International Journal of Advanced Research and Review

SMS Spam Classifier

[1]Dr. G. Prasuna, [2]Kondru Anikshema, [3]Mathi Mohan Krishna Siva Sairam, [4]Kavuri Efefra

[1]Associate professor, COMPUTER SCIENCE AND ENGINEERING, St.Ann's College of Engineering and Technology, Nayunipalli (V), Vetapalem (M), Chirala, Bapatla Dist, Andhra Pradesh – 523187, India

[2,3,4]U. G Student, Dept COMPUTER SCIENCE AND ENGINEERING, St.Ann's College of Engineering and Technology, Nayunipalli (V), Vetapalem (M), Chirala, Bapatla Dist, Andhra Pradesh – 523187, India

ABSTRACT

Mobile communication is widely used for personal and professional purpose. spam SMS message have become a major problem, causing annoyance and security risks. These messages can lead to privacy issues and financial losses. The SMS Spam Classifier project aims to automatically detect and block such messages. It uses Machine Learning and NLP techniques for accurate classification. A naive bayes model is trained on SMS data after preprocessing text with tokenization, stop word removal, and TF-IDF vectorization. Users can access the system through a web interface to check messages and view spam statistics. The main goal is to provide a fast, reliable, and user-friendly solution to reduce spam SMS.

Keywords SMS Spam Detection, Machine learning, NLP, naive bayes classifier, Spam filtering, Text Classification, Flask Web Application

INTRODUCTION

In today's digital world of era, mobile communication is essential for personal and professional interactions. However, the rise of unsolicited and fraudulent text messages, commonly known as spam SMS, poses significant challenges for users. These messages can cause inconvenience and may lead to privacy breaches, financial losses, and security threats. The SMS Spam Classifier project addresses this issue by leveraging Machine Learning and Natural Language Processing techniques to automatically detect and filter spam messages. The system uses a Multinomial Naive Bayes classifier trained on a labeled SMS dataset. Text data is preprocessed through tokenization, stop word removal, punctuation stripping, and TF-IDF vectorization to convert it into a numerical form suitable for ML. The trained model is deployed on a Flask backend, providing REST APIs for real-time classification and analytics. An intuitive web interface built with HTML, CSS, Bootstrap, and JavaScript allows users to input messages and instantly view classification results with confidence scores. The dashboard displays prediction statistics, spam/ham distribution, and recent classification history. By combining robust ML

algorithms with a user-friendly interface, the SMS Spam Classifier provides an efficient, scalable, and real-time solution to combat spam messages effectively.

LITERATURE SURVEY

Several studies have explored SMS spam detection using different methods. Almeida et al. (2011) created the widely used SMS Spam Collection Dataset. Keyword-based approaches often missed contextual spam, while classical ML models required manual feature engineering. Hidalgo et al. (2006) applied tokenization, stopword removal, and TF-IDF for spam filtering. Deep learning methods like LSTM (Zhang et al., 2018) improved accuracy but needed large datasets and high computation. The proposed SMS Spam Classifier uses NLP, TF-IDF, and a Multinomial Naive Bayes model to provide accurate, efficient, and real-time spam detection suitable for web applications. It can adapt to evolving spam patterns and offers a user-friendly interface for instant message classification, making it practical for daily use.

RELATED WORK

Over the years, several approaches have been proposed for SMS and email spam detection. Almeida et al. (2011) introduced the UCI SMS Spam Collection Dataset, which has become a benchmark for research. Early methods relied on keyword-based filtering, but these often failed to capture the context of messages. Classical machine learning approaches, such as SVM, Decision Trees, and Naive Bayes, improved accuracy but required manual feature selection and engineering (Sakkis et al., 2003; Hidalgo et al., 2006). Deep learning techniques, including LSTM networks, have been applied to detect sequential patterns in spam messages (Zhang et al., 2018), achieving high accuracy but demanding large datasets and high computational resources. Other studies, like Cormack (2008), explored email spam filtering methods that are applicable to SMS, such as token frequency analysis, n-grams, and Bayesian learning. In comparison, the proposed SMS Spam Classifier leverages NLP, TF-IDF vectorization, and Multinomial Naive Bayes to deliver an efficient, real-time, and user-friendly spam detection system suitable for web-based applications.

EXISTING METHOD

Over the years, several methods have been developed to detect SMS spam. Early approaches used keyword-based filtering, which identified spam messages containing specific words but often failed with obfuscated or new spam patterns. Classical machine learning techniques, such as Support Vector Machines, Decision Trees, and Naive Bayes, improved detection accuracy but required manual feature extraction. Text preprocessing methods, including tokenization, stopword removal, stemming, and TF-IDF vectorization, converted messages into numerical features suitable for classification. Deep learning approaches, like LSTM networks, captured sequential dependencies in messages and achieved high accuracy but needed large datasets and substantial computational resources. While these methods laid the foundation for spam detection, many faced limitations in adapting to evolving spam patterns or providing real- time results. The proposed SMS Spam Classifier overcomes these challenges by using NLP, TF-IDF, and a Multinomial Naive Bayes model, offering efficient, accurate, and scalable spam detection suitable for web-based applications.

