

Diabetes Mellitus Detection System Using Machine Learning with the Support Vector Machine (SVM) Method

Syahira Ratu Fadilla^{1*)}, Sayed Fachrurrazi²⁾, Desvina Yulisda³⁾

¹⁾²⁾³⁾Program Studi Sistem Informasi, Jurusan Teknik Elektro, Fakultas Teknik, Universitas Malikussaleh

^{*)}Correspondence author: syahira.220180052@mhs.unimal.ac.id, Lhokseumawe, Indonesia

DOI: <https://doi.org/10.37012/jtik.v12i2.3596>

Abstract

Background: Diabetes Mellitus (DM) remains a major non-communicable disease with a continuously rising prevalence, and delayed diagnosis increases the risk of severe complications such as kidney failure, cardiovascular disease, and neuropathy. Objective: This study aims to design and build a web-based Diabetes Mellitus detection system using the Support Vector Machine (SVM) algorithm and to measure the classification performance of the resulting model. Methods: The dataset consisted of 170 medical records from Rumah Sakit Cut Meutia collected over the 2023-2025 period, comprising eight predictor variables (gender, blood pressure, glucose level, insulin, skin thickness, Body Mass Index, diabetes family-history score, and age). The data were preprocessed through label encoding, median-based missing-value imputation, Interquartile Range (IQR) outlier clipping, and Min-Max normalization, then split into 80% training data (132 records) and 20% testing data (34 records). The SVM model, comparing Linear and Radial Basis Function (RBF) kernels, was optimized using GridSearchCV with Stratified 5-Fold Cross Validation. Results: The optimal model used an RBF kernel with $C = 1$ and $\gamma = 0.01$, achieving a cross-validation ROC-AUC of 96.43%. On the test set, the model achieved an accuracy of 88.24%, precision of 94.12%, recall of 84.21%, and an F1-score of 88.89%, with a test-set AUC of 96.1%. Usability testing using the System Usability Scale (SUS) on 10 respondents produced an average score of 71.75, placed in the "Good" category. Conclusion: The SVM-based web system was able to classify Diabetes Mellitus status accurately and was assessed by users as easy to operate, indicating its feasibility as a decision-support tool for the early detection of Diabetes Mellitus.

Keywords: Diabetes Mellitus, Support Vector Machine, GridSearchCV, Classification, Web-Based System

Abstrak

Diabetes Mellitus (DM) merupakan salah satu penyakit tidak menular yang jumlah penderitanya terus meningkat, dan keterlambatan diagnosis dapat memperbesar risiko komplikasi serius seperti gagal ginjal, penyakit kardiovaskular, dan neuropati. Penelitian ini bertujuan untuk merancang dan membangun sistem pendeteksi Diabetes Mellitus berbasis web menggunakan algoritma Support Vector Machine (SVM) serta mengukur kinerja klasifikasi model yang dihasilkan. Dataset yang digunakan berjumlah 170 data rekam medis pasien Rumah Sakit Cut Meutia [PERLU KONFIRMASI: rentang tahun data], terdiri atas delapan variabel prediktor (jenis kelamin, tekanan darah, kadar glukosa, insulin, ketebalan kulit, Body Mass Index, skor riwayat genetika diabetes, dan usia). Data melalui tahapan preprocessing berupa label encoding, imputasi missing value berbasis median, penanganan outlier dengan Interquartile Range (IQR), dan normalisasi Min-Max, kemudian dibagi menjadi 80% data latih (132 data) dan 20% data uji (34 data). Model SVM, dengan perbandingan kernel Linear dan Radial Basis Function (RBF), dioptimasi menggunakan GridSearchCV dan Stratified 5-Fold Cross Validation. Model terbaik menggunakan kernel RBF dengan $C = 1$ dan $\gamma = 0,01$, memperoleh ROC-AUC cross validation sebesar 96,43%. Pada data uji, model memperoleh accuracy 88,24%, precision 94,12%, recall 84,21%, F1-score 88,89%, dan AUC uji sebesar 96,1%. Hasil pengujian usability menggunakan System Usability Scale (SUS) terhadap 10 responden memperoleh skor rata-rata 71,75 yang termasuk kategori Baik. Sistem berbasis SVM yang dibangun mampu mengklasifikasikan status Diabetes Mellitus dengan baik dan dinilai pengguna mudah dioperasikan, sehingga layak digunakan sebagai pendukung keputusan deteksi dini Diabetes Mellitus.

Kata Kunci: Diabetes Mellitus, Support Vector Machine, GridSearchCV, Klasifikasi, Sistem Berbasis Web

PENDAHULUAN

Diabetes Mellitus (DM) merupakan penyakit yang tidak menular ditandai dengan kenaikan kadar glukosa dalam darah (hiperglikemia) akibat gangguan pada sekresi insulin, masalah fungsi insulin, atau gabungan keduanya. Jika tidak dikelola dengan baik, DM bisa berujung pada komplikasi serius seperti gagal ginjal, penyakit kardiovaskular, neuropati, retinopati, bahkan amputasi, sehingga deteksi dini merupakan langkah penting untuk mengurangi risiko komplikasi dan meningkatkan kualitas hidup penderita (Novianti et al., 2022).

