Towards Filtering of SMS Spam Messages Using Machine Learning Based Technique

Neelam Choudhary(✉) and Ankit Kumar Jain

Computer Engineering Department, National Institute of Technology, Kurukshetra, Haryana, India
neelamchoudhary197@gmail.com, ankitjain@nitkkr.ac.in

Abstract. The popularity of mobile devices is increasing day by day as they provide a large variety of services by reducing the cost of services. Short Message Service (SMS) is considered one of the widely used communication service. However, this has led to an increase in mobile devices attacks like SMS Spam. In this paper, we present a novel approach that can detect and filter the spam messages using machine learning classification algorithms. We study the characteristics of spam messages in depth and then found ten features, which can efficiently filter SMS spam messages from ham messages. Our proposed approach achieved 96.5% true positive rate and 1.02% false positive rate for Random Forest classification algorithm.

Keywords: SMS spam · Mobile devices · Machine learning · Feature selection

1 Introduction

Short Message Service (SMS) is one of the popular communication services in which a message is sent electronically. The reduction in the cost of SMS services by telecom companies has led to the increased use of SMS. This rise attracted attackers which have resulted in SMS Spam problem. A spam message is generally any unwanted message that is sent to user's mobile phone. Spam messages include advertisements, free services, promotions, awards, etc. People are using SMS messages to communicate rather than emails because while sending SMS message there is no need of internet connection and it is simple and efficient [1].

The SMS Spam problem is increasing day by day with the increase in the use of text messaging. There are various security measures available to control SMS Spam problem but they are not so mature. Many android apps [2–4] are also on play store to block spam messages but people are not aware of these apps due to lack of knowledge. Other than apps the filtering techniques available mainly focuses on email spam as email spam is one of the oldest problem [5] but with the popularity of mobile devices, SMS spam is the one of the major issue these days.

SMS is one of the cheapest ways to communicate and can be considered as the simplest way to perform phishing attacks as mobile devices contain sensitive and personal information like card details, username, password, etc. [6–8]. Attackers are finding different ways to steal this information from mobile devices and SMS is one of the easiest ways. Smishing i.e. SMS based Phishing is more popular these days in which

D. Singh et al. (Eds.): ICAICR 2017, CCIS 712, pp. 18–30, 2017.
DOI: 10.1007/978-981-10-5780-9_2

user sends malicious link via SMS and asks user to visit that link and steals sensitive information from user's mobile device. There are various detection approaches available for detecting mobile phishing like QR code, machine learning based, biometric based, matrix code reader based, knowledge based and authentication based [9].

SMS spammers can purchase any mobile number with any area code to send spam messages so that it becomes difficult to identify the attacker. US tatango learning center provided the list of top 25 SMS Spam area codes used by spammers [10]. Moreover, top 5 SMS Spam messages [11] are shown in Fig. 1.

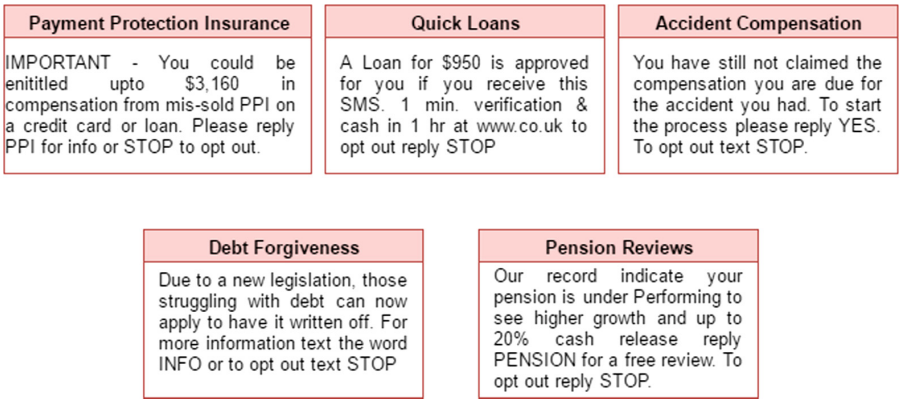

Fig. 1. Top 5 popular SMS spam messages

In 2014, a report was released by Cloud mark in which they stated that how spammers used Twilio to send 385,000 spam messages [12]. National Fraud Intelligence Bureau (NFIB) published a media report about the latest scams which was analyzed by action fraud in 2016 [13]. Spammers are targeting bank customers these days by sending spam messages for asking their bank account details, ATM pin number, password, etc. and the customer thinks that the message is coming from the bank and he/she may give all the details to the spammer. A report was published by ACMA that how bank customers are becoming the victim of SMS Spam attacks [14].

In our proposed approach main aim is to filter the spam and ham SMS using machine learning algorithms. We have used a feature set of 10 features for classification. These features can differentiate a spam SMS from ham SMS. Machine learning techniques were effective in email spam filtering as it helps in preventing zero-day attacks and provides the high level of security. The Same approach is being used for mobile devices in order to prevent from SMS Spam problem but in the case of SMS Spam features will be different from email spam as the size of the text message is small and the user uses less formal language for text messages. And text message is simple without any graphic content and attachments.

