International of Advanced Research and Multidisciplinary Journal Trends (IJARMT)
An International Open Access, Peer-Reviewed Refereed Journal
Impact Factor: 7.2 Website: https://ijarmt.com ISSN No.: 3048-9458

Optimization Accuracy of Distributed Denial of Service Attacks in Cybersecurity based on Machine Learning

[1]Geeta Singh, [2]Mr. Sudhir Goswami
M. Tech. Scholar, Department of Computer Science and Engineering, SORT, People's University, Bhopal, India[1]
Assistant Professor, Department of Computer Science and Engineering, SORT, People's University, Bhopal, India[2]

ABSTRACT
Distributed Denial of Service (DDoS) attacks remain one of the most disruptive threats in modern cybersecurity, causing severe service outages and financial losses across networked systems. This study presents an optimized machine learning-based framework for the accurate detection and classification of DDoS attacks using four widely adopted algorithms: Logistic Regression (LR), Decision Tree (DT), Random Forest (RF), and Extreme Gradient Boosting (XGBoost). The proposed approach emphasizes improving detection accuracy through effective data preprocessing, feature selection, and model tuning techniques. A benchmark network traffic dataset is utilized, where features are normalized and transformed to enhance model performance. Comparative analysis is conducted using key evaluation metrics, including training accuracy, testing accuracy, precision, and recall, to ensure robust and unbiased assessment. Experimental results demonstrate that ensemble-based models, particularly Random Forest and XGBoost, significantly outperform traditional methods such as Logistic Regression and Decision Tree in terms of accuracy and generalization capability. Among all models, XGBoost achieves the highest detection accuracy with improved precision and recall, indicating its effectiveness in handling complex and imbalanced traffic patterns. The findings highlight the importance of optimized machine learning techniques in strengthening intrusion detection systems and mitigating DDoS threats in real-time environments. This research contributes to the development of intelligent, scalable, and efficient cybersecurity solutions capable of enhancing network resilience against evolving attack vectors.

Keywords- SDN, DDoS Attack, Accuracy, AN, Boosting Algorithm

1. INTRODUCTION

The rapid expansion of internet-based services, cloud computing, and interconnected devices has significantly increased the vulnerability of modern networks to cyber threats. Among these threats, Distributed Denial of Service (DDoS) attacks have emerged as one of the most critical and challenging issues in cybersecurity. A DDoS attack aims to overwhelm a target system, server, or network with an excessive volume of traffic, rendering it inaccessible to legitimate users. With the growing sophistication of attack strategies and the availability of botnets, attackers can now launch highly coordinated and large-scale DDoS attacks, leading to severe disruptions in online services, financial losses, and compromised user trust.

International of Advanced Research and Multidisciplinary Journal Trends (IJARMT)
An International Open Access, Peer-Reviewed Refereed Journal
Impact Factor: 7.2 Website: https://ijarmt.com ISSN No.: 3048-9458

Traditional security mechanisms such as firewalls and signature-based intrusion detection systems often struggle to detect and mitigate DDoS attacks effectively, particularly when dealing with unknown or evolving attack patterns. These conventional approaches rely heavily on predefined rules and lack the adaptability required to respond to dynamic and high-volume network traffic. As a result, there is an increasing need for intelligent and automated solutions that can accurately identify malicious traffic while maintaining minimal impact on normal network operations.

Machine learning (ML) has emerged as a powerful tool in addressing cybersecurity challenges due to its ability to learn patterns from data and make accurate predictions. By analyzing network traffic features, ML algorithms can distinguish between legitimate and malicious activities, enabling early detection of DDoS attacks. In this context, various supervised learning techniques such as Logistic Regression (LR), Decision Tree (DT), Random Forest (RF), and Extreme Gradient Boosting (XGBoost) have gained significant attention for their effectiveness in classification tasks. While Logistic Regression provides a simple and interpretable baseline model, Decision Tree offers rule-based classification, and ensemble methods like Random Forest and XGBoost enhance prediction accuracy by combining multiple learners and reducing overfitting.

This study focuses on the optimization of detection accuracy for DDoS attacks using these machine learning techniques. The proposed framework involves data preprocessing, feature extraction, and model training, followed by performance evaluation using key metrics such as accuracy, precision, and recall. By comparing the performance of LR, DT, RF, and XGBoost, the research aims to identify the most effective model for real-time DDoS detection. The results highlight the superiority of ensemble-based methods in capturing complex traffic patterns and improving overall system performance.

The significance of this work lies in its contribution to the development of intelligent intrusion detection systems that are capable of adapting to evolving cyber threats. By leveraging optimized machine learning algorithms, the proposed approach enhances the reliability, scalability, and efficiency of cybersecurity solutions, ultimately strengthening network defense mechanisms against DDoS attacks in modern digital environments.

