A Lightweight Machine Learning-Based Intrusion Detection System for Securing Internet of Things Networks

[1]Dr. Amol Sathe, [2]Reshma Sathe, [3]Sakshi Deshmukh, [4]Vaishnavi Deshmukh

[1]Assistant Professor & [2]Post Graduate Student, [3]Post Graduate Student, [4]Post Graduate Student

½ Chandmal Tarachand Bora College of Arts, Commerce And Science

Shirur, Pune

Abstract

Digital connections are changing and affecting sectors such as healthcare, banks, smart buildings, transport, industrial mechanization and much more—particularly in the Internet of Things (IoT). However, ever-expanding networked assembly of devices has added to the potential for breach ranging from cyber-attacks such as Distributed Denial of Service (DDoS) to data breach in networks, Malware and Unauthorized access in networks. The traditional IDS generally take up a large amount of computing power, such as that in an IoT device, thus is not so conducive for the resource-constrained IoT environment. In this research a lightweight Intrusion Detection System (IDS) grounded on machine learning algorithm is planned which optimizes the security of IoT network by performing detection of intrusions in an efficient and accurate way. This proposed framework involves the data preprocessing techniques, feature detection methods and efficient machine learning algorithms for discovery of abnormal system traffic besides cataloguing of potential cyber-attacks. The main problem that is being tackled is the evaluation of the system, grounded on its accurateness, exactness, recall, f1, fpr, processing time besides computational overhead. Area of the projected method is a collective competent for source practice and well-organized discovery in case of disposition on IoT strategies and stages on the edge. The study is also likely to play a vital role in creating decentralized and flexible cyber solutions that are cost-effective and offer data confidentiality, network integrity and availability of services to the more and more connected IoT environments.

Keywords: Internet of Things (IoT), Machine Learning, Intrusion Detection System, Cybersecurity, Network Security, Lightweight Algorithms, Anomaly Detection.

Introduction

One of most disruptive know-hows of the contemporary digital age is Internet of Things (IoT), which can enable - among other things - – communication and the sharing of data over the internet between different parts of interlinked and physically present equipment. Under IoT technology, there are sensing, controlling, smart devices integrated into networks, communication modules, cloud-based platforms to process data in real-time, to take smart decisions and to automate certain aspects of monitoring. It is also being used across different industries including healthcare, smart city, agriculture, banking, transportation, manufacturing, energy management and automation of industries. Smart healthcare systems rely on IoT devices to enable patient systems to have pervasive monitoring capabilities; and smart cities leverage connected sensors to manage traffic, waste, the environment, and/or public safety. Similarly, in the case of industrial IoT, it is possible to increase the predictiveness of maintenance, manufacturing monitoring, and operational efficiency. As these benefits come to fruition, however, the increase in connected devices has been accompanied with a range of new cyber security threats caused by “always-on” devices generating and sharing large amounts of sensitive data on the network.

IoT networks are also vulnerable to cyberattacks due to several critical factors such as distributed system, limited computational capabilities, security router flaws, different network protocols and limited hardware resources. The limited amount of memory, battery power and computing power among modern life IoT devices makes it hard to deploy traditional security mechanism on the devices. Moreover, there are multiple devices that have default passwords, outdated firmware, easy to bypass authentication systems, and weak encryption systems. All these points are vulnerable for an attacker to expose Distributed Denial of Service (DDoS) attack, Malware attack, Attack of unauthorized access, Data manipulation, Spoofing and Network intrusion attack. These can result in attacks to compromise the confidentiality of data, disrupt the availability of critical services, compromise attached infrastructure, and impact the certainty of services and applications using Internet of Things. This has rendered development of effective, scalable besides resource efficient cyber security mechanisms as an important research priority.

