IoT NETWORK TRAFFIC ANALYSIS USING MACHINE LEARNING BASED TECHNIQUES

A MAJOR PROJECT-II REPORT

SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE AWARD OF THE DEGREE
OF

MASTER OF TECHNOLOGY
IN
INFORMATION SYSTEMS

Submitted by:
ABHISHEK SHUKLA
2K22/ISY/02

Under the supervision of
MR. RAHUL GUPTA

DEPARTMENT OF INFORMATION AND TECHNOLOGY
DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Bawana Road, Delhi-110042

May, 2024

DELHI TECHNOLOGICAL UNIVERSITY
(FORMERLY DELHI COLLEGE OF ENGINEERING)
Bawana Road, Delhi-110042

CANDIDATE'S DECLARATION

I, Abhishek Shukla, 2K22/ISY/02 student of M.Tech in Information Systems, hereby declare that the Major Project-II dissertation titled "**IoT NETWORK TRAFFIC ANALYSIS USING MACHINE LEARNING BASED TECHNIQUES**" which is submitted by me to the Department of Information Technology, Delhi Technological University, Delhi in partial fulfillment of the requirement for the award of the degree of Master of Technology, is original and not copied from any source without proper citation. This work has not previously formed the basis for the award of any degree, Diploma Associateship, Fellowship, or other similar title or recognition.

Place: Delhi
Date: 28/05/2024

ABHISHEK SHUKLA
2K22/ISY/02

DEPARTMENT OF INFORMATION TECHNOLOGY
DELHI TECHNOLOGICAL UNIVERSITY
(FORMERLY DELHI COLLEGE OF ENGINEERING)
Bawana Road, Delhi-10042

CERTIFICATE

I hereby certify that the Major Project-II dissertation titled "**IoT NETWORK TRAFFIC ANALYSIS USING MACHINE LEARNING BASED TECHNIQUES**" which is submitted by Abhishek Shukla, 2K22/ISY/12, Department of Information Technology, Delhi Technological University, Delhi in partial fulfillment of the requirement for the award of the degree of Master of Technology, is a record of the project work carried out by the student under my supervision. To the best of my knowledge, this work has not been submitted inpart or full for any degree or diploma to this University or elsewhere.

Place: Delhi
Date: 28/05/2024

Mr. RAHUL GUPTA
SUPERVISOR
Assistant Professor
Department of Information Technology
DELHI TECHNOLOGICAL UNIVERSITY

ABSTRACT

Despite the advantages of the IoT, a rising amount of malware designed specifically for IoT devices poses a serious danger to the Internet's environment. These malware attacks have created the need to evaluate the IoT system's security and the need to create defenses against potential threats. It is crucial to stop IoT malwares from spreading. Furthermore, for educating people and evaluating the accuracy of cyber security the gathering and researching of data from many sources of IoT data is essential. Regarding this, we need to analyze the network traffic. Previous research papers suggested some models, which is based on Long Short-Term Memory (LSTM), successfully completes two tasks 1) Identifying the benign nature of the given traffic and 2) Identifying the sort of malware to look for in malicious network data. For this, there is need for a sizable amount of traffic data from the number of files of both good and bad traffic which can be gathered from different distinct IoT devices. Flowrelated, traffic flag-related, and packet payload related characteristics were the three modalities into which the features that were retrieved from the datasets at the feature and modality levels, a feature selection technique was used, and the best modalities and features were applied for performance improvement. After we apply a number of Machine Learning algorithm for analyzing the traffic to find if it is benign or malicious.

ACKNOWLEDGEMENT

I would like to express my sincere gratitude towards my supervisor Mr. Rahul Gupta for providing her invaluable guidance, comments, and suggestions throughout the course of the project.

The results of this thesis would not have been possible without support from all who directly or indirectly, have lent their hand throughout the course of the project. I would like to thank my parents and faculties of the Department of Information Technology, Delhi Technological University, for their kind cooperation and encouragement which helped me complete this thesis. I hope that this project will serve its purpose to the fullest extent possible.

ABHISHEK SHUKLA
2K22/ISY/02

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

1.1 IoT

In the field of telecommunications, out of many technologies the one which stood out is the Internet of Things (IoT). The Internet of Things is the physical objects of any network, vehicle, device, building, and other objects that are embedded with things such as electronics, sensors, software and connectivity to the network. These objects are able to collect and exchange information through the network. The Internet of Things is the next big step in this era of new technology and its exponential growth in the number of devices connected to networks are testimony to its popularity and success.

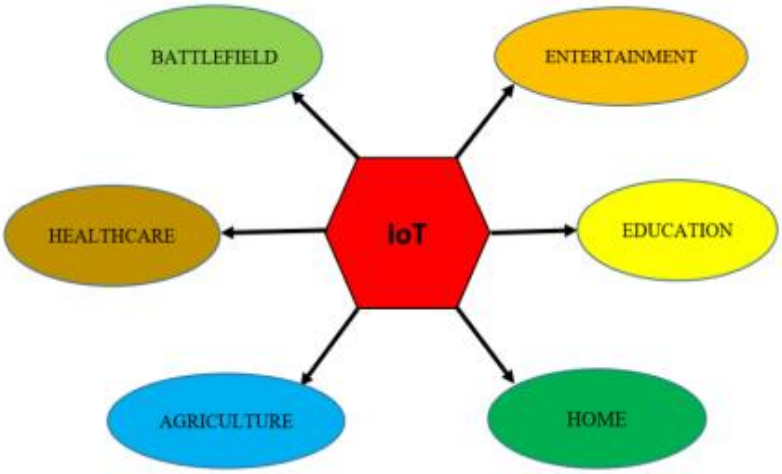


Fig. 1.1. IoT and its uses.

1.2 Need for Analysis

But with the growing number of devices, it raise about concern of dangerous online activities. These type of attacks has the potential to harm someone's privacy. Many has to compromise with their data and information as it may be altered by the attackers. Sometimes there are financial losses happen as malign actors get the access to user's bank credentials [1]. Researchers found out that IoT device commonly use

the open-source component and framework that have various security vulnerabilities and attackers shifted their target towards these badly secures smart devices.

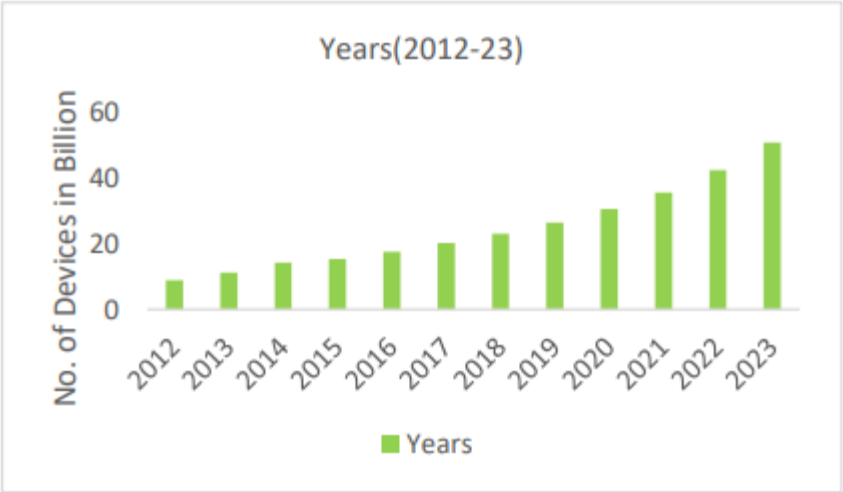


Fig. 1.2. No. of IoT Devices (in Billion)

To minimize and protect from such kinds of malicious activities much research work has been done in this field using machine learning[2]. But it was not up to mark with respect to accuracy or only compatible with specific kind of malware or devices. In this study we delve into the realm of IoT malware detection using an ensemble learning framework. Our approach combines three distinct classifiers: k-Nearest Neighbours (k-NN), Random Forrest, Decision Tree, and logistic regression [3][4]. By leveraging their complementary strengths, we aim to create a robust and adaptive system capable of identifying both known and novel malware variants. The ensemble learning paradigm capitalizes on the wisdom of crowds, aggregating predictions from multiple models to achieve superior performance Each classifier contributes its unique perspective, and their collective decision-making process enhances overall accuracy making it well-suited for the IoT landscape.

CHAPTER 2

RELATED WORK

This paper [6] states that the increasing dependence on the Internet and the importance of protecting information from theft are critical concerns, to address this, an advanced NIDS based on DL methodology is being developed using the "NSL-KDD" dataset. The proposed CNN model has a learning accuracy of 95.5% and performs better than the forward neural network due to its larger hidden layer. The model is efficient in time and accuracy and can be optimized by reorganizing parameters and hyper parameters. Preprocessing is crucial for improving the accuracy of attack labelled data. The model also includes a dropout layer to avoid exploding and vanishing gradients.

