INTERNATIONAL JOURNAL OF SCIENTIFIC RESEARCH IN ENGINEERING AND MANAGEMENT (IJSREM)
VOLUME: 10 ISSUE: 03 | MARCH - 2026 SJIF RATING: 8.659 ISSN: 2582-3930

Spam Message Filtering Using Machine Learning

Dharshanaa Sree T[1], Gana Sri MS[2], SwethaM. J[3], Vishmitha. T[4], Thamaraiselvi K[5]

Department of CSE, School of Engineering, Avinashilingam Institute for Home Science and Higher Education for Women, Coimbatore 18.
22ueo014@avinuty.ac.in, 22ueo016@avinuty.ac.in, 22ueo057@avinuty.ac.in 22ueo062@avinuty.ac.in, thamaraiselvi_cse@avinuty.ac.in

ABSTRACT
Spam messages have significantly increased as a result of the quick growth of digital communication platforms, posing major security risks like phishing attacks, malware distribution, fraudulent links, and invasive ads. These risks have an adverse effect on system dependability, user privacy, and general communication effectiveness. The design and implementation of an intelligent spam message filtering system that can automatically identify and categorise spam in a variety of input formats, such as plain text messages, URLs, and uploaded documents, is presented in this mini-project. To find linguistic and probabilistic patterns frequently linked to spam content, the suggested system combines Natural Language Processing methods with the Naive Bayes machine learning algorithm. To improve classification performance and lower dataset noise, text pre-processing techniques like tokenisation, normalisation, stop-word removal, and TF-IDF feature extraction are used. A lightweight web application with an interactive and user-friendly interface is created to offer real-time predictions. Experimental findings outperform traditional rule-based filtering techniques in terms of accuracy, precision, and recall. The system can adjust to changing spam patterns and is scalable and computationally efficient. All things considered, the suggested solution enhances the security of digital communications, reduces undesired content, and provides a useful, affordable framework for automated spam detection in contemporary messaging environments.
Keywords: Web application, machine learning, TF-IDF, NLP, Naive Bayes, and spam detection.

1.INTRODUCTION

In today's digital communication environment, spam messages have grown to be a significant problem that affects online services, messaging apps, and emails. These messages frequently include time-wasting and security-threatening links, phishing attempts, fraudulent schemes, or pointless ads. Because spammers are always changing their strategies to get around these filters, traditional spam detection systems that rely on set rules or keyword matching are no longer effective.

Machine learning and natural language processing methods offer clever and flexible ways to get around these restrictions. These systems automatically categorise messages as legitimate or spam by identifying patterns in massive datasets. Because of its ease of use, speed, and high accuracy in text classification tasks, the Naive Bayes classifier is a popular choice among different classification algorithms.

The project creates a web-based spam filtering system that can examine documents, URLs, and plain text. In order to predict spam in real time, the application preprocesses inputs, extracts significant features, and applies trained models. The goal of the solution is to increase overall communication efficiency while offering safe, scalable, and dependable spam detection.

2.OBJECTIVE

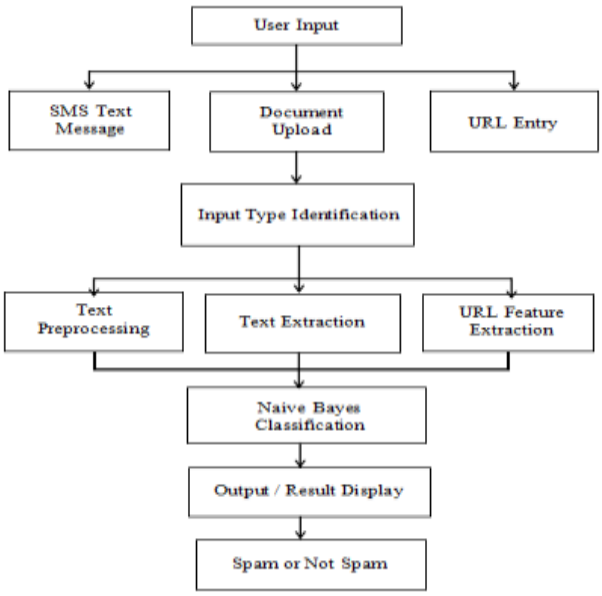
The Spam Message Filtering system's main goal is to create an intelligent, dependable, and automated system that correctly identifies and categorises spam content in a variety of formats, such as text messages, URLs, and documents. By stopping malicious links, phishing attempts, and intrusive ads from reaching users, the system seeks to improve the security of digital communications. The application guarantees quick, accurate predictions in real time by utilising machine learning, natural language processing, and effective feature extraction techniques. For

better long-term performance, it also aims to offer a user-friendly web interface, scalability, and adaptability to changing spam patterns.

