Department of Digital Business
Journal of Artificial Intelligence and Digital Business (RIGGS)
Homepage: https://journal.ilmudata.co.id/index.php/RIGGS
Vol. 4 No. 3 (2025) pp: 3621-3626
P-ISSN: 2963-9298, e-ISSN: 2963-914X

Optimalisasi Preprocessing Data Menggunakan Pendekatan CRISP-DM untuk Meningkatkan Kualitas Klasifikasi Penyakit Jantung

Wahyu Nursahid[1], Bangkit Indarmawan Nugroho[2], Syefudin Syefudin[3]
[1,3]Teknik Informatika, STMIK YMI Tegal
[2]Sistem Informasi, STMIK YMI Tegal
[1]wahjoenoershaheed@gmail.com, [2]bangkit_in@stmik-tegal.ac.id, [3]syefudin@stmik-tegal.ac.id

Abstrak

Penyakit jantung masih menjadi penyebab kematian utama di dunia sehingga deteksi dini berbasis data medis menjadi sangat penting. Penelitian ini menerapkan kerangka CRISP-DM untuk membangun klasifikasi penyakit jantung dengan pendekatan sistematis yang menekankan optimalisasi preprocessing data. Algoritma k-Nearest Neighbor (KNN) digunakan sebagai model dasar, dengan empat skema preprocessing yang dibandingkan: baseline dengan encoding, penambahan normalisasi, integrasi seleksi fitur berbasis Information Gain, serta kombinasi normalisasi, SMOTE dan seleksi fitur. Proses evaluasi dilakukan menggunakan 10 fold cross validation dengan metrik akurasi, presisi, recall, F1-score dan AUC. Hasil eksperimen menunjukkan bahwa skema keempat memberikan performa terbaik dengan akurasi 81,26 persen dan AUC 0,8460, melampaui skema lainnya. Fakta ini menegaskan bahwa strategi preprocessing yang tepat berkontribusi signifikan terhadap peningkatan performa model. Implikasi penelitian ini adalah perlunya menempatkan preprocessing sebagai bagian integral dari kerangka CRISP-DM, bukan sekadar langkah tambahan, serta membuka peluang penelitian lanjutan untuk mengeksplorasi variasi teknik preprocessing yang lebih adaptif. Pada tataran implementasi nyata, kombinasi preprocessing terbaik dengan algoritma yang lebih kuat dapat dipertimbangkan guna menghasilkan sistem prediksi penyakit jantung yang akurat dan andal.

Kata kunci: *Preprocessing, CRISP-DM, Klasifikasi, KNN, Penyakit Jantung*

1. Latar Belakang

Penyakit jantung merupakan salah satu penyebab kematian tertinggi secara global maupun nasional. Menurut *World Health Organization* (WHO), penyakit kardiovaskular menyebabkan sekitar 17,9 juta kematian setiap tahunnya di seluruh dunia [1]. Di Indonesia, jumlah kematian akibat penyakit jantung diperkirakan melebihi 295.000 jiwa per tahun [2]. Kondisi ini menuntut adanya strategi deteksi dini yang lebih akurat dan efisien untuk membantu tenaga medis dalam mengidentifikasi risiko sejak awal. Diagnosis konvensional sering kali terlambat karena gejala baru muncul pada tahap lanjut, sehingga pendekatan berbasis analisis data menjadi semakin penting.

Kemajuan teknologi pembelajaran mesin telah membuka peluang baru dalam mendukung diagnosis medis. Model klasifikasi memungkinkan identifikasi pola pada data pasien yang tidak mudah terlihat melalui pemeriksaan klinis biasa [3]. Salah satu metodologi yang paling terkenal dan banyak digunakan dalam pembelajaran mesin yaitu kerangka CRISP-DM (*Cross Industry Standard Process for Data Mining***)** karena fleksibilitasnya dalam berbagai kasus *data mining* serta kemampuannya menjamin penelitian dapat direplikasi dengan jelas [4].

Dalam penelitian ini digunakan algoritma *K-Nearest Neighbor* (KNN) karena kesederhanaannya, kemudahan interpretasi, serta sifatnya yang sensitif terhadap data [5]. Karakteristik tersebut menjadikan KNN sesuai dengan fokus penelitian ini yaitu mengevaluasi pengaruh berbagai teknik preprocessing terhadap kualitas prediksi penyakit jantung. Sehingga KNN dipilih bukan hanya sebagai algoritma klasifikasi, tetapi juga sebagai alat evaluasi untuk menilai sejauh mana strategi preprocessing dapat meningkatkan performa model.

