**Journal of Science and Social Research** ISSN 2615 – 4307 (Print)
August 2025, VIII (3): 3482 – 3490 ISSN 2615 – 3262 (Online)
Available online at http://jurnal.goretanpena.com/index.php/JSSR

# SELEKSI FITUR MENGGUNAKAN MUTUAL INFORMATION UNTUK DETEKSI INTRUSI

**Riki Andri Yusda[1], Sahren[2], Mustika Fitri Larasti Sibuea[3], Nadira Meutia Arifin[4], Bima Aditya[5]**
**Universitas Royal, Kisaran**
email: [1]rikiandriyusda@gmail.com, [2]sahren.one@gmail.com, [3]bukmus.inaction@gmail.com

***Abstract:*** *This study explores the use of Mutual Information (MI) for feature selection in intrusion detection, focusing on the CICIDS 2017 dataset. Given the complexity and large volume of data in intrusion detection systems, this research aims to identify the most informative features. The methodology includes data preprocessing, MI calculation, and feature selection based on the highest MI values. The analysis results indicate that using MI contributes to improving model accuracy and reducing the false positive rate. These findings underscore the importance of feature selection in enhancing the effectiveness of intrusion detection systems and provide significant contributions to developing more efficient cybersecurity strategies.*

***Keyword:*** *IDS, Mutual Information, CICIDS2017, Feature selection*

**Abstrak:** Penelitian ini mengeksplorasi penggunaan Mutual Information (MI) untuk seleksi fitur dalam deteksi intrusi, dengan fokus pada dataset CICIDS 2017. Mengingat kompleksitas dan volume data yang besar dalam sistem deteksi intrusi, penelitian ini bertujuan untuk mengidentifikasi fitur-fitur yang paling informatif. Metodologi yang diterapkan mencakup preprocessing data, perhitungan MI, dan seleksi fitur berdasarkan nilai MI tertinggi. Hasil analisis menunjukkan bahwa penggunaan MI berkontribusi pada peningkatan akurasi model serta pengurangan tingkat false positive. Temuan ini menegaskan pentingnya seleksi fitur dalam meningkatkan efektivitas sistem deteksi intrusi dan memberikan kontribusi signifikan dalam pengembangan strategi keamanan siber yang lebih efisien.

**Kata kunci:** IDS, Mutual Information, CICIDS2017, Seleksi fitur

## PENDAHULUAN

Di era digital yang semakin maju, keamanan informasi telah menjadi salah satu prioritas utama bagi organisasi, pemerintah, dan individu. Dengan meningkatnya ketergantungan pada teknologi informasi, ancaman siber juga semakin kompleks dan beragam(Yang & Peng, 2025). Serangan seperti malware, phishing, dan denial-of-service (DoS) dapat menyebabkan kerugian yang signifikan baik secara finansial maupun reputasi. Oleh karena itu, sistem deteksi intrusi (IDS) menjadi sangat penting untuk melindungi infrastruktur dan data dari ancaman yang tidak diinginkan (Ahmad et al., 2021)(Latif et al., 2024).

Deteksi intrusi berperan penting dalam mengidentifikasi serta melaporkan aktivitas yang mencurigakan di dalam jaringan komputer (Prazeres et al., 2023). Namun, proses ini menghadapi tantangan besar akibat besarnya volume dan kompleksitas data yang dihasilkan oleh berbagai perangkat serta aplikasi. Data tersebut sering kali memuat banyak atribut, di antaranya terdapat fitur yang tidak relevan atau bersifat redundan (berulang), yang dapat menurunkan efektivitas sistem deteksi. Oleh karena itu, proses seleksi fitur yang tepat

menjadi sangat esensial untuk meningkatkan performa sistem IDS (Sahu et al., 2024).

Seleksi fitur merupakan teknik untuk memilih atribut yang paling signifikan terhadap permasalahan yang sedang dianalisis dari sekumpulan besar fitur yang tersedia. Pendekatan ini tidak hanya berguna dalam mengurangi dimensi data, tetapi juga berperan dalam meningkatkan akurasi model serta efisiensi waktu pemrosesan (Kushwaha et al., 2023)(Ogwara et al., 2022). Salah satu metode seleksi fitur yang umum digunakan adalah Mutual Information (MI), yang mengukur tingkat ketergantungan antara dua variabel, sehingga mampu mengidentifikasi fitur yang paling relevan terhadap label target (Zhou et al., 2024).

