

Analisis Komparatif Algoritma Klasifikasi Machine Learning untuk Memprediksi Diabetes

Comparative Analysis of Classification Algorithms in Machine Learning for Predicting Diabetes

Alfa Saleh^{*1}, Ria Eka Sari², Ramadani³, Fujiati⁴, Ratna Lestari⁵

¹Informatika, Universitas Samudra

²Sistem Informasi, Universitas Potensi Utama

³Informatika, Universitas Muhammadiyah Asahan

⁴Kewirausahaan, Universitas Satya Terra Bhinneka

⁵Biologi, Universitas Samudra

E-mail: [1alfasaleh@unsam.ac.id](mailto:alfasaleh@unsam.ac.id), [2ladiespure@gmail.com](mailto:ladiespure@gmail.com), [3ramadans.ordinary@gmail.com](mailto:ramadans.ordinary@gmail.com),
[4fujiati@satyaterrabhinneka.ac.id](mailto:fujiati@satyaterrabhinneka.ac.id), [5ratnalestari@unsam.ac.id](mailto:ratnalestari@unsam.ac.id)

Abstrak

Diabetes Mellitus merupakan salah satu penyakit kronis yang paling umum dijumpai, penyakit ini juga menjadi perhatian utama dalam isu kesehatan masyarakat secara global. pada penelitian ini, dilakukan pendekatan Machine Learning untuk membantu memprediksi penyakit diabetes di masyarakat. Machine learning sangat berguna dalam menganalisis data kesehatan sebab kemampuannya yang baik untuk mengolah data dalam jumlah yang besar. Studi komparatif dengan beberapa algoritma klasifikasi machine learning seperti K-Nearest Neighbors (KNN), Support Vector Machine (SVM), Naive Bayes dan Decision Tree (C4.5) telah dilakukan untuk menentukan algoritma mana yang memberikan hasil terbaik dalam hal memprediksi penyakit diabetes. Dimana, fitur yang digunakan dalam memprediksi penyakit diabetes meliputi jenis kelamin, usia, riwayat hipertensi, riwayat penyakit jantung, riwayat merokok, BMI, level dari HbA1c dan kadar glukosa dalam darah. Dari hasil penelitian ini, diperoleh tingkat akurasi prediksi penyakit diabetes untuk algoritma K-Nearest Neighbors (KNN) sebesar 92,5 %, selanjutnya algoritma Support Vector Machine (SVM) sebesar 94,5%, kemudian algoritma Naive Bayes sebesar 90% dan yang terakhir algoritma Decision Tree (C4.5) sebesar 93,5%. Sehingga, dari hasil pengujian beberapa algoritma klasifikasi machine learning tersebut dapat disimpulkan bahwa Algoritma Support Vector Machine (SVM) merupakan algoritma yang paling optimal dalam hal memprediksi penyakit diabetes.

Kata kunci: *Machine Learning, KNN, SVM, Naive Bayes, Decision Tree, Diabetes*

Abstract

Diabetes Mellitus is one of the most common chronic diseases, this disease is also a major concern in global public health issues. in this study, a Machine Learning approach was carried out to help predict diabetes in the community. Machine learning is very useful in analyzing health data because of its good ability to process large amounts of data. A comparative study with several machine learning classification algorithms such as K-Nearest Neighbors (KNN), Support Vector Machine (SVM), Naive Bayes and Decision Tree (C4.5) has been conducted to determine which algorithm gives the best results in terms of predicting diabetes. Where, the features used in predicting diabetes include gender, age, history of hypertension, history of heart disease,



history of smoking, BMI, level of HbA1c and blood glucose levels. From the results of this study, the accuracy rate of diabetes prediction for the K-Nearest Neighbors (KNN) algorithm is 92.5%, the Support Vector Machine (SVM) algorithm is 94.5%, the Naive Bayes algorithm is 90% and the last for the Decision Tree (C4.5) algorithm is 93.5%. So, from the test results of several machine learning classification algorithms it can be concluded that the Support Vector Machine (SVM) algorithm is the most optimal algorithm in terms of predicting diabetes.

Keywords: Machine Learning, KNN, SVM, Naive Bayes, Decision Tree, Diabetes.

