Available Online at www.ijcsmc.com

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X
IMPACT FACTOR: 7.056

IJCSMC, Vol. 10, Issue. 6, June 2021, pg.10 – 19

SMS Spam Detection Framework Using Machine Learning Algorithms and Neural Networks

[1]T. Jhansi Rani; [2]T. Jaya Vumesh; [3]P. Saiteja; [4]V. Ajay Kumar Reddy; [5]M. Meghana
[1]Assistant Professor, Dept of CSE, GITAM School of Technology, Hyderabad, Telangana, INDIA
[2,3,4,5]B-Tech, Dept of CSE, GITAM School of Technology, Hyderabad, Telangana, INDIA
DOI: 10.47760/ijcsmc.2021.v10i06.002

Abstract: ***In our current generation we are very much habituated to many mobile services like communication, ecommerce etc. In mobile communication services 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 attacks which is trap to users to access their personal information or attracting them to purchase a product from unauthorized websites. It is very easy for companies send any information or service or alert to their customers/users with these SMS API's. Based on these services it is also possible for sending spam messages. So in this system we are using advance Machine Learning concepts for detection of the spam filtering in the SMS's. 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), K-Nearest Neighbor (KNN), and Neural Networks (NN) algorithms and with their metrics like accuracy, precision, recall and f-score. We calculate performances between there algorithms as well as we show the experiment results with visualization techniques and analyses which algorithm is best for spam SMS detection.***

Keywords: Short Message Service, Support Vector Machine, K-Nearest Neighbor, Naïve Bayees

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 financial effect [1][3], 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 Trojans etc [5]. 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 sessions [4]. 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.

The rest of the paper in section 2: literature survey, we discuss the related work of this concept. Section 3: proposed methodology: we discuss about the system model, about the data collection, preprocessing, features extractions, and classifications of our algorithms steps. Section 4: results,

we show the comparison graph results in between our three algorithms. Section 5: conclusion. Section 6: Future work, we mention the scope of the future work. Section 7: References.

2. LITERATURE SURVEY

Collin Mulliner and Charlie described in [10], how inject the spam messages in to the communication layer, and how to prevent the spam message injections. In this they proposed a framework with Radio Interface Layer uses monitor telephony. This can be only deployed in the Android based operating system mobiles for monitoring and testing the SMS injections. Bhowmick et al. proposed concept called TREC (Text Retrieval Conference) [6] is for content based spam filter methodology. This spam detection methodology is developed based on context based because of SMS's and Emails mainly depends on the text related data. In this spam detection is proposed by using four steps. Tokenization, Stop words removal, Stemming and Feature Extraction. In Tokenization, it is a process of the splitting the statement into number of words. Next in Stop words removal process is to remove the non- informative words like 'an', 'is', 'the' etc will remove here. In next process Stemming, in this converting the word to their morphological format. In the next main step Feature Extraction, it is comparison and calculation of input words and spam and ham data corpus. It is used BoW (Bag of Words) model. It is finalize the result of the SMS is spam or ham.

Kanaris et al, proposed the concept called Anti-Spam filtering [4], by using the traditional data mining technique called N-Grams. Here input statement will tokenize with type on N-grams. Here types are may be 1-gram, 2-gram, 3-gram etc. After conversion of N-grams the tokens will compare to the Spam and Ham data corpus. *Yerazunis et al.* proposed concept called SBPH (Sparse Binary Polynomial Hashing) [7], main motive behind this is, if we compare the input statement with Spam or Ham data corpus, this may leads to heavy computation cost because of the text to text comparison. So in this survey they convert to text to hashing before comparison between input data and data corpus. Both input data and corpus data we need to covert to hashing, with this we can achieve computation cost. For detection of Spam Messages survey used concepts like BoW, after this proposed N-Grams, instead of text to text comparison in other survey proposed SBPH, to enhancing this concept and for reduction of the computation cost. *Siefkes et al.* proposed a concept called OSP (Orthogonal Sparse Bigrams) [8], in SBPH we need to take the word combinations of 2^{N-1} for matching with the corpus. But in OSP, by using

proposed concepts we can reduce combinations to N-1. By using proposed skip feature we can reduce the number of feature sets.

