International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 04 | April - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

SMS Spam Detection using Naive Bayes and Test Processing

Md Manjar Alam [1], Kumari Sonam [2], Rakesh Kumar [3]

[1] B. Tech, CSE Dept., R. V. S College of Engg. & Tech., Jamshedpur
[2] Professor, CSE Dept., R. V. S College of Engg. & Tech., Jamshedpur
[3] B.Tech, CSE Dept., R. V. S College of Engg. & Tech., Jamshedpur

Abstract
The mobile user growth is happening at a rapid pace. This leads to an increased trend in mobile messaging, which in turn strengthens the existence of SMS spam. Unlike messaging apps like WhatsApp or Facebook Messenger, SMS does not require an active internet connection. Unsolicited and sometimes dangerous, spam messages pose a significant problem within mobile communications; their primary intent is the dissemination of electronic content aimed at commercial promotion or monetary prize. For this reason, now known as principle protection of ensuring true integrity within telecom communication channels, providing solutions for SMS spam is very important. The inadequacy of current email filtering techniques would be attributed to several problems like the lack of complete folders for storing all your SMS spams and having less message structure formality and features available.The proposed method includes multiple components such as dataset integration, data preprocessing, exploratory data analysis, and feature selection. Furthermore, various machine learning models, including Naive Bayes and Support Vector Machine (SVM), are evaluated for constructing an efficient classifier. The primary objective of SMS spam detection is to safeguard users from the risks associated with spam messages.

Keywords: Spam SMS, Social Media, WhatsApp, Internet Connectivity, Financial Fraud, Datasets, Data Preprocessing, Feature Engineering, Naïve Bayes, Model Development.

1. Introduction

The convenience of access and popularity of SMS have rendered it a favorite target for malicious activities, resulting in unnecessary expenses for mobile users and a threat to secure mobile message communication. Numerous individuals and companies exploit this medium to send bulk unsolicited messages, popularity known as Spam SMS.

This project is aimed at designing an efficient SMS spam filter system based on Machine Learning approaches. Different ML algorithms, such as Naïve Bayes, Support Vector Machines (SVM), and Random Forests, will be used to investigate in order to analyze and classify SMS messages on the basic of their textual content, linguistic style, and other corresponding features. By using strict training and testing procedures, the objective is to create a very accurate and effective spam filter model capable of identifying nuanced feature specific to spam messages.

Machine learning offers a very feasible technique for the automation of spam detection through pattern recognition and characteristics gained from labeled data. Numerous models, including Naïve Bayes, SVM, and neural networks, can be trained on text-based features like word frequency analysis, n-grams, and semantic structures. Normalization and tokenization are part of the preprocessing that helps in feature engineering too. In other words, they help shape how much better or more effective a system to find spam is.

By continuously updating these models with new data, the system of SMS spam detection can evolve along with the changing techniques of spamming and provide the users a dependable mechanism for filtering out unwanted messages while ensuring smooth and secure communication.

2. Literature Survey

The convenience of access and popularity of SMS have rendered it a favorite target for malicious activities, resulting in unnecessary expenses for mobile users and a threat to secure mobile message communication. Numerous individuals and companies exploit this medium to send bulk unsolicited messages, popularity known as Spam SMS.

This project is aimed at designing an efficient SMS spam filter system based on Machine Learning approaches. Different ML algorithms, such as Naïve Bayes, Support Vector Machines (SVM), and Random Forests, will be used to investigate in order to analyze and classify SMS messages on the basic of their textual content, linguistic style, and other corresponding features. By using strict training and testing procedures, the objective is to create a very accurate and effective spam filter model capable of identifying nuanced feature specific to spam messages.

