

Optimasi Performa Prediksi Penyakit Jantung Menggunakan Teknik Stacking Classifier

Eka Amelya, Erliyan Redy Susanto*

Fakultas Teknik dan Ilmu Komputer, Sistem Informasi, Universitas Teknokrat Indonesia, Bandar Lampung, Indonesia

Email: ¹eka_amelya@teknokrat.ac.id, ^{2,*}erliyan_redy@teknokrat.ac.id

Email Penulis Korespondensi: erliyan_redy@teknokrat.ac.id

Submitted: 24/01/2025; Accepted: 26/02/2025; Published: 01/03/2025

Abstrak—Penyakit kardiovaskular, termasuk penyakit jantung, merupakan salah satu penyebab kematian tertinggi di Indonesia. Penyakit jantung adalah kondisi yang mengganggu fungsi jantung dan pembuluh darah, yang dapat disebabkan oleh penyumbatan atau penyempitan pembuluh darah arteri. Arteri sendiri berperan dalam mengalirkan darah yang kaya oksigen dari jantung ke seluruh tubuh, termasuk otot jantung melalui arteri koroner. Hal tersebut dapat terjadi karena adanya berbagai faktor seperti sumbatan pembuluh darah, peradangan, infeksi, atau kelainan bawaan. Kondisi ini dapat menghambat kemampuan jantung dalam memompa darah secara optimal, sehingga berpotensi menjadi ancaman serius bagi kesehatan seseorang. Penelitian ini bertujuan untuk meningkatkan akurasi prediksi penyakit jantung dengan menerapkan teknik stacking classifier, sebuah metode ensemble learning yang menggabungkan beberapa algoritma pembelajaran mesin, yaitu *Support Vector Machine (SVM)*, *Logistic Regression*, dan *Decision Tree*. Dataset yang digunakan telah melalui proses standarisasi dan divalidasi menggunakan metode stratified k-fold cross-validation guna memastikan kestabilan hasil prediksi. Kontribusi utama penelitian ini terletak pada peningkatan akurasi dan efisiensi diagnosis penyakit jantung melalui penerapan teknik stacking classifier yang mampu mengatasi dataset kompleks dan tidak seimbang. Dalam penelitian sebelumnya, teknik SMOTEEN digunakan dalam prediksi penyakit jantung, namun hasil penelitian ini menunjukkan bahwa pendekatan stacking classifier memiliki performa yang lebih baik. Hasil evaluasi menunjukkan bahwa metode ini mencapai akurasi sebesar 88,52%, precision 87,88%, recall 90,62%, dan ROC-AUC 94,18%, yang membuktikan efektivitasnya dalam meningkatkan keandalan diagnosis medis serta mengurangi kesalahan prediksi yang berisiko dalam dunia kesehatan.

Kata Kunci: Analitik Prediktif; Ensemble Learning; Diagnosis Penyakit Jantung; Stacking Classifier

Abstract—Cardiovascular diseases, including heart disease, are among the leading causes of death in Indonesia. Heart disease is a condition that disrupts the function of the heart and blood vessels, often caused by blockages or narrowing of the arteries. Arteries play a crucial role in delivering oxygen-rich blood from the heart to the entire body, including the heart muscles through the coronary arteries. This condition can result from various factors such as vascular blockages, inflammation, infections, or congenital abnormalities. Such issues can impair the heart's ability to pump blood efficiently, posing a serious threat to an individual's health. This study aims to improve the accuracy of heart disease prediction by implementing the stacking classifier technique—an ensemble learning method that combines multiple machine learning algorithms, namely Support Vector Machine (SVM), Logistic Regression, and Decision Tree. The dataset used has undergone a standardization process and has been validated using the stratified k-fold cross-validation method to ensure stable predictive results. The primary contribution of this research lies in enhancing the accuracy and efficiency of heart disease diagnosis through the application of the stacking classifier, which effectively handles complex and imbalanced datasets. Previous studies have utilized the SMOTEEN technique for heart disease prediction. However, the findings of this study demonstrate that the stacking classifier approach performs better. Evaluation results show that this method achieves an accuracy of 88.52%, precision of 87.88%, recall of 90.62%, and an ROC-AUC of 94.18%, proving its effectiveness in improving medical diagnosis reliability and reducing prediction errors that could pose risks in the healthcare field.

Keywords: Predictive Analytics; Ensemble Learning; Heart Disease Diagnosis; Stacking Classifier

1. PENDAHULUAN

Sistem kardiovaskular adalah salah satu sistem terpenting dalam tubuh manusia, berfungsi untuk menjaga sirkulasi darah dan memenuhi kebutuhan oksigen serta nutrisi seluruh organ. Namun, penyakit kardiovaskular tetap menjadi penyebab utama kematian global. Menurut laporan *American Heart Association (AHA)* pada tahun 2017, terdapat 17,8 juta kematian akibat penyakit kardiovaskular, meningkat 21,1% dibandingkan tahun 2007 [1]. Data dari WHO pada tahun 2021 menunjukkan bahwa penyakit jantung menyumbang 17,8 juta kematian per tahun, atau sekitar satu dari tiga kematian secara global. Di Indonesia, prevalensi penyakit jantung meningkat dari 0,5% pada tahun 2013 menjadi 1,5% pada tahun 2018, dengan tren peningkatan serangan jantung pada individu di bawah 40 tahun yang bertambah sekitar 2% setiap tahunnya [2][3].

Penyakit jantung ditandai dengan gangguan fungsi jantung dalam menjalankan siklus aliran darah, yang dapat disebabkan oleh penyempitan atau penyumbatan pembuluh darah, gangguan katup jantung, dan aritmia. Penumpukan kolesterol di pembuluh darah menjadi salah satu penyebab utama, yang mengakibatkan kenaikan tekanan darah dan menurunnya efisiensi fungsi jantung [2]. Oleh karena itu, deteksi dini dan prediksi penyakit jantung sangat penting untuk mencegah komplikasi lebih lanjut dan mengurangi angka kematian.

Dalam era kemajuan teknologi informasi, pengolahan data dan analisis prediktif menjadi komponen utama dalam mendukung keputusan di berbagai bidang, termasuk kesehatan. Salah satu pendekatan yang menjanjikan untuk meningkatkan akurasi prediksi adalah metode *ensemble learning*, yang menggabungkan beberapa model pembelajaran mesin untuk meningkatkan performa keseluruhan. Teknik *stacking classifier*, yang memadukan prediksi

dari beberapa model dasar menggunakan model meta, telah terbukti efektif dalam menangkap pola data yang kompleks [4][5].

