Terakreditasi SINTA Peringkat 4
Surat Keputusan Direktur Jenderal Pendidikan Tinggi, Riset, dan Teknologi, Nomor: 158/E/KPT/2021
masa berlaku mulai Volume 5 Nomor 2 Tahun 2018 sampai Volume 10 Nomor 1 Tahun 2023

Terbit online pada laman web jurnal: **https://jurnal.plb.ac.id/index.php/tematik/index**

TEMATIK
Jurnal **Teknologi Informasi Komunikasi (e-Journal)**
Vol. 9 No. 2 (2022) 161 – 166 ISSN Media Elektronik: 2443-3640

Perbandingan Algoritma Machine Learning Dalam Mendeteksi Serangan DDOS

Comparison of Machine Learning Algorithms in Detecting DDOS Attacks

Wahyuni[1], Pitrasacha Adytia[2]
[1]Teknik Informatika, STMIK Widya Cipta Dharma
[2]Sistem Informasi, STMIK Widya Cipta Dharma
[1]wahyuni@wicida.ac.id, [2]pitra@wicida.ac.id

Abstract

Ddos is an attack method by sending a lot of packets into a network that causes the device not to run according to its function. This attack will result in machine or network resources cannot be accessed or used by the user. Various methods are used to detect DDOS attacks on SDN, namely statistical methods, machine learning, SDN architecture, blockchain, Network Function Virtualization, honeynets, network slicing, and moving target defense. Because so many people use machine learning to detect DDoS attacks, it is necessary to do further research to find out which one is the best and has high accuracy. Therefore, a research entitled "Comparison of Machine Learning Algorithms in Detecting DDoS Attacks was made. In this study, three machine learning algorithms will be compared, namely XGBoost, Decision Tree and ANN. The methods used are data acquisition, data understanding, data preparation, modeling, performance evaluation, and conclusions. In this study it can be said that for accuracy, the highest model is XGBoost in determining attacks, but to execute it requires the longest time among other models tested. While Decision tree also has high accuracy, slightly below XGBoost, but the time required to execute is fast or short. Therefore, in this study it can be said that the Decision Tree is the best model in detecting and classifying DDoS attacks.Keywords: Ddos Attack, Machine Learning, Decision Tree, XGBoost, ANN.

Keywords: DdoS attack, machine learning, decision tree, XGBoost, ANN

Abstrak

Ddos adalah sebuah metode serangan dengan mengirimkan banyak paket kedalam sebuah jaringan yang menyebabkan perangkat jaringan tidak berjalan sesuai fungsinya. Serangan ini akan mengakibatkan sumber daya mesin ataupun jaringan tidak bisa diakses atau digunakan oleh pengguna. Berbagai macam metode dilakukan dalam pendeteksian serangan DDOS pada SDN , yaitu dengan cara statistikal, machine learning, arsitektur SDN, *blockchain*, *Network Function Virtualization*, *honeynets*, *network slicing*, dan *moving target defense*. Karena begitu banyak yang menggunakan *machine learning* dalam mendeteksi serangan DDoS, maka perlu dilakukan penelitian lanjutan untuk mengetahui algoritma mana yang paling baik dan memiliki akurasi yang tinggi. Oleh karena itu dibuatlah penelitian yang berjudul "Perbandingan Algoritma Machine Learning Dalam Mendeteksi Serangan DDoS. Pada penelitian ini akan dibandingkan sebanyak tiga buah algoritma machine learning, yaitu XGBoost, Decision Tree dan ANN. Metode yang digunakan adalah *data acquisition, data understanding, data preparation, modelling*, evaluasi performansi, dan kesimpulan. Dalam penelitian ini dapat disimpulkan bahwa untuk *accuracy*, model yang tertinggi adalah XGBoost dalam menentukan serangan, namun untuk mengeksekusinya membutuhkan waktu yang paling lama di antara model lain yang diuji. Sedangkan Decision tree juga memiliki accuracy yang tinggi, sedikit di bawah XGBoost, namun waktu yang dibutuhkan untuk mengeksekusi sangatlah cepat atau singkat. Oleh karena itu dalam penelitian ini dapat disimpulkan bahwa Decision Tree adalah model yang paling baik dalam mendeteksi dan mengklasifikasi kan serangan DdoS.

