Journal of System & Technology
ISSN: 3110-0171
Vol. 1, No. 2, Desember 2025

ANALISIS PERBANDINGAN KINERJA ALGORITMA K-NEAREST NEIGHBORS DAN SUPPORT VECTOR MACHINE UNTUK KLASIFIKASI PENYAKIT DIABETES

Hidayat Hatta Irsyad[1], Muhammad Ikran Syafwan[1], Dian Ramadhani[1]*
[1]Dept.of Informatics Engineering, Faculty of Engineering, Universitas Riau, Pekanbaru,Indonesia
hidayathatta834@gmail.com, ikram.syafwan@gmail.com email, dianramadhani@lecturer.unri.ac.id
*Corresponding author: dianramadhani@lecturer.unri.ac.id

Abstract: Penyakit diabetes merupakan masalah kesehatan global yang jumlah penderitanya terus meningkat. Deteksi dini sangat penting untuk mencegah komplikasi serius, namun metode diagnosis konvensional seringkali membutuhkan waktu dan sumber daya besar. Penelitian ini bertujuan membandingkan kinerja algoritma K-Nearest Neighbors (K-NN) dan Support Vector Machine (SVM) untuk klasifikasi penyakit diabetes. Metode yang diterapkan mencakup preprocessing, penanganan imbalance data dengan SMOTE, serta normalisasi fitur pada dataset Healthcare-Diabetes. Hasil evaluasi menunjukkan bahwa algoritma K-NN memiliki performa lebih baik dengan akurasi 92,20% dan F1-score 0,93, mengungguli SVM yang mencapai akurasi 88,39% dan F1-score 0,89. Dengan demikian, K-NN terbukti lebih optimal dalam mendeteksi kasus diabetes pada dataset ini. Untuk pengembangan selanjutnya, disarankan melakukan penyetelan hyperparameter dan eksplorasi teknik seleksi fitur guna meningkatkan akurasi serta efisiensi model.

Keywords—Akurasi, Diabetes, Klasifikasi, K-NN, SVM

I. PENDAHULUAN

Diabetes Melitus merupakan penyakit kronis yang ditandai oleh kadar gula darah tinggi dan berisiko menimbulkan komplikasi serius seperti penyakit jantung dan stroke. World Health Organization (WHO) mencatat bahwa pada tahun 2021, lebih dari 10% populasi dewasa dunia mengidap diabetes, dan hampir setengahnya tidak menyadari kondisinya [1]. Deteksi dini menjadi faktor penting dalam menurunkan risiko komplikasi tersebut. Seiring perkembangan teknologi, pemanfaatan machine learning dalam sistem pendukung diagnosis penyakit menjadi pendekatan yang menjanjikan [2].

Algoritma klasifikasi seperti K-Nearest Neighbor (K-NN) dan Support Vector Machine (SVM) merupakan dua metode yang banyak digunakan dalam klasifikasi penyakit diabetes. K-NN bekerja berdasarkan kemiripan data berdasarkan jarak terdekat, sedangkan SVM membangun hyperplane optimal untuk memisahkan kelas [3]. Berbagai studi telah membandingkan kinerja kedua algoritma ini dengan hasil yang beragam. Penelitian oleh Mulyani et al. menunjukkan bahwa K-NN dengan k=3 menghasilkan akurasi 83,33% dalam klasifikasi diabetes, lebih tinggi dari SVM dengan kernel RBF yang menghasilkan 81,67% [4]. Namun, penelitian lain justru menunjukkan performa lebih tinggi pada SVM dibanding K-NN saat menggunakan teknik oversampling dan data preprocessing tertentu [5].

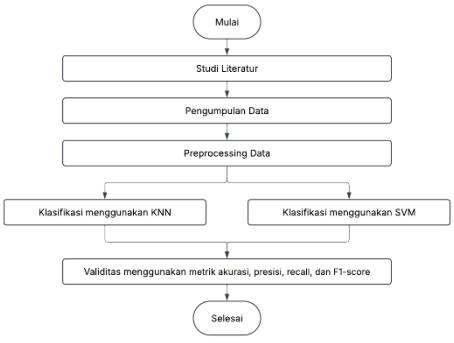
Selain itu, Alzboon et al. menekankan bahwa meskipun algoritma kompleks seperti neural networks menunjukkan performa tinggi, metode seperti SVM dan KNN tetap kompetitif dan efektif dalam deteksi dini diabetes, khususnya saat digunakan dalam sistem skrining berbasis data historis [6]. Dengan kemampuan mengolah data dari perangkat medis dan rekam kesehatan elektronik, machine learning menjadi tulang punggung dalam membangun sistem klasifikasi prediktif yang akurat dan dapat diandalkan.