2. DISTRIBUTED DENIAL OF SERVICE ATTACKS

A Distributed Denial of Service (DDoS) attack is a highly disruptive form of cyberattack in which multiple compromised systems, often geographically distributed and connected through the internet, are used to flood a target server, network, or application with an overwhelming volume of traffic, ultimately making it inaccessible to legitimate users. These attacks are typically carried out using botnets—large networks of infected devices such as computers, servers, or Internet of Things (IoT) devices—that are remotely controlled by an attacker. Unlike traditional Denial of Service (DoS) attacks that originate from a single source, DDoS attacks are more complex and difficult to mitigate due to their distributed nature and the sheer scale of traffic they generate. The attack process generally begins with the exploitation of vulnerabilities in devices, allowing attackers to install malicious software and recruit them into a botnet. Once activated, these compromised devices simultaneously

International of Advanced Research and Multidisciplinary Journal Trends (IJARMT)
An International Open Access, Peer-Reviewed Refereed Journal
Impact Factor: 7.2 Website: https://ijarmt.com ISSN No.: 3048-9458

send a massive number of requests or data packets to the target system, consuming its bandwidth, processing power, or memory resources, and causing service degradation or complete failure. DDoS attacks can be broadly categorized into volume-based attacks that aim to saturate network bandwidth, protocol-based attacks that exploit weaknesses in network protocols, and application-layer attacks that target specific services such as web servers by mimicking legitimate user behavior. The consequences of such attacks are severe, including service downtime, financial losses, reputational damage, and reduced customer trust, particularly for organizations that rely heavily on online operations. Detecting DDoS attacks is challenging because malicious traffic often closely resembles normal traffic patterns, and attackers continuously evolve their strategies to bypass traditional security mechanisms like firewalls and signature-based intrusion detection systems. In this context, machine learning techniques have emerged as effective solutions for identifying and mitigating DDoS attacks by analyzing large volumes of network traffic data and detecting anomalies or suspicious patterns. Algorithms such as Logistic Regression, Decision Tree, Random Forest, and XGBoost are widely used to classify traffic as benign or malicious, offering improved accuracy, adaptability, and real-time detection capabilities. Overall, DDoS attacks represent a significant threat in modern cybersecurity, necessitating the adoption of intelligent, scalable, and automated defense mechanisms to ensure the availability and reliability of network services.

How DDoS Attacks Work

In a typical DDoS attack:

1. The attacker infects multiple devices (computers, IoT devices, servers) with malware.
2. These infected devices form a botnet.
3. The botnet simultaneously sends massive traffic or requests to a target system.
4. The target becomes overwhelmed and denies service to normal users.

Types of DDoS Attacks

DDoS attacks are generally classified into three categories:

1. Volume-Based Attacks
    - Aim: Consume bandwidth
    - Example: UDP Flood, ICMP Flood
    - Effect: Network congestion
2. Protocol-Based Attacks
    - Aim: Exploit server or network protocol weaknesses
    - Example: SYN Flood, Ping of Death
    - Effect: Exhaust server resources
3. Application Layer Attacks
    - Aim: Target specific applications (Layer 7)
    - Example: HTTP Flood
    - Effect: Crash web servers or APIs

Impacts of DDoS Attacks

International of Advanced Research and Multidisciplinary Journal Trends (IJARMT)
An International Open Access, Peer-Reviewed Refereed Journal
Impact Factor: 7.2 Website: https://ijarmt.com ISSN No.: 3048-9458

- Website or service downtime
- Financial losses (especially for e-commerce platforms)
- Damage to brand reputation
- Loss of customer trust
- Increased infrastructure costs