Kata kunci: : serangan DdoS, machine learning, decision tree, XGBoost, ANN

1. Pendahuluan

Ddos *(Distributed Denial Of Service*) merupakan jenis serangan *Denial of Service* yang menggunakan banyak host penyerang baik itu menggunakan komputer yang didedikasikan untuk penyerangan atau komputer yang dipaksa menjadi zombie untuk menyerang satu buah host target dalam sebuah jaringan [1]. Ddos adalah

Diterima Redaksi: 28-10-2022 | Selesai Revisi: 07-11-2022 | Diterbitkan Online: 01-12-2022

sebuah metode serangan dengan mengirimkan banyak paket kedalam sebuah jaringan yang menyebabkan perangkat jaringan tidak berjalan sesuai fungsinya [2]. Serangan *Distributed Denial-of-Service* (DDoS) dianggap sebagai ancaman keamanan utama bagi server online dan penyedia *cloud*[]. Ada beberapa jenis serangan ddos yang sering terjadi seperti *UDP Flooding, SYN Flooding, Ping Of Death*, dan *Remote Controlled Attack* [3]. Pada bulan Oktober hingga akhir Desember 2021 para peneliti Kaspersky mengamati peningkatan besar-besaran dalam bentuk jumlah serangan ddos. Serangan ddos mencapai rekor tertinggi di Q4 2021 dibandingkan dengan Q3 2021, jumlah total serangan ddos menunjukkan peningkatan 52%. Serangan ddos Q4 dilaporkan di beberapa negara seperti Amerika Serikat (43,55%), China (9,96%), Hong Kong (8,80%), Jerman (4,85%), dan Prancis (3,75%).

Serangan DDoS adalah bentuk serangan yang dilakukan dengan mengirim paket secara terus menerus kepada mesin bahkan jaringan komputer. Serangan ini akan mengakibatkan sumber daya mesin ataupun jaringan tidak bisa diakses atau digunakan oleh pengguna. Serangan DDoS merupakan varian dari serangan DOS, dimana perbedaanya terletak pada dispersi sumber serangan [4]. Serangan DDOS adalah serangan yang berbahaya dan mengancam dalam suatu jaringan, karena dapat membanjiri jaringan dan memblokir akses ke server dengan mengirim paket dalam jumlah besar dan menggunakan sumber daya jaringan untuk menolak akses lainnya [5]. Pada serangan DDoS jalur serangan digenerate dari beberapa sumber, sedangkan serangan DOS hanya bersumber dari satu tempat. Berbagai macam metode dilakukan dalam pendeteksian serangan DDOS pada SDN [6] , yaitu dengan cara statistikal, *machine learning*, arsitektur *SDN, blockchain, Network Function Virtualization, honeynets, network slicing, dan moving target defense*. Selain itu sudah banyak pula pada penelitian sebelumnya yang mengangkat topik mendeteksi serangan DDOS pada SDN menggunakan *machine learning*.

Pada penelitian [6], dijelaskan mengenai pendekatan Machine Learning untuk melawan serangan DDOS. Dilakukan analisis mengenai pendekatan *single* dan *hybrid machine learning* dalam mendeteksi serangan DDOS. Penelitian tersebut juga mendiskusikan mengenai perbedaan sistem penangkal DDOS berbasis machine learning yang menggunakan virtual environment, termasuk *cloud computing, software defined network* dan *network function virtualization.*

Pada penelitian [7] telah dibandingkan beberapa model algoritma machine learning (XGBoost, KNN, Stochastic gradient descent, dan Naïve Bayes) untuk mendeteksi dan mengklasifikasi serangan DDoS. Dan hasil dari penelitian tersebut adalah XGBoost menghasilkan nilai akurasi tertinggi.

