© April 2025 | IJIRT | Volume 11 Issue 11 | ISSN: 2349-6002

Public And Private E-Mail Services for Spam Spoiler Detection System

Vinothini T[1], Syed Suhel A[2], Theju Kv[3], Yeshwanth Reddy L[4].

[1] *Assistant Professor,Department of Computer Science and Engineering,Adhiyamaan College of Engineering (Autonomous), Tamil Nadu, India.*

[2,3,4] *Department of Computer Science and Engineering, Adhiyamaan College of Engineering (Autonomous), Tamil Nadu, India.*

***Abstract*—In recent years, cyber security incidents have occurred frequently. In most of these incidents, attackers have used different types of spam email as a knock-on to successfully invade government systems, well-known companies, and websites of politicians and social organizations in many countries. This project proposed to design a novel efficient approach named Spam Spoiler for big e-mail data classification into four different classes: Normal, Fraudulent, Harassment, and Suspicious E-mails by using LSTM-based Algorithm**

I. INTRODUCTION

With the increasing reliance on email communication, cyber threats such as phishing, fraud, and harassment have become major concerns. Attackers frequently use spam emails as an entry point to compromise government systems, corporations, and social institutions, leading to severe financial and reputational risks. Traditional spam filters often fail to detect sophisticated threats, necessitating a more advanced approach.

To address this issue, this project proposes Spam Spoiler, an efficient LSTM-based deep learning model for classifying big email data into four categories: Normal, Fraudulent, Harassment, and Suspicious Emails. Leveraging LSTM's ability to analyze sequential patterns in text, this system aims to improve accuracy in detecting evolving spam threats.

By employing advanced text processing, feature extraction, and real-time classification, Spam Spoiler enhances email security while minimizing false positives. The model will be trained on a diverse dataset and evaluated using key performance metrics, with potential deployment in real-world email filtering systems to combat cyber threats effectively.

II. LITERATURE REVIEW

The literature survey provides an overview of recent research studies focusing on spam email detection using various machine learning techniques. These studies, published in the IEEE International Conference on Computing, explore different methodologies to improve spam filtering and cybersecurity.

In 2024, Pallavi Jain proposed a Natural Language Processing (NLP)-based machine learning approach for detecting spam emails. NLP techniques enable the system to analyze the content of emails, helping to differentiate between legitimate and spam messages more effectively. Another study by Lokaiah Pullagura (2024) examined various suspicious email detection techniques, focusing on identifying patterns that indicate fraudulent or harmful emails.

In 2023, Livia Shreenithi S. explored spam email detection using a Logistic Regression algorithm, demonstrating how traditional machine learning models can still be effective in filtering out unwanted emails. Meanwhile, Akanksha Dhar and K.O. Vedasree Anusha conducted a comparative analysis of multiple machines learning models, including Deep Learning, Support Vector Machine (SVM), Random Forest, and XGBoost. Their study aimed to evaluate and compare the efficiency of these models in detecting spam emails, providing insights into the strengths and weaknesses of different approaches.

Overall, these studies highlight the growing significance of machine learning and deep learning techniques in combating spam and cyber threats. By leveraging advanced classification methods, researchers aim to enhance email security, reduce false positives, and improve the accuracy of spam detection systems. The findings from this survey contribute to the development of more intelligent, adaptive, and efficient email filtering mechanisms for real-world applications.

LSTM-based models have demonstrated high accuracy in spam detection, outperforming traditional methods. However, Spam Spoiler aims to further improve classification performance by incorporating advanced pre-processing, word embeddings, and deep learning techniques to combat evolving cyber threats.

III. EXITING SYSTEM

The current spam email detection systems primarily rely on traditional rule-based filtering and machine learning classifiers to identify unwanted emails. These systems use predefined keyword lists, blacklists, and statistical models to differentiate between spam and legitimate emails. Rule-based filtering works by applying a set of predefined rules to detect spam based on keywords, sender reputation, and email structure. Similarly, blacklist and whitelist methods allow or block emails based on previously identified senders.

Machine learning classifiers such as Naïve Bayes, Support Vector Machines (SVM), and Random Forest analyze email content, metadata, and subject lines to identify spam patterns. Some systems also employ content-based filtering, where emails are scanned for suspicious words, links, or attachments. While these methods have been widely used, they suffer from significant limitations. Traditional spam filters often generate a high number of false positives, misclassifying legitimate emails as spam. Additionally, as cybercriminals constantly evolve their techniques, rule-based and conventional machine learning models struggle to adapt, leading to inefficiencies in detecting newly emerging threats.