This paper [7] states that as technology advances, society relies on computers and the internet for data storage and retrieval. Intrusion Detection Systems (IDSs) are essential to protect these networks, particularly the Internet of Things (IoT). IDSs are needed to protect physical objects from unauthorized access and vulnerabilities. This paper explores existing IDSs for IoT, compares different IDSs based on features, and implements two known attacks, DIS and Version, on Cooja. The study aims to analyse the impact of these attacks on the network, identify abnormal behaviour, and predict the location of malicious nodes.

This paper [8] paper examines static IoT malware detection, focusing on its definition, development, and security risks. It reviews existing IoT malware location techniques and provides strategies for future research. The paper also provides a comprehensive audit of IoT malware, revealing its increasing frequency and increasing complexity. It also presents a comparative analysis of the malware's components, location, and handling time. The authors propose a lightweight identification strategy to manage recognized malicious executable files in IoT devices.

This paper [9] paper explores static IoT malware detection, focusing on its definition, development, and security risks. It reviews existing techniques and proposes a lightweight identification strategy for managing malicious executable files in IoT devices. Cybercrimes are prevalent in IoTs, posing security, privacy, and identity

verification issues. The paper suggests future research strategies and addressing vulnerabilities in IoT devices.

This paper [10] paper explores static IoT malware detection, focusing on its definition, development, and security risks. It reviews existing techniques and proposes a lightweight identification strategy for managing malicious executable files. The paper highlights the need for best practices in IoT devices and suggests application whitelisting as a first line of defense against malware. Other mechanisms like verified boot and remote attestation are also crucial.

This paper [11] states that due to its ability to reduce hazards and enable real-time monitoring, the Internet of Things, or IoT, has become essential for industrial organizations. However possible abnormalities in IoT networks give rise to worries about safety and security. This study presents a way for quickly classifying malicious software assaults on Internet of Things networks by utilizing deep learning and machine learning techniques. The Decision Tree (DT) method is used because of its cheap computing time cost and great accuracy. The study found that Gaussian Naive Bayes performed the least well when it came to machine learning malware identification on the Avast IoT-23 challenge dataset.

This paper [12] study examines IoT malware and suggests a scalable architecture for IoT honeypots called HoneyIoT. HoneyIoT uses firmware support and adjustable vulnerabilities to record harmful actions and draw in assaults. During seven-day industrial testing, 3,423 unsuccessful login attempts and over 12,500 fraudulent connections were recorded by Honeypots. The study highlights how crucial back-end implementation and front-end interaction capabilities are for IoT honeypots. IoT botnet activity has been greatly influenced by the Mozi virus, which blends elements of three different IoT malware families. According to the study, Mozi's influence on other IoT malware variants is expected to last for years.

This paper [13] presents HoneyIoT, a scalable framework for IoT honeypots, designed to detect and record malicious behaviours. The framework uses Federated Learning (FL) to create and evaluate guided and solo models without disclosing sensitive information. It is designed for use on network nodes that grant access to IoT devices on Wi-Fi, 5G, or B5G networks. The N-BaIoT dataset was used to

demonstrate its applicability in an actual IoT environment. The federated strategy enhances model performance in both supervised and unsupervised scenarios while ensuring data privacy. The federated models' resistance to malicious clients was assessed using various attacks. The paper emphasizes the need for robust aggregation methods to protect against adversarial attacks

The paper [14] proposes a cross-architecture IoT malware detection system using GAN to address the issue of cross-architecture malware. The system uses Opcode and PSI feature characteristics of graph nodes and a CFG retrieved from a binary executable file. It uses GAT to acquire neighbourhood attributes, assign varying weights to neighbours, and perform distinct training phases to enhance system efficiency and protect user privacy. The system achieves a 99.67% detection accuracy.

The AASH technique (IoT Malware Detection) is a cutting-edge approach that finds malware at the source code level by utilizing the Fibonacci search with the Adler-32 hash function. This method [15] works better than earlier approaches like DROIDMD and SQVDT, which have longer detection times and lower accuracy but are scalable and may be used on IoT devices. AASH outperformed DROIDMD and SQVDT by 19.8 seconds and 20.7 seconds, respectively, in detecting harmful pieces in 17.3 seconds. It performed better in accuracy as well, hitting 91.2% accuracy. To find harmful code in third-party code, the AASH approach can be used more widely.

This article [16] investigates the definition, evolution, and security hazards of static IoT malware detection. It talks about the rising incidence of IoT malware and evaluates 0 20 40 60 No. of Devices in Billion Years (2012-23) Years 3 current methods. A comparative analysis of IoT malware and its effectiveness is also presented in the paper. The authors stress the significance of tackling the particular difficulties in recognizing and managing IoT malware in the context of the Internet and private information by proposing a lightweight identification technique to manage recognized dangerous executable files in IoT devices.

The proliferation of malware poses a serious security threat to the increasing number of IoT networks. Malware detection and isolation on their own are insufficient, and conventional control ideas don't work [17]. A two-pronged method is suggested: a stochastic model predictive controller and a HaRM that uses HPC values. Compared to previous systems, the suggested technique provides a quicker detection of malware

with an accuracy of 92.21% and a 10ns latency. As a result, the average network performance is over 200% higher than in networks of IoT devices without inbuilt defence. In an IoT network with 20 nodes, the technique is tested, and it achieves an average virus detection accuracy of 92% with a 10ns delay.

The IoT has given rise to several security risks, including new malware that uses leaked source code to target IoT devices [18]. A distributed modular solution called EDIMA has been created by researchers to identify IoT malware network activities in large-scale networks during the scanning/infecting stage. To classify edge device traffic, EDIMA uses machine learning techniques in conjunction with a policy module, a packet traffic feature vector database, and an optional packet sub-sampling module. Through tests, the study assessed the performance of EDIMA's classification. It also plans to keep improving software-based implementation and performance evaluation. Upcoming projects will involve modifying cutting-edge botnet detection methods to detect malware activities.

Due to its ability to reduce hazards and enable real-time monitoring, the Internet of Things, or IoT, has become essential for industrial organizations [19].This is the also become the reason for their attack. However possible anomalies in IoT networks give rise to worries about safety and security. With the help of machine learning and deep learning techniques, this research presents a novel method for quickly classifying malicious software attacks on Internet of Things networks. The Decision Tree (DT) algorithm is selected because of its low computing time cost and great accuracy. Five models are used in the study, comprising one deep learning approach and four machine learning techniques. To compare approaches and broaden the area of the investigation, more research is required.

The Internet of Things, or IoT, is expanding quickly, opening up new application possibilities but also raising security concerns [20]. Due to resource limitations, antivirus software designed for desktop computers and servers is not immediately relevant to IoT devices, making malware detection difficult. This paper introduces SIMBIoTA++, a resource-requirement-optimized version of SIMBIoTA. For malware identification, SIMBIoTA and SIMBIoTA++ both use a similarity metric on binary files. The suggested enhancement is a new dominating set updating algorithm that, at the same computation time, produces smaller sets of TLSH hash values for

Internet of Things devices. It is discovered that the 40 TLSH difference threshold is a meaningful threshold number since it maximizes the malware samples' similarity graph's clustering coefficient.

Global adoption of the Internet of Things (IoT) is accelerating, yet security is still lacking. Through the scalability of machine learning techniques to enable near-real-time network traffic anomaly detection and packet classification as benign or malicious, this study [21] investigates the viability of malware detection in a single Internet of Things device. An ESP32 device can be used to implement the recommended software for classifying data points from the IoT-23 dataset. This could allow IoT devices to determine if a network connection is a part of a malware attack or a routine connection. Operational code sequences are transformed into vector space using the study's deep learning technique, which then classifies them into harmful and hazardous applications. The effectiveness of the suggested strategy against garbage and virus detection Code insertion attacks is shown, and the classifier maintains good metrics while being fast, small, and precise. Subsequent efforts could concentrate on putting in place a comprehensive software that gathers internet packets, separates features, and classes them

Because of its computing power, malware has made the Internet of Things (IoT) a frequent target. Owing to domain knowledge requirements, detecting IoT malware with machine learning and deep learning approaches is difficult. This work [22] eliminates the need for domain expertise by proposing a novel Convolution model that solves the problem using raw byte sequences. With a 99.01% accuracy rate, the model can correctly identify IoT binaries as malicious or benign. The study compared three models and discovered that spatial content is more effective in identifying IoT malware than temporal dynamics. This leads to improved algorithms by improving our understanding of malware binary files. The results can suggest future lines of inquiry for further study.

