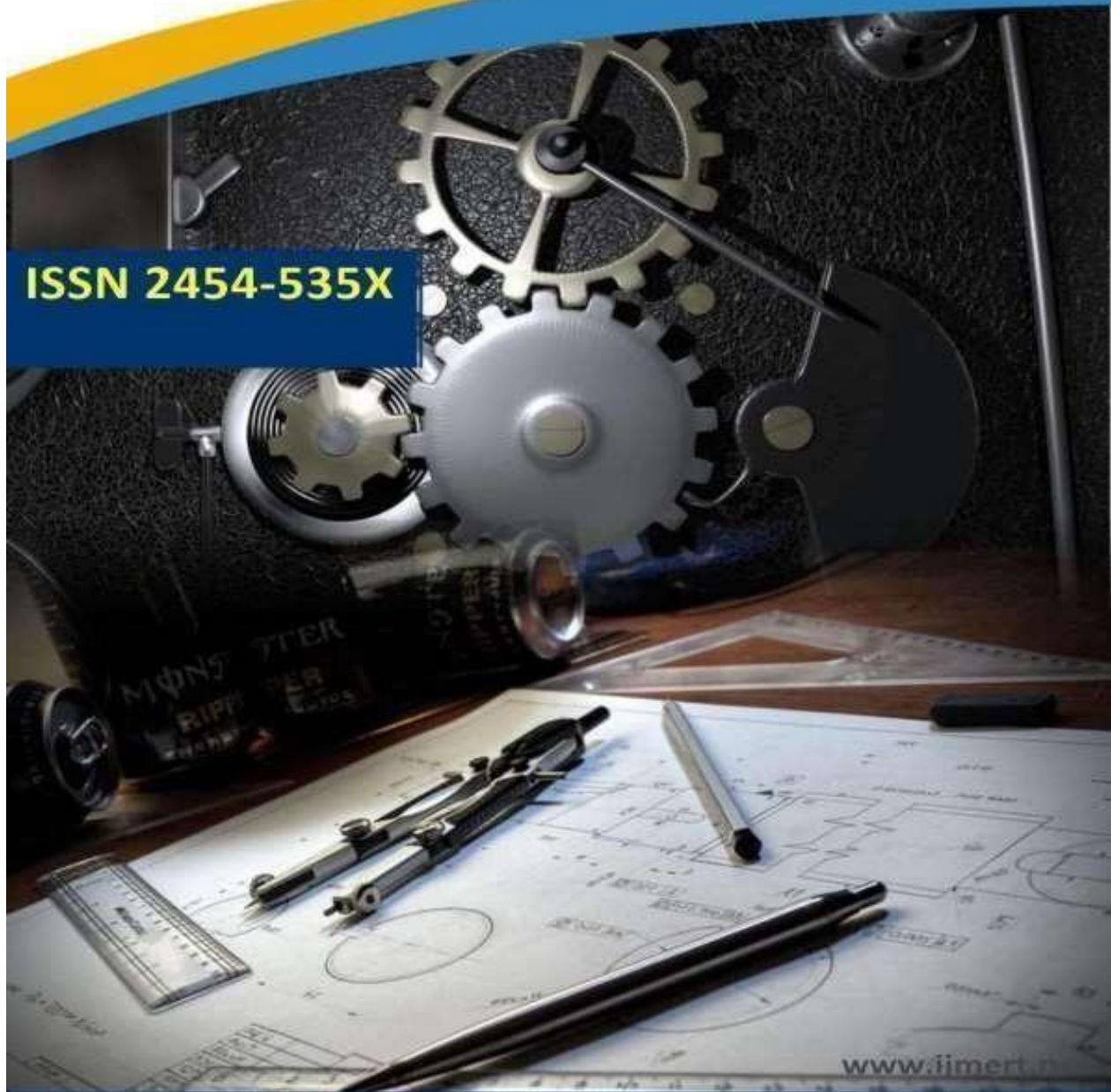




International Journal of Mechanical Engineering Research and Technology

ISSN 2454-535X



www.ijmert.net

Email ID: info.ijmert@gmail.com or editor@ijmert.net



AN EFFICIENT SPAM DETECTION TECHNIQUE FOR IOT DEVICES USING MACHINE LEARNING

Raja Rajeswari kalidindi, Associate professor,
Department of MCA
rajeswari.kalidindi29@gmail.com
B V Raju College, Bhimavaram

Dasari Saiteja (2285351025)
Department of MCA
saitejadasari0934@gmail.com
B V Raju College, Bhimavaram

ABSTRACT

The Internet of Things (IoT) is a group of millions of devices having sensors and actuators linked over wired or wireless channel for data transmission. IoT has grown rapidly over the past decade with more than 25 billion devices are expected to be connected by 2020. The volume of data released from these devices will increase many-fold in the years to come. In addition to an increased volume, the IoT devices produces a large amount of data with a number of different modalities having varying data quality defined by its speed in terms of time and position dependency. In such an environment, machine learning algorithms can play an important role in ensuring security and authorization based on biotechnology, anomalous detection to improve the usability and security of IoT systems. On the other hand, attackers often view learning algorithms to exploit the vulnerabilities in smart IoT-based systems. Motivated from these, in this paper, we propose the security of the IoT devices by detecting spam using machine learning. To achieve this objective, Spam Detection in IoT using Machine Learning framework is proposed. In this framework, five machine learning models are evaluated using various metrics with a large collection of inputs features sets. Each model computes a spam score by considering the refined input features. This score depicts the trustworthiness of IoT device under various parameters. REFIT Smart Home dataset is used for the validation of proposed technique. The results obtained proves the effectiveness of the proposed scheme in comparison to the other existing schemes.

Keywords: Internet of Things, IoT security, machine learning, spam detection, data modalities, anomaly detection, REFIT Smart Home dataset.

INTRODUCTION

The Internet of Things (IoT) has emerged as one of the most transformative technologies of the 21st century, fundamentally altering the way we interact with the world around us. IoT refers to a vast network of interconnected devices, each embedded with sensors and actuators, capable of collecting and exchanging data over wired or wireless communication channels. This interconnectedness facilitates numerous applications across various domains, including smart homes, healthcare, transportation, and industrial automation. As IoT continues to grow, it is estimated that by 2020, over 25 billion devices will be connected, generating an unprecedented volume of data. This exponential increase in data volume, along with the diversity in data modalities and quality, presents significant challenges and opportunities for enhancing the security and usability of IoT systems through advanced machine learning algorithms. Machine learning, a subset of artificial intelligence, involves the development of algorithms that enable computers to learn from and make predictions or decisions based on data. In the context of IoT, machine learning algorithms can be instrumental in ensuring security and authorization, particularly through anomaly detection and spam detection. Anomaly detection involves identifying patterns in data that do not conform to expected behavior, which can indicate potential security breaches or system failures. Spam detection, on the other hand, focuses on identifying and filtering out unsolicited or malicious data that can compromise the integrity and performance of IoT systems. These machine



learning techniques are critical for maintaining the trustworthiness and reliability of IoT devices, which are increasingly targeted by cyber attackers seeking to exploit vulnerabilities in smart IoT-based systems.