Lebih lanjut, Tanveer et al. menjelaskan bahwa machine learning kini digunakan tidak hanya untuk diagnosis, tetapi juga untuk memberikan rekomendasi pengobatan personal yang disesuaikan dengan data fisiologis dan gaya hidup pasien [7]. Pendekatan ini mendorong munculnya model manajemen diabetes yang bersifat individual dan adaptif, sekaligus meningkatkan kepatuhan pasien terhadap pengobatan.

Melihat perbedaan hasil tersebut, penelitian ini dilakukan untuk membandingkan kembali performa algoritma K-NN dan SVM dalam mengklasifikasikan penyakit diabetes menggunakan dataset Healthcare-Diabetes. Evaluasi dilakukan menggunakan metrik akurasi, presisi, recall, dan f1-score untuk menentukan metode yang paling efektif dan efisien sebagai sistem pendukung diagnosis dini.

II. METODOLOGI PENELITIAN

Terdapat tujuh tahapan utama yang dilakukan dalam penelitian ini, yakni studi literatur, pengumpulan data, preprocessing data, klasifikasi menggunakan algoritma K-Nearest Neighbor (K-NN), klasifikasi menggunakan algoritma Support Vector Machine (SVM), validasi hasil menggunakan metrik evaluasi berupa akurasi, presisi, recall, dan F1-score, serta penyimpulan hasil penelitian. Metode penelitian dapat dilihat pada Gambar 1.



Gambar 1. Metode Penelitian

Dikirim	Revisi	Selesai Revisi	Diterima
21-Jul-2025	10-Nov-2025	25-Nov-2025	25-Nov-2025

A. Data Mining

Data mining adalah proses untuk menemukan pola-pola tersembunyi atau pengetahuan baru dari kumpulan data besar melalui teknik statistik, pemrograman, atau kecerdasan buatan [8][8]. Dalam bidang kesehatan, khususnya prediksi penyakit seperti diabetes, data mining digunakan untuk mengolah data rekam medis guna menghasilkan informasi yang berguna bagi diagnosa [8], [9]. Teknik ini juga merupakan bagian penting dari proses Knowledge Discovery in Databases (KDD) yang melibatkan tahapan seleksi data, praproses, transformasi, hingga evaluasi hasil [8].

Model data mining seperti klasifikasi telah diterapkan dalam berbagai penelitian prediksi diabetes, menunjukkan bahwa metode ini mampu meningkatkan akurasi dan efisiensi sistem pendukung Keputusan [10]. Selain itu, pemanfaatan data publik seperti Pima Indian Diabetes Dataset juga sering digunakan untuk membangun model prediksi yang kuat [9].

B. Classification

Klasifikasi merupakan salah satu metode utama dalam data mining yang bertujuan untuk memetakan data ke dalam kategori atau kelas tertentu berdasarkan nilai atribut-atribut input [11]. Dalam konteks medis, klasifikasi dapat digunakan untuk menentukan apakah pasien beresiko mengidap diabetes berdasarkan atribut seperti usia, tekanan darah, kadar glukosa, dan indeks massa tubuh [8], [11].

Beberapa algoritma klasifikasi yang populer dalam prediksi diabetes adalah Naïve Bayes, K-Nearest Neighbor (KNN), dan Support Vector Machine (SVM). Penelitian Hidayatunnisa'i et al. menunjukkan bahwa metode SVM mampu menghasilkan akurasi hingga 97% dalam klasifikasi soal atau data medis, mengungguli Naïve Bayes yang hanya mencapai 95% [11]. Metode klasifikasi juga dievaluasi menggunakan metrik akurasi, presisi, recall, dan F1-score, yang semuanya dapat diperoleh dari confusion matrix [12].

C. K-Nearest Neighbors

Algoritma K-Nearest Neighbor (KNN) adalah metode klasifikasi yang bekerja berdasarkan jarak terdekat antara data uji dan data latih. KNN termasuk dalam kategori lazy learner karena tidak membangun model eksplisit pada saat pelatihan, melainkan melakukan klasifikasi langsung saat ada data baru [9], [10].

