WWW.IJRPR.COM

International Journal of Research Publication and Reviews

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

SMS Spam Detection Using Deep Learning

M.Lavanya[#1], K.R.Aruna[#2]

[#1] M.Sc, Department of Computer Science, Kamban College of Arts and Science for Women, Tiruvannamalai-606603

[#2] Head of the department, Department of Computer Science, Kamban College of Arts and Science for Women, Tiruvannamalai- 606603.

ABSTRACT

Overrecentyears,asthepopularityofmobilephonedeviceshasincreased,ShortMessage Service (SMS) has grown into a multi-billion dollars industry. At the same time,reductioninthecostofmessagingserviceshasresultedingrowthinunsolicitedcommercial advertisements (spams) being sent to mobile phones. *SMS spamming is anactivity of sending 'unwanted messages' through text messaging or other communicationservices; normally using mobile phones. The SMS spam problem can be approached withlegal, economic or technical measures.* Nowadays there are many methods for SMS spamdetection, ranging from the list-based, statistical algorithm, IP-based and using machinelearning. However, an optimum method for SMS spam detection is difficult to find due toissues of SMS length, battery and memory performances. *A database of real SMS Spamsfrom UCI Machine Learning repository is used, and after preprocessing and featureextraction, different machine learning techniques are applied to the database*. Among thewide range of technical measures, Bayesian filters are playing a key role in stopping smsspam. Here, we analyze to what extent Bayesian filtering techniques can be applied to theproblem of detecting and stopping mobile spam. In particular, we have built SMS spamtest collections of significant size in English.We have tested on them a number ofmessagesrepresentationtechniquesandMachineLearningalgorithms,intermsofeffectiveness. The effectiveness of the proposed features is empirically validated usingmultiple classification methods. The results demonstrate that the proposed features canimprovetheperformanceofSMSspamdetection.

KEYWORDS: Spam Sms, Machine Learning ,CNN

1.INTRODUCTION:

In our online world, technology use for the many utilities of our life styles. Mainly used for communications, business, data storage, and data security etc. In communication domain, our current gadgets or electronic devices use SMS (Short Message Service), Emails services, and online chat apps for communicating the information of personal data, professional data, media data etc. In these very commonly we use SMS services for personal and professional information sharing. Almost all multination companies like Insurance, Banking, E-commerce companies are spreading their services, offers, promotions etc. to the users through the SMS″s. Traditionally SMS are the short message which is delivered between mobile devices using via mobile operating network. This 224 character messages will transferred by the user by creating message by typing group of characters in the mobile devices. This type of the SMS″s we call it as traditional SMS. We have another type of the SMS service, which is Auto SMS service. In this type of service, the instead of human being a program will send the SMS″s based on the type of programmed, that may be time series or completion of the task or alerts etc. For these types of services we have so many third party services or API″s to send the bulk of SMS″s to the users by clicking on a single button. It is very easy for companies send any information or service or alert to their customers/users with these SMS API″s. Other hand we are facing problem of malicious incoming SMS from the different types of malicious attackers. Due to these advanced API″s attackers can also send the bulk of SMS″s which consists of malicious behavior, when we respond to the SMS″s positively or negatively the targeted attacks will occurs. These attacks may effect to the users in various types like power consumption, leakage of the data, and ransomware etc. In android mobiles, most of the security attacks occur by sharing the vulnerable messages or click on the vulnerable messages. Once user clicks on the malware SMS, the device will compromise the security with few attacks like mobile botnets, spyware, and Mobile botnets are kind of security attacks in to the mobile operating systems which are unpatched. This attack targets the smart phones and gets complete access to the mobile data like contacts and photos etc. This attack also self-driven to forward malware message from compromised mobile to its contacts through messages and emails. Spyware attacks are designed for stealing the user information of user internet movements by collection of cookies and It can steal very sensitive information like user″s baking information, security keys, and user browser history. It may also cause to pop up the ads in the devices. Trojan attacks may affect the mobiles in different types like inserting the malicious code to operating systems to lock the phones, to send the messages which may cause to heavy pay charges etc. Which may cause to ransomware, attackers will pressure on the victims to pay the amount to unlock the mobile devices or unlock the data. In this paper, we discuss the main

