


A Novel Approach to Detect Spam and Smishing SMS using Machine Learning Techniques

Ankit Kumar Jain, National Institute of Technology, Kurukshetra, India

 <https://orcid.org/0000-0002-9482-6991>

Sumit Kumar Yadav, Income Tax Department, Government of India, India

Neelam Choudhary, National Institute of Technology, Kurukshetra, India

ABSTRACT

Smishing attack is generally performed by sending a fake short message service (SMS) that contains a link of the malicious webpage or application. Smishing messages are the subclass of spam SMS and these are more harmful compared to spam messages. There are various solutions available to detect the spam messages. However, no existing solution, filters the smishing message from the spam message. Therefore, this article presents a novel method to filter smishing message from spam message. The proposed approach is divided into two phases. The first phase filters the spam messages and ham messages. The second phase filters smishing messages from spam messages. The performance of the proposed method is evaluated on various machine learning classifiers using the dataset of ham and spam messages. The simulation results indicate that the proposed approach can detect spam messages with the accuracy of 94.9% and it can filter smishing messages with the accuracy of 96% on neural network classifier.

KEYWORDS

Machine Learning, Mobile Phishing, Short Message Service, Smishing, Spam, Text Classification

1. INTRODUCTION

1.1. Contextualization

Smishing is a cyber-security attack in which the mobile user is deceived into installing malicious software into their mobile phone. Smishing word is constructed by the combination of two words i.e. SMS + Phishing = Smishing. In other words, smishing message is a harmful spam message which aims at stealing mobile users' sensitive data (Goel & Jain, 2018). These messages may contain a link. On following the link, the user is asked to enter their details for verification purpose (Choudhary & Jain, 2017). Attackers also lure victims by sending messages that look like they originate from an authentic bank, stating that their account has been locked and to unlock the account, the victim is asked to follow the link in the message. The purpose of spam messages is generally to promote some product or to disturb the user with useless messages. On the other hand, phishing message always have some criminal intent. The effect of smishing attack is financial loss and identity theft.

DOI: 10.4018/IJESMA.2020010102

1.2. Importance/Relevance of the Theme

Smishing attacks are increasing day by day as attackers find SMS, a cheaper and more convenient way to communicate with victims (“Why do phishing attacks work better on mobile phones,” 2011). Smartphone users are considered to be three times more likely to fall victims of phishing attacks than desktop users (MOBILE THREAT REPORT, 2012). In 2014, cloud mark report (Hacked Hotel Phones Fueled Bank Phishing Scams, 2015) identified a phishing SMS, which attempted to steal user’s secret information like credit card number, bank account details, etc. A smishing message was sent to thousands of people with different bank affiliations (SMiShing & Vishing News, 2017).

1.3. Research Question

There are various security measures available to control SMS Spam problem, but these are not so mature. Many Android apps (Spam Blocker App; Mr. Number - Caller ID & Spam Protection) are also available on play store. However, their detection accuracy is not up to the mark. Moreover, existing text filtering techniques mainly focus on email spam (Diale, Celik, & Walt, 2019). However, with the popularity of mobile devices, SMS spam and smishing is the one of the major issues these days. Existing approaches mainly focus only on SMS Spam detection. However, there is no efficient technique developed which can efficiently filter out smishing text messages from SMS Spam.

1.4. Objectives

This paper presents a novel machine learning based approach to detect the spam and smishing messages. The paper used effective feature set for detection of spam and smishing messages. The proposed approach is divided into two phases. The first phase distinguishes the spam messages from ham messages using eleven basic features. The second phase filters smishing message from the spam message using four phishing features. The performance of our proposed method is evaluated on various machine learning classifiers using the dataset of ham and spam messages. Followings are the major contributions of the proposed approach:

- Implementation of a classifier for spam and smishing messages.
- Identification of four outstanding phishing features suitable for mobile smishing detection.
- Detection of zero day mobile phishing attacks using the proposed features.
- Utilization of information gain values to get the best features for SMS spam and smishing detection.

1.5. Structure of the Paper

The rest of the paper is organized as follows. The related work is presented in Section 2. The system architecture of our proposed scheme is described in Section 3. The experimental details including various experimental results are presented in Section 4. Finally, the conclusion and future work is presented in Section 5.

2. RELATED WORK

These days various tools and techniques are developed to ensure safety and security in the mobile devices. However, problem of security still arises. Smishing attacks are rapidly growing on mobile devices (Goel & Jain, 2017).

A mechanism to normalize and expand short and noisy messages is given by Almeida et al. (2011). It is a novel approach that can highly improve the quality of the text messages before classification, due to which better attributes are obtained in turn improving the classification accuracy. Two dictionaries

- lexicographic and semantic dictionaries are used. It has been seen that text processing removes the basic text representation problems, improving the classification accuracy.

To filter spam SMS text messages, Ma et al. (2016) proposed a message topic model (MTM). MTM is based on probability topic model. To appropriately represent SMS message, pre-processing is used along with some background and symbol term. K-means algorithm is used to avoid sparse problem. They clustered spam text messages into rough classes and then accumulated spam messages in single file to identify word co-occurrence patterns.

Joo et al. (2017) used Naïve Bayes classification algorithm to improve the efficiency of the smishing detection scheme. They have statistical learning method to filter simple text messages and smishing text messages. They have used morphological analyzer to extract noun words and Bayesian classifier to represent the uncertainties related to the models and parameter values.

Chan et al. (2015) presented two methods for SMS Spam filtering, namely, feature reweighting method and good word attack. Both approaches focus on the length of the message along with considering the weight of the message.

Etaiwi and Awajan (2017) investigated the comparison of two features implementing them separately on the same database. The two features used in this paper include a bag of words and word count, which are compared using various algorithms like SVM, random forests, etc. However, this approach achieved only 80-90% detection accuracy in the case of a bag of words and 60-70% for word count.

Yadav et al. (2011) developed an application “SMSAssassin” which can be used over mobile phones for filtering spam messages in real time. Naïve Bayesian Classifier and SVM are two classifiers used to classify the messages as ham or spam. Crowd-sourcing is used to keep track of new patterns and keywords, and to update the list of spam patterns and keywords. Authors have used two lists - GlobalSpamKeywordsFreq and SpamKeywordsFreq list at the server and the mobile phones, respectively to keep track of spam keywords. There is a UserPreferencesList in the application where a user can specify ham and spam keywords of their choice.