Secara formal, untuk mengklasifikasikan sebuah titik data *x* menggunakan KNN dengan *K* tetangga, langkah-langkahnya adalah sebagai berikut:

Menghitung Jarak: Hitung jarak antara titik data x dan semua titik data dalam set pelatihan. Jarak Euclidean umumnya digunakan sebagai metrik jarak:

$$d(x, x_i) = \sqrt{\sum_{j=1}^{n} (x_j - x_{ij})^2} \quad (1)$$

i mana $x=(x_1,x_2,...,x_n)$ adalah titik data yang akan diklasifikasikan, dan $x_i=(x_{i1},x_{i2},...,x_{in})$ adalah titik data ke-*i* dalam set pelatihan dengan *n* fitur. Kemudian mengidentifikasi tetangga terdekat, yaitu memilih *K* titik data dalam set pelatihan yang memiliki jarak terkecil ke *x*. Dalam penelitian ini, nilai *K* ditetapkan sebesar 5. Selanjutnya menentukan kelas prediksi, yaitu menentukan kelas dari titik data *x* berdasarkan kelas mayoritas dari *K* tetangga terdekat yang telah diidentifikasi. Jika ada seri (misalnya, jumlah tetangga dari dua kelas berbeda sama), strategi tie-breaking tertentu (yang diimplementasikan dalam scikit-learn) akan digunakan untuk menentukan kelas akhir. Dalam implementasinya untuk prediksi diabetes, KNN terbukti memberikan hasil akurasi tinggi, misalnya sebesar 96% dengan nilai k=23 [10].

Penelitian lain menunjukkan bahwa normalisasi data menggunakan metode Z-Score atau Min-Max dapat meningkatkan akurasi model KNN [9]. Evaluasi yang dilakukan pada dataset diabetes menghasilkan variasi nilai k optimal tergantung pada metode normalisasi yang digunakan. Sebagai contoh, dengan Min-Max normalization dan k=3, akurasi mencapai 74% [9].

D. Support Vector Machine (SVM)

Support Vector Machine (SVM) adalah algoritma klasifikasi yang bertujuan mencari hyperplane terbaik untuk memisahkan dua kelas data secara maksimal [11], [13]. SVM sangat efektif digunakan untuk data berdimensi tinggi dan dapat menangani kasus klasifikasi linear maupun non-linear melalui pemanfaatan fungsi kernel [13]. Klasifikasi dilakukan menggunakan model Support Vector Machine (SVM) dari pustaka scikit-learn. Model SVM bertujuan untuk menemukan hyperplane optimal yang memisahkan data antar kelas. Fungsi keputusan untuk SVM dengan kernel didefinisikan sebagai berikut:

$$f(x) = \sum_{i=1}^{n_s} \alpha_i \, y_i K(x_i, x) + b \quad (2)$$

di mana *x* adalah vektor fitur input, x_i adalah support vector, n_s adalah jumlah support vector, α_i adalah bobot yang terkait dengan setiap support vector, $y_i \in \{+1, -1\}$ adalah label kelas dari support vector, $K(x_i, x)$ adalah fungsi kernel, dan *b* adalah istilah bias. Kernel Radial Basis Function (RBF) digunakan dalam penelitian ini.

Dalam konteks prediksi penyakit diabetes, SVM telah dibandingkan dengan berbagai algoritma lain dan menunjukkan hasil akurasi yang kompetitif. Misalnya, dalam penelitian oleh Yang [14], SVM memperoleh area under curve (AUC) sebesar 0.79, menunjukkan performa klasifikasi yang layak dibandingkan model lainnya seperti Logistic Regression atau XGBoost [14]. Sementara itu, penelitian lain yang menggunakan validasi silang 10-fold menemukan bahwa kernel linear memberikan performa terbaik pada dataset diabetes [13], [14].

Penelitian yang dilakukan oleh Lumbanraja et al. mengimplementasikan SVM untuk klasifikasi penderita diabetes dan menghasilkan performa akurasi yang sangat baik ketika diuji dengan data pasien dari rumah sakit [15]. Evaluasi tersebut menggunakan metode validasi silang dan menunjukkan bahwa SVM dapat diterapkan secara andal

dalam lingkungan dunia nyata, seperti klinik atau sistem rekam medis.

Sementara itu, dalam penelitian lain yang membandingkan model SVM dan Binary Logistic Regression dalam klasifikasi penderita diabetes, SVM kembali unggul dalam hal akurasi dan efektivitas prediksi, yaitu sebesar 97.75% [16]. Hasil ini menunjukkan bahwa SVM tidak hanya unggul secara teoritis, tetapi juga dalam praktiknya lebih stabil dibandingkan model statistik tradisional.

