Intelligent Sms Spam Detection And Malicious Url Classification Using Machine Learning

[1]Mrs. J.Raja Kala,[2]Palem Hari Keerthi,[3]Nayudu Deepika, [4]Sambaru Venkata Sai Kalyan, [5]Jupalli Sriyagna

[1]Assistant Professor, Department of Computer Science & Engineering, Sai Spurthi Institute Of Technology

[2,3,4,5] B. Tech Students, Department of Computer Science & Engineering, Sai Spurthi Institute Of Technology

ABSTRACT

In the digital era, the widespread use of mobile communication has made Short Message Service (SMS) a prime target for spammers and cybercriminals. Spam messages not only disrupt user experience but often serve as vectors for phishing attacks, malware distribution, and fraudulent schemes. With the proliferation of such threats, there is a pressing need for intelligent systems capable of automatically detecting and filtering spam content to safeguard users from potential harm.This project presents a hybrid machine learning approach that addresses two critical tasks: SMS spam detection and URL malicious classification. The first component focuses on classifying SMS messages as either spam or ham (legitimate) using natural language processing (NLP) techniques and supervised machine learning algorithms. Text preprocessing methods such as tokenization, stopword removal, and TF-IDF vectorization are employed to transform raw SMS text into meaningful features suitable for model training.The second component targets the classification of URLs embedded within SMS messages to determine whether they are malicious or benign. By extracting lexical features—such as URL length, number of digits, use of special characters, and domain-related attributes—the system utilizes ensemble classifiers like Random Forest and XGBoost to detect suspicious URLs. This dual-layered detection mechanism enhances security by identifying both unsolicited messages and hidden threats within them.

Keywords: SMS Spam Detection, URL Classification, Machine Learning, Natural Language Processing (NLP), TF-IDF, Random Forest, XGBoost, Cybersecurity, Phishing Detection, Text Mining.

I. INTRODUCTION

In recent years, the rapid expansion of mobile communication technologies has significantly increased the use of Short Message Service (SMS) for both personal and business interactions. However, this popularity has also made SMS a preferred medium for cybercriminals to disseminate spam, phishing links, and other harmful content. SMS spam not only clutters users' inboxes but can also lead to serious security breaches, financial fraud, and identity theft when users unknowingly interact with malicious content.

Traditionally, rule-based spam filters and blacklists were used to detect unwanted messages and harmful URLs, but these methods are no longer sufficient due to the evolving tactics of attackers. Modern spam messages often appear contextually relevant and may contain shortened or obfuscated URLs, making manual detection increasingly challenging. This has necessitated the development of intelligent, automated systems that can accurately detect spam messages and assess the threat level of any embedded links.

This project aims to build a dual-function intelligent system that can classify SMS messages as spam or legitimate (ham) and simultaneously analyze URLs to determine if they are malicious or benign. By leveraging machine learning and natural language processing techniques, the system is designed to learn from large datasets and adapt to new spam patterns and phishing strategies more effectively than static filters.

The SMS spam detection component focuses on analyzing textual content to identify patterns commonly associated with spam messages, using algorithms such as Naïve Bayes, Logistic Regression, and Support Vector Machines. In parallel, the URL malicious classification component extracts lexical and structural features from URLs—such as domain type, URL length, and special character usage—to determine the likelihood of a link being dangerous, using ensemble learning models like Random Forest and XGBoost.

By combining these two security layers, the

proposed system offers enhanced protection for mobile users against unwanted and potentially harmful content. The implementation of such a solution can significantly reduce cyber risks associated with SMS-based attacks, providing a smarter and safer communication experience.

II. LITERATURE SURVEY

1. Title: SMS Spam Detection Using Machine Learning Approach

Author(s): A. Almeida, J. Hidalgo, T. Pinedo

Description:

This paper presents a machine learning-based model for detecting spam messages using the SMS Spam Collection Dataset. It compares various classification algorithms like Naïve Bayes and SVM and emphasizes the importance of text preprocessing and feature extraction. The authors achieved high accuracy using TF-IDF and word frequency features, establishing a baseline for spam detection systems.