Angka prevalensi DM terus bertambah setiap tahunnya. Federasi Diabetes Internasional (IDF) memperkirakan jumlah pengidap diabetes di dunia untuk rentang usia 20–79 tahun akan mengalami peningkatan signifikan dalam beberapa dekade ke depan jika tidak didukung oleh langkah pencegahan dan pengelolaan yang efektif, dengan Indonesia termasuk dalam kategori negara yang memiliki jumlah pengidap diabetes tertinggi di dunia [PERLU KONFIRMASI: angka proyeksi IDF terbaru]. Kenaikan prevalensi ini akibat perubahan gaya hidup, tingginya asupan makanan yang mengandung gula dan lemak, rendahnya aktivitas fisik, obesitas, serta faktor genetik (Fauzi & Yunial, 2022).

Salah satu masalah utama dalam penanganan DM adalah diagnosis yang terlambat. Pada fase awal, penyakit ini sering kali tidak menunjukkan gejala yang jelas, sehingga banyak pasien baru menyadari kondisi mereka setelah komplikasi muncul. Situasi ini semakin buruk akibat keterbatasan akses terhadap layanan kesehatan dan tingginya biaya pemeriksaan laboratorium, khususnya di wilayah dengan fasilitas kesehatan yang minim (Azmi et al., 2023). Perkembangan Artificial Intelligence dan Machine Learning memberikan kesempatan untuk mendukung tenaga kesehatan dalam menganalisis data pasien dengan lebih cepat dan objektif sebagai bantuan untuk mengambil keputusan klinis.

Support Vector Machine (SVM) merupakan salah satu algoritma Machine Learning yang sering digunakan dalam klasifikasi data medis karena kemampuannya untuk membentuk hyperplane optimal dengan margin terbesar antar kelas, yang menghasilkan generalisasi baik meskipun pada dataset dengan ukuran yang relatif kecil (Fachrurrazi, 2018). Studi internasional juga menunjukkan keampuhan SVM dalam berbagai situasi diagnosis klinis: Abbas et al. (2019) menggunakan SVM untuk meramalkan risiko diabetes tipe 2 jangka panjang berdasarkan data uji toleransi glukosa oral, sedangkan tinjauan sistematis oleh Guido et al. (2024) menegaskan bahwa SVM tetap menjadi salah satu model pembelajaran mesin yang paling banyak digunakan di berbagai bidang layanan kesehatan karena

ketahanan performanya pada data berdimensi tinggi. Penelitian lain oleh Khaleel & Al-Bakry (2023) juga menunjukkan bahwa penggabungan algoritma klasifikasi dengan proses preprocessing yang tepat dapat secara signifikan meningkatkan akurasi diagnosis diabetes.

Penelitian terdahulu menunjukkan bahwa metode SVM dapat menghasilkan klasifikasi yang baik pada data kesehatan, contohnya akurasi mencapai 82,36% dalam klasifikasi penyakit hati setelah penyeimbangan data dengan SMOTE (Usman & Prasetyo, 2024), dan akurasi 95% pada klasifikasi penyakit jantung koroner yang didasarkan pada data rekam medis pasien (Kausar & Sari, 2025). Walaupun demikian, sebagian besar studi sebelumnya terbatas pada pengembangan model klasifikasi tanpa mengintegrasikannya ke dalam sistem yang dapat langsung digunakan oleh pengguna, dan banyak di antaranya menggunakan dataset publik yang belum sepenuhnya mencerminkan karakteristik pasien di fasilitas layanan kesehatan di Indonesia. Penelitian sebelumnya juga belum banyak mengintegrasikan tahapan preprocessing komprehensif, optimasi hyperparameter dengan GridSearchCV, dan validasi Stratified Cross Validation dalam satu rangkaian pembangunan model klasifikasi DM yang juga diimplementasikan ke dalam sistem berbasis web.

Berdasarkan kesenjangan tersebut, penelitian ini memanfaatkan 170 data rekam medis pasien Rumah Sakit Cut Meutia yang dikumpulkan selama periode tahun 2023–2025, meliputi delapan variabel prediktor, yakni jenis kelamin, tekanan darah, kadar glukosa, kadar insulin, ketebalan kulit, Indeks Massa Tubuh (IMT), riwayat genetik diabetes, dan usia. Keunikan penelitian ini terletak pada kombinasi langkah preprocessing menyeluruh (enkoding, imputasi nilai hilang berdasarkan median per kelas, penanganan outlier dengan IQR, dan normalisasi Min-Max), optimasi hyperparameter SVM melalui GridSearchCV dengan Stratified 5-Fold Cross Validation, serta penerapan langsung model terbaik ke dalam sistem deteksi berbasis web yang diperiksa keefektifannya menggunakan System Usability Scale (SUS) di kalangan pengguna nyata di lingkungan Rumah Sakit Cut Meutia. Penelitian ini bertujuan untuk (1) merancang dan membangun sistem pendeteksi Diabetes Mellitus berbasis web menggunakan algoritma SVM, (2) mengimplementasikan optimasi hyperparameter SVM menggunakan GridSearchCV dan Stratified 5-Fold Cross Validation, dan (3) mengukur kinerja klasifikasi serta tingkat usability sistem yang dihasilkan. Penelitian ini diharapkan memberikan kontribusi teoretis terhadap penggunaan optimasi hyperparameter SVM pada data rekam medis lokal, serta kontribusi praktis berupa sistem

dukungan keputusan untuk deteksi dini DM yang dapat langsung digunakan oleh tenaga medis dan masyarakat.