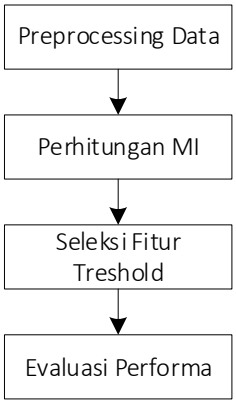
Dalam kerangka deteksi intrusi, penerapan MI dapat membantu mengidentifikasi atribut yang memiliki nilai informasi tinggi dan relevan terhadap klasifikasi serangan (Al-E'mari et al., 2024)(Saq et al., 2024)(Afolabi & Akinola, 2024). Penelitian ini memanfaatkan dataset CICIDS 2017, yang dikenal sebagai salah satu dataset paling lengkap dalam studi deteksi intrusi karena mencakup berbagai jenis serangan dan trafik jaringan normal (Sharafaldin et al., 2018)(Kurniabudi et al., 2020)(Barkah et al., 2023)(Oyelakin et al., 2024). Dengan menerapkan metode MI pada dataset ini, dimungkinkan untuk mengurangi gangguan data (noise) dan menekankan pada fitur-fitur yang secara signifikan berkontribusi terhadap pendeteksian ancaman. Pendekatan ini diharapkan dapat meningkatkan akurasi deteksi serta menurunkan tingkat kesalahan positif (false positive).

Penelitian ini bertujuan untuk mengevaluasi dan menganalisis efektivitas MI dalam proses seleksi fitur untuk sistem deteksi intrusi, dengan memanfaatkan dataset CICIDS 2017. Fokus utama dari penelitian ini mencakup pemilihan dataset, penerapan algoritma MI, serta evaluasi performa sistem deteksi setelah proses seleksi fitur dilakukan. Selain itu, penelitian ini juga membahas dampak penggunaan MI terhadap pengembangan IDS yang lebih akurat dan efisien. Diharapkan hasil dari studi ini dapat memberikan kontribusi yang berarti dalam bidang keamanan siber serta menjadi referensi untuk pengembangan sistem deteksi yang lebih canggih dan andal.

Dengan pemahaman yang lebih mendalam mengenai penerapan MI dalam konteks seleksi fitur, diharapkan penelitian ini dapat membuka jalan bagi peningkatan sistem keamanan dan proteksi data dalam ekosistem digital yang semakin terkoneksi.

## METODE

Proses penelitian dalam konteks seleksi fitur menggunakan Mutual Information (MI) untuk deteksi intrusi melibatkan beberapa langkah sistematis, sebagai berikut.




**Gambar 1 Tahapan Proses Penelitian**

### Preprocessing Data

Preprocessing dilakukan untuk mempersiapkan data agar siap digunakan dalam proses seleksi fitur. Dataset CICIDS 2017 dibersihkan dari nilai hilang, duplikasi, dan outlier, kemudian fitur numerik dinormalisasi agar berada pada skala yang seragam. Fitur kategorikal dikonversi ke format numerik melalui teknik encoding, sedangkan label target disederhanakan menjadi dua kelas

yaitu normal dan serangan. Langkah ini memastikan bahwa perhitungan Mutual Information dapat berjalan optimal dalam mengidentifikasi fitur yang paling relevan untuk deteksi intrusi.

**Perhitungan MI**

Perhitungan MI dilakukan untuk mengukur tingkat ketergantungan antara setiap fitur dengan label target, sehingga dapat mengidentifikasi fitur-fitur yang paling relevan terhadap klasifikasi intrusi. MI memberikan nilai yang mencerminkan seberapa besar informasi yang dimiliki suatu fitur tentang kelas target, tanpa mengasumsikan hubungan linier. Hasil perhitungan berupa skor MI untuk masing-masing fitur, yang selanjutnya digunakan sebagai dasar dalam proses seleksi fitur dengan memilih fitur-fitur yang memiliki nilai informasi tertinggi terhadap label klasifikasi.

