International Journal of Scientific Research in Science, Engineering and Technology
Print ISSN: 2395-1990 | Online ISSN : 2394-4099 (www.ijsrset.com)
doi : https://doi.org/10.32628/IJSRSET

Effective Spam Filtration and Fraud Identification Mechanism in Android Phones using Deep Learning and Artificial Intelligence

Ms. S. S. Wankhede[1], Pradnya Khobragade*[2], Shivani Bhoyar[3], Trupti Kawale[4], Sahil Raut[5]

[1]Assistant Professor, Department of Computer Science & Engineering, Priyadarshini Bhagwati College of Engineering, Nagpur, Maharashtra, India

[2-5] Department of Computer Science & Engineering, Priyadarshini Bhagwati College of Engineering, Nagpur, , Maharashtra, India

Article Info
Volume 9, Issue 2
Page Number : 132-142
Publication Issue :
March-April-2022
Article History
Accepted : 20 March 2022
Published: 30 March 2022

ABSTRACT

The widely used and mostly accessible communication medium to reach large volume of users in low cost is the "Short Message Service" i.e. SMS. These communication even though are useful for the advertisements in various sectors like banking, agriculture or even for the governmental schemes but sometimes they create a nuisances for those users which are not intended audience for that message. Some messages even may contain malicious links too. The efforts are proposed to restrict these spam messages using the hybrid mechanism deploying the deep learning and artificial intelligence.

Keywords: Short message services, deep learning, Artificial Intelligence, Spam

I. INTRODUCTION

In today's era, the mobile smartphones have occupied the human life in the society as the inevitable entity. The users utilize their smartphones to complete all the important tasks of their daily routine as well as official work. There is one important application in every mobile phone that is "Short message Service" which is called as SMS. The intention of this utility is to communicate with each other but now a day it has become an important means of carrying out the advertisements of various sectors like banking, some gaming mobile applications or even some fraudulent services or adult messages too. If the message is not intended for the user or if the message is not useful for the user where the user has never shown his interest, then it is spam. Even though the mobile smartphone technology has skyrocketed but still the spam filtration mechanism still pose a challenge. The spam can also be the unsolicited message which is sent to the user without their consent. The statistics shows that now a days, out of the total messages received by the user, 85% are the spam one. This shows the seriousness of the problem and the need to resolve the same effectively without losing the authentic messages. The popularity of spam in the malicious individuals is because the cost incurred to send the spam message is very low and the upshots of the same for the receiver can be disastrous. The authors proposed the Machine Learning approach to address the problem by building the spam classifier with Machine Learning Tensor Flow. Even the telecom companies are not able to control the spam

Copyright: © the author(s), publisher and licensee Technoscience Academy. This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial License, which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited

messages and research area is still open to control this problem. According to one shocking statistics, the practice of spamming often known as Grey Traffic causes revenue loss to the tune of $5 billion annually for telecommunication providers globally and in India alone there is a potential $795 million loss of capita income due to time spent in reading spam SMS every year. Telecom Regulatory Authority of India (TRAI) has introduced severe penalties for telemarketers who violate customer privacy and spam their inboxes without permission. But still the issue seems to be as it is because still almost everyone is getting the spam messages today also. These facts clearly underline the importance of the Tools for Spam Detection. Some AI tools can actually read the contents of the message to declare it spam by deeply looking into the meaning and the intention of the text present in it. Many algorithms have been proposed in Artificial Intelligence that can study the textual patterns and perform the classification of the SMS towards its intention for the receiver. If the use of traditional filters is deployed, they may mark the authentic SMS as spam and thus may hurt the vendor as well as audience. Hence the requisite for more intelligent mechanism for the pseudo SMS detection is gaining more importance that can balance the vendors and the companies sending SMS.

II. RELATED WORK

Naughton et al in [6] classifies the sentences in natural language such as Question Answering (QA) and Text Summarisation.

Huang et al in [7] proposed a complex method based on SMS network with a phone calling.

Narayan A et al in [8] proposed SMS spam classifiers on short text message and he developed a two level classifier for short message services.

HoushmandShirani-Mehral et al in [9] proposed an UCI machine repository which is used for real SMS spam database after pre-processing and feature extraction. Amir Karami et al in [0] proposed a new content based features to improve the performance of SMS spam detection.

Mukherjee A et al in [11] proposes an unsupervised approach for opinion spam detection which can exploit both linguistic and behavioural footprints left behind by spammers.

Fernandes, et al in [12] introduced the Optimum-Path Forest classifier to the context of spam filtering in SMS messages, as well as they compared it against with some state-of-the-art supervised pattern recognition techniques. Zainal et al in [13] study the discriminatory control of the features and considering its informative or influence factor in classifying SMS spam messages.

Etaiwi W et al in [14] used various features Word Count, n-gram feature sets and number of pronouns. In order to extract such features, many types of pre-processing steps could be performed applying the classification method, this steps may include POS tagging, n-gram term frequencies.

