P- ISSN: 2355-6684, E-ISSN: 2776-6640
https://ojs.unpkediri.ac.id/index.php/noe

Hybrid Ensemble Learning Sistem Keamanan Jaringan untuk Meningkatkan Performa Deteksi Anomali

Nico Adi Saputra[1], Rony Heri Irawan[2], Umi Mahdiyah[3]
Teknik Informatika Universitas Nusantara PGRI Kediri[1,2,3]
nicoas441@gmail.com[1], rony@unpkediri.ac.id[2], umimahdiyah@unpkediri.ac.id[3]

Abstrak

Serangan siber seperti zero-day attacks dan APT menjadi tantangan serius bagi sistem deteksi intrusi jaringan, terutama yang masih mengandalkan metode berbasis tanda tangan. Penelitian ini bertujuan merancang sistem deteksi anomali jaringan berbasis hybrid ensemble learning dengan menggabungkan algoritma Isolation Forest, K-Means, dan Random Forest menggunakan metode majority voting. Proses penelitian meliputi preprocessing data, pelatihan dan evaluasi model menggunakan dataset publik CSE-CIC-IDS2018. Evaluasi dilakukan dengan metrik akurasi, precision, recall, F1-score, dan AUC. Hasil menunjukkan bahwa pendekatan hybrid ini meningkatkan akurasi deteksi hingga 99,9% dan menurunkan false positive secara signifikan dibanding pendekatan tunggal. Sistem yang diusulkan terbukti lebih adaptif dan efisien dalam mengidentifikasi berbagai pola serangan siber, serta memberikan kontribusi terhadap pengembangan teknologi keamanan jaringan yang lebih andal.

Kata Kunci : *Deteksi Anomali, Hybrid Ensemble, Keamanan Siber, Serangan Siber, Network Intrusion Detection Sistem*

A. PENDAHULUAN

Perkembangan teknologi digital telah menjadikan keamanan siber sebagai faktor krusial dalam melindungi data dan sistem informasi. Berbagai sektor, seperti pemerintahan, kesehatan, dan bisnis, kini semakin mengandalkan jaringan komputer untuk mendukung operasi sehari-hari mereka. Ketergantungan ini, meskipun meningkatkan efisiensi dan produktivitas, juga membuka peluang bagi penyerang untuk mengeksploitasi kelemahan sistem melalui serangan siber yang canggih, seperti serangan *zero-day* dan *Advanced Persistent Threats* (APT) (Alserhani & Aljared, 2023; Goswami, 2024.). Untuk menghadapi ancaman tersebut, *Network Intrusion Detection Systems* (NIDS) telah menjadi alat yang sangat penting dalam sistem keamanan jaringan. NIDS dirancang untuk memantau aliran data dalam jaringan dan mendeteksi aktivitas yang mencurigakan yang bisa menandakan adanya serangan. Namun, metode tradisional yang berbasis tanda tangan *(signature-based detection)* memiliki kekurangan besar karena hanya dapat mendeteksi serangan yang sudah dikenal sebelumnya, sehingga tidak efektif dalam menghadapi serangan baru yang belum memiliki tanda tangan (Alserhani & Aljared, 2023).

Sebagai alternatif, pendekatan deteksi anomali *(anomaly-based detection)* menawarkan solusi yang lebih fleksibel dengan menggunakan model pembelajaran untuk mengenali pola lalu lintas jaringan yang normal dan mendeteksi anomali sebagai tanda serangan. Meskipun metode ini menjanjikan, sering kali menghasilkan *false positives* yang tinggi, yang membebani pengguna dengan analisis manual dan menurunkan tingkat kepercayaan terhadap sistem deteksi. Beberapa penelitian telah mengeksplorasi pendekatan deteksi anomali dalam sistem keamanan jaringan. (Santoso et al., 2024) menggunakan algoritma *Naive Bayes* yang mencapai akurasi 86%, namun masih menghadapi masalah *false positives*. (Agustina et al., 2024) menerapkan algoritma *Random Forest* dengan metode pemilihan fitur *wrapper*, yang menghasilkan akurasi lebih tinggi yaitu 91,51%, dibandingkan dengan metode filter. (Bakhare., et al, 2024) mengevaluasi beberapa algoritma dan menemukan bahwa *Random Forest* memberikan akurasi terbaik sebesar 87%. Namun, pendekatan-pendekatan tersebut masih memiliki kekurangan, seperti ketergantungan pada algoritma tunggal dan tingginya *false positives*. Hal ini menunjukkan perlunya pendekatan yang lebih menyeluruh, seperti *hybrid ensemble learning*, yang dapat mengatasi keterbatasan tersebut.

