
PERBANDINGAN AKURASI ALGORITMA SUPPORT VECTOR MACHINE (SVM) DAN LOGISTIC REGRESSION (LR) DALAM APLIKASI PREDIKSI DIABETES BERBASIS WEB MENGGUNAKAN PYTHON

Ahmad Firmansyah¹⁾, Amak Yunus Ep²⁾, Heri Santoso³⁾

Universitas PGRI Kanjuruhan Malang^{1,2,3)}
firs070801@gmail.com

Abstrak

Penelitian ini membandingkan akurasi antara algoritma Support Vector Machine (SVM) dan Logistic Regression (LR) dalam prediksi diabetes berbasis web menggunakan Python. Tujuannya adalah menilai kinerja kedua algoritma dalam mengklasifikasikan data pasien berdasarkan parameter medis. Pengujian dilakukan dengan dataset dan mengukur kinerja menggunakan Accuracy, Precision, Recall, dan F1-score. Hasilnya, SVM memperoleh Accuracy 77.27%, Precision 75.68%, Recall 51.85%, dan F1-score 61.54%, sedangkan Logistic Regression mendapatkan Accuracy 75.32%, Precision 64.91%, Recall 67.27%, dan F1-score 66.07%. SVM lebih unggul dalam Accuracy dan Precision, sedangkan Logistic Regression lebih baik dalam Recall dan F1-score. Pemilihan algoritma bergantung pada tujuan aplikasi: jika prioritasnya adalah mendeteksi lebih banyak kasus positif, maka Logistic Regression lebih disarankan, sementara jika fokusnya mengurangi kesalahan prediksi positif, SVM lebih tepat. Aplikasi web yang dikembangkan memudahkan pengguna dalam prediksi diabetes secara cepat, membantu tenaga medis atau individu dalam deteksi dini serta pengambilan keputusan untuk pencegahan dan penanganan penyakit.

Kata Kunci : Support Vector Machine; Logistic Regression; Prediksi Diabetes; Web; Python.

Abstract

This study compares the accuracy of the Support Vector Machine (SVM) and Logistic Regression (LR) algorithms in web-based diabetes prediction using Python. The goal is to assess the performance of both algorithms in classifying patient data based on medical parameters. Testing was conducted with a dataset and measuring performance using Accuracy, Precision, Recall, and F1-score. The results showed that SVM obtained Accuracy of 77.27%, Precision of 75.68%, Recall of 51.85%, and F1-score of 61.54%, while Logistic Regression obtained Accuracy of 75.32%, Precision of 64.91%, Recall of 67.27%, and F1-score of 66.07%. SVM was superior in Accuracy and Precision, while Logistic Regression was better in Recall and F1-score. The choice of algorithm depends on the purpose of the application: if the priority is to detect more positive cases, then Logistic Regression is more recommended, while if the focus is to reduce positive prediction errors, SVM is more appropriate. The developed web application makes it easier for users to predict diabetes quickly, helping medical personnel or individuals in early detection and decision making for disease prevention and treatment.

Keywords : Support Vector Machine; Logistic Regression; Diabetes Prediction; Web; Python.

1. PENDAHULUAN

Diabetes merupakan penyakit fatal yang disebabkan oleh kadar gula darah dalam tubuh yang tidak terkontrol dan ketidakmampuan pankreas memproduksi insulin dalam jumlah yang cukup. Jika tidak diobati, diabetes dapat memicu berbagai komplikasi serius seperti penyakit jantung, gangren, dan kerusakan organ lainnya. Menurut Atlas 10 *International Diabetes Federation* (IDF) 2021, pada tahun 2021 sekitar 537 juta orang berusia antara 20 dan 79 tahun menderita diabetes. Jumlah penderita diabetes diperkirakan akan meningkat menjadi 643 juta pada tahun 2030 dan mencapai 783 juta pada tahun 2045. Diabetes menyebabkan sekitar 6,7 juta kematian dan pengeluaran layanan kesehatan sebesar \$966 miliar. Di Indonesia,

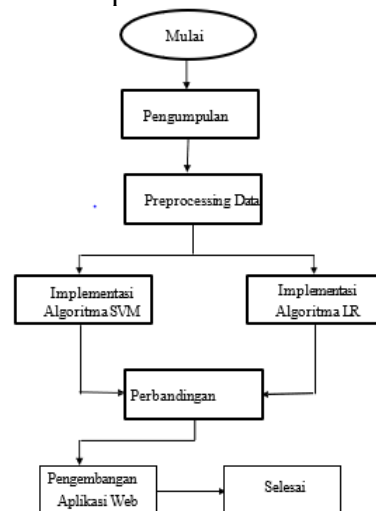
berdasarkan hasil Riskesdas 2018, prevalensi diabetes berdasarkan diagnosis medis pada penduduk usia 15 tahun ke atas adalah 2%.