METODE PENELITIAN

Penelitian ini adalah penelitian terapan dalam bidang Sistem Informasi yang mengintegrasikan pengembangan model klasifikasi Machine Learning dengan pembuatan sistem informasi berbasis web. Penelitian dilakukan dengan memanfaatkan data rekam medis pasien di Rumah Sakit Cut Meutia, Lhokseumawe. Tahapan penelitian meliputi pengumpulan data, preprocessing data, pemodelan dan optimasi algoritma Support Vector Machine (SVM), penerapan model ke dalam sistem berbasis web, serta pengujian kebermanfaatan sistem.

Dataset Penelitian

Dataset yang digunakan terdiri atas 170 data rekam medis pasien Rumah Sakit Cut Meutia yang dikumpulkan pada rentang tahun 2023–2025, dengan delapan variabel prediktor yaitu jenis kelamin, tekanan darah diastolik, kadar glukosa plasma, kadar insulin serum, ketebalan lipatan kulit triceps, Body Mass Index (BMI), skor riwayat genetika diabetes, dan usia, serta satu variabel target berupa status Diabetes Mellitus (0 = tidak terindikasi, 1 = terindikasi). Sebagai gambaran umum karakteristik dataset, Tabel 1 menyajikan statistik deskriptif dari seluruh variabel numerik. Demi menjaga kerahasiaan data medis pasien, identitas pasien pada dataset penelitian ini telah dianonimkan dan tidak ditampilkan dalam bentuk nama individu pada naskah ini.

Tabel 1.
Statistik Deskriptif Dataset (N = 170)

Variabel	Minimum	Maksimum	Rata-rata	Std. Deviasi
Glukosa (mg/dl)	25,00	498,90	174,27	98,45
Tekanan Darah (mmHg)	30,00	212,00	102,95	40,64
Ketebalan Kulit (mm)	10,00	63,00	32,05	10,80
Insulin (μ U/ml)	14,00	846,00	168,91	133,06
BMI	19,30	67,10	34,82	7,63
Riwayat Genetika	0,09	2,42	0,54	0,39
Usia (tahun)	21,00	80,00	38,09	15,22

Distribusi kelas target menunjukkan bahwa 95 pasien (55,9%) terdiagnosis Diabetes Mellitus dan 75 pasien (44,1%) tidak terdiagnosis, sedangkan berdasarkan jenis kelamin, dataset terdiri dari 116 pasien perempuan (68,2%) dan 54 pasien laki-laki (31,8%).

Komposisi ini memperlihatkan dataset yang cukup seimbang di antara kelas-kelas sehingga mendukung pelatihan model klasifikasi.

Preprocessing Data

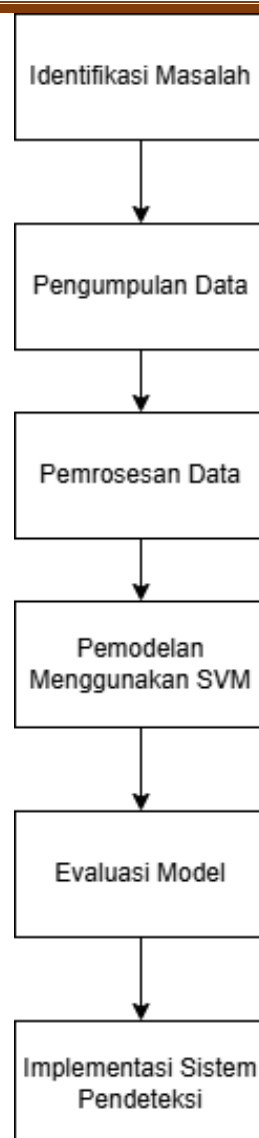
Sebelum tahapan klasifikasi dijalankan, dataset melalui empat tahapan preprocessing. Pertama, encoding data kategorikal pada atribut jenis kelamin menggunakan teknik Label Encoding (Laki-laki = 1, Perempuan = 0). Kedua, penanganan missing value pada atribut tekanan darah, ketebalan kulit, dan insulin yang bernilai 0 — nilai-nilai ini dianggap tidak mencerminkan kondisi fisiologis pasien dan diimputasi dengan median yang didasarkan pada kelas target. Ketiga, penanganan outlier menggunakan metode Interquartile Range (IQR) dengan teknik clipping, sehingga nilai ekstrem dibatasi pada batas bawah dan batas atas tanpa mengurangi jumlah data. Keempat, Min-Max Normalization untuk menyamakan skala semua variabel numerik ke dalam rentang 0–1, agar tidak ada variabel yang mendominasi proses pembelajaran model karena perbedaan satuan ukur.

