



Feature Selection pada Dataset NSL-KDD Menggunakan Algoritma Genetic Algorithm untuk Deteksi Serangan Jaringan

Freyro Dobry Sianipar^{1*}, Ruth Amelia Vega S Meliala², Yoseph Christian Sitanggang³, Adidtya Perdana⁴

^{1,2,3,4}Program Studi Ilmu Komputer, Universitas Negeri Medan, Indonesia

*Penulis Korespondensi: freyrodobry@mhs.unimed.ac.id

Abstract. Information system security faces serious challenges due to increasingly complex cyber attacks. Intrusion Detection Systems (IDS) require efficient approaches to handle high-dimensional data such as the NSL-KDD dataset with 41 features. This study aims to implement the Genetic Algorithm (GA) for feature selection on the NSL-KDD dataset to improve the efficiency and accuracy of network attack detection. The method used is computational experimental research, involving data preprocessing, GA implementation for feature selection, building a classification model using Random Forest, and performance evaluation based on accuracy, precision, recall, F1-score, and computation time. The results show that GA successfully reduced features from 41 to 12 features (70.7% reduction), significantly improving computational efficiency. However, model accuracy slightly decreased from 0.4973 to 0.4951, indicating that while GA is effective for feature selection, the elimination of certain features may reduce classification capability. The implication of this study is that GA can be used as a tool to simplify intrusion detection models, but it should be combined with parameter optimization and data imbalance handling to achieve more optimal performance.

Keywords: Feature Selection; Genetic Algorithm; Intrusion Detection; Network Security; NSL-KDD.

Abstrak. Keamanan sistem informasi menghadapi tantangan serius akibat serangan siber yang semakin kompleks. *Intrusion Detection System* (IDS) memerlukan pendekatan efisien untuk menangani data berdimensi tinggi seperti pada dataset NSL-KDD yang memiliki 41 fitur. Penelitian ini bertujuan mengimplementasikan *Genetic Algorithm* (GA) untuk seleksi fitur pada dataset NSL-KDD guna meningkatkan efisiensi dan akurasi deteksi serangan jaringan. Metode yang digunakan adalah eksperimental komputasional dengan tahapan preprocessing data, implementasi GA untuk seleksi fitur, pembangunan model klasifikasi menggunakan *Random Forest*, serta evaluasi performa berdasarkan akurasi, presisi, recall, F1-score, dan waktu komputasi. Hasil penelitian menunjukkan bahwa GA berhasil mereduksi fitur dari 41 menjadi 12 fitur (reduksi 70,7%), yang meningkatkan efisiensi komputasi secara signifikan. Namun, akurasi model sedikit menurun dari 0,4973 menjadi 0,4951, mengindikasikan bahwa meskipun GA efektif dalam seleksi fitur, eliminasi beberapa fitur dapat mengurangi kemampuan klasifikasi. Implikasi dari penelitian ini adalah GA dapat digunakan sebagai alat bantu dalam menyederhanakan model deteksi intrusi, namun perlu dikombinasikan dengan optimasi parameter dan penanganan ketidakseimbangan data untuk mencapai performa yang lebih optimal.

Kata kunci: Deteksi Intrusi; *Genetic Algorithm*; Keamanan Jaringan; NSL-KDD; Seleksi Fitur.

1. LATAR BELAKANG

Keamanan sistem informasi di era digital menghadapi tantangan besar akibat serangan siber yang semakin kompleks dan dinamis. *Intrusion Detection System* (IDS) merupakan komponen vital dalam arsitektur keamanan jaringan untuk mendeteksi aktivitas mencurigakan. Namun, efektivitas IDS sering terhambat oleh dimensi data yang tinggi, seperti pada dataset NSL-KDD yang memiliki 41 fitur. Penggunaan seluruh fitur tanpa seleksi dapat menyebabkan beban komputasi yang berat dan penurunan akurasi deteksi (Putra & Amarudin, 2025). Oleh karena itu, diperlukan metode komputasi cerdas yang mampu menyeleksi fitur penting secara efisien.

Pendekatan algoritma *metaheuristic* telah diakui secara luas sebagai solusi ampuh untuk menyelesaikan masalah optimasi yang rumit dan *NP-hard* di berbagai bidang (Tohidi &

Rustamov, 2022);(Ali et al., 2022). Berbagai algoritma baru terus dikembangkan, seperti *Sculptor Optimization Algorithm* (SOA) yang terinspirasi dari proses memahat untuk masalah teknik (Hamadneh et al., 2024), hingga penerapan algoritma optimasi pada desain berbasis reliabilitas (*Reliability-Based Design Optimization*) yang membuktikan ketangguhan metode ini dalam menangani variabel campuran (Meng et al., 2021).

