A

PROJECT REPORT

ON

**SMS SPAM DETECTION**

**USING**

**MACHINE LEARNING(ML)**

**SUBMITTED BY:**

**BBRCP192 G.VIJAY**

**BBRCP193 K. JAYANTH**

**BBRCP194 SD.KHAZA ALI**

**BBRCP195 P.KOMAL**

**UNDER THE GUIDANCE OF**

**Y.R. SAI KRISHNA SIR**

**CONTENTS**





**ABSTRACT**

In the modern world where digitization is everywhere, SMS has become one of the most vital forms of communications, unlike other chatting-based messaging systems like Facebook, WhatsApp etc, SMS does not require active internet connection at all. As we all know that Hackers / Spammer tries to intrude in Mobile Computing Device, and SMS support for mobile devices had become vulnerable, as attacker tries to intrude to the system by sending unwanted link, with which on clicking those link the attacker can gain remote access over the mobile computing device. So, to identify those messages Authors have developed a system which will identify such malicious messages and will identify whether or not the message is SPAM or HAM (malicious or not malicious). Authors have created a dictionary using the TF-IDF Vectorizer algorithm, which will include all the features of words a SPAM SMS possess, based on content of message and referring to this dictionary the system will be classifying the SMS as spam or ham.



The Internet has become an inseparable part of human lives, where more than four and half billion Internet users ﬁnd it a convenient to use it for their facilitation. Moreover, emails are considered as a reliable form of communication by the Internet users [[1].](#_bookmark22) Over the decades, e-mail services have been evolved into a powerful tool for the exchange of dif- ferent kind of information. The increased use of the e-mail also entails more spam attacks for the Internet users. Spam can be sent from anywhere on the planet from users having deceptive intentions that has access to the Internet. Spams are unsolicited and unwanted emails sent to recipients who do not want or need them. These spam emails have fake content with mostly links for phishing attacks and other threats , and these emails are sent in bulk to a largenumberof recipients [[2].](#_bookmark24) The intention behind them is to steal users’ personal information and then use them against their will to gain materialistic beneﬁts [[3].](#_bookmark25) These emails either contain malicious content or have URLs that lead to malicious content. Such emails are also sometimes referred to as phishing emails.

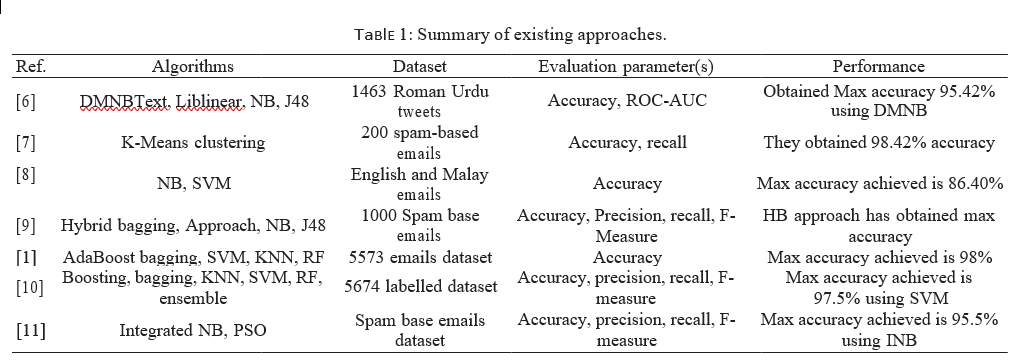
Despite the advancement of spam ﬁltering applications and services, there is no deﬁnitive way to distinguish be- tween legitimate and malicious emails because of the ever- changing content of such emails. Spams have been sent for over three or four decades now, and with the availability of various antispam services, even today, nonexpert end-users get trapped into such hideous pitfall [[4].](#_bookmark26) In e-mail managers, spam ﬁlters detect spam and forward it to a dedicated space, spam folder, allowing the user to choose whether or not to access them. Spam ﬁltering tools such as corporate e-mail systems, e-mail ﬁltering gateways, contracted antispam services, and end-user training can deal with spam emails in English or any other language [[4].](#_bookmark26) However, they are in- eﬀective at ﬁltering spam emails in other languages that recently have been digitized, such as Urdu Language. The proposed study exploits the existing artiﬁcial intelligence models to detect spam emails written in Urdu. This article describes how machine learning (ML) and deep learning (DL) models such as Support Vector Machine (SVM), Naive Bayes, Convolutional Neural Network (CNN), and Long Short-Term Memory (LSTM), a recurrent neural network, can be trained to detect Urdu spam emails. Moreover, as there is no dataset for spam emails, this article also explains its creation and training of various machine learning models. Precision, recall, and f-measure are considered key eval- uating measures to compare Naive Bayes and SVM, while the evaluation parameters, i.e., Model Loss and ROC-AUC, are calculated for deep learning models such as CNN and LSTM. Finally, a comparison is made between all models for the best accuracy and values of evaluation parameters obtained by DL and ML models [[5].](#_bookmark27) As we all know, a lot of work has been done in the ﬁeld of e-mail spam detection in English or any other foreign language.