**Seleksi Fitur Treshold**

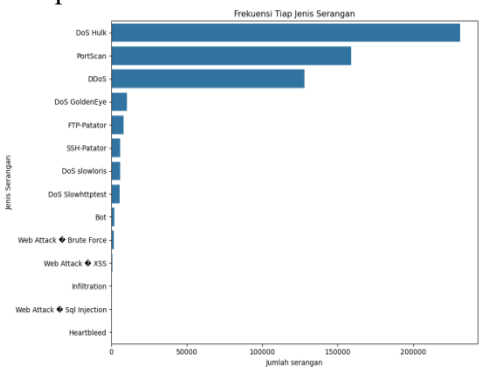
Setelah nilai MI untuk setiap fitur diperoleh, proses seleksi fitur dilakukan dengan dua pendekatan umum, yaitu berdasarkan threshold. Pada pendekatan threshold, fitur yang memiliki nilai MI di atas ambang batas tertentu dipertahankan, sementara fitur dengan nilai di bawah ambang tersebut dieliminasi karena dianggap kurang informatif. Pendekatan ini bertujuan untuk menyaring fitur yang paling relevan terhadap klasifikasi, sehingga dapat mengurangi kompleksitas model, mempercepat waktu pemrosesan, serta meningkatkan akurasi dan generalisasi sistem deteksi intrusi.

**Evaluasi Performa**

Evaluasi performa dilakukan untuk mengukur efektivitas sistem deteksi intrusi setelah proses seleksi fitur menggunakan MI. Kinerja model dievaluasi menggunakan metrik umum dalam deteksi intrusi, yaitu akurasi, presisi, recall, F1-score, dan false positive rate. Perbandingan hasil evaluasi sebelum dan sesudah seleksi fitur dilakukan untuk menilai dampak positif dari pengurangan dimensi terhadap kualitas deteksi. Evaluasi ini memberikan gambaran sejauh mana fitur-fitur terpilih mampu meningkatkan kemampuan sistem dalam membedakan antara trafik normal dan serangan secara efektif.

## HASIL DAN PEMBAHASAN

Dataset CICIDS 2017 merupakan dataset yang banyak digunakan pada penelitian sistem deteksi intrusi dan pengembangan algoritma deteksi anomali di jaringan. Dataset ini memiliki sekitar 2.8 juta data entri dan 79 fitur. Fitur-fitur ini mencakup informasi mengenai koneksi jaringan seperti IP sumber dan tujuan, port jaringan, jenis protokol, durasi koneksi, jumlah paket dikirim dan diterima, panjang paket, hingga kecepatan aliran data.

