International Journal of Research in Engineering and Science (IJRES)
ISSN (Online): 2320-9364, ISSN (Print): 2320-9356
www.ijres.org Volume 11 Issue 4 ǁ April 2023 ǁ PP. 470-473

A Survey Paper on Smishing Prediction System

Smitha C
Computer Science and Engineering Dept.
Global Academy of Technology
Banglore,India

Anusuva
Computer Science and Engineering Dept.
Global Academy of Technology
Banglore,India

Vidhisha M S
Computer Science and Engineering Dept.
Global Academy of Technology
Banglore,India

Vaishnavi R
Computer Science and Engineering Dept.
Global Academy of Technology
Banglore,India

Prof. Dr Kanagavalli R
Computer Science and Engineering Dept
Global Academy of Technology Bangalore,

Abstract-*Worldwide, SMS spam is a problem that is becoming worse. Unwanted and unwelcome SMS messages are those that are sent to mobile devices. These mails may contain advertisements, phishing scams, or even scams.SMS spam is a problem since it may be exceedingly unpleasant and, in some situations, even harmful. For instance, some spam communications may contain links to phishing or dangerous websites that seek to steal personal data.In order to filter out these spam communications and prevent them from getting to the end consumers, many experts have created various procedures. The majority of the solutions used machine learning techniques to filter spam messages, copying the strategy that sms spam filtering had been using successfully. Logistical regression, Naive Bayes algorithms are some of the well-liked machine learning methods that have been employed successfully. The goal of the current study is to identify the most effective method of spam message filtering by implementing these strategies and evaluating their efficacy. According to the results, the trained classifier model that uses a neural network to categorise incoming communications as spam or ham operates well.*

Keywords—*Review, SMS Spam, Spam, Mobile SMS.*

Date of Submission: 08-04-2023 Date of acceptance: 22-04-2023

I. INTRODUCTION

Generally, short message service(sms) is a one of the liked and also inexpensive communication service package. sms users are increasing day by day,sms services are available without any internet also that is the reason sms services are available in the smart phones and basic mobile phones also. spam messages in the sms is one of the trouble,spam messages will harm our device. A spammer is a human being or business which is responsible for uninvited messages, for their business benifits spammers sends a enormous amount of messages. Spam is uninvited and unwanted messages sent technologically and whose content may be dangerous.in sms length of the text messages are short unlike e-mail. The problem of spam messages are unwanted advertisement,subjection of personal information, becoming a sufferer of a deceit or commercial scheme, being attracted into adware and hackimg websites,etc. A common danger that compromises phone security by disseminating adware on phone devices is spam in SMS, WhatsApp, and other messaging services.A phone gadget that sends spam could also be the result of a security breach. Similar to e-mail spam, mobile SMS spam annoys consumers of mobile devices and damage to modern common gratings for portable electronics. Unlike SMS spam, which is typically disseminated through a phone network, scam emails can be sent or received globally. Current research indicates that mobile SMS spam filtering strategies are still in the early stages of categorization. It still takes a lot of time to reduce SMS spam using traditional filtration methods like the Bayesian classification filter, logistic regression, and decision tree algorithms. The various kinds of current techniques for filtering and reducing mobile SMS spam messages have been studied in detail.In our article, we provide a summary of the currently available techniques, as well as some speculation and potential directions for future research on techniques for detecting, filtering, and reducing mobile SMSspam. While it is possible to evaluate the spam messages using multiple data sets. Several publications have used numerous supervised and unsupervised algorithms, but we are only using one supervised algorithm in this paper.

II. Literature Survey

[1] We can conclude from the following experiments that naive Bayes outperforms random forest algorithm and SMS spam classification using the logistic regression algorithm. Using just the information gain matrix, the naive Bayes algorithm categorised the text as spam or not with an accuracy of 98.445%. The naive Bayes algorithm has the shortest running time of these algorithms, despite the fact that we haven't provided a full study of their running timings. With both features, the Random Forest method likewise performed brilliantly, and it may very well be a good replacement. The Naive Bayes algorithm beats Random Forest and Logistic Regression, as we have empirically demonstrated.

[2] In this study, our primary objectives were to explain and evaluate machine learning techniques for identifying spam SMS. Using 8 different classifiers, we conducted contrasts. Convolutional neural network classifier achieves the highest accuracy for the two datasets, 99.19% and 98.25 percent, and an AR value of 0.9926 and 0.9994, according to the results of our analysis of the classifiers. Despite being frequently used in the classification of data related to images, CNN outperforms traditional classifiers and also has the best accuracy among themwhen it comes to text-related data.

[3] The method most frequently used to categorise communications as spam or ham is machine learning. It is a potential alternative for classifying mobile spam messages because of its success in building an email spam classification system. Therefore it's one of the methods that may be used to lessen the amount of SMS spam that individual mobile phone customers have to deal with. So, the system of spam identification, classification, and blocking has been implemented using machine learning approaches.