Beberapa penelitian terdahulu telah mengaplikasikan pembelajaran mesin untuk prediksi dengan algoritma seperti *Decision Tree*, *Random Forest* maupun *Logistic Regression* dan melaporkan akurasi yang cukup baik [6], [7], [8]. Namun, sebagian besar studi lebih berfokus pada perbandingan algoritma dan belum menelaah secara mendalam bagaimana desain *preprocessing* mempengaruhi performa model, khususnya dalam kerangka CRISP-DM dengan KNN. Kesenjangan tersebut membuka peluang penelitian untuk mengeksplorasi strategi *preprocessing* secara sistematis. Penelitian ini bertujuan menganalisis pengaruh berbagai skema *preprocessing* terhadap performa KNN dalam prediksi penyakit jantung dengan kerangka CRISP-DM, sekaligus menegaskan bahwa *preprocessing* merupakan komponen strategis yang menentukan kualitas model. Hasil yang diperoleh diharapkan berkontribusi

pada pengembangan sistem prediksi medis yang lebih akurat dan dapat diimplementasikan dalam sistem pendukung keputusan di bidang kesehatan. Penelitian ini dibatasi pada penggunaan algoritma *K-Nearest Neighbors* (KNN) dengan variasi *preprocessing data*, menggunakan dataset penyakit jantung dari Kaggle. Implementasi dilakukan dengan RapidMiner dan Google Colab, sehingga temuan penelitian ini masih terbatas pada karakteristik data serta perangkat lunak yang digunakan.

2. Metode Penelitian

2.1. Kerangka Penelitian

Penelitian ini mengacu pada metodologi CRISP-DM (*Cross Industry Standard Process for Data Mining*) sebagai kerangka kerja utama. Tahapan CRISP-DM dipilih karena memberikan alur sistematis dalam proses *data mining* mulai dari pemahaman masalah hingga evaluasi model [9]. Alur penelitian berdasarkan tahapan tersebut ditunjukkan pada Gambar 1.

Gambar 1. Alir Penelitian Berbasis CRISP-DM

Gambar 1 menggambarkan enam tahapan penelitian yang diterapkan. Pada tahap *business understanding*, masalah yang diangkat adalah tingginya prevalensi penyakit jantung dan solusi yang ditawarkan berupa penerapan pembelajaran mesin untuk prediksi dini. Tahap *data understanding* mengeksplorasi dataset yang digunakan pada penelitian ini. Tahap *data preparation* mencakup penerapan empat skema *preprocessing* yang terdiri dari pembersihan data, *encoding*, normalisasi, seleksi fitur *Information Gain* dan penyeimbang data dengan SMOTE. Pada tahap *modeling*, algoritma *K-Nearest Neighbor* (KNN) dipilih karena kesederhanaan dan sensitivitasnya terhadap preprocessing. Tahap *evaluation* dilakukan dengan validasi silang 10 bagian menggunakan metrik akurasi, presisi, *recall*, *F1-score* dan AUC. Akhirnya, pada tahap *deployment*, hasil penelitian disajikan sebagai kontribusi akademik yang dapat menjadi dasar pengembangan sistem prediksi kesehatan berbasis aplikasi.

2.2. *Dataset*

Dataset yang digunakan bersumber dari *website* Kaggle (https://www.kaggle.com/datasets/fedesoriano/heart-failure-prediction) terdiri dari 918 catatan pasien dengan 11 atribut prediktor dan satu atribut target. Atribut meliputi *Age*, *ChesTPainType*, *RestingBP*, *FastingBP*, *Cholesterol*, *MaxHR* dan *Oldpeak* serta fitur kategorikal seperti *Sex*, *ExerciseAngina*, *RestingECG* dan *ST_Slope*. Fitur target *HeartDisease* menunjukkan status penyakit (1 = positif, 0 = negatif). Beberapa nilai nol pada kolom *RestingBP* dan *Cholesterol* mengindikasikan data tidak valid yang ditangani melalui proses imputasi pada tahap *cleaning* [10]. Berikut sebagian *dataset* ditunjukkan pada Tabel 1.

