Covenant Journal of Informatics & Communication Technology. Vol. 12, No. 2, December, 2024
ISSN: *print* 2354 – 3566 *electronics* 2354 – 3507 DOI:

An Open Access Journal Available Online

Performance Analysis of selected Machine Learning Algorithms in the prediction of Man in the Middle in Internet of Things Environment

Stephen A. Mogaji, Olaiya Folorunsho, Yetunde Daramola, Timothy T. Odufuwa

Department of Computer Sciences, Federal University, Oye Ekiti, Nigeria.
stephen.mogaji@fuoye.edu.ng, olaiya.folorunsho@fuoye.edu.ng,
comfort.daramola@fuoye.edu.ng, tolulope.odufuwa@fuoye.edu.ng

Received: xx.xx.xxxx **Accepted: xx.xx.xxxx**

Publication: December 2024

Abstract— Concerns have been expressed over Internet of Things (IoT) devices' growing prevalence and susceptibility to cyberattacks, namely Man-in-the-Middle (MitM) assaults. The performance of selected machine learning algorithms: Logistic Regression, Decision Trees, and K-Nearest Neighbors were analyzed and compared using accuracy, precision, recall, F1-score, and error rate using a dataset comprising normal and attacked data sets from Kaggle. According to the research findings, the Decision Tree algorithm outperformed other selected algorithms in terms of MitM attack prediction accuracy of 99.42% and a good balance between precision, F1-score, and recall, with the lowest error rate of 0. 0058. The results of the study improve the security and reliability of IoT applications by aiding in the creation of efficient MitM attack prediction systems for IoT environments. The findings also emphasize how crucial it is to choose the best machine-learning algorithm for a given IoT security task. Investigating the use of transfer other techniques in MitM attack detection for IoT contexts is one area of future research.

Keywords/Index Terms— Prediction, MiTM, Decision Tree, Logistic regression, KNN

1. Introduction

The Internet of Things (IoT) permits seamless communications, from smart household appliances to industrial machines enabling them to communicate and interact with each other over the internet. According to (Alexander, 2024), the Internet of Things (IoT) is a system of intelligent things that includes sensors, actuators, programmable central microcontrollers, and other processors that may be wirelessly connected to routers and gateways. The core of IoT networks are sensors, which are in charge of monitoring and collecting data on changes in the physical environment, such as temperature, motion, humidity, and pressure, and convert the detected changes into a format that can be understood and processed by the microcontroller or processor, transmitting the data in either digital or analog signals, enabling the IoT network to make decisions and operate efficiently. The microcontrollers or processors utilized in both embedded systems and various applications and IoT networks are designed to be efficient and are often battery-powered, which also form the core of the network, are typically battery-powered and resource-constrained in terms of power consumption, Random Access Memory (RAM), and Read-Only Memory (ROM). These resource constraints require efficient processing, memory management, and power optimization to ensure reliable performance and extended battery life in IoT devices (Aeris, 2024). This network is made up of various connected "things" including devices like networked household devices, portable technology, and networked vehicles, all of which communicate with each other and with centralized systems to automate and streamline processes.

When a felonious individual places himself in the center of a discussion between a user and an application, either to eavesdrop or to pretend to be one of the parties, creating the impression that a legitimate information exchange is taking place, this is known as a man-in-the-middle (MiTM)

In the situation of a Man-in-the-Middle attack, the message communicated between two devices is forwarded through a rogue harmful gadget controlled by the attacker. In the field of cybersecurity, detecting and preventing MiTM attacks has become a top priority. As these attacks continue to jeopardize the confidentiality, integrity, and availability of sensitive information, the need for a dependable detection system has grown. To identify MiTM attacks, an Intrusion Detection System (IDS) is been developed and implemented. An intrusion detection system (IDS) is a type of security system that monitors network traffic to identify and stop hostile or unauthorized activity. The primary purpose of an IDS is to monitor network traffic and system activities to identify potential threats or intrusions. (Zeeshan, 2020).

A common attack on IoT networks is the Man-in-the-Middle (MitM) attack. However, traditional preventive measure approaches such as encryption, to tackle MitM attacks which are effective in IT networks, are not ideal for IoT devices resulting from their limited resources and battery-powered operation. Although encryption can enhance security, it also drains battery life and increases computational performance overhead on microcontrollers (MCUs). This results in the direct application of IT network security measures to IoT systems is often not feasible. To overcome this challenge,

machine learning models have been created to detect and recognize threats based on data patterns and sensor records within IoT networks (Xiao et al., 2018).

Machine learning is a subdivision of artificial intelligence that emphasizes data analysis and pattern recognition utilizing computer algorithms, statistical analysis, and computational analysis to mimic human learning and ultimately improve accuracy (IBM, 2020). Machine learning enables computers to study and develop automatically based on experience without the need for explicit programming. The learning phase begins with data observation, followed by analysis, pattern discovery, and prediction based on the information gained from data training.

A subset of machine learning known as supervised machine learning involves using a labeled training dataset to build a model that an algorithm subsequently uses to generate predictions. As a result, in general, the goal is to learn a mapping between input data and output labels presence, so the algorithm can be developed and is capable of making correct predictions.

According to a report (Statista, 2022), the IoT landscape will grow by 2025, it is projected that the number of devices linked to the internet will expand to 30.9 billion, highlighting a rapid adoption and combination of IoT technology across different sectors and applications. The security of IoT networks remains a neglected concern, even as they become increasingly prevalent. The resource-constrained nature of the microcontrollers and processors that power these networks, due to their battery-powered design and limited task functionality, poses significant challenges to implementing effective security solutions, thereby leaving IoT networks vulnerable to exploitation. Due to the resource constraints and battery-powered nature of IoT devices, minimizing power consumption and prolonging battery life are top priorities in IoT application design. Consequently, IoT chip manufacturers focus on developing smaller, faster, and more power-efficient chips that consume minimal current, thereby extending battery life and reducing power costs. According to (Zhang et al, 2018), IoT devices and networks are vulnerable to a range of attacks, including jamming, spoofing, and exploitation of vulnerabilities, which may pose a threat to the security of the network and allow unauthorized access. These attacks pose significant risks to the confidentiality and integrity of IoT systems, highlighting the importance of implementing strong security processes to protect against such threats. A common attack on IoT networks is the Man-in-the-Middle (MitM) attack. However, traditional preventive measure approaches such as encryption, to tackle MitM attacks which are effective in IT networks, are not ideal for IoT devices resulting from their limited resources and battery-powered operation. Although encryption can enhance security, it also drains battery life and increases computational performance overhead on microcontrollers (MCUs). This results in the direct application of IT network security measures to IoT systems is often not feasible. To overcome this challenge, machine learning models have been created to detect and recognize threats based on data patterns and sensor records within IoT networks (Xiao et al., 2018). Although the authors (Kiran et al, 2020) used an IoT testbed to simulate MitM attacks and used the data to train different machine learning algorithms to detect the IoT network data behavior based on cyber-attacks. However,

