CyberSecurity dan Forensik Digital e-ISSN: 2615-8442
Vol. 6, No. 1, Mei 2023, hlm. 6 of 11

Deteksi Serangan Low Rate DDoS pada Jaringan Tradisional menggunakan Algoritma Decision Tree

Fadil Febriansyah[1], Zian Asti Dwiyanti[2], Diash Firdaus[3]

[1,2] Teknik Informatika, [3]Informatika
[1,2] Universitas Logistik dan Bisnis Internasional, [3]Institut Teknologi Nasional, Indonesia
Email: [1]fadilfebriansyah76@gmail.com , [2]ziandwiasti23@gmail.com, [3]diashfirdaus@gmail.com

Abstrak

Serangan Distributed Denial of Service (DDoS) menjadi salah satu serangan yang dapat melumpuhkan lalu lintas dan servis jaringan dengan cara membebani server, network link dan perangkat jaringan (switch, router, dll.) dengan traffic network yang sangat tinggi. Deteksi DDoS dapat dilakukan dengan menggunakan machine learning, salah satunya dengan menggunakan algoritma Decision Tree. Decision Tree adalah salah satu metode yang sering digunakan dalam data mining dan machine learning untuk memprediksi hasil atau mengambil keputusan berdasarkan input yang diberikan. Dalam aplikasinya untuk mendeteksi serangan Low Rate DDoS (Distributed Denial of Service) pada jaringan tradisional, decision tree dapat digunakan untuk memprediksi kemungkinan terjadinya serangan Low rate DDoS berdasarkan beberapa fitur yang dianggap penting dalam mengidentifikasi serangan tersebut. Fitur-fitur tersebut bisa berupa jumlah traffic yang masuk ke jaringan, tipe traffic yang masuk, atau karakteristik traffic lainnya. Untuk mendeteksi serangan low rate DDoS pada jaringan tradisional dengan menggunakan dataset CICIDS 2017. Hasil analisis menunjukkan bahwa metode decision tree dengan algoritma Gini Index lebih baik dari Entropy untuk mendeteksi low rate DDoS (Distributed Denial of Servcice) pada jaringan tradisional berdasarkan nilai Accuracy, Precision , dan F1 Score, yaitu dengan nilai 99,740%, 99,113%, dan 99,231%.

Kata kunci: *Decision tree*, DDoS, *Machine learning*, CICIDS2017, Gini Index , Entropy

Detection of Low Rate DDoS Attacks on Traditional Networks using the Decision Tree Algorithm

Abstract

Distributed Denial of Service (DDoS) attacks are attacks that can paralyze network traffic and services by overloading servers, network links and network devices (switches, routers, etc.) with very high network traffic. DDoS detection can be done using machine learning, one of which is using the Decision Tree algorithm. Decision Tree is a method that is often used in data mining and machine learning to predict results or make decisions based on the input provided. In its application to detect Low Rate DDoS (Distributed Denial of Service) attacks on traditional networks, decision trees can be used to predict the possibility of Low rate DDoS attacks based on several features that are considered important in identifying such attacks. These features can be the amount of traffic that enters the network, the type of traffic that comes in, or other traffic characteristics. To detect low rate DDoS attacks on traditional networks using the CICIDS 2017 dataset. The results of the analysis show that the decision tree method with the Gini Index algorithm is better than Entropy for detecting low rate DDoS (Distributed Denial of Service) on traditional networks based on Accuracy, Precision, and F1 Score, with values of 99.740%, 99.113% and 99.231%.

