Jurnal Sains Informatika Terapan (JSIT)

E-ISSN: 2828-1659, Volume: 5, Issue: 1, Month: February, Year: 2026
Page: 190 - 194, Available online at: https://rcf-indonesia.org/home/

PENGEMBANGAN MODEL STACKING MACHINE LEARNING DENGAN OPTIMASI HYPERPARAMETER UNTUK DETEKSI SERANGAN MALWARE PADA SERVER

Elvira Sawitri [1)], Nadya Alinda Rahmi [2)], Ilmawati [3)], Pradani Ayu Widya Purnama [4)]

[1,2,3]Sistem Informasi, Fakultas Ilmu Komputer, Universitas Putra Indonesia YPTK Padang
[4]Teknik Informatika, Fakultas Ilmu Komputer, Universitas Putra Indonesia YPTK Padang
Corresponding Author: [1] elvirasawitri114@gmail.com

Article Info

Article history:

Received: Jan 07, 2026
Revised: Jan 20, 2026
Accepted: Jan 25, 2026
Published: Feb 01, 2026

Keywords:

Akurasi
GridSearchCV Malware
Naive
Bayes
Stacking

ABSTRACT

Penelitian ini dilatarbelakangi oleh meningkatnya ancaman serangan malware yang mengganggu sistem server di lingkungan institusi pendidikan, termasuk Universitas Muhammadiyah Muara Bungo (UMMUBA). Server kampus menjadi pusat layanan akademik, administrasi, serta penyimpanan data penting, sehingga rentan terhadap gangguan keamanan yang disebabkan oleh aktivitas malware. Penanganan serangan ini tidak bisa lagi dilakukan secara manual karena keterbatasan waktu, sumber daya manusia, dan kompleksitas serangan yang terus berkembang. Oleh karena itu, dibutuhkan suatu sistem deteksi malware yang dapat bekerja secara otomatis, akurat, real-time, dan mudah dioperasikan oleh pengguna non-teknis. Penelitian ini bertujuan untuk mengembangkan model stacking machine learning dengan algoritma Naive Bayes (Gaussian, Multinomial, Bernoulli, Complement) sebagai base learners dan Logistic Regression sebagai meta learner. Untuk mengatasi masalah ketidakseimbangan data antara kelas malware dan non-malware, digunakan teknik SMOTE (Synthetic Minority Over-sampling Technique). Selain itu, performa model ditingkatkan dengan optimasi hyperparameter menggunakan GridSearchCV, sehingga diperoleh konfigurasi terbaik. Penelitian dilakukan melalui beberapa tahap, mulai dari pengumpulan dan pelabelan data dari log server UMMUBA, preprocessing data, pelatihan model dasar, pengembangan arsitektur stacking, evaluasi menggunakan metrik akurasi, precision, recall, dan F1-score, hingga implementasi ke dalam antarmuka aplikasi web berbasis Streamlit. Aplikasi ini memungkinkan pengguna untuk mengunggah file log dan mendapatkan hasil klasifikasi serta visualisasi performa model secara langsung. Hasil penelitian diharapkan mampu memberikan solusi nyata untuk deteksi malware pada server institusi pendidikan.

1. INTRODUCTION

Perkembangan teknologi digital yang sangat pesat memberikan dampak besar terhadap sistem informasi, terutama pada lingkungan server yang menjadi tulang punggung layanan digital (1). Namun, pesatnya adopsi teknologi juga diiringi dengan meningkatnya ancaman keamanan siber, salah satunya adalah serangan malware yang dapat menyebabkan kebocoran data, kerusakan sistem, hingga terganggunya layanan public (2). Deteksi malware secara otomatis dan akurat menjadi sangat krusial sebagai upaya mitigasi serangan lebih lanjut.

Dalam konteks ini, pemanfaatan machine learning untuk mendeteksi pola serangan menjadi pendekatan yang menjanjikan. Beberapa penelitian terdahulu telah mengimplementasikan model Naive Bayes seperti Gaussian Naive Bayes (GNB), Multinomial Naive Bayes (MNB), Bernoulli Naive Bayes (BNB), dan Complement Naive Bayes (CNB) sebagai model dasar untuk klasifikasi malware (3–6). Namun, pendekatan ini masih memiliki keterbatasan dalam hal generalisasi dan ketahanan terhadap data yang tidak seimbang. Oleh karena itu, pengembangan model stacking machine learning yang menggabungkan kekuatan beberapa model dasar (base learners) dengan meta learner, serta dioptimalkan menggunakan teknik hyperparameter tuning seperti GridSearchCV, dapat meningkatkan akurasi dan efisiensi deteksi (7,8).