Howard et al in [15] proactivated and proposed Universal Language Model Fine-tuning (ULMFiT), an effective transfer learning method that can be applied to any task in NLP.

Widiastuti et al in [16] proactivated CNN Method for solving text mining domain and NLP. CNN that is proficient in image classification has proven its ability to process text.

Xia et al in [17] examines new method based on the discrete hidden Markov model (HMM) to use the word order information and to solve the low term frequency issue in SMS spam detection.

Ghourabi et al in [18] explained s detection model is based on the combination of two deep learning methods CNN and LSTM. It is intended to deal with mixed text messages that are written in Arabic or English.

III. AIM AND OBJECTIVES

The proposed work is motivated with the following objectives:

a. To report a review of various machine learning and hybrid algorithms for detecting SMS spam messages and comparing them according to accuracy criterion.
b. To deal with machine learning and hybrid approaches for SMS spam filtering.
c. Data extraction: Many articles extracted by searching a predefined string and the outcome was reviewed by one author and checked by the second.
d. Main objective of the Spam SMS filtering to reduce or blocks the unwanted message sent by the Spammer.
e. To study machine learning and hybrid methods for detecting SMS spam messages.
f. The objective of proposed approach is to employ two deep model together i.e. Support Vector Machine and Naive Bayes.

IV. RESEARCH METHODOLOGY

When a user receives an SMS or MMS message from an unknown sender, the Messages app can ask the proposed Message Filter app extension to determine whether the message is unsolicited or otherwise unwanted. The proposed application extension can make this determination by using its own built-in data and logic or by deferring to analysis done by the associated server. The following figure shows main work flow diagram.

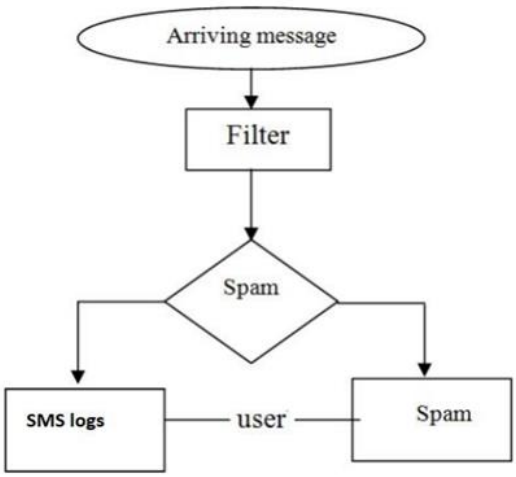

Fig.1 Main Work Flow Diagram

The data in the SMS i.e. text messages is pre-processed as follows:

Converted the messages to lowercase to capture the unique tokens and not differentiate between alphabet cases. The messages contained phone numbers, SMS addresses, http links, money symbols, and other numbers which were substituted by a fixed string. Eg. xyz@abc.com was replaced with "SMSaddr" using regex. Implemented regular expressions to remove punctuations and replace them with whitespace. To correct typos, a spell corrector was implemented which basically decides the best word based on edit distance. Spellchecker library was used for the same. Implemented stemming on words to bring them to their root form. The important words are the tokens other than the stop words. So, the stopwords are removed from the messages as they do not act as the differentiating factor. Finally, implemented Count Vectorizer and TfIdf Vectorizer for the data modelling exercise.

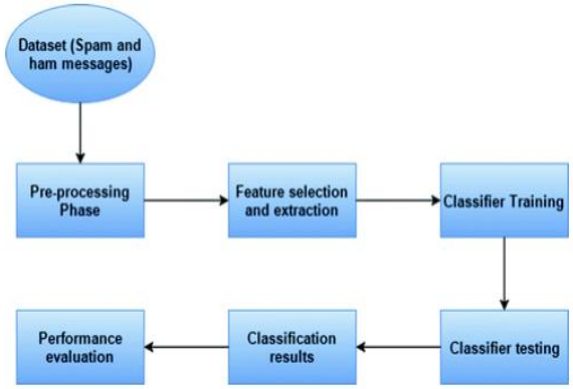

Fig.2 Spam and Ham Dataset Classification

Bayesian filtering and SVM are reported to be most successful techniques for SMS spam filtering [8]. Bayesian filtering has been reported as one of best techniques in SMS spam classification systems also. Apart from that, Bayesian spam filtering approach does minimal computation for classification unlike other classification algorithms like SVM, thus making it a preferred choice to use at mobile handheld devices. Like all other supervised machine learning techniques, Bayesian learning also needs a seed dataset to be

trained. In the training stage of Bayesian learning, it computes the occurrence of a word in spam as well as legitimate SMS to learn the probability of finding that word into spam / ham. For example, words like “sms”, “free”, “call” has higher probability due to their frequent occurrence in spam SMS.