Beberapa penelitian sebelumnya menunjukkan keberhasilan metode stacking dalam berbagai aplikasi. Misalnya, penelitian oleh Diky Djafar Sidik (2024) [6]. Menggunakan stacking untuk prediksi curah hujan dengan akurasi 78,25%. Penelitian Budi Sunarko dan Uswatun Hasanah (2023) menunjukkan bahwa stacking ensemble learning dapat mencapai akurasi hingga 99,9% dalam klasifikasi dampak polutan terhadap kesehatan [7]. Penelitian ini berfokus pada optimasi performa prediksi penyakit jantung menggunakan teknik *stacking classifier*, yang tidak hanya mengutamakan akurasi tetapi juga efisiensi waktu pelatihan, kemampuan generalisasi, dan relevansi klinis.

Tujuan utama dari penelitian ini adalah mengoptimasi performa prediksi penyakit jantung dengan memanfaatkan model machine learning, khususnya melalui penerapan teknik *stacking classifier*. Pendekatan ini dirancang untuk meningkatkan akurasi, efisiensi waktu pelatihan, dan kemampuan generalisasi model dalam mengidentifikasi faktor risiko penyakit jantung. Dengan menggabungkan beberapa algoritma klasifikasi seperti *Logistic Regression* [22], *Decision Tree* [23], *K-Nearest Neighbor* [24], *Support Vector Machine* [25], dan *Multi-Layer Perceptron* [26], penelitian ini bertujuan untuk menghasilkan model prediktif yang lebih unggul dibandingkan dengan model tunggal. Selain itu, penelitian ini juga berfokus pada relevansi klinis dengan menyediakan alat bantu prediksi yang dapat diimplementasikan secara praktis oleh tenaga medis untuk diagnosis dini dan penanganan penyakit jantung [8]. Penyakit kardiovaskular, termasuk penyakit jantung, merupakan penyebab utama kematian global, dengan angka kematian yang terus meningkat setiap tahunnya. Di Indonesia, prevalensi penyakit jantung juga menunjukkan tren peningkatan, terutama pada kelompok usia di bawah 40 tahun. Deteksi dini dan prediksi yang akurat sangat penting untuk mencegah komplikasi dan mengurangi angka kematian. Namun, metode prediksi konvensional seringkali memiliki keterbatasan dalam hal akurasi dan kemampuan menangkap pola data yang kompleks. Oleh karena itu, penelitian ini hadir untuk menjawab tantangan tersebut dengan mengoptimasi model prediksi menggunakan teknik *stacking classifier*, yang diharapkan dapat memberikan hasil yang lebih akurat dan relevan secara klinis [9].

Kebaharuan penelitian ini terletak pada penerapan teknik *stacking classifier* yang menggabungkan beberapa model machine learning untuk prediksi penyakit jantung. Berbeda dengan penelitian sebelumnya yang umumnya menggunakan model tunggal, pendekatan ini memanfaatkan kekuatan dari berbagai algoritma klasifikasi untuk meningkatkan performa prediksi. Selain itu, penelitian ini tidak hanya fokus pada akurasi, tetapi juga mengoptimalkan efisiensi waktu pelatihan dan kemampuan generalisasi model. Validasi silang yang mendalam dan pendekatan yang efisien dalam pelatihan model juga menjadi aspek kebaruan yang ditawarkan, sehingga hasil penelitian ini dapat diimplementasikan secara praktis dalam konteks klinis.

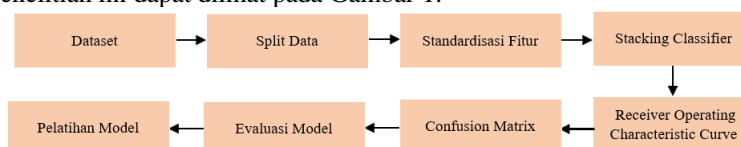
Penelitian ini memiliki beberapa keunggulan dibandingkan studi sebelumnya dalam memprediksi penyakit jantung. Salah satu aspek pembeda utamanya adalah penggunaan teknik *stacking classifier*, yang mengintegrasikan berbagai model machine learning, berbeda dari penelitian terdahulu yang umumnya hanya menerapkan model tunggal. Keunggulan lainnya mencakup peningkatan akurasi, efisiensi waktu pelatihan, dan kemampuan generalisasi, yang memungkinkan model mengenali pola data yang lebih kompleks serta meningkatkan keakuratan prediksi dengan hasil evaluasi menunjukkan bahwa metode ini mencapai akurasi sebesar 88,52%, precision 87,88%, recall 90,62%, dan ROC-AUC 94,18%. Selain itu, penelitian ini menitikberatkan pada relevansi klinis, dengan tujuan mengembangkan alat bantu diagnosis yang dapat diterapkan dalam dunia medis. Berbeda dari metode konvensional, penelitian ini menerapkan validasi silang yang lebih mendalam serta mengombinasikan algoritma seperti *logistic regression*, *decision tree*, *K-Nearest neighbor*, *support vector machine*, dan *multi-layer perceptron*, sehingga menghasilkan prediksi yang lebih akurat dan dapat dijadikan referensi dalam pengembangan sistem prediksi kesehatan berbasis data.

Dengan menggunakan teknik *stacking classifier*, penelitian ini berhasil menghasilkan model prediktif dengan akurasi yang lebih tinggi dibandingkan model tunggal, sehingga dapat membantu tenaga medis dalam mendiagnosis penyakit jantung secara lebih tepat. Hasil prediksi yang dihasilkan dapat digunakan sebagai alat bantu diagnosis oleh tenaga medis, memungkinkan deteksi dini dan penanganan yang lebih cepat terhadap pasien berisiko tinggi. Penerapan teknik *stacking classifier* dalam prediksi penyakit jantung merupakan inovasi metodologi yang dapat menjadi referensi bagi penelitian selanjutnya di bidang kesehatan berbasis data.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Penelitian ini menggunakan tiga algoritma dalam prediksi penyakit jantung dengan Teknik *stacking classifier*, yaitu SVM, *Logistic Regression*, dan *Decision Tree*, untuk memprediksi penyakit jantung. Dataset yang diperoleh dibagi berdasarkan kriteria tertentu, kemudian diuji menggunakan ketiga metode tersebut untuk mengevaluasi performa masing-masing. Alur penelitian ini dapat dilihat pada Gambar 1.



