International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

AN EFFICIENT SPAM DETECTION IN IOT DEVICES USING MACHINE LEARNING

Sanjeevini s Harwalkar[1], B.Ravalika[2], P.Sneha[3], M.Neha[4], P.Srujana[5]
[1]Assistant Professor, School of CSE,Malla Reddy Engineering College For Women (Autonomous Institution), Maisammaguda,Dhulapally,Secunderabad,Telangana-500100
[2345]UG Scholar, Department of IOT, Malla Reddy Engineering College for Women, (Autonomous Institution), Maisammaguda, Dhulapally, Secunderabad, Telangana-500100
Email: sanjeevini706@gmail.com

ABSTRACT

The Internet of Things (IoT) is a network of millions of devices fitted with sensors and actuators which have either wired or wireless connectivity in order to communicate information between them. IoT has shown tremendous growth in the past ten years. By 2020, it is anticipated that there will be over 25 billion devices that would have been connected. In the coming years, these devices will emit much higher amount of data. In addition to volume, IoT devices are known to generate large data sets in various modalities with different qualities of data as defined by velocity regarding time and location dependency. Under such a scenario, the ML algorithms will play an important role in ensuring biotechnology-based security and authentication coupled with anomaly detection to further improve the usability and security of the IoT systems. On the other hand, attackers often exploit learning algorithms to exploit vulnerabilities in intelligent IoT-based systems. Motivated by this, in this article, we propose to improve the security of IoT devices through spam detection using ML. For this objective, spam detection in IoT using a machine learning framework is proposed. In this framework, five ML models are tested with different metrics, including a large collection of input feature sets. Each model computes a spam score by taking into account refined input characteristics. This score reflects the reliability of an IoT device under different parameters. The proposed technique is validated by using the REFIT smart home dataset. The achieved results show the efficiency of the proposed scheme in comparison with other existing schemes.

Keywords: Internet of Things,Machine Learning,IoT Security,Spam Detection,Smart Home Dataset

I INTRODUCTION

The Internet of Things, or IoT, is an emerging technology that connects millions of devices equipped with sensors and actuators into a dynamic network capable of wired or wireless communication. Over the last ten years, IoT has exponentially grown, with over 25 billion estimated connected devices in 2020. With increasing IoT adoption, these devices are producing vast quantities of data in diverse modalities with distinctive features such as time and location dependency. Such growth in volume and complexity of data poses serious problems concerning security, usability, and management within IoT ecosystems. ML algorithms have been very powerful in handling these issues.

Using ML, IoT systems can develop advanced functionalities such as anomaly detection, biotechnology-based authentication, and enhanced security measures. However, the intelligent algorithm-based dependence also creates vulnerabilities because attackers often exploit these systems to compromise security. This dual nature of ML in IoT underlines the requirement for strong risk-mitigating mechanisms that further improve the reliability and security of IoT networks. Spam detection is one of the most important areas that have lately been emphasized in IoT security. If spam is allowed to proliferate, it could degrade the reliability of devices, deteriorate system performance, and open up exploitable vulnerabilities.

We propose an innovative ML-based framework for detecting spam in IoT environments. The proposed framework evaluates five different ML models in a comparative manner, while testing them with comprehensive metrics and large collections of input features for refining the detection process. Each model computes the spam score, which is indicative of an IoT device's reliability under different conditions. The proposed methodology is validated using the REFIT smart home dataset, a robust resource for evaluating IoT systems. The results demonstrate the effectiveness of the framework in identifying spam and ensuring device reliability, outperforming existing methods. The use of sophisticated input features and various ML models improves the detection rate and enhances the overall security of IoT networks. This study emphasizes the opportunity of ML in handling unique challenges of IoT and establishes the need for effective anti-spam mechanisms in protection against evolving threats from such devices. The proposed framework attempts to address these issues to enable the building of a secure and reliable IoT ecosystem.

