

SUPPORT VECTOR MACHINE UNTUK KLASIFIKASI PENYAKIT DIABETES MELLITUS

Dedi Triyanto

Universitas Bina Sarana Informatika
Jl. Kramat Raya No. 98, Kwitang Jakarta Pusat, Indonesia

email korespondensi: dedi.triyanto@bsi.ac.id

Abstrak

Diabetes mellitus adalah penyakit metabolik kronis yang semakin umum dan dapat menyebabkan komplikasi serius jika tidak didiagnosis dan diklasifikasikan dengan tepat. Salah satu masalah utama yang dihadapi adalah kurangnya metode diagnosis yang cepat dan akurat untuk mendeteksi diabetes pada tahap awal. Pendekatan manual seringkali memakan waktu dan tidak memberikan hasil yang optimal, sehingga diperlukan metode berbasis teknologi yang lebih efisien. Penelitian ini bertujuan untuk mengembangkan model pembelajaran mesin menggunakan algoritma Support Vector Machine (SVM) yang dapat membantu dalam klasifikasi diabetes mellitus dengan akurasi yang lebih tinggi. Data yang digunakan berasal dari database kasus diabetes yang mencakup berbagai parameter klinis pasien. Sebelum data digunakan untuk pelatihan model, dilakukan pra-pemrosesan yang mencakup analisis komponen utama dan normalisasi, sehingga fitur yang paling relevan dapat dipilih. Model SVM kemudian digunakan untuk melakukan klasifikasi biner, yaitu menentukan apakah seorang pasien memiliki diabetes mellitus atau tidak. Untuk mengevaluasi kinerja model ini, digunakan beberapa metrik, termasuk akurasi, presisi, recall, dan skor f1. Hasil penelitian menunjukkan bahwa model SVM mencapai tingkat akurasi sebesar 88%, yang menandakan bahwa algoritma ini memiliki potensi besar dalam membantu proses diagnosis diabetes dengan cepat dan akurat. Dengan demikian, model ini diharapkan dapat menjadi solusi dalam mengatasi keterbatasan metode diagnosis tradisional, serta membantu tenaga medis dalam memberikan diagnosis yang lebih tepat, sehingga komplikasi yang diakibatkan oleh diabetes dapat dicegah lebih efektif.

Kata Kunci : Klasifikasi Diabetes, Support Vector Machine, Diagnosis Akurat, Algoritma Pembelajaran Mesin.

Abstract

Diabetes mellitus is a chronic metabolic disease that is becoming increasingly common and can lead to serious complications if not diagnosed and classified correctly. One of the major challenges faced is the lack of fast and accurate diagnostic methods for detecting diabetes at an early stage. Manual approaches are often time-consuming and do not provide optimal results, highlighting the need for more efficient, technology-based methods. This study aims to develop a machine learning model using the Support Vector Machine (SVM) algorithm to assist in the classification of diabetes mellitus with higher accuracy. The data used in this study comes from a diabetes case database containing various clinical parameters of patients. Before the data is used for model training, pre-processing steps are performed, including principal component analysis and normalization, to select the most relevant features. The SVM model is then applied to perform binary classification, determining whether or not a patient has diabetes mellitus. Several metrics, including accuracy, precision, recall, and F1-score, are used to evaluate the model's performance. The results show that the SVM model achieves an accuracy rate of 88%, indicating that this algorithm has great potential to help in diagnosing diabetes quickly and accurately. Thus, this model is expected to be a solution to overcome the limitations of traditional diagnostic methods and to assist medical professionals in providing more precise diagnoses, thereby preventing complications caused by diabetes more effectively.

Keywords: Diabetes Classification, Support Vector Machine, Accurate Diagnosis, Machine Learning Algorithm.

