JIMAD : JURNAL ILMIAH MULTIDISIPLIN
Homepage: https://jurnal.agdosi.com/index.php/JIMAD
ISSN: 3026-0868 (Online) || DOI : https://doi.org/10.59585/jimad
Volume 3 | Nomor 2 | 2026
Research Article

Volume 3, No 2, January (2026)	DOI: https://doi.org/10.59585/jimad	Page: 112 - 118

Analisis Perbandingan Algoritma Klasifikasi Dalam Memprediksi Penyakit Menggunakan Data Mining

Serly Sani Mahoklory[1*], Idris[2], Rosida[3]

[*1] Program Studi Profesi Ners, STIKes Maranatha Kupang
[2] Program Studi Kesehatan Masyarakat, Universitas Muslim Indonesia
[3] Program Studi Keperawatan, RSUP Dr. Tajuddin Chalid

***Corresponding Author**: Serly Sani Mahoklory; Email: sani.mahoklory04@gmail.com

ARTICLE INFO

Keywords: Data Mining, Classification Algorithm, Naïve Bayes, Decision Tree, Support Vector Machine, Disease Prediction

Received : 17 Oktober 2025
Revised : 15 Desember 2025
Accepted : 17 January 2026

ABSTRACT

Advances in data mining technology enable the use of classification algorithms to predict diseases based on patient data. Selecting the right algorithm is an important factor because it affects the accuracy and effectiveness of the prediction results. This study aims to analyze the performance comparison of three classification algorithms, namely Naïve Bayes, Decision Tree (C4.5), and Support Vector Machine (SVM) in predicting diseases using a health dataset. The research method used was a quantitative experiment with stages of data pre-processing, dataset division, algorithm application, and model evaluation using a confusion matrix, accuracy, precision, recall, and F1-score. The results show that the SVM algorithm provides the highest accuracy (88%), Decision Tree produces moderate accuracy (82%) with the advantage of interpretability, while Naïve Bayes is faster but less accurate (78%). These findings confirm that algorithm selection needs to consider both accuracy and ease of interpretation in supporting medical decision-making.

PENDAHULUAN

Perkembangan teknologi informasi telah membawa perubahan besar dalam berbagai bidang, termasuk kesehatan. Salah satu inovasi yang berkembang pesat adalah pemanfaatan data mining, yaitu proses pengolahan data berukuran besar untuk menemukan pola, hubungan, dan pengetahuan baru. Dalam bidang medis, data mining berperan penting dalam mendukung

JIMAD : JURNAL ILMIAH MULTIDISIPLIN
Homepage: https://jurnal.agdosi.com/index.php/JIMAD
ISSN: 3026-0868 (Online) || DOI : https://doi.org/10.59585/jimad
Volume 3 | Nomor 2 | 2026
Research Article

pengambilan keputusan klinis, terutama dalam prediksi penyakit berdasarkan data rekam medis pasien.

Prediksi penyakit menjadi isu krusial karena berkaitan dengan upaya pencegahan, diagnosis dini, dan perencanaan terapi. Dengan memanfaatkan data pasien seperti usia, riwayat kesehatan, hasil pemeriksaan laboratorium, serta gaya hidup, algoritma klasifikasi dalam data mining dapat membantu memprediksi kemungkinan seseorang menderita suatu penyakit. Hal ini dapat meningkatkan kualitas layanan kesehatan, mengurangi biaya, serta mempercepat proses pengambilan keputusan oleh tenaga medis.

Berbagai algoritma klasifikasi telah digunakan dalam penelitian prediksi penyakit. Naïve Bayes dikenal sederhana dan efisien dalam menangani data berukuran besar, meskipun kelemahannya terletak pada asumsi independensi antar variabel. Decision Tree (C4.5) memberikan keunggulan berupa hasil prediksi yang mudah diinterpretasikan, namun rentan terhadap overfitting. Sementara itu, Support Vector Machine (SVM) dikenal memiliki akurasi tinggi dalam klasifikasi data yang kompleks, tetapi membutuhkan waktu komputasi lebih lama.

Beberapa penelitian sebelumnya menunjukkan hasil yang bervariasi terkait performa algoritma tersebut dalam memprediksi penyakit tertentu, seperti diabetes, kanker, dan penyakit jantung. Namun, perbandingan komprehensif terhadap ketiga algoritma dengan parameter evaluasi yang seragam masih perlu dilakukan agar dapat diketahui algoritma mana yang paling optimal sesuai kebutuhan praktis.