Cara untuk meminimalisir serangan ini salah satunya dengan menggunakan solusi mitigasi tradisional seperti teknik analisis trafik jaringan dengan bantuan manusia [8], akan tetapi mengalami beberapa batasan dan masalah kinerja. Untuk mengatasi keterbatasan ini, *Machine Learning* telah menjadi salah satu teknik utama untuk memperkaya, melengkapi, dan meningkatkan pengalaman keamanan tradisional. *Machine learning* memiliki banyak sekali model atau algoritma yang bisa dimanfaatkan untuk mendeteksi serangan DDoS, namun kita masih perlu untuk menguji tingkat akurasi yang dihasilkan oleh masing-masing model tersebut. *Decision Tree* dan ANN juga dapat digunakan untuk mendeteksi serangan tersebut. Namun, kita masih belum bisa memastikan apakah model tersebut baik atau tidak untuk diaplikasikan.

Decision tree merupakan algoritma yang umum digunakan [9]. Perhitungannya cukup mudah, namun memiliki keakuratan yang cukup baik dalam berbagai kasus. Ide dasar dari algoritma ini adalah mencari akar sebagai fitur yang memiliki pengaruh tertinggi pada kasus. Kemudian fitur lainnya akan berada di bawah fitur akar sebagai cabang hingga sampai daun. Pada tingkatan daun, perhitungan tidak bisa lagi dilakukan karena itu adalah akhir dari struktur pohon. Setelah model dibuat, akan menghasilkan aturan IF-Then yang dapat digunakan dan dipahami tanpa membutuhkan pengetahuan statistika sama sekali. Salah satu contohnya adalah adalah penggunaan *Decision Tree* untuk mengidentifikasi potensi caloj kreditur yang berpotensi bermasalah atau lancar [10]. Selain menggunakan *Decision Tree* juga digunakan pengklasifikasian *Naïve Bayes* sebagai pembanding. Hasil penelitian menyimpulkan bahwa *Decision Tree* lebih superior dan lebih sesuai untuk permasalahan dan dataset yang digunakan.

Artificial Neural Network (ANN) adalah sebuah teknologi komputasi parallel yang meniru system kerja jaringan syaraf otak manusia [11]. Seperti pada neuron manusia, ANN juga terdiri dari saluran input, komponen pemrosesan, dan saluran output. Pada AN juga terdapat fungsi aktivasi dan threshold. Neuron-neuron ini juga saling terhubung dan membentuk ANN dan hubungan tersebut disebut bobot (*weight). Artificial Neural Network (*ANN) merupakan metode *machine learning* yang memiliki tiga hingga empat layer [12]. Layer-layer tersebut dinamakan multilayer perceptron (MLP). Bagian dari MLP yaitu *input layer, hidden layer,* dan *output layer*. Tiap layer memiliki banyak neuron. Neuron dapat memiliki input yang sama, namun bobotnya berbeda.

XGBoost merupakan salah satu algoritma yang sangat *powerfull* di antara algoritma *machine learning* yang

DOI: 10.38204/tematik.v9i2.1070

ada [13]. Perbedaan antara *Gradient Boosting* dan *XGBoost* adalah pada cara kerja matematis untuk tiap model dan masing-masing implementasi. *XGBoost* memberikan 3 peningkatan dibanding *AdaBoost* dan *Reguler Gradient Boosting* yaitu, lebih cepat, secara umum lebih baik, dan bisa menggunakan lebih banyak parameter untuk optimasi.