Jain and Gupta (2018) proposed a rule-based framework for detection and prevention of smishing messages. The authors have identified nine rules that effectively differentiate smishing messages from ham messages. These rules are extracted from the content of the message. The scheme was analyzed on three different classification algorithms – decision tree, PRISM and RIPPER. RIPPER gave best results as compared to other algorithms. Additionally, the scheme is able to detect zero hour smishing attacks.

El-Alfy and AlHasan (2016) proposed a text based message filtering mechanism for SMS as well emails. They have used 11 features for the detection of malicious messages and evaluated their approach on five standard SMS and email dataset.

Silva et al. (2017) proposed a technique for detecting spam text messages on the basis of Minimum Description Length (MDL) principle. MDL based technique provides effective and efficient message classification approach. This technique gives low computational cost even with large amount of data.

Ali and Maqsood (2018) developed a .NET library written in C# which provides a cross-platform solution for spam detection. They have discussed some algorithms and have implemented Random Forests. The .NET library developed can run on any platform as per requirements. They have trained the C# library efficiently resulting in a good accuracy.

Idris et al. (2015) proposed an email spam detection algorithm that uses particle swarm optimization algorithm to detect spam messages. The designed system is an improvement of random detector generation in negative selection algorithm. The accuracy achieved by the system is 83.20%, while, for negative selection algorithm, the accuracy achieved is 68.86%.

3. PROPOSED APPROACH

Smishing message is a harmful spam message that steals personal credentials. As per our observation, our research work finds the followings characteristics of a smishing message:

1. Contains the malicious fake links or mobile number
2. Advertises something like providing free minutes, etc.
3. Asks the user to subscribe to some service.
4. Announces, the user as a winner of some fake contest and luring him using the prize money.
5. Intends to spread some fake news.

Therefore, this article considers these characteristics while making the proposed feature set. It is difficult to detect smishing message, therefore, we have analyzed fake messages in depth to filter out smishing messages from spam messages. Our scheme is able to effectively detect spam as well as smishing messages with high accuracy. The proposed approach first filters out the spam messages and then filters the smishing messages from these spam messages with the help of identified features. The architecture and working of the proposed framework is discussed below:

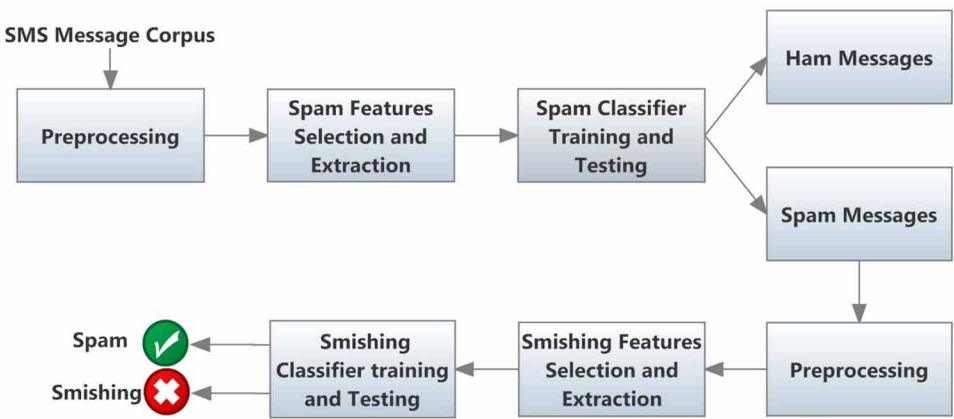
3.1. System Architecture

Figure 1 shows the architecture of proposed scheme. The first phase of our method extracts spam features from the SMS. After filtering spam SMS, the second phase classifies smishing messages. Detailed flow chart of proposed approach is presented in figure 2.

Phase 1: Filtering Spam messages - In the first phase, each incoming message is classified into spam and ham categories. The mechanism of filtering spam messages is discussed as follows:-

1. Preprocessing - Preprocessing the entire SMS Spam Corpus is done to remove the redundancy. The SMS Spam Corpus is stored and handled using Microsoft Excel to clean the data from noise.
2. Spam Features Selection and Extraction – The proposed approach uses 11 basic features to filter out spam messages. A feature vector value 1 is set for spam messages and 0 for ham messages.
3. Classification Training- The labeled dataset is used to train the machine learning classifier. Further, the cross validation set improves the parameters, and evaluates the performance of the method.

Figure 1. Architecture of proposed scheme



Phase 2: Filtering Smishing messages – Once the approach classifies ham and spam messages, the next phase is used for filtering the smishing SMS from spam ones. The mechanism of filtering smishing messages from spam messages is discussed as follows:-

1. **Preprocessing** - This phase labels the message as spam or smishing. We select smishing messages manually by reading all spam messages as no benchmark dataset is available for smishing messages. Afterwards, a feature vector value 1 is set for smishing messages and 0 is set for spam messages.
2. **Smishing Features Selection and Extraction**- In this, our approach selects four outstanding features based on Smishing characteristics in order to filter out Smishing messages from spam messages.
3. **Smishing Classifier Training and Testing**- The trained classifier separates the message into two classes namely, spam and smishing. The rest of the process is similar for all the stages. In this, a message is checked for smishing. First, the URL is checked for the message. If the URL is present, it is a phishing message otherwise, the second feature is checked to determine whether award token is present or not and so on. After filtering all our features, if a message is smishing, then a warning message is shown to the user otherwise the message is declared as spam.

3.2. Feature Selection

Feature selection is a critical task for our approach. It does not require sophisticated analysis to avoid the delay in message service. Moreover, to increase the spam and smishing detection, accuracy features extraction should be highly correlated with the message type. In our proposed approach, the feature set is divided into two categories i.e. Basic Features and Phishing Features. The diagrammatic view of proposed features is presented in figure 3. These features are discussed in detail below:-

3.2.1. Basic Features

These features are used to filter SMS spam from message corpus. Our scheme used a set of 11 features $\{F_1, F_2, F_3, \dots, F_{11}\}$. Following are the basic feature set:

F_1 : *Greetings Token*- This feature checks for the presence of greeting keywords like Good Morning, Good Night, Dear, etc. If a message contains greetings token, then it falls under the category of ham message. 3% messages in our dataset contain greeting tokens.