attacks of the mobiles spreading through the Short Message Service (SMS). Machine Learning is a concept of learn the things to make decisions, for predictions, for clustering based on the given data. And it will improve itself to make better results in various aspects. The learning means it is a process of learn the model, learn the behavior, learn the grouping objects based on the given data. We have two types of the machine learning concepts are there, one is Supervised Machine learning and another one is Unsupervised Machine learning methodology. In supervised machine learning concept, the learning process is depends on the existed data, classified with labels. When we have the existing data with decision made, for example, an email from the userid, IP Address, port no we have these details and we have the labeled answer is it˝s a SPAM or it˝s a NOT SPAM. Based on the features data and based on the labels supervised machine learning algorithms will predict the answer based on the features data. In unsupervised machine learning concept, the learning process is not depends on the existed data, it can˝t train readymade. It will prepare the structured data from unstructured corpus of the data. It will process which are unlabeled, hidden structure. It will form the group data, or cluster data from the unstructured data. In this paper we are detecting the spam messages by using supervised machine learning algorithms based on content. Here we are predicting the spam filtering using labeled dataset by applying the supervised algorithms called k-Nearest Neighbor (KNN), Support Vector Machine (SVM) and Naïve Bayes (NB). Our goal paper is detection spam messages, but our main goal of the paper is detection of the high accuracy performance of the supervised machine learning algorithm. Here we train the dataset with three methods of the machine learning algorithms and we test the results with labeled dataset, then we calculate the accuracy of these three algorithms and we show the best prediction algorithm for detection of the spam messages and ham messages.

2. LITERATURE SURVEY

[1]. **Krishnaveni et al.** proposed a methodology for identifying the Spam Reviews using the Natural Language Processing for Preprocessing techniques and Neural Networks Classifier. The Features of Dataset is considered for classification with Multiple Features based on NLP and the Reviewers characteristics. The Polarity of the Text is also considered as a Feature

[2]. **Dipak et al.** have used the dataset of spam SMS to predict whether the messages is spam or ham. Natural language processing (NLP) steps like are done in the case of content based text message to detect whether the messages send is spam or ham. Error messages are predicted using different Statistical Techniques. The algorithms like Support Vector Machine, Neural Network and Relevance vector Machine are used and analyzed the best accuracy rate

[3]. **Jabbar et al.** has predicted the spam e-mails which can contain the phishing or malware that can harm the system or it can steal confidential information, so it is important to analysis the fake emails. Negative Selection Algorithm (NSA) is used for the anomaly detection for spam filtering techniques. E-mails are scanned to analyses the text content and the process of tokenization and stop word removal process are implemented for the analysis of the e-mail. Based on the content and based on the true positive and true negative values spam emails are detected.

[4]. **Alkahtani et al.** used the spam SMS dataset to analyses the messages based on the text content and by using filtering techniques. Based on techniques like backlist, white list, challenge response system and origin diversity analysis the detection of spam SMS are found. Filters like Heuristic filter, Rule based filter and Genetic algorithm, Artificial Neural Networks, Decision tree techniques and Clustering Techniques are used for the prediction of the spam SMS messages.

[5]. **SaritChakraborty et al.** has analyzed the e-mails and predicted out of which it is spam or ham e-mails. The detection is based on the machine learning techniques and used Cumulative Weighted Sum (CWS) for the higher level of accuracy rate. Emails are classified into content based and image based mails. Basis of weight fixation techniques like Frequency based weight fixation, Matrix based weight fixation, Tree based weight fixation are used for Cumulative Weighted Sum techniques for analysis of spam e-mails.

3.PROPOSED SYSTEM:

In the proposed system, we mainly focus on characterize users and their rating behaviors and the helpfulness scores received from others and the correlation of their reviews with product popularity.
In this system we implement the system of spam detection methods is presented which uses machine learning algorithm. It was shown that using different datasets yields extremely different results.

Architecture diagram:

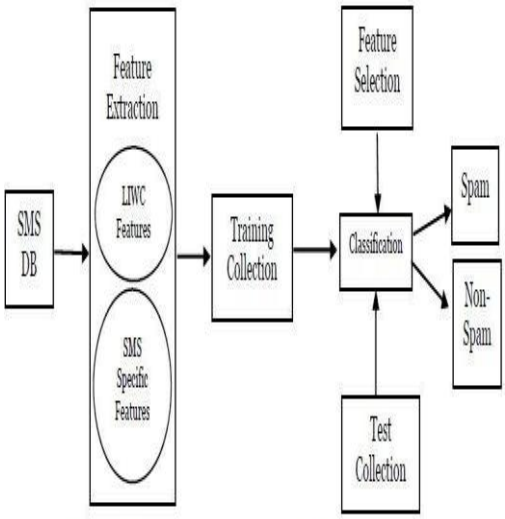

Figure 2: A Framework of Content based SMS Spam Detection

4. METHODOLOGY:

DATASET:

The dataset is sourced from ULB Machine Learning Group and description is found. The dataset contains credit card transactions made by European cardholders. This dataset presents transactions that occurred in two days, consisting of 284,807 transactions. The positive class (fraud spamcases) make up 0.172% of the transactions data. The dataset is highly unbalanced and skewed towards the positive class. It contains only numerical (continuous) input variables which are as a result of a Principal Component Analysis (PCA) feature selection transformation resulting to 28 principal components. Thus a total of 30 input features are utilized in this study. The details and background information of the features cannot be presented due to confidentiality issues. The time feature contains the seconds elapsed between each transaction and the first transaction in the dataset. The 'amount' feature is the transaction amount. Feature 'class' is the target class for the binary classification and it takes value 1 for positive case (fraud) and 0 for negative case (non fraud).

DATASET GATHERING

UCI Machine Learning Repository has a collection of sms messages – SMS Spam Dataset. It contains:

A collection of 425 spam messages from Grumbletext web site.

A subset of 3375 ham messages from the NUS Sms Corpus

A list of 450 ham messages from Caroline Tag's Phd Thesis.

A total of 4827 ham and 747 spam = 5574 messages The dataset consists of one message per line. Each line is prefixed with a ham/spam label separated by a tabspace.

Evaluation

There are a variety of measures for various algorithms and these measures have been developed to evaluate very different things. So it should be criteria for evaluation of various proposedmethod. False Positive (FP), False Negative (FN), True Positive (TP), and True Negative (TN) and the 19 relation between them are quantities which usually adopted by credit card fraud spamdetection researcher to compare the accuracy of different approaches. The definitions of mentioned parameters are presented below:

FP: the false positive rate indicates the portion of the non-fraudulent transactions wrongly being classified as fraudulent transactions.

FN: the false negative rate indicates the portion of the fraudulent transactions wrongly being classified as normal transactions.

TP: the true positive rate represents the portion of the fraudulent transactions correctly being classified as fraudulent transactions.
TN: the true negative rate represents the portion of the normal transactions correctly being classified as normal transactions.

Table 4shows the details of the most common formulas which are used by researchers for evaluation of their proposed methods. As can be seen in this table some researchers had been used multiple formulas in order to evaluated their proposed model.

CREDIT CARD FRAUD:

Credit card frauds have been partitioned into two types: inner card fraud spamand external fraud, while a broader classification have been done in three categories, that is, traditional card related frauds (application, stolen, account takeover, fake and counterfeit), merchant related

frauds (merchant collusion and triangulation) and Internet frauds (site cloning, credit card generators and false merchant sites) Credit card transactions data are mainly characterized by an unusual phenomenon. Both legitimate transactions and fraudulent ones tend to share the same profile. Fraudsters learn new ways to mimic the spending behaviour of legitimate card (or cardholder). Thus, the profiles of normal and fraudulent behaviours are constantly dynamic. This inherent characteristic leads to a decrease in the number of truefraudulent cases identified in a pool of credit card transactions data leading to a highly skewed distribution towards the negative class (legitimate transactions). The credit card data investigated in contains 20% of the positive cases, 0.025% positive cases and below 0.005% positive cases . The data used in this study has positive class (frauds) accounting for 0.172% of all transactions.

Fraud spam/ non-fraud spamtraining data generate classifiers with the highest true positive rate and low false positive rate. The paper uses stratified sampling to under sample the legitimate records to a meaningful number. It experiment on 50:50, 10:90 and 1:99 distributions of fraud spamto legitimate cases reports that 10:90 distribution has the best performance (regarding the performance comparisons on the 1:99 set) as it is closest to the real distribution of frauds and legitimates. Stratified sampling is also applied in . In this study, a hybrid of under-sampling the negative cases and oversampling the positive cases is carried in order to preserve valuable patterns from the data.

FEATURE SELECTION:

The basis of credit card fraud spamdetection lies in the analysis of cardholder's spending behaviour. This spending profile is analysed using optimal selection of variables that capture the unique behaviour of a credit card. The profile of both a legitimate and fraudulent transaction tends to be constantly changing. Thus, optimal selection of variables that greatly differentiates both profiles is needed to achieve efficient classification of credit card transaction. The variables that form the card usage profile and techniques used affect the performance of credit card fraud spamdetection systems. These variables are derived from a combination of transaction and
past transaction history of a credit card.