Gambar 1. Alur Penelitian

2.2 Dataset

Dataset merupakan kumpulan data yang diatur dalam format yang terstruktur, seperti tabel atau file, dan berisi informasi dari berbagai sumber. Dataset dapat berupa data numerik, teks, gambar, atau gabungan dari semuanya. Keberagaman jenis data set ini memungkinkan aplikasi dalam berbagai industri dan disiplin ilmu. Data set dikumpulkan oleh profesional di bidang data, seperti data analyst. Untuk bisa digunakan, data set perlu melewati beberapa tahapan pengolahan data, seperti data cleaning dan kategorisasi. Sehingga, data set yang dapat digunakan oleh profesional biasanya terkumpul berdasarkan kategorinya masing-masing, dan di dalamnya terdapat variabel-variabel yang saling berhubungan [10]. Dataset prediksi penyakit jantung yang akan digunakan pada penelitian ini bersumber dari *Kaggle*.

2.3 Split Data

Pada penelitian ini data kemudian dibagi menjadi dua bagian utama menggunakan teknik split data, di mana 80% dialokasikan untuk pelatihan (*training*) dan 20% untuk pengujian (*testing*). Pembagian ini dilakukan secara seimbang dengan menerapkan metode *stratified k-fold cross-validation*.

2.4 Standardisasi Fitur

Pada penelitian ini, standarisasi fitur diterapkan untuk memastikan semua fitur dalam dataset memiliki skala yang konsisten, sehingga meningkatkan kinerja dan efisiensi model pembelajaran mesin yang digunakan. Fitur-fitur seperti usia, tekanan darah, dan kadar kolesterol memiliki rentang nilai yang bervariasi. Untuk mengatasi perbedaan ini, standarisasi dilakukan menggunakan metode *Z-score normalization*, yang mengubah setiap fitur menjadi distribusi dengan rata-rata nol dan standar deviasi satu. Langkah ini sangat penting, terutama untuk algoritma seperti SVM, *Logistic Regression*, dan *Decision Tree* yang sangat peka terhadap perbedaan skala fitur. Berikut ini adalah rumus *Z-score normalization*:

$$z = \frac{x - \mu}{\sigma} \quad (1)$$

Keterangan: X: Nilai asli fitur, μ : Rata-rata nilai fitur, σ : Standar deviasi fitur.

2.5 Stacking Classifier

Stacking classifier merupakan metode ensemble learning dalam machine learning yang menggabungkan prediksi dari sejumlah model dasar (*base classifiers*) menggunakan model tambahan yang disebut meta-model atau meta-learner. Teknik ini bertujuan untuk meningkatkan akurasi dan kemampuan generalisasi model dengan memanfaatkan keunggulan berbagai algoritma pembelajaran [11]. Penelitian ini menerapkan teknik *stacking classifier* yang bertujuan untuk meningkatkan keakuratan prediksi penyakit jantung dengan menggabungkan beberapa algoritma pada *machine learning*. Pada tahap penelitian ini, algoritma *machine learning* seperti SVM, dan *Decision Tree* digunakan sebagai base models. Prediksi yang dihasilkan oleh base models kemudian digabungkan oleh meta-classifier, yang dalam penelitian ini berupa *Logistic Regression*, untuk menghasilkan prediksi akhir. Kinerja model dievaluasi menggunakan data uji dengan metrik akurasi, *precision*, *recall*, *F1-score*, dan *receiver operating characteristic curve*, untuk membuktikan keunggulan teknik *stacking classifier* dibandingkan model individual dalam meningkatkan performa prediksi. Model ini dapat dimanfaatkan sebagai alat bantu diagnosis bagi tenaga medis dalam mendeteksi penyakit jantung dengan lebih cepat dan akurat. Penerapannya dapat dilakukan melalui integrasi dengan sistem rekam medis elektronik (EMR) di rumah sakit atau klinik, memungkinkan analisis otomatis terhadap data pasien untuk memprediksi risiko penyakit jantung. Selain itu, model ini berpotensi dikembangkan menjadi aplikasi web atau *mobile*, sehingga pasien dapat melakukan skrining dini berdasarkan riwayat kesehatan serta parameter medis seperti tekanan darah, kadar kolesterol, dan riwayat pada keluarga. Model ini juga bisa dikoneksikan dengan perangkat pemantau kesehatan *wearable*, seperti *smartwatch* atau alat pemantau jantung, untuk menganalisis data secara *real-time* serta memberikan peringatan dini jika terdeteksi risiko penyakit jantung. Dengan penerapan ini, kesadaran masyarakat terhadap kesehatan jantung dapat meningkat, memungkinkan intervensi medis lebih cepat, dan berkontribusi dalam menekan angka kematian akibat penyakit jantung.

2.6 Pelatihan Model

Pada tahap pelatihan model, teknik *stacking classifier* digunakan untuk meningkatkan akurasi prediksi dengan menggabungkan output dari beberapa algoritma pembelajaran mesin. Tiga algoritma yang digunakan sebagai base models adalah SVM, *Logistic Regression*, dan *Decision Tree*.

2.6.1 Support Vector Machine (SVM)

Algoritma SVM adalah algoritma yang bertujuan untuk menemukan hyperplane maksimal, hyperplane adalah suatu fungsi yang dapat memisahkan antara dua kelas. SVM adalah algoritma *machine learning* yang bekerja atas prinsip *Structural Risk Minimization* (SRM) dengan tujuan menemukan *hyperplane* terbaik yang memisahkan dua buah *Class* pada input *space* [27]. Pada prosesnya SVM akan memaksimalkan margin atau jarak antara pola pelatihan dan batas keputusan [12].

$$w \cdot x + b = 0 \tag{2}$$

Dari rumus tersebut, w merupakan vektor bobot (koefisien) yang menggambarkan arah dari *hyperplane*, w : vektor bobot (koefisien) yang menggambarkan arah dari *hyperplane*, x : vektor fitur (input data), b : bias, yang menentukan posisi *hyperplane* relatif terhadap titik asal.

