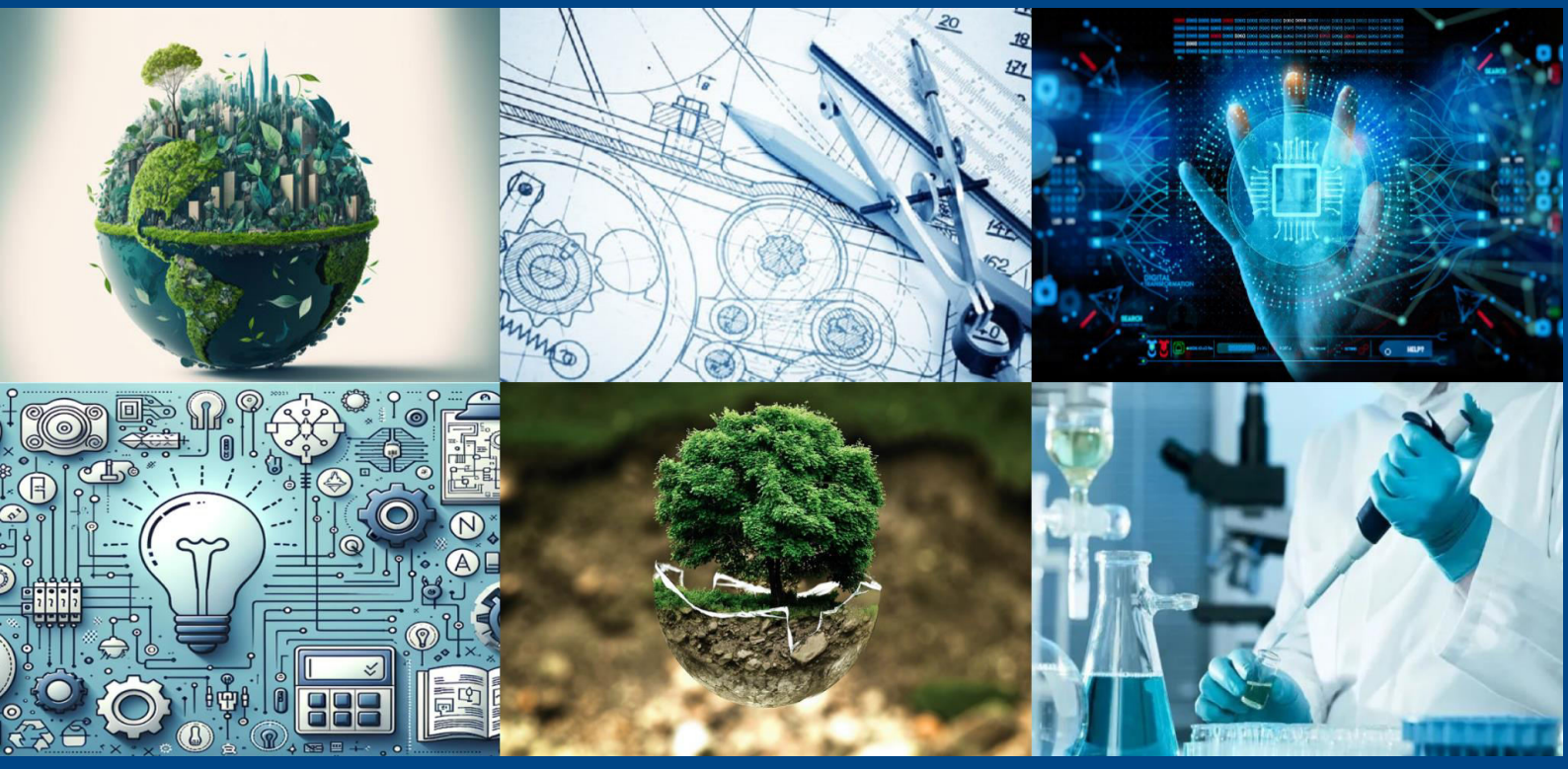
ISSN: 2582-7219

International Journal of Multidisciplinary Research in Science, Engineering and Technology