Penelitian oleh Pinandito et al. lebih lanjut menunjukkan bahwa penerapan machine learning termasuk SVM dalam skrining risiko tinggi diabetes melitus pada kehamilan mampu membantu tenaga medis untuk melakukan deteksi dini [17]. Hal ini penting karena faktor hormonal selama kehamilan dapat meningkatkan kompleksitas diagnosis, dan penggunaan algoritma klasifikasi otomatis dapat meningkatkan efisiensi serta akurasi deteksi).

E. Confusion Matrix

Confusion matrix adalah representasi tabular dari hasil prediksi model klasifikasi yang terdiri atas empat elemen utama: True Positive (TP), False Positive (FP), True Negative (TN), dan False Negative (FN) [12]. Matriks ini digunakan untuk menghitung berbagai metrik evaluasi seperti akurasi, presisi, recall, dan F1-score, yang penting dalam menilai kinerja algoritma klasifikasi.

Dalam penelitian oleh Vamsikrishna et al., metrik ini digunakan untuk membandingkan performa metode SVM dan Decision Tree dalam model ensemble, yang menghasilkan akurasi hingga 97% [12]. Confusion matrix juga digunakan untuk membandingkan performa KNN dan Naïve Bayes terhadap dataset yang berbeda ukuran misalnya, pada dataset 2000 pasien, KNN mencapai 99% akurasi sementara Naïve Bayes hanya 75% [8]. Evaluasi semacam ini menunjukkan pentingnya alat ukur objektif dalam validasi model klasifikasi medis.

Studi yang dilakukan oleh AlShammari menunjukkan bagaimana confusion matrix diimplementasikan dalam Python untuk mengevaluasi performa model secara sistematis. Penelitian ini menekankan pentingnya penghitungan akurasi, presisi, recall, dan F1-score secara terprogram, yang memungkinkan analisis model secara efisien menggunakan data aktual dan hasil prediksi [18].

Sementara itu, Swaminathan dan Tantri dalam studi yang diterbitkan di African Journal of Biomedical Research membahas secara mendalam tentang metrik evaluasi berbasis confusion matrix, termasuk AUC, ROC, G-mean, dan balanced accuracy. Mereka menyimpulkan bahwa confusion matrix adalah pusat dari hampir seluruh metrik klasifikasi dan memiliki peran penting dalam menyesuaikan model terhadap karakteristik dataset, terutama ketika menangani data yang tidak seimbang [19].

III. HASIL DAN PEMBAHASAN

Bagian ini menyajikan hasil pengujian dan pembahasan terhadap performa algoritma K-Nearest Neighbors (K-NN) dan Support Vector Machine (SVM) dalam mengklasifikasikan penyakit diabetes berdasarkan dataset Healthcare-Diabetes. Setiap tahapan, mulai dari pengambilan data, preprocessing, hingga evaluasi model, dijelaskan secara rinci untuk memberikan gambaran menyeluruh terhadap proses dan hasil penelitian

A. Data Retrieval

Data yang digunakan dalam penelitian ini berasal dari dataset Healthcare Diabetes yang tersedia di platform Kaggle. Dataset ini disusun oleh Nandita Pore dan dirancang untuk mendukung analisis risiko diabetes serta pengembangan model prediksi berbasis pembelajaran mesin. Dataset ini berisi satu file dalam format CSV yang mencakup berbagai atribut medis yang relevan untuk klasifikasi diabetes. Beberapa fitur utama dalam dataset ini meliputi Glucose, BloodPressure, BMI, Age, dan Outcome (Label target yang menunjukkan apakah pasien menderita diabetes (1) atau tidak (0)). Atribut-atribut yang tersedia dalam dataset ini memberikan dasar yang kuat untuk membangun dan mengevaluasi model klasifikasi.

Dengan menggunakan data medis yang representatif, penelitian ini dapat menguji dan membandingkan efektivitas algoritma K-Nearest Neighbors dan Support Vector Machine dalam mengidentifikasi pasien yang berisiko menderita diabetes. Dataset ini juga memungkinkan analisis performa model secara objektif menggunakan matrik evaluasi standar dalam pembelajaran mesin.

