WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Guarding Your Inbox: SMS Spam Detection through Naive Bayes Classification

Dora Devi Prasad*

*UG Student, GMR Institute of Technology, Rajam, Andhra Pradesh-532127, India. Email: 21341A1235@gmrit.edu.in

A B S T R A C T

One of the best services for exchanging messages with others is this SMS (Short Message Service). It is now functioning as unwanted or unsolicited communications that are received on mobile devices. These text message spams could include content threats, unsolicited promotions, and harmful links. The mobile subscribers may also experience a serious annoyance as a result of this spam. It also results in the unsolicited SMS mailbox filling up. Because of this, the service providers are also concerned that it may annoy their customers or perhaps result in them losing subscribers. Researchers have suggested many methods to detect and filter SMS spam, including naïve bayes (NB), Convolutional Neural Network (CNN) model, and Support Vector Machine (SVM), in an effort to limit this activity. In this paper, I evaluate the state of the art, issues, and potential approaches for Naïve Bayes (NB)-based spam detection systems. The purpose of the work is to support future researchers who wish to investigate the topic of SMS spam detection through machine learning.

Keywords: SMS (Short Message Service), Convolutional Neural Networks (CNN), Machine Learning, Support Vector Machine (SVM), Naïve Bayes (NB), Accuracy, SMS Spam detection, mitigation, filtration.

1. Introduction

In today's digital environment, when unsolicited and undesired SMS messages can be annoying and even dangerous, SMS spam identification is an essential responsibility. By filtering out these unsolicited texts, an efficient SMS spam detection system can shield consumers against malware, phishing scams, and other fraudulent activity. Classifying SMS communications as spam or ham (legal) is the aim of SMS spam detection. This entails dissecting the message's content to find trends and characteristics that set spam apart from legitimate communications. For this objective, machine learning and natural language processing (NLP) approaches are frequently used.

A dataset of labelled SMS messages, with each message assigned a label designating whether it is spam or ham, is used to train machine learning algorithms. The algorithm gains the ability to recognize links and patterns in the data that help it differentiate between the two categories. The model can be used to categorize new SMS messages and determine whether they are likely to be spam or ham once it has been trained. Language characteristics are extracted and analyzed from the SMS messages using NLP techniques. Word frequencies, sentiment analysis, n-grams (sequences of n words), and other language patterns are a few examples of these attributes. Spam identification is aided by the model's ability to analyze these traits, which provide deeper insights into the message's meaning and intent.

A dataset of labelled SMS messages, with each message assigned a label designating whether it is spam or ham, is used to train machine learning algorithms. The algorithm gains the ability to recognize links and patterns in the data that help it differentiate between the two categories. The model can be used to categorize new SMS messages and determine whether they are likely to be spam or ham once it has been trained. Language characteristics are extracted and analyzed from the SMS messages using NLP techniques.

1.A Spam Transformer Model for SMS Spam Detection

1.1 Objective

The identification of SMS spam has made extensive use of traditional machine learning techniques like logistic regression (LR), Random Forest (RF), Support Vector Machine (SVM), Naïve Bayes (NB), and Decision Trees (DT). In order to address the SMS spam issues, deep learning techniques such as Long Short-Term Memory (LSTM), Recurrent Neural Networks (RNN), and Convolutional Neural Networks (CNN) have also been used. Here, the researchers investigated the effectiveness of several classifiers, such as CNN, Neural Network, and Naïve Bayes, and discovered that CNN, NB, DT, LR, RF, AdaBoost, and SVM, all obtained excellent accuracy on the SMS spam collection.

1.2 Model Explanation

A basic model that is simply understood by everybody is the main architecture of the updated transformer model for SMS spam detection. Positional encoding is used for both the memory (trainable parameters) and the input message embeddings. Subsequently, the message vectors that have been processed are sent to the encoder layers, where self-attention is executed. The decoder layers receive the output from the encoder layers. Based on the outcomes of the encoder layers and the processed memory, the decoder layers perform Multi-Head Attention. Subsequently, the decoded vectors are routed through a few fully-connected liner layers and a final classification activation function.

1.3 Results would be

With values of 98.92% and 0.9451%, the suggested updated spam transformer model produced the best results for SMS spam identification in terms of accuracy, recall, and F1-score. Additionally, the model performed better than other contenders on the Twitter dataset, achieving accuracy, precision, recall, and F1-score values of 87.06%, 0.8576%, and 0.8576%, respectively. In comparison to earlier methods for SMS spam detection, the experimental findings demonstrated that the suggested spam transformed model performed better.