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.206 **Volume 8, Issue 5, May 2025**

© 2025 IJMRSET | Volume 8, Issue 5, May 2025| DOI:10.15680/IJMRSET.2025.0805227

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Detecting Unknown Distributed Denial of Service (DDoS) Attacks in Server-Side Network Traffic using Machine Learning

Dr D. Eshwara Chaitanya[1], B. Sai Rohith[2], D. Sai Venkatesh[3], B. Hemesh[4], B. Kalvik[5]

Associate Professor, Department of ECE, R.V.R&J.C.C. E, Chowdavaram, Guntur, A.P., India[1]

Undergraduate Students, Department of ECE, R.V.R&J.C.C. E, Chowdavaram, Guntur, A.P., India[2-5]

ABSTRACT: Distributed Denial of Service (DDoS) attacks continue to pose an ongoing challenge to maintaining online services available and unscathed. Traditional signature-based systems would likely be undermined by novel or evolving channels of attack. This research presents a machine learning-based detection system that is able to identify known and hitherto unknown DDoS attacks from network traffic features. The system employs four classifiers—Decision Tree, Random Forest, Passive-Aggressive, and XG Boost—on learning behavioral features from the CIC-DDoS2019 dataset. Through rigorous preprocessing with feature selection, normalization, and class balancing via SMOTE, training and testing the models were achieved on key performance metrics. Performance results show that the Random Forest classifier performed best in terms of overall accuracy and resilience, while the XG Boost model offered superior performance in terms of AUC, indicating its excellent ability to distinguish between classes. The Passive-Aggressive model provided a lightweight, real-time solution. The methodology developed is promising for deployment to live network monitoring infrastructure, offering a scalable, adaptive defense for DDoS attacks.

KEYWORDS: DDoS, machine learning, network security, anomaly detection, Random Forest, Decision Tree, Passive Aggressive, XG Boost SMOTE, cybersecurity.

I. INTRODUCTION

In today's age of pervasive cloud interconnectivity, security and protection of communication networks are of utmost concern. As the businesses, governments, and people of the world become more dependent on cloud computing, Internet of Things (IoT) devices, and 5G technologies, it is imperative to provide continuous availability of service. One of the biggest dangers to network availability is Distributed Denial of Service (DDoS) attacks, which deliberately bombard the affected systems with a deluge of traffic, resulting in service disruptions, financial losses, and reputational harm.

DDoS attacks have increased in size and sophistication in terms of scale and complexity. Initial attacks were mostly volumetric and based on simple flooding methods like SYN. Legacy detection methods rely overwhelmingly on signature-based or threshold-based methods, which compare network traffic against known attack signatures or look for out-of-pattern traffic-volume spikes. Even though these are effective at flagging known attacks, they fail miserably when it comes to flagging new, zero-day, or quiet varieties of attack. Furthermore, they generate enormous false positives upon encountering real traffic spikes, including flash crowds, and thus aren't well-suited to the modern dynamic network environment. To meet these challenges, researchers have resorted to machine learning (ML) methods, which allow for the creation of adaptive models that learn from network traffic data and identify non-normal patterns. Detection systems based on ML can generalize to unknown attacks, which is appropriate for the constantly evolving dynamics of the threat landscape. The techniques, however, need to be applied in the right manner by well-selecting models, feature engineering, and preprocessing for dealing with imbalanced and noisy data.