2.6.2 Decision Tree

Decision tree merupakan algoritma *machine learning* yang populer karena kemampuannya dalam memodelkan dan memahami hubungan kompleks di antara variabel-variabel yang terlibat dalam suatu keputusan [13]. *Decision Tree* merupakan struktur pohon yang terdiri dari node-node yang merepresentasikan keputusan dan cabang-cabang yang merepresentasikan konsekuensi dari sebuah keputusan [14]. *Decision Tree* terbuat dari tiga simpul yaitu *leaf*, lalu terdiri juga dari simpul *root* yang merupakan titik awal dari suatu *decision tree*, dan yang terakhir adalah simpul perantara yang berhubungan dengan suatu pengujian [28].

$$Entropy(S) = -\sum_{i=1}^k p_i \cdot \log_2(p_i) \tag{3}$$

Dimana S merupakan dataset yang sedang dievaluasi, K merupakan jumlah kelas dalam dataset, p_i merupakan proporsi data dalam kelas ke- i (probabilitas kelas i), dan \log_2 merupakan logaritma basis 2.

2.6.3 Logistic Regression

Logistic regression merupakan metode yang menganalisis hubungan antara variabel independen dan variable dependen dalam data nominal atau ordinal [15]. Pada beberapa kasus, variabel tidak bebas bersifat kualitatif dan dideskripsikan dalam satu atau dua kategori [29].

$$p(y = 1|x) = \frac{1}{1+e^{-(w \cdot x+b)}} \tag{4}$$

Keterangan: $p(y = 1|x)$ Probabilitas bahwa $y =$, diberikan fitur x , w : Vektor bobot untuk fitur, x : Vektor fitur input, e : Bilangan Euler (≈ 2.718).

2.7 Confusion Matrix

Confusion Matrix adalah tabel yang digunakan untuk mengevaluasi kinerja model klasifikasi dengan membandingkan hasil prediksi model dengan nilai aktual (data sebenarnya). *Confusion matrix* memberikan informasi tentang prediksi benar dan salah yang dilakukan model untuk setiap kelas [16].

$$\begin{bmatrix} TP & FP \\ FN & TN \end{bmatrix} \tag{5}$$

Dimana TP (*True Positive*) adalah banyaknya dokumen positif yang diklasifikasikan oleh sistem sebagai kelas positif, sedangkan FP (*False Negative*) diklasifikasikan sebagai kelas negatif. FN (*False Negative*) adalah banyaknya dokumen negatif yang diklasifikasikan oleh sistem sebagai kelas positif, sedangkan TN (*True Negative*) diklasifikasikan sebagai kelas negatif oleh sistem [30].

Confusion matrix terdiri dari beberapa perhitungan.

a. Akurasi

Mengukur seberapa sering model membuat prediksi yang benar.

$$\frac{TP+TN}{TP+TN+FP+FN} \tag{6}$$

b. Presisi

Mengukur proporsi prediksi positif yang benar-benar positif.

$$\frac{TP}{TP+FP} \tag{7}$$

c. Recall

Mengukur kemampuan model untuk mendeteksi semua kasus positif yang sebenarnya.

$$\frac{TP}{TP+FN} \tag{8}$$

d. Specifycity

Mengukur kemampuan model untuk mengenali semua kasus negatif yang sebenarnya.

$$\frac{TN}{TN+FP} \tag{9}$$

e. F1-Score

Menggabungkan Precision dan Recall dalam satu metrik dengan mengambil rata-rata harmonisnya.

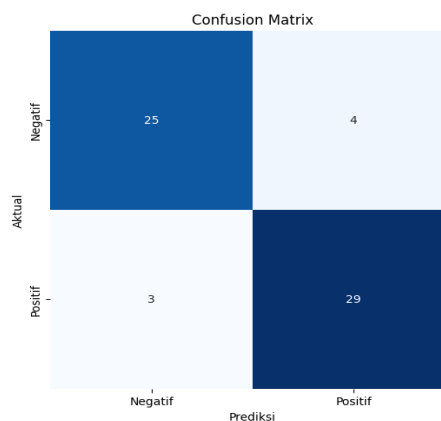
$$F1 = 2 \times \frac{\text{Presisi} \times \text{Recall}}{\text{Presisi} + \text{Recall}} \tag{10}$$

3. HASIL DAN PEMBAHASAN

Dalam penelitian menggunakan metode *Support Vector Machine (SVM)*, *Logistic Regression*, dan *Decision Tree*. Dengan teknik *STACKING CLASSIFIER*.

3.1 Confusion matrix dengan teknik *Stacking Classifier*

Berikut ini optimasi model *SVM*, *Logistic Regression*, dan *Decision Tree*. untuk prediksi penyakit jantung menggunakan teknik *STACKING CLASSIFIER* dengan *Confusion Matrix* algoritma *SVM*, *Logistic Regression*, dan *Decision Tree* dapat dilihat pada Gambar 2.



Gambar 2. *Confusion matrix* dengan teknik *Stacking Classifier*

Confusion matrix yang ditunjukkan pada Gambar 2 di atas menunjukkan hasil evaluasi model prediksi yang menggunakan teknik *stacking classifier*. Teknik ini bekerja dengan mengintegrasikan prediksi dari beberapa algoritma *machine learning* sebagai base models, seperti *Logistic Regression*, *Decision Tree*, dan *SVM*. Prediksi yang dihasilkan oleh base models kemudian digabungkan menggunakan *meta-classifier*, yang umumnya menggunakan algoritma seperti *Logistic Regression* untuk menentukan prediksi akhir.

Pada matriks tersebut, terdapat 25 prediksi benar untuk kelas negatif (*True Negative*), menunjukkan kemampuan model dalam mengenali data negatif secara akurat, dan 29 prediksi benar untuk kelas positif (*True Positive*), yang mencerminkan akurasi tinggi dalam mengidentifikasi data positif. Di sisi lain, terdapat 4 prediksi salah untuk kelas negatif (*False Positive*), di mana data yang sebenarnya negatif diprediksi sebagai positif, serta 3 prediksi salah untuk kelas positif (*False Negative*), di mana data yang sebenarnya positif diprediksi sebagai negatif.

Secara keseluruhan, tingginya jumlah prediksi benar dibandingkan prediksi salah menunjukkan bahwa *stacking classifier* berhasil meningkatkan kinerja model. Dengan memanfaatkan keunggulan dari berbagai algoritma pembelajaran mesin, teknik ini mampu menghasilkan model yang lebih tangguh dan akurat dibandingkan model individual, khususnya dalam menyelesaikan masalah klasifikasi yang kompleks, seperti prediksi penyakit jantung. Performa model ini biasanya juga diukur dengan metrik tambahan, seperti akurasi, *precision*, *recall*, *F1-score*, *ROC* dan *AUC*, untuk memberikan penilaian yang lebih komprehensif terhadap keefektifan pendekatan *stacking classifier*.