Support Vector Machine (SVM) adalah algoritma machine learning yang efektif untuk klasifikasi dan regresi. SVM bekerja dengan menemukan hyperplane terbaik yang memisahkan data ke dalam kelas-kelas yang berbeda. Dalam penelitian oleh Febyan Sulastris dkk., SVM dengan kernel rbfdot dan badselldot mencapai nilai akurasi tertinggi yaitu 72,73%(Sulastris et al., 2023). dengan akurasi maksimum mencapai 88%. Algoritma ini dikenal karena kemampuannya dalam mengelola data yang kompleks dan non-linear, sehingga sering dipilih dalam berbagai studi prediksi penyakit, termasuk diabetes.

Algoritma Logistic Regression (LR) dipilih dalam penelitian ini karena kemampuannya yang terbukti handal dalam menangani klasifikasi dua kategori, seperti membedakan antara individu yang menderita dan tidak menderita diabetes. LR juga dikenal mudah dipahami dan memungkinkan peneliti untuk menilai sejauh mana masing-masing faktor risiko berkontribusi terhadap peluang terjadinya Diabetes Mellitus Tipe 2 (DMT2). Sejumlah studi terdahulu menunjukkan bahwa LR mampu menghasilkan prediksi yang akurat dan konsisten, dengan tingkat akurasi antara 84% hingga 85% pada berbagai jenis data kesehatan. Di samping itu, algoritma ini cukup sederhana, hemat sumber daya komputasi, dan tetap efektif meskipun digunakan pada dataset yang terbatas ukurannya—sehingga sangat sesuai digunakan dalam konteks data lokal seperti yang tersedia di Puskesmas, yang umumnya memiliki keterbatasan data. Oleh karena itu, LR dijadikan algoritma utama dalam penelitian ini untuk memprediksi potensi risiko DMT2 dan membantu proses deteksi dini di layanan kesehatan tingkat dasa.

2. METODE / ALGORITMA

Tahapan kerja penelitian ini mencakup identifikasi masalah, persiapan data, modeling, evaluasi, dan deployment, disertai beberapa visualisasi data.



Gambar 1. Metode Penelitian

Berikut ini merupakan tahapan penelitian:

1. Pengumpulan Data: Data yang digunakan adalah dari Pima Indians Diabetes Dataset yang diambil dari Kaggle.
2. Preprocessing Data: Tahap ini melibatkan pembersihan dan persiapan data sebelum digunakan dalam model, seperti penanganan nilai yang hilang, normalisasi, dan pemilihan fitur relevan.
3. Implementasi Algoritma:

- Support Vector Machine (SVM): Algoritma SVM diterapkan pada data yang telah diproses untuk membangun model prediksi diabetes.
 - Logistic Regression (LR): Algoritma LR juga diterapkan pada data yang sama untuk membangun model prediksi diabetes.
4. Evaluasi dan Perbandingan Akurasi: Kedua model (Support Vector Machine dan Logistic Regression) diuji dengan menggunakan data uji. Hasil prediksi dari kedua model dievaluasi dan dibandingkan untuk menentukan akurasi masing-masing model.
 5. Pengembangan Aplikasi Web
 - Menggunakan Visual Studio Code (VSC), Python, GitHub, dan Streamlit untuk mengembangkan aplikasi web yang menampilkan hasil prediksi diabetes.
 - Streamlit digunakan untuk membuat antarmuka pengguna dan Streamlit Share digunakan untuk mempublikasikan aplikasi.
 - Website 2 APK Builder Pro digunakan untuk mengonversi aplikasi web menjadi aplikasi mobile dalam format APK, sehingga dapat diakses lebih luas melalui perangkat Android.
 6. Pelaporan Hasil: Laporan yang menyajikan hasil penelitian, termasuk temuan dari perbandingan akurasi SVM dan LR serta hasil klasifikasi dari kedua model.