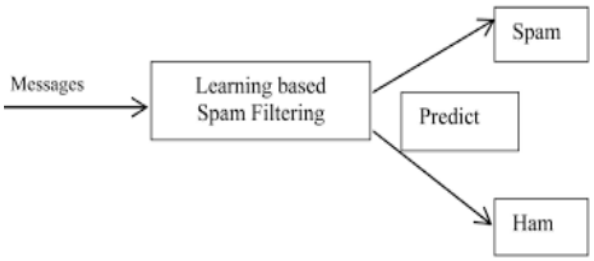

Fig.3 Dataflow diagram of Overall project

A data set (or dataset) is a collection of data. In the case of tabular data, a data set corresponds to one or more database tables, where every column of a table represents a particular variable, and each row corresponds to a given record of the data set in question.

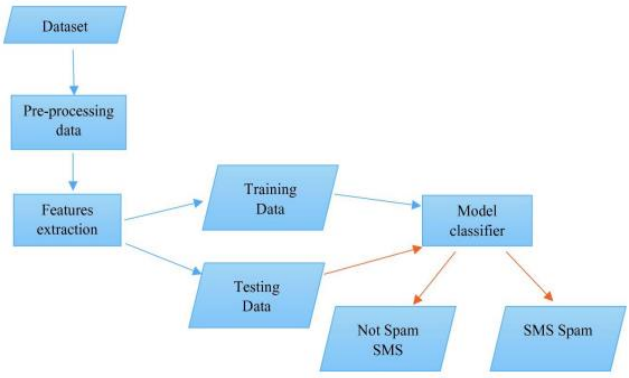

Fig.4 Flow diagram of Dataset collection

It is the graphical representation of information and data. By using visual elements like charts, graphs, and maps, data visualization tools provide an accessible way to see and understand trends and patterns in data.

a. In proposed mechanism, the Tf-idf vectorizer is used to convert the alphabetical messages into numeric form.

b. After vectoring the messages, fitting and feeding them as input for the model and feature label as output.

c. Here, two algorithms Naïve Bayes and SVM are used to compare the accuracy percentage.

d. In Naive Bayes, the accuracy percentage raised from 83% to max 88% even after minimizing the test size in the train test split. But in SVM the accuracy percentage went up to 98% showing better results than Naïve Bayes. SVM with parameters : (C=1.0 , kernel = ‘linear’, degree=3 , gamma=’auto’)

e. After applying all the suitable algorithms on the dataset, it is observed that SVM is performing better than Naive Bayes therefore it fits best for proposed objective. Hence, all the further process is done in SVM.

Following is the Flow chart of dataset and overall detection model.

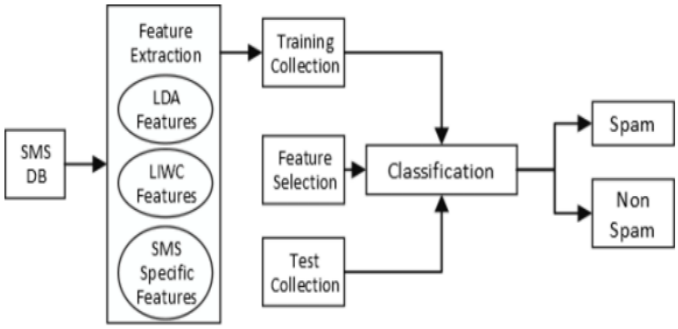

Fig.5 Flow chart of dataset and overall detection model.

The overall methodology can be depicted as follows:

Text classification is a machine learning technique that assigns a set of predefined categories to open-ended text. Text classifiers can be used to organize, structure, and categorize pretty much any kind of text i.e. from documents, medical studies and files, and all over the web. For example, new articles can be organized by topics, support tickets can be organized by urgency, chat conversations can be organized by language, brand mentions can be organized by sentiment, and so

on. Text classification is one of the fundamental tasks in natural language processing with broad applications such as sentiment analysis, topic labelling, spam detection, and intent detection.

A. Dataset Collection:

A data set (or dataset) is a collection of data. In the case of tabular data, a data set corresponds to one or more database tables, where every column of a table represents a particular variable, and each row corresponds to a given record of the data set in question. The data set lists values for each of the variables, such as height and weight of an object, for each member of the data set. Each value is known as a datum. Data sets can also consist of a collection of documents or files.

B. Big data and data collection

Big data describes voluminous amounts of structured, semi-structured and unstructured data collected by organizations. But because it takes a lot of time and money to load big data into a traditional relational database for analysis, new approaches for collecting and analysing data have emerged. To gather and then mine big data for information, raw data with extended metadata is aggregated in a data lake. From there, machine learning and artificial intelligence programs use complex algorithms to look for repeatable patterns.

