SMS SPAM DETECTION & URL MALICIOUS CLASSIFICATION

Mr. SHAIK HIMAM BASHA[1], INAKOLLU VENKATA LIKHITHA[2]
1.Assistant Professor, #2. Pursuing M.C.A Department of Master of Computer Application
QIS College of Engineering &Technology
Vengamukkapalem(V), Ongole, Prakasam dist., Andhra Pradesh-523272

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.

Evaluation of both models was performed using publicly available datasets, and the results demonstrated high accuracy, precision, and recall, proving the effectiveness of the proposed approach. The integration of spam detection with malicious URL classification provides a more robust solution compared to traditional standalone filters, significantly reducing the risk of user exploitation.

Overall, this project contributes a comprehensive solution for enhancing digital communication security. It can be deployed in mobile applications, messaging platforms, or enterprise systems to provide real-time protection against spam and malicious attacks, thereby fostering a safer messaging ecosystem for users.

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.

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, QingtianZhan
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.

6. Title: "LSTM-Based SMS Spam Detection"Authors**:** S. Hochreiter, J. Schmidhuber (1997)
Description:
• Applied recurrent neural networks for sequence modeling.
• Captured contextual dependencies in SMS text.
• Improved classification of ambiguous messages.

7. Title: "A Hybrid Approach for SMS Spam Detection"
Authors: K. S. Adewole, A. Azeta (2019)
Description:
• Combined TF-IDF with Random Forest classifier.
• Enhanced robustness to noisy text.
• Reduced misclassification rate.

8. Title: “Machine Learning Approaches for URL Classification”

Authors: J. Ma, L. Saul, S. Savage (2009)
Description**:**

- Proposed lexical feature-based URL detection.
- Used machine learning for malicious URL prediction.
- Demonstrated real-time detection capability.

9. Title: “Beyond Blacklists: Learning to Detect Malicious URLs”

Authors**:** J. Ma, L. Saul, S. Savage (2009)
Description**:**

- Highlighted limitations of blacklist systems.
- Proposed predictive ML model.
- Improved early detection of phishing links.

10. Title: “Phishing URL Detection Using Machine Learning”

Authors**:** A. A. Abdelhamid, A. Ayesh, F. Thabtah (2014)
Description:

- Used URL-based features like length and domain age.
- Applied Decision Tree and SVM.
- Achieved high phishing detection accuracy.

SYSTEM ANALYSIS

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.

Disadvantages of Existing Systems

- Lack of Adaptability
  Traditional rule-based and keyword-based filters cannot adapt to new types of spam or malicious content. Attackers frequently change message patterns and obfuscate URLs, which static systems fail to detect effectively.
- High False Positive and False Negative Rates
  Existing systems often incorrectly classify legitimate messages as spam (false positives) or fail to detect actual spam (false negatives), leading to poor user experience and security risks.
- Blacklist Dependence
  URL detection systems heavily rely on blacklists of known malicious domains. These lists are reactive and do not detect new or zero-day malicious URLs, allowing threats to bypass detection until they are reported and added to the list.
- No Contextual Understanding
  Rule-based systems do not understand the context or semantics of SMS content. For instance, a benign message with suspicious words may be flagged, while a cleverly worded phishing message may go undetected.
- Fragmented Approach
  Spam filtering and malicious URL detection are often handled by separate tools or services, resulting in fragmented protection. This lack of integration limits the system's ability to provide holistic threat analysis and response.
- Inflexibility and Maintenance Overhead
  Updating rules, blacklists, and signatures manually is time-consuming and labor-intensive. It also introduces human error and delays in adapting to new threats.
- Limited Scalability
  Many traditional systems are not designed to scale efficiently with increasing message volumes or URL traffic, which can degrade performance in large-scale applications.
- No Learning Mechanism
  Existing systems do not improve over time since they lack machine learning capabilities. They cannot learn from new patterns or user feedback, which limits long-term effectiveness.

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.