Setelah tahapan preprocessing selesai, dataset dibagi menjadi data latih dan data uji dengan rasio 80:20 menggunakan teknik train-test split, sehingga diperoleh 132 data latih dan 34 data uji.

Pemodelan dan Optimasi Support Vector Machine

Model klasifikasi dikembangkan dengan algoritma Support Vector Machine (SVM) dengan membandingkan dua jenis fungsi kernel, yaitu Linear Kernel dan Radial Basis Function (RBF). SVM berfungsi dengan menemukan hyperplane terbaik yang memaksimalkan jarak antara kelas-kelas data (Fachrurrazi, 2018). Untuk mendapatkan kombinasi hyperparameter yang optimal, dilakukan optimasi menggunakan GridSearchCV pada parameter kernel, Cost (C), dan gamma. Setiap kombinasi parameter dievaluasi menggunakan Stratified 5-Fold Cross Validation dengan metrik ROC-AUC sebagai dasar pemilihan model terbaik, sehingga proporsi kelas pada setiap fold tetap seimbang.

Alur penelitian secara keseluruhan disajikan pada Gambar 1, yang menggambarkan tahapan mulai dari pengumpulan data hingga implementasi model ke dalam sistem berbasis web.



Gambar 1. Alur Penelitian

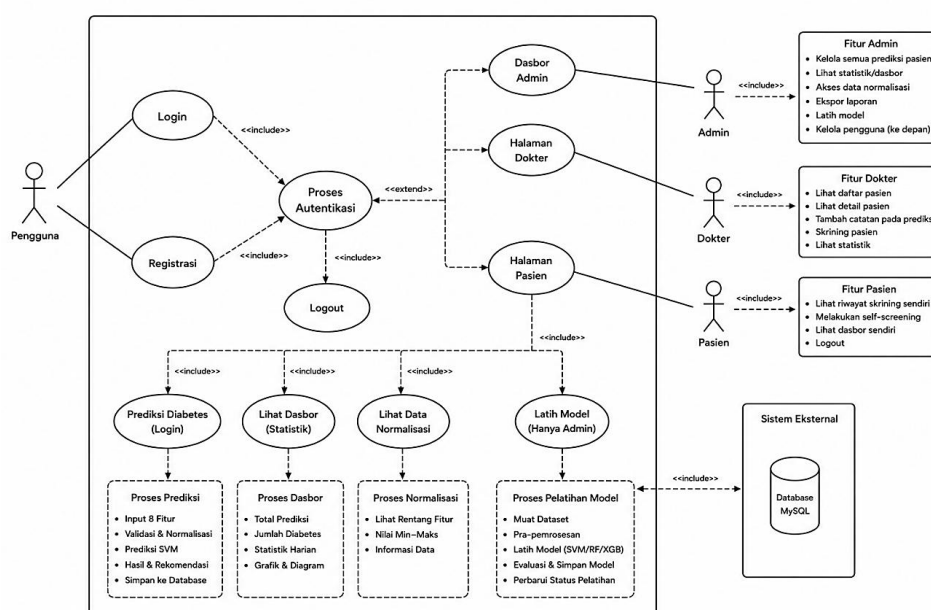
Evaluasi Model

Kinerja model diukur menggunakan Confusion Matrix untuk memperoleh nilai True Positive (TP), True Negative (TN), False Positive (FP), dan False Negative (FN), yang selanjutnya digunakan untuk menghitung metrik Accuracy, Precision, Recall, F1-Score, dan Area Under the ROC Curve (AUC).

Implementasi Sistem dan Pengujian Usability

Model SVM dengan parameter optimal diterapkan dalam sistem berbasis web menggunakan bahasa pemrograman Python dan database MySQL, dengan Visual Studio Code sebagai alat pengembangan. Sistem menawarkan peran pengguna admin, dokter, dan pasien, yang dilengkapi dengan fitur registrasi dan autentikasi, masukan data pemeriksaan, proses

klasifikasi otomatis, penyimpanan riwayat hasil deteksi, dasbor statistik, serta ekspor data. Sistem diuji tingkat kebergunaannya menggunakan metode System Usability Scale (SUS) melalui website glucolab.my.id yang melibatkan 10 responden (1 admin, 1 dokter, dan 8 pasien). Responden diminta menjalankan skenario penggunaan sistem secara lengkap — mulai dari registrasi, login, pengisian data pemeriksaan, hingga melihat hasil deteksi — sebelum mengisi 10 pernyataan kuesioner SUS. Skor tiap pernyataan dikonversi sesuai aturan SUS, dijumlahkan, dan dikalikan 2,5 untuk memperoleh skor akhir pada rentang 0–100, kemudian diinterpretasikan berdasarkan kategori skor SUS standar.