1. PENDAHULUAN

Diabetes mellitus merupakan salah satu penyakit kronis yang paling sering ditemui dan menjadi perhatian utama dalam kesehatan masyarakat di seluruh dunia. Berdasarkan informasi dari American Diabetes Association, Diabetes tidak hanya menyerang orang dewasa namun juga beresiko untuk menyerang anak-anak muda[1]. Jumlah penderita diabetes meningkat dari 200 juta pada tahun 1990 menjadi sekitar 830 juta hingga saat ini[2]. Diagnosa dan klasifikasi yang akurat sangat penting untuk pengelolaan penyakit yang efektif dan pencegahan komplikasi lebih lanjut [3]. Seiring dengan meningkatnya jumlah penderita diabetes, tentu penting untuk mengembangkan metode yang lebih efisien dan akurat dalam mendeteksi dan memprediksi risiko diabetes pada populasi yang lebih luas.

Machine learning telah menjadi alat yang sangat berguna dalam analisis data[4], [5], [6], [7], [8], salah satunya untuk menganalisis data kesehatan, dikarenakan kemampuannya untuk mengolah data dalam jumlah yang besar dan menemukan pola yang tidak terdeteksi oleh metode tradisional[9], [10]. *Machine learning* memiliki beberapa model algoritma yang dikategorikan ke dalam *Supervised learning* dan *Unsupervised learning*. Fokus pada penelitian ini adalah penerapan beberapa algoritma klasifikasi, seperti *Naive Bayes*, *K-Nearest Neighbors* (KNN), *Support Vector Machine* (SVM) dan C4.5. algoritma tersebut sudah banyak diterapkan dalam berbagai penelitian untuk tujuan klasifikasi dan juga prediksi.

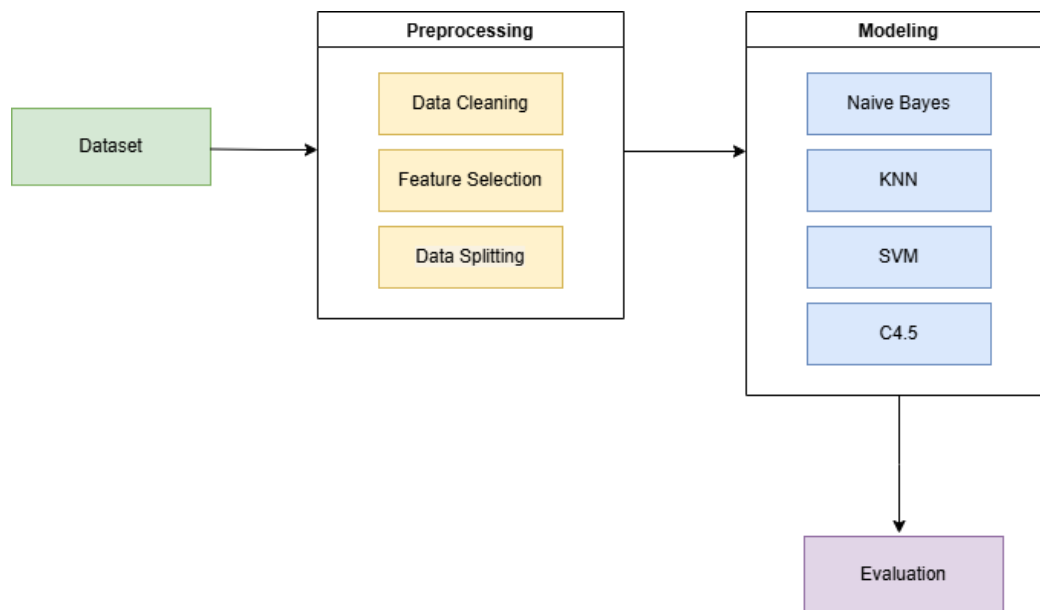
Proses klasifikasi dan prediksi terhadap penyakit diabetes menjadi tujuan dalam penelitian ini, untuk memastikan hasil yang optimal maka perlu melakukan studi komparatif terhadap algoritma-algoritma tersebut untuk menentukan algoritma mana yang memberikan hasil terbaik dalam memprediksi diabetes. Penelitian sebelumnya menunjukkan bahwa kombinasi teknik *machine learning* dapat meningkatkan akurasi prediksi penyakit [11], [12], [13], [14]. Namun, penelitian yang secara khusus membandingkan efektivitas *Naive Bayes*, KNN, SVM, dan C4.5 dalam memprediksi diabetes masih terbatas.