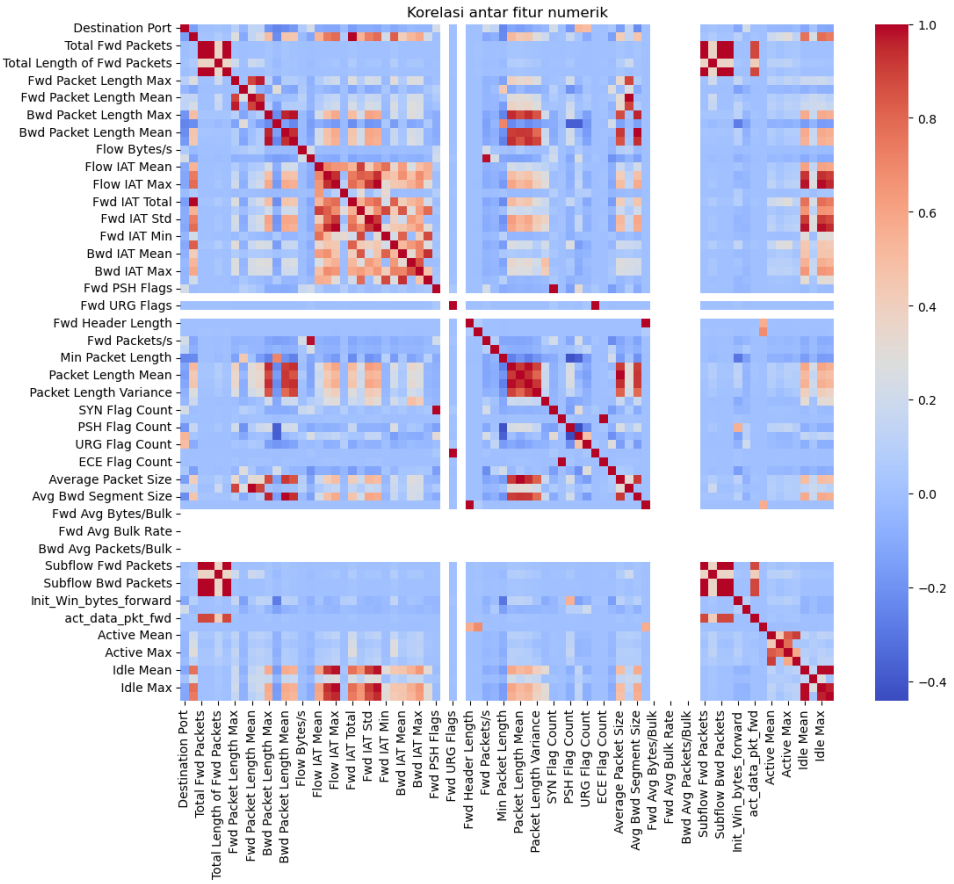



**Gambar 2 Frekuensi Jenis Serangan**

Pada gambar 2 dapat diketahui pola serangan yang sering terjadi, yang dapat membantu untuk merancang langkah-langkah untuk mengalokasikan sumber daya dan perhatian secara tepat berdasarkan prevelensi masing-masing serangan. Ini dibutuhkan untuk perlindungan yang lebih efektif terhadap berbagai jenis ancaman serangan siber. Jenis-jenis serangan dipengaruhi oleh distribusi fitur numerik yang ada pada dataset.

Hubungan antar fitur numerik terhadap jenis-jenis serangan yang terjadi pada dataset CICIDS 2017 dapat dilihat pada gambar 3 berikut.

**Journal of Science and Social Research** ISSN 2615 – 4307 (Print)
August 2025, VIII (3): 3482 – 3490 ISSN 2615 – 3262 (Online)
Available online at http://jurnal.goretanpena.com/index.php/JSSR




**Gambar 3 Korelasi antar fitur pada Dataset CICIDS 2017**

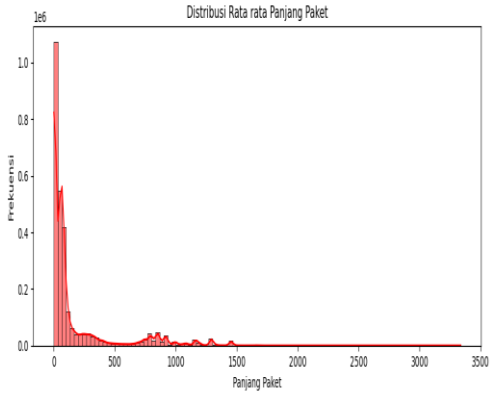
Visualisasi pada gambar memperlihatkan hubungan korelasi antar fitur numerik, nilai korelasi berkisar antara -1 (negatif sempurna) hingga 1 (positif sempurna). Warna merah tua merupakan tanda untuk keterkaitan positif yang kuat, sementara biru tua menunjukkan hubungan negatif yang kuat. Fitur seperti "Average Packet Size" dan "Packet Length Mean" yang memiliki korelasi kuat dan positif dengan serangan jenis DDoS atau DoS, menunjukkan bahwa serangan ini cenderung memanfaatkan paket besar dan trafik yang tinggi. Sebaliknya, fitur yang menunjukkan nilai rendah atau bahkan korelasi negatif bisa menunjukkan bahwa jenis serangan tertentu seperti port scanning atau serangan web mungkin memiliki karakteristik yang berbeda, seperti lalu lintas yang lebih kecil dan tersebar.

Dengan menganalisis pola ini, dapat disimpulkan bahwa beberapa fitur secara kolektif mencerminkan karakteristik dari jenis serangan tertentu. Misalnya, serangan DDoS dan DoS biasanya menunjukkan tingginya nilai fitur seperti "Average Packet Size" dan "Total Length of Bwd Packets", yang menunjukkan trafik besar dan paket panjang. Sementara fitur lain yang berkorelasi rendah atau negatif mungkin berkaitan dengan serangan yang lebih kecil atau berbeda polanya, seperti port scan atau serangan web.