These variables fall under five main variable types, namely all transactions statistics, regional statistics, merchant type statistics, time-based amount statistics and time-based number of transactions statistics

While credit card fraud spamdetection has gained wide-scale attention in the literature, there are yet some issues (a number of significant open issues) that face researchers and have not been addressed beforeadequately. Wehope this overview focuses the direction of future research to provide more efficient and trustable fraud spamdetectionsystems.These issues are as follow:

Nonexistence of standard and comprehensive credit card benchmark or dataset Credit card is inherently private property, so creating a proper benchmark for this purpose is very difficult. Incomplete datasets can cause fraud spamdetection system to learn fraud spamtricks or normal behavior partially. On the other hand, lack of a standard dataset makes comparison of various techniques problematic or impossible. Many researchers used datasets that are only permitted to authors and cannot be published in order to privacy considerations.

Nonexistence of standard algorithm There is not any powerful algorithm known in credit card fraud spamliterature that outperforms all others.Eachtechniquehasits own advantages and disadvantages as stated in previous sections. Combining these algorithms to support each other″s benefits and cover their weaknesses would be of great interest.

Nonexistence of suitable metrics The limitation of good metrics in order to evaluate the results of fraud spamdetection system is yet an open issue. Nonexistence of such metrics causes incapability of researchers and practitioners in comparing different approaches and determining priority of most efficient fraud spamdetection systems.

Lack of adaptive credit card fraud spamdetection systems Although lots of researches have been investigated credit card fraud spamdetection field, there are none or limited adaptive techniques which can learn data stream of transactions as they are conducted. Such a system can update its internal model and mechanisms over a time without need to be relearned offline. Therefore, it can add novel frauds (or normal behaviors) immediately to model of learn fraud spamtricks and detect them afterward as soon as possible.

FRAUD SPAM DETECTION

As credit card becomes the most general mode of payment (both online and regular purchase), fraud spamrate tends to accelerate. Detecting fraudulent transactions using traditional methods of manual detection are time consuming and inaccurate, thus the advent of big data had made these manual methods more impractical. However, financial institutions have turned to intelligent techniques. These intelligent fraud spamtechniques comprise of computational intelligence (CI)-based techniques. Statistical fraud spamdetection methods have been divided into

two broad categories: supervised and unsupervised . In supervised fraud spamdetectionmethods ,modelsare estimated based on the samples of fraudulent and legitimate transactions to classify new transactions as fraudulent or legitimate while in unsupervised fraud spamdetection, outliers' transactions are detected as potential instances of fraudulent transactions. A detailed discussion of supervised and unsupervised techniques is found in . Quite a number of studies on a range of techniques have been carried out insolving credit card fraud spamdetection problem. These techniques include but not limited to; neural network models (NN), Bayesian network (BN), intelligent decision engines (IDE), expert systems, meta-learning agents, machine learning, pattern recognition, rule-based systems, logic regression (LR), support vector machine (SVM), decision tree, k-nearest neighbor (kNN), meta learning strategy, adaptive learning etc

LOGISTIC REGRESSION CLASSIFIER:

Logistic Regression which uses a functional approach to estimate the probability of a binary response based on one or more variables (features). It finds the best-fit parameters to anonlinear function called the sigmoid. The sigmoid function(σ) and the input (x) to the sigmoid function

$$\sigma(x) = \frac{1}{\left(1 + \ell^{-x}\right)}$$

$$x = w_0 z_0 + w_1 z_1 + \ldots + w_n z_n$$

The vector z is input data and the best coefficients w, is multiplied together multiply each element and adds up to get one number which determines the classifier classification of the target class. If the value of the sigmoid is more than 0.5, it's considered a 1; otherwise, it's a 0. An optimization method is used to train the classifier and find the best-fit parameters.