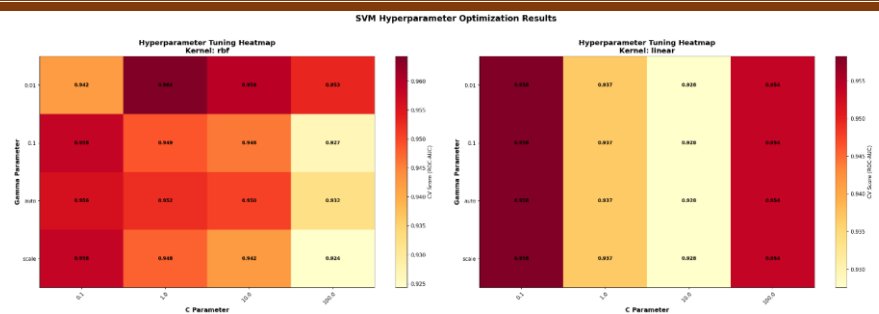


Gambar 2. Use Case Diagram Sistem Pendeteksi Diabetes Mellitus

HASIL DAN PEMBAHASAN

Optimasi Hyperparameter Model

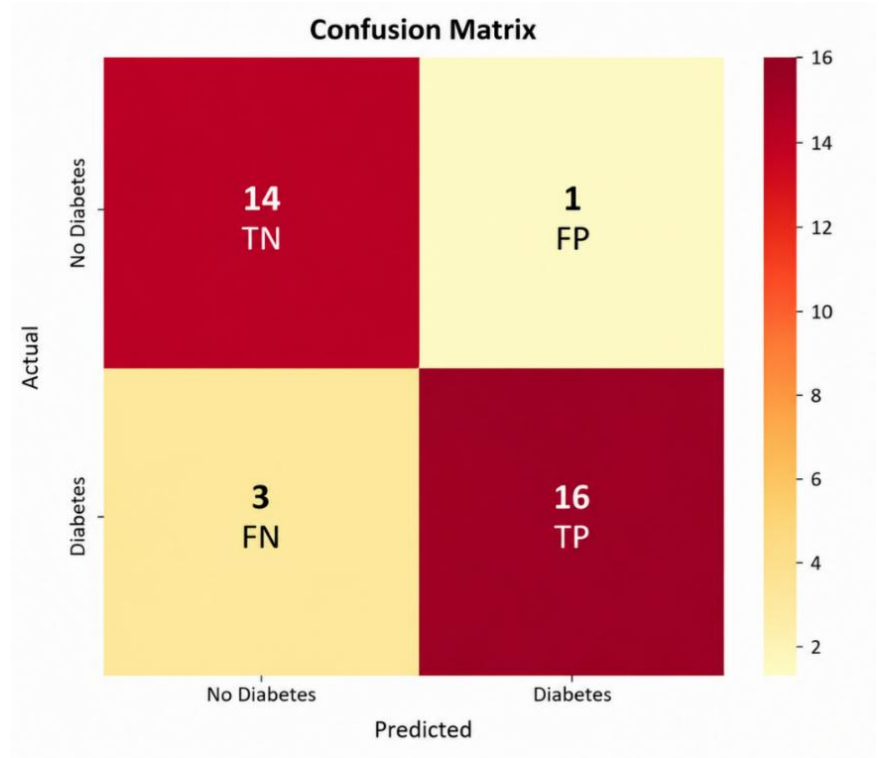
Proses optimasi menggunakan GridSearchCV dengan Stratified 5-Fold Cross Validation menghasilkan kombinasi parameter terbaik, yaitu kernel Radial Basis Function (RBF), Cost (C) sebesar 1, dan gamma sebesar 0,01, dengan nilai ROC-AUC cross validation sebesar 96,43%. Hasil ini menunjukkan bahwa kernel RBF lebih unggul dibanding kernel Linear dalam menangkap pola hubungan nonlinier antarvariabel pada dataset penelitian ini, sejalan dengan karakteristik data rekam medis yang umumnya tidak dapat dipisahkan secara linear (Fachrurrazi, 2018).



Gambar 3. Hyperparameter Heatmap GridSearchCV

Hasil Evaluasi Model pada Data Uji

Model SVM dengan parameter terbaik selanjutnya diuji menggunakan 34 data uji. Tabel 2 menyajikan Confusion Matrix hasil pengujian, sedangkan Tabel 3 menyajikan ringkasan metrik evaluasi.



Gambar 4. Confusion Matrix Model SVM pada Data Uji

Tabel 2.