© 2025 IJMRSET | Volume 8, Issue 5, May 2025| DOI:10.15680/IJMRSET.2025.0805227

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

This work aims at building a robust DDoS detection model using four different machine learning classifiers, namely Decision Tree (DT), Random Forest (RF), Passive-Aggressive (PA), and XG Boost. These classifiers were used due to their complementary advantages in interpretation, accuracy, robustness, and real-time processability. Training and testing are carried out using the CIC-DDoS2019 dataset offering labelled network traffic samples with a rich variety of DDoS attack types. The key contributions presented in this manuscript include a comparative study of DT, RF, PA, and XG Boost classifiers to identify unknown DDoS attacks, the application of data pre-processing techniques such as SMOTE for class imbalance problems, an in-depth analysis of classifiers based on various performance parameters such as accuracy, precision, recall, F1-score, and ROC curves, and a discussion on what is suited for each model for real-time and scalable network security systems.

The remainder of the paper is organized as follows. Section 2 presents related work in DDoS detection and machine learning solutions. Section 3 describes suggested system architecture including data preprocessing and model training. Section 4 presents experimental setup and evaluation results. Section 5 concludes with future direction of work
.

II. LITERATURE SURVEY

DDoS attack detection is a critical area of research, driven by the escalating impact of cyber threats on online service availability. As businesses and individuals increasingly rely on interconnected technologies, adaptive defense mechanisms are vital to counter sophisticated and evolving DDoS attacks that overwhelm traditional security systems. This section reviews key studies and methodologies, particularly those utilizing machine learning, for advancing DDoS threat detection.

Industry reports consistently highlight the growing scale and complexity of DDoS attacks. Cloudflare's 2023 Q3 DDoS Threat Report details current trends and attack vectors, underscoring the continuous evolution of these threats [1]. This ongoing challenge necessitates dynamic and intelligent detection systems.
Major technology providers have also documented significant DDoS incidents. Microsoft Security Response Center's response to HTTP/2 DDoS Attacks illustrates the impact on critical internet protocols [2]. Similarly, Cloudflare's analysis of the HTTP/2 Zero-Day Vulnerability [3] and Google Cloud's explanation of the 'Rapid Reset' DDoS Attack [4] provide insights into the technical specifics and unprecedented scale of recent sophisticated attacks.
Machine learning has emerged as a promising approach for automated DDoS detection. Ahuja et al. (2021) explored ML applications in Software-Defined Networking (SDN) for enhanced security against DDoS, demonstrating how these techniques can be integrated into modern network infrastructures [5].

Awan et al. (2021) addressed real-time DDoS detection using big data, emphasizing the scalability required for processing massive volumes of network traffic [6]. Their work highlights the importance of big data frameworks in handling the velocity and volume of network data for effective threat identification.
Doriguzzi-Corin et al. (2020) introduced LUCID, a lightweight deep learning solution for DDoS attack detection, showcasing its effectiveness with reduced computational overhead, suitable for various deployment environments [7].
Bansal and Kaur (2018) focused on optimizing XG Boost for intrusion detection systems through hyperparameter tuning [8]. Their research is crucial for maximizing the performance of this powerful gradient boosting framework in classifying network anomalies.

Beyond traditional ML, Kim, Shin, and Choi (2020) demonstrate d the capability of CNNs for intrusion detection, showing how deep learning can learn complex features from raw network data to identify cyberattacks, including DDoS [9].

Finally, foundational work by Cesa-Bianchi, Conconi, and Gentile (2005) on algorithms like the Second-order Perceptron provides a theoretical basis for understanding adaptive, real-time learning. This underpins the utility of models such as the Passive-Aggressive Classifier in the current project for its responsive nature [10].

In summary, the literature underscores the shift towards advanced machine learning and deep learning for DDoS detection, moving beyond static methods. These studies collectively inform the development of adaptive and intelligent systems crucial for protecting network availability against evolving DDoS threats.

© 2025 IJMRSET | Volume 8, Issue 5, May 2025| DOI:10.15680/IJMRSET.2025.0805227

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

III. PROPOSED METHODOLOGY

The proposed methodology for Distributed Denial of Service (DDoS) attack detection employs a robust machine learning-based approach to study network traffic features. The overall process, illustrated in the block diagram, involves systematic data handling, model training, and a comprehensive evaluation pipeline designed to accurately identify various types of DDoS attacks.