DOI: https://doi.org/10.31004/riggs.v4i3.2514

Tabel 1. Dataset Penyakit Jantung

No.	*Age*	*Sex*	*ChestPainType*	...	*Oldpeak*	*ST_Slope*	*HeartDisease*
1	55	*M*	*NAP*	…	1.5	*FLAT*	1
2	53	*M*	*ASY*	…	2	*DOWN*	0
3	38	*M*	*ASY*	…	2.5	*FLAT*	1
4	39	*F*	*NAP*	…	0	*UP*	0
5	51	*M*	*NAP*	…	0	*UP*	0
6	32	*M*	*TA*	…	0.7	*UP*	1
7	51	*M*	*ASY*	…	2.2	*FLAT*	1
8	57	*M*	*ASY*	…	0.7	*DOWN*	1
9	52	*M*	*ASY*	…	0.8	*FLAT*	1
10	40	*M*	*ASY*	…	0	*UP*	1
11	64	*F*	*ASY*	…	1.1	*DOWN*	1
12	63	*M*	*ASY*	…	1	*UP*	1
13	34	*M*	*ATA*	…	0	*UP*	0
14	60	*M*	*ASY*	…	1	*FLAT*	1
15	43	*F*	*TA*	…	0	*UP*	0
16	49	*M*	*ATA*	…	0	*UP*	0
⋮	⋮	⋮	⋮	…	⋮	⋮	⋮
916	64	*F*	*ASY*	…	1	*FLAT*	1
917	61	*M*	*NAP*	…	0	*FLAT*	1
918	56	*F*	*ASY*	…	4	*DOWN*	1

2.3. *Preprocessing Data*

Tahapan *preprocessing* dirancang dalam empat skema berbeda untuk menguji kontribusi masing-masing teknik terhadap kinerja KNN. Rancangan keempat skema tersebut ditampilkan pada Tabel 2.

Tabel 2. Skema Preprocessing

Skema	*Cleaning*	Transformasi	Reduksi
1	Tidak	*Encoding*	Tidak
2	Ya	*Encoding*, Normalisasi	Tidak
3	Ya	*Encoding*, Normalisasi	*Information Gain*
4	Ya	*Encoding*, Normalisasi, SMOTE	*Information Gain*

Tabel 2 memperlihatkan variasi kombinasi *preprocessing* yang diterapkan. *Encoding* diterapkan sebagai *baseline* karena atribut kategorikal harus diubah ke bentuk numerik agar dapat diproses algoritma KNN [11]. Normalisasi ditambahkan pada skema berikutnya untuk menyeragamkan skala fitur, mengingat KNN berbasis perhitungan jarak yang sangat sensitif terhadap perbedaan rentang nilai [12]. Seleksi fitur berbasis *Information Gain* digunakan untuk mempertahankan atribut yang paling relevan dengan target, sehingga kompleksitas model berkurang tanpa mengorbankan performa [13]. Sementara itu, pada skema terakhir dilakukan kombinasi normalisasi, *Synthetic Minority Over-sampling Technique* (SMOTE), dan seleksi fitur, karena distribusi kelas pada data tidak seimbang [14]. Dengan pendekatan ini, model tidak hanya belajar dari data yang lebih proporsional, tetapi juga fokus pada fitur yang paling bermakna sehingga performanya dapat ditingkatkan secara optimal.

2.4. Algoritma Klasifikasi KNN

Model klasifikasi penelitian ini dibangun menggunakan algoritma *K-Nearest Neighbor* (KNN) dengan nilai k = 5 yang merupakan nilai *default* pada aplikasi RapidMiner . Prinsip kerja KNN adalah mengklasifikasikan sampel baru berdasarkan mayoritas kelas dari sejumlah tetangga terdekat, dengan kedekatan antar data dihitung menggunakan *Euclidean Distance*. Jarak antar *instance* dihitung menggunakan rumus *Euclidean Distance* seperti rumus 1.

$$d(p,q) = \sqrt{\sum_{i=1}^{n}(p_i - q_i)^2} \qquad (1)$$

di mana $p = (p_i, p_2, \dots, p_n)$ adalah vektor fitur dari data pelatihan, $q = (q_i, q_2, \dots, q_n)$ adalah vektor fitur dari data uji, n = jumlah fitur (dimensi data) dan $d(p,q)$ = jarak *Euclidean* antara data pelatihan dan data uji [15].

