STM JOURNALS

International Journal of
Wireless Security and Networks

ISSN: 2584-0703
Volume 2, Issue 2, 2024
July–December
DOI (Journal): 10.37591/IJWSN

https://journals.stmjournals.com/ijwsn

Review IJWSN

Unified Ensemble Techniques for Enhanced DDoS Attack Prevention and Detection

Purushotam Naidu K.[1,*], Telukala Anitha[2], Inaganti Poshitha[3], Yetendriya Lamani K.[4], G. Satvika[5], G. Sireesha[6]

Abstract

Today's world is entirely reliant on the internet. The internet is a worldwide information source that all users rely on, hence its accessibility is critical. There have been reports in recent years, particularly in the information and technology division of significant organizations worldwide, of data breaches where the terms denial-of-service (DoS) and DDoS are consistently present in the stolen material. Network security is seriously threatened by DoS attacks. They have the ability to overburden a server or network, rendering it inaccessible to authorized users. DoS assaults are identified and stopped by means of network intrusion detection systems, or Network Detection Systems (NDS). However, because DoS assaults frequently employ dispersed and coordinated techniques, conventional network intrusion detection system (NIDS) have trouble identifying them. NIDS's ability to identify DoS attacks can be enhanced by the use of ensemble methods, a machine learning methodology. In order to provide a more accurate forecast, ensemble approaches integrate the predictions of numerous classifiers. They may thereby overcome the shortcomings of individual classifiers, making them ideal for identifying DoS assaults. The UNSW-NB15 dataset, developed by the Australian Cyber Security Centre in 2015, is used in this model's evaluation of the effectiveness of employing well-known ensemble techniques, including bagging, AdaBoost, stacking, decorate, Random Forest, and voting, to detect DoS attacks.

Keywords: Machine learning, ensemble classifier, stacking, AdaBoost, bagging, Distributed Denial of Service (DDoS), UNSW-NB15 dataset

INTRODUCTION

Distributed Denial of Service (DDoS) attacks are malevolent attempts to interfere with a targeted system or network's regular operation by flooding it with traffic from numerous sources [1]. Distributed Denial of Service (DDoS) assaults use the power of botnets—networks of hijacked devices under the direction of a central attacker—to increase the volume and intensity of the attack, in contrast to typical DoS attacks, which come from a single source. These attacks are especially difficult to counter because they use strength in number theory.

Distributed Denial of Service (DDoS) assaults can be carried out for a variety of reasons such as monetary gain, ideological disagreements, competitive advantage, or just the sheer desire to wreak havoc. Hacktivist groups, cybercriminal organizations, nation-states, and even disgruntled individuals may orchestrate Distributed Denial of Service (DDoS) attacks for diverse purposes ranging from extortion to political activism.

Developing successful mitigation strategies requires an understanding of the anatomy of DDoS

***Author for Correspondence**
Purushotam Naidu K.
E-mail: purushotam.k30@gmail.com

[1]Assistant Professor, Department of Computer Science and Engineering (Artificial Intelligence and Machine Learning), GVP College of Engineering for Women, Visakhapatnam, Andhra Pradesh, India
[2-6]Student, Department of Computer Science and Engineering (Artificial Intelligence and Machine Learning), GVP College of Engineering for Women, Visakhapatnam, Andhra Pradesh, India

Received Date: October 05, 2024
Accepted Date: October 22, 2024
Published Date: November 08, 2024

Citation: Purushotam Naidu K., Telukala Anitha, Inaganti Poshitha, Yetendriya Lamani K., G. Satvika, G. Sireesha. Unified Ensemble Techniques for Enhanced DDoS Attack Prevention and Detection. International Journal of Wireless Security and Networks. 2024; 2(2): 20–27p.

attacks [2]. Frequently used DDoS attack vectors are application-layer attacks that target particular applications or services to use up server capacity, protocol attacks, which use network protocol vulnerabilities to consume server resources, and volumetric attacks, which seek to saturate the bandwidth of the target. Every offensive style has different difficulties and calls for different defense strategies.

Despite advancements in DDoS mitigation technologies, combating these attacks remains an ongoing challenge. Organizations must adopt a multi-layered approach to defense, incorporating network infrastructure hardening, traffic filtering, rate limiting, and the deployment of specialized DDoS mitigation services. Collaboration among stakeholders, information sharing, and proactive monitoring are essential elements in the fight against DDoS attacks.