Oleh karena itu, penelitian ini bertujuan untuk mengevaluasi dan membandingkan optimalitas algoritma *Naive Bayes*, KNN, SVM, dan C4.5 dalam memprediksi diabetes. Dengan melakukan analisis komparatif ini, diharapkan dapat memberikan wawasan yang lebih mendalam tentang keunggulan dan keterbatasan masing-masing algoritma, serta membantu praktisi kesehatan dan peneliti dalam memilih alat prediksi yang paling tepat untuk digunakan dalam diagnosis penyakit diabetes.



2. METODE PENELITIAN

Untuk memudahkan dalam melakukan setiap tahapan dalam penerapan algoritma machine learning untuk memprediksi penyakit diabetes, maka diperlukan metodologi penelitian sebagai panduan dalam mengerjakan setiap tahapannya, adapun ilustrasi dari metodologi penelitian dalam studi ini dapat dilihat pada gambar 1 berikut.



Gambar 1. Flowchart Metodologi Penelitian

2.1 Mempersiapkan dataset

Dataset diabetes yang digunakan dalam penelitian ini diperoleh dari Kaggle yang terdiri dari 10646 data. Dalam dataset tersebut terdapat beberapa fitur numerik dan juga fitur kategorikal yang berkaitan dengan diabetes. Dari data tersebut juga, ada sebanyak 9477 data merupakan data para penderita diabetes dan 1169 data merupakan data yang bukan penderita diabetes.

2.2 Tahapan Preprocessing

Tahapan ini merupakan tahapan penting dalam *machine learning*, dataset yang telah disediakan sering kali tidak tersusun dengan struktur yang baik, masih terdapat data yang hilang, data yang tidak lengkap bahkan data yang mengandung *noise*[15]. Pada tahapan ini, akan dilakukan pembersihan terhadap data yang mengandung noise. Setelah dibersihkan, selanjutnya akan dilakukan pemilihan fitur yang berpengaruh atau berdampak pada proses klasifikasi untuk tujuan prediksi dan menghapus fitur-fitur yang tidak berpengaruh atau berdampak. Pada pemilihan fitur ini, didapatkan beberapa fitur yang dijadikan acuan dalam proses klasifikasi seperti jenis kelamin, usia, riwayat hipertensi, riwayat penyakit jantung, riwayat merokok, BMI, level dari HbA1c dan kadar glukosa dalam darah. Berikutnya, dari dataset yang telah dibersihkan dan dipilih fitur yang berpengaruh, akan dilakukan pemisahan dataset menjadi data latih dan data uji. Berdasarkan dari

10646 dataset yang ada, akan dibagi porsinya menjadi 70:30. Dimana 70% dari dataset akan dijadikan data latih lalu 30% dari dataset akan dijadikan sebagai data uji.

2.3 Modeling

Naive bayes

Algoritma naive bayes merupakan algoritma dari machine learning yang digunakan untuk tujuan klasifikasi berdasarkan teorema bayes, algoritma ini menggunakan prinsip estimasi kemungkinan maksimum untuk mengklasifikasikan sampel ke dalam label yang paling mungkin[16] Adapun persamaan yang terdapat pada algoritma naive bayes adalah sebagai berikut.

$$P(C_i|X) = \frac{\prod_{j=1}^k P(A_j = x_j|C_i) \cdot P(C_i)}{P(X)} \quad (1)$$

Dalam penerapan algoritma naive bayes ini, selalu ada kemungkinan dimana probabilitas suatu kategori bernilai 0, sehingga untuk mengatasi masalah tersebut digunakan lah persamaan Laplacian sebagai berikut.

$$P(C_i|X) = \alpha \frac{N(C_i)}{N(D)} \prod_{j=1}^k \frac{N(A_j = x_j) + 1}{N(D) + q_j} \cdot \frac{N(C = C_i, A_j = x_j) + 1}{N(C_i) + q_j} \quad i = 1, 2, \dots, n \quad (2)$$

K-Nearest Neighbour (KNN)