***Keywords**: Decision tree, DDoS, Machine learning, CICIDS2017, Gini Index , Entropy*

1. PENDAHULUAN

Perkembangan teknologi merupakan salah satu hal yang paling banyak dibahas ketika mengkaji isu industri 4.0 yang erat dengan konektivitas internet, penggunaan teknologi yang lebih banyak dalam kehidupan sehari-hari, dan kegiatan ekonomi.Terlepas dari peran teknologi dalam jaringan internet, manfaat dengan adanya revolusi industri 4.0 pertukaran informasi dapat dengan mudah dilakukan kapan saja serta dimana saja tanpa batasan waktu hanya dengan jaringan internet.(Jose, 2021)

Dengan semakin banyak memanfaatkan teknologi jaringan komputer selain mendapatkan banyak manfaat seperti mempermudah pekerjaan, penggunaan jaringan komputer juga rentan terhadap serangan. Serangan *Distributed Denial of Service* (DDoS) menjadi salah satu serangan yang dapat melumpuhkan lalu lintas dan servis jaringan dengan cara membebani server, network link dan perangkat

jaringan (switch, router, dll.) dengan *traffic network* yang sangat tinggi.(Syahputra et al., 2020)

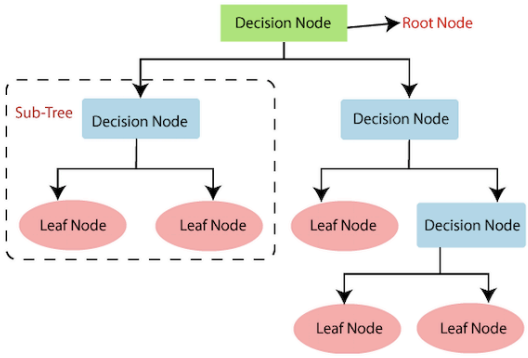
Permasalahan yang sering terjadi di perusahaan atau instansi biasanya terdapat *server down* sehingga ketika kita mengakses jaringan tersebut jaringan tidak dapat diakses. Hal ini tentu sangat mengganggu kegiatan yang ada di perusahaan.Dengan semakin banyak permasalahan pada jaringan komputer di era 4.0 ini, maka diperlukan sebuah model *machine learning* untuk melakukan pendeteksian adanya serangan *low rate* DDoS. Diharapkan model *machine learning* ini mampu bekerja dengan baik dalam mendeteksi serangan *low rate* DDoS.(Pangestu & Solichin, 2022)

Oleh sebab itu, sangat penting untuk merealisasikan sistem deteksi *low rate* DDoS untuk menjaga keamanan jaringan. Dalam penelitian ini akan dilakukan deteksi serangan *low rate* DDoS pada Jaringan Tradisional menggunakan algoritma *Decision Tree*. Manfaat utama menggunakan algoritma *Decision tree* adalah kemampuan untuk memecah proses pengambilan keputusan yang kompleks menjadi lebih sederhana, sehingga pengambil keputusan lebih baik menafsirkan solusi untuk masalah.(Charbuty & Abdulazeez, 2021)

Dalam beberapa tahun terakhir, teknologi *machine learning* (ML) telah digunakan untuk mengembangkan sistem cerdas dengan cara melatih mesin untuk membuat keputusan. Dengan menggunakan dataset sebagai input dan classifier sebagai metodenya, ML dapat mengidentifikasi data baru yang memiliki kemiripan dengan data yang telah dikenal. Ada banyak algoritma ML yang dapat digunakan untuk membangun framework atau model ML, dan masing-masing algoritma memiliki kelebihan dan kekurangan yang berbeda tergantung pada dataset dan fitur yang digunakan.

1.A. *Decision Tree*

Decision tree merupakan metode klasifikasi dan prediksi yang sangat kuat dan terkenal. Metode *decision tree* mengubah fakta yang sangat besar menjadi pohon keputusan yang mempresentasikan aturan. Aturan dapat dengan mudah dipahami dengan bahasa alami.(Lakshmi et al., 2016) *Decision tree* adalah algoritma yang digunakan dalam data mining dan *machine learning* untuk membuat pohon keputusan yang digunakan untuk memprediksi hasil atau mengambil keputusan berdasarkan sejumlah input. Algoritma ini bekerja dengan membuat pohon keputusan yang terdiri dari node-node yang mewakili pertanyaan atau kondisi, dan edge-edge yang menghubungkan node-node tersebut.