IoT devices are on the target of malware assaults because of their constrained computational power, low security index, along with resource availability. Botnet attacks, which cause distributed denial of service (DDoS) attacks, can affect these devices. Although these threats can be detected by machine learning techniques, lengthy processing times provide a problem. This research [23] presents an up to

100% accurate malware attack detection system based on the CART learning algorithm. The findings indicate that Naïve Bayes outperforms other learning algorithms when given specific features. Further research will examine these additional options.

CHAPTER 3

MATERIAL & METHODOLOGY

3.1 Dataset

An organized set of data items put together for analysis, modelling, or decision-making is called a dataset. The suggested model makes use of "The Bot-IoT Dataset," which was created by UNSW Canberra Cyber Range Lab. It consists of a mix of malicious (botnet) and benign (normal) traffic. The dataset may be found as produced argus files, csv files, and original pcap files. To facilitate dataset administration, we employed specific MySQL queries in this work to extract 5% of the original dataset. This dataset has 32 attribute and out of them the ‘attack’ attribute is going to provide us the information about whether it is normal or malicious.

pkSeqID	daddr	Dur	Sbytes
Stime	dport	Mean	Dbytes
flgs	pkts	stddev	Rate
flgs_number	bytes	Sum	Srate
Proto	state	Min	Drate
proto_number	state_number	Max	Attack
saddr	ltime	Spkts	Category
sport	seq	dpkts	subcategory

Table 3.1. Attributes of Bot-IoT Dataset

3.2 Data Pre-processing

3.2.1 Data Encoding.

In data encoding the column with categorical data is converted into numerical type data so that they can be fitted into machine learning models which only accept numerical type data. To perform encoding we have used level encoding and its output is numpy array.

3.2.2 Correlation Matrix

It gives an idea about data set. This tells us how two columns are related to each other. If the similarity score is 1 then relationship is strong and out of them one can be dropped from the dataset to make it more optimal. If the similarity score is

0 then the relationship between them is neutral but if the score is -1 then the relationship between them is weak. Here if the similarity score is greater than 0.80(i.e. 80%) then we drop one of the column to make it easier for training and testing of models.

3.2.3 Data Balancing

Predictive modelling has difficulties when dealing with unbalanced datasets however, these issues are expected as there are many imbalanced cases in real life. Balancing the dataset makes it easier to train a model since it prevents the model from being biased toward a certain class. Stated differently, the approach does not automatically benefit the majority class just because it has access to more data. We oversample the dataset in this study due to the small sample size of the minority class. We increase the duplicate data fields of the minority class to make 70:30 ratio for dataset.

3.2.4 Reshuffling

One essential pre-processing method that is often used to enhance model learning is data shuffling. Data shuffling is intended to address any problems that may arise from patterns in the training sample sequence, which may cause overfitting.

3.3 Models Used

3.3.1 K-Nearest Neighbour

K-Nearest Neighbours is a non-parametric classification and regression approach in machine learning that uses the similarity principle to classify or forecast a new data point's label or value. The algorithm's performance and generalization capacity are influenced by the chosen K value, the higher the value of K higher the complexity. Despite the need for distance computations between new and existing data points, K-NN is resilient to noisy data and performs well with large datasets. Generally, the value of K is chosen to be odd to form the majority and here we took the value of K as 3.

3.3.2 Random Forest

Using random data sets, the Random Forest method is a machine learning technique that reduces overfitting and enhances prediction accuracy by building

Decision Trees. For reliable and accurate findings, it averages or combines the vote results from each tree. Regression and classification problems make extensive use of it.

3.3.3 Decision Tree

Decision trees are a machine learning model used for regression and classification applications. They consist of a hierarchy of decision nodes and leaf nodes, each representing a feature or class label. They divide feature space iteratively to maximize homogeneity and purity. Decision trees are interpretable and useful for human-readable explanations. However, they can overfit and catch noise in data. Ensemble techniques like Random Forests and Adaboost reduce this problem. Despite their drawbacks, decision trees remain essential for adaptability, transparency, and efficiency.

3.3.4 Logistic Regression

It is a supervised machine learning technique that is used to forecast the likelihood that an instance will fall into a specific class in binary classification. A probability value between 0 and 1 is obtained by applying a sigmoid function to examine the connection between two data components.

3.3.5 Ensemble Learning

We use Adaboost, decision trees, and K-Nearest Neighbors (K-NN) ensemble learning paradigms to demonstrate the adaptability of ensemble learning. It demonstrates how different algorithms can be used synergistically to address challenging learning tasks, improving prediction accuracy, robustness, and generalization.

3.4 Methodology

In this paper first, we applied data encoding to make our dataset numerical, the reason behind this is that some of the models that we are using only accept numerical values. Here we used level encoding for encoding the data. After that, we made a correlation matrix which tells us about how much two columns are related to each other and from that if the similarity between two columns comes out 80% or more than that then we eliminated one out of those two column so that

it will be easier for training the models. This process is repeated until a similarity score of 80% is achievable between any two attributes.

After that, this preprocessed dataset is split into training and testing sets, which are in the ratio of 70:30.The training class will be used to train the models and once the models gets trained the other part of dataset i.e. testing class is used on these models to get their results and evaluation.

Now, as our data is highly imbalanced the size of minority class is quite low so we performed data balancing (oversampling) to increase the minority class to make the ratio up to 70:30 in favor of the majority class.

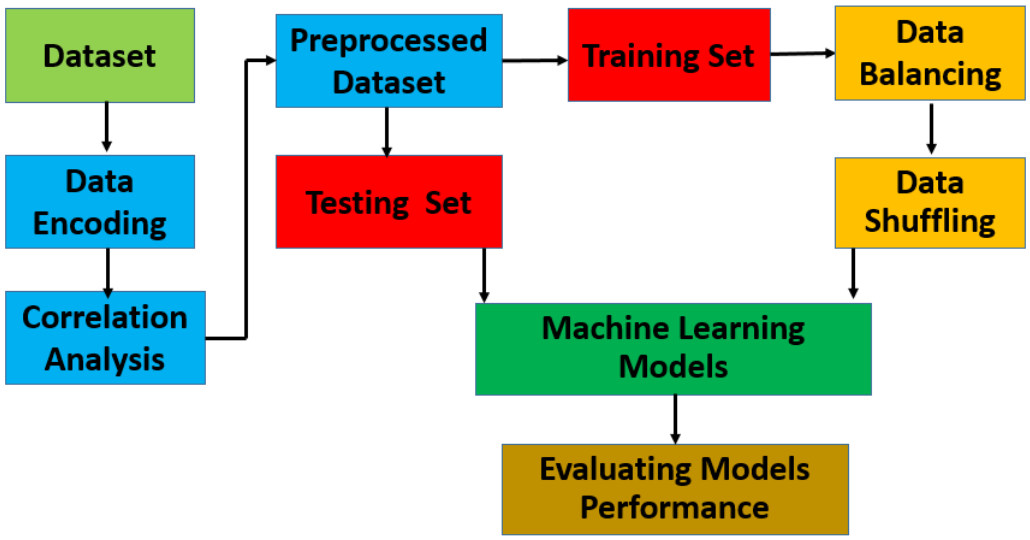


Fig. 3.1. Model Architecture.

Data shuffling is performed to make it susceptible to overfitting and enhance the learning of models. It also addresses any problems that arise from patterns in the training datasets. And at last, all the machine learning algorithms used in this research are trained using this shuffled dataset soon after the testing process is initiated to evaluate the model performance.

KNN algorithm's performance and generalization capacity are influenced by the chosen K value, the higher the value of K higher the complexity. Despite the need for distance computations between new and existing data points, K-NN is resilient to noisy data and performs well with large datasets. Generally, the value of K is chosen to be odd to form the majority and here we took the value of K as 3.

Two components make up Random Forest. The first step is to combine N decision trees to create a random forest, and the second is to forecast each tree that was created in the first phase. Select K data points at random from the training set. Construct the decision trees that correspond to the selected data points. If you want to build decision trees, choose the number N. Locate the predictions for the new data points in each decision tree by following the preceding procedures again, and then assign them to the group that received the majority of votes.

Decision Trees consist of a hierarchy of decision nodes and leaf nodes, each representing a feature or class label. They divide feature space iteratively to maximize homogeneity and purity. Decision trees are interpretable and useful for human-readable explanations. However, they can overfit and catch noise in data. Ensemble techniques like Random Forests and Adaboost reduce this problem. Despite their drawbacks, decision trees remain essential for adaptability, transparency, and efficiency.