Secara spesifik, *Genetic Algorithm* (GA) yang berbasis pada prinsip evolusi alam, telah menunjukkan fleksibilitas dan keandalan yang tinggi dalam berbagai studi kasus. Pada bidang penjadwalan, (Pangestu et al., 2023) berhasil menggunakan GA untuk penjadwalan mata pelajaran yang optimal, sementara (Cirua & Cokrowibowo, 2023), mengembangkan representasi kromosom *Gray Code* pada GA untuk menyelesaikan masalah *job shop scheduling* dengan hasil *makespan* yang lebih baik.

Di bidang kesehatan, peran algoritma *metaheuristic* juga sangat signifikan. (Kaur et al., 2023) mengulas bagaimana teknik ini efektif untuk seleksi fitur dalam diagnosis penyakit. Hal ini diperkuat oleh (Buani, 2021) yang membuktikan bahwa penerapan GA sebagai metode seleksi fitur pada algoritma *Naïve Bayes* mampu meningkatkan akurasi prediksi gagal jantung secara drastis dari 69,60% menjadi 96,67%. Selain itu, (Putri, 2025) menunjukkan bahwa integrasi GA untuk mengoptimalkan parameter *Backpropagation* mampu meningkatkan akurasi prediksi tingkat pengangguran dengan nilai MAPE yang rendah (3,27%).

Meskipun berbagai studi di atas menunjukkan keunggulan GA dalam optimasi dan seleksi fitur di bidang kesehatan dan industri, penerapannya secara spesifik untuk reduksi dimensi pada dataset NSL-KDD guna meningkatkan deteksi serangan jaringan masih perlu dieksplorasi lebih lanjut. Studi terbaru oleh (Putra & Amarudin, 2025) pada dataset NSL-KDD menunjukkan bahwa algoritma *Boosting* mencapai akurasi tinggi (99,36%), namun memiliki kelemahan pada waktu pelatihan yang lama. Di sisi lain, penggunaan fitur yang tidak relevan sering kali menjadi penyebab inefisiensi tersebut. Kesenjangan penelitian ini terletak pada integrasi kemampuan seleksi fitur *Genetic Algorithm*—yang terbukti sukses di bidang lain (Buani, 2021);(Putri, 2025) ke dalam domain keamanan jaringan untuk memilih subset fitur NSL-KDD yang paling optimal, sehingga diharapkan dapat mempertahankan akurasi tinggi sekaligus mengefisienkan komputasi.

Penelitian ini bertujuan untuk mengimplementasikan *feature selection* menggunakan algoritma *Genetic Algorithm* (GA) pada dataset NSL-KDD. Tujuan utamanya adalah mendapatkan subset fitur yang relevan untuk mendeteksi serangan jaringan, serta menganalisis dampak reduksi fitur tersebut terhadap akurasi dan efisiensi model deteksi intrusi.

2. KAJIAN TEORITIS

Intrusion Detection System (IDS) dan Keamanan Jaringan

Keamanan jaringan merupakan aspek vital dalam infrastruktur teknologi informasi. Salah satu mekanisme pertahanan utama adalah *Intrusion Detection System (IDS)*, yang berfungsi sebagai sistem deteksi dini terhadap serangan jaringan. (Khaliq & Sari, 2022) mendefinisikan IDS sebagai teknologi yang tidak hanya memberikan peringatan kepada administrator saat terjadi serangan, tetapi juga mencatat aktivitas yang bertujuan mengganggu jaringan. IDS dapat beroperasi berdasarkan basis aturan (*rule-based*) atau deteksi anomali. Namun, kompleksitas serangan siber modern menuntut sistem yang lebih adaptif dibandingkan metode tradisional yang statis. (Prabuningrat et al., 2024) menekankan bahwa deteksi anomali bertujuan untuk menemukan perilaku menyimpang dari pola normal dalam jaringan, yang sangat krusial untuk mengantisipasi serangan baru yang belum teridentifikasi oleh basis data *signature*.

Machine Learning dan Klasifikasi Serangan

Penerapan *Machine Learning (ML)* dalam IDS telah menjadi solusi populer untuk mengatasi kelemahan sistem berbasis aturan. (Sari et al., 2024) menjelaskan bahwa teknik *supervised learning* menggunakan dataset lalu lintas jaringan normal dan serangan untuk melatih algoritma agar mampu mendeteksi intrusi secara *real-time*. Dalam studi komparatif pada dataset NSL-KDD, (Putra & Amarudin, 2025) mengevaluasi algoritma seperti *Decision Tree*, *Random Forest*, *SVM*, dan *Boosting*. Hasil penelitian mereka menunjukkan bahwa meskipun algoritma *Boosting* mencapai akurasi tertinggi (99,36%), algoritma ini membutuhkan waktu pelatihan yang signifikan. Hal ini mengindikasikan adanya *trade-off* antara akurasi dan efisiensi komputasi, yang sering kali disebabkan oleh tingginya dimensi fitur pada dataset serangan jaringan seperti NSL-KDD.