Tabel 1. hasil Teknik *Stacking Classifier*

Matrix	Nilai
Akurasi (Accuracy)	0.8852
Presisi (Precision)	0.8788
Recall	0.9062
Spesifisitas (Specificity)	0.8788
F1-Score	0.8923
ROC-AUC	0.9418

Pada Tabel 1, evaluasi kinerja model dengan teknik *stacking classifier* menunjukkan hasil yang sangat memuaskan berdasarkan berbagai metrik evaluasi. Model memiliki akurasi sebesar 88.52%, yang mengindikasikan bahwa sebagian besar prediksi yang dilakukan oleh model adalah benar. Akurasi ini mencerminkan kemampuan model untuk menangani data dengan berbagai karakteristik secara andal. *Precision* sebesar 87.88% mengungkapkan bahwa sebagian besar prediksi positif yang dihasilkan oleh model adalah benar. Hal ini menjadi sangat penting dalam konteks medis, seperti prediksi penyakit jantung, di mana mengurangi kesalahan prediksi positif (*False Positive*) sangat diperlukan guna menghindari diagnosa yang tidak diperlukan. Selain itu, *recall* sebesar 90.62% menunjukkan

kemampuan model dalam mendeteksi sebagian besar data positif secara akurat, sehingga memastikan tidak ada kasus penting yang terlewatkan, yang menjadi prioritas dalam pengelolaan risiko penyakit.

Specificity model mencapai 87.88%, yang menunjukkan efektivitas model dalam mengenali data negatif dan meminimalkan kesalahan dalam memprediksi individu sehat sebagai sakit (*False Positive*). *F1-score* sebesar 89.23% menunjukkan keseimbangan yang baik antara *precision* dan *recall*, memungkinkan model untuk menangani ketidakseimbangan antara kesalahan yang mungkin terjadi. Hal ini sangat penting, terutama ketika dataset memiliki distribusi kelas yang tidak merata. Nilai ROC-AUC yang mencapai 94.18% menunjukkan kemampuan luar biasa model dalam membedakan antara kelas positif (penderita penyakit) dan kelas negatif (individu sehat). ROC-AUC yang tinggi mengindikasikan bahwa model dapat diandalkan untuk membuat keputusan prediksi dengan tingkat keakuratan yang konsisten. Dapat dilihat pada Tabel 1.

Berdasarkan penelitian yang telah dilakukan sebelumnya dengan menggunakan 3 metode algoritma pembelajaran mesin dalam prediksi penyakit jantung yang terdiri dari SVM, *Logistic Regression* dan *Artificial Neural Network* (ANN) dengan pembagian data uji 80% : 20% [18]. Diperoleh hasil seperti pada Tabel 2.

Tabel 2. Hasil perbandingan kinerja algoritma

Confusion Matrix	Precision	Recall	F1 Score	Accuracy
SVM	0.70	0.89	0.78	0.72
Logistic Regression	0.86	0.86	0.86	0.83
ANN	0.86	0.81	0.84	0.82

Pada Tabel 2, menyajikan tabel perbandingan kinerja tiga algoritma klasifikasi, yaitu SVM, *Logistic Regression*, dan ANN, berdasarkan empat metrik evaluasi: *Precision*, *recall*, *f1-score*, dan *Accuracy*. Dari tabel tersebut, *logistic regression* menunjukkan performa terbaik dengan *precision*, *recall*, dan *f1-score* masing-masing sebesar 0.86, serta akurasi tertinggi sebesar 0.83, yang menunjukkan bahwa model ini mampu mengidentifikasi kelas positif dan negatif dengan seimbang.

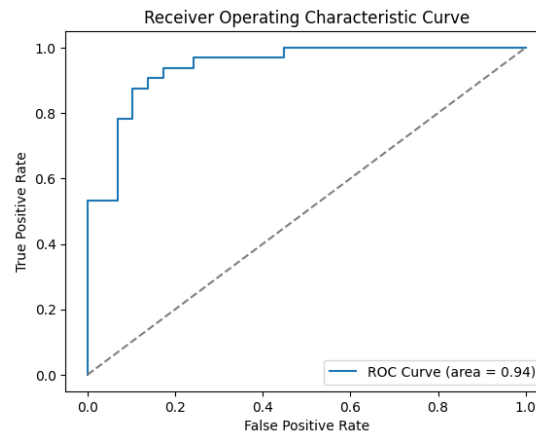
Sementara itu, ANN juga memiliki *precision* yang tinggi sebesar 0.86, namun *recall* yang lebih rendah di angka 0.81, menghasilkan *F1-score* sebesar 0.84 dan akurasi 0.82, yang sedikit di bawah *logistic regression*.

Di sisi lain, SVM memiliki *recall* tertinggi di antara ketiga model dengan nilai 0.89, tetapi *precision* yang lebih rendah sebesar 0.70, sehingga menghasilkan *F1-score* 0.78 dan akurasi terendah, yaitu 0.72. Hal ini menunjukkan bahwa meskipun SVM mampu mendeteksi kelas positif dengan baik, model ini cenderung menghasilkan lebih banyak kesalahan dalam mengklasifikasikan data positif yang sebenarnya negatif.

3.2 ROC Curve

Receiver Operating Characteristic (ROC) adalah sebuah grafik yang berfungsi untuk menilai performa model klasifikasi, terutama dalam membedakan dua kategori (klasifikasi biner). Grafik ini menampilkan *True Positive Rate* (TPR) atau sensitivitas pada sumbu y dan *False Positive Rate* (FPR) pada sumbu x dengan berbagai nilai ambang (*threshold*) [17]. ROC digunakan untuk mengukur kemampuan model dalam membedakan kelas positif dan negatif. Semakin dekat kurva ROC ke sudut kiri atas, semakin baik kinerja model. *Area Under the Curve* (AUC) dari ROC merepresentasikan performa total model, di mana nilai 1 menunjukkan model sempurna, sedangkan nilai 0,5 mencerminkan performa acak. Kurva *Receiver Operating Characteristic* (ROC) digunakan untuk mengevaluasi kinerja model klasifikasi, dalam konteks ini berdasarkan teknik *stacking classifier*. ROC menggambarkan hubungan antara *True Positive Rate* (TPR), yaitu persentase prediksi benar untuk kelas positif, dan *False Positive Rate* (FPR), yaitu persentase kesalahan prediksi untuk kelas negatif, pada berbagai ambang keputusan. Nilai area di bawah kurva (AUC) sebesar 0,94 menunjukkan bahwa model memiliki kinerja klasifikasi yang sangat baik, karena AUC mendekati nilai maksimum 1.0. Semakin besar nilai AUC, semakin baik kemampuan model membedakan antara kelas positif dan negatif, sementara nilai AUC 0,5 mengindikasikan model tidak lebih baik dari tebakan acak.

