IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 11, November 2025

DOI: 10.17148/IJARCCE.2025.141176

Spam or Ham Message Detection Model

Dipali Gulab Mali[1], Prof. Shivam Limbare[2] , Manoj V. Nikum*[3]

Student Of MCA, Shri Jaykumar Rawal Institute of Technology Dondaicha, KBC NMU Jalgaon,Maharashtra, India[1]

Assistant Professor, MCA Department, SJRIT DONDAICHA, KBC NMU JALGAON, Maharashtra, India[2]

Assistant Professor & HOD, MCA Department, SJRIT DONDAICHA,KBC NMU JALGAON, Maharashtra, India[3]

Abstract: Correct identification of spam messages is a critical component of modern digital communication. With the rapid growth of mobile devices and instant messaging platforms, users are increasingly exposed to unsolicited messages that often contain phishing links, fraudulent schemes, or promotional content. These spam messages not only pose security threats but also result in time loss, decreased productivity, and potential financial damage. Therefore, developing automated systems capable of accurately detecting and filtering spam messages is essential for safeguarding users and maintaining the integrity of communication networks.

In this research, we propose a machine learning-based system to classify SMS messages as spam or non-spam (ham). The system relies on textual features extracted from messages, including content analysis, message length, presence of specific keywords, and frequency patterns. These features serve as inputs for machine learning models that learn patterns distinguishing spam from legitimate messages. The dataset utilized for this study is the widely recognized SMS Spam Collection Dataset, which consists of thousands of labeled messages. To ensure high-quality model performance, the dataset undergoes preprocessing to handle missing values, remove inconsistencies, standardize formats, and normalize text data. This preprocessing phase also involves tokenization, stopword removal, and transformation of text into numerical representations using techniques like TF-IDF or Bag-of-Words.

For predictive modeling, three widely used machine learning algorithms are employed: Naive Bayes, Decision Trees, and Support Vector Machines (SVM). Naive Bayes is a probabilistic classifier well-suited for text-based data due to its efficiency and simplicity. Decision Trees provide a transparent, rule-based approach that can capture nonlinear relationships in the data, while SVM is a robust classifier that maximizes the margin between spam and ham messages, often resulting in high accuracy. These models are trained and evaluated using standard performance metrics such as accuracy, precision, recall, and F1-score, allowing for a comprehensive assessment of their predictive capabilities.

The results of this study demonstrate that machine learning models can achieve high accuracy in spam detection, highlighting their practical utility in real-world communication systems. SVM, in particular, showed superior performance, although Naive Bayes and Decision Trees also provided competitive results. By systematically addressing challenges such as feature extraction, preprocessing, and model selection, this research lays a solid foundation for the development of automated spam detection systems.

Future work can enhance the system further by integrating ensemble learning techniques, which combine multiple classifiers to improve robustness and prediction accuracy. Moreover, real-time message stream analysis can enable the system to detect and block spam messages instantaneously. Advanced natural language processing (NLP) methods, including deep learning models like LSTM networks or Transformers, can be employed to capture complex sequential and semantic patterns in SMS text. Expanding the dataset to include multilingual messages and adapting the models to handle cross-language spam can also increase the applicability of the system in global communication environments. Overall, this study emphasizes the critical role of machine learning in protecting users from spam and securing digital communication platforms.

I. INTRODUCTION

The exponential growth of mobile communication has led to a surge in unsolicited messages, commonly referred to as spam. These messages are not only an inconvenience but also a potential threat to digital security. Spam messages can contain phishing attempts, malware links, or fraudulent offers that may compromise user privacy or result in financial loss. According to recent studies, millions of SMS messages are sent daily, and a significant portion of them constitutes spam. For instance, the Cellular Telecommunications Industry Association (CTIA) reported that over 50% of mobile users have received spam messages at least once a week, highlighting the scale and pervasiveness of the problem. As

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 11, November 2025
DOI: 10.17148/IJARCCE.2025.141176

mobile messaging continues to grow as a primary communication channel, the need for effective spam detection mechanisms becomes increasingly critical.

Traditional spam filtering techniques rely on rule-based systems that detect messages based on predefined keywords, patterns, or sender addresses. While these systems can block basic spam effectively, they have several limitations. Spammers frequently modify their message content, use obfuscation techniques, or adopt sophisticated linguistic patterns that evade static rules. As a result, rule-based filters often suffer from low adaptability and can generate high rates of false positives, where legitimate messages are incorrectly flagged as spam. Additionally, manual updates to rules and patterns are labor-intensive and may not keep pace with evolving spam tactics. These challenges necessitate the development of more intelligent, adaptive approaches for spam detection.