Algoritma Metaheuristik dan Genetic Algorithm

Untuk menangani permasalahan optimasi yang kompleks, algoritma metaheuristik menawarkan pendekatan yang efisien. (Ali et al., 2022) mengklasifikasikan metaheuristik sebagai algoritma cerdas yang mampu memecahkan masalah optimasi *NP-hard* dengan memprediksi solusi di ruang pencarian yang luas. Salah satu jenis metaheuristik yang paling populer adalah *Evolutionary Algorithms*, di mana *Genetic Algorithm (GA)* menjadi representasi utamanya.

Algoritma Genetika didasarkan pada prinsip seleksi alam Darwin, yang melibatkan operator evolusi seperti seleksi, *crossover* (pindah silang), dan mutasi untuk menghasilkan solusi yang lebih baik dari generasi ke generasi. Keunggulan GA terletak pada kemampuannya untuk menghindari jebakan optima lokal (*local optima*) dan menemukan solusi global yang

optimal (Tohidi & Rustamov, 2022). Efektivitas tahapan GA—mulai dari inialisasi populasi hingga evaluasi fungsi *fitness*—dalam menyelesaikan masalah penjadwalan yang memiliki banyak batasan (*constraints*), menunjukkan fleksibilitas GA untuk diterapkan pada berbagai domain masalah kombinatorial (Pangestu et al., 2023).

Dalam konteks *data mining* dan *machine learning*, GA sering digunakan untuk *feature selection* atau seleksi fitur guna mereduksi dimensi data yang tidak relevan. Teknik metaheuristik, termasuk GA, sangat efektif dalam menyeleksi fitur optimal untuk diagnosis penyakit, yang berdampak langsung pada peningkatan akurasi prediksi (Kaur et al., 2023).

Efektivitas GA dalam meningkatkan kinerja model klasifikasi telah dibuktikan dalam beberapa studi empiris. (Septiani & Rohwadi, 2021) menerapkan optimasi GA sebagai fitur seleksi pada algoritma C4.5 untuk deteksi dini penyakit diabetes. Hasil penelitian mereka menunjukkan peningkatan akurasi dari 95,96% menjadi 96,54% setelah penerapan GA, membuktikan bahwa reduksi fitur yang tepat dapat meningkatkan performa model. Sejalan dengan hal tersebut, GA untuk mengoptimalkan parameter pada Jaringan Syaraf Tiruan *Backpropagation*, yang menghasilkan penurunan *error* dan peningkatan koefisien determinasi (R^2) secara signifikan (Putri, 2025). Selain itu, modifikasi representasi kromosom pada GA (seperti *Gray Code*) dapat memberikan solusi optimasi yang lebih baik dengan nilai rata-rata *makespan* yang lebih kecil dalam masalah penjadwalan (Shodiq et al., 2024).

Berdasarkan tinjauan literatur di atas, dataset NSL-KDD yang digunakan dalam deteksi intrusi memiliki dimensi fitur yang tinggi, yang dapat membebani komputasi dan menurunkan kinerja algoritma klasifikasi (Aljammal et al., 2024). Mengingat keberhasilan GA dalam melakukan optimasi dan seleksi fitur di berbagai bidang seperti kesehatan (Fauziah et al., 2022);(Nugroho et al., 2021) dan penjadwalan (Saputra & Sukmono, 2024), penelitian ini berlandaskan pada pemikiran bahwa penerapan *Genetic Algorithm* sebagai metode *feature selection* pada dataset NSL-KDD akan mampu mengeliminasi fitur yang redundan atau tidak relevan. Hal ini secara teoritis akan menghasilkan subset fitur yang lebih optimal, yang pada gilirannya dapat meningkatkan akurasi deteksi serangan jaringan sekaligus mengefisienkan waktu komputasi model *machine learning* (Shodiq et al., 2024) ; (Pamungkas et al., 2022).

3. METODE PENELITIAN

Jenis Penelitian

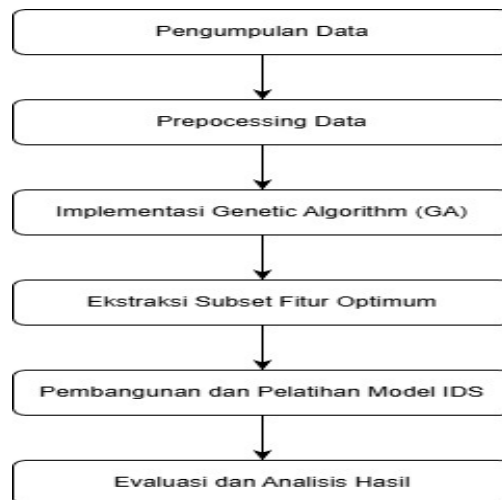
Penelitian ini merupakan penelitian eksperimental komputasional yang berfokus pada penerapan algoritma Genetic Algorithm (GA) untuk proses seleksi fitur (*feature selection*) pada dataset NSL-KDD dalam konteks deteksi serangan jaringan (Intrusion Detection System/IDS).