A. Dataset Preparation

The study commences with Data Acquisition, where the network traffic dataset, specifically the CIC-DDoS2019 dataset, is obtained. This comprehensive dataset provides labelled samples of both benign and various DDoS attack types. Following acquisition, the data undergoes rigorous Data Preprocessing. This critical phase involves several steps to ensure the data's quality and suitability for machine learning. Initially, any rows containing missing values are dropped to maintain data integrity. The 'Label' column, which categorizes traffic as 'BENIGN' or 'Attack', is then converted into a binary numerical format (0 for BENIGN, 1 for Attack) to facilitate model training.

A key aspect of preprocessing is Feature Selection, where a specific set of impactful features relevant to DDoS detection is chosen to reduce dimensionality and focus the models on critical attributes. These selected features then undergo Feature Scaling using Standard Scaler to normalize their range, preventing any single feature from dominating the learning process due to its magnitude. This prepares the data for effective model training.

After preprocessing, the refined dataset is split into training (80%) and testing (20%) sets using a Train-Test Split strategy. Stratified sampling is employed during this split to ensure that the proportions of benign and attack classes are maintained in both subsets, which is crucial given the typical class imbalance in network traffic data. To further address this imbalance, the Synthetic Minority Over-sampling Technique (SMOTE) is applied to the training data. This Handling Imbalance step synthesizes new samples for the minority (attack) class, resulting in a more balanced training dataset that enhances the models' ability to learn and detect rare attack instances effectively.

B. Model Implementation

The core of the detection system lies in the implementation of several machine learning classifiers. Building upon the pre-processed and balanced feature set, the Model Training phase involves the application of four distinct models: Decision Tree (DT), Random Forest (RF), Passive Aggressive Classifier (PA), and XG Boost. Each model is selected for its complementary advantages in terms of interpretability, accuracy, robustness, and suitability for real-time processing. The Decision Tree and Random Forest models are ensemble methods known for their ability to capture complex non-linear relationships in data. The Passive Aggressive Classifier is an online learning algorithm, making it particularly adaptive to streaming data. XG Boost, a powerful gradient boosting framework, is chosen for its high performance and efficiency in classification tasks, often excelling in distinguishing between classes. Each model is initialized with a fixed random state for reproducibility during the training process.

C. Training and Evaluation

The training process involves fitting each of the selected machine learning models to the SMOTE-resampled training data. Once trained, the models proceed to the Attack Detection phase, where they are used to make predictions on the unseen, scaled test dataset.

For Results and Analysis, the performance of each trained model is rigorously evaluated using a suite of standard classification metrics:

- Accuracy: Overall correctness of predictions.
- Precision: Proportion of positive identifications that were actually correct.
- Recall: Proportion of actual positives that were identified correctly.
- F1-Score: Harmonic mean of precision and recall, providing a balanced measure.
- AUC (Area Under the Receiver Operating Characteristic Curve): Measures the model's ability to distinguish between attack and benign classes.

In addition to numerical metrics, the evaluation includes comprehensive visualizations. A bar graph is generated to visually compare the key performance metrics across all trained models, offering an intuitive overview of their comparative strengths. Furthermore, detailed Confusion Matrices are plotted for each classifier, providing a granular breakdown of True Positives, True Negatives, False Positives, and False Negatives, which is essential for understanding each model's specific classification behaviour and its effectiveness in distinguishing between benign and attack traffic.

© 2025 IJMRSET | Volume 8, Issue 5, May 2025| DOI:10.15680/IJMRSET.2025.0805227

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

This systematic approach ensures a thorough assessment of each model's suitability for deployment in real-world DDoS detection scenarios.

