2026. [Online]. Available: https://idf.org/our-network/regions-and-members/western-pacific/members/indonesia/

[3] D. A. Suryandari, L. Yunaini, D. Kristanty, and A. Prawiningrum, "Exploring Differentially Expressed Genes to Identify Biomarkers of Cervical Cancer : A Bioinformatics Approach," *Indones. J. Med. Chem. Bioinforma.*, vol. 4, no. 1, pp. 1–9, 2025, doi: 10.7454/ijmcb.v4i1.1045.

[4] E. F. Nopitasari, S. P. A. Alkadri, and R. W. S. Insani, "Implementasi Data Mining untuk Klasifikasi Penyakit Stroke Menggunakan Algoritma K Nearest Neighbor," *Kreat. J. Pengabdi. Masy. Nusant.*, vol. 5, no. 3, pp. 344–365, 2025, doi: https://doi.org/10.55606/kreatif.v5i3.8215.

[5] A. Gunawan and I. Fenriana, "Design of Diabetes Prediction Application Using K-Nearest Neighbor Algorithm," *Bit-Tech (Binary Digit. - Technol.*, vol. 6, no. 2, pp. 110–117, 2023, doi: 10.32877/bt.v6i2.939.

[6] G. F. Ibanez, G. W. Wiriasto, and Rosmaliati, "Kombinasi Principal Component Analysis dengan Algoritma K-Means untuk Klasterisasi Data Stunting," *KLIK Kaji. Ilm. Inform. dan Komput.*, vol. 5, no. 1, pp. 131–141, 2024, doi: 10.30865/klik.v5i1.1977.

[7] A. Oktaviana, D. P. Wijaya, A. Pramuntadi, and D. Heksaputra, "Prediction of Type 2 Diabetes Mellitus Using The K-Nearest Neighbor (K-NN) Algorithm Prediksi Penyakit Diabetes Melitus Tipe 2 Menggunakan Algoritma K-Nearest Neighbor (K-NN)," *Inst. Ris. dan Publ. Indones.*, vol. 4, no. 3, pp. 812–818, 2024, doi: https://doi.org/10.57152/malcom.v4i3.1268.

[8] R. Pratama, A. M. Siregar, S. A. P. Lestari, and S. Faisal, "IMPLEMENTATION OF DIABETES PREDICTION MODEL USING RANDOM IMPLEMENTASI MODEL PREDIKSI DIABETES MENGGUNAKAN ALGORITMA," *J. Tek. Inform.*, vol. 5, no. 4, pp. 1165–1174, 2024, doi: https://doi.org/10.52436/1.jutif.2024.5.4.2593.

[9] A. S. Firmansyah, A. Aziz, and M. Ahsan, "OPTIMASI K-NEAREST NEIGHBOR MENGGUNAKAN ALGORITMA SMOTE UNTUK MENGATASI IMBALANCE CLASS PADA KLASIFIKASI ANALISIS SENTIMEN," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 7, no. 6, 2023.

[10] U. M. Learning and K. T. Bot, "Pima Indians Diabetes Database," kaggle. Accessed: Jan. 21, 2026. [Online]. Available: https://www.kaggle.com/datasets/uciml/pima-indians-diabetes-database

[11] R. R. Hallan and I. N. Fajri, "Prediksi Harga Rumah menggunakan Machine Learning Algoritma Regresi Linier," *J. Teknol. Dan Sist. Inf. Bisnis*, vol. 7, no. 1, pp. 57–62, 2025, doi: https://doi.org/10.47233/jteksis.v7i1.1732.

[12] M. R. A. Prasetya, A. M. Priyatno, and Nurhaeni, "Jurnal Informasi dan Teknologi Penanganan Imputasi Missing Values pada Data Time Series dengan Menggunakan Metode Data Mining," *J. Inf. dan Teknol.*, vol. 5, no. 2, pp. 56–62, 2023, doi: 10.37034/jidt.v5i1.324.

[13] A. Setiawan, Arnita, D. Yusuf, N. Syafira, and T. Tania, "ANALISIS DESKRIPTIF UMP (UPAH MINIMUM

ISSN: 3110-0864 (Media Online)
JURNAL ILMU KOMPUTER DAN TEKNIK INFROMATIKA
Volume 2 No 1 Januari 2026 | pp: 119-130
https://journals.raskhamedia.or.id/index.php/juikti
DOI: https://doi.org/10.64803/juikti.v2i1.116

PROVINSI) SEINDONESIA (2002-2022) MENGGUNAKAN METODE FUZZY C MEANS CLUSTERING," *J. Ilmu Komput. Revolusioner*, vol. 8, no. 12, pp. 1–13, 2024.

[14] F. D. Astuti and F. N. Lenti, "Implementasi SMOTE untuk mengatasi," *J. JUPITER*, vol. 13, no. 1, pp. 89–98, 2021.

[15] Ridwan, E. H. Hermaliani, and M. Ernawati, "Penerapan Metode SMOTE Untuk Mengatasi Imbalanced Data Pada Klasifikasi Ujaran Kebencian," *Comput. Sci.*, vol. 4, no. 1, pp. 80–88, 2024.

[16] T. Gori, A. Sunyoto, and H. Al Fatta, "PREPROCESSING DATA DAN KLASIFIKASI UNTUK PREDIKSI KINERJA AKADEMIK SISWA," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 11, no. 1, pp. 215–224, 2024, doi: 10.25126/jtiik.20241118074.

[17] B. N. Azmi, A. Hermawan, and D. Avianto, "Analisis Pengaruh Komposisi Data Training dan Data Testing pada Penggunaan PCA dan Algoritma Decision Tree untuk Klasifikasi Penderita Penyakit Liver," *JTIM J. Teknol. Inf. dan Multimed.*, vol. 4, no. 4, pp. 281–290, 2023, doi: https://doi.org/10.35746/jtim.v4i4.298.