Tabel 1. Datasets of Healthcare Diabetes.

z	Pregnancies	Glucose	…	Outcome
1	6	148	…	1
2	1	85	…	0
3	8	183	…	1
4	1	89	…	0
5	0	137	…	1
…	…	…	…	…
2768	2	81		0

B. Pre-Processing Data

1) Data Selection
   Pada tahap ini, data yang digunakan diperoleh dari sumber yang relevan dan memiliki atribut yang mendukung proses klasifikasi diabetes. Data terdiri dari beberapa fitur seperti jumlah kehamilan, kadar glukosa, tekanan darah, ketebalan kulit, kadar insulin, indeks massa tubuh (BMI), fungsi keturunan diabetes, dan usia
2) Cleaning Data
   Proses pembersihan data melibatkan penghapusan kolom yang tidak memiliki kontribusi terhadap proses klasifikasi dan juga penghapusan outlier. Kolom “Id” dihapus karena hanya berfungsi sebagai identifikasi dan tidak mengandung informasi penting untuk pelatihan model. Selain itu, dilakukan proses iterative outlier removal menggunakan metode Interquartile Range (IQR) pada fitur numerik seperti Pregnancies, Glucose, BloodPressure, dan lainnya. Nilai yang berada di luar rentang normal ($Q1 - 1.5 \times IQR$ hingga $Q3 - 1.5 \times IQR$) dihapus secara berulang hingga tidak ditemukan

lagi outlier atau batas maksimum iterasi tercapai. Tujuannya adalah menghilangkan data ekstrem yang dapat mengganggu akurasi model klasifikasi.

3) Handling Data Imbalance
Distribusi label dalam dataset menunjukkan bahwa jumlah data antara penderita diabetes (kelas 1) dan non-penderita (kelas 0) tidak seimbang. Untuk mengatasi hal ini digunakan teknik SMOTE (Synthetic Minority Over-sampling Technique) yang menghasilkan data sintetis pada kelas minoritas.
4) Feature and Label Separation
Data dibagi menjadi dua bagian: fitur (X) dan label (y). Kolom "Outcome" digunakan sebagai label target yang menunjukkan status diabetes, sedangkan kolom lainnya menjadi fitur input bagi model klasifikasi.
5) Feature Normalization
Untuk memastikan distribusi fitur seragam dan meningkatkan performa algoritma seperti SVM yang sensitif terhadap skala data, dilakukan proses normalisasi menggunakan StandardScaler.
6) Train and Testing Data Splitting
Dataset yang telah diseimbangkan dan dinormalisasi kemudian dibagi menjadi data latih dan data uji dengan perbandingan 80:20. Parameter stratify digunakan untuk menjaga proporsi kelas pada masing-masing subset agar representative.

Tabel 2. Cleaned Datasets of Healthcare Diabetes

z	Pregnancies	Glucose	…	Outcome
1	6	148	…	1
2	1	85	…	0
4	1	89	…	0
6	5	116	…	0
11	3	78	…	1
…	…	…	…	…
1943	2	81		0

C. *Classification with the K-Nearest Neighbors Algorithm*

Berdasarkan hasil proses klasifikasi, algoritma K-Nearest Neighbors (K-NN) menunjukkan performa yang sangat baik dengan tingkat akurasi sebesar 92,20%. Nilai precision untuk kelas 0 adalah 0.98, sedangkan untuk kelas 1 sebesar 0.88. Sementara itu, nilai recall untuk kelas 0 adalah 0.86, dan untuk kelas 1 sebesar 0.98.

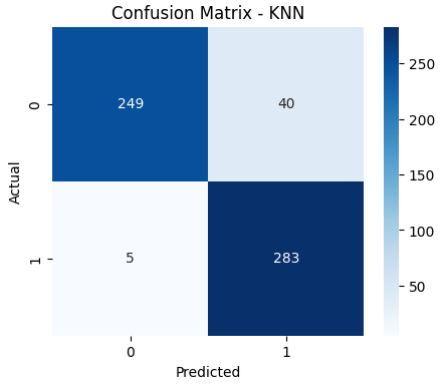
```
--- KNN ---
Accuracy: 0.9220103986135182
Confusion Matrix:
 [[249  40]
 [  5 283]]
Classification Report:
              precision    recall  f1-score   support

           0       0.98      0.86      0.92       289
           1       0.88      0.98      0.93       288

    accuracy                           0.92       577
   macro avg       0.93      0.92      0.92       577
weighted avg       0.93      0.92      0.92       577
```

Gambar 2. KNN Classification Model Evaluation

Rata-rata nilai F1-score yang diperoleh sebesar 0.92, mengindikasikan bahwa algoritma K-NN sangat efektif dan seimbang dalam aspek presisi maupun sensitivitas terhadap kedua kelas. Hasil ini menunjukkan bahwa K-NN mampu mengidentifikasi data penderita diabetes dengan baik serta menjaga keseimbangan performa klasifikasi antar kelas.