Untuk mengatasi kelemahan tersebut, penelitian ini mengusulkan penggunaan *hybrid ensemble learning* yang menggabungkan kekuatan beberapa algoritma pembelajaran mesin, seperti *IForest-K-Means, Random Forest,* dan penerapan *Majority Voting*. Kombinasi algoritma ini diharapkan dapat meningkatkan kinerja model dan memperkuat keputusan deteksi, sehingga

menghasilkan sistem deteksi yang lebih andal dan efisien. Dengan demikian, penelitian ini bertujuan untuk merancang dan mengevaluasi metode *hybrid ensemble learning* yang lebih efektif dalam mendeteksi ancaman siber, khususnya *zero-day attacks* dan APT.

B. LANDASAN TEORI

Dalam penelitian ini, landasan teori akan mencakup pemahaman tentang konsep keamanan siber, deteksi anomali, dan berbagai algoritma yang digunakan dalam sistem deteksi intrusi, seperti *Isolation Forest, K-Means, Random Forest,* dan teknik *ensemble learning*. Selanjutnya, akan mengulas berbagai penelitian terdahulu yang relevan, untuk mengetahui perkembangan metode yang digunakan serta untuk mengidentifikasi kekurangan-kekurangan yang perlu diperbaiki dalam penelitian ini. Berikut adalah dasar-dasar teori yang diterapkan dalam penelitian ini sebagai berikut:

1. Serangan Siber *(Cyber Attack)*

Serangan siber adalah tindakan yang dilakukan dengan sengaja untuk mencuri, mengungkap, mengubah, merusak, atau menonaktifkan data, aplikasi, atau aset lainnya melalui akses yang tidak sah ke sistem komputer, jaringan, atau perangkat digital (IBM, 2024). Para pelaku ancaman ini memiliki berbagai motif, mulai dari pencurian kecil hingga aktivitas yang dapat disamakan dengan perang siber. Mereka memanfaatkan berbagai metode, termasuk penggunaan *malware*, manipulasi sosial, dan pencurian kata sandi, untuk memperoleh akses ilegal ke sistem yang menjadi target.

2. Keamanan Siber *(Cyber Security)*

Keamanan siber mencakup serangkaian teknik, teknologi, dan prosedur yang digunakan untuk melindungi sistem digital, jaringan, dan data dari akses yang tidak sah, serangan, serta kerusakan. Dalam era digital yang saling terhubung saat ini, keamanan siber menjadi sangat vital. Ini mencakup berbagai cara untuk mendeteksi dan menangani ancaman siber, seperti *malware, phishing, ransomware,* dan lain-lain. Tujuan utama dari keamanan siber adalah untuk mengembangkan perangkat yang dilengkapi dengan aturan yang ketat serta membangun berbagai lapisan pertahanan terhadap serangan berbasis internet (Kumar, 2023).

3. Sistem Deteksi Intrusi Jaringan *(Network Intrusion Detection System)*

NIDS adalah alat yang sangat penting untuk melindungi infrastruktur digital di dunia yang semakin terhubung. Fungsi utamanya adalah untuk mendeteksi dan mencegah akses ilegal, aktivitas berbahaya, serta ancaman yang dapat merusak integritas dan kerahasiaan jaringan. Dengan meningkatnya ketergantungan organisasi pada jaringan digital untuk operasi penting dan penyimpanan data sensitif, perlindungan terhadap ancaman siber menjadi semakin krusial (Bhelkar, 2024).

4. Deteksi Anomali

Deteksi anomali adalah komponen penting dalam analisis data yang berfokus pada pengidentifikasian pola yang menyimpang dari perilaku yang diharapkan. Seiring dengan berkembangnya volume data besar dan sistem yang semakin kompleks, metode deteksi anomali konvensional menjadi tidak lagi memadai. Deteksi anomali yang menggunakan kecerdasan buatan memanfaatkan algoritma pembelajaran mesin untuk meningkatkan akurasi dan efisiensi dalam mengidentifikasi data yang tidak biasa (Anglen, 2024).