Stacking classifier adalah metode *ensemble learning* yang mengkombinasikan prediksi dari beberapa model dasar (*base classifiers*), seperti *decision tree*, SVM, atau *logistic regression*, menggunakan *meta-classifier* untuk menghasilkan prediksi akhir yang lebih akurat. Pendekatan ini memanfaatkan keunggulan masing-masing model dasar untuk menangani kerumitan data, sehingga sering kali memberikan hasil yang lebih unggul dibandingkan model tunggal atau metode *ensemble* lainnya seperti *bagging* atau *boosting*. Berdasarkan kurva ROC, terlihat bahwa model *stacking* ini mampu menekan tingkat kesalahan prediksi (*false positives*) serendah mungkin, sambil mempertahankan tingkat keberhasilan deteksi (*true positives*) yang tinggi. Hal ini menunjukkan bahwa teknik *stacking* berhasil meningkatkan performa model klasifikasi dengan memanfaatkan keunggulan dari berbagai model dasar yang digunakan. Dapat dilihat pada Gambar 3.



Gambar 3. ROC Curve

3.3 Pembahasan

Berdasarkan hasil penelitian yang telah dilakukan sebelumnya terhadap dataset UCI mengenai klasifikasi penyakit jantung menggunakan tiga metode yang berbeda, yaitu SVM, Logistic Regression dan ANN ditemukan hasil sebagai berikut, saat menggunakan pembagian data latih dan data uji dengan komposisi 90:10 ketika metode tersebut menghasilkan akurasi yang sama yaitu sebesar 80%, saat menggunakan komposisi 80:20 hasil terbaik diperoleh ketika menggunakan metode Logistic Regression sebesar 80%, kemudian ketika data dibagi dengan persentase 70:30 hasil performa terbaik didapatkan saat menggunakan metode ANN yaitu sebesar 82%, dan terakhir saat menggunakan perbandingan 60:40 performa terbaik diperoleh ketika menggunakan metode Logistic Regression dengan akurasi 83% [18]. Berdasarkan hasil penelitian yang telah dilakukan sebelumnya menunjukkan bahwa algoritma stacking mampu memberikan kinerja yang lebih baik dalam hal akurasi, TPR, TNR, *G-Mean* dan AUC dibandingkan dengan *classifier* tunggal lainnya. Peningkatan nilai-nilai tersebut diharapkan dapat menjadi referensi dalam pengembangan berbagai sistem yang mendukung dan meningkatkan keberhasilan proses deteksi dini penyakit jantung menggunakan teknik data mining [19]. Berdasarkan hasil penelitian yang telah dilakukan sebelumnya pengembangan sistem *diagnostic* berbasis Machine Learning (ML) digunakan untuk meramalkan penyakit jantung dengan dataset gangguan jantung. Proses prapemrosesan data yang diterapkan mencakup deteksi dan penghapusan outlier, penanganan data yang hilang, normalisasi fitur yaitu RF, MLP, KNN, ETC, XGB, SVC, ADB, DT dan GBM. Selain itu, kinerja model evaluasi menggunakan delapan metric, termasuk akurasi, klasifikasi, presisi, f1-score, spesifisitas, ROC, sensitivitas, log-loss serta koefisien korelasi Matthews, bersama dengan delapan metode evaluasi kinerja lainnya [20]. Berdasarkan penelitian yang telah dilakukan dengan mengevaluasi empat metode pemilihan sampel acak sederhana, acak berstrata proporsional, acak berstrata tidak proporsional dan pengambilan sampel deliberative serta tiga pendekatan validasi silang *k-fold*, *leave-on-out* dan metode *Monte Carlo*. Selain itu, penelitian ini juga menyelidiki dampak lokalisasi pemilihan sampel pada subset geografis kecil dari keseluruhan area, sebuah pendekatan yang sering digunakan untuk menekan biaya pengumpulan data pelatihan. Metode yang digunakan diuji dalam konteks klasifikasi menggunakan *Support Vector Machine* (SVM) dan analisis citra berbasis objek geografis (GEOBIA). Data yang digunakan terdiri dari *orthoimagery* beresolusi tinggi dari *National Agricultural Imagery Program* (NAIP) serta raster hasil LIDAR, yang mencakup area seluas 2.609 km² di wilayah timur laut *West Virginia, AS*. Hasil penelitian menunjukkan bahwa metode pengambilan sampel berbasis stratifikasi statistik menghasilkan akurasi klasifikasi tertinggi. Selain itu, penggunaan sejumlah kecil sampel pelatihan yang diambil hanya dari subset wilayah studi memberikan tingkat akurasi keseluruhan yang sebanding dengan sampel berukuran serupa yang dikumpulkan secara tersebar diseluruh dataset skala regional [21]. Sedangkan pada penelitian ini dengan menggunakan teknik stacking *classifier* menggunakan metode SVM, *Decision Tree*, dan *Logistic Regression* jauh lebih unggul.

Model dasar yang digunakan dalam penelitian ini mencakup SVM, *decision tree*, dan *logistic regression*, yang dipilih berdasarkan karakteristik serta keunggulan masing-masing algoritma. SVM mampu menangani data berdimensi tinggi dengan menemukan hyperplane optimal, sedangkan *decision tree* unggul dalam kemudahan interpretasi dan kemampuannya menangani data tidak terstruktur. *Logistic regression*, meskipun sederhana, memberikan hasil yang andal dalam klasifikasi biner serta menghasilkan probabilitas yang jelas untuk setiap kelas. Kombinasi ketiga model ini memungkinkan *stacking classifier* memanfaatkan kelebihan masing-masing, sehingga meningkatkan akurasi dan performa prediksi secara keseluruhan. Keunggulan utama dari *stacking classifier* dibandingkan model individu adalah kemampuannya mengurangi bias dan *varians*. Dengan mengombinasikan beberapa algoritma, teknik ini mengurangi risiko *overfitting* yang sering terjadi pada model tunggal dan meningkatkan kemampuan generalisasi terhadap data baru. Oleh karena itu, pemahaman lebih lanjut mengenai mekanisme *stacking classifier*, pemilihan base models, dan peran *meta-classifier* akan memperkaya pembahasan serta menjelaskan mengapa teknik ini unggul dalam klasifikasi penyakit jantung.