their study only compared ML algorithm performance and identified the best one for detecting attacked sensor records in the network. However, (Kiran et al, 2020) took a limited approach, using only a single sensor and a small-scale dataset comprising of 480 records for training, testing, and evaluation of multiple machine learning frameworks.

To achieve highly accurate predictions and robust ML models, use an extensive training dataset, a high-quality dataset with clear and diverse data. As a general guideline, the quantity and quality of training data directly impact the performance of ML models, with more comprehensive datasets yielding better results. Hence, the research questions that guides this study are:

i Can the machine learning approach effectively detect MiTM attacks in IoT networks?

ii. What is the best machine learning technique for identifying MiTM attacks in IoT systems using network data patterns?

Designing a machine learning-based MiTM attack prediction model in IoT networks lies in the urgent need to discuss the security concerns posed by the extensive utilization of IoT devices and the increasing sophistication of cyber threats. By utilizing machine learning technologies, researchers aim to develop proactive, adaptive, and effective security solutions that maintain the trustworthiness of IoT networks in the face of growing threats.

This research aimed at the careful selection of a suitable supervised machine learning approach for the prediction of Man-in-the-Middle (MiTM)attacks in IoT networks using a comprehensive dataset.

The specific objectives of this study include:

i. To design an MITM attack prediction system
ii. To determine the optimal approach for identifying and preventing MiTM attacks in real-time in order to cut down the swift increase in MiTM attacks to a minimal level
iii. To safeguard the concern and confidentiality of users while navigating around the Internet of Things (IoT) Environment.

2. Methodology

A selected Machine learning algorithms were used to carry out this research. The phases taken to implement this are:

1. Dataset collection: The dataset, obtained from the Kaggle dataset Network (https://www.kaggle.com/datasets/sampadab17/ using a dataset with 25,192 records that included both "Normal" and "Attacked" data.
2. Data Preprocess: This involved an important phase in preparing the text data for machine learning, ensuring that errors were reduced and results were accurate.
3. Feature Selection: Recursive Feature Elimination (RFE) was employed to optimize model complexity and enhance interpretability by identifying and selecting the most informative features.
4. Data Splitting: The dataset will be partitioned into training and testing sets, with a 70 to 30 split ratio. This indicates that the algorithm will get trained on 70% of the data, and 30%, set aside for evaluation of performance.
5. Model Selection and Development:

Dataset were trained and tested using some selected machine learning algorithms (Logistic Regression, Decision Trees, and K-Nearest Neighbor (KNN).

6. Model Performance analysis: The performance of the model was determined by making use of the listed metrics: True Positive, True Negative, False Positive, and False Negative. The evaluation was performed using Confusion Matrix, Precision, Recall, F1 score, and Error rate.
7. Comparative Performance Analysis: The models were compared based on their performances in Predicting MiTM attacks in IoT networks.

2.1 Dataset and Dataset Collection

This stage is the first stage in the execution of the solution. It involves gathering several datasets on Man-in-the-Middle (MiTM) attacks using a dataset that includes both "Normal" and "Attacked" data. The dataset, obtained from the Kaggle dataset Network /(https://www.kaggle.com/datasets/sampadab17/ network-intrusion-detection, was downloaded in .csv format and consisted of 25,192 records. The data was divided into a training dataset of 17,634 records and a testing dataset of 7,558 records. 'Attacked' records and 'Normal' records (non-attack data) are used to train and evaluate the potential machine learning models.

2.2 Data Preprocessing

This involved a vital step in preparing the text data for machine learning, ensuring that errors were reduced and results were accurate. To achieve this, the datasets underwent several key processes. To assess and validate the model, the data was distributed into training and testing sets. Next, a check for null values was conducted to prevent errors and ensure the model could process all available data. Missing values can significantly impact model accuracy, so this step was crucial. Categorical features were then handled using Label Encoding, converting them into numerical labels to enable machine parsing. More precisely, attributes like 'normality' in the second dataset and 'flag, service, count, and protocol type' in the first dataset, were encoded into numerical labels to facilitate machine learning processing.

2.3 Feature Selection

This is the process of isolating the most consistent, non-redundant, and relevant features to use in model construction.

In this research, the Recursive Feature Elimination (RFE) combined with a Random Forest classifier was used for feature selection, which involves recursively removing features and building models to determine the key features that drive the outcome or results of the algorithm. Recursive Feature Elimination (RFE) was employed to optimize model complexity and enhance interpretability by identifying and selecting the most informative features. RFE was utilized to identify the most important characteristics for outcome prediction in the MiTM dataset, thereby eliminating redundant or non-essential features and retaining only those that significantly contribute to the algorithm's performance.

2.4 Data Splitting

Splitting the data is an important processing machine learning for evaluating model

performance and generalizability. The dataset was partitioned into training and testing set, with a 70 to 30 split ratio. This indicates that the algorithm was trained on 70% of the data, and 30%, for testing and also performance analysis.

2.5 Model Selection and Model Development

The selection of the three algorithms was based on their popularity as well as obvious conflicting findings from earlier studies.

The type of problem and the properties of the dataset are taken into consideration when evaluating different machine-learning techniques. This comprises techniques like Logistic Regression, Decision Trees, and K-Nearest Neighbor (KNN), each algorithm offers unique strengths and weaknesses, making it essential to experiment with multiple approaches to determine the most suitable one for the task.

