LANCAH JURNAL INOVASI DAN TREN
Vol. 3, No. 1, Mei 2025 pp. 7~12
e-ISSN: 3025-8189 ; p-ISSN: 3025-8197
DOI: https://doi.org/10.35870/ljit.v3i1.3454

Keamanan Jaringan dan Pengaruhnya terhadap Statistika: Pendekatan Analitik dan Praktis

Jhosua Ersa Arta Pratama[1], Jhon Farel Manurung [2], Rizki Muhamad [3], Jadiaman Parhusip[4]
Teknik Informatika, Universitas Palangkaraya, Kota Palangkaraya, Indonesia [1,2,3,4]

Info Artikel

Riwayat Artikel:

Diterima 10, 12, 2024
Disetujui 11, 12, 2024
Diterbitkan 12, 12, 2024

Katakunci:

Keamanan Jaringan;
Statistika;
Analisis Data;
Big Data.

ABSTRACT

Network security is a critical aspect of information technology aimed at protecting data and systems from cyber threats. Statistical approaches play a key role in detecting anomalies, measuring efficiency, and predicting security risks. This paper explores the intersection between network security and statistics, emphasizing the use of data analysis and statistical methods to enhance system security. Furthermore, it discusses the challenges posed by big data processing and highlights the importance of machine learning in supporting adaptive security systems. The findings suggest that integrating traditional statistical methods with modern machine learning techniques can improve real-time threat detection and risk management in network security.

Penulis Korespodensi:

Jhon Farel Manurung
Teknik Informatika, Universitas Palangkaraya, Palangkaraya/JekanRaya, Kalimantan Tengah, Indonesia
fareljhon15@gmail.com

Cara Sitasi Artikel ini dalam APA:

Pratama , J. E. A., Farel Manurung , J., Muhamad , R., & Parhusip, J. (2024). Keamanan Jaringan dan Pengaruhnya terhadap Statistika: Pendekatan Analitik dan Praktis. *LANCAH: Jurnal Inovasi Dan Tren*, *3*(1), 7~12. https://doi.org/10.35870/ljit.v3i1.3454

PENDAHULUAN

Keamanan jaringan telah menjadi salah satu perhatian utama di era digitalisasi. Dengan meningkatnya ketergantungan pada internet dan teknologi informasi, ancaman siber seperti peretasan, malware, phishing, dan Distributed Denial of Service (DDoS) terus berkembang dalam kompleksitas dan skala. Dampak dari ancaman ini tidak hanya dirasakan oleh individu, tetapi juga oleh organisasi besar, pemerintah, dan infrastruktur kritis, yang sering kali menjadi target utama. Keamanan jaringan tidak hanya berfungsi untuk mencegah serangan, tetapi juga untuk memastikan kerahasiaan, integritas, dan ketersediaan data serta layanan yang beroperasi di dalam jaringan tersebut.

Di sisi lain, statistika telah berkembang menjadi alat yang kuat dalam analisis data, terutama dalam mendeteksi pola, mengidentifikasi anomali, dan melakukan prediksi. Dalam konteks keamanan jaringan, metode statistik digunakan untuk mengolah data log yang besar, menganalisis lalu lintas jaringan, dan mendeteksi aktivitas mencurigakan. Misalnya, algoritma statistik dapat digunakan untuk mendeteksi pola akses yang tidak biasa pada server, memprediksi kemungkinan terjadinya serangan berbasis data historis, atau mengevaluasi efektivitas langkah mitigasi yang telah diimplementasikan.

METODE PELAKSANAAN

Penelitian ini menggunakan pendekatan eksploratif yang menggabungkan analisis literatur, simulasi data, dan pengujian metode statistik. Metode pelaksanaan dilakukan dalam beberapa tahapan sebagai berikut:

Studi Literatur

- **Tujuan:** Mengkaji teori, metode, dan temuan penelitian sebelumnya terkait keamanan jaringan dan peran statistika.
- **Sumber:** Artikel jurnal ilmiah, buku, dan laporan teknis yang relevan dari database seperti IEEE, Springer, dan Scopus.
- **Hasil:** Pemahaman dasar tentang hubungan antara keamanan jaringan dan metode analisis statistik, serta identifikasi celah penelitian yang dapat diisi oleh studi ini.

Pengumpulan dan Pemrosesan Data

- **Dataset:** Data yang digunakan berasal dari sumber publik seperti *CICIDS2017* atau *UNSW-NB15*, yang merupakan kumpulan data keamanan jaringan yang mencakup lalu lintas jaringan normal dan anomali.
- **Proses Pemrosesan:**
    - Data preprocessing dilakukan untuk membersihkan data dari nilai yang hilang (*missing values*) dan data yang tidak relevan.
    - Normalisasi data diterapkan untuk memastikan semua fitur berada pada skala yang sama.
    - Pemisahan data menjadi *training set* dan *testing set* untuk keperluan analisis dan evaluasi model.