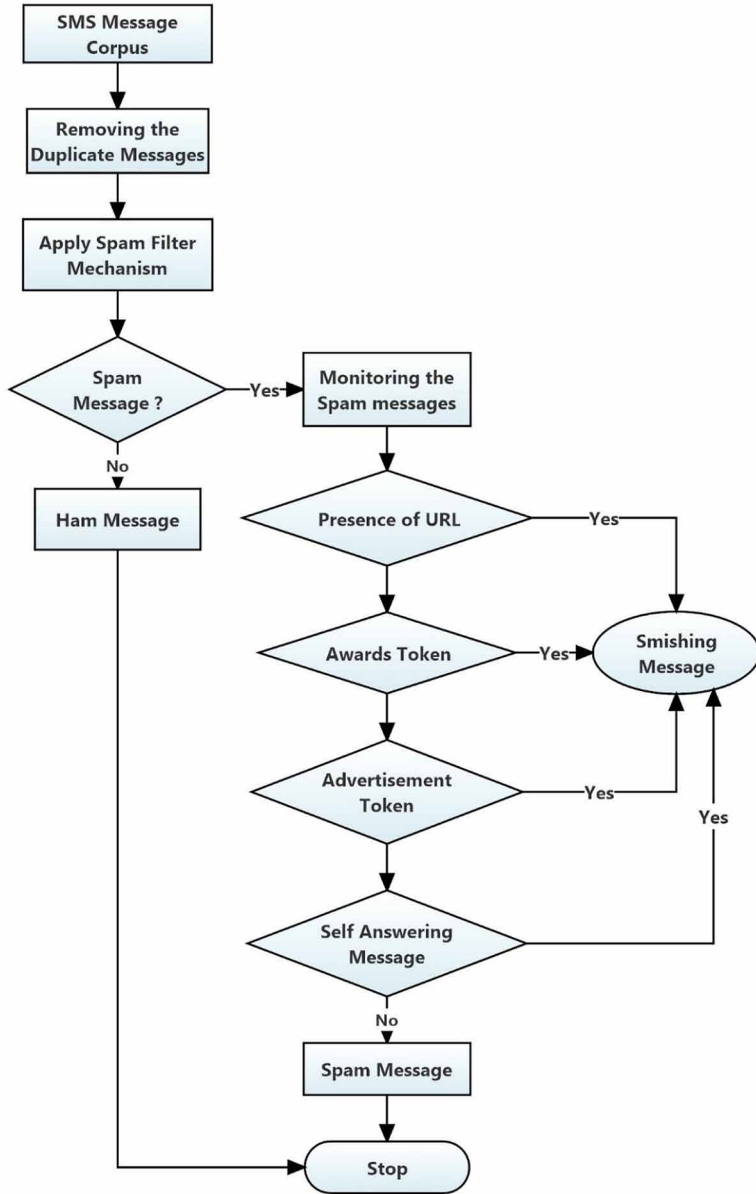
$$F_1 = \text{If } \begin{cases} \text{message contains any greeting token} \rightarrow Ham \\ \text{otherwise} \rightarrow Spam \end{cases} \quad (1)$$

F_2 : *Presence of dot symbol*- The presence of dot symbol is a good sign for ham messages as users use dots while chatting over social networking sites or on various chatting apps. Moreover, users often use dots to separate the sentences.

$$F_2 = \text{If } \begin{cases} \text{message contains dot symbol} \rightarrow Ham \\ \text{otherwise} \rightarrow spam \end{cases} \quad (2)$$

F_3 : *Presence of emotions*- These symbols are generally used while chatting. Emotions can be of two types i.e. positive and negative. Positive emotions like love, nice, sweet, happy, etc. and negative emotions like sad, angry, hurt, nasty, etc. People uses these emotions in the form of symbols to express their feelings or views about the message.