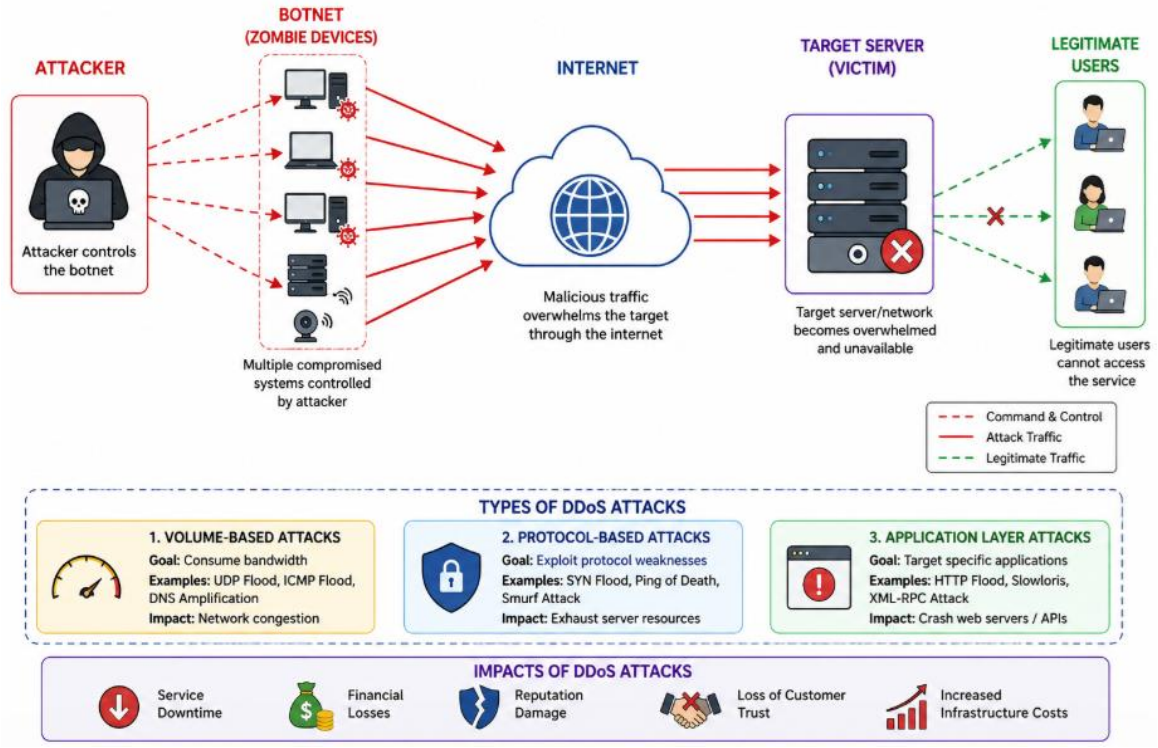

Figure 1: DDoS attacks

3. PROPOSED METHODOLOGY

The proposed methodology for detecting Distributed Denial of Service (DDoS) attacks using machine learning follows a systematic and structured pipeline designed to maximize detection accuracy and reliability. The process begins with data collection, where network traffic data is obtained from benchmark datasets such as CICIDS or similar sources containing both normal and DDoS traffic instances. This raw data is then passed through a data preprocessing stage, which is crucial for improving model performance; it includes handling missing values, removing duplicate records, encoding categorical features, and applying normalization or standardization techniques to ensure uniform data distribution. After preprocessing, the methodology proceeds to feature engineering and selection, where relevant features are extracted using techniques like correlation analysis, information gain, or variance thresholding, helping reduce dimensionality and improve learning efficiency. The dataset is then divided into training, validation, and testing subsets (commonly in 80:20 ratio) during the data splitting phase, ensuring unbiased model evaluation.

In the next stage, model training, multiple machine learning algorithms—Logistic Regression (LR), Decision Tree (DT), Random Forest (RF), and XGBoost—are trained on the processed dataset. Each model learns patterns that distinguish normal traffic from malicious DDoS

International of Advanced Research and Multidisciplinary Journal Trends (IJARMT)
An International Open Access, Peer-Reviewed Refereed Journal
Impact Factor: 7.2 Website: https://ijarmt.com ISSN No.: 3048-9458

traffic. To further enhance performance, hyperparameter tuning techniques such as Grid Search or Random Search are applied, optimizing parameters like tree depth, number of estimators, and learning rate. Once trained, the models are evaluated in the model evaluation phase using performance metrics including accuracy, precision, recall, F1-score, confusion matrix, and ROC-AUC curve. This step ensures a comprehensive assessment of each model's capability in correctly classifying traffic while minimizing false positives and false negatives.