Sebagai bagian dari upaya penerapan hasil penelitian ke dalam sistem nyata, model yang dikembangkan juga diuji melalui antarmuka pengguna grafis (GUI) berbasis web menggunakan Streamlit. Aplikasi ini dirancang agar dapat

digunakan oleh pengguna non-teknis seperti administrator server untuk melakukan deteksi malware secara otomatis, cepat, dan interaktif. Pengujian model melalui GUI tidak hanya membantu validasi fungsionalitas, tetapi juga memastikan bahwa sistem mampu beroperasi dalam skenario penggunaan nyata dan memberikan pengalaman pengguna yang mudah diakses (9).

Selain memiliki nilai teknis, penelitian ini juga memiliki kontribusi terhadap pembangunan nasional dan global. Dalam Asta Cita Pembangunan Nasional, khususnya cita ke-6 yaitu "Meningkatkan produktivitas rakyat dan daya saing di pasar internasional", keamanan digital menjadi syarat mutlak dalam menciptakan iklim teknologi yang produktif dan kompetitif. Dalam dokumen Rencana Riset Indonesia Nasional (RRIN) Tahun 2017–2045, penelitian ini selaras dengan klaster riset bidang Teknologi Informasi dan Komunikasi (TIK), khususnya dalam subbidang cyber security dan pengembangan sistem cerdas untuk perlindungan data.

Lebih jauh lagi, penelitian ini turut mendukung pencapaian Sustainable Development Goals (SDGs), khususnya tujuan ke-9 (Industry, Innovation, and Infrastructure) dan tujuan ke-16 (Peace, Justice and Strong Institutions), melalui penguatan infrastruktur digital yang aman dan inovatif. Dengan pendekatan yang holistik dan terstruktur, penelitian ini diharapkan memberikan kontribusi nyata dalam pengembangan solusi keamanan server berbasis kecerdasan buatan yang handal, adaptif, dan mudah digunakan melalui aplikasi GUI berbasis web yang dapat langsung dimanfaatkan oleh pemangku kepentingan.

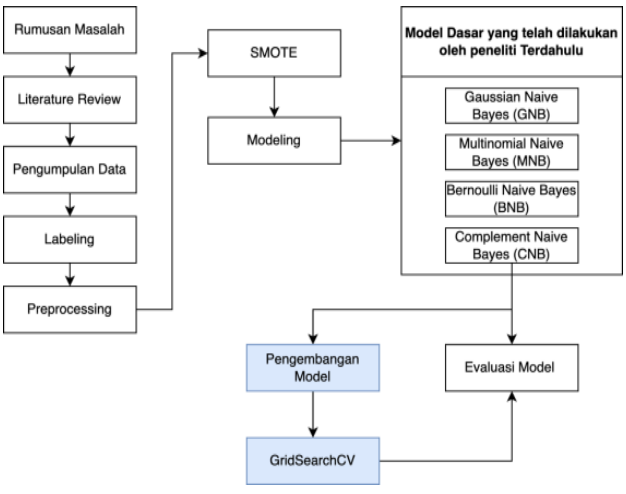
2. MATERIALS AND METHODS

Sejumlah penelitian telah membuktikan keberhasilan algoritma machine learning dalam mendeteksi serangan malware dan anomali jaringan, meskipun masih terdapat keterbatasan dari sisi keandalan, penanganan data tidak seimbang, dan optimasi parameter model. (11) menerapkan stacking ensemble dengan kombinasi model Neural Network, Random Forest, dan k-NN, menghasilkan akurasi tinggi 98,7%, namun tanpa penanganan terhadap class imbalance maupun tuning hyperparameter. Penelitian oleh (12) menggunakan teknik stacking dan random oversampling untuk deteksi penipuan keuangan, membuktikan bahwa metode ensemble meningkatkan AUC secara signifikan. Sayangnya, penelitian ini tidak mengaplikasikan modelnya dalam bentuk sistem otomatis berbasis web, serta tidak menggunakan GridSearchCV untuk pencarian parameter optimal. Penelitian lain oleh (13–15) menunjukkan efektivitas model seperti Decision Tree, Random Forest, dan MLP dalam mendeteksi malware maupun serangan DDoS, tetapi terbatas pada model tunggal, tanpa integrasi metode ensemble, SMOTE, atau deployment aplikasi.