Machine learning (ML) offers a dynamic alternative by enabling systems to automatically learn distinguishing patterns from historical message data. ML-based spam detection models can capture complex relationships between message features, such as word frequency, syntax, message length, and presence of hyperlinks or numbers. Unlike rule-based systems, these models are capable of generalizing from past data and recognizing new forms of spam that may not have been explicitly seen before. Furthermore, ML models are scalable, allowing them to process vast amounts of SMS data efficiently, making them suitable for deployment in high-volume messaging platforms.

In this study, we develop a machine learning-based spam detection system that classifies SMS messages into spam or ham categories. The research emphasizes several key aspects of building an effective ML system. Data preprocessing is essential to clean and normalize message text, remove noise, and handle missing values. Feature engineering plays a crucial role in improving model performance by identifying the most informative attributes of messages, such as n-grams, term frequency-inverse document frequency (TF-IDF) scores, and patterns of common spam keywords. Model selection is another critical factor, as different algorithms offer varying trade-offs in terms of accuracy, interpretability, and computational efficiency. In this study, we employ three widely used classifiers—Naive Bayes, Decision Trees, and Support Vector Machines (SVM)—to explore their effectiveness in detecting spam messages.

The objectives of this research include:

1. Developing a reliable and scalable machine learning-based system for SMS spam detection.
2. Evaluating and comparing the performance of Naive Bayes, Decision Trees, and SVM classifiers using accuracy, precision, recall, and F1-score metrics.
3. Exploring data preprocessing and feature engineering strategies to enhance model performance.
4. Laying the groundwork for future improvements, such as real-time message detection, ensemble learning, and the integration of deep learning methods like LSTM networks or Transformer models for advanced text analysis.

By addressing these objectives, this study aims to provide a robust foundation for automated spam detection systems that are capable of adapting to evolving spam tactics while minimizing false positives. The implementation of such systems not only improves the user experience by reducing unwanted interruptions but also enhances the overall security and reliability of mobile communication platforms. Moreover, the research highlights the potential of machine learning in real-world applications, demonstrating its ability to combine efficiency, adaptability, and accuracy in addressing a pressing digital security challenge.

II. LITERATURE SURVEY

Spam detection has been extensively studied, with various approaches proposed to enhance accuracy and efficiency. This section provides a detailed review of key contributions in this field.

Naive Bayes for SMS Spam Detection
In [1], Almeida et al. explored the use of Naive Bayes algorithms for SMS spam detection. The study showed that Naive Bayes, despite its simplicity, can achieve high accuracy when combined with appropriate preprocessing. Techniques like tokenization, stemming, and stop-word removal were highlighted as critical for improving model performance.

Decision Trees and Random Forests
Cormack [2] focused on Decision Trees and Random Forests for spam classification. Tree-based models demonstrated

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 11, November 2025
DOI: 10.17148/IJARCCE.2025.141176

robustness in handling feature-rich datasets and provided interpretability, allowing users to understand which features contributed most to spam classification. Compared to rule-based systems, tree models showed improved precision and recall.

Feature Engineering Techniques
Jain and Kumar [3] emphasized the role of text preprocessing and feature engineering, using techniques such as TF-IDF, word embeddings, and n-grams to represent message content numerically. Their study indicated that high-quality features significantly influence model accuracy.

SVM-Based Approaches
Huang and Tsai [4] implemented Support Vector Machines (SVM) for SMS spam detection. SVM was effective in separating spam and ham messages in high-dimensional feature space. Their research highlighted the importance of kernel selection, parameter tuning, and regularization for optimal performance.

Ensemble and Hybrid Models
Ghosh and Sen [5] explored hybrid approaches combining multiple classifiers to enhance detection accuracy. Ensemble methods, such as bagging and boosting, proved beneficial in handling dynamic messaging environments where message patterns evolve rapidly.

Real-Time Spam Detection
Almeida and Hidalgo [6] proposed frameworks for real-time spam detection, emphasizing incremental learning and integration with messaging platforms. Real-time detection ensures immediate identification and blocking of spam, improving user experience and system reliability.

III. RESEARCH METHODOLOGY

The research methodology involves multiple stages, including data collection, preprocessing, feature engineering, model selection, and evaluation.

1. Data Collection

The study uses the **SMS Spam Collection Dataset**, which contains 5,574 SMS messages labeled as spam or ham. The dataset includes message content and corresponding labels, serving as the basis for training and evaluating the models.