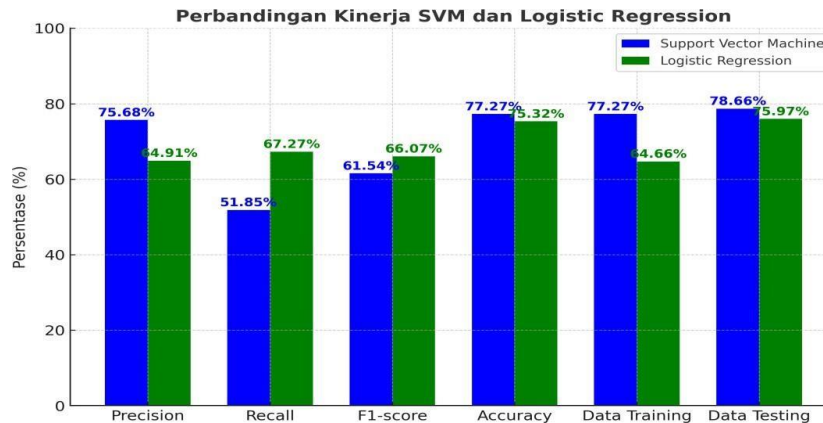
3. HASIL DAN PEMBAHASAN

Pada penelitian ini, algoritma Support Vector Machine (SVM) diterapkan untuk mengklasifikasikan data dengan menggunakan berbagai metrik evaluasi seperti Precision, Recall, F1-score, dan Accuracy. Hasil uji menunjukkan bahwa algoritma SVM menghasilkan nilai sebagai berikut:

Tabel 1. Hasil Pengujian

Metrix	Suport Vector Machine
Precision	75.68%
Reca l	51.85%
F1-score	61.54%
Acurracy	77.97%
Data Training	77.27%
Data Testing	78.66%

Pada pengujian algoritma ANN dengan menggunakan data training 70% dan data testing 30% menghasilkan akurasi 88,20% sedangkan untuk algoritma Random Forest menghasilkan akurasi sebesar 97,28%, kemudian untuk data training 80% dan data testing 20% algoritma ANN menghasilkan akurasi 88,76% dan Random forest menghasilkan akurasi 97,37%, dan yang terakhir untuk data training 90% dan data testing 10% algoritma ANN menghasilkan akurasi 91,39% dan Random forest menghasilkan akurasi 95,69%, di bawah ini merupakan grafik dari hasil perbandingan algoritma ANN dan Random Forest : Metrik ini menggambarkan performa SVM dalam melakukan klasifikasi, di mana Precision yang lebih tinggi menunjukkan bahwa model lebih tepat dalam mengidentifikasi kategori yang benar, meskipun Recall yang lebih rendah mengindikasikan beberapa data positif mungkin tidak terdeteksi. F1-score mencerminkan keseimbangan antara Precision dan Recall, sementara Accuracy menunjukkan tingkat keberhasilan keseluruhan dari model.



4. KESIMPULAN

Penelitian ini melakukan perbandingan antara algoritma Support Vector Machine (SVM) dan Logistic Regression (LR) dalam prediksi diabetes berbasis web menggunakan Python. Berdasarkan hasil pengujian, SVM memperoleh akurasi 77,27% dengan presisi 75,68%, recall 51,85%, dan F1-score 61,54%. Kekuatan utama SVM terletak pada presisi yang tinggi, yang menunjukkan kemampuan lebih baik dalam mengidentifikasi kasus positif dengan tingkat kesalahan lebih rendah, meskipun kelemahannya adalah recall yang relatif rendah. Di sisi lain, Logistic Regression mencatat akurasi 75,32%, presisi 64,91%, recall 67,27%, dan F1-score 66,07%. Hal ini mengindikasikan bahwa LR lebih mampu mendeteksi kasus positif secara menyeluruh, meskipun tingkat presisinya tidak setinggi SVM.

Perbandingan ini menegaskan bahwa masing-masing algoritma memiliki keunggulan yang berbeda: SVM lebih efektif dalam menekan kesalahan prediksi positif, sementara LR lebih unggul dalam menjaga keseimbangan antara presisi dan recall, sehingga lebih adaptif dalam mendeteksi berbagai kemungkinan kasus. Dengan demikian, pemilihan algoritma sebaiknya disesuaikan dengan kebutuhan penelitian atau aplikasi, apakah lebih menekankan pada akurasi prediksi atau pada kemampuan mendeteksi sebanyak mungkin kasus positif.

Selain itu, penelitian ini juga membuktikan bahwa integrasi algoritma machine learning ke dalam aplikasi berbasis web dapat memberikan nilai praktis, karena mempermudah pengguna maupun tenaga medis dalam melakukan prediksi diabetes secara cepat, efisien, dan berbasis data. Hasil ini diharapkan dapat menjadi landasan bagi pengembangan sistem prediksi medis berbasis teknologi di masa depan.