2.5. Evaluasi Model

Evaluasi model dilakukan menggunakan teknik *10-fold cross-validation*, di mana *dataset* dibagi menjadi sepuluh *subse*t dengan satu *subset* secara bergantian digunakan sebagai data uji dan sisanya sebagai data latih. Pendekatan

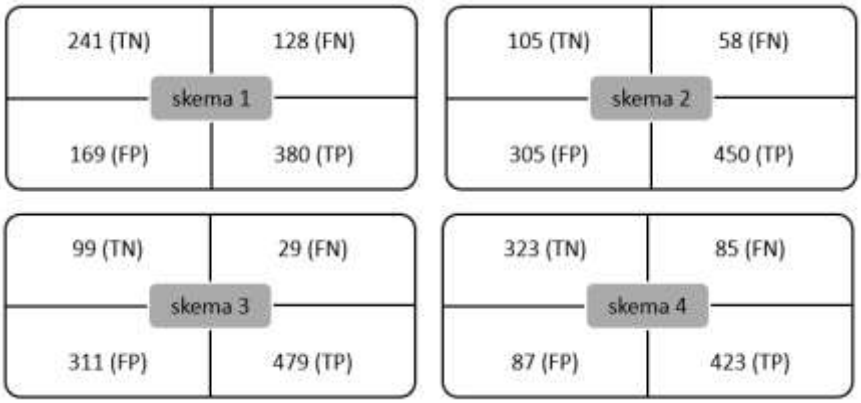
DOI: https://doi.org/10.31004/riggs.v4i3.2514

ini dipilih untuk menghasilkan estimasi performa yang lebih stabil sekaligus meminimalkan risiko bias maupun *overfitting* [16]. Kinerja model diukur menggunakan beberapa metrik utama, yaitu akurasi, presisi, *recall*, F1-score, dan *area under the curve* (AUC), sehingga penilaian tidak hanya berfokus pada ketepatan klasifikasi secara umum, tetapi juga pada keseimbangan kemampuan model dalam mendeteksi kasus positif maupun negatif.

3. Hasil dan Diskusi

3.1. Hasil Eksperimen

Hasil eksperimen menunjukkan bahwa empat skema *preprocessing* memberikan keluaran berbeda sebagaimana dirangkum pada Tabel 2. Evaluasi menggunakan *10-fold cross-validation* menghasilkan *confusion matrix* yang ditampilkan pada Gambar 2 menampilkan *confusion matrix* dari keempat skema *preprocessing*. Tampak bahwa Skema 3 memiliki jumlah *True Positive* (TP) tertinggi, sementara Skema 4 menunjukkan keseimbangan lebih baik antara *True Negative* (TN) dan *True Positive* dibanding skema lainnya.



Gambar 2. Hasil *Confusion Matrix*

Sementara performa model dihitung dengan metrik akurasi, presisi, *recall*, F1-score dan AUC. Ringkasan nilai dari kelima metrik tersebut ditunjukkan pada Tabel 3, yang memperlihatkan variasi kinerja antar skema serta mengungkap adanya perbedaan kontribusi *preprocessing* terhadap efektivitas model.