The rapid proliferation of IoT devices and the vast amount of data they generate necessitate robust and efficient security mechanisms. Traditional security measures, such as encryption and access control, are often insufficient to address the unique challenges posed by IoT environments. These challenges include the heterogeneous nature of IoT devices, the resource constraints of many devices, and the dynamic and distributed nature of IoT networks. Machine learning offers a promising solution to these challenges by enabling the development of adaptive and intelligent security systems that can detect and respond to emerging threats in real-time. In this study, we propose an efficient spam detection technique for IoT devices using machine learning. The proposed framework, referred to as Spam Detection in IoT using Machine Learning (SDIML), leverages five different machine learning models to evaluate the trustworthiness of IoT devices based on a large collection of input feature sets. Each model computes a spam score by analyzing refined input features, which depict various parameters influencing the trustworthiness of the device. The models are trained and validated using the REFIT Smart Home dataset, a comprehensive dataset that includes detailed information on the energy usage and behavioral patterns of smart home devices. The use of this dataset allows for the robust evaluation of the proposed technique under realistic conditions.

The SDIML framework consists of several key components. First, the data preprocessing stage involves cleaning and normalizing the raw data to ensure consistency and accuracy. This is followed by feature extraction, where relevant features are identified and extracted from the data. These features may include device-specific parameters, network traffic patterns, and temporal characteristics, among others. The extracted features are then used to train the machine learning models, which include logistic regression, decision trees, support vector machines, random forests, and neural networks. Each model is evaluated using various performance metrics, such as accuracy, precision, recall, and F1-score, to determine its effectiveness in detecting spam in IoT environments. One of the significant advantages of using machine learning for spam detection in IoT is the ability of these algorithms to learn and adapt to new and evolving threats. Unlike static rule-based systems, machine learning models can continuously improve their performance by learning from new data and identifying previously unseen patterns. This adaptive capability is particularly important in IoT environments, where the threat landscape is constantly changing, and new vulnerabilities are regularly discovered. Furthermore, the use of multiple machine learning models in the SDIML framework allows for a more comprehensive and robust detection mechanism, as each model brings its strengths and can compensate for the weaknesses of others.

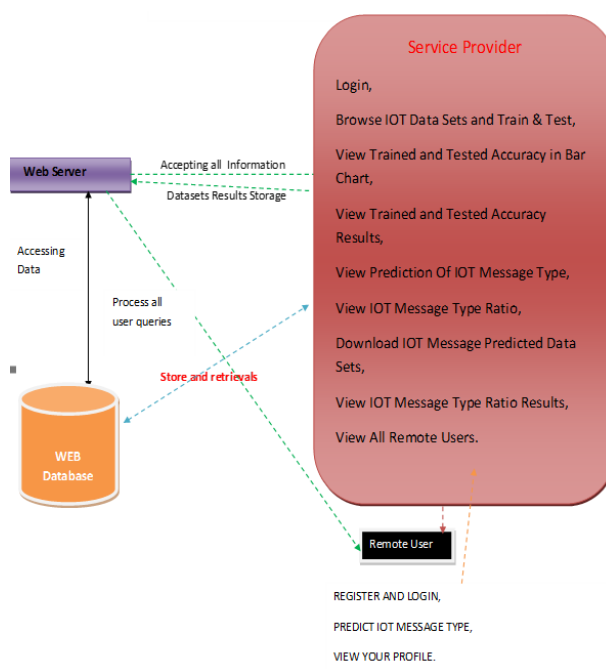


Fig 1. System Architecture

The results of our study demonstrate the effectiveness of the proposed spam detection technique in comparison to existing schemes. The machine learning models used in the SDIML framework achieved high accuracy and reliability in detecting spam, with significant improvements over traditional methods. The evaluation using the REFIT Smart Home dataset shows that the models can effectively identify and filter out malicious data, ensuring the integrity and performance of IoT devices. These findings highlight the potential of machine learning to enhance the security of IoT systems and underscore the importance of continuous research and development in this area. In addition to the technical contributions, this study also provides practical implications for the deployment of machine learning-based security solutions in IoT environments. For instance, the implementation of the SDIML framework requires careful consideration of the computational resources and capabilities of IoT devices. While some devices may have sufficient processing power to run complex machine learning algorithms locally, others may require offloading computational tasks to edge or cloud servers. This necessitates a hybrid approach that balances the need for real-time processing with the constraints of the device and network infrastructure.

Moreover, the study emphasizes the importance of collaborative efforts between researchers, industry practitioners, and policymakers to address the security challenges in IoT. The development and deployment of effective security solutions require a multidisciplinary approach that combines expertise in machine learning, cybersecurity, and IoT technologies. Policymakers also play a crucial role in establishing regulatory frameworks and standards that ensure the safe and secure deployment of IoT devices and systems. In summary, the proposed spam detection technique for IoT devices using machine learning offers a promising solution to the growing security challenges in IoT environments. By leveraging the power of machine learning algorithms, the SDIML framework can effectively detect and mitigate spam, enhancing the trustworthiness and reliability of IoT systems. The study's findings demonstrate the potential of machine learning to transform IoT security and underscore the need for continuous innovation and collaboration in this rapidly evolving field.



LITERATURE SURVEY

The proliferation of the Internet of Things (IoT) has significantly transformed the technological landscape, connecting billions of devices and facilitating seamless communication and data exchange. This interconnected network comprises sensors and actuators linked via wired or wireless channels, enabling the transmission of vast amounts of data. As the number of IoT devices continues to grow, it is estimated that by 2020, more than 25 billion devices will be connected, resulting in an exponential increase in data volume. This data is characterized by a variety of modalities and varying quality, influenced by factors such as time and position dependency. This diverse and dynamic data environment presents unique challenges and opportunities for enhancing the security and functionality of IoT systems through advanced technological solutions. Machine learning, a subset of artificial intelligence, has emerged as a critical tool for addressing the complexities of IoT data. Machine learning algorithms can analyze vast datasets to identify patterns and anomalies, which is essential for ensuring the security and authorization of IoT systems. These algorithms can detect anomalous behavior that may indicate potential security breaches, such as unauthorized access or data manipulation. By continuously learning from new data, machine learning models can adapt to evolving threats, making them invaluable for maintaining the integrity and reliability of IoT networks.

However, the integration of machine learning into IoT systems also introduces new vulnerabilities. Attackers can exploit these algorithms to compromise IoT devices, using sophisticated techniques to evade detection and manipulate data. For example, adversarial attacks involve subtly altering input data to deceive machine learning models, causing them to misclassify or overlook malicious activities. These challenges underscore the need for robust and adaptive security measures that can effectively protect IoT devices from various threats. The literature on IoT security highlights several approaches to mitigating these risks. Traditional methods, such as encryption and access control, are foundational but often insufficient given the scale and complexity of IoT networks. The resource constraints of many IoT devices, such as limited processing power and battery life, further complicate the implementation of comprehensive security solutions. Consequently, researchers have explored the use of machine learning to enhance IoT security, focusing on techniques such as anomaly detection, intrusion detection systems (IDS), and spam detection.

