International Journal of Scientific Research & Engineering Trends
Volume 12, Issue 2, Mar-Apr-2026, ISSN (Online): 2395-566X

SpamShield: A Robust Machine Learning Framework for Intelligent SMS and Email Spam Detection via Hybrid Text Analytics

Mrs. T.Swapna Sridevi[1], Peddireddy Pattabhi Rama Lingeswar[2]
[1]Associate Professor, [2]M. tech Student
Department of CSE Artificial Intelligence (AI), Pydah College of Engineering,
Yanam Road, Tallarevu, Patavala, Andhra Pradesh, 533461

Abstract- The rapid growth of digital communication platforms has significantly increased the volume of SMS and email messages exchanged daily. While these technologies enhance connectivity and information sharing, they have also become primary channels for spam, phishing, and fraudulent activities. Spam messages not only cause inconvenience but also pose serious security and privacy risks to individuals and organizations. Therefore, developing an accurate and efficient automated spam detection system has become an essential requirement. This study proposes a robust machine learning framework for intelligent classification of spam and legitimate (ham) SMS and email messages using advanced text analytics techniques. The system incorporates comprehensive preprocessing methods, including text cleaning, tokenization, stop-word removal, and normalization, followed by feature extraction using techniques such as TF-IDF and word embeddings. Multiple machine learning algorithms, including Naïve Bayes, Support Vector Machines, Logistic Regression, Random Forest, and Gradient Boosting, are implemented and comparatively evaluated. To further enhance predictive performance, ensemble learning strategies are employed to combine the strengths of individual classifiers. Experimental results demonstrate that the proposed hybrid framework achieves high accuracy, precision, recall, and F1-score across benchmark datasets. The system effectively minimizes false positives and false negatives, thereby improving reliability in real-world applications. The proposed approach contributes to the development of scalable, intelligent, and adaptive spam filtering systems capable of handling evolving spam patterns in modern communication networks.

Keywords – Spam Detection, SMS Classification, Email Filtering, Machine Learning, Text Analytics, Natural Language Processing (NLP), Ensemble Learning, TF-IDF, Binary Classification.

I. INTRODUCTION

In today's digital era, electronic communication through SMS and email has become an integral part of personal, academic, and professional interactions. Billions of messages are transmitted daily across mobile networks and internet-based platforms, facilitating instant communication worldwide. However, alongside these benefits, the proliferation of spam messages has emerged as a serious challenge. Spam SMS and emails are often used to promote unsolicited advertisements, spread malicious links, conduct phishing attacks, or perform financial fraud. These activities not only disrupt communication but also compromise user privacy and cybersecurity. Traditional spam filtering mechanisms initially relied on rule-based systems and manually defined keyword filters. Although such approaches were effective in early stages, they lack adaptability and fail to handle the rapidly evolving tactics employed by spammers.

Modern spam messages often use obfuscation techniques, misleading content, and contextual manipulation to bypass simple filtering systems. As a result, more intelligent and adaptive detection mechanisms are required. Machine learning has gained significant attention as a powerful tool for automated text classification tasks. By learning patterns from labelled datasets, machine learning models can distinguish between spam and legitimate (ham) messages with improved accuracy. Algorithms such as Naïve Bayes, Support Vector Machines, Logistic Regression, and Random Forest have shown promising results in binary classification problems. Furthermore, advances in Natural Language Processing (NLP) enable systems to understand contextual and semantic information within text data, enhancing detection capabilities. Despite these advancements, challenges such as class imbalance, feature selection complexity, evolving spam patterns, and false positive errors continue to affect system performance. In real-world scenarios, minimizing false

positives is especially critical, as misclassifying legitimate messages can lead to important communication loss. Motivated by these challenges, this work proposes a robust machine learning framework that integrates advanced text preprocessing techniques, optimized feature extraction, and ensemble-based classification strategies. The primary objective is to design an intelligent, scalable, and high-performance spam detection system capable of accurately identifying spam messages across both SMS and email platforms. By leveraging hybrid text analytics and machine learning techniques, the proposed system aims to enhance filtering accuracy, reduce security risks, and contribute to safer digital communication environments.