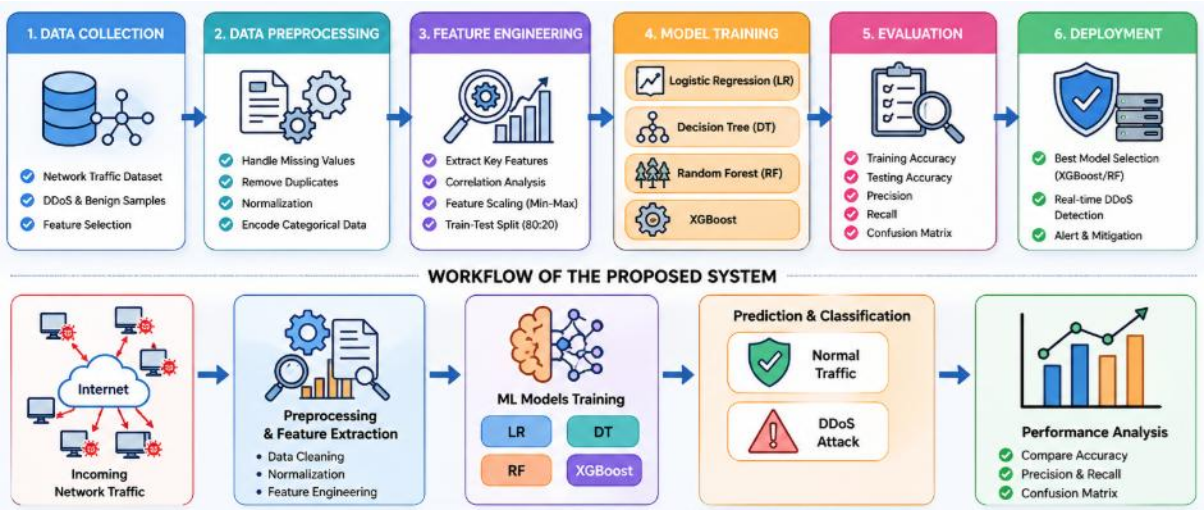

Figure 2: Flow Chart

Based on the evaluation results, the best-performing model—typically XGBoost or Random Forest due to their superior ability to handle complex and high-dimensional data—is selected in the optimized model selection stage. This model is then deployed for real-time prediction, where incoming network traffic is continuously monitored and classified as either normal or DDoS attack. If malicious activity is detected, the system can trigger alerts or initiate mitigation strategies to prevent service disruption. Overall, this methodology integrates data preprocessing, intelligent model selection, and performance optimization to build a robust, scalable, and efficient DDoS detection system capable of adapting to evolving cybersecurity threats.

4. RESULT ANALYSIS

Generally, the performance of a classification model is evaluated in terms of accuracy, sensitivity and specificity to calculate which the values of true positives (TP), true negatives (TN), false positives (FP) and false negatives (FN) need to be considered. A good machine learning model requires high accuracy and low false alarm rates. A confusion matrix is used to determine these parameters. In the confusion matrix, true positive is the number of normal records correctly identified as normal records; false positive is the number of normal records incorrectly identified as attacks; true negative is the number of attack records correctly identified as attacks and false negative is the number of attack records incorrectly identified as normal records.

International of Advanced Research and Multidisciplinary Journal Trends (IJARMT)
An International Open Access, Peer-Reviewed Refereed Journal
Impact Factor: 7.2 Website: https://ijarmt.com ISSN No.: 3048-9458

	duration	protocol_type	service	flag	src_bytes	dst_bytes	land	wrong_fragment	urgent	hot	...	dst_host_srv_count	dst_host_same_srv_rate	dst_host_diff_srv_rate	dst_host_same_src_port_
0	0	tcp	ftp_data	SF	491	0	0	0	0	0	...	25	0.17	0.03	
1	0	udp	other	SF	146	0	0	0	0	0	...	1	0.00	0.60	
2	0	tcp	private	S0	0	0	0	0	0	0	...	26	0.10	0.05	
3	0	tcp	http	SF	232	8153	0	0	0	0	...	255	1.00	0.00	
4	0	tcp	http	SF	199	420	0	0	0	0	...	255	1.00	0.00	
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
25187	0	tcp	exec	RSTO	0	0	0	0	0	0	...	7	0.03	0.06	
25188	0	tcp	ftp_data	SF	334	0	0	0	0	0	...	39	1.00	0.00	
25189	0	tcp	private	REJ	0	0	0	0	0	0	...	13	0.05	0.07	
25190	0	tcp	nnsp	S0	0	0	0	0	0	0	...	20	0.08	0.06	
25191	0	tcp	finger	S0	0	0	0	0	0	0	...	49	0.19	0.03	

25192 rows × 42 columns

Figure 3: Dataset

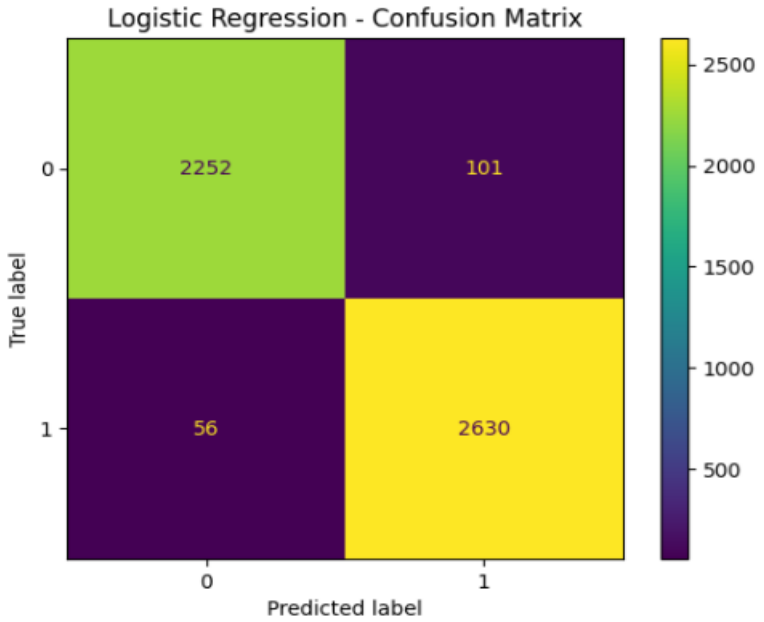

Figure 4: CM of LR

The confusion matrix for the Logistic Regression model illustrates its classification performance in distinguishing between normal traffic and DDoS attacks. In this matrix, the top-left value (2252) represents true negatives (TN), indicating that 2252 normal instances were correctly classified as normal. The top-right value (101) corresponds to false positives (FP), meaning 101 normal instances were incorrectly classified as DDoS attacks. The bottom-left value (56) represents false negatives (FN), where 56 DDoS attack instances were misclassified as normal, which is critical since undetected attacks can compromise system security. Finally, the bottom-right value (2630) shows true positives (TP), indicating that 2630 DDoS instances were correctly identified.