Hubungan antara distribusi ukuran paket dan pola serangan sangat penting dalam analisis keamanan siber. Gambar 4 menunjukkan distribusi rata-rata panjang paket dalam trafik jaringan dapat memberikan wawasan awal mengenai jenis aktivitas yang sedang berlangsung, termasuk potensi adanya serangan. Umumnya, paket-paket yang berukuran pendek mendominasi lalu lintas karena komunikasi normal seperti permintaan HTTP, DNS, atau sinyal kontrol TCP

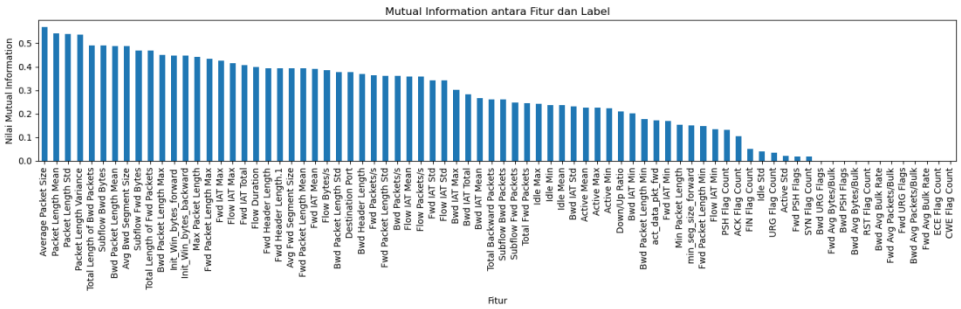
**Journal of Science and Social Research** ISSN 2615 – 4307 (Print)
August 2025, VIII (3): 3482 – 3490 ISSN 2615 – 3262 (Online)
Available online at http://jurnal.goretanpena.com/index.php/JSSR

cenderung menggunakan paket kecil. Namun, jika distribusi menunjukkan pola yang tidak biasa, misalnya lonjakan ekstrem pada ukuran tertentu atau munculnya paket-paket dengan ukuran sangat besar atau sangat kecil yang mana hal ini bisa menjadi tanda adanya aktivitas berbahaya.




**Gambar 4 Distribusi Rerata Panjang Paket**

Beberapa tipe serangan memiliki pola panjang paket yang khas. Contohnya, serangan DoS atau DDoS biasanya melibatkan pengiriman paket-paket pendek secara massal dalam waktu singkat guna membanjiri target. Pola ini dapat terlihat dari tingginya frekuensi paket kecil dalam rentang ukuran tertentu. Serangan brute-force atau port scanning juga menunjukkan pola serupa dengan banyak paket kecil yang dikirim ke berbagai port. Sebaliknya, serangan yang bertujuan pengambilan data atau malware biasanya melibatkan paket-paket besar yang menyimpang dari distribusi normal karena mengandung muatan yang tidak biasa. Dengan demikian, pola distribusi panjang paket tidak hanya mencerminkan karakteristik umum lalu lintas jaringan, tetapi juga dapat menjadi indikator penting dalam mendeteksi dan mengklasifikasi aktivitas serangan yang sedang berlangsung.




**Gambar 5 Nilai MI antara Fitur dan Label**

Dari grafik pada gambar 5, dapat terlihat bahwa beberapa fitur memiliki nilai MI yang tinggi. Ini menunjukkan bahwa fitur-fitur tersebut sangat berpengaruh untuk membantu dalam membedakan hubungan antara kategori dan label. Sebaliknya, nilai fitur yang lebih rendah menunjukkan bahwa fitur-fitur tersebut kurang relevan dan dapat dipertimbangkan untuk dihapus. Perhitungan MI dalam konteks analisis data merupakan langkah yang penting untuk memahami seberapa baik fitur dapat memprediksi label atau variabel target.