PROPOSED METHOD

The proposed SMS Spam Classifier detects spam messages efficiently and accurately in real time. Incoming SMS text is preprocessed using NLP techniques, including lowercasing, punctuation and stopword removal, tokenization, and stemming. TF- IDF vectorization converts the text into numerical features, which are fed into a Multinomial Naive Bayes classifier trained on a labeled SMS dataset. The model learns patterns and word frequencies to distinguish Spam from Ham messages. A Flask-based backend hosts the model and provides APIs for real-time classification. The web interface, built with HTML, CSS, Bootstrap, and JavaScript, allows users to input messages and view predictions with confidence scores. A dashboard displays spam/ham distribution, statistics, and recent classification. This method combines NLP, ML, and a user-friendly interface to provide scalable and accurate spam detection.

SYSTEM ARCHITECTURE

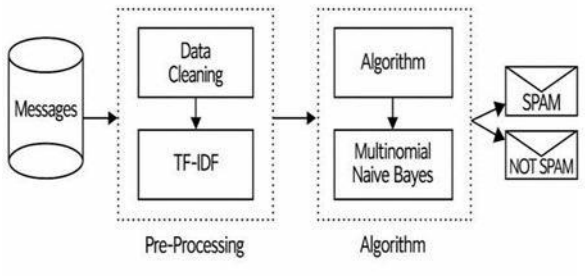

Fig.1: Architecture of SMS Spam Classifier

METHODOLOGY DESCRIPTION

Input Collection: SMS messages are collected from users or publicly available datasets for analysis.

Preprocessing: The text is cleaned by converting to lowercase, removing punctuation, special characters, numbers, and stopwords, followed by tokenization and stemming or lemmatization.

Feature Extraction: The cleaned text is transformed into numerical features using TF-IDF (Term Frequency–Inverse Document Frequency) to identify the most significant words in each message.

Model Training: The numerical features are used to train a Multinomial Naive Bayes classifier, which learns patterns that distinguish Spam from Ham messages.

Database Storage: All processed messages, predictions, and confidence scores are securely stored in a database for future reference and analysis.

Prediction: New SMS messages are preprocessed and vectorized similarly, and the trained model predicts whether each message is Spam or Ham along with a confidence score.

Web Application Interface: A user-friendly web interface built with HTML, CSS, Bootstrap, and JavaScript allows users to input messages and view instant classification results.

Dashboard & Analytics: The system provides a dashboard showing spam/ham distribution, prediction statistics, and recent message history.

Outcome: This approach ensures fast, accurate, and scalable spam detection, improving user safety and

communication efficiency.

RESULTS AND DISCUSSION

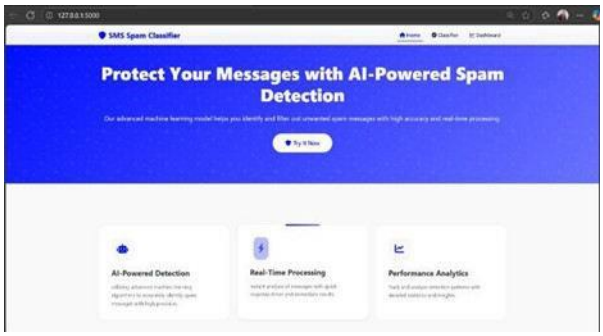

Fig.2: Application Home Page

This is the homepage of the SMS Spam Classifier web application, featuring a clean interface with AI-powered spam detection, real-time processing, and performance analytics. Users can easily access the classifier and dashboard to analyze messages and view detection results.

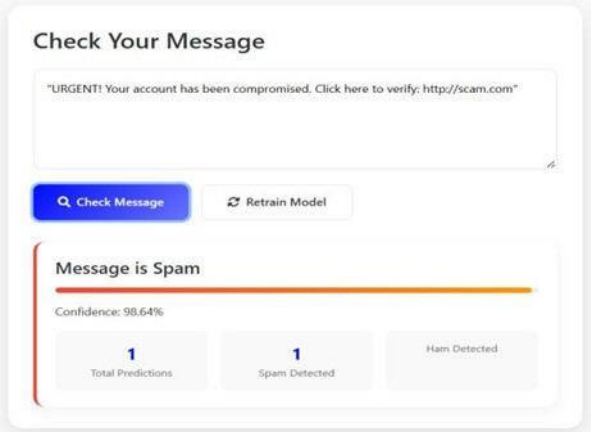

Fig. 3: Spam Message Detection

The model predicts "Message is Spam" with 98.64% confidence, updating the statistics to reflect total predictions and spam detections.