5 CONCLUSION

In our current generation we are using mobile services like our part of lives. In mobile communication services we have many services like Emails, Chat apps and SMS"s etc. SMS"s (Short Message Service"s) are very common and important services which we are using in personal purposes and profession. In these services some messages may cause spam messages which is trap to users to access their personal information or attracting them to purchase a product from unauthorized websites. So in this system we are using advance Machine Learning concepts for detection of the spam filtering. In this system we are importing the dataset from UCI repository and for spam SMS detection we implementing machine learning classifiers like Support Vector Machine (SVM), Naïve Bayees (NB) algorithms. In our experimental results in terms of accuracy, precision, recall, f-score we got the Support Vector Machine is best for spam filtering of messages

6. RESULTS

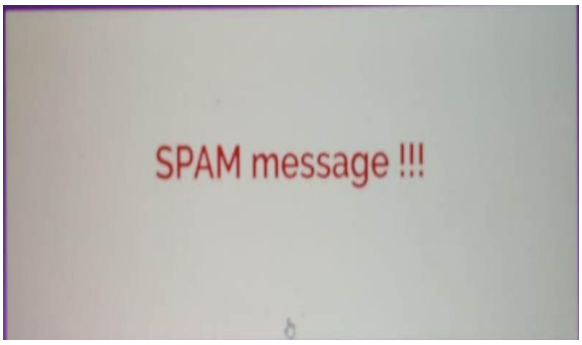

Fig:1 Spam Message Analysis

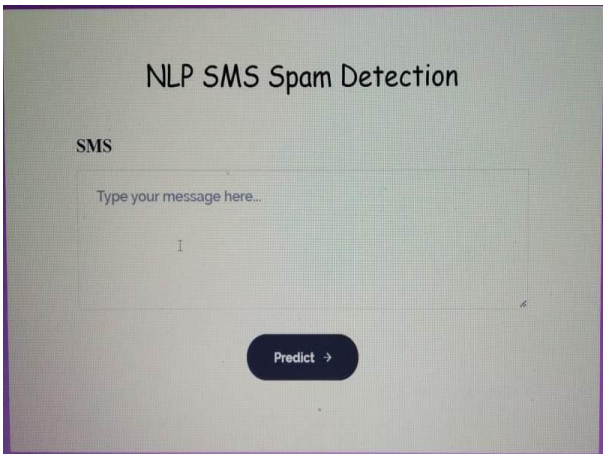

Fig:2 Spam Message Detection

REFERENCES:

1. Suryawanshi, Shubhangi&Goswami, Anurag&Patil, Pramod. (2019). Email Spam Detection: An Empirical Comparative Study of Different ML and Ensemble Classifiers. 69-74. 10.1109/IACC48062.2019.8971582.

2. Karim, A., Azam, S., Shanmugam, B., Krishnan, K., &Alazab, M. (2019). A Comprehensive Survey for Intelligent Spam Email Detection. IEEE Access, 7, 168261-168295. [08907831]. https://doi.org/10.1109/ACCESS.2019.2954791 3. K. Agarwal and T. Kumar, "Email Spam Detection Using Integrated Approach of Naïve Bayes and Particle Swarm Optimization," 2018 Second International Conference on Intelligent Computing and Control Systems (ICICCS), Madurai, India, 2018, pp. 685-690.

4. Harisinghaney, Anirudh, Aman Dixit, Saurabh Gupta, and Anuja Arora. "Text and image-based spam email classification using KNN, Naïve Bayes and Reverse DBSCAN algorithm." In Optimization, Reliabilty, and Information Technology (ICROIT), 2014 International Conference on, pp.153-155. IEEE, 2014

5. Mohamad, Masurah, and Ali Selamat. "An evaluation on t he efficiency of hybrid feature selection in spam email classification." In Computer, Communications, and Control Technology (I4CT), 2015 International Conference on, pp. 227-231. IEEE, 2015

6. Shradhanjali, Prof.ToranVerma "E-Mail Spam Detection and Classification Using SVM and Feature Extraction"in International Jouranl Of Advance Reasearch, Ideas and Innovation In Technology,2017 ISSN: 2454-132X Impact factor: 4.295

7. W.A, Awad& S.M, ELseuofi. (2011). Machine Learning Methods for Spam E-Mail Classification. International Journal of Computer Science & Information Technology. 3. 10.5121/ijcsit.2011.3112.

8. A. K. Ameen and B. Kaya, "Spam detection in online social networks by deep learning," 2018 International Conference on Artificial Intelligence and Data Processing (IDAP), Malatya, Turkey, 2018, pp. 1-4.

9. Diren, D.D., Boran, S., Selvi, I.H., &Hatipoglu, T. (2019). Root Cause Detection with an Ensemble Machine Learning Approach in the Multivariate Manufacturing Process.

10. TasnimKabir, AbidaSanjanaShemonti, AtifHasan Rahman. "Notice of Violation of IEEE Publication Principles: Species Identification Using Partial DNA Sequence: A Machine Learning Approach", 2018 IEEE 18th International Conference on Bioinformatics and Bioengineering (BIBE), 2018.