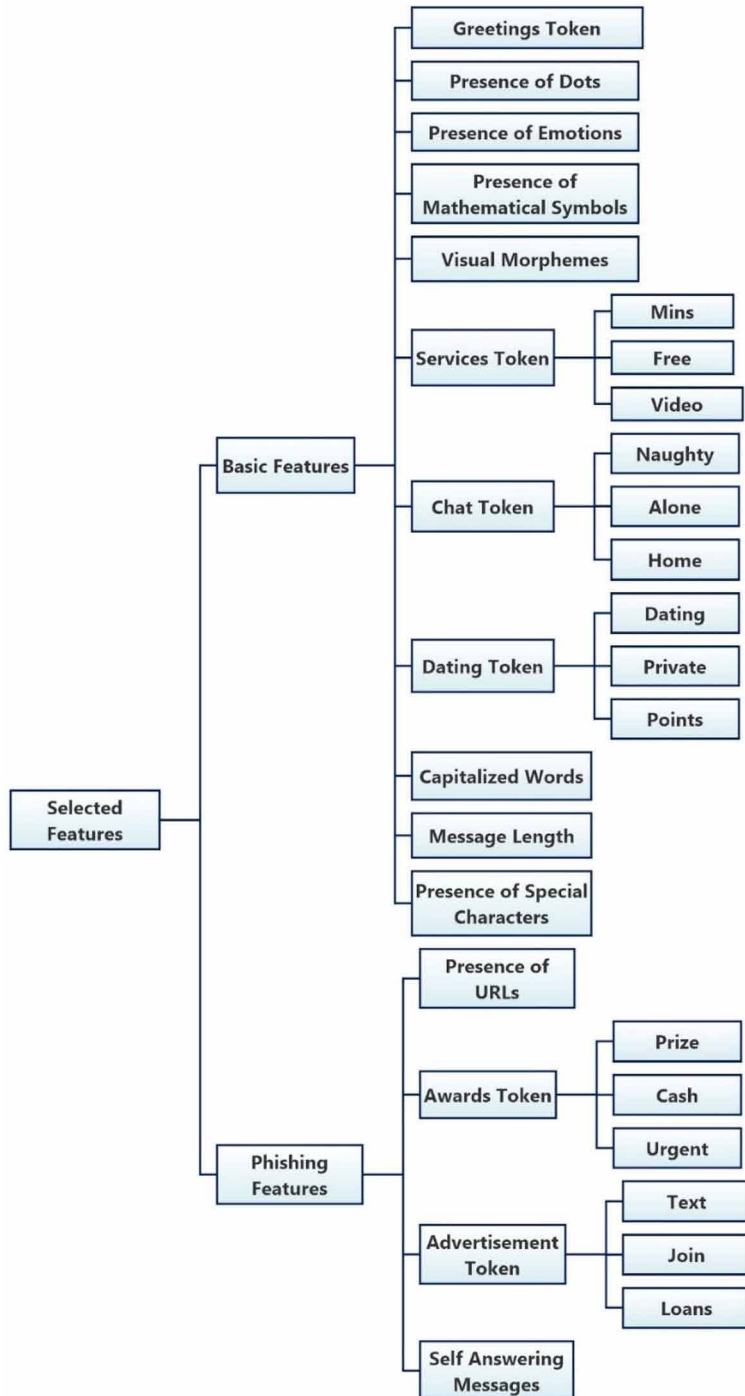
Figure 2. Flowchart of proposed scheme



$$F_3 = \text{If} \begin{cases} \text{message contains positive or negative emotions} \rightarrow \text{Ham} \\ \text{otherwise} \rightarrow \text{spam} \end{cases} \quad (3)$$

F_4 : Presence of special character - The special characters which our approach checks are !, \$, &, # and ~. The character "\$" denotes currency in the fake award SMS, and character "!" shows special attention of user like "CONGRATULATIONS!", "WINNER!", etc.

Figure 3. Proposed feature set



$$F_4 = \text{If} \begin{cases} \text{message contains special characters} \rightarrow \text{spam} \\ \text{otherwise} \rightarrow \text{Ham} \end{cases} \quad (4)$$

F_5 : *Presence of Mathematical Symbols*- Mathematical symbols are not used in genuine text messages because people use simple language to communicate rather than mathematical symbols. In this feature, the approach checks whether the SMS contains +, -, /, ^ and % mathematical symbols.

$$F_5 = \text{If} \begin{cases} \text{message contains mathematical symbol} \rightarrow \text{spam} \\ \text{otherwise} \rightarrow \text{Ham} \end{cases} \quad (5)$$

F_6 : *Services Token*- The presence of suspicious keywords like text, join, loans are considered to be spam keywords. These keywords are very common in spam messages. Spammers lure users by using these keywords, asking them to join their services or obtain loans through them. After reviewing our dataset, the approach was able to find that 5% of the messages contains services token.

$$F_6 = \text{If} \begin{cases} \text{message contains services token} \rightarrow \text{spam} \\ \text{otherwise} \rightarrow \text{Ham} \end{cases} \quad (6)$$

F_7 : *Visual Morphemes*- The visual morphemes like xx, xxx, xxxx are considered to be spam keywords. In our dataset, only 2% of messages contain the visual morphemes tokens.

$$F_7 = \text{If} \begin{cases} \text{message contains visual morphemes} \rightarrow \text{spam} \\ \text{otherwise} \rightarrow \text{Ham} \end{cases} \quad (7)$$

F_8 : *Message length*- In most of the countries, text messages contain a maximum of 160 characters only. Chatting SMS are generally small as compared to spam message. This feature results in spam if the message length is greater than 160 characters.

$$F_8 = \text{If} \begin{cases} \text{message length} \geq 160 \rightarrow \text{spam} \\ \text{otherwise} \rightarrow \text{Ham} \end{cases} \quad (8)$$

F_9 : *Chat Token*- The presence of suspicious keywords like naughty, alone, and home are considered to be spam keywords. These kinds of words are not present in normal genuine messages.

$$F_9 = \text{If} \begin{cases} \text{message contains chat token} \rightarrow \text{spam} \\ \text{otherwise} \rightarrow \text{Ham} \end{cases} \quad (9)$$

F_{10} : *Dating Token*- The presence of suspicious keywords like dating, points and private are considered to be spam keywords. After reviewing our dataset, only 1% of spam messages contains dating token.

$$F_{10} = \text{If} \begin{cases} \text{message contains dating token} \rightarrow \text{spam} \\ \text{otherwise} \rightarrow \text{Ham} \end{cases} \quad (10)$$

F_{11} : *Capitalized Words*- The spammer generally uses uppercase keywords to pursue mobile user's attention.

$$F_{11} = \text{If} \begin{cases} \text{message uses uppercase word} \rightarrow \text{spam} \\ \text{otherwise} \rightarrow \text{Ham} \end{cases} \quad (11)$$

3.2.2. Smishing Features

These features are used to filter smishing messages from spam messages. Proposed scheme used four smishing features (F_{12} - F_{15}). These features are as follows:

F_{12} : *Presence of URL*- Approach considers the presence of URL in the message as a smishing feature, since phishers may ask the user to click on a link to visit a website for winning a prize, fake update or to provide some personal information on the fake webpage. After reviewing our dataset, 14% smishing messages contain URL.

$$F_{12} = \text{If} \begin{cases} \text{URL is present in the } S \rightarrow \text{Smishing} \\ \text{otherwise} \rightarrow \text{Spam} \end{cases} \quad (12)$$

F_{13} : *Awards token*- The token like prize, cash and urgent are used to lure users in terms of cash prize or award. In our dataset, 21% of smishing messages contains awards keywords.

$$F_{13} = \text{If} \begin{cases} \text{message contains awards token} \rightarrow \text{Smishing} \\ \text{otherwise} \rightarrow \text{Spam} \end{cases} \quad (13)$$

F_{14} : *Advertisement Token*- The presence of suspicious keywords like mins, free and video are considered to be smishing keywords. Phishers use these keywords to attract a user by advertising some product or service for free. In our dataset 21% of smishing messages contains advertisement token.

$$F_{14} = \text{If} \begin{cases} \text{message contains advertisement token} \rightarrow \text{Smishing} \\ \text{otherwise} \rightarrow \text{Spam} \end{cases} \quad (14)$$

F_{15} : *Self Answering SMS*- The existence of self-answering message like asking user to subscribe to any facility falls under the category of smishing message.

$$F_{15} = \text{If} \begin{cases} \text{message is self answered} \rightarrow \text{Smishing} \\ \text{otherwise} \rightarrow \text{Spam} \end{cases} \quad (15)$$

4. IMPLEMENTATION DETAILS AND EXPERIMENTAL RESULTS

4.1. Performance Metrics

The evaluation measures set for our experiment are true positive rate (TPR), true negative rate (TNR), false positive rate (FPR), false negative rate (FNR), precision, recall, accuracy, and receiver operating characteristics (ROC) area (Jain & Gupta, 2019). The way to calculate these values can be explained through a confusion matrix as displayed in Table 1.