The availability and integrity of online services are seriously threatened by DDoS attacks, which makes it necessary to conduct ongoing studies, innovate, and work together to create strong defenses. We can lessen the effects of DDoS attacks and protect the digital ecosystem from malevolent interruptions by improving our knowledge of these attacks and implementing proactive security measures.

In the ever-evolving landscape of machine learning, where the pursuit of predictive accuracy and generalization performance reign supreme, ensemble methods have emerged as powerful tools to unlock the potential of collective intelligence within diverse models. Ensemble methods, a cornerstone of modern machine learning, represent a paradigm shift from conventional single-model approaches by leveraging the wisdom of crowds to enhance prediction robustness, stability, and accuracy.

Ensemble methods operate on the premise that diverse models, when judiciously combined, can collectively outperform any individual constituent model. The fundamental idea stems from the concept of the ‘wisdom of the crowd,’ wherein the aggregation of multiple perspectives leads to more informed decisions. In the context of machine learning, ensemble methods capitalize on this principle by aggregating predictions from multiple base models to generate a final prediction that often surpasses the performance of individual components.

One defining characteristic of ensemble methods is their ability to harness the complementary strengths of diverse models, thereby mitigating individual model weaknesses and biases. Techniques, such as bagging, boosting, and stacking, orchestrate collaboration among base learners, each contributing unique insights and perspectives to the collective decision-making process. Through intelligent aggregation strategies, ensemble methods effectively exploit the ‘wisdom’ embedded within an ensemble to produce superior predictive outcomes. Ensemble methods are shown in Figures 1–3.

Furthermore, ensemble methods offer robustness against outliers, noise, and data imbalances, making them particularly well suited for handling real-world datasets characterized by inherent complexities and uncertainties. By aggregating predictions from multiple models, ensemble methods exhibit remarkable resilience to perturbations in the data, thereby enhancing the stability and reliability of the overall prediction.

Distributed Denial of Service (DDoS) attacks, characterized by their distributed and coordinated nature, exploit vulnerabilities in network infrastructure to overwhelm targeted systems with a deluge of malicious traffic, rendering them inaccessible to legitimate users. Traditional detection methods often struggle to cope with the dynamic and sophisticated nature of Distributed Denial of Service (DDoS) DDoS attacks, necessitating more adaptive and resilient approaches to mitigate their impacts. Ensemble techniques rooted in the principle of collective decision-making and diversity present a compelling solution to this challenge.

Fundamentally, ensemble methods work by combining predictions from several detection models, each of which is trained using a distinct collection of characteristics, an algorithmic paradigm, or data

representation. Ensemble approaches improve detection performance, robustness, and generalization accuracy by utilizing the complementary characteristics of various models, increasing the effectiveness of DoS assault detection. Moreover, ensemble techniques exhibit inherent resilience against evasion tactics employed by attackers because they require adversaries to subvert multiple detection mechanisms simultaneously to evade detection effectively.

The application of ensemble techniques in DDoS attack detection encompasses various algorithmic paradigms, including but not limited to an ensemble of classifiers, feature representations, and detection strategies [3]. Techniques such as bagging, boosting, random forests, and stacking facilitate the integration of diverse detection models, each offering unique insights into the underlying patterns and anomalies that are indicative of DDoS attacks. Through intelligent aggregation strategies, ensemble techniques enable the synthesis of more accurate and reliable detection decisions, thereby empowering cybersecurity practitioners to proactively identify and mitigate DDoS attacks.

An ensemble learning model was constructed by including five different types of machine learning models as learners. Such Random Forest, decision tree, logistic regression, Gaussian Naïve Bayes, and K-nearest neighbors algorithms are models that can be used for Distributed Denial of Service (DDoS) attacks, as shown in Figure 4.

LITERATURE SURVEY

Ensemble methods are used to detect DoS attacks in network intrusion detection systems [4]—This study assessed the effectiveness of ensemble methods for identifying denial-of-service attacks on the UNSW-NB15 dataset, including voting, decorating, stacking, AdaBoost, and bagging. The best single classifier, Random Forest, achieved 99.02%, but stacking, which uses heterogeneous classifiers, provided the highest classification quality with an F-measure of 99.28%.