K-Nearest Neighbour (KNN) juga merupakan algoritma machine learning yang digunakan untuk tujuan klasifikasi, KNN itu sendiri mampu mengklasifikasikan dataset dengan menggunakan model pelatihan yang mencari kemiripan data dengan mempertimbangkan titik data terdekat dengan data yang sedang diuji[17]. Berikut persamaan menghitung jarak terdekat pada algoritma KNN menggunakan rumus Euclidean distance[18].

$$d_i = \sqrt{\sum_{i=1}^n (p_i - q_i)^2} \quad (3)$$

Support Vector Machine (SVM)

Support Vector Machine atau disingkat dengan SVM merupakan algoritma machine learning yang digunakan untuk mengklasifikasikan dataset. Algoritma ini memiliki beberapa variasi persamaan yang bertujuan untuk menyelesaikan masalah yang berdimensi besar sehingga hasil klasifikasinya bergantung pada sejumlah kecil vektor[19]. Berikut persamaan algoritma SVM linier yang digunakan pada penelitian ini.

$$f(x) = \sum_{i=0}^N a_i y_i x_i^T \cdot x + \beta_0 \quad (4)$$

Sementara itu, untuk masalah yang tidak dapat dipisahkan secara linear, persamaan di atas dapat dimodifikasi untuk kernel SVM sebagai berikut.



$$f(x) = \sum_{i=0}^N a_i y_i K(x_i, x) + \beta_0 \quad (5)$$

Decision Tree (C4.5)

Algoritma decision tree model C4.5 merupakan salah satu algoritma pada machine learning yang digunakan untuk membangun pohon keputusan. Cara kerja algoritma ini yaitu dengan membagi dataset menjadi beberapa subset yang lebih kecil berdasarkan fitur yang memberikan informasi paling banyak[20]. Tahapan pada algoritma C4.5 berfokus pada penentuan nilai entropy dengan menggunakan persamaan berikut[21].

$$Entropy(S) = - \sum_{j=1}^c p(S, j) \log p(S, j) \quad (6)$$

Kemudian menentukan nilai gain dari nilai entropy yang sudah didapatkan sebelumnya.

$$Gain(S, T) = Entropy(S) - \sum_{values(T_j)} \frac{|T_{S,v}|}{|T_S|} Entropy(S_v) \quad (7)$$

Pengacuan pustaka dilakukan dengan menuliskan [nomor urut pada daftar pustaka] mis. [1], [1,2], [1,2,3]. Sitasi kepustakaan harus ada dalam Daftar Pustaka dan Daftar Pustaka harus ada sitasinya dalam naskah. Pustaka yang disitasi pertama kali pada naskah [1], harus ada pada daftar pustaka no satu, yang disitasi ke dua, muncul pada daftar pustaka no 2, begitu seterusnya. Daftar pustaka urut kemunculan sitasi, bukan urut nama belakang. Daftar pustaka hanya memuat pustaka yang benar benar disitasi pada naskah.

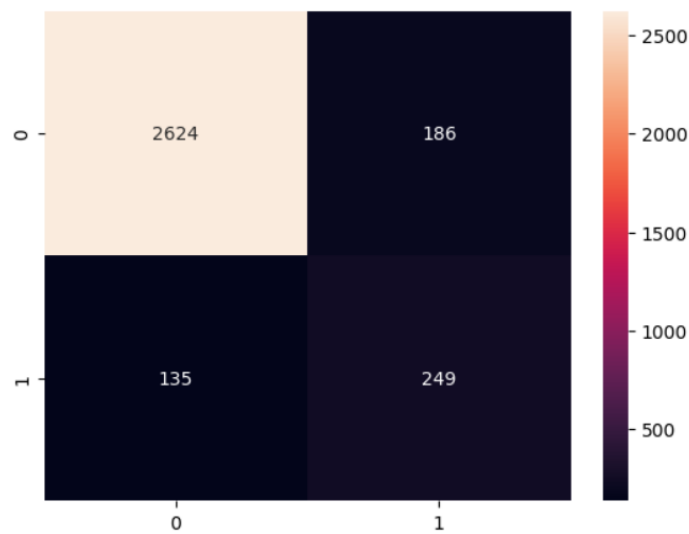
2.4 Evaluation

Pada tahapan ini, hasil pengujian dari beberapa algoritma klasifikasi yang telah ditentukan akan dievaluasi dengan menggunakan beberapak metrik pengujian seperti akurasi, presisi, recall, dan skor F1 serta matrik confusion untuk menentukan algoritma mana yang memiliki performa paling optimal dalam memprediksi diabetes.