Gambar 3. Confusion Matrix KNN

D. *Classification with the SVM Algorithm*

Hasil proses klasifikasi menggunakan algoritma Support Vector Machine (SVM) menunjukkan bahwa model mampu mencapai akurasi sebesar 88,38%. Nilai evaluasi performa model juga menunjukkan hasil yang cukup baik, dengan nilai precision dan recall masing-masing sebesar 0.93 dan 0.83 untuk kelas 0, serta 0.85 dan 0.94 untuk kelas 1. Rata-rata F1-score yang dihasilkan dari model ini adalah 0.88, menunjukkan keseimbangan antara precision dan recall yang memadai.

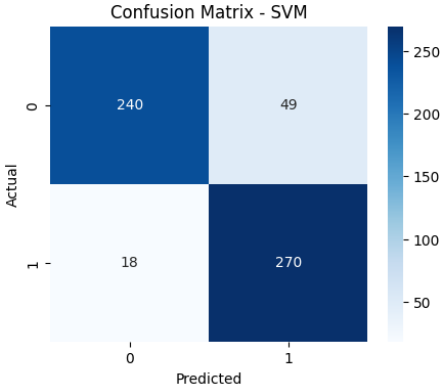
```
--- SVM ---
Accuracy: 0.8838821490467937
Confusion Matrix:
 [[240  49]
 [ 18 270]]
Classification Report:
              precision    recall  f1-score   support

           0       0.93      0.83      0.88       289
           1       0.85      0.94      0.89       288

    accuracy                           0.88       577
   macro avg       0.89      0.88      0.88       577
weighted avg       0.89      0.88      0.88       577
```

Gambar 4. SVM Classification Model Evaluation

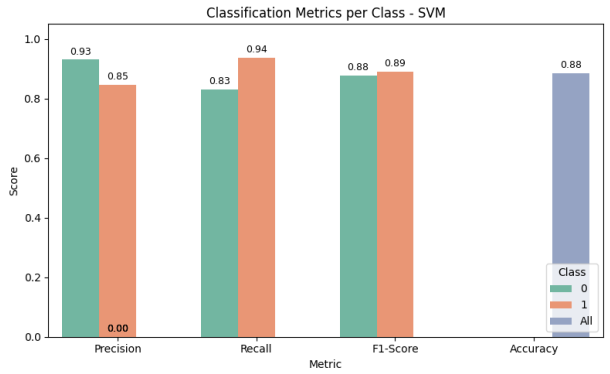
Berdasarkan analisis confusion matrix, model berhasil mengklasifikasikan dengan benar sebanyak 240 data pada kelas 0 (non-diabetes) dan 270 data pada kelas 1 (diabetes), sementara terdapat sebanyak 67 data yang salah diklasifikasikan.



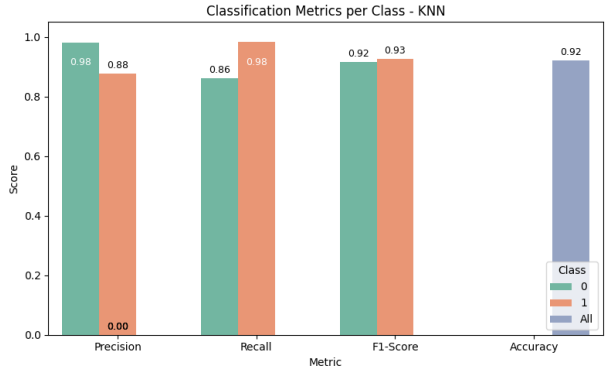
Gambar 5. Confusion Matrix SVM

E. *Comparison of svm and knn algorithms*

Perbandingan antara algoritma Support Vector Machine (SVM) dan K-Nearest Neighbors (K-NN) ditampilkan secara visual pada Gambar 6 (SVM) dan Gambar 7 (K-NN). Kedua gambar menunjukkan performa masing-masing algoritma dalam hal precision, recall, F1-score per kelas, serta akurasi keseluruhan.



Gambar 6. Hasil evaluasi SVM untuk setiap kelas diabetes dan non diabetes



Gambar 7. Hasil evaluasi SVM untuk setiap kelas diabetes dan non diabetes

Berdasarkan Gambar 7, algoritma K-NN secara konsisten menunjukkan nilai metrik yang lebih tinggi dibandingkan dengan SVM (Gambar 1). Pada metrik precision, K-NN mencatat nilai sebesar 0,98 untuk kelas 0 dan 0,88 untuk kelas 1, sedangkan SVM hanya mencapai 0,93 dan 0,85. Pada metrik recall, K-NN mencatat 0,86 (kelas 0) dan 0,98 (kelas 1), lebih baik dibandingkan SVM yang memperoleh 0,83 dan 0,94. Nilai F1-score untuk K-NN juga lebih tinggi dengan 0,92 pada kelas 0 dan 0,93 pada kelas 1, sedangkan SVM mencatat 0,88 dan 0,89.