Berdasarkan penelitian yang telah dilakukan penggunaan metode stacking classifier, yang mengintegrasikan algoritma pembelajaran mesin seperti SVM, logistic regression, dan decision tree, mampu meningkatkan akurasi

prediksi penyakit jantung secara signifikan. Dengan memanfaatkan dataset yang telah melalui proses standarisasi serta teknik validasi yang cermat, model yang dihasilkan menunjukkan performa tinggi dengan akurasi 88,52%, precision 87,88%, recall 90,62%, dan ROC-AUC 94,18%.

4. KESIMPULAN

Penelitian ini mengungkapkan bahwa penggunaan metode stacking classifier, yang mengintegrasikan algoritma pembelajaran mesin seperti SVM, logistic regression, dan decision tree, mampu meningkatkan akurasi prediksi penyakit jantung secara signifikan. Dengan memanfaatkan dataset yang telah melalui proses standarisasi serta teknik validasi yang cermat, model yang dihasilkan menunjukkan performa tinggi dengan akurasi 88,52%, precision 87,88%, recall 90,62%, dan ROC-AUC 94,18%. Pendekatan ini tidak hanya efektif dalam mengatasi dataset yang kompleks dan tidak seimbang, tetapi juga mampu mengurangi kesalahan prediksi yang penting dalam dunia medis. Dengan kombinasi algoritma yang tepat, penelitian ini memberikan kontribusi besar dalam penerapan analitik prediktif di sektor kesehatan, terutama untuk meningkatkan efisiensi dan akurasi diagnosis. Ke depan, terdapat beberapa peluang riset yang dapat dikembangkan untuk memperluas temuan ini. Penelitian dapat diarahkan untuk menguji metode stacking classifier pada dataset yang lebih besar dan beragam, termasuk data dari populasi yang berbeda, untuk memvalidasi generalisasi model. Eksplorasi terhadap algoritma pembelajaran mesin lain, seperti deep learning atau ensemble methods (contohnya XGBoost, LightGBM), dapat dilakukan untuk melihat apakah performa prediksi dapat ditingkatkan lebih lanjut. Integrasi data multimodal, seperti data genetik, gaya hidup, dan riwayat medis lengkap, dapat menjadi area riset menarik untuk meningkatkan akurasi prediksi. Pengembangan sistem prediksi *real-time* yang dapat diintegrasikan dengan perangkat medis atau aplikasi kesehatan digital dapat menjadi terobosan dalam praktik klinis sehari-hari. Penelitian lebih lanjut juga dapat fokus pada interpretabilitas model untuk memastikan bahwa prediksi yang dihasilkan dapat dipahami dan dipercaya oleh tenaga medis, sehingga memudahkan penerapannya dalam pengambilan keputusan klinis. Dengan mengeksplorasi peluang-peluang ini, penelitian di bidang prediksi penyakit jantung dapat terus berkembang dan memberikan dampak yang lebih besar dalam dunia kesehatan.

REFERENCES

- [1] A. A. Maulani, S. Winarno, J. Zeniarja, R. T. E. Putri, and A. N. Cahyani, "Comparison of Hyperparameter Optimization Techniques in Hybrid CNN-LSTM Model for Heart Disease Classification," *Sinkron*, vol. 9, no. 1, pp. 455–465, 2024, doi: 10.33395/sinkron.v9i1.13219.
- [2] D. A. Ryfai, N. Hidayat, and E. Santoso, "Klasifikasi Tingkat Resiko Serangan Penyakit Jantung Menggunakan Metode K-Nearest Neighbor," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 6, no. 10, pp. 4701–4707, 2022, [Online]. Available: <https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/11662>
- [3] M. T. M. H. Medan, "Cardio Update 2024." Access Date 20 Dec 2024, [Online]. Available: <https://lms.kemkes.go.id/courses/35bff824-437e-4557-b37a-94b128c43333>
- [4] T. Pipit Mulyah, Dyah Aminatun, Sukma Septian Nasution, Tommy Hastomo, Setiana Sri Wahyuni Sitepu, "濟無No Title No Title No Title," *J. GEEJ*, vol. 7, no. 2, pp. 10–30, 2020.
- [5] A. Zubair and R. Umamit, "Penerapan Metode Holt-Winters Untuk Peramalan Penjualan pada Industri Makanan Ringan," *Techno.com J. Teknol. Inf.*, vol. 20, no. 4, pp. 499–507, 2021.
- [6] M. Murad, S. Sukmawaty, A. Ansar, R. Sabani, and S. Hidayat, "Sistem Pendeteksi Kerusakan Buah Mangga Menggunakan Sensor Gas Dengan Metode DCS - LCA," *JTIM J. Teknol. Inf. dan Multimed.*, vol. 3, no. 4, pp. 186–194, 2021, doi: 10.35746/jtim.v3i4.169.
- [7] B. M. Karomah, "Penerapan Metode Stacking Dalam Mengklasifikasikan Penderita Penyakit Diabetes," *J. Publ. Ilmu Komput. dan Multimed.*, vol. 1, no. 3, pp. 188–194, 1922, doi: 10.55606/jupikom.v1i3.522.
- [8] D. D. Sidik and R. W. Sen, "Penggunaan Stacking Classifier Untuk Prediksi Curah Hujan," *IT Soc.*, vol. 4, no. 1, pp. 21–27, 2019, doi: 10.33021/itfs.v4i1.1180.
- [9] B. Sunarko et al., "Penerapan Stacking Ensemble Learning untuk Klasifikasi Efek Kesehatan Akibat Pencemaran Udara," *Edu Komputika J.*, vol. 10, no. 1, pp. 55–63, 2023, doi: 10.15294/edukomputika.v10i1.72080.
- [10] S. A. Risyad, "Data Set: Pengertian, Jenis, dan Contohnya." Access Date 20 Dec 2024, [Online]. Available: <https://dibimbing.id/blog/detail/pengertian-data-sheet-jenis-dan-contoh>
- [11] A. K. Putri and Hari Suparwito, "Uji Algoritma Stacking Ensemble Classifier pada Kemampuan Adaptasi Mahasiswa Baru dalam Pembelajaran Online," *KONSTELASI Konvergensi Teknol. dan Sist. Inf.*, vol. 3, no. 1, pp. 1–12, 2023, doi: 10.24002/konstelasi.v3i1.7009.
- [12] N. Fitriyah, B. Warsito, and D. A. I. Maruddani, "Analisis Sentimen Gojek Pada Media Sosial Twitter Dengan Klasifikasi Support Vector Machine (Svm)," *J. Gaussian*, vol. 9, no. 3, pp. 376–390, 2020, doi: 10.14710/j.gauss.v9i3.28932.
- [13] R. N. Ramadhon, A. Ogi, A. P. Agung, R. Putra, S. S. Febrihartina, and U. Firdaus, "Implementasi Algoritma Decision Tree untuk Klasifikasi Pelanggan Aktif atau Tidak Aktif pada Data Bank," *Karimah Tauhid*, vol. 3, no. 2, pp. 1860–1874, 2024, doi: 10.30997/karimahtauhid.v3i2.11952.
- [14] D. Feby, "Apa Itu Decision Tree di Machine Learning Model?" Access Date 20 Dec 2024, [Online]. Available: <https://dqlab.id/apa-itu-decision-tree-di-machine-%0Alearning-model>
- [15] Findasari and A. I. A. Himayati, "Analisis Regresi Logistik Biner Pada Faktor Resiko Kejadian Tuberkulosis," *J. Mat. Sains dan Teknol.*, vol. 24, no. 1, pp. 01–14, 2023, doi: 10.33830/jmst.v24i1.4666.2023.
- [16] M. A. Tamaza and S. Defit, "Jurnal Computer Science and Information Technology (CoSciTech) Implementasi Naïve Bayes dalam M-Series 4 Mobile Legends untuk Prediksi Kemenangan Implementation of Naïve Bayes in M-Series 4 Mobile Legends for Winning Prediction," vol. 5, no. 1, pp. 205–214, 2024.