II. LITERATURE SURVEY

Spam detection has been an active area of research for more than two decades, evolving alongside advancements in communication technologies and machine learning techniques. Early spam filtering systems primarily relied on rule-based approaches and keyword matching methods. These systems used manually defined rules to detect suspicious words or phrases commonly associated with spam. Although simple to implement, rule-based systems lacked adaptability and struggled to handle new or cleverly disguised spam messages. With the advancement of machine learning, researchers began exploring statistical classification algorithms for automated spam detection. Naïve Bayes became one of the earliest and most widely adopted methods due to its simplicity, computational efficiency, and relatively strong performance in text classification tasks. Support Vector Machines (SVM) were later introduced to improve classification boundaries by maximizing margin separation between spam and legitimate messages. Logistic Regression also demonstrated effectiveness in binary classification scenarios due to its probabilistic output and interpretability. As datasets grew larger and more complex, ensemble learning techniques such as Random Forest and Gradient Boosting gained popularity. These models combine multiple weak learners to improve predictive accuracy and robustness. Studies have shown that ensemble methods often outperform single classifiers, particularly in scenarios involving noisy or imbalanced datasets. Hybrid models that integrate different machine learning algorithms have also been proposed to enhance detection performance. In recent years, the application of Natural Language Processing (NLP) techniques has significantly improved spam classification systems. Feature extraction methods such as Term Frequency–Inverse Document Frequency (TF-IDF), n-grams, and word embeddings enable better representation of textual data. Deep learning approaches, including Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN), have further enhanced performance by capturing contextual and sequential information within messages. Despite these advancements, several challenges remain. Spam messages continuously evolve, employing sophisticated obfuscation techniques to bypass filters. Additionally, class imbalance issues often affect detection accuracy, as legitimate messages typically outnumber spam messages in real-world datasets. High false positive rates can reduce user trust and system reliability. Therefore, there is a need for more adaptive, scalable, and hybrid frameworks that combine effective preprocessing, optimized feature extraction, and ensemble learning strategies. The present study builds upon these existing works by integrating advanced text analytics with multiple machine learning algorithms to design a robust and intelligent spam detection framework capable of achieving improved performance and real-world applicability.

III. SYSTEM ANALYSIS

A. EXISTING SYSTEM

Existing spam detection systems primarily rely on traditional machine learning algorithms for binary text classification. In earlier approaches, researchers implemented conventional classifiers such as Naïve Bayes, Decision Trees, Logistic Regression, Support Vector Machines (SVM), and k-Nearest Neighbours (k-NN) to distinguish between spam and legitimate (ham) messages. These models were trained using manually extracted textual features such as word frequency, keyword occurrence, and n-gram representations. Some studies extended these methods by incorporating ensemble techniques such as Random Forest and AdaBoost to improve predictive performance. Hybrid approaches combining multiple classifiers through majority voting mechanisms were also proposed to enhance classification robustness. In addition, certain models were evaluated under noisy conditions to assess their resilience against message obfuscation techniques commonly used by spammers. Although these approaches demonstrated reasonable performance on benchmark datasets, they often depend heavily on static feature engineering and lack adaptability to evolving spam patterns. Many existing systems focus on either SMS or email datasets individually, limiting their generalization capability across different communication platforms.

DISADVANTAGES OF THE EXISTING SYSTEM

Despite notable progress in spam detection research, several limitations remain in current systems:

- **Limited Interpretability:** Complex machine learning models, particularly ensemble and deep learning approaches, may produce high accuracy but lack transparency. Understanding why a message is classified as spam is important for user trust and system accountability.
- **Overfitting and Underfitting Issues:** Models may overfit training data, capturing noise instead of meaningful textual patterns, or underfit when unable to fully learn message characteristics. Proper validation and parameter tuning are required to address these concerns.

- **Feature Dependency:** Traditional systems rely heavily on manually engineered features, which may not effectively capture contextual or semantic relationships within text data.
- **Computational Overhead:** Advanced ensemble or deep learning models can require significant computational resources, making real-time deployment challenging in resource-constrained environments.
- **Evolving Spam Techniques:** Spammers continuously modify message structures using abbreviations, symbols, and obfuscation methods to bypass filters. Static models struggle to adapt quickly to such dynamic changes.
- **Scalability Challenges:** As communication traffic increases, spam detection systems must efficiently handle large volumes of messages without performance degradation.