3. HASIL DAN PEMBAHASAN

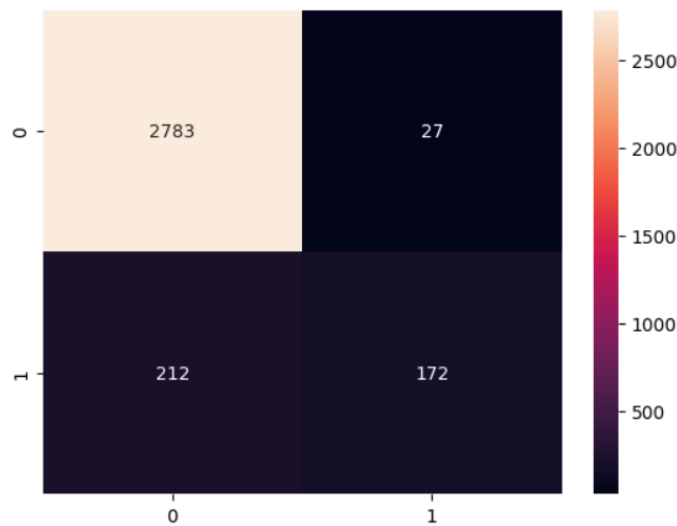
Pada hasil dan pembahasan ini, akan dilakukan evaluasi dengan menunjukkan hasil confusion matrix dari setiap algoritma untuk menilai kinerja setiap model. Dimana hasil evaluasi ini memberikan gambaran terperinci tentang bagaimana algoritma-algoritma tersebut melakukan klasifikasi dan juga membantu dalam memahami kelemahan setiap algoritma lebih mendalam. Berikut confusion matrix dari algortima Naive Bayes.





Gambar 2. Confusion Matrix dari Algoritma Naive Bayes

Dari gambar 2 di atas dapat ditarik kesimpulan bahwa dari total jumlah data yang diuji yaitu sebanyak 3194 data, terdapat 249 data yang berhasil diklasifikasikan dengan benar sebagai data penderita diabetes, 2624 data berhasil diklasifikasikan dengan benar sebagai data yang bukan penderita diabetes. sementara itu, terdapat 186 data yang gagal diklasifikasikan sebagai data penderita diabetes dan 135 data yang juga gagal diklasifikasikan sebagai data yang bukan penderita diabetes. Berdasarkan confusion matrix tersebut di atas diketahui bahwa Algoritma Naive Bayes ini bekerja cukup baik untuk mengklasifikasi data diabetes.

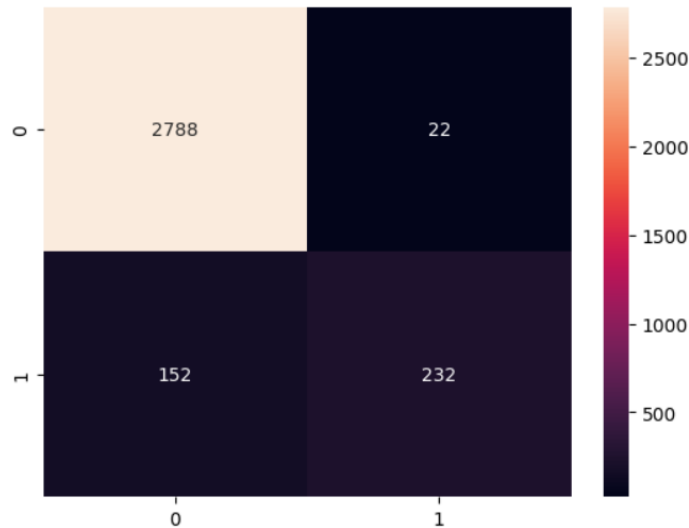


Gambar 3. Confusion Matrix dari Algoritma KNN

Dari gambar 3 di atas dapat ditarik kesimpulan bahwa dari total jumlah data yang diuji yaitu sebanyak 3194 data, terdapat 172 data yang berhasil diklasifikasikan dengan benar sebagai data penderita diabetes, 2783 data berhasil diklasifikasikan dengan benar sebagai data yang bukan penderita diabetes. sementara itu, terdapat 27 data yang gagal diklasifikasikan sebagai data