IV. PROPOSED SYSTEM

To overcome the limitations of existing spam detection methods, this project introduces Spam Spoiler, a deep learning-based email classification system using Long Short-Term Memory (LSTM) networks. The proposed system aims to enhance spam detection accuracy by leveraging LSTM’s ability to analyze the sequential nature of email content and detect hidden patterns indicative of fraudulent, suspicious, and harmful emails. Unlike traditional rule-based filters and basic machine learning models, this system focuses on understanding the context of email text, making it more adaptable to evolving spam techniques.

The Spam Spoiler system will classify emails into four distinct categories: Normal, Fraudulent, Harassment, and Suspicious Emails. To achieve this, the system will preprocess email data using text-cleaning techniques, tokenize the content, and convert it into meaningful numerical representations using word embeddings. The LSTM-based model will then analyze these representations to identify subtle patterns that differentiate between legitimate and harmful emails.

By integrating deep learning with real-time email classification, the proposed system aims to minimize false positives and improve detection rates. The model will be trained on a diverse dataset to ensure robustness against new spam tactics, and performance will be evaluated using key metrics such as accuracy, precision, recall, and F1-score. Additionally, the system can be deployed as a scalable solution, capable of handling large email volumes efficiently.

V. METHODOLGY

The Spam Spoiler system follows a structured methodology to classify emails using an LSTM-based deep learning model. The process begins with data collection and preprocessing, where emails are cleaned, tokenized, and transformed into numerical representations using word embeddings like Word2Vec or GloVe. This step ensures that the text data is properly formatted for deep learning analysis.

Next, the LSTM model is trained to identify sequential patterns in email content, enabling it to distinguish between Normal,
Fraudulent, Harassment, and Suspicious Emails. The model undergoes optimization through hyperparameter tuning and dropout regularization to enhance accuracy and prevent overfitting.

Once trained, the model is evaluated using key performance metrics such as accuracy, precision, recall, and F1-score, ensuring its
effectiveness. Finally, the system is deployed for real-time email classification, offering a scalable and adaptive spam detection solution that continuously improves with new data.