Penelitian ini bersifat kuantitatif, karena seluruh analisis dan evaluasi dilakukan berdasarkan data numerik dan hasil perhitungan metrik performa seperti akurasi, precision, recall, dan F1-score. Melalui pendekatan ini, dilakukan pengujian secara sistematis terhadap bagaimana penerapan metode GA mampu meningkatkan kinerja model deteksi serangan jaringan dibandingkan dengan penggunaan seluruh fitur tanpa seleksi.

Selain itu, penelitian ini juga termasuk dalam kategori eksperimen berbasis data (data-driven experiment) di bidang keamanan siber dan optimasi algoritmik. Fokus utama penelitian ini adalah mengamati pengaruh proses optimasi fitur terhadap efisiensi model pembelajaran mesin (machine learning) dalam mengenali pola serangan pada jaringan komputer. Dengan menggunakan dataset NSL-KDD sebagai studi kasus, penelitian ini tidak hanya menilai kemampuan GA dalam menemukan subset fitur yang relevan, tetapi juga membuktikan sejauh mana pendekatan metaheuristik dapat digunakan sebagai solusi efektif untuk mengatasi masalah kompleksitas tinggi dan redundansi data pada sistem deteksi intrusi berbasis machine learning.

Tahapan Penelitian

Penelitian ini akan terdiri dari enam tahap sebagaimana ditunjukkan pada gambar berikut.



Gambar 1. Alur penelitian.

Pengumpulan Data

Tahapan pertama adalah pengumpulan dataset yang digunakan sebagai dasar dalam penelitian. Dataset yang digunakan yaitu NSL-KDD, yang merupakan hasil penyempurnaan dari KDD Cup 1999 dataset. Dataset ini banyak digunakan dalam penelitian sistem deteksi intrusi (IDS) karena telah menghilangkan duplikasi data dan ketidakseimbangan kelas yang ada pada versi sebelumnya. Dataset ini berisi 41 fitur yang mempersentasikan berbagai karakteristik lalu lintas jaringan seperti durasi koneksi, jenis protokol, status flag hingga

jumlah paket yang dikirim yang dikelompokkan menjadi 5 kelas yakni Normal , DoS , Probe , U2R, dan R2L.

Preprocessing Data

Tahap *preprocessing* dilakukan untuk mempersiapkan data agar layak digunakan dalam proses analisis dan pelatihan model. Langkah pertama adalah *data cleaning*, yaitu menghapus data duplikat, menangani nilai kosong (*missing values*), serta memastikan konsistensi tipe data sehingga tidak terjadi kesalahan saat pemrosesan. Selanjutnya dilakukan proses *encoding* terhadap fitur kategorikal, seperti *protocol_type*, *service*, dan *flag*, menggunakan teknik *Label Encoding* atau *One-Hot Encoding* agar fitur-fitur tersebut dapat diolah oleh algoritma pembelajaran mesin yang hanya menerima nilai numerik. Tahap berikutnya adalah normalisasi menggunakan metode *Min-Max Normalization* untuk menstandarkan nilai pada rentang 0–1 sehingga seluruh fitur numerik berada pada skala yang seimbang. Proses *preprocessing* ini sangat penting untuk memastikan bahwa data yang digunakan tidak mengandung bias serta memungkinkan setiap fitur memberikan kontribusi yang proporsional dalam proses optimasi dan klasifikasi model.

Implementasi Genetic Algorithm (GA)

Pada tahap ini diterapkan *Genetic Algorithm* (GA) untuk melakukan pencarian fitur yang paling relevan terhadap proses klasifikasi. GA merupakan algoritma evolusioner yang mengadopsi prinsip seleksi alam melalui mekanisme seleksi, *crossover*, dan mutasi. Dalam penelitian ini, setiap solusi direpresentasikan dalam bentuk kromosom biner sepanjang 41 bit, di mana nilai “1” menunjukkan bahwa fitur dipilih dan nilai “0” menunjukkan bahwa fitur diabaikan. Populasi awal dibangkitkan secara acak dengan sejumlah individu tertentu, misalnya 50 kromosom. Setiap kromosom kemudian dievaluasi berdasarkan nilai *fitness* yang dihitung dari akurasi model klasifikasi—seperti *Random Forest*—yang dilatih menggunakan subset fitur yang ditunjukkan oleh kromosom tersebut. Proses evolusi dilanjutkan melalui tahap seleksi menggunakan metode seperti *roulette wheel*, *crossover* tipe *single-point*, serta mutasi dengan probabilitas sekitar 0,1 untuk menjaga keragaman populasi. Tahap evolusi ini berlangsung hingga memenuhi kriteria terminasi, baik setelah mencapai jumlah generasi tertentu (misalnya 100 generasi) maupun ketika tidak terjadi peningkatan signifikan pada nilai *fitness*. Melalui mekanisme tersebut, GA diharapkan mampu mengidentifikasi kombinasi fitur yang paling optimal dalam mendukung performa model klasifikasi.