2. Data Preprocessing

Data preprocessing is a critical step to ensure the quality of input data:

- **Text Cleaning:** Removal of punctuation, special characters, and stop words to reduce noise.
- **Tokenization:** Splitting messages into words or tokens.
- **Normalization:** Conversion of all text to lowercase for uniformity.
- **Vectorization:** Transformation of textual data into numerical features using **TF-IDF** or **Bag-of-Words**.

3. Feature Engineering

Feature engineering creates meaningful variables that enhance model accuracy:

- Message length (number of words or characters)
- Frequency of common spam keywords (e.g., “win”, “free”, “prize”)
- Presence of URLs or numeric sequences
- N-gram features capturing sequential patterns of words

4. Model Selection

Three ML algorithms were selected for this study:

DOI: 10.17148/IJARCCE.2025.141176

- **Naive Bayes:** Probabilistic model suitable for text classification due to its simplicity and efficiency.
- **Decision Trees:** Splits data based on feature thresholds to make predictions, offering interpretability.
- **Support Vector Machines (SVM):** Finds an optimal hyperplane that separates spam and ham messages with maximum margin.

5. Model Evaluation Metrics

- **Accuracy:** Ratio of correctly predicted messages to total messages.
- **Precision:** Fraction of predicted spam messages that are truly spam.
- **Recall:** Fraction of actual spam messages correctly identified.
- **F1-Score:** Harmonic mean of precision and recall, balancing both metrics.

IV. RESULTS AND DISCUSSION

Model	Accuracy	Precision	Recall	F1-Score
Naive Bayes	0.97	0.94	0.88	0.91
Decision Tree	0.96	0.92	0.89	0.90
SVM	0.98	0.95	0.90	0.92

The **SVM model** achieved the highest overall accuracy of 98%, demonstrating strong performance in classifying SMS messages. Naive Bayes also performed well, particularly in handling text-based features, while Decision Trees offered interpretability with slightly lower accuracy.

The results indicate that machine learning can reliably detect spam messages. The choice of features, preprocessing techniques, and model selection significantly impacts the system's effectiveness. Additionally, SVM's robustness in high-dimensional spaces makes it particularly suitable for text classification tasks.

Analysis

- **Confusion Matrix Insights:** The SVM model had fewer false positives compared to Naive Bayes, making it preferable in applications where mistakenly labeling ham messages as spam can cause user inconvenience.
- **Feature Importance:** Features like message length, presence of URLs, and spam keyword frequency were highly influential across models.

V. CONCLUSION AND FUTURE SCOPE

The rapid growth of mobile communication and the proliferation of digital messaging platforms have made SMS spam a critical concern for users, service providers, and organizations. Spam messages not only cause inconvenience but can also result in financial fraud, phishing attacks, and leakage of sensitive information. Therefore, developing reliable automated methods to detect and filter spam messages is of significant practical importance.

This study developed a **machine learning-based system for SMS spam detection**, leveraging features extracted from message text such as content, length, frequency of keywords, and presence of links or numbers. Among the machine learning models evaluated—**Naive Bayes, Decision Trees, and Support Vector Machines (SVM)**—the SVM model achieved the highest overall accuracy of 98%. This demonstrates the model's ability to handle high-dimensional textual data effectively, making it particularly suitable for detecting patterns in SMS messages that distinguish spam from legitimate messages (ham).

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 11, November 2025
DOI: 10.17148/IJARCCE.2025.141176

The **Naive Bayes** model performed well due to its efficiency in handling text classification problems and its ability to operate effectively even with relatively small datasets. The **Decision Tree** model, while slightly lower in accuracy, provided transparent decision-making rules, which can be advantageous in understanding why specific messages are flagged as spam. Overall, these results indicate that machine learning models, particularly when selected and optimized carefully, can form the foundation for robust and reliable automated spam detection systems.

Future Improvements

Despite the strong performance achieved in this study, several areas of improvement remain, which could further enhance system accuracy, adaptability, and practical deployment. These improvements are discussed in detail below.

1. Using Ensemble Methods to Combine Multiple Classifiers

Ensemble learning has emerged as one of the most effective strategies in modern machine learning for improving predictive performance. Instead of relying on a single classifier, ensemble methods combine multiple models to leverage their individual strengths while mitigating weaknesses.

Importance:

- Combining classifiers can significantly reduce false positives (legitimate messages misclassified as spam) and false negatives (spam messages not detected).
- Ensemble methods are better suited to handle complex and feature-rich datasets, such as SMS messages that contain diverse structures, slang, and abbreviations.
- They provide greater stability in dynamic environments, where spam patterns frequently evolve.