C. Types of data

Generally, there are two types of data: quantitative data and qualitative data. Quantitative data is any data that is in numerical form e.g., statistics and percentages. Qualitative data is descriptive data e.g., colour, smell, appearance and quality. In addition to have quantitative and qualitative data, some organizations might also make use of secondary data to help drive business decisions. Secondary data is typically quantitative in nature and has already been collected by another party for a different purpose. For example, a company might use U.S. Census data to make decisions about marketing campaigns. In media, a news team might use government health statistics or health studies to drive content strategy. As technology evolves, so does data collection. Recent advancements in mobile technology and the Internet of Things are forcing organizations to think about how to collect, analyse and monetize new data. At the same time, privacy and security issues surrounding data collection heat up.

D. Machine Learning Text Classification Algorithm

Some of the most popular text classification algorithms include the Naive Bayes family of algorithms, support vector machines (SVM), and deep learning.

a. Naive Bayes:

The Naive Bayes family of statistical algorithms are some of the most used algorithms in text classification and text analysis, overall. One of the members of that family is Multinomial Naive Bayes (MNB) with a huge advantage, that one can get really good results even when your dataset isn't very large (a couple of thousand tagged samples) and computational resources are scarce. Naive Bayes is based on Bayes's Theorem, which helps to compute the conditional probabilities of the occurrence of two events, based on the probabilities of the occurrence of each individual event. So the probability of each tag for a given text is calculated, and then outputting the tag with the highest probability. The probability of A, if B is true, is equal to the probability of B, if A is true, times the probability of A being true, divided by the probability of B being true. This means that any vector that represents a text will have to contain information about the probabilities of the appearance of certain words within the texts of a given category, so that the algorithm can compute the likelihood of that text belonging to the category.

b. Support Vector Machine:

Support Vector Machines (SVM) is another powerful text classification machine learning algorithm, because like Naive Bayes, SVM doesn't need much training data to start providing accurate results. SVM does, however, require more computational resources than Naive Bayes, but the results are even faster and more accurate. In short, SVM draws a line or "hyper plane" that divides a space into two subspaces. One subspace contains vectors (tags) that belong to a group, and another subspace contains vectors that do not belong to that group.

V. IMPLEMENTATION

Text feature extraction and pre-processing for classification algorithms are very significant. In this section, discussion is made about text cleaning since most of the documents contain a lot of noise. In this part, two primary methods of text feature extractions-word embedding and weighted word are made. In android, after loading the tflite model as soon as one clicks on predict button the model predicts whether the message is spam or ham. In Natural Language Processing (NLP), most of the text and documents contain many words that are redundant for text classification, such as stopwords, mis-spellings, slangs, and etc. In this section, a brief explanation of some techniques and methods for text cleaning and pre-processing text documents is given. In many algorithms like statistical and probabilistic learning methods, noise and unnecessary features can negatively affect the overall performance. So, the elimination of these features is extremely important. The TensorFlow Lite Model Maker library simplifies the process of adapting and converting a TensorFlow model to particular input data when deploying this model for on-device ML applications.

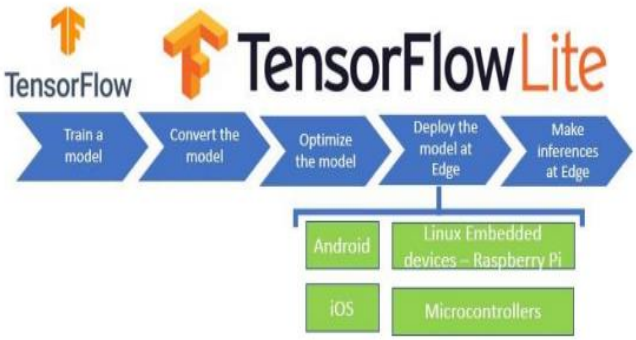

Figure 6. TFLITE model

The below mentioned notebook shows an end-to-end example that utilizes the Model Maker library to illustrate the adaptation and conversion of a commonly-used text classification model to classify movie reviews on a mobile device. The text classification model classifies text into predefined categories. The inputs should be pre-processed text and the outputs are the probabilities of the categories. The dataset used in this tutorial are positive and negative movie reviews. To create one's own dataset, create a CSV file containing two columns – sentence and label respectively. Add the texts and their respective labels in both these columns. The first column is the index column which is later created by the Pandas library automatically. Simply ignore it while creating the dataset. Note: If the dataset contains many labels, make sure that collection is made for enough data for each label so that the dataset doesn't become biased.