Overall, the model demonstrates strong performance, with a high number of correct predictions (both TN and TP) and relatively low misclassification rates. The small number of false negatives suggests that the model is effective in detecting most DDoS attacks, while the moderate number of false positives indicates some tendency to incorrectly flag normal traffic as malicious. This balance reflects a reliable classification model, although further optimization could reduce false alarms and improve precision without compromising recall.

International of Advanced Research and Multidisciplinary Journal Trends (IJARMT)
An International Open Access, Peer-Reviewed Refereed Journal
Impact Factor: 7.2 Website: https://ijarmt.com ISSN No.: 3048-9458

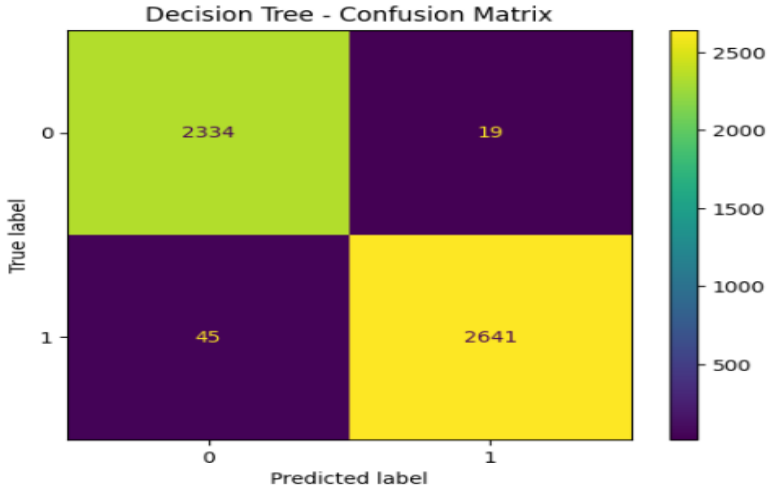

Figure 5: CM of DT

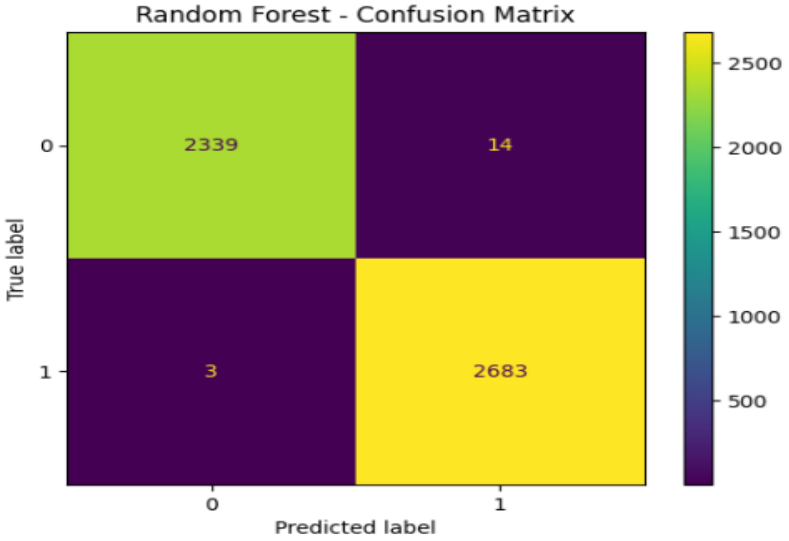

Figure 6: CM of RF

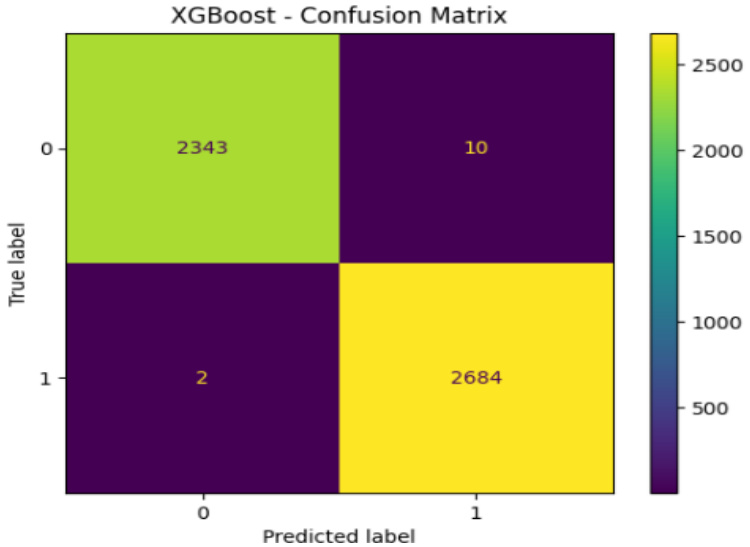

Figure 7: CM of XG Boost

International of Advanced Research and Multidisciplinary Journal Trends (IJARMT)
An International Open Access, Peer-Reviewed Refereed Journal
Impact Factor: 7.2 Website: https://ijarmt.com ISSN No.: 3048-9458