Advantages of the Proposed System

Dual-LevelProtection
The system provides comprehensive security by detecting both spam messages and malicious URLs within SMS content. This two-tiered defense mechanism significantly reduces the chances of phishing, fraud, and malware infections.

 Machine Learning-Based Accuracy
By using advanced machine learning algorithms and natural language processing, the system offers higher accuracy in classification compared to traditional rule-based or blacklist methods. It can detect subtle spam patterns and zero-day threats.

 Adaptive and Self-Improving
The system is designed to learn from new data through periodic retraining. This adaptive nature enables it to evolve with emerging trends in spam and phishing attacks, keeping detection effective over time.

 Reduced False Positives/Negatives
With intelligent feature extraction and model training, the system minimizes misclassification, leading to fewer false alarms and better user experience.

 Independence from Blacklists
Unlike traditional URL detection systems, this model does not rely solely on static blacklists. Instead, it uses URL feature analysis, making it more capable of identifying previously unknown or disguised malicious URLs.

 Scalable and Fast
The modular architecture allows the system to scale efficiently for large volumes of messages. It's also optimized for real-time processing, making it suitable for mobile and enterprise applications.

 Automated Workflow
The end-to-end process — from SMS analysis to URL verification — is fully automated, requiring minimal manual intervention. This ensures quick and reliable threat detection.

 Platform Agnostic Deployment
The solution can be integrated into various platforms such as mobile SMS apps, email gateways, and enterprise communication tools, enhancing its applicability in real-world scenarios.

 User Feedback Integration
The inclusion of a feedback loop allows users to flag incorrect classifications, which can be used to improve the model over time — making the system user-aware and continuously improving.

 Enhanced Privacy and Security
By processing data locally or in a secure cloud environment and not solely relying on third-party services, the system ensures better data privacy and information security for users.

Implementation

The implementation of the SMS Spam Detection and URL Malicious Classification System focuses on identifying spam messages and detecting malicious URLs using Machine Learning and Natural Language Processing (NLP) techniques. The system analyzes SMS content, sender behavior, and URL characteristics to classify messages as legitimate or spam and determine whether URLs are safe or malicious.

The proposed system helps protect users from phishing attacks, scams, malware distribution, and unwanted spam communications.

1. Data Collection

The first stage involves collecting SMS and URL datasets from various sources such as:

- Public SMS Spam Datasets
- Telecom Service Providers
- Cybersecurity Databases

- Phishing URL Repositories
- User Reports
- Web Traffic Logs

The collected dataset may include:

SMS Features

- Message Content
- Sender Information
- Message Length
- Special Characters
- Keyword Frequency
- Timestamp

URL Features

- URL Length
- Domain Name
- IP Address Usage
- HTTPS Availability
- Special Symbols in URL
- Domain Age
- Redirection Information

These attributes help identify spam messages and malicious websites.

2. Data Preprocessing

The collected SMS and URL data is cleaned and prepared before analysis.

Preprocessing Steps

SMS Data Processing

- Lowercase conversion
- Stop-word removal
- Tokenization
- Stemming and Lemmatization
- Noise removal

URL Data Processing

- URL normalization
- Removing duplicate URLs
- Domain extraction
- Feature encoding

This improves model accuracy and processing efficiency.

3. Feature Extraction

Important features are extracted from SMS messages and URLs.

SMS Features

Textual Features

- TF-IDF values
- Word frequency
- N-gram patterns
- Spam keywords

Behavioral Features

- Sender frequency
- Message timing
- Repeated patterns

URL Features

Structural Features

- URL length
- Number of dots and slashes
- Presence of suspicious characters

Domain Features

- Domain age
- WHOIS information
- SSL certificate status

Content Features

- JavaScript usage
- Embedded suspicious scripts
- Redirection behavior

Feature extraction improves spam and malicious URL detection performance.

4. Machine Learning Model Development

Machine Learning algorithms are used for SMS spam classification and malicious URL detection.