Gambar 1. Struktur Decision Tree
Sumber : https://glints.com/id/lowongan/decision-tree-adalah/

Setiap node pada pohon keputusan akan membagi data menjadi dua kelompok: yang memenuhi kondisi node tersebut dan yang tidak memenuhi kondisi node tersebut. Proses ini akan dilakukan kembali untuk setiap kelompok data yang terbentuk hingga terdapat keputusan yang dapat diambil. Dengan demikian, *decision tree* dapat digunakan untuk memprediksi hasil atau mengambil keputusan berdasarkan pertanyaan-pertanyaan yang diajukan kepada data yang tersedia.

Ada beberapa teknik yang biasa digunakan dalam membangun pohon keputusan, di antaranya:

- ID3 (*Iterative Dichotomiser* 3): Algoritma ini menggunakan metode entropy dan information gain untuk memutuskan pertanyaan yang paling tepat untuk diajukan pada setiap node pada pohon keputusan.(Yusa et al., 2016)
- C4.5: Algoritma ini mirip dengan ID3, namun menggunakan metode gain ratio yang lebih advance untuk memutuskan pertanyaan yang paling tepat untuk diajukan pada setiap node.
- CART (*Classification and Regression Trees*): Algoritma ini menggunakan metode Gini Index untuk memutuskan pertanyaan yang paling tepat untuk diajukan pada setiap node.(Yusa et al., 2016)

1.B. Gini & Entropy

Gini adalah ukuran ketidakseimbangan dalam distribusi dari sebuah atribut dari sekumpulan objek data. Nilai gini yang lebih rendah menunjukkan distribusi yang lebih merata (less skewed).(Mauludin Rohman & Adinugroho, 2021)

Entropy adalah ukuran ketidakpastian (impuritas) dari sebuah atribut dari sekumpulan objek data dalam satuan bit. Entropy biasanya digunakan dalam pengklasifikasian data di mana entropy yang lebih rendah menunjukkan kelas yang lebih homogen (kurang bervariasi).(Mauludin Rohman & Adinugroho, 2021)

1.C. Pengukuran *Performa*

Untuk pengukuran performa pada penelitian ini akan menggunakan *confusion matrix* seperti yang disajikan pada tabel 1. *Confusion matrix* adalah salah satu cara yang sering digunakan untuk mengevaluasi kinerja model *machine learning*, termasuk model yang dibuat menggunakan algoritma *decision tree*. *Confusion matrix* adalah tabel yang menggambarkan hasil prediksi model dibandingkan dengan hasil yang sebenarnya.

Untuk menggunakan *confusion matrix*, kita perlu memilih dua kelas yang akan diprediksi oleh model, biasanya disebut sebagai "positif" dan "negatif". Kemudian, kita akan menghitung jumlah *true positive* (TP), *false positive* (FP), *true negative* (TN), dan *false negative* (FN). Jumlah-jumlah tersebut akan digunakan untuk menghitung beberapa metrik yang umum digunakan untuk mengevaluasi kinerja model, seperti akurasi, presisi, dan recall.(Mauludin Rohman & Adinugroho, 2021)

Tabel 1. *Confusion Matrix*

	Hasil Serangan	**Hasil Normal**
Prediksi Serangan	TP (*True Positive*)	FP (*False Positive*)
Prediksi Normal	FN (*False Negative*)	TN (*True Negative*)

- Akurasi adalah persentase dari prediksi yang benar dari seluruh prediksi yang dilakukan model. Akurasi dapat dihitung dengan menggunakan rumus :

$$Accuracy = \frac{TP + TN}{TP+TN+FP+FN} \times 100\%$$

- Presisi adalah persentase dari prediksi positif yang benar dari seluruh prediksi positif yang dilakukan model. Presisi dapat dihitung dengan menggunakan rumus :

$$Precision = \frac{TP}{FP+TP} \times 100\%$$

- *Recall* adalah persentase dari prediksi positif yang benar dari seluruh data positif yang ada pada dataset. *Recall* dapat dihitung dengan menggunakan rumus :

$$Recall = \frac{TP}{TP+FN} \times 100\%$$

- *F1 score* merupakan rata-rata harmonis dari presisi dan *recall*. *F1 score* dapat dihitung dengan menggunakan rumus :

$$F1\ Score = \frac{2 \times (precision \times recall)}{precision + recall} \times 100\%$$