One common technique for estimating the logistic regression model is maximum likelihood estimation (MLE). This approach iterates by trying various values of beta several times to get the best match for the log odds. Each of these repetitions produces a log-likelihood function, which logistic regression seeks to maximize to provide the best possible parameter estimate. If the ideal coefficient—or coefficients, if there are several independent variables—are found, the conditional probabilities for each observation may then be calculated, logged, and summed to get a prediction probability. In binary classification, a probability of less than 0.5 predicts 0 while a probability of more than 0.50 predicts 1.

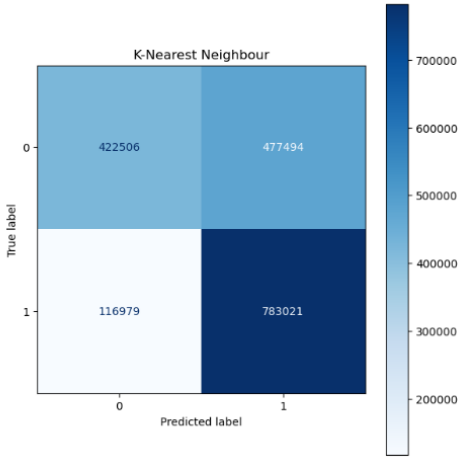
And at last, we use the ensemble learning technique which uses different machine learning paradigms i.e. K-NN, Random Forest, Decision Tree, and logistic regression to demonstrate how different algorithms can be used synergistically to address challenging learning tasks, improving prediction accuracy, robustness, and generalization.

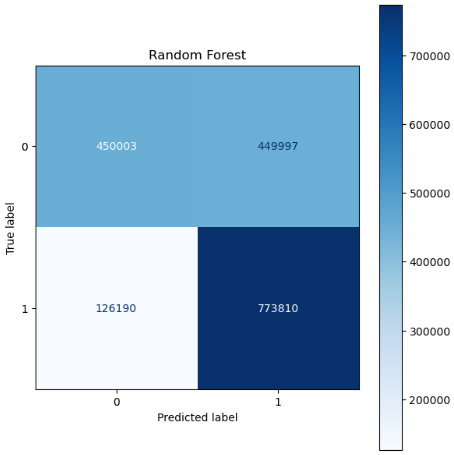
CHAPTER 4

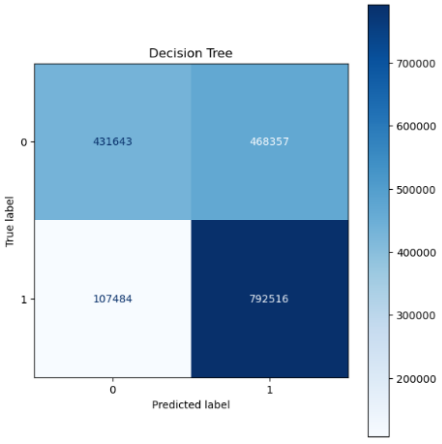
RESULTS AND DISCUSSION

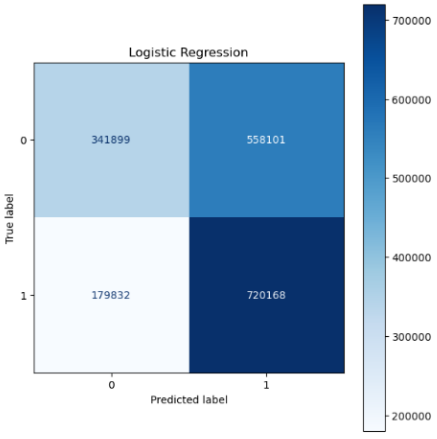
4.2 Confusion Matrix

A confusion matrix tabulates the counts of TP, FP, TN and FN to give a clear breakdown of the performance of a classification model. It functions as a fundamental tool for assessing a classifier's efficacy and comprehending the kinds of mistakes it produces.









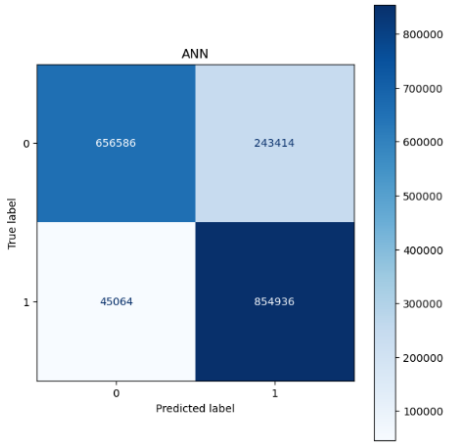


Fig. 4.1. Confusion Matrix of **(a).** K-NN **(b).** Random Forest **(c).** Decision Tree **(d)** Logistic Regression and **(e)** Ensemble Learning on Bot-IoT Dataset

4.2 Accuracy

In machine learning, accuracy is a critical indicator that shows the percentage of properly identified or predicted instances relative to the total. It is typically given as a percentage and is computed by dividing the total number of correct guesses by the total number of forecasts. Accuracy, while simple, might not adequately convey subtleties of performance, particularly in datasets that are unbalanced or where distinct errors have varying weights. Even though it's straightforward, accuracy is a useful metric for assessing a model's overall efficacy.

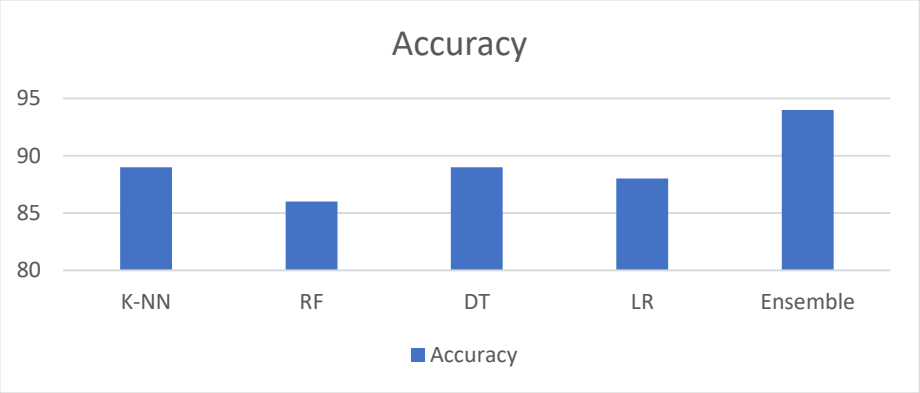


Fig. 4.2. Comparison of Accuracy on Bot-IOT dataset

4.3 Precision

Precision is a crucial metric in classification issues that evaluates a model's ability to forecast successful outcomes. It is calculated as the ratio of TP forecasts to all positive predictions, or TP and FP. A low FPR is shown by high precision, which

makes the model especially helpful in situations where reducing false positives is essential.

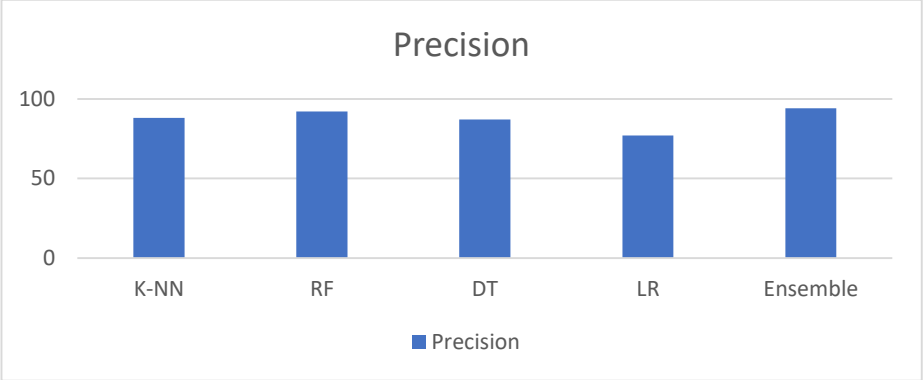


Fig. 4.3. Comparison of Precision on Bot-IOT dataset

4.4 Recall

Recall, which is often referred to as sensitivity or TPR, assesses how well a model can distinguish TP cases from all of the positive instances that actually occurred. It is computed as the ratio of TP instances (TP plus FN) to TPR. High recall means that, even at the cost of more FP, the model successfully captures a significant percentage of positive cases, which makes it useful in situations when finding every positive is essential.