2.5.1 Logistic Regression

A sigmoid function, which transfers any real-valued collection of input autonomous variables into a value between 0 and 1, is used by the logistic regression model to convert the continuous value output of the linear regression function into a definite value output. This technique is straightforward but powerful, also known for its interpretability and effectiveness. Its simplicity and efficiency make it an appropriate standard model for evaluation throughout the model selection process, generating information about the efficiency of more complicated algorithms.

2.5.2 Decision Trees

Decision trees are an algorithm that classifies or regresses data based on a series of true or false answers to specific questions. When visualized, consists of a hierarchical arrangement of nodes, the resulting structure resembles a tree, with various types of nodes including the root, internal nodes, and leaf nodes. The root node represents the initial decision point, internal nodes represent subsequent questions or decisions, and leaf nodes indicate the final classification or regression outcomes. Decision trees are popular because of their capacity to capture non-linear correlations in data, making them ideal for detection classification tasks with complicated patterns. Decision trees are interpretable, giving a better knowledge of the decision-making process, which can be useful when assessing clues in data. However, decision tree models are prone to overfitting, particularly with high-dimensional data, and might require scaling or other normalization approaches to maximize their generalization efficiency. Notwithstanding these drawbacks, Decision Trees remain a reasonable choice for emotion identification models due to their ease of use and interpretability.

2.5.3 K-Nearest Neighbor (KNN)

KNN is a model that uses the points that are most similar to it to classify data points. It makes an "educated guess" on the proper classification for an unclassified point based on test results. It is predicated on the notion that comparable data points typically have comparable labels or values.

It has the ability to capture non-linear relationships between features and attack detection allowing it to identify complex patterns in network traffic. The KNN provides insight into patterns and

relationships between data points, allowing for a better understanding of network traffic behaviors.

2.6: Model Performance and Evaluation

The model outcomes were retrieved for each of the used machine learning techniques.

Assessing the effectiveness of machine learning models is an important step in ensuring their efficacy and trustworthiness. The following metrics and methods are utilized to assess a model's ability to predict the outcome of unobserved data.

1. **Confusion Matrix:** The accuracy of a classification model is evaluated using a table-based evaluation method called a confusion matrix. A comparison of the real and expected classifications, enables for more thorough examination of the model's effectiveness.

   TP is true positives, measures correctly identified cases as positive.

   TN is true negatives; measures correctly identified cases as negative.

   FP is false positives, measures incorrectly identified cases as positive.

   FN is false negatives; incorrectly identified the case as negative.

2 **Precision**: The ratio of accurately predicted positives to the total number of positive predictions indicates the prediction accuracy of a classifier. It is one of the model's primary parameters that is used to evaluate the model's accuracy in classifying positive values.

$$Precision = \frac{True\ \ Positive}{True\ Positive + False\ Positive} \quad \ldots (1)$$

3. **Recall:** This shows how many positive samples were correctly classified as positive compared to how many positive samples were incorrectly classified. Recall is a metric used to assess a model's ability to identify positive samples. Higher recall rates result in a higher detection rate of positive samples.

$$Recall = \frac{True\ \ Positive}{True\ Positive + False\ Negetive} \quad \ldots (2)$$

4. **Accuracy**: While accuracy is a simple statistic, it may not be sufficient for datasets that are not balanced. It is defined as the fraction of correctly predicted occurrences to all instances in the dataset.

$$Accuracy = \frac{True\ \ Positive + True\ Negetive}{Total} \quad \ldots (3)$$

5. **F1 Score:** It uses a weighted average to describe recall and precision. As F1 has a range of 0 to 1, 1 represents the most favorable value and 0 is the least favorable. This component determines the classifier's resilience and precision.

$$F1\ Score = \frac{2*(Precision*Recall)}{Precision+Recall} \quad (4)$$

Error Rate

it refers to the proportion of incorrect predictions or classifications made by a

machine learning model. A lower error rate indicates better model performance.

$$Error\ Rate = \frac{False\ Positive + False\ Negetive}{Total} \quad (5)$$

3.0: System Design

The development environment for the model included Python 3.8 as the programming language. Jupyter Notebook was used for interactive coding and Jupyter Lab provided an enhanced interface, and Visual Studio Code (VS Code) was used for advanced code editing. This setup ensured efficient development, debugging, and visualization for the model.

Several Python libraries such as Pyplotlib, SciKit-Learn, Matplotlib, Pandas, NumPy, and software extensions were used to facilitate the various stages of data processing, model building, and evaluation. The choice of libraries was driven by their compatibility with the tasks at hand, their community support, and their effectiveness in handling large datasets and complex computations.

4. Model Performance Analysis in the Prediction of MiTM Attack using selected Machine Learning Algorithms

The performance analysis of the model was carried out using the selected machine learning algorithms (Logistic Regression, Decision Tree, and K-Nearest Neighbor Forest).

The confusion matrix of the three selected algorithms with the greatest Performance using the 7,558 dataset were examined. The algorithms' Performance evaluation in Predicting MiTM attacks in IoT networks is analysed here. To get a detailed understanding of each algorithm's performance, recall, accuracy, FI Score, precision, and error rate were examined.

4.1 Logistic Regression

The confusion matrix for Logistic Regression was able to accurately predicted MiTM 3228 attacks (True positive), correctly classified just 3819 as non-MiTM attacks (True negatives), wrongly predicted 241 MiTM attacks as correct (False positive), and wrongly classified 270 MiTM attacks (False negatives).

The results show that the Logistic Regression algorithm has a True Positive prediction of yielded an accuracy and precision of 0.9324 and 0.9305 respectively with an error rate of 0.0676 as shown in Figure 1.