Ekstraksi Subset Fitur Optimum

Tahapan ini bertujuan untuk memperoleh subset fitur optimum berdasarkan hasil optimasi dari Genetic Algorithm. Fitur-fitur yang terpilih dianggap memiliki kontribusi paling

besar dalam membedakan antara trafik normal dan trafik serangan. Proses ini disebut juga Optimal Feature Subset Extraction, yang menghasilkan fitur-fitur esensial untuk digunakan pada tahap modeling berikutnya. Dengan pengurangan fitur yang tidak relevan, diharapkan terjadi peningkatan efisiensi komputasi, pengurangan overfitting, dan peningkatan akurasi model klasifikasi. Tahap ini juga melibatkan analisis terhadap fitur yang terpilih untuk memastikan relevansinya dengan domain keamanan jaringan.

Pembangunan dan Pelatihan Model IDS

Setelah subset fitur optimum diperoleh, tahap selanjutnya adalah membangun model deteksi serangan jaringan (Intrusion Detection System) menggunakan algoritma klasifikasi. Algoritma yang digunakan Random Forest (RF). Pada tahap ini, dataset dibagi menjadi dua bagian, yaitu data latih dan data uji, dengan proporsi 80:20. Model kemudian dilatih menggunakan data latih yang telah direduksi berdasarkan subset fitur hasil Genetic Algorithm, sedangkan pengujian dilakukan menggunakan data uji untuk mengevaluasi kemampuan generalisasi model terhadap data yang belum pernah dilihat sebelumnya. Tahap ini bertujuan untuk memastikan bahwa model tidak hanya mampu mengenali pola pada data pelatihan, tetapi juga dapat melakukan klasifikasi secara akurat pada data baru.

Evaluasi dan Analisis Hasil

Tahap akhir penelitian adalah melakukan evaluasi terhadap performa model sebelum dan sesudah proses seleksi fitur untuk menilai efektivitas penerapan Genetic Algorithm. Evaluasi dilakukan menggunakan beberapa metrik, antara lain accuracy, precision, recall, F1-score, serta waktu komputasi pelatihan. Selain itu, nilai *confusion matrix* dihitung untuk setiap model guna memperoleh gambaran yang lebih mendetail mengenai kemampuan klasifikasi. Analisis hasil mencakup perbandingan jumlah fitur yang digunakan, waktu pelatihan yang dibutuhkan, serta perubahan performa model setelah penerapan seleksi fitur berbasis Genetic Algorithm. Seluruh temuan dari proses evaluasi ini kemudian menjadi dasar dalam penyusunan kesimpulan penelitian.

4. HASIL DAN PEMBAHASAN

Analisis Data Awal

Analisis data awal dilakukan untuk memahami karakteristik dasar dari dataset NSL-KDD sebelum dilakukan proses preprocessing dan feature selection. Dataset ini terdiri dari 41 fitur sebagai variabel independen dan 1 label sebagai variabel target yang merepresentasikan kategori serangan maupun trafik normal. Fitur pada dataset mencakup berbagai jenis informasi, mulai dari fitur numerik seperti jumlah koneksi, durasi koneksi, hingga fitur kategorikal seperti

protocol_type, service, dan flag yang menggambarkan kondisi koneksi jaringan. Selain itu, dataset memiliki distribusi kelas yang tidak seimbang, di mana kelas “normal” dan beberapa jenis serangan (misalnya DoS) memiliki jumlah sampel jauh lebih besar dibandingkan kelas serangan lain seperti R2L dan U2R. Ketidakeimbangan ini menjadi perhatian penting karena dapat memengaruhi performa model dalam mengenali serangan minoritas. Analisis ini juga mencakup identifikasi nilai yang hilang, data redundan, dan struktur data keseluruhan untuk memastikan bahwa proses preprocessing berikutnya dapat dilakukan dengan benar. Dengan memahami karakteristik awal dataset, peneliti dapat merencanakan strategi feature selection dan model deteksi yang lebih tepat sasaran.

Tabel 1. Karakteristik dataset.