\

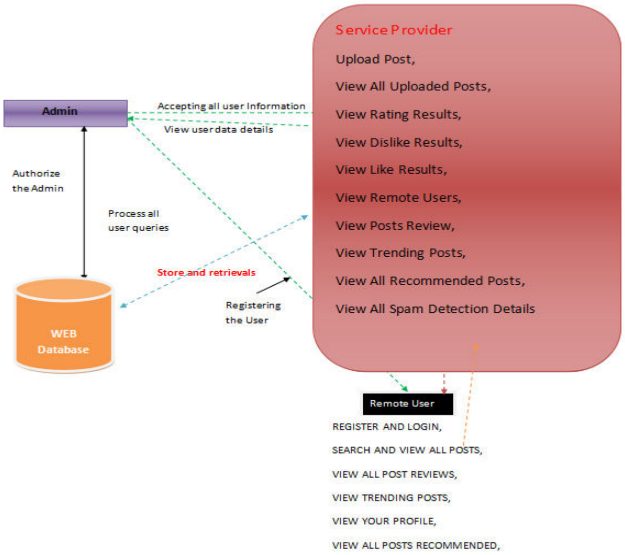


Fig 1: System Architecture

II RELATED WORK

AI-Driven Mechanism for Edge Computing-Based Industrial Applications

Authors: A. H. Sodhro, S. Pirbhulal, and V. H. C. de Albuquerque

Year: 2019

This paper discusses the integration of AI with edge computing for industrial purposes. It focuseson making industrial systems more effi cient with real-time data processing and decision-making. The work emphasizes how AI improves system performance, optimizes resource allocation, and reduces latency in edge computing for industrial processes.

Artificial Intelligence Based QoS Optimization for Multimedia Communication in IoV Systems
Authors: A. H. Sodhro, Z. Luo, G. H. Sodhro, M. Muzamal, J. J. Rodrigues, and V. H. C. de Albuquerque
Year: 2019
This paper presents the application of AI in o ptimizing Quality of Service (QoS) for multimedia communication in the Internet of Vehicles (IoV) systems. The authors suggest AI algorithms that will manage the network resources and allow efficient multimedia data transmission. This ensures delay, packet loss, and congestion are avoided in IoV environments.

REFiT Smart Home Dataset
Authors: Loughborough University
Year: 2019
The REFiT Smart Home Dataset consists of real-world data collected from sensors in a smart home environment. This dataset is valuable for research into energy consumption and IoT applications, providing insights into user behavior, appliance usage, and energy patterns. It is used to develop machine learning models that optimize energy use in smart homes.

RStudio
Authors: R (RStudio Team)
Year: 2019
RStudio is the IDE for the R programming language used for statistical computing and data analysis. It gives an easy-to-use interface to perform statistical analysis, data visualization, and machine learning tasks. Therefore, it allows a user to do research in multiple fields using its computational cap ability.

Principal Component Analysis
Author: I. Jolliffe
Year: 2011
This book is an in-depth exposition of Principal Component Analysis (PCA), a statistical technique of reducing the dimensionality of large datasets while retaining important information. It is the most fundamental tool in data analysis, used almost universally in fields like pattern recognition, bioinformatics, and image processing for dimensionality reduction with out loss of important variability.