5. *Ensemble Learning*

Ensemble learning adalah pendekatan dalam pembelajaran mesin di mana beberapa model dilatih untuk menyelesaikan masalah yang sama, dan hasil prediksi dari setiap model digabungkan untuk meningkatkan kinerja secara keseluruhan. Konsep dasar dari *ensemble learning* adalah dengan menggabungkan berbagai model yang memiliki kekuatan dan kelemahan masing-masing. Teknik *ensemble* dapat diterapkan pada berbagai jenis tugas pembelajaran mesin, seperti klasifikasi, regresi, dan klasterisasi. Beberapa metode *ensemble* yang umum digunakan antara lain *bagging, boosting,* dan *stacking* (Simplilearn, 2024).

6. *Hybrid Ensemble Learning*

Hybrid ensemble learning adalah metode yang melibatkan pembuatan dan penggabungan beberapa model secara terkoordinasi, seperti para ahli atau klasifikasi, untuk menyelesaikan masalah tertentu dalam kecerdasan komputasional. Pendekatan ini digunakan untuk meningkatkan kinerja model atau mengurangi kemungkinan pemilihan model yang kurang optimal. *Hybrid ensemble learning* menghasilkan keputusan yang lebih valid dengan menggabungkan beberapa *weak regressors,* sehingga menghasilkan prediksi yang lebih akurat dengan tingkat keandalan yang lebih tinggi (Asad et al., 2023).

7. *Isolation Forest*

Isolation Forest adalah teknik *unsupervised learning* yang digunakan untuk mendeteksi anomali, yang pertama kali diperkenalkan oleh Liu et al. pada tahun 2008. Algoritma ini bersifat nonparametrik dan bekerja berdasarkan prinsip dasar pohon keputusan *(decision trees)*. Proses deteksi anomali dengan menggunakan *Isolation Forest* terdiri dari dua tahap utama, tahap pelatihan dan tahap evaluasi. Pada tahap pelatihan, algoritma membangun pohon isolasi menggunakan sampel data yang ada. Kemudian, pada tahap evaluasi, setiap data point diproses melalui pohon isolasi untuk menghitung skor anomali masing-masing (Zulkifar, Rahmani, & Azizah, 2023). Berikut rumus *Isolation Forest* sebagai berikut:

$$s(x,n) = 2 - \frac{E(h(x))}{c(n)} \quad (1)$$

8. *K-Means*

K-Means adalah metode *unsupervised learning* yang banyak digunakan untuk mengatasi masalah klasterisasi. Pendekatan ini mengelompokkan data ke dalam sejumlah klaster yang telah ditentukan sebelumnya dengan cara menghitung jarak *Euclidean* kuadrat terdekat ke pusat klaster *(centroid)*. Setelah itu, *centroid* diperbarui dengan menghitung rata-rata posisi data dalam klaster tersebut (Saputra & Nataliani, 2021). *K-Means* bertujuan untuk meminimalkan jarak antara data dan *centroid* kelompoknya. Berikut rumus *K-Means* sebagai berikut:

$$d(x,C) = j = \sum_{j=1}^{n}(x_j - c_j)^2 \quad (2)$$

9. *Random Forest*

Random Forest adalah metode pembelajaran mesin terawasi yang diperkenalkan oleh Breiman pada tahun 2001. Metode ini efektif dalam mencegah *overfitting* dan mengurangi *noise,* serta mudah diterapkan dan cepat dalam pelatihan. Karena keunggulannya, *Random Forest* banyak digunakan dalam tugas klasifikasi dan regresi. Dalam *Random Forest,* pengambilan sampel acak diterapkan pada data untuk menghasilkan beberapa sub-set pelatihan dan pengujian. Proses kerjanya melibatkan pemilihan fitur secara acak, pemisahan *node* berdasarkan fitur terbaik, dan pembagian *node* menjadi *child node* untuk membentuk pohon keputusan. Proses ini diulang hingga membentuk sejumlah pohon keputusan dalam hutan (Elmahalwy et al., 2023). Berikut rumus *Random Forest* sebagai berikut:

a. *Random Feature Selection*

$$F_k = k\ feature\ from\ m\ total\ features \quad (3)$$

b. *Best Split Selection*

$$Best\ Split = arg\ max_{split}\left(\frac{Gini\ index}{Entropy}\right) \quad (4)$$