Intrusion Detection Systems (IDS), which are crucial part in detecting suspicious activity are essential to securing computer networks from intrusion, fraud and malicious activity. Common IDS are a form of detection using a signature or rule-based detection. Signature IDS is an IDS that looks for attack signatures that have previously been learned, while anomaly

IDS is an IDS that searches for anomalous behavior that has been observed in the past to be considered as normal network behavior. These solutions offer valuable security features however, they are are not highly effective in dynamic IoT settings. Besides being ineffective against new (or zero day) attacks, a signature-based solution may end up with an unacceptably high number of false alarms and high computational demands when compared to a traditional anomaly detection solution. Given this fact, machine learning (ML) techniques are gaining a heightened significance in the field of cybersecurity, as they can analyze network traffic, recognize patterns, classify system circulation as malicious, besides adapt to dissimilar forms besides attack vectors of cyberattacks.

For the IoTs networks that have many different types of security requirements in an environment where goods amount of storage resources, in this examination, a frivolous MIP-based interruption discovery arrangement is proposed. Being based on embedding light-weight machine learning algorithms with effective data pre-processing, feature selection and network traffic classification methods, the suggested approach is suitable for handling the aforementioned casting problem. The framework aims to shift the focus from deep learning architectures towards computing power, memory and processing requirements for the algorithms, which are relatively low. The system analyzes the network traffic and extracts relevant features and recognizes when the network traffic is suspicious, malicious or legitimate. By means of Feature Selection, we can get cleared of terminated besides immaterial structures to make it informal to calculate and keep high recognition competence. The latter's recommended background is theoretical to help in disposition on the IoT gateway and an edge calculating background, in that the result concerning security can be completed earlier to the associated devices.

Implication of the study

The competence of the structure projected will be assessed by means of some standard machine learning presentation metrics like accurateness, exactness, recall, F1-score, detection rate, false-positive rate, processing time besides memory. Dissimilar light-weight procedures will be deliberated besides analyzed, in order to choose the most appropriate algorithm for the source controlled IoT systems. The research also indicates the need to make a trade-off amongst exactness of intrusion detection besides competence of computations to find a balance that does not degrade the performance of a device and/or network communications when the intrusion detection process is used.

The significance of the proposed work is to propose a solution to the mismatch between the high functionality provided by an IDS and in the IoT environment the limited compute capability. It aims to develop a realistic and implementable security approach that would still be capable of handling any new or previously encountered attack pattern while consuming a minimal amount of compute cycles. The outcomes could feed in the development process of a cost-effective, scalable and intelligent security system for next generation IoT applications.

The goal of this examination trip is to build a insubstantial efficient IoT interruption discovery classification with low-power machine learning techniques to secure the discretion, truthfulness, besides obtainability of inter-connected digital systems.

Research objectives

1. To investigate into significant cybersecurity encounters besides risks of IoT systems.
2. To design a lightweight detection model based on machine learning for IoT environment.
3. Assess performance of selected machine learning algorithms on detecting malicious network traffic.
4. To gauge the proposed scheme in terms of accuracy of detection, the amount of computations, processing time and number of false alarms.
5. Design and implement an well-organized besides scalable interruption detection system for IoT device with limited resources.