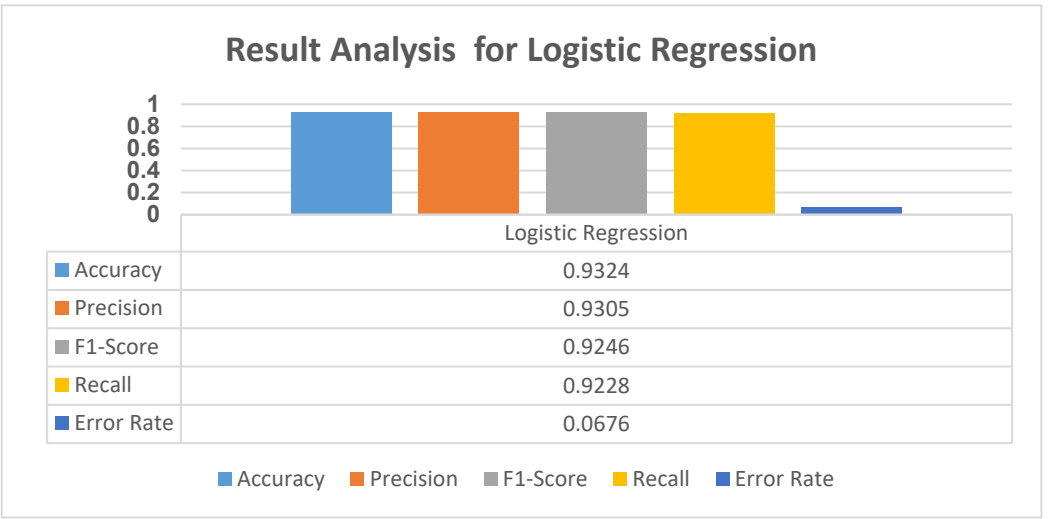


Figure 1: Performance Analysis of Logistic Regression

4.2 Decision Tree

The confusion matrix for Decision Tree accurately predicted MiTM 3479 attacks (True positive), correctly classified just 4035 as non-MiTM attacks (True negatives), wrongly predicted 25 MiTM attacks as correct (False positive), and wrongly classified 19 MiTM attacks (False negatives)

The decision Tree algorithm had an accuracy and precision of 0.9942 and 0.9927 respectively with an error rate of 0.0058 as shown in Figure 2.

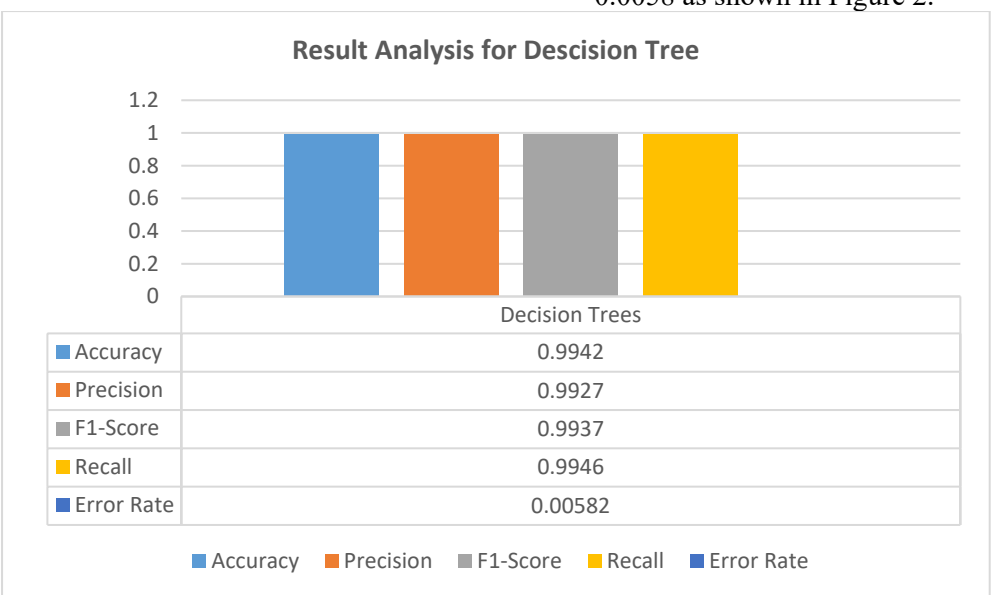


Figure 2: Performance Analysis of Decision Tree

4.3 k-nearest neighbors (KNN)

The confusion matrix for k-nearest neighbors (KNN) was able to correctly predicted MiTM 3479 attacks (True positive), correctly classified just 4035 as non-MiTM attacks (True negatives), wrongly predicted 25 MiTM attacks as correct (False positive), and wrongly classified 19 MiTM attacks (False negatives)

The k-nearest neighbors (KNN) algorithm had an accuracy and precision of 0.9845 and 0.9812 respectively with an error rate of 0.0155 as shown Figure 2.