Sementara itu, (16) mengkaji performa SVM dan Neural Network dalam mendeteksi malware, dan menekankan pentingnya model adaptif terhadap serangan yang terus berevolusi. Namun, pendekatan mereka belum mencakup solusi implementatif yang langsung dapat digunakan oleh pengguna. Secara umum, mayoritas penelitian belum mengintegrasikan stacking model yang adaptif, penanganan ketidakseimbangan data menggunakan SMOTE, dan tuning hyperparameter secara sistematis menggunakan GridSearchCV, serta belum banyak yang menyediakan implementasi akhir dalam bentuk aplikasi web berbasis Streamlit untuk penggunaan otomatis oleh pengguna akhir seperti admin server atau tim keamanan TI.

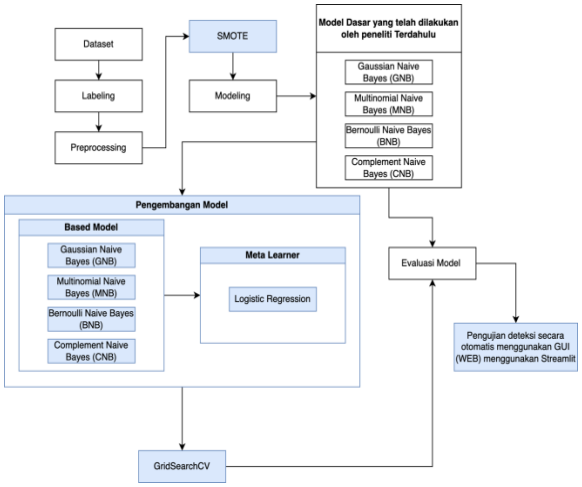
Penelitian ini menghadirkan kebaruan dari berbagai aspek. Pertama, model dikembangkan menggunakan stacking machine learning dengan base learner dari keluarga Naive Bayes yakni Gaussian, Multinomial, Bernoulli, dan Complement yang jarang digunakan dalam konfigurasi ensemble. Kedua, penelitian ini mengintegrasikan teknik SMOTE untuk menangani ketidakseimbangan data malware, sehingga hasil klasifikasi menjadi lebih stabil dan tidak bias terhadap kelas mayoritas. Ketiga, untuk memperoleh konfigurasi model terbaik, dilakukan optimasi hyperparameter secara sistematis menggunakan GridSearchCV, baik pada model dasar maupun meta learner. Keempat, hasil model diimplementasikan ke dalam bentuk aplikasi web yang otomatis dan interaktif berbasis Streamlit, sehingga pengguna seperti admin server dapat melakukan deteksi malware secara real-time melalui antarmuka yang mudah digunakan. Kombinasi dari pendekatan teknis dan aplikasi nyata ini menjadikan penelitian ini unggul dan relevan untuk mendukung keamanan infrastruktur server yang terus menghadapi risiko serangan siber.

Berikut ini merupakan alur penelitian yang digunakan pada penelitian ini.



3. RESULTS AND DISCUSSION

Berikut ini merupakan pengembangan yang dilakukan



Tahapan pengembangan dimulai setelah data dari server UMMUBA selesai melalui proses labeling dan preprocessing. Ketidakseimbangan data antara kelas malware dan non-malware diatasi menggunakan SMOTE, sehingga model dapat belajar secara lebih adil terhadap kedua kelas. Selanjutnya dilakukan pemodelan awal menggunakan empat algoritma Naive Bayes (GNB, MNB, BNB, CNB) sebagai model dasar yang telah umum digunakan dalam penelitian sebelumnya.