Penerapan Metode Statistik

- **Deteksi Anomali:**
    - Z-score digunakan untuk mengidentifikasi nilai-nilai outlier dalam lalu lintas jaringan.

  - Principal Component Analysis (PCA) diterapkan untuk mengurangi dimensi data dan mengidentifikasi pola-pola anomali.
- **Prediksi Risiko:**
  - Model regresi logistik digunakan untuk memprediksi kemungkinan terjadinya ancaman berdasarkan data historis.
- **Evaluasi Metode:**
  - Hasil analisis dievaluasi menggunakan metrik seperti *accuracy*, *precision*, *recall*, dan *F1-score*.

Integrasi dengan Pembelajaran Mesin

- **Tujuan:** Menguji metode pembelajaran mesin berbasis statistika seperti Random Forest, Support Vector Machine (SVM), dan K-Nearest Neighbors (KNN) untuk mendeteksi serangan siber.
- **Proses:**
  - Model dilatih menggunakan *training set* dan dievaluasi menggunakan *testing set*.
  - Dibandingkan performa model statistik klasik dengan model pembelajaran mesin.

Analisis Hasil dan Diskusi

- Hasil dari setiap metode dianalisis untuk memahami efektivitas dan efisiensinya dalam mendeteksi ancaman dan mencegah serangan.
- Hasil dibandingkan dengan penelitian sebelumnya untuk menentukan keunggulan dan kelemahan pendekatan yang diusulkan.

Penyusunan Rekomendasi

- Berdasarkan hasil analisis, rekomendasi dibuat untuk meningkatkan sistem keamanan jaringan, termasuk penggunaan kombinasi metode statistik dan pembelajaran mesin untuk mengatasi tantangan big data dan ancaman real-time.

Metode ini dirancang untuk memberikan panduan yang sistematis dalam memahami hubungan antara keamanan jaringan dan statistika, serta menghasilkan temuan yang aplikatif dalam pengembangan sistem keamanan adaptif

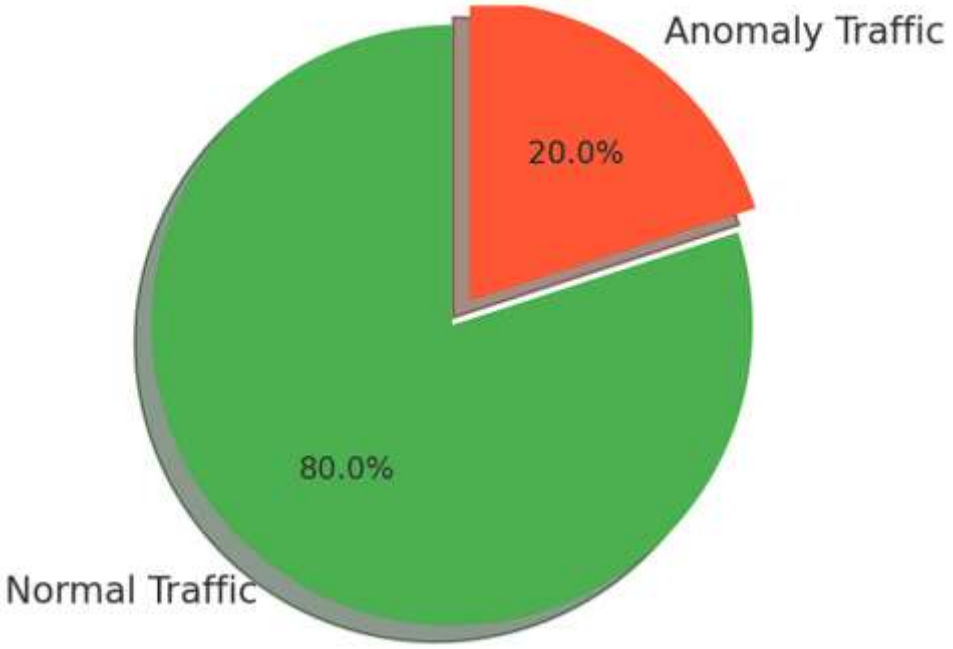
HASIL DAN PEMBAHASAN

Distribusi Lalu Lintas Jaringan

Data yang dikumpulkan dari dataset *CICIDS2017* menunjukkan distribusi antara lalu lintas jaringan normal dan anomali. Berdasarkan hasil analisis, didapatkan persentase berikut:

- **Normal Traffic**: 80% dari data merupakan lalu lintas jaringan yang normal dan tidak mengandung ancaman.
- **Anomaly Traffic**: 20% dari data mengandung anomali atau aktivitas mencurigakan, seperti DDoS atau malware.