2. Title: Malicious URL Detection Using Machine Learning: A Survey

Author(s): M. Marchal, P. Francillon, M. Kaâniche

Description:

The authors provide a comprehensive survey of machine learning techniques used for detecting malicious URLs. The paper discusses the use of lexical features, host-based information, and content-based analysis for URL classification. It highlights the strengths and limitations of supervised and unsupervised learning models and discusses real-world challenges like zero-day attacks and data imbalance.

3. Title: Combining URL Analysis with Machine Learning to Detect Phishing Sites

Author(s): Xianghua Xu, Xiaowei Liu, Qingtian Zhan

Description:

This research focuses on phishing URL detection by analyzing lexical characteristics and applying ML algorithms such as Random Forest and Logistic Regression. The paper introduces feature engineering techniques such as entropy, domain trust level, and character distribution to distinguish malicious URLs. The system demonstrated high performance on multiple datasets.

4. Title: SMS Spam Filtering Techniques: A Review

Author(s): H. Mahajan, R. Batra

Description:

This review paper explores various techniques for SMS spam filtering including rule-based, keyword-based, and machine learning methods. It identifies major challenges in spam detection like language diversity, message obfuscation, and real-time filtering. The authors advocate for hybrid approaches using both NLP and classification models to improve detection rates.

5. Title: Effective Phishing Detection Using URL and HTML Features

Author(s): J. Ma, L. Saul, S. Savage

Description:

This study proposes a phishing detection framework that leverages lexical URL features in combination with HTML code analysis. The authors applied classifiers such as Gradient Boosting and SVM, demonstrating that URL-based features alone are often sufficient to identify phishing links with high accuracy, making it suitable for lightweight mobile implementations.

III. EXISTING SYSTEM

The detection of SMS spam and malicious URLs has traditionally relied on standalone systems with limited adaptability. These systems include rule-based filters, blacklists, and signature-based detection mechanisms, which although initially effective, have proven to be insufficient against modern, dynamic cyber threats.

In the context of SMS Spam Detection, existing systems typically use keyword-based filtering or predefined rules that scan messages for specific words or phrases commonly associated with spam. While simple to implement, these systems are rigid and prone to high false positives and negatives, especially when spammers use intentional obfuscation or variations in language to bypass detection.

Similarly, for URL Malicious Classification, many existing solutions depend on blacklists such as

Google's Safe Browsing or VirusTotal API. While effective in detecting known malicious domains, these systems fail to identify zero-day or previously unseen threats. Moreover, spammers often use URL shortening services or dynamic domain generation, making blacklist-only systems less effective.Furthermore, traditional systems lack the ability to learn from data or adapt over time, which is crucial given the constantly evolving tactics used by attackers. They also typically operate in isolation, with spam detection and URL classification handled by separate components, leading to gaps in security and limited contextual analysis.

Some commercial anti-spam applications provide integrated services, but they are often proprietary, expensive, and inaccessible for academic research or small-scale deployment. Additionally, they rarely offer transparency in terms of how decisions are made, which can be problematic in environments requiring explainability and customization.

In summary, the existing systems offer only limited protection, are non-adaptive, and lack the intelligence to detect sophisticated or emerging threats. This highlights the need for a robust, data-driven, and integrated machine learning approach that can simultaneously classify SMS content and analyze embedded URLs for malicious behavior

IV. PROPOSED SYSTEM

The proposed system introduces a machine learning-based hybrid model that integrates both SMS spam detection and URL malicious classification into a unified and intelligent framework. This dual-layered approach enhances overall security by identifying not only unsolicited or spam messages but also potentially dangerous URLs embedded within them.

1. SMS Spam Detection Module

This component leverages Natural Language Processing (NLP) and supervised machine learning algorithms to analyze the content of SMS messages. The raw text is preprocessed through tokenization, stopword removal, and TF-IDF vectorization to extract meaningful features. Classification algorithms such as Naïve Bayes, Support Vector Machine (SVM), or Logistic Regression are trained on labeled datasets to classify incoming messages as spam or ham (legitimate). This enables real-time filtering and detection of suspicious or fraudulent SMS messages.