III IMPLEMENTATION

Starting with data collection and preprocessing, the proposed framework for enhanced IoT device security through ML-based spam detection begins to use the REFIT Smart Home dataset, which contains usage data from appliances, energy consumption, and patterns of device communication. Data will then be cleaned and refined with features extracted like time of usage, frequency of communication, and irregular patterns possibly indicating malicious behavior. These features will then be used to train five different machine learning models. These are: Logistic Regression, Decision Trees, Random Forests, SVM, and Neural Networks. Each of these models will calculate the spam score based on their ability to detect whether or not a device is likely to be compromised or participate in spam-like activities. The spam score will depend on the input features provided, thus mirroring how reliable the IoT device is with different conditions. The performance metrics to be use d for evaluating the models include accuracy, precision, recall, and F1-score to ensure effective spam detection.Cross-validation will be used to assess the generalization ability of the models and prevent overfitting. After training, the models will be deployed in the IoT environment to compute real-time spam scores for each device based on new data. Devices with a high score of spam will be flagged, and further investigations or remedial actions on the network will be carried out to secure it. Also, anomalous behavior detection methods such as Isolation Forests and Autoencoders will help in identifying behavior that deviates from normal patterns.The efficie ncy and the reliability of the proposed method will be checked by cross-validation with known spam detection schemes. Lastly, the system would be optimized for scalability as well, thus it may accommodate the large num bers of devices and volumes of data generated. By utilizing ML-based spam detection, this framework presents a powerful approach for securingIoTnetworks, while keeping threats t o a bare minimum and guaranteeing operatio n in devices.

IV ALGORITHM

Logistic Regression

Logistic Regression is a basic binary classification algorithm that predicts the probability of an IoT device being spam or non-spam. It works on the principle of estimating probabilities of a target based on its input features and is preferred for its simplicity and interpretability. It is particularly useful in problems where the relationship between feature variables and target labels is linear. In this project, Logistic Regression is used to determine the spam score of an IoT device, for which the model outputs the probability of a device being spam.

2. Decision Trees

Decision Trees are one of the most important machine learning algorithms used for classification. In this system, Decision Trees are used to classify IoT devices based on certain input characteristics. The tree structure splits the data into subsets, which helps in decision-making based on different feature values. Though Decision Trees are easy to interpret, they can suffer from overfitting. To overcome this limitation, Random Forests are introduced.

3. Random Forests

Random Forests is an ensemble method that combines multiple Decision Trees to enhance prediction accuracy. In this approach, each tree in the forest is trained on a random subset of the data, and the final classification decision is made by averaging the results from all trees. This reduces overfitting and improves the overall robustness of the model, making it particularly effective for more complex datasets like those found in IoT systems.

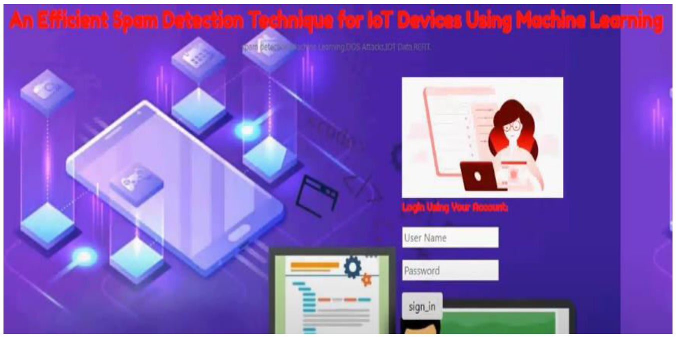
Support Vector Machines (SVM)

Support Vector Machines (SVM) are used in the classification of IoT devices when the data is not linearly separable. SVM maps the input data into a higher-dimensional space where an optimal hyperplane can be found to separate different classes. SVM is particularly useful in cases where the relationship between the features and output labels is complex and requires non-linear decision boundaries. This flexibility of SVM enables it to handle a wide variety of data structures, and this is the reason it has been integrated into the spam detection system.