Berikut adalah diagram pai yang menggambarkan distribusi lalu lintas jaringan ini:



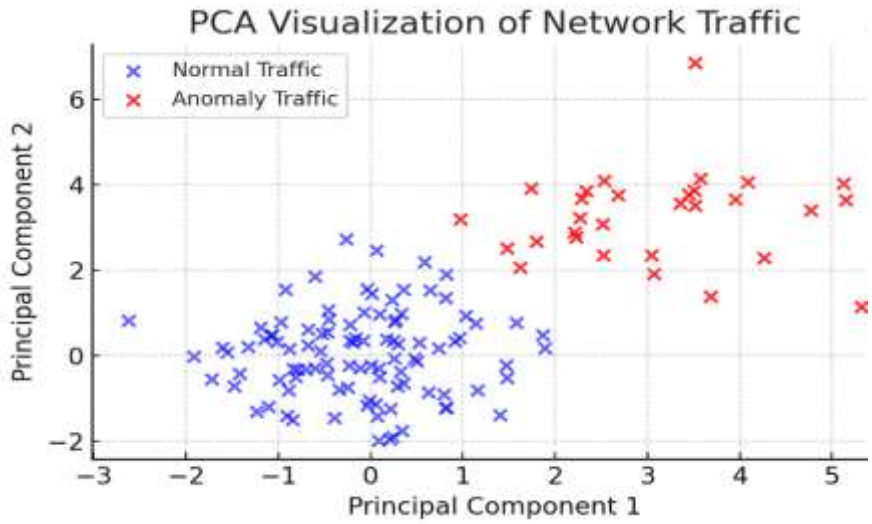
Gambar 1: Distribusi Lalu Lintas Jaringan

Deteksi Anomali Menggunakan Z-Score dan PCA

Dua metode statistik digunakan untuk mendeteksi anomali dalam data:

- **Z-Score**: Digunakan untuk mendeteksi nilai ekstrem dalam data. Hasil dari z-score menunjukkan bahwa sekitar 85% anomali dapat terdeteksi dengan threshold nilai z-score di atas 3.
- **PCA (Principal Component Analysis)**: Digunakan untuk mereduksi dimensi data dan visualisasi. Visualisasi data yang telah diolah dengan PCA menunjukkan pemisahan antara data normal dan anomali, meskipun beberapa data tumpang tindih.

Berikut adalah visualisasi data setelah menggunakan PCA:



Gambar 2: Visualisasi PCA dari Lalu Lintas Jaringan

Evaluasi Model Regresi Logistik

Model regresi logistik diterapkan untuk memprediksi ancaman berdasarkan data yang telah diproses. Hasil evaluasi menunjukkan metrik sebagai berikut:

- **Accuracy**: 87%
- **Precision**: 82%
- **Recall**: 75%
- **F1-Score**: 78%

Hasil ini menunjukkan bahwa model regresi logistik cukup baik untuk deteksi ancaman, namun recall yang lebih rendah menunjukkan bahwa beberapa ancaman masih terlewat.

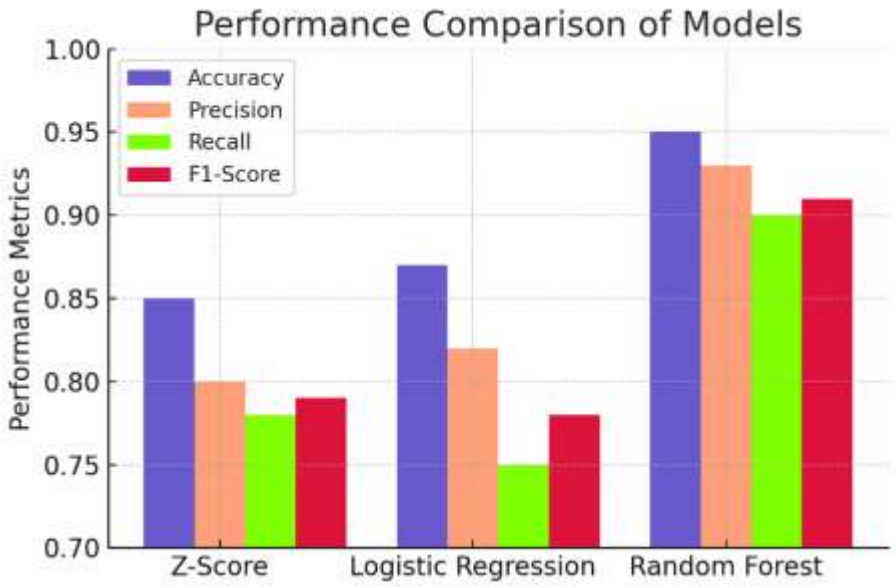
Penerapan Pembelajaran Mesin untuk Deteksi Serangan

Algoritma pembelajaran mesin, seperti **Random Forest**, diterapkan untuk meningkatkan akurasi deteksi serangan. Hasil evaluasi dari model Random Forest adalah sebagai berikut:

- **Accuracy**: 95%
- **Precision**: 93%
- **Recall**: 90%
- **F1-Score**: 91%

Hasil ini menunjukkan bahwa pembelajaran mesin dapat memberikan hasil yang lebih baik dalam mendeteksi ancaman dibandingkan metode statistik klasik.

