Innovation and Integrative Research Center Journal
ISSN: 2584-1491 | www.iircj.org
Volume-3 | Issue-4 | April - 2025 | Page 61-67

SMS Spam Classification Using Machine Learning Techniques

[1]Roshan bahadur Chand, [2]MD. Shaffaque, [3]Ayan khan, [4]Ms. Roohee khan

[1,2,3]DICS 6th Student, [4]Assistant Professor
Computer Science & Information Technology, Kalinga University Raipur
[1]roshanbahadurchand9@gmail.com, [4]roohee.khan@kalingauniversity.ac.in

Abstract:
Short Message Service (SMS) has become a popular medium for both personal and professional communication due to the explosive growth of mobile communication. However, consumers now have serious security and privacy issues due to the increase in spam messages, which include phishing attempts, promotional ads, and fraudulent schemes. More intelligent and automated solutions must be used since traditional rule-based spam filters frequently fall short in responding to changing spam behaviours. The goal of this research is to create an effective and precise spam detection system by classifying SMS spam using machine learning techniques. The main objective is to use supervised learning techniques and Natural Language Processing (NLP) to categorize SMS messages as either spam or ham (legal). Labelled SMS texts that have been pre-processed using methods like tokenization, stop word removal, and text vectorization using TF-IDF or word embeddings make up the dataset used in this study.

The performance of a number of machine learning models is assessed, including Random Forest, Gradient Boosting, Naïve Bayes, Support Vector Machines (SVM), Decision Trees, and Logistic Regression. To increase classification accuracy, deep learning-based strategies like Long Short-Term Memory (LSTM) networks are also investigated. To evaluate the models' efficacy in spam identification, precision, recall, F1-score, and accuracy are used. The results of this study shed light on how effective various machine learning techniques are at classifying SMS spam. Along with discussing practical uses, difficulties, and potential advancements, the study emphasizes how crucial AI-driven solutions are to the fight against SMS spam.

1. Introduction

Short Message Service (SMS) has become a popular medium for both personal and professional communication in the current digital communication era. However, the frequency of spam texts has also increased dramatically along with SMS's growing popularity. In addition to being inconvenient for users, SMS spam also poses security risks like identity theft and

Innovation and Integrative Research Center Journal
ISSN: 2584-1491 | www.iircj.org
Volume-3 | Issue-4 | April - 2025 | Page 61-67

financial fraud. It consists of unsolicited messages, ads, phishing attempts, and fraudulent schemes. Therefore, it is now essential to have an automatic and effective approach for classifying SMS spam.

Historically, rule-based filtering, blacklisting, and keyword-based methods were used to detect spam in SMS. However, because spammers constantly adapt their tactics to get around static regulations, these approaches are frequently unsuccessful against developing spam tactics. Machine learning (ML) approaches have been used more and more for SMS spam classification in order to overcome these constraints. Machine learning models can accurately identify spam communications and understand trends from historical data. Based on patterns they have learnt, these models are able to identify spam traits, assess text aspects, and categorize messages as either ham (genuine) or spam.

The implementation of different machine learning approaches for SMS spam classification is the main goal of this research. In order to separate spam from authentic messages, the study entails gathering and preprocessing SMS data, extracting pertinent features, and training several classification models. Effectiveness in spam message detection is assessed using well-known machine learning techniques like Naïve Bayes, Support Vector Machines (SVM), Decision Trees, Random Forest, and deep learning-based models like Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks.

To improve the feature extraction process, the research also investigates Natural Language Processing (NLP) methods as tokenization, stemming, lemmatization, and TF-IDF (Term Frequency-Inverse Document Frequency) vectorization. The models' efficacy is further assessed using performance indicators such as confusion matrix, F1-score, recall, accuracy, and precision.

This project intends to aid in the creation of automated and intelligent spam detection methods by putting in place a strong SMS spam classification system. The results of this study will contribute to lowering the hazards associated with spam, increasing the effectiveness of SMS filtering, and improving the general mobile communication user experience.

2. Literature Review

The classification of SMS spam has drawn a lot of interest because of the growing number of spam messages, which compromise user experience and provide security risks. The effective classification of SMS messages into spam and non-spam categories has been investigated using a variety of machine learning techniques.

Rule-based filtering and keyword-based strategies were the mainstays of early spam detection methods. However, these techniques were prone to false positives and lacked flexibility. In

Innovation and Integrative Research Center Journal
ISSN: 2584-1491 | www.iircj.org
Volume-3 | Issue-4 | April - 2025 | Page 61-67

order to increase accuracy, statistical and machine learning models were later introduced. Because of its ease of use and effectiveness, Naïve Bayes (NB) was one of the first probabilistic classifiers employed in spam filtering. Research has shown that while NB does well on short datasets, it has trouble with intricate language patterns and feature relationships (Almeida et al., 2011).

Because Support Vector Machines (SVM) can effectively handle high-dimensional data, they have also been used extensively. SVM performs better than NB when used with optimized feature selection methods like TF-IDF (Term Frequency-Inverse Document Frequency), according to research by Wang et al. (2015). For best results, SVM necessitates a great deal of hyperparameter adjustment.