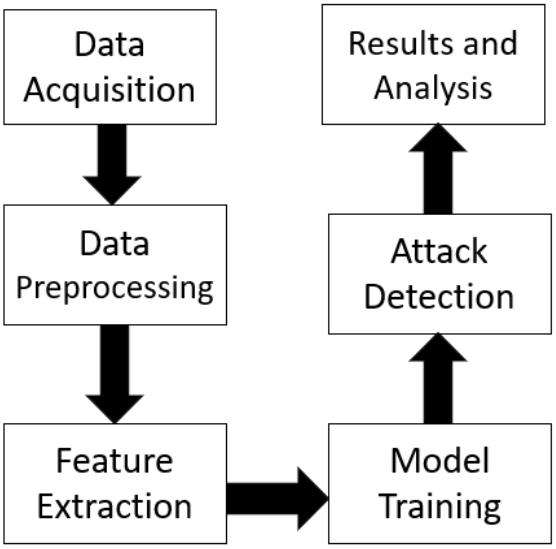


Figure 1: Block diagram for proposed methodology

IV. TESTING AND RESULTS

The below table represents a comparision between all the metric parameters of all the algorithms

--- Performance Metrics Summary (on the test set) ---

	Accuracy	Precision	Recall	F1 Score	AUC
Decision Tree	0.9776	0.9947	0.9824	0.9885	0.8988
Random Forest	0.9785	0.9947	0.9833	0.9890	0.9014
Passive Aggressive	0.8165	0.9997	0.8134	0.8970	0.9512
XGBoost	0.8895	0.9974	0.8898	0.9406	0.9706

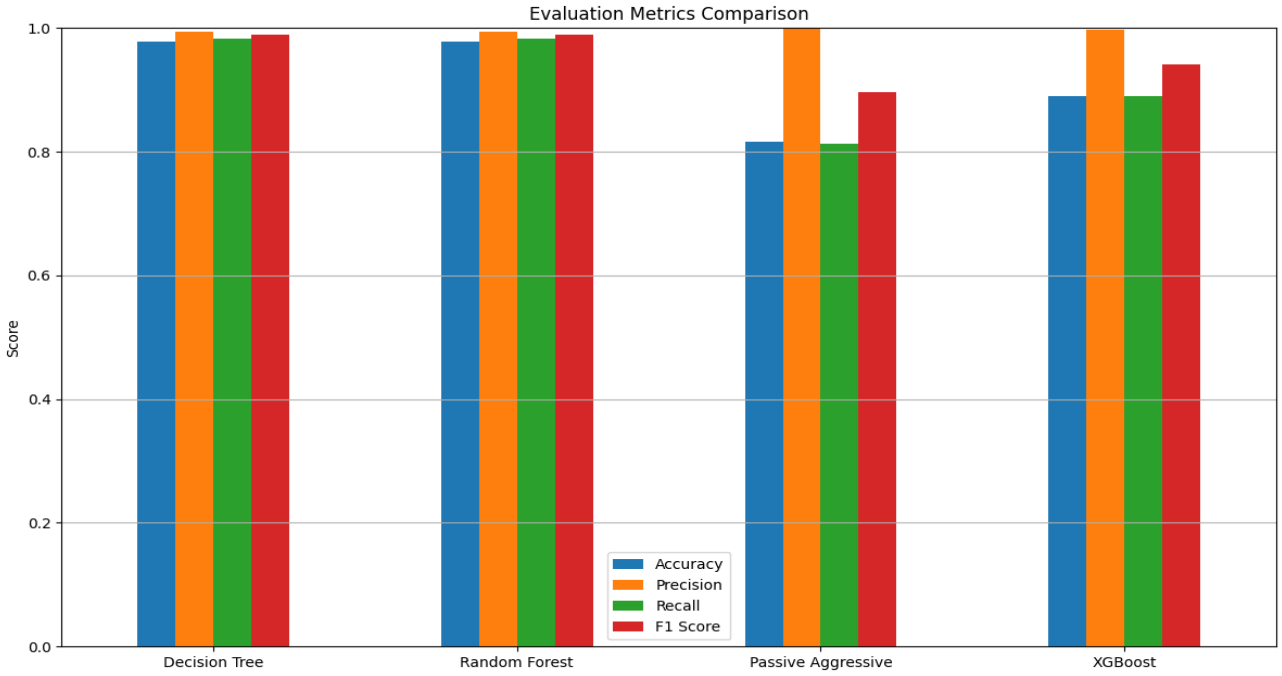
From the results, both Decision Tree and Random Forest models demonstrated high accuracy and F1-scores, with Random Forest showing a slight edge in overall performance due to its resilience. The Passive Aggressive Classifier achieved very high precision, indicating minimal false positives, though its recall and F1-score were comparatively lower. Notably, XG Boost provided a balanced and strong performance across metrics, achieving the highest AUC score, which signifies its superior ability to distinguish between benign and attack traffic. To further visualize and understand the comparative performance across all models, a bar graph illustrating the key evaluation metrics is provided below.