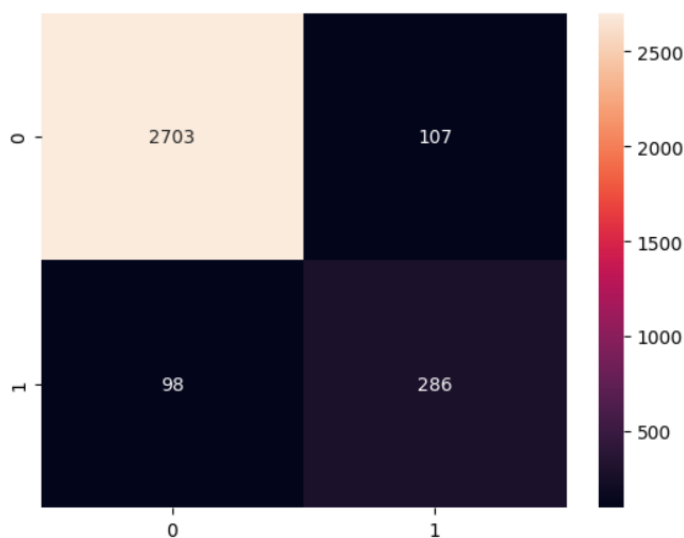


penderita diabetes dan 212 data yang juga gagal diklasifikasikan sebagai data yang bukan penderita diabetes. Berdasarkan confusion matrix tersebut di atas diketahui bahwa Algoritma KNN ini menunjukkan kinerja yang cukup baik untuk mengklasifikasi data diabetes.



Gambar 4. Confusion Matrix dari Algoritma SVM

Dari gambar 4 di atas dapat ditarik kesimpulan bahwa dari total jumlah data yang diuji yaitu sebanyak 3194 data, terdapat 232 data yang berhasil diklasifikasikan dengan benar sebagai data penderita diabetes, 2788 data berhasil diklasifikasikan dengan benar sebagai data yang bukan penderita diabetes. sementara itu, terdapat 22 data yang gagal diklasifikasikan sebagai data penderita diabetes dan 152 data yang juga gagal diklasifikasikan sebagai data yang bukan penderita diabetes. Berdasarkan confusion matrix tersebut di atas diketahui bahwa Algoritma SVM ini menunjukkan performa yang baik untuk mengklasifikasi data diabetes. Selanjutnya, confusion matrix dari algoritma C4.5 sebagai berikut.



Gambar 5. Confusion Matrix dari Algoritma C4.5

Dari gambar 5 di atas dapat ditarik kesimpulan bahwa dari total jumlah data yang diuji yaitu sebanyak 3194 data, terdapat 286 data yang berhasil diklasifikasikan dengan benar sebagai data penderita diabetes, 2703 data berhasil diklasifikasikan dengan benar sebagai data yang bukan penderita diabetes. sementara itu, terdapat 89 data yang gagal diklasifikasikan sebagai data penderita diabetes dan 107 data yang juga gagal diklasifikasikan sebagai data yang bukan penderita diabetes. Berdasarkan confusion matrix tersebut di atas diketahui bahwa Algoritma C4.5 ini juga menunjukkan performa yang baik untuk mengklasifikasi data diabetes. Selanjutnya, Untuk melengkapi evaluasi pada penelitian ini, berikut gambaran nilai precision, recall, f1-score dan persentase akurasi untuk setiap algoritma yang telah diuji.

Tabel 1. Evaluasi Matrik

Algoritma	Label	Precision	Recall	F1-Score	Akurasi
Naive Bayes	Diabetes	0.95	0.93	0.94	90%
	No Diabetes	0.57	0.65	0.61	
KNN	Diabetes	0.93	0.99	0.96	92,5%
	No Diabetes	0.86	0.45	0.59	
SVM	Diabetes	0.95	0.99	0.97	94,5%
	No Diabetes	0.91	0.60	0.73	
C4.5	Diabetes	0.97	0.96	0.96	93,5%
	No Diabetes	0.73	0.74	0.74	

Berdasarkan tabel 1 di atas, semua algoritma memiliki performa yang baik, terbukti hasil akurasi setiap algoritma berada pada nilai lebih besar sama dengan 90%. Dimana algoritma yang memiliki performa yang terbaik adalah algoritma Support Vector Machine (SVM) dengan persentase akurasi sebesar 94,5%.