Jumlah TP menunjukkan jumlah paket yang benar-benar merupakan serangan DDoS dan diprediksi dengan benar oleh model sebagai serangan. Jumlah FP menunjukkan jumlah paket yang normal namun salah diprediksi sebagai serangan oleh model. Jumlah FN menunjukkan jumlah paket yang merupakan serangan namun salah diprediksi sebagai normal oleh model. Jumlah TN menunjukkan jumlah paket yang benar-benar normal dan diprediksi dengan benar oleh model sebagai normal.(NOR et al., 2022)

Dengan menghitung jumlah-jumlah tersebut, kita dapat menghitung beberapa metrik yang biasa digunakan dalam mengevaluasi performa model, seperti akurasi, presisi, recall dan f1 score. Semakin tinggi nilai metrik-metrik tersebut, semakin baik performa model tersebut dalam mendeteksi serangan DDoS pada jaringan tradisional.

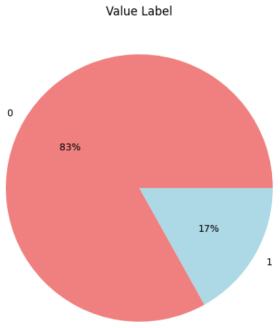
1.D. Dataset

Dataset yang digunakan dalam penelitian ini adalah CICIDS2017. CICIDS2017 adalah dataset yang digunakan dalam konferensi "*Canadian Conference on Computer and Information Science* (CICIS)" pada tahun 2017. Dataset ini merupakan kumpulan data yang dikumpulkan dari jaringan tradisional yang mengalami serangan DDoS (*Distributed Denial of Service*). Dataset ini terdiri dari beberapa tipe serangan DDoS yang umum, seperti HTTP Flood, UDP Flood, SYN Flood, dan lainnya.(Panigrahi & Borah, 2018)

Dataset CICIDS2017 biasa digunakan dalam penelitian dan pengembangan algoritma untuk mendeteksi serangan *low rate* DDoS pada jaringan tradisional. Dataset ini juga dapat digunakan untuk melatih model *machine learning*, seperti model yang dibangun dengan menggunakan algoritma *decision tree*, untuk memprediksi serangan *low rate* DDoS. Dataset yang digunakan dalam penelitian ini adalah :

Tabel 2. Data Penelitian

Jenis Data	**Jumlah**	**Persen**
Serangan	425878	83%
Normal	2096484	17%
Total	2522362	100%



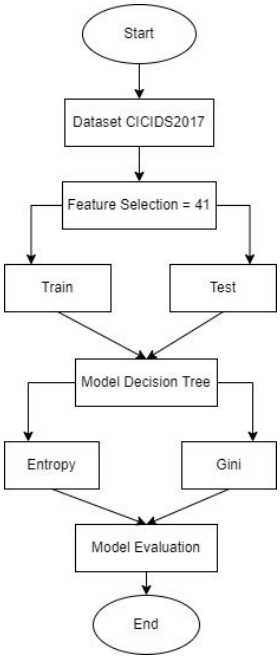
Gambar 2. Pie Chart Data Penelitian

2. METODOLOGI PENELITIAN

Pada bagian ini peneliti memaparkan kerangka penelitian, metoda dan *experiment setup* yang digunakan pada penelitian.