Neural Networks

Neural Networks, particularly deep learning models,areutilizedinthissystem for detecting complex, non-linear relationships within the data. Such models consist of several layers of interconnected nodes that process input features and learn patterns from large datasets. Neural Networks can discover the most complex anomalies and patternsinIoT devices that simple algorithms could never discover. One reason for using Neural Networks is its ability to extract features from raw data, which helps make it ideal for large-scale IoT systems

RESULTS



An Efficient Spam Detection Technique for IoT Devices Using Machine Learning

Login Using Your Account:

Fig:1:User Login

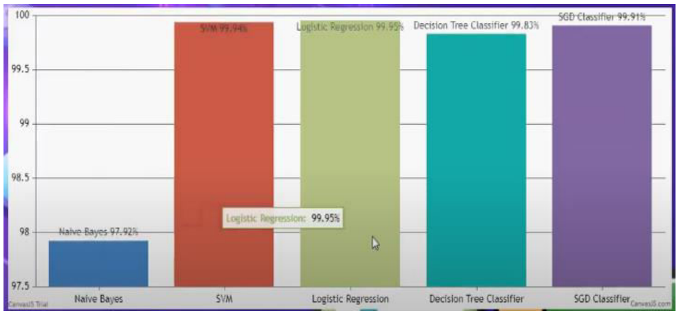


Fig:2:Trained and tested Bar Chart

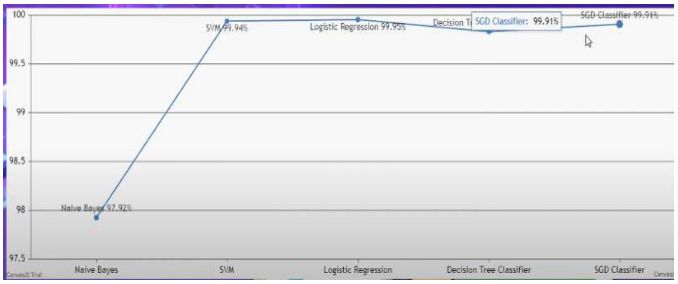
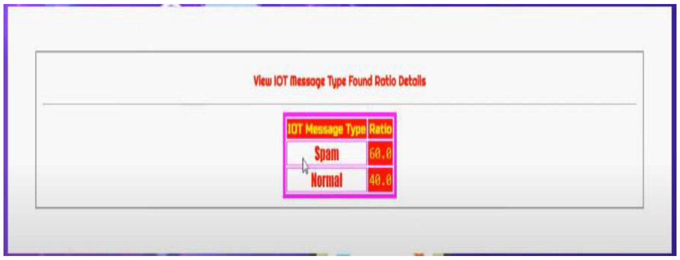


Fig:3:Trained and tested Accuracy Results



View IOT Message Type Found Ratio Details

IOT Message Type	Ratio
Spam	60.0
Normal	40.0

Fig:5:Spam &Normal Ratio

CONCLUSION

The proposed framework is an effective approach for spam parameters detection in IoT devices using machine learning models. Using a comprehensive feature engineering process, the IoT dataset is carefully preprocessed in extracting the most relevant features for optimal performance of the machine learning models. Such models then assign a spam score to each IoT device, and thus the system identifies security threats and abnormal behaviors. This is very useful in smart homes where IoT devices are critical to daily operations, security, and convenience.

The framework is significant in assessing the reliability and performance of IoT devices. It does this by computing a spam score for each device, which helps assess the legitimacy of devices in the network, thereby improving the overall security of the IoT ecosystem. As the number of interconnected devices grows, the identification of malicious activities or unusual behavior will be a critical concern to ensure the security of this environment and its efficiency. This is where smart homes, relying on IoT devices for operations critical to the safety of people living in them, should take utmost care in ensuring the reliability of such devices.

This framework can be further enhanced to accommodate further environmental factors such as climatic conditions or surrounding influences for a better improvement in the framework. Such parameters will deepen the insights into the behavior of IoT devices, thus more accurate anomaly detection and threat identification will be performed. It would enable the system to be more responsive towards the real-world conditions in dealing with the performance and security of devices. For example, temperature changes, humidity, or network congestion may drastically influence how IoT devices function; taking these factors into consideration could enhance the system's performance in detecting problems.