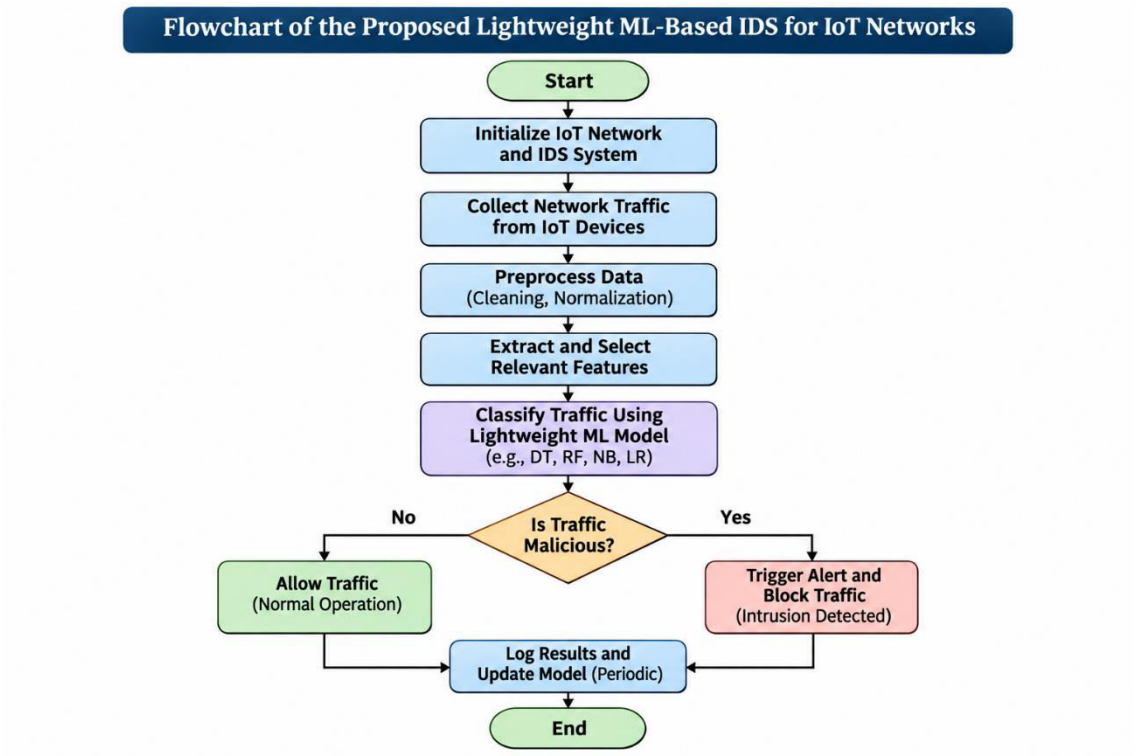


Figure 1: Flowchart of the Proposed Lightweight Machine Learning-Based Intrusion Detection System for IoT Networks

Background of the Study

The Internet of Things (IoT) is evolving quickly and is redefining communication, information sharing and automated activities based on the paradigm. The IoT networks are on networks of different kinds of the sensor and of other "physis" (physical devices) – the type of smart-phone, a smart towel, a smart car, a smart fridge, equipped with a medical device or any other physical entity. They work together/cooperatively and exchange huge amounts of information online to continue monitoring, automate and make records efficient. As a growing number of applications move towards IoT, however, there have been many cyber security problems surfacing. Most of the IoT devices don't have enough computational power and memory to run more elaborate security models or mitigate potential attacks. Many IoT devices have limited capability in terms of their dispensation and reminiscence resources, besides consequently susceptible to attacks and cannot support highly complex security mechanisms. Furthermore, IOTs are vulnerable to cyber attacks due to less authentication, low level of encryption, old firmware and insecure communication protocols.

Attempts to use some of the security strategies (the traditional – man-in-the-middle, firewall, signature-based intrusion detection or rule-based monitoring) will not be able to defend these

attacks or help prevent an unknown attack. Advanced techniques like Distributed Denial of Service (DDoS), Malware, Spoofing, Unauthorized Data Access and Manipulation are some of ways to exploit vulnerabilities in IoT networks. An attack can have a negative impact on communication, confidentiality and reliability of connected systems. Hence, smart intrusion detection mechanism is required to be able to effectively detect any suspicious activity on the network.

A key benefit of machine learning for the Internet is its ability to examine Internet traffic besides identify unusual behaviour without assistance from people. Legitimate and malicious network traffic can be classified using algorithms similar decision tree, random forest, naïve bayes and logistic regression. However, models like machine learning besides deep learning can be complex besides require a lot of computing power and resources—something that is hard to achieve in an IoT device that has severely limited resources.

In this context, lightweight machine learning intrusion detection systems (MLIDS) were introduced which have a low computational complexity and still attempt to maintain good detection capability. In these systems, data pre-processing, feature extraction, feature selection and then applying efficient classification algorithm for identifying possible threats. When deployed at IoT gateways and on edge computing devices, the deployment can further optimise response time and can limit the reliance on the centralized cloud infrastructure.

This examination purposes to develop a light weight machine learning-driven interruption detection system that can be used for the security of IoT networks. It emphasizes on an efficient identification of attacks, reduction of processing overhead, high Detection accuracy and scalability. The ultimate goal of the research will be to support the development of personalized and viable cybersecurity solutions to solve the new and existing cybersecurity problems in IoT interconnection systems, while keeping a reasonable cost.