2.A. Kerangka Penelitian

Pada penelitian ini penulis menggunakan beberapa tahapan kegiatan, yang disusun dalam kerangka penelitian pada gambar berikut :



Gambar 3. Tahapan Penelitian

Berdasarkan gambar 3, dapat dijelaskan tahapan-tahapan yang dilakukan dalam penelitian ini adalah sebagai berikut :

- Pengumpulan data, pada tahap pengumpulan data pada penelitian ini yaitu menggunakan dataset dari CICIDS2017.
- Dalam data set tersebut terdapat 79 atribut , namun yang digunakan dalam penelitian ini hanya 41 atribut dengan menggunakan *library sklearn.feature_selection* yaitu selectKBest dan chi2.
- Setelah dilakukan *feature selection*, dataset tersebut akan dibagi menjadi dua bagian, yaitu 70% untuk pelatihan (*train*) dan 30% untuk pengujian (*test*).
- Kemudian, dataset *train* akan dilatih menggunakan algoritma *decision tree* (DT). Dua algoritma yang digunakan dalam pembuatan pohon keputusan (*decision tree*) untuk mengukur kemurnian suatu kelas dalam sebuah node yaitu Gini dan Entropy.
- Model evaluation. Untuk mengevaluasi hasil pengujian, beberapa metrik untuk mengukur kinerja model yang umum digunakan adalah akurasi, presisi, recall, dan f1 score. Dengan menghitung metrik-metrik tersebut, kita dapat mengetahui seberapa baik model yang telah dibuat dapat digunakan untuk memprediksi serangan *low rate* DDoS pada jaringan tradisional.

2.B. Experiment Setup

Experiment Setup yang dibutuhkan untuk mendeteksi Low Rate DDoS menggunakan Machine Learning sebagai berikut.

Tabel 3. *Experiment Setup*

No.	Kebutuhan	Keterangan
1	System OS Virtual	Ubuntu dan Kali Linux
2	Simulation Tools	Mininet, VirtualBox
3	Dataset	CICIDS2017
4	CPU	Intel core i3
5	RAM	8GB
6	Operating System	Windows 10
7	Supporting Tools	Jupyter Notebook Sklearn Library Matplotlib Library Python3

3. HASIL DAN PEMBAHASAN

Pada hasil dan pembahasan merupakan hasil dari *decision tree* gini index dan entropy untuk mendeteksi serangan *low rate* DDoS pada jaringan tradisional. Berikut merupakan atribut yang digunakan dalam penelitian :

Tabel 4. *Features*

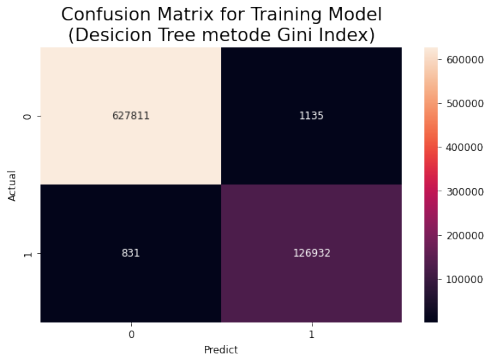
No.	Features	No.	Features
1	Destination Port	22	Bwd IAT Std
2	Flow Duration	23	Bwd IAT Max
3	Total Length of Bwd Packets	24	Bwd IAT Min
4	Fwd Packet Length Mean	25	Fwd Packets/s
5	Fwd Packet Length Std	26	Bwd Packets/s
6	Bwd Packet Length Max	27	Max Packet Length
7	Bwd Packet Length Mean	28	Packet Length Mean
8	Bwd Packet Length Std	29	Packet Length Std
9	Flow Bytes/s	30	Packet Length Variance
10	Flow Packets/s	31	Average Packet Size
11	Flow IAT Mean	32	Avg Fwd Segment Size
12	Flow IAT Std	33	Avg Bwd Segment Size

No.	Features	No.	Features
13	Flow IAT Max	34	Subflow Bwd Bytes
14	Flow IAT Min	35	Active Mean
15	Fwd IAT Total	36	Active Std
16	Fwd IAT Mean	37	Active Max
17	Fwd IAT Std	38	Idle Mean
18	Fwd IAT Max	39	Idle Max
19	Fwd IAT Min	40	Idle Min
20	Bwd IAT Total	41	Label
21	Bwd IAT Mean		

Setelah mengevaluasi kedua algoritma yang telah dilakukan, tahap selanjutnya adalah membandingkan performa masing-masing metode dengan menghitung nilai *accuracy*, *precission*, *recall*, dan *F1 Score*.