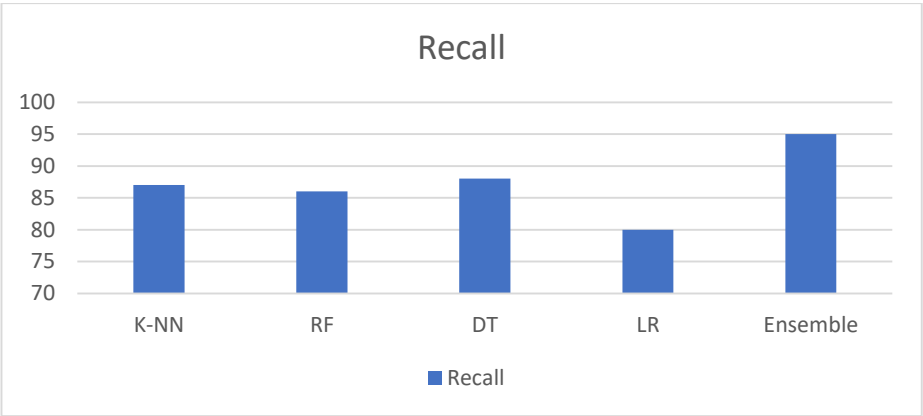


Fig. 4.4. Comparison of Recall on Bot-IOT dataset

4.5 F-1 Score

By balancing accuracy and recall, the F1 score is a single statistic that offers a comprehensive assessment of a classifier's performance. Recall and accuracy are

harmonically meaned to get a single value that combines the two metrics. A higher score indicates better overall performance since it finds a balance between recall and accuracy. The F1 score is between 0 and 1.

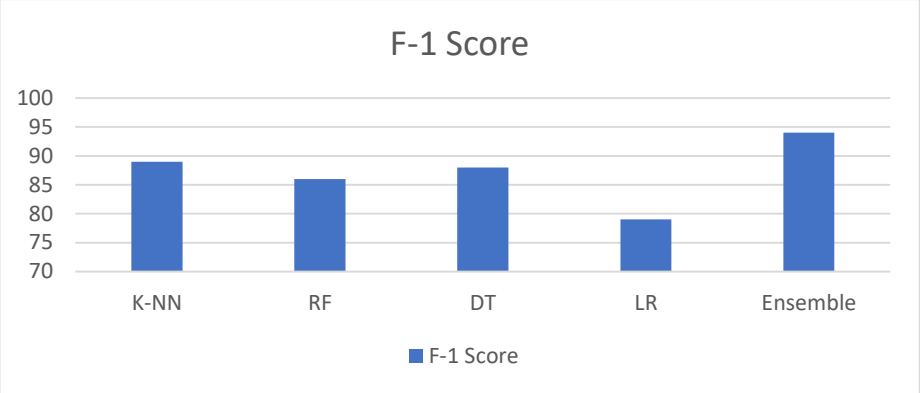


Fig. 4.5. Comparison of F-1 Score on Bot-IOT dataset

ML Models	Accuracy	Precision	Recall	F-1 Score
K-Nearest Neighbour	0.89	0.88	0.87	0.89
Random Forest	0.86	0.92	0.86	0.86
Decision Tree	0.89	0.87	0.88	0.88
Logistic Regression	0.88	0.77	0.80	0.79
Ensemble Learning	0.94	0.95	0.95	0.94

Table 4.1 Performance Matrices Table

As we can see from the table these five models including ensemble learning i.e. K-Nearest Neighbour, Random Forest, Decision Tree, and Logistic Regression have a learning accuracy of 89%, 86%, 89%, 88%, and 94% respectively. Their Precision Score is 88%, 92%, 87%, 77%, and 95% respectively. Their Recall Score is 87%, 86%, 88%, 80% and 95% respectively. Their F_1Score Score is 89%, 86%, 88%, 79% and 94% respectively. Hence from here, we conclude that

based on the learning of all models, Ensemble Learning outperformed all other models individually in all performance matrices.

4.6 ROC Curve

The AUC is a performance measure in binary classification, evaluating a model's ability to differentiate between positive and negative classes. The ROC curve, plotted against the genuine positive rate, indicates a model's enhanced discriminatory capacity.

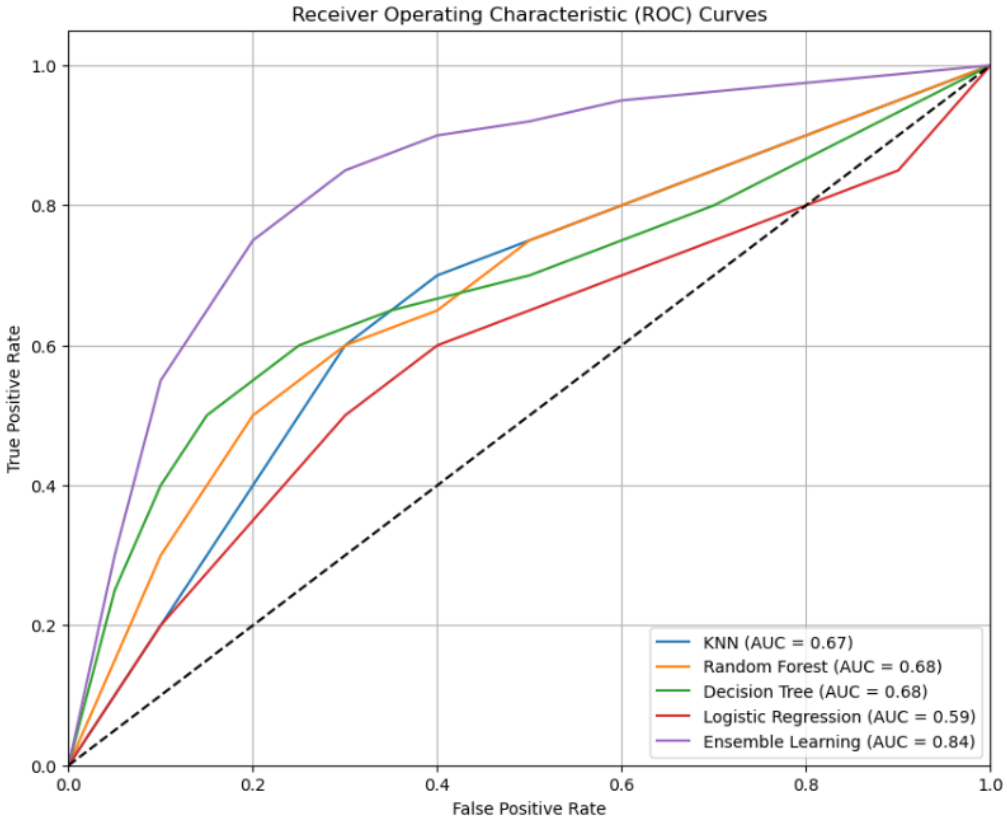


Fig. 4.6. ROC Curve

CHAPTER 5

CONCLUSION

As the IoT gadgets numbers are on the ascent, so is the abuse of them for malicious reasons and there is a requirement for a proficient method of distinguishing such threats from the normal class of networks. First of all, we preprocess our dataset to make it more effective towards machine learning models and after that we make the dataset balanced (oversampling) and then shuffled it to make it ready for the training of the dataset. For solving this problem this paper proposed an ensemble learning model using adaboost, decision tree and K-NN with a learning accuracy of 94% and observed that there is a slight increase in accuracy and precision when previously they were used individually on the “The Bot-IoT dataset”. Ensemble Learning techniques are robust in nature with the help of which our model can easily analyze whether the network is normal or malign.