Confusion Matrix Hasil Pengujian Model SVM

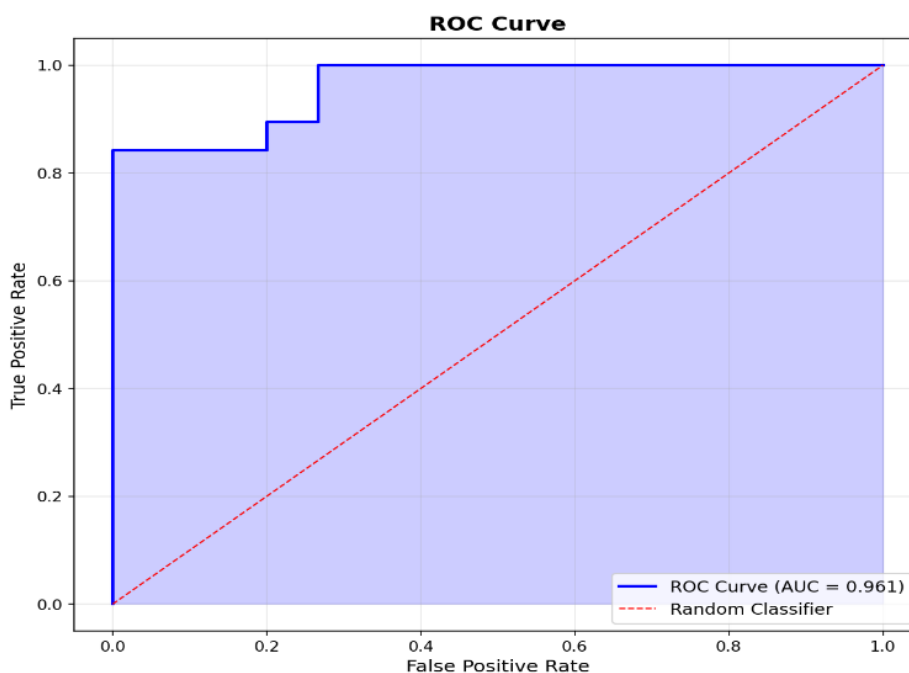
	Prediksi Positif (DM)	Prediksi Negatif (Tidak DM)
Aktual Positif (DM)	16 (TP)	3 (FN)
Aktual Negatif (Tidak DM)	1 (FP)	14 (TN)

Tabel 3.

Ringkasan Metrik Evaluasi Model SVM

Metrik	Nilai
Accuracy	88,24%
Precision	94,12%
Recall	84,21%
F1-Score	88,89%
AUC (Data Uji)	96,1%

Berdasarkan Tabel 2, model berhasil mengklasifikasikan 30 dari 34 data uji dengan benar (16 True Positive dan 14 True Negative), dengan 1 False Positive dan 3 False Negative. Nilai accuracy sebesar 88,24% menunjukkan bahwa mayoritas data uji diklasifikasikan dengan benar. Nilai precision sebesar 94,12% mengindikasikan bahwa hampir seluruh pasien yang diprediksi terindikasi DM memang benar terindikasi, sehingga model menghasilkan risiko positif palsu yang rendah — karakteristik yang penting pada konteks skrining kesehatan agar pasien tidak menerima diagnosis awal yang keliru. Nilai recall sebesar 84,21% menunjukkan model masih melewatkan sebagian kecil pasien yang sebenarnya terindikasi DM (3 dari 19 kasus positif aktual), yang perlu menjadi perhatian karena false negative pada konteks skrining kesehatan berisiko menunda penanganan pasien. Nilai F1-Score sebesar 88,89% memperlihatkan keseimbangan yang baik antara precision dan recall, sementara nilai AUC menunjukkan model memiliki kemampuan diskriminasi yang baik antara kelas DM dan tidak DM.



Gambar 5. Kurva ROC Model SVM pada Data Uji

Pengujian Usability Sistem

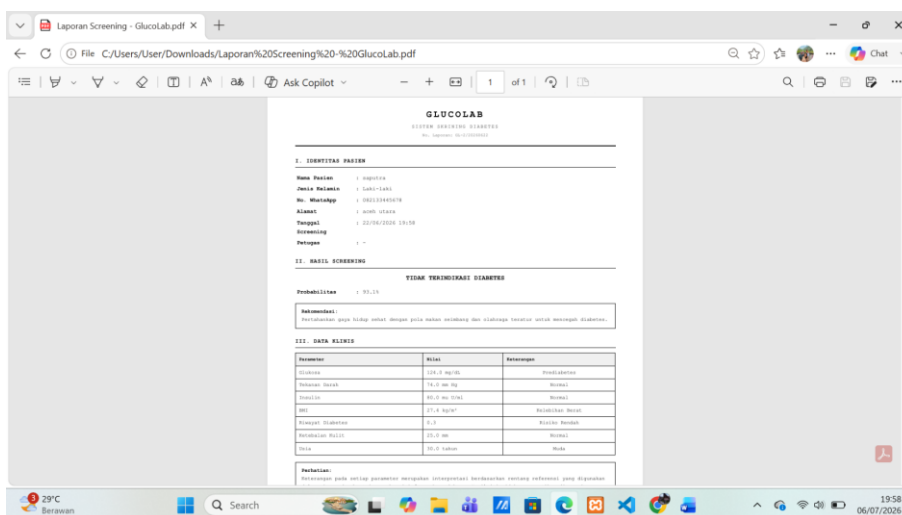
Pengujian usability menggunakan System Usability Scale (SUS) melibatkan 10 responden dengan skor individu berada pada rentang 55,0 hingga 97,5. Dari 10 orang responden, 5 orang mendapat kategori Kurang, 2 orang kategori Baik, dan 3 orang kategori Sangat Baik. Rata-rata nilai SUS yang didapatkan adalah 71,75, yang termasuk dalam kategori Baik (rentang 68–80,3). Hasil tersebut menunjukkan bahwa sistem yang dibuat umumnya gampang dipelajari dan dijalankan oleh pengguna dari berbagai peran, meskipun perbedaan skor antarresponden menunjukkan adanya peluang untuk perbaikan pada aspek kemudahan penggunaan bagi sebagian pengguna.