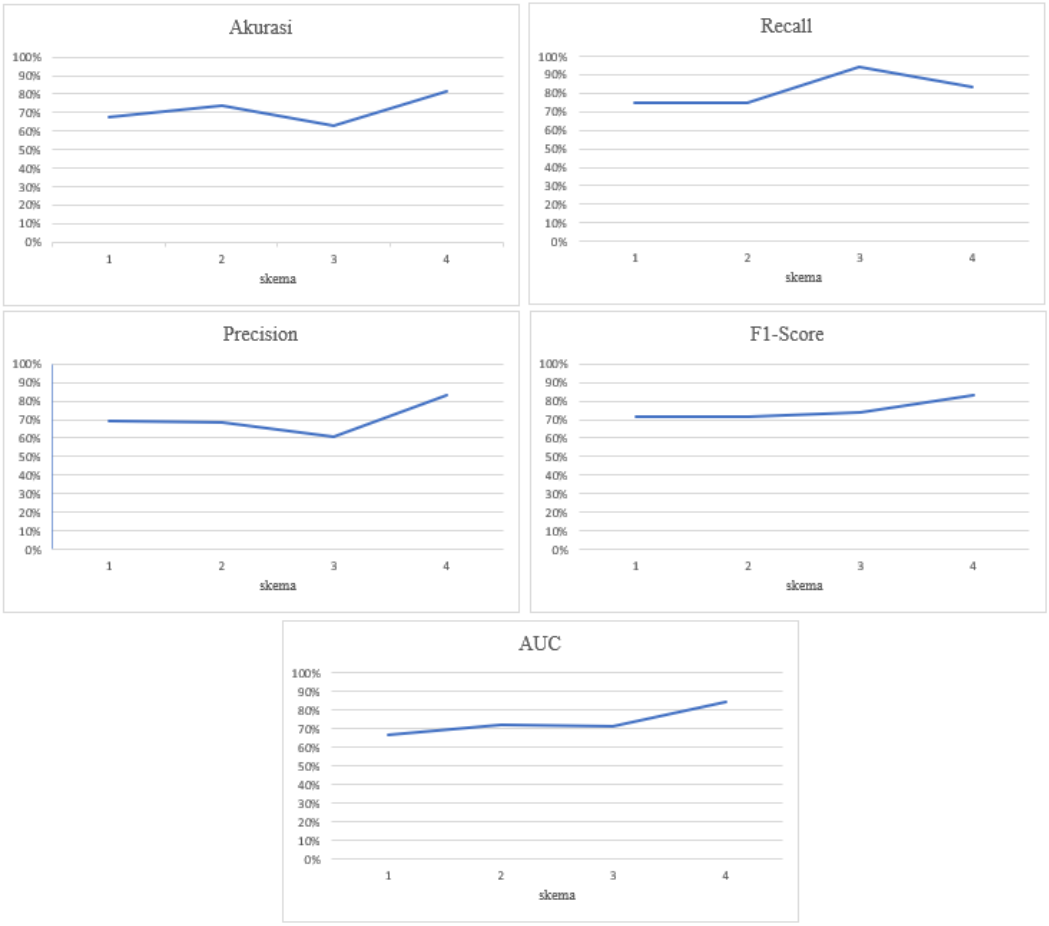
Tabel 3. Hasil Evaluasi

Skema	Akurasi	Presisi	*Recall*	*F1-score*	AUC
1	67.64%	69.22%	74.80%	71.91%	0.6679
2	73.89%	68.65%	75.00%	71.69%	0.7170
3	62.96%	61.06%	94.27%	73.91%	0.7114
4	81.26%	83.36%	83.25%	83.00%	0.8460

Hasil pada Tabel 3 menunjukkan adanya variasi kinerja antar skema. Skema pertama sebagai *baseline* hanya menghasilkan akurasi 67,64 persen dengan AUC 0,6679. Penambahan normalisasi pada skema kedua meningkatkan akurasi hingga 73,89 persen, meskipun presisi belum optimal. Skema ketiga memperlihatkan *recall* tertinggi yaitu 94,27 persen, tetapi akurasi turun menjadi 62,96 persen akibat dominasi prediksi positif. Peningkatan menyeluruh diperoleh pada skema keempat, yang mencapai akurasi 81,26 persen dan AUC 0,8460, sekaligus memberikan keseimbangan terbaik antara presisi dan *recall.*

DOI: https://doi.org/10.31004/riggs.v4i3.2514

3.2. Analisis Perbandingan Skema



Gambar 3. Perbandingan Tiap Matriks

Perbandingan antar skema memperlihatkan bagaimana setiap teknik *preprocessing* berkontribusi pada kinerja model. Penambahan normalisasi pada skema kedua meningkatkan akurasi dari 67,64 persen menjadi 73,89 persen, meskipun presisi belum meningkat signifikan. Hal ini membuktikan bahwa perbedaan skala fitur berpengaruh langsung pada algoritma KNN, sehingga normalisasi menjadi tahapan penting.

Skema ketiga menunjukkan fenomena menarik, di mana *recall* melonjak hingga 94,27 persen tetapi akurasi turun ke 62,96 persen. Kondisi ini terjadi karena model lebih sering mengklasifikasikan data ke kelas positif, yang menghasilkan sensitivitas tinggi namun menurunkan keseimbangan prediksi. Temuan ini menegaskan bahwa seleksi fitur tanpa *balancing* data justru dapat menciptakan bias terhadap salah satu kelas.