The given graph presents a comparative analysis of four machine learning models—Logistic Regression, Decision Tree, Random Forest, and XGBoost—based on their training and testing accuracy. From the visualization, it is evident that all models achieve very high accuracy, with values close to 1.0, indicating strong classification performance for DDoS attack detection. Logistic Regression shows slightly lower accuracy compared to the other models, though it still performs reliably with minimal difference between training and testing accuracy, suggesting good generalization. Decision Tree improves upon this with higher accuracy, but may still carry a slight risk of overfitting depending on model depth. Random Forest demonstrates near-perfect accuracy for both training and testing datasets, highlighting its robustness and ability to handle complex patterns effectively through ensemble learning. Among all models, XGBoost achieves the highest accuracy, with almost identical training and testing scores, indicating excellent generalization and minimal overfitting. Overall, the graph clearly shows that ensemble methods, particularly Random Forest and XGBoost, outperform traditional models, making them more suitable for accurate and reliable DDoS attack detection in cybersecurity applications.

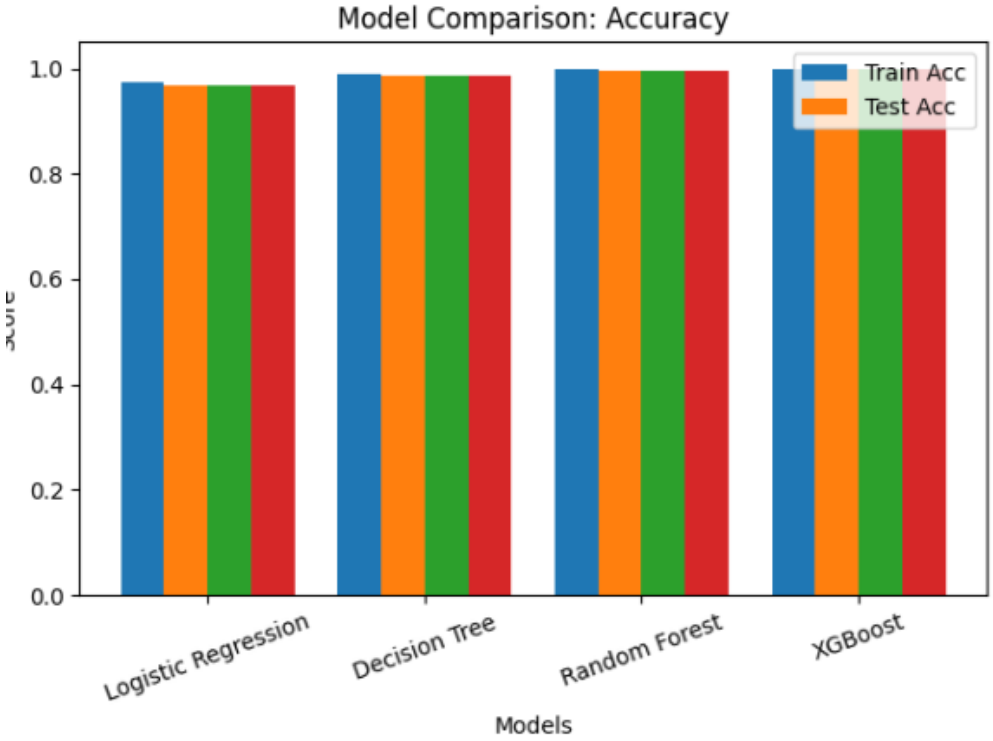

Figure 8: Accuracy

International of Advanced Research and Multidisciplinary Journal Trends (IJARMT)
An International Open Access, Peer-Reviewed Refereed Journal
Impact Factor: 7.2 Website: https://ijarmt.com ISSN No.: 3048-9458