3.LITERATURE SURVEY

From simple keyword filtering to sophisticated machine learning and deep learning techniques, spam detection research has progressed. Although early rule-based systems were straightforward, they were unable to adjust to new spam tactics. Subsequently, detection accuracy was greatly increased by machine learning algorithms like
Random Forest, Support Vector Machines, and Naive Bayes.Semantic comprehension of messages was improved by natural language processing methods such as TF-IDF, Bag-of-Words, and word embedding's. The main goals of URL analysis techniques are to find phishing links, unusual patterns, and dubious domains. Additionally, recent research emphasizes the use of text extraction tools for document-based spam detection.
Overall, research shows that integrating machine learning with intelligent preprocessing and real-time deployment enhances spam filtering systems' accuracy, automation, and dependability.

4.ARCHITECTURE



5. METHODOLOGY

5.1 DATA COLLECTION:

Several trustworthy sources, such as SMS spam collections, malicious URL repositories, and sample document files, provided pertinent datasets. Supervised learning is made possible by the labelled spam and legitimate content in these datasets. Gathering balanced and varied data guarantees that the model learns a variety of spam patterns and enhances overall classification accuracy and generalization capacity.

5.2 PREPROCESSING:

Prior to analysis, raw input data is cleaned and standardized. Text is changed to lowercase and special characters, numbers, and unnecessary symbols are eliminated. Words are normalized using tokenization, stop-word removal, stemming, and lemmatization. These procedures improve consistency, lower noise, and boost the effectiveness of further feature extraction.
Feature Extraction:
Statistical analysis and methods like TF-IDF are used to produce meaningful features. Numerical vectors are created from text frequency, keyword presence, URL properties, and document attributes. By capturing significant patterns and relationships in the data, this representation makes it possible for machine learning algorithms to distinguish between spam and authentic content.

5.3 MODEL TRAINING:

A Naive Bayes classifier that learns probabilistic relationships between words and spam labels is trained using the extracted features. The model finds patterns connected to spam messages and computes class probabilities. It works well for text classification tasks and real-time spam detection due to its ease of use, speed, and efficiency.

5.4 EVALUATION:

Metrics like accuracy, precision, recall, and F1-score are used to gauge the model's performance. False positives and negatives are found by analyzing a confusion matrix. These evaluation methods help maintain balanced and consistent performance across various input types by ensuring the classifier reliably detects spam while minimizing errors.

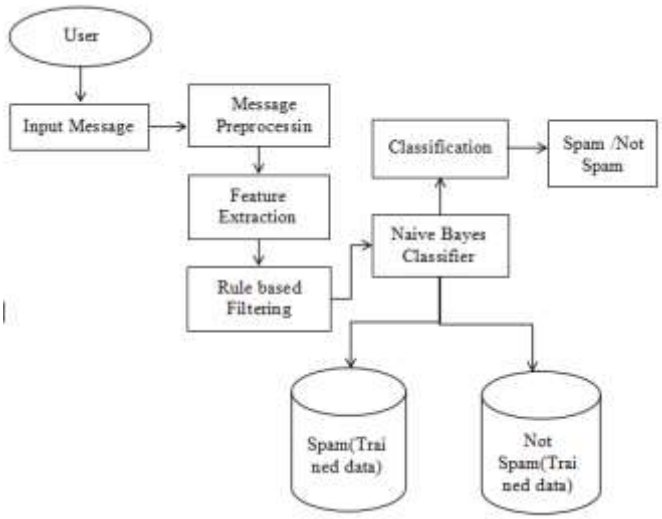
5.5 WEB INTEGRATION:

A Flask-developed web application incorporates the trained model. While the frontend offers users an interactive interface, the backend manages data processing and prediction. Through a user-friendly platform, this integration facilitates real-time spam detection and allows smooth communication between components.

5.6 TESTING:

Real-time user inputs, such as text messages, URLs, and document uploads, are used for thorough testing. To make sure the system is robust, different edge cases and uncommon formats are assessed. Continuous validation confirms that the application functions well in useful, real-world communication environments by verifying accuracy, dependability, and responsiveness.

6. BLOCK DIAGRAM:



7.TECHNICAL STACK

Python is used in the development of the Spam Message Filtering system for backend processing and model implementation; Flask is used to build the web application framework; SQLite is used for lightweight database management; and OpensCV is used to support document and content analysis. To guarantee dynamic rendering and responsiveness, the frontend interface is developed using HTML, CSS, JavaScript, and Jinja templating. Together, these technologies offer effective computation, smooth integration, scalability, and a user-friendly environment, allowing for dependable spam detection and real-time interaction while preserving high performance and low resource consumption.