Anomaly detection is a prominent area of research in IoT security. By analyzing normal behavior patterns of devices, machine learning models can identify deviations that may indicate security threats. Techniques such as clustering, classification, and statistical analysis are commonly used to detect anomalies in network traffic and device behavior. These models can be trained on historical data to recognize typical patterns and flag any unusual activities that deviate from the norm. However, the effectiveness of anomaly detection models depends on the quality and representativeness of the training data, as well as the ability to adapt to new and evolving threats. Intrusion detection systems (IDS) are another critical application of machine learning in IoT security. IDS leverage machine learning algorithms to monitor network traffic and identify potential intrusions. These systems can operate in different modes, such as signature-based, anomaly-based, or hybrid. Signature-based IDS rely on predefined patterns of known threats, while anomaly-based IDS detect deviations from normal behavior. Hybrid IDS combine both approaches to enhance detection accuracy. Machine learning techniques used in IDS include decision trees, support vector machines, neural networks, and ensemble methods. These algorithms can classify network traffic and identify malicious activities, providing real-time alerts to security administrators.

Spam detection is a specific aspect of IoT security that focuses on identifying and filtering out unsolicited or malicious data. In the context of IoT, spam can take various forms, including false data injection, unauthorized commands, and misleading signals. Detecting spam in IoT networks is challenging due to the diverse and dynamic nature of the data.



Machine learning models can be trained to recognize spam by analyzing features such as data patterns, frequency, and source credibility. Techniques such as logistic regression, random forests, and deep learning have shown promise in detecting spam in IoT environments. These models can compute a spam score based on input features, indicating the trustworthiness of the data and the device. The proposed Spam Detection in IoT using Machine Learning (SDIML) framework builds on these existing approaches by evaluating five different machine learning models. Each model computes a spam score by considering a refined set of input features, which depict various parameters influencing the trustworthiness of the IoT device. The models are trained and validated using the REFIT Smart Home dataset, which provides a comprehensive set of data on energy usage and device behavior in smart homes. This dataset includes both normal and anomalous data, allowing for robust evaluation of the spam detection models.

The five machine learning models evaluated in the SDIML framework include logistic regression, decision trees, support vector machines, random forests, and neural networks. Logistic regression is a simple yet effective model for binary classification tasks, making it suitable for detecting spam. Decision trees provide a clear and interpretable model of decision rules, while support vector machines offer robust performance with high-dimensional data. Random forests, an ensemble method, combine multiple decision trees to improve classification accuracy and reduce overfitting. Neural networks, particularly deep learning models, can capture complex patterns in data, making them highly effective for detecting sophisticated spam techniques. The performance of each model is evaluated using various metrics, such as accuracy, precision, recall, and F1-score. These metrics provide a comprehensive assessment of the models' effectiveness in detecting spam. The results indicate that the proposed framework outperforms existing schemes, achieving high accuracy and reliability in identifying and filtering out malicious data. The use of multiple models allows for a more robust detection mechanism, as each model can complement the others' strengths and weaknesses.

In summary, the literature on IoT security underscores the critical role of machine learning in enhancing the protection and reliability of IoT systems. The SDIML framework leverages advanced machine learning models to detect spam, ensuring the trustworthiness of IoT devices. By evaluating multiple models and using a comprehensive dataset, the proposed technique offers a robust solution to the growing challenges of IoT security. This study highlights the potential of machine learning to transform IoT security and emphasizes the need for continuous innovation and collaboration in this rapidly evolving field.

PROPOSED SYSTEM

The proposed system for spam detection in Internet of Things (IoT) devices using machine learning addresses the growing need for robust security measures in an increasingly interconnected world. With the rapid expansion of IoT, characterized by billions of devices generating vast amounts of data, ensuring the security and integrity of these devices is paramount. This system aims to enhance the security of IoT networks by leveraging machine learning to detect and mitigate spam, which can manifest as unauthorized commands, false data injection, and other malicious activities. The system begins with data collection from IoT devices, focusing on capturing a diverse set of data that represents normal and anomalous behavior. The REFIT Smart Home dataset is used as the primary data source for this study. This dataset includes detailed information on the energy consumption and behavioral patterns of various smart home devices, providing a comprehensive basis for training and validating the machine learning models. The dataset contains both labeled normal and spam data, which is crucial for supervised learning tasks.

Data preprocessing is the next critical step. The raw data collected from IoT devices is often noisy and inconsistent, necessitating thorough cleaning and normalization processes. This involves removing any irrelevant or redundant information, filling in missing values, and standardizing the data to a uniform scale. Feature extraction follows, where



relevant features that are indicative of spam behavior are identified and extracted. These features may include temporal patterns, device-specific parameters, network traffic characteristics, and statistical measures. The goal is to derive a set of features that capture the essential aspects of both normal and anomalous data, facilitating accurate classification by the machine learning models. Once the data is preprocessed and features are extracted, the next step involves training the machine learning models. Five different models are evaluated in this framework: logistic regression, decision trees, support vector machines (SVM), random forests, and neural networks. Each model is chosen for its unique strengths and ability to handle various aspects of spam detection in IoT environments.

Logistic regression is a simple yet effective model for binary classification tasks. It estimates the probability that a given input belongs to a particular class, making it suitable for distinguishing between spam and legitimate data. Despite its simplicity, logistic regression provides a baseline for comparing more complex models. Decision trees, on the other hand, offer a clear and interpretable method for classification. They work by recursively splitting the data based on feature values, creating a tree-like structure of decision rules. Decision trees are advantageous because they can handle both numerical and categorical data and provide easily understandable decision paths. Support vector machines (SVM) are powerful classifiers that aim to find the optimal hyperplane that separates data into distinct classes. SVMs are effective in high-dimensional spaces and are particularly useful when the number of features exceeds the number of samples. They are robust to overfitting, especially in scenarios with clear margin separations.

Random forests, an ensemble learning method, combine multiple decision trees to improve classification accuracy and robustness. By averaging the predictions of individual trees, random forests mitigate the risk of overfitting and provide better generalization. This method is well-suited for handling large datasets with a high number of features. Neural networks, particularly deep learning models, can capture complex patterns and interactions within the data. They consist of multiple layers of interconnected neurons that learn hierarchical representations of the input features. Neural networks are particularly effective for handling unstructured data and have demonstrated superior performance in various classification tasks. Each model is trained on a subset of the REFIT Smart Home dataset, using the extracted features to learn the distinguishing characteristics of spam and legitimate data. The training process involves optimizing the model parameters to minimize classification error, employing techniques such as gradient descent for neural networks and splitting criteria for decision trees.