1. Pendahuluan

Diabetes adalah penyakit kronis yang mempengaruhi metabolisme tubuh, ditandai dengan kadar gula darah yang lebih tinggi dari normal karena adanya masalah pada produksi atau fungsi insulin. Penyakit ini dapat menyebabkan berbagai komplikasi serius, meningkatkan risiko kematian dini, dan memengaruhi beberapa bagian tubuh. Komplikasi yang umum terjadi antara lain gagal ginjal, amputasi kaki, kehilangan penglihatan, dan kerusakan saraf. Selain itu, orang dewasa yang mengidap diabetes berisiko lebih tinggi mengalami serangan



jantung dan stroke. Pada kehamilan, diabetes yang tidak terkontrol dapat meningkatkan risiko kematian janin serta komplikasi lainnya (Rizky et al., 2024)

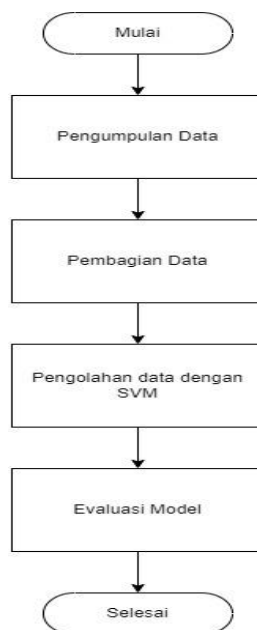
Penderita Diabetes Mellitus seringkali tidak menyadari penyakitnya, yang menyebabkan pengobatan tertunda. Karena Diabetes Mellitus dapat mengenai semua organ tubuh dan menimbulkan banyak keluhan, penyakit ini biasanya disebut sebagai pembunuh diam (ASFAW, 2019; Santoso et al., 2020). Gangguan penglihatan mata, katarak, penyakit jantung, sakit ginjal, impotensi seksual, luka membusuk atau gangren, infeksi paru-paru, gangguan pembuluh darah, stroke, dan lain-lain adalah penyakit yang dapat menyebabkan kematian (ASFAW, 2019; Jogo et al., 2023; Mucholladin et al., 2021). Faktor penyebab Diabetes Mellitus termasuk faktor keturunan (genetika), faktor pengaruh lingkungan, dan gaya hidup yang tidak sehat (Dwinnie & Dwyne, 2024; Maulidah et al., 2021)

Penunjang yang dapat digunakan untuk menegakan diagnosa terhadap suatu penyakit adalah pemeriksaan klinis, seperti yang dilakukan pada penderita diabetes melitus, di mana mereka dapat memeriksa kadar gula darah mereka dengan glukometer (Ginting et al., 2022). Salah satu cara penderita diabetes dapat mencegah komplikasi adalah dengan melakukan deteksi dini. Dalam banyak kasus, informasi awal tentang penderita diabetes membantu mencegah penyakit, menyembuhkan, dan mengobati penyakit. Banyak sistem komputer menggunakan kecerdasan manusia untuk membantu penderita mengelola penyakit (Junus et al., 2023; Maulidah et al., 2021). Metode pembelajaran mesin seperti Support Vector Machine dapat digunakan untuk membuat sistem berbasis kecerdasan buatan dalam implementasi. Metode ini dapat ditanamkan pada aplikasi yang dirancang khusus untuk diagnosis penyakit diabetes (A. D. Cahyani & Basuki, 2019; Jogo et al., 2023; Maulidah et al., 2021; Rizky et al., 2024)

Kekhawatiran tentang efek negatif diabetes mellitus, yang terus meningkat seiring dengan peningkatan kasusnya, menjadi latar belakang utama penelitian ini. Banyak pasien diabetes tidak menyadari risiko pra-diabetes yang mereka alami, yang pada akhirnya dapat memperburuk kondisi mereka. Penelitian ini bertujuan untuk mengembangkan model klasifikasi diabetes menggunakan pendekatan pembelajaran mesin. Proses penelitian mencakup pengumpulan data, pra-pemrosesan data, dan pengembangan model klasifikasi yang akurat (Cahyani et al., 2022; Ginting et al., 2022). Tujuan utama dari penelitian ini adalah untuk meningkatkan akurasi model prediksi diabetes dan menciptakan model yang dapat diterapkan pada berbagai dataset, bukan hanya satu. Dalam penelitian ini, dataset yang digunakan diambil dari situs web Kaggle, yang menyediakan data patokan untuk klasifikasi diabetes. Penelitian ini diharapkan dapat membantu dalam mendeteksi diabetes secara lebih awal dan akurat, sehingga komplikasi yang lebih serius dapat dicegah.