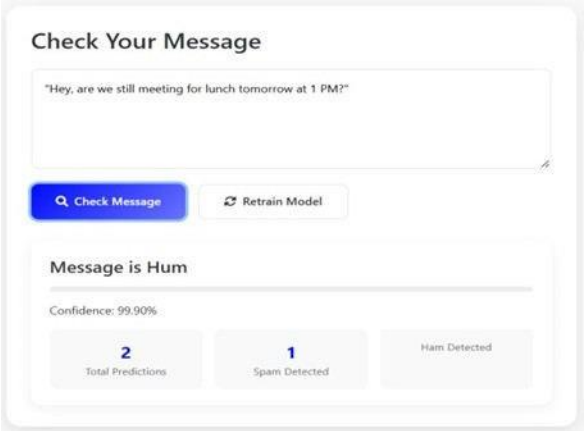

Fig. 4: Hum Message Detection

This figure shows the classification output for a non-spam (ham) message entered by the user. The system predicts the message category along with a confidence score and updates the dashboard with the total predictions, spam count, and ham count.

CONCLUSION

The SMS Spam Classifier effectively applies Machine Learning (ML) and Natural Language Processing (NLP) to automatically detect and filter spam messages, converting raw SMS data into meaningful features using tokenization, stopword removal, and TF-IDF vectorization. Results analysis shows high accuracy in real-time testing, correctly classifying spam and ham messages while providing instant predictions for users.

FUTURE SCOPE

For future enhancement, advanced techniques such as BERT transformers or LSTM/GRU models can improve contextual understanding, while continuous dataset updates, cloud deployment, and a mobile app interface can enhance adaptability, scalability, and accessibility for evolving spam patterns.

REFERENCES

1. Harini, D. P. (2012/9). codes: A Collaborative spam Detection system with a novel E-mail abstraction scheme. IDSR Journal of Engineering.

2. Bishi, M. R., Manikanta, N. S., Bharadwaj, G. H. S., Teja, P. S. K., & Rao, G. R. K. "Optimizing SMS Spam Detection: Leveraging the Strength of a Voting Classifier Ensemble." *International Journal of Intelligent Systems and Applications in Engineering (IJISAE)*, 2022.
3. Ranjith Reddy, K., & Joshi, G. "Spam Detection of SMS Messages Using Random Forest Classifier Algorithm." *Journal of Harbin Engineering University*, vol. 44 no. 8, 2023.
4. Lakshman, T., Sanjay Kumar, S., Satish Kumar, U., Sri Sekhar, Y., & Suresh, Y. (2023). SMS spam detection in machine learning using natural language processing. *International Journal of Advance Research, Ideas and Innovations in Technology*, 9(5).
5. Pandey, R., Prajapati, P., Singh, V. K., Tyagi, M., & Amb, C. A. (2024). SMS spam filtration using text features and supervised machine learning algorithms. *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*.
6. Gadde, S. (2021). SMS spam detection using machine learning and deep learning techniques. Project Report, Sathyabama University.
7. Methre, A., & Veena, K. (2021). SMS spam detection. *EasyChair Preprint*, 5166.
8. Choudhary, E., Verma, B., Choudhary, A., Sengar, A., & Agarwal, A. (2023). Spam SMS prediction using machine learning. *International Journal for Research in Applied Science and Engineering Technology (IJRASET)*.
9. Li, Y., Zhang, R., Rong, W., & Mi, X.
10. (2024). SpamDam: Towards privacy-preserving and adversary-resistant SMS spam detection. *arXiv Preprint*.
11. Ahmadi, M., Khajavi, M., Varmaghani, A., Ala, A., Danesh, K., & Javaheri, D. (2025). Leveraging large language models for cybersecurity: Enhancing SMS spam detection with robust and context-aware text classification. *arXiv Preprint*.
12. Resende, A., Railsback, D., Dowsley, R., Nascimento, A. C. A., & Aranha, D. F. (2021). Fast privacy-preserving text classification based on secure multiparty computation. *arXiv Preprint*.
13. Message spam identification by Naive Bayes classifier algorithm using machine learning. (2024). *International Journal of Innovative Science and Research Technology (IJISRT)*, 9(3).
14. Machine learning techniques on mobile SMS spam. (2024). *Atlantis Press*.
15. An improved machine learning-based short message service spam detection system. (2019–2024). *I.J. Computer Network and Information Security*, 12.
16. SMS spam detection using TensorFlow in Python. (2025). *GeeksforGeeks Tutorial*.
17. SMS spam classifier – Machine Learning. Kishan Kumar. Project Blog.
18. SMS spam detector – sms-spam-classifier. GitHub Project Repository.
19. Simbal, H. K., et al. (2024). SMS spam detection. *International Journal for Multidisciplinary Research (IJFMR)*.
20. SMS spam detection – Secure classification for short texts. General Survey Article.