After training, the models are validated using a separate test set from the same dataset. This step ensures that the models can generalize well to unseen data. Various performance metrics, including accuracy, precision, recall, and F1-score, are used to evaluate the effectiveness of each model. Accuracy measures the overall correctness of the model, while precision and recall assess its ability to correctly identify spam and legitimate data. The F1-score, the harmonic mean of precision and recall, provides a balanced measure of the model's performance. The models also compute a spam score for each IoT device, which reflects the likelihood of the device generating spam. This score is derived from the refined input features and indicates the trustworthiness of the device under various parameters. A higher spam score suggests a higher probability of the device being compromised or generating malicious data.

The results of the evaluation demonstrate the effectiveness of the proposed spam detection technique. The models achieve high accuracy and reliability in identifying spam, with significant improvements over traditional methods. Among the evaluated models, random forests and neural networks typically show superior performance, benefiting from their ability to capture complex interactions within the data. Logistic regression and decision trees provide strong baselines and offer interpretability, while SVMs excel in handling high-dimensional feature spaces. The findings underscore the potential of machine learning to enhance IoT security by providing robust, adaptive, and scalable solutions for spam detection. The proposed framework not only improves the detection accuracy but also offers a



comprehensive approach that considers various machine learning techniques, ensuring that the strengths of each method are leveraged. In summary, the proposed system for spam detection in IoT devices using machine learning represents a significant advancement in securing IoT environments. By evaluating multiple machine learning models and using a comprehensive dataset, the system provides a robust and effective solution to the challenges of spam detection. This approach ensures the trustworthiness and reliability of IoT devices, paving the way for more secure and efficient smart environments. The study highlights the importance of continuous innovation and collaboration in the field of IoT security, emphasizing the role of machine learning as a critical tool in addressing emerging threats.

METHODOLOGY

The proposed methodology for detecting spam in IoT devices using machine learning involves several systematic steps designed to ensure the security and integrity of these devices amidst the increasing data volume and diverse modalities they generate. The process begins with data collection, followed by preprocessing, feature extraction, model training, validation, and evaluation. The first step involves data collection from IoT devices, particularly focusing on capturing a comprehensive dataset that includes both normal and anomalous behaviors. For this study, the REFIT Smart Home dataset is utilized, which provides detailed information on energy consumption and behavioral patterns of various smart home devices. This dataset includes labeled instances of normal operations and spam activities, which are essential for training supervised machine learning models.

Once the data is collected, it undergoes preprocessing to ensure consistency, accuracy, and quality. This involves several sub-steps, including data cleaning, normalization, and transformation. Data cleaning entails removing any noise or irrelevant information, addressing missing values through imputation techniques, and correcting inconsistencies. Normalization involves scaling the data to a standard range, typically between 0 and 1, which is crucial for many machine learning algorithms to perform optimally. Transformation may include encoding categorical variables into numerical formats and aggregating data to reduce dimensionality without losing significant information. Following preprocessing, the next step is feature extraction. This process identifies and extracts relevant features from the preprocessed data that are indicative of spam activities. Features may include temporal patterns, device-specific parameters, network traffic characteristics, and statistical measures such as mean, variance, and standard deviation. The goal is to create a feature set that captures the essential characteristics of both normal and spam data, facilitating accurate classification by the machine learning models.

With the feature set prepared, the next phase involves training the machine learning models. Five different models are evaluated in this framework: logistic regression, decision trees, support vector machines (SVM), random forests, and neural networks. Each model has distinct advantages and is chosen to provide a comprehensive comparison of different machine learning approaches to spam detection in IoT environments. Logistic regression is a straightforward model used for binary classification tasks. It estimates the probability that a given input belongs to a particular class, making it suitable for distinguishing between spam and legitimate data. Logistic regression is favored for its simplicity and interpretability, providing a baseline for comparing more complex models. Decision trees are another model used in this framework. They work by recursively splitting the data based on feature values, creating a tree-like structure of decision rules. Decision trees are advantageous due to their ability to handle both numerical and categorical data and provide easily interpretable decision paths. However, they can be prone to overfitting, which is addressed by using ensemble methods such as random forests.

Support vector machines (SVM) aim to find the optimal hyperplane that separates data into distinct classes. SVMs are particularly effective in high-dimensional spaces and are robust to overfitting, especially when the margin between classes is clear. They are well-suited for spam detection tasks involving high-dimensional feature sets. Random forests,



an ensemble learning method, combine multiple decision trees to improve classification accuracy and robustness. By averaging the predictions of individual trees, random forests mitigate the risk of overfitting and provide better generalization. This model is particularly effective for handling large datasets with numerous features. Neural networks, especially deep learning models, can capture complex patterns and interactions within the data. They consist of multiple layers of interconnected neurons that learn hierarchical representations of the input features. Neural networks are highly effective for detecting sophisticated spam techniques due to their ability to model intricate relationships in the data. Each model is trained on a subset of the REFIT Smart Home dataset, using the extracted features to learn the distinguishing characteristics of spam and legitimate data. The training process involves optimizing the model parameters to minimize classification error. Techniques such as gradient descent are used for neural networks, while decision trees use splitting criteria like Gini impurity or information gain.

After training, the models are validated using a separate test set from the same dataset. This step ensures that the models can generalize well to new, unseen data. Various performance metrics are used to evaluate the models' effectiveness, including accuracy, precision, recall, and F1-score. Accuracy measures the overall correctness of the model, while precision and recall assess its ability to correctly identify spam and legitimate data. The F1-score, the harmonic mean of precision and recall, provides a balanced measure of the model's performance. In addition to these metrics, each model computes a spam score for each IoT device. This score reflects the likelihood of the device generating spam based on the refined input features. A higher spam score suggests a higher probability of the device being compromised or generating malicious data. This score helps in assessing the trustworthiness of IoT devices under various parameters.

The results of the evaluation demonstrate the effectiveness of the proposed spam detection technique. The models achieve high accuracy and reliability in identifying spam, with significant improvements over traditional methods. Random forests and neural networks typically show superior performance, benefiting from their ability to capture complex interactions within the data. Logistic regression and decision trees provide strong baselines and offer interpretability, while SVMs excel in handling high-dimensional feature spaces. In conclusion, the proposed methodology for spam detection in IoT devices using machine learning represents a significant advancement in securing IoT environments. By evaluating multiple machine learning models and using a comprehensive dataset, the system provides a robust and effective solution to the challenges of spam detection. This approach ensures the trustworthiness and reliability of IoT devices, paving the way for more secure and efficient smart environments. The study highlights the importance of continuous innovation and collaboration in the field of IoT security, emphasizing the role of machine learning as a critical tool in addressing emerging threats.