2. Metode

Metode yang akan digunakan dalam penelitian ini mengikuti beberapa tahap penting yang digambarkan dalam diagram alir. Proses dimulai dengan pengumpulan data, yang merupakan langkah awal untuk mendapatkan dataset yang relevan terkait diabetes. Setelah data dikumpulkan, langkah berikutnya adalah membagi data tersebut menjadi data pelatihan dan pengujian. Data kemudian diolah menggunakan algoritma Support Vector Machine (SVM) untuk melakukan klasifikasi. Selanjutnya, model yang telah dilatih akan dievaluasi berdasarkan metrik yang relevan untuk mengukur kinerjanya. Tahapan ini diakhiri dengan evaluasi model, yang bertujuan untuk menentukan seberapa baik model dapat mengklasifikasikan data dengan akurasi yang tinggi.



Gambar 1. Metodologi Penelitian

Berikut adalah keterangan untuk setiap proses yang telah disesuaikan:

1. **Pengumpulan Data:** Data yang digunakan dalam penelitian ini dikumpulkan dari sumber yang relevan, seperti dataset yang tersedia di situs Kaggle. Dataset ini berisi berbagai parameter klinis yang berhubungan dengan diabetes. Untuk penelitian ini, data yang dikumpulkan dari jenis wanita keturunan Indian pima digunakan. Jenis ini memiliki 9 atribut dan satu label, yang terdiri dari `n_pregnant`, `glucose_conc`, `bp`, `skin_len`, `insulin`, `bmi`, fungsi keturunan, usia, dan label output seperti terlihat pada Tabel 1. Data yang terkumpul akan digunakan untuk melatih dan menguji model dalam tahap berikutnya.

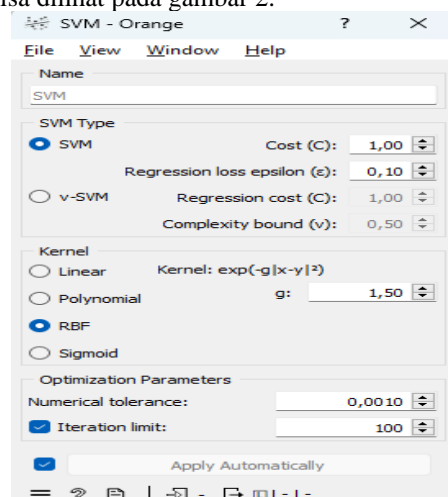
Tabel 1. Atribut dan deskripsi Dataset

Atribut	Deskripsi
Preg	Jumlah Berapa Kali Hamil
Glu	Konsentrasi Glukosa Setelah 2 jam Test
Blo	Tekanan Darah Diastolik
Skin	Ketebalan Kulit Bagian Triceps
Ins	2 jam Setelah Serum Insulin
BMI	Indeks massa tubuh (berat dalam kg/tinggi dalam m) ²
Ped	Fungsi Pedigree Diabetes
Age	Umur
Class	Kelas untuk klasifikasi, 1 artinya terkena diabetes dan 0 berarti Tidak

2. **Pembagian Data:** Setelah data terkumpul, langkah selanjutnya adalah membagi data menjadi dua bagian, yaitu 80% untuk data pelatihan dan 20% untuk data pengujian. Data pelatihan digunakan untuk mengajarkan model cara mengenali pola dari data, sedangkan data pengujian digunakan untuk menilai performa model pada data yang belum pernah dilihat sebelumnya. Pembagian ini bertujuan untuk memastikan bahwa model dapat mempelajari dan diuji secara seimbang.
3. **Pengolahan Data dengan SVM:** Pada tahap ini, algoritma Support Vector Machine (SVM) digunakan untuk mengolah data. Algoritma ini akan melakukan klasifikasi biner, yang bertujuan untuk menentukan apakah seorang pasien memiliki diabetes atau tidak. Algoritma SVM dipilih karena kemampuannya dalam menangani data dengan dimensi tinggi dan menghasilkan model klasifikasi yang akurat.
4. **Evaluasi Model:** Setelah model SVM selesai dilatih menggunakan data pelatihan, kinerjanya dievaluasi menggunakan Confusion Matrix. Confusion Matrix membantu dalam mengukur akurasi, presisi, recall, dan skor f1 dari model. Dengan menggunakan metode ini, evaluasi akan memperlihatkan seberapa baik model dapat mengklasifikasikan data secara tepat dan seberapa besar tingkat kesalahan yang terjadi pada prediksi diabetes.