1.E. Hasil Implementasi Metode *Decission Tree* Algoritma Gini Index

Hasil yang diperoleh dari implementasi menggunakan metode *decision tree* dengan algoritma gini index untuk deteksi *low rate* DDoS dapat dilihat pada gambar *confusion matrix* dibawah ini :



Gambar 4. *Confusion Matrik* Gini Index

Keterangan :

Tabel 5. *Confusion Matrix* Gini Index

Jenis Metrix	Hasil Score
Accuracy	99.739 %
Precision	99.111 %
Recall	99.344 %
F1 Score	99.228 %

Dari *confusion matrix* untuk *decision tree* yang menggunakan algoritma gini index terdapat data normal yang diperoleh nilai *true positif* sebanyak 627811 dan *true negative* sebanyak 1135, ini berarti bahwa dari data normal teridentifikasi sebanyak 1135. Sementara itu, untuk data serangan, diperoleh nilai *true positif* sebanyak 831 dan *true negative* sebanyak 126932. Ini berarti bahwa dari data serangan teridentifikasi sebanyak 126932.

1.F. Hasil Implementasi Metode *Decission Tree* Algoritma Entropy

Hasil yang di peroleh dari implementasi menggunakan metode *decision tree* dengan algoritma entropy untuk deteksi *low rate* DDos dapat di lihat pada gambar *confusion matrix* di bawah ini :

Gambar 5. *Confusion Matrix* Entropy

Keterangan :

Tabel 6. Confusion Matrix Entropy

Jenis Metrix	Hasil Score
Accuracy	99.739 %
Precision	99.108 %
Recall	99.351 %
F1 Score	99.229 %

Dari *confusion matrix* untuk *decision tree* yang menggunakan algoritma gini index terdapat data normal yang diperoleh nilai *true positif* sebanyak 627813 dan *true negative* sebanyak 1142, ini berarti bahwa dari data normal teridentifikasi sebanyak 1142. Sementara itu, untuk data serangan, diperoleh nilai *true positif* sebanyak 829 dan *true negative* sebanyak 126925. Ini berarti bahwa dari data serangan teridentifikasi sebanyak 126925.

1.G. Hasil Perbandingan

Setelah melakukan implementasi terhadap semua metode selanjutnya melakukan perbandingan kedua algoritma dalam perhitungan *Accuracy*, *Precision* , *Recall* dan *F1 score*.

Tabel 7. Hasil Perbandingan

Metrik	Gini Index	Entropy
Accuracy	99.740 %	99.739 %
Precision	99.113 %	99.108 %
Recall	99.349 %	99.351 %
F1 Score	99.231 %	99.229 %

Dapat disimpulkan dari tabel diatas bahwa algoritma Gini Index memiliki nilai *Accuracy Precission* dan *F1 Score* yang lebih tinggi, sementara algoritma Entropy memiliki nilai Recall yang lebih tinggi. Oleh karena itu, pilihan terbaik tergantung pada prioritas yang ingin dicapai. Jika yang diutamakan adalah keakuratan dalam memprediksi, maka algoritma Gini Index mungkin lebih sesuai. Namun, jika yang diutamakan adalah Recall, maka metode Entropy mungkin lebih sesuai.

4. KESIMPULAN DAN SARAN

Berdasarkan dari hasil analisis dapat disimpulkan sebagai berikut.