DDoS attack and detection methods in internet-enabled networks [5]—For nascent researchers, this survey offers an overview of attack detection techniques in Internet of Things (IoT) applications, encompassing conventional, machine learning, and deep learning approaches. The following datasets were used: KDD'99, DARPA'99, CICDDoS2019, and ISCX2012. This study investigated anomaly-based techniques, such as entropy-based and queue modeling detection. One strategy that combined Autoregressive Integrated Moving Average (ARIMA) with a chaotic system produced a 94.4% true positive rate and provided network managers with early warning capabilities.

DDoS attacks and machine learning based detection methods—A survey and taxonomy [6]. This assessment highlights the advantages and disadvantages of the dataset and provides a taxonomy of machine learning based techniques for DDoS attack detection, including supervised, unsupervised, and hybrid approaches [7]. To improve cybersecurity against these various and serious network attacks, recommendations are made for future research topics and attempts to assist researchers in understanding DDoS detection techniques.

DDoS intrusion detection through machine learning ensemble [8]—To improve robustness against a variety of intrusion types, this research provides a network intrusion detection system (NIDS) that detects both current and emerging DDoS attacks by combining several classifiers through ensemble models. By utilizing the NSL-KDD dataset, the NIDS surpasses previous techniques in terms of accuracy and flexibility for novel attack patterns, achieving a 99.1% detection rate with a smaller feature set.

Effective network intrusion detection using stacking [9] by addressing the shortcomings of anomaly-based NIDS datasets, this study produced a new dataset called CIPMAIDS2023-1. A suggested ensemble strategy based on stacking was evaluated using CICFlowMeter-extracted features in a GNS-3 simulated network under a range of assault scenarios. With a weighted F1-score of 98.24%, the strategy exceeds the current approaches and demonstrates its usefulness for NIDS.

Novel ensemble learning based model for network intrusion detection [10]. A new dataset, CIPMAIDS2023-1, was created as a result of this study, which overcomes the shortcomings of anomaly-based NIDS datasets. Using features retrieved by a CICFlowMeter, an ensemble technique based on stacking was presented and evaluated for multiple assault scenarios in a GNS-3 simulated network. The strategy outperforms current techniques, obtaining a weighted F1-score of 98.24%, proving its efficacy for NIDS.

This study used the CIC-IDS2017 dataset and K-fold cross-validation for training and testing eight supervised machine learning algorithms for DDoS attack detection. Accuracy, precision, recall, and false alarm rates are examples of performance measures [11]. According to the results, Random Forest is the most successful model for early-stage DDoS detection because it performs better than the other models on all criteria**.** It produced 99.885% accuracy, 99.88% precision, 100% recall, and 0.05% false alarm rate to detect DDoS attacks.

This study distinguishes between DoS and DDoS attacks, emphasizing that the latter frequently uses botnets as their primary source. Using K-fold cross-validation and the CIC-IDS2017 dataset, it assesses eight supervised machine learning methods for DDoS attack detection [12]. The Random Forest model is the best option for early DDoS attack detection, according to the results, which show that it performs better than the other models in terms of accuracy, precision, recall, and false alarm rate, with 99.885% accuracy, 99.88% precision, 100% recall, and 0.05% false alarm rate to detect DDoS attacks at the earliest.

To detect denial-of-service (DoS) assaults, we used two machine learning techniques, Random Forest (RF) and MultiLayer Perceptron (MLP), using the Scikit ML library and the large data framework Spark ML library. KDD, and NSL-KDD datasets. 99.5 Of the models, 99.5% had a mean accuracy that we were able to achieve, both with and without big data approaches [13].

METHODOLOGY

Bagging Classifier

An ensemble technique called bootstrap aggregation, or "bagging," was created to address the problem of overfitting in classification or regression problems, thereby increasing the precision and efficiency of machine learning algorithms [14]. Replacement is used to fit a classifier (for classification tasks) or regressor (for regression tasks) to each random subset of the original dataset. To improve the overall prediction accuracy, the results from each subset were then combined, either by averaging for regression or by majority voting for classification. Random Forest is a prime example of a Bagging Classifier, employing multiple decision trees trained on different subsets, reducing overfitting and increasing robustness by combining diverse predictions, as shown in Figure 1.

AdaBoost Classifier

An ensemble learning algorithm called Adaptive Boosting (AdaBoost) combines several weak learners to produce a powerful classifier. It operates by sequentially training weak classifiers, with each new classifier focused on instances that were misclassified by the preceding ones. Each weak learner is assigned a weight by the algorithm based on how accurate it is; the alpha parameter, which is inversely proportional to the error rate of each learner, shows how much each learner contributed to the final model. The model performed better overall and was more resilient to misclassifications because of this adaptive strategy. AdaBoost is versatile, effective in handling various types of data, and tends to perform well, even with simple base classifiers.

Random Forest

Using random selections of data and characteristics, Random Forest is an ensemble learning technique in machine learning that creates several decision trees. The prediction performance is improved, and overfitting is decreased by this randomness. To make predictions, the algorithm averages the data from each tree or uses majority voting for classification and regression. Random Forest is a

popular technique for both classification and regression tasks in complex datasets owing to its reliability and accuracy [15]. This is illustrated in Figure 2.

Stacking Classifier

Stacking combines multiple diverse base models by training one meta-learner (another model) on their predictions. The base models were trained separately, and their outputs served as inputs for the meta-learner. Stacking leverages the diverse predictions of individual models, allowing the meta-learner to learn when to trust the output of each base model. It aims to capture complementary strengths from different models, potentially improving overall predictive performance and generalization. This is illustrated in Figure 3.

These methods, bagging, AdaBoost, Random Forest, and stacking, represent diverse approaches within ensemble learning, each with its unique way of leveraging multiple models to enhance predictive accuracy and robustness in different scenarios, as shown in Figure 4.

Majority Voting

Voting ensembles are a class of machine learning models that incorporate predictions from many other models. The improved models can outperform any one model in an ensemble using this method. The predictions of the multiple models were combined in an ensemble voting system.

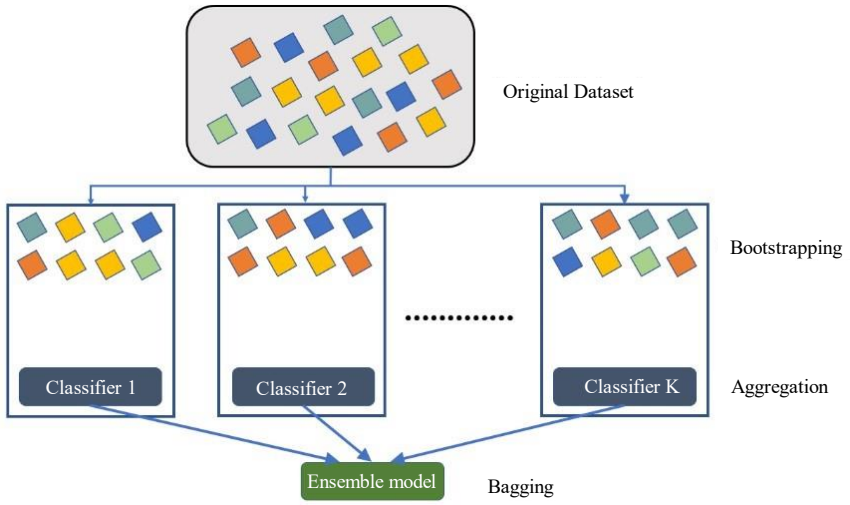

Figure 1. Bagging classifier.

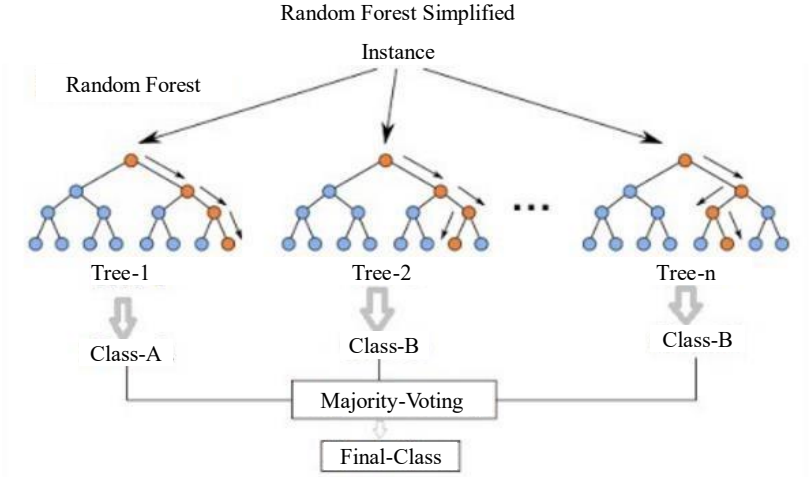

Figure 2. Random Forest classifier.

Classification or regression is accomplished by averaging the forecasts from each model, which is accomplished by regression. A total of predictions was collected for each label, and categorization was assigned to the label with the most votes. It is possible to consider voting ensembles in the form of a meta-model. Every trained machine learning model can be utilized in an ensemble with its meta-model without the model, even when it is in the ensemble.

RESULTS AND DISCUSSION

A stacking classifier was created using the stacking classifier class from scikit-learn. This classifier combines multiple base estimators; in this case, Random Forest classifiers, and then a final estimator (another Random Forest classifier) is trained on the outputs of these base estimators. Approximately 97.89% of the model predictions were accurate, with an accuracy of approximately 97.89%. The percentage of true positive forecasts among all positive predictions is known as precision. This means that when the model predicts a favorable result, it is accurate approximately 97.42% of the time. The percentage of true positive predictions among all real positive events was measured using recall. This was approximately 98.33%, indicating that the model correctly identified approximately 98.33% of the actual positive instances. The harmonic mean of recall and precision was the F1 score. It offers a single metric that strikes a balance between recall and precision to assess a model's performance. The F1 score was approximately 97.87%, as shown in Table 1.

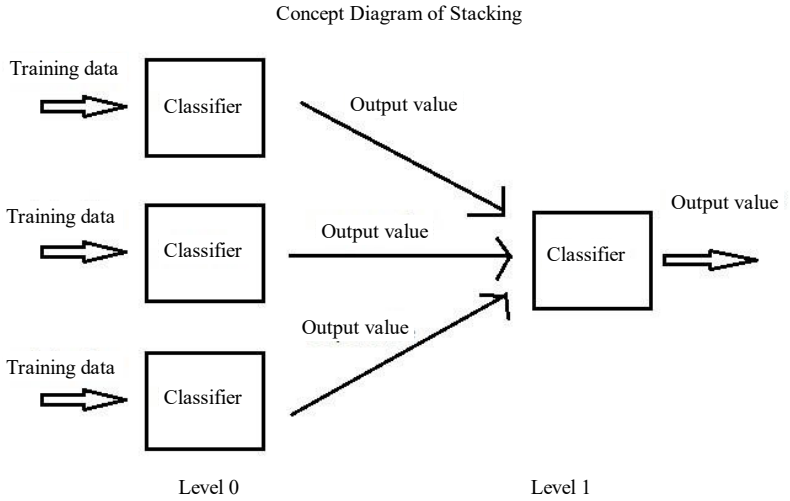

Figure 3. Stacking classifier.

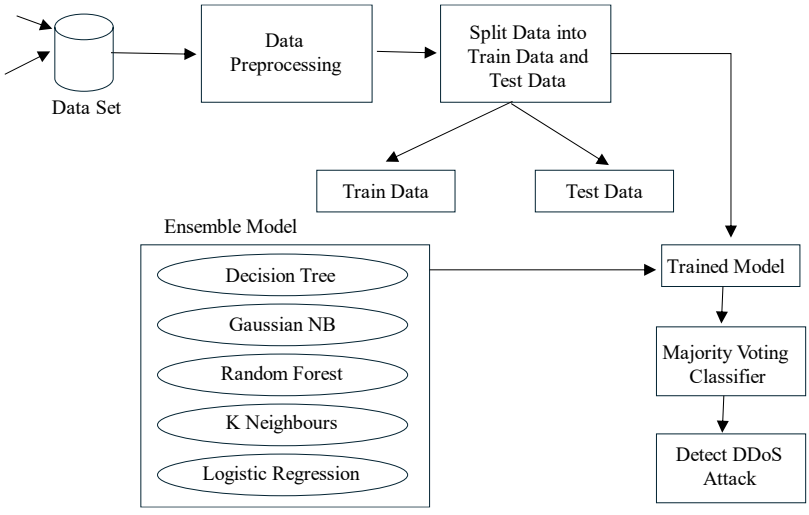

Figure 4. System architecture.

Table 1. Models' accuracy in %.

Model	Accuracy	Precision	Recall	F1 score
Decision tree	96.57	96.51	96.60	96.56
Gaussian Naïve Bayes	71.30	77.31	68.86	72.84
Random Forest	97.89	97.42	98.32	97.87
K-nearest neighbors	82.26	78.85	84.49	81.57
Logistic regression	96.57	96.51	96.60	96.56

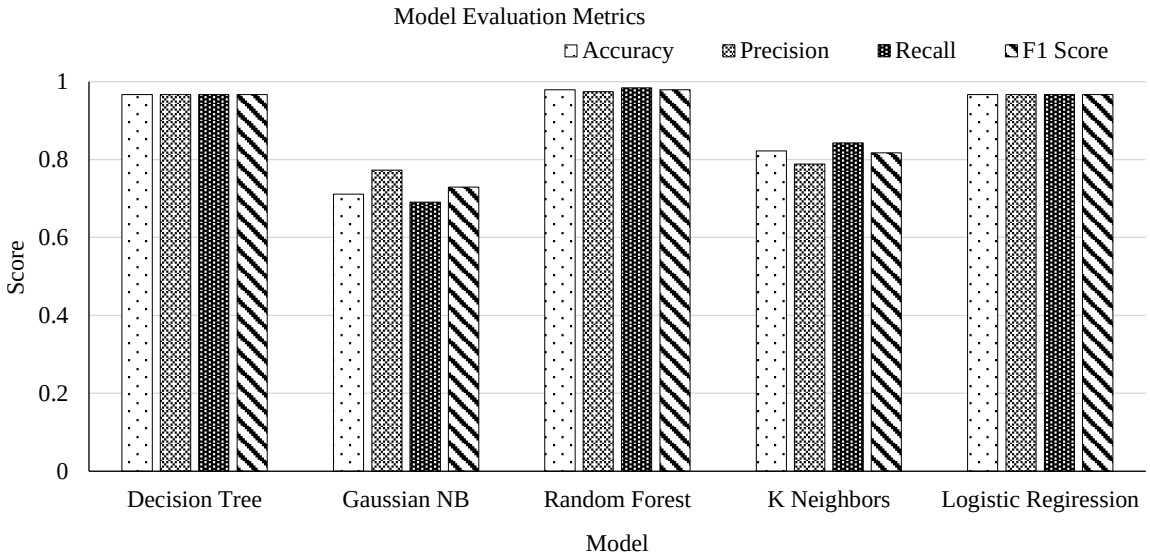

Figure 5. Algorithms accuracy, precision, recall, and F1 score.

The code trains and evaluates a stacking classifier using logistic regression as both the base and final estimators. The accuracy of this ensemble model was approximately 96.58%. The percentage of correctly classified instances relative to all instances is known as accuracy. The trains evaluate a stacking classifier using K-nearest neighbors (KNN) as both the base and final estimators. The accuracy of this ensemble model was approximately 82.26%. The percentage of correctly classified instances relative to all instances is known as accuracy. Precision, which is defined as the percentage of actual positive forecasts among all positive predictions, was approximately 78.85%. The recall is approximately 84.50%, meaning that it represents the percentage of accurate positive predictions among all actual positive instances. The F1 score, representing the harmonic mean of precision and recall, was approximately 81.58%.

AdaBoost (Adaptive Boosting) is an ensemble technique that focuses on sequentially training weak learners and adjusting weights to emphasize misclassified instances, thereby improving the overall performance. While logistic regression is a linear model, AdaBoost improves its performance by adjusting weights to focus on misclassified instances. AdaBoost has proven to be effective in improving the accuracy of models, especially those prone to high bias or underfitting, such as decision trees and Naïve Bayes. The AdaBoost Accuracy of the Decision Tree was 96.67%, the AdaBoost Accuracy of Naïve Bayes was 49.35%, the AdaBoost Accuracy of Logistic Regression was 76.42%, and the AdaBoost Accuracy of Random Forest was 97.17. The accuracy of the proposed model was compared with five different machine algorithms, and the results are shown in Figure 5, which shows that the proposed ensemble Random Forest model is more accurate.

CONCLUSION

In this work, a machine learning ensemble model was proposed for DDoS attacks using the ensemble technique, and five different algorithms, Random Forest, decision tree, logistic regression, Gaussian

Naïve Bayes, and KNN algorithms, were used to achieve the best results. The accuracy of the proposed model was compared with five different machine algorithms, and the proposed ensemble Random Forest model was found to be more accurate.

Research Contributions

This article presents an ensemble-based machine learning model for detecting DDoS attacks by leveraging techniques such as bagging, AdaBoost, stacking, and Random Forest. This demonstrates that ensemble methods outperform single classifiers by combining diverse models and achieving higher accuracy, precision, and robustness against DDoS threats using the UNSW-NB15 dataset.

REFERENCES

1. Rahamathullah U, Karthikeyan E. Distributed denial of service attacks prevention, detection and mitigation – A review. Proceedings of the International Conference on Smart Data Intelligence (ICSMDI 2021). 2021 May 25. p. 16. DOI: 10.2139/ssrn.3852902. Available at SSRN: https://ssrn.com/abstract=3852902.
2. Mittal M, Kumar K, Behal S. Deep learning approaches for detecting DDoS attacks: A systematic review. Soft Comput. 2022;1–37. DOI: 10.1007/s00500-021-06608-1.
3. Awan MJ, Farooq U, Babar HMA, Yasin A, Nobanee H, Hussain M, et al. Real-time DDoS attack detection system using big data approach. Sustainability. 2021;13:10743. DOI: 10.3390/su131910743.
4. Thanh HN, Van Lang T. Use the ensemble methods when detecting dos attacks in network intrusion detection systems. EAI Endorsed Trans Context-Aware Syst Appl. 2019;6(1–7). DOI: 10.4108/eai.29-11-2019.163484.
5. Adedeji KB, Abu-Mahfouz AM, Kurien AM. DDoS attack and detection methods in internet-enabled networks: Concept, research perspectives, and challenges. J Sensor Actuator Netw. 2023;12:51. DOI: 10.3390/jsan12040051.
6. Najafimehr M, Zarifzadeh S, Mostafavi S. DDoS attacks and machine-learning-based detection methods: A survey and taxonomy. Eng Rep. 2023;5:e12697. DOI: 10.1002/eng2.12697.
7. Prriyadarshini MA, Renuka Devi SR. Detection of DDoS attacks using supervised learning technique. J Phys Conf Ser. IOP Publishing; 2020;1716. DOI: 10.1088/1742-6596/1716/1/012057.
8. Das S, Mahfouz AM, Venugopal D, Shiva S. DDoS intrusion detection through machine learning ensemble. 2019 IEEE 19th International Conference on Software Quality, Reliability and Security Companion (QRS-C), Sofia, Bulgaria. 2019. p. 471–7. DOI: 10.1109/QRS-C.2019.00090.
9. Ali M, Haque MU, Durad MH, Usman A, Mohsin SM, Mujlid H, et al. Effective network intrusion detection using stacking-based ensemble approach. Int J Inf Secur. 2023;22:1781–98. DOI: 10.1007/s10207-023-00718-7.
10. Thockchom N, Singh MM, Nandi U. A novel ensemble learning-based model for network intrusion detection. Complex Intell Syst. 2023;9:5693–714. DOI: 10.1007/s40747-023-01013-7.
11. Nalayini CM, Katiravan J, Geetha S, Ji CE. A novel dual optimized IDS to detect DDoS attack in SDN using hyper tuned RFE and deep grid network. Cyber Secur Appl. 2024;2:100042. DOI: 10.1016/j.csa.2024.100042.
12. El-Sofany HF, El-Seoud SA, Taj-Eddin IATF. A case study of the impact of denial of service attacks in cloud applications. J Commun. 2019;14(2):153–8. doi:10.12720/jcm.14.2.153-158.
13. Sharma N, Gupta KD. Everything on DDoS attacks, DDoS incidents & DDoS defense mechanisms! Turk J Comput Math Educ. 2020;11:608–16.
14. Nalayini M, Katiravan J. Detection of DDoS Attack Using Machine Learning Algorithms. J Emerg Technol Innov Res. 2022 Jul;9(7). Available at SSRN: https://ssrn.com/abstract=4173187.
15. Alotaibi Y, Ilyas M. Enhancing IoT attack detection through ensemble-based multiclass attacks classification. 2023 IEEE 20th International Conference on Smart Communities: Improving Quality of Life using AI, Robotics and IoT (HONET), Boca Raton, FL, USA. 2023. p. 1–6. DOI: 10.1109/HONET59747.2023.10374636.