3. Hasil dan Pembahasan

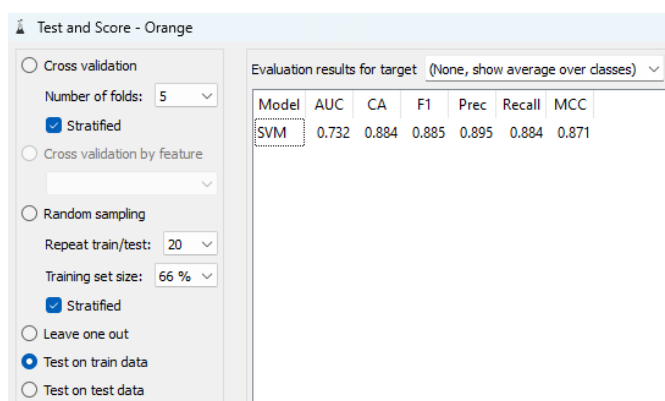
Untuk penelitian ini, parameter pengujian yang digunakan adalah biaya (C) = 1,00 dan epsilon untuk kerugian regresi (ϵ) = 0,10. Penelitian menggunakan kernel RBF (Function Basis Radial), yang diiterasi 100 kali. Detail algoritma SVM untuk parameter bisa dilihat pada gambar 2.



Gambar 2. Hasil Algoritma SVM

Pengujian Model

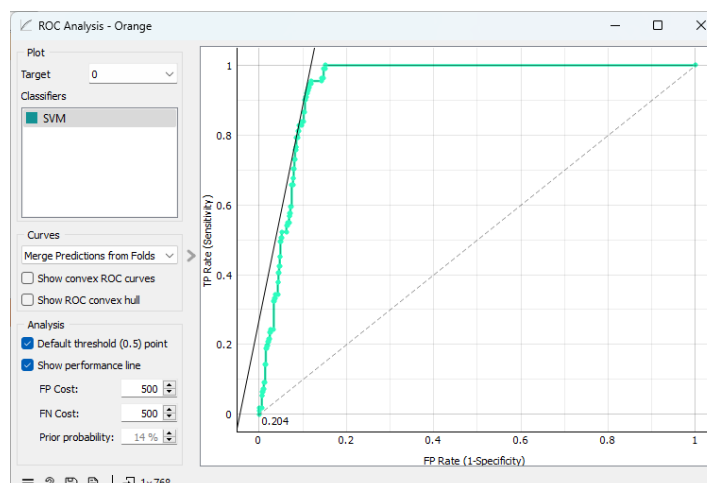
Hasil evaluasi performa model Support Vector Machine (SVM) yang ditampilkan dalam Gambar 3 menunjukkan beberapa metrik penting yang digunakan untuk menilai efektivitas model. Model ini dievaluasi menggunakan validasi silang (cross validation) sebanyak 5 lipatan, dengan hasil yang cukup baik. Nilai AUC (Area Under the Curve) adalah 0.732, yang menunjukkan bahwa model memiliki kemampuan moderat dalam membedakan antara kelas positif dan negatif. Akurasi keseluruhan model (CA) mencapai 88.4%, yang menunjukkan bahwa model mampu memprediksi dengan tepat dalam sebagian besar kasus. Nilai F1-score yang diperoleh adalah 0.885, mengindikasikan keseimbangan yang baik antara presisi dan recall. Nilai presisi mencapai 0.895, yang berarti dari semua prediksi positif yang dibuat oleh model, sekitar 89.5% benar. Recall berada di angka 0.884, yang berarti model berhasil menangkap sekitar 88.4% dari semua kasus positif sebenarnya. Selain itu, nilai MCC (Matthews Correlation Coefficient) yang cukup tinggi yaitu 0.871, menunjukkan bahwa model memiliki performa yang sangat baik dan konsisten dalam klasifikasi. Secara keseluruhan, metrik-metrik ini menunjukkan bahwa model SVM yang dibangun memiliki performa yang cukup baik dalam mengklasifikasikan data diabetes, dengan keseimbangan yang kuat antara akurasi dan kemampuan prediktif.