c. *Child Node Creation*

$$Node\ Split \rightarrow Child\ Node \quad (5)$$

d. *Recursion*

$$Repeat\ steps\ 1-3 \quad (6)$$

e. *Forest Creation*

$$RF = \{Tree_1, Tree_2, \dots, Tree_n\} \quad (7)$$

10. *Majority Voting*

Majority voting adalah metode *ensemble* yang menggabungkan hasil prediksi dari beberapa model untuk meningkatkan akurasi. Dalam *hard voting*, hasil akhir adalah kelas dengan jumlah suara terbanyak dari semua model, sedangkan dalam *soft voting*, hasil prediksi ditentukan berdasarkan probabilitas tertinggi yang dihitung dari semua model (Elmahalwy et al., 2023). Berikut rumus *Majority Voting* sebagai berikut:

a. *Hard Voting*

$$y\hat{} = mode\{h1(x), h2(x), \dots, hn(x)\} \quad (8)$$

b. *Soft Voting*

$$y = arg\ max_{c \in C} \frac{1}{n} \sum_{i=1}^{n} P_i(c|x) \quad . \quad (9)$$

11. Evaluasi *Matrix*

Evaluasi model dalam machine learning menggunakan berbagai metrik untuk menilai kinerjanya. Metrik yang sering digunakan antara lain *accuracy, precision, recall,* dan *F1-score*. Metrik seperti *accuracy, precision,* dan *recall,* digunakan untuk menilai proporsi prediksi yang benar, akurasi prediksi positif, dan kemampuan model dalam mendeteksi data positif secara akurat (Elmahalwy et al., 2023). Berikut rumus Evaluasi Matrix sebagai berikut:

a. AUC

$$TPR = \frac{TP}{TP+FN}, FPR = \frac{FP}{FP+TN} \quad (10)$$

b. Akurasi

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (11)$$

c. Presisi

$$Precision = \frac{TP}{TP+FP} \quad (12)$$

d. *Recall*

$$\frac{TP}{TP+FN} \quad (13)$$

e. *F1-Score*

$$F1 - Score = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \quad (14)$$

12. *Flask*

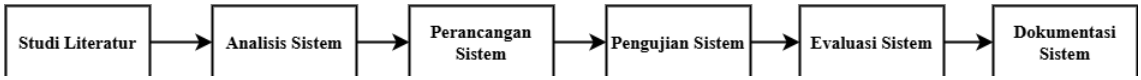
Flask adalah *framework mikro* berbasis *Python* yang dirancang untuk aplikasi yang ringan dan efisien. *Flask* dipilih untuk meningkatkan efisiensi dalam pengembangan aplikasi karena tidak memerlukan banyak sumber daya. Meskipun minimalis, *Flask* tetap mampu menjalankan fungsinya sesuai kebutuhan (Ngantung, 2021).

13. *Python*

Python adalah bahasa pemrograman tingkat tinggi yang pertama kali diperkenalkan pada tahun 1991. *Python* dirancang untuk mendukung berbagai paradigma pemrograman dan memiliki desain yang menekankan keterbacaan kode. *Python* juga memiliki sistem tipe dinamis dan manajemen memori otomatis, yang menjadikannya pilihan yang fleksibel untuk berbagai jenis aplikasi (A, 2021).

C. METODE PENELITIAN

Penelitian ini menggunakan menggunakan pendekatan Dessain Simulasi untuk merancang, mengimplementasikan, dan mengevaluasi sistem deteksi anomali berbasis *Hybrid Ensemble Learning*. Metode yang digunakan adalah *Waterfall,* yang terdiri dari tahapan studi literatur, analisis kebutuhan, perancangan sistem, implementasi, pengujian, evaluasi hasil, dan dokumentasi. Setiap tahapan akan dijelaskan sebagai berikut:



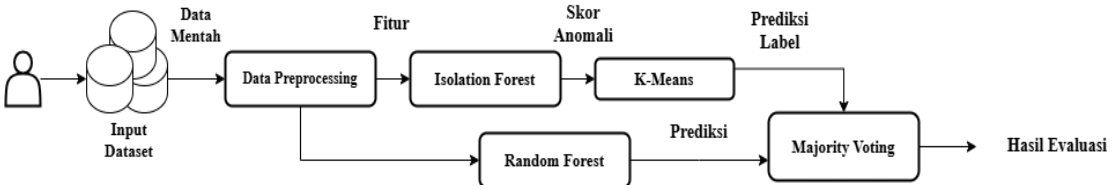
Gambar 1. Prosedur Penelitian

Tahap pertama dilakukan kajian pustaka untuk memahami konsep dasar, algoritma, dan metode yang relevan dengan penelitian ini. Literatur yang dikaji mencakup penelitian sebelumnya tentang sistem deteksi anomali berbasis machine learning, kekuatan dan kelemahan algoritma seperti *Isolation Forest, K-Means,* dan Random Forest, serta penerapan hybrid ensemble learning dengan *Majority Voting*. Selain itu, dataset *CSE-CIC-IDS2018* juga dikaji untuk memahami karakteristiknya serta relevansi dengan penelitian. Tahap kedua, yang bertujuan untuk memastikan sistem memenuhi tujuan penelitian. Kebutuhan sistem diidentifikasi berdasarkan aspek fungsional, seperti kemampuan sistem untuk memproses dataset jaringan dan menghasilkan prediksi deteksi. Selain itu, kebutuhan non-fungsional, seperti efisiensi waktu pemrosesan, keamanan, dan kegunaan sistem, juga diperhatikan. Dataset *CSE-CIC-IDS2018* dikumpulkan sebagai bahan utama penelitian, dan subset data yang relevan dipilih untuk proses pelatihan dan pengujian model.

Pada tahap ketiga, dilakukan perancangan arsitektur sistem deteksi anomali yang akan dibangun. Proses ini melibatkan integrasi algoritma *Isolation Forest, K-Means,* dan *Random Forest,* serta penentuan parameter utama seperti jumlah *cluster* dalam *K-Means* dan tingkat kontaminasi dalam *Isolation Forest*. Teknik *Majority Voting* dirancang sebagai metode pengambilan keputusan untuk menghasilkan prediksi akhir dari *IForest-KMeans*, *Random Forest*. Tahap keempat, di mana sistem dirancang dan diimplementasikan menggunakan bahasa pemrograman Python. Library seperti *Scikit-learn* digunakan untuk mengimplementasikan algoritma *machine learning*, sementara *Pandas* digunakan untuk *preprocessing* data dan manipulasi *dataset*. *Backend* sederhana dibangun menggunakan *Flask* untuk menampilkan hasil deteksi, dan metode penyimpanan serta pemuatan model menggunakan *joblib* diimplementasikan untuk meningkatkan efisiensi. Tahap kelima, yang bertujuan untuk memastikan sistem berfungsi sesuai desain dan menghasilkan performa yang diharapkan. Pengujian dilakukan untuk mengevaluasi kinerja sistem menggunakan metrik seperti akurasi, *precision, recall, F1-score, dan AUC*. *Dataset* uji yang dipilih dari *CSE-CIC-IDS2018* digunakan untuk mengukur efektivitas sistem dalam mendeteksi anomali jaringan.

Tahap keenam, bertujuan untuk menganalisis performa sistem secara keseluruhan. Hasil pengujian dibandingkan antara *hybrid ensemble learning* dan algoritma tunggal untuk mengukur peningkatan akurasi dan pengurangan *false positives*. Visualisasi hasil seperti *charts* and *table* dibuat untuk memudahkan interpretasi data. Efisiensi sistem juga dievaluasi berdasarkan waktu pemrosesan dan konsumsi sumber daya. Tahap terakhir, di mana semua hasil penelitian disusun dalam laporan akhir. Laporan ini mencakup analisis kebutuhan, desain, implementasi, pengujian, dan evaluasi sistem. Visualisasi hasil dan tabel metrik evaluasi disertakan untuk mendukung kesimpulan penelitian. Selain itu, laporan memberikan rekomendasi untuk pengembangan sistem deteksi anomali di masa depan.

P- ISSN: 2355-6684, E-ISSN: 2776-6640
https://ojs.unpkediri.ac.id/index.php/noe



Gambar 2. Kerangka Model

Tahapan awal, *dataset CSE-CIC-IDS2018* diinputkan oleh pengguna. Lalu, analisis mencakup perhitungan rata-rata (μ) dan simpangan baku (σ) untuk setiap fitur, seperti jumlah koneksi dan login gagal. Kemudian, dilakukan standarisasi data menggunakan *Z-Score Normalization,* yang mengubah nilai asli menjadi distribusi dengan *mean* nol dan standar deviasi satu. Standarisasi ini memastikan semua fitur memiliki skala yang sebanding sehingga lebih mudah dianalisis dalam model pembelajaran mesin. Metode *Isolation Forest* digunakan untuk menghitung skor anomali, yaitu angka yang menunjukkan seberapa jauh sebuah data menyimpang dari normal. Model ini bekerja dengan cara membangun pohon isolasi dan mengukur panjang jalur rata-rata suatu titik data dalam pohon tersebut. Semakin pendek jalur yang ditempuh oleh suatu data, semakin besar kemungkinan data tersebut adalah anomali.