Ying et al. proposed a concept and proved that, it is better solution than the OSP. The proposed methodology is LC(Local Concentration) [9]. In this they are calculated weighted score instead of the matched score. Here they proposed two sliding window concepts called Fixed Length and Variable Length. *Hamandi et al.* explained the SMS monitoring system [10] in the Android operating system kernel level. In this they explained +CMT command is used for the monitoring the forward and Receive messages from the Android Mobiles. But these types of concepts are limited to certain operating systems of the mobiles only.

3. PROPOSED METHODOLOGY

3.1 System Model:

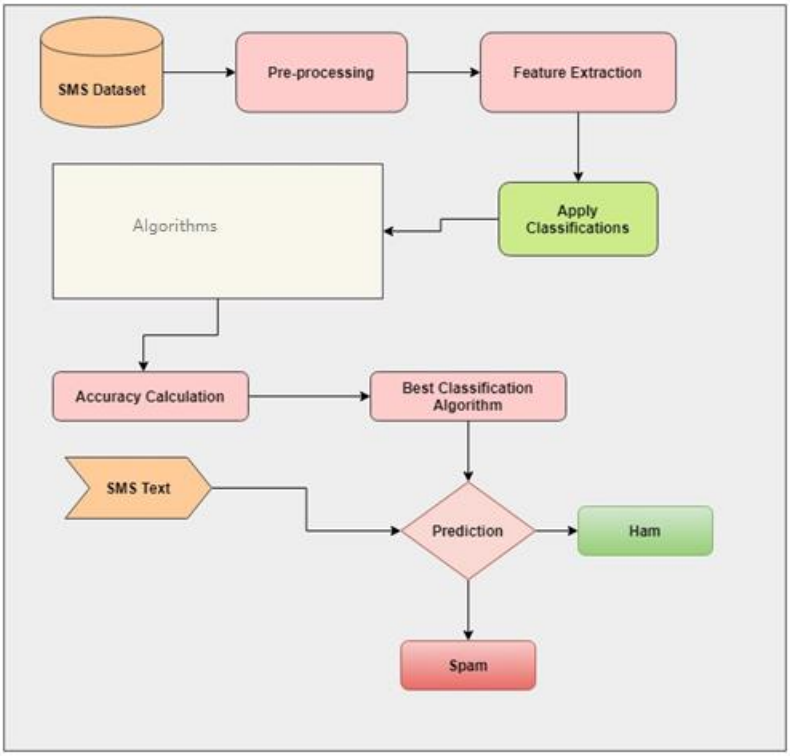


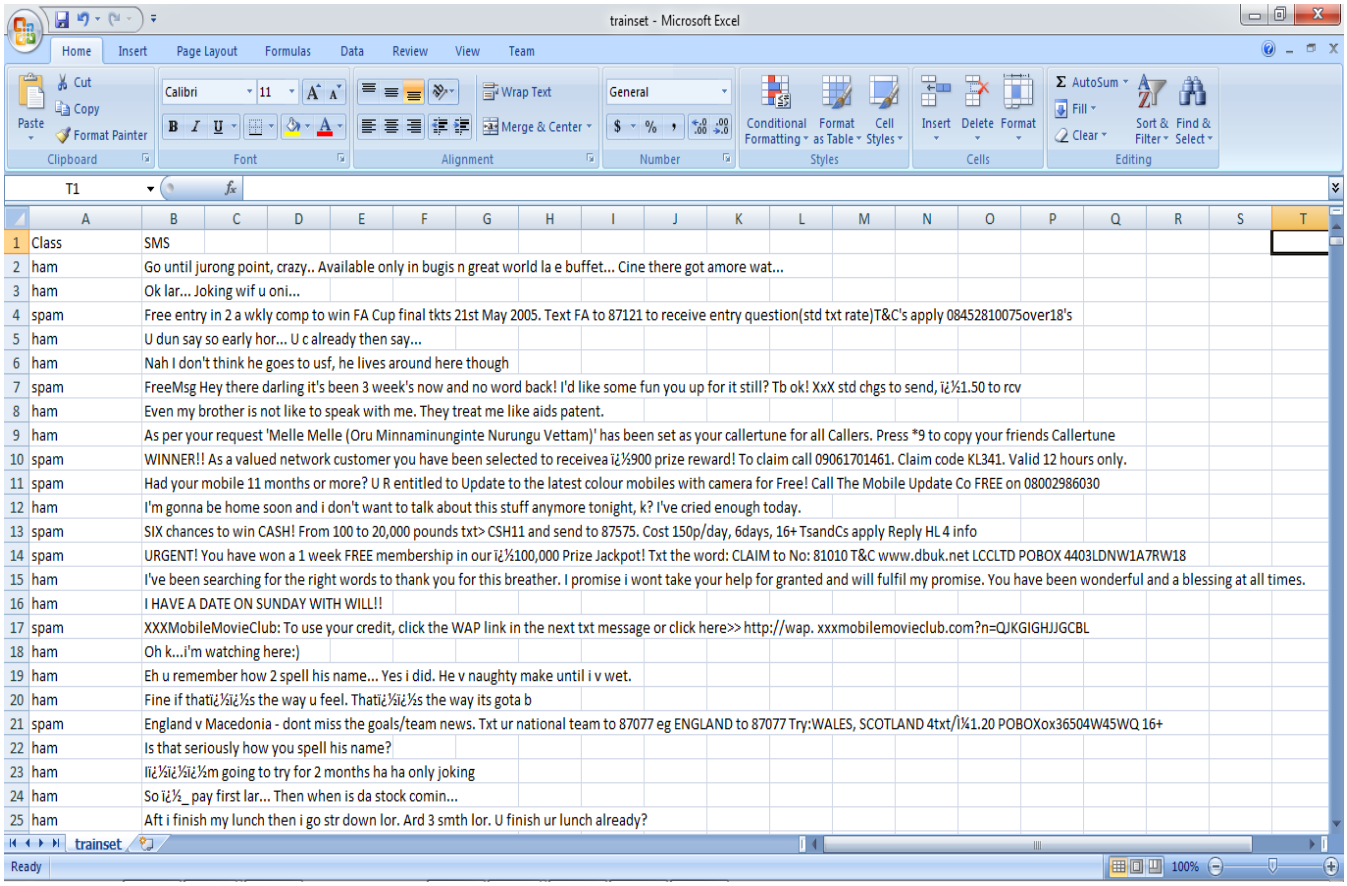
Fig.1 System Architecture

In Figure.1, the system needs load training SMS dataset which is access from UCI repository and performing preprocessing for removing stop and stemming words as well as using of feature extraction we can retrieve most repeated keywords as dataset attributes with help of

TfidfVectorizer class which is inbuilt python library. Once finishing of preprocessing and feature extraction then it can split the dataset as training and testing and we choose machine learning classifiers like SVM, KNN, NB for SMS Spam detection, while the prediction process it can take test data as input and it returns output as spam or ham with help of training dataset.

3.2 Dataset Collection:

In this system the SMS dataset is shown in Figure 2 which is contain ham and spam messages with two columns 'Class' and 'SMS' has imported from UCI repository [11]. These dataset contains 5574 messages and among them 4827 were ham messages, 747 were spam messages. The dataset is stored in CSV (Comma Separated Value) file format where each row represents a single message.