Gambar 3. Pengujian Model

Hasil Nilai Area Under Curve (AUC)

Hasil Klasifikasi penyakit Diabetes Mellitus ditunjukkan melalui kurva ROC. Bisa dilihat pada gambar 4. Dalam Kurva ROC tersebut ada di titik 0.204.



Gambar 4. Hasil AUC

Grafik di atas menunjukkan Receiver Operating Characteristic (ROC) curve dari model Support Vector Machine (SVM) yang dievaluasi untuk klasifikasi kasus diabetes. Sumbu vertikal (y-axis) menunjukkan nilai True Positive Rate (sensitivitas), sementara sumbu horizontal (x-axis) menggambarkan False Positive Rate (1 - spesifisitas). Kurva ROC ini menunjukkan bahwa model memiliki performa yang baik, karena sebagian besar plot berada mendekati sisi kiri atas grafik, yang menunjukkan sensitivitas tinggi dan tingkat kesalahan rendah. AUC (Area Under the Curve) dari kurva ini adalah 0.732, yang berarti model memiliki kemampuan moderat dalam membedakan antara kelas positif dan negatif. Titik-titik pada kurva menunjukkan variasi threshold dalam <http://jurnal.bsi.ac.id/index.php/mti>

menentukan prediksi positif, dengan titik optimal pada posisi threshold sekitar 0.204. Ini menunjukkan bahwa pada titik ini, model dapat menghasilkan keseimbangan yang baik antara mendeteksi positif sebenarnya (true positives) dan mengurangi positif palsu (false positives). Secara keseluruhan, ROC curve ini memberikan indikasi bahwa model SVM yang digunakan cukup handal untuk klasifikasi diabetes, dengan keseimbangan yang baik antara sensitivitas dan spesifisitas.

4. Kesimpulan

Kesimpulan dari penelitian ini berdasarkan evaluasi model Support Vector Machine (SVM) menunjukkan bahwa model memiliki performa yang cukup memuaskan dalam klasifikasi diabetes. Dengan nilai AUC (Area Under the Curve) sebesar 0.732 yang diperoleh dari kurva ROC, dapat disimpulkan bahwa model memiliki kemampuan moderat dalam membedakan antara pasien dengan diabetes dan yang tidak. Nilai akurasi model sebesar 88.4% mengindikasikan bahwa model berhasil mengklasifikasikan sebagian besar data dengan tepat. Hal ini didukung oleh nilai F1-score sebesar 0.885, yang menandakan keseimbangan yang baik antara presisi dan recall. Presisi model sebesar 0.895 menunjukkan bahwa hampir 90% dari prediksi positif yang dihasilkan oleh model adalah benar, sedangkan recall sebesar 0.884 menegaskan bahwa model mampu mengidentifikasi sekitar 88.4% dari semua kasus positif yang ada. Nilai MCC (Matthews Correlation Coefficient) yang cukup tinggi, yakni 0.871, juga menguatkan bahwa model ini memiliki kinerja yang baik dalam hal prediksi dan klasifikasi, dengan distribusi kesalahan yang rendah. Selain itu, dari hasil analisis kurva ROC, model ini menunjukkan keseimbangan yang baik antara True Positive Rate (sensitivitas) dan False Positive Rate (spesifisitas). Dengan titik threshold optimal sekitar 0.204, model SVM berhasil menunjukkan kinerja yang cukup efisien dalam mendeteksi kasus diabetes dan mengurangi kemungkinan kesalahan prediksi.

Secara keseluruhan, model SVM ini dapat dianggap andal untuk digunakan dalam klasifikasi diabetes. Kombinasi antara nilai akurasi yang tinggi, presisi yang baik, dan kemampuan dalam membedakan kelas positif dan negatif membuat model ini sangat potensial untuk diterapkan sebagai alat bantu dalam diagnosis klinis diabetes secara cepat dan tepat. Penelitian selanjutnya juga dapat difokuskan pada pengembangan sistem berbasis web atau aplikasi yang memanfaatkan model ini, sehingga dapat digunakan oleh praktisi medis atau masyarakat umum untuk membantu proses diagnosis diabetes secara lebih cepat dan akurat. Penelitian lebih lanjut juga dapat mengeksplorasi integrasi model prediktif dengan teknologi kesehatan terkini, seperti wearable devices, untuk mendapatkan data real-time yang lebih akurat dalam mendeteksi diabetes.

