Performance of Various ML Algorithms for detection of DDoS Attack

REPORT on B. Tech Project
(CS4272)

By

Jaykishan Padia (2020CSB032)
Amrita Kesh (2020CSB036)
Komal Gupta (2020CSB087)

under the guidance of

Prof. Sipra Das Bit

DEPARTMENT OF COMPUTER SCIENCE AND TECHNOLOGY,
IIEST, SHIBPUR
HOWRAH – 711103

2023-2024

Declaration

I hereby declare that this thesis is the record of bona fide research work carried out by us under the supervision of Dr. Sipra Das Bit, Professor, Department of Computer Science and Technology, Bengal Engineering and Science University, Shibpur. I further declare that this thesis has not previously formed the basis for the award of any degree, diploma, associate ship, fellowship or other similar title of recognition.

Jaykishan Padia
Enrollment No. 2020CSB032

Amrita Kesh
Enrollment No. 2020CSB036

Komal Gupta
Enrollment No. 2020CSB087

This page is intentionally left blank

Acknowledgements

I would like to thank my supervisor Prof. Sipra Das Bit for her guidance, support and advice at all stages of our work. I would also like to thank the head Prof. Apurba Sarkar and all other professors of the Department of Computer Science and Technology
for their valuable suggestions.

Contents

List of figures

Abstract

Distributed Denial of Service (DDoS) attack is a menace to network security that aims at exhausting the target network with malicious traffic. This research explores the effectiveness of different machine learning (ML) classification algorithms in detection of Distributed Denial of Service (DDoS) attacks. We compare the performance of logistic regression, random forest and Naive-Bayes algorithms in identifying DDoS attacks. Our work primarily focuses on preprocessing datasets, training machine learning models, and making predictions based on the trained models. The study evaluates the algorithms based on various metrics such as **accuracy, precision, recall, and F1-score.**

Chapter-1

Introduction

Network security is one of the most important challenges that we face today. Among the many threats, Distributed Denial of Service (DDoS) attack is a very powerful technique to attack internet resources.

A DDoS attack is a malicious attempt to disrupt the normal traffic of a targeted server, service, or network by overwhelming the target or its infrastructure with a flood of internet traffic.

DDoS attacks are carried out with networks of Internet-connected machines.

These networks consist of computers and other devices (such as IoT devices) which have been infected with malware, allowing them to be controlled remotely by an attacker. These individual devices are referred to as bots (or zombies), and a group of bots is called a botnet.

ML techniques are good as they do not have any prior known data distribution, but defining the best feature-set is one of the main concerns for them.

Chapter-2

Related Works

2.1 Literature Review

Distributed Denial of Service (DDoS) attacks pose significant challenges to network security and availability, prompting researchers to develop sophisticated detection mechanisms to mitigate their impact. In recent years, various approaches have been proposed to detect and mitigate DDoS attacks, leveraging advanced machine learning and statistical techniques. This literature review examines two notable studies in the field of DDoS attack detection, highlighting their methodologies, contributions, and findings.

Jin and Yeung (2004) proposed a covariance analysis model for DDoS attack detection, as presented in [1]. The study focused on analyzing the covariance structure of network traffic features to identify patterns indicative of DDoS attacks. By capturing correlations and relationships between different traffic attributes, such as packet sizes, transmission rates, and protocol types, the model aimed to distinguish between normal and attack traffic effectively. Experimental results demonstrated the effectiveness of the covariance analysis model in detecting various forms of DDoS attacks, showcasing its potential for enhancing network security.

In [2], Subbulakshmi et al. (2011) introduced an approach for DDoS attack detection using enhanced support vector machines (SVMs) with real-time generated datasets. The study leveraged SVMs, a popular machine learning algorithm known for its ability to classify complex and high-dimensional data, to identify DDoS attack patterns from network traffic data. To address the challenges of limited and imbalanced training data, the researchers proposed a method for generating synthetic datasets in real-time, enabling the SVM model to adapt and learn from dynamic network environments. Experimental evaluations demonstrated the effectiveness of the proposed approach in accurately detecting DDoS attacks while minimizing false positives.

In a similar vein, Singh and De (2015) presented an approach for DDoS attack detection using classifiers, as discussed in [6]. The study explored the use of various classification algorithms to distinguish between normal and attack traffic based on features extracted from network packets. By training classifiers on labeled datasets and evaluating their performance, the researchers aimed to identify the most effective algorithm for detecting DDoS attacks in real-world scenarios.

Both studies underscore the importance of leveraging advanced machine learning and statistical techniques for DDoS attack detection. While Jin and Yeung (2004) focused on covariance analysis to identify attack patterns, Subbulakshmi et al. (2011) explored the use of SVMs with real-time generated datasets for adaptive detection. These contributions highlight the diverse strategies and methodologies employed by researchers to combat the evolving threat landscape of DDoS attacks, paving the way for more robust and adaptive detection mechanisms in the future.

Prior research has investigated the use of statistical methods[1][3] and machine learning methods[2][6][11]. Previous works have demonstrated the effectiveness of ML-based approaches in detecting anomalies and identifying malicious traffic patterns. However, there is a need for further research to evaluate the performance of different ML algorithms under DDoS attacks.

2.2 Motivation

The increasing prevalence and sophistication of Distributed Denial of Service (DDoS) attacks pose a significant threat to modern network infrastructures, including emerging IoT environments. With the proliferation of interconnected devices the susceptibility of networks to DDoS attacks has heightened, necessitating robust detection and mitigation strategies. This project aims to address this critical need by evaluating the performance of different classification algorithms in detecting and mitigating various types of DDoS attacks, thereby enhancing the security and resilience of networks.

2.3 Objective

The primary objective of this project is to compare the effectiveness of logistic regression, random forest, and naive Bayes classification algorithms in detecting and mitigating DDoS attacks targeting IoT devices. Specifically, the project aims to assess the accuracy, precision, and F1 score of each algorithm in distinguishing between normal network traffic and different types of DDoS attacks, including UDP flood, SYN flood, LDAP reflection, and others. By achieving these objectives, the project seeks to provide valuable insights into the strengths and limitations of different machine learning techniques for combating DDoS threats in IoT environments, ultimately contributing to the development of more adaptive and resilient cybersecurity solutions.

2.4 Relevance

While this project may not be directly based on 5G IoT devices, its findings hold significant relevance in the context of 5G IoT environments. With the rapid deployment of 5G technology and the proliferation of interconnected IoT devices, the threat landscape for DDoS attacks has expanded exponentially. Therefore, understanding the performance of classification algorithms in detecting and mitigating DDoS attacks is crucial for ensuring the security and resilience of 5G IoT networks. By benchmarking logistic regression, random forest, and naive Bayes algorithms against various types of DDoS attacks, this work provides valuable insights into their effectiveness and applicability in 5G IoT environments. The findings of this project can inform the development of tailored cybersecurity solutions and strategies to safeguard 5G IoT networks.

Chapter-3

System Model

3.1 Architecture

The architecture of our proposed system will include our computer/server system connected to the network.

Our computer/server system will do the task of network data collection and DDoS attack detection.

The data consisting of 80 traffic features will be extracted using CICFlowMeter[9]. Afterwards, this data will be passed through our trained ML models which will detect whether or not our network is experiencing a DDoS attack.

The detection of network will be done when:

- There is a suspicious amount of traffic originating from a single IP address or IP range.
- A flood of traffic from users who share a single behavioral profile, such as device type or geolocation, or web browser version.
- An immediate surge in requests to a single page or endpoint.
- Odd traffic patterns of sudden spikes which appear unnatural

The monitoring of network traffic will be done continuously and in suspicious cases, we will use the ML models to detect whether there is a DDoS attack on the system.

3.2 Proposed Model

We propose to evaluate the performance of three ML algorithms (logistic regression, random forest and Naive Bayes) regarding detection of DDoS attacks. These algorithms are chosen for their simplicity, scalability, and interpretability. To train the models, we will use the CIC-DDoS2019 dataset. **[10]**

3.2.1 Processing Model

Before training the models, we will preprocess the data to ensure its suitability for ML algorithms. This preprocessing phase involves tasks such as data cleaning, feature selection, normalization, and handling of missing values. We used CICFlow Meter to analyse the network traffic data **[12]**. By preparing the data in a structured and standardized format, we aim to enhance the robustness and effectiveness of the ML models.

3.2.2 Tools for analyzing

Libraries utilized for data preprocessing, model training, and performance evaluation include popular Python libraries such as pandas, scikit-learn, and matplotlib for data manipulation, machine learning, and visualization, respectively. **Running Location -** The machine learning algorithms are deployed on a centralized server or cloud platform, where they analyze incoming data streams.

3.2.3 Input and Output

The input to the algorithms consists of features extracted from network traffic, system logs, and device telemetry data. These features include packet headers, traffic volume, packet length, protocol type, and other relevant attributes. The output of the algorithms is a binary classification indicating whether a DDoS attack is detected or not. This output provides actionable insights for network administrators and security personnel to respond to potential security threats in real-time.

3.3 Various DDOS attacks considered

DDoS attacks can take various forms, including volumetric, protocol, and application layer attacks. Volumetric attacks flood the network with a high volume of traffic, consuming bandwidth and resources. Protocol attacks target vulnerabilities in network protocols, exploiting weaknesses in packet handling and processing. Application layer attacks focus on specific services or applications, aiming to exhaust server resources or disrupt communication channels.

3.3.1 UDP FLOOD

UDP flood attacks involve sending a massive volume of UDP (User Datagram Protocol) packets to a target. These packets may be sent from a single source (single-flow UDP flood) or from multiple distributed sources (distributed UDP flood). he goal is to consume the target's network bandwidth, exhaust its computational resources (such as CPU and memory), or overwhelm its network infrastructure (such as routers and switches).

3.3.2 UDP Lag Attack

UDP Lag attacks, the flood of UDP packets can introduce significant network latency or lag, causing delays in packet delivery and increasing round-trip times for network communication. While the primary goal is still to overwhelm the target's resources, UDP Lag attacks may emphasize the disruption of real-time applications that are sensitive to latency, such as online gaming, VoIP (Voice over IP), and streaming media.

3.3.3 NetBIOS Amplification

NetBIOS services can be leveraged in reflection and amplification attacks. Attackers send spoofed NetBIOS queries to vulnerable servers, which then respond with larger NetBIOS responses to the victim's IP address. This amplification of response traffic can consume the victim's bandwidth and exhaust its network resources, resulting in a denial of service.

3.3.4 SYN FLOOD (Synchronize)

SYN flood attacks exploit the three-way handshake process of the TCP protocol. Attackers flood the target server with a large number of SYN requests, but they do not complete the handshake by sending the final ACK packet. This results in the target server maintaining half-open connections, eventually exhausting its resources and preventing legitimate users from establishing connections.

3.3.5 LDAP Reflection

LDAP servers can be exploited in reflection attacks similar to DNS and NTP reflection attacks. Attackers send spoofed LDAP queries to vulnerable servers, which then respond with larger LDAP responses to the victim's IP address. This amplification of response traffic can overwhelm the victim's network infrastructure, leading to service disruption.

3.3.6 MSSQL

Attackers can exploit vulnerabilities in Microsoft SQL Server (MSSQL) to launch DDoS attacks. By sending specially crafted SQL queries or exploiting known vulnerabilities in MSSQL services, attackers can cause the target server to become unresponsive or crash, resulting in denial of service for legitimate users.

3.4 Algorithms Used

3.4.1 Logistic regression

Logistic regression is a linear model suitable for binary classification tasks and it forms a S shaped curve (sigmoid). In machine Learning, we use sigmoid to map predictions to probabilities.

$$h\theta(X) = \frac{1}{1 + e^{-\left(\beta_0 + \beta_1 X\right)}}$$

[8]

3.4.2 Random Forest

Random forest is an ensemble learning method that combines multiple decision trees to improve accuracy and robustness. Random forests are a combination of tree predictors such that each tree depends on the values of a random vector sampled independently and with the same distribution for all trees in the forest. The generalisation error for forests converges as to a limit as the number of trees in the forest becomes large. The generalisation error of a forest of tree classifiers depends on the strength of the individual trees in the forest and the correlation between them. Using a random selection of features to split each node yields error rates that compare favourably to Adaboost, but are more robust with respect to noise. **[4]**

3.4.3 Naive-Bayes

Naive Bayes is simple, scalable and can handle high dimensional data. Naïve Bayes is part of a family of generative learning algorithms, meaning that it seeks to model the distribution of inputs of a given class or category. Unlike discriminative classifiers, like logistic regression, it does not learn which features are most important to differentiate between classes.**[13]**

$$\underbrace{P(c|x)}_{\text{Posterior Probability}} = \frac{\overbrace{P(x|c)}^{\text{Likelihood}}\overbrace{P(c)}^{\text{Class Prior Probability}}}{\underbrace{P(x)}_{\text{Predictor Prior Probability}}}$$

$$P(c\,|\,X) = P(x_1\,|\,c) \times P(x_2\,|\,c) \times \cdots \times P(x_n\,|\,c) \times P(c)$$

[5]

3.5 Metrics Used

3.5.1 Accuracy

The proportion of correctly classified instances by the machine learning algorithms.

$$\text{Accuracy} = \frac{(TP + TN)}{(TP + FP + TN + FN)}$$

[7]

3.5.2 Precision

The ratio of true positive predictions to the total number of positive predictions, indicating the accuracy of positive predictions.

$$precision = \frac{tp}{tp + fp}$$ [7]

3.5.3 Recall

The ratio of true positive predictions to the total number of actual positive instances, measuring the algorithm's ability to identify all positive instances.

$$Recall = \frac{TP}{TP + FN}$$ [7]

3.5.4 F1 Score

The harmonic mean of precision and recall, providing a balanced measure of the algorithm's performance.

$$\text{F1 Score} = \frac{2}{\frac{1}{\text{Precision}} + \frac{1}{\text{Recall}}} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$ [7]

Chapter-4

Analysis on Performance of classification

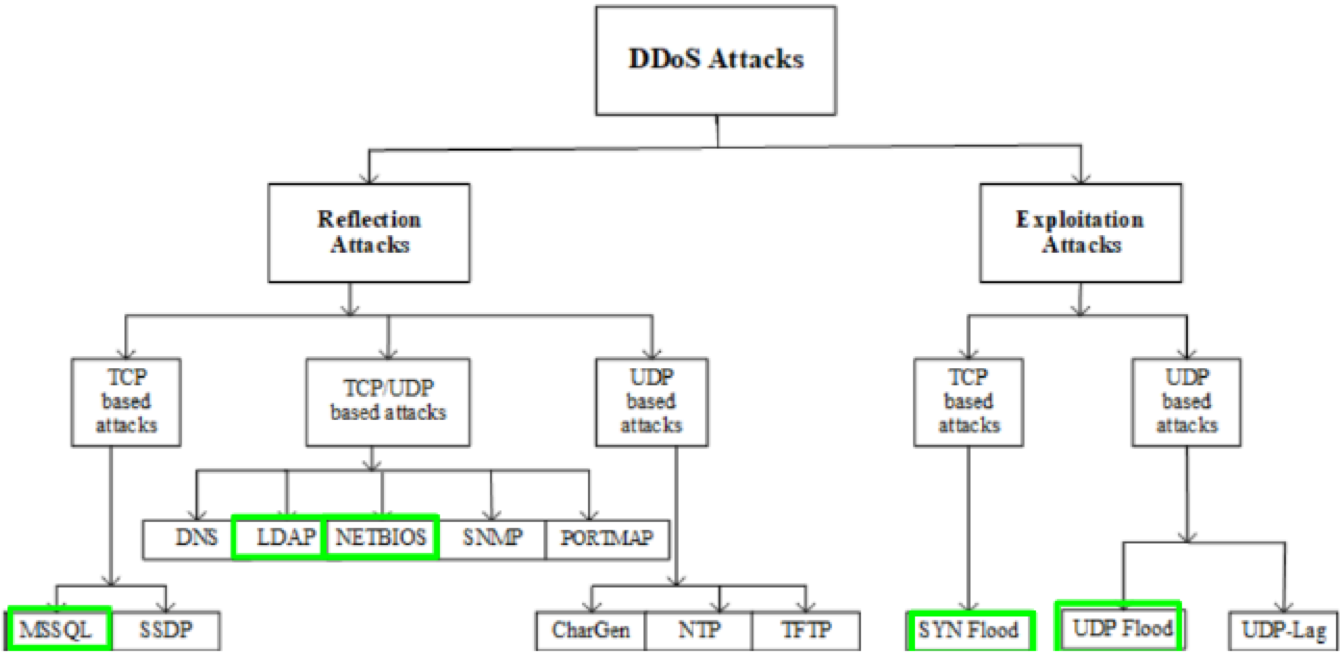

Figure 1:DDoS Attack Taxonomy

4.1 UDP FLOOD

This was plotted using the Random Forest Regression algorithm. This is useful for gaining insights regarding the important features for a particular type of DDoS attack.

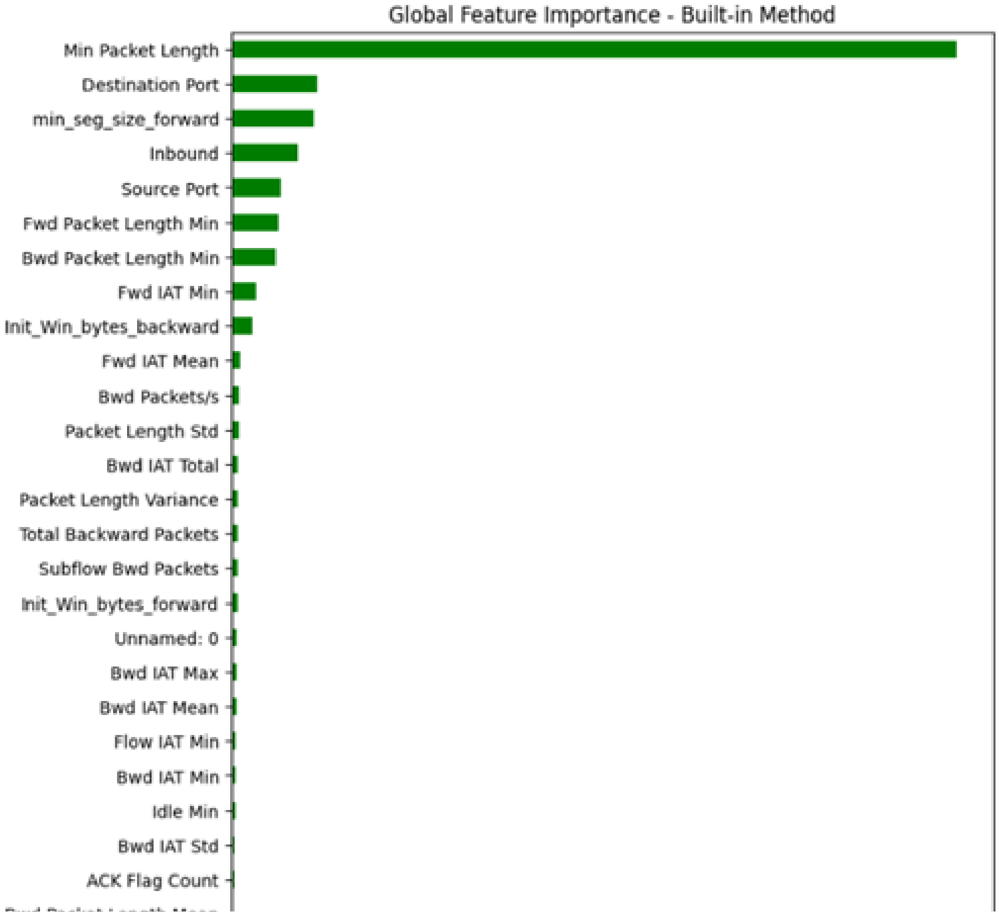

Figure 2:relative importance of different parameters for a UDP attack

4.1.1. Random Forest Classifier

- Accuracy: 0.9935479152176735
- Precision: 0.993548146621589
- Recall: 0.999994834821675
- F1 Score: 0.9967610671217557

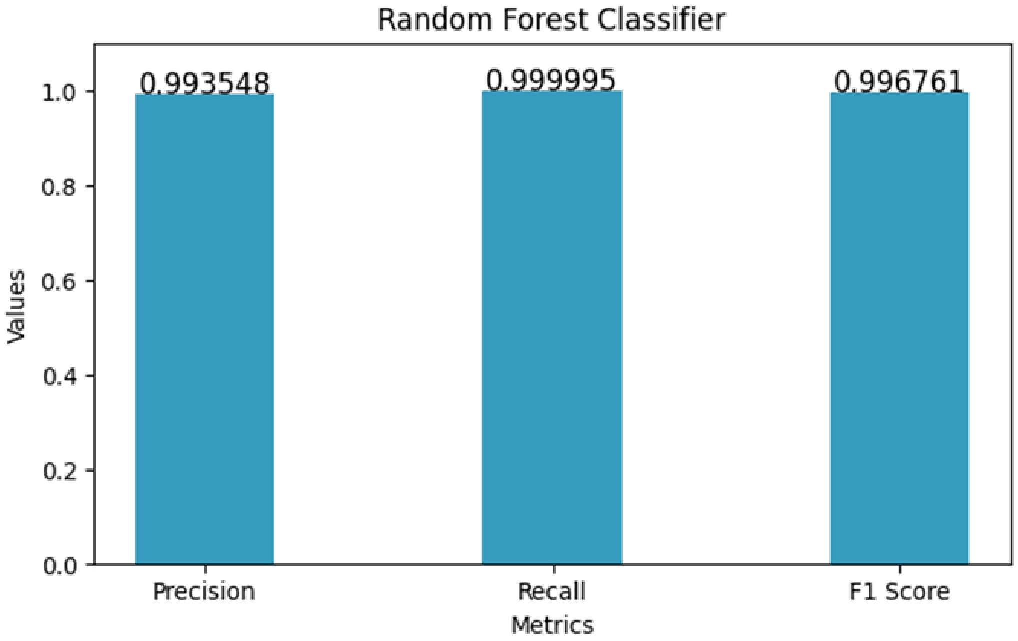

Figure 3: Bar graph for comparing Precision, Recall and F1 Score for Random Forest Classifier

4.1.2. Logistic Regression

- Accuracy: 0.9928200112329862
- Precision: 0.9928203940645784
- Recall: 0.9999994562970184
- F1 Score: 0.9963969940233413

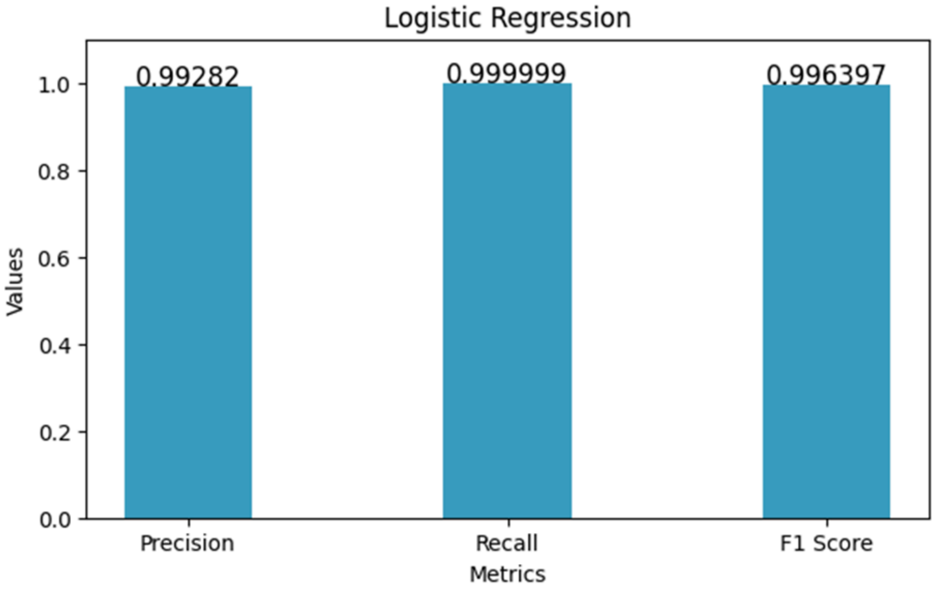

Figure 4: Bar graph for comparing Precision, Recall and F1 Score for Logistic Regression

4.1.3. Naive Bayes

- Accuracy: 0.9930048885891267
- Precision: 0.993015693902095
- Recall: 0.9999874948314236
- F1 Score: 0.996489400204582

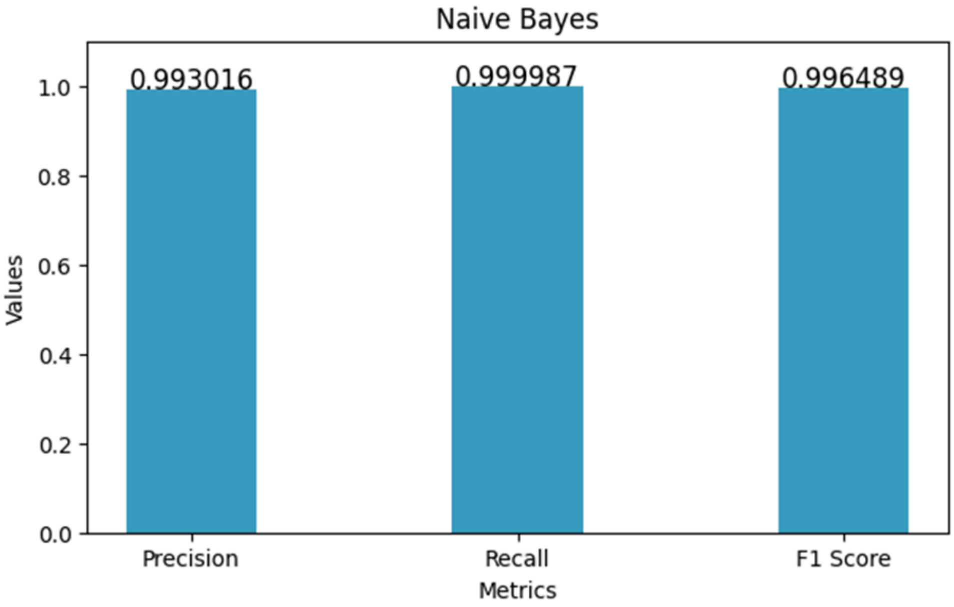

Figure 5: Bar graph for comparing Precision, Recall and F1 Score

4.2 NetBIOS AMPLIFICATION

This was plotted using the Random Forest Regression algorithm. This is useful for gaining insights regarding the important features for a particular type of DDoS attack.

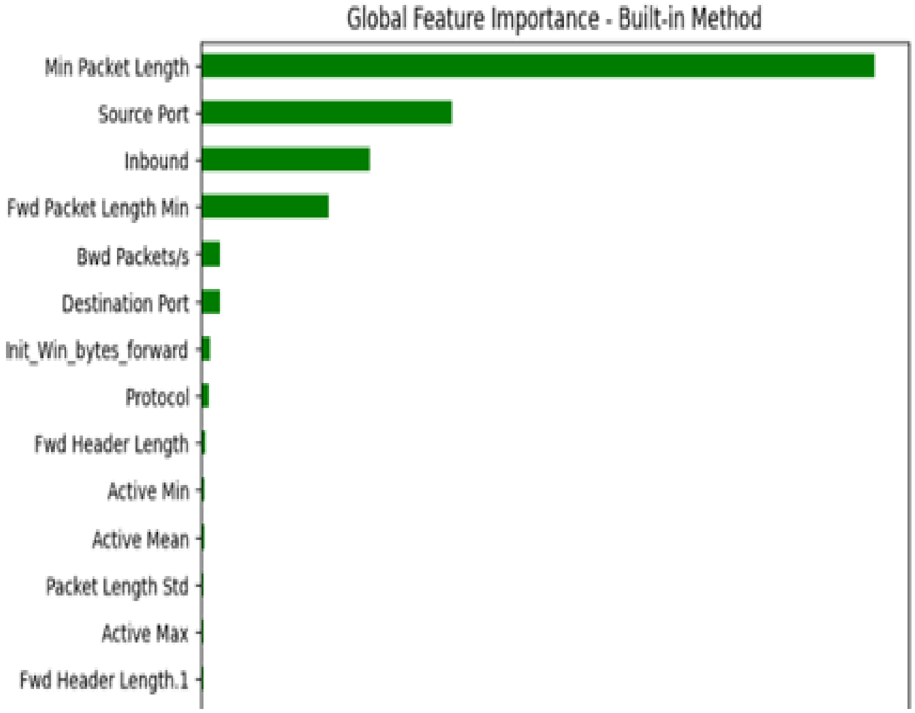

Figure 6: Relative importance of different features for a NetBios Attack

4.2.1. Random Forest Classifier

- Accuracy: 0.999976844466083
- Precision: 0.9999771367717423
- Recall: 0.9999996991612616
- F1 Score: 0.9999884178392351

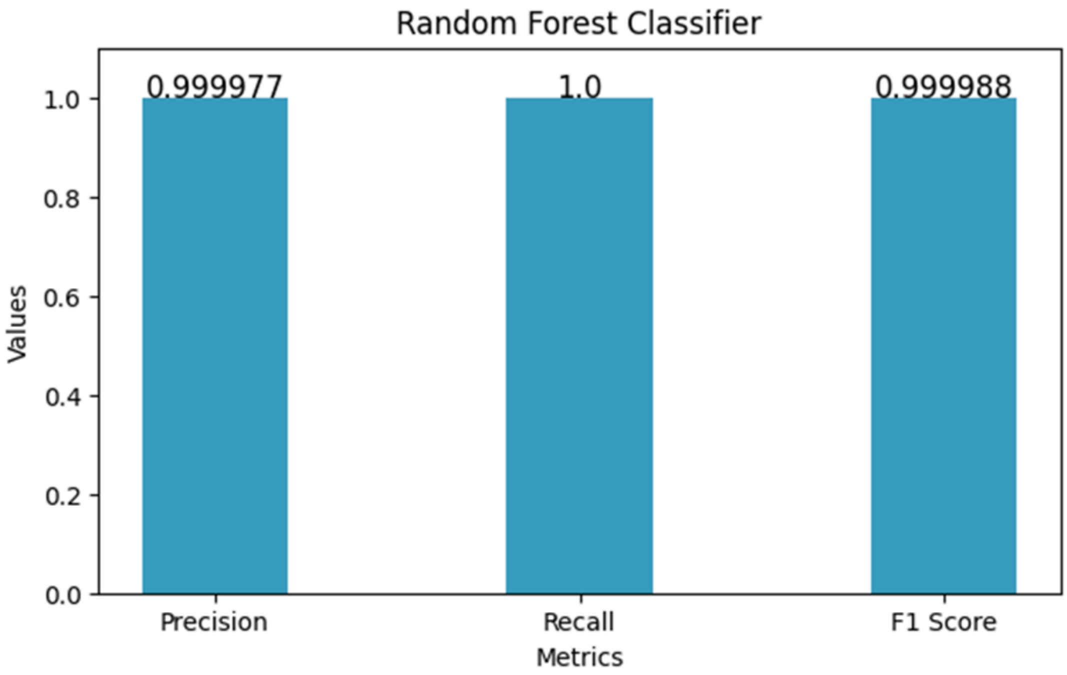

Figure 7: Bar graph for comparing Precision, Recall and F1 Score

4.2.2. Logistic Regression

- Accuracy: 0.9583733645982876
- Precision: 0.9942776481173973
- Recall: 0.9635509138501749
- F1 Score: 0.9786739129643197

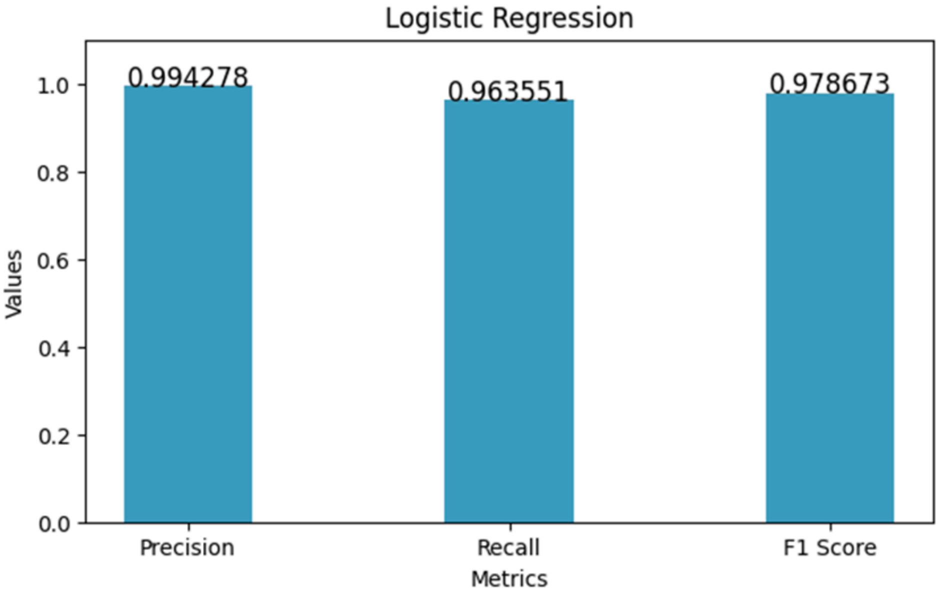

Figure 8: Bar graph for comparing Precision, Recall and F1 Score

4.2.3. Naive Bayes

- Accuracy: 0.9985535309332372
- Precision: 0.9997380283471395
- Recall: 0.9988146953706935
- F1 Score:0.9992761485686252

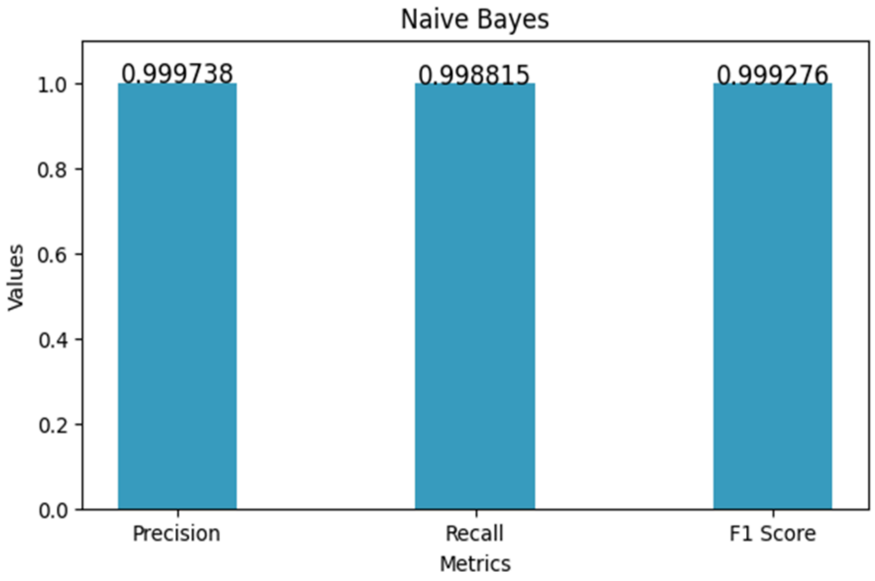

Figure 9: Bar graph for comparing Precision, Recall and F1 Score

4.3 SYN FLOOD

This was plotted using the Random Forest Regression algorithm. This is useful for gaining insights regarding the important features for a particular type of DDoS attack.

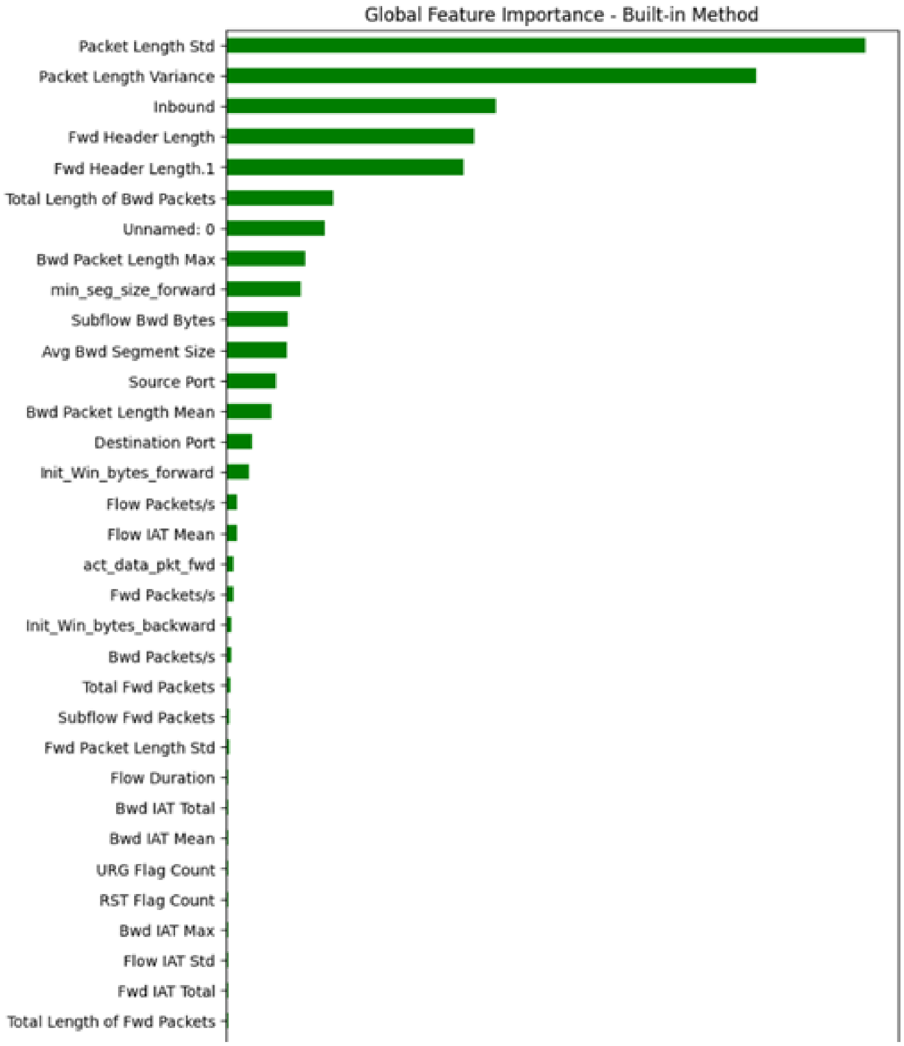

Figure 10: Relative importance of different features for a Syn Flood attack

4.3.1. Random Forest Classifier

- Accuracy: 0.9990102700580686
- Precision: 0.9990887261190069
- Recall: 0.9999135449134174
- F1 Score: 0.9995009653498113

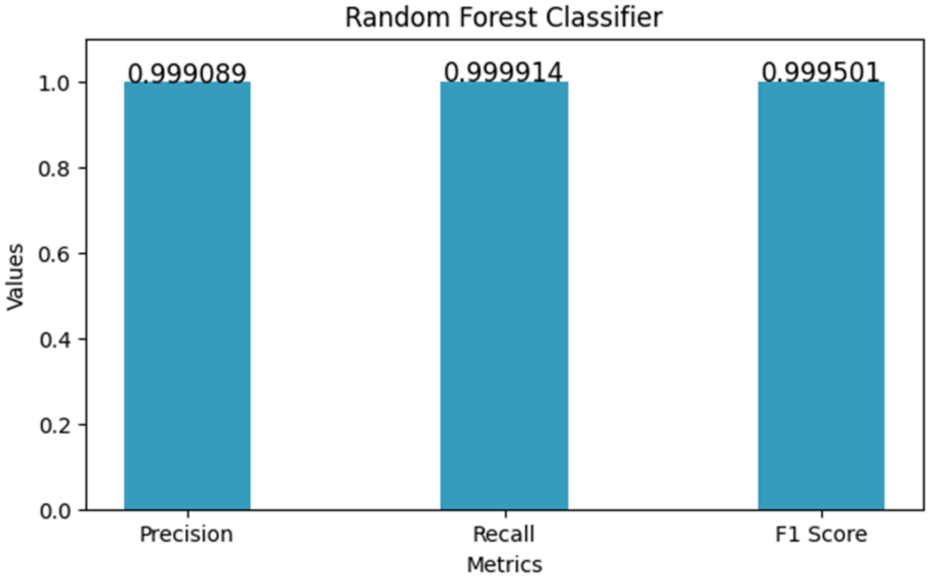

Figure 11: Bar graph for comparing Precision, Recall and F1 Score

4.3.2. Logistic Regression

- Accuracy: 0.9583733645982876
- Precision: 0.9942776481173973
- Recall: 0.9635509138501749
- F1 Score: 0.9786739129643197

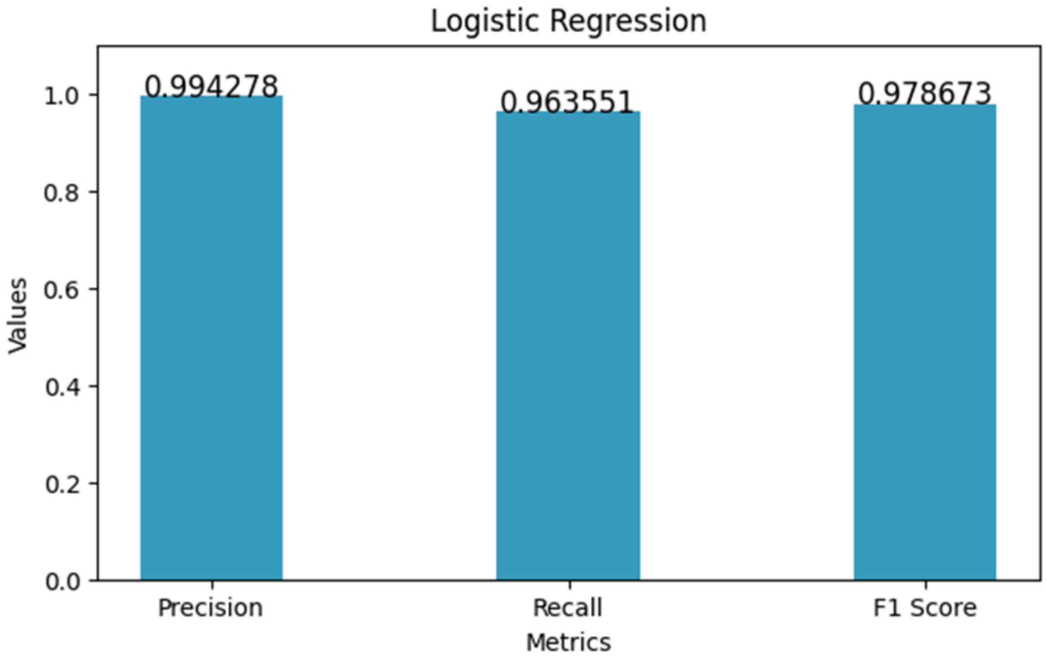

Figure 12: Bar graph for comparing Precision, Recall and F1 Score

4.3.3. Naive Bayes

- Accuracy: 0.9583731375999545
- Precision: 0.9942775356532425
- Recall: 0.9635506854189249
- F1 Score: 0.9786729914953117

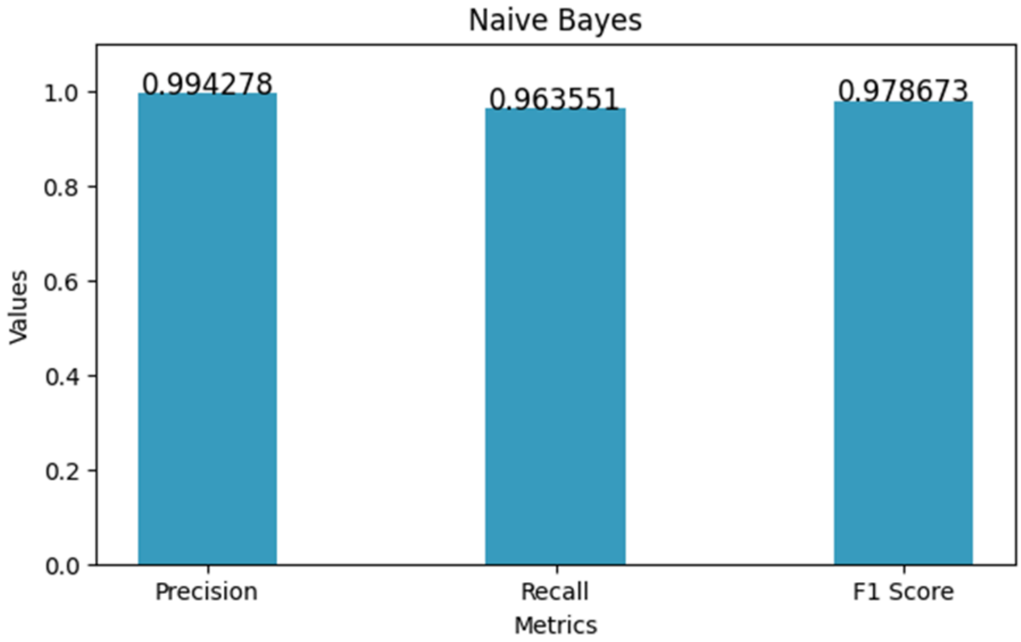

Figure 13: Bar graph for comparing Precision, Recall and F1 Score

4.4 LDAP REFLECTION

This was plotted using the Random Forest Regression algorithm. This is useful for gaining insights regarding the important features for a particular type of DDoS attack.

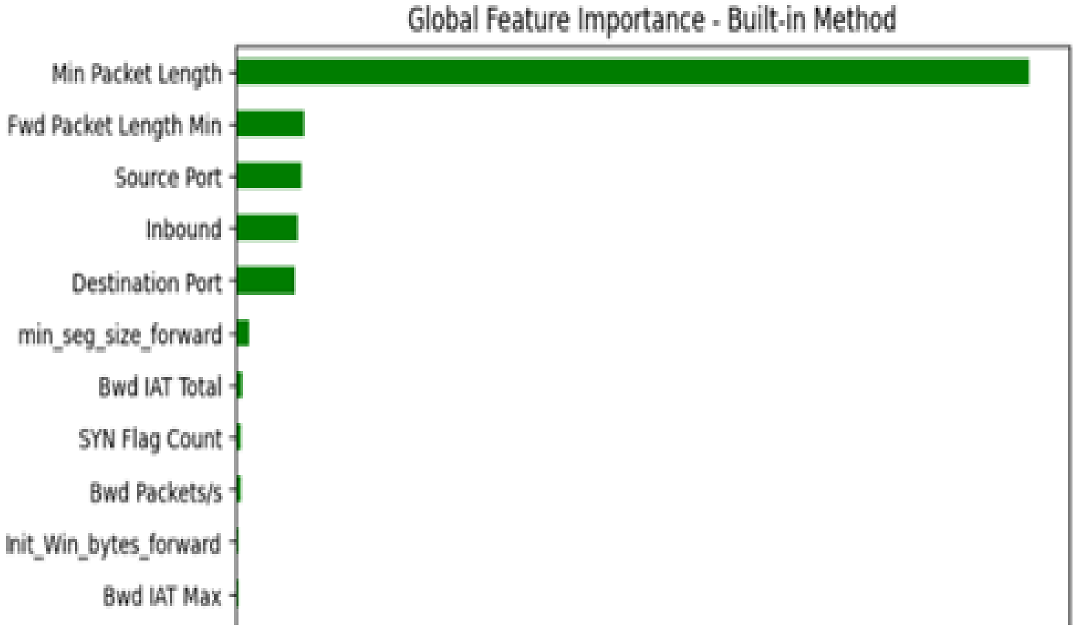

Figure 14: Relative importance of different features for a Ldap attack

4.4.1. Random Forest Classifier

- Accuracy: 0.992175791429551
- Precision: 0.9915325900231428
- Recall: 0.9998872310662302
- F1 Score: 0.9956923853533081

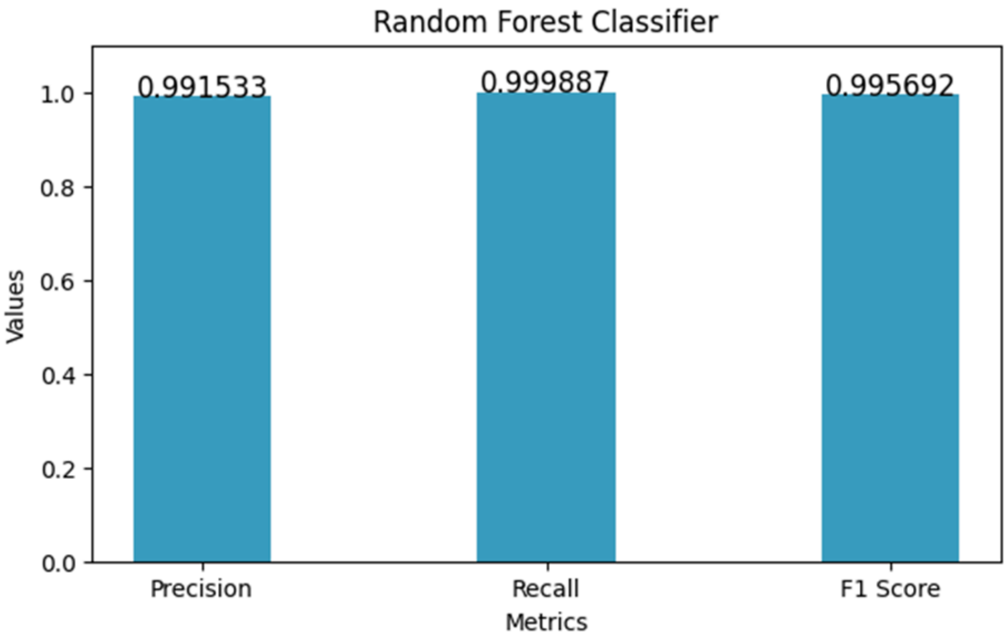

Figure 15: Bar graph for comparing Precision, Recall and F1 Score

4.4.2. Logistic Regression

- Accuracy: 0.9057088409526002
- Precision: 0.9056206168720637
- Recall: 0.9999484484874196
- F1 Score: 0.9504498653147526

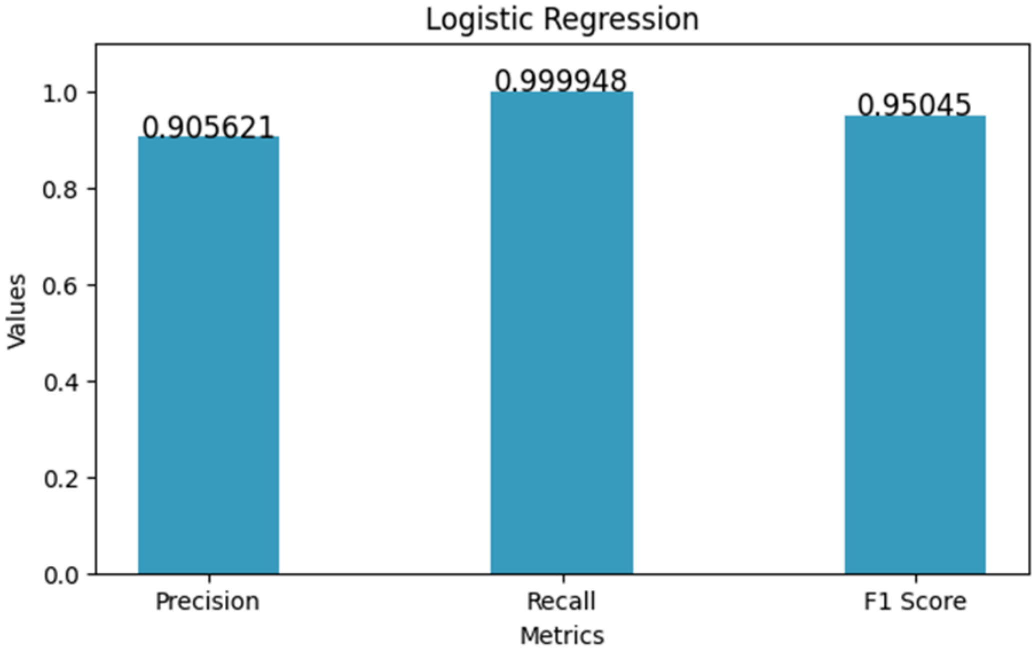

Figure 16: Bar graph for comparing Precision, Recall and F1 Score

4.4.3. Naive Bayes

- Accuracy: 0.9049857172134531
- Precision: 0.9049888969837413
- Recall: 0.9999162287920568
- F1 Score: 0.9500873011742544

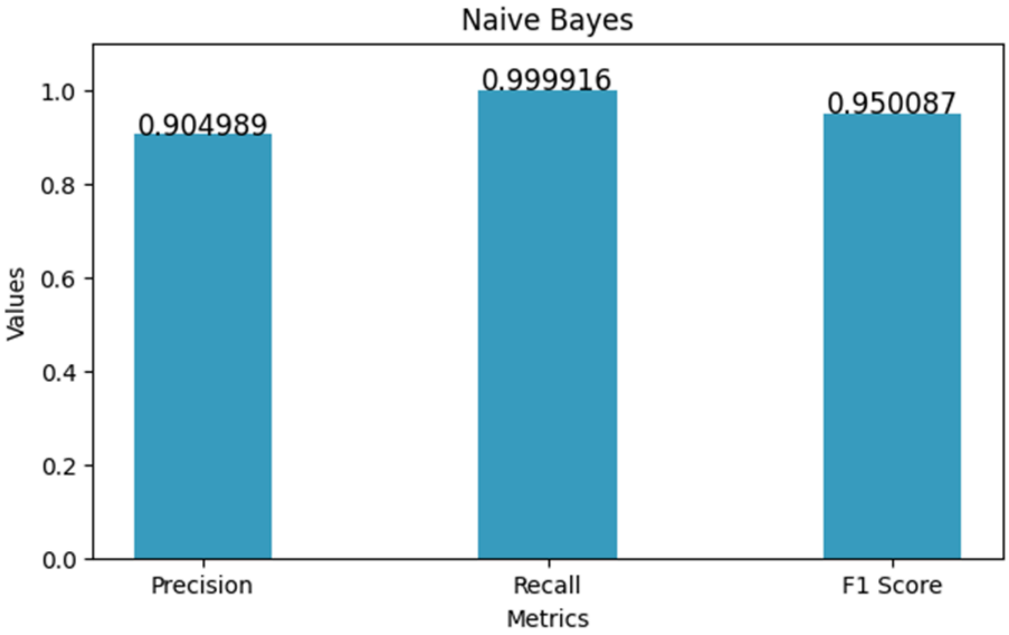

Figure 17: Bar graph for comparing Precision, Recall and F1 Score

4.5 MSSQL EXPLOIT

This was plotted using the Random Forest Regression algorithm. This is useful for gaining insights regarding the important features for a particular type of DDoS attack.

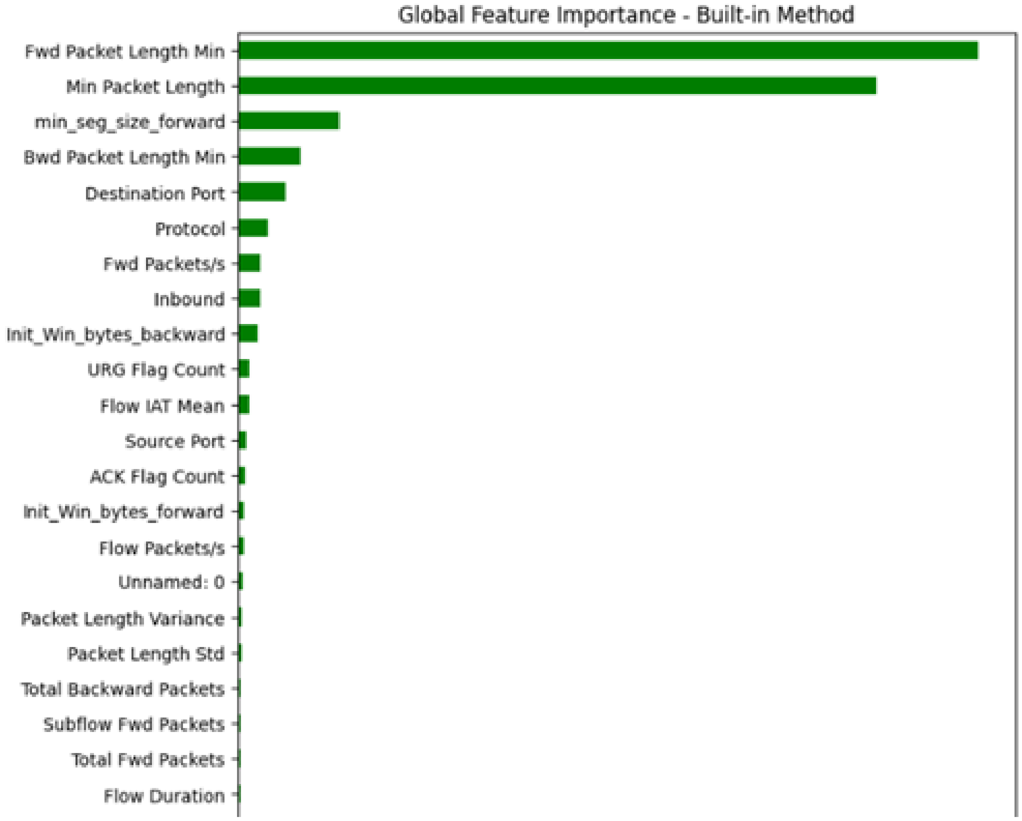

Figure 18: Relative importance of different features for a Mssql attack

4.5.1. Random Forest Classifier

- Accuracy: 0.9981453535427659
- Precision: 0.9981470721747966
- Recall: 0.9999974824101391
- F1 Score: 0.9990714204930858

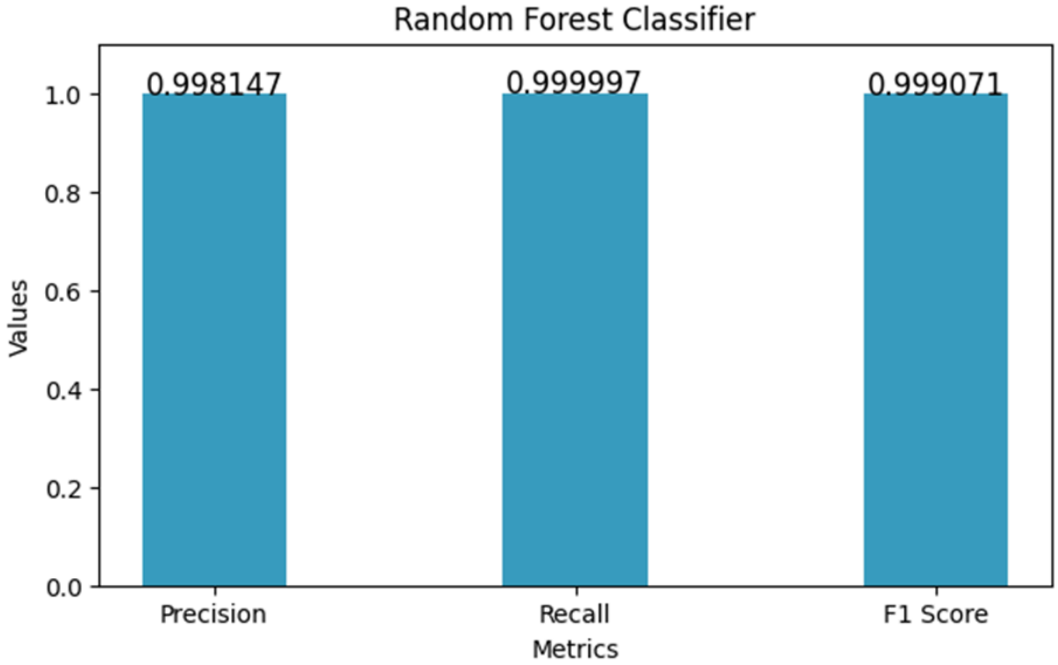

Figure 19: Bar graph for comparing Precision, Recall and F1 Score

4.5.2. Logistic Regression

- Accuracy: 0.9979361518624781
- Precision: 0.9979752096892731
- Recall: 0.9999602580457676
- F1 Score: 0.9989667477453353

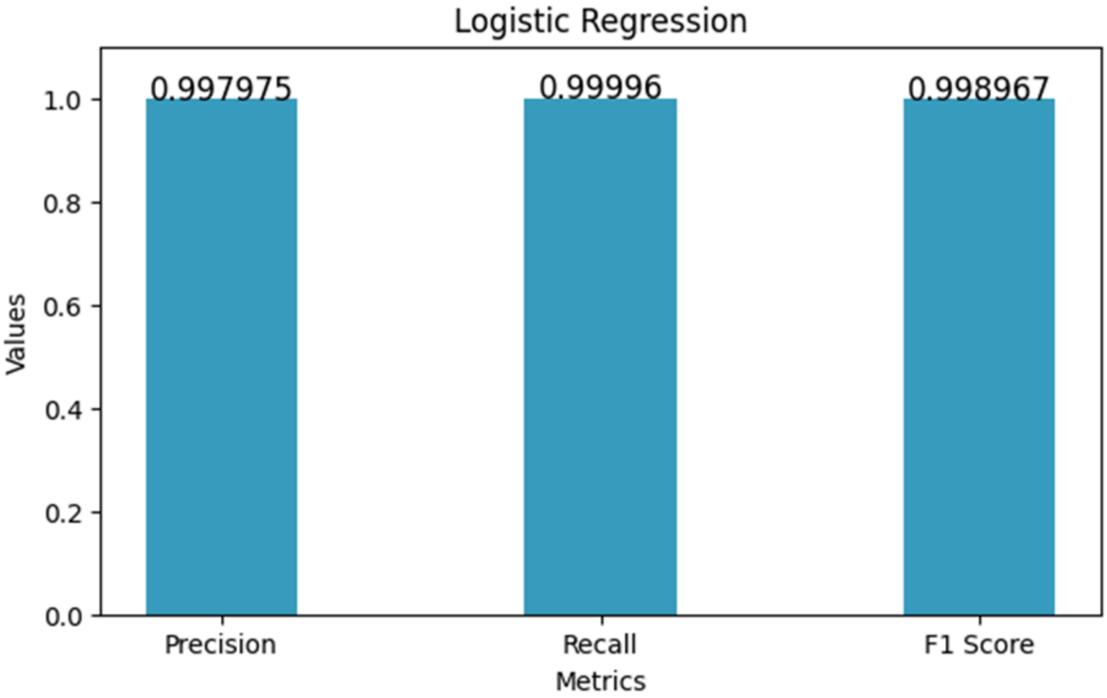

Figure 20: Bar graph for comparing Precision, Recall and F1 Score

4.5.3. Naive Bayes

- Accuracy: 0.9971901307761687
- Precision: 0.997849051501561
- Recall: 0.99933787386659
- F1 Score: 0.9985929077555391

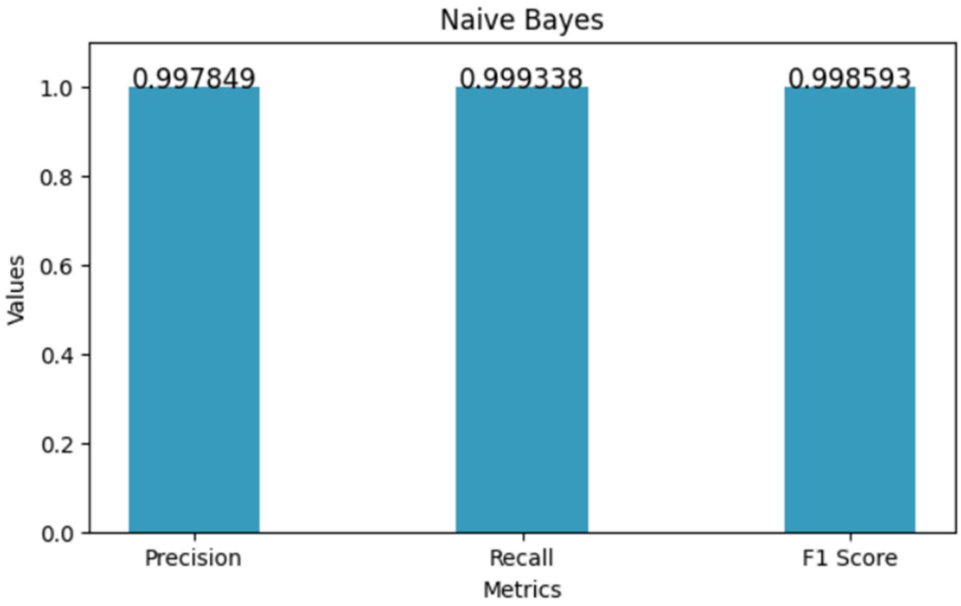

Figure 21: Bar graph for comparing Precision, Recall and F1 Score

Chapter-5

PERFORMANCE EVALUATION

Table 1: Preferred Algorithm for different DDoS attacks

DDoS Attack	ML Algorithm
UDP Flood	Logistic Regression
MSSQL	Random Forest Classification
SYN	Random Forest Classification
LDAP	Random Forest Classification
NetBIOS	Random Forest Classification

Chapter-6

Conclusion

In conclusion, this project delves into an extensive exploration of the performance of logistic regression, random forest, and Naive Bayes algorithms in the detection of Distributed Denial of Service (DDoS) attacks. Through comprehensive analysis and experimentation, the study sheds light on the efficacy of machine learning (ML)-based methodologies in fortifying cybersecurity measures.

The findings underscore the significance of employing ML approaches for bolstering security protocols, particularly in the realm of DDoS attack detection. Among the algorithms investigated, random forest emerges as the frontrunner, exhibiting remarkable accuracy and robustness in identifying and mitigating DDoS threats. Its ability to handle complex data patterns and maintain high performance levels under diverse conditions positions it as a formidable tool in the arsenal against cyber-attacks.

Nevertheless, it is imperative to acknowledge the nuanced strengths and limitations inherent in each algorithm. While random forest excels in various aspects, logistic regression and Naive Bayes algorithms demonstrate notable potential in real-time DDoS detection. Their relatively simpler structures and computational efficiency make them viable options, particularly in scenarios where resource constraints or latency considerations are paramount.

Looking ahead, the trajectory of research in this domain should prioritize continual refinement of existing algorithms and exploration of alternative ML techniques. Fine-tuning the parameters and architectures of logistic regression, random forest, and Naive Bayes models could yield further enhancements in detection accuracy and response efficacy. Additionally, investigating novel ML methodologies, such as deep learning or ensemble techniques, holds promise for extending the frontier of cybersecurity defense mechanisms.

In essence, this project serves as a stepping stone towards a more comprehensive understanding of ML-driven cybersecurity paradigms. By leveraging the insights garnered herein, stakeholders can proactively fortify their systems against evolving cyber threats, thereby fostering a safer and more resilient digital ecosystem.

5.1 Future Scope

While our current implementation provides a foundation for detecting DDoS attacks using machine learning algorithms, there are several enhancements and additional features that can be explored to make the system more effective, scalable, and real-time. Here are some future scope considerations.

5.1 Real-Time Data Ingestion

Implement mechanisms to ingest live network traffic and system logs in real-time, allowing for continuous monitoring and analysis of incoming data streams.

5.2 Stream Processing Frameworks

Explore the use of stream processing frameworks such as Apache Kafka or Apache Flink to handle high-volume, real-time data streams and enable parallelized processing and analysis.

5.3 User Interface and Visualization

Develop a user-friendly interface and visualization tools to facilitate monitoring, analysis, and reporting of DDoS attack detection results, enabling security analysts to gain insights and take appropriate actions effectively.

5.4 Trigger-Based Alerting

Implement a trigger mechanism that monitors the output of the detection system and alerts security personnel in real-time when ongoing DDoS attacks are detected, providing timely notifications for proactive response and mitigation efforts.

REFERENCES

[1] S. Jin and D. S. Yeung, “A covariance analysis model for ddos attack detection,” in 2004 IEEE International Conference on Communications, vol. 4, pp. 1882–1886 Vol.4, 2004.

[2] T. Subbulakshmi, K. BalaKrishnan, S. M. Shalinie, D. AnandKumar, V.GanapathiSubramanian, and K. Kannathal, “Detection of ddos attacks using enhanced support vector machines with real time generated dataset,” in Third International Conference on Advanced Computing, pp. 17–22, 2011.

[3] Freedman, D. A. (2008). Logistic Regression: Why We Cannot Do What We Think We Can Do, and What We Can Do About It. *Journal of the American Statistical Association*, 95(450), 1-4.

[4] L. Breiman, “Random forests,” Machine learning, vol. 45, no. 1, pp. 5–

32, 2001.

[5] Domingos, P., & Pazzani, M. (1997). The Optimality of Naive Bayes. *Proceedings of the 13th International Conference on Machine Learning*, 118-126. - Naive Bayes

[6] K. J. Singh and T. De, “An approach of ddos attack detection using classifiers,” Emerging Research in Computing, Information, Communication and Applications, 2015.

[7] Powers, D. M. (2011). Evaluation: From Precision, Recall and F-Measure to ROC, Informedness, Markedness & Correlation. *Journal of Machine Learning Technologies*, 2(1), 37-63..

[8]Joseph Berkson, (1944) Logistic Regression

[9] CICFlowMeter, 2021 https://github.com/ahlashkari/CICFlowMeter.

[10] CIC-DDoS2019 https://www.unb.ca/cic/datasets/ddos-2019.html

[11]Developing Realistic Distributed Denial of Service (DDoS) Attack Dataset and Taxonomy https://ieeexplore.ieee.org/document/8888419

[12]B. H. Ali, N. Sulaiman, S. A. R. Al-Haddad, R. Atan and S. L. M. Hassan, "DDoS Detection Using Active and Idle Features of Revised CICFlowMeter and Statistical Approaches," 2022 4th International Conference on Advanced Science and Engineering (ICOASE), Zakho, Iraq, 2022, pp. 148-153, doi: 10.1109/ICOASE56293.2022.10075591. keywords: {Sensitivity;Databases;Scalability;Bidirectional control;Denial-of-service attack;Feature extraction;Entropy;Sequential probability ratio test;Shannon Entropy;Confusion Matrix;CICFlowMeter;DDoS},

[13] https://www.ibm.com/topics/naive-bayes