VI. ARCHITECTURE DESIGN

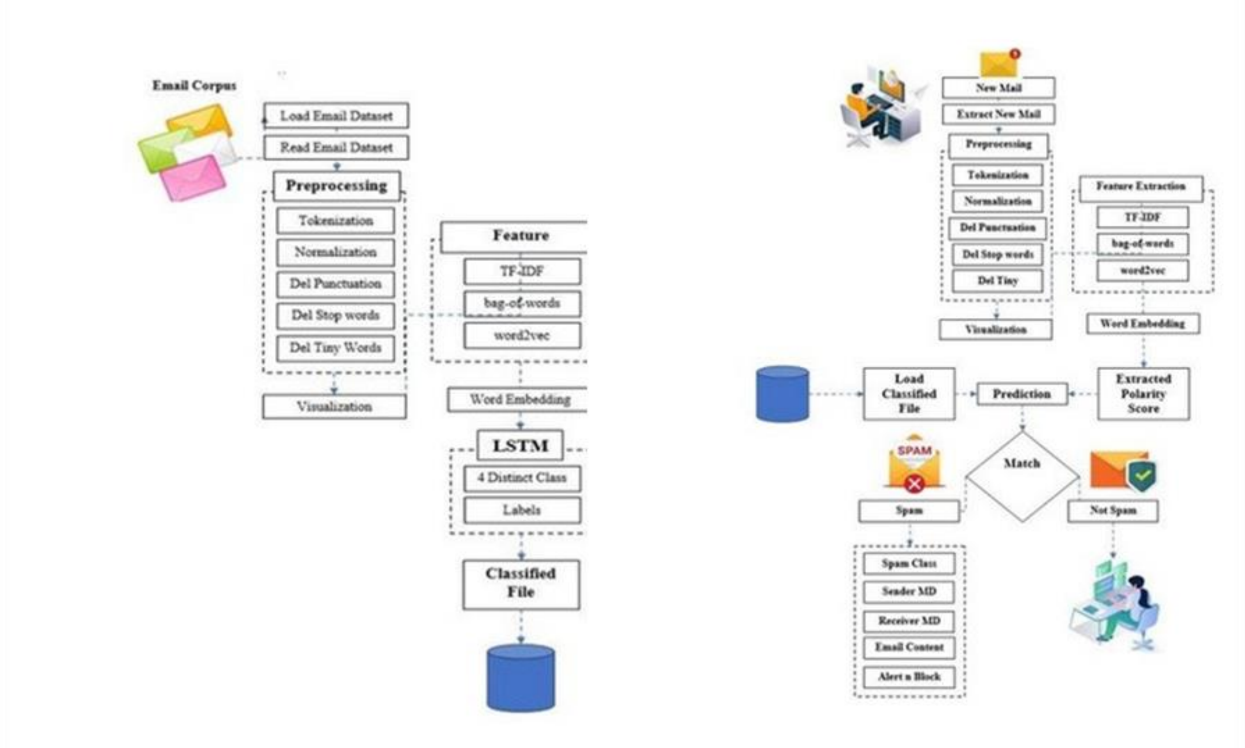
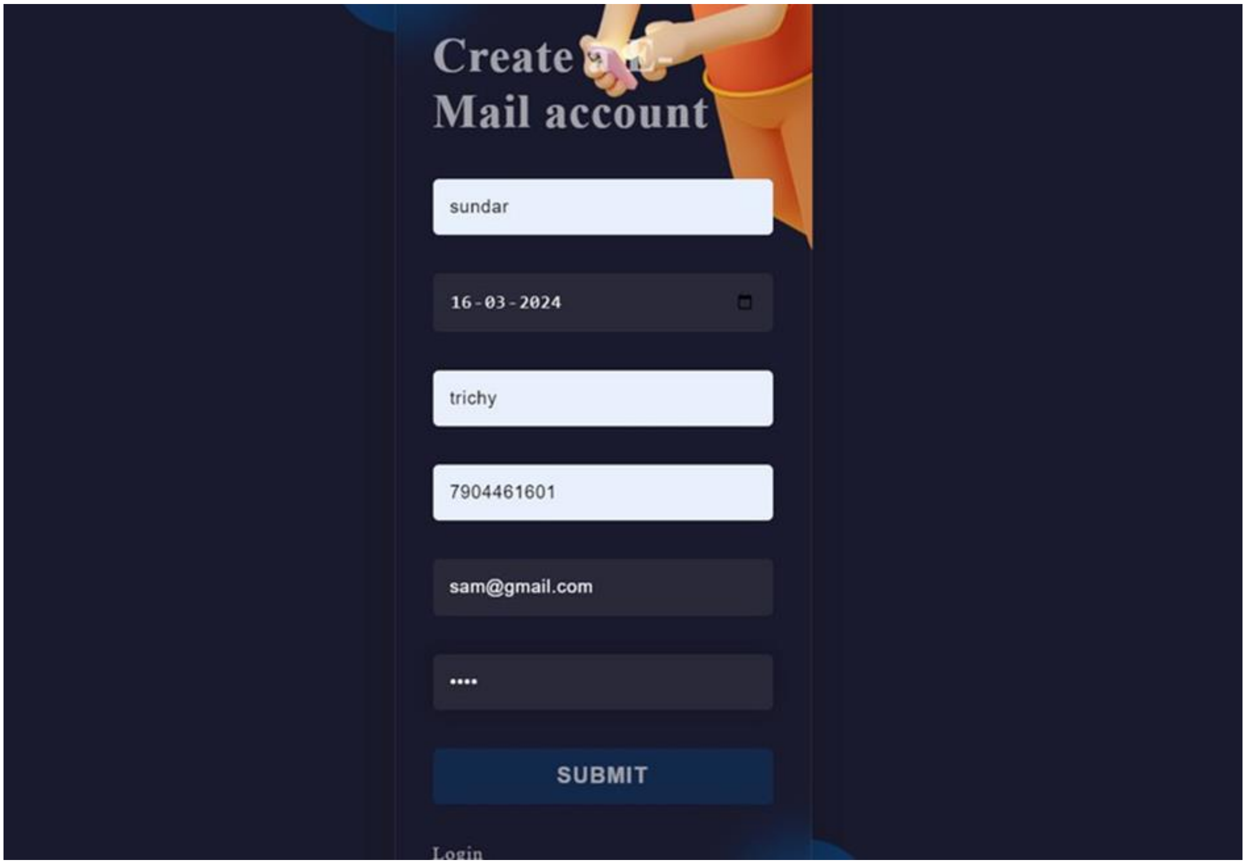


Fig 1. Architecture diagram

The architecture diagram represents the workflow of the Spam Spoiler system, divided into two main sections: the Training Phase and the Prediction Phase. In the training phase, the system starts by loading a dataset of emails, which undergo preprocessing steps such as tokenization, normalization, punctuation removal, stop-word removal, and tiny word filtering. After preprocessing, feature extraction techniques like TF-IDF, Bag-of-Words, and Word2Vec are applied, followed by word embedding to convert text into numerical representations. The LSTM-based deep learning model is then trained to classify emails into four categories: Normal, Fraudulent, Harassment, and Suspicious Emails. Once trained, the classified results are stored in a database.