B. PROPOSED SYSTEM

To overcome the limitations of existing approaches, the proposed system introduces a robust and hybrid machine learning framework for intelligent spam detection across both SMS and email platforms. Initially, the dataset undergoes comprehensive preprocessing, including text cleaning, removal of special characters, stop-word elimination, tokenization, normalization, and stemming/lemmatization. This step ensures high-quality input for model training. Feature extraction techniques such as TF-IDF vectorization and advanced text representations are applied to transform textual data into meaningful numerical formats. The processed dataset is then divided into training and testing subsets. Multiple machine learning classifiers including Naïve Bayes, Support Vector Machines, Logistic Regression, Random Forest, and Gradient Boosting are trained and evaluated. To enhance predictive performance, ensemble learning strategies are employed to combine the strengths of individual models. Hyperparameter optimization techniques are applied to fine-tune model parameters for improved generalization. Cross-validation is performed to ensure robustness and prevent overfitting. The system performance is evaluated using comprehensive metrics such as accuracy, precision, recall, F1-score, confusion matrix analysis, and Receiver Operating Characteristic (ROC) curves. The proposed hybrid framework aims to deliver improved detection accuracy, reduced false positive rates, enhanced adaptability to evolving spam patterns, and scalable performance suitable for real-world communication environments.

IV. SYSTEM DESIGN

System Architecture

Below diagram depicts the whole system architecture.

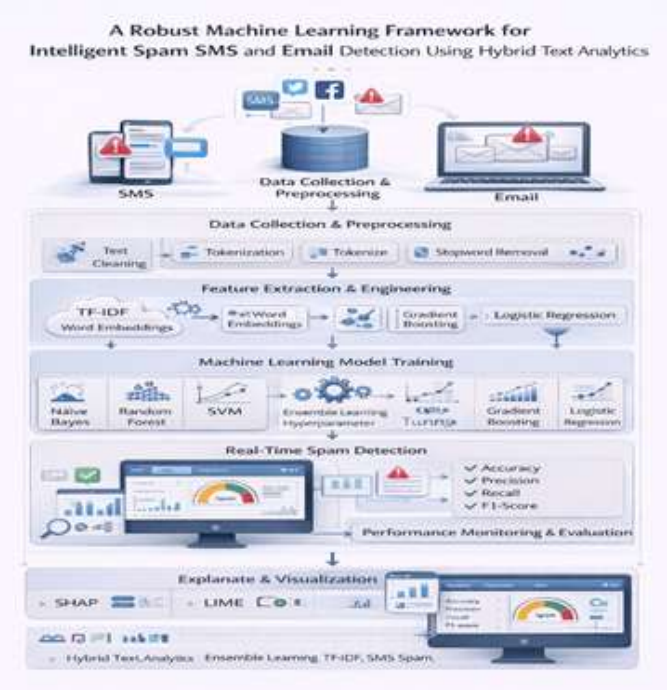

Fig 1. Methodology followed for proposed model

V. SYSTEM IMPLEMENTATION

MODULES

1. Data Collection and Preprocessing

The first stage of implementation involves collecting relevant datasets containing labelled SMS and email messages categorized as spam or legitimate (ham). The raw textual data undergoes extensive preprocessing to improve quality and consistency. This includes removing punctuation, special characters, URLs, and unnecessary symbols, converting text to lowercase, eliminating stop words, and performing tokenization and stemming or lemmatization. These preprocessing steps help standardize the text and reduce noise, enabling more effective feature extraction and model training.

2. Feature Extraction and Engineering

After preprocessing, meaningful numerical representations of textual data are generated using feature extraction techniques such as Term Frequency–Inverse Document Frequency (TF-IDF) and n-gram modelling. These techniques transform unstructured text into structured feature vectors suitable for machine learning algorithms. Feature selection methods are also applied to retain the most informative attributes while reducing dimensionality, thereby improving computational efficiency and classification performance.

3. Machine Learning Model Training
Multiple machine learning algorithms are implemented to classify messages as spam or ham. These include Naïve Bayes, Logistic Regression, Support Vector Machines (SVM), Random Forest, and Gradient Boosting classifiers. Each model is trained using the pre-processed and vectorized dataset. Hyperparameter tuning is performed to optimize model performance and improve generalization capability. Ensemble learning strategies are further incorporated to combine predictions from multiple models, enhancing robustness and overall accuracy.

4. Real-Time Spam Detection Module
A real-time classification component is developed using the trained model to automatically analyse incoming SMS or email messages. When a new message is received, it undergoes the same preprocessing and feature extraction steps before being classified. This module ensures quick and efficient filtering of spam messages, reducing exposure to malicious content and improving user experience.

5. Model Evaluation and Monitoring
The performance of the trained models is evaluated using standard classification metrics, including accuracy, precision, recall, F1-score, and confusion matrix analysis. Receiver Operating Characteristic (ROC) curves are also analysed to assess model discrimination capability. Continuous monitoring mechanisms can be incorporated to evaluate system performance over time and update the model when new spam patterns emerge.