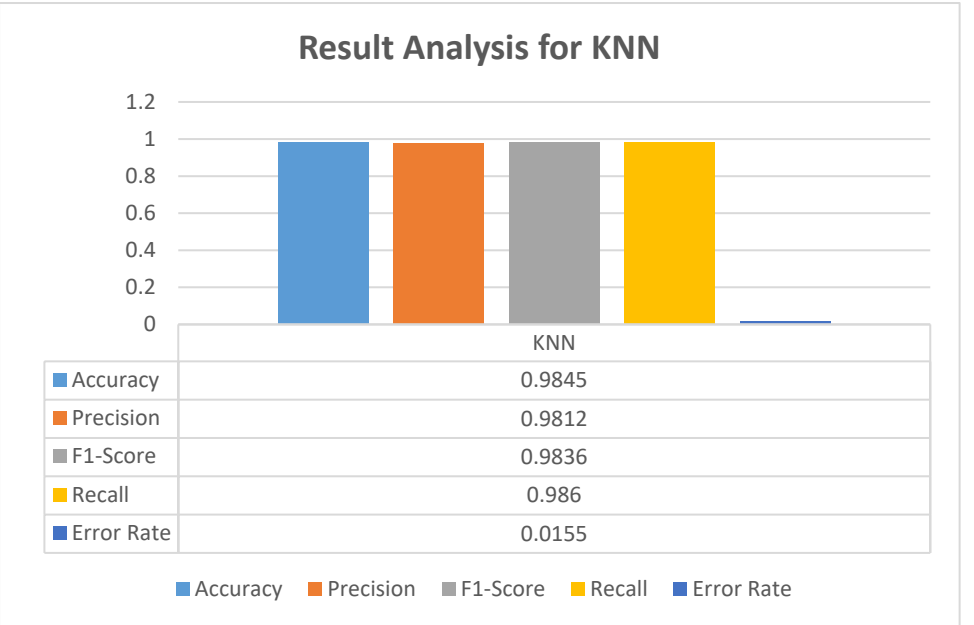


Figure 3: Performance analysis of KNN

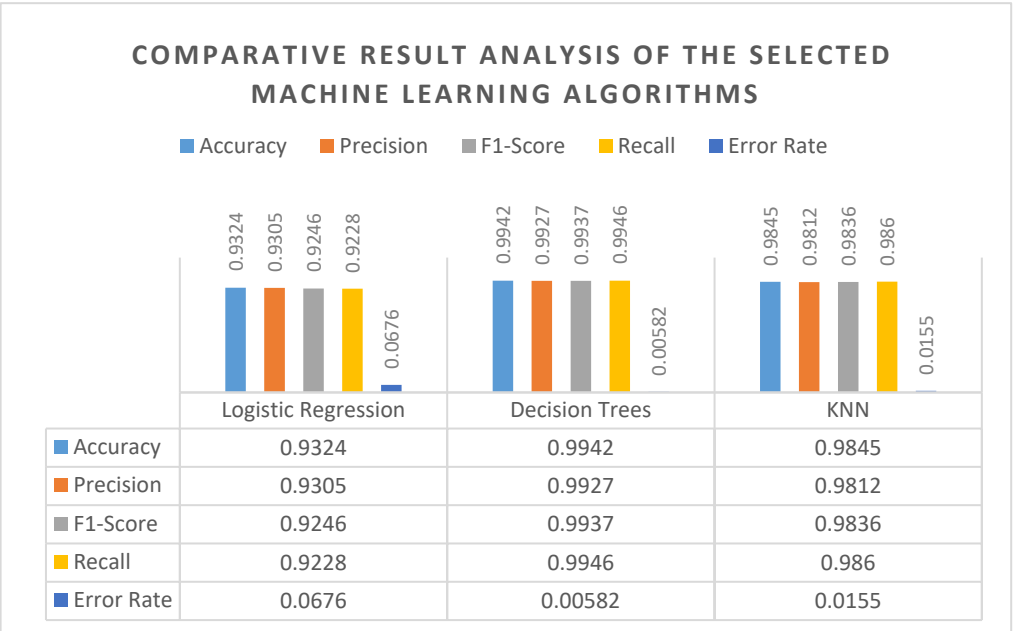
5. Comparative Performance analysis of the three selected Machine Learning Algorithms

This section shows a Comparative performance analysis of the system using the three selected Machine learning algorithms in the prediction of a Man in the Middle (MiTM) attack in the Internet of Things (IoT) network environment. Figure 4 gives a pictorial comparative analysis and Performance Evaluation for the three selected Machine Learning Algorithms.

The Logistic Regression while still competent, showed lower performance compared to the other two algorithms, and yielded an accuracy of 93.24% it correctly classified fewer instances than the Decision Tree and KNN. It also has 99.42% and 98.45% in both precision and recall, respectively suggesting good performance but with some trade-offs in identifying all positives compared to KNN and Decision Trees.

The F1-score of 92.46% indicates a balanced performance but also highlights that Logistic Regression is not as effective as the other models in this case.

The K-Nearest Neighbors (KNN) model achieved an impressive 98.45% accuracy, demonstrating great overall performance. Its recall rate of 98.45% is particularly noteworthy, indicating exceptional ability to identify true positive cases. While precision and F1-score also showed strong results at 98.36%, the model fell slightly short of the Decision Tree's performance in balancing the identification of true positives with the minimization of false positives. With an error rate of 1.53%, the KNN model occasionally misclassified instances.



	Logistic Regression	Decision Trees	KNN
Accuracy	0.9324	0.9942	0.9845
Precision	0.9305	0.9927	0.9812
F1-Score	0.9246	0.9937	0.9836
Recall	0.9228	0.9946	0.986
Error Rate	0.0676	0.00582	0.0155

Figure 4: Comparative Performance analysis of the three selected machine Learning Algorithms

The decision tree algorithm is clearly the best performer in this case, as illustrated in Figure 4, with a high accuracy rate of 99.42% and the lowest error rate of 0.58%. Both precision and recall are well-balanced, with 99.27% and 99.47%, respectively, and the algorithm's minimal error rate of 0.58% emphasizes its overall reliability and low misclassification.

In view of the Comparative analysis shown above, Decision Trees provided the highest performance across all metrics, other machine learning algorithms made a good impact on the prediction. These findings demonstrate that the model selection can have a substantial impact on the Network attack's ability to detect and predict MitM attacks accurately and that combining several algorithms can frequently produce better results.

Each of the selected algorithms has its strengths, but Decision Trees clearly excelled in this evaluation.

6. Related Work

Vlajic & Zhou's (2018) study demonstrated experiments on Internet of Things-based cameras about the execution of DDoS attacks on the webcams and proposed ways to stop the DDoS assaults. Physical, network, and application layers were the levels of attack origin that were used to categorize the attacks in their study.

The authors Kiran et., al. (2020) used an IoT testbed to simulate MitM attacks and used the data to train different machine learning algorithms to detect the IoT network data behavior based on cyber-attacks. However, their study only compared ML algorithm performance and identified the best one for detecting attacked sensor records in the network.

However, Kiran *et. al.,* (2020) took a limited approach, using only a single sensor and a small-scale dataset comprising of 480 records for training, testing, and evaluation of multiple machine learning frameworks.

In order to detect botnet assaults on IoT devices, Mohammad *et., al.* (2020) developed an optimized machine learning (ML)-based framework that integrated a decision tree classification model with the Bayesian Optimization Gaussian Process (BO-GP). The study's primary goal was to develop a dynamic and effective framework for IoT devices to identify botnet attacks.

Jones & Kumar (2019) discovered the MiTM attack using a deep learning group of instructions with the network simulator 2 (NS2) simulation platform, which is known as synthetic artificial neural networks (ANN). For a few attacks, they employed a dataset that included the mobility patterns and network-number-of-site visitors' conditions. To analyze the ANN model, they used four evaluation metrics: recall, f1-score, accuracy, and precision. They found that the accuracy fee of the ANN was 88.235%.