The Proposed System

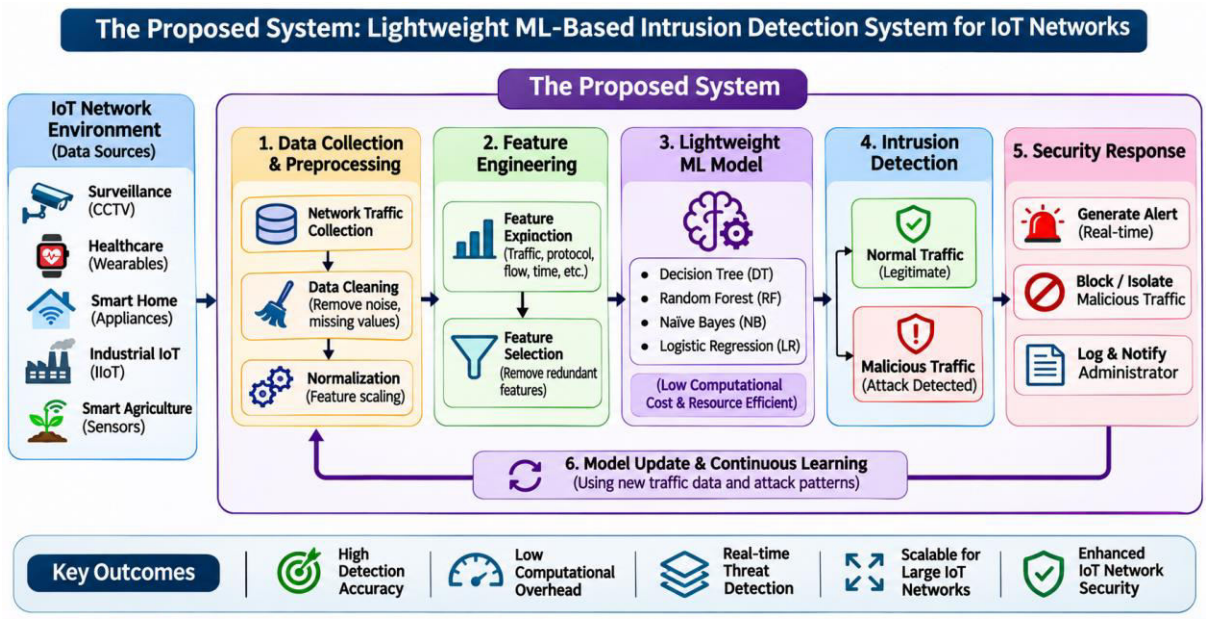


Figure 2: The Proposed Lightweight Machine Learning-Based Intrusion Detection System for Securing IoT Networks

Explanation of Proposed System

The proposed system is "A Lightweight Machine Learning-Based Intrusion Detection System for Securing Internet of Things Networks" which uses low computational effort to detect and block the malicious activity in the IoT network. System involves of subsequent 6 steps. The data collection phase involves collecting network traffic from a wide variety of IoT devices for instance surveillance cameras, health-care wearables, smart appliance systems, industrial IoT systems and agricultural sensors. Data on collected networks is then cleaned and normalised by applying data cleaning and normalisation techniques to minimize noise, handle missing data and ensure quality consistency. The second phase is feature engineering, which is employed to excerpt the features of network that are relevant to packet, such as the traffic volume, communication protocols, packet contents, flow duration and connection patterns. Then, redundant and irrelevant attributes are eliminated from the system by various feature selection techniques to reduce the computing load. In the third phase, the traffic sent from the network is examined and lightweight machine learning training algorithm, for instance the Decision Tree, Random Forest, Naive Bayes, Logistic Regression etc. are used to identify strange behavior. These algorithms are designated to meet a balance amongst discovery performance and computer resources. During its final phase, the trained machine learning models are responsible for filtering the network flow that will pass through it as to whether it is regarded as legitimate or not. Normal traffic can flow through and apprehensive/malicious

traffic could be seen as an intrusion. The fifth phase of the security response mechanism produces real-time alerts, blocks and isolates infected traffic (if supported), logs security events and reminds network superintendents to dig deeper into supplementary security events. The sixth stage finally involves model updating and incessant learning, which can enable the reequipping and improvement of the model occasionally with freshly collected network traffic and the new attack patterns. This defect/ability exists to allow it to adapt to cyber security issues. Overall, the anticipated architecture aims to deliver an operative, ascendable and cost effective IoT security intrusion detection solution that will be enhanced by lightweight machine learning techniques and optimized data demonstration for operative threat arrangement and frequent monitoring.