To mitigate these issues, researchers have come up with pattern-based ML approaches that sift through SMS messages and classify them into either spam or legit messages. A number of ML models have been implemented into spam classification systems including Naïve Bayes, SVM, Decision Tree, Random Forest, and even Neural Networks. They rely on text-based features like word counts, n-grams, and semantic analysis for detection accuracy.
One of the benchmarks datasets in SMS spam classification research is SMS Spam Collection Dataset, which consists of spam and ham messages that are tagged. Research conducted on this dataset has shown that Naïve Bayes' classifiers are efficient owing to their use of probability, but fail at sophisticated spam messages with hidden keywords. SVMs have been seen to give better generalization with the use of feature selection techniques like TF-IDF.
Besides, Recurrent Neural Networks (RNNs) and Convolutional Neural Networks (CNNs) have also been studied for SMS spam detection using deep learning approaches. These models are able to learn the representation of the text and extract complex spam features automatically. Nonetheless, they demand enormous computational resources and a great deal of labeled data for training. Other researchers have used NLP techniques such as word embeddings and sentiment analysis to improve spam detection performance.
Despite the high accuracy achieved by the ML-based SMS spam detection systems, there are still some issues to overcome like dealing with imbalanced datasets, adapting to changes in spam patterns, and reducing the rate of legitimate messages incorrectly flagged as spam. To enhance detection robustness, other researchers have proposed combining several ML models, referred to as ensemble methods. Also, some have adopted hybrid methods that use rule-based systems alongside ML approaches to improve spam classification efficiency.
Over time, some improvements have been proposed. In particular, Gómez Hidalgo et al. fused Naïve Bayes with a rule-based filtering system to increase spam detection accuracy. Almeida et al. showed that the performance of ML models is greatly improved through the use of feature selection. To tackle the problem of class imbalance in spam datasets, Bahnsen et al. introduced a learning approach that is sensitive to costs. Furthermore, there has been research into cloud-based real-time spam detection systems to improve scalability and speed of processing.
According to the literature, current approaches to classifying SMS spam can be segmented into three main groups: keyword-based techniques, ML techniques, and mixed-approach techniques. Traditional methods always rely on specific templates, forgetting to adjust to the changing nature of spam. The statistical nature of ML approach enables greater adaptability. Employing multiple approaches to achieve a better overall result defines hybrid methods.
To develop an efficient SMS spam detection system, researchers emphasize the importance of feature engineering, dataset quality, and model interpretability. Future research directions include adversarial learning to combat changing spam strategies, real-time detection frameworks for broad deployment, and explainable AI for model transparency.
By utilizing machine learning methods and continuously improving categorization models, SMS spam detection systems can provide a workable solution to combat the growing problem of unsolicited messages while preserving safe and convenient communication.

3. Methodology

The development of the SMS Spam Classifier is carried out using Python and machine learning techniques to distinguish between spam and non-spam (ham) messages. The methodology involves multiple stages, including data collection, preprocessing, feature extraction, model selection, training, evaluation, and deployment.

A. Data Collection

To construct a reliable spam classifier, a dataset consisting of labeled SMS messages is gathered. Commonly used publicly available datasets include:

- SMS Spam Collection Dataset: A widely referenced dataset containing SMS messages labeled as spam or ham.
- NUS SMS Corpus: Frequently utilized in text classification research, this dataset includes spam and ham-labeled messages.
- Kaggle Datasets: Several datasets on Kaggle are explored to ensure high data integrity and accurate classification labels.

B. Data Preprocessing

To enhance data quality and improve model performance, several preprocessing steps are applied:

- Tokenization: Each SMS message is split into individual words or tokens.
- Text Cleaning: Removal of stopwords, punctuation, and unnecessary characters to minimize noise.
- Handling Missing Values: While missing values are uncommon in SMS datasets, appropriate techniques are used to manage them if encountered.
- Vectorization: Textual data is converted into numerical representations using methods like TF-IDF (Term Frequency-Inverse Document Frequency) or word embeddings.

C. Feature Engineering

Feature extraction plays a crucial role in distinguishing spam from non-spam messages. Key features considered include:

- Message Length: Spam messages often contain more words than non-spam messages.
- Presence of Specific Keywords or Patterns: Words such as "win," "free," "offer," and "urgent" frequently appear in spam messages.
- Word Frequency Analysis: Identifying frequently occurring terms that may indicate spam.