KESIMPULAN DAN SARAN

Berdasarkan analisis komparasi dari algoritma klasifikasi machine learning yang dilakukan, dapat dilihat bahwa setiap algoritma memiliki kinerja yang bervariasi. Algoritma K-Nearest Neighbors (KNN) memiliki performa sebesar 92,5 %, selanjutnya algoritma Support Vector Machine (SVM) memiliki performa sebesar 94,5%, kemudian algoritma Naive Bayes memiliki performa sebesar 90% dan yang terakhir algoritma Decision Tree (C4.5) memiliki performa sebesar 93,5%.

Pada algoritma SVM memiliki akurasi tertinggi di 94,5% dengan f1-score 73% untuk label Diabetes. Meskipun terdapat peningkatan, namun jumlah kesalahan dalam klasifikasi pada label diabetes masih menunjukkan bahwa ada potensi untuk perbaikan. Dengan mempertimbangkan berbagai metrik, parameter algoritma dapat lebih dioptimalkan dan melakukan eksplorasi menggunakan teknik lain sehingga mampu meningkatkan kemampuan prediksi pada label minoritas untuk mendapatkan hasil yang lebih seimbang dan akurat.



UCAPAN TERIMA KASIH

Terima kasih peneliti sampaikan kepada Universitas Samudra yang telah memberikan banyak dukungan dan kepada rekan-rekan peneliti yang juga telah berkontribusi memberikan sumbangsuhnya sehingga artikel ilmiah ini dapat diselesaikan serta dipublikasikan tepat waktu,

DAFTAR PUSTAKA

- [1] H. Feldman *et al.*, “Standards of Care in Diabetes—2023 Abridged for Primary Care Providers,” *Clinical Diabetes*, vol. 41, no. 1, pp. 4–31, 2023, doi: 10.2337/CD23-AS01.
- [2] “Diabetes.” Accessed: Mar. 14, 2025. [Online]. Available: <https://www.who.int/news-room/fact-sheets/detail/diabetes>
- [3] “Diagnosis and classification of diabetes mellitus,” *Diabetes Care*, vol. 37, no. SUPPL.1, Jan. 2014, doi: 10.2337/DC14-S081.
- [4] S. Rafatirad, H. Homayoun, Z. Chen, and S. M. P. Dinakarrao, “Machine learning for computer scientists and data analysts: From an applied perspective,” *Machine Learning for Computer Scientists and Data Analysts: From an Applied Perspective*, pp. 1–458, Jul. 2022, doi: 10.1007/978-3-030-96756-7/COVER.
- [5] L. Oluwaseye, J. * A1, W. Doorsamy, and S. Paul, “A Review of Missing Data Handling Techniques for Machine Learning,” *International Journal of Innovative Technology and Interdisciplinary Sciences*, vol. 5, no. 3, pp. 971–1005, Sep. 2022, doi: 10.15157/IJITIS.2022.5.3.971-1005.
- [6] M. Unik, M. E. Aldo Fahrurrozi, H. Harmaini, and C. Kusuma, “MODEL PREDIKSI CURAH HUJAN MENGGUNAKAN ALGORITMA MACHINE LEARNING YANG DITINGKATKAN,” *Jurnal CoSciTech (Computer Science and Information Technology)*, vol. 5, no. 2, pp. 439–448, Aug. 2024, doi: 10.37859/COSCITECH.V5I2.7629.
- [7] M. M. İlman, Ş. Yavuz, and P. Y. Taser, “Generalized Input Preshaping Vibration Control Approach for Multi-Link Flexible Manipulators using Machine Intelligence,” *Mechatronics*, vol. 82, Apr. 2022, doi: 10.1016/J.MECHATRONICS.2021.102735.
- [8] L. Jen and Y. H. Lin, “A Brief Overview of the Accuracy of Classification Algorithms for Data Prediction in Machine Learning Applications,” *Journal of Applied Data Sciences*, vol. 2, no. 3, pp. 84–92, Sep. 2021, doi: 10.47738/JADS.V2I3.38.
- [9] A. Alanazi, “Using machine learning for healthcare challenges and opportunities,” *Inform Med Unlocked*, vol. 30, p. 100924, Jan. 2022, doi: 10.1016/J.IMU.2022.100924.
- [10] M. Javaid, A. Haleem, R. Pratap Singh, R. Suman, and S. Rab, “Significance of machine learning in healthcare: Features, pillars and applications,” *International Journal of Intelligent Networks*, vol. 3, pp. 58–73, Jan. 2022, doi: 10.1016/J.IJIN.2022.05.002.
- [11] M. Bari Antor *et al.*, “A Comparative Analysis of Machine Learning Algorithms to Predict Alzheimer’s Disease,” *J Healthc Eng*, vol. 2021, no. 1, p. 9917919, Jan. 2021, doi: 10.1155/2021/9917919.
- [12] M. Swathy and K. Saruladha, “A comparative study of classification and prediction of Cardio-Vascular Diseases (CVD) using Machine Learning and Deep Learning techniques,” *ICT Express*, vol. 8, no. 1, pp. 109–116, Mar. 2022, doi: 10.1016/J.ICTE.2021.08.021.
- [13] A. Dairi, F. Harrou, A. Zeroual, M. M. Hittawe, and Y. Sun, “Comparative study of machine learning methods for COVID-19 transmission forecasting,” *J Biomed Inform*, vol. 118, p. 103791, Jun. 2021, doi: 10.1016/J.JBI.2021.103791.
- [14] R. Katarya and S. K. Meena, “Machine Learning Techniques for Heart Disease Prediction: A Comparative Study and Analysis,” *Health Technol (Berl)*, vol. 11, no. 1, pp. 87–97, Jan. 2021, doi: 10.1007/S12553-020-00505-7/METRICS.