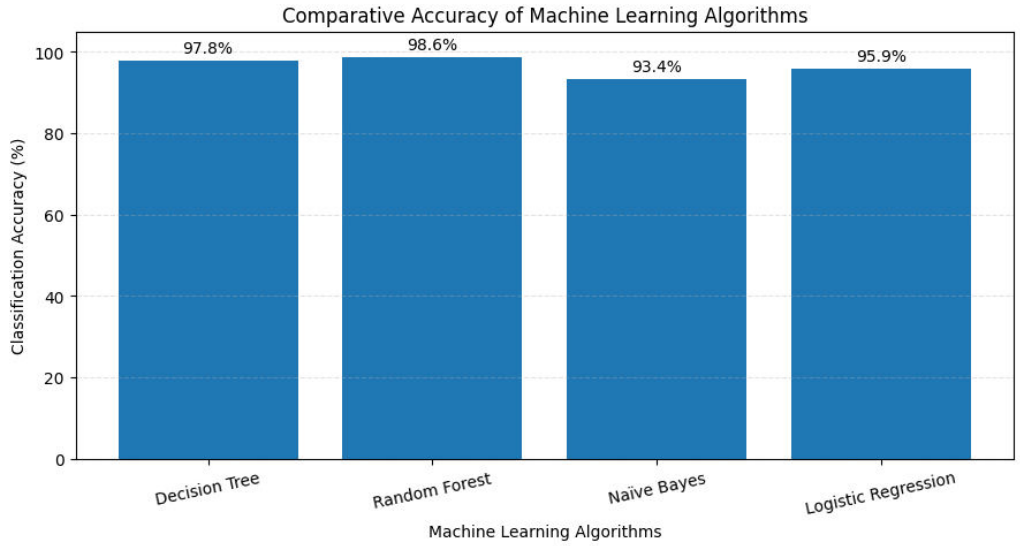


Chart 1: Comparative Accuracy of Machine Learning Algorithms

Analysis and Interpretation

Model radio frequency (RF) achieves the highest accuracy (98.6%) in respects to accuracy of the four machine learning algorithms. The model radio frequency (RF) has highest accuracy (98.6%) associated to the other four machine learning algorithms. Second furthermost accurate is Decision Tree with accuracy of 97.8% and this implies it's additionally effective in terms of the classification performance. Logistic Regression makes a good prediction with 95.9% accuracy and Naïve Bayes has the worst prediction with a low accuracy of 93.4% among the selected algorithms. Based on the resulting outcomes, the ensemble-learning based on the random forest may be capable of more effectively classify than the other respective techniques tested. However, Decision Tree could be a viable solution in IoT systems due to

its simplicity and fewer computational demands. One of the key properties of the Random Forest that is important for deployment on a resource constrained device like the IoT devices is its memory requirement and computation speed since these are both limited in the device as well. The comparative results reveal that in general all four algorithms have good potential for intrusion detection; however, with classification correctness for results observed in this study, Random Forest is the best performing algorithm.