Sedangkan sisi akurasinya, K-NN mencatat skor akurasi keseluruhan sebesar 0,92 (Gambar 7), sedangkan SVM mencatat 0,88 (Gambar 6). Hal ini menunjukkan bahwa K-NN tidak hanya lebih akurat secara keseluruhan, tetapi juga lebih seimbang dalam hal presisi dan sensitivitas terhadap kedua kelas, terutama kelas positif (penderita diabetes).

Berdasarkan analisis visual dan metrik evaluasi dari Gambar 6 dan Gambar 7, algoritma K-NN menunjukkan performa yang lebih unggul dibandingkan SVM dalam klasifikasi penyakit diabetes pada dataset ini.

IV. KESIMPULAN

Berdasarkan hasil pengujian dan evaluasi terhadap algoritma K-Nearest Neighbors (K-NN) dan Support Vector Machine (SVM) dalam klasifikasi penyakit diabetes, diperoleh beberapa temuan penting. Algoritma K-NN menunjukkan kinerja terbaik dengan akurasi sebesar 92,20%, serta nilai presisi, recall, dan F1-score yang tinggi, khususnya pada kelas positif (penderita diabetes). Model K-NN menghasilkan recall sebesar 0,98 dan F1-score sebesar 0,93 pada kelas 1, yang mengindikasikan kemampuan yang sangat baik dalam mendeteksi pasien berisiko diabetes.

Sementara itu, algoritma SVM juga menunjukkan performa yang cukup baik dengan akurasi sebesar 88,39%. Model ini mencatat nilai recall sebesar 0,94 dan F1-score sebesar 0,89 pada kelas 1. Namun demikian, metrik evaluasi lainnya seperti presisi dan akurasi secara keseluruhan masih berada di bawah performa K-NN.

Secara umum, algoritma K-NN lebih unggul dalam mengidentifikasi kasus diabetes, terutama pada kelas positif, dan lebih direkomendasikan untuk digunakan dalam sistem pendukung keputusan medis berbasis data. Untuk pengembangan selanjutnya, disarankan untuk melakukan penyetelan nilai hyperparameter, seperti jumlah tetangga (k) pada K-NN, serta pemilihan kernel dan parameter C pada SVM. Selain itu, penerapan teknik validasi silang (cross-validation) dapat meningkatkan reliabilitas evaluasi model. Eksplorasi terhadap teknik seleksi dan rekayasa fitur juga perlu dipertimbangkan guna meningkatkan akurasi dan efisiensi model klasifikasi yang dikembangkan.

DAFTAR PUSTAKA

[1] Asri Mulyani, Sarah Khoerunisa, and Dede Kurniadi, "Perbandingan Kinerja Algoritma KNN dan SVM Menggunakan SMOTE untuk Klasifikasi Penyakit Diabetes," *Jurnal Nasional Teknik Elektro dan Teknologi Informasi*, vol. 14, no. 1, pp. 25–34, Feb. 2025, doi: 10.22146/jnteti.v14i1.15198.

[2] N. Maulidah, R. Supriyadi, D. Y. Utami, F. N. Hasan, A. Fauzi, and A. Christian, "Prediksi Penyakit Diabetes Melitus Menggunakan Metode Support Vector Machine dan Naive Bayes," *Indonesian Journal on Software Engineering (IJSE)*, vol. 7, no. 1, pp. 63–68, 2021, doi: 10.31294/ijse.v7i1.10279.

[3] S. Arifin and R. Satria, "Pengembangan Algoritma Support Vector Machine (SVM) untuk Mengklasifikasi Penyakit Diabetes," JIITE, Aug. 2024. doi: 10.63547/jiite.v1i2.39.

[4] D. Nasien *et al.*, "Perbandingan Implementasi Machine Learning Menggunakan Metode KNN, Naive Bayes, Dan Logistik Regression Untuk Mengklasifikasi Penyakit Diabetes," 2024. doi: 10.58794/jekin.v4i1.640.

[5] A. Wibowo, A. F. N. Masruriyah, and S. Rahmawati, "Refining Diabetes Diagnosis Models: The Impact of SMOTE on SVM, Logistic Regression, and Naïve Bayes for Imbalanced Datasets," *Journal of Electronics, Electromedical Engineering, and Medical Informatics*, vol. 7, no. 1, pp. 197–207, Jan. 2025, doi: 10.35882/jeeemi.v7i1.596.