In the prediction phase, when a new email is received, it undergoes the same preprocessing and feature extraction steps. The system then loads the classified dataset and uses the trained LSTM model to compare and classify the email. If the email matches spam characteristics, it is labeled as Spam; otherwise, it is marked as Not Spam. If detected as spam, additional details such as Sender ID, Receiver ID, and Email Content are analyzed, and appropriate actions like blocking or alerting the user are taken. This architecture ensures an efficient, real-time, and highly accurate spam detection system by leveraging deep learning techniques.

VII. RESULT OUTCOME



sundar

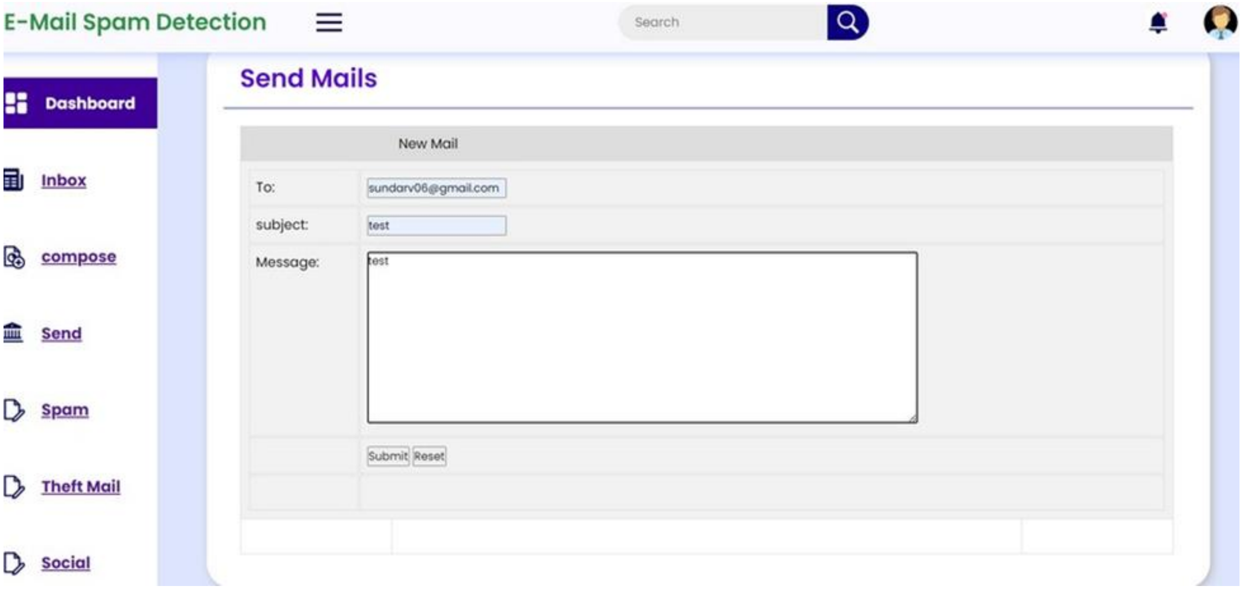
16-03-2024

trichy

7904461601

sam@gmail.com

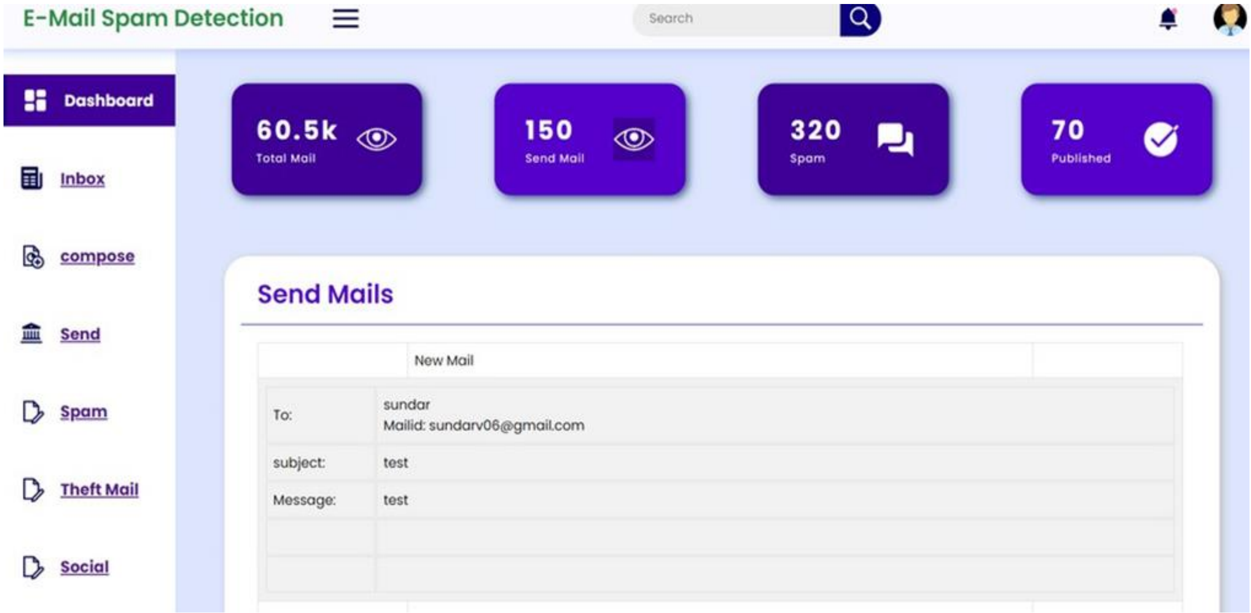
••••



To: sundarv06@gmail.com

subject: test

Message: test



To:	sundar Mailid: sundarv06@gmail.com
subject:	test
Message:	test

VIII. FUTURE SCOPE

The future scope of the Spam Spoiler system is vast, with several potential enhancements and applications. One major improvement is the integration of reinforcement learning and self-learning AI models that continuously adapt to evolving spam patterns, making detection more accurate and dynamic. The system can also be extended to include multimodal spam detection, incorporating not just email text but also attachments, images, and links to identify phishing attempts and malware.

Another promising direction is the real-time spam filtering system for enterprises and government organizations, ensuring immediate detection and prevention of cyber threats. Implementing cloud-based and distributed computing will enhance scalability, allowing the system to handle vast volumes of email traffic efficiently. Additionally, developing cross-platform compatibility for mobile and web applications will provide seamless spam protection across devices.