Setelah mendapatkan skor anomali dari *Isolation Forest*, teknik *K-Means Clustering* diterapkan untuk mengelompokkan data ke dalam dua kategori, normal dan anomali. Proses ini dimulai dengan pemilihan *centroid* awal dan berlanjut hingga tidak ada lagi perubahan pada *centroid*, menandakan bahwa klaster telah stabil. Langkah selanjutnya adalah melakukan *hard voting* antara model *IForest-K-Means*, dan *Random Forest* untuk memperoleh hasil klasifikasi yang lebih akurat. Model *Random Forest* digunakan untuk klasifikasi, yang merupakan algoritma pembelajaran mesin berbasis pohon keputusan. Selain itu, sistem menghasilkan *AUC*, yang memberikan gambaran lebih jelas tentang performa model dalam mengklasifikasikan data secara benar. Hasil deteksi ini dikemas dalam bentuk visualisasi dan log aktivitas sistem untuk pemantauan serta *debugging*. Pengguna dapat mengunduh laporan hasil deteksi dalam format *CSV* guna melakukan analisis lebih lanjut. Dengan fitur yang disediakan, sistem ini membantu pengguna dalam mendeteksi, mengevaluasi, dan mengambil tindakan tepat terhadap potensi ancaman keamanan jaringan.

D. HASIL DAN PEMBAHASAN

1. Pengolahan Data

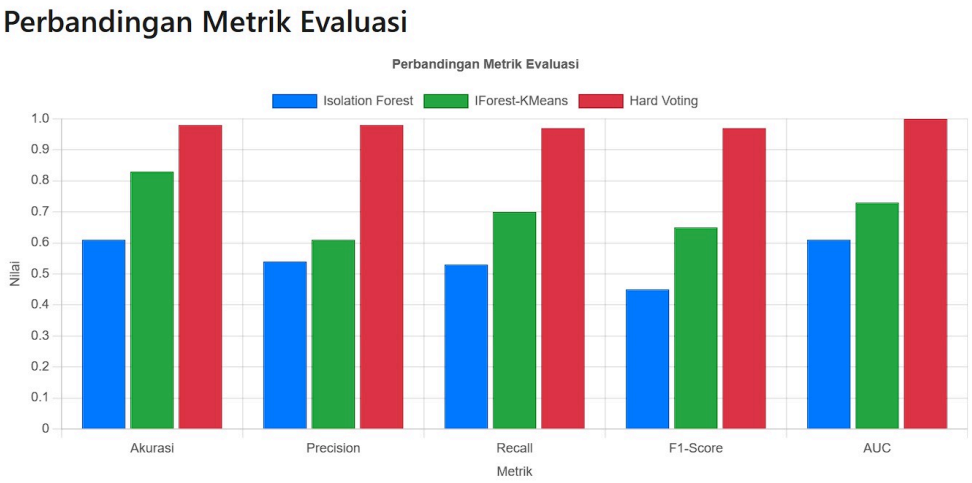
Proses pengolahan data dimulai dengan mempersiapkan subset dataset CSE-CIC-IDS2018 yang terdiri dari 16.815.481 *record* untuk pelatihan model deteksi anomali. Tahapan pertama adalah melakukan encoding pada kolom-kolom kategorikal menggunakan *LabelEncoder*, yang mengubah nilai kategorikal menjadi angka. Selanjutnya, untuk menangani nilai hilang, data yang mengandung NaN atau inf diganti dengan NaN, dan baris yang mengandung nilai tersebut dihapus agar dataset menjadi bersih. Setelah itu, dilakukan normalisasi data menggunakan *StandardScaler* dengan metode *Z-Score*, yang menyelaraskan distribusi fitur sehingga memiliki rata-rata 0 dan deviasi standar 1, sehingga mencegah dominasi fitur dengan skala besar. Di samping itu, fitur baru *FeatureX_Interaction* ditambahkan, yang merupakan hasil penjumlahan seluruh fitur numerik, untuk memberikan informasi tambahan bagi model. Setelah tahapan *preprocessing*, *Isolation Forest* digunakan untuk mendeteksi anomali dalam data, menghasilkan skor yang digunakan untuk klasifikasi lebih lanjut. Kemudian, teknik *K-Means Clustering* diterapkan untuk mengelompokkan data menjadi dua kategori: normal dan anomali, berdasarkan skor dari *Isolation Forest*. Selanjutnya, model *Random Forest* dilatih menggunakan hasil klasifikasi yang diperoleh dari *Isolation Forest* dan *K-Means*, dengan menggunakan label asli jika tersedia, atau label yang dihasilkan oleh *clustering* jika label asli tidak ada. Dataset dibagi menjadi data latih (70%) dan data uji (30%). Proses pelatihan dan pengujian ini dilakukan di perangkat dengan spesifikasi CPU AMD Ryzen 5 6600H dan RAM 16GB untuk mendukung proses komputasi yang cepat dan efisien. Terakhir, model yang sudah terlatih beserta *scaler* dan *encoder* disimpan menggunakan *joblib* untuk digunakan dalam aplikasi berikutnya.