The rest of the paper is organized as follows: First of all, Sect. 2 discusses the related work, Sect. 3 presents our proposed model including features that we have

selected for our experiment. Section 4 presents the experimental detail including the dataset collection, results of our experiment. Finally, Sect. 5 presents conclusion and future work.

2 Related Work

A number of SMS Spam messages detection techniques are available these days like android apps to block spam messages, filtering spam messages using classification algorithms, etc. In this section, we will review the SMS Spam detection techniques by filtering spam messages based on feature selection using machine learning techniques.

El-Alfy and AlHasan [15] have proposed a model for filtering text messages for both email and SMS. They have analyzed different methods in order to finalize a feature set such that complexity can be reduced. They have used two classification algorithms i.e. Support Vector Machine (SVM) and Naïve Bayes and 11 features i.e. URLs, likely spam words, emotion symbols, special characters, gappy words, message metadata, JavaScript code, function words, recipient address, subject field and spam domain. They have evaluated their proposed model on five email and SMS datasets.

Jialin et al. [16] have proposed a message topic model (MTM) for filtering Spam messages. Messages Topic Model (MTM) considers symbol terms, background terms and topic terms to represent spam messages and it is based on the probability guess of latent semantic analysis. They have used k-means algorithm to remove the sparse problem by training SMS spam messages into random irregular classes and then aggregating all SMS spam messages as a single file such that to capture word co-occurrence patterns.

Chan et al. [17] have presented two methods for SMS Spam filtering i.e. feature reweighting method and good word attack. Both methods focus on the length of the message along with considering the weight of message. Good word attack focuses on deceiving the output of classifier by using least number of characters while for feature reweighting method they have introduced a new rescaling function for rescaling the weights. They have evaluated the experiment on two datasets i.e. SMS and comment.

Delany et al. [18] discuss different approaches available for SMS Spam filtering and the problems associated with the dataset collection. They have analyzed a large dataset of SMS spam and used ten clusters i.e. ringtones, competitions, dating, prizes, services, finance, claims chat, voicemail and others.

Xu et al. [19] have detected SMS Spam messages using content-less features. They have used 2 classification algorithms i.e. SVM and k-nearest neighbor (KNN) and feature set consisting of 3 features i.e. static, temporal and network for their experiment. They found that by combining temporal and network features SMS Spam messages can be detected more accurately and with good performance. Moreover, they also found the ways filter SMS Spam messages by using features that contain graph-topology and temporal information thus excluding the content of the message.

Nuruzzaman et al. [20] used Text Classification techniques on independent mobile phones to evaluate their performance of filtering SMS spam. The training, filtering, and updating processes were performed on an independent mobile phone. Their proposed model was able to filter SMS spam with some good accuracy, less storage

consumption, and good enough processing time without using a large amount of dataset or any support from computer.

Uysal et al. [21] have proposed a method for SMS Spam filtering by using two feature selection based approaches i.e. chi-square metrics and information gain in order to select discriminative features. They have also developed a real time mobile application for SMS Spam filtering based on android application. They have used two different Bayesian based classification algorithms i.e. probabilistic and binary. According to the authors, their proposed system is highly accurate in detecting both spam and legitimate messages.

Yadav et al. [22] developed a model SMSAssassin for SMS Spam filtering. They have used a feature set of 20 lightweight features and two machine learning algorithms i.e. Support Vector Machine (SVM) and Bayesian learning. They have collected a dataset of 2000 messages from users within the time span of two months. In their proposed model whenever the user gets some message over his phone, then SMSAssassin initially captures that message without user's knowledge, fetches feature values, and sends theses values to the server for classification. If the messages are reported as spam, then the user will not be able to see that message and it will be redirected to spam folder.

Hidalgo et al. [23] have analyzed that how Bayesian filtering technique can be used to detect SMS Spam. They have built two datasets one in English and another in Spanish. Their analysis shows that Bayesian filtering techniques that were earlier used in detecting email spam can also be used to block SMS Spam.

3 Proposed Methodology

The main objective of our approach is to classify the spam SMS messages as soon as it received on the mobile phone, regardless of newly created spam message (zero-hour attack). In this, we firstly collected dataset and finalized the features for our experiment. After finalizing features, we extracted the features from the messages (ham and spam) to create a feature vector. These feature vectors are used for training and testing purposes. Our proposed system takes the decision based on ten features. Figure 2 shows the system architecture of our proposed approach. In the training phase, a binary classifier is generated by applying the feature vectors of spam and ham messages. In the testing phase, the classifier determines whether a new message is a spam or not. At the end we get classification results for different machine learning algorithms and performance is evaluated for each machine learning algorithm such that we can get the best algorithm for our proposed approach.

3.1 