IX. CONCLUSION

The Spam Spoiler system provides an efficient and intelligent solution for detecting and classifying emails into Normal, Fraudulent, Harassment, and Suspicious categories using an LSTM-based deep learning model. By leveraging advanced text preprocessing, feature extraction, and sequential pattern recognition, the system significantly enhances the accuracy of spam detection. The integration of deep learning enables adaptive learning, ensuring the system stays effective against evolving spam techniques.

In conclusion, Spam Spoiler is a powerful and adaptable solution that not only enhances cybersecurity but also ensures a safer and more reliable communication environment.

REFERENCES

[1] S. Sinha, I. Ghosh, and S. C. Satapathy, ``A study for ANN model for spam classification," in Intelligent Data Engineering and Analytics. Singapore: Springer, 2021, pp. 331-343.

[2] Q. Li, M. Cheng, J. Wang, and B. Sun, ``LSTM based phishing detection for big email data," IEEE Trans. Big Data, early access, Mar. 12, 2020, doi: v10.1109/TBDATA.2020.2978915.

[3] T. Gangavarapu, C. D. Jaidhar, and B. Shanduka, ``Applicability of machine learning in spam and phishing email filtering: Review and approaches," Artif. Intell. Rev., vol. 53, no. 7, pp. 5019-5081, Oct. 2020, doi: 10.1007/s10462-020-09814-9.

[4] E. Bauer. 15 Outrageous Email Spam Statistics

That Still Ring True in 2018, RSS. Accessed: Oct. 10, 2020. [Online]. Available: https://www.propellercrm.com/blog/email-spam-ststatistics.

[5] Karim, S. Azam, B. Shanmugam, K. Kannoorpatti, and M. Alazab, ``A comprehensive survey for intelligent spam email detection," IEEE Access, vol. 7, pp. 168261-168295, 2019.