Karena begitu banyak yang menggunakan machine learning dalam mendeteksi serangan DDoS, maka perlu dilakukan penelitian lanjutan untuk mengetahui algoritma mana yang paling baik dan memiliki akurasi yang tinggi. Oleh karena itu dibuatlah penelitian yang berjudul "Perbandingan Algoritma *Machine Learning* Dalam Mendeteksi Serangan DDoS. Pada penelitian ini akan dibandingkan sebanyak tiga buah algoritma machine learning, yaitu *XGBoost, Decision Tree* dan ANN. Melalui penelitian ini diharapkan kita dapat mengetahui algoritma mana yang paling baik dan yang memiliki akurasi paling tinggi dalam mendeteksi serangan DDoS.

2. Metode Penelitian

Adapun metode penelitian yang digunakan pada penelitian ini adalah:

Data Acquisition : Penelitian diawali dengan tahapan akusisi data. Pada penelitian ini menggunakan data CICIDS 2019. Dataset ini digunakan dipilih karena memiliki serangan DDoS yang paling mutahir dan menyerupai data dunia nyata..

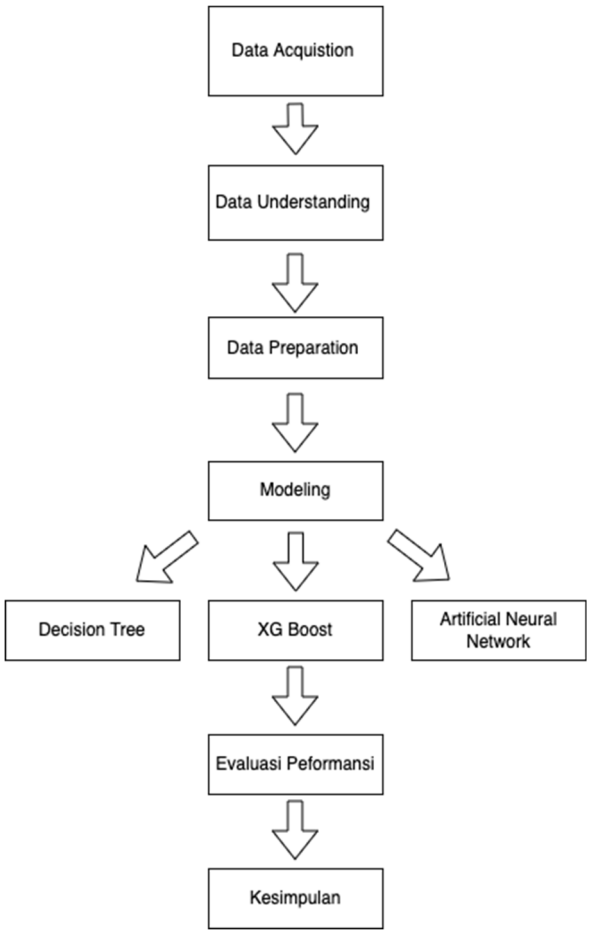
Data Understanding : Tahapan ini merupakan tahapan untuk proses pengumpulan data, mengidentifikasi data untuk menemukan pola yang menarik yang dapat digunakan untuk membuat hipotesis [14]. Pada tahapan ini dilakukan investigasi awal, dan pemahaman terhadap data sehingga dapat menemukan karakter dari data. Adapun metode yang digunakan adalah menggunakan *Exploratory Data Analysis (EDA)*

Data Preparation : Pada tahap ini data akan dipersiapkan sehingga mempermudah proses mining [15]. Proses preparation ini mencakup tiga hal utama yaitu *data selection, data preprocessing* dan *data transformation.*

Modeling : Pada tahapan ini dilakukan permodelan menggunakan algoritma klasifikasi. Pada penelitian ini menggunakan 3 algoritma klasifikasi yaitu : (1) Decision tree, (2) XGBoost, dan (3) ANN.