[6] M. Salem Alzboon, M. Alqaraleh, and M. Subhi Al-Batah, "Diabetes Prediction and Management Using Machine Learning Approaches," *Data and Metadata*, vol. 4, Jan. 2025, doi: 10.56294/dm2025545.

[7] H. Tanveer, M. Faheem, A. H. Khan, and M. A. Adam, "Machine Learning Models for Personalized Treatment Recommendations in Diabetes Management," *International Journal of Scientific Research and Management (IJSRM)*, vol. 13, no. 04, pp. 2145–2154, Apr. 2025, doi: 10.18535/ijsrm/v13i04.ec03.

Dikirim	Revisi	Selesai Revisi	Diterima
21-Jul-2025	10-Nov-2025	25-Nov-2025	25-Nov-2025

[8] F. Nur Ikhromr, I. Sugiyarto, U. Faddillah, and B. Sudarsono, "Implementasi Data Mining Untuk Memprediksi Penyakit Diabetes Menggunakan Algoritma Naives Bayes dan K-Nearest Neighbor," *Journal of Information Technology and Computer Science (INTECOMS)*, vol. 6, no. 1, 2023, doi: 10.31539/intecoms.v6i1.5916.

[9] M. Sholeh, D. Andayati, and Rr. Y. Rachmawati, "DATA MINING MODEL KLASIFIKASI MENGGUNAKAN ALGORITMA K-NEAREST NEIGHBOR DENGAN NORMALISASI UNTUK PREDIKSI PENYAKIT DIABETES," *TeIKa*, vol. 12, no. 02, pp. 77–87, Oct. 2022, doi: 10.36342/teika.v12i02.2911.

[10] U. I. Lestari, "Penerapan Metode K-Nearest Neighbor untuk Sistem Pendukung Keputusan Identifikasi Penyakit Diabetes Melitus," *JATISI (Jurnal Teknik Informatika dan Sistem Informasi)*, vol. 8, no. 4, pp. 2071–2082, Dec. 2021, doi: 10.35957/jatisi.v8i4.1235.

[11] Hidayatunnisa'i, Kusrini, and Kusnawi, "Perbandingan Kinerja Metode Naïve Bayes dan Support Machine (SVM) dalam Analisis Kualitas Butir Soal Pilihan Ganda," *JURNAL FASILKOM*, vol. 13, no. 02, pp. 173–180, Aug. 2023, doi: 10.37859/jf.v13i02.5087.

[12] M. Vamsikrishna *et al.*, "An ensemble approach for detection of diabetes using SVM and DT," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 38, no. 1, p. 689, Apr. 2025, doi: 10.11591/ijeecs.v38.i1.pp689-698.

[13] A. A. Abdillah, A. Azwardi, I. Wahyudi, S. Arifin, M. A. Ibrahim, and T. Mauritsus, "LDA and SVM performance evaluation for diabetes prediction," in *AIP Conference Proceedings*, American Institute of Physics, Apr. 2024. doi: 10.1063/5.0198841.

[14] S. Yang, "Diabetes Prediction Based on KNN, XGBoost, SVM and LR model," *Applied and Computational Engineering*, vol. 104, no. 1, pp. 91–95, Nov. 2024, doi: 10.54254/2755-2721/104/20241142.

[15] F. R. Lumbanraja, F. Lufiana, Y. Heningtyas, and K. Muludi, "Implementasi Support Vector Machine (SVM) untuk Klasifikasi Pederita Diabetes Mellitus" *Jurnal Komputasi*, vol. 10, no. 1, pp. 75–83, Apr. 2022, doi: 10.23960/komputasi.v10i1.2940.

[16] Fibia Sentauri Cahyaningrum, "Comparison of Binary Logistic Regression and SVM to Classify Diabetes Sufferers," *Journal of Intelligent Systems and Information Technology*, vol. 1, no. 2, pp. 74–84, Jul. 2024, doi: 10.61971/jisit.v1i2.76.

[17] A. Pinandito, S. A. Wicaksono, and S. H. Wijoyo, "Implementasi Machine Learning dalam Deteksi Risiko Tinggi Diabetes Melitus pada Kehamilan," *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 10, no. 4, pp. 739–746, Aug. 2023, doi: 10.25126/jtiik.20241047005.

[18] A. Farhan AlShammari, "Implementation of Model Evaluation using Confusion Matrix in Python," 2024. doi: 10.5120/ijca2024924236.

[19] S. Sathyanarayanan, "Confusion Matrix-Based Performance Evaluation Metrics," *African Journal of Biomedical Research*, pp. 4023–4031, Nov. 2024, doi: 10.53555/AJBR.v27i4S.4345.