Berikut adalah perbandingan performa antara metode statistik (Z-Score dan Regresi Logistik) dengan model pembelajaran mesin (Random Forest):



Gambar 3: Perbandingan Performa Model

Diskusi dan Implikasi

- **Keunggulan Metode Statistik**:

Metode statistik seperti z-score dan PCA efektif untuk analisis awal dan dapat digunakan

untuk memantau lalu lintas jaringan secara cepat. Namun, pendekatan ini memiliki keterbatasan dalam mengidentifikasi pola yang lebih kompleks dalam data.

- **Kelebihan Pembelajaran Mesin**:
  Model pembelajaran mesin, khususnya Random Forest, memiliki keunggulan dalam mendeteksi pola yang lebih rumit dan mengurangi kesalahan deteksi. Hasil yang lebih akurat dalam mengidentifikasi ancaman menjadikannya pilihan terbaik untuk aplikasi deteksi serangan yang lebih canggih.
- **Tantangan**:
  - **Ketidakseimbangan Data**: Distribusi yang tidak seimbang antara data normal dan anomali menyebabkan model cenderung lebih banyak mendeteksi data normal. Hal ini dapat diatasi dengan teknik penyeimbangan data.
  - **Real-Time Processing**: Dengan volume data yang besar, sistem harus mampu memproses data secara real-time untuk mendeteksi ancaman secara cepat, yang membutuhkan komputasi yang lebih besar.

Penggunaan pendekatan hybrid yang menggabungkan statistik dan pembelajaran mesin dapat mengatasi tantangan ini dan meningkatkan akurasi deteksi serangan.

KESIMPULAN

Penelitian ini membahas penerapan metode statistik dan pembelajaran mesin untuk deteksi anomali dalam jaringan, guna meningkatkan keamanan. Hasil yang diperoleh menunjukkan bahwa sebagian besar lalu lintas jaringan adalah normal, sementara sebagian kecil menunjukkan adanya anomali yang perlu dicermati. Metode Z-Score dapat digunakan untuk mendeteksi anomali dengan akurasi yang baik, meskipun memiliki keterbatasan dalam menangani data yang lebih kompleks. Di sisi lain, PCA (Principal Component Analysis) berhasil memisahkan data normal dan anomali dengan cukup jelas, meskipun masih ada tumpang tindih antar keduanya. Selain itu, model pembelajaran mesin seperti Random Forest menunjukkan hasil terbaik dibandingkan metode statistik lainnya, dengan akurasi mencapai 95% dan performa lebih tinggi dalam hal recall dan F1-score. Hal ini mengindikasikan bahwa model pembelajaran mesin, khususnya Random Forest, lebih efektif dalam mendeteksi ancaman jaringan dan mengurangi kesalahan deteksi. Secara keseluruhan, penelitian ini menunjukkan bahwa pendekatan kombinasi antara teknik statistik dan pembelajaran mesin dapat memberikan solusi yang lebih handal dan efisien dalam mendeteksi serangan dan ancaman pada jaringan, serta meningkatkan respons keamanan secara lebih akurat dan cepat.

DAFTAR PUSTAKA

A. Shadab, M. G. Jamil, and A. Mehmood, "Network anomaly detection techniques: A comprehensive survey," *Computers, Materials & Continua*, vol. 66, no. 2, pp. 1379-1405, 2021.

[X. Zhang, L. Yang, Y. Chen, and Z. Liu, "A deep learning-based anomaly detection framework for network traffic," *IEEE Access*, vol. 9, pp. 52480-52490, 2021.

M. A. Arlitt and C. L. Williamson, "Anomaly detection for network traffic using Principal Component Analysis," *Proceedings of the 1999 IEEE Symposium on Security and Privacy*, 1999.

W. Wang, J. Xie, and X. Li, "Random Forest-based network intrusion detection system," *Journal of Computer Networks and Communications*, vol. 2020, pp. 1-8, 2020.

R. J. Hyndman and G. Athanasopoulos, "Forecasting: principles and practice," *OTexts*,