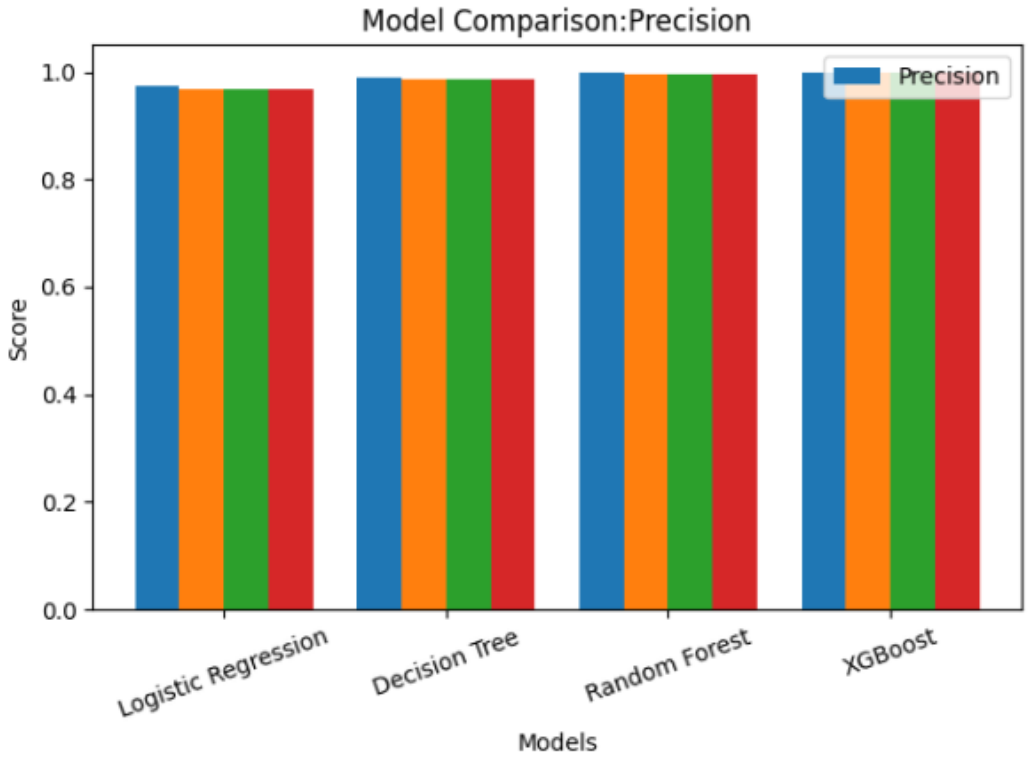

Figure 9: Precision

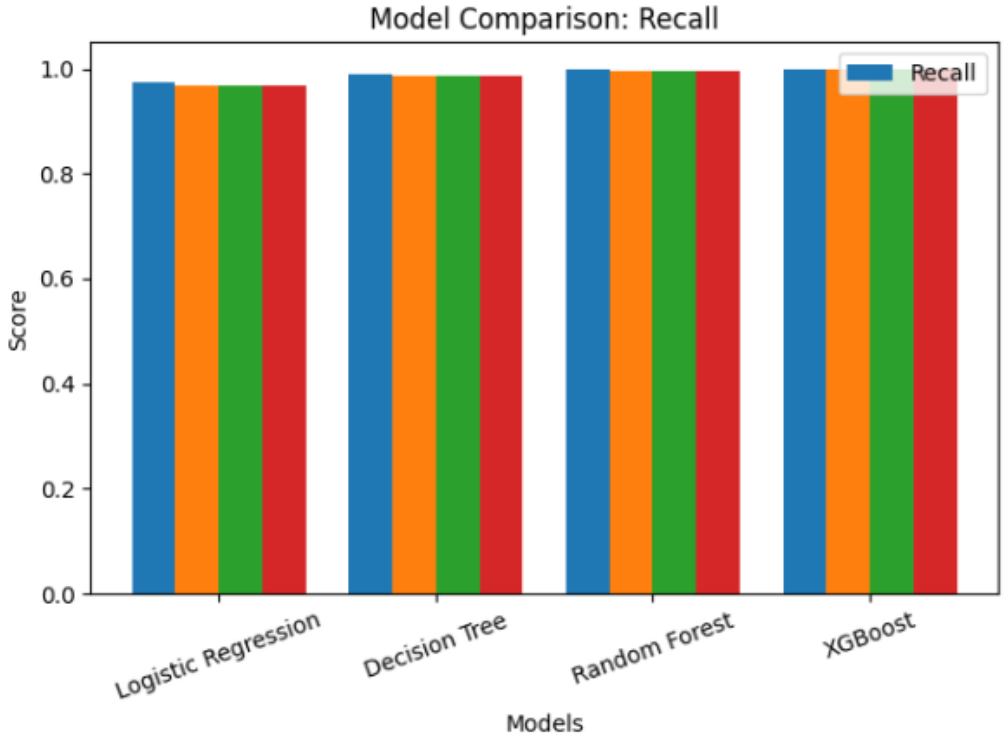

Figure 10: Recall

International of Advanced Research and Multidisciplinary Journal Trends (IJARMT)
An International Open Access, Peer-Reviewed Refereed Journal
Impact Factor: 7.2 Website: https://ijarmt.com ISSN No.: 3048-9458

5. CONCLUSION

This study presented an optimized machine learning-based framework for the detection and classification of Distributed Denial of Service (DDoS) attacks in cybersecurity environments. By employing and comparing four widely used algorithms—Logistic Regression (LR), Decision Tree (DT), Random Forest (RF), and Extreme Gradient Boosting (XGBoost)—the research aimed to enhance detection accuracy and overall system performance. The implementation involved effective data preprocessing, feature transformation, and model training to ensure reliable and consistent evaluation.