Evaluasi Performansi *: Performansi dilihat dari* confusion matrix *yang terdiri dari* accuracy, precision, recall, *dan* f1 score. *Selain itu pada penelitian ini juga melihat lama waktu yang digunakan* training *dan prediksi.*

Kesimpulan : Pada tahap ini membuat kesimpulan berdasarkan evaluasi pefomansi yang dilakukan pada tahap sebelumnya



Gambar 1. Metode Penelitian

3. Hasil dan Pembahasan

3.1. Spesifikasi

Penelitian ini dilakukan pada komputer dengan spesifikasi processor 3.1 GHz Dual Coe Intel Core i7 dengan memory sebesar 16 GB1857 MHz DDR3 pada system operasi macOS Big Sur.

3.2. Data Understanding

Dataset CICIDS 2019 terdiri dari 2.827.808 baris dan 81 fitur. Terdiri dari dari 2.271.320 data Bening dan 556. 488 serangan DDoS. Detail Jenis label serangan DDoS dapat dilihat pada tabel 1.

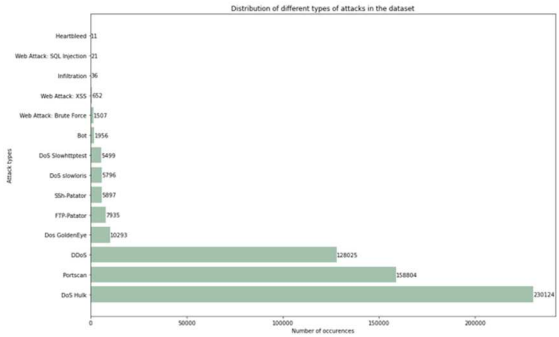
Tabel 1. Jenis serangan DDoS

Label	Jumlah
BENIGN	2.271.320
DoS Hulk	230.124
PortScan	158.804
DdoS	128.025
DoS GoldenEye	10.293

DOI: 10.38204/tematik.v9i2.1070

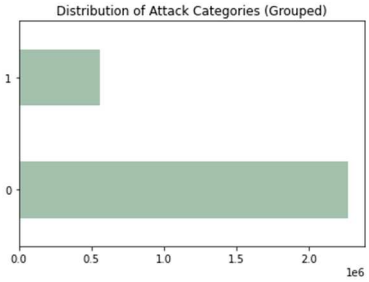
Label	Jumlah
FTP-Patator	7.935
SSH-Patator	5.897
DoS slowloris	5.796
DoS Slowhttptest	5.499
Bot	1.956
Brute Force	1.507
XSS	652

Jika disajikan dalam bentuk grafik, maka distribusi data serangan pada dataset dapat dilihat pada gambar 2.



Gambar 2 Jenis Serangan DDoS

Label BENIGN berarti data bukan serangan DDoS , selain itu adalah serangan DDoS. Untuk dapat membedakan maka dibuat kolom baru diberi nama Attack dimana nilai dari kolom adalah 0 dan 1. Nilai 0 berarti BENIGN sedangkan nilai 1 berarti serangan. Sehingga jika dibuat dalam bentuk grafik maka distribusi BENIGN dan serangan DDos dapat dilihat pada gambar 3.



Gambar 3 Distribusi Serangan

3.3. Data Preparation

Dataset dipisahkan menjadi data train , test dan validate Kolom yang hanya memiliki 1 nilai yang sama dianggap tidak relevan. Untuk itu pada penelitian ini kolom tersebut dihapus tidak diikutsertakan pada pembuatan model. Setelah itu dilakukan proses normalisasi. Normalisasi data dilakukan dengan menggunakan metode min-max. Untuk memastikan semua data diantara nilai 0 dan 1. Dapat dilihat pada tabel 2.

Tabel 2

Count	1.187679e+08
mean	1.006497e-01
std	2.714226e-01
min	0.000000e+00
25%	0.000000e+00
50%	1.085271e-06
75%	7.543457e-03
Max	1.000000e+00

Setelah dilakukan normalisasi, selanjutnya dilakukan dengan menggunakan SelectKBest dengan fungsi yang digunakan adalah Chi-Squared dengan nilai k=40. Sehingga didapatkan 40 fitur yang akan digunakan dalam proses permodelan.