Informasi	Nilai
Jumlah Fitur	41
Jumlah Data Training	125971
Jumlah Data Testing	22542
Jumlah Kelas	2

Implementasi Genetic Algorithm (GA)

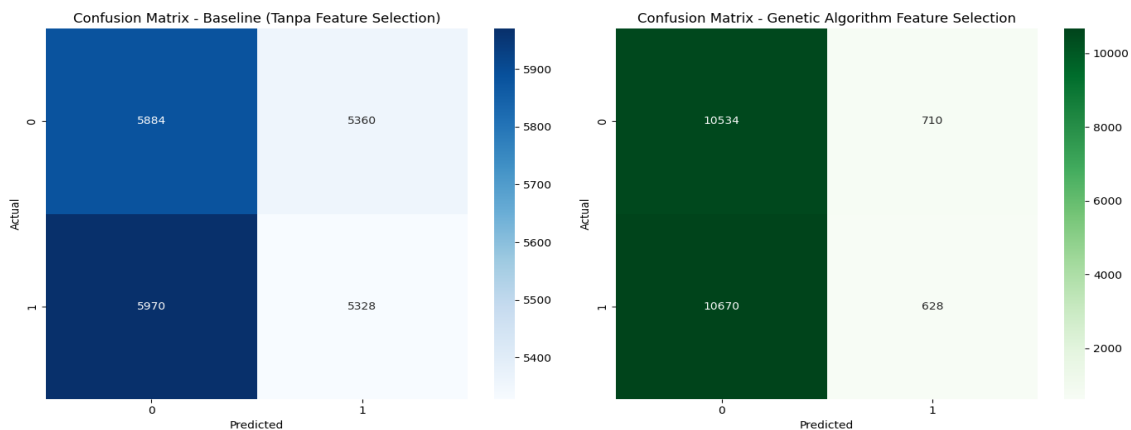
Implementasi algoritma Genetic Algorithm (GA) pada penelitian ini dilakukan untuk melakukan seleksi fitur terhadap dataset NSL-KDD dengan tujuan memperoleh subset fitur yang paling relevan dalam mendeteksi serangan jaringan. Proses GA diawali dengan membangkitkan populasi awal berupa kromosom biner yang merepresentasikan pemilihan fitur, di mana nilai “1” menandakan fitur dipilih dan nilai “0” menandakan fitur tidak digunakan. Setiap kromosom kemudian dievaluasi menggunakan nilai akurasi dari model *Random Forest Classifier* sebagai fungsi *fitness*, sehingga semakin tinggi akurasi maka semakin tinggi nilai *fitness*-nya. Proses ini berlangsung selama beberapa generasi sampai GA menemukan kombinasi fitur optimal atau mencapai batas iterasi yang ditentukan. GA menghasilkan subset fitur terpilih sebanyak 12 fitur dari total 41 fitur awal. Fitur-fitur ini dianggap memiliki kontribusi terbesar dalam membedakan trafik normal dan serangan berdasarkan evaluasi fungsi *fitness*. Hasil seleksi fitur ini kemudian digunakan untuk pelatihan model, sehingga dapat dibandingkan performanya dengan model baseline tanpa seleksi fitur. Implementasi GA pada penelitian ini menjadi fase penting dalam upaya meningkatkan efisiensi model dengan mengurangi jumlah fitur tanpa mengorbankan performa secara signifikan.

Tabel 2. Fitur terpilih.

No	Fitur
1	flag
2	land
3	urgent
4	hot
5	logged_in
6	num_compromised
7	root_shell
8	num_file_creations
9	num_access_files
10	num_outbound_cmds
11	dst_host_srv_diff_host_rate
12	dst_host_srv_rerror_rate

Evaluasi Model

Evaluasi model dilakukan dengan membandingkan kinerja algoritma klasifikasi sebelum dan sesudah penerapan *Genetic Algorithm* (GA) sebagai metode *feature selection*. Pada kondisi awal, model menggunakan seluruh 41 fitur dan memperoleh akurasi sebesar 0,4973. Setelah dilakukan seleksi fitur menggunakan GA, jumlah fitur berkurang secara signifikan menjadi 12 fitur terpilih. Meskipun proses seleksi ini berhasil mereduksi dimensi data hingga sekitar 70%, performa model tidak menunjukkan peningkatan. Akurasi model setelah menggunakan fitur hasil seleksi justru sedikit menurun menjadi 0,4951, dengan selisih akurasi sebesar -0,0022 dibandingkan baseline.