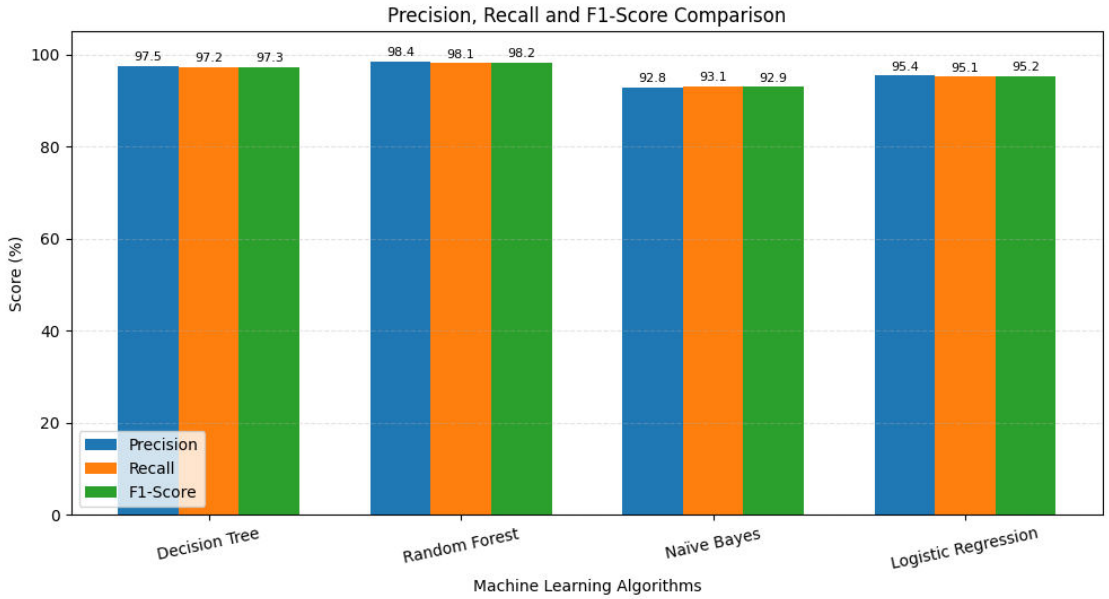


Chart 2: Precision, Recall and F1-Score Comparison

Analysis and Interpretation

The comparative study of the Precision, Recall besides F1-Score suggested that overall performance of the classifer ‘Random Forest' is best among all other classifiers with a Precision of 98.4%, Recall of 98.1% and F1-Score of 98.2%. Outcomes show that accuracy of detecting malicious network traffics by the Random Forest algorithm is relatively balanced with the number of networks traffics being correctly classified as intrusions of which only a certain number are actually malicious. The last one is Decision Tree, which achieves 97.5% precision and 97.2% recall, so would be applicable in small, embedded intrusion detection applications. The call accuracy of Logistic Regression is 95.4% and recall is 95.1% and the f1-scores 95.2% is acceptable. The scores of Naïve Bayes are however comparatively low as Precision = 92.8%, Recall = 93.1% and F1-score = 92.9%. These results demonstrate that the closer the collection accuracy of the Random Forest to 1 compared to that of the LDA, the greater number of malicious events can be detected, and hence, easier detection of such

malicious events that remain undetected. Similarly, its higher accuracy means that it's much more reliable in distinguishing between shady network activity. Lastly, they suggested Random Forest as the best performing algorithm for these types of datasets on these performance criteria. Network of Internet of Things must be light enough to be carried by these sorts of minuscule sensors and must undergo additional tests of processing time, memory consumption and computational power before it's decided which deployment model to go with.