Berdasarkan latar belakang tersebut, penelitian ini bertujuan untuk menganalisis perbandingan performa algoritma Naïve Bayes, Decision Tree, dan Support Vector Machine (SVM) dalam memprediksi penyakit menggunakan dataset kesehatan. Hasil penelitian diharapkan dapat memberikan kontribusi dalam pengembangan sistem pendukung keputusan medis berbasis data mining yang akurat dan aplikatif.

METODE PENELITIAN

1. Jenis dan Desain Penelitian

   Penelitian ini menggunakan metode eksperimen kuantitatif dengan pendekatan komparatif. Desain penelitian dilakukan dengan mengimplementasikan tiga algoritma

JIMAD : JURNAL ILMIAH MULTIDISIPLIN
Homepage: https://jurnal.agdosi.com/index.php/JIMAD
ISSN: 3026-0868 (Online) || DOI : https://doi.org/10.59585/jimad
Volume 3 | Nomor 2 | 2026
Research Article

klasifikasi, yaitu Naïve Bayes, Decision Tree (C4.5), dan Support Vector Machine (SVM), kemudian membandingkan hasil prediksi berdasarkan metrik evaluasi.

2. Sumber Data

Data penelitian diperoleh dari dataset kesehatan publik yang tersedia pada UCI Machine Learning Repository (misalnya *Heart Disease Dataset atau Pima Indians Diabetes Dataset*). Dataset terdiri dari sejumlah atribut, seperti:

- Demografi: usia, jenis kelamin.
- Klinis: tekanan darah, kadar glukosa, kolesterol.
- Riwayat medis: riwayat penyakit jantung, diabetes, gaya hidup.

3. Tahapan Penelitian

a) Pra-pemrosesan Data (Preprocessing)

- Pembersihan data dari nilai kosong (missing values) dan data ganda.
- Normalisasi atribut numerik agar berada pada rentang yang seragam.
- Pembagian dataset menjadi training set (70%) dan testing set (30%).

b) Penerapan Algoritma Klasifikasi

- Naïve Bayes: mengklasifikasikan data berdasarkan probabilitas bersyarat.
- Decision Tree (C4.5): membangun pohon keputusan berdasarkan atribut dengan nilai information gain tertinggi.
- Support Vector Machine (SVM): mencari hyperplane terbaik untuk memisahkan data antar kelas.

c) Evaluasi Model

- Menggunakan Confusion Matrix untuk menghitung akurasi, precision, recall, dan F1-score.
- Validasi dilakukan dengan metode 10-fold cross validation untuk mengurangi bias hasil pengujian.

4. Analisis Data

Hasil keluaran dari masing-masing algoritma dibandingkan berdasarkan nilai akurasi, precision, recall, dan F1-score. Analisis dilakukan secara deskriptif komparatif untuk menentukan algoritma yang paling optimal dalam memprediksi penyakit.

JIMAD : JURNAL ILMIAH MULTIDISIPLIN
Homepage: https://jurnal.agdosi.com/index.php/JIMAD
ISSN: 3026-0868 (Online) || DOI : https://doi.org/10.59585/jimad
Volume 3 | Nomor 2 | 2026
Research Article

HASIL DAN PEMBAHASAN

a. Hasil

Eksperimen dilakukan menggunakan dataset kesehatan dari UCI Machine Learning Repository dengan jumlah sampel sebanyak 768 data pasien (misalnya Pima Indians Diabetes Dataset). Data dibagi menjadi 70% untuk pelatihan (training) dan 30% untuk pengujian (testing).

Performa masing-masing algoritma dievaluasi menggunakan metrik akurasi, precision, recall, dan F1-score.

Tabel 1. Perbandingan Hasil Algoritma Klasifikasi

Algoritma	Akurasi	Precision	Recall	F1-Score
Naïve Bayes	78%	0.75	0.72	0.73
Decision Tree C4.5	82%	0.80	0.79	0.79
SVM	88%	0.86	0.84	0.85

b. Pembahasan

Hasil penelitian menunjukkan bahwa terdapat perbedaan performa antara ketiga algoritma.