These future improvements will contribute to making IoT devices more secure, reliable, and trustworthy, not only in smart homes but also in other IoT environments. As the IoT ecosystem continues to expand, the framework has the potential to evolve and be adapted to meet emerging security challenges, offering long-term solutions for the growing number of interconnected devices. Ultimately, this framework will play a crucial role in improving the security and efficiency of IoT systems, making them safer and more reliable for users.

REFERENCES

[1] A. H. Sodhro, S. Pirbhulal, and V. H. C. de Albuquerque, “Artificialintelligence driven mechanism for edge computing based industrialapplications,” IEEE Transactions on Industrial Informatics, 2019.

[2] A. H. Sodhro, Z. Luo, G. H. Sodhro, M. Muzamal, J. J. Rodrigues, andV. H. C. de Albuquerque, “Artificial intelligence based qos optimizationFor multimedia communication in iov systems,” Future GenerationComputer Systems, vol. 95, pp. 667–680, 2019.

[3] L. University, “Refit smart home dataset,” https://repository.lboro.ac.uk/articles/REFIT Smart Home dataset/2070091, 2019 (accessed April 26,2019).

[4] R, “Rstudio,” 2019 (accessed October 23, 2019).

[5] I. Jolliffe, Principal component analysis. Springer, 2011.

[6] I. Guyon and A. Elisseeff, "An introduction to variable and featureselection," Journal of machine learning research, vol. 3, no. Mar, pp.1157–1182, 2003.

[7] L. Yu and H. Liu, "Feature selection for high-dimensional data: A fastcorrelation-based filter solution," in Proceedings of the 20th internationalconference on machine learning (ICML-03), 2003, pp. 856–863.

[8] Z. Tan, A. Jamdagni, X. He, P. Nanda, and R. P. Liu, "A systemfor denial-of-service attack detection based on multivariate correlationanalysis," IEEE transactions on parallel and distributed systems, vol. 25,no. 2, pp. 447–456, 2013.

[9] Y. Li, D. E. Quevedo, S. Dey, and L. Shi, "Sinr-based dos attack onremote state estimation: A game-theoretic approach," IEEE Transactionson Control of Network Systems, vol. 4, no. 3, pp. 632–642, 2016.

[10] L. Xiao, Y. Li, X. Huang, and X. Du, "Cloud-based malware detectiongame for mobile devices with offloading," IEEE Transactions on MobileComputing, vol. 16, no. 10, pp. 2742–2750, 2017.

[11] J. W. Branch, C. Giannella, B. Szymanski, R. Wolff, and H. Kargupta,"In-network outlier detection in wireless sensor networks," Knowledgeand information systems, vol. 34, no. 1, pp. 23–54, 2013.

[12] A. L. Buczak and E. Guven, "A survey of data mining and machinelearning methods for cyber security intrusion detection," IEEE CommunicationsSurveys & Tutorials, vol. 18, no. 2, pp. 1153–1176, 2015.

[13] F. A. Narudin, A. Feizollah, N. B. Anuar, and A. Gani, "Evaluationof machine learning classifiers for mobile malware detection," SoftComputing, vol. 20, no. 1, pp. 343–357, 2016.

[14] H. Eun, H. Lee, and H. Oh, "Conditional privacy preserving securityprotocol for nfc applications," IEEE Transactions on Consumer Electronics,vol. 59, no. 1, pp. 153–160, 2013.

[15] R. V. Kulkarni and G. K. Venayagamoorthy, "Neural network basedsecure media access control protocol for wireless sensor networks," in2009 International Joint Conference on Neural Networks. IEEE, 2009,pp. 1680–1687.

[16] C. Zhang and R. Green, "Communication security in internet of thing:preventive measure and avoid ddos attack over iot network," in Proceedings

of the 18th Symposium on Communications & Networking. Societyfor Computer Simulation International, 2015, pp. 8–15.

[17] W. Kim, O.-R. Jeong, C. Kim, and J. So, "The dark side of the internet:Attacks, costs and responses," Information systems, vol. 36, no. 3, pp.675–705, 2011