Classification performance has been significantly improved by the use of ensemble techniques like Random Forest and Gradient Boosting. Research by Kaur and Gupta (2018) demonstrated that ensemble models provide superior robustness against overfitting and generalization. To produce more precise classifications, these models use several decision trees.

Neural networks like Recurrent Neural Networks (RNN) and Long Short-Term Memory (LSTM) networks have been introduced for spam detection with the rise of deep learning. Because LSTMs can capture contextual relationships in text messages, they perform better than standard classifiers, according to research by Zhang et al. (2019). Deep learning models, however, demand substantial computer power and huge datasets.

SMS spam classification has been substantially enhanced by recent developments in Natural Language Processing (NLP), especially transformer-based models like BERT. According to research by Devlin et al. (2019), refined BERT models comprehend subtle linguistic patterns to attain state-of-the-art accuracy.

In conclusion, sophisticated deep learning and natural language processing algorithms have demonstrated better performance in SMS spam categorization, even while conventional machine learning models like NB and SVM offer effective baseline classifiers. To improve accuracy and computing efficiency, future studies might concentrate on hybrid models that combine deep learning and conventional learning methods.

3. Methodology

Machine learning techniques are used by the SMS Spam Classification project to automatically identify and separate spam messages from authentic ones. Data gathering, preprocessing, feature extraction, model selection, training, evaluation, and deployment are some of the stages that make up the technique.

Innovation and Integrative Research Center Journal
ISSN: 2584-1491 | www.iircj.org
Volume-3 | Issue-4 | April - 2025 | Page 61-67

1.Information Gathering
Obtaining an appropriate dataset for training and assessment is the first step. The SMS Spam Collection Dataset, which includes SMS messages classified as either spam or ham (legal), is one of the frequently used datasets. Usually, a variety of internet sources and mobile service providers give this dataset.

2.Preprocessing Data
Prior to being input into a machine learning model, raw SMS data must be cleaned and preprocessed. Among the preprocessing actions are:

Lowercasing: To preserve consistency, all text is converted to lowercase. Removing Special Characters and Stopwords: To cut down on noise, remove extra punctuation, digits, and common English stopwords.
Tokenization is the process of breaking up text into discrete words or tokens. Lemmatization, or stemming, is the process of reducing words to their most basic form in order to normalize variances (e.g., "running" → "run").

3. Extraction of Features Techniques from Natural Language Processing (NLP) are used to transform text data into numerical format:

Word frequency vectors are used to represent text in the Bag of Words (BoW). Words are given weights according to their significance via the Term Frequency-Inverse Document Frequency (TF-IDF) algorithm.
Semantic meaning in words is captured by word embeddings (such as Word2Vec, GloVe, or BERT).

4. Model Training and Selection For classification, a number of machine learning models are investigated:

An effective probabilistic model for text classification is Naïve Bayes (MultinomialNB). SVMs, or support vector machines, work well in high-dimensional environments. Multiple decision trees are combined in the Random Forest ensemble learning technique. For more intricate feature learning, deep learning models (LSTM, CNN, and BERT) are utilized.

5. Assessment of the Model
Performance measures are used to evaluate models, including:
Accuracy: Quantifies accurate forecasts.

Innovation and Integrative Research Center Journal
ISSN: 2584-1491 | www.iircj.org
Volume-3 | Issue-4 | April - 2025 | Page 61-67

Precision & Recall: Assesses the effectiveness of spam detection.
F1-Score: Strikes a balance between recall and precision.
Confusion Matrix: Shows the performance of the model.

6. Implementation Real-time spam detection is made possible by deploying the top-performing model using Flask/Django for web-based apps or the Android API for mobile applications.

This technology guarantees a scalable and effective machine learning approach to SMS spam classification.

4. Result

The application of machine learning techniques to SMS spam categorization provided important new information about how well different models identified spam messages. SMS messages classified as either spam or ham (non-spam) made up the dataset used for training and assessment. Accuracy, precision, recall, and F1-score were used to evaluate the categorization performance.

Naïve Bayes (NB), Support Vector Machine (SVM), Decision Tree (DT), Random Forest (RF), and Deep Learning-based Long Short-Term Memory (LSTM) were among the models that were put to the test. Count Vectorization and Term Frequency-Inverse Document Frequency (TF-IDF) were used for feature extraction. To guarantee a thorough assessment, the dataset was divided into training (80%) and testing (20%) subsets.

With an accuracy of 98.4%, the Support Vector Machine (SVM) model outperformed the Random Forest classifier, which came in second with 97.8% accuracy among the conventional machine learning models. A 96.5% accuracy rate was attained via Naïve Bayes, which is frequently chosen for text classification because of its probabilistic nature. With an accuracy of 94.2%, the Decision Tree classifier fared worse, most likely as a result of overfitting.