	A	B
1	Class	SMS
2	ham	Go until jurong point, crazy.. Available only in bugis n great world la e buffet... Cine there got amore wat...
3	ham	Ok lar... Joking wif u oni...
4	spam	Free entry in 2 a wkly comp to win FA Cup final tkts 21st May 2005. Text FA to 87121 to receive entry question(std txt rate)T&C's apply 08452810075over18's
5	ham	U dun say so early hor... U c already then say...
6	ham	Nah I don't think he goes to usf, he lives around here though
7	spam	FreeMsg Hey there darling it's been 3 week's now and no word back! I'd like some fun you up for it still? Tb ok! XxX std chgs to send, ï¿½1.50 to rcv
8	ham	Even my brother is not like to speak with me. They treat me like aids patent.
9	ham	As per your request 'Melle Melle (Oru Minnaminunginte Nurungu Vettam)' has been set as your callertune for all Callers. Press *9 to copy your friends Callertune
10	spam	WINNER!! As a valued network customer you have been selected to receivea ï¿½900 prize reward! To claim call 09061701461. Claim code KL341. Valid 12 hours only.
11	spam	Had your mobile 11 months or more? U R entitled to Update to the latest colour mobiles with camera for Free! Call The Mobile Update Co FREE on 08002986030
12	ham	I'm gonna be home soon and i don't want to talk about this stuff anymore tonight, k? I've cried enough today.
13	spam	SIX chances to win CASH! From 100 to 20,000 pounds txt> CSH11 and send to 87575. Cost 150p/day, 6days, 16+ TsandCs apply Reply HL 4 info
14	spam	URGENT! You have won a 1 week FREE membership in our ï¿½100,000 Prize Jackpot! Txt the word: CLAIM to No: 81010 T&C www.dbuk.net LCCLTD POBOX 4403LDNW1A7RW18
15	ham	I've been searching for the right words to thank you for this breather. I promise i wont take your help for granted and will fulfil my promise. You have been wonderful and a blessing at all times.
16	ham	I HAVE A DATE ON SUNDAY WITH WILL!!
17	spam	XXXMobileMovieClub: To use your credit, click the WAP link in the next txt message or click here>> http://wap. xxxmobilemovieclub.com?n=QJKGIGHJJGCBL
18	ham	Oh k...i'm watching here:)
19	ham	Eh u remember how 2 spell his name... Yes i did. He v naughty make until i v wet.
20	ham	Fine if thatï¿½ï¿½s the way u feel. Thatï¿½ï¿½s the way its gota b
21	spam	England v Macedonia - dont miss the goals/team news. Txt ur national team to 87077 eg ENGLAND to 87077 Try:WALES, SCOTLAND 4txt/Ì¼1.20 POBOXox36504W45WQ 16+
22	ham	Is that seriously how you spell his name?
23	ham	Iï¿½ï¿½ï¿½m going to try for 2 months ha ha only joking
24	ham	So ï¿½_ pay first lar... Then when is da stock comin...
25	ham	Aft i finish my lunch then i go str down lor. Ard 3 smth lor. U finish ur lunch already?

Fig 2. SMS Dataset

3.3 Preprocessing:

In this system first we need to gather the training dataset which is in CSV file format so that we need to read the file data with help of pandas library for convert to list array as well as the input message which is given by user that one should be append to list array, because the machine

cannot understand the file format data. The finishing of converting process then it can remove the stop words and stemming words from training and testing data which is reduced irrelevant data.

3.4 Features Extraction:

In this system we are using SMS dataset which is in text format so that for comparison word to word it takes long time for spam SMS detection which is not recommended. To overcome this after finishing of preprocessing stage we need to convert from text (messages) to numeric data. So in this system we are implementing TfidfVectorizer python library for generating numerical data from text data which is available in list format. Here we need to apply TfidfVectorizer for both training and testing with n-gram weights which is useful to splitting keywords. The TfidfVectorizer class contains fit_transform() function which is take input as out number of messages in list format and generate features like attributes with help of get_feature_names() function. Finally these features are become independent values and Class column become a dependent value and apply the any machine learning classifier then predict the given message is spam or ham.