P- ISSN: 2355-6684, E-ISSN: 2776-6640
https://ojs.unpkediri.ac.id/index.php/noe

2. Hasil Evaluasi

Total Data: 16,815,481

Metode	Akurasi	Precision	Recall	F1-Score	AUC
Isolation Forest	0.61	0.54	0.53	0.35	0.61
IForest-KMeans	0.83	0.61	0.70	0.65	0.73
Hard Voting	0.98	0.98	0.97	0.97	1.0



Gambar 3. Hasil Metrik Evaluasi

Gambar 3 menunjukkan hasil evaluasi dan perbandingan kinerja tiga metode deteksi anomali jaringan, yaitu *Isolation Forest, IForest-KMeans,* dan *Hard Voting (Hybrid Ensemble).* Evaluasi dilakukan terhadap dataset berukuran besar yang terdiri dari 16.815.481 data, menggunakan lima metrik utama: akurasi, *precision, recall, F1-score,* dan *AUC*. Berdasarkan tabel dan grafik yang ditampilkan, metode *Isolation Forest* memiliki performa terendah, dengan akurasi sebesar 61%,

precision 54%, *recall* 53%, *F1-score* 35%, dan *AUC* hanya 61%. Hal ini menunjukkan bahwa model ini kurang mampu mengidentifikasi anomali secara akurat dan konsisten. Kinerja yang lebih baik ditunjukkan oleh metode *IForest-KMeans,* dengan peningkatan akurasi menjadi 83% dan *F1-score* sebesar 65%. Kombinasi ini memperbaiki kelemahan dari *Isolation Forest* dengan menambahkan proses *clustering* menggunakan *K-Means*. Namun, metode yang paling unggul ditunjukkan oleh pendekatan *Hard Voting,* yaitu gabungan dari *IForest-KMeans* dan *Random Forest* model melalui teknik *majority voting. Hard Voting* mencapai akurasi sebesar 98%, *precision* 98%, *recall* 97%, *F1-score* 97%, dan *AUC* sempurna sebesar 1,00. Hasil ini menunjukkan bahwa pendekatan *hybrid ensemble* mampu secara signifikan meningkatkan kinerja deteksi anomali, baik dari sisi akurasi maupun kestabilan prediksi. Visualisasi grafik batang turut memperjelas dominasi *Hard Voting* pada semua metrik evaluasi. Dengan demikian, dapat disimpulkan bahwa sistem deteksi anomali berbasis *hybrid ensemble* merupakan solusi yang lebih efektif dan andal dalam menghadapi ancaman siber pada jaringan berskala besar.

E. Kesimpulan dan Saran

Penelitian ini berhasil merancang dan mengevaluasi sistem deteksi anomali jaringan berbasis *hybrid ensemble* yang menggabungkan algoritma *Isolation Forest, K-Means,* dan *Random Forest* melalui teknik *majority voting*. Berdasarkan hasil pengujian terhadap *dataset CSE-CIC-IDS2018* yang berjumlah lebih dari 16 juta data, sistem menunjukkan performa yang sangat baik. Metode *Hard Voting* mencapai akurasi 98%, *precision* 98%, *recall* 97%, dan *AUC* sebesar 1.00, yang berarti model mampu membedakan data normal dan anomali secara hampir sempurna.

Dibandingkan dengan metode tunggal seperti *Isolation Forest* dan *IForest-KMeans,* pendekatan *hybrid* terbukti jauh lebih unggul dalam hal akurasi dan kestabilan deteksi, serta mampu mengurangi tingkat *false positive* secara signifikan. Hal ini menunjukkan bahwa pendekatan *hybrid ensemble* sangat potensial untuk diterapkan dalam sistem keamanan jaringan guna menghadapi berbagai pola serangan siber yang kompleks dan dinamis. Berdasarkan hasil yang telah diperoleh, terdapat beberapa saran yang dapat diberikan untuk pengembangan lebih lanjut. Pertama, sistem deteksi anomali yang dikembangkan masih bersifat simulasi dan belum diimplementasikan secara real-time.