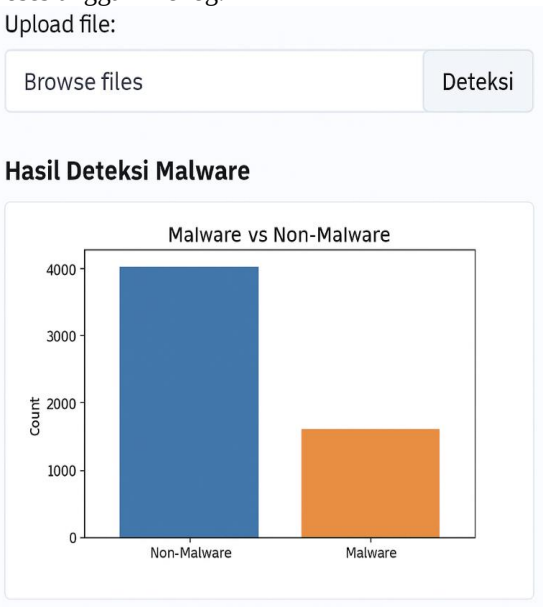
Untuk meningkatkan akurasi dan generalisasi, keempat model tersebut digabungkan dalam arsitektur stacking, dengan Logistic Regression sebagai meta learner. Setelah stacking terbentuk, dilakukan optimasi hyperparameter menggunakan GridSearchCV untuk memperoleh kombinasi parameter terbaik. Model akhir kemudian dievaluasi menggunakan metrik akurasi, precision, recall, dan F1-score untuk menilai kinerjanya dibandingkan model dasar tunggal.

Tahap akhir adalah implementasi ke dalam aplikasi GUI berbasis Streamlit, yang memungkinkan pengguna seperti admin server mengunggah log data dan mendapatkan hasil klasifikasi secara otomatis serta visualisasi performa model.

Gambar dibawah Rancangan Deteksi Malware Perfield

Gambar diatas menampilkan antarmuka GUI deteksi malware berbasis web, di mana pengguna dapat mengunggah dataset dan menekan tombol “Deteksi Malware” untuk memproses data. Hasil yang ditampilkan meliputi prediksi, probabilitas, dan akurasi

Gambar dibawah menunjukkan tampilan deteksi serangan malware dalam jumlah banyak melalui proses unggah file log.



Gambar diatas menampilkan contoh GUI dalam mendeteksi malware, di mana pengguna dapat mengunggah file dan menekan tombol "Deteksi". Hasil klasifikasi ditampilkan dalam grafik batang Malware vs Non-Malware yang memvisualisasikan jumlah data berdasarkan jenisnya secara ringkas dan informatif.

4. CONCLUSION

Hasil pembahasan

Tabel 1. Hasil pembahasan

No	Hasil yang diharapkan	Indoikator Ketercapaian
1	Terciptanya Model Deteksi Malware yang Akurat dan Andal	- Nilai akurasi model mencapai minimal 90%. - Nilai F1-score pada kelas malware ≥ 0,85. - Performa model stacking lebih baik dibandingkan model Naive Bayes individu (GNB, MNB, BNB, CNB).
2	Optimalisasi Performa Model Melalui GridSearchCV	- GridSearchCV berhasil menghasilkan parameter dengan skor validasi tertinggi.

		- Terjadi peningkatan performa ≥ 5% dibanding parameter default.
3	Penanganan Ketidakseimbangan Data dengan SMOTE	- Rasio jumlah data antara kelas mayoritas dan minoritas mendekati 1:1. - Performa model tidak mengalami penurunan setelah diterapkannya SMOTE.
4	Implementasi Sistem Deteksi Malware dalam Bentuk Aplikasi Web (GUI)	- Aplikasi Streamlit dapat dijalankan. - Pengguna dapat mengunggah file log dan mendapatkan hasil klasifikasi secara otomatis. - Output mencakup informasi prediksi, visualisasi hasil, dan evaluasi model.