MI bekerja dengan mengukur seberapa besar informasi yang diberikan oleh suatu variabel (fitur) terhadap variabel target (label), yang secara matematis dapat dituliskan:

$$MI(X;Y) = \sum_{x \in X} \sum_{y \in Y} P(x,y).\log(\frac{P(x,y)}{P(x)P(y)})$$

Di mana:

$X$ = Fitur

$Y$ = Label (misalnya 0=normal, 1=serangan)

$P(x,y)$ = Probabilitas bersama

$P(x)P(y)$ = Probabilitas marginal

Berikut contoh perhitungan manual pada MI pada Dataset CICIDS 2017 dengan fitur yang diambil adalah Protocol Type. Nilai yang diambil hanya

**Journal of Science and Social Research** ISSN 2615 – 4307 (Print)
August 2025, VIII (3): 3482 – 3490 ISSN 2615 – 3262 (Online)
Available online at http://jurnal.goretanpena.com/index.php/JSSR

dua nilai yaitu TCP dan UDP, dengan Label target yaitu Normal (0) dan Attack (1). Misalkan dari 100 sampel, distribusinya:

**Tabel 1 Contoh Data CICIDS 2017**

| Protocol | Label (0 = Normal) | Label (1 = Attack) | Total |
|---|---|---|---|
| TCP | 30 | 40 | 70 |
| UDP | 20 | 10 | 30 |
| Total | 50 | 50 | 100 |

Langkah pertama yang dilakukan adalah menghitung probabilitas, yaitu:

- P(TCP) = 70/100 = 0.7
- P(UDP) = 30/100 = 0.3
- P(0) = 50/100 = 0.5, P(1) = 50/100 = 0.5
- P(TCP, 0) = 30/100 = 0.3, P(TCP, 1) = 40/100 = 0.4
- P(UDP, 0) = 20/100 = 0.2, P(UDP, 1) = 10/100 = 0.1

Langkah selanjutnya menghitung MI.

$$MI = \sum_{x \in \{TCP,UDP\}} \sum_{y \in \{0,1\}} P(x,y).\log_2\left(\frac{P(x,y)}{P(x)P(y)}\right)$$

Hitung tiap komponen:

$P(TCP,0) = 0.3, P(TCP) * P(0)$

- $= 0.7 * 0.5 = 0.35 * 0.3 * \log_2\left(\frac{0.3}{0.35}\right)$

$= 0.3 * \log_2(0.857) \approx 0.3 * (-0.222) = -0.0666$

$P(TCP,1) = 0.4, P(TCP) * P(1)$

- $= 0.7 * 0.5 = 0.35 * 0.4 * \log_2\left(\frac{0.4}{0.35}\right)$

$= 0.4 * \log_2(1.143) \approx 0.4 * 0.192 = 0.0768$

$P(UDP,0) = 0.2, P(UDP) * P(0)$

- $= 0.3 * 0.5 = 0.15 * 0.2 * \log_2\left(\frac{0.2}{0.15}\right)$

$= 0.2 * \log_2(1.333) \approx 0.2 * 0.415 = 0.083$

$P(UDP,1) = 0.1, P(UDP) * P(1)$

- $= 0.3 * 0.5 = 0.15 * 0.1 * \log_2\left(\frac{0.1}{0.15}\right)$

$= 0.2 * \log_2(0.667) \approx 0.1 * (-0.585) = -0.0585$

Jumlahkan seluruh hasil, maka:

$$MI \approx -0.0666 + 0.0768 + 0.083 - 0.0585 = 0.0347$$

Dari hasil perhitungan, didapatkan nilai MI antara fitur Protocol dan label adalah ~0.035 yang mana nilai ini relatif kecil, berarti fitur ini hanya memberikan sedikit informasi terhadap label.

Untuk menghitung seluruh nilai fitur dengan MI menggunakan kode python seperti berikut ini.

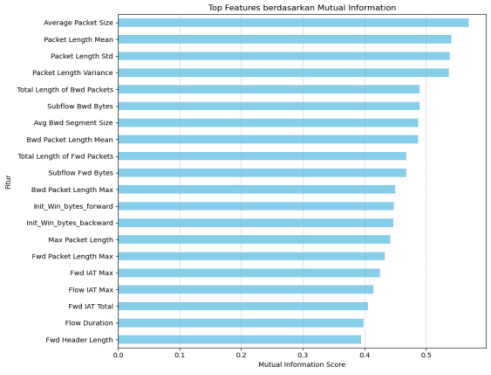
```
mi_scores = mutual_info_classif(X, y)
mi_series = pd.Series(mi_scores, index=X.columns).sort_values(ascending=False)
```

Kode ini digunakan untuk menghitung nilai MI antara fitur dan label dalam suatu dataset. Pertama, fungsi mutual_info_classif(X, y) dipanggil untuk menghitung nilai MI dari setiap fitur di X terhadap label yang ada di y.