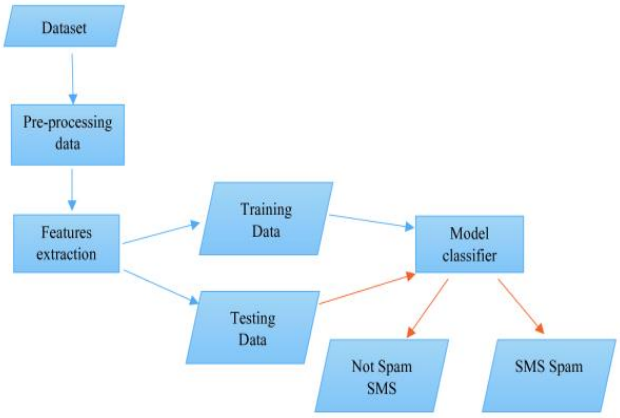

Figure 7. Training model

The proposed filter detects and filters out SMS spam messages in a smart manner rather than black/white

list approaches that require intervention of phone users. In the study, index based approach is preferred as the feature selection method. The feature vectors consisting of the selected discriminative features are then fed into two well- known pattern classifiers, namely Naive Bayes and k-Nearest Neighbour, for recognition process. Furthermore, a mobile application, which exploits the proposed detection scheme, is developed particularly for the mobile phones with Android™ operating system. Thus, SMS spam messages are automatically filtered out without disturbing the phone user. The proposed detection scheme is evaluated on a large SMS message dataset consisting of spam and legitimate messages. The results of the experimental work reveal that the proposed system is considerably successful in filtering SMS spam messages. Text classification, also known as text categorization, is a classical problem in natural language processing (NLP), which aims to assign labels or tags to textual units such as sentences, queries, paragraphs, and documents. It has a wide range of applications including question answering, spam detection, sentiment analysis, news categorization, user intent classification, content moderation, and so on. Text data can come from different sources, including web data, SMSs, chats, social media, tickets, insurance claims, user reviews, and questions and answers from customer services, to name a few. Text is an extremely rich source of information. But extracting insights from text can be challenging and time-consuming, due to its unstructured nature. Text classification can be performed either through manual annotation or by automatic labelling. With the growing scale of text data in industrial applications, automatic text classification is becoming increasingly important. Approaches to automatic text classification can be grouped into two categories:

- Rule-based methods
- Machine learning (data-driven) based method

Rule-based methods classify text into different categories using a set of pre-defined rules, and require deep domain knowledge. On the other hand, machine learning based approaches learn to classify text based on observations of data. Most classical machine learning based models follow the two-step procedure. In the first step, some hand-crafted features are extracted from the documents (or any other textual unit). In the second step, those features are fed to a classifier to make a prediction. Popular hand-crafted features include bag of words (BoW) and their extensions. Popular choices of classification algorithms include Naïve Bayes, support vector machines (SVM), hidden Markov model (HMM), gradient boosting trees, and random forests. The two-step approach has several limitations. For example, reliance on the handcrafted features requires tedious feature engineering and analysis to obtain good performance. In addition, the strong dependence on domain knowledge for designing features makes the method difficult to generalize to new tasks. Finally, these models cannot take full advantage of large amounts of training data because the features (or feature templates) are pre-defined.

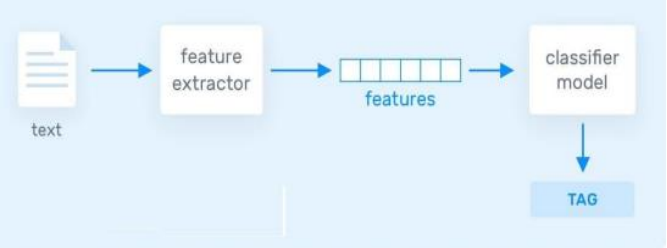

Figure 8. Prediction mode

- Software Requirement : Google Chrome, Firebase Account, Android Studio, Windows / Ubuntu Operating System, Android Phone (SDK 30 is recommended), Java, Kotlin, Python, Xml
- Hardware Requirement : 64bit based processor, 128 GB HDD (SSD is recommended), 8 GB Ram, Intel Processor (i3 5th Generation processor recommended)

VI. EXPERIMENTAL RESULTS

Recent developments in electronic communication, which is one of the crucial communication tools owing to wide spreading internet technologies, heightened the requirements for a solution of the problem of circulation of unsolicited bulk messages on the internet, which is referred to as Spam. In this approach, ELM (extreme learning machine), which is a training method for single hidden layer feed-forward artificial neural network, was employed as a filter to spam messages. The experimental results showed that, the proposed method achieved higher performance in terms of detection speed and classification accuracy, 91.655%, than other machine learning methods such as SVM (support vector machines), NB (naive Bayes), MLP (multi layer perceptron). The results of the experiments demonstrated that the classification performance of LR is high as compared with K-NN and DT, and the LR achieved a high accuracy of 99%. Additionally, the proposed method performance is good as compared with the existing state-of-the-art methods.

$$P(A|B) = \frac{P(B|A) \times P(A)}{P(B)}$$

Diverse approaches are used for spam detection. Abayomi-Alli et al. presented a comprehensive review of the soft techniques in spam classifications. Acceptability of users of SMS spam application on the store of Android App was assessed. Here a technique to identify short-text spam messages is proposed. The proposed model is helpful for different strategies of business and presented a novel method for evaluating the crowd security of OSN trustworthiness. Mao et al. made security network of dependency from the access behaviour to measure the significance of object security from with broad perspectives. We performed experiments to classify the ham and spam using the SMS spam collection dataset. Classifiers LR, decision tree, and k-nearest neighbour were used for the classification in this approach. The dataset is divided as follows: 30% for validation and 70% for training. The results obtained from experiments are shown in Figure No.9 graphically. The python on an Intel (R) Core™ i5 -2400 CPU and Windows 10 were used for the experiments and setup to obtain the computation results of the experimental work.