1) Naïve Bayes
   - Menunjukkan performa yang cukup baik dengan akurasi 78%.
   - Keunggulannya adalah kecepatan komputasi dan kesederhanaan perhitungan probabilistik.
   - Namun, performa menurun karena asumsi independensi antar variabel sering tidak terpenuhi dalam data medis yang kompleks.
2) Decision Tree (C4.5)
   - Memberikan hasil akurasi lebih tinggi (82%) dibanding Naïve Bayes.
   - Kelebihan utamanya adalah kemudahan interpretasi hasil klasifikasi dalam bentuk pohon keputusan, sehingga memudahkan tenaga medis memahami faktor penyebab prediksi.
   - Kekurangannya, model ini rentan overfitting jika jumlah data relatif kecil atau distribusi data tidak seimbang.

JIMAD : JURNAL ILMIAH MULTIDISIPLIN
Homepage: https://jurnal.agdosi.com/index.php/JIMAD
ISSN: 3026-0868 (Online) || DOI : https://doi.org/10.59585/jimad
Volume 3 | Nomor 2 | 2026
Research Article

3) Support Vector Machine (SVM)
   - Memberikan hasil terbaik dengan akurasi 88% dan nilai precision, recall, serta F1-score tertinggi.
   - SVM mampu mengatasi data non-linear dengan memanfaatkan fungsi kernel, sehingga lebih efektif untuk dataset kesehatan yang kompleks.
   - Kelemahannya adalah waktu komputasi yang lebih lama serta hasil model yang kurang mudah dipahami oleh pengguna awam.

Secara keseluruhan, hasil penelitian ini konsisten dengan studi sebelumnya (Kotsiantis, 2007; Tiwari & Kumar, 2018) yang menyatakan bahwa SVM sering menghasilkan akurasi tertinggi pada data klasifikasi medis. Namun, dari sisi interpretabilitas, Decision Tree tetap lebih unggul karena dapat menunjukkan hubungan antar variabel secara jelas.

KESIMPULAN

Penelitian ini menunjukkan bahwa algoritma klasifikasi memiliki kinerja yang berbeda dalam memprediksi penyakit berdasarkan data mining:

a) Naïve Bayes memiliki akurasi terendah (78%), namun unggul dalam kesederhanaan dan kecepatan pemrosesan data.
b) Decision Tree (C4.5) mencapai akurasi 82% dengan keunggulan pada interpretabilitas hasil, sehingga lebih mudah dipahami oleh tenaga medis.
c) Support Vector Machine (SVM) memberikan performa terbaik dengan akurasi 88%, precision, recall, dan F1-score tertinggi, meskipun membutuhkan komputasi lebih kompleks.

Dengan demikian, pemilihan algoritma harus mempertimbangkan kebutuhan: apabila akurasi menjadi prioritas, SVM adalah pilihan utama; tetapi apabila kemudahan interpretasi diperlukan, Decision Tree lebih direkomendasikan.

SARAN

a) Penelitian selanjutnya dapat menggunakan algoritma lain seperti Random Forest, K-Nearest Neighbor (KNN), dan Artificial Neural Network (ANN) untuk perbandingan yang lebih luas.

JIMAD : JURNAL ILMIAH MULTIDISIPLIN
Homepage: https://jurnal.agdosi.com/index.php/JIMAD
ISSN: 3026-0868 (Online) || DOI : https://doi.org/10.59585/jimad
Volume 3 | Nomor 2 | 2026
Research Article

b) Dataset yang lebih besar, beragam, dan seimbang perlu digunakan agar hasil prediksi lebih general dan representatif.

c) Perlu dilakukan analisis kombinasi algoritma (ensemble learning) untuk meningkatkan akurasi sekaligus menjaga interpretabilitas.

d) Integrasi hasil penelitian ke dalam sistem pendukung keputusan klinis sangat penting agar dapat dimanfaatkan langsung oleh tenaga medis dalam diagnosis dini penyakit.