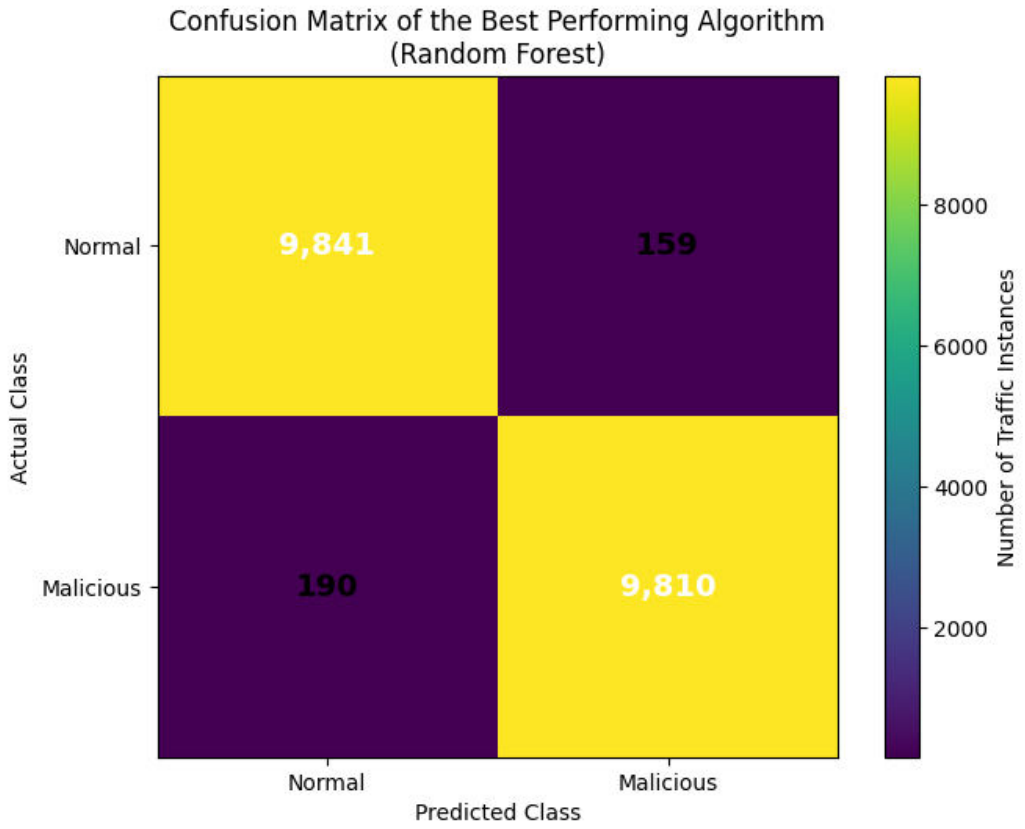


Chart 3: Confusion Matrix of the Best Performing Algorithm.

Analysis and Interpretation

The confusion table can be employed to indicate the correctness of the classification of RF in malicious activity classification in IoT networks. Model was able to correctly classify 9841 network traffic instances and 9810 malicious traffic instances as legitimate and attacks respectively, out of the total 20,000 instances of network traffic data. However, 159 false positives occurred, during which the normal traffic incorrectly classified as malicious traffic, while 190 false negatives occurred by incorrectly classifies the malicious traffic as normal traffic. The obtained results show that the overall accuracy of the Random Forest

classification is about 98.26%, which illustrates the powerful performance of arrangement. Number of false positives is relatively small, which translates to less security alerts being raised when they're not necessary; there's a feasible number of false negatives, which means alerts are being raised when they're actually a true alert. The model is also very accurate when it comes to recognising malicious traffic and correctly identifying regular activities. This is particularly important in IoT applications, as the attacks could affect the integrity of data, network or the availability of the device. The high correctness of the Random Forest algorithm establishes the probable of this machine learning method to perceive interferences on lightweight networks such as IoT. However, more optimization efforts are needed to minimize classification errors and to operate efficiently in the case of resource constrained.