**MOTIVATION**

A number of SMS Spam messages detection techniques are available these days like android app

spam messages, filtering spam messages using classification algorithms, etc. In this section, we will review the SMS Spam detection techniques by filtering spam messages based on feature selection using machine learning techniques. El-Alfy and AlHasan have proposed a model for filtering text messages for both email and SMS. They have analyzed different methods in order to finalize a feature set such that complexity can be reduced.They have used two classification algorithms i.e. Support Vector Machine (SVM) and Naïve Bayes and 11 features i.e. URLs, likely spam words, emotion symbols, special characters, gappy words, message metadata, JavaScript code, function words, recipient address, subject field and spam domain. They have evaluated their proposed model on five email and SMS datasets. Jialin et al. [16] have proposed a message topic model (MTM) for filtering Spam messages. Messages Topic Model (MTM) considers symbol terms, background terms and topic terms to represent spam messages and it is based on the probability guess of latent semantic analysis. They have used k-means algorithm to remove the sparse problem by training SMS spam messages into random irregular classes and then aggregating all SMS spam messages as a single file such that to capture word co-occurrence patterns. Chan et al. [17] have presented two methods for SMS Spam filtering i.e. feature reweighting method and good word attack. Both methods focus on the length of the message along with considering the weight of message. Good word attack focuses on deceiving the output of classifier by using least number of characters while for feature reweighting method they have introduced a new rescaling function for rescaling the weights. They have evaluated the experiment on two datasets i.e. SMS and comment. Delany et al. [18] discuss different approaches available for SMS Spam filtering and the problems associated with the dataset collection. They have analyzed a large dataset of SMS spam and used ten clusters i.e. ringtones, competitions, dating, prizes, services, finance, claims chat, voicemail and others. Xu et al. [19] have detected SMS Spam messages using content-less features. They have used 2 classification algorithms i.e. SVM and k-nearest neighbor (KNN) and feature set consisting of 3 features i.e. static, temporal and network for their experiment. They found that by combining temporal and network features SMS Spam messages can be detected more accurately and with good performance. Moreover, they also found the ways filter SMS Spam messages by using features that contain graph-topology and temporal information thus excluding the content of the message.





Boosting technique could be used for additional study in order to improve the system’s outcomes.

The authors in [[1]](#_bookmark22) used machine learning algorithms to detect spam emails. They compiled a dataset using online tools such as ‘kaggle’ and others. They have collected 5573 emails and used that data to train seven machine learning models. The greatest result is 98.5% accuracy with Multi- nomial Nave Bayes; however, it has obvious limitations as class-conditional dependency, which causes the system to misidentify some data items. On the other hand, ensemble approaches have been shown to be eﬀective since they use many learners to predict categories.

In [[10],](#_bookmark32) the authors have also made signiﬁcant contri- butions to the ﬁeld of spam e-mail detection. They used kaggle and the UCI machine learning repository to collect 5674 emails and deﬁne them as spam or ham. They estimated accuracy using six machine learning classiﬁers. They ex- plored a variety of ml algorithms; however, it was discovered that Ensemble Filter produces more remarkable outcomes and has accuracy of 98.5%, which is higher than the other learners, as well as faster testing. The article’s limitation is that testing was done on an e-mail sample without taking into account evolving trends in the mails, which could impair a classiﬁer’s eﬀectiveness.

For the ﬁltration of spam emails, an integrated Naive Bayes algorithm along with particle swarm optimization (PSO) is deﬁned in [[11].](#_bookmark33) They used NB to train and classify emails and PSO for swarm behavior property distribution. Finally, they used the proposed integrated concept NB and PSO to achieve evaluation steps. They employed a combined NB and PSO method. PSO is utilized to optimize the pa- rameters of the NB technique. Naive Bayes is employed as a separator among spam and ham emails based on the key- words. They achieved a maximum accuracy of 96.42% after using an integrated NB method. It would be better if the Naive Bayes approach was used in combination with ant colony or artiﬁcial bee colony optimization

.