Pembahasan

Hasil penelitian ini menunjukkan bahwa algoritma SVM dengan kernel RBF yang dioptimalkan menggunakan GridSearchCV dan Stratified 5-Fold Cross Validation dapat menghasilkan kinerja klasifikasi yang bersaing pada data rekam medis pasien DM. Nilai akurasi 88,24% yang dicapai dalam penelitian ini setara dengan penelitian klasifikasi penyakit hati yang menggunakan SVM dan SMOTE dengan akurasi 82,36% (Usman & Prasetyo, 2024), tetapi masih lebih rendah dibandingkan penelitian klasifikasi penyakit jantung koroner yang mencapai akurasi 95% dengan kernel Linear pada data rekam medis numerik (Kausar & Sari, 2025). Perbedaan ini mungkin dipengaruhi oleh sifat dataset, jumlah contoh, serta kerumitan hubungan antarvariabel pada masing-masing kasus penyakit. Hasil ini juga sejalan dengan ulasan literatur oleh Guido et al. (2024) yang menekankan bahwa kinerja SVM dalam aplikasi kesehatan sangat bergantung pada kualitas tahap preprocessing dan pemilihan kernel yang sesuai dengan karakteristik data.

Kontribusi utama penelitian ini dibanding penelitian terdahulu terletak pada penggabungan tahapan preprocessing secara menyeluruh dengan optimasi hyperparameter menggunakan GridSearchCV dan validasi Stratified Cross Validation, yang belum banyak diterapkan secara komprehensif dalam satu rangkaian pembangunan model klasifikasi DM di penelitian-penelitian sebelumnya (Usman & Prasetyo, 2024; Kausar & Sari, 2025). Selain itu, tidak seperti kebanyakan penelitian yang hanya berhenti pada proses pembangunan model (Laela & Baihaqi, 2024; Nugraha & Rahayu, 2023), studi ini juga mengaplikasikan model terbaik ke dalam sistem berbasis web yang diuji secara langsung tingkat kegunaannya oleh pengguna nyata di sektor layanan kesehatan, sehingga hasil penelitian menawarkan nilai manfaat praktis yang lebih jelas.

Ditinjau dari sisi implikasi untuk bidang Sistem Informasi, penelitian ini menunjukkan bahwa memasukkan algoritma Machine Learning ke dalam sistem informasi berbasis web dapat berfungsi sebagai salah satu jenis sistem pendukung keputusan (decision support system) dalam layanan kesehatan, seiring dengan tren penggunaan teknologi digital untuk meningkatkan akses skrining kesehatan (Rehman et al., 2020). Nilai recall yang tetap di bawah precision dalam penelitian ini menjadi perhatian penting: dalam konteks skrining kesehatan, false negative menimbulkan risiko lebih besar dibanding false positive, sehingga pengembangan lebih lanjut sebaiknya mempertimbangkan penyesuaian ambang klasifikasi atau penambahan data pelatihan untuk kelas positif demi mengurangi angka false negative.



Gambar 6. Antarmuka Halaman Hasil Deteksi pada Sistem Berbasis Web

KESIMPULAN DAN REKOMENDASI

Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan beberapa hal berikut. Pertama, sistem pendeteksi Diabetes Mellitus berbasis web berhasil dirancang dan diimplementasikan menggunakan algoritma Support Vector Machine (SVM), dengan tahapan preprocessing berupa encoding, imputasi missing value, penanganan outlier menggunakan IQR, dan normalisasi Min-Max. Kedua, optimasi hyperparameter SVM menggunakan GridSearchCV dengan Stratified 5-Fold Cross Validation berhasil diimplementasikan dan menghasilkan model terbaik dengan kernel RBF, $C = 1$, dan $\gamma = 0,01$, yang memperoleh ROC-AUC cross validation sebesar 96,43%. Ketiga, model yang dihasilkan memperoleh accuracy 88,24%, precision 94,12%, recall 84,21%, dan F1-score 88,89% pada data uji, yang menunjukkan bahwa model memiliki kemampuan klasifikasi

yang baik meskipun masih terdapat ruang perbaikan pada nilai recall. Keempat, hasil pengujian usability menggunakan System Usability Scale (SUS) memperoleh rata-rata skor 71,75 yang termasuk kategori Baik, menunjukkan bahwa sistem yang dikembangkan relatif mudah digunakan oleh pengguna dari berbagai peran.

Penelitian selanjutnya disarankan untuk meningkatkan jumlah dan variasi dataset guna memperbaiki kemampuan generalisasi model, membandingkan performa SVM dengan algoritma Machine Learning lain seperti Random Forest atau XGBoost, mempertimbangkan teknik penyeimbangan kelas atau penyesuaian threshold untuk menekan tingkat false negative, mengembangkan sistem ke dalam versi aplikasi mobile agar lebih mudah diakses publik, serta melibatkan kolaborasi dengan tenaga medis dalam menentukan parameter dan variabel tambahan yang relevan secara klinis.