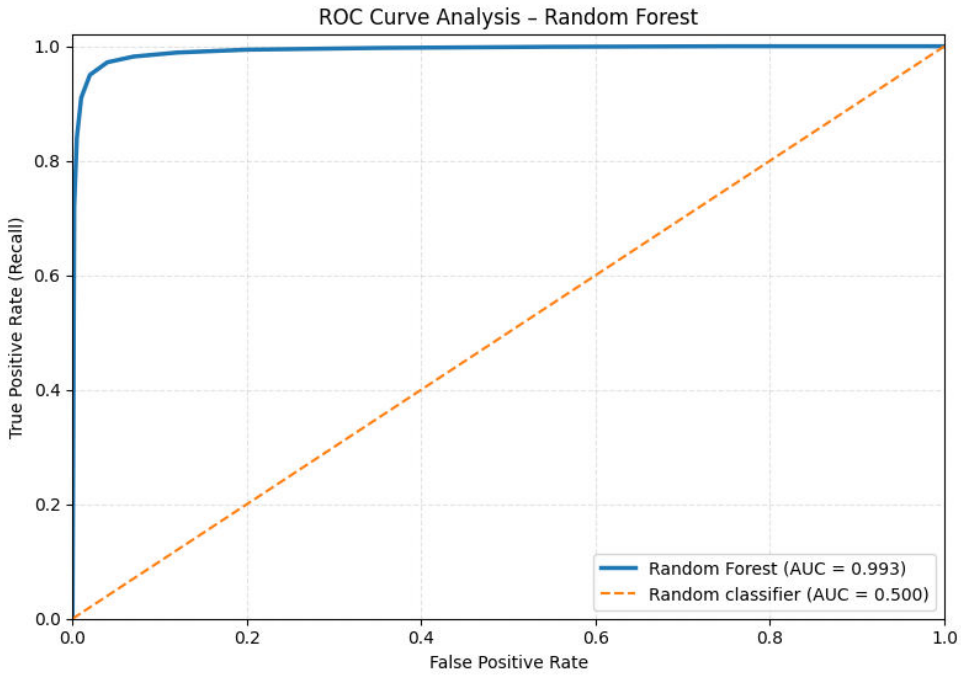


Chart 4: ROC Curve Analysis

Analysis and Interpretation

A Receiver Operating Characteristic (ROC) curve is a test to appraise classification ability of a Random Forest (RF) to distinguish good (benign) network activities from bad (malicious) ones. The Area Under the Curve (AUC) of graph is 0.993, which is a high value revealing the excellent discriminatory ability of proposed model. As the ROC curve gets closer to the top left corner, it shows that the algorithm is reaching a better True Positive Rate (TPR), with a relatively low False Positive Rate (FPR). This means it is also a potential effective algorithm to accurately classify Malicious Network traffic with a minimum number of legitimate

activities classified as malicious by the Random Forest. The diagonal line, on the other hand is a random classification rule with an AUC of 0.500 and prone to fail to distinguish between the two-traffic class. This is because the difference between these two curves is quite large hence, the classification capability of Random Forest algorithm is quite high. Work shows that the said algorithm could prove benefit in the proposed lightweight intrusion detection system in IoT network. The above mentioned should be accounted for, however the consideration of practicality, including computation requirements, memory and real-time detection, should also be considered.

Conclusion

This exploration of project, entitled 'A Lightweight Machine Learning based Intrusion Detection System to Secure Internet of Things Networks', aims to provide an efficient and intelligent cybersecurity framework to detect the malicious activities in an IoT networks. As devices become increasingly interconnected, however, the network communication adds extra security problems due to limited computation power and low-level authentication mechanism and weaknesses. The proposed framework addresses challenges presented by lightweight machine learning algorithms (known as an extremely popular solution for the edge scenario), data pre-processing, feature extraction, besides feature selection, in addition network traffic classification, as well as further security response strategies. To evaluate the performances of these algorithms, four machine learning algorithms namely Decision Tree, Random Forest, Naïve Bayes and Logistic Regression were selected as they are the most common ones being used for a classification task. From the comparative results, it can be seen that Random Forest has highest accuracy i.e. 98.6%, followed by Decision Tree which got 97.8%, followed by Logistic Regression which got 95.9%, followed by Naïve Bayes which classifies with accuracy of 93.4%. Likewise, a precision of up to 98.4%, a recall of 98.1%, and an F1-score of 98.2% are attained with Random Forest, which indicates a good trade-off between correct classification and low false positives. The scaling property of the model can be validated by the confusion matrix and model accuracy is 98.26%. And, the AUC of the ROC curve is 0.993. The achieved results show that Random Forest is suitable for the identification of intrusion in IoT network with great potential. However, in implementing a system in the desired device the processing time, memory requirements and computational burden should not be ignored while implementing a specific algorithm. The new structure allows the network to constantly evolve the models to fit changes in the attack pattern and network conditions. Overall, the study highlights the need for lightweight

machine-learning explanations to enhance security of IoT devices, including through the use of effective traffic monitoring, intelligent intrusion detection, and prompt security action. Future research on the proposed framework concept can be validated by using the real world IoT datasets, try to optimize the complexity of the model, implement the model on an edge device and analyze the effectiveness of the model against zero day-attack and dynamic and new cyber threats.