International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Volume: 04 Issue: 03 | March – 2025 DOI: 10.55041/ISJEM02759
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

Phishguard: ML- based phishing detection system

Bhagyashri Satarkar[*1], Dhavalsigh Vibhute[*2], Vishal Wadgoankar[*3], Yasar Sayyad[*4]

Under the Guidance of Prof. PradnyaKothawade

Genba Moze College of Engineering,Balewadi

Abstract - Spam messages in SMS and email systems pose significant security and productivity risks. Traditional detection methods, such as rule-based filters, struggle to adapt to evolving spam techniques. This paper explores machine learning techniques for spam detection, emphasizing algorithms like Naïve Bayes and Support Vector Machines (SVM), along with their applications, strengths, and limitations. Key challenges, including scalability, dataset diversity, and evolving spam patterns, are identified. The study highlights future directions such as real-time classification, deep learning models, and improved feature engineering to enhance adaptive and robust spam detection systems..

Keywords: Spam Detection, Machine Learning, Text Classification, Naive Bayes, Feature Engineering, Cybersecurity

1.INTRODUCTION

Spam messages significantly impact cybersecurity, leading to data breaches, financial fraud, and wasted resources. Traditional rule-based filtering techniques have been the primary defense mechanism, but they fail to adapt to new and sophisticated spam tactics. Machine learning (ML) presents an adaptive and scalable approach to spam detection by leveraging data-driven classification methods.
This survey aims to:
1)Critically analyze existing spam detection methodologies.
2)Identify research gaps and challenges in ML-based spam detection.
3)Propose advancements for improved spam detection systems.

2. Literature Review

2.1 Approaches to Spam Detection

Spam detection techniques can be broadly categorized as follows:

1)URL-based Approaches: These methods focus on detecting malicious links within messages by analyzing domain reputation, URL shortening services, and blacklisted domains.

2)Content-based Approaches: These leverage keyword matching, linguistic analysis, and natural language processing (NLP) to identify spam patterns.

3)Text Classifiers: Machine learning models analyze email or SMS text to classify messages as spam or legitimate, utilizing word frequency, sentiment analysis, and metadata.

Traditional static filters, while initially effective, struggle to adapt to newly emerging spam patterns, necessitating machine learning-based solutions.

2.2 Machine Learning Algorithms for Spam Detection

Several machine learning algorithms have been applied to spam detection:

Naïve Bayes: A probabilistic classifier widely used in text classification due to its efficiency and simplicity.

Support Vector Machines (SVM): Well-suited for high-dimensional text data, SVM effectively separates spam and non-spam messages using optimized decision boundaries.

Decision Trees & Gradient Boosting: Decision trees provide interpretable classification, while boosting techniques like XGBoost improve accuracy by iteratively refining predictions.

Ensemble and Hybrid Models: These combine multiple algorithms (e.g., Random Forest, stacking) to enhance accuracy and generalization.

2.3 Feature Engineering in Spam Detection

Feature extraction plays a critical role in improving model accuracy. Key techniques include:

i)TF-IDF (Term Frequency-Inverse Document Frequency): Highlights important words by evaluating their relative importance in a message.

ii)Principal Component Analysis (PCA): Reduces feature dimensionality for improved computational efficiency.

iii)Custom Features: Includes character n-grams, bigrams, sender metadata, and email structure analysis.

2.4 Summary of Findings and Challenges

Despite advancements in ML-based spam detection, several challenges persist:

Scalability: Many ML models struggle with real-time spam classification at scale.

Dataset Diversity: Existing datasets lack multilingual and real-world spam diversity.

Evolving Spam Patterns: Attackers frequently change spam characteristics, making static detection models ineffective.

3. Methodology

3.1 Dataset

The **SMS Spam Collection** dataset is utilized for this study, consisting of labeled spam and non-spam messages.

3.2 Preprocessing Steps

Tokenization: Splitting text into individual words or phrases.

Vectorization: Converting text into numerical representations using CountVectorizer and TF-IDF techniques.

3.3 Model Selection

Naïve Bayes is selected for its effectiveness in text classification, offering a balance between accuracy and computational efficiency.