4.2. Dataset Collection

The proposed approach uses SMS dataset v.1 from SMS Spam research work (Almeida, Hidalgo, & Yamakami, 2011). This dataset contains 5574 text messages in the English language. This corpus has been collected from following sources:-

1. A set of 425 spam SMS extracted from the Grumbletext Website.
2. A list of 450 ham SMS collected from Caroline Tag's PhD Thesis.
3. A subset of 3,375 ham SMS collected from NUS SMS Corpus.
4. A set of 1,002 ham and 322 spam messages are included from the SMS Spam Corpus v.0.1 Big.

Moreover, we have also collected 71 unique smishing message images from pinterest.com and converted these images into text and added them into the dataset.

4.3. Classification Algorithms

Various classification algorithms are used in our approach to find out best prediction accuracy. These classification algorithms are discussed as follows:

1. Naïve Bayes (NB): It is a simple probabilistic classifier based on the Bayes theorem that uses independent feature model along with a decision rule.
2. Neural Network (NN): It is a feed forward neural network which is used to map the input dataset to an output dataset. It uses back propagation which is a supervised learning technique. There are multiple layers of hidden nodes with a nonlinear activation function.
3. Logistic Regression (LR): It assumes a logistic and a normal distribution function, it finds the relationship of the dependent variable with respect to one or more independent variables.

Table 1. Confusion Matrix

	True Results	
	Positive Classification	Negative Classification
Prediction		
Positive	True Positive Rate	False Positive Rate
Negative	False Negative Rate	True Negative Rate

4.4. Experiment 1: Filter Spam messages from Ham messages

There are six classification algorithms used for spam classification, namely NB, LR, NN, DT, J48 and RF. Proposed work used 10-fold cross-validation in the experiment. Table 2 shows the outcomes of our scheme for filtering spam messages on various classification algorithms. From Table 2, we find that Neural Network gives best results in filtering of spam messages.

As our approach gets the best results for Neural Network, so we have analyzed this classification algorithm, for various cross folds. Table 3 presents the numeric values of various folds for Neural Network and Figure 4 shows the performance results.

The histograms of the basic feature set are shown in Figure 5. Each feature of histogram shows every possible value of each feature instance. The feature value is divided into two types - spam messages are colored as red and ham messages are colored as blue. The overlapping between feature values can be easily noticed. This indicates that it is not possible to differentiate between spam and ham messages using a single feature. Thus, each feature has its own importance in detecting spam and ham messages.

4.5. Experiment 2: Filtering Smishing Messages

In this experiment, proposed approach filters smishing messages from spam messages. Table 4 presents classification results of various machine learning classifiers for filtering smishing messages.

Figure 6 shows the ROC curve for Neural Network where x-axis shows the FPR and y-axis presents the TPR for the class type. Here, class type differentiates smishing from spam messages.

In this, our approach gets the best results for Neural Network with an accuracy of 96%. We have further analyzed Neural Network for various cross folds. Table 5 shows the numeric results for different cross folds and Figure 7 shows the comparison of the different rates for various cross folds.

The importance of three main smishing features for smishing message classification is shown in Figure 8. After extracting features we get the list of spam, ham and phishing messages.

Table 2. Classification Results for spam classification

Algorithm	TPR	FPR	Precision	Recall	ROC	Accuracy
Naïve Bayes (NB)	0.979	0.292	0.959	0.979	0.961	0.945
Logistic Regression (LR)	0.982	0.300	0.958	0.982	0.957	0.946
Neural Network (NN)	0.985	0.294	0.959	0.985	0.953	0.949
Decision Table (DT)	0.988	0.351	0.951	0.988	0.945	0.946
J48	0.989	0.335	0.953	0.989	0.916	0.948
Random Forest (RF)	0.977	0.288	0.959	0.977	0.929	0.947

Table 3. Performance Results for Neural Network

Neural Network	TPR	FPR	Precision	Recall	ROC	Accuracy
5:5	0.988	0.315	0.956	0.988	0.955	0.950
4:6	0.988	0.315	0.956	0.988	0.958	0.949
3:7	0.990	0.309	0.957	0.990	0.955	0.952
2:8	0.989	0.315	0.956	0.989	0.955	0.951
1:9	0.985	0.294	0.959	0.985	0.953	0.949