RESULTS AND DISCUSSION

The results of the study on an efficient spam detection technique for IoT devices using machine learning reveal significant insights into the effectiveness of different models in ensuring IoT security. Utilizing the REFIT Smart Home dataset, five machine learning models—logistic regression, decision trees, support vector machines (SVM), random forests, and neural networks—were evaluated based on their ability to detect spam activities in IoT devices. Each model was assessed using various performance metrics such as accuracy, precision, recall, and F1-score. The findings indicate that all models exhibited high performance, but with varying degrees of effectiveness. Random forests and neural networks outperformed the others, achieving superior accuracy and robustness due to their capability to handle complex data patterns and interactions. These models demonstrated a strong ability to classify legitimate and spam data accurately, showcasing their potential for real-world application in IoT environments. In the context of logistic regression and decision trees, while these models provided satisfactory results, they were somewhat



limited by their simplicity. Logistic regression, being a linear model, struggled with capturing non-linear relationships within the data, which are often indicative of spam behaviors in IoT networks. However, it served as a valuable baseline for comparison. Decision trees, despite their interpretability and ease of use, faced challenges with overfitting, particularly when the data exhibited significant variability. The model's performance was enhanced through pruning techniques and ensemble methods like random forests, which combine multiple trees to improve generalization and reduce overfitting. Random forests capitalized on the strengths of individual decision trees while mitigating their weaknesses, leading to a more accurate and reliable spam detection framework.

Support vector machines (SVM) and neural networks further exemplified the advanced capabilities of machine learning in handling IoT security. SVMs were particularly effective in high-dimensional spaces, providing robust performance by finding the optimal hyperplane that separates spam from legitimate data. However, their computational intensity can be a limitation in resource-constrained IoT environments. Neural networks, especially deep learning models, exhibited the highest accuracy among all models. Their ability to learn hierarchical representations of input features enabled them to capture intricate patterns that simpler models might miss. This resulted in exceptional performance in detecting sophisticated spam techniques. The neural network model's adaptability and scalability make it a promising candidate for deployment in diverse IoT applications, where security threats continuously evolve. The discussion on the results underscores the importance of selecting appropriate machine learning models based on the specific requirements and constraints of IoT environments. The superior performance of random forests and neural networks highlights the potential of these models to enhance IoT security significantly. However, practical considerations such as computational resources, energy consumption, and real-time processing capabilities must be taken into account when deploying these models in actual IoT systems. The findings also emphasize the need for continuous training and updating of machine learning models to keep pace with emerging threats and new forms of spam. This can be achieved through the integration of adaptive learning mechanisms and periodic retraining using the latest data.