The experimental results demonstrated that while Logistic Regression and Decision Tree provide satisfactory baseline performance, they are limited in capturing complex and non-linear patterns present in network traffic data. In contrast, ensemble-based models such as Random Forest and XGBoost showed superior performance in terms of training accuracy, testing accuracy, precision, and recall. Among all the models, XGBoost achieved the highest overall accuracy and demonstrated strong generalization capability, making it the most effective approach for DDoS detection in this study.

Furthermore, the analysis highlighted the importance of optimizing model parameters and selecting relevant features to improve classification outcomes and reduce false positives and false negatives. The comparative evaluation confirmed that advanced ensemble techniques are better suited for handling large-scale, high-dimensional, and imbalanced datasets commonly found in cybersecurity applications.

In conclusion, the integration of optimized machine learning algorithms significantly enhances the capability of intrusion detection systems to identify and mitigate DDoS attacks in real-time. The proposed approach contributes to building robust, scalable, and intelligent cybersecurity solutions. Future work can focus on incorporating deep learning models, real-time deployment, and hybrid approaches to further improve detection performance and adapt to evolving cyber threats.

REFERENCES

[1] A. A. Alashhab, M. S. Zahid, B. Isyaku, A. A. Elnour, W. Nagmeldin, and A. Abdelmaboud, “Enhancing DDoS Attack Detection and Mitigation in SDN Using an Ensemble Online Machine Learning Model,” IEEE Access, vol. 12, pp. 51630–51649, Apr. 2024.

[2] A. Hussain, E. M. Tordera, X. Masip-Bruin, and H. C. Leligou, “Rule-Based With Machine Learning IDS for DDoS Attack Detection in Cyber-Physical Production Systems (CPPS),” IEEE Access, vol. 12, pp. 114894–114911, Aug. 2024.

[3] C. S. Shieh, F.-A. Ho, M.-F. Horng, T.-T. Nguyen, and P. Chakrabarti, “Open-Set Recognition in Unknown DDoS Attack Detection With Reciprocal Points Learning,” IEEE Access, vol. 12, pp. 56461–56476, Apr. 2024.

[4] S. Naiem, A. E. Khedr, A. M. Idrees, and M. I. Marie, “Enhancing the Efficiency of Gaussian Naïve Bayes Machine Learning Classifier in the Detection of DDoS in Cloud Computing,” IEEE Access, vol. 11, pp. 124597–124608, Oct. 2023.

[5] G. W. de Oliveira, M. Nogueira, A. L. dos Santos, and D. M. Batista, “Intelligent VNF Placement to Mitigate DDoS Attacks on Industrial IoT,” IEEE Trans. Network and Service Management, vol. 20, no. 2, pp. 1319–1331, Jun. 2023.

[6] K. Muthamil Sudar, M. Beulah and P. Deepalakshmi, “Detection of Distributed Denial of Service Attacks in SDN using Machine learning techniques”, International Conference on Computer Communication and Informatics (ICCCI), Jan. 27 – 29, 2021, Coimbatore, INDIA.

[7] Muthamil Sudar, K., & Deepalakshmi, P. (2020). A two level security mechanism to detect a DDoS flooding attack in software-defined networks using entropy-based and C4. 5 technique. Journal of High Speed Networks, (Preprint), 1- 22.

[8] Dong, S., & Sarem, M. (2019). DDoS Attack Detection Method Based on Improved KNN With the Degree of DDoS Attack in Software-Defined Networks. IEEE Access, 8, 5039-5048.

[9] Dong, S., Abbas, K., & Jain, R. (2019). A survey on distributed denial of service (DDoS) attacks in SDN and cloud computing environments. IEEE Access, 7, 80813-80828.

[10] Gu, Y., Li, K., Guo, Z., & Wang, Y. (2019). Semisupervised K-means DDoS detection method using hybrid feature selection algorithm. IEEE Access, 7, 64351- 64365.

[11] A. Raghavan, F. D. Troia, and M. Stamp, ``Hidden Markov models with random restarts versus boosting for malware detection," J. Comput. Virol. Hacking Techn., vol. 15, no. 2, pp. 97107, Jun. 2019.

[12] T. Young, D. Hazarika, S. Poria, and E. Cambria, ``Recent trends in deep learning based natural language processing [review article]," IEEE Comput. Intell. Mag., vol. 13, no. 3, pp. 5575, Aug. 2018.

[13] X. Lei and Y. Xie, ``Improved XGBoost model based on genetic algorithm for hypertension recipe recognition," Comput. Sci, vol. 45, pp. 476481, 2018.

[14] Y. Guo, Y. Liu, A. Oerlemans, S. Lao, S. Wu, and M. S. Lew, ``Deep learning for visual understanding: A review," Neurocomputing, vol. 187, pp. 2748, Apr. 2016.

[15] Abduvaliyev, A., Pathan, A.-S. K., Zhou, J., Roman, R., and Wong, W.-C. “On the Vital areas of Intrusion Detection Systems in Wireless Sensor Networks”, IEEE Communications Surveys & Tutorials, Vol. 15, Issue 3, pp. no. 1223–1237, 2015.