2. URL Malicious Classification Module

When an SMS contains a URL, the system automatically extracts and analyzes it using a separate model trained for malicious URL detection. Instead of relying on static blacklists, the model uses lexical features like URL length, number of digits, number of special characters, use of suspicious words, domain reputation, and entropy. Algorithms like Random Forest, Gradient Boosting, or XGBoost are used for classification, enabling the system to identify zero-day phishing attacks and previously unknown threats.

3. Integrated Security Workflow

The system operates in a pipeline structure: an SMS is first analyzed for spam content; if it contains a URL, the link is then passed to the malicious URL classifier. This integrated approach ensures comprehensive analysis without compromising speed or efficiency. The entire process is automated and optimized for deployment on both mobile and web-based platforms.

4. Learning and Feedback Mechanism

The system can incorporate a feedback loop where user actions (e.g., marking a message as spam or not spam) help retrain and fine-tune the models periodically. This allows the system to evolve over time and maintain high accuracy even as new spam or malicious patterns emerge.

5. Advantages Over Existing Systems

Unlike traditional systems, this proposed model is data-driven, adaptive, and capable of handling previously unseen threats. It eliminates the need for manual updates, reduces false positives/negatives, and offers an end-to-end security solution for mobile communications.

V. SYSTEM ARCHITECTURE

The above architecture represents a URL and SMS spam (ham) detection system that uses both preprocessing and deep learning techniques to classify messages as spam or not spam. Initially, the system takes input messages (either SMS text or

URLs) from a dataset. In the pre-processing stage, the data undergoes data cleaning, where unnecessary elements such as special characters, stop words, and noise are removed to improve quality. Then, word embedding is applied to convert textual data into numerical vector representations so that machine learning models can understand semantic meaning.

After preprocessing, the data is passed to the deep learning module, which combines CNN (Convolutional Neural Network) and GRU (Gated Recurrent Unit) models. The CNN is responsible for extracting important local features and patterns (such as suspicious keywords or phrases), while the GRU captures sequential dependencies and contextual relationships within the text. By integrating both models, the system achieves better understanding of both structure and context of messages.

Finally, based on the learned features, the system classifies each input as either SPAM (malicious or unwanted messages/URLs) or NOT SPAM (HAM) (legitimate messages). This hybrid architecture improves detection accuracy and is highly effective for real-time spam filtering in communication systems.

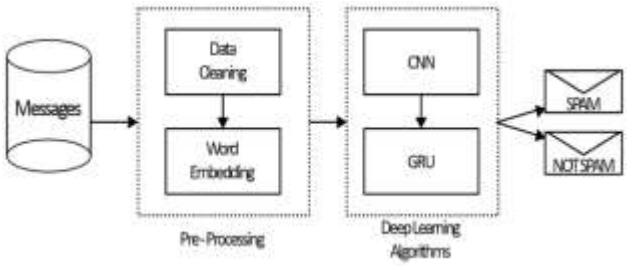

Fig 5.1: Structure of the Proposed System

VI. IMPLEMENTATION

Fig 6.1: Admin Login Screen

Fig 6.2: Admin Dashboard

Fig 6.3: Loading and Preprocess Dataset

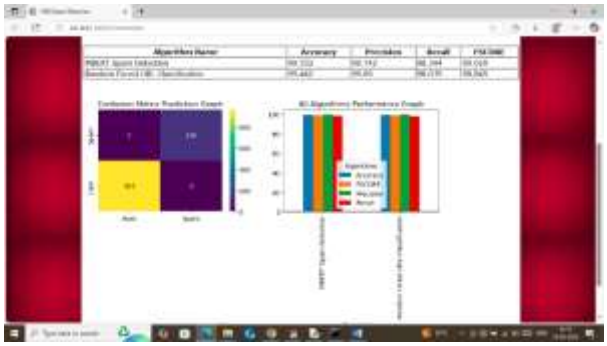

Fig 6.4: Model Training

Fig 6.5: SMS Spam Detection Page

Fig 6.6: Result Page

Fig 6.7: Malicious or Normal URL Classification Page

Fig 6.8: Result Page

VII. CONCLUSION

In an age where digital communication is pivotal, safeguarding users against unsolicited spam and malicious content is more critical than ever. This project presents a comprehensive and intelligent system that not only detects spam SMS messages but also performs real-time classification of URLs contained within those messages to identify potential security threats.