A prediction version-based totally-system mastering technique was presented by Benter & Kuhlang (2021) to identify MiTM from business management structures. They gathered real-time MiTM data, which includes a variety of functions such as temp max, cntt avg, cntt stdev, temp stdev, temp min, and temp avg. The gadget is based on the KNN version. They have demonstrated that the model based on k-nearest neighbor provided excellent performance for MiTM attack detection.

Hammad *et al.* (2020) used a strategy that includes four distinct approaches, CFS feature selection, and k-means clustering to target this data set. Zero, J48, RF, and SVM. The performance of the majority of classifiers has been successfully improved by the suggested method. J48 has the highest reported accuracy, with 96.7%, and 10-fold cross-validation is used. This paper is aimed at designing a Machine Learning Model to predict Man in the middle (MitM) attacks in IoT environment and implemented using supervised machine learning algorithm. This is justified due to the rapid increase associated with IOT devices, there has been a crucial security risk, especially the threat of MiTM attacks, which compromise the confidentiality and integrity of data. By creating a strong model, it hopes to improve applications like smart home security, mobile devices, and autonomous vehicle security, promoting better security, integrity, and reliability of these systems. The results of this investigation are useful for researchers to understand the threat of MiTM attacks in

IoT networks. The technological advances have the potential to advance the overall security of IoT networks and devices and also protect against MiTM and other cyber threats.

This research also carefully selected a suitable supervised machine learning method for detecting Man-in-the-Middle (MiTM)attacks in IoT networks using a comprehensive dataset, by evaluating the performance of various algorithms and determining the optimal approach for identifying and preventing MiTM attacks in real-time The authors Kiran *et al.* (2020) used an IoT testbed to simulate MitM attacks and used the data to train different machine learning algorithms to detect the IoT network data behavior based on cyber-attacks. However, their study only compared ML algorithm performance and identified the best one for detecting attacked sensor records in the network. However, Kiran et., al. (2020) took a limited approach, using only a single sensor and a small-scale dataset comprising of 480 records for training, testing, and evaluation of multiple machine learning frameworks.

Vlajic & Zhou's (2018) study demonstrated experiments on Internet of Things-based cameras about the execution of DDoS attacks on the webcams and proposed ways to stop the DDoS assaults. Physical, network, and application layers were the levels of attack origin that were used to categorize the attacks in their study.

In order to detect botnet assaults on IoT devices, Mohammad et., al. (2020) developed an optimized machine learning (ML)-based framework that integrated a decision tree classification model with the Bayesian Optimization Gaussian Process (BO-GP). The study's primary goal was to develop a dynamic and effective framework for IoT devices to identify botnet attacks.

The authors of Sudhanshu & Bichitrananda (2023) concentrated on utilizing machine learning (ML) techniques to analyze intrusion detection systems (IDSs). Network attacks can be accurately and efficiently detected by IDSs that use machine learning techniques. Nevertheless, the effectiveness of these systems deteriorates in data with huge dimensional spaces. Accordingly, it is crucial to implement a workable feature reduction method that can exclude traits that don't significantly impact the categorization process. The researcher analyzed the KDD CUP-'99' intrusion detection dataset used for training and authenticating ML models. Then implemented ML classifiers such as "Logistic Regression, Decision Tree, KNearest Neighbour, Naïve Bayes, Bernoulli Naïve Bayes, Multinomial Naïve Bayes, XG-Boost Classifier, AdaBoost, Random Forest, SVM, Rocchio classifier, Ridge, Passive-Aggressive classifier, ANN besides Perceptron (PPN), the optimal classifiers were determined by comparing the results of Stochastic Gradient Descent and backpropagation neural networks for IDS", Conventional classification indicators, such as "accuracy, precision, recall, and the f1-measure", were used to evaluate the performance of the ML classification algorithms.

Jones & Kumar (2019) discovered the MiTM attack using a deep learning procedures with the network simulator 2 (NS2) simulation platform, which is known as synthetic artificial neural networks (ANN). For a few attacks, they employed a dataset that included the mobility patterns and network-number-of-site visitors' conditions. To analyze the ANN model, they used four evaluation metrics: recall, f1-score, accuracy, and precision. They

found that the accuracy fee of the ANN was 88.235%.

A detection version-based totally-system mastering technique was presented by Benter & Kuhlang (2021) to identify MiTM from business management structures. They gathered real-time MiTM data, which includes a variety of functions such as temp max, cntt avg, cntt stdev, temp stdev, temp min, and temp avg. The gadget is based on the KNN version.

The authors of Hammad *et al.* (2020) used a strategy that includes four distinct approaches, CFS feature selection, and k-means clustering to target this data set. Zero, J48, RF, and SVM. The performance of the majority of classifiers has been successfully improved by the suggested method. J48 has the highest reported accuracy, with 96.7%, and 10-fold cross validation is used.This paper is aimed at designing a Machine Learning Model to detect Man in the middle (MitM) attacks in IoT environment and implemented using supervised machine learning algorithm. This is justified due to the rapid increase associated with IOT devices, there has been a crucial security risk, especially the threat of MiTM attacks, which compromise the confidentiality and integrity of data. By creating a strong model, it hopes to improve applications like smart home security, mobile devices, and autonomous vehicle security, promoting better security, integrity, and reliability of these systems. The results of this investigation are useful for researchers to understand the threat of MiTM attacks in IoT networks. The technological advances have the potential to advance the overall security of IoT networks and devices and also protection against MiTM and other cyber threats.

Several machine learning models were used by Su et al. (2021) to forecast network assaults on Internet of Things devices. The models that were put to the test were the gradient-boosting machine (GBM), decision tree, and random forest. According to the results, the random forest method had higher AUC scores, but the decision tree approach had the highest accuracy.

This research also carefully selected a suitable supervised machine learning method for recognizing Man-in-the-Middle (MiTM)attacks in IoT networks using a comprehensive dataset, by assessing the performance of various algorithms and determining the optimal approach for identifying and preventing MiTM attacks in real-time.