Algorithms Used

- Naive Bayes
- Logistic Regression
- Decision Tree
- Random Forest
- Support Vector Machine (SVM)
- XGBoost
- Artificial Neural Networks (ANN)

The models are trained using labeled spam/ham messages and malicious/benign URL datasets.

5. Natural Language Processing (NLP)

NLP techniques are used for analyzing SMS content.

NLP Functions

- Text classification
- Sentiment analysis
- Keyword extraction
- Context understanding
- Pattern detection

NLP improves the system's ability to understand spam message characteristics.

6. Model Training and Testing

The dataset is divided into:

- Training Dataset
- Validation Dataset
- Testing Dataset

Training Phase

The model learns patterns from historical spam messages and malicious URLs.

Testing Phase

The trained model is evaluated using unseen SMS and URL samples.

Performance metrics include:

- Accuracy
- Precision
- Recall
- F1-Score
- ROC-AUC Score
- Confusion Matrix

Methodology

The methodology of the proposed SMS Spam Detection and URL Malicious Classification System follows a Machine Learning and NLP-based cybersecurity approach.

Step 1: Problem Identification

Spam messages and malicious URLs are major cybersecurity threats that lead to phishing attacks, financial fraud, malware infections, and privacy breaches. Traditional filtering methods may fail to detect advanced spam and malicious links effectively. The proposed system aims to provide automated detection using Machine Learning and NLP techniques.

Step 2: Requirement Analysis

The following requirements are analyzed:

- SMS and URL datasets
- NLP processing requirements
- Machine Learning framework requirements
- Real-time security monitoring requirements
- Threat alert system requirements

Step 3: Dataset Preparation

Spam SMS and malicious URL datasets are collected and divided into:

- Training Dataset
- Validation Dataset
- Testing Dataset

Relevant textual and URL-based features are selected for analysis.

Step 4: NLP and Feature Engineering

The methodology includes:

1. Preprocess SMS text data
2. Extract spam keywords and patterns
3. Analyze URL structural characteristics
4. Generate feature vectors
5. Prepare data for ML classification

Step 5: Machine Learning Implementation

The Machine Learning workflow includes:

1. Train spam detection model
2. Train malicious URL classification model
3. Analyze incoming SMS and URLs
4. Predict spam and malicious threats
5. Generate security alerts

Step 6: Performance Evaluation

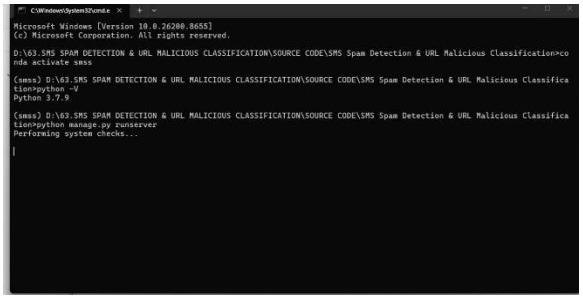
The system is evaluated based on:

- Spam detection accuracy
- URL classification efficiency
- False positive rate
- Real-time processing speed
- Cyber threat detection capability

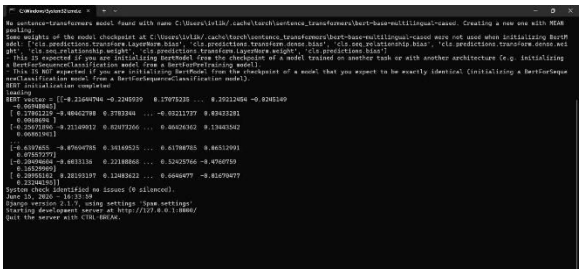
Technologies Used

- Python
- Machine Learning Algorithms
- Natural Language Processing (NLP)
- Scikit-learn
- TensorFlow / Keras
- Pandas & NumPy
- Flask / Django
- MySQL / MongoDB

Result



The screenshot shows a Windows Command Prompt with the smss virtual environment activated and Python 3.7.9 installed. The user executes python manage.py runserver, and Django begins startup by performing system checks before launching the development server.