CHAPTER 6

FUTURE WORK

In future research, we will pay attention to improve the Accuracy and Detection Rate of IoT network to make it free from all possible attacks of malicious actors and their cybercrimes. It will be necessary to extend the study's reach by evaluating the ML/DL techniques on a variety of datasets sourced from different contexts. A more thorough grasp of performance, time efficiency, and method comparisons will be possible with such an extension

CHAPTER 7

REFERENCES

1. V. Wahi, S. Yadav, Y. Thenuia and A. Chauhan, "Anomaly Based Intrusion Detection For IoT," in *2022 3rd International Conference for Emerging Technology (INCET)*, 2022.
2. J. Lai, D. Hu, A. Yin and L. Lu, "Edge Intelligence (EI)-Enabled Malware Internet of Things (IoT) Detection System," in *2021 IEEE 4th International Conference on Computer and Communication Engineering Technology (CCET)*, 2021.
3. Y. Glani, L. Ping and S. A. Shah, "AASH: A Lightweight and Efficient Static IoT Malware Detection Technique at Source Code Level," in *2022 3rd Asia Conference on Computers and Communications (ACCC)*, 2022.
4. S. Gaba, S. Nagpal, A. Aggarwal, R. Kumar and S. Kumar, "An Analysis of Internet of Things (IoT) Malwares and detection based on Static and Dynamic Techniques," in *2022 Seventh International Conference on Parallel, Distributed and Grid Computing (PDGC)*, 2022.
5. S. M. Pudukotai Dinakarrao, H. Sayadi, H. M. Makrani, C. Nowzari, S. Rafatirad and H. Homayoun, "Lightweight Node-level Malware Detection and Network-level Malware Confinement in IoT Networks," in *2019 Design, Automation & Test in Europe Conference & Exhibition (DATE)*, 2019.
6. A. Kumar and T. J. Lim, "EDIMA: Early Detection of IoT Malware Network Activity Using Machine Learning Techniques," in *2019 IEEE 5th World Forum on Internet of Things (WF-IoT)*, 2019.
7. M. T. Jahangir, M. Wakeel, H. Asif and A. Ateeq, "Systematic Approach to Analyze The Avast IOT-23 Challenge Dataset For Malware Detection Using Machine Learning," in *2023 18th International Conference on Emerging Technologies (ICET)*, 2023.
8. L. Buttyán, R. Nagy and D. Papp, "SIMBIoTA++: Improved Similarity-based IoT Malware Detection," in *2022 IEEE 2nd Conference on Information Technology and Data Science (CITDS)*, 2022.
9. B. Sharma, R. Kumar, A. Kumar, M. Chhabra and S. Chaturvedi, "A Systematic Review of IoT Malware Detection using Machine Learning," in *2023 10th*

International Conference on Computing for Sustainable Global Development (INDIACom), 2023.

10. R. Kumar and G. Geethakumari, "Temporal Dynamics and Spatial Content in IoT Malware detection," in *TENCON 2019 - 2019 IEEE Region 10 Conference (TENCON)*, 2019.
11. C. S. Htwe, M. M. Su Thwin and Y. M. Thant, "Malware Attack Detection using Machine Learning Methods for IoT Smart Devices," in *2023 IEEE Conference on Computer Applications (ICCA)*, 2023.
12. Y. Xu, Y. Jiang, L. Yu and J. Li, "Brief Industry Paper: Catching IoT Malware in the Wild Using HoneyIoT," in *2021 IEEE 27th Real-Time and Embedded Technology and Applications Symposium (RTAS)*, 2021.
13. B. Sharma, R. Kumar, A. Kumar, M. Chhabra and S. Chaturvedi, "A Systematic Review of IoT Malware Detection using Machine Learning," in *2023 10th International Conference on Computing for Sustainable Global Development (INDIACom)*, 2023.
14. J. Sahota and N. Vlajic, "Mozi IoT Malware and Its Botnets: From Theory To Real-World Observations," in *2021 International Conference on Computational Science and Computational Intelligence (CSCI)*, 2021.
15. A. Meena, D. Nigam, D. Sharma and A. Chauhan, "Anomaly Based Intrusion Detection For IoT: (A Deep Learning Approach)," in *2021 3rd International Conference on Advances in Computing, Communication Control and Networking (ICAC3N)*, 2021.
16 M. Kumar, M. Yadav and A. Chauhan, "Outlier Analysis Based Intrusion Detection for IoT," in *2021 3rd International Conference on Advances in Computing, Communication Control and Networking (ICAC3N)*, 2021.
17. M. T. Jahangir, M. Wakeel, H. Asif and A. Ateeq, "Systematic Approach to Analyze The Avast IOT-23 Challenge Dataset For Malware Detection Using Machine Learning," in *2023 18th International Conference on Emerging Technologies (ICET)*, 2023.
18. T. S. Gopal, M. Meerolla, G. Jyostna, P. Reddy Lakshmi Eswari and E. Magesh, "Mitigating Mirai Malware Spreading in IoT Environment," in *2018 International*

Conference on Advances in Computing, Communications and Informatics (ICACCI), 2018.

19. S. Gaba, S. Nagpal, A. Aggarwal, R. Kumar and S. Kumar, "An Analysis of Internet of Things (IoT) Malwares and detection based on Static and Dynamic Techniques," in *2022 Seventh International Conference on Parallel, Distributed and Grid Computing (PDGC)*, 2022.
20. V. Clincy and H. Shahriar, "IoT Malware Analysis," in *2019 IEEE 43rd Annual ComputerSoftware and Applications Conference (COMPSAC)*, 2019.
21. A. Borys, A. Kamruzzaman, H. N. Thakur, J. C. Brickley, M. L. Ali and K. Thakur, "An Evaluation of IoT DDoS Cryptojacking Malware and Mirai Botnet," in *2022 IEEE World AI IoT Congress (AIIoT)*, 2022.
22. C. Q. Qiang, L. J. Ping, S. Gang and W. Z. Hui, "Ensemble Method For Net Traffic Classification Based On Deep Learning," in 2021 18th International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP), 2021.
23. L. Buttyán, R. Nagy and D. Papp, "SIMBIoTA++: Improved Similarity-based IoT Malware Detection," in *2022 IEEE 2nd Conference on Information Technology and Data Science (CITDS)*, 2022.

PUBLICATIONS

[1] A. Shukla and R. Gupta, "IoT Network Traffic Analysis Using Ensemble Learning" 2024, 4th International Conference on Machine Learning and Big Data Analytics (ICCMLBDA), Kurukshetra, India, 2024.

[2] A. Shukla and R. Gupta, " Important Feature Filtering in IoT Network Traffic to Enhance ML-based Classification" 2024, 5th IEEE India Council International Subsection Conference (INDISCON), Chandigarh, India, 2024.

Gmail abhi shukla <02rksa12@gmail.com>

Accepted paper in the EquinOCS system

EquinOCS <equinocs-admins@springernature.com> 28 March 2024 at 16:11
To: Abhishek Shukla <02rksa12@gmail.com>

This message has been sent by the EquinOCS system
https://equinocs.springernature.com/

PLEASE DO NOT REPLY

========

Dear Abhishek Shukla,

We are pleased to inform you that your paper

059: "IoT Network Traffic Analysis Using Ensemble Learning"

has been accepted for

4th International Conference on Machine Learning and Big Data Analytics (ICMLBDA2024)

Please find the reports beneath.

=== Comment ===

Congratulations! Your submitted paper has been reviewed and accepted for presentation at the IAASSE Technically Sponsored 4th International Conference on Machine Learning and Big Data Analytics (ICMLBDA 2024), to be held from May 9-11, 2024, in National Institute of Technology kurukshetra.

All accepted registered and presented papers will be submitted for possible inclusion Springer Proceedings in Mathematics & Statistics. (Format Available at following link: https://www.springer.com/kr/authors-editors/conference-proceedings/conference-proceedings-guidelines)

1. Register (transfer online paper registration fee) your accepted paper to the conference by following instructions of registration available at the conference website (Early Bird Registration Date: 10 April, 2024) : Registration Details Available at: https://icmlbda2024.iaasse.org/register.html

2. Please strictly follow the Springer Template and revise as per REVIEWERS' COMMENTS for your paper, which are intended to help you to improve your paper before the final publication. The listed comments should be addressed carefully in your revision.

3. Your final paper MUST be submitted by April 08, 2024.

4. In order for your paper to be published in the ICMLBDA- 2024 conference proceedings, ensure the following checklist before Camera Ready Paper submission:

i) Organization of the paper follows; title, abstract, keywords, introduction, related work, proposed work, result analysis,

09-11 May 2024

4th International Conference on

Machine Learning and Big Data Analytics (ICMLBDA) 2024 (Hybrid Mode)

Venue:

Department of Electronics and Communication Engineering
National Institute of Technology, Kurukshetra, Haryana

Important Dates

Special Session Proposals:
Feb 05, 2024

Full Paper Submission :
Feb 29, 2024

Acceptance Notification:
March 25 , 2024

Early Bird Registration:
April 10, 2024

Camera ready Paper:
April 15, 2024

ABOUT CONFERENCE

ICMLBDA 2024 provides a platform for researchers and professionals to share their research and reports of new technologies and applications in ML and Big Data Analytics like biometric recognition systems, medical diagnosis, industries, telecommunications, AI Petri Nets model-based diagnosis, gaming, stock trading, intelligent aerospace systems, robot control, law, remote sensing and scientific discovery agents and multiage systems, and natural language and Web intelligence. The conference program will include special sessions, presentations delivered by researchers from the international community, including presentations from keynote speakers and state-of-the-art lectures.
The ICMLBDA aims to bridge the gap between these non-coherent disciplines of knowledge and fosters unified development in next generation computational models for machine intelligence.