**Gambar 2.** Perbandingan confusion matrix.

Interpretasi Hasil

Penerapan *Genetic Algorithm* (GA) pada proses seleksi fitur memberikan dampak yang cukup signifikan terhadap struktur data, meskipun tidak menghasilkan peningkatan akurasi model. Akurasi model mengalami penurunan kecil dari 0,4973 menjadi 0,4951 setelah jumlah fitur direduksi dari 41 menjadi 12 fitur. Penurunan ini mengindikasikan bahwa beberapa fitur yang dieliminasi GA kemungkinan memiliki kontribusi penting terhadap proses klasifikasi, sehingga kehilangan fitur tersebut sedikit mengurangi kemampuan model membedakan serangan dan trafik normal.

Pemilihan fitur tertentu oleh GA menunjukkan bahwa algoritma ini lebih menekankan pada fitur-fitur yang memiliki korelasi kuat terhadap pola serangan dalam dataset NSL-KDD, seperti *num_compromised*, *root_shell*, *logged_in*, atau *dst_host_srv_diff_host_rate*. Fitur-fitur tersebut secara historis memang dikenal sebagai indikator utama perilaku anomali pada intrusi jaringan. GA memilih fitur berdasarkan evaluasi fitness, sehingga fitur yang berulang kali terbukti meningkatkan nilai fitness selama proses evolusi akan bertahan hingga generasi akhir. Terkait efektivitas GA pada dataset NSL-KDD, hasil ini menunjukkan bahwa GA berhasil melakukan reduksi fitur secara drastis, yaitu sebesar **70,7%**, sehingga meningkatkan efisiensi model dari sisi kompleksitas, waktu komputasi, dan kebutuhan sumber daya.

Secara keseluruhan, meskipun akurasi sedikit menurun, penggunaan Genetic Algorithm tetap memberikan manfaat, terutama dalam mengurangi jumlah fitur dari 41 menjadi 12 fitur dan meningkatkan efisiensi model. Reduksi fitur ini tidak hanya mempercepat proses pelatihan, namun juga membantu mengidentifikasi fitur-fitur yang paling berpengaruh dalam mendeteksi serangan.

5. KESIMPULAN DAN SARAN

Berdasarkan hasil penelitian, dapat disimpulkan bahwa penerapan Genetic Algorithm (GA) dalam proses seleksi fitur pada dataset NSL-KDD mampu mereduksi jumlah fitur secara signifikan dari 41 menjadi 12 fitur, sehingga meningkatkan efisiensi komputasi dan menyederhanakan kompleksitas model deteksi intrusi. Namun, reduksi fitur tersebut belum menunjukkan peningkatan akurasi model, ditandai dengan sedikitnya penurunan performa dari akurasi baseline sebesar 0,4973 menjadi 0,4951 setelah menggunakan fitur hasil seleksi GA. Temuan ini menunjukkan bahwa meskipun GA efektif dalam mengidentifikasi fitur-fitur yang paling relevan, tidak semua eliminasi fitur berkontribusi positif terhadap performa klasifikasi, terutama pada dataset yang memiliki ketergantungan fitur yang kompleks seperti NSL-KDD. Keterbatasan penelitian ini terletak pada penggunaan satu jenis algoritma klasifikasi serta

parameter GA yang belum dieksplorasi secara mendalam. Oleh karena itu, penelitian selanjutnya direkomendasikan untuk melakukan optimasi parameter GA yang lebih komprehensif, membandingkan beberapa algoritma klasifikasi, serta menerapkan teknik penanganan ketidakseimbangan data seperti SMOTE untuk memperoleh performa yang lebih representatif dan generalisasi yang lebih baik. Dengan demikian, penelitian ini memberikan dasar yang kuat bahwa GA memiliki potensi dalam seleksi fitur pada sistem deteksi intrusi, namun perlu dikombinasikan dengan pendekatan tambahan agar dapat menghasilkan peningkatan performa yang lebih signifikan.