Potential Techniques:

- **Bagging (Bootstrap Aggregating):** Multiple models are trained on random subsets of the dataset, and predictions are averaged. This reduces variance and prevents overfitting.
- **Boosting:** Models are trained sequentially, with each new model focusing on correcting errors made by previous models. Techniques like AdaBoost or Gradient Boosting can improve sensitivity to subtle spam features.
- **Stacking:** Predictions from multiple models serve as input to a higher-level model that learns the optimal way to combine predictions. This can yield better performance than any single model.

By integrating ensemble methods, the system could achieve higher accuracy and robustness, particularly in environments with diverse and constantly evolving spam patterns.

2. Incorporating Deep Learning Techniques such as LSTM or Transformers

While traditional machine learning models rely heavily on manually engineered features, deep learning models such as **Long Short-Term Memory (LSTM) networks** and **Transformers (e.g., BERT)** can automatically capture complex sequential patterns and contextual relationships in text data.

Advantages:

- **Contextual Understanding:** LSTM networks can retain information from previous words in a message, helping detect spam messages that use subtle phrasing or indirect patterns. Transformers analyze entire messages to identify contextual relationships between words, improving detection of sophisticated spam.
- **Scalability:** Deep learning models can be fine-tuned on larger datasets to detect new types of spam messages without requiring extensive manual feature engineering.
- **Adaptability:** Such models are capable of generalizing better to unseen or evolving spam messages, including those that attempt to evade traditional keyword-based filters.

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 11, November 2025
DOI: 10.17148/IJARCCE.2025.141176

Challenges:

- Deep learning models require more computational resources and memory compared to classical models.
- Large datasets are necessary to prevent overfitting and to ensure reliable generalization.
- Optimization is needed to maintain low latency for real-time applications.

Incorporating deep learning could elevate the performance of the system, particularly in detecting complex, context-dependent spam patterns that traditional models might miss.

REFERENCES

[1]. Hedieh Sajedi, Golazin Zarghami Parast, Fatemeh Akbari, "SMS Spam Filtering Using Machine Learning Techniques: A Survey," *Machine Learning Research*, Vol. 1, Issue 1, pp. 1–14, 2016.
[2]. Noura Al Moubayed, Toby Breckon, Peter Matthews, A. Stephen McGough, "SMS Spam Filtering using Probabilistic Topic Modelling and Stacked Denoising Autoencoder," *arXiv preprint*, 2016.
[3]. Sergio Rojas-Galeano, "Using BERT Encoding to Tackle the Mad-lib Attack in SMS Spam Detection," *arXiv preprint*, 2021.
[4]. Mohammad Amaz Uddin, Muhammad Nazrul Islam, Leandros Maglaras, Helge Janicke, Iqbal H. Sarker, "ExplainableDetector: Exploring Transformer-based Language Modeling Approach for SMS Spam Detection with Explainability Analysis," *arXiv preprint*, 2024.
[5]. Muhammad Salman, Muhammad Ikram, Nardine Basta, Mohamed Ali Kaafar, "SpaLLM-Guard: Pairing SMS Spam Detection Using Open-source and Commercial LLMs," *arXiv preprint*, 2025.
[6]. Thanniru Lakshman, Singarapu Sanjay Kumar, Ulligaddala Satish Kumar, Yenikepalli Sri Sekhar, Yellamati Suresh, "SMS spam detection in Machine Learning using Natural Language Processing," *International Journal of Advance Research, Ideas and Innovations in Technology (IJARIIT)*, Vol. 9, Issue 5, 2023.
[7]. C. Naveenkumar, K. Venkataramana, "SMS SPAM Filtering With Machine Learning Model DistilBERT," *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, Vol. 11, No. 3, pp. 170–175, 2025.
[8]. Manjunatha P.V., Sri Narahari C.N., Sriram Lakshmi Narasimha, Tarun Muthyala, Rakshith R., "SMS Spam Detection Using Deep Learning," *International Journal of Advanced Research in Computer and Communication Engineering (IJARCCE)*, (peer-reviewed), 2023.
[9]. Manas Ranjan Bishi, N. Sardhak Manikanta, G. Hari Surya Bharadwaj, P. Siva Krishna Teja, G. Rama Koteswara Rao, "Optimizing SMS Spam Detection: Leveraging the Strength of a Voting Classifier Ensemble," *International Journal of Intelligent Systems and Applications in Engineering (IJISAE)*.