VI. RESULTS AND DISCUSSION

To assess the effectiveness of the proposed framework, multiple machine learning algorithms are evaluated using stratified cross-validation techniques to ensure balanced representation of spam and legitimate messages. Hyperparameter tuning and ensemble strategies are applied to enhance predictive performance. The experimental results indicate that ensemble-based models outperform individual classifiers in terms of accuracy and robustness. The proposed hybrid framework achieves high precision and recall values, effectively reducing false positives and false negatives. Minimizing false positives is particularly important in spam detection systems to prevent legitimate messages from being incorrectly filtered. Comparative analysis demonstrates that the integration of optimized feature extraction and ensemble learning significantly improves classification performance over traditional single-model approaches. The system exhibits scalability and adaptability, making it suitable for real-world SMS and email filtering applications.

VII. CONCLUSION AND FUTURE WORK

This study presented a robust machine learning framework for intelligent spam SMS and email detection using hybrid text analytics techniques. The proposed system integrates comprehensive preprocessing, advanced feature extraction, multiple classification algorithms, and ensemble learning strategies to achieve high detection accuracy and reliability.
Experimental evaluation demonstrates that the framework effectively distinguishes between spam and legitimate messages while minimizing classification errors. The integration of ensemble methods enhances robustness against evolving spam tactics and improves generalization performance. In future work, the system can be extended by incorporating deep learning architectures such as Convolutional Neural Networks (CNN) or Transformer-based models to capture deeper semantic relationships within text data. Additionally, integrating real-time adaptive learning mechanisms would further improve the system's ability to handle continuously evolving spam patterns. Expanding the dataset to multilingual and cross-platform communication environments could also enhance system versatility and global applicability.

REFERENCES

1. L. N. Lota et al., "A systematic literature review on SMS spam detection techniques," I.J. Information Technology and Computer Science, vol. 7, pp. 42–50, 2017.
2. P. Sethi et al., "SMS spam detection and comparison of various machine learning algorithms," in Proc. Int. Conf. Computing and Communication Technologies for Smart Nation (IC3TSN), 2017, pp. 28–31.
3. S. M. Abdulhamid et al., "A review on mobile SMS spam filtering techniques," IEEE Access, vol. 5, pp. 15650–15666, 2017.
4. M. Rubin Julis et al., "Spam detection in SMS using machine learning through text mining," Int. J. Scientific & Technology Research, vol. 9, no. 2, 2020.
5. A. Alzahrani et al., "Comparative study of machine learning algorithms for SMS spam detection," in Proc. SoutheastCon, 2019, pp. 1–6.
6. N. Nisar et al., "Voting-ensemble classification for email spam detection," in Proc. Int. Conf. Communication Information and Computing Technology (ICCICT), 2021, pp. 1–6.
7. S. Agarwal et al., "SMS spam detection for Indian messages," in Proc. Int. Conf. Next Generation Computing Technologies (NGCT), 2015, pp. 634–638.
8. M. Crawford et al., "Survey of review spam detection using machine learning techniques," Journal of Big Data, 2015.

9. A. Radhakrishnan et al., “Email classification using machine learning algorithms,” Int. J. Engineering and Technology (IJET), pp. 335–340, 2017.
10. N. Govil et al., “A machine learning based spam detection mechanism,” in Proc. Int. Conf. Computing Methodologies and Communication (ICCMC), 2020, pp. 954–957.
11. Shirani-Mehr et al., “SMS spam detection using machine learning approach,” 2013, pp. 1–4.
12. A. Tomasic et al., “Learning to detect phishing emails,” in Proc. 16th Int. Conf. World Wide Web, 2007, pp. 649–656.
13. D. D. Arifin et al., “Enhancing spam detection on mobile phone short message service (SMS) performance using FP-growth and Naive Bayes classifier,” in Proc. IEEE Asia Pacific Conf. Wireless and Mobile (APWiMob), 2016, pp. 80–84.
14. A. Andronicus et al., “Classification of phishing email using random forest machine learning technique,” Journal of Applied Mathematics, Hindawi, 2014.
15. G. Tripathi et al., “Feature selection and classification approach for machine learning applications,” Machine Learning and Applications: An International Journal, vol. 2, no. 2, pp. 1–16, 2015.
16. A. Vikram et al., “Anomaly detection in network traffic using unsupervised machine learning approach,” in Proc. 5th Int. Conf. Communication and Electronics Systems (ICCES), 2020, pp. 476–479.
17. C. V. Krishna et al., “A review of artificial intelligence methods for data science and data analytics: Applications and research challenges,” in Proc. 2nd Int. Conf. I-SMAC (IoT in Social, Mobile, Analytics and Cloud), 2018, pp. 591–594.