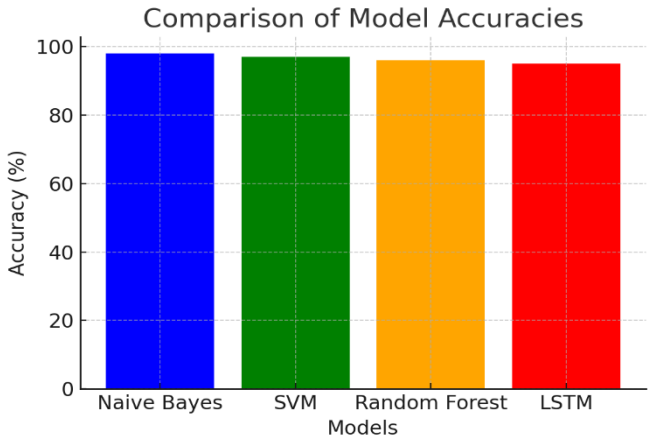
3.4 Evaluation Metrics

Model performance is assessed using:

Accuracy: Measures the overall classification correctness.

Precision & Recall: Evaluates spam detection effectiveness and false positives.

- **F1-score**: Ensures a balanced assessment of precision and recall.



4. Discussion

4.1 Strengths of ML-based Spam Detection

a)Adaptability: ML models learn from evolving spam patterns.

b)Efficiency: Text classification algorithms process large datasets quickly.

c)Integration: ML-based spam detection can be seamlessly integrated into email and messaging systems.

4.2 Limitations

a)Dependence on Labeled Data: High-quality labeled datasets are required for effective training.

b)Computational Costs: Complex models may require high processing power.

c)False Positives: Some legitimate messages may be mistakenly classified as spam.

4.3 Emerging Trends

i)Deep Learning for Spam Detection: Techniques like LSTMs and CNNs enhance contextual understanding.

ii)Real-time Spam Filtering: Cloud-based systems can process messages instantly for rapid classification.

ii)Multilingual Spam Detection: Future models will support spam detection in multiple languages.

5. Future Scope

The phishing detection system has much scope for development and improvement. The possible enhancements and areas for further development on this project are as follows:

5.1.Communication with Real-time Messaging Systems:

With its integration with the most popular means of messaging, such as SMS, WhatsApp, or email, the system can detect phishing in real time. This will also assist users in detecting phishing messages in real time as they receive their messages, hence ensuring more proactive protection.

5.2.Advanced Machine Learning Models:

Future work may include investigating more complex and sophisticated machine learning models, such as deep learning (e.g., LSTM, CNNs), to improve phishing detection accuracy with more varied and complex datasets.

Ensemble models that combine the predictions of multiple algorithms (such as Random Forest or XGBoost) may further improve performance.

5.3.Cross-Platform Integration:

The web-based interface can be enhanced to be extended into a mobile application. People can view the messages from any phone or even other devices themselves.

5.4.Multilinguality Support

The model would be further enriched by support to multiple languages as it will provide detection of the phishing message beyond the English and used in Atm machines to detect frauds. Also, the system can be used at the time of elections where the voter can be identified by recognizing the face.

6. Conclusion

Machine learning has significantly advanced spam detection by enabling adaptive, scalable, and accurate classification. While traditional approaches struggle with evolving spam patterns, ML-based techniques, including Naïve Bayes, SVM, and ensemble models, have shown promising results. However, scalability, dataset diversity, and real-time processing remain key challenges. Future research should focus on deep learning, cross-platform integration, and multilingual spam detection to further improve security and efficiency.

6. References

1. The Answer is in the Text: Multi-Stage Methods for Phishing Detection Based on Feature Engineering
2. Phishing Detection System Through Hybrid Machine Learning Based on URLUCI Machine Learning Repository. "SMS Spam Collection Dataset".
3. Robust Ensemble Machine Learning Model for Filtering Phishing URLs: Expandable Random Gradient Stacked Voting Classifier (ERG-SVC);
4. A Comprehensive Taxonomy of Social Engineering Attacks and Defense Mechanisms: Toward Effective Mitigation Strategies