Fungsi ini mengembalikan array yang berisi nilai MI, yang menunjukkan seberapa banyak informasi yang dibawa oleh setiap fitur tentang label.

Selanjutnya, nilai MI tersebut diubah menjadi sebuah Series menggunakan pd.Series, dengan indeks yang diatur ke nama kolom dari X. Ini memudahkan untuk mengaitkan setiap nilai MI dengan fitur yang sesuai. Akhirnya, metode .sort_values(ascending=False) digunakan untuk mengurutkan Series berdasarkan nilai MI dalam urutan menurun.

Dengan demikian, fitur-fitur yang paling informatif muncul di atas, membantu dalam pemilihan fitur yang relevan untuk model klasifikasi.




**Gambar 6 Fitur Teratas Berdasarkan Mutual Information**

Gambar 6 tersebut menunjukkan grafik batang yang menyajikan fitur-fitur teratas berdasarkan nilai MI dengan label. Nilai MI ini mengukur seberapa banyak informasi yang diberikan oleh setiap fitur dalam memprediksi label tertentu. Dari grafik, fitur-fitur seperti "Average Packet Size," "Packet Length Mean," dan "Packet Length Std" memiliki nilai MI yang tinggi. Ini menunjukkan bahwa ukuran paket rata-rata dan panjang paket sangat berpengaruh dalam membedakan antara kategori label, mungkin terkait dengan aktivitas jaringan atau jenis serangan yang terjadi.

Fitur-fitur lain seperti "Total Length of Bwd Packets" dan "Max Packet Length" juga mencatat nilai yang signifikan, menunjukkan bahwa karakteristik dari paket yang dikirim memiliki relevansi yang kuat untuk analisis. Hal ini dapat mengindikasikan bahwa banyak serangan, seperti DDoS, mungkin melibatkan pola tertentu dalam ukuran dan jumlah paket. Dengan informasi ini, analisis dapat difokuskan pada fitur-fitur tersebut saat membangun model. Fitur yang memiliki nilai MI rendah mungkin dapat diabaikan, yang dapat memudahkan model dan mengurangi kompleksitas. Ini juga membantu dalam mendeteksi anomali, di mana pola lalu lintas yang tidak biasa dapat menandakan serangan atau kegiatan mencurigakan. Analisis dari grafik ini membantu dalam memahami hubungan antara fitur dan label, serta dalam merumuskan strategi yang lebih efektif untuk pengawasan dan respons terhadap insiden keamanan.

## SIMPULAN

Penelitian ini telah mengeksplorasi penggunaan MI sebagai metode seleksi fitur dalam konteks deteksi intrusi, dengan fokus pada dataset CICIDS 2017. Hasil analisis menunjukkan bahwa MI efektif dalam mengidentifikasi fitur-fitur yang paling relevan dan informatif untuk mendeteksi aktivitas yang mencurigakan dalam jaringan. Dengan mengurangi dimensi data dan memfokuskan perhatian pada fitur yang signifikan, sistem deteksi intrusi dapat meningkatkan akurasi dan mengurangi jumlah false positive. Implementasi MI dalam seleksi fitur tidak hanya memperbaiki kinerja model deteksi intrusi, tetapi juga mengefisiensikan proses pemrosesan data. Temuan ini memberikan wawasan yang berharga dalam pengembangan sistem keamanan siber yang lebih handal dan responsif terhadap ancaman yang terus berkembang. Ke depan, penelitian lebih lanjut dapat dilakukan untuk mengeksplorasi kombinasi metode seleksi fitur lainnya dan penerapan teknik pembelajaran mesin yang lebih canggih, dengan harapan dapat terus meningkatkan efektivitas sistem deteksi intrusi dalam menghadapi tantangan keamanan informasi yang semakin kompleks.