- [17] E. Richardson, R. Trevizani, J. A. Greenbaum, H. Carter, M. Nielsen, and B. Peters, “The receiver operating characteristic curve accurately assesses imbalanced datasets,” *Patterns*, vol. 5, no. 6, p. 100994, 2024, doi: 10.1016/j.patter.2024.100994.
- [18] F. Handayani, “Komparasi Support Vector Machine, Logistic Regression Dan Artificial Neural Network Dalam Prediksi Penyakit Jantung,” *J. Edukasi dan Penelit. Inform.*, vol. 7, no. 3, p. 329, 2021, doi: 10.26418/jp.v7i3.48053.
- [19] A. Nurmasani and Y. Pristyanto, “Algoritme Stacking Untuk Klasifikasi Penyakit Jantung Pada Dataset Imbalanced Class,” *Pseudocode*, vol. 8, no. 1, pp. 21–26, 2021, doi: 10.33369/pseudocode.8.1.21-26.
- [20] M. M. Islam, T. N. Tania, S. Akter, and K. H. Shakib, “An Improved Heart Disease Prediction Using Stacked Ensemble Method,” *Lect. Notes Inst. Comput. Sci. Soc. Telecommun. Eng. LNICST*, vol. 490 LNICST, pp. 84–97, 2023, doi: 10.1007/978-3-031-34619-4_8.
- [21] C. A. Ramezan, T. A. Warner, and A. E. Maxwell, “Evaluation of sampling and cross-validation tuning strategies for regional-scale machine learning classification,” *Remote Sens.*, vol. 11, no. 2, 2019, doi: 10.3390/rs11020185.
- [22] H. Jindal, S. Agrawal, R. Khera, R. Jain, and P. Nagrath, “Heart disease prediction using machine learning algorithms,” *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 1022, no. 1, 2021, doi: 10.1088/1757-899X/1022/1/012072.
- [23] K. M. Shiwangi, J. K. Sandhu, and R. Sahu, “Effective Heart-Disease Prediction by Using Hybrid Machine Learning Technique,” *Proc. Int. Conf. Circuit Power Comput. Technol. ICCPCT 2023*, pp. 1670–1675, 2023, doi: 10.1109/ICCPCT58313.2023.10245785.
- [24] 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,” *Sci. Rep.*, vol. 12, no. 1, pp. 1–11, 2022, doi: 10.1038/s41598-022-10358-x.
- [25] M. E. Farooqui and D. J. Ahmad, “Disease Prediction System Using Support Vector Machine and Multilinear Regression,” *Int. J. Innov. Res. Comput. Sci. Technol.*, vol. 8, no. 4, pp. 331–336, 2020, doi: 10.21276/ijircst.2020.8.4.15.
- [26] M. P. Romero *et al.*, “Decision tree machine learning applied to bovine tuberculosis risk factors to aid disease control decision making,” *Prev. Vet. Med.*, vol. 175, no. November, 2020, doi: 10.1016/j.prevetmed.2019.104860.
- [27] D. Irawan, E. B. Perkasa, Y. Yurindra, D. Wahyuningsih, and E. Helmud, “Perbandingan Klasifikasi SMS Berbasis Support Vector Machine, Naive Bayes Classifier, Random Forest dan Bagging Classifier,” *J. Sisfokom (Sistem Inf. dan Komputer)*, vol. 10, no. 3, pp. 432–437, 2021, doi: 10.32736/sisfokom.v10i3.1302.
- [28] R. Puspita and A. Widodo, “Perbandingan Metode KNN, Decision Tree, dan Naïve Bayes Terhadap Analisis Sentimen Pengguna Layanan BPJS,” *J. Inform. Univ. Pamulang*, vol. 5, no. 4, p. 646, 2021, doi: 10.32493/informatika.v5i4.7622.
- [29] F. D. Pramakrisna, F. D. Adhinata, and N. A. F. Tanjung, “Aplikasi Klasifikasi SMS Berbasis Web Menggunakan Algoritma Logistic Regression,” *Teknika*, vol. 11, no. 2, pp. 90–97, 2022, doi: 10.34148/teknika.v11i2.466.
- [30] G. F. Grandis, Y. Arumsari, and Indriati, “Seleksi Fitur Gain Ratio pada Analisis Sentimen Kebijakan Pemerintah Mengenai Pembelajaran Jarak Jauh dengan K-Nearest Neighbor,” *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 5, no. 8, pp. 3507–3514, 2021.