3.4. Modeling

Selanjutnya proses permodelan dilakukan untuk decision tree, XGBoost dan juga ANN. Berikut merupakan gambar dari implementasi model *Decision Tree*.

```
Program Model Decision Tree
# fit the model
start = time.time()
classifier.fit(x_train, y_train.Attack)
end = time.time()
training_time = end - start

# predict validation
start = time.time()
y_predicted = classifier.predict(x_validate)
end = time.time()
predict_time = end  - start
```

Gambar 4. Program Model Decision Tree

Setelah melakukan pemodelan menggunakan model *Decision Tree*, maka dilanjutkan dengan melakukan pemodelan dengan menggunakan model *XGBoost* dapat dilihat pada gambar 5.

```
Program Model XGBoost
# fit the model
start = time.time()
model = XGBClassifier()
model.fit(x_train, y_train.Attack)
end = time.time()
training_time = end - start

# Predict validation
start = time.time()
y_pred = model.predict(x_validate)
end = time.time()
predict_time = end - start
```

Gambar 5. Program Model XGBoost

Lalu setelah itu dilakukan lagi pemodelan dengan menggunakan model ANN. Adapaun implementasinya dapat dilihat pada gambar 6.

DOI: 10.38204/tematik.v9i2.1070

Program Model ANN Classifier

```
model = Sequential()

# input layer
model.add(Dense(40,  activation='relu'))
model.add(Dropout(0.2))

# hidden layer
model.add(Dense(20, activation='relu'))
model.add(Dropout(0.2))


# output layer
model.add(Dense(units=1,activation='sigmoid'))

# Compile model
model.compile(loss='binary_crossentropy',
optimizer='adam')

# Fit the model
start = time.time()
model.fit(x=x_train,
            y=y_train.Attack,
            epochs=25,
            batch_size=256,
            validation_data=(x_validate,
y_validate.Attack),
            )
end = time.time()
training_time = end - start


#predict validation
start = time.time()
y_pred = (model.predict(x_validate) >
0.5).astype("int32")
end = time.time()
predict_time = end - start
```

Gambar 6. Program Model ANN Classifier

3.5 Evaluasi Model

Evaluasi model menggunakan confussion matrix dan waktu eksekusi training dan predict. Perbandingan dari ketiga algoritma dapat dilihat pada dilihat pada tabel 3

Tabel 3. Perbandingan dari ketiga algoritma

Model	Accuracy	Precision	Recall	F1 Score	Training Time	Predict Time
Decision Tree	99,87	0,99	0,99	0,99	73,12	0,12
XGBoost	99,92	0,99	0,99	0,99	506,46	1,177
ANN	98,27	0,98	0,98	0,97	422,60	23,46

4. Kesimpulan

Dari hasil evaluasi model yang menggunakan confussion matrix, kita dapat melihat *accuracy*, *precision*, *recall* dan F1 *score*. Untuk *accuracy*, dapat dilihat bahwa *Decision Tree* memiliki *accuracy* sebesar 99,87%, *XGBoost* memiliki *accuracy* sebesar 99,92%, dan ANN sebesar 98,27%. Dari hasil di atas, dapat dilihat bahwa yang memiliki accuracy tertinggi adalah model *XGBoost*. Untuk *precision*, *Recall*, dan F1 *Score*, *Decision Tree* dan *XGBoost* memiliki nilai yang sama sebesar 0,99% sedangkan ANN memiliki *Precision* 0,98 % *Recall* 0,98% dan *F1Score* 0,97%. Untuk *Training Time*, *Decision Tree* membutuhkan waktu 73,12s, *XGBoost* 506,46s dan ANN membutuhkan waktu 422,60s. Sedangkan untuk *Predict Time Decision Tree* membutuhkan waktu 0,12s, *XGBoost* membutuhkan waktu 1,177s dan ANN membutuhkan waktu 23,46s. Berdasarkan dari paparan di atas, dapat disimpulkan bahwa untuk *accuracy*, model yang tertinggi adalah *XGBoost* dalam menentukan serangan, namun untuk mengeksekusinya membutuhkan waktu yang paling lama di antara model lain yang diuji. Sedangkan *Decision tree* juga memiliki *accuracy* yang tinggi, sedikit di bawah XGBoost, namun waktu yang dibutuhkan untuk mengeksekusi sangatlah cepat atau singkat. Oleh karena itu dalam penelitian ini dapat disimpulkan bahwa Decision Tree adalah model yang paling baik dalam mendeteksi dan mengklasifikasi kan serangan DdoS.