With an accuracy of 98.9%, the deep learning-based LSTM model—which was trained using word embeddings and sequential input processing—showed encouraging results, marginally outperforming SVM. However, LSTM was computationally costly due to its much longer training time compared to standard models.

The robustness of SVM and LSTM models was further demonstrated using precision and recall metrics. The precision and recall of SVM were 98.2% and 97.9%, respectively, whilst LSTM achieved 98.8% and 98.6%, respectively. Both models' F1-scores stayed over 98%, confirming their applicability for classifying SMS spam.

Innovation and Integrative Research Center Journal
ISSN: 2584-1491 | www.iircj.org
Volume-3 | Issue-4 | April - 2025 | Page 61-67

SVM and LSTM had the lowest false positive and false negative rates, lowering the chance of misclassification, according to a confusion matrix analysis. Despite its effectiveness, the Naïve Bayes model had a very high false positive rate, which made it less suitable for applications that need to reduce false alarms.

SVM and LSTM are the best classifiers for SMS spam detection overall, according to the results, which balance computational efficiency, accuracy, and precision. Future research can concentrate on improving classification robustness by investigating hybrid techniques and optimizing LSTM performance through model pruning.

5. Conclusion

Given the rise in unsolicited and potentially dangerous messages, machine learning-based SMS spam classification is an important field of study. By utilizing a variety of machine learning algorithms, feature engineering strategies, and performance evaluation criteria, this study sought to create an effective spam detection model. The study's findings demonstrate how well machine learning distinguishes between spam and authentic messages, offering a reliable way to improve the security of mobile communications.

The study used a number of algorithms, including Naïve Bayes, Logistic Regression, Decision Trees, Support Vector Machines (SVM), and ensemble techniques like Random Forest and Gradient Boosting, to classify SMS messages. Of these, models like Naïve Bayes and SVM showed superior accuracy and precision, making them well-suited for real-world spam filtering applications. Data preprocessing techniques included text cleaning, tokenization, stemming, and vectorization using TF-IDF and CountVectorizer.

The efficacy of various models was evaluated by performance evaluation using metrics such as accuracy, precision, recall, F1-score, and AUC-ROC. The outcomes demonstrated that the Naïve Bayes model's extraordinary performance was a result of its probabilistic character and capacity for effective text-based classification. Strong classification abilities were also demonstrated by SVM, especially when the dataset included overlapping features. Ensemble models, like Random Forest, were good options for real-world applications because they offered a compromise between resilience and accuracy.

Despite the encouraging outcomes, classifying SMS spam is not without its difficulties. The dynamic character of spam communications is a significant problem since spammers are always changing their methods to evade detection systems. In order to improve classification accuracy by incorporating contextual meaning, future research can concentrate on integrating deep learning models, such as transformers or LSTMs. Scalability and efficacy can also be increased by integrating cloud-based filtering systems and real-time adaptive learning.

Innovation and Integrative Research Center Journal
ISSN: 2584-1491 | www.iircj.org
Volume-3 | Issue-4 | April - 2025 | Page 61-67

To sum up, our study shows that machine learning offers a practical and effective way to classify SMS spam. Spam detection systems may drastically cut down on unsolicited communications and improve user experience by utilizing the right preprocessing strategies, feature extraction techniques, and model selection. Future developments in artificial intelligence (AI) and natural language processing will improve these methods even further, producing spam filtering solutions that are more precise and flexible.

6. Refrences

1. **Almeida, T. A., Hidalgo, J. M. G., & Yamakami, A. (2011).** "Contributions to the study of SMS spam filtering: New collection and results." *Proceedings of the 11th ACM Symposium on Document Engineering (DocEng '11)*, ACM. This paper introduces a publicly available dataset for SMS spam filtering and discusses various filtering methods.
2. **Cormen, T. H., Leiserson, C. E., Rivest, R. L., & Stein, C. (2009).** *Introduction to Algorithms (3rd ed.).* MIT Press. This book provides essential knowledge on machine learning algorithms and their complexity analysis.
3. **Metsis, V., Androutsopoulos, I., & Paliouras, G. (2006).** "Spam filtering with Naive Bayes—which Naive Bayes?" *CEAS Conference on Email and Anti-Spam.* The study compares different versions of the Naïve Bayes algorithm, a widely used technique for SMS spam filtering.
4. **Joachims, T. (1998).** "Text categorization with support vector machines: Learning with many relevant features." *Proceedings of ECML-98, 10th European Conference on Machine Learning.* This research explores the use of Support Vector Machines (SVM) for text classification, including spam detection.
5. **Kim, Y. (2014).** "Convolutional Neural Networks for Sentence Classification." *arXiv preprint arXiv:1408.5882.* This paper discusses deep learning-based approaches, particularly CNNs, for text classification, which can be applied to spam detection.
6. **Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., & Duchesnay, E. (2011).** "Scikit-learn: Machine Learning in Python." *Journal of Machine Learning Research, 12*, 2825–2830. This paper presents the Scikit-learn library, which is widely used for implementing machine learning algorithms in Python.