DAFTAR REFERENSI

- Ali, K. W., Hawezi, R. S., Kareem, S. W., Khoshabai, F. S., & Askar, S. K. (2022). Metaheuristic algorithms in optimization and its application: A review. *Journal of Applied Research in Engineering and Engineering Education*, 6(1), 7–12. <https://doi.org/10.12962/jaree.v6i1.216>
- Aljammal, A. H., Al-Oqily, I., Obiedat, M., Qawasmeh, A., Taamneh, S., & Wedyan, F. I. (2024). Anomaly intrusion detection using machine learning–IG-R based on NSL-KDD dataset. *Bulletin of Electrical Engineering and Informatics*, 13(6), 4466–4474. <https://doi.org/10.11591/eei.v13i6.7308>
- Buani, D. C. P. (2021). Penerapan algoritma naïve Bayes dengan seleksi fitur algoritma genetika untuk prediksi gagal jantung. *Jurnal Evolusi*, 9(2), 43–48. <https://doi.org/10.31294/evolusi.v9i2.11141>
- Cirua, A. A. A., & Cokrowibowo, S. (2023). Representasi chromosome gray code algoritma genetika pada job shop scheduling problem. *Senarai*, 184–189.
- Fauziah, A. S., Cholissodin, I., & Rahayudi, B. (2022). Optimasi pendistribusian air mineral menggunakan algoritma genetika. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 6(2), 966–972. <http://j-ptiik.ub.ac.id>
- Hamadneh, T., Kaabneh, K., Alsayed, O., Montazeri, Z., Dehghani, M., Bektemyssova, G., & Eguchi, K. (2024). Sculptor optimization algorithm: A new human-inspired metaheuristic algorithm for solving optimization problems. *International Journal of Intelligent Engineering and Systems*, 17(4), 564–575. <https://doi.org/10.22266/ijies2024.0831.43>
- Kaur, S., Kumar, Y., Koul, A., & Kamboj, S. K. (2023). A systematic review on metaheuristic optimization techniques for feature selections in disease diagnosis: Open issues and challenges. *Archives of Computational Methods in Engineering*, 30(3), 1847–1875. <https://doi.org/10.1007/s11831-022-09853-1>
- Khaliq, A., & Sari, S. N. (2022). Lisensi internasional Creative Commons Attribution-ShareAlike 4.0. *Jurnal Nasional Teknologi Komputer*, 2(3), 150–158. <https://doi.org/10.61306/jnastek.v2i3.52>
- Meng, Z., Li, G., Wang, X., Sait, S. M., & Yıldız, A. R. (2021). A comparative study of metaheuristic algorithms for reliability-based design optimization problems. *Archives of Computational Methods in Engineering*, 28(3), 1853–1869. <https://doi.org/10.1007/s11831-020-09443-z>

- Nugroho, B., Puspaningrum, E. Y., & Munir, M. S. (2021). Kinerja algoritma optimasi root-mean-square propagation dan stochastic gradient descent pada klasifikasi pneumonia COVID-19 menggunakan CNN. *Jurnal Edukasi dan Penelitian Informatika (JEPIN)*, 7(3), 420–427. <https://doi.org/10.26418/jp.v7i3.49172>
- Pamungkas, P. Y., Mayke, N., & Normasari, N. M. E. (2022). Algoritma spotted hyena optimizer pada capacitated vehicle routing problem. Institut Teknologi Adhi Tama Surabaya.
- Pangestu, L. A., Suryawan, S. H., & Latipah, A. J. (2023). Penerapan algoritma genetika dalam penjadwalan mata pelajaran. *Jurnal Informatika*, 10(2), 194–205. <https://doi.org/10.31294/inf.v10i2.16701>
- Prabuningrat, G. S. W., Hostiadi, D. P., & Srinadi, N. L. P. (2024). Klasifikasi deteksi anomali menggunakan metode machine learning. *Spinter*, 1(2).
- Putra, R. P., & Amarudin. (2025). Perbandingan algoritma machine learning untuk intrusion detection system pada dataset NSL-KDD. *Jurnal Teknologi dan Manajemen Sistem Informasi*, 14, 1654–1664. <https://doi.org/10.32520/stmsi.v14i4.5246>
- Putri, N. W. (2025). Optimasi jaringan syaraf tiruan metode backpropagation dengan algoritma genetika untuk prediksi tingkat pengangguran di Provinsi Sumatera Utara. *Jurnal Manajemen, Pendidikan dan Ilmu Komputer*, 2(1), 29–36. <https://doi.org/10.65309/yjwq8x86>
- Saputra, N. Q., & Sukmono, T. (2024). Analisis optimalisasi rute distribusi untuk mengefisiensikan logistik menggunakan algoritma genetika. *Matrik: Jurnal Manajemen dan Teknik Industri Produksi*, 25(1), 67–75. <https://doi.org/10.30587/matrik.v25i1.7989>
- Sari, D. P., Halim, Z., Irlon, I., Waseso, B., & Saromah, S. (2024). Implementasi machine learning untuk deteksi intrusi pada jaringan komputer. *Jurnal Minfo Polgan*, 13(2), 1389–1394. <https://doi.org/10.33395/jmp.v13i2.14074>
- Septiani, W. D., & Rohwadi, U. (2021). Optimasi algoritma genetika pada algoritma C4.5 untuk deteksi dini penyakit diabetes. *Jurnal Akrab Juara*, 6, 1–6.
- Shodiq, M. F., Sasongko, B. T., Ardiansa, M. A., & Haq, M. A. (2024). Memprediksi kerusakan peralatan pada pembangkit listrik Jawa Bali menggunakan algoritma genetika. *Prosiding Seminar Nasional Sistem Informasi (SENASIF)*, 8, 5141–5150.
- Tohidi, N., & Rustamov, R. B. (2022). Short overview of advanced metaheuristic methods.