Daftar Rujukan

[1] R. F. Junaedi, D. Stiawan and A. Heryanto, "Deteksi Serangan DDoS (Distributed Denial Of Service) Di Cloud Computing Dengan Menggunakan Metode Rule Base," Undergraduate Thesis, Universitas Sriwijaya, 2019.

[2] I. Riadi, R. Umar and F. D. Aini, "Analisis Perbandingan Detecsion Traffic Anomaly Dengan Metode Naïve Bayes dan Support Vector Machine (SVM)," *ILKOM Jurnal Ilmiah*, vol. 11, p. 17, 2019.

[3] P. Darryl and M. Subali, "Perbandingan Algoritma SVM dan Algoritma KNN dalam Menghasilkan Klasifikasi DDoS dan Benign.," *Jurnal Ilmiah KOMPUTASI*, vol. 20, no. https://doi.org/http://dx.doi.org/10.32409/jikstik.20.4.2799, pp. 491-500, 2021.

[4] B. B. Gupta and A. Dahiya, Distributed Denial of Service (DDoS) Attacks, CRC Press, 2021.

[5] F. Rahmatullah, "Deteksi Distributed Denial Of Service (DDOS) Dalam Jaringan Software Defined Network Dengan Metode Support Vector Machine," UPN "Veteran", Yogyakarta, 2022.

[6] I. A. Valdinos, J. A. Perez-Diaz, K.-K. R. Choo and J. F. Botero, "Emerging DDOS attack detection and mitigation strategies in software-defined network: Taxonomy, Challenges and Future Directions," *ELSEVIER:Journal of Network and Computer Applications*, vol. 187, 2021.

[7] G. Usha, M. Narang and A. Kumar, "Detection and Classification of Distributed DoS Attacks Using Machine Learning," in *Computer Networks and Inventive Communication Technologies*, India, 2021.

[8] M. N. Faiz, O. Somantri, A. R. Supriyono and A. W. MUhammad, "Impact of Feature Selection Methods on Machine Learning-based for Detecting DDoS Attacks : Literature Review," *JITE (Journal of Informatics and Telecommunication Engineering)*, p. 305, 2022.

[9] I. Saputra and D. A. Kristiyanti, "Machine Learning Untuk Pemula".

[10] M. Sadikin, Teknik Machine Learning Untuk Menangani Permasalahan Data Bias Pada Transaksi Aplikasi POS, Jawa Tengah: Zahira Media Publisher, 2022.

[11] A. Kuswantori, Memahami ANN, Deep Learning, CNN dan YOLO, Malang: Medza Media, 2022.

[12] W. Setiawan, Deep Learning Menggunakan Convolutional Neural Network: Teori dan Aplikasi, Malang: Media Nusa Creative, 2020.

DOI: 10.38204/tematik.v9i2.1070

[13] N. Vandeput, Data Science for Supply Chain Forecasting, Boston: Walter de Gruyter GmbH, 2021.

[14] N. L. W. S. R. Ginantra and dkk., Data Mining dan Penerapan Algoritma, Yayasan Kita Menulis, 2021.

[15] N. W. Wardani, Penerapan Data Mining Dalam Analytic CRM, Yayasan Kita Menulis, 2020.

DOI: 10.38204/tematik.v9i2.1070