REFERENCES

1. Vărzaru AA, Bocean CG. Digital Transformation and Innovation: The Influence of Digital Technologies on Turnover from Innovation Activities and Types of Innovation. Systems [Internet]. 2024 Sep 11;12(9):1–25. Available from: https://www.mdpi.com/2079-8954/12/9/359
2. Maraveas C, Rajarajan M, Arvanitis KG, Vatsanidou A. Cybersecurity threats and mitigation measures in agriculture 4.0 and 5.0. Smart Agricultural Technology. 2024 Dec 1;9:1–30.
3. Pratama R, Akbi DR, Nastiti VRS. Classification Of Malware Families Using Naïve Bayes Classifier. REPOSITOR. 2021;3(4):367–74.
4. Sunarto. Sistem Deteksi Malware Menggunakan Algoritma Naive Bayes pada Jaringan IoT. Jurnal BATIRSI [Internet]. 2025 Jan;8(2):2044–8. Available from: https://jurnal.polgan.ac.id/index.php/jmp/article/view/14356
5. Budiana FP, Shelviani H. Kombinasi Algoritma Decision Tree Dan Naive Bayes Untuk Meningkatkan Akurasi Dan Kecepatan Waktu Deteksi Malware. In: Seminar Nasional Riset dan Inovasi Teknologi (SEMNAS RISTEK). 2025. p. 419–25.
6. Anggraini I, Kunang YN, Firdaus. Penerapan Naive Bayes Pada Detection Malware dengan Diskritisasi Variabel. Telematika. 2020 Feb 28;13(1):11–21.
7. Syarif I, Prugel-bennett A, Wills G. SVM Parameter Optimization Using Grid Search and Genetic Algorithm to Improve Classification Performance. TELKOMNIKA. 2016;14(4):1502–9.
8. Krishna BLVSR, Mahalakshmi V, Nukala GKM. A Stacking Model for Outlier Prediction using Learning Approaches. International Journal of Intelligent Systems and Applications in Engineering [Internet]. 2023;12(2s):629–38. Available from: https://www.kaggle.com/datasets/johnsmith88/heart-
9. Putranto A, Azizah NL, Astutik IRI. Web-based Heart Disease Prediction System Using SVM Method and Streamlit Framework. Available from: https://archive.ics.uci.edu/ml/datasets/heart+disease
10. Mubarak MMR, Chrisnanto YH, Sabrina PN. Implementation of Random Forest Using Smote and Smoteenn in Customer Churn Classification in E-Commerce. Enrichment: Journal of Multidisciplinary Research and Development. 2023 Nov 28;1(8):463–77.
11. Rafrastara FA, Supriyanto C, Paramita C, Astuti YP. Deteksi Malware menggunakan Metode Stacking berbasis Ensemble. Jurnal Informatika: Jurnal pengembangan IT [Internet]. 2023;8(1):11–6. Available from: https://orangedatamining.com/
12. Saputra DRK, Via YV, Sihananto AN. Deteksi Anomali Menggunakan Ensemble Learning Dan Random Oversampling Pada Penipuan Transaksi Keuangan. Jurnal Informatika dan Teknik Elektro Terapan. 2024 Aug 3;12(3):2779–88.
13. Arifin MAS, Susilo AAT, Martadinata AT, Santoso B. KLIK: Kajian Ilmiah Informatika dan Komputer Deteksi Aktifitas Malware pada Internet of Things menggunakan Algoritma Decision Tree dan Random Forest. Media Online). 2024;4(6):3073–9.
14. Rahman H, Sutabri T. Analisis Serangan DDOS Menggunakan Machine Learning Pada Arsitektur Software-Define Network Article Info ABSTRAK. JSAI: Journal Scientific and Applied Informatics. 2024;7(3):531–6.
15. Sari L, Faiz MN, Muhammad AW. Perbandingan Pendekatan Machine Learning dalam Deteksi Serangan DDoS Jaringan Komputer. INFOTEKMESIN. 2025;16(1):153–9.
16. Bororing GMG. Evaluasi Kinerja Algoritma Machine Learning Dalam Prediksi Serangan Malware. Jurnal Review Pendidikan dan Pengajaran. 2024;7(1):3060–6.
17. Herianto, Kurniawan B, Hartomi ZH, Irawan Y, Anam MK. Machine Learning Algorithm Optimization using Stacking Technique for Graduation Prediction. Journal of Applied Data Sciences. 2024 Sep 1;5(3):1272–85.

18. Erlin E, Desnelita Y, Nasution N, Suryati L, Zoromi F. Dampak SMOTE terhadap Kinerja Random Forest Classifier berdasarkan Data Tidak seimbang. MATRIK : Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer. 2022 Jul 31;21(3):677–90.