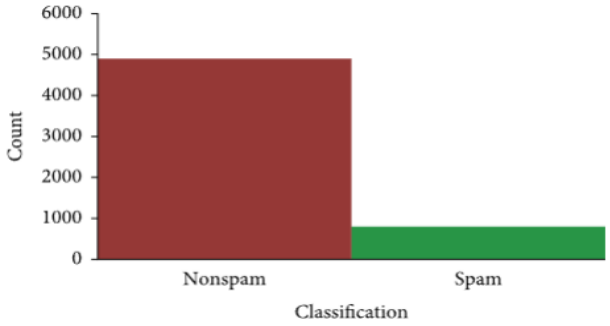

Figure No. 9 Results

Text classification is the task of assigning a sentence or documents an appropriate category. The categories depend on the chosen dataset and can range from topics. Text classification problems include emotion classification, news classification, and citation intent classification, among others. Benchmark datasets for evaluating text classification capabilities include GLUE, AGNews, among others. In recent years, deep learning techniques like XLNet and RoBERTa have attained some of the biggest performance jumps for text classification problems.

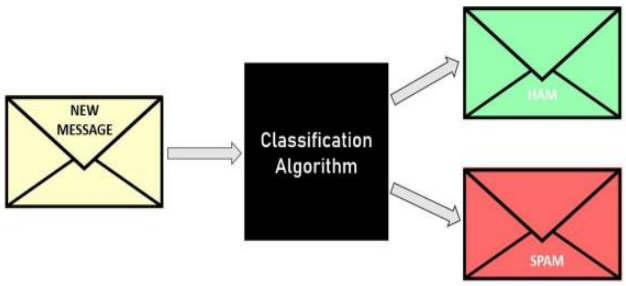

Figure No. 10 Classifier

In order to build a model for future classification of SPAM / HAM messages, first there is a need to prepare the data in the format as desired by proposed algorithm. We cannot feed text data to our algorithm as it won't understand it.

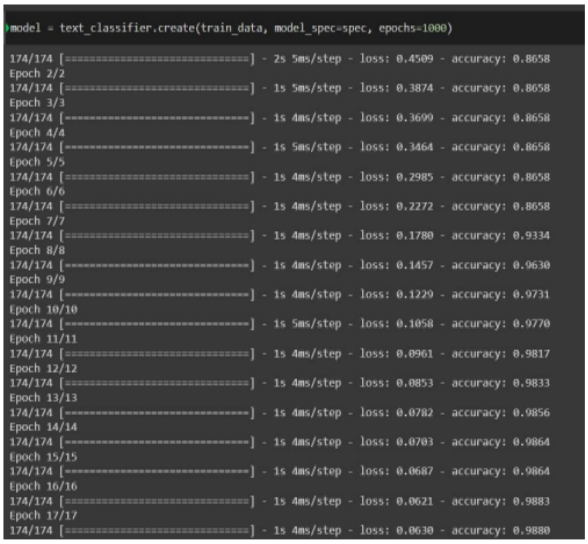

Figure No. 11 Training of model in python

The performance of the proposed models is compared to the models based on machine learning algorithms including Support Vector Machine and Naïve Bayes. The experimental results show that the model built from the Long Short-Term Memory technique provides the best overall accuracy as high as 98.18%. On accurately screening spam messages, this model shows the ability that it can detect spam messages with the 90.96% accuracy rate, while the error percentage that it misclassifies a normal message as a spam message is only 0.74%.