Figure 4. Neural Network result for various folds

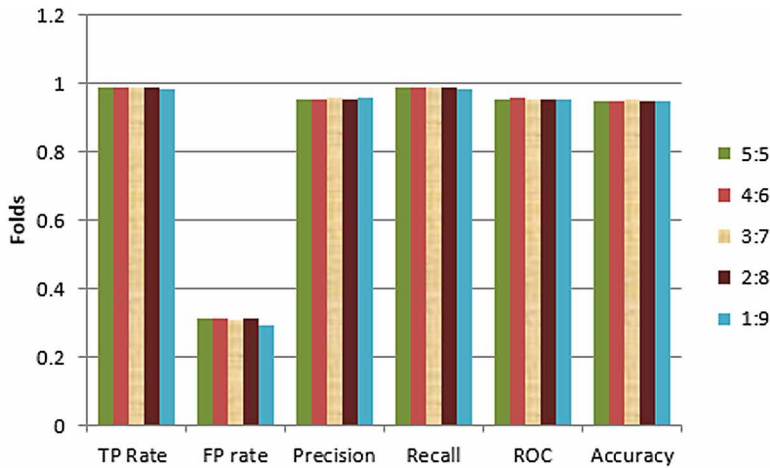
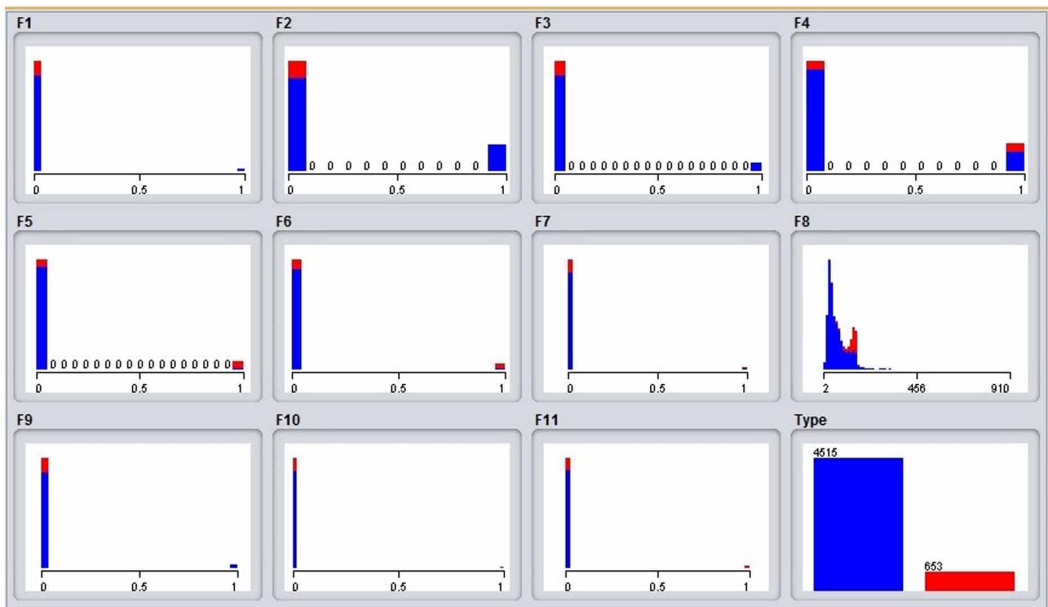


Figure 5. Histograms of spam features



4.6. Importance of Features

Information gain helps in analyzing the quality of each feature and removing redundant features, thus, helping in selecting the best feature. We run information gain feature selection algorithm to choose the ten important features for classification of ham and spam messages and three important features for smishing and spam classification. In Table 6, we present top ten features for filtering spam messages with their respective information gain values and Table 7 shows the top three features for filtering smishing messages with their IG values.

Table 4. Classification results for smishing messages classification

Algorithm	TPR	FPR	Precision	Recall	ROC	Accuracy
Naïve Bayes	0.923	0.010	0.994	0.923	0.960	0.950
Logistic	0.910	0.035	0.989	0.910	0.945	0.946
Neural Network	0.932	0.007	0.994	0.932	0.958	0.960
Decision Table	0.922	0.013	0.990	0.922	0.953	0.952
J48	0.922	0.013	0.990	0.922	0.953	0.952
Random Forest	0.925	0.015	0.991	0.925	0.954	0.958

Figure 6. ROC curve for Neural Network classification algorithm

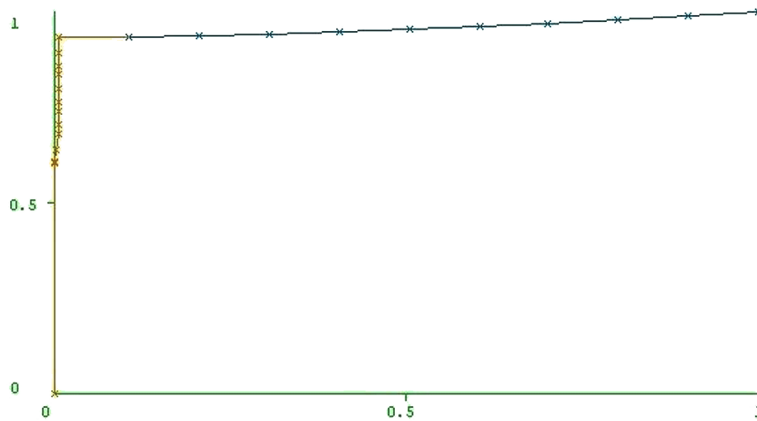


Table 5. Performance Results for Neural Network

Neural Network	TPR	FPR	Precision	Recall	ROC	Accuracy
5:5	0.988	0.315	0.956	0.988	0.955	0.950
4:6	0.988	0.315	0.956	0.988	0.958	0.949
3:7	0.990	0.309	0.957	0.990	0.955	0.952
2:8	0.932	0.007	0.994	0.932	0.958	0.960
1:9	0.932	0.007	0.994	0.932	0.958	0.960

4.7. Comparisons with Existing Approaches

Table 8 shows a comparative analysis between existing and proposed approach for mobile devices. Our approach detects both spam as well as smishing messages whereas techniques by Yadav et al. (2011), Delany et al. (2012), Eshmawi et al. (2013) detect spam messages only. We have employed feature reweighting mechanism which is not used by Bottazzi et al. (2015), Delany et al. (2012). Although we have used various algorithms for spam and smishing detection but the maximum performance is achieved by neural network algorithm. Accuracy achieved by our scheme is 96% which is comparatively more than the accuracy achieved by other techniques.