© 2025 IJMRSET | Volume 8, Issue 5, May 2025| DOI:10.15680/IJMRSET.2025.0805227

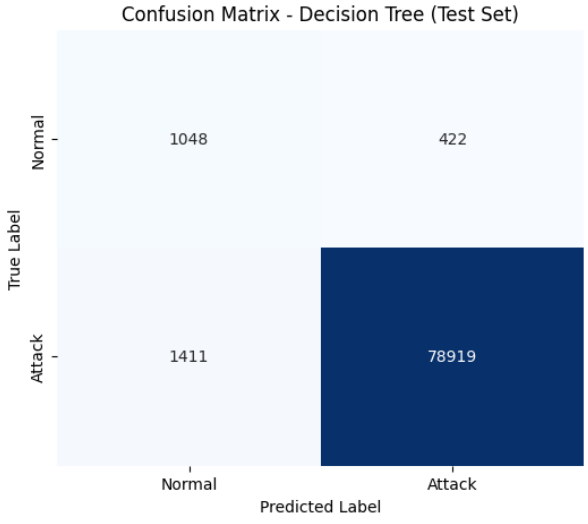
ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

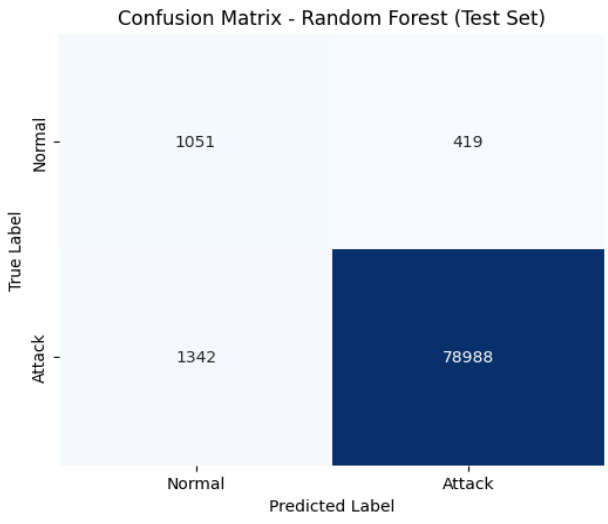
International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



A detailed breakdown of correct and incorrect classifications for each model is presented through confusion matrices. These matrices are crucial for understanding the True Positives, True Negatives, False Positives, and False Negatives, providing insights into each model's specific classification behavior and its tendency for errors.