Chief Patron

Dr. B.V. Ramana Reddy
Director, NIT Kurukshetra

Patron

Dr. Brahmjit Singh
Professor ECE Dept. NIT Kurukshetra

Co-Patron

Mr. Karan Sharma
HoD ECE Dept. NIT Kurukshetra

Organizing Chair(s)

Dr. Pankaj Verma
Dr. Chhagan Charan
ECE, NIT Kurukshetra
Dr. Mohit Dua
CoE, NIT Kurukshetra

Organizing Secretary(s)

Dr. T N Sasamal
ECE, NIT Kurukshetra
Dr. Ankit Kumar Jain
CoE, NIT Kurukshetra

Conference Tracks

Track 1: Machine Learning
Track 2: Big Data Analytics

Submission Guidelines

Papers reporting original and unpublished research results pertaining to the related topics are solicited.
Full paper manuscripts must be in English of up to 10 pages as per Springer format.

PAPER SUBMISSION

Indexing

ALL ACCEPTED & PRESENTED papers will be published in SCOPUS indexed Springer Proceedings in Mathematics & Statistics (PROMS).
ISI Conference Proceedings Citation Index – ISI Web of Science, Scopus, DBLP.

https://icmlbda2024.iaasse.org/

All accepted and Presented Papers will be published in Springer Book Series

June 05, 2024

Publication (Tentative)

December 2024

SUBMIT PAPER NOW

Indexing

Post conference, proceedings will be made available to the following indexing services for possible inclusion:

Scopus®

dblp computer science bibliography

Clarivate Analytics

Google Scholar

Conference Tracks

Springer

National Institute of Technology Kurukshetra, Haryana

4th International Conference on

Machine Learning and Big Data Analytics (ICMLBDA) 2024

Certificate of Appreciation

This is to certify that Dr./Mr./Ms. **Abhishek Shukla**

has presented a paper entitled **IoT Network Traffic Analysis Using Ensemble Learning**

in the 4th International Conference on Machine Learning and Big Data Analytics (ICMLBDA) 2024 organized by Department of Electronics and Communication Engineering, National Institute of Technology kurukshetra on May 09-11, 2024.

Conference Secretary

Conference Chair

early makers | em lyon business school

CENTER FOR Global Management
CALIFORNIA STATE UNIVERSITY, SAN BERNARDINO

IAASSE

Gmail abhi shukla <02rksa12@gmail.com>

Decision on Paper ID 1017 of INDISCON 2024.

Microsoft CMT <email@msr-cmt.org> 7 June 2024 at 21:42
Reply-To: Arun Kumar Singh <ieeeindiscon2024@gmail.com>
To: Abhishek Shukla <02rksa12@gmail.com>

Dear Author,

We are pleased to inform you that your Paper ID 1017 Titled "Important Feature Filtering in IoT Network Traffic to Enhance ML-based Classification" has been accepted for oral/poster presentation at the INDISCON 2024. Further details can be found on conference website (https://ieeeindiscon.org). The detailed reviews/comments given by the reviewers are available in your Microsoft CMT Account.

Please consider the comments provided by the reviewers and revise your paper based on the comments.

Please note the following:
1. At least one of the authors of every accepted paper must register for the conference as author and present the paper in order for it to be included in the conference proceedings of INDISCON 2024, and subsequent submission to IEEE Xplore digital library.
2. For the paper to be accepted in the Final programme, it is expected that at least one author is registered, and Camera Ready Paper is submitted. Non-presented papers will not be submitted to IEEE Xplore digital library as per IEEE no-show policy.
3. Papers presented in the Conference will be eligible for submission for further consideration of publishing in the IEEE Xplore, subject to maintenance of quality, and post-conference scrutiny of response of Conference Organizers to Technical Program of Questionnaire (TPQ) on the Conference.
4. Instructions for submission of Camera Ready Paper will be notified shortly on conference website.

Looking forward to seeing you at INDISCON 2024 at PEC Chandigarh.

Best Wishes,
TPC Chairs, INDISCON 2024

CALL FOR PAPERS [EXTENDED DEADLINE]

5th IEEE India Council International Subsections Conference
INDISCON 2024

August 22-24, 2024,

Venue- Punjab Engineering College (Deemed to be University), Chandigarh, India

Theme- Science, Technology and Society

Patron:
Prof. Baldev Setia, *Director PEC Chandigarh*
Prof. Debabrata Das, *Chair IEEE India Council*

Co-Patron:
Prof. Prerna Gaur, *Chair-Elect, IEEE India Council*
Prof. A. Q. Ansari, *Chair, IEEE Delhi Section*

Honorary Chair:
Prof. Rudra Pratap, *VC Plaksha University*
Prof. Lalit Awasthi, *Director NIT UK*
Prof. Anupam Shukla, *Director NIT Surat*
Prof. B. K. Panigrahi, *IIT Delhi*
Mr. B. A. Sawale, *DG, CPRI, Bengaluru*

General Chair:
Prof. Arun Kumar Singh, *PEC Chandigarh*
Prof. Manish Hooda, *SCL Mohali*
Dr. Puneet Mishra, *URSC Bengaluru*

Organizing Secretary:
Dr. Padmavati, *PEC Chandigarh*
Dr. Simranjit Singh, *PEC Chandigarh*
Dr. Manohar Singh, *PEC Chandigarh*

TPC Chair:
Prof. Sudeb Das Gupta, *IIT Roorkee*
Prof. N. S. Chaudhari, *IIT Indore*
Prof. G. Bhuvaneswari, *Mahindra University*

Finance Chair:
Dr. Vijayalata Yellasiri, *Treasurer, IEEE*
Dr. Sneha Kabra, *Delhi University*
Mr. Mayank Gupta, *PEC Chandigarh*

Publication Chair:
Prof. Jawar Singh, *IIT Patna*
Prof. Jagdish Kumar, *PEC Chandigarh*
Prof. Balwinder Raj, *NIT Jalandhar*

Executive Steering Committee:
Sh. Deepak Mathur, *2024 IEEE VP MGA*
Prof. Preeti Bajaj, *VC-SA IEEE India Council*
Dr. K. R. Suresh Nair, *IEEE India Council*
Prof. M. N. Hooda, VC *IEEE Delhi Section*
**other committees are mentioned on the conference website.*

About the Conference

INDISCON is the flagship International Conference organised by IEEE India Council and IEEE Subsections in India to bring together researchers from academia and industries on various aspects of Sciences, Engineering and Technology. The Conference provides an excellent international platform for sharing of state-of-the-art research/technologies in the field of Electronics, Electrical, Information Technology etc., wherein many national/international eminent personalities will share their vision, expertise and knowledge.

INDISCON 2024 is organised by IEEE Chandigarh Subsection and hosted by Punjab Engineering College (Deemed to be University), Chandigarh along with IEEE India Council. INDISCON 2024 will include a wide range of technical sessions, invited talks, workshops, tutorials, special sessions, industry sessions, exhibits etc.

Technical Tracks

Track 1: Power and Energy Systems
Track 2: Power Electronics, Drives and Intelligent Control
Track 3: Instrumentation, Control and Signal processing
Track 4: Artificial Intelligence and Data Science
Track 5: Communication, Networks & IOT
Track 6: Next Generation Computing and applications
Track 7: Security & Privacy
Track 8: RF/Microwave/Terahertz Technologies
Track 9: Semiconductor Devices
Track 10: VLSI & Embedded Systems
Track 11 : Nanotechnology Materials and Devices
Track 12 : Education Technologies
Track 13 : Women in Engineering

Submission Guidelines

Paper submission instructions and template will be available at http://ieeeindiscon.org/
Paper submission link: https://cmt3.research.microsoft.com/INDISCON2024/

Papers (upto 6 pages in .pdf) presented in the Conference, duly accepted after peer review, will be eligible for submission for further consideration of publishing in the IEEE Xplore, subject to maintenance of quality, and post-conference scrutiny of response of Conference Organizers to Technical Program of Questionnaire (TPQ) on the Conference.

Note: Based on the significance of the work, novelty and technical contents, papers will be selected for the Best Poster Award and Best Paper Award. Travel grant will be awarded to a limited number of applicants on a highly competitive basis, for more details visit conference website.