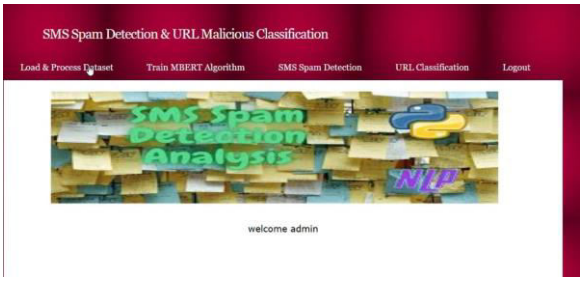


The screenshot shows the Django application successfully loading the BERT model and generating vector embeddings during startup. After completing system checks with no errors, the Django development server launches successfully at http://127.0.0.1:8000/, indicating that the SMS spam and malicious URL classification system is ready for use.

The screenshot displays the home page of the SMS Spam Detection and URL Malicious Classification system, providing navigation options such as Home and User Login. It serves as the main interface where users can access the application and begin the spam and malicious URL detection process.

detection features.

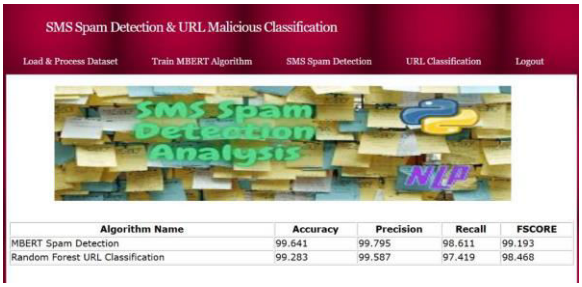
This is the User Login Page of the SMS Spam Detection & URL Malicious Classification system. It allows users to securely enter their username and password to access the application and use spam message and malicious URL detection features.



This is the Admin Dashboard of the SMS

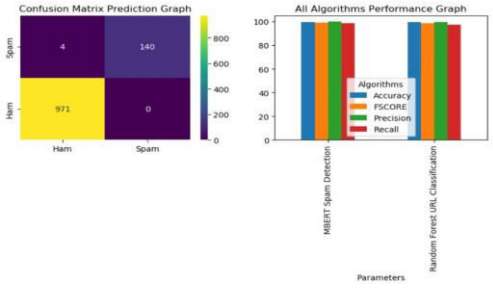
Spam Detection & URL Malicious Classification system. It provides options to load datasets, train the MBERT algorithm, perform SMS spam detection, classify URLs, and manage the system efficiently.

This page shows the Dataset Loading and Processing results for SMS Spam Detection. It displays the total records, spam/ham labels, train-test split details, and sample messages used for model training and testing.



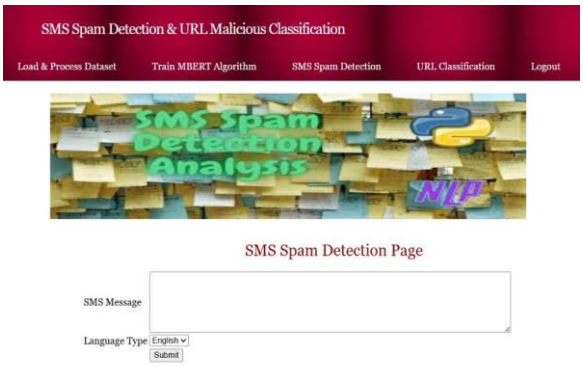
This page shows the final results of the SMS Spam Detection and URL Malicious Classification system. The MBERT algorithm achieved 99.64% accuracy for spam detection, while the Random Forest model achieved 99.28% accuracy for URL classification.

The performance metrics such as Accuracy, Precision, Recall, and F1-Score indicate that both models provide highly reliable and effective detection of spam messages and malicious URLs.