Skema keempat memberikan hasil paling seimbang dengan presisi 83,36 persen, *recall* 83,25 persen dan *F1-score* 83,00 persen. Keberhasilan ini berasal dari kombinasi SMOTE, normalisasi dan seleksi fitur. SMOTE memperbaiki distribusi kelas, normalisasi menyeragamkan skala dan seleksi fitur memastikan hanya atribut relevan yang dipertahankan.

3.3 Diskusi

Perbandingan Hasil penelitian ini menegaskan bahwa *preprocessing* merupakan komponen strategis dalam kerangka CRISP-DM. Setiap teknik memberikan kontribusi berbeda terhadap kinerja model dan kombinasi yang tepat dapat meningkatkan hasil secara signifikan. Dalam konteks klasifikasi penyakit jantung, *preprocessing* yang efektif meningkatkan reliabilitas sistem, sehingga model lebih layak digunakan sebagai alat bantu deteksi dini.

Implikasi dari penelitian ini adalah bahwa keberhasilan sistem prediksi medis tidak hanya ditentukan oleh algoritma, melainkan juga oleh kualitas *preprocessing*. Dengan demikian, penelitian lebih lanjut sebaiknya mengeksplorasi strategi *preprocessing* lanjutan, seperti teknik *balancing* berbasis *generative* model atau

DOI: https://doi.org/10.31004/riggs.v4i3.2514

normalisasi adaptif. Temuan ini juga menunjukkan bahwa ketika *preprocessing* dirancang dengan matang, algoritma sederhana seperti KNN pun mampu bersaing, meskipun algoritma lebih kompleks dapat dipertimbangkan untuk aplikasi nyata guna menghasilkan performa terbaik.

4. Kesimpulan

Penelitian ini membuktikan bahwa desain *preprocessing* yang tepat dalam kerangka CRISP-DM berperan penting dalam meningkatkan kinerja klasifikasi penyakit jantung menggunakan algoritma KNN. Skema *baseline* hanya menghasilkan akurasi 67,64 persen, sedangkan kombinasi normalisasi, SMOTE dan seleksi fitur pada skema keempat mampu meningkatkan akurasi hingga 81,26 persen dengan AUC 0,8460. Fakta ini menegaskan bahwa *preprocessing* bukan sekadar tahap teknis tambahan, melainkan komponen strategis yang menentukan reliabilitas model prediksi medis. Implikasinya, sistem deteksi dini penyakit jantung dapat menjadi lebih andal jika *preprocessing* dirancang secara sistematis, sementara penelitian lanjutan dapat diarahkan pada eksplorasi teknik *preprocessing* yang lebih adaptif serta integrasinya dengan algoritma yang lebih kompleks untuk mencapai performa optimal dalam aplikasi nyata.

Referensi

[1] A. Wahyu *et al.*, "PEDULI KESEHATAN JANTUNG UPAYA MEMBANGUN MASYARAKAT SADAR KESEHATAN JANTUNG DI DESA NGAWI JAWA TIMUR," vol. 6, no. 3, 2022, doi: 10.31764/jpmb.v6i3.9823.

[2] "Peningkatan Pengetahuan dan Keterampilan Kader dalam Penanganan Korban Gawat Darurat Henti Jantung Prehospital," *Cardiology (Switzerland)*, vol. 7, no. 2. pp. 121–127, 2025. doi: 10.1159/000165558.

[3] I. Y. Pramunsyi, J. Dharmawan, and R. Widiastutik, "Analis Diagnosa Anak Berkebutuhan Khusus Menggunakan Metode Decision Tree (Studi Kasus di Sekolah Luar Biasa Sumenep)," vol. 3, no. 2, pp. 299–306, 2024, doi: 10.24929/jars.v3i1.3785.

[4] D. Kurniawan and M. Yasir, "Optimization Sentimen Analysis using CRISP-DM and Naive Bayes Methods Implemented on Social Media," *Cybersp. J. Pendidik. Teknol. Inf.*, vol. 6, no. 2, p. 74, 2022, doi: 10.22373/cj.v6i2.12793.

[5] S. Lestari, M. Mupaat, and A. Erfina, "Analisis Sentimen Masyarakat Indonesia terhadap Pemindahan Ibu Kota Negara Indonesia pada Twitter," *JUSIFO (Jurnal Sist. Informasi)*, vol. 8, no. 1, pp. 13–22, 2022, doi: 10.19109/jusifo.v8i1.12116.