2. Spam filtering Method Based on Multi-Modal Fusion

2.1 Objective

Text-based, image-based, and multi-modal spam detection are the three categories into which the method divides the current approaches for detecting spam. Decision tree techniques, numerous weak classifiers, Bayesian algorithms, K-Nearest Neighbor algorithms, and deep learning algorithms are a few examples of text-based spam detection techniques. Convolutional neural networks, multi model combination techniques, supported vector machines, and multi feature combination techniques are used in image-based spam detection techniques. P-SVM, logistic regression, and multimodal feature fusion are some of the techniques used in multimodal spam detection. In order to efficiently filter spam in both the text and image sections of an email, the model suggests a novel model called multi modal architecture based on model fusion that combines a CNN model and an LSTM model. The model attains precision in the range of 92.64-98.48% and outperforms traditional spam filtering systems.

2.2 Model Explanation

Email preprocessing is a feature that extracts the text and image data from an email and creates separate datasets for the text and images. After that, the best classifiers are obtained, and the LSTM and CNN models are trained and optimized using the text and picture datasets, respectively—resulting in the best CNN and LSTM models. The image dataset may then be re-entered into the best CNN model to determine the classification probability values of the image dataset as spam. From there, we can obtain the classification probability values. To determine the text dataset's categorization probability values as spam, the text dataset is similarly re-entered into the best LSTM model. For an email with text-only content or Using dropout ideology and image data, we determine the relevant model output probability value, $p = 0.5$. The two classification probability values are then given into the fusion model in order to train and optimize it, resulting in the eventual creation of the optimal fusion model.

2.3 Results

Whether spam is buried in text or an image, the MMA-MF model achieved accuracies in the range of 92.64-98.48% in filtering it out. With AUC indications better than 0.93, the MMA-MF model demonstrated exceptional performance for spam identification on text, picture, and mixed datasets. Evaluation criteria, including F1-score, accuracy, recall, and precision, were employed to gauge how successful the suggested approach was.

3. Conclusion:

In conclusion, machine learning (ML) models for SMS spam identification have become a potent tool in the fight against the ubiquitous problem of unsolicited and undesired SMS messages. Support Vector Machines (SVMs), Naive Bayes, Random Forest, and neural networks are just a few of the machine learning (ML) models that have shown encouraging results in identifying SMS texts as real or spam. SMS spam detection performance has been significantly improved by deep learning approaches, with recurrent neural networks (RNNs) and convolutional neural networks (CNNs) exhibiting remarkably high accuracy rates. Additionally, we observed many hybrid and spam transformer model kinds. Hybrid models have also demonstrated promise, combining various ML techniques and utilizing the advantages of each model independently to increase detection accuracy.

4. REFERENCES

1.“A Review on Mobile SMS Spam Filtering Techniques” by Shafi’I Mahammad Abdulhamid’ ,(Member, IEEE), Muhammad Shafie ABD Latiff, Haruna Chiroma, (Members, IEEE), Oluwafemi Osho, Gaddafi Abdul-Salaam, Adamu I. Abubakar, (Member, IEEE), and Tutut Herawan. Zhou, Shaocong, et al. "Dynamic wireless power transfer system for electric vehicles employing multiplexing LCC modules with individual transmitters." IEEE Access 6 (2018): 62514- 62527.

2. “A Spam Transformer Model for SMS Spam Detection” by Xiaoxu LIU, Haoye LU, (Member, IEEE), and Amiya Nayak, (Senior Member, IEEE).

3. “A Spam Filtering Method Based on Multi-Modal Fusion” by Hong Yang, Qihe Liu, Shijie Zhouand Yang Luo.

4. “Short Message Service (Sms) Spam Detection and Classification Using Naïve Bayes” by Christine Bukola Asaju, Ekuma, James Ekorabon, Richard Ojochegbe Orah.

5. “A Hybrid CNN-LSTM Model for SMS Spam Detection in Arabic and English Messages” by Abdallah Ghourabi, Mahmood A. Mahmood and Qusay M. Alzubi.

6. “A Discrete Hidden Markov Model for SMS Spam Detection” Tian Xia 1 and Xuemin Chen.