## DAFTAR PUSTAKA

Afolabi, A. S., & Akinola, O. A. (2024). Network Intrusion Detection Using Knapsack Optimization, Mutual Information Gain, and Machine Learning. *Journal of Electrical and Computer Engineering*, *2024*, 1–21. https://doi.org/10.1155/2024/7302909

Ahmad, Z., Shahid Khan, A., Wai Shiang, C., Abdullah, J., & Ahmad, F. (2021). Network intrusion detection system: A systematic study of machine learning and deep learning approaches. *Transactions on Emerging Telecommunications Technologies*, *32*(1). https://doi.org/10.1002/ett.4150

Al-E'mari, S., Sanjalawe, Y., Alsmadi, D., Alduweib, E., & Alharbi, A. (2024). EMPLOYING MUTUAL

INFORMATION FEATURE SELECTION AND LIGHTGBM FOR INTRUSION DETECTION IN IOT. *ICIC Express Letters*, *18*(6), 597–606. https://doi.org/10.24507/icicel.18.06.597

Barkah, A. S., Selamat, S. R., Abidin, Z. Z., & Wahyudi, R. (2023). Data Generative Model to Detect the Anomalies for IDS Imbalance CICIDS2017 Dataset. *TEM Journal*, *12*(1), 80–89. https://doi.org/10.18421/TEM121-11

Kurniabudi, Stiawan, D., Darmawijoyo, Bin Idris, M. Y. Bin, Bamhdi, A. M., & Budiarto, R. (2020). CICIDS-2017 Dataset Feature Analysis with Information Gain for Anomaly Detection. *IEEE Access*, *8*, 132911–132921. https://doi.org/10.1109/ACCESS.2020.3009843

Kushwaha, J. P., Bhadauria, S., & Tapaswi, S. (2023). Multi-Method Stacked Feature Selection Approach based IDS for IoT Networks. *Procedia Computer Science*, *230*, 564–573. https://doi.org/10.1016/j.procs.2023.12.112

Latif, S., Boulila, W., Koubaa, A., Zou, Z., & Ahmad, J. (2024). DTL-IDS: An optimized Intrusion Detection Framework using Deep Transfer Learning and Genetic Algorithm. *Journal of Network and Computer Applications*, *221*. https://doi.org/10.1016/j.jnca.2023.103784

Ogwara, N. O., Petrova, K., Yang, M. L., & Tan, L. (2022). Towards the Development of a Cloud Computing Intrusion Detection Framework Using an Ensemble Hybrid Feature Selection Approach. *Journal of Computer Networks and Communications*, *2022*. https://doi.org/10.1155/2022/5988567

Oyelakin, A. M., Ameen, A. O., Ogendele, T. S., Salau-Ibrahim, T., Abdulrauf, U. T., Olufadi, H. I., & Ajiboye, I. K. (2024). Overview and Exploratory Analyses of CICIDS2017 Intrusion Detection Dataset. *Indonesian Journal of Data and Science*, *4*(3). https://doi.org/10.56705/ijodas.v4i3.80

Prazeres, N., Costa, R. L. de C., Santos, L., & Rabadão, C. (2023). Engineering the application of machine learning in an IDS based on IoT traffic flow. *Intelligent Systems with Applications*, *17*. https://doi.org/10.1016/j.iswa.2023.200189

Sahu, D. P., Tripathy, B., & Samantaray, L. (2024). FogNet: Custom CNN with optimal feature selection-based combat model for secured fog computing environment. *E-Prime - Advances in Electrical Engineering, Electronics and Energy*, *8*. https://doi.org/10.1016/j.prime.2024.100604

Saq, A. H. A., Zainal, A., Al-Rimy, B. A. S., Alyami, A., & Abosaq, H. A. (2024). Intrusion Detection in IoT using Gaussian Fuzzy Mutual Information-based Feature Selection. *Engineering, Technology and Applied Science Research*, *14*(6), 17564–17571. https://doi.org/10.48084/etasr.8268

Sharafaldin, I., Lashkari, A. H., & Ghorbani, A. A. (2018). Toward generating a new intrusion detection dataset and intrusion traffic characterization. *ICISSP 2018 - Proceedings of the 4th International Conference on Information Systems Security and Privacy*, *2018-January*, 108 116. https://doi.org/10.5220/0006639801080116

Yang, Y., & Peng, X. (2025). BERT-based network for intrusion detection system. *Eurasip Journal on Information Security*, *2025*(1).

Zhou, H., Wang, X., & Zhang, Y. (2024). Feature selection based on weighted conditional mutual information. *Applied Computing and*