D. Model Selection

Several machine learning algorithms are evaluated for text classification, including:

- Naïve Bayes: Well-suited for text-based classification problems.
- Support Vector Machines (SVM): Effective in handling high-dimensional data.
- Logistic Regression: A straightforward yet efficient model for spam detection.
- Decision Trees: Useful for capturing patterns in text data.
- Random Forests: An ensemble technique that improves prediction accuracy.
- Gradient Boosting Machines (GBM): Enhances model performance through iterative learning.

E. Model Training

The dataset is divided into training and testing subsets. The selected machine learning algorithm is trained on the training data to learn patterns and distinguish spam from non-spam messages.

F. Model Evaluation

The performance of the trained model is assessed using various evaluation metrics, including:

- Accuracy: Measures the proportion of correctly classified messages.
- Precision: Indicates the fraction of correctly identified spam messages.
- Recall: Evaluates the model's ability to detect spam messages.
- F1-Score: Provides a balance between precision and recall.
- ROC-AUC Score: Measures the model’s ability to differentiate spam from ham.

G. Hyperparameter Tuning

To optimize model performance, hyperparameter tuning techniques are employed. Cross-validation is used to test various configurations, minimizing overfitting and improving generalization.

H. Model Deployment

The trained model is deployed in a real-world environment for classifying SMS messages. Key deployment aspects include:

- User Interface: A web or mobile application to display classification results.
- User Feedback System: Allowing users to provide feedback on misclassified messages.
- Performance Monitoring: Generating insights and analytical reports to assess classification accuracy and trends.

The security aspects of the SMS Spam Classifier, including data integrity, reliability, and robustness, are carefully considered during its development. Ensuring a balance between accuracy and efficiency is crucial to maintaining the effectiveness and reliability of the system.

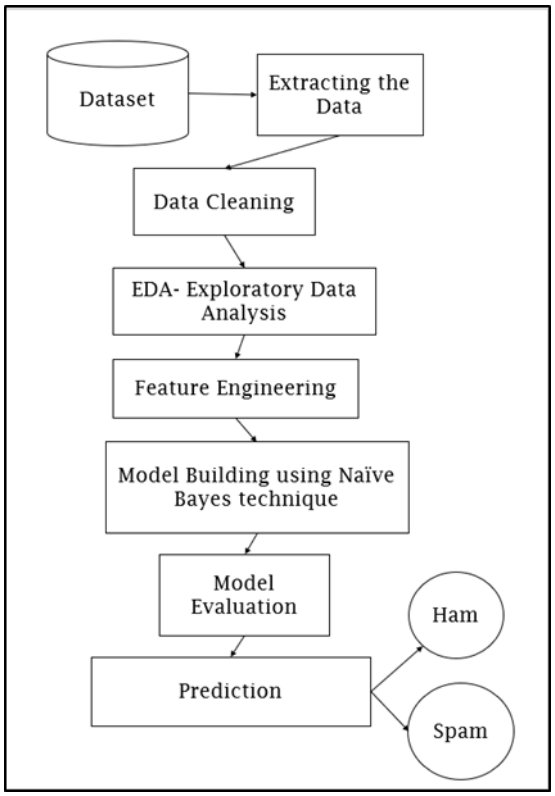

Fig1:Block Diagram Of SMS Spam Classifier

It is important to note that while the SMS Spam Classifier enhances message filtering accuracy, certain challenges remain, such as reducing false positives and ensuring adaptability to evolving spam techniques. To address these challenges and improve the classifier's effectiveness for real-world applications, including secure communication, fraud prevention, and automated message filtering, continuous research and refinement are necessary.

4. Conclusion and Future Work:

4.1 Conclusion :

Spam detection is very important for securing SMS and email communication. Achieving high accuracy in spam classification, though, is a challenging task, prompting researchers to investigate multiple detection methods. Machine learning based SMS spam detection offers a flexible and resilient solution that can be customized according to the requirements. Through ongoing model refinement and enhanced detection methods, we can keep pace with the dynamic nature of spam message so that users can enjoy increased security and a hassle-free messaging experience.