[6] M. R. Baihaqi, T. N. Padilah, M. Jajuli, J. M. H. Y. Al-Afghoni, Wahyudi Setiawan, and Y. Dwi Putra Negara, "Klasifikasi Jenis Benih Kacang Menggunakan Smote Dan Decision Tree C4.5," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 7, no. 1, pp. 661–671, 2024, doi: 10.36040/jati.v9i1.12366.

[7] A. M. A. Rahim, Inggrid Yanuar Risca Pratiwi, and Muhammad Ainul Fikri, "Klasifikasi Penyakit Jantung Menggunakan Metode Synthetic Minority Over-Sampling Technique Dan Random Forest Clasifier," *Indones. J. Comput. Sci.*, vol. 12, no. 5, pp. 2995–3011, 2023, doi: 10.33022/ijcs.v12i5.3413.

[8] M. R. Baihaqi, T. N. Padilah, and M. Jajuli, "Implementasi Metode Imputasi Mean dan Single Center Imputation Chained Equation (SICE) Terhadap Hasil Prediksi Linear Regression pada Data Numerik," *J. JTIK (Jurnal Teknol. Inf. dan Komunikasi)*, vol. 7, no. 4, pp. 661–671, 2023, doi: 10.35870/jtik.v7i4.1169.

[9] A. A. Baskara, N. M. Piranti, and M. F. Romdendine, "FRAMEWORK DATA MINING : SEBUAH SURVEI," vol. 9, no. 3, pp. 4886–4895, 2025, doi: 10.36040/jati.v9i3.13803.

[10] F. Yulian Pamuji, A. Rofiqul Muslikh, R. Muhammad Arief, and D. Muti, "JIP (Jurnal Informatika Polinema) KOMPARASI METODE MEAN DAN KNN IMPUTATION DALAM MENGATASI MISSING VALUE PADA DATASET KECIL", doi: 10.33795/jip.v10i2.5031.

[11] N. Suhandi, R. Gustriansyah, A. Destria, M. Amalia, and V. Kris, "Prediksi Kualitas Susu Menggunakan Metode K-Nearest Neighbors Milk Quality Prediction Using The K-Nearest Neighbors Method," vol. 14, no. 2, 2024, doi: 10.30700/sisfotenika.v14i2.430.

[12] M. A. M. Setiawan, K. Kusrini, and A. D. Hartono, "Menggunakan Metode Machine Learning Untuk Memprediksi Nilai Mahasiswa Dengan Model Prediksi Multiclass," *J. Inform. J. Pengemb. IT*, vol. 10, no. 1, pp. 190–204, 2025, doi: 10.30591/jpit.v10i1.8334.

[13] S. Zulaikhah Hariyanti Rukmana, A. Aziz, and W. Harianto, "Optimasi Algoritma K-Nearest Neighbor (Knn) Dengan Normalisasi Dan Seleksi Fitur Untuk Klasifikasi Penyakit Liver," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 6, no. 2, pp. 439–445, 2022, doi: 10.36040/jati.v6i2.4722.

[14] J. T. Kumalasari, A. Merdekawati, and A. Hidayati, "Klasifikasi Multi Class Pada Metode Kerja Jarak Jauh Menggunakan Algoritma Decision Tree dan Imbalance Data," *J. Inf. Syst. Applied, Manag. Account. Res.*, vol. 8, no. 1, p. 109, Jan. 2024, doi: 10.52362/jisamar.v8i1.1350.

[15] E. Sahelvi, P. Cikita, and R. M. Sapitri, "Comparison of K-Nearest Neighbors and Random Forest Algorithms for Recommendations for a Healthy Lifestyle in Prevent Heart Disease Perbandingan Algoritma K-Nearest Neighbors dan Random Forest untuk Rekomendasi Gaya Hidup Sehat dalam Mencegah Penyakit Jan," vol. 5, no. July, pp. 830–840, 2025, doi: 10.57152/malcom.v5i3.1972.

[16] W. A. Firmansyach, U. Hayati, and Y. Arie Wijaya, "Analisa Terjadinya Overfitting Dan Underfitting Pada Algoritma Naive Bayes Dan Decision Tree Dengan Teknik Cross Validation," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 7, no. 1, pp. 262–269, 2023, doi: 10.36040/jati.v7i1.6329.

DOI: https://doi.org/10.31004/riggs.v4i3.2514