Figure 7. Smishing detection result on various folds for neural network

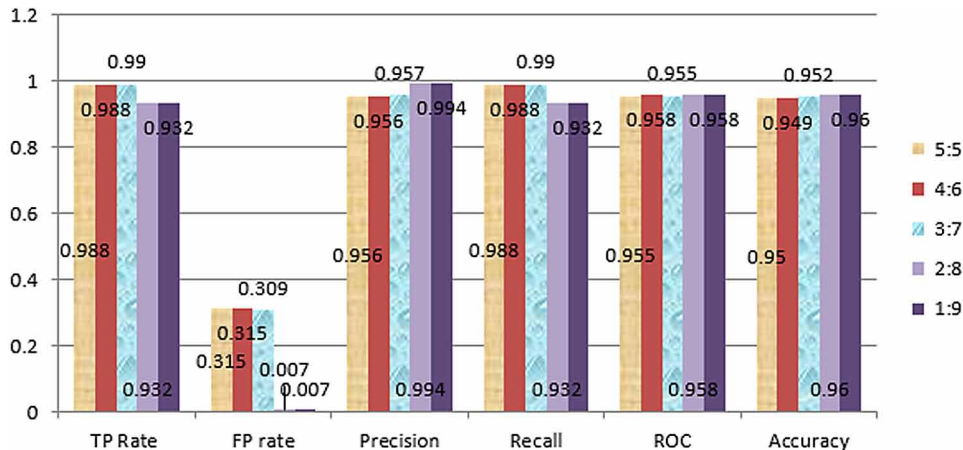


Figure 8. Importance of smishing feature set for smishing message classification

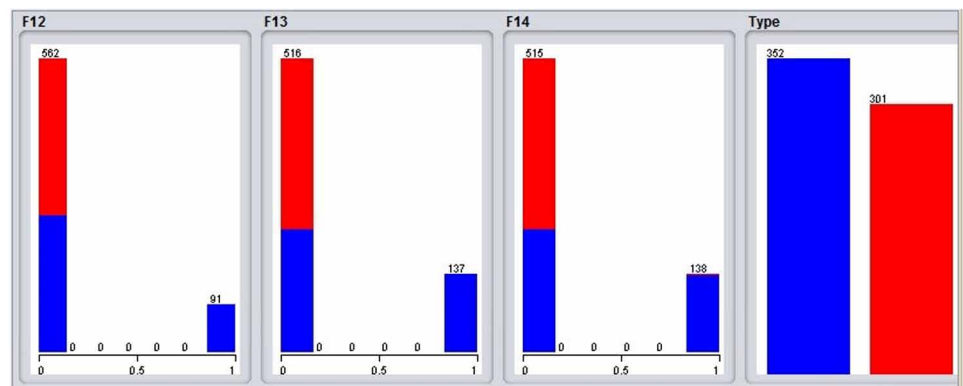


Table 6. Feature set for filtering spam messages

Feature Name	IG Value
Message length	0.20866
Presence of Mathematical Symbols	0.16135
Services Token	0.08021
Capitalized Words	0.05418
Presence of special characters	0.04318
Presence of Dots	0.2786
Dating Token	0.01531
Presence of Emotions	0.01505
Visual Morphemes	0.00578
Chat Token	0.00322

Table 7. Feature set for filtering phishing messages

Feature	IG Value
Awards Token	0.2213
Advertisements Token	0.1987
Presence of URL	0.1381

4.8. Limitations of Approach

Following are the few limitations of our proposed approach:

- The feature set of our phishing detection approach depends entirely on the text content of the message, and it can detect the smishing message only if it is written in English language.
- Nowadays, various malicious applications can steal the user's data, and send it to the phishers. Consequently, our approach is not capable enough to detect such malicious applications.

5. CONCLUSION AND FUTURE WORK

5.1. Conclusion

This paper presented a scheme to classify spam and smishing messages efficiently. The main contribution of this article is the filtering of smishing messages from spam messages, which has never been considered before. The existing approaches either detect only spam messages or smishing messages but our proposed approach is capable of detecting both spam and smishing messages. Moreover, this research work identified various new features for detection of spam and smishing messages. We have experimentally determined the importance of proposed features using the Information gain. Our scheme can also identify the zero-day smishing attack. Proposed method trains the neural network classifier and accurately filters spam and smishing messages. Our scheme is able to detect spam messages with an accuracy of 94.9% and smishing messages with an accuracy of 96%. The proposed work can be used as a mobile application, which blocks or shows a warning message whenever user receives the malicious messages.

Table 8. Comparison of proposed approach with existing approaches

Author	Accuracy	Spam Detection	Smishing Detection	Feature Reweighting	Classification Algorithm
Bottazzi et al. (2015)	89.2%	√	√	*	J48, SMO, BayesNet, IBK, SGD
Yadav et al. (2011)	86%	√	*	√	Bayesian, SVM
Delany et al. (2012)	93.31%	√	*	*	SVM
Eshmawi and Nai (2013)	88%	√	*	√	SVM, RF NB, CART
Idris et al. (2015)	83.20%	√	*	*	PSO
Etaiwi and Awajan (2017)	87.30	√	*	√	SVM, NB, RF, DT
Ma et al. (2016)	96%	√	*	*	SVM
Proposed Approach	96%	√	√	√	Neural Network

5.2. Future Work

Several approaches have been proposed in order to protect users from SMS phishing attacks, but still, the threat is not elevated and demands more attention towards the improvement of defense solutions. Future work that can be done in the proposed mechanism includes –

- Analysis of the URL present in the message, so as to determine if this URL redirects the user to a malicious login page or leads to the download of some malicious application.
- The performance of our scheme may be improved by integrating various more features, which effectively detect spam and smishing messages, in turn, improving the accuracy of the scheme.
- Size of the dataset can be increased in order to enhance the richness of the scheme.