8.FEATURES AND OUTCOMES

By fusing machine learning with clever preprocessing methods, the Spam Message Filtering system provides efficient and trustworthy unwanted message detection. In addition to its main features, the program has an easy-to-use web interface, multi-format input support for text, URLs, and documents, and real-time prediction. These

features give users a smooth and reliable communication experience by increasing usability, strengthening security, minimising manual labour, and guaranteeing consistent performance.

9.CONCLUSION

The Spam Message Filtering system shows how machine learning and natural language processing can be used to identify unwanted and harmful content. Real-time applications can benefit from the Naive Bayes classifier's quick and precise predictions. The system improves usability and dependability by providing a straightforward web interface and supporting several input formats. The project demonstrates how intelligent automation can greatly enhance communication security and lower the risks associated with spam. All things considered, the solution supports safe online interactions and provides a solid basis for further improvements.

9.ACKNOWLEDGEMENT

We, the students of CSE, want to express our heartfelt thanks to our guide, Assistant. prof. **Thamaraiselvi k,** ***for their essential help, support, and encouragement during this project. We also appreciate the Department of CSE at Avinashilingam Institute for giving us the resources and facilities we needed. Lastly, we would like to recognize our team members for their support and teamwork, which were crucial to finishing this work successfully.***

10.REFERENCES

1."SMS Spam Filteration Using Text Features and Supervised Machine Learning Algorithms," International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSCSEIT), Vol. 10, No. 6, 2024, pp. 641–651, Pandey Rashmi, Pushpendra Prajapati, Vibhanshu Kumar Singh, Mayank Tyagi, and Chetan Anand Amb.

2."An Enhanced SMS Spam Detection Framework Using Blockchain and Machine Learning," International Journal of Intelligent Systems and Applications in Engineering, Vol. 12, No. 22s, 2024, pp. 728–736, Gedam Ravi H.

3."Optimizing SMS Spam Detection: Leveraging the Strength of a Voting Classifier Ensemble," International Journal of Intelligent Systems and Applications in Engineering, Vol. 12, No. 3, 2024, pp. 2458–2469, Sardhak Manikanta N., G. Hari Surya Bharadwaj, P. Siva Krishna Teja, G. Rama Koteswara Rao, and M. R. B.

4.Applied Sciences, Vol. 14, No. 24, 2024; Altunay Hakan Can and Zafer Albayrak, "SMS Spam Detection System Based on Deep Learning Architectures for Turkish and English Messages."

5.Electronics, Vol. 15, No. 4, 2026. Şahin Meryem Soysaldı, Durmuş Özkan Şahin, and Areej Fateh Salah, "Revisiting SMS Spam Detection: The Impact of Feature Representation on Classical Machine Learning Models."

6.Applied Sciences, 2024; Altunay Hakan Can and Zafer Albayrak, "SMS Spam Detection System Based on Deep Learning Architectures for Turkish and English Messages."

7."SMS Scam Detection Application Based on Optical Character Recognition for Image Data Using Unsupervised and Deep Semi-Supervised Learning," Sensors, Vol. 24, 2024. Shinde Anjali, Essa Q. Shahra,

Shadi Basurra, Faisal Saeed, Abdulrahman A. AlSewari, and Waheb A. Jabbar.

8.Gudkova, D., N. Demidova, M. Vergelis, and T. Shcherbakova. (2017) Phishing and Spam in Q3 2017 12.Kaspersky Lab's Cyberthreat Research and Reports are called Securelist.Accessed on April 10, 2018, from https://securelist.com/spam-and-phishing-in-q3-2017/82901/.

9.Shruti Ranjan, Prayati Garhwal, Anupama Bhan, Monika Arora, and Anu Mehra, "Framework For Image Forgery Detection And Classification Using Machine Learning," 2nd International Conference on Trends in... 2018.

10.Androutsopoulos et al. An Experimental Comparison of Naïve Bayesian and Keyword-Based Anti-Spam Filtering with Personal Email Messages. http://www.aueb.gr/users/ion/publications.html.The URL is http://www.aueb.gr/users/ion/publications.html.

11.Salvi Siddhi, Ravindra, et al. "Phishing Website Detection Based on URL." International Journal of Scientific Research in Computer Science, Engineering, and Information Technology (USRCSEIT) 7.3 (2021): 589 594.

12.Androutsopoulos et al. An Experimental Comparison of Naïve Bayesian and Keyword-Based Anti-Spam Filtering with Personal Email Messages.