- Berdasarkan nilai Gini Index *Accuracy* ,*Precision* dan *F1 Score* untuk mendeteksi *low rate* DDoS pada jaringan tradisional lebih baik dari Entropy karena *Accuracy*, *Precision* dan *F1 Score* yaitu dengan nilai 99.740 % , 99.113 % dan 99.231 %.
- Berdasarkan nilai *Recall* untuk mendeteksi *low rate* DDoS pada jaringan tradisional, Entropy lebih baik dari pada Gini Index. Dengan nilai *Recall* sebesar 99.351 % .

Berdasarkan hasil analisis tersebut, disarankan untuk menggunakan algoritma Gini Index dalam mendeteksi *low rate* DDoS pada jaringan tradisional jika prioritasnya adalah memperoleh hasil yang akurat dalam memprediksi . Namun, jika yang diutamakan adalah *Recall* , maka disarankan untuk menggunakan algoritma Entropy. Selain itu, disarankan untuk mengevaluasi kembali hasil analisis tersebut dengan menggunakan data yang lebih banyak dan melakukan perbandingan dengan algoritma-algoritma lain untuk memastikan hasil yang optimal.

DAFTAR PUSTAKA

CHARBUTY, B., & ABDULAZEEZ, A. (2021). Classification Based on Decision Tree Algorithm for Machine Learning. *Journal of Applied Science and Technology Trends*, *2*(01), 20–28. https://doi.org/10.38094/jastt20165

JOSE, H. S. (2021). Politisasi Agenda Keamanan Siber Pada Era Industri 4.0 di Forum Multilateral. *Populika*, *9*(2), 70–85. https://doi.org/10.37631/populika.v9i2.390

LAKSHMI, B. N., INDUMATHI, T. S., & RAVI, N. (2016). A Study on C.5 Decision Tree Classification Algorithm for Risk Predictions During Pregnancy. *Procedia Technology*, *24*, 1542–1549. https://doi.org/10.1016/j.protcy.2016.05.128

MAULUDIN ROHMAN, M., & ADINUGROHO, S. (2021). Analisis Sentimen pada Ulasan Aplikasi Mobile JKN Menggunakan Metode Maximum Entropy dan Seleksi Fitur Gini Index Text. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, *5*(6), 2646–2654. http://j-ptiik.ub.ac.id

NOR, S., MUSLIM, M. A., & ASWIN, M. (2022). Pengenalan Pola Dasar Angka berdasarkan Gerakan Tangan menggunakan Machine Learning. *ELKOMIKA: Jurnal Teknik Energi Elektrik, Teknik Telekomunikasi, & Teknik Elektronika*, *10*(3), 595. https://doi.org/10.26760/elkomika.v10i3.595

PANGESTU, R., & SOLICHIN, A. (2022). Klasifikasi Serangan Jaringan Menggunakan Metode Decision Tree Berbasis Website Classification Of Network Attacks Using Website-Based Decision Tree Method. *Klasifikasi Serangan Jaringan Menggunakan Metode Decision Tree Berbasis Website Classification Of Network Attacks Using Website-Based Decision Tree Methodng Website-Based Decision Tree Method*, *September*, 614–620.

PANIGRAHI, R., & BORAH, S. (2018). A detailed analysis of CICIDS2017 dataset for designing Intrusion Detection Systems. *International Journal of Engineering and Technology(UAE)*, *7*(3.24 Special Issue 24), 479–482.

SYAHPUTRA, M. Q., AKBI, D. R., & RISQIWATI, D. (2020). Deteksi Dan Mitigasi Serangan DDoS Pada Software Defined Network Menggunakan Algoritma Decision Tree. *Jurnal Repositor*, *2*(11), 1491. https://doi.org/10.22219/repositor.v2i11.795

YUSA, M., UTAMI, E., & LUTHFI. TAUFIQ, E. (2016). Evaluasi Performa Algoritma Klasifikasi Decision Tree ID3, C4.5, dan CART Pada Dataset Readmisi Pasien Diabetes. *Infosys (Information System) Journal*, *4*(1), 23–34.