Important Dates

• Last Date of Paper Submission **[Extended]**	**April 15, 2024**
• Notification of Acceptance	May 15, 2024
• Camera Ready/Final Paper Submission	June 10, 2024
• Last Date of Registration	June 15, 2024

+91-7814171121 ieeeindiscon2024@gmail.com http://ieeeindiscon.org/

About us

INDISCON is a flagship annual international conference of the IEEE India Council organized by an IEEE Subsection in INDIA. INDISCON 2024 scheduled during **August 22-24, 2024**, is being organized by IEEE Chandigarh Subsection along with IEEE India Council. The conference will be hosted by **Punjab Engineering College (Deemed to be University), Chandigarh**. The conference aims to provide an interdisciplinary platform for the academicians, researchers, industry professionals and research scholars to exchange and share their knowledge, experience & research.
Proceedings of previous versions of the conference are available here

Previous Edition	Dates	Venue	Theme
IEEE INDISCON 2023	August 5-7, 2023	GSSS Institute of Engineering & Technology for Women, Mysuru	Computational Intelligence and Learning Systems
IEEE INDISCON 2022	July 15-17, 2022	KIIT Deemed to be University, Bhubaneswar	Impactful Innovations for Benefits of Society and Industry
IEEE INDISCON 2021	August 27-29, 2021	Visvesvaraya National Institute of Technology, Nagpur	Impactful innovations for the benefit of industry and society
IEEE INDISCON 2020	October 3-4, 2020	Gayatri Vidya Parishad College of Engineering, Visakhapatnam	Smart and Sustainable Systems - Decade Ahead

IEEE India Council

IEEE is the world's largest professional association dedicated to advancing technological innovation and excellence for the benefit of humanity. IEEE and its members inspire a global community through IEEE's highly cited publications, conferences, technology standards, and professional & educational activities. IEEE India Council is the umbrella organisation which coordinates IEEE activities in India. Its primary aim is to assist and coordinate the activities of local "Sections", in order to benefit mutually, and avoid duplication of effort and resources. IEEE India Council was established on May 20, 1976 and is one of the five councils in the Asia Pacific Region (Region #10 of IEEE).
Details

IEEE Chandigarh Subsection

IEEE Chandigarh Subsection is a technical society that was established on June 18, 2005, under IEEE Delhi Section of IEEE India Council. It provides a platform for the students to enhance their technical skills and professional growth. The subsection organizes various events and technical extravaganzas, such as Techadroit, which is an annual event organized by IEEE PEC Student Branch in association with IEEE Chandigarh Subsection for students. In 2020, the subsection organized the first-ever Chandigarh Subsection Congress with the participation of more than 1700 delegates.
Details

ticketdude.

INVOICE

Place of supply: Karnataka

NO: d9b03655-3cb2-466c-ba2f-4dc996a507d6

5TH IEEE INDIA COUNCIL INTERNATIONAL SUBSECTIONS CONFERENCE 1 (INDISCON 2024)

IEEE India Council, Bangalore & IEEE Chandigarh Sub-section, Chandigarh

---

Date Issued: 2024-06-29

Invoice to: **Abhishek Shukla**

DESCRIPTION	QUANTITY	AMOUNT
Event : IEEE INDISCON 2024 Early Bird Non-IEEE Student Members (Indian)	1	INR 6000.00
Discount		INR 0.00
Processing Fee		INR 300.00
Surcharge Fee		INR 0.00
GRAND TOTAL		**INR 6300.00**

Six Thousand, Three Hundred Rupees, Zero Paise

***All Currency in INR**

This is a system generated invoice and does not need any signature. Thank you

ANNEXURE-IV

DELHI TECHNOLOGICAL UNIVERSITY

(Formerly Delhi College of Engineering)
Shahbad Daulatpur, Main Bawana Road, Delhi-42

PLAGIARISM VERIFICATION

Title of the Thesis IOT Malware Detection Using Machine Learning Based Techniques Total Pages 24 Name of the Scholar Abhishek Shukla

Supervisor (s)

(1) Mr. Rahul Gupta

(2)_____

(3)_____

Department INFORMATION TECHNOLOGY, DELHI TECHNOLOGICAL UNIVERSITY

This is to report that the above thesis was scanned for similarity detection. Process and outcome is given below:

Software used: Turnitin Similarity Index: 16% , Total Word Count: 5338

Date: 31/05/2024

Candidate's Signature **Signature of Supervisor(s)**

PAPER NAME
final.pdf

AUTHOR
Abhishek Shukla

WORD COUNT
5338 Words

CHARACTER COUNT
28827 Characters

PAGE COUNT
24 Pages

FILE SIZE
683.7KB

SUBMISSION DATE
May 31, 2024 9:39 AM GMT+5:30

REPORT DATE
May 31, 2024 9:40 AM GMT+5:30

16% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

- 6% Internet database
- 5% Publications database
- Crossref database
- Crossref Posted Content database
- 9% Submitted Works database

Excluded from Similarity Report

- Bibliographic material
- Small Matches (Less then 8 words)

PAPER NAME
final.pdf

AUTHOR
Abhishek Shukla

WORD COUNT
5338 Words

CHARACTER COUNT
28827 Characters

PAGE COUNT
24 Pages

FILE SIZE
683.7KB

SUBMISSION DATE
May 31, 2024 9:39 AM GMT+5:30

REPORT DATE
May 31, 2024 9:40 AM GMT+5:30

16% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

- 6% Internet database
- 5% Publications database
- Crossref database
- Crossref Posted Content database
- 9% Submitted Works database

Excluded from Similarity Report

- Bibliographic material
- Small Matches (Less then 8 words)

16% Overall Similarity

Top sources found in the following databases:

- 6% Internet database
- Crossref database
- 9% Submitted Works database
- 5% Publications database
- Crossref Posted Content database

TOP SOURCES

The sources with the highest number of matches within the submission. Overlapping sources will not be displayed.

1	Abhishek Meena, Deepanshu Nigam, Deepesh Sharma, Anamika Chau... Crossref	2%
2	Aston University on 2024-03-24 Submitted works	2%
3	Asia Pacific University College of Technology and Innovation (UCTI) on... Submitted works	<1%
4	Veermata Jijabai Technological Institute on 2023-11-08 Submitted works	<1%
5	ebin.pub Internet	<1%
6	researchgate.net Internet	<1%
7	Rochester Institute of Technology on 2023-12-19 Submitted works	<1%
8	assets.researchsquare.com Internet	<1%

9	Sydney Institute of Higher Education on 2024-05-16 Submitted works	<1%
10	arxiv.org Internet	<1%
11	Middle East Technical University on 2023-09-25 Submitted works	<1%
12	scilit.net Internet	<1%
13	Sheffield Hallam University on 2022-09-07 Submitted works	<1%
14	University of Hertfordshire on 2024-01-11 Submitted works	<1%
15	s2.smu.edu Internet	<1%
16	Pentecost University College on 2022-04-23 Submitted works	<1%
17	Sulyman Age Abdulkareem, Chuan Heng Foh, Francois Carrez, Klaus M... Crossref	<1%
18	Yasir Glani, Luo Ping, Syed Asad Shah. "AASH: A Lightweight and Effici... Crossref	<1%
19	researchid.co Internet	<1%
20	Siyabonga Matchaba, Rafik Fellague-Chebra, Purushottam Purushotta... Crossref	<1%

21	repository.msa.edu.eg Internet	<1%
22	sersc.org Internet	<1%
23	turcomat.org Internet	<1%
24	ijisae.org Internet	<1%
25	Coventry University on 2021-08-27 Submitted works	<1%
26	Laith Farhan, Sinan T. Shukur, Ali E. Alissa, Mohmad Alrweg, Umar Raz... Crossref	<1%
27	Leevy, Joffrey. "Machine Learning Algorithms for Predicting Botnet Att... Publication	<1%
28	Liverpool John Moores University on 2020-05-29 Submitted works	<1%
29	K. Radha, G. Sivagamidevi, N. Juliet, S. Niranjana, Nimmalaharathi Nim... Crossref	<1%
30	Mukul Kumar, Mridul Yadav, Anamika Chauhan. "Outlier Analysis Base... Crossref	<1%
31	National College of Ireland on 2023-04-23 Submitted works	<1%
32	University College London on 2024-04-21 Submitted works	<1%

33	University of Surrey on 2023-09-19 Submitted works	<1%
34	University of Warwick on 2018-04-26 Submitted works	<1%
35	academic-accelerator.com Internet	<1%
36	mospace.umsystem.edu Internet	<1%
37	onlineresource.ucsy.edu.mm Internet	<1%
38	iariajournals.org Internet	<1%
39	opastpublishers.com Internet	<1%
40	tnsroindia.org.in Internet	<1%