REFERENSI

1. Abbas, H. T., Alic, L., Erraguntla, M., Ji, J. X., Abdul-Ghani, M., Abbasi, Q. H., & Qaraqe, M. K. (2019). Predicting long-term type 2 diabetes with support vector machine using oral glucose tolerance test. *PLoS ONE*, 14(12), e0219636. <https://doi.org/10.1371/journal.pone.0219636>
2. Azmi, G. J., et al. (2023). Analysis of factors affecting diabetes melitus prevalence in cities/districts of East Java in 2018. *Jurnal Sains dan Kesehatan (J. Sains Kes.)*, 5(5), 623.
3. Bhat, S. S., Selvam, V., Ansari, G. A., Ansari, M. D., & Rahman, M. H. (2022). Prevalence and early prediction of diabetes using machine learning in North Kashmir: A case study of District Bandipora. *Computational Intelligence and Neuroscience*, 2022, Article 2789760. <https://doi.org/10.1155/2022/2789760>
4. Fachrurrazi, S. (2018). Penggunaan metode support vector machine untuk mengklasifikasi dan memprediksi. *Jurnal Sistem Informasi*, 2(2), 1–10.
5. Fauzi, A., & Yunial, A. H. (2022). Optimasi algoritma klasifikasi Naive Bayes, Decision Tree, K-Nearest Neighbor, dan Random Forest menggunakan algoritma Particle Swarm Optimization pada diabetes dataset. *Jurnal Edukasi dan Penelitian Informatika*, 8(3), 470. <https://doi.org/10.26418/jp.v8i3.56656>

6. Guido, R., Ferrisi, S., Lofaro, D., & Conforti, D. (2024). An overview on the advancements of support vector machine models in healthcare applications: A review. *Information*, 15(4), 235. <https://doi.org/10.3390/info15040235>
7. Iparraguirre-Villanueva, O., Espinola-Linares, K., Flores Castañeda, R. O., & Cabanillas-Carbonell, M. (2023). Application of machine learning models for early detection and accurate classification of type 2 diabetes. *Diagnostics*, 13(14), 2383. <https://doi.org/10.3390/diagnostics13142383>
8. Kausar, M. R., Fuadi, W., & Fitri, Z. (2025). Classification of coronary heart disease based on community health centre medical record data using SVM algorithm. *INOVTEK Polbeng - Seri Informatika*, 10(2), 262–272. <https://doi.org/10.35314/isi.v10i2.635>
9. Khaleel, F. A., & Al-Bakry, A. M. (2023). Diagnosis of diabetes using machine learning algorithms. *Materials Today: Proceedings*, 80, 3200–3202. <https://doi.org/10.1016/j.matpr.2021.07.196>
10. Laela, I. N., & Baihaqi, W. M. (2024). Analisis klasifikasi penyakit multiple sclerosis menggunakan algoritma logistic regression dan support vector machine. *Generation Journal*, 8(1), 27–33. <https://doi.org/10.29407/gj.v8i1.20646>
11. Laila, U. E., Mahboob, K., Khan, A. W., Khan, F., & Taekeun, W. (2022). An ensemble approach to predict early-stage diabetes risk using machine learning: An empirical study. *Sensors*, 22(14), 5247. <https://doi.org/10.3390/s22145247>
12. Novianti, N., Zarlis, M., & Sihombing, P. (2022). Penerapan algoritma Adaboost untuk peningkatan kinerja klasifikasi data mining pada imbalance dataset diabetes. *Jurnal Media Informatika Budidarma*, 6(2), 1200. <https://doi.org/10.30865/mib.v6i2.4017>
13. Oladimeji, O. O., Oladimeji, A., & Oladimeji, O. (2024). Classification models for likelihood prediction of diabetes at early stage using feature selection. *Applied Computing and Informatics*, 20(3/4), 279–286.
14. Prasetyo, T. M., Amrullah, A., Syahrir, S., & Sari, B. N. (2022). Klasifikasi penyakit paru-paru menggunakan support vector machine. *JurTI (Jurnal Teknologi Informasi)*, 6(2), 205–212. <https://doi.org/10.36294/jurti.v6i2.242>
15. Rahmah, N., Purnawansyah, & Umar, F. (2024). Metode support vector machine untuk klasifikasi data penyakit hati yang imbalance. *Buletin Sistem Informasi dan Teknologi Islam (BUSITI)*, 5(1), 55–64. <https://doi.org/10.33096/busiti.v5i1.2189>

-
16. Rehman, A., Athar, A., Khan, M. A., et al. (2020). Modelling, simulation, and optimization of diabetes type II prediction using deep extreme learning machine. *Journal of Ambient Intelligence and Smart Environments*, 12(2), 125–138.
 17. Samet, S., Laouar, M. R., Bendib, I., & Eom, S. (2022). Analysis and prediction of diabetes disease using machine learning methods. *International Journal of Decision Support System Technology*, 14(1), 1–18. <https://doi.org/10.4018/IJDSST.303943>
 18. Sneha, N., & Gangil, T. (2019). Analysis of diabetes mellitus for early prediction using optimal features selection. *Journal of Big Data*, 6(1), Article 13. <https://doi.org/10.1186/s40537-019-0175-6>
 19. Tasin, I., Nabil, T. U., Islam, S., & Khan, R. (2023). Diabetes prediction using machine learning and explainable AI techniques. *Healthcare Technology Letters*, 10(1–2), 1–10. <https://doi.org/10.1049/htl2.12039>