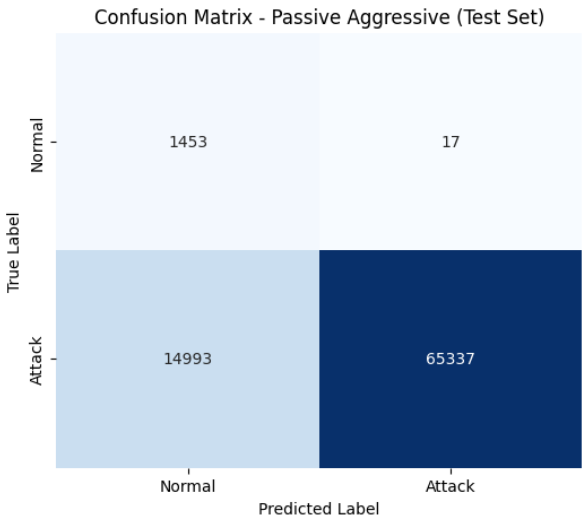


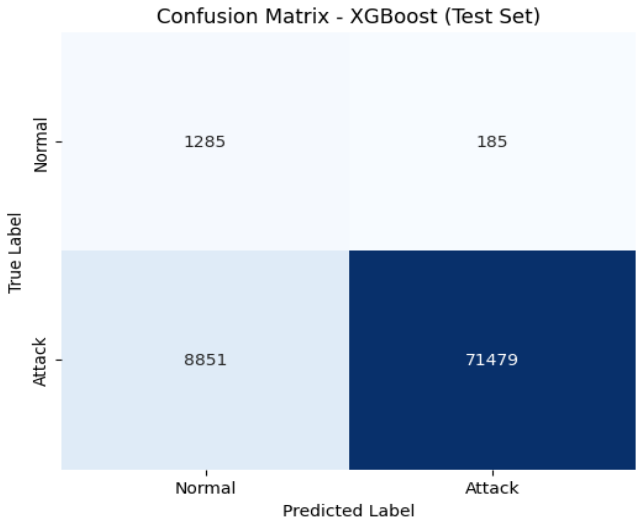
© 2025 IJMRSET | Volume 8, Issue 5, May 2025| DOI:10.15680/IJMRSET.2025.0805227

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

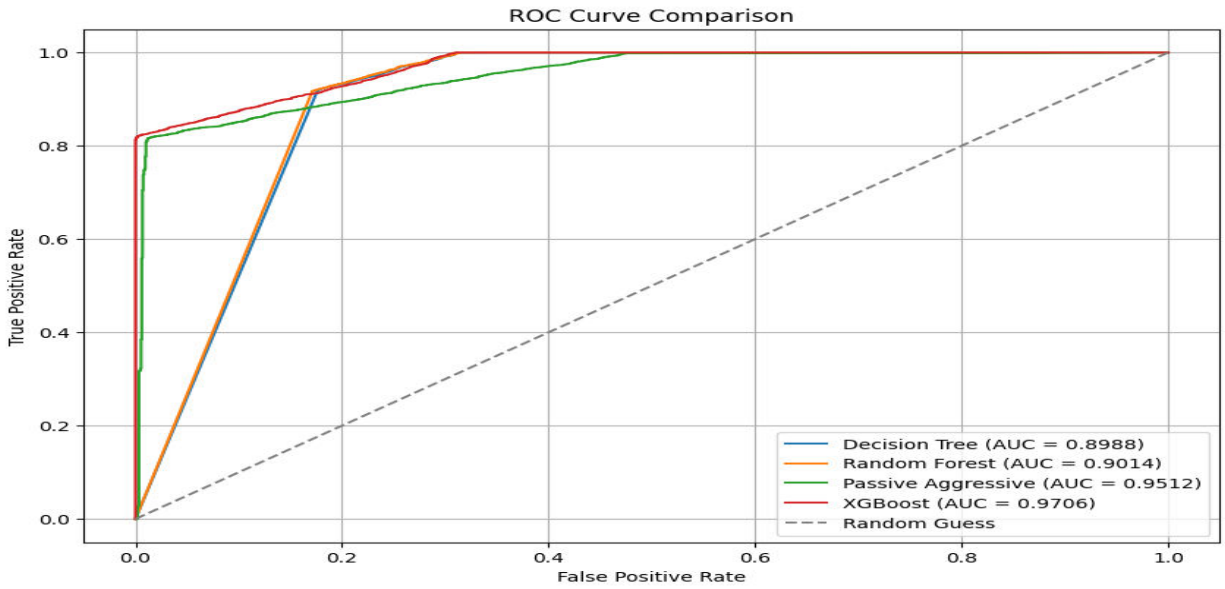
International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)