3.5 Classifications:

NN:

The neural network classifier is advance of all classifiers, because it was follow the brain neurons working process. The NN classifier takes the three layers such as input layer, hidden layer and output layer. Here the input layer collect the dataset features and passing to hidden layers to features classification and finally the classification results can share to output layer. When the classifier return the predictable output where it matches highest percentage with output layers. This classifier also can import *sklearn.neural_network* package *MLPClassifier* for heart disease prediction. Follow the below snippet code:

```
from sklearn.neural_network import MLPClassifier
rf = MLPClassifier()
rf.fit(x_train, y_train)
pre_cls = rf.predict(x_test)
```

SVM:

The support vector machine classifiers in an important classifier because of it's classification advantages. The SVM classifier while classification of features, first it can draw the margins between different classes and the hyper plane line can separate with support vectors which means

the nearest classes to that hyper plane line. Here this system can separate hyper plane with POSITIVE and NEGATIVE features and select the nearest support vectors and build the training model to predict the heart disease status. The below syntax is following the code of heart disease prediction.

```
from sklearn import svm
svm = svm.SVC()
svm.fit(x_train, y_train)
pre_cls = svm.predict(x_test)
```

KNN:

The K-nearest neighbor classifier is a different learning classifier compare with another machine learning classifier, because it follow the Euclidian distance formula to calculate the distance. This classifier while prediction it calculates the distance between with each records then it returns the distance and store it like this follow the last record and it can return the predictable output value which distance is less to compare with all distances and that one become our heart disease predictable output like POSITIVE or NEGATIVE. It is also import the *KNeighborsClassifier* module from this *sklearn.neighbors*. Here we took *K* value is 1 for pick the nearest distance value as output.

```
from sklearn.neighbors import KNeighborsClassifier
knn=KNeighborsClassifier()
knn.fit(x_train, y_train)
pre_cls = knn.predict(x_test)
```

4. RESULTS

In our experimental results of SMS spam detection model, we compare our algorithms in terms of the accuracy, precision, recall and f-score.

Build Model: Build model is used to train the model using testing dataset

Performance evaluation: Performance evaluation is used to see the results

Results

	Algorithm	Accuracy	Precision	Recall	F1_Score
1	SVM	82.0	0.81	0.76	0.73
2	KNN	84.5	1.0	0.69	0.82
3	NN	95.0	0.95	0.91	0.92