These sophisticated models exhibit outstanding accuracy and dependability in detecting spam message , , rendering them invaluable tools for successful SMS filtering.

4.2Future Work:

We plan to improve the SMS Spam classifier by employing sophisticated deep learning models and enhanced future selection methods. Improvement in the classification process will make it less expensive computationally, thus more suitable for real time applications. Further, incorporating privacy-preserving approaches will enhance security of data. The classifier may also be generalized to indentify spam in emails and social media with emphasis on mass deployment in businesses and mobile network.

4. References:

1.Modupe, A., O. O. Olugbara, and S. O. Ojo. (2014) ―Filtering of Mobile Short Messaging Communication Using Latent Dirichlet Allocation with Social Network Analysis‖, in Transactions on Engineering Technologies: Special Volume of the World Congress on Engineering 2013, G.-C. Yang, S.-I. Ao, and L. Gelman, Eds. Springer Science & Business. pp. 671–686.

2. Shirani-Mehr, H. (2013) ―SMS Spam Detection using Machine Learning Approach. p. 4.

3. Gudkova, D., M. Vergelis, T. Shcherbakova, and N. Demidova. (2017) ―Spam and Phishing in Q3 2017.‖ Securelist - Kaspersky Lab‘s Cyberthreat Research and Reports. Available from: https://securelist.com/spam-and-phishing-in-q3-2017/82901/. [Accessed: 10th April 2018].

4 Abdulhamid, S. M. et al., (2017) ―A Review on Mobile SMS Spam Filtering Techniques.‖ IEEE Access 5: 15650–15666.

5.Aski, A. S., and N. K. Sourati. (2016) ―Proposed Efficient Algorithm to Filter Spam Using Machine Learning Techniques.‖ Pac. Sci. Rev. Nat. Sci. Eng. 18 (2):145–149.

6. Narayan, A., and P. Saxena. (2013) ―The Curse of 140 Characters: Evaluating The Efficacy of SMS Spam Detection on Android.‖ p. 33– 42.

7. Almeida, T. A., J. M. Gómez, and A. Yamakami. (2011) ―Contributions to the Study of SMS Spam Filtering: New Collection and Results.‖ p. 4.

8. Mujtaba, D. G., and M. Yasin. (2014) ―SMS Spam Detection Using Simple Message Content Features.‖ J. Basic Appl. Sci. Res. 4 (4): 5.

9. Choudhary, N., and A. K. Jain. (2017) ―Towards Filtering of SMS Spam Messages Using Machine Learning Based Technique‖, in Advanced Informatics for Computing Research 712: 18-30.

10. Bauza R, Gozalvez J, Sanchez-Soriano J. Road traffic congestion detection through cooperative vehicle-to-vehicle communications. Paper presented at: Proceedings of the 2010 IEEE 35th Conference on IEEE Local Computer Networks (LCN); 2010:606–612.

11. Sajedi, H., G. Z. Parast, and F. Akbari. (2016) ―SMS Spam Filtering Using Machine Learning Techniques: A Survey.‖ Machine Learning, 1 (1): 14.

12. Xu, E., W. Xiang, Q. Yang, J. Du, and J. Zhong. (2012) ―SMS Spam Detection Using Noncontent Features.‖ IEEE Intell. Syst. 27(6): 44–51.

13. Chan, P. P. K., C. Yang, D. S. Yeung, and W. W. Y. Ng. (2015) ―Spam Filtering for Short Messages in Adversarial Environment.‖ Neurocomputing 155: 167–176.

14.Almeida. T. A., and J. M. G. Hidalgo. (2018) ―SMS Spam Collection.‖ Available from: http://www.dt.fee.unicamp.br/~tiago/smsspamcollection/. [Accessed: 11st April 2018].

15. Chan, P. P. K., C. Yang, D. S. Yeung, and W. W. Y. Ng. (2015) ―Spam Filtering for Short Messages in Adversarial Environment.‖ Neurocomputing 155: 167–176.