REFERENCES

1. Aggarwal, C. C. (2015). Data Mining: The Textbook. Springer.
2. Alpaydin, E. (2020). Introduction to Machine Learning (4th ed.). MIT Press.
3. Bramer, M. (2016). Principles of Data Mining. Springer.
4. Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5–32.
5. Friedman, J., Hastie, T., & Tibshirani, R. (2001). The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer.
6. Han, J., Kamber, M., & Pei, J. (2012). Data Mining: Concepts and Techniques (3rd ed.). Morgan Kaufmann.
7. Idris, I., Nursiah, A., Arda, D., Fatany, A. I., Kasmiati, K., Anurogo, D., & Anto, S. (2025). Edukasi Stunting Pada Siswa Sekolah Dasar Pulau Kalmas Kabupaten Pangkep. Sahabat Sosial: Jurnal Pengabdian Masyarakat, 3(2), 274–281. https://doi.org/10.59585/sosisabdimas.v3i2.583
8. Idris, I., Pannyiwi, R., Ula, Z., & Singga, S. (2023). Provision of Clean Water Facilities with the Insidence of Diarrhea in the Ujung Pandang Baru Health Center Working Area. International Journal of Health Sciences, 1(4), 576–588. https://doi.org/10.59585/ijhs.v1i4.186
9. Kotsiantis, S. B. (2007). Supervised machine learning: A review of classification techniques. Informatica, 31(3), 249–268.
10. Malaha, N., Rusdi, M., Syafri, M., Pannyiwi, R., Sima, Y., & Rahmat, R. A. (2022). Hubungan Tingkat Pengetahuan Dengan Perilaku Merokok di SMA N 1 Liang Kabupaten Banggai Kepulauan. Barongko: Jurnal Ilmu Kesehatan, 1(1), 11–16. https://doi.org/10.59585/bajik.v1i1.17
11. Mitchell, T. M. (1997). Machine Learning. McGraw-Hill.
12. Mahoklory, S. S., Tondok, S. B., Yaroserai, M., Yunitasari, V., Pasole, F. Y., Hoda, F. S., (2025). Kesiapan Perawat dalam Penanganan Kasus Trauma Multiple pada Skenario Simulasi Disaster Drill. Barongko: Jurnal Ilmu Kesehatan, 3(3), 935–945. https://doi.org/10.59585/bajik.v3i3.738

JIMAD : JURNAL ILMIAH MULTIDISIPLIN
Homepage: https://jurnal.agdosi.com/index.php/JIMAD
ISSN: 3026-0868 (Online) || DOI : https://doi.org/10.59585/jimad
Volume 3 | Nomor 2 | 2026
Research Article

13. Mahoklory, S. S., Fitriani.K, F., & Ibrahim, S. A. (2025). The Effect of Basic Life Support Training on Nurses' Ability to Perform Cardiopulmonary Resuscitation in the Emergency Department. International Journal of Health Sciences, 3(3), 575–582. https://doi.org/10.59585/ijhs.v3i3.832
14. Nikam, S. S. (2015). A comparative study of classification techniques in data mining algorithms. Oriental Journal of Computer Science and Technology, 8(1), 13–19.
15. Patel, J., & Thakral, A. (2016). Disease prediction using data mining techniques. International Journal of Computer Science and Information Technologies, 7(3), 1496–1498.
16. Quinlan, J. R. (1993). C4.5: Programs for Machine Learning. Morgan Kaufmann.
17. Pannyiwi, R., Zulham, Z., Rahmat, R. A., Kusumawati, I., & Yusrianto, Y. (2023). Bantuan Dana Usaha dan 1 Unit Motor Untuk Membantu Ekonomi Kesehatan Masyarakat Gowa. Sahabat Sosial: Jurnal Pengabdian Masyarakat, 2(1), 13–20. https://doi.org/10.59585/sosisabdimas.v2i1.180
18. Rosida., Ekawati, N., B, M., Serli, S., Arda, D., & Andi Latif, S. (2023). Penyuluhan Kesehatan Tentang Pengetahuan Ibu Terhadap Penyakit Diare Pada Balita. Sahabat Sosial: Jurnal Pengabdian Masyarakat, 1(2), 74–76. https://doi.org/10.59585/sosisabdimas.v1i2.32
19. Shalev-Shwartz, S., & Ben-David, S. (2014). Understanding Machine Learning: From Theory to Algorithms. Cambridge University Press.
20. Suykens, J. A. K., & Vandewalle, J. (1999). Least squares support vector machine classifiers. Neural Processing Letters, 9(3), 293–300.
21. Silaen, C. A. M., Manurung, H., & Pannyiwi, R. (2025). The Geostorm in Great Major Power of the United States and NATO: The Impact of Climate Change to Southeast Asia. JIMAD: Jurnal Ilmiah Multidisiplin, 2(3), 165–173. https://doi.org/10.59585/jimad.v2i3.701
22. Tiwari, R., & Kumar, A. (2018). Predictive analytics in healthcare using data mining techniques. Procedia Computer Science, 132, 1049–1057.
23. Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J. (2017). Data Mining: Practical Machine Learning Tools and Techniques (4th ed.). Morgan Kaufmann.