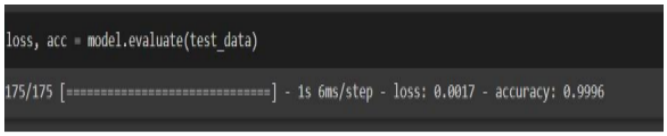

Figure No. 12 Evaluation of accuracy

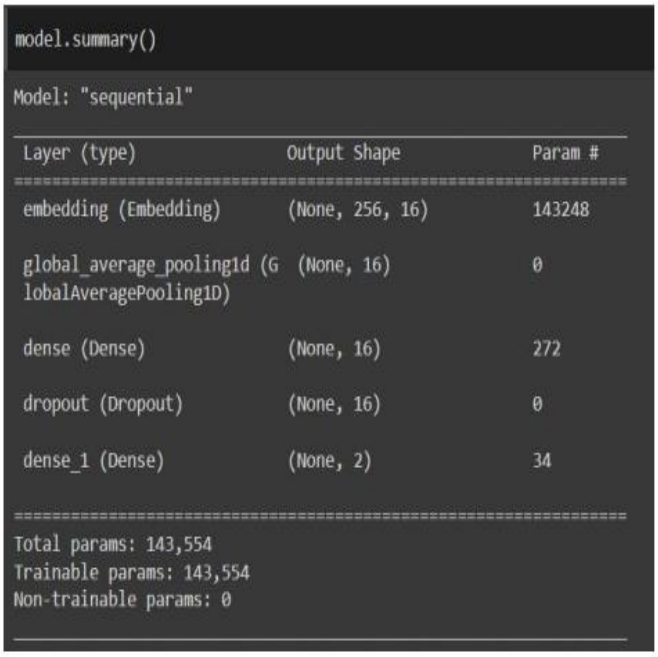

Figure No. 13 Summary of train mode

The describe () method from pandas provide a summary statistics. Such as, there are 5,572 labels and messages. There are two unique labels indicating for "ham" and "spam". It has also been observed that there are less unique messages (5,169) than total message count (5,572) indicating some repeated messages. The top label is "ham" and the top message in the data is "Sorry, I'll call later".

	label	message
count	5572	5572
unique	2	5169
top	ham	Sorry, I'll call later
freq	4825	30

Figure No. 14 Summary Statistics

Model Evaluation

After fitting both NB and SVC on the train data, now there is a need to evaluate the models on the unseen data (test data) in order to understand the accuracy and generalization power of the models. However, blindly using accuracy would not be good here as dataset used here is imbalanced. It will always be a bit biased towards the majority class (Ham). Hence, one can look at precision, recall or f1 score to gain better insights

about the models' performances. And all these metrics can be calculated using the famous "Confusion Matrix".

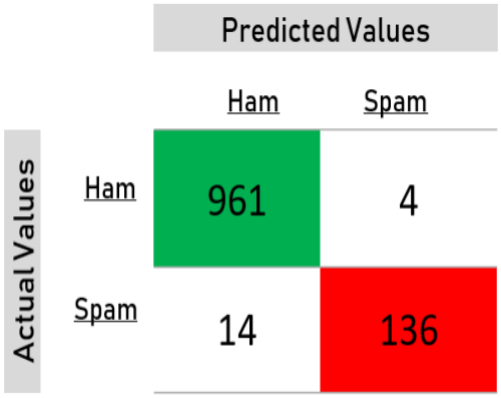

Figure No. 15 Confusion Matrix — Naive Bayes

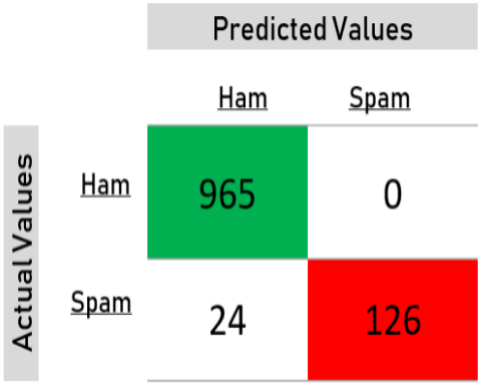

Figure No. 16 Confusion Matrix — Support Vector Classifier

Accuracy for both NB and SVC are approximately 98%. However, we are interested in either precision, recall or f1 score. Now, f1 is same for both the models but "Recall" (predicting spam as spam) is higher for Naive Bayes (91% to 84%). Hence, if our objective is to correctly predict spam as spam even if few ham messages are incorrectly classified as spam, we will go for Naive Bayes. So, overall, we can conclude that Naive Bayes is the way forward for this classification problem. But again, it depends on our objective. If reducing FP and FN are both important, we will choose a model which has the highest f1 score on test.

VII. CONCLUSION AND FUTURE SCOPE

Detection of spam is important for securing message and e-mail communication. The accurate detection of spam is a big issue, and many detection methods have been proposed by various researchers. However, these methods have a lack of capability to detect the spam accurately and efficiently. To solve this issue, a method is proposed for spam detection using machine learning predictive models. The method is applied for the purpose of detection of spam. The experimental results obtained show that the proposed method has a high capability to detect spam. The proposed method achieved 99% accuracy which is high as compared with the other existing methods. Thus, the results suggest that the proposed method is more reliable for accurate and on-time detection of spam, and it will secure the communication systems of messages and emails.

In our future work, we will try to add more features as best spam features that will help in detecting spam messages more accurately. We will also try to collect more and more datasets from the real world. Future scope will be but not limited to:

- Implementation of feedback
- Machine learning using feedback received from user
- Update the existing spam dataset
- Optimizing the dataset so that accuracy will be increased.

VIII. REFERENCES

[1]. M. M. A. a. A. Q. M. Ghourabi, "A Hybrid CNN- LSTM Model for SMS Spam Detection in Arabic and English Messages", 2020.

[2]. M. R. a. C. M. U. a. o, "Spam filtering using ML algorithms", 2005.