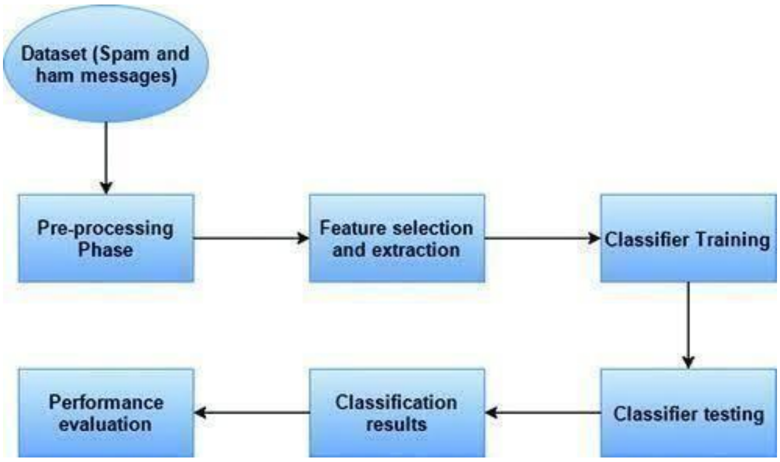
[4] Smart cities and communities are fostering unparalleled socioeconomic and technical advancement in day-to-day living, but they also make us more susceptible to an unlimited number of unfathomably varied threats. Infected mobile devices can spread malware, which is one way that Short Messaging Service (SMS) spam can compromise mobile securityAdditionally, a security flaw in a mobile device could allow it to transmit spam messages. In order to identify the best process configurations a system should employ in different layers of a distributed system, such as cloud, fog, and edge layers, we developed and studied the performance of various datasets, preprocessing methods, and feature extraction methods on spam and benign SMS classification.

[5] The outcomes of the testing set show that the model uses a pick from the dataset with a proportionate for each class. Using the 10-fold cross validation as a benchmark, the MLR and XGB algorithms continued to produce the best outcomes. The best accuracy findings came from MLR, XGB, and stochastic gradient descent. (SGD).The accuracy results from the test set and 10-fold cross validation demonstrate a sharp drop, which is caused by the dataset itself.

III. METHODOLOGY

1. System Design

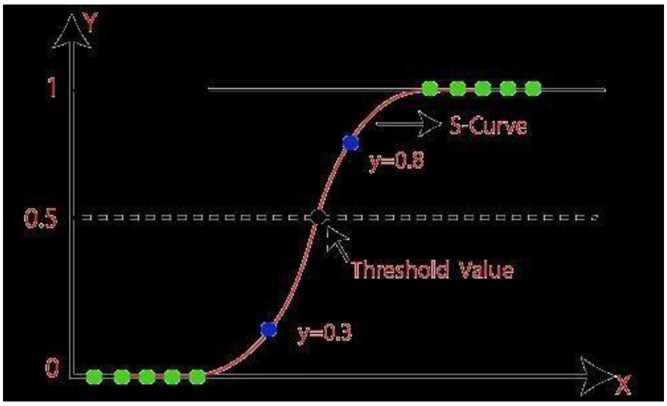
The dataset is collected from the Kaggle,this dataset contains labeled messages that have been collected from mobile phones,and labeled either spam or ham, the dataset contains 5574 English, real and non-encoded messages. We can use this dataset to predict the messages are spam or ham, and can be used to develop a tool that can automatically identify and block spam messages. dataset could be used to study the characteristics of spam messages and develop strategies for identifying and avoiding them. Dataset is splited into two parts one for testing and other part for training and 80% os the data isused for training and 20% for testing.



2. Algorithms

i. Logistic Regression

One of the most well-known machine learning algorithms, logistic regression is a supervised learning method in which the outcome is predicted as either True or False, 0 or 1, Yes or No, or 0 or 1. Because of this, the result must be categorical or derive value. Although logistic regression and linear regression are comparable, logistic regression is used to address classification issues while linear regression addresses regression issues. Instead of fitting a regression line, we fit a "S"-shaped logistic function in logistic regression, which forecasts values of 0 or 1.Dependent values are categorical in form in logistic regression



ii. Naïve Bayes

The Naive Bayes algorithm is a type of supervised learning algorithm that uses the Bayes theorem to solve categorization problems. The Naive Bayes classifier is a probabilistic classifier, which means it makes output predictions based on the likelihood that an item exists. Text categorization heavily utilizesthe Bayes theorem.

IV. CONCLUSION AND FUTURE WORK

In this paper we dicussed on machine acquiring knowledge of methods for sms spam identification, andthe dataset description collected from the kaggle and the various algorithms used evaluate the accuracy,we compare the Logistic Regression and Naive bayes algorithms and find out which algorithm will give betteraccuracy to predict the spam messages. This project will help to users to identify easily whether the message is spam.

REFERENCES

[1]. P. Sethi, V. Bhandari and B. Kohli, "SMS spam detection and comparison of various machine learning algorithms," 2017 International Conference on Computing and Communication Technologies for Smart Nation (IC3TSN), 2017, pp. 28-31, doi: 10.1109/IC3TSN.2017.8284445.

[2]. M. Gupta, A. Bakliwal, S. Agarwal and P. Mehndiratta, "A Comparative Study of Spam SMS Detection Using Machine Learning Classifiers," 2018 Eleventh International Conference on Contemporary Computing (IC3), 2018, pp. 1-7, doi: 10.1109/IC3.2018.8530469.

[3]. A. Alzahrani and D. B. Rawat, "Comparative Study of Machine Learning Algorithms for SMS Spam Detection," 2019 SoutheastCon, 2019, pp. 1-6, doi: 10.1109/SoutheastCon42311.2019.9020530.

[4]. S. Bosaeed, I. Katib and R. Mehmood, "A Fog- Augmented Machine Learning based SMS Spam Detection and Classification System," 2020 Fifth International Conference on Fog and Mobile EdgeComputing (FMEC), 2020, pp. 325-330, doi: 10.1109/FMEC49853.2020.9144833.

[5]. Theodorus, T. K. Prasetyo, R. Hartono and D. Suhartono, "Short Message Service (SMS) SpamFiltering using Machine Learning in Bahasa Indonesia," 2021 3rd East Indonesia Conference on Computer and Information Technology (EIConCIT), 2021, pp. 199-203, doi: 10.1109/EIConCIT50028.2021.9431859.