Fig 3. Performance Table

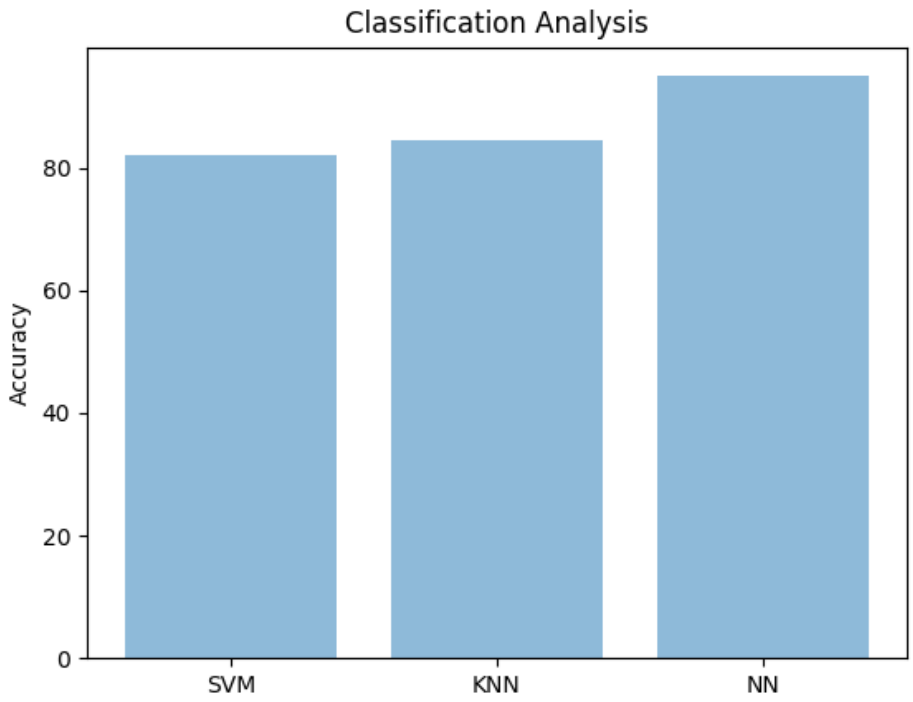


Fig 4. Accuracy Graph

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), K-Nearest Neighbor (KNN), and 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. FUTURE WORK

In our paper we introduced Machine Learning concepts for prediction of spam based on the content data like only message content. In future work we can elaborate this topic to prediction by using content and context data like Host address of the SMS, sender, number of times received, URL's in the messages etc.

REFERENCES

[1]. D. Barrera et al., "A Methodology for Empirical Analysis of Permission-Based Security Models and Its Application to Android," Proc. 17th ACM Conf. Computer and Communications Security (CCS 10), ACM, pp. 73-84, 2010.

[2]. Stefan Frei, Thomas Duebendofer, Gunter Ollman, and Martin May, Understanding the Web browser threat: Examination of vulnerable online Web browser populations and the insecurity iceberg Archived September 11, 2016, at the Wayback Machine, Communication Systems Group, 200.

[3]. I. Burguera, U. Zurutuza, and S. Nadjm-Tehrani, "Crowdroid: Behavior-Based Malware Detection System for Android," Proc. ACM Workshop Security and Privacy in Mobile Devices (SPMD 11), ACM, pp. 15-26, 2011.

[4]. Kanaris I, Kanaris K, Houvardas I, Stamatatos E (2006) Words vs. Character n-grams for Anti-spam Filtering. International Journal on Artificial Intelligence Tools XX(X):1–20.

[5]. J. Drew, Malware growth maintains rapid pace as mobile threats surge, found 10 September 2012, Accessible via: http://www.journalofaccountancy.com/News/20126400.htm.

[6]. Bhowmick, Alexy & Hazarika, Shyamanta. (2018). E-Mail Spam Filtering: A Review of Techniques and Trends. 10.1007/978-981-10-4765-7_61.

[7]. Yerazunis WS (2003) Sparse Binary Polynomial Hashing and the CRM114 Discriminator Rough Guide to this Talk. In: MIT Spam Conference.

[8]. Siefkes C, Assis F, Chhabra S, Yerazunis WS (2004) Combining Winnow and Orthogonal Sparse Bigrams for Incremental Spam Filtering. In: European Conference on Machine Learning (ECML), pp 410–421.

[9]. Zhu, Yuanchun & Tan, Ying. (2011). A Local-Concentration-Based Feature Extraction Approach for Spam Filtering. Information Forensics and Security, IEEE Transactions on. 6. 486 - 497. 10.1109/TIFS.2010.2103060.

[10]. C. Mulliner, C. Miller, "Injecting SMS Messages into Smart Phones for Security Analysis," in Proceedings of the 3rd USENIX Workshop on Offensive Technologies (WOOT), 2009.

[11]. SMS Spam Collection Data Set from UCI Machine Learning Repository, http://archive.ics.uci.edu/ml/datasets/SMS+Spam+Collection.

[12]. F. Peng, "Augmenting Naive Bayes Classifiers with Statistical Language Models", Computer Science Department Faculty Publication Series", Paper 91, 2003.

[13]. M. Akhil, B. L. Deekshatulu, and P. Chandra, "Classification of Heart Disease Using K- Nearest Neighbor and Genetic Algorithm," Procedia Technol., vol. 10, pp. 85–94, 2013.