Some machine learning techniques, comprising logistic regression (LR), decision tree (DT), support vector machine (SVM), random forest (RF), and artificial neural network (ANN), were used by Hassan et al. (2019) to predict attacks and anomalies on IoT devices. The outcomes showed that RF, DT, and ANN outperformed the other techniques, attaining an accuracy of 99.4%.

According to Obonna *et al.* (2023), the incorporation of open network to operation technology (OT) as an upshot of low-cost network expansion may have been the primary cause of the amorphous cyberattacks that have affected the process control network (PCN) of oil and gas installations. These attacks include Distributed-Denial-of-Service (DDoS), Denial-of-Service (DoS), and Man-in-the-Middle (MitM) attacks. Process control was simulated using MATLAB, Allen Bradley RSLogic 5000 PLC Emulator software, Deep-Learning Toolkit, and Python 3.0 Libraries. With notable, accurate attack detections found utilizing a coarse tree approach, the trials' results validated the dependability and effectiveness of the

various machine learning methods in identifying these abnormalities.

Wang *et al*., (2021) reviewed the application of machine learning in anomaly detection under various network situations and introduced the difficulties of anomaly detection in both traditional and next-generation networks. The methods and benefits of each machine learning category are described, along with an explanation of the process. Additionally, a summary of the comparison of various machine learning models was provided.

Reddy *et al*., (2021) worked on prediction strategy for fog node structure protection that uses the XGBoost ensemble technique and Exact Greedy Boosting for hyperparameter fine-tuning. Based on network data, the suggested framework classifies normal and abnormal behavior. XGBoost improves trees using efficient split-finding technique to avoid overfitting and boost performance in comparison to traditional GB. Additionally, the technique uses a straightforward standard feature selection mechanism known as Variance Threshold, which eliminates features with low variance.

Ennaji *et al*. (2024) used the IoTID20 dataset to **study** the effectiveness of many machine learning techniques for creating IDS, such as support vector machines (SVM), k-Nearest Neighbor (k-NN), decision trees (DT), random forests (RF), and Ada Boost. The dataset contains three target classes: classes for **sets and subsets** of the binary class, as well as a binary class for normal or deviant behavior. In order to reduce execution time and increase accuracy, the writers chose the most pertinent elements. According to the overall results, the decision tree algorithm outperformed the other algorithms in terms of accuracy, achieving 99.80% with the lowest error rate.

7. Conclusion

The design of MiTM attack detection model for IoT networks using a machine learning approach demonstrates the potential for effectively identifying and mitigating security threats in connected environments. The project provided a comprehensive overview of IoT's impact on daily activities, highlighting its background, benefits, drawbacks, and future potential. The demonstration illustrates the profound impact of IoT on our daily lives, reshaping our daily experience, work practice and technological interactions. This machine learning-based detection model not only enhances the security of IoT networks by identifying potential MiTM attacks but also breaks new grounds on the approach that enables defense against MitM attacks, ensuring confidentiality and integrity of IoT devices offering a scalable and adaptable solution for future advancements in cybersecurity. In order to improve IoT security, we investigated MiTM Attack detection in the IoT network devices and the MiTM Attack dataset making use of several supervised ML algorithm approaches. Our results demonstrated exceptional performance in terms of accuracy and efficiency. Among the tested algorithms, the Decision Tree classifier stood out, achieving an exceptional accuracy rate of 99%, thereby addressing the research question “What is the best machine learning algorithm for identifying MiTM attacks in IoT systems using network data patterns?” This finding has significant

implications for enhancing the security of IoT networks.

The suggested approach for detecting MiTM attacks employs a decision tree algorithm and was evaluated using the MiTM Attack Detection dataset. The decision tree algorithm was compared against KNN and Logistic Regression algorithms. With an accuracy rate of 99% on the dataset, the decision tree algorithm emerged as the top-performing model, along with the fastest runtime of 412 milliseconds, one of its most notable advantages is its ability to reduce computing complexity.

In network security, MiTM attack detection systems are essential. They help organizations comply with security policy and meet compliance requirements while enhancing overall network visibility throughout the entire network and by learning and recognizing triggers and behavioral trends that enhance the detection of intrusions efficiently.

References

Aeris, (2024) “What is IoT? Defining the Internet of Things (IoT)” [online]. Available: https://www.aeris.com/in/what-is-iot/ Accessed on February 2024.

Alexander S. Gillis (2024). IoT Agenda, “What is the internet of things (IoT)?” Online]. Available:https://internetofthingsagenda.techtarget.com/definition/Internet- of Things-IoT Accessed on February 2024.

Benter, M. and Kuhlang,P. (2021) MTM-HWD – integration of ergonomic evaluation into production planning. ASU Arbeitsmedizin Sozialmedizin Umweltmedizin, vol. 2021, no. 11, pp. 703–706, doi: 10.17147/asu-2111-9792.

Ennaji, E., Elhajla, S., Maleh, Y., & Mounir, S. (2024) Machine Learning Algorithms for Intrusion Detection in IoT Prediction and Performance Analysis. International Symposium on Green Technologies and Applications (ISGTA’2023). Available online at www.sciencedirect.com. 10.1016/j.procs.2024.05.054. 461-467.

Hammad, M., El-medany, W., & Ismail, Y. (2020). Intrusion Detection System using Feature Selection with Clustering and Classification Machine Learning Algorithms on the UNSW-NB15 dataset. In 2020 *International Conference on Innovation and Intelligence for Informatics, Computing and Technologies* (3ICT) (pp. 1-6). IEEE.

Hasan, M., Islam, M. M., Zarif, M. I. I., & M. M. A. Hashem. M. A. (1019) Attack and anomaly detection in IoT sensors in IoT sites using machine learning approaches, Internet of Things, vol. 7, p. 100059, doi: 10.1016/j.iot.2019.100059.

IBM, (2020) “Machine Learning” [online]. Available: https://www.ibm.com/cloud/learn/machine-learning Accessed on February 2024.

Imperva (2024)"Man in the middle (MITM) attack "https://www.imperva.com/learn/ap

plication-security/man-in-the-middle-attack- mitm/ Accessed July 2024

Imperva, (2024), "What is the ARP Protocol" [online]. Available: https://www.imperva.com/learn/application-security/arp-spoofing/ Accessed on June 2024.