[3]. S. a. M. D. Youn, "A comparative study for SMS classification", 2007.

[4]. E. a. B. A. Blanzieri, "A survey of learning-based techniques of SMS spam filtering", 2008.

[5]. J. a. H. T. a. T. P. Gbel, "Towards proactive spam filtering", 2009.

[6]. W. a. W. T. Liu, “Index-based online text classification for sms spam filtering”, 2010.
[7]. M. a. S. N. a. C. J. Naughton, “Sentence-level event classification in unstructured texts”, 2010.
[8]. Xu, “Sms spam detection using non content features”, 2012.
[9]. S. P. Narayan, “The curse of 140 characters: evaluating the efficacy of SMS spam detection on android”, 2013.
[10]. H. Shirani-Mehr, “SMS spam detection using machine learning approach”, 2013.
[11]. Z. L. Karami, “Improving static SMS spam detection by using new content-based features”, 2014.
[12]. V. V. Mukherjee, “Opinion spam detection: An unsupervised approach using generative models”, 2014.
[13]. K. a. J. M. Z. Zainal, “A review of feature extraction optimization in SMS spam messages classification”, 2016.
[14]. D. a. d. C. K. A. a. A. T. A. a. P. J. P. Fernandes, “SMS spam filtering through optimum-path forest based classifiers”, 2015.
[15]. N. a. B. T. a. M. P. a. M. A. S. Al Moubayed, “Sms spam filtering using probabilistic topic modelling and stacked denoising autoencoder”, 2016.
[16]. W. a. N. G. Etaiwi, The impact of applying different preprocessing steps on review spam detection, 2017.
[17]. J. a. R. S. Howard, “Universal language model fine-tuning for text classification”, 2018.
[18]. N. Widiastuti, “Convolution Neural Network for Text Mining and Natural Language Processing”, 2019.
[19]. T. a. C. X. Xia, “A discrete hidden Markov model for SMS spam detection”, 2020.
[20]. S. K. Tuteja, “Classification Algorithms for SMS Spam Filtering", 2016.
[21]. G. Mujtaba, L. Shuib, R. G. Raj, N. Majeed, and M.
[22]. Al-Garadi, "SMS Classification Research Trends: Review and Open Issues", 2017.
[23]. S. Ajaz, M. T. Nafis, and V. Sharma, "Spam Mail Detection Using Hybrid Secure Hash Based Naive Classifier, 2017
[24]. Shafi'i Muhammad Abdul Hamid, M. S., Osho, O., Ismaila, I., & Alhassan, J. K. “Comparative Analysis of Classification Algorithms for SMS Spam Detection”, 2018.
[25]. Rusland, N. F., Wahid, N., Kasim, S., & Hafit, H.,“Analysis of Naive Bayes Algorithm for SMS Spam Filtering across Multiple Datasets”, 2017.
[26]. Yuksel, A. S., Cankaya, S. F., & Üncü, İ. S. “Design of a Machine Learning Based Predictive Analytics System for Spam Problem”, 2017Verma, T,”E-Mail Spam Detection and Classification Using SVM and Feature Extraction”, 2017.
[27]. Singh, V. K., & Bhardwaj, S,”Spam Mail Detection Using Classification Techniques and Global Training Set”, 2018.
[28]. Priti Sharma,Uma Bhardwaj,”Machine learning based Spam SMSdetection”,2017
[29]. Manmohan Singh, Rajendra Pamula, Shudhanshu Kumar shekhar”.SMS Spam Classification by Support VectorMachine”,2018.
[30]. Linda Huang, Julia Jia,Emma Ingram, Wuxu Peng “Enhancing the Naive Bayes Spam Filtering through Intelligent Text ModificationDetection”,2018
[31]. Prachi Gupta, Ratnesh Kumar Dubey,Dr. Sadhna Mishra,”Detecting Spam SMSs/Sms Using Naive Bayes And Support VectorMachine”,2019.
[32]. Sah, U. K., & Parmar, N,”An approach for Malicious Spam Detection in SMS with comparison of different classifiers”,2017.
[33]. D. Ruano-Ordas, F. Fdez-Riverola, J.R Mendez,”Using evolutionary computation for discovering spam patterns from e-mail samples”, 2018.
[34]. A. SAski,N.KSourati,”Efficient algorithm to filter spam using machine learning techniques”,2016

Cite this article as :

Ms. S. S. Wankhede, Pradnya Khobragade, Shivani Bhoyar, Trupti Kawale, Sahil Raut, "Effective Spam Filtration and Fraud Identification Mechanism in Android Phones using Deep Learning and Artificial Intelligence", International Journal of Scientific Research in Science, Engineering and Technology (IJSRSET), Online ISSN : 2394-4099, Print ISSN : 2395-1990, Volume 9 Issue 2, pp. 132-142, March-April 2022.
Journal URL : https://ijsrset.com/IJSRSET229226