REFERENSI

- [1] Abbi Rokhman K, Berliana, & Arsi Primadani. (2021). Perbandingan Metode Support Vector Machine Dan Decision Tree Untuk Analisis Sentimen Review Komentar Pada Aplikasi Transportasi Online. *Jurnal Of Information System Management*, Vol. 2, No. 2.
- [2] Almufqi, F. M., & Voutama, A. (2023). Perbandingan Metode Data Mining Untuk Memprediksi Prestasi Akademik Siswa. *Jurnal Teknika*, 15(1), 61–66. <https://doi.org/10.30736/jt.v15i1.929>
- [3] Apriyani, H. (2020). Perbandingan Metode Naïve Bayes Dan Support Vector Machine Dalam Klasifikasi Penyakit Diabetes Melitus. In *Journal of Information Technology Ampera* (Vol. 1, Issue 3). <https://journal-computing.org/index.php/journal-ita/index>
- [4] Bidang SDK (ags). (2023). Permasalahan Obesitas di Indonesia. <https://Dinkes.Jogjaprovo.go.id/Berita/Detail/Permasalahan-Obesitas-Di-Indonesia?Utm>.

- [5] Fattya Ariani, & Andi Taufik. (2020). Perbandingan Metode Klasifikasi Data Mining untuk Prediksi Tingkat Kepuasan Pelanggan Telkomsel Prabayar. *SATIN - Sains Dan Teknologi Informasi*, 6(2), 46–55. <https://doi.org/10.33372/stn.v6i2.666>
- [6] Haganta Depari, D., Widiastiwi, Y., Mega Santoni, M., Ilmu Komputer, F., Pembangunan Nasional Veteran Jakarta, U., Fatmawati Raya, J. R., & Labu, P. (2022). Perbandingan Model Decision Tree, Naive Bayes dan Random Forest untuk Prediksi Klasifikasi Penyakit Jantung. *Jurnal Informatik Edisi Ke*, 18, 2022.
- [7] Badan Kebijakan Pembangunan Kesehatan. (2023). <https://www.badankebijakan.kemkes.go.id/daftar-frequently-asked-question-seputar-hasil-utama-ski-2023/hasil-utama-ski-2023/?utm>.
- [8] Jaringan Pangan & Gizi Indonesia. (2023). Pencegahan dan Pengendalian Obesitas di Indonesia. <https://jpg-indonesia.net/2023/11/pencegahan-dan-pengendalian-obesitas-di-indonesia/?utm>.
- [9] Kusumarini, A. I., Hogantara, P. A., & Chamidah, N. (2021). Perbandingan Algoritma Random Forest, Naïve Bayes, Dan Decision Tree Dengan Oversampling Untuk Klasifikasi Bakteri E. Coli. In *Seminar Nasional Mahasiswa Ilmu Komputer dan Aplikasinya (SENAMIKA) Jakarta-Indonesia*.
- [10] Marito Putry, N., & Nurina Sari, B. (2022). Komparasi Algoritma Knn Dan Naïve Bayes Untuk Klasifikasi Diagnosis Penyakit Diabetes Melitus. *Jurnal Sains Dan Manajemen*, 10(1).
- [11] Mega Santoni, M., Chamidah, N., & Matondang, N. (2020). Prediction of Hypertension using Decision Tree, Naïve Bayes and Artificial Neural Networks in KNIME Analytics Platform (Vol. 19, Issue 4).
- [12] Obesity and overweight. (2024). <https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight>.
- [13] Subarkah, P., Pambudi, E. P., & Hidayah, S. O. N. (2020). Perbandingan Metode Klasifikasi Data Mining untuk Nasabah Bank Telemarketing. *MATRIK: Jurnal Manajemen, Teknik Informatika Dan Rekayasa Komputer*, 20(1), 139–148. <https://doi.org/10.30812/matrik.v20i1.826>
- [14] Wibowo, M., & Ramadhani, R. (2021). Perbandingan Metode Klasifikasi Data Mining Untuk Rekomendasi Tanaman Pangan. *Jurnal Media Informatika Budidarma*, 5(3), 913. <https://doi.org/10.30865/mib.v5i3.3086>
- [15] Yoga Religia, Agung Nugroho, & Wahyu Hadikristanto. (2021). Klasifikasi Analisis Perbandingan Algoritma Optimasi pada Random Forest untuk Klasifikasi Data Bank Marketing. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 5(1), 187–192. <https://doi.org/10.29207/resti.v5i1.2813>
- Arifianto, T., & Moonlight, L. S. (2022). Penggunaan Metode Support Vector Machine (SVM). 2(2), 3–6.