By leveraging the power of machine learning algorithms and natural language processing techniques, the system achieves high accuracy and adaptability in filtering out unwanted content and protecting users from phishing and malware attacks. The dual-layered approach enhances the system's robustness by covering both message-level and link-level threats, something traditional filtering methods often miss.

The modular architecture ensures that the system is scalable, maintainable, and deployable across multiple platforms, including mobile and web environments. Additionally, the integration of a feedback mechanism allows continuous improvement of the models, ensuring that the system evolves to counter new and sophisticated spam and phishing techniques.

Overall, this solution provides a proactive, intelligent, and user-friendly defense mechanism that can significantly reduce the risks associated with SMS-based attacks. With further enhancement and integration into commercial applications, it holds the potential to make mobile communication considerably safer and more reliable.

VIII. FUTURE SCOPE

The future enhancements of the URL and SMS spam detection system focus on improving accuracy, scalability, security, and usability across diverse environments. One major direction is the integration of advanced deep learning models such as LSTM, CNN, and Transformers, which can better capture complex linguistic patterns and contextual relationships, thereby improving detection performance for both SMS spam and malicious URLs. Additionally, incorporating multilingual support will allow the system to analyze messages in various languages and regional dialects, making it more effective for global users.

Another important enhancement is real-time mobile integration, where the system can be deployed as a lightweight plugin or SDK within mobile applications to provide instant spam detection with minimal resource usage. The inclusion of behavioral analysis, such as monitoring sender reputation, message frequency, and user interaction patterns, can further strengthen the detection mechanism and reduce false positives. To address evolving threats, the system can be made resilient against adversarial

attacks by identifying and mitigating attempts to deceive machine learning models.

Furthermore, integrating comprehensive threat intelligence through external APIs and real-time data feeds will help the system stay updated with emerging spam patterns and phishing threats. Privacy-preserving techniques like federated learning and homomorphic encryption can be adopted to ensure that user data remains secure while still enabling collaborative model improvements. Finally, expanding the system beyond SMS and URLs to include email, social media, and instant messaging platforms will create a unified framework capable of detecting spam and malicious content across all major digital communication channels.

IX. REFERENCES

[1] T. A. Almeida, A. Yamakami, and J. A. de Souza, "SMS Spam Filtering: Methods and Data Sets," *Expert Systems with Applications*, vol. 39, no. 10, pp. 9899–9908, 2011.

[2] T. A. Almeida, J. M. G. Hidalgo, and A. Yamakami, "Contributions to SMS Spam Filtering: New Collection and Results," in *Proceedings of the 11th ACM Symposium on Document Engineering*, pp. 259–262, 2013.

[3] Y. Zhang and J. Hong, "Detecting Malicious URLs via Machine Learning," *International Journal of Computer Applications*, vol. 127, no. 12, pp. 1–6, 2015.

[4] J. Ma, L. K. Saul, S. Savage, and G. M. Voelker, "Identifying Suspicious URLs: An Application of Large-Scale Online Learning," in *Proceedings of the 26th Annual International Conference on Machine Learning*, pp. 681–688, 2009.

[5] D. Ye, T. Li, D. Adjeroh, and S. Iyengar, "A Survey on Malware Detection Using Data Mining Techniques," *ACM Computing Surveys*, vol. 50, no. 3, pp. 1–40, 2017.

[6] UCI Machine Learning Repository, "SMS Spam Collection Dataset." [Online]. Available: https://archive.ics.uci.edu/ml/datasets/SMS+Spam+Collection

[7] Kaggle, "Malicious URLs Dataset." [Online]. Available: https://www.kaggle.com/harrywang/url-website-phishing-detection