Jiang, L., Zhang, S., & Wu, J. (2021). Location-based services and smart home systems: A comprehensive review. ACM Computing Surveys, 54(3), 1-35. DOI: 10.1145/3448311

Jones, S. B. R. & Kumar,N. (2019) Unraveling the security pitfalls that stem from core cloud benefits through analyzing various DoS attacks, detection and prevention, *Journal of Advanced Research in Dynamical and Control Systems*, vol. 11, no. 9, pp. 541–553, doi: 10.5373/JARDCS/V11/20192603.

Liu, X., Wang, Y., & Chen, Y. (2022). Autonomous driving: A comprehensive review of sensor technologies and data fusion methods. Sensors, 22(1), 123. DOI: 10.3390/s220100123

M. Saed, M.& Aljuhani, A., (2022) Detection of Man in The Middle Attack using Machine learning," *2022 2nd International Conference on Computing and Information Technology (ICCIT),* Tabuk, Saudi Arabia, pp. 388-393, doi: 10.1109/ICCIT52419.2022.9711555

Micro.ai, (2023) "10 Types of Cyber Security Attacks in IoT" [online]. Available: https://www.micro.ai/blog/10-types-of-cyber-security-attacks-in-the-iot Accessed on May 2024.

Misra S., A (2021) Step-by-Step Guide for Choosing Project Topics and Writing Research Papers in ICT Related Disciplines, Communications in Computer and Information Science, Volume : 1350 Page : 727-744 the Publication year 2021 (pp. 727-744). Springer International Publishing.

Mohammad, N.I., Adballah, M., Abdallah, S (2020), "Detecting BotnetAttacks in IoT Environments: An Optimized Machine Learning Approach"

Obonna, U.O., Opara, F.K., Mbaocha, C.C., Jude-Kennedy, C.O., Akwukwegbu, I.O., & Amaefule, M.M. (2023) Detection of Man-in-the-Middle (MitM) Cyber-Attacks in Oil and Gas Process Control Networks using Machine Learning Algorithms. *Preprints.org.* 1-13. doi:10.20944/preprints202307.0747.v1.

Oracle, (2022), "What is IoT" [online]. Available: https://www.oracle.com/internet-of things/what-is-iot/ Accessed May 2024.

Reddy, D. K. K., Behera, H. S., Nayak, J., Naik, B., Ghosh, U., & Sharma, P. K. (2021) Exact greedy algorithm based split finding approach for intrusion detection in fog-enabled IoT environment, J. Inf. Secur. Appl., vol. 60, no. June, p. 102866, doi: 10.1016/j.jisa.2021.102866.

Ross J. Anderson, (2020), Security Engineering: A guide to building dependable distributed systems (3rd ed.). Hoboken, NJ: Wiley"

Sai, K. V. V. N. L., Kiran, R. N., Kamakshi D. N., Pavan Kalyan, K. &. Mukundini, R.K. (2020) Building an Intrusion Detection System for IoT Environment using machine Learning Techniques. Dept of Computer Science and Engineering, Amirita School of Engineering, Combatore, Amirita Vishwa Bidyapeetham, India

Statista. (2022). Internet of Things (IoT) and non-IoT active device connections worldwide from 2010-2015. Retrieved from https://www.statista.com/statistics/1101442/iot- number-of-connected-devices-worldwide/ Accessed February 2024.

Su, J., He, S., & Wu, Y. (2021) Features selection and prediction for IoT attacks, High-Confidence Comput., vol. 2, no. 2, p. 100047, doi: 10.1016/j.hcc.2021.100047.

Sudhanshu S.T., & Bichitrananda B. (2023) Performance Evaluation of Machine Learning Algorithms for Intrusion Detection System. *Journal of Biomechanical Science and Engineering* Japan Society of Mechanical Engineers ISSN: 1880-9863. 621-640. DOI 10.17605/OSF.IO/WX6CS

Ullah, I. & Mahmoud, Q. H. (2020) A Technique for Generating a Botnet Dataset for Anomalous Activity Detection in IoT Networks, vol. 2020- Octob, no. May. Springer International Publishing, doi: 10.1109/SMC42975.2020.9283220.

Vijay Kanade, (2022) What Is Machine Learning? Definition, Types, Applications, and Trends. https://www.spiceworks.com/tech/artificial-intelligence/articles/what-is-ml/ Accessed on June 2024

Vijay, (2022). Machine Learning Algorithms. https://www.spiceworks.com/tech/artificial-intelligence/articles/top-ml-algorithms/ Accessed on June 2024

Vlajic ,N., & Zhou, D.(2018) IoT as a Land of Opportunity for DDoS Hackers," in Computer, vol. 51, no. 7, pp. 26-34, doi: 10.1109/MC.2018.3011046.

Wang, S., Balarezo, J. F., Kandeepan, S., Al-Hourani, A., Chavez, K. G., & Rubinstein, B (2021) Machine Learning in Network Anomaly Detection: *A Survey. IEEE Access: Multidisciplinary Open Access Journal.* Accessible on https://creativecommons.org/licenses/by/4.0/. 9, 152379-152395, doi:10.1109/ACCESS.2021.3126834.

Xiao, L., Wan, X., Lu, X., Zhang, Y., & Wu, D. (2018). IoT security techniques based on machine learning. arXiv preprint arXiv:1801.06275.

Yin, Y., Xu, J., & Wang, C. (2020). A survey of smart home technology and applications. Computer Networks, 179, 107413. DOI: 10.1016/j.comnet.2020.107413

ZDNet, (2020) "What is IoT? Everything you need to know about the Internet of Things right now" [online]. What is the IoT? Everything you need to know about the Internet of Things right now ZDNET Accessed

on February 2024

Zeeshan Ahmad, Adnan Shahid Khan, Cheah Wai Shiang, Johari Abdullah& Farhan Ahmad (2020). Network intrusion detection system: A systematic study of machine learning and deep learning approaches. doi.org/10.1002/ett.4150

Zhang, Y., Li, Q., & Gao, L. (2023). Sensor-based data analysis for autonomous vehicles: Challenges and solutions. IEEE Transactions on Intelligent Vehicles, 8(2), DOI:10.1109/TIV.2022.3217500.