REFERENCES

- Ali, S. S., & Maqsood, J. (2018). Net library for SMS spam detection using machine learning: A cross platform solution. In *Proceedings of the 2018 15th International Bhurban Conference on Applied Sciences and Technology (IBCAST)* (pp. 470-476). IEEE.
- Almeida, T. A., Hidalgo, J. M., & Yamakami, A. (2011). Contributions to the study of SMS spam filtering: new collection and results. In *Proceedings of the 11th ACM symposium on Document engineering* (pp. 259-262). Mountain View, CA: ACM. doi:10.1145/2034691.2034742
- Bottazzi, G., Casalicchio, E., Cingolani, D., Marturana, F., & Piu, M. (2015). MP-Shield: A Framework for Phishing Detection in Mobile Devices. In *Proceeding of IEEE International Conference on Computer and Information Technology; Ubiquitous Computing and Communications; Dependable, Autonomic and Secure Computing; Pervasive Intelligence and Computing*, Liverpool, UK (pp. 1977-1983). Academic Press. doi:10.1109/CIT/IUCC/DASC/PICOM.2015.293
- Chan, P. P., Yang, C., Yeung, D. S., & Ng, W. W. (2015). Spam filtering for short messages in adversarial environment. *Neurocomputing*, 155, 167–176. doi:10.1016/j.neucom.2014.12.034
- Choudhary, N., & Jain, A. K. (2017). Comparative Analysis of Mobile Phishing Detection and Prevention Approaches. In *Proceedings of International Conference on Information and Communication Technology for Intelligent Systems*, Ahmedabad, India (pp. 349-356). Springer.
- Delany, S. J., Buckley, M., & Greene, D. (2012). SMS spam filtering: Methods and data. *Expert Systems with Applications*, 39(10), 9899–9908. doi:10.1016/j.eswa.2012.02.053
- Diale, M., Celik, T., & Walt, C. V. (2019). Unsupervised feature learning for spam email filtering. *Computers & Electrical Engineering*, 74, 89–104. doi:10.1016/j.compeleceng.2019.01.004
- El-Alfy, E. S., & AlHasan, A. A. (2016). Spam filtering framework for multimodal mobile communication based on dendritic cell algorithm. *Future Generation Computer Systems*, 64, 98–107. doi:10.1016/j.future.2016.02.018
- Eshmawi, A., & Nai, S. (2013). Feature reduction for optimum sms spam filtering using domain knowledge. In *Proceedings of the International Conference on Security and Management (SAM)*. Academic Press.
- Etaiwi, W., & Awajan, A. (2017). The Effects of Features Selection Methods on Spam Review Detection Performance. In *Proceedings of the 2017 International Conference on New Trends in Computing Sciences (ICTCS)*, Amman (pp. 116-120). IEEE. doi:10.1109/ICTCS.2017.50
- Goel, D., & Jain, A. K. (2017). Smishing-Classifer: A Novel Framework for Detection of Smishing Attack in Mobile Environment. In *Proceedings of the International Conference on Next Generation Computing Technologies*, Dehradun (pp. 502-512). Springer.
- Goel, D., & Jain, A. K. (2018). Mobile phishing attacks and defence mechanisms: State of art and open research challenges. *Computers & Security*, 73, 519–544. doi:10.1016/j.cose.2017.12.006
- Hacked Hotel Phones Fueled Bank Phishing Scams*. (2015, Feb 4). Krebson Security. Retrieved from <https://krebsonsecurity.com/2015/02/hacked-hotel-phones-fueled-bank-phishing-scams/>
- Idris, I., Selama, A., Nguyen, N. T., Omatu, S., Krejcar, O., Kuca, K., & Penhaker, M. (2015). A combined negative selection algorithm–particle swarm optimization for an email spam detection system. *Engineering Applications of Artificial Intelligence*, 39, 33–44. doi:10.1016/j.engappai.2014.11.001
- Jain, A. K., & Gupta, B. B. (2018). Rule-Based Framework for Detection of Smishing Messages in Mobile Environment. *Procedia Computer Science*, 125, 617–623. doi:10.1016/j.procs.2017.12.079
- Jain, A. K., & Gupta, B. B. (2019). Feature Based Approach for Detection of Smishing Messages in the Mobile Environment. *Journal of Information Technology Research*, 12(2), 17–35. doi:10.4018/JITR.2019040102
- Joo, J. W., Moon, S. Y., Singh, S., & Park, J. H. (2017). S-Detector: An enhanced security model for detecting Smishing attack for mobile computing. *Telecommunication Systems*, 66(1), 29–38. doi:10.1007/s11235-016-0269-9

Ma, J., Zhang, Y., Liu, J., Yu, K., & Wang, X. (2016). Intelligent SMS Spam Filtering Using Topic Model. In *Proceedings of the 2016 International Conference on Intelligent Networking and Collaborative Systems (INCoS)* m Ostrawva, Czech Republic (pp. 380-383). IEEE. doi:10.1109/INCoS.2016.47

MOBILE THREAT REPORT. (2012, September). F Secure. Retrieved from <https://www.f-secure.com/documents/996508/1030743/Mobile+Threat+Report+Q3+2012.pdf>

Mr. Number - Caller ID & Spam Protection. (n.d.). Google Store. Retrieved from <https://play.google.com/store/apps/details?id=com.mrnumber.blocker&hl=en>

Silva, R. M., Almeida, T. A., & Yamakamia, A. (2017). MDLText: An efficient and lightweight text classifier. *Knowledge-Based Systems*, 118, 152–164. doi:10.1016/j.knosys.2016.11.018

SMiShing & Vishing News. (2017). Tumblr Numbercop. Retrieved from <https://numbercop.tumblr.com/post/119761088780/weekly-summary-54-517>

Spam Blocker App. (n.d.). Google Store. Retrieved from <https://play.google.com/store/apps/details?id=com.smsBlocker&hl=en>

Why do phishing attacks work better on mobile phones? (2011, January 20). Retrieved May 23, 2019, from <https://www.welivesecurity.com/2011/01/20/why-do-phishing-attacks-work-better-on-mobile-phones>

Yadav, K., Kumaraguru, P., Goyal, A., Gupta, A., & Naik, V. (2011). SMSAssassin: crowdsourcing driven mobile-based system for SMS spam filtering. In *Proceedings of the 12th Workshop on Mobile Computing Systems and Applications*, Phoenix, AZ (pp. 1-6). ACM. doi:10.1145/2184489.2184491

Ankit Kumar Jain is presently working as Assistant Professor in National Institute of Technology, Kurukshetra, India. He received a Master of technology from the Indian Institute of Information Technology Allahabad (IIIT) India and PhD degree from the National Institute of Technology, Kurukshetra, India. He published more than 25 research papers in international journals and conferences of high repute including IEEE, Elsevier, IGI Global, Springer, Taylor & Francis, Inderscience, etc. His general research interest is in the area of Information and cyber security, spam filtering, phishing detection, web security, mobile security, IoT security, online social network and machine learning.

Sumit Kumar Yadav is Assistant Director in Income Tax Department, Government of India before joining Income Tax Department he was Assistant Professor in Department of Computer Science and Engineering, Indira Gandhi Delhi Technical University for Women, Kashmere Gate, Government of Delhi. He has vast experience of 8 years in research and teaching. His research domain includes data science, data mining, sentiment analysis and opinion mining, fuzzy logic and information security. He has published more than 25 research papers in international conferences and journals of repute. He has also served as reviewer of many international journals.

Neelam Choudhary is completed Master of Technology in Cyber Security from National Institute of Technology, Kurukshetra. She received the Bachelor of Engineering in computer engineering. Her general research interests are in the areas of information and cybersecurity, spam detection, e-mail and web security.