- [15] C. C. Olisah, L. Smith, and M. Smith, "Diabetes mellitus prediction and diagnosis from a data preprocessing and machine learning perspective," *Comput Methods Programs Biomed*, vol. 220, p. 106773, Jun. 2022, doi: 10.1016/J.CMPB.2022.106773.
- [16] H. Chen, S. Hu, R. Hua, and X. Zhao, "Improved naive Bayes classification algorithm for traffic risk management," *EURASIP J Adv Signal Process*, vol. 2021, no. 1, pp. 1–12, Dec. 2021, doi: 10.1186/S13634-021-00742-6/TABLES/5.
- [17] S. Uddin, I. Haque, H. Lu, M. A. Moni, and E. Gide, "Comparative performance analysis of K-nearest neighbour (KNN) algorithm and its different variants for disease prediction," *Scientific Reports 2022 12:1*, vol. 12, no. 1, pp. 1–11, Apr. 2022, doi: 10.1038/s41598-022-10358-x.
- [18] O. Saeful Bachri, R. Mohamad, and H. Bhakti, "Penentuan Status Stunting pada Anak dengan Menggunakan Algoritma KNN," *Jurnal Ilmiah Intech : Information Technology Journal of UMUS*, vol. 3, no. 02, pp. 130–137, Nov. 2021, doi: 10.46772/INTECH.V3I02.533.
- [19] B. Gaye, D. Zhang, and A. Wulamu, "Improvement of Support Vector Machine Algorithm in Big Data Background," *Math Probl Eng*, vol. 2021, no. 1, p. 5594899, Jan. 2021, doi: 10.1155/2021/5594899.
- [20] X. Zheng, W. Feng, M. Huang, and S. Feng, "Optimization of PBFT Algorithm Based on Improved C4.5," *Math Probl Eng*, vol. 2021, no. 1, p. 5542078, Jan. 2021, doi: 10.1155/2021/5542078.
- [21] A. Saleh, M. Maryam, and K. Puspita, "Determination of Corn Quality using the Decision Tree of C 4.5 Algorithm," *2019 7th International Conference on Cyber and IT Service Management, CITSM 2019*, Nov. 2019, doi: 10.1109/CITSM47753.2019.8965334.