Finally, the Receiver Operating Characteristic (ROC) curves for all models are presented. The ROC curve, along with the AUC score, graphically represents the diagnostic ability of a binary classifier system as its discrimination threshold is varied. A higher AUC value, reflected by a curve closer to the top-left corner, indicates a better performance in distinguishing between the two classes. As observed in the figure, XG Boost exhibits the most favourable ROC curve, corroborating its superior AUC score and strong discriminative power.



V. CONCLUSION

This project successfully implemented and evaluated various machine learning models for DDoS attack detection, leveraging comprehensive preprocessing including SMOTE for class imbalance. Among the classifiers, Random Forest and Decision Tree demonstrated robust performance, while XG Boost achieved superior AUC, proving highly effective for precise attack differentiation. The Passive Aggressive Classifier, though high in precision, showed comparatively lower recall. The use of bar graphs, confusion matrices, and ROC curves provided critical insights into model performance. Future work will explore advanced hyperparameter tuning, feature engineering, and deep learning architectures to further enhance detection capabilities against evolving DDoS threats.

REFERENCES

1. O. Yoachimik and J. Pacheco, DDoS Threat Report for 2023 Q3,Cloudflare, 2023. [Online]. Available: https://www.cloudflare.com
2. Microsoft Security Response Center, Microsoft Response to Distributed Denial of Service (DDoS) Attacks Against HTTP/2, 2023. [Online]. Available: https://msrc.microsoft.com
3. G. Bourzikas, HTTP/2 Zero-Day Vulnerability Results in Record Breaking DDoS Attacks, Cloudflare, 2023. [Online]. Available: https://blog.cloudflare.com
4. Google Cloud, How It Works: The Novel HTTP/2 Rapid Reset DDoS Attack, 2023. [Online]. Available: https://cloud.google.com
5. N. Ahuja, G. Singal, D. Mukhopadhyay, and N. Kumar, Automated DDoS attack detection in software-defined networking, Journal of Network and Computer Applications, vol. 187, Art. no. 103108, Aug. 2021.
6. M. J. Awan, U. Farooq, H. M. A. Babar, A. Yasin, H. Nobanee, M. Hussain, O. Hakeem, and A. M. Zain, Real-time DDoS attack detection system using big data approach, Sustainability, vol. 13, no. 19, p. 10743, Sep. 2021.
7. R. Doriguzzi-Corin, S. Millar, S. Scott-Hayward, J. Martnez-del Rinc on, and D. Siracusa, LUCID: A practical, lightweight deep learning solution for DDoS attack detection, IEEE Transactions on Network and Service Management, vol. 17, no. 2, pp. 876-889, Jun. 2020.
8. A. Bansal and S. Kaur, XG Boost tuning for intrusion detection systems,in Proc. Int. Conf. on Communication Systems and Network Technologies, 2018.
9. J. Kim, Y. Shin, and E. Choi, CNN-based intrusion detection model, Electronics, vol. 9, no. 6, p. 916, Jun. 2020.
10. F. Cesa-Bianchi, A. Conconi, and C. Gentile, A Second-order Perceptron Algorithm, SIAM Journal on Computing, vol. 34, no. 3, pp. 640∟668, 2005.

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH

IN SCIENCE, ENGINEERING AND TECHNOLOGY

| Mobile No: +91-6381907438 | Whatsapp: +91-6381907438 | ijmrset@gmail.com |

www.ijmrset.com