Referensi

- Asfaw, T. A. (2019). Prediction Of Diabetes Mellitus Using Machine Learning Techniques . *International Journal Of Computer Engineering And Technology*, 10(4), 145–148. <https://doi.org/10.34218/Ijcet.10.4.2019.004>
- Cahyani, A. D., & Basuki, A. (2019). Klasifikasi Diabetes Mellitus Menggunakan Support Vector Machine (Studi Kasus: Puskesmas Modopuro, Mojokerto). *Rekayasa*, 12(2), 174–182. <https://doi.org/10.21107/Rekayasa.V12i2.19763>
- Cahyani, Q. R., Finandi, M. J., Rianti, J., Arianti, D. L., Dwi, A., Putra, P., & Artikel, G. (2022). Prediksi Risiko Penyakit Diabetes Menggunakan Algoritma Regresi Logistik Diabetes Risk Prediction Using Logistic Regression Algorithm Article Info Abstrak. *Jomlai: Journal Of Machine Learning And Artificial Intelligence*, 1(2), 2828–9099. <https://doi.org/10.55123/Jomlai.V1i2.598>
- Dwinnie, Z. C., & Dwyne, Z. C. (2024). Comparison Of Machine Learning Algorithms In Diabetes Risk Classification. *Institut Riset Dan Publikasi Indonesia (Irpi)*, 1(August), 54–60.
- Ginting, J., Ginting, R., & Hartono, H. (2022). Deteksi Dan Prediksi Penyakit Diabetes Melitus Tipe 2 Menggunakan Machine Learning (Scooping Review). *Jurnal Keperawatan Priority*, 5(2), 93–105. <https://doi.org/10.34012/Jukep.V5i2.2671>
- Jogo, M. M. S., Biddinika, M. K., & Fadlil, A. (2023). Klasifikasi Penyakit Diabetes Dengan Algoritma Decision Tree Dan Naïve Bayes. *Resistor (Elektronika Kendali Telekomunikasi Tenaga Listrik Komputer) Vol.*, 6(2), 113–118.
- Junus, C. Z. V., Tarno, T., & Kartikasari, P. (2023). Klasifikasi Menggunakan Metode Support Vector Machine Dan Random Forest Untuk Deteksi Awal Risiko Diabetes Melitus. *Jurnal Gaussian*, 11(3), 386–396. <https://doi.org/10.14710/J.Gauss.11.3.386-396>
- Maulidah, N., Supriyadi, R., Utami, D. Y., Hasan, F. N., Fauzi, A., & Christian, A. (2021). Prediksi Penyakit Diabetes Melitus Menggunakan Metode Support Vector Machine Dan Naive Bayes. *Indonesian Journal On Software Engineering (Ijse)*, 7(1), 63–68. <https://doi.org/10.31294/Ijse.V7i1.10279>
- Mucholladin, A. W., Bachtiar, F. A., & Furqon, M. T. (2021). Klasifikasi Penyakit Diabetes Menggunakan Metode Support Vector Machine. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 5(2), 622–633. <http://j-ptiik.ub.ac.id>

- Rizky, M., Pramuntadi, A., Prastowo, D., & Hardan Gutama, D. (2024). Implementation Of Deep Neural Network Method On Classification Of Type 2 Diabetes Mellitus Disease Implementasi Metode Deep Neural Network Pada Klasifikasi Penyakit Diabetes Melitus Tipe 2. *Malcom: Indonesian Journal Of Machine Learning And Computer Science* , 4(3), 1043–1050.
- Santoso, R. R., Megasari, R., & Hambali, Y. A. (2020). Implementasi Metode Machine Learning. *Jurnal Aplikasi Dan Teori Ilmu Komputer*, 3(2), 85–97. <https://ejournal.upi.edu/index.php/jatikom>