Oleh karena itu, disarankan agar penelitian selanjutnya mengembangkan sistem ini ke dalam bentuk implementasi langsung pada jaringan aktif agar dapat mendeteksi serangan secara cepat dan responsif. Kedua, model *hybrid ensemble* yang digunakan dapat ditingkatkan lagi dengan mengintegrasikan algoritma tambahan seperti *XGBoost, SVM,* atau metode deep learning seperti *LSTM* untuk menangani pola serangan yang lebih kompleks dan dinamis. Ketiga, perlu dilakukan tuning parameter secara lebih optimal dengan pendekatan sistematis seperti *Grid Search* atau *Bayesian Optimization* agar performa model benar-benar maksimal. Keempat, agar model lebih general dan dapat digunakan pada berbagai skenario, sebaiknya dilakukan pengujian terhadap beberapa dataset jaringan lain seperti *NSL-KDD* atau *UNSW-NB15*. Terakhir, sebelum diimplementasikan secara luas, sistem perlu diuji dalam aspek efisiensi sumber daya komputasi seperti waktu pemrosesan, penggunaan *CPU*, dan memori, terutama jika akan digunakan pada lingkungan jaringan berskala besar dan berkecepatan tinggi.

DAFTAR PUSTAKA

Agustina, T., Masrizal, M., & Irmayanti, I. 2024. Performance Analysis of Random Forest Algorithm for Network Anomaly Detection using Feature Selection. *Sinkron*, *8*(2). https://doi.org/10.33395/sinkron.v8i2.13625

Alserhani, F., & Aljared, A. 2023. Evaluating Ensemble Learning Mechanisms for Predicting Advanced Cyber Attacks. *Applied Sciences*, *13*(24), 13310. https://doi.org/10.3390/app132413310

Anglen, J. 2024. AI in Anomaly Detection for Businesses. Retrieved January 14, 2025, from Rapidinnovation.io website: https://www.rapidinnovation.io/post/ai-in-anomaly-detection-for-businesses?form=MG0AV3

Asad, R., Altaf, S., Ahmad, S., Mahmoud, H., Huda, S., & Iqbal, S. 2023. Machine Learning-Based Hybrid Ensemble Model Achieving Precision Education for Online Education Amid the Lockdown Period of COVID-19 Pandemic in Pakistan. *Sustainability*, *15*(6), 5431. https://doi.org/10.3390/su15065431

Dept. of Cyber Security, Shah & Anchor Kutchhi Engineering College Mumbai, India, & Bhelkar, Mr. S. 2024. Network Intrusion Detection System. *Interantional Journal of Scientific Research in Engineering and Management*, *08*(04), 1–5. https://doi.org/10.55041/IJSREM31278

Goswami, M. J. 2024. *AI-Based Anomaly Detection for Real-Time Cybersecurity*.

Kumar, A. Prof. R. 2023. A Review Article on Cyber Security. *Interantional Journal of Scientific Research in Engineering and Management*, *07*(08). https://doi.org/10.55041/IJSREM25428

Mohamed Elmahalwy, A., Mousa, H. M., & Amin, K. M. 2023. New hybrid ensemble method for anomaly detection in data science. *International Journal of Electrical and Computer Engineering (IJECE)*, *13*(3), 3498. https://doi.org/10.11591/ijece.v13i3.pp3498-3508

Sakshi Bakhare & Dr. Sudhir W. Mohod. 2024. Evaluating the Performance and Challenges of Machine Learning Models in Network Anomaly Detection. *International Journal of Scientific Research in Science, Engineering and Technology*, *11*(3), 42–52. https://doi.org/10.32628/IJSRSET5241134

Santoso, N. A., Lutfayza, R., Nughroho, B. I., & Gunawan, G. 2024. Anomaly detection in network security systems using machine learning. *Journal of Intelligent Decision Support System (IDSS)*, *7*(2), 113–120. https://doi.org/10.35335/idss.v7i2.238

Simplilearn. 2021. Ensemble Learning: Boost Accuracy with Multiple Models. Retrieved January 14, 2025, from Simplilearn.com website: https://www.simplilearn.com/ensemble-learning-article?form=MG0AV3