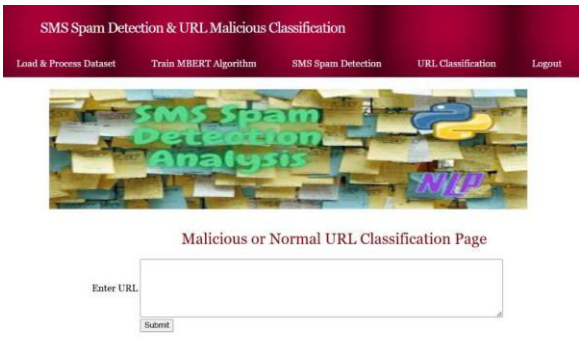
The confusion matrix graph shows the classification results of spam and ham messages, helping evaluate the model's prediction accuracy. The performance graph compares Accuracy, Precision, Recall, and F1-Score of MBERT Spam Detection and Random Forest URL Classification, showing excellent results for both models.

Both algorithms achieved performance values close to 100%, indicating high reliability in detecting spam messages and malicious URLs effectively.



This screenshot shows a web application for SMS Spam Detection and URL Malicious Classification using machine learning and NLP techniques.

Users can load datasets, train the MBERT algorithm, and classify SMS messages or URLs as spam, safe, or malicious.



This web application is an SMS Spam Detection & URL Malicious Classification system built using Python and NLP techniques, featuring an MBERT (Multilingual BERT) algorithm for training. It provides a user-friendly interface allowing users to load datasets, train the model, detect spam SMS messages, and classify URLs as either malicious or normal through a dedicated classification page.

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.

References

1. Almeida, T. A., Yamakami, A., & de Souza, J. A. (2011). SMS Spam Filtering: Methods and Data Sets. *Expert Systems with Applications*, 39(10), 9899-9908. https://doi.org/10.1016/j.eswa.2012.03.011
2. Almeida, T. A., Hidalgo, J. M. G., &Yamakami, A. (2013). Contributions to SMS Spam Filtering: New Collection and Results. *Proceedings of the 11th ACM Symposium on Document Engineering*, 259-262. https://doi.org/10.1145/2494266.2494291
3. Zhang, Y., & Hong, J. (2015). Detecting Malicious URLs via

Machine Learning. *International Journal of Computer Applications*, 127(12),1-6. https://doi.org/10.5120/ijca2015906499

4. Ma, J., Saul, L. K., Savage, S., & Voelker, G. M. (2009). Identifying Suspicious URLs: An Application of Large-Scale Online Learning. *Proceedings of the 26th Annual International Conference on Machine Learning*, 681-688.
5. Ye, D., Li, T., Adjeroh, D., & Iyengar, S. (2017). A Survey on Malware Detection Using Data Mining Techniques. *ACM Computing Surveys*, 50(3), 1-40. https://doi.org/10.1145/3054920
6. UCI Machine Learning Repository: SMS Spam Collection Dataset. Retrieved from https://archive.ics.uci.edu/ml/datasets/SMS+Spam+Collection
7. Kaggle. Malicious URLs Dataset. Retrieved from https://www.kaggle.com/harrywang/url-website-phishing-detection

Mr. Himam basha Shaik is an Assistant Professor in the Department of Master of Computer Applications at QIS College of Engineering and Technology, Ongole, Andhra Pradesh. He earned his Master of Computer Applications (MCA) from Anna University, Chennai. With a strong research background, He has authored and co-authored research papers published in reputed peer-reviewed journals. His research interests include Machine Learning, Artificial Intelligence, Cloud Computing, and Programming Languages. He is committed to advancing research and fostering innovation while mentoring students to excel in both academic and professional pursuits.

Ms. I. Venkata Likhitha is a postgraduate student pursuing a Master of Computer Applications (MCA) in the Department of Computer Applications at QIS College of Engineering & Technology, Ongole — an autonomous institution in Prakasam District. She completed her undergraduate degree in **B.Sc. (Statistics)** from Acharya Nagarjuna University. With a strong interest in research and practical learning, she actively engages in academic projects and technical activities, contributing to advancements in her field.