Detection of DDOS Attack Using Machine Learning

Supraja.M[1], Ankitha.M[2], Mohammed Sadiya[3], Harshitha.S[4] and B Revathi[5*]

[1,2,3,4]Student,Department of CSE(AIML),CMR Engineering College,Hyderabad,Telangana

[5]Asst Professor,Department of CSE(AIML),CMR Engineering College,Hyderabad,Telangana

ABSTRACT:*The increasing number of internet users globally has posed a huge threat to security of internet resources, and the spreading nature of Distributed Denial of Service (DDOS) attacks has made the threat rampant. This study, therefore, focused on the need to detect DDOS attacks using machine learning techniques. K-Nearest Neighbour, Logistic Regression, and Random Forest models, among other tools, could detect DDOS attacks under the NSL KDD dataset to 98% initially. Finally, aiming to leverage these outcomes, more sophisticated ensemble models, for instance, the Voting Classifier with RF + Adaboost and Stacking Classifier, were investigated. As a result, they demonstrated a 100% error accuracy in DDOS detection. This research could thus contribute to improving security by improving the detectability of DDOS incidents*

Keywords: *Distributed Denial of Service (DDOS), K-Nearest Neighbour, Logistic Regression, Random Forest, Voting Classifier(RF+Adaboost), Stacking Classifier.*

I. INTRODUCTION

The Distributed Denial of Service attacks, DDoS, remain a considerable threat to the security of the network. Attackers are targeting the critical infrastructure and disturbing online services with these attacks, which are triggered with stylish DoS attacks overrun the boundary of a solitary system's effectiveness. However, the conventional mechanisms of DDoS detection are frequently ineffective in determining when and how these attacks will emerge. Given the challenges in which accurate DDoS identification can be a difficult goal to achieve. This study broadenens the scope in which the aim of DDoS detection isn't the truth but instead achieves by including ensemble agency. Ensemble methods, such as the Voting Classifier via Random Forest and Adaboost classifier, and the Stacking classifier provide a method of combining multiple separate agents and allow us to accomplish looking for the DDoS repercussion consistently.

Based on the studies, we build on this work to widen the research scope and utilize ensemble methods to enhance the detection accuracy. Ensemble techniques like the Voting Classifier with Random Forest and Adaboost and the Stacking Classifier provide the ability to enhance the prediction of DDoS detection by merging the advantages of several individual models into a more dependable and accurate prediction

Apart from improving the detection performance, this research also emphasizes the usability of the detection system. Utilizing the Flask framework to develop the front-end interface makes the platform more readily available to users for testing and validating the detection system. With the incorporation of user authentication servers, the system can be assessed securely, and DDoS attacks can be evaluated and monitored effortlessly. This research strives to contribute to the improvement of DDoS detection by integrating ensemble techniques and implementing a user-friendly front-end interface for improved usability and accessibility.

II. RELATED WORK

Author in [1] proposed a Classification model to detect DDoS Attacks using Machine Learning Algorithms. The work in the literatures presented the various existing work of machine learning-based approaches for detecting DDoS attack. The existing models used the Classification methods. That quantitative research applies Logistic Regression, Decision Tree, Random Forest, Ada Boost, Gradient Boost, KNN and Navie Bayes Classification algorithms to detect DDoS attack on the CIC-DDoS2019 dataset, the Experimental results show that AdaBoost and Gradient Boost algorithms give the best classification results, Logistic Regression, KNN, and Navie Bayes give good classification results, decision Tree and Random Forest poor Classification results.

The Result in research proved that machine learning algorithms detect DDoS attacks accurately in very less time by evaluating the effectiveness of the classification algorithms for detecting DDoS attacks on the CICDDoS2019 dataset.

Author [2] It provides a comparison of various machine learning algorithms for detecting DDoS attacks. The study seeks to investigate how well different machine learning techniques can work when it comes to detecting and mitigating DDoS attacks in real time network environments. The authors also discuss the difficulty of distinguishing between authorized traffic and DDoS attack traffic, focusing on the incapability of classifying streaming data in real-time due to large memory requirements. This is complemented by stressing that there is a need for designing an integrated autonomous model that will be responsible for identifying different kinds of intrusions or network anomalies in future.

Author in[3] proposed a model that is effectively identify and mitigate DDoS attacks,machine learning techniques have been extensively employedin intrusion detection systems.Machine learning models or approaches offer the advantage of automating the deteection process by learning patterns and characteristics of DDoS attacks from dataset.Many researchers have explored various machine learning algorithms such as KNN,SVM,Random forest and navis bayes to classify and detect DDoS attcaks.These algorithms leverage featuresextracted from network traffic dataa,including properties of network and traffic behaviour to differentiate between normal and malicious traffic.

And also it examines how machine learning techniques including Naive Bayes, Support Vector Machines (SVM), K-Nearest Neighbors (KNN), Random Forest, Convolutional Neural Network (CNN), Logistic Regression, ADA Boost, and Decision Tree perform as far as discerning DDos attacks are concerned, Moreover, it acknowledges that while machine learning holds great promise for improving DDoS detection capabilities, there are still significant challenges that require more research.

III. METHODOLOGY

The revised NSL KDD dataset is used to classify attacks scenarios. The ML Models used in the proposed system are Logistic Regression, KNN and Random forest and voting classifier (RF+AdaBoost) and stacking classifier (RF+MLP with LightGBM) and we are using open-source library Scikit-learn, numpy and pandas to validate the model and compare the results.

The dataset contains 12 important features namely Service(set of functionalities provided by servers, or applications to fulfill specific requirements within a network) Flag(refers to a special attribute within a network protocol or packet header) Src_bytes(number of bytes transmitted by the source) dst_bytes(total number of bytes received by the destination entity) count(refers to the number of occurrences of a particular event, attribute within a specified time period) Serror_rate(which is a metric used to measure the rate of connection attempts that resulted in an error originating from the source) Same_srv_rate(refers to the rate of connections to the same service among all connections initiated by the source) diff_srv_rate(refers to the rate of connections to different services among all connections initiated by the source) Dst_host_srv_count(measures the total number of unique services or ports available on the destination host)dst_host_srv_rate(refers to the rate of connections to services on the destination host among all connections initiated by the source host) dst_host_diff_srv_rate(measures the proportion of connections initiated by the source host to different services) dst_host_serror_rate(refers to the rate of connections that resulted in an error at the destination host among all connections initiated by the source host to that destination host)error_rate(rate of errors encountered during network communication or data transfer)

The two level filtering is done on the data to fit it as per the requirements of model. The filtering later organized as pre-processing and Exploring of dataset. Exploring a dataset typically involves conducting various analyses and visualizations to gain insights into its structure, patterns, and relationships between variables. After loading the dataset we need to understand dataset and Review the dataset's metadata, including column names, data types, and descriptions, and we used functions like info() and describe() to get summary statistics and information about the dataset's structure and we need to Handle Missing Values by Check for missing values in the dataset using functions like isnull() or isna().Decide on a strategy for handling missing values, such as imputation or removal, based on the nature of the data and the extent of missingness.

We need use the Label encoding is a process of converting categorical labels into numerical labels that can be provided to machine learning algorithms for training. And next Feature selection is a crucial step in building a machine learning model and data analysis where you choose a subset of relevant features (variables, predictors) from our dataset to build out model. we are working on Correlation-based Feature Selection by considering both the correlation between features and the class label and the correlation between features themselves, CFS aims to select features that are not only individually predictive but also complementary to each other. This can help reduce redundancy and overfitting in the feature set, leading to better generalization performance of the machine learning model. Splitting the data to train and test we are using 70-80% to training set and we allocated 20-30% to testing set.

In this case, we used different classifiers namely Logistic Regression, Random Forests, and K-Nearest Neighbor is conducted to showcase the performance of the algorithms. As an extension we applied an ensemble method combining the predictions of multiple individual models to produce a more robust and accurate final prediction.

However, we can further enhance the performance by exploring other ensemble techniques such as Voting Classifier with RF + Adaboost and Stacking Classifier which got 100% accuracy, as an extension we can build the front end using the flask framework for user

testing and with user authentication. Furthermore, it includes the Flask web framework and SQLite, for user authentication. Users input network traffic. Specific scenarios for DDoS analysis. The input is preprocessed to match the trained machine learning model involving data cleaning and feature extraction. The processed data is then inputted into the model to predict network behavior or detect a DDoS attack. Users can view prediction results, on the web interface offering insights and mitigation recommendations. To access predictions users must sign up. Log in for system entry. Security measures safeguard user data.

IV.SYSTEM DESIGN

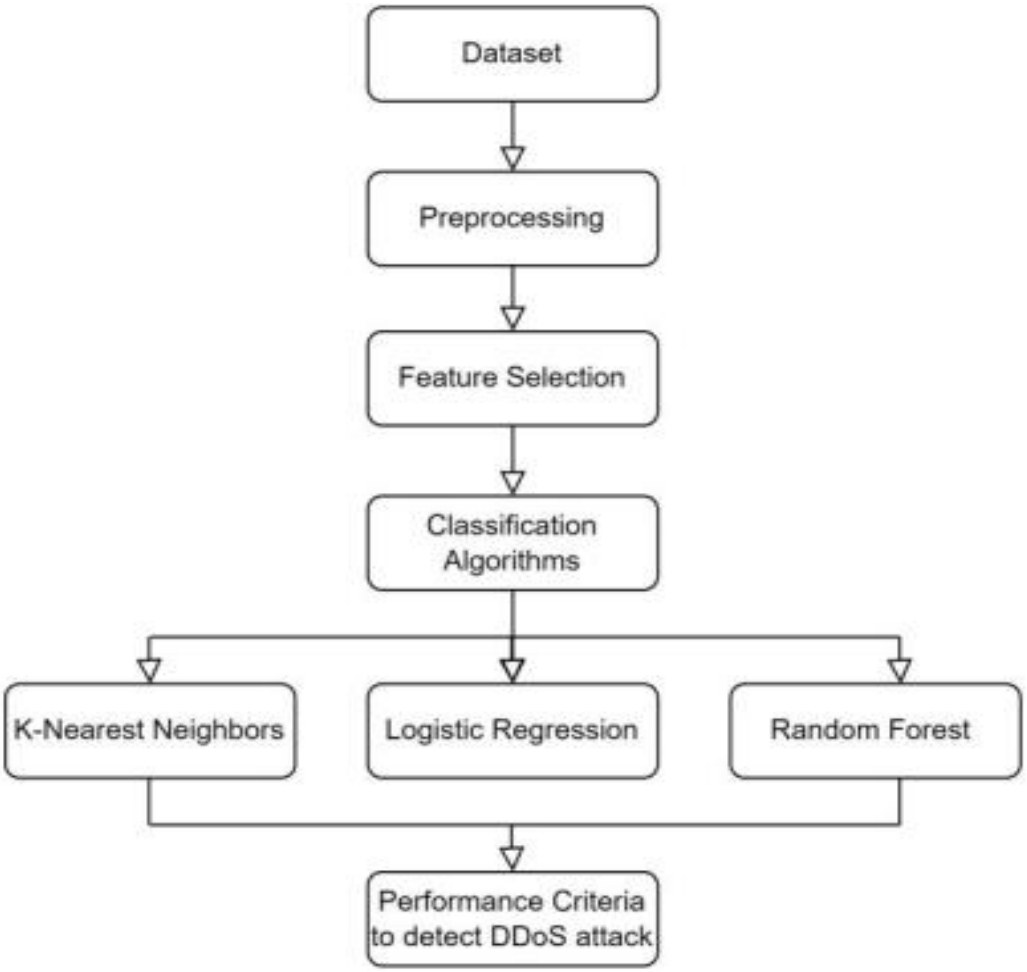


Figure 1.System design

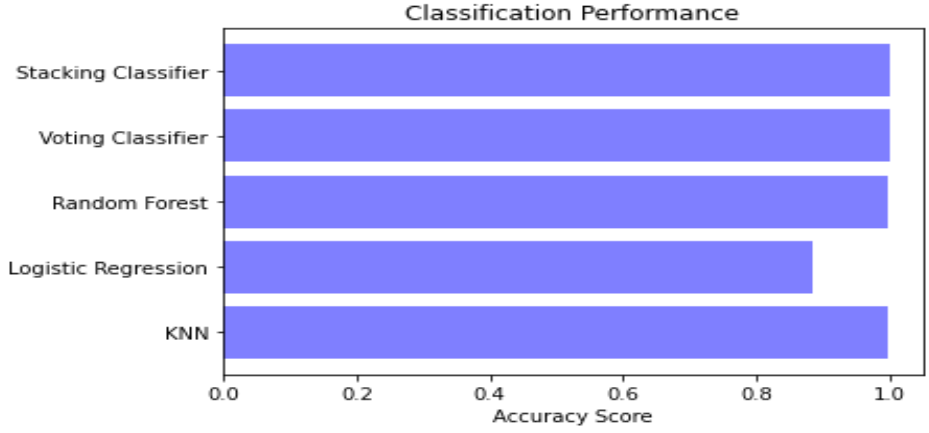
V. RESULT

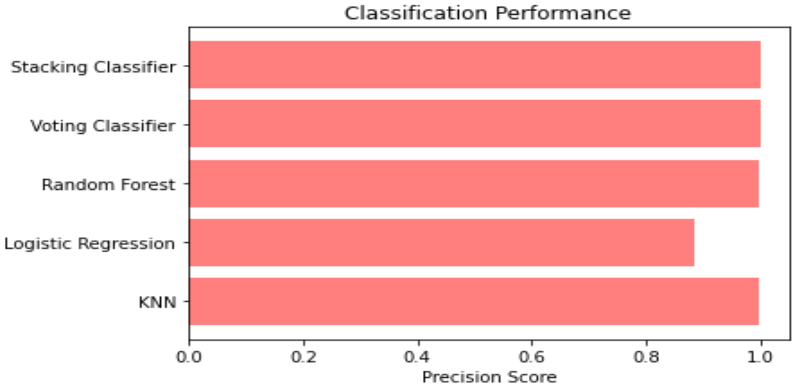
The dataset that contains all the network properties or features such as packet sizes, frequencies, protocols and other relevant information. This model is to detect whether there is DDOS attack is there not. The result of this detection could be presented in terms of accuracy, recall, precision and F1 score metrics. According to algorithms we use three algorithms which are random forest, knn, logistic regression and we use voting classifier which says the highest probability of the algorithms and stacking classifier.

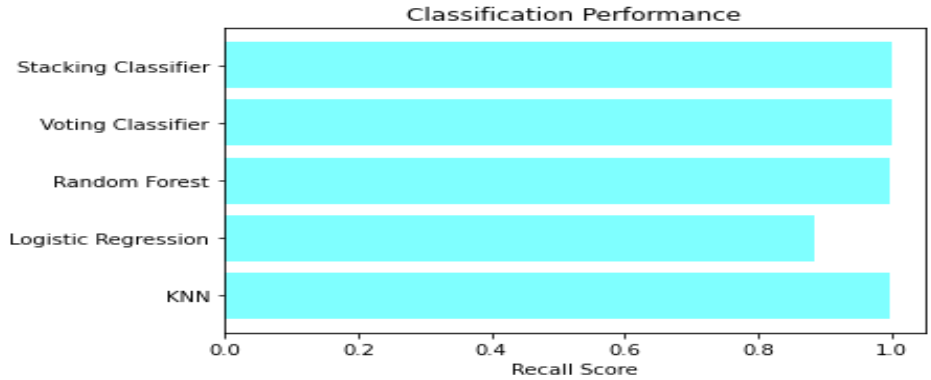
The outcome of the project and model is to detect the ddos attack. This suggests a comprehensive analysis of different algorithms (Random forest, knn, logistic regression), and ensemble (voting classifier and stacking classifier) methods which might provide a more well-rounded understanding of the DDos detection performance.

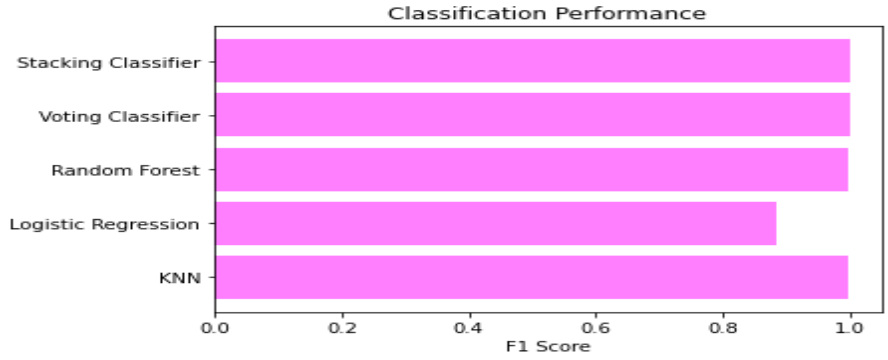
ML Model	KNN	Logistic Regression	Random Forest	Voting Classifier	Stacking Classifier
Accuracy	0.997	0.883	0.998	1.000	1.000
Precision	0.997	0.885	0.998	1.000	1.000
Recall	0.997	0.883	0.998	1.000	1.000
F1-score	0.997	0.884	0.998	1.000	1.000

The ensemble methods (voting classifier and stacking classifier outperform individual classifiers with perfect accuracy, precision, recall and F1 score. RF and knn performs exceptionally well, while logistic regression showed slightly lower performance.









V.CONCLUSION AND FUTURE SCOPE

The DDoS attack, which uses a very potent technique to attack network devices and services, is typically understood as one of the most serious attacks. As a result, we explore testing, analysing and developing a machine learning model to detect DDoS attacks in this work. The most important features that can be utilized to accurately forecast DDoS attacks were chosen using a variety of feature selection techniques in this. Three machine learning algorithms have been used with selected features from the dataset. According to the results, KNN and Random forest exhibit the best performance, while logistic regression only obtains less accuracy. There by enhancing the security posture of organizations and minimizing the impact of such cyber threats. Addressing adversarial threats through robust defense mechanisms and exploring real-time detection and response strategies are critical to ensuring the resilience of detection systems in dynamic network environments. We intend to use real-

time DDOS detection tools in our upcoming work so that we can identify DDoS attacks in real time.

REFERENCES

[1] Kishore Babu Dasari and Nagaraju Devarakonda," Detection of DDoS Attacks Using Machine Learning Classification Algorithms", I.J. Computer Network and Information Security, 2022, 6, 89-97.

[2] Zerin Hasan Sahosh, Azraf Faheem, Marzana Bintay Tuba, Md. Istiaq Ahmed, and Syeda Anika Tasnim.A Comparative Review on DDoS Attack Detection Using Machine Learning Techniques. Malaysian Journal of Science and Advanced Technology.2023.

[3] Mahmood A. Al-Shareeda, Selvakumar Manickam, Murtaja Ali Saare,"DDoS attacks detection using machine learning and deep learning techniques: analysis and comparison". Bulletin of Electrical Engineering and Informatics Vol. 12, No. 2, April 2023, pp. 930~939Journal homepage: http://beei.org ISSN: 2302-9285, DOI: 10.11591/eei. v12i2.4466

[4] Swathi Sambangi and Lakshmeeswari Gondi," A Machine Learning Approach for DDoSDistributed Denial of Service) Attack Detection Using Multiple Linear Regression". 20 December 2020.