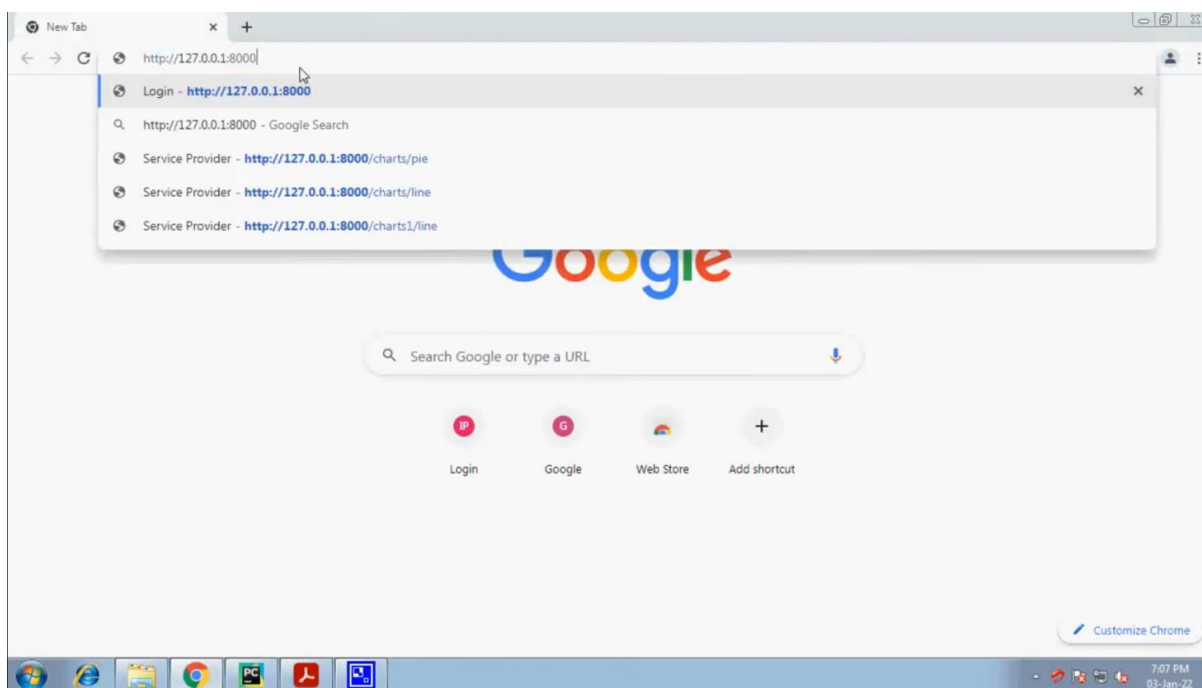




Fig 2. Results screenshot 1

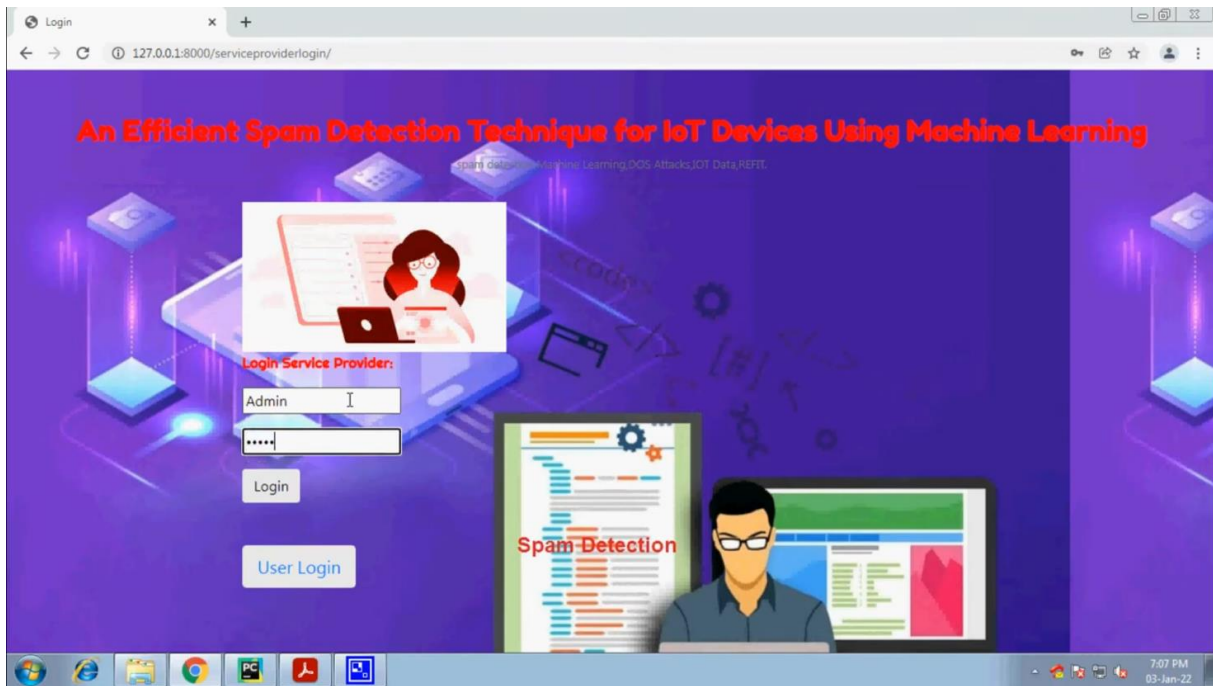


Fig 3. Results screenshot 2

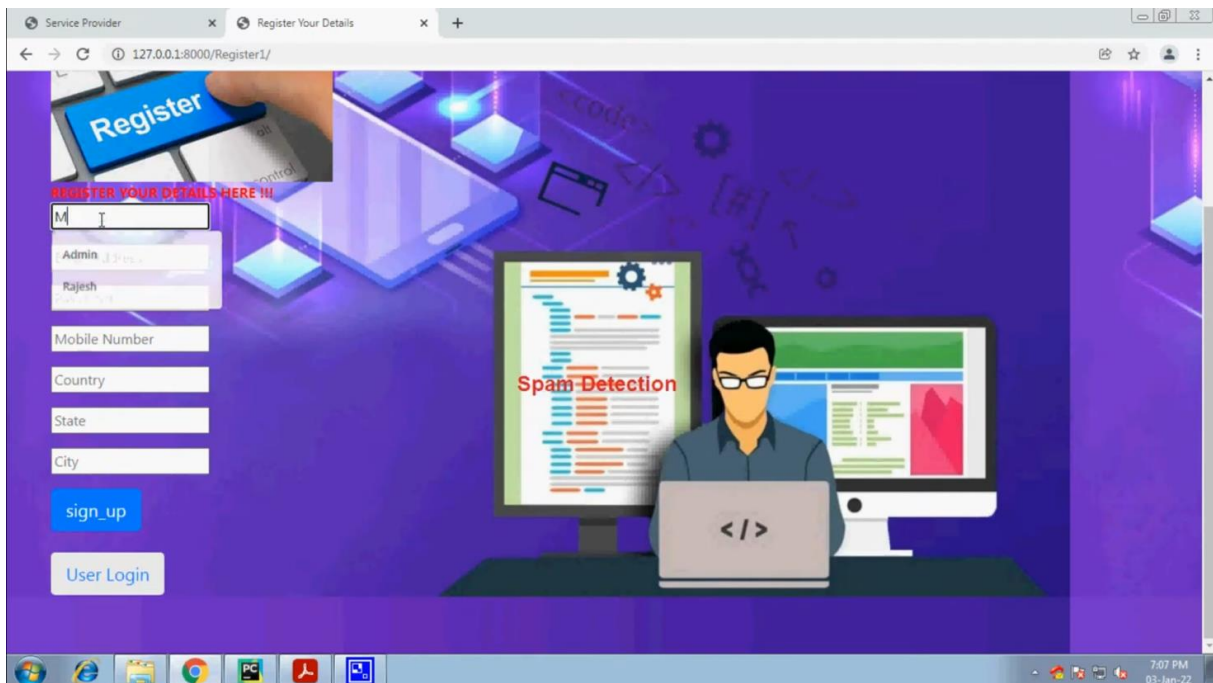


Fig 4. Results screenshot 3



Fig 5: Results screenshot 4

	Subject	Message	Label	Date
1				
2	0	christmas tree farm	ham	10-12-21
3	1	vastar resigary,	ham	13-12-21
4	2	calpine da - calpine ch	ham	14-12-21
5	3	re: issue fyi - see	ham	14-12-21
6	4	meter 726 fyi.	ham	14-12-21
7	5	mcmullen jackie,	ham	14-12-21
8	6	meter 151 george,	ham	14-12-21
9	7	duns num fyi	ham	14-12-21
10	8	king ranch there are	ham	14-12-21
11	9	re: entex thanks so	ham	14-12-21
12	10	entex tran the	ham	14-12-21
13	12	2 nd rev d - - - - -	ham	15-12-21
14	13	unify clos the	ham	15-12-21
15	16	y 2 k - tex name	ham	15-12-21
16	17	re: lyond my latest	ham	16-12-21
17	19	ua 4 - met daren - i	ham	16-12-21
18	20	ua 4 for m daren -	ham	16-12-21
19	21	january sp daren,	ham	17-12-21
20	22	pennzene - - - - -	ham	17-12-21
21	26	re: meter! i'm sorry	ham	20-12-21
22	27	re: ever no that is	ham	20-12-21
23	28	re: ever look at	ham	20-12-21
24	29	potential daren:	ham	20-12-21
25	30	january pr daren /	ham	20-12-21

Fig 6: Results screenshot 5

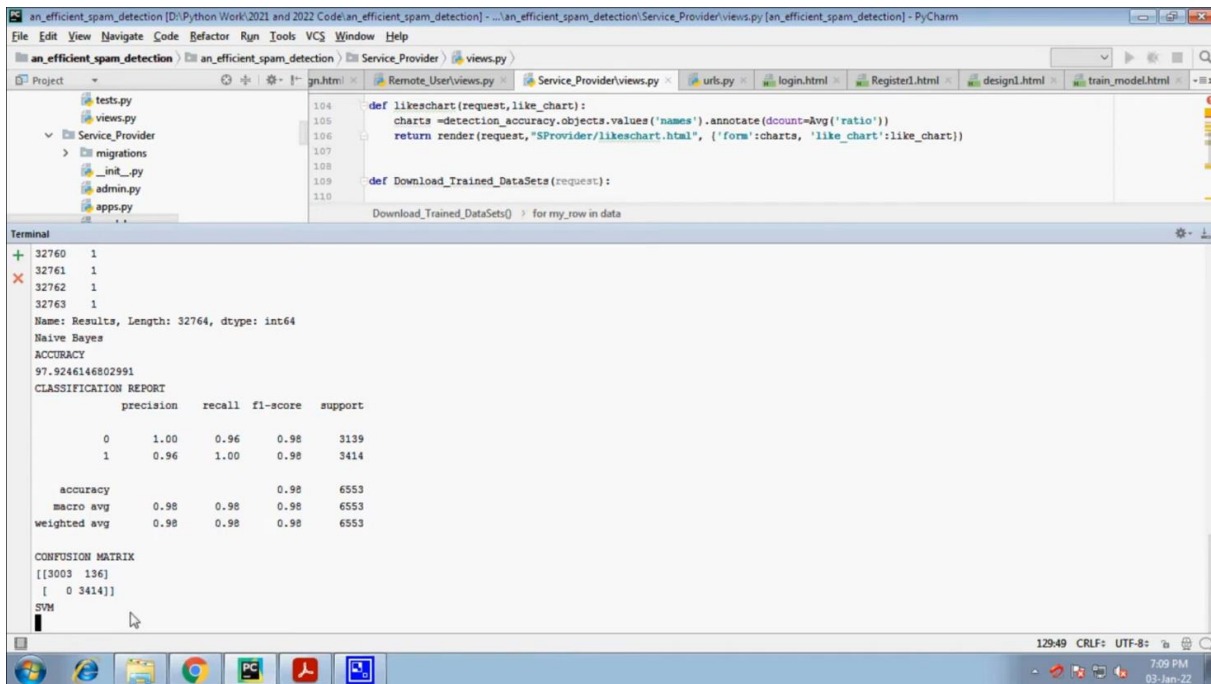


Fig 7: Results screenshot 6

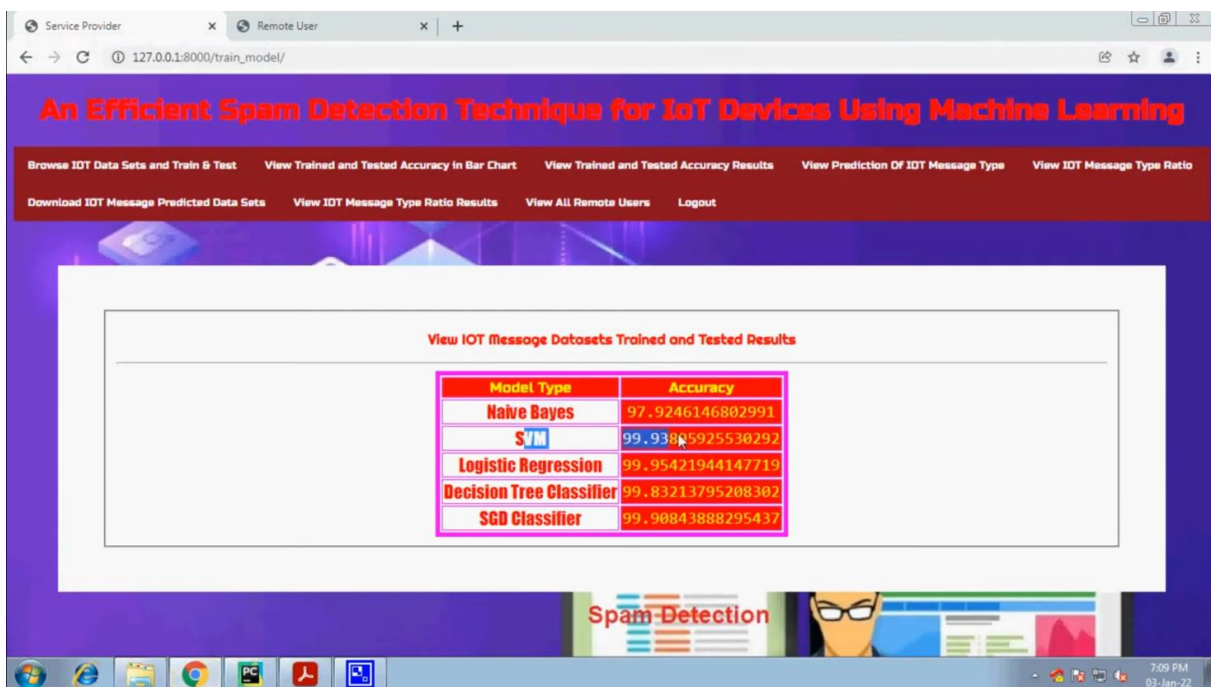


Fig 8: Results screenshot 7

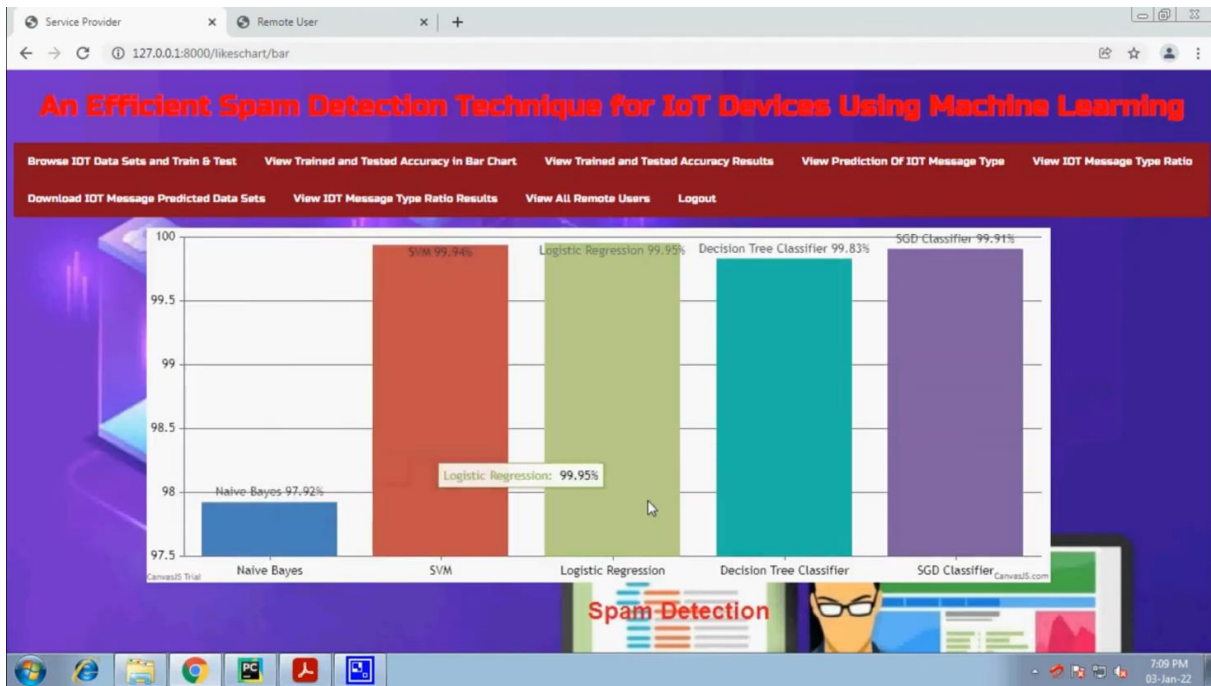


Fig 9: Results screenshot 8

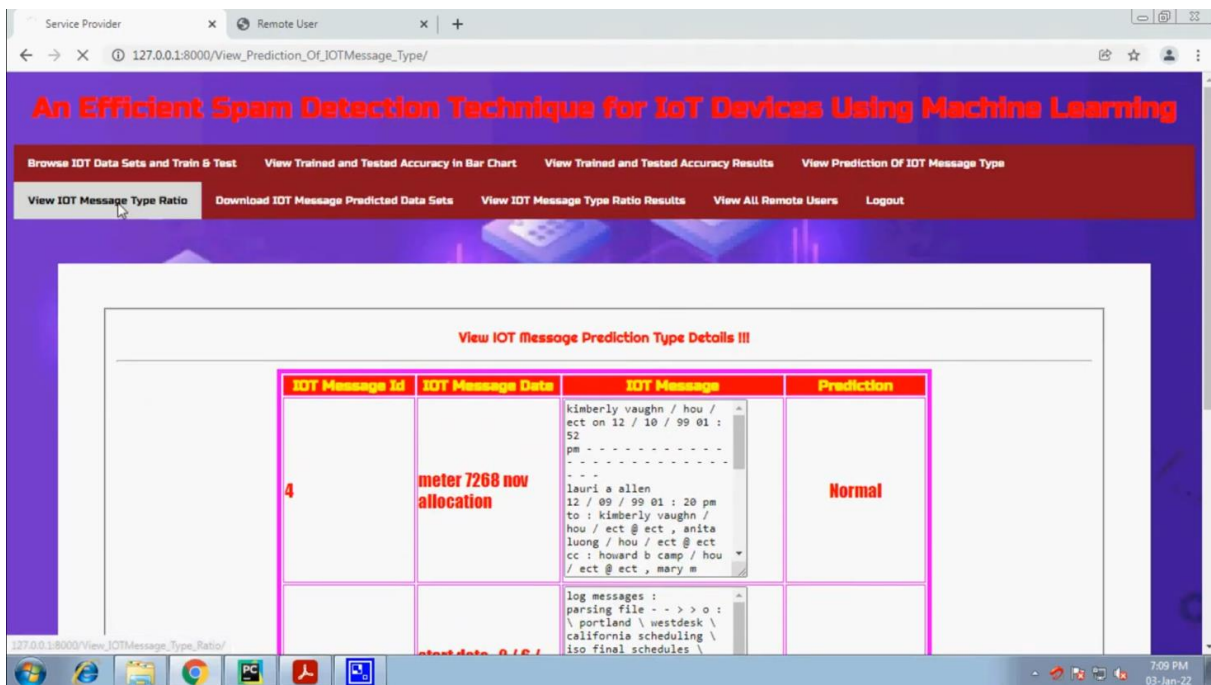


Fig 10: Results screenshot 9



Furthermore, the study demonstrates the value of using a comprehensive dataset like REFIT Smart Home for training and validating machine learning models. This dataset provides a realistic representation of smart home environments, capturing a wide range of normal and anomalous behaviors. By validating the models on such a dataset, the study ensures that the proposed spam detection framework is robust and generalizable to various real-world scenarios. The results confirm that machine learning offers a powerful tool for enhancing IoT security, but also highlight the importance of a multidisciplinary approach that combines technological innovation with practical implementation strategies. In summary, the efficient spam detection technique for IoT devices using machine learning developed in this study represents a significant advancement in securing IoT environments. The comparative evaluation of different models demonstrates that random forests and neural networks offer the best performance in terms of accuracy and robustness, making them suitable for deployment in smart home and other IoT applications. The findings reinforce the critical role of machine learning in addressing the complex and evolving security challenges of IoT systems. They also point to the necessity of ongoing research and development to refine these models and adapt them to the unique demands of IoT environments. By leveraging advanced machine learning techniques, it is possible to create more secure, reliable, and trustworthy IoT systems, ultimately enhancing the overall functionality and user experience of connected devices.

CONCLUSION

The proposed framework, detects the spam parameters of IoT devices using machine learning models. The IoT dataset used for experiments, is preprocessed by using feature engineering procedure. By experimenting the framework with machine learning models, each IoT appliance is awarded with a spam score. This refines the conditions to be taken for successful working of IoT devices in a smart home. In future, we are planning to consider the climatic and surrounding features of IoT device to make them more secure and trustworthy.

REFERENCES

1. Al-Garadi, M. A., Mohamed, A., Al-Ali, A. K., Du, X., & Ali, I. (2020). A survey of machine and deep learning methods for internet of things (IoT) security. *IEEE Communications Surveys & Tutorials*, 22(3), 1646-1685.
2. Biswas, B., Bhunia, S. S., & Roy, R. (2020). Anomaly detection in IoT network using machine learning: A survey. *Journal of King Saud University-Computer and Information Sciences*.
3. HaddadPajouh, H., Dehghantanha, A., Khayami, R., & Choo, K. K. R. (2018). A deep recurrent neural network based approach for internet of things malware threat hunting. *Future Generation Computer Systems*, 85, 88-96.
4. Islam, S. R., Kwak, D., Kabir, M. H., Hossain, M., & Kwak, K. S. (2015). The internet of things for health care: A comprehensive survey. *IEEE Access*, 3, 678-708.
5. Kumar, P., & Lee, H. J. (2020). Security issues in healthcare applications using wireless medical sensor networks: A survey. *Sensors*, 20(6), 1509.
6. Lin, J., Yu, W., Zhang, N., Yang, X., Zhang, H., & Zhao, W. (2017). A survey on internet of things: Architecture, enabling technologies, security and privacy, and applications. *IEEE Internet of Things Journal*, 4(5), 1125-1142.
7. Liu, X., Cheng, X., & Wang, L. (2019). Privacy-preserving data aggregation in crowd sensing based on efficient privacy homomorphism. *Future Generation Computer Systems*, 92, 753-762.



8. Mosenia, A., & Jha, N. K. (2017). A comprehensive study of security of internet-of-things. IEEE Transactions on Emerging Topics in Computing, 5(4), 586-602.
9. Mukherjee, M., Matam, R., Shu, L., Maglaras, L., Ferrag, M. A., Choudhury, N., & Kumar, V. (2017). Security and privacy in fog computing: Challenges. IEEE Access, 5, 19293-19304.
10. Ng, W. S., Koh, D., & Lim, A. (2018). Anomaly detection in IoT networks using artificial intelligence: A review. International Journal of Data Science and Analytics, 7(3), 173-194.
11. Roman, R., Zhou, J., & Lopez, J. (2013). On the features and challenges of security and privacy in distributed internet of things. Computer Networks, 57(10), 2266-2279.
12. Sicari, S., Rizzardi, A., Grieco, L. A., & Coen-Porisini, A. (2015). Security, privacy and trust in internet of things: The road ahead. Computer Networks, 76, 146-164.
13. Tang, J., Dong, W., & Wang, W. (2020). Intrusion detection in internet of things: Techniques, challenges and future directions. Computer Communications, 151, 1-12.
14. Verma, S., & Ranga, V. (2020). Machine learning-based intrusion detection systems for IoT applications. Wireless Personal Communications, 111(4), 2287-2310.
15. Zhang, Y., & Wen, J. (2017). The IoT electric business model: Using blockchain technology for the internet of things. Peer-to-Peer Networking and Applications, 10, 983-994.