



REPUBLIC OF TÜRKİYE
ALTINBAŞ UNIVERSITY
Institute of Graduate Studies
Information Technologies

**ARTIFICIAL INTELLIGENCE MODEL FOR
DDOS-DETECTION**

Ali Mohammed Saadoon BRAICHAT

Master's Thesis

Supervisor

Prof. Dr. Osman Nuri UÇAN

Istanbul, 2023

ARTIFICIAL INTELLIGENCE MODEL FOR DDOS-DETECTION

Ali Mohammed Saadoon BRAICHAT

Information Technologies

Master`s Thesis

ALTINBAŞ UNIVERSITY

2023

The thesis titled ARTIFICIAL INTELLIGENCE MODEL FOR DDOS-DETECTION prepared by ALI MOHAMMED SAADOON BRAICHAT and submitted on 00/03/2023 has been **accepted unanimously** for the degree of Master of Science information technologies.

Prof. Dr.Osman Nuri UÇAN
the Supervisor

Thesis Defense Jury Members:

Prof. Dr. Osman Nuri UÇAN	Department - , Altinbas University	_____
Asst. Prof. Dr. Name SURNAME	Depatment of , Altinbas University	_____
Asst. Prof. Dr. Name SURNAME	Depatment of , University	_____

I hereby declare that this thesis meets all format and submission requirements of a Master`s Thesis.

Submission date of the thesis to the Graduate Education Institute: ____/____/.

I hereby declare that all information data presented in this graduation project has been obtained in full accordance with academic rules and ethical conduct. I also declare all unoriginal materials and conclusions have been cited in the text and all references mentioned in the Reference List have been cited in the text, and vice versa as required by the abovementioned rules and conduct.

Ali Mohammed Saadoon BRAICHAT

Signature

DEDICATION

With love and gratitude, I dedicate this thesis:

To my father, who has been my rock and my inspiration, I dedicate this thesis. Your unwavering faith in me has been a constant reminder of what is possible, and I am forever grateful for your love and guidance.

To my mother, who has always been my biggest cheerleader, thank you for your unending encouragement, your wise counsel, and your unrelenting support. You have shown me what it means to be strong, and I am so proud to be your son.

To my siblings, who have always been my partners in crime, thank you for making my life so much more colorful and fun. Your unwavering support and belief in me have been invaluable, and I am so lucky to have you in my life.

To my friends, who have been my sounding boards, my confidants, and my source of laughter and joy, thank you for always being there for me. Your unwavering support and encouragement have meant the world to me, and I am so grateful to have you in my life.

And finally, to my supervisor, Prof. Dr. Osman Nuri UÇAN, who has been my guide and mentor, thank you for your patience, wisdom, and expertise. Your guidance and support have been instrumental in shaping this thesis, and I am so grateful for the opportunity to have worked with you.

Thank you all for being a part of my journey, and for making this moment all the more meaningful.

ABSTRACT

ARTIFICIAL INTELLIGENCE MODEL FOR DDOS-DETECTION

BRAICHAT, Ali Mohammed Saadon

M.Sc., Information Technologies, Altınbaş University,

Supervisor: Prof. Dr. Osman Nuri UÇAN

Date: April / 2023

Pages: 96

This master's study is guided at investigating and examining the beneficial contributions and crucial rationale of ML principles and intelligent DL algorithms in detecting cyber threats in organizations and network systems that might result in significant damage to financial resources and private databases. The work relied on employing four ML models (including Logistic Regression (LR), Support Vector Machine (SVM), Stochastic Gradient Descent (SGD), and Machine Learning Model (MLM)). The mathematical numeric Python code was developed and simulations were executed. The work outputs revealed that the average accuracy of the four ML schemes obtained via the performance assessment techniques ranges from 74.54% to 96.20% for training. The minor ratio was for the SGD classifier. Also, the maximum percentage was associated with MLM. Meanwhile, the average accuracy related to the four algorithms for the testing process varies between 74.23% (minimum for the SGD classifier) and 95.93% (maximum for MLM). Using intelligent ML and DL algorithms provided practical means of cyber threat detection for the thesis's case study in terms of performance, reliability, and accuracy. The precision rates of the four ML algorithms range from 75.83% (minimum for the SGD classifier) to 98.27% (maximum for MLM), concerning the '0' results of binary classifications in the training process. For the binary classification results of '1,' the minimum portions were between 69.14% (minimum

for the LR algorithm) and 93.16% (maximum for MLM), respectively, in the training phase. For the testing procedure, the precision rate varies between 75.00% (minimum for the SGD classifier) and 97.94% (maximum for MLM) for the '0' binary classification findings. In comparison, the '1' binary classification outputs range between 70.15% (the lowest for the LR model) and 93.08% (ultimate for MLM) in the testing procedure.

Keywords: Cyber Threats, Network Attacks, Accuracy, Detection Performance, DDOS, Artificial Intelligence.



TABLE OF CONTENTS

	<u>Pages</u>
ABSTRACT	vi
LIST OF TABLES.....	xi
LIST OF FIGURES.....	xi
ABBREVIATIONS.....	xv
LIST OF SYMBOLS.....	xvii
1. INTRODUCTION	1
1.1 RESEARCH BACKGROUND	1
1.2 PROBLEM STATEMENT.....	5
1.3 RESEARCH SIGNIFICANCE.....	5
1.4 RESEARCH MAJOR AIM AND MINOR OBJECTIVES.....	6
1.5 THESIS STATEMENT	6
1.6 RESEARCH QUESTION.....	6
1.7 RESEARCH ORGANIZATION	7
2. LITERATURE REVIEW	8
2.1 GENERAL.....	8
2.2 CRITICAL CONTRIBUTIONS OF AI ALGORITHMS FOR DDOS DETECTION	8
2.3 LITERATURE REVIEW SUMMARY.....	19
3. RESEARCH METHODOLOGY	28
3.1 CHAPTER GOAL	28
3.2 RESEARCH METHOD	28
3.3 RESEARCH PHILOSOPHY.....	29
3.4 DATASET DEFINITION AND PREPARATION	30

3.5 RESEARCH TOOLS AND REQUIRED INSTRUMENTS	33
3.5.1 LR Algorithm.....	33
3.5.2 SVM Algorithm	37
3.5.3 SGD Classifier	39
3.5.4 MLM of Optimization.....	40
3.6 PERFORMANCE EXAMINATION OF ML ALGORITHMS	41
3.6.1 Accuracy Assessment	41
3.6.2 F1-Score.....	42
3.6.3 Precision.....	42
3.6.4 Recall	43
3.8 CHAPTER SUMMARY.....	43
4. RESULTS AND DISCUSSIONS.....	44
4.1 CHAPTER GOAL	44
4.2 STATISTICAL FINDINGS OF DDOS ATTACKS	44
4.3 THE MISSING VARIABLES IN THE RESEARCH.....	46
4.4 THE NUMERICAL OUTPUTS LINKED TO THE NETWORK REQUESTS FROM THE IP ADDRESS	47
4.5 NUMERICAL RESULTS OF VARIABLES CORRELATIONS	54
4.6 ML ALGORITHMS PERFORMANCE EVALUATIONS	55
4.6.1 LR Algorithm Performance Evaluation	56
4.6.2 SVM Algorithm Effectiveness Assessment.....	59
4.6.3 EVALUATION OF SGD CLASSIFIER’S PERFORMANCE	64
4.6.4 EFFECTIVENESS EVALUATION OF MLM	67
4.7 DISCUSSIONS.....	72
5. CONCLUSIONS AND RECOMMENDATIONS	73
5.1 CHAPTER GOAL	73
5.2 CONCLUSIONS	73
5.3 RECOMMENDATIONS.....	74

REFERENCES 75



LIST OF TABLES

	<u>Pages</u>
Table 2.1: Critical Findings Obtained From The Articles Reviewed In Chapter Two.	19
Table 3.1: An Illustration Of Critical Features And Description Of Each Part Linked To SDN-Based Attacks [20].	31
Table 4.1: The Numerical Outcomes Of The Training Process Related To The LR Algorithm.	56
Table 4.2: The Numerical Outcomes Of The Testing Process Related To The LR Algorithm.	58
Table 4.3: The Numerical Outcomes Of The Training Process Related To The SVM Algorithm.	60
Table 4.4: The Numerical Outcomes Of The Testing Process Related To The SVM Algorithm.	62
Table 4.5: The Numerical Outcomes Of The Training Process Related To The SGD Classifier.....	64
Table 4.6: The Numerical Outcomes Of The Testing Process Related To The SGD Classifier.	66
Table 4.7: The Numerical Outcomes Of The Training Process Related To The MLM.....	68
Table 4.8: The Numerical Outcomes Of The Testing Process Related To The MLM.....	70

LIST OF FIGURES

	<u>Pages</u>
Figure 1.1: Ddos Attack Components [2].....	2
Figure 1.2: Denial Of Service Attack Types [3].....	3
Figure 1.3: Ddos Attack Detection System [6].	4
Figure 2.1: The Workflow Of Machine Learning Classification Techniques [3].	9
Figure 2.2: The Main Stages Of AIMM [5].	10
Figure 2.3: The Main Steps Flow In The Detection Model Based On The Ensemble [10].	12
Figure 2.4: A Taxonomy Of Primary Ddos Defense Sorting [8].	13
Figure 2.5: A Ddos Detection System With Deep Hybrid Learning [18].	15
Figure 2.6: A Classification Of Deep Learning Models [19].	18
Figure 3.1: The Research Methodology Of This Work.....	29
Figure 3.2: The Research's Independent And Dependent Variables Addressed In This Work.	30
Figure 3.3: A Schematic Block Diagram Associated With The Central Principles Of The LR Model Used In Binary Classification [28].....	34
Figure 3.4: A Comparative Analysis Between (A) Single LR Classification Process And (B) Multiple LR Identification Procedure [29].	35
Figure 3.5: A Curve Resulting From The Sigmoid Function Plotting [31].....	37
Figure 3.6: A Configuration Of A Depositional Microfacies Classification Process Via The SVM Algorithm [39].	38

Figure 3.7: A Schematic Architecture Related To The Optimum Classification, Support Vectors, Hyperplane, And Optimal SVM Decision Planes [39].	38
Figure 3.8: The Adoption Of The SGD Algorithm In Making A Binary Classification Process, Including (A) The Concentric Circle Dataset And (B) The Network Structure [45].	40
Figure 3.9: The Implementation Of MLM To Detect Cyber Threats In Networks [50].	41
Figure 4.1: The Numerical Results Of The Binary Classification Of Zeros And Ones Related To The Ddos Attacks In The Dataset.	45
Figure 4.2: The Frequency Results (In Numbers) Linked To The Binary Classification Of Zeros And Ones Of The Ddos Attacks.	46
Figure 4.3: The Number Of Features With Null (Zero) And Nonzero Values.	47
Figure 4.4: The Number Of Requests Associated With Various IP Addresses Based On The ‘IP Address Of Sender.’	48
Figure 4.5: The Number Of Requests Associated With Various IP Addresses Based On The ‘IP Address Of Sender.’	49
Figure 4.6: The Number Of Requests According To Different Protocols Employed In This Work.	51
Figure 4.7: The Density Of Date And Time Implementation In The Numerical Analysis.	52
Figure 4.8: The Variation Between The Number Of Requests And The Number Of Switches.	53
Figure 4.9: The Variation Between The Number Of Requests And The Number Of Ports.	54
Figure 4.10: The Correlations Between The Dataset Variables Investigated In This Study.	55

Figure 4.11: The Numerical Simulation Findings Of The Confusion Matrix Linked To The LR Algorithm – Training.....	57
Figure 4.12: The Numerical Simulation Findings Of The Confusion Matrix Linked To The LR Algorithm – Testing.	59
Figure 4.13: The Numerical Simulation Findings Of The Confusion Matrix Linked To The SVM Model – Training Process.	61
Figure 4.14: The Numerical Simulation Findings Of The Confusion Matrix Linked To The SVM Model – Testing Process.....	63
Figure 4.15: The Numerical Simulation Findings Of The Confusion Matrix Linked To The SGD Model – Training Process.....	65
Figure 4.16: The Numerical Simulation Findings Of The Confusion Matrix Linked To The SGD Model – Testing Process.	67
Figure 4.17: The Numerical Simulation Findings Of The Confusion Matrix Linked To The MLM – Training Process.....	69
Figure 4.18: The Numerical Simulation Findings Of The Confusion Matrix Linked To The MLM – Testing Process.	71

ABBREVIATIONS

AML	:	Adversarial Machine Learning
ANN	:	Artificial Neural Network
Bagging	:	Bootstrap Aggregating
BayesNet	:	Bayesian Network
BiLSTM	:	Bidirectional Long Short-Time Memory
CNN	:	Convolutional Neural Network
DDoS	:	Distributed Denial Of Service
DrDoS	:	Distributed Reflection Denial Of Service
GAN	:	Generative Adversarial Network
ICMP	:	Internet Control Message Protocol
IDS	:	Intrusion Detection System
IG	:	Information Gain
kNN	:	K-Nearest Neighbors
LLC	:	Large Linear Classification
LR	:	Logistic Regression
LSTM	:	Long Short-Time Memory
MI	:	Mutual Information
MISP	:	Malware Information Sharing Platform
MLM	:	Machine Learning Model
RF	:	Recursive Feature

RNN	:	Recurrent Neural Network
SAG	:	Stochastic Average Gradient
SAGAA	:	Stochastic Average Gradient Adaptive Algorithm
SDN	:	Software-Defined Network
SGD	:	Stochastic Gradient Descent
SL	:	Simple Logistic
SMO	:	Sequential Minimal Optimization
SVM	:	Support Vector Machine
TCP	:	Transmission Control Protocol
UDP	:	User Datagram Protocol

LIST OF SYMBOLS

NaN	:	The Number Of Missing Values
x_i	:	The Feature Vectors Linked To The Training Samples
FN	:	False Negatives
FP	:	False Positives
KF	:	Kernel Function
TN	:	True Negatives
TP	:	True Positives
X	:	The Explanatory Variable
Z	:	The Sigmoid Correlation Or Logistic Function
i	:	The Initial Number Of Terms
n	:	The Final Number Of Terms
r	:	The Optimal Kernel Function Parameter
x	:	The Prediction Samples
β	:	The Intercept Index
π	:	Pi Function
$\sigma(Z)$	:	The Explanatory Sigmoid Function

1. INTRODUCTION

1.1 RESEARCH BACKGROUND

Distributed Denial of Service (DDoS) attack has become a global challenge and a new danger to the accessibility of Internet services. DDoS attack happens when attackers make companies that do business using the internet unavailable. It tries to avoid legitimate traffic flow between clients and servers. DDoS attacks use several distributed resources against targets. Attack by DDoS traffic is similar to regular traffic in most situations, so it is hard to be detected.

DDoS attack has a massive effect on the victims. It has two specific episodes, which are IP spoofing and flooding. In IP spoofing, attackers impersonate a trusted source. In IP spoofing attacks disguise its identity to disallow easy trace back or deceive the Cloud Provider into enjoying certain services accrued to a trusted host [1]. While in a flooding attack, attackers send several packets to mess up the accessibility of services. DDoS flooding attacks can be classified into two categories based on the protocol level that is targeted [2]:

- i. Network/transport-level DDoS flooding attacks
- ii. Application-level DDoS flooding attacks

Attackers in network-level attacks use TCP, UDP, ICMP and DNS protocol packets. During application-level attacks, the server resources have been exhausted to disrupt legitimate users' services.

There are four components involved in a DDoS attack [1]. The first component is the attacker, who scans millions of pieces of equipment over the internet to find weak equipment with poor security. The second one is control masters or handlers controlled by the attacker directly. The attacker indirectly controls the third agent, an agent or enslaved person. The attacker uses the control master and agent to organize an attack on the final component, target or victim. Figure 1.1 shows these components.

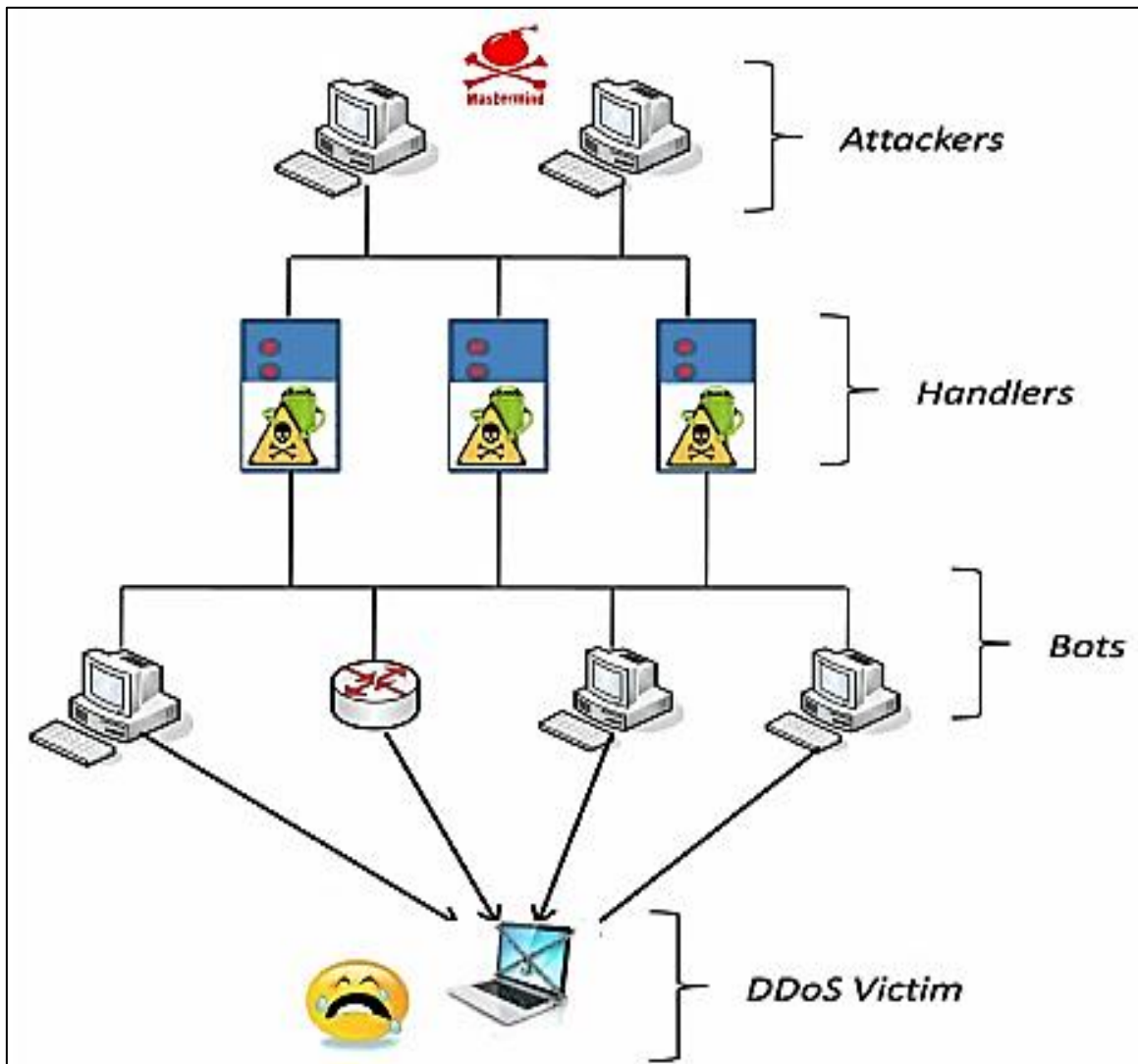


Figure 1.1: Ddos Attack Components [2].

Also, Figure 1.2 exhibits the classification of DDoS attacks. The network device level includes attacks because of the software's weakness or the power of bugs. The OS-level attacks depended on implementing protocols in the operating system. In application-based attacks are tried to settle the services using the advantages of particular bugs in network applications. In data flooding attacks, an attacker tries to use the bandwidth offered to a network by sending vast quantities of data. The protocol features take advantage of standard protocol features [3].

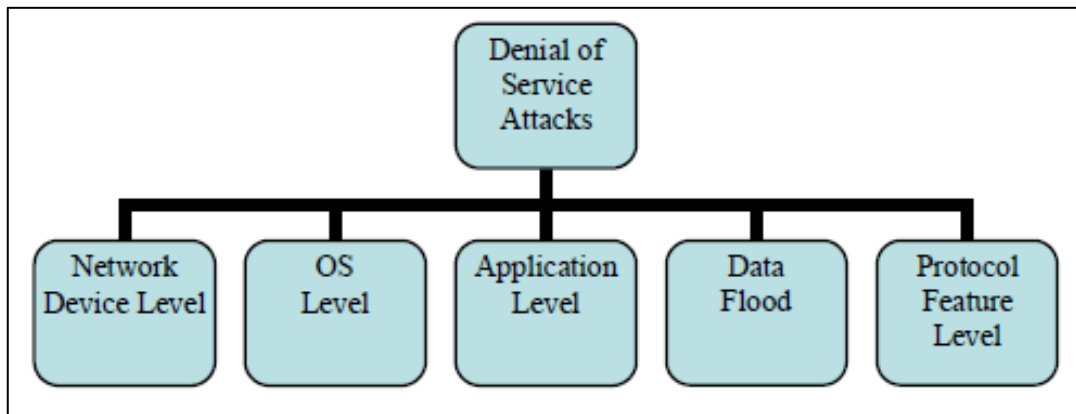


Figure 1.2: Denial Of Service Attack Types [3].

Attack detection is discovering an activity that attempts to compromise the confidentiality, integrity, or availability of resources. Monitoring computer networks and systems from external attacks is essential for security. As shown in [4], there are three methodologies to detect attacks by DDoS:

- i. Detection based on the known particulars of the attack,
- ii. Detection based on the significant anomaly detection,
- iii. Specification-based Detection.

Researchers in [5] have identified the most requirements for detecting and protecting from DDoS attacks:

- a. The DDoS protection mechanism must not disturb legitimate user's activities
- b. The DDoS detection mechanism can prevent attacks from outside and inside the network.
- c. The mechanism should meet the performance and scalability requirements of modern data centers.
- d. The DDoS protection mechanism should be low-cost.
- e. The mechanism should be solid and adaptive to any changes.
- f. Mitigation support should be provided.
- g. DDoS protection mechanism must have Low false-positive and high detection rates.

Figure 1.3 shows the architecture of the detection system.

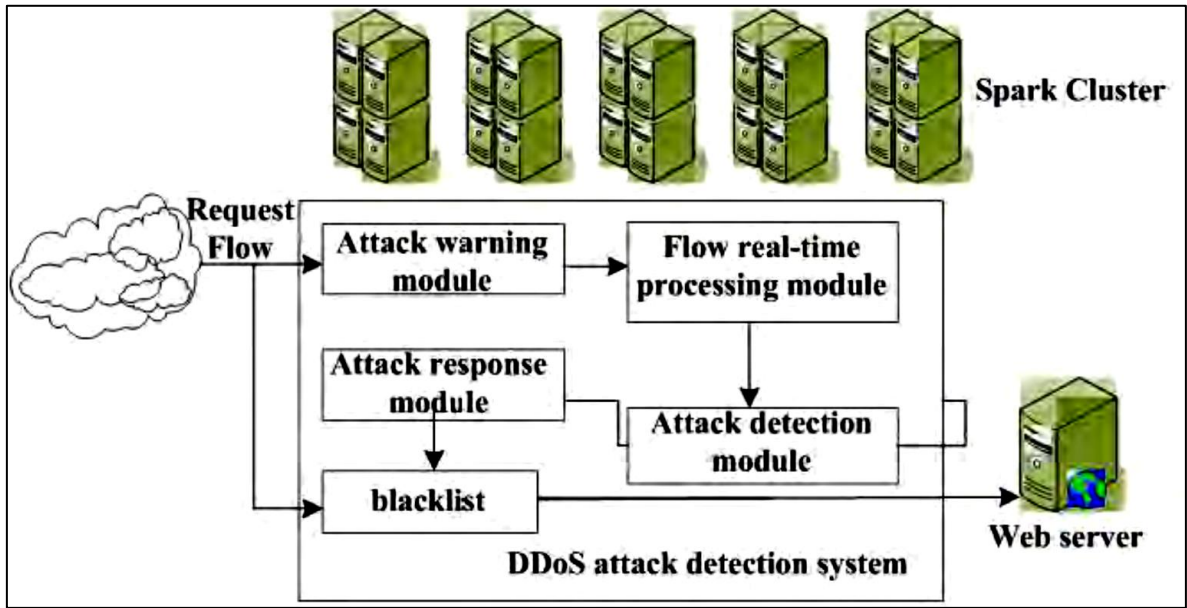


Figure 1.3: Ddos Attack Detection System [6].

Artificial Intelligence is a branch of computer science that produces intelligent machines, a system that can monitor their environment and takes over activities that increase succeeding computer models [5]. With the evolution of artificial intelligence techniques, they have been used in DDoS detection and defense, especially anomaly detection.

The most frequently artificial techniques used in DDoS detection are [6]:

- i. Bayes classification. Bayes classifiers are used machine learning methods based on the application of the Bayes theorem.
- ii. Artificial neuron network. It is composed of artificial neurons that can communicate with other neurons.
- iii. Support vector machine: it is a supervised learning model with associated learning algorithms that analyze data used for classification and regression analysis.

In addition, decision trees which are an effective and general tool for classification and prediction [7], have been lately used to detect DDoS attacks, such as in [8]–[10]. Their models showed excellent performance in the intrusion detection domain. Decision tree classifier provides a high accuracy separating the attack traffic from legitimate traffic

1.2 PROBLEM STATEMENT

The issue of DDoS attacks has been studied for a long time. These attacks aim to consume the network's resources, such as stack space, the server processor's capacity and internet connection power [11]. In addition, DDoS attacks aim to cause harm to the victim. The reasons behind these attacks might be personal reasons or for prestige. Conversely, DDoS attacks might be performed for political or material gain reasons. For its harm, these attacks must be identified, managed and looked for ways to separate legitimate

The way to rescue the network resources from these attacks is by detecting them, and then routers can be configured at the network layer to prevent DDoS. The previous studies show that there are two troubles with DDoS detection systems. First is how to separate regular packets from attack packets, and the other is how to accurately identify the different attack types [12].

Many different approaches have been proposed for DDoS attack classification and prevention. In [13], researchers build a deep learning model for DDoS attack detection. A strategy that has been presented is a complete system for detecting DDoS. This approach is based on classification algorithms like Random Forest and XGBoost.

Based on previous information, this master's research is carried out to build a model with a quality performance in the detection process using various classification algorithms. In addition, [14] addresses DDoS detection issues by developing a general model to protect a particular environment [15]–[19]. The model was designed with Sklearn tools and Python.

1.3 RESEARCH SIGNIFICANCE

The research has focused on determining DDoS attacks that consider the most significant security challenges in internet network environments. Several solutions have been proposed in this context, using different techniques such as machine learning. All solutions focused on having remarkable accuracy, but this caused the controller's overhead and false alarms to increase. Therefore, there is a need to discover a new solution which detects attacks and enhances performance. Artificial intelligent techniques have offered new ways to defend against the threat of invasions. These ways are perfectly apposite for traffic monitoring applications. The benefits of using artificial intelligence are the ability to build flexible, adaptable, and able-to-learn detection models. Classification of a large amount of different

information is the primary need for DDoS attacks. This classification is complex, so using artificial intelligent classification is the best solution for detecting several attacks.

The need for new performance ways and solutions is the starting point of this master's thesis, which looks at the different classification models used to detect the DDoS attack and preview the performance and accuracy of its model.

1.4 RESEARCH MAJOR AIM AND MINOR OBJECTIVES

The main aim of this study is to design and develop a new model for detecting different DDoS attacks using some frequent artificial intelligent techniques. This model can be applied to several network environments. The following minor objectives are performed to reach the main aim of this master's thesis:

- i. To present the existing DDoS protection mechanisms implemented on different network environments. And address the practical, intelligent algorithms used in the detection process.
- ii. To design the model to detect various DDoS attacks and analyze parameters used by the existing solutions.
- iii. To use Python code and numerical analysis to help the algorithm function correctly and accurately.
- iv. To assess the performance of the proposed model in defending different DDoS attacks.

1.5 THESIS STATEMENT

The title of this study is "ARTIFICIAL INTELLIGENT MODEL FOR DDOS-DETECTION." This study helped the detection process by developing a formal model used for defense against DDoS attacks. This model can be applied in several network environments, so it is built in with flexibility and adaptability. It is also meant to explain various stratagems that utilize protection mechanisms. This work will rely on numerical code and simulations by Python software tools.

1.6 RESEARCH QUESTION

This master's thesis seeks to answer the following research questions:

- a. What are the different approaches to making a DDoS attack? Why is DDoS a complex problem to handle?
- b. What attacks have been handled effectively by existing defence systems? What attacks remain unaddressed and why?
- c. How can the gained information about DDoS protections and detections be applied in a specific network environment?

1.7 RESEARCH ORGANIZATION

The architecture of this master's thesis is ordered as the following consequence:

- a. Chapter One has the title INTRODUCTION. It shows detailed data about the research background, significance, problem statement, primary goal and specific objectives, and research questions.
- b. Chapter Two has the title of LITERATURE REVIEW. It includes the methodology used in previous literature in the field of DDoS attack detection. It emphasizes the research which used artificial intelligence to build a model with high performance in detecting DDoS attacks.
- c. Chapter Three has the title RESEARCH METHODOLOGY. It presents the procedure steps followed to build the model by different classification algorithms. Python code is being designed to simulate the detection of DDoS attacks.
- d. Chapter Four has the title RESULTS AND DISCUSSIONS. It illustrates the substantial numerical research's outputs and provides discussions linked to the outcomes attained from the research findings from other scientists.
- e. Chapter Five has the title CONCLUSIONS AND RECOMMENDATION. It encompasses the study conclusion based on the numerical results and simulations. Also, it suggests and recommends vital aspects for the future development of the resulting model.

2. LITERATURE REVIEW

2.1 GENERAL

This chapter offers detailed information that aims to provide the groundwork for DDoS attacks and defense techniques. Displayed information was focused on the artificial intelligence algorithms used in the detection process. Besides that, it addresses how different artificial intelligence algorithms are used for DDoS detection. The material generated by the literature review gives the foundation for the thesis research base, theoretical background, and related work. The information was collected from articles, research papers, and academic publications that can be found in web journals. Even though identifying viruses and web attacks that are difficult, complex, and challenging to detect, modern algorithms could provide solutions for this difficult task.

2.2 CRITICAL CONTRIBUTIONS OF AI ALGORITHMS FOR DDOS DETECTION

Researchers in [20] presented a survey which showed that innovative artificial methods had improved cyber threats. The paper also showed several techniques had contributed to improving intrusion detection systems. Most research focused on intrusion detection and prevention systems using support vector machines. This paper gives brief information about quantitative and qualitative methods which are used in the area of IoT, intrusion detection systems, software-defined network attacks, prevention of cyber attack, and crimes. The research mentioned in this survey has used machine learning algorithms that use training data for building mathematical methods. In addition, it addressed the three main types of machine learning supervised learning, unsupervised learning, and reinforcement learning. In cloud computing, a new DDoS attack is called EDoS, and [21] proposed a system to detect this attack. The detection system was produced depending on SVM, KNN, and RF tree algorithms. This system deals with two situations for the detection process, the first one is binary classification, and the other is multi-classification. Binary classification has two classes (attack and regular), and multi-classification has nine categories. The main aim of this study is to find a performance model for detecting DDoS attacks by combining machine learning with deep learning. RF tree algorithm is the best in binary situations. However, the SVM is the best algorithm in multi cases.

Experiments in [22] have been done with several machine learning algorithms to show the strength and weaknesses of DDoS detection. There are four experiments in this paper first one is comparative. The next is the impact of observable traffic proportions. The third is the impact of attack intensities and the final experiment for the effect of the class imbalance problem. Most of the models in DDoS detection and classification follow the workflow presented in Figure 2.1.

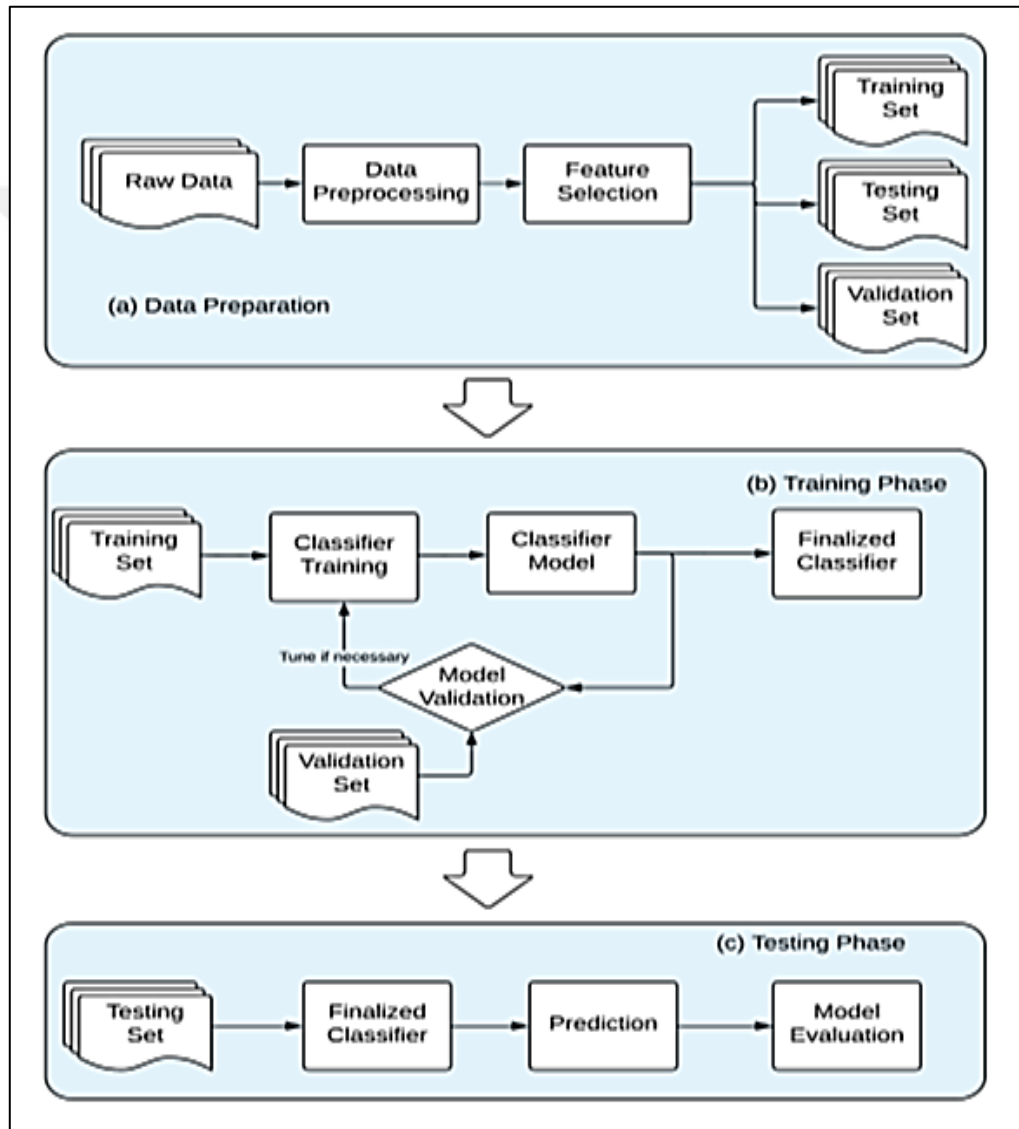


Figure 2.1: The Workflow of Machine Learning classification Techniques [22]. To evaluate the performance of detection models, the researchers used the following measures: Accuracy, Sensitivity and Specificity.

In [23], researchers built a model using binary classification. The model is based on deep learning and selects a group of features to evaluate DDoS detection. The proposed model in this study is called CyDDoS. This model combines the ensemble of feature engineering algorithms with a deep neural network. Twelve experiments are done to choose the best group of features to get the top classification performance. Four machine learning algorithms have been trained. They are random forest, Logistic Regression, Perceptron, and Gaussian Naive Bayes. This paper also compares the new model with classical machine learning and deep learning systems. They used the recall, precision, f1-score and ROC curve for comparison. The results showed that their model has best results than the classical machine learning and deep learning systems. The model in this paper was a Software-Defined Network (SDN) application. It predicted the DDoS attacks with significant rates of accuracy.

[24] proposed a new framework which focuses on the fast detection process. The framework is AIMM which means Artificial Intelligence Merged Methods. This framework is flexible to adding any machine learning algorithm in the training process. They used two machine learning algorithms: KNN and ANN. The types of attacks detected with this framework are TCP-flood and UDP-flood. Figure 2.2 shows three stages for the proposed model: preprocessing data, classifying data using artificial intelligence and making decisions.

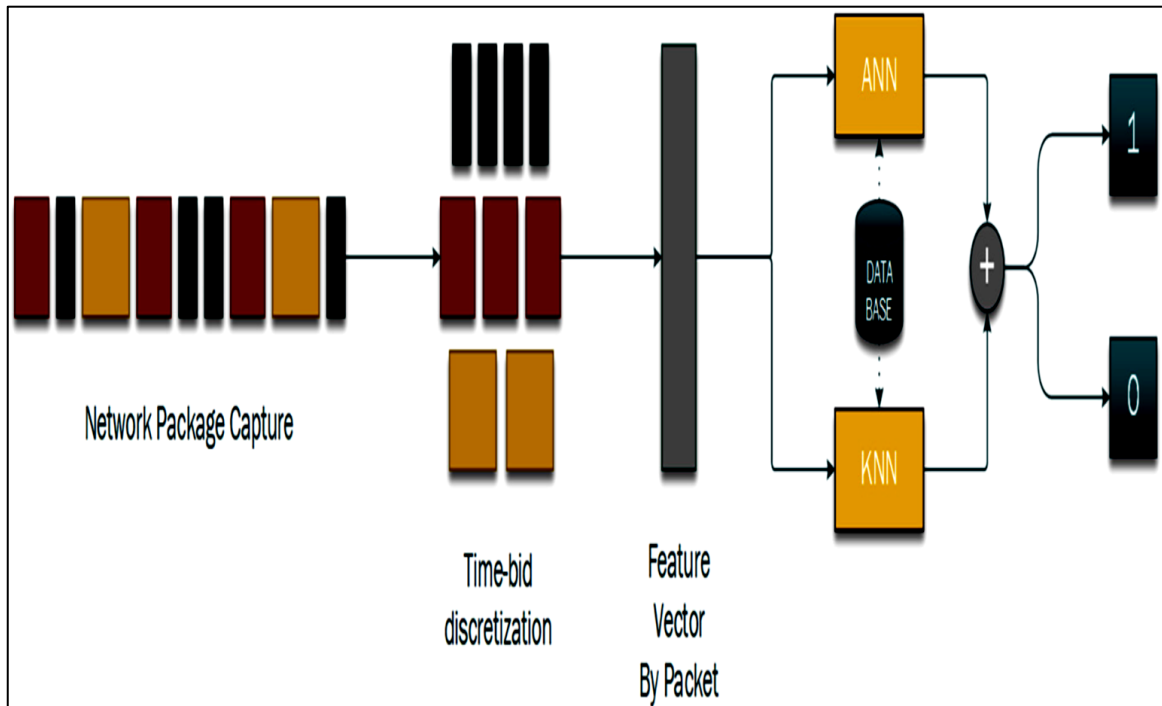


Figure 2.2: The Main Stages of AIMM [24].

[25] discussed using different machine learning algorithms to detect several DDoS attacks, such as NTP, DNS, and LDAP. In addition, it compared between K-Nearest Neighbors, Decision Tree, Random Forests and Artificial Neural Networks. The result of the experiments noted that the Artificial Neural Network algorithm is the best, with an accuracy of 99.95%. For feature selection, the researchers used the ExtraTreesClassifier.

[26] presented a framework which depends on a deep neural network classifier. It involves one classifier and two generators. This research employs the Generative Adversarial Network (GAN) as its framework. In this paper, the researchers tried to avoid the problem using classical machine learning and deep learning to detect DDoS. This problem appears when the input feature space is changed. They used Adversarial Machine Learning (AML) techniques. The generator in this search tested the classifiers to check reliability in detecting attacks.

[27] focused on Artificial Neural Network (ANN) classifiers. This paper highlights a significant point that is successful in any detection system based on selecting the correct algorithm. Five machine learning algorithms used with big datasets have been implemented in this search: Support Vector Machine (SVM), Decision Tree, Random Forest, Naïve Bayes and ANN. The results in this paper showed that SVM has a significant delay. However, the probabilistic accuracy for all algorithms was 99%. One of the reasons behind this success in probabilistic accuracy is the ratio of training and testing data. The ANN algorithm also had the best result using two neural network layers.

In [28], five supervised machine learning algorithms have been trained. These algorithms are Bayesian Network (BayesNet), Bootstrap Aggregating (Bagging), k-Nearest Neighbors (kNN), Sequential Minimal Optimization (SMO), and Simple Logistic (SL). Training the algorithms was done by using the same dataset called CICDDoS2019. The proposed system in this paper used 24 features to detect 12 DDoS attacks, and the accuracy of each classifier has been measured in the experiments done by the WEKA tool. The model was proposed in [10] based on an ensemble classifier consisting of four machine-learning classifiers that work in parallel. Figure 3 shows the building model in this search.

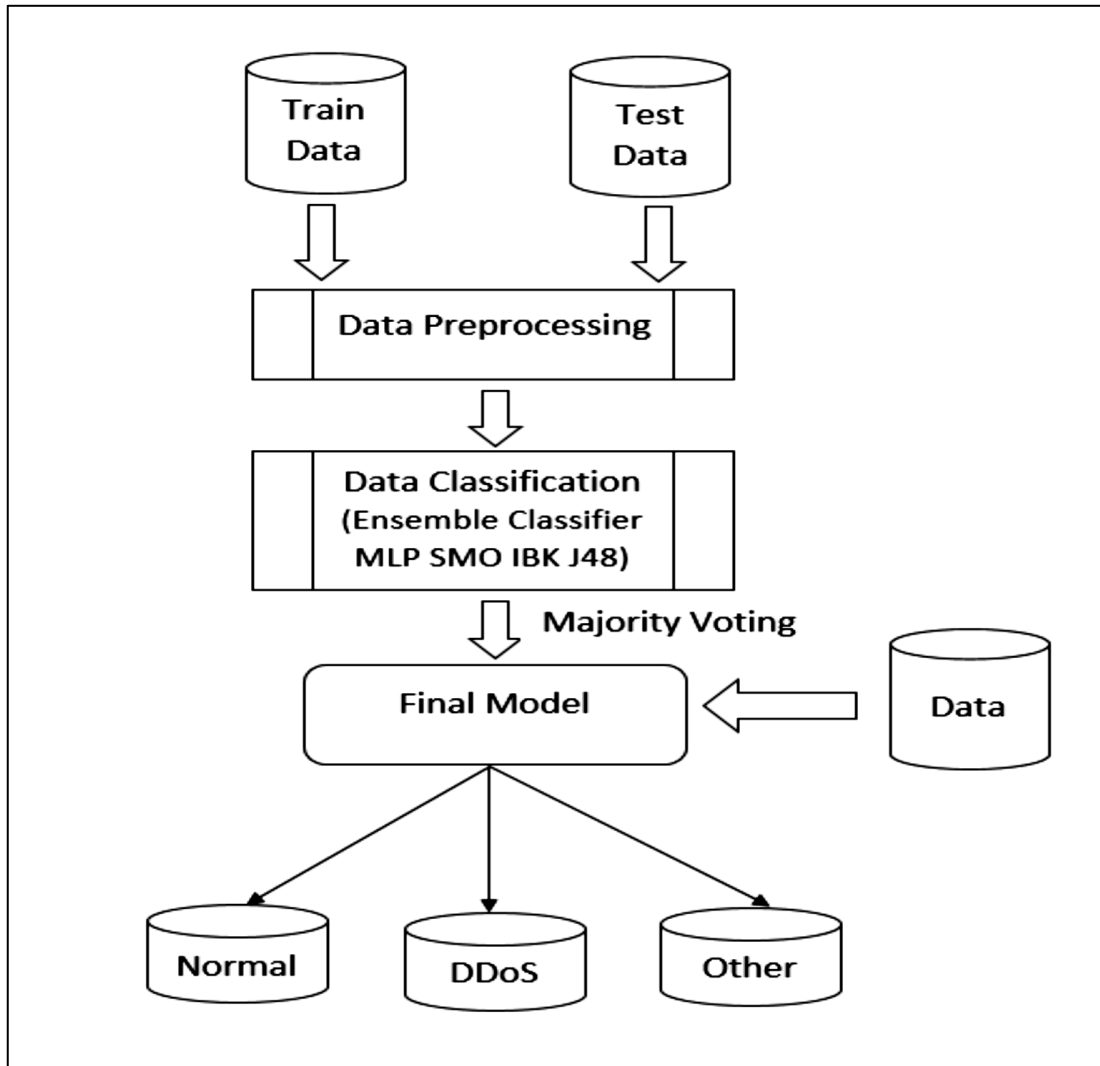


Figure 2.3: The Main Steps Flow In The Detection Model Based On The Ensemble [29].

The model has two primary processes, data preprocessing and data classification. The dataset used in this paper is NSL-KDD, which is divided into three groups: a training dataset, a test dataset, and a validation dataset. This dataset had been classified on each classifier individually. The machine learning classifiers chosen for use in this model are MLP (NN), SMO (SVM), IBK (KNN) and J48 (DT-C4.5). These classifiers were from different classifier families. To combine the classifier, Majority voting in the WEKA tool was used. Their experiments showed that the detection DDoS system with an ensemble model is better than the detection system with a single classifier.

[30] gives a review of existing DDoS attack detection systems. It focused on artificial intelligence and statistical approaches illustrated in Figure 2.4.

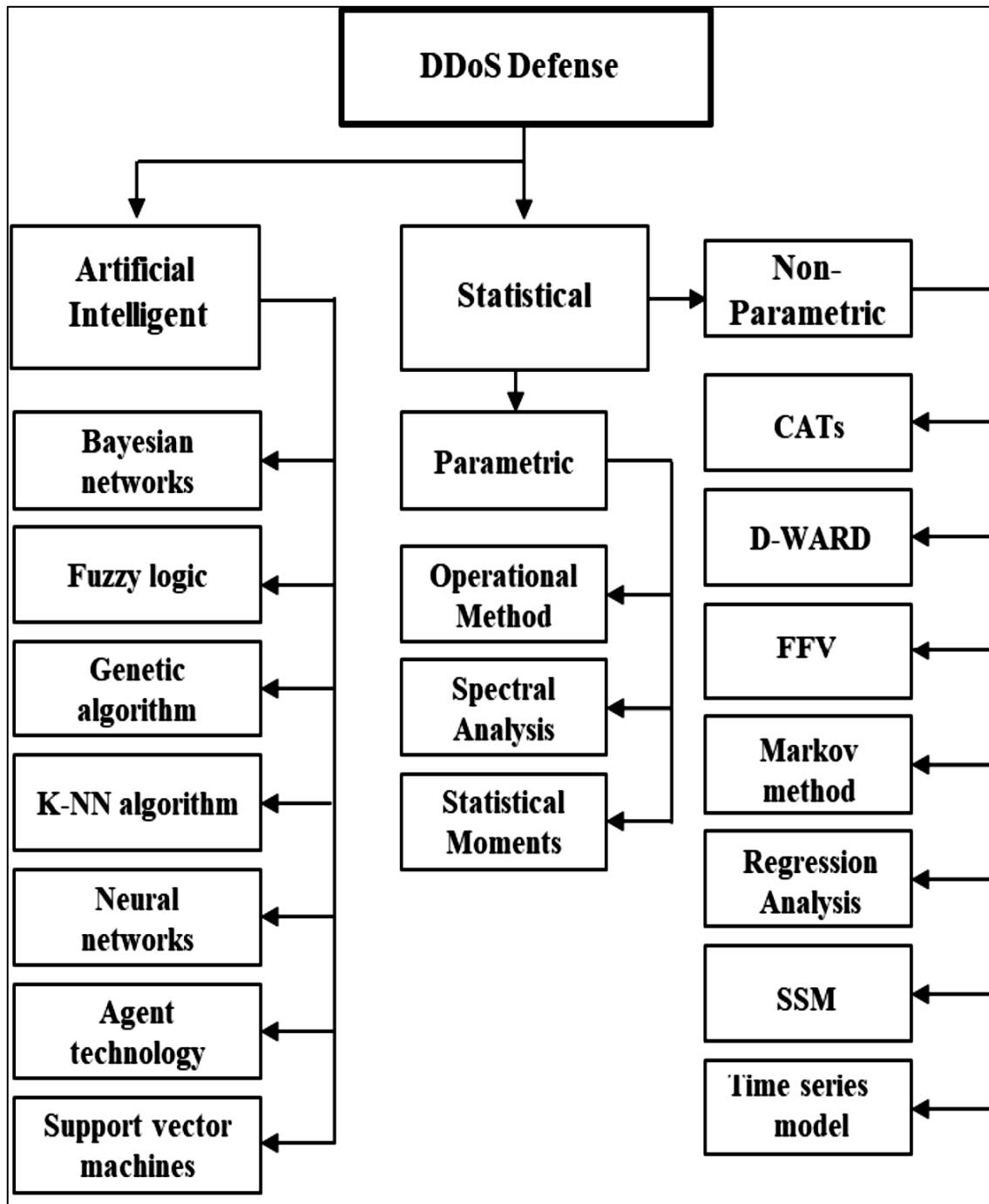


Figure 2.4: A Taxonomy of Primary Ddos Defense Sorting [30].

Researchers [31] presented a new model-sharing machine learning with attribute-based IoC. Their MISP (Malware Information Sharing Platform) has described the input feature vector and the output labels. Searching aims to find the best model that uses machine learning or deep learning. The first classifier that was shared is the decision tree. A random forest

classifier was trained using 20 decision trees. In addition, they introduced the SVM classifier with a group of benign feature vectors.

In many other real-life applications, detection systems with artificial intelligence have been used. [32] and [33] used artificial intelligence to detect disease in plants and detect attacks in banking sectors. The model presented for plant disease [32], [33] has the following workflow: first, the image acquisition. Second, the image preprocessing, then image segmentation. After feature extraction, the final is the detection and classification of plant diseases. The classifier used in this model is Convolutional Neural Network (CNN) with deep learning. The results appeared that artificial intelligence models are great at disease detection. In [33], researchers built their models using the banking dataset to monitor network intrusions and train the choosing classifiers with this dataset. Researchers used KNN, SVM, and Random Forests to classify DDOS attacks in the Banking Sector Dataset. In the banking sector, the time complexity is the main feature, and the results show that SVM is better than other used classifiers. In [34] deep learning model was built to detect DDoS in social media applications. The dataset used in this paper is from Twitter. It trained the improving CNN model and LSTM neural network model. The architecture of the proposed CNN consists of 13 layers. The architecture of improved LSTM has three layers: input layer, hidden layer and output layer. The memory block is the primary unit of the hidden layer. This block contains memory cells and two adaptive, multiplicative gating units. The experiments presented good results with the improved models based on the following evaluation matrices recall, precision, F-measure, training loss, training accuracy, testing loss, and testing accuracy. Three classifiers have been used in the detection system [35]. The machine learning classifiers used in the detection system are Decision Tree, Naïve Bayes and Artificial Neural Network. In their model, they used an open-source dataset of network traffic. Each row of data has the following: time, sender IP address, data flow, number of packages and bytes per packet sent—the data preprocessing by sorting it using simulation tools. In addition, this paper compared the three classifiers by using performance measures: predictive accuracy, recall, precision, and F-measure. Based on the accuracy results, an artificial neural network classifier is the best choice for the detection system. The problem of DDoS attacks has been studied in the cloud environment [36]. The proposed model in this paper used multiple linear regression analysis. They used CICIDS 2017 dataset, and for

feature selection, they used Information Gain (IG) approach. Selecting features are used in the machine learning model, built using linear regression analysis. The main aim of using IG is to choose the critical attributes of the dataset to deliver it to the detect model.

[37] presented a new model related to deep hybrid learning. It contained a convolutional neural network (CNN). It also comprised Bidirectional Long Short-Time Memory (BiLSTM) learning algorithms. The framework of the proposed model is: applying the χ^2 test to select the high-rate features, then using CNN to extract the particular features, and finally, the elements into the BiLSTM. Figure 5 shows deep learning with a hybrid technique.

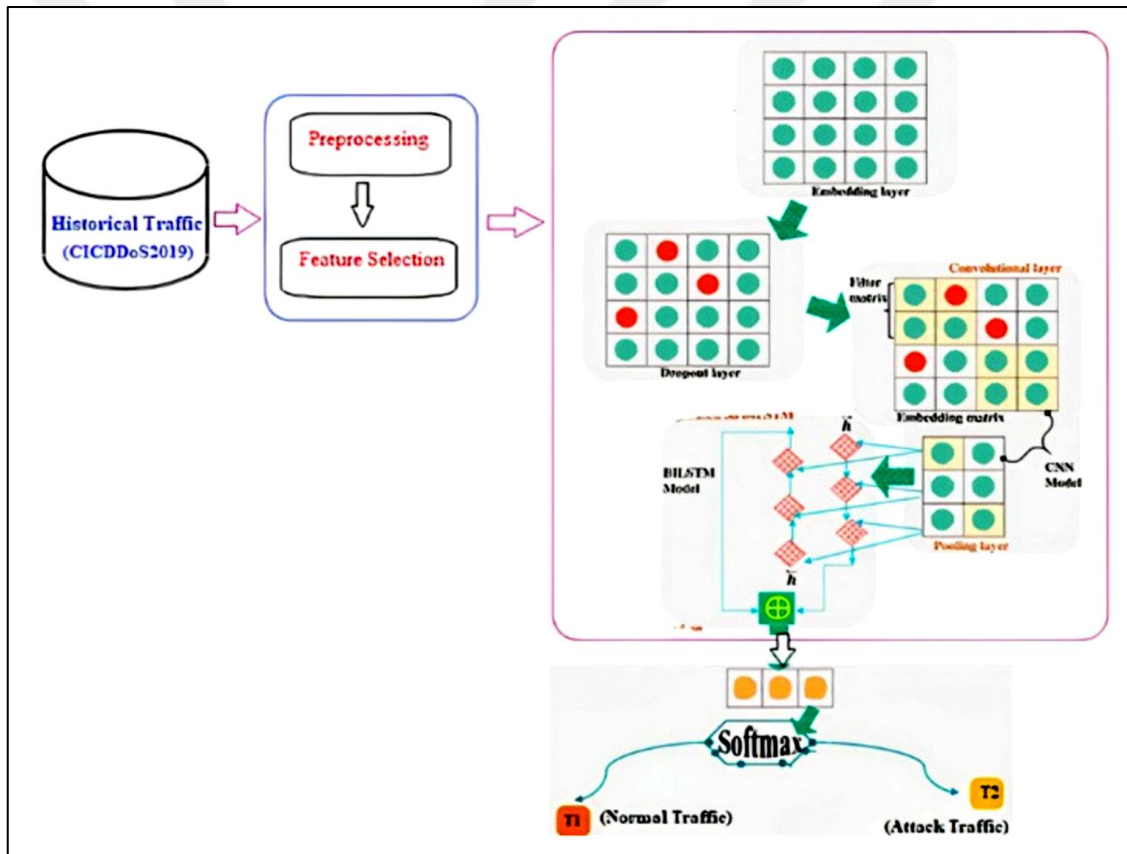


Figure 2.5: A Ddos Detection System With Deep Hybrid Learning [37].

For getting high performance at prediction DDoS using this model, searchers built a user-friendly interface. The experiments with this model have been done using Python on the TensorFlow 2.0 and Keras platforms. The results showed that the accuracy would increase if the unit size of BiLSTM was less. To get the answer to how CNN and BiLSTM models can function, diverse parameters were considered in their testing process. These parameters

include vocabulary size, input vector size, embedding dimension, unit size of BiLSTM and many others. In this paper, also a comparison between their model of deep hybrid learning and classical machine learning has been made. CNN with BiLSTM gets a higher accuracy than Linear Regression, SVM, XGBoost, Random Forest, and Logistic Regression.

Two methodologies have been proposed in [38], one using a hybrid Intrusion Detection System (IDS) and the other using a deep learning model. This paper focuses on detecting a special kind of DDoS attack: Distributed Reflection Denial of Service (DrDoS) attack. The researchers highlighted the advantages and weaknesses of each IDS class: signature-based method and anomaly-based. They suggested combining the signature and anomaly to avoid the weak point of each one. In the hybrid methodology, if the sensor doesn't detect the attack and passes the signature system without noticing, then this attack will monitor carefully with the anomaly-based detector. This model has been implemented using Java programming language. In the deep learning model, they used Long Short-Time Memory (LSTM) and Recurrent Neural Network (RNN). The reason behind choosing these methods is the capability of solving the vanishing gradients problem with networks.

[39] proposed a DDoS detection model using an SVM classifier and a Logistic Regression classifier, and they compared them. The dataset CIC-DDoS2019 was used in their experiments. They used Recursive Feature (RF) algorithm to select small features from the data. At the preprocessing, all rows of data that are null or non-informative are dropped. It is noted from their results that the SVM in the detection system is more accurate than Logistic Regression, but the researcher focused on the point that this comparison cannot be the conclusion. The best accuracy depends on the selected features.

In [40], DDoS attacks detected using wavelet analysis depended on the fact that abnormal traffic can be founded using energy distribution. Wavelet analysis can catch complex temporal correlations. In this search, the network traffic simulation has been done with the NS simulator. Energy distribution was used in the simulation successfully. The results showed a noticeable change in the energy distribution at the attack traffic.

In [41], [42] new method for DDoS attack detection has been proposed. The workflow of the method has four significant steps they are: data collection, dataset preprocessing, classification and finally, and evaluation. The datasets used in the experiments are CICIDS

2017 and CICDDoS 2019, and the classifiers are Logistic Regression, K Nearest Neighbor, Gradient Boosting, Random Forest, and Weighted Voting Ensemble Classifier. Features used with the classifier were selected by Mutual Information (MI) method. The aims of establishing characteristics are: to improve the performance, give robust generalization, and understand the data. [42] presented a model of attack detection using seven machine-learning classifiers. The proposed model has the following workflow: first static monitors, second feature extractor, and finally, machine learning classifier. They used both types of machine learning, supervised and unsupervised. The confusion matrix is used to view the performance of the machine learning classifiers. In this paper, four types of DDoS attacks have been used in the confusion matrix. These are UDP Flood, HTTP Flood, Smurf and SiDDoS. The results showed that Linear Regression is the best with high accuracy, and Naïve Bayes is the worst.

[20] presented a review of different approaches in detection, which are quantitative and qualitative in the fields of IoT, Blockchain, Cybercrimes, Business, IDS, Software-defined networks and cyber forensics. It summarized the artificial intelligent approaches to help researchers in this field. The [43] survey of using deep learning in a DDoS detection system has also been presented. The main objectives of this survey are to: identify the deep learning approaches for DDoS detection, highlight the strong and weak points in the existing deep learning approaches, summarise the datasets and the classes of attacks in them, and finally points to the future directions to improve the using of deep learning.

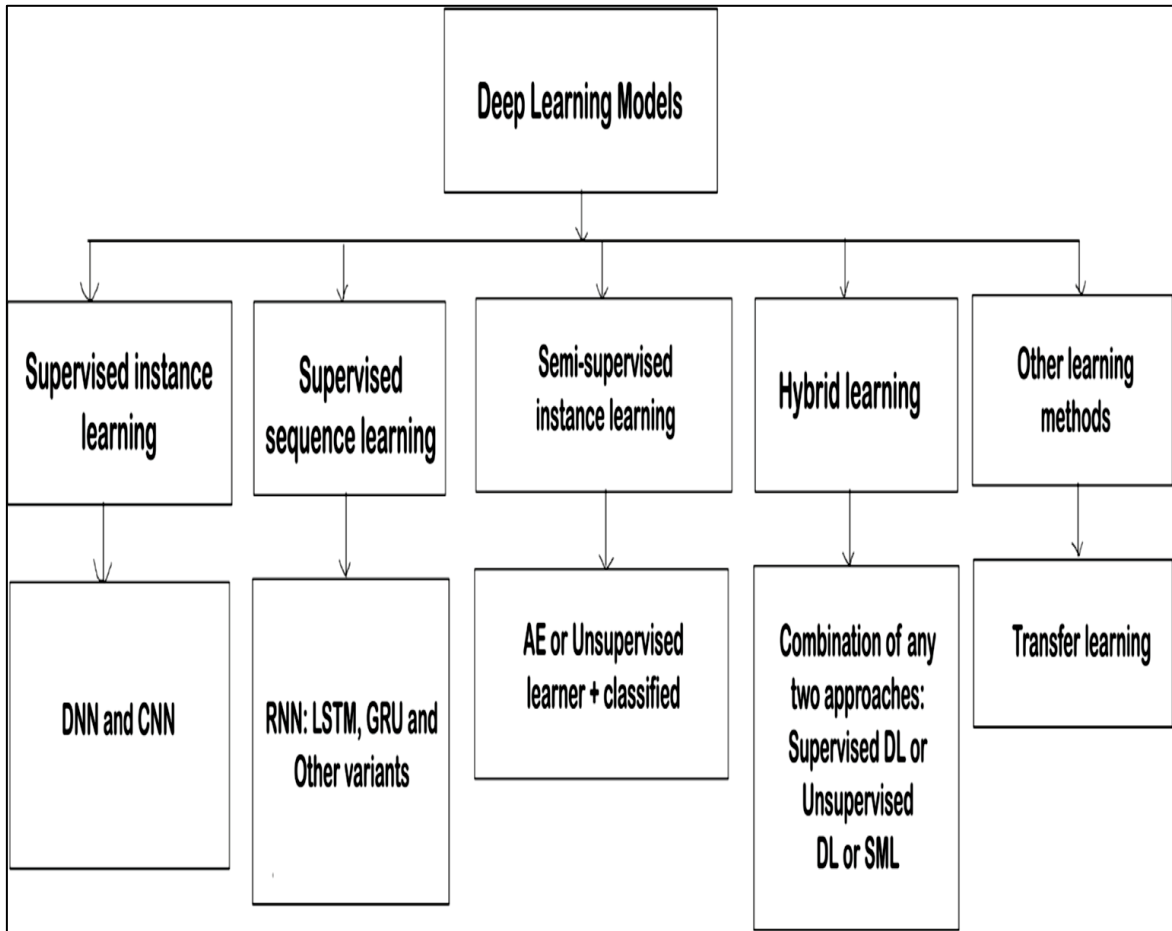


Figure 2.6: A Classification of Deep Learning Models [43].

[44] answered the question about the frequent use of artificial intelligence in the security environment, specifically in detecting and predicting attacks. They mentioned three types of DDoS attack detection and the searches that show the strengths and weaknesses of each. They discussed the following artificial intelligent approaches: supervised learning, unsupervised learning, Probabilistic learning, soft computing, Knowledge-based anomaly detection, and Combination learners. They mentioned an important point: the performance and accuracy of any learning classifier depend on the monitoring parameters by the detection system. In [45], researchers highlighted the attacks at the application layer and reviewed the existing detection systems. The types of DDoS attacks mentioned in this survey are Session Flooding Attacks, Request Flooding Attacks, Asymmetric Attacks, and Slow Request/Response Attacks. This survey also proposed the main phases of the detection systems: 1) Prevention, 2) Mitigation, 3) Detection 4) Monitoring.

[46] helped the researchers in the DoS/DDoS detection field. They showed studies in different classifications and divided them into general DoS classification and network flooding DoS-based, highlighting each limitation. They mentioned several issues for getting reliable detection systems, such as the location of the machine doing the detection process. Also, the kind of services that the detection system gives, a choice for the methodology, such as anomaly-based or signature-based, and choosing the proper statistic method among the wide available range present further issues associated with DDoS cyber threats.

2.3 LITERATURE REVIEW SUMMARY

The literature review publications reviewed in this work can be summarized in Table 2.1.

Table 2.1: Critical Findings Obtained From The Articles Reviewed in Chapter Two.

#	Author(s)	Year	Paper Title	Major Contributions
1	Tao, Feng, Muhammad Shoaib Akhtar, and Zhang Jiayuan	2021	The future of artificial intelligence in cybersecurity: A comprehensive survey	It helps researchers to gain knowledge about the critical research that has been done to study and confront cybercrimes.

Table 2.1: Critical Findings Obtained from The Articles Reviewed in Chapter Two. “Table continued.”

2	Aldhyani, Theyazn HH, and Hasan Alkahtani	2022	Artificial Intelligence Algorithm-Based Economic Denial of Sustainability Attack Detection Systems: Cloud Computing Environments.	The results of the experiments that have been done on the building model in this paper showed that the Rain Forest algorithm is the best at binary classification with an accuracy of 98%. In addition, it was found that the SVM algorithm had the best performance in multi-classification with an accuracy of 97.54%.
3	Liang, Xiaoyu, and Taieb Znati	2019	On the performance of intelligent techniques for intensive and stealthy DDos detection	This study presented details about the advantages and disadvantages of many techniques in DDos detection
4	Ivandro Ortet Lopes, Deqing Zou, Francis A Ruambo, Saeed Akbar, and Bin Yuan	2021	Towards effective detection of recent DDos attacks: A deep learning approach	CyDDoS can be used as an SDN application. The evaluation results presented that the accuracy of the constructed model CyDDoS is 99.6%. Also, it was found that the ROC value is nearly 1.

Table 2.1: Critical Findings Obtained from the Articles Reviewed in Chapter Two.“Table continued.”

5	Jaszcz, Antoni, and Dawid Połap.	2022	AIMM: Artificial Intelligence Merged Methods for flood DDoS attacks detection	This study focuses on the fast detection process. Modifying the framework obtained in this search is easy by adding any classifier and increasing it.
6	R. Amrish, K. Bavapriyan, V. Gopinaath, A. Jawahar, C. Vinoth Kumar	2022	DDoS detection using machine learning techniques	From the results of this study, it was found that the accuracy of the ANN classifier was the best compared with other classifiers. The accuracy of this algorithm recorded a value of 99.95%.
7	Jugal Shroff, Rahee Walambe, Sunil Kumar Singh, and Ketan Kotecha	2022	Enhanced security against volumetric DDoS attacks using adversarial machine learning	The model building in this study focuses on detecting volumetric DDoS attacks. The results showed the prediction reliability of this model.
8	Khalaf, B. A., Mostafa, S. A., Mustapha, A., Mohammed, M. A., & Abdullah, W. M.	2019	Comprehensive review of artificial intelligence and statistical approaches in distributed denial of service attack and defense methods	This review provides the researchers with many studies methods in DDoS attack detection and comparative among them. It mentioned the classes of DDoS attacks according to the type of vulnerability, the degree of automation, impact, and dynamics.

Table 2.1: Critical Findings Obtained from The Articles Reviewed in Chapter Two. “Table continued.”

9	Hussain, Yasar Shahid.	2020	Network Intrusion Detection for Distributed Denial- of-Service (DDoS) Attacks using Machine Learning Classification Techniques.	This study did a lot of experiments on many classifiers then the average accuracy was calculated. The average shows that the Bagging classifier is the best. And Simple Logistics is the worst.
10	Saikat Das, Ahmed M. Mahfouz, Deepak Venugopal, Sajjan Shiva	2019	DDoS intrusion detection through machine learning ensemble.	The experiments indicated that combining the machine learning classifier using the 10 -folds cross- validation has better detection accuracy with a lower false positive rate than the model that depends on individual classifiers.
11	Veranyurt, Ozan.	2019	Usage of Artificial Intelligence in DOS/DDOS Attack Detection.	The main point of this study is the detection of attacks at application layers. The result showed that ANN, Decision Tree and Random Forest have perfect accuracy. At the execution time, ANN demonstrated the margin result with a time of 1.39 sec.

Table 2.1: Critical Findings Obtained From The Articles Reviewed In Chapter Two. “Table continued.”

12	Sujawat, Gyan Singh.	2021	Application of Artificial Intelligence in detection of diseases in plants: A Survey	The survey highlighted most of the research that talked about the models that were built to detect plant diseases.
13	Umar Islam, Ali Muhammad, Rafiq Mansoor, Md Shamim Hossain, Ijaz Ahmad, Elsayed Tag Eldin Javed Ali Khan, Ateeq Ur Rehman and Muhammad Shafiq	2022	Detection of distributed denial of service (DDoS) attacks in IOT based monitoring system of banking sector using machine learning models.	It can be seen from the results that SVM is more robust as compared to KNN, RF and existing Machine Learning (ML/DL) approaches. And it's the best choice for the banking dataset.
14	Tao, Feng, Muhammad Shoaib Akhtar, and Zhang Jiayuan.	2021	The future of artificial intelligence in cybersecurity: A comprehensive survey.	This study presented most of the models that have been built for cyber-security. The main aim is to help researchers who are studying cybercrimes.

Table 2.1: Critical Findings Obtained From The Articles Reviewed In Chapter Two. “Table continued.”

15	Preuveneens, Davy, and Wouter Joosen.	2021	Sharing machine learning models as indicators of compromise for cyber threat intelligence.	This work proposed a framework for sharing machine learning models. It is clear from the experimental results that traditional machine learning and deep learning classifiers misclassify malicious traffic
16	Alguliyev, Rasim M., Ramiz M. Aliguliyev, and Fargana J. Abdullayeva.	2019	Deep learning method for prediction of DDoS attacks on social media.	This work referred to the sentiment analysis of the text on social media and the detection of the negative and positive classes. The results showed the model using CNN, and LSTM has a high performance in the detection system.
17	Liu, Zhongbang, Likun Qian, and Shiwen Tang.	2022	The prediction of DDoS attack by machine learning.	The results presented that the model in this search, PCA-SVM, has high accuracy at a low cost. The main point of building this model is to search for a low-cost system and high performance

Table 2.1: Critical Findings Obtained From The Articles Reviewed In Chapter Two. “Table continued.”

18	Daniyal Alghazzawi, Omaimah Bamasag, Hayat Ullah and Muhammad Zubair Asghar	2021	Efficient detection of DDoS attacks using a hybrid deep learning model with improved feature selection.	The results of many experiments in this search showed that combining two deep learning like, CNN and BiLSTM, outperforms the other deep learning models and many classical machine learning models.
19	Mittal, Meenakshi, Krishan Kumar, and Sunny Behal.	2022	Deep learning approaches for detecting DDoS attacks: A systematic review	This study summarized many models that used deep learning in the detection system and pointed to different models’ weaknesses.
20	Mona Alduailij, Qazi Waqas Khan Muhammad Tahir, Muhammad Sardaraz, Mai Alduailij and Fazila Malik	2022	Machine-Learning-Based DDoS Attack Detection Using Mutual Information and Random Forest Feature Importance Method.	The experiment results illustrate that different machine learning classifiers give high accuracy and performance when using feature selection techniques

Table 2.1: Critical Findings Obtained From The Articles Reviewed In Chapter Two. “Table continued.”

21	Li, Lan, and Gyungho Lee.	2005	DDoS attack detection and wavelets	This work focused on the energy distribution analysis and showed it could characterize traffic behavior under dynamic conditions and outperform other existing schemes.
22	Alenezi, Mohammed, and Martin J. Reed.	2012	Methodologies for detecting DoS/DDoS attacks against network servers	This literature review can provide the researchers with a recognition of various techniques for detecting attacks.
23	Vuković, Igor.	2020	Application of Artificial Intelligence in Detection of DDoS Attacks	It pointed to the answer question of the importance of AI in the detection system
24	Jaafar, Ghafar A., Shahidan M. Abdullah, and Saifuladli Ismail.	2019	Review of recent detection methods for HTTP DDoS attack	This literature review summarized 12 recent detections of DDoS attacks at the application layer

Table 2.1: Critical Findings Obtained From The Articles Reviewed In Chapter Two. “Table continued.”

25	Kshira Sagar Sahoo,Amaan Iqbal, Prasenjit Maiti, Bibhudatta Sahoo	2018	A machine learning approach for predicting DDoS traffic in software defined networks.	The results of the experiment exhibited that the accurate attack detection system depended on the use of the proper machine learning
26	Shurman, Mohammad M., Rami M. Khrais, and Abdulrahman A. Yateem.	2020	DoS and DDoS attack detection using deep learning and IDS	The experiments result showed that two methods proposed in this study, which is: a hybrid-based IDS for IoT networks and a deep learning model based on LSTM have high accuracy
27	Sanmorino, Ahmad.	2019	A study for DDOS attack classification method.	This study compared machine learning in term of DDoS attack detection. The results showed that the artificial neural network is the best choice.
28	Sambangi, Swathi, and Lakshmeeswari Gondi.	2020	A machine learning approach for DDoS (distributed denial of service) attack detection using multiple linear Regression.	It focused on measuring the model’s accuracyel with selected attributes from the dataset. The results showed that the model with 16 characteristics has the high accuracy

3. RESEARCH METHODOLOGY

3.1 CHAPTER GOAL

The third chapter illustrates and elaborates on the stages of a research project and primary practices and strategies used and implemented to carry out the numerical investigation and simulation method through the use of a Python code established to validate the vital contributions of DL and ML in forecasting cybersecurity risks and internet attacks using intelligent AI algorithms.

3.2 RESEARCH METHOD

For this master's thesis, the first stage adopted was the definition of the problem or (Problem Analysis). In this phase, the kind of attack and its major merits were analyzed and defined. A case study is considered, which depicts the airplane network influenced and attacked by a cyber threat that may harm the airplane system, passengers' data privacy, flight employees' and managers' information security, and the safety of some critical data associated with all parties and individuals linked to the aviation sector. In this context, intelligent ML algorithms and DL methods are implemented to help detect severe cyber threats and internet attacks that may result in dangerous financial resources or database damage pertaining to the aviation field. The Python programming language will be used to create, write, and process the numerical code that will carry out the numerical investigation and mathematical simulation, verifying and validating the beneficial impacts of ML and DL algorithms. Following this research step, the numerical outputs and mathematical work will be refined, corrected, and validated via a panel of ML experts, AI professionals, and DL engineers. After conducting the validation of the thesis's results, the conclusions and vital recommendations and suggestions will be drawn to help scholars implement novel ML and AI techniques in detecting cyber threats flexibly. Also, those recommendations are substantial to help high officials and managers working in aviation and pilots protect their airplane networks and secure them from cyber threats that could damage their information security and the private data of passengers. The research methodology flowchart is illustrated in Figure 3.1.

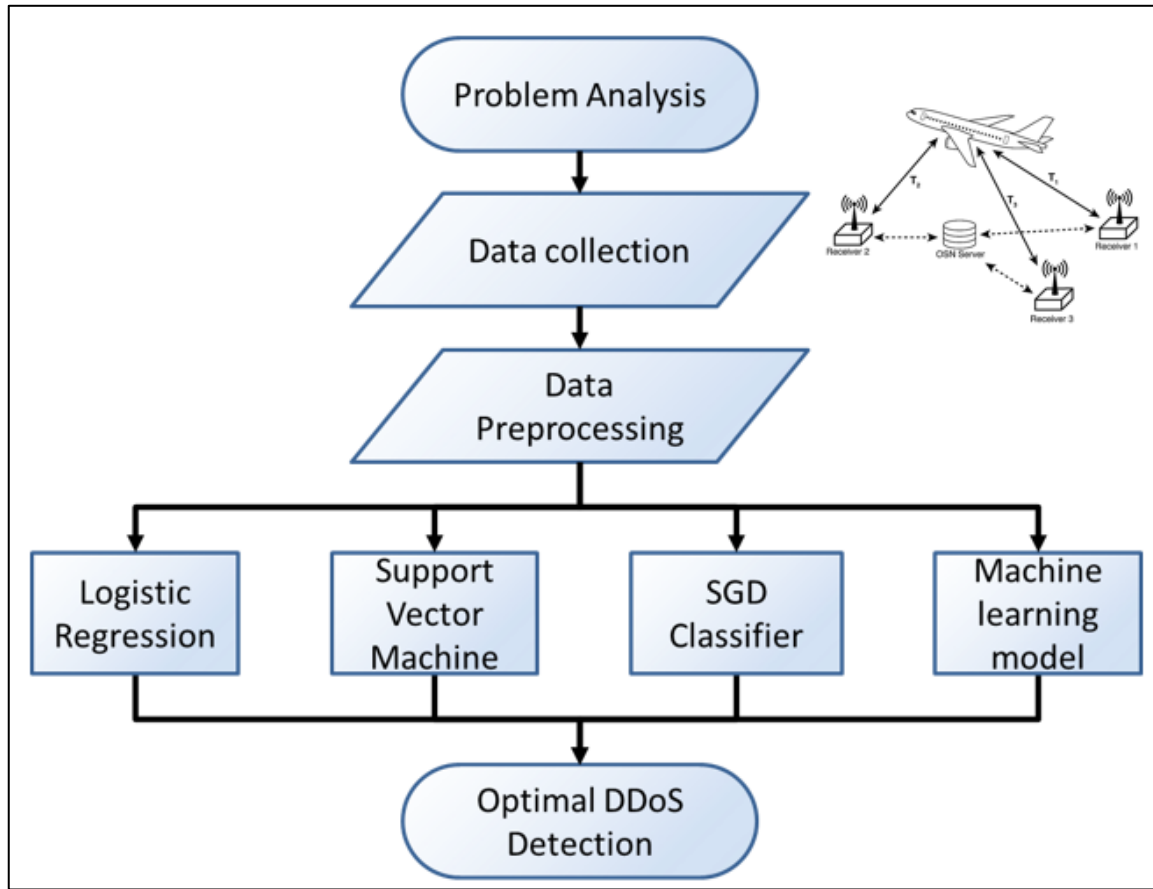


Figure 3.1: The Research Methodology of This Work.

3.3 RESEARCH PHILOSOPHY

The philosophy of this work is mirrored by gathering two major categories of data, which include primary and secondary data. According to the secondary data collection, different databases and critical information will be collected from global journals and various peer-reviewed publications in the web to address and highlight the beneficial roles of DL and ML algorithms and intelligent AI principles in securing the aviation databases and protecting airplane internet networks from harmful cyber attacks.

On the other hand, the primary data collection process is carried out to analyze the active contributions and crucial rationale of DL approaches and ML principles in detecting complex internet threats and difficult-to-distinguish aviation attacks that may put the databases and private information of passengers, pilots, managers, and other aviation employees in critical and challenging situations. The dataset will be chosen and numerically analyzed to examine the advantageous role of those intelligent ML algorithms in identifying different harmful attacks affecting the aviation industry.

It is worth mentioning that this master's research will depend on various dependent and independent variables which can illustrate the efficacy and performance of vital DL and ML models in achieving accurate and practical means of cyber threat detection. The research's dependent and independent variables can be described in Figure 3.2.

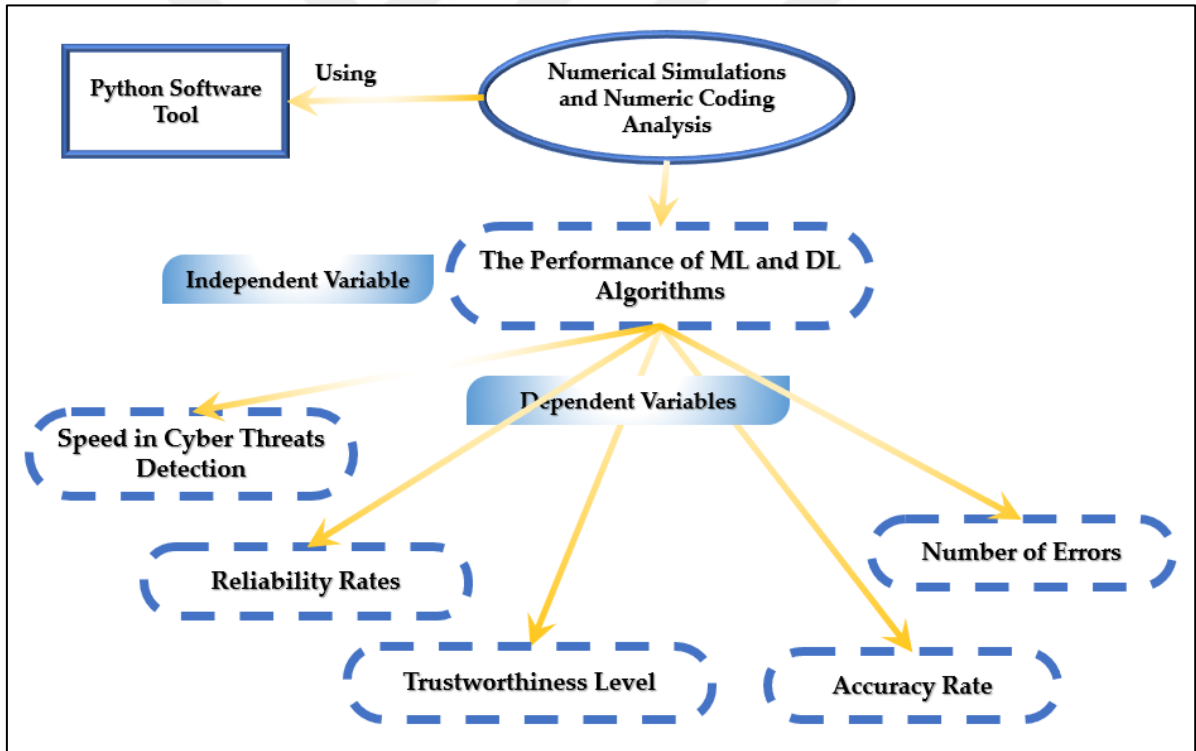


Figure 3.2: The Research's Independent and Dependent Variables Addressed in This Work.

3.4 DATASET DEFINITION AND PREPARATION

One of the worst and most common forms of a cyberattack on online services is the Dedicated Denial of Service Attack (DDoS). The attackers' goal is to disable the victim's system so that the service they were providing cannot be maintained. In this research, the

identification of cyber threats related to DDoS will be executed with the help of ML and DL principles. False features and harmful signals will be considered in this investigation [47].

There are a variety of available databases collecting information on different DDoS assaults. However, the present study will focus on analyzing the beneficial impacts of ML and DL implementation in identifying SDN-based attacks. The associated dataset was obtained using the link . This dataset contains both normal TCP, UDP, and ICMP traffic and malicious traffic with TCPS Syn, UDP flood, and ICMP assaults. This dataset comprises 23 extracted switch characteristics. Table 3.1 offers a detailed list associated with the characteristics and descriptions of different elements in this dataset. There are 104,345 rows in the dataset, with 63,335 benign and 40,504 harmful cyber accidents [48].

Table 3.1: An Illustration of Critical Features and Description of Each Part Linked to SDN-Based Attacks [47].

No.	Feature name	Description
1	Dt	The date and time which has been converted into a number
2	Switch	Switch number
3	Src	Source IP
4	Dst	Destination IP
5	Pktcount	Packets in a flow
6	Bytecount	Bytes in a flow

Table 3.1: An Illustration of Critical Features and Description of Each Part Linked to SDN-Based Attacks. “Table continued.”

7	dur	Duration in seconds
8	dur_nsec	Duration in nanoseconds
9	tot_dur	A sum of duration in seconds and nanoseconds
10	Flows	Number of flows
11	Packetins	Number of packets in a message
12	Pktperflow	Packet count during a single flow
13	Byteperflow	Byte count during a single flow
14	Pktrate	Number of packets sent per second
15	Pairflow	Number of flow’s pairs
16	Protocol	Type of protocol (TCP, UDP, ICMP)
17	Port_no	Port number
18	tx_bytes	Data transfer rate in bytes
19	rx_bytes	Data receiving rate in bytes
20	tx_kbps	Data transfer rate in kilobytes
21	rx_kbps	Data receiving rate in kilobytes
22	tot_kbps	The sum of tx_kbps and rx_kbps
23	Label	Label indicates the traffic type is false DDoS (0) or harm (1)

3.5 RESEARCH TOOLS AND REQUIRED INSTRUMENTS

This master's analysis will adopt some research tools and critical instruments that would help achieve the goal of this work, mirrored by investigating the beneficial role of ML and DL principles in detecting cyber threats. The major research tools that will be employed include intelligent ML and DL algorithms, which contains the following models:

- i. Logistic Regression (LR),
- ii. Support Vector Machine (SVM),
- iii. Stochastic Gradient Descent (SGD), and
- iv. Machine Learning Model (MLM).

In addition, the computer used in this work to analyze the performance and efficacy of ML algorithms and DL concepts in identifying cyber threats is the LENOVO ThinkBook Core i7, with a RAM of 16 GB and a processor of 64 bits. The windows system used in this numerical analysis is Windows 10.

At the same time, 'Colab Pro,' which stands for Google Collaboratory, is employed in this master's research. It is a product developed by Google company. Google Collaboratory utilizes Python Software Package Version 3.0. The Jupyter notebook service is implemented to run the numeric code in the Python computer program. This service is especially suited to ML principles and data analysis. Moreover, Google Collaboratory needs no setup to use. It can provide free access to conduct numerical analysis and simulations by virtue of the Python program.

The following paragraphs will illustrate some details related to the four algorithms that operate depending on ML and DL principles.

3.5.1 LR Algorithm

Logistic Regression (LR) can be defined as "A statistical analysis approach that is adopted to forecast a binary output, including 'yes' or 'no,' according to previous observation and definition of a dataset [49]–[51]. The LR algorithm can predict a dependent variable depending on the analysis of the correlation between one or multiple independent data parameters. For instance, the LR model can be employed to identify and predict whether a scholar or researcher will win a prize in a competition [52]–[54]. Figure 3.3 represents a

schematic block diagram associated with the central principles of the LR model used in binary classification.

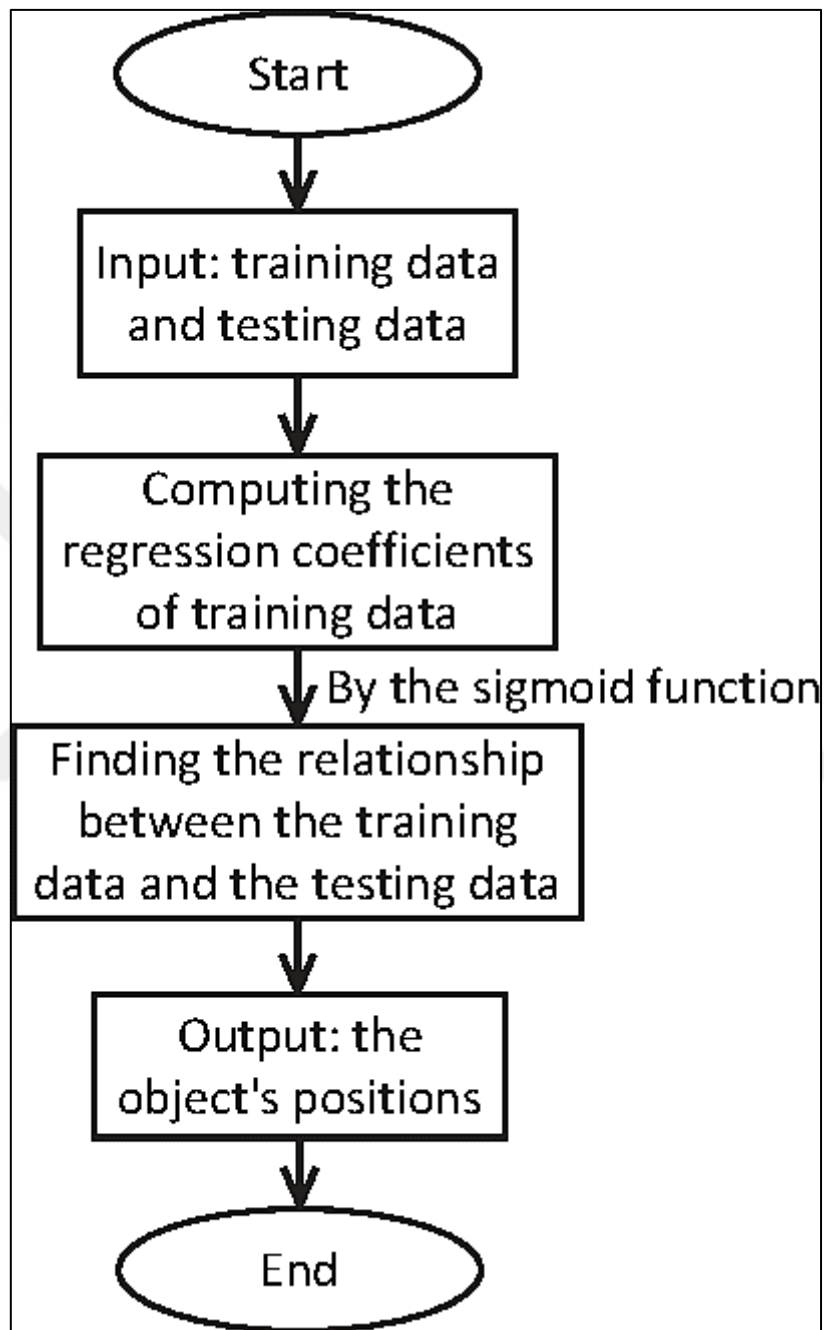


Figure 3.3: A Schematic Block Diagram Associated with The Central Principles of The LR Model Used in Binary Classification [55].

Those binary outputs enable flexible and straightforward classification of the decision between two alternatives. Furthermore, the LR algorithm takes into account more than input constraints. In the case of a PhD student's acceptance into a university, the LR scheme considers a set of factors and criteria associated with the acceptance chance, including the number of extracurricular tasks, GRE and GMAT scores, and the rate of GDP. Depending on previous data and similar results and decisions made based on the data related to historical students, the LR can provide an accurate decision on the opportunity for this PhD student to be accepted.

At the same time, Figure 3.4 indicates a comparative analysis between (I) the single LR classification process and (II) the multiple LR identification phase.

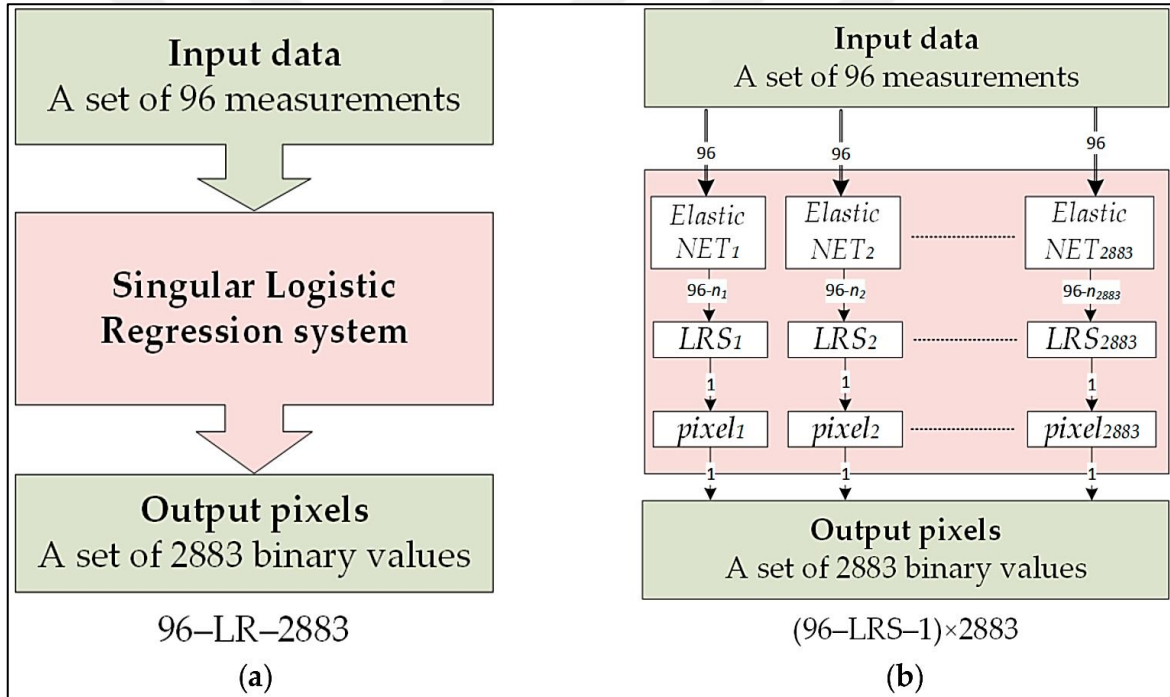


Figure 3.4: A Comparative Analysis Between (A) Single LR Classification Process and (B) Multiple LR Identification Procedure [56].

Consequently, the LR model has been broadly utilized and implemented in various ML problems and AI cases to help researchers classify some data and make an optimum decision regarding a possibility/ case depending on the historical dataset. With more historical data defined in the pre-simulation process, the effectiveness and reliability of the LR algorithm can be enhanced [57].

Various scholars studied the LR algorithm based on theoretical modeling and some mathematical formulas. For instance, [58] conducted an analysis examining the performance and characteristics of the LR algorithm. They stated that the LR model follows a linear statistical trend, which is commonly described using the following relationship:

- | | |
|---------|---------------------------------|
| π | : Pi function |
| β | : The intercept index [59] |
| i | : The initial number of terms |
| n | : The final number of terms |
| X | : The explanatory variable [59] |

$$\sigma(Z) = \frac{1}{1 + \exp(-z)} \dots \dots \dots (3.2)$$

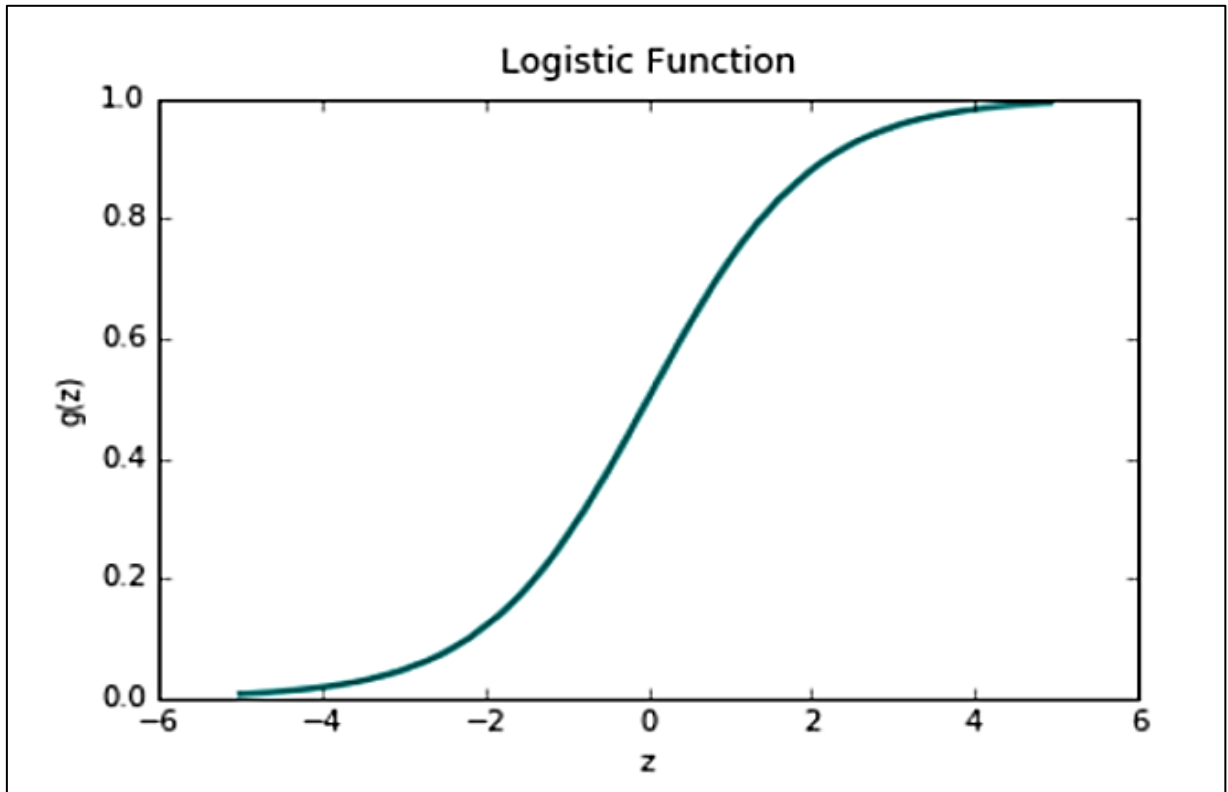


Figure 3.5: A Curve Resulting from The Sigmoid Function Plotting [58].

3.5.2 SVM Algorithm

Support Vector Machine (SVM) can be described as “One of the most familiar and common examples of supervised ML schemes that are employed for regression purposes and classification applications. Nonetheless, this type of algorithm is typically implemented for problems associated with classification goals that are related to ML principles [60]–[62]. SVM can create the optimum decision or the most appropriate line, which can include and segregate n-dimensional space into various classes. Therefore. The new dataset and elements can be flexibly defined and identified for future simulation problems. The selection and development of the boundary that has the best decision can be recognized with the hyperplane [63]–[66]. Further, the SVM algorithm depends on ML concepts to select the extreme vectors or maximum points that support the creation of hyperplanes. Those extreme vectors and maximum points are known with support vectors. For this reason, this algorithm is denoted as the SVM. Figure 3.6 indicates a configuration pertaining to a depositional microfacies classification procedure by virtue of the SVM algorithm.

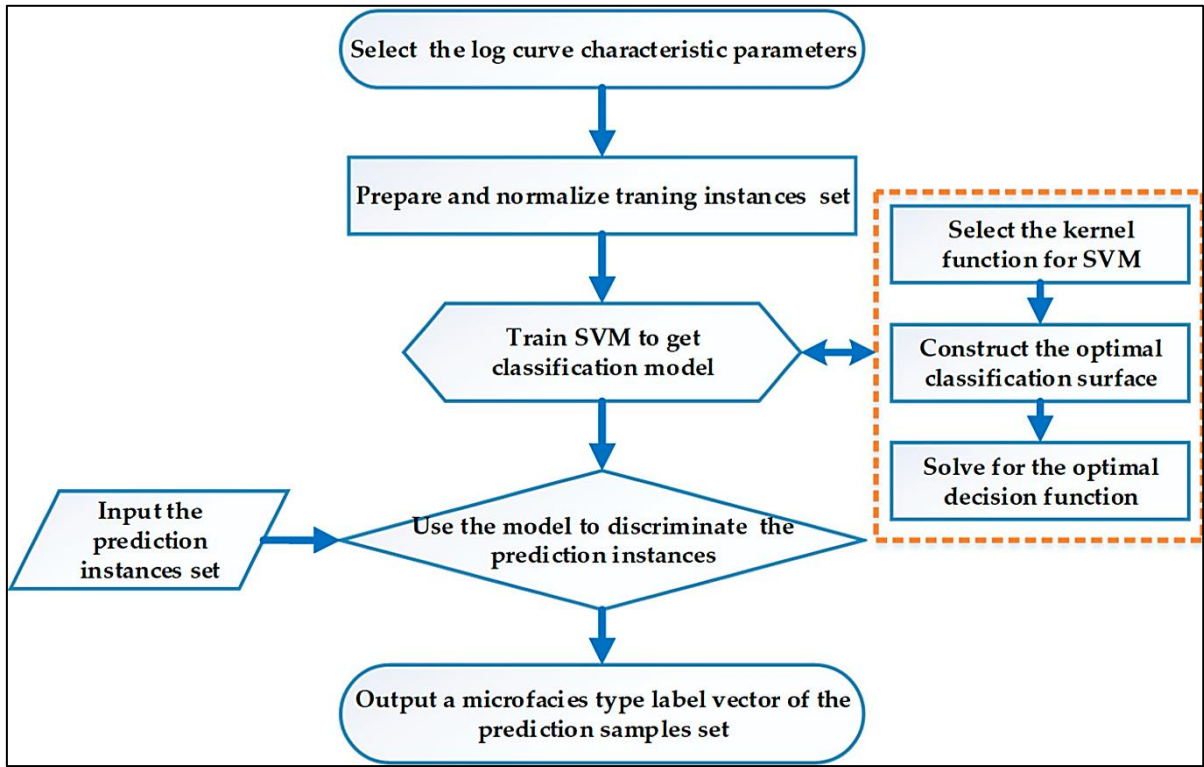


Figure 3.6: A Configuration of A Depositional Microfacies Classification Process Via The Svm Algorithm [66].

In addition, Figure 3.7 illustrates a schematic architecture associated with the optimum classification, support vectors, hyperplane, and optimal decision surfaces using the SVM algorithm.

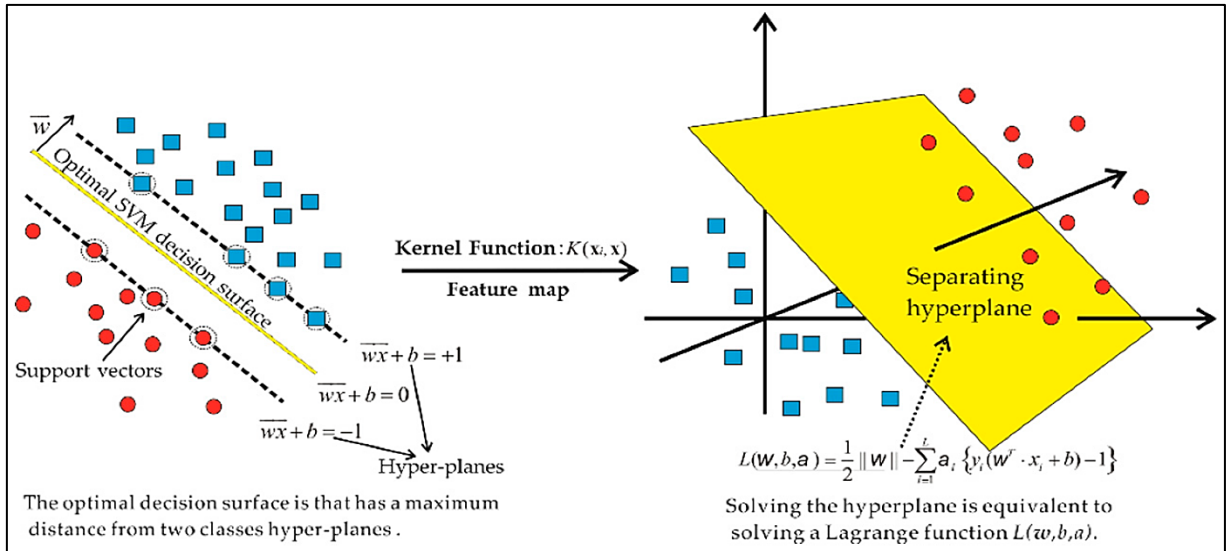


Figure 3.7: A Schematic Architecture Related to The Optimum Classification, Support Vectors, Hyperplane, and Optimal SVM Decision Planes [66].

[66] mentioned that the SVM scheme depends on a critical mathematical formula to solve problems associated with the inner product operation within a high-dimensional feature space. This formula is known as the Kernel Function, KF . KF can be expressed via the following relationship:

$$KF = K(x_i, x) \dots \dots \dots (3.3)$$

Where:

$$K(x_i, x) = \exp(-r \times \|x_i - x\|^2) \dots \dots \dots (3.4)$$

From those two equations:

- x_i : The feature vectors linked to the training samples
- x : The prediction samples
- r : The optimal Kernel function parameter

3.5.3 SGD Classifier

Stochastic Gradient Descent (SGD) can be depicted as “An optimization algorithm often employed in several ML applications to find the model parameters corresponding to the best fitting between forecasted and actual outcomes. This algorithm provides inexact results. However, it is greatly powerful and robust [67], [68]. Besides, this ML model is broadly employed in various ML applications as it can combine its scheme and working principles with backpropagation. Hence, it is flexibly implemented in variant neural network training applications [69]–[71]. Figure 3.8 indicates an architecture of the adoption of the SGD algorithm in making binary classification, including the concentric circle dataset and the network structure.

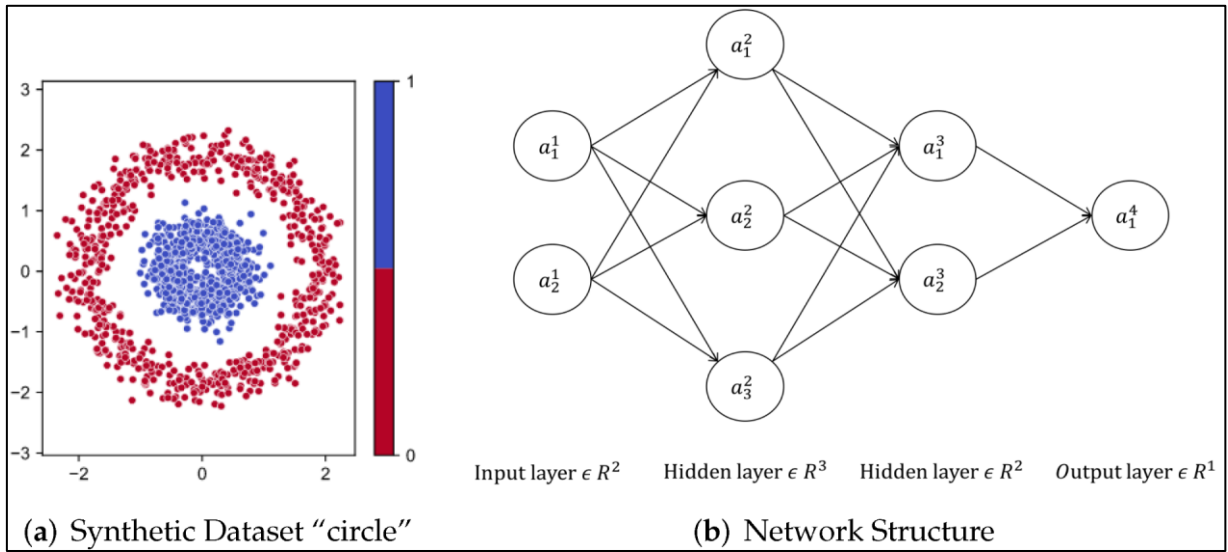


Figure 3.8: The Adoption of The SGD Algorithm In Making A Binary Classification Process, Including (A) The Concentric Circle Dataset and (B) The Network Structure [72].

3.5.4 MLM of Optimization

The Machine Learning Model (MLM) can be illustrated as "A file that has been trained to recognize specific categories and characteristics related to patterns. The researcher can depend on this algorithm to train a model over a group of datasets. This aspect can offer an active approach through which the MLM can reason over and learn from those data [73]–[76]. Figure 3.9 explains a schematic of implementing MLM to detect cyber threats in networks

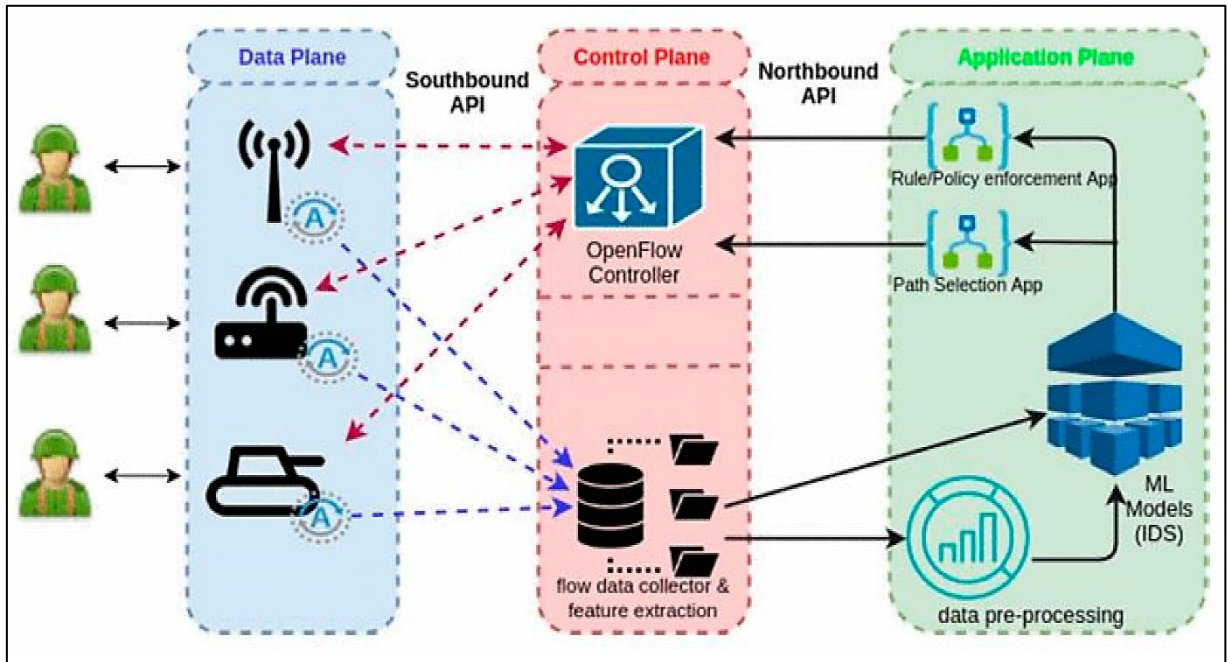


Figure 3.9: The Implementation of MLM to Detect Cyber Threats in Networks [77].

3.6 PERFORMANCE EXAMINATION OF ML ALGORITHMS

In this study, a set of assessment techniques and evaluation metrics were implemented to validate the performance and efficiency of the four ML algorithms employed to make accurate cyber threat detection. To conduct these critical examination measures, four aspects of evaluation criteria were considered, which include the following metrics:

- a. Accuracy,
- b. F1-Score,
- c. Precision, and
- d. Recall.

The paragraphs below will illustrate further information related to each evaluation method used to verify the performance of the four ML algorithms addressed in this work.

3.6.1 Accuracy Assessment

The accuracy can be defined as “The ratio of correctly forecasted data to all samples, or the percentage of correctly forecasted samples, including all observations, is the most intuitive performance statistic.” In addition, accuracy can be stated as “The proportion of correctly predicted samples with all models.”

Accuracy can be illustrated as “The most intuitive performance statistic.” In other words, it can be described as “The ratio of correctly projected data to all samples, or the percentage related to correctly forecasted samples, encompassing all observations.” “The proportion of samples that are accurately predicted by all models” can be considered another definition of accuracy. This index can be assessed and computed with the help of the following formula:

$$Accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)} \times 100\% \dots \dots \dots (3.5)$$

Where:

- TP : True Positives
- FP : False Positives
- TN : True Negatives
- FN : False Negatives

3.6.2 F1-Score

Another index employed to examine the effectiveness and dependability of the four DL schemes and ML algorithms considered in this work is the F1-Score. The F1-Score can be defined as “The weighted mean value related to the recall and precision portions. The evaluation of this parameter can be conducted via the following relationship:

$$'F1 - Score' = \frac{2 \times (Precision \times Recall)}{(Precision + Recall)} \times 100\% \dots \dots \dots (3.6)$$

Where the recall and precision are represented and illustrated in the following sections.

3.6.3 Precision

A third indicator is implemented to evaluate and describe the efficacy pertaining to the DL and ML algorithms addressed in this master’s research. This metric is recognized with ‘Precision.’ This index can be depicted as “The ratio related to accurately and precisely forecasted positive outcomes in comparison with the overall number of expected positive samples.” The precision quantity (as a percentage ratio) can be calculated using the principles of the following formula:

$$Precision = \frac{(TP)}{(TP + FP)} \times 100\% \dots \dots \dots (3.7)$$

3.6.4 Recall

A fourth technique adopted to assess and explore the performance of ML algorithms in terms of active cyber threat detection is the recall. The recall can be defined as “The proportion of correctly anticipated positive samples concerning the overall number of samples that were predicted to be positive. The relationship below can be exploited to examine and calculate this performance indicator:

$$Recall = \frac{(TP)}{(TP + FN)} \times 100\% \dots \dots \dots (3.8)$$

3.8 CHAPTER SUMMARY

This chapter was written and designed to describe the fundamental approaches and critical research phases used in this thesis to validate and identify some smart ML algorithms and DL fundamentals in detecting complicated cyber attacks and challenging internet and network threats which can harm the data privacy and information security related to aeroplane system and interconnected networks and web apps corresponding to airplane authorities and employees working on airports, whose data privacy and the information security of passengers may be adversely affected due to dangerous cyber attacks. Additionally, this chapter included examples and some illustrations regarding the independent and dependent variables addressed in evaluating the efficacy of ML methods to detect cyber threats in airplane networks. Furthermore, the dataset definition and description were conducted to help clarify the database considered through the numerical study to prove the ML algorithms’ effectiveness in identifying cyber attacks. The sorts of ML algorithms and full descriptions were also proposed in the third chapter to further show the critical traits and advantageous attributes of such DL methods that can be used to provide active remedies to cyber-attacks. The following chapter represents the main research findings and outcomes based on numerical analysis and Python code processing.

4. RESULTS AND DISCUSSIONS

4.1 CHAPTER GOAL

This chapter describes the major numerical research outputs and mathematical findings obtained based on the processing and analysis of the numeric Python code and simulations process of the four DL and ML algorithms considered and investigated in this master's thesis. Some graphical illustrations and tables are used to express those numerical outcomes associated with the effectiveness of cyber threat detection.

4.2 STATISTICAL FINDINGS OF DDOS ATTACKS

Based on the numerical findings of the Python simulation process, it was found that the case study assault comprised forty percent of the overall database, according to the dataset investigation and the target variable analysis. Furthermore, the outputs revealed that for the target class variable, the distribution was significantly balanced, ranging between 40% and 60%. As a result, the number of bogus DDoS assaults was almost equal to the number of harmful attacks. This fact could result in an issue with achieving perfect balance when fitting our models. Figure 4.1 represents the numerical outputs pertaining to the binary classification of zeros and ones associated with the DDoS attacks in the dataset.

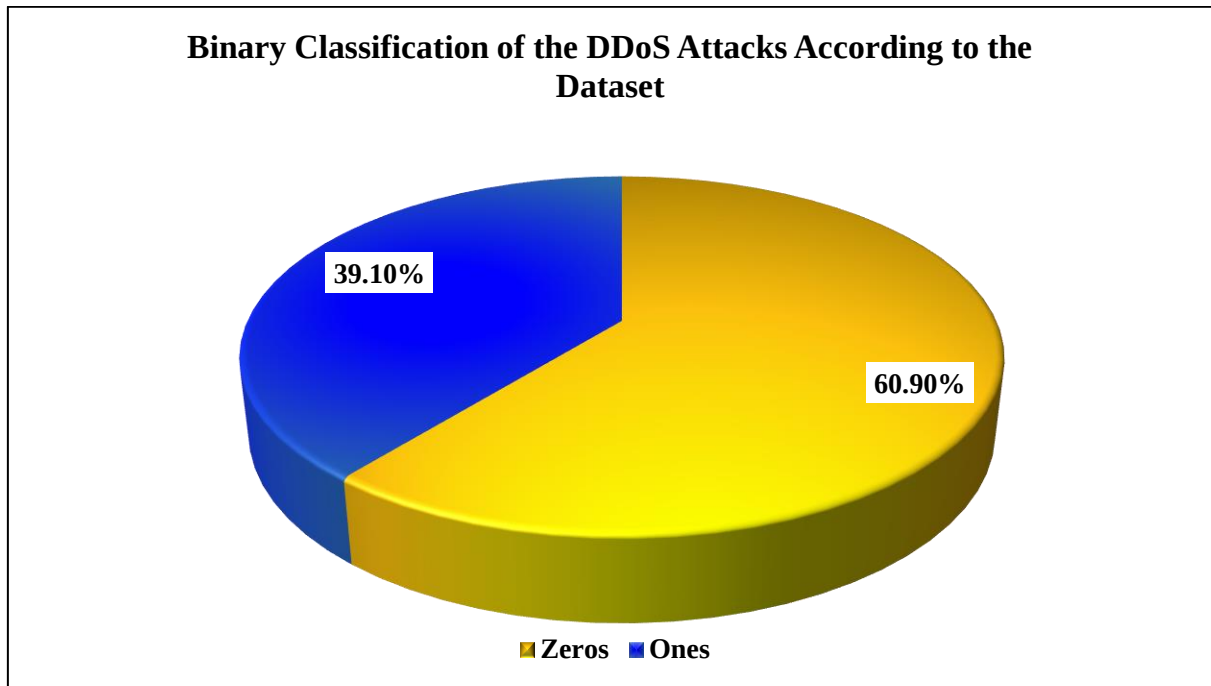


Figure 4.1: The Numerical Results of The Binary Classification of Zeros and Ones Related To The Ddos Attacks in The Dataset.

It can be inferred from Figure 4.1 that the percentage of zeros detected using ML and DL algorithms by virtue of binary classification principles exceeded the portion of ones, corresponding to approximately 60.90% compared with 39.10%, respectively. Furthermore, Figure 4.2 indicates the same numerical outcomes of this research but represents numeric values in terms of frequency instead of ratios, as in Figure 4.1.

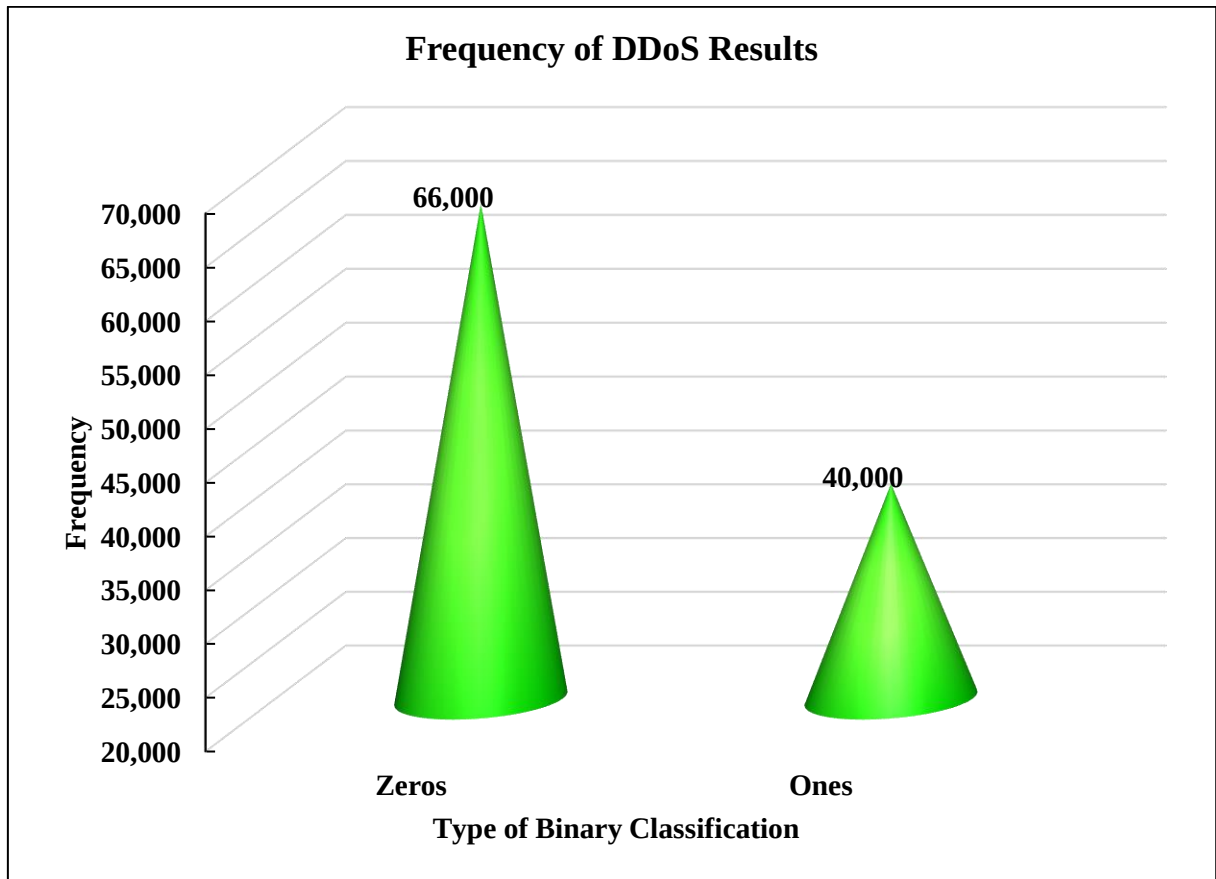


Figure 4.2: The Frequency Results (In Numbers) Linked to The Binary Classification of Zeros And Ones of The Ddos Attacks.

It can be indicated from the numeric data represented in Figure 4.2 that roughly the number of 'Ones' equal two-thirds of that related to the 'Zeros,' corresponding to frequencies of 66,000 and 40,000 for Zeros and Ones, respectively.

4.3 THE MISSING VARIABLES IN THE RESEARCH

Based on the numerical analysis and Python code processing, it was discovered that the NaN (or missing values of the cyber threat detection) associated with this case study were identified in the 'rx' kbps and tot kbps features. In addition, the number of those missing values (NaN) was identified, as illustrated in Figure 4.3.

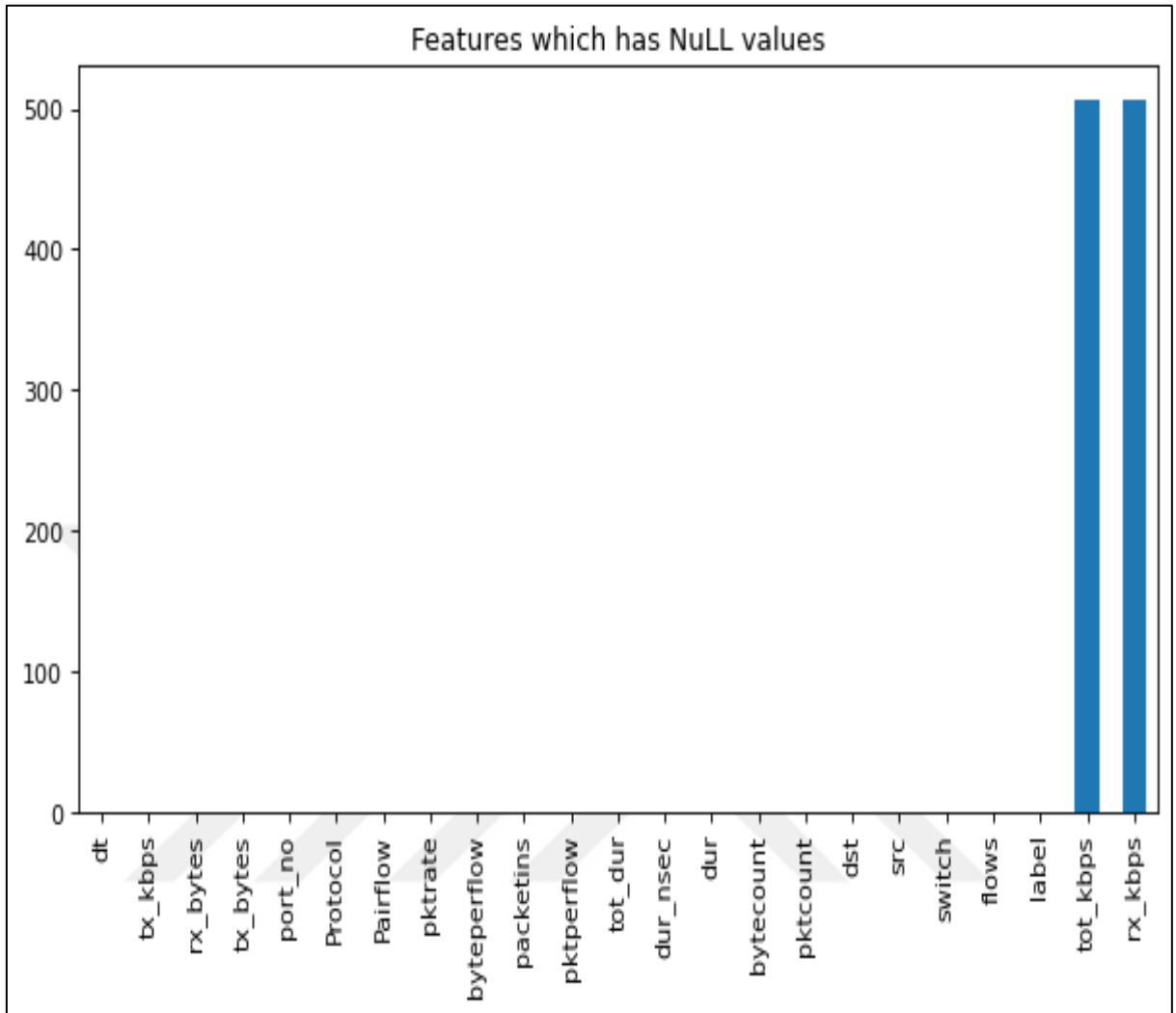


Figure 4.3: The Number of Features With Null (Zero) and Nonzero Values.

It can be concluded from the statistical information represented in Figure 4.3 that there are an overall of roughly 506 missing values for the ‘tot_kbps’ and ‘rx_kbps’ features. For this reason, the dataset considered in this study was modified and cleaned up to remove the missing data obtained in this numerical simulation process, helping enhance the performance and practicality of the mathematical investigation.

4.4 THE NUMERICAL OUTPUTS LINKED TO THE NETWORK REQUESTS FROM THE IP ADDRESS

After cleaning the dataset, it was determined that the dimension (103839, 23) attained specific features, including the following aspects:

- i. float64(3),

- ii. int64(17), and
- iii. object (3).

Object features ‘src,’ ‘dst,’ and ‘Protocol’ corresponded to the sender IP, destination IP, and internet protocol. Figure 4.4 expresses the number of requests associated with various IP addresses based on the ‘IP Address of Sender.’

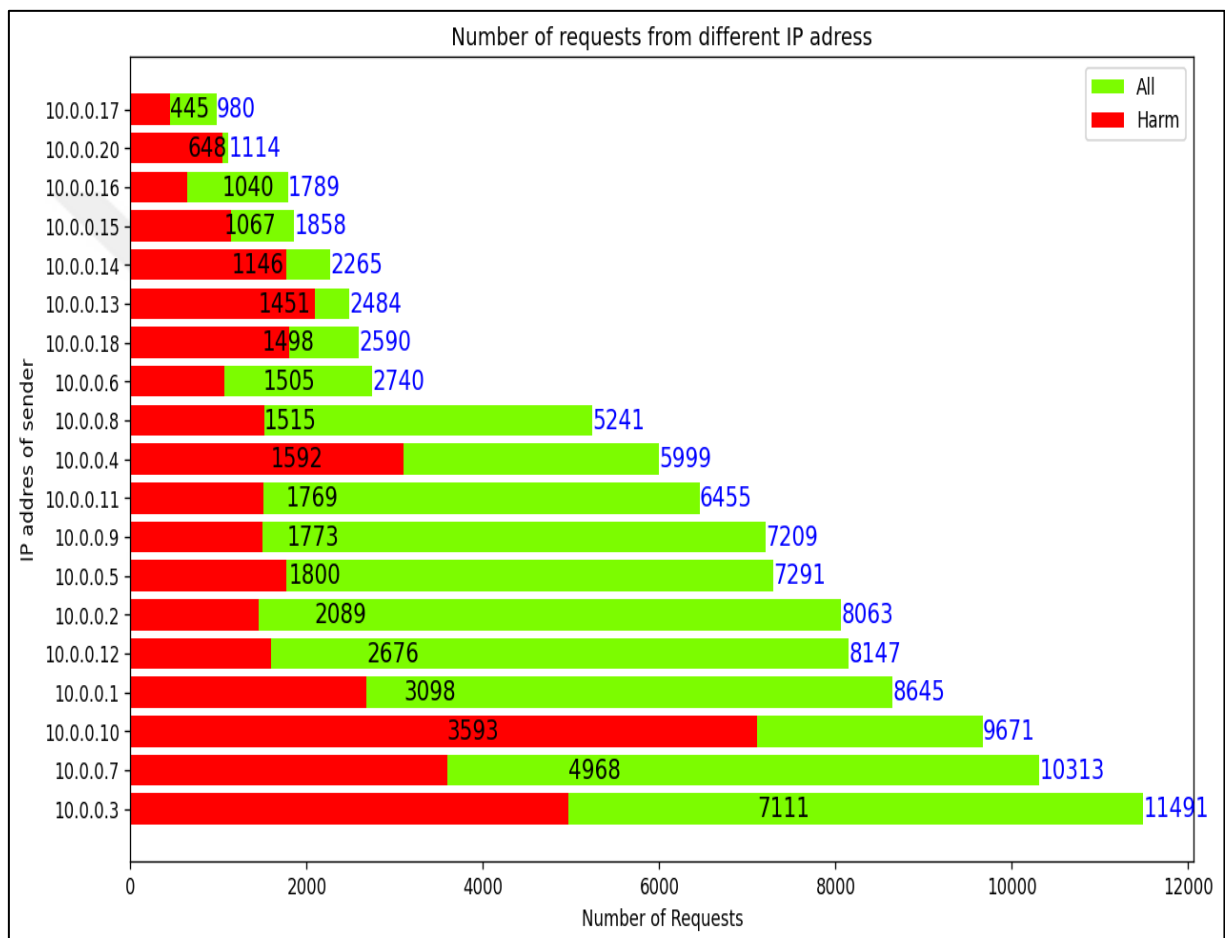


Figure 4.4: The Number of Requests Associated With Various IP Addresses Based on The ‘IP Address of Sender.’

It can be shown from Figure 4.4 that in the DDoS attack tactics and random senders and receivers, the IP address has little influence on detecting damage assaults. Similarly, Figure 4.5 explains the number of requests pertaining to various IP addresses based on the ‘IP Address of Destination.’

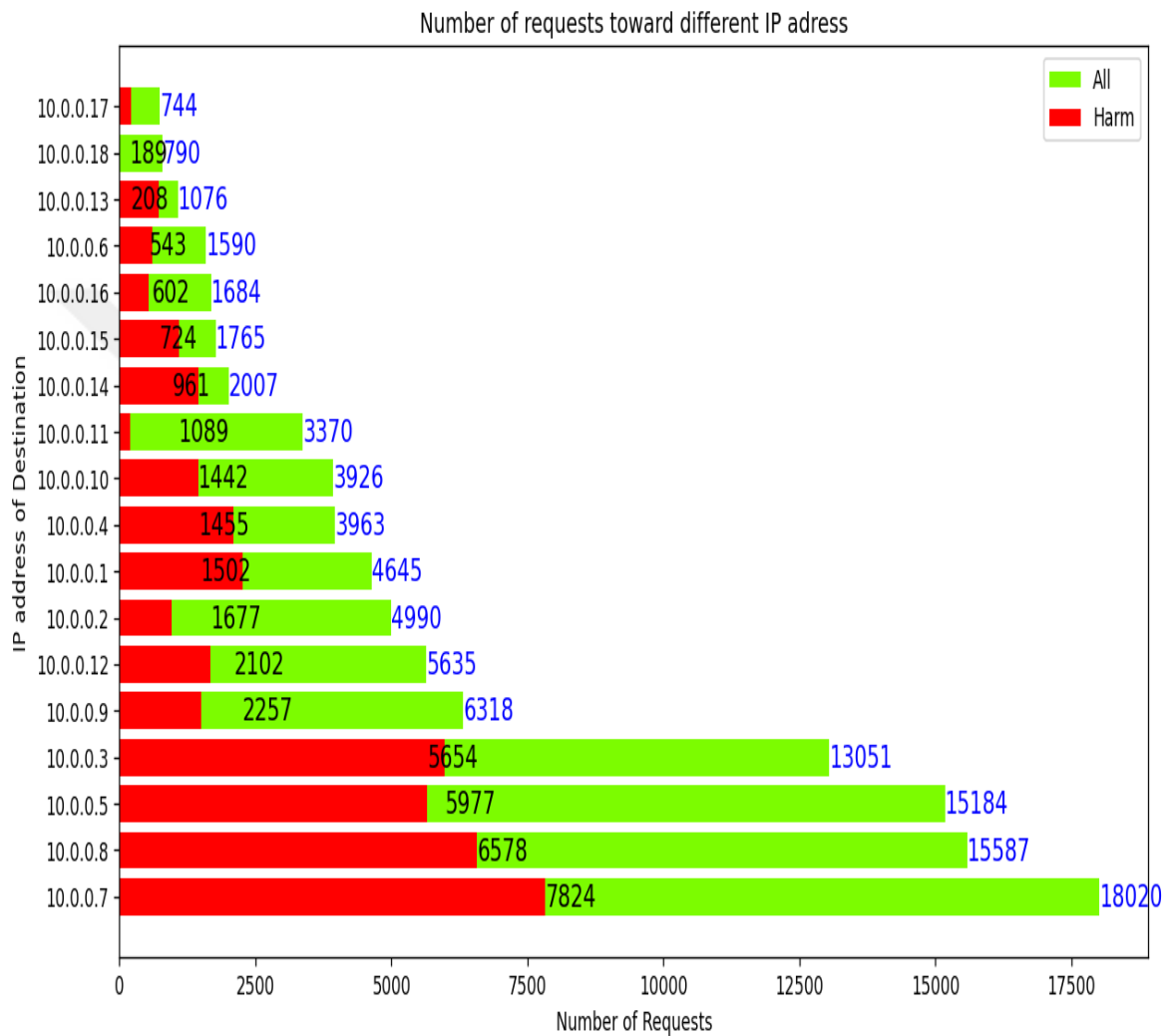


Figure 4.5: The number of requests associated with various IP addresses based on the ‘IP Address of Sender.’

According to the research findings related to the ‘IP Address of Destination,’ represented in Figure 4.5, it can be concluded that the IP address had a negligible impact on detecting cyber

threats and internet attacks based on the DDoS attack mechanisms and random senders and receivers.

Nonetheless, it was found that unlike the Transmission Control Protocol (TCP), User Datagram Protocol (UDP), and Internet Control Message Protocol (ICMP), the ICMP helped send an Echo Request message (and gets an Echo Reply) to detect whether the target computer is accessible and how long the packet takes to be responded to net protocol. Networked operating systems can employ ICMP to transmit error signals, such as “target computer inaccessible.” Because UDP and TCP behave similarly, the Protocol variable was



eliminated. Figure 4.6 illustrates the number of requests according to different protocols implemented and investigated in this work.

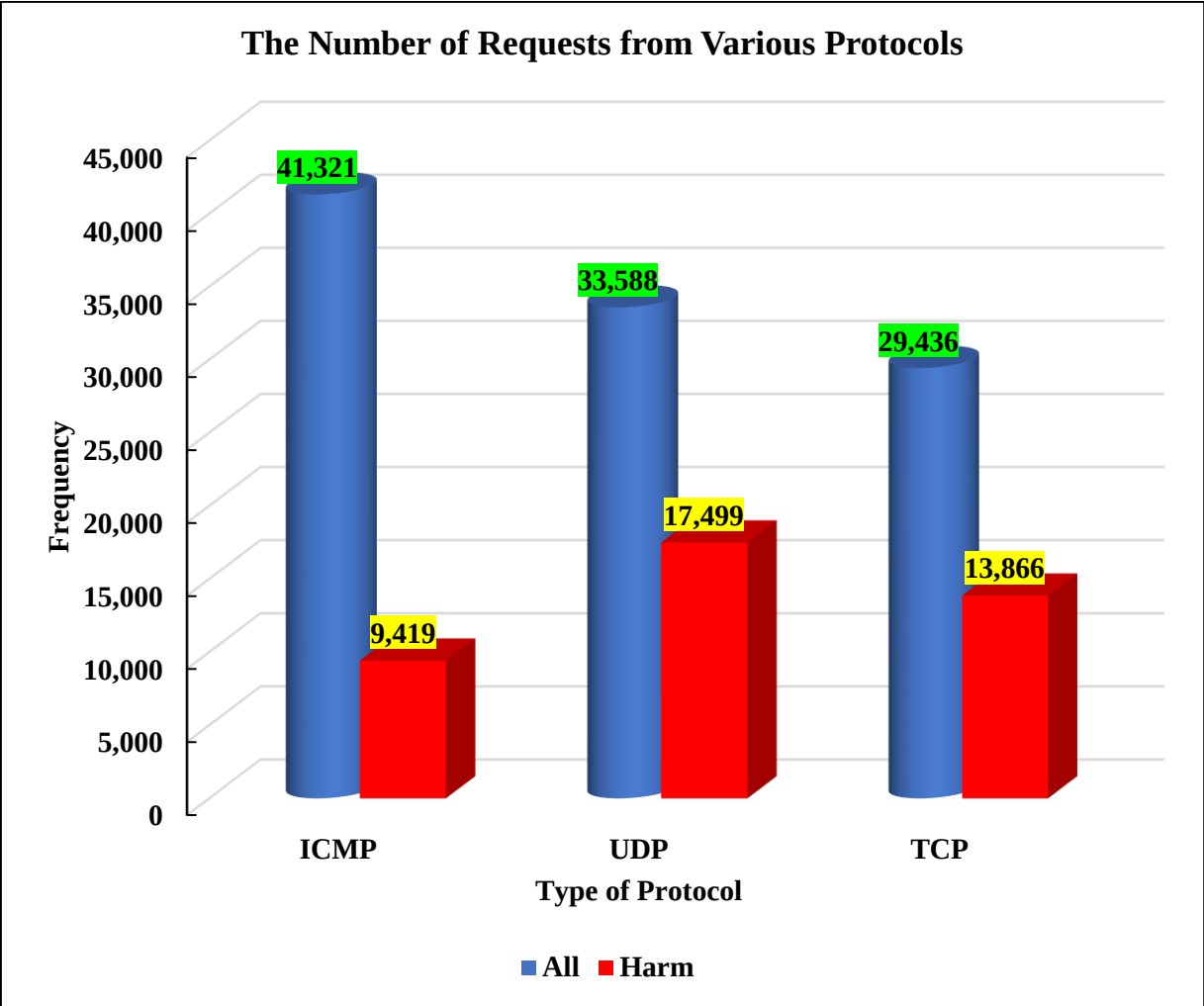


Figure 4.6: The Number of Requests According To Different Protocols Employed in This Work.

It can be indicated from the numeric statistics represented in Figure 4.6 that the ICMP had the maximum number of ‘All Requests,’ recording a frequency of 41,320 compared with

UDP and TCP methods. On the other hand, the UDP approach attained the ultimate value of ‘Harm Requests,’ corresponding to a frequency of 17,499.

Besides, the mathematical simulation process confirmed that the correlations associated with the ‘Time’ and ‘Date’ variables, which are denoted with ‘dt,’ had a remarkably low impact on the numerical analysis, as shown in Figure 4.7.

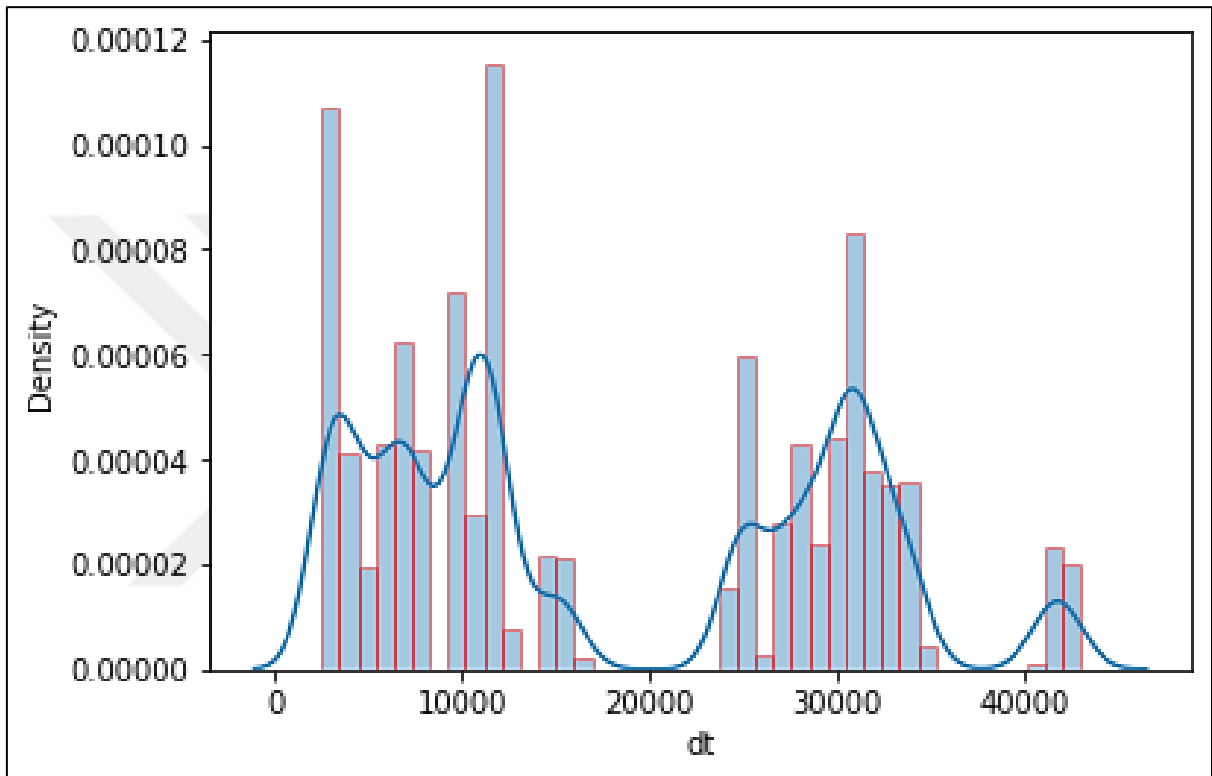


Figure 4.7: The Density of Date and Time Implementation in The Numerical Analysis.

It can be indicated from the data graphed in Figure 4.7 that the density of the ‘dt’ correlation ranges between approximately 0.00000 and 0.00011, explaining that the change in time and variation in data have negligible effects on the research findings. Figure 4.8 expresses the number of requests related to the number of switches for the ‘All’ and ‘Harm’ data.

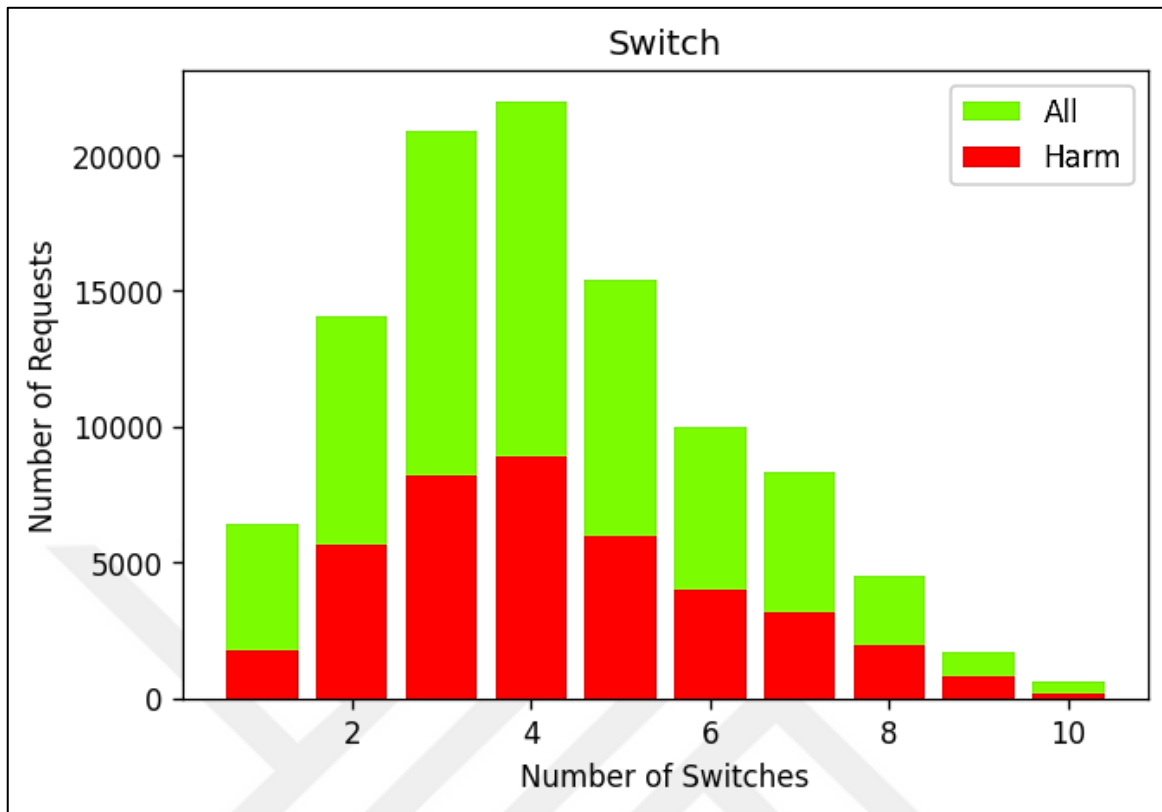


Figure 4.8: The Variation Between The Number of Requests and The Number of Switches.

It can be inferred from Figure 4.8 that the number of switches of four had the maximum number of requests in terms of ‘All’ and ‘Harm.’ In contrast, ten switches provided the least number of requests.

A similar graphical representation of the relationship between the number of ports and the number of requests is described in Figure 4.9.

Figure 4.9 expresses the variation between the number of requests and the number of ports.

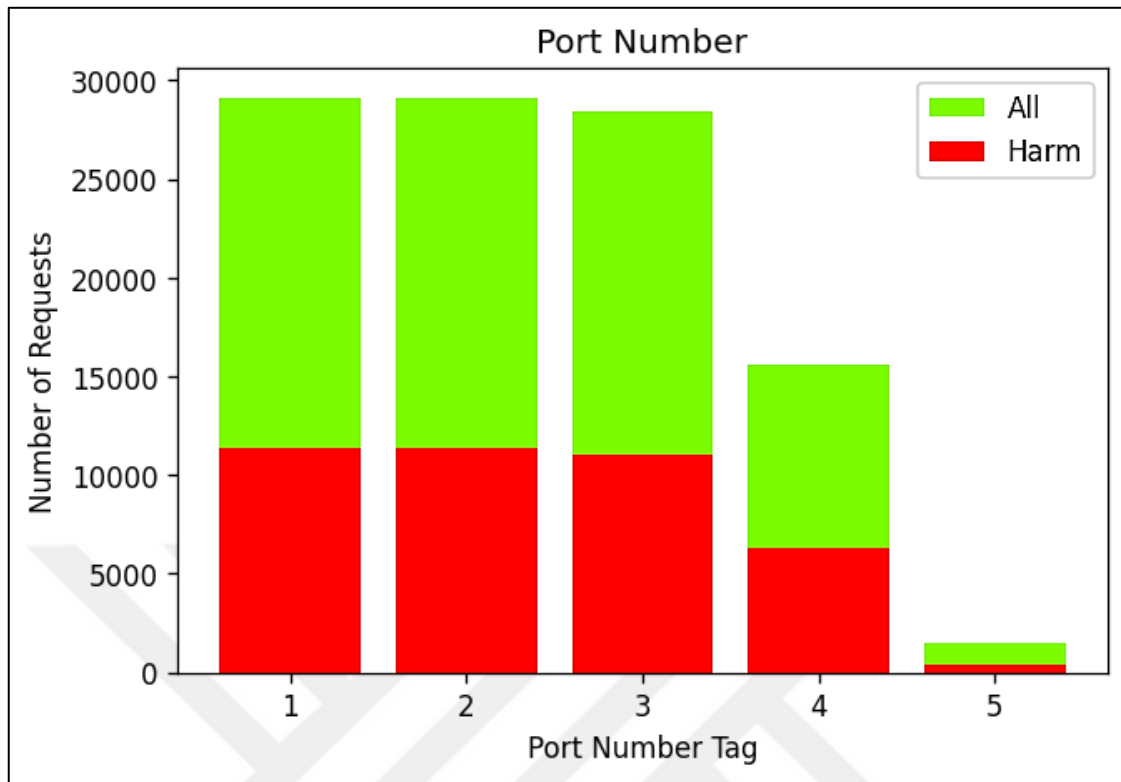


Figure 4.9: The Variation Between The Number of Requests and The Number of Ports.

It can be deduced from the numeric data represented in Figure 4.9 that approximately one, two, and three ports offered the maximum number of requests in terms of ‘All’ and ‘Harm.’ However, five ports provided the least number of requests compared with all port numbers.

4.5 NUMERICAL RESULTS OF VARIABLES CORRELATIONS

The correlations between the research variables mentioned in the dataset were achieved depending on the findings attained from the numerical analysis and simulation process. It was found that the correlation between the dataset parameters is characterized by its linear-based properties. The correlation coefficient value ranges between -1 and 1 . Figure 4.10 indicates the correlations between the dataset variables investigated in this master’s thesis.

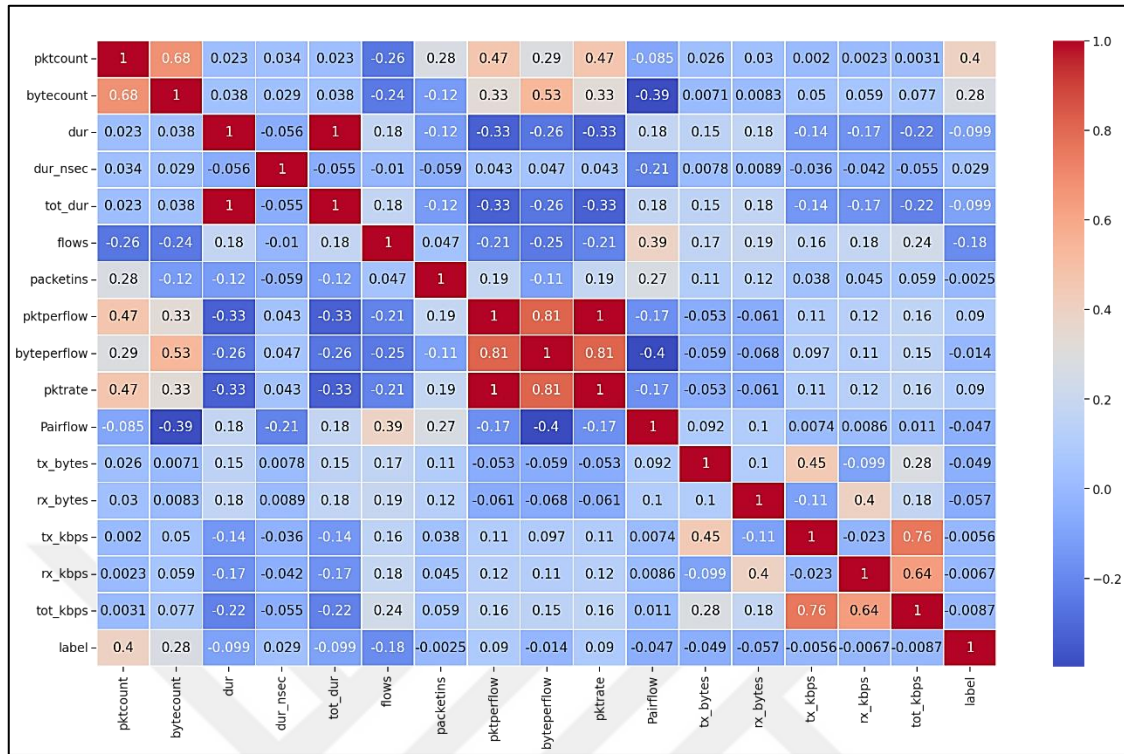


Figure 4.10: The Correlations Between The Dataset Variables Investigated in This Study.

It can be inferred from the results obtained from the numerical simulation procedures that the correlation between the dataset's variables can be expressed as a matrix. This enables researchers to identify and determine which pairings have a correlation flexibly. At the same time, Figure 4.10 can represent the degree to which two variables rise or fall simultaneously based on a clear illustration.

In addition, it can be concluded that the variable 'pktcount' and the target variable 'label' are the most correlated, with a value of (0.4). Meanwhile, the variables 'packetins,' 'tx_kbps,' 'rx_kbps,' and 'tot_kbps' are not correlated (0.008) like other parameters due to the fact that they are dropped.

4.6 ML ALGORITHMS PERFORMANCE EVALUATIONS

Before presenting the findings associated with the ML and DL models assessment and examination via performance metrics, it is worth mentioning that the dataset was divided into a training set (representing about 75% of the total data). On the other hand, the rest of the data was employed for the testing set.

4.6.1 LR Algorithm Performance Evaluation

Experiments were conducted to determine the optimum logistic regression solver performance compared with other algorithms. A Library for Large Linear Classification (LLC), Stochastic Average Gradient (SAG), and Stochastic Average Gradient Adaptive Algorithm (SAGAA). It was determined that the optimal solution was LLC. The results of the training process related to the LR are represented in Table 4.1.

Table 4.1: The Numerical Outcomes of The Training Process Related to The LR Algorithm.

Metric Category	Classification Outputs			
	'0'	'1'	Accuracy	Weighted
Precision	78.07%	69.14%	74.89%	74.61%
Recall	82.05%	63.57%	74.89%	74.89%
F1-Score	80.01%	66.24%	74.89%	74.68%
Support	47,703	30,176	74.89%	77,879
- Accuracy Score = 74.89%				
Confusion Matrix: [[39,142 8,561] [10,994 19,182]]				

Figure 4.11 indicates a graphical representation of the numerical simulation findings of the confusion matrix associated with the LR algorithm for the training process.

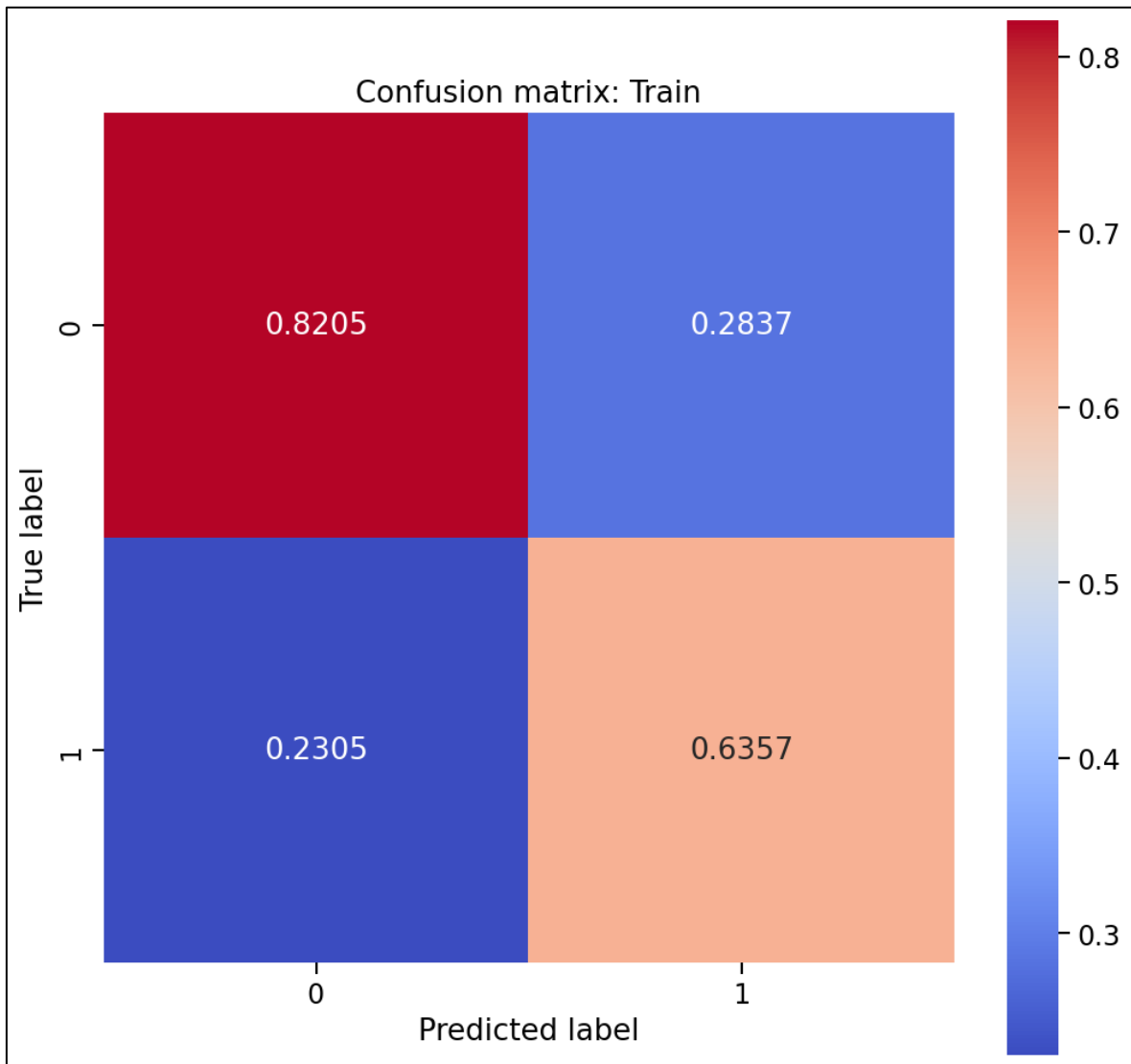


Figure 4.11: The Numerical Simulation Findings of The Confusion Matrix Linked to The LR Algorithm – Training.

It can be inferred that the ‘Zero’-‘Zero’ correlation had the most significant value, corresponding to a value of 0.8205. In contrast, the Predicted Label value of ‘0’ with the True Label value of ‘1’ had the most negligible correlation with a quantity of 0.2305.

In addition, Table 4.2 represents the numerical research findings related to the testing process of the LR algorithm.

Table 4.2: The Numerical Outcomes of The Testing Process Related to The LR Algorithm.

Metric Category	Classification Outputs			
	'0'	'1'	Accuracy	Weighted
Precision	77.23%	70.15%	74.69%	74.41%
Recall	82.20%	63.31%	74.69%	74.69%
F1-Score	79.64%	66.56%	74.69%	74.43%
Support	15,632	10,328	74.69%	25,960
- Accuracy Score = 74.69%				
Confusion Matrix: [[12,850 2,782] [3,789 6,539]]				

Figure 4.12 indicates a graphical representation of the numerical simulation findings of the confusion matrix associated with the LR algorithm for the testing process.

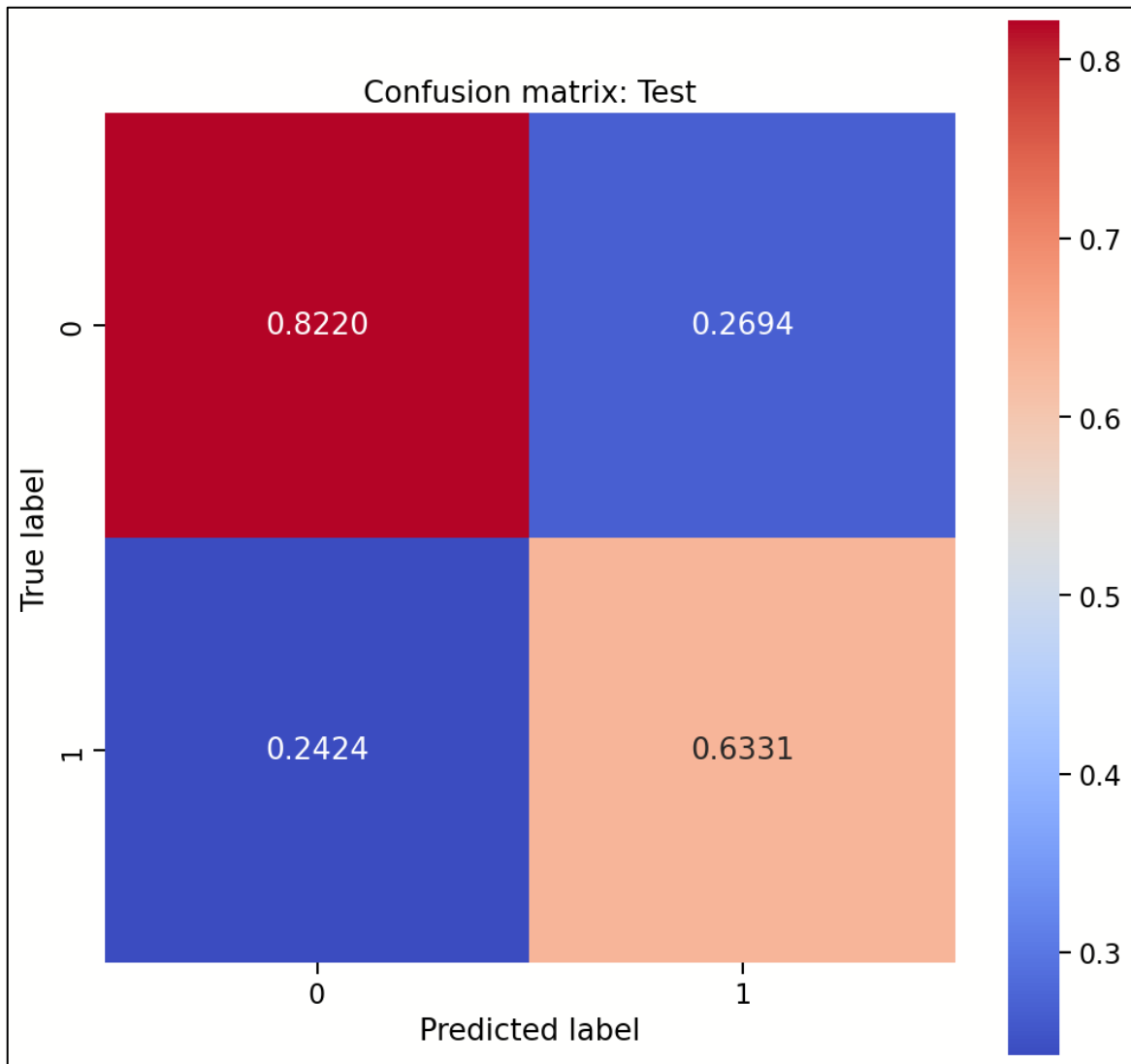


Figure 4.12: The Numerical Simulation Findings of the Confusion Matrix Linked to The LR Algorithm – Testing.

It can be inferred from Figure 4.12 that the ‘Zero’-‘Zero’ in Predicted and True Labels had the most significant value, corresponding to a value of 0.8220. In contrast, the Predicted Label value of ‘0’ with the True Label value of ‘1’ had the most negligible correlation with a quantity of 0.2424.

4.6.2 SVM Algorithm Effectiveness Assessment

To determine the reliability and performance associated with the optimal kernel principles for the SVM algorithm, the four assessment metrics were also used in this model to examine the effectiveness of this ML algorithm in performing its task of detecting cyber threats.

The results of the training process related to the SVM model are shown in Table 4.3.

Table 4.3: The Numerical Outcomes of The Training Process Related to The SVM Algorithm.

Metric Category	Classification Outputs			
	'0'	'1'	Accuracy	Weighted
Precision	95.42%	89.80%	93.17%	93.24%
Recall	93.32%	92.92%	93.17%	93.17%
F1-Score	94.36%	91.33%	93.17%	93,19%
Support	47,703	30,176	93.17%	77,879
- Accuracy Score = 93.17%				
Confusion Matrix: {[44,518 3,185] [2,137 28,039]}				

Furthermore, based on the data represented in Table 4.3, Figure 4.13 expresses a graphical illustration related to the numerical simulation outcomes of the confusion matrix associated with the SVM algorithm for the training process.

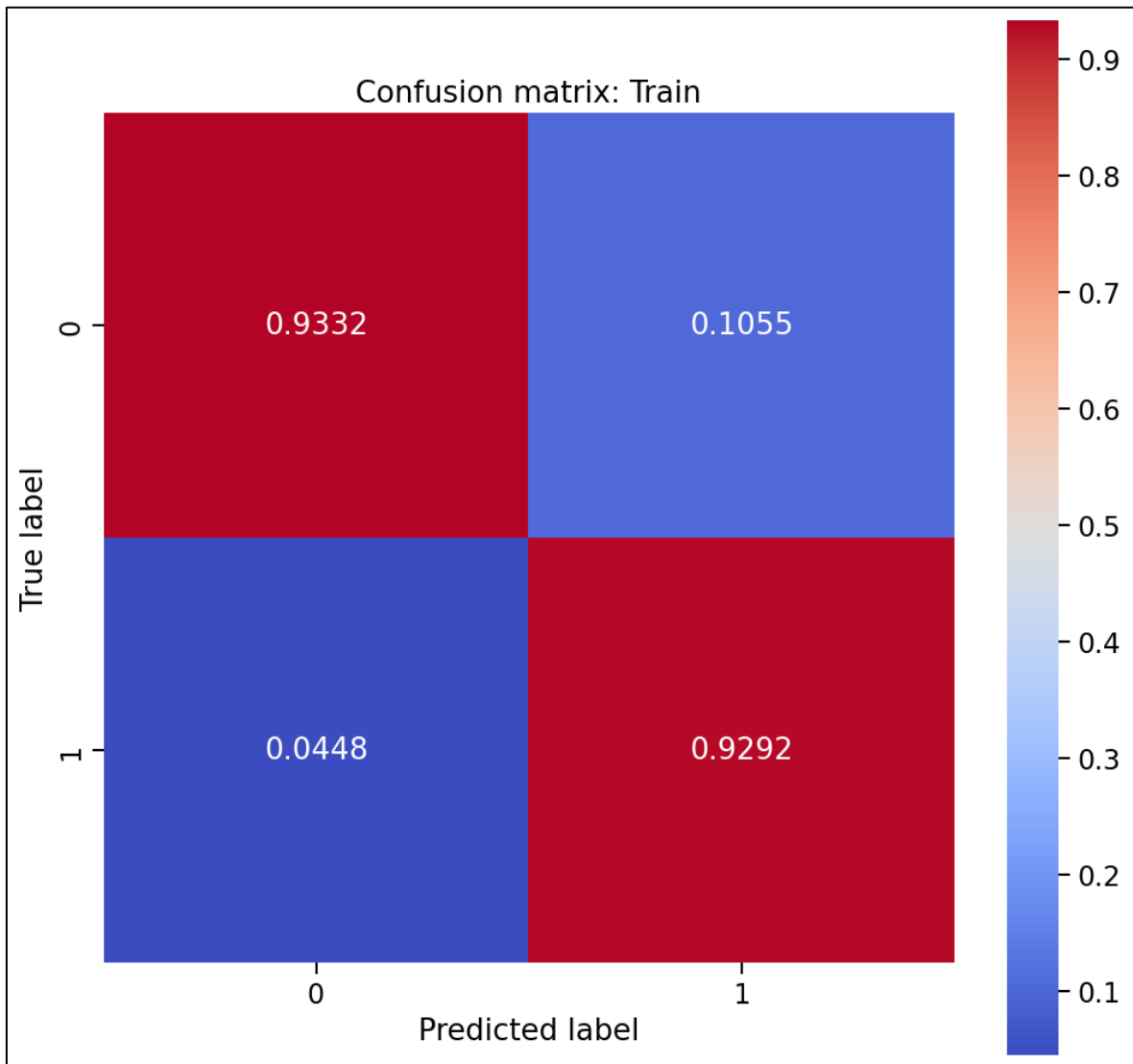


Figure 4.13: The Numerical Simulation Findings of The Confusion Matrix Linked to The Svm Model – Training Process.

It can be inferred from Figure 4.13 that the ‘Zero’ Predicted Label with the ‘One’ True Label had the smallest value of correlation, contributing to 0.0448. At the same time, the Predicted Label value of ‘0’ with the True Label value of ‘0’ had the most significant amount of correlation with a rate of 0.9332.

In addition, Table 4.4 represents the numerical research findings related to the testing process of the SVM model.

Table 4.4: The Numerical Outcomes of The Testing Process Related to The SVM Algorithm.

Metric Category	Classification Outputs			
	'0'	'1'	Accuracy	Weighted
Precision	95.01%	90.00%	92.96%	93.02%
Recall	93.21%	92.59%	92.96%	92.96%
F1-Score	94.10%	91.28%	92.96%	92.98%
Support	15,632	10,328	92.96%	25,960
- Accuracy Score = 92.96%				
Confusion Matrix: [[14,570 1,062] [765 9,563]]				

According to the results expressed in Table 4.4, Figure 4.14 expresses a graphical illustration related to the numerical simulation outcomes of the confusion matrix associated with the SVM algorithm for the training process.

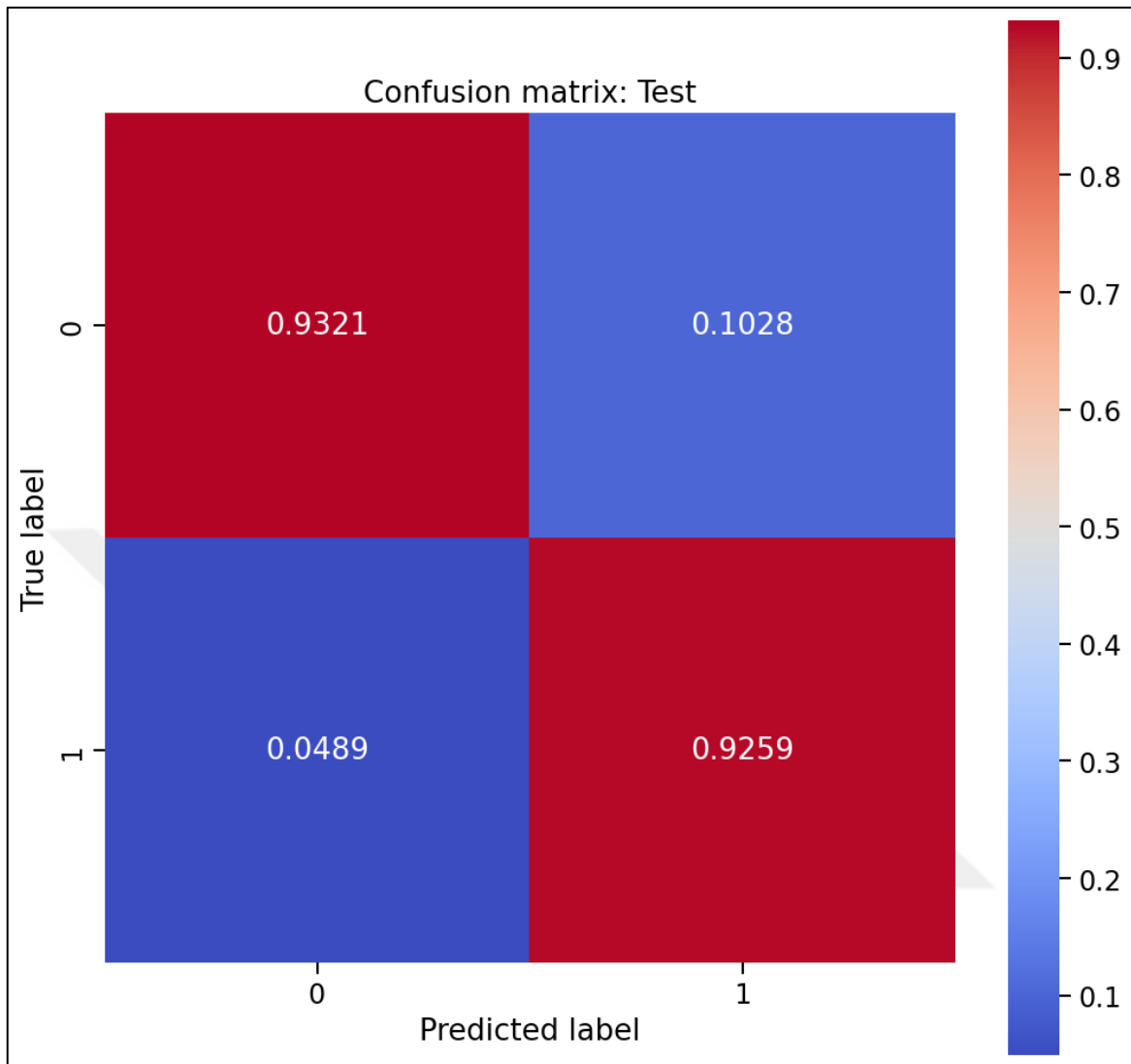


Figure 4.14: The Numerical Simulation Findings of The Confusion Matrix Linked to The SVM Model – Testing Process.

It can be inferred from Figure 4.14 that the 'Zero' Predicted Label with the 'One' True Label had the most negligible amount of correlation, contributing to approximately 0.0489. On the other hand, the Predicted Label value of '0' with the True Label value of '0' provided the most remarkable quantity of correlation with a measure of 0.9321.

4.6.3 EVALUATION OF SGD CLASSIFIER'S PERFORMANCE

To determine the reliability and performance pertaining to the SGD classifier, the four assessment metrics were implemented again as they were applied in previous and later assessment processes to explore and identify the effectiveness of this ML scheme in performing its role of detecting cyber attacks.

The numerical outcomes linked to the training process related to the SGD classifier are shown in Table 4.5.

Table 4.5: The Numerical Outcomes of The Training Process Related to The SGD Classifier.

Metric Category	Classification Outputs			
	'0'	'1'	Accuracy	Weighted
Precision	75.83%	71.64%	74.54%	74.20%
Recall	85.78%	56.77%	74.54%	74.54%
F1-Score	80.50%	63.34%	74.54%	73.85%
Support	47,703	30,176	74.54%	77,879
- Accuracy Score = 74.54%				
Confusion Matrix: [[40,921 6,782] [13,045 17,131]]				

According to the numeric data of assessment findings shown in Table 4.5, Figure 4.15 describes a graphical illustration related to the numerical simulation outcomes of the confusion matrix associated with the SGD classifier for the training process.

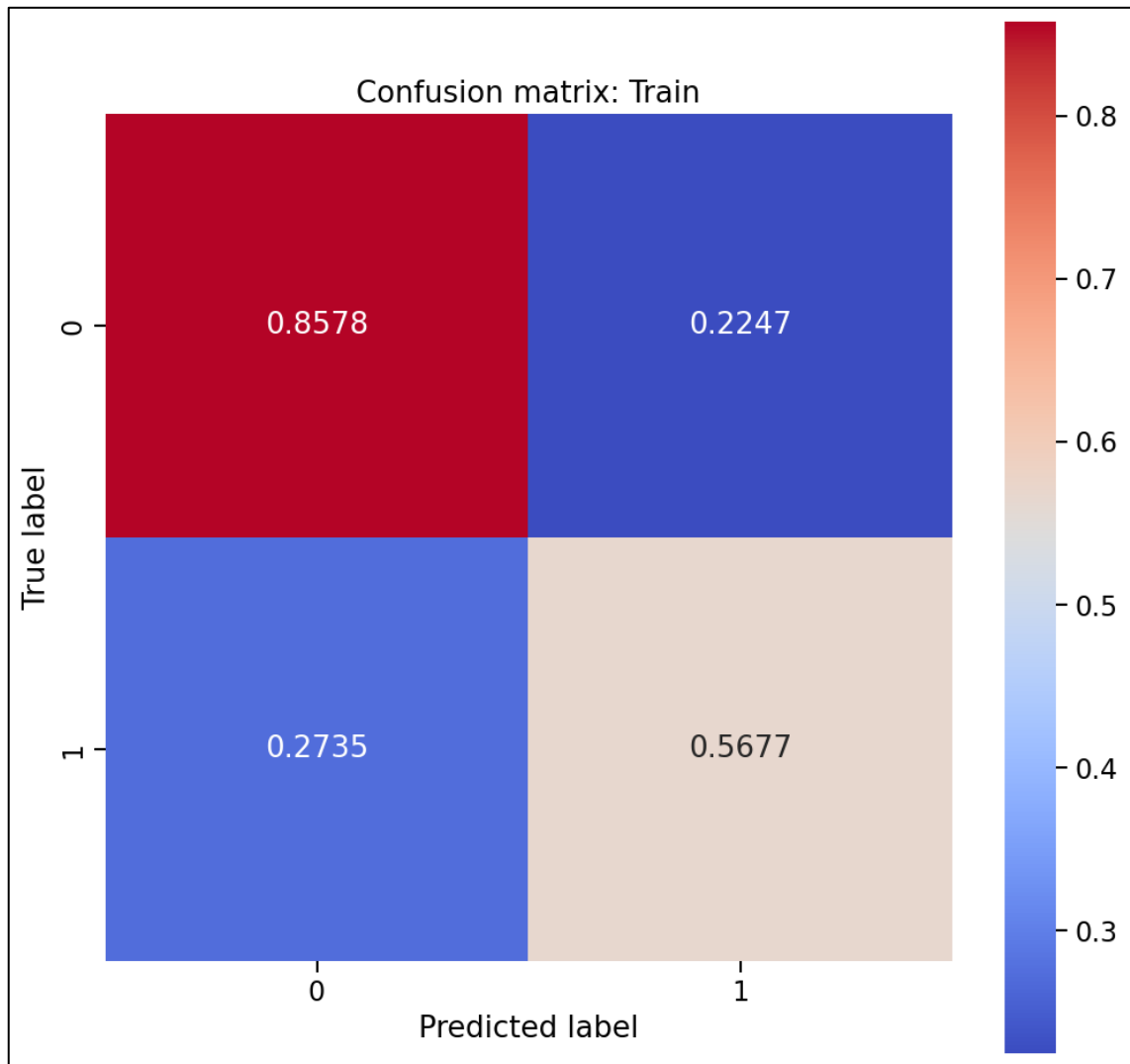


Figure 4.15: The Numerical Simulation Findings of The Confusion Matrix Linked to The SGD Model – Training Process.

It is concluded from the results illustrated in Figure 4.15 that the ‘1’ Predicted Label with the ‘0’ True Label contributed to the most negligible value of correlation, corresponding to roughly 0.2247. Meantime, the Predicted Label value of ‘0’ with the True Label value of ‘0’ gave the most considerable value of correlation with a rate of 0.8578.

Moreover, Table 4.6 represents the numerical research outputs related to the testing process of the SGD classifier.

Table 4.6: The Numerical Outcomes of The Testing Process Related To The SGD Classifier.

Metric Category	Classification Outputs			
	'0'	'1'	Accuracy	Weighted
Precision	75.00%	72.50%	74.23%	74.01%
Recall	85.79%	56.73%	74.23%	74.23%
F1-Score	80.03%	63.65%	74.23%	73.52%
Support	15,632	10,328	74.23%	25,960
- Accuracy Score = 74.23%				
Confusion Matrix: [[13,410 2,222] [4,469 5,859]]				

Depending on those metric data, Figure 4.16 describes the numerical simulation findings of the confusion matrix linked to the SGD classifier for the testing process.

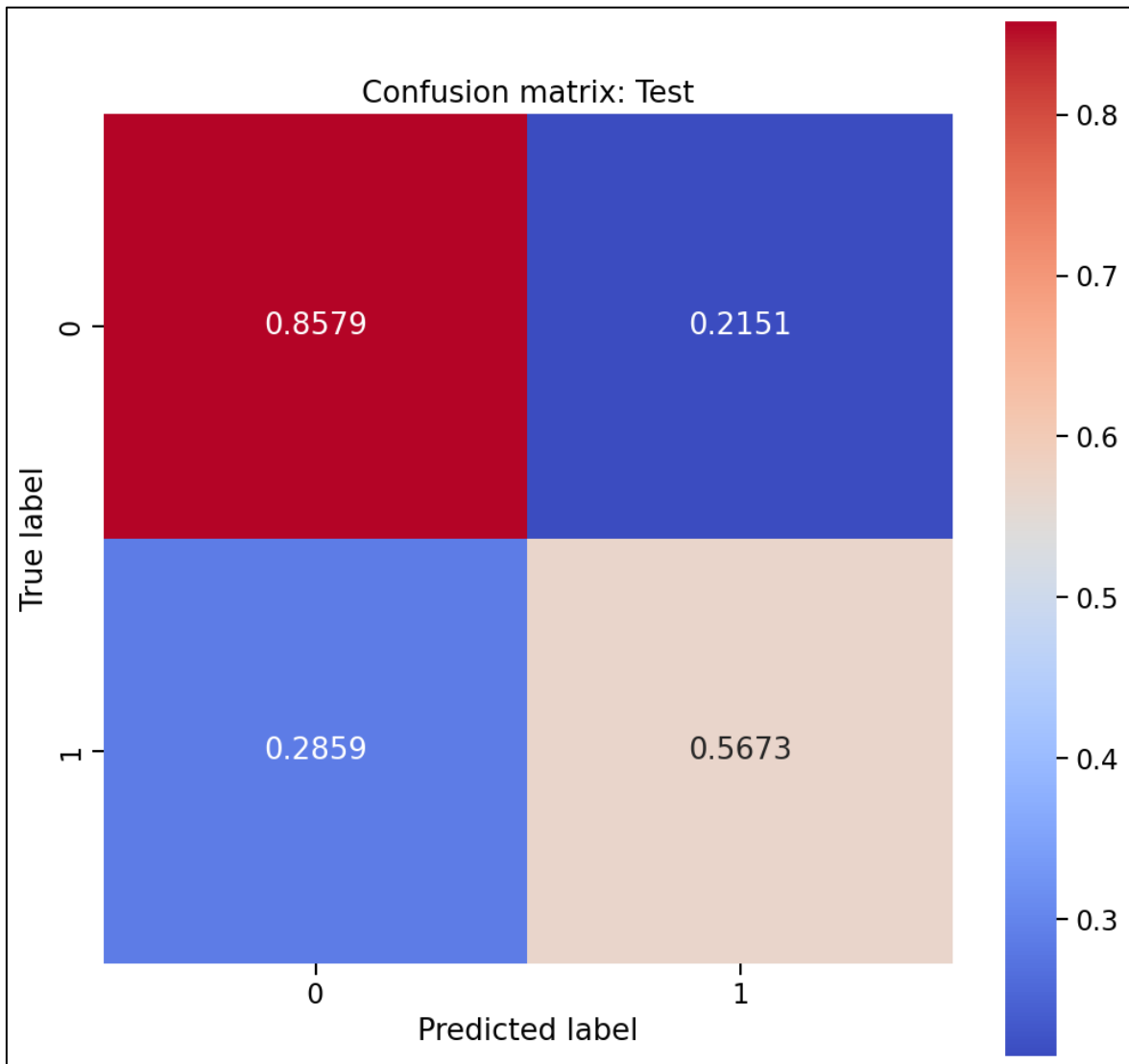


Figure 4.16: The Numerical Simulation Findings of The Confusion Matrix Linked To The SGD Model – Testing Process.

It is concluded from the results illustrated in Figure 4.16 that the ‘1’ Predicted Label with the ‘0’ True Label contributed to the most negligible value of correlation, corresponding to roughly 0.2247. Meantime, the Predicted Label value of ‘0’ with the True Label value of ‘0’ gave the most considerable value of correlation with a rate of 0.8578.

4.6.4 EFFECTIVENESS EVALUATION OF MLM

As implemented in the previous sections, the same procedure of performance assessment metrics was applied and used to examine the reliability and performance linked to the MLM.

Thus, the same four assessment metrics were employed as they were utilized in prior ML schemes. Those examination procedures are vital to identify the performance of this MLM in executing cyber threat detection actively.

The numerical outcomes linked to the training process related to the MLM classifier are shown in Table 4.7.

Table 4.7: The Numerical Outcomes of The Training Process Related To The MLM.

Metric Category	Classification Outputs			
	'0'	'1'	Accuracy	Weighted
Precision	98.27%	93.16%	96.20%	96.29%
Recall	95.48%	97.34%	96.20%	96.20%
F1-Score	96.85%	95.20%	96.20%	96.21%
Support	47,703	30,176	96.20%	77,879
- Accuracy Score = 96.20%				
Confusion Matrix: {[45,547 2,156] [803 29,373]}				

Relying on the numeric data of assessment findings shown in Table 4.7, Figure 4.17 describes a graphical illustration related to the numerical simulation outcomes of the confusion matrix referring to the MLM, considering the training process.

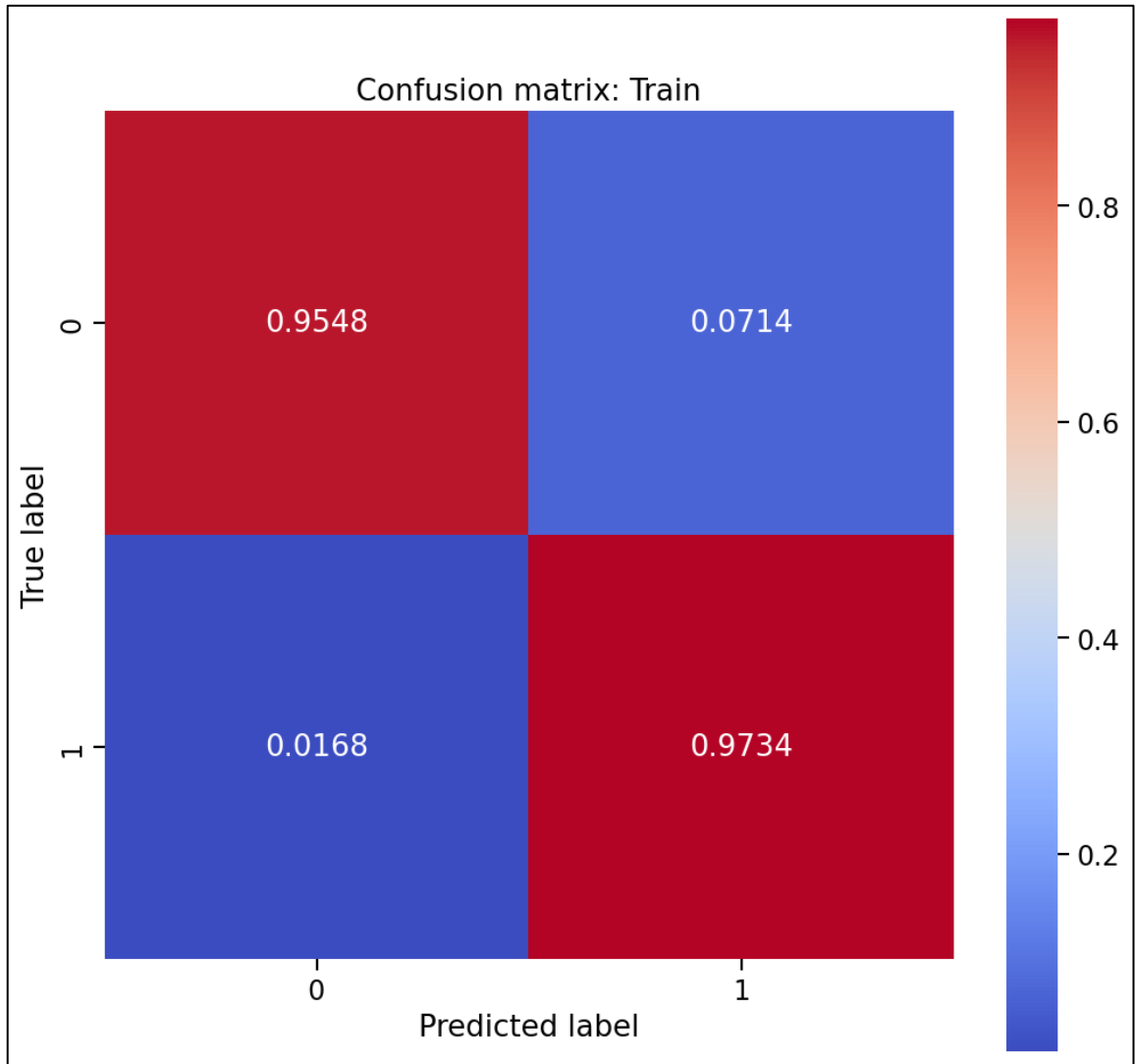


Figure 4.17: The Numerical Simulation Findings of The Confusion Matrix Linked to The MLM – Training Process.

It can be deduced from the results presented in Figure 4.17 that the ‘1’ Predicted Label with the ‘1’ True Label contributed to the most considerable amount of correlation, with a rough value of 0.9734. Meanwhile, the Predicted Label value of ‘0’ with the True Label value of ‘1’ provided the most negligible correlation rate with 0.0168.

Besides, Table 4.8 illustrates the numeric findings associated with a similar DL scheme but for the testing scenario.

Table 4.8: The Numerical Outcomes of The Testing Process Related to The MLM.

Metric Category	Classification Outputs			
	'0'	'1'	Accuracy	Weighted
Precision	97.94%	93.08%	95.51%	96.01%
Recall	95.24%	96.97%	95.51%	95.93%
F1-Score	96.57%	94.99%	95.51%	95.94%
Support	15,632	10,328	95.51%	25,960
- Accuracy Score = 95.93%				
Confusion Matrix: [[14,888 744] [313 10,015]]				

Based on the numeric data of MLM assessment outcomes expressed in Table 4.8, Figure 4.18 describes a graphical illustration related to the numerical simulation outcomes of the confusion matrix referring to the MLM, considering the testing phase.

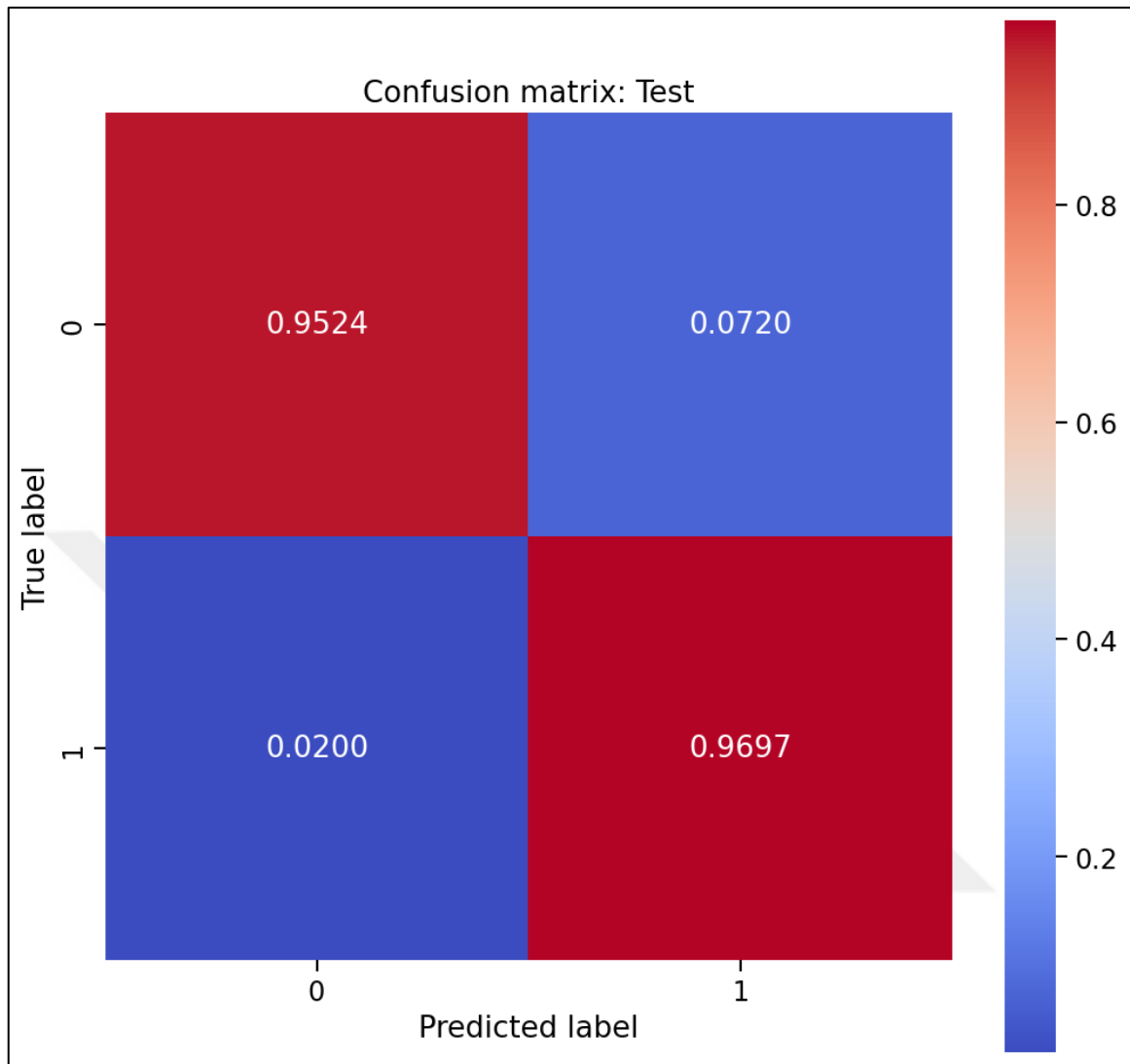


Figure 4.18: The Numerical Simulation Findings of The Confusion Matrix Linked to The MLM – Testing Process.

It can be inferred from the numerical findings presented in Figure 4.18 that the ‘1’ Predicted Label with the ‘1’ True Label contributed to the most remarkable value of correlation, with an approximate rate of 0.9697. At the same time, the Predicted Label value of ‘0’ with the True Label value of ‘1’ provided the lowest correlation value of 0.0200.

4.7 DISCUSSIONS

The critical research findings associated with the ML models and DL algorithms investigated in this work revealed that the average accuracy of the four ML schemes obtained via the performance assessment techniques ranges from 74.54% to 96.20% for training. The minor ratio was for the SGD classifier. In comparison, the maximum percentage was associated with MLM. On the other hand, the average accuracy related to the four algorithms but for the testing process varies between 74.23% (minimum for the SGD classifier) and 95.93% (maximum for MLM). Furthermore, it was discovered that using intelligent ML and DL algorithms provided practical means of cyber threat detection for the thesis's case study in terms of performance, reliability, and accuracy. Moreover, the precision rates of the four ML algorithms range from 75.83% (minimum for the SGD classifier) to 98.27% (maximum for MLM), concerning the '0' results of binary classifications in the training process. For the binary classification results of '1,' the minimum portions were between 69.14% (minimum for the LR algorithm) and 93.16% (maximum for MLM), respectively, in the training phase. For the testing procedure, the precision rate varies between 75.00% (minimum for the SGD classifier) and 97.94% (maximum for MLM) for the '0' binary classification findings. In comparison, the '1' binary classification outputs range between 70.15% (the lowest for the LR model) and 93.08% (ultimate for MLM) in the testing procedure. The results of this master's work are consistent with the research findings of [73]–[76], who carried out an analysis exploring the critical benefits of ML in making accurate and high-performance classification and detection of various threats and aspects associated with different areas, including cyber security, healthcare, data privacy, and big data classification. In addition, the numerical findings of this work are compatible with the study outputs of [67], [68], who reported that the SGD classifier has lower performance and efficacy compared with other intelligent ML algorithms.

5. CONCLUSIONS AND RECOMMENDATIONS

5.1 CHAPTER GOAL

After the numerical research findings are represented and described in the fourth chapter, it is essential to mention some critical remarks and concluding points that summarize all the simulation outputs carried out with the processing of Python numeric code. Furthermore, it is crucial to introduce some practical recommendations and critical suggestions to help researchers and business owners protect their organizations from severe cyber threats that may cause dangerous financial and database damage. This chapter provides all those essential points.

5.2 CONCLUSIONS

This master's research is directed at investigating and examining the beneficial contributions and crucial rationale of ML principles and intelligent DL algorithms in detecting cyber threats in organizations and network systems that might result in significant damage to financial resources and private databases. The work relied on employing four ML models (including Logistic Regression, Support Vector Machine, Stochastic Gradient Descent, and Machine Learning Model). Depending on the mathematical numeric Python code simulations, the research findings can be summarized in the following points:

- a. The average accuracy of the four ML schemes obtained via the performance assessment techniques ranges from 74.54% to 96.20% for training. The minor ratio was for the SGD classifier.
- b. The maximum percentage was associated with MLM. On the other hand, the average accuracy related to the four algorithms for the testing process varies between 74.23% (minimum for the SGD classifier) and 95.93% (maximum for MLM).
- c. Using intelligent ML and DL algorithms provided practical means of cyber threat detection for the thesis's case study in terms of performance, reliability, and accuracy.

- d. The precision rates of the four ML algorithms range from 75.83% (minimum for the SGD classifier) to 98.27% (maximum for MLM), concerning the '0' results of binary classifications in the training process.
- e. For the binary classification results of '1,' the minimum portions were between 69.14% (minimum for the LR algorithm) and 93.16% (maximum for MLM), respectively, in the training phase.
- f. For the testing procedure, the precision rate varies between 75.00% (minimum for the SGD classifier) and 97.94% (maximum for MLM) for the '0' binary classification findings. In comparison, the '1' binary classification outputs range between 70.15% (the lowest for the LR model) and 93.08% (ultimate for MLM) in the testing procedure.

5.3 RECOMMENDATIONS

Relying on the critical numerical simulation findings and mathematical analysis in this master's thesis, it is recommended to follow some vital suggestions that could help protect the financial resources and data privacy in different sectors, including the following aspects:

- i. Elevate the levels of awareness and attention regarding the severity and danger of cyber threats and internet attacks, especially for business owners and organizations' managers.
- ii. Carry out training sessions and educational courses for employees and senior personnel in different organizations to help protect their data privacy and information security by virtue of ML tactics.
- iii. Organize various workshops, conferences, and meetings with scholars and business owners to figure out some practical techniques that could alleviate and stop the cyber threats with the help of ML concepts and DL approaches.

REFERENCES

- [1] A. , G. B. B. , T. A. , S. A. , & M. A. Srivastava, “A recent survey on DDoS attacks and defense mechanisms.,” *In Advances in Parallel Distributed Computing: First International Conference on Parallel, Distributed Computing Technologies and Applications, PDCTA 2011, Tirunelveli, India, September 23-25, 2011*, pp. 570–580, 2011.
- [2] S. Tan, J. M. Guerrero, P. Xie, R. Han, and J. C. Vasquez, “Brief Survey on Attack Detection Methods for Cyber-Physical Systems,” *IEEE Syst J*, vol. 14, no. 4, pp. 5329–5339, Dec. 2020, doi: 10.1109/JSYST.2020.2991258.
- [3] S. Tripathi, B. Gupta, A. Almomani, A. Mishra, and S. Veluru, “Hadoop Based Defense Solution to Handle Distributed Denial of Service (DDoS) Attacks,” *Journal of Information Security*, vol. 04, no. 03, pp. 150–164, 2013, doi: 10.4236/jis.2013.43018.
- [4] Naveen Bindra and Manu Sood, “Detecting DDoS Attacks Using Machine Learning Techniques and Contemporary Intrusion Detection Dataset,” *Automatic Control and Computer Sciences*, vol. 53, no. 5, pp. 419–428, Sep. 2019, doi: 10.3103/S0146411619050043.
- [5] K. , S. H. , & K. L. Napanda, “Artificial intelligence techniques for network intrusion detection.,” *International Journal of Engineering Research & Technology (IJERT)*, ISSN, 2278-0181., 2015.
- [6] D. , R. C. , M. R. , & A. N. F. Glăvan, “DDoS detection and prevention based on artificial intelligence techniques.,” *Scientific Bulletin" Mircea cel Batran" Naval Academy*, 22(1), 1-11., 2019.
- [7] O. A. Osanaiye, “Short Paper: IP spoofing detection for preventing DDoS attack in Cloud Computing,” in *2015 18th International Conference on Intelligence in Next Generation Networks*, 2015, pp. 139–141. doi: 10.1109/ICIN.2015.7073820.

- [8] G. Lucky, F. Jjunju, and A. Marshall, "A Lightweight Decision-Tree Algorithm for detecting DDoS flooding attacks," in *2020 IEEE 20th International Conference on Software Quality, Reliability and Security Companion (QRS-C)*, Dec. 2020, pp. 382–389. doi: 10.1109/QRS-C51114.2020.00072.
- [9] S. T. , J. J. , & T. D. Zargar, "A survey of defense mechanisms against distributed denial of service (DDoS) flooding attacks.," *IEEE communications surveys & tutorials*, 15(4), pp. 2046–2069, 2013.
- [10] N. Z. Bawany, J. A. Shamsi, and K. Salah, "DDoS Attack Detection and Mitigation Using SDN: Methods, Practices, and Solutions," *Arab J Sci Eng*, vol. 42, no. 2, pp. 425–441, Feb. 2017, doi: 10.1007/s13369-017-2414-5.
- [11] S. , A. M. , H. O. , F. A. , & B. P. Elgendy, "DTEXNet: Artificial Intelligence-Based Combination Scheme for DDoS Attacks Detection.," 2022.
- [12] T. Peng, C. Leckie, and K. Ramamohanarao, "Survey of network-based defense mechanisms countering the DoS and DDoS problems," *ACM Comput Surv*, vol. 39, no. 1, p. 3, Apr. 2007, doi: 10.1145/1216370.1216373.
- [13] M. Aamir and M. Arif, "Study and Performance Evaluation on Recent DDoS Trends of Attack & Defense," *International Journal of Information Technology and Computer Science*, vol. 5, no. 8, pp. 54–65, Jul. 2013, doi: 10.5815/ijitcs.2013.08.06.
- [14] C. Douligeris and A. Mitrokotsa, "DDoS attacks and defense mechanisms: classification and state-of-the-art," *Computer Networks*, vol. 44, no. 5, pp. 643–666, Apr. 2004, doi: 10.1016/j.comnet.2003.10.003.
- [15] J. Mirkovic and P. Reiher, "A taxonomy of DDoS attack and DDoS defense mechanisms," *ACM SIGCOMM Computer Communication Review*, vol. 34, no. 2, pp. 39–53, Apr. 2004, doi: 10.1145/997150.997156.
- [16] H. Jiang, Z. He, G. Ye, and H. Zhang, "Network Intrusion Detection Based on PSO-Xgboost Model," *IEEE Access*, vol. 8, pp. 58392–58401, 2020, doi: 10.1109/ACCESS.2020.2982418.

- [17] Ismail *et al.*, “A Machine Learning-Based Classification and Prediction Technique for DDoS Attacks,” *IEEE Access*, vol. 10, pp. 21443–21454, 2022, doi: 10.1109/ACCESS.2022.3152577.
- [18] G. Kumar, K. Kumar, and M. Sachdeva, “The use of artificial intelligence based techniques for intrusion detection: a review,” *Artif Intell Rev*, vol. 34, no. 4, pp. 369–387, Dec. 2010, doi: 10.1007/s10462-010-9179-5.
- [19] D. Han, K. Bi, H. Liu, and J. Jia, “A DDoS attack detection system based on spark framework,” *Computer Science and Information Systems*, vol. 14, no. 3, pp. 769–788, 2017, doi: 10.2298/CSIS161217028H.
- [20] F. Tao, M. Akhtar, and Z. Jiayuan, “The future of Artificial Intelligence in Cybersecurity: A Comprehensive Survey,” *EAI Endorsed Transactions on Creative Technologies*, vol. 8, no. 28, p. 170285, Aug. 2021, doi: 10.4108/eai.7-7-2021.170285.
- [21] T. H. H. Aldhyani and H. Alkahtani, “Artificial Intelligence Algorithm-Based Economic Denial of Sustainability Attack Detection Systems: Cloud Computing Environments,” *Sensors*, vol. 22, no. 13, p. 4685, Jun. 2022, doi: 10.3390/s22134685.
- [22] X. Liang and T. Znati, “On the performance of intelligent techniques for intensive and stealthy DDos detection,” *Computer Networks*, vol. 164, p. 106906, Dec. 2019, doi: 10.1016/j.comnet.2019.106906.
- [23] I. Ortet Lopes, D. Zou, F. A. Ruambo, S. Akbar, and B. Yuan, “Towards Effective Detection of Recent DDoS Attacks: A Deep Learning Approach,” *Security and Communication Networks*, vol. 2021, pp. 1–14, Nov. 2021, doi: 10.1155/2021/5710028.
- [24] A. Jaszcz and D. Połap, “AIMM: Artificial Intelligence Merged Methods for flood DDoS attacks detection,” *Journal of King Saud University - Computer and Information Sciences*, vol. 34, no. 10, pp. 8090–8101, Nov. 2022, doi: 10.1016/j.jksuci.2022.07.021.

- [25] R. Amrish, K. Bavapriyan, V. Gopinaath, A. Jawahar, and C. Vinoth Kumar, "DDoS Detection using Machine Learning Techniques," *Journal of ISMAC*, vol. 4, no. 1, pp. 24–32, May 2022, doi: 10.36548/jismac.2022.1.003.
- [26] J. Shroff, R. Walambe, S. K. Singh, and K. Kotecha, "Enhanced Security Against Volumetric DDoS Attacks Using Adversarial Machine Learning," *Wirel Commun Mob Comput*, vol. 2022, pp. 1–10, Mar. 2022, doi: 10.1155/2022/5757164.
- [27] O. Veranyurt, "Usage of Artificial Intelligence in DOS/DDOS Attack Detection," 2019.
- [28] Y. S. Hussain, "Network Intrusion Detection for Distributed Denial-of-Service (DDoS) Attacks using Machine Learning Classification Techniques," 2020.
- [29] S. Das, A. M. Mahfouz, D. Venugopal, and S. Shiva, "DDoS Intrusion Detection Through Machine Learning Ensemble," in *2019 IEEE 19th International Conference on Software Quality, Reliability and Security Companion (QRS-C)*, Jul. 2019, pp. 471–477. doi: 10.1109/QRS-C.2019.00090.
- [30] B. A. Khalaf, S. A. Mostafa, A. Mustapha, M. A. Mohammed, and W. M. Abdullallah, "Comprehensive Review of Artificial Intelligence and Statistical Approaches in Distributed Denial of Service Attack and Defense Methods," *IEEE Access*, vol. 7, pp. 51691–51713, 2019, doi: 10.1109/ACCESS.2019.2908998.
- [31] D. Preuveneers and W. Joosen, "Sharing Machine Learning Models as Indicators of Compromise for Cyber Threat Intelligence," *Journal of Cybersecurity and Privacy*, vol. 1, no. 1, pp. 140–163, Feb. 2021, doi: 10.3390/jcp1010008.
- [32] G. S. S. Et.al, "Application of Artificial Intelligence in detection of diseases in plants: A Survey," *Turkish Journal of Computer and Mathematics Education (TURCOMAT)*, vol. 12, no. 3, pp. 3301–3305, Apr. 2021, doi: 10.17762/turcomat.v12i3.1581.
- [33] U. Islam *et al.*, "Detection of Distributed Denial of Service (DDoS) Attacks in IOT Based Monitoring System of Banking Sector Using Machine Learning Models," *Sustainability*, vol. 14, no. 14, p. 8374, Jul. 2022, doi: 10.3390/su14148374.

- [34] R. M. Alguliyev, R. M. Aliguliyev, and F. J. Abdullayeva, "Deep Learning Method for Prediction of DDoS Attacks on Social Media," *Advances in Data Science and Adaptive Analysis*, vol. 11, no. 01n02, p. 1950002, Apr. 2019, doi: 10.1142/S2424922X19500025.
- [35] A. Sanmorino, "A study for DDOS attack classification method," *J Phys Conf Ser*, vol. 1175, p. 012025, Mar. 2019, doi: 10.1088/1742-6596/1175/1/012025.
- [36] S. Sambangi and L. Gondi, "A Machine Learning Approach for DDoS (Distributed Denial of Service) Attack Detection Using Multiple Linear Regression," in *The 14th International Conference on Interdisciplinarity in Engineering—INTER-ENG 2020*, Dec. 2020, p. 51. doi: 10.3390/proceedings2020063051.
- [37] D. Alghazzawi, O. Bamasag, H. Ullah, and M. Z. Asghar, "Efficient Detection of DDoS Attacks Using a Hybrid Deep Learning Model with Improved Feature Selection," *Applied Sciences*, vol. 11, no. 24, p. 11634, Dec. 2021, doi: 10.3390/app112411634.
- [38] M. Shurman, R. Khrais, and A. Yateem, "DoS and DDoS Attack Detection Using Deep Learning and IDS," *The International Arab Journal of Information Technology*, vol. 17, no. 4A, pp. 655–661, Jul. 2020, doi: 10.34028/iajit/17/4A/10.
- [39] Z. Liu, L. Qian, and S. Tang, "The prediction of DDoS attack by machine learning," in *Third International Conference on Electronics and Communication; Network and Computer Technology (ECNCT 2021)*, Mar. 2022, p. 55. doi: 10.1117/12.2628658.
- [40] L. Li and G. Lee, "DDoS Attack Detection and Wavelets," *Telecommun Syst*, vol. 28, no. 3–4, pp. 435–451, Mar. 2005, doi: 10.1007/s11235-004-5581-0.
- [41] M. Alduailij, Q. W. Khan, M. Tahir, M. Sardaraz, M. Alduailij, and F. Malik, "Machine-Learning-Based DDoS Attack Detection Using Mutual Information and Random Forest Feature Importance Method," *Symmetry (Basel)*, vol. 14, no. 6, p. 1095, May 2022, doi: 10.3390/sym14061095.
- [42] K. S. Sahoo, A. Iqbal, P. Maiti, and B. Sahoo, "A Machine Learning Approach for Predicting DDoS Traffic in Software Defined Networks," in *2018 International*

- Conference on Information Technology (ICIT)*, Dec. 2018, pp. 199–203. doi: 10.1109/ICIT.2018.00049.
- [43] M. Mittal, K. Kumar, and S. Behal, “Deep learning approaches for detecting DDoS attacks: a systematic review,” *Soft comput*, Jan. 2022, doi: 10.1007/s00500-021-06608-1.
 - [44] IGOR. VUKOVIĆ, “APPLICATION OF ARTIFICIAL INTELLIGENCE IN DETECTION OF DDOS ATTACKS,” 2020.
 - [45] G. A. Jaafar, S. M. Abdullah, and S. Ismail, “Review of Recent Detection Methods for HTTP DDoS Attack,” *Journal of Computer Networks and Communications*, vol. 2019, pp. 1–10, Jan. 2019, doi: 10.1155/2019/1283472.
 - [46] M. , & R. M. J. Alenezi, “Methodologies for detecting DoS/DDoS attacks against network servers. ,” *In The Seventh International Conference on Systems and Networks Communications ICSNC (pp. 92-98).* , 2012.
 - [47] N. Ahuja, G. Singal, D. Mukhopadhyay, and N. Kumar, “Automated DDOS attack detection in software defined networking,” *Journal of Network and Computer Applications*, vol. 187, p. 103108, Aug. 2021, doi: 10.1016/j.jnca.2021.103108.
 - [48] N. Ahuja, G. Singal, and D. Mukhopadhyay, “DDOS attack SDN dataset.” Mendeley Data, The data set is created as a part of the research work at Bennett University, India, 2020.
 - [49] X. Zou, Y. Hu, Z. Tian, and K. Shen, “Logistic Regression Model Optimization and Case Analysis,” in *2019 IEEE 7th International Conference on Computer Science and Network Technology (ICCSNT)*, Oct. 2019, pp. 135–139. doi: 10.1109/ICCSNT47585.2019.8962457.
 - [50] S. U. Hassan, J. Ahamed, and K. Ahmad, “Analytics of machine learning-based algorithms for text classification,” *Sustainable Operations and Computers*, vol. 3, pp. 238–248, 2022, doi: 10.1016/j.susoc.2022.03.001.

- [51] S. Alqadhi *et al.*, “Combining logistic regression-based hybrid optimized machine learning algorithms with sensitivity analysis to achieve robust landslide susceptibility mapping,” *Geocarto Int*, pp. 1–26, Jan. 2022, doi: 10.1080/10106049.2021.2022009.
- [52] F. Itoo, Meenakshi, and S. Singh, “Comparison and analysis of logistic regression, Naïve Bayes and KNN machine learning algorithms for credit card fraud detection,” *International Journal of Information Technology*, vol. 13, no. 4, pp. 1503–1511, Aug. 2021, doi: 10.1007/s41870-020-00430-y.
- [53] F. Thabtah, N. Abdelhamid, and D. Peebles, “A machine learning autism classification based on logistic regression analysis,” *Health Inf Sci Syst*, vol. 7, no. 1, p. 12, Dec. 2019, doi: 10.1007/s13755-019-0073-5.
- [54] M. Rezapour, A. Mehrara Molan, and K. Ksaibati, “Analyzing injury severity of motorcycle at-fault crashes using machine learning techniques, decision tree and logistic regression models,” *International Journal of Transportation Science and Technology*, vol. 9, no. 2, pp. 89–99, Jun. 2020, doi: 10.1016/j.ijtst.2019.10.002.
- [55] Q. Lei, H. Lv, H. Zhang, H. Sun, and L. Tang, “Logistic regression-based device-free localization in changeable environments,” in *2016 IEEE 13th International Conference on Signal Processing (ICSP)*, Nov. 2016, pp. 1062–1066. doi: 10.1109/ICSP.2016.7877992.
- [56] T. Rymarczyk, E. Kozłowski, G. Kłosowski, and K. Niderla, “Logistic Regression for Machine Learning in Process Tomography,” *Sensors*, vol. 19, no. 15, p. 3400, Aug. 2019, doi: 10.3390/s19153400.
- [57] J. H. Cheon, D. Kim, Y. Kim, and Y. Song, “Ensemble Method for Privacy-Preserving Logistic Regression Based on Homomorphic Encryption,” *IEEE Access*, vol. 6, pp. 46938–46948, 2018, doi: 10.1109/ACCESS.2018.2866697.
- [58] K. Kirasich, T. Smith, and B. Sadler, “Random forest vs logistic regression: binary classification for heterogeneous datasets,” *SMU Data Science Review*, vol. 1, no. 3, pp. 1–24, 2018.

- [59] R. Balina, M. Idasz-Balina, and N. A. Achسانی, “Predicting Insolvency of the Construction Companies in the Creditworthiness Assessment Process—Empirical Evidence from Poland,” *Journal of Risk and Financial Management*, vol. 14, no. 10, p. 453, Sep. 2021, doi: 10.3390/jrfm14100453.
- [60] M. Ashrafuzzaman, S. Das, Y. Chakhchoukh, S. Shiva, and F. T. Sheldon, “Detecting stealthy false data injection attacks in the smart grid using ensemble-based machine learning,” *Comput Secur*, vol. 97, p. 101994, Oct. 2020, doi: 10.1016/j.cose.2020.101994.
- [61] T. T. Khoei, A. Gasimova, M. A. Ahajjam, K. al Shamaileh, V. Devabhaktuni, and N. Kaabouch, “A Comparative Analysis of Supervised and Unsupervised Models for Detecting GPS Spoofing Attack on UAVs,” in *2022 IEEE International Conference on Electro Information Technology (eIT)*, May 2022, pp. 279–284. doi: 10.1109/eIT53891.2022.9813826.
- [62] M. S. Khan, L. Khan, N. Gul, M. Amir, J. Kim, and S. M. Kim, “Support Vector Machine-Based Classification of Malicious Users in Cognitive Radio Networks,” *Wirel Commun Mob Comput*, vol. 2020, pp. 1–11, Jul. 2020, doi: 10.1155/2020/8846948.
- [63] I. H. Sarker, “Machine Learning: Algorithms, Real-World Applications and Research Directions,” *SN Comput Sci*, vol. 2, no. 3, p. 160, May 2021, doi: 10.1007/s42979-021-00592-x.
- [64] G. E. I. Selim, E. E.-D. Hemdan, A. M. Shehata, and N. A. El-Fishawy, “Anomaly events classification and detection system in critical industrial internet of things infrastructure using machine learning algorithms,” *Multimed Tools Appl*, vol. 80, no. 8, pp. 12619–12640, Mar. 2021, doi: 10.1007/s11042-020-10354-1.
- [65] M. A. Almaiah *et al.*, “Performance Investigation of Principal Component Analysis for Intrusion Detection System Using Different Support Vector Machine Kernels,” *Electronics (Basel)*, vol. 11, no. 21, p. 3571, Nov. 2022, doi: 10.3390/electronics11213571.

- [66] D. Wang, J. Peng, Q. Yu, Y. Chen, and H. Yu, "Support Vector Machine Algorithm for Automatically Identifying Depositional Microfacies Using Well Logs," *Sustainability*, vol. 11, no. 7, p. 1919, Mar. 2019, doi: 10.3390/su11071919.
- [67] A. McCarthy, E. Ghadafi, P. Andriotis, and P. Legg, "Functionality-Preserving Adversarial Machine Learning for Robust Classification in Cybersecurity and Intrusion Detection Domains: A Survey," *Journal of Cybersecurity and Privacy*, vol. 2, no. 1, pp. 154–190, Mar. 2022, doi: 10.3390/jcp2010010.
- [68] D. T. Anh *et al.*, "Effect of Gradient Descent Optimizers and Dropout Technique on Deep Learning LSTM Performance in Rainfall-runoff Modeling," *Water Resources Management*, vol. 37, no. 2, pp. 639–657, Jan. 2023, doi: 10.1007/s11269-022-03393-w.
- [69] M. Payvand, M. E. Fouda, F. Kurdahi, A. M. Eltawil, and E. O. Neftci, "On-Chip Error-Triggered Learning of Multi-Layer Memristive Spiking Neural Networks," *IEEE J Emerg Sel Top Circuits Syst*, vol. 10, no. 4, pp. 522–535, Dec. 2020, doi: 10.1109/JETCAS.2020.3040248.
- [70] S. Ray, "A Quick Review of Machine Learning Algorithms," in *2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon)*, Feb. 2019, pp. 35–39. doi: 10.1109/COMITCon.2019.8862451.
- [71] T. Le-Duc, Q.-H. Nguyen, J. Lee, and H. Nguyen-Xuan, "Strengthening Gradient Descent by Sequential Motion Optimization for Deep Neural Networks," *IEEE Transactions on Evolutionary Computation*, pp. 1–1, 2022, doi: 10.1109/TEVC.2022.3171052.
- [72] H. Shi, N. Yang, H. Tang, and X. Yang, "aSGD: Stochastic Gradient Descent with Adaptive Batch Size for Every Parameter," *Mathematics*, vol. 10, no. 6, p. 863, Mar. 2022, doi: 10.3390/math10060863.
- [73] K.-A. Tait *et al.*, "Intrusion Detection using Machine Learning Techniques: An Experimental Comparison," in *2021 International Congress of Advanced Technology*

- and Engineering (ICOTEN)*, Jul. 2021, pp. 1–10. doi: 10.1109/ICOTEN52080.2021.9493543.
- [74] T. M. Ghazal *et al.*, “IoT for Smart Cities: Machine Learning Approaches in Smart Healthcare—A Review,” *Future Internet*, vol. 13, no. 8, p. 218, Aug. 2021, doi: 10.3390/fi13080218.
- [75] M. I. Tariq *et al.*, “A Review of Deep Learning Security and Privacy Defensive Techniques,” *Mobile Information Systems*, vol. 2020, pp. 1–18, Apr. 2020, doi: 10.1155/2020/6535834.
- [76] M. A. Amanullah *et al.*, “Deep learning and big data technologies for IoT security,” *Comput Commun*, vol. 151, pp. 495–517, Feb. 2020, doi: 10.1016/j.comcom.2020.01.016.
- [77] N. Ahmed *et al.*, “Network Threat Detection Using Machine/Deep Learning in SDN-Based Platforms: A Comprehensive Analysis of State-of-the-Art Solutions, Discussion, Challenges, and Future Research Direction,” *Sensors*, vol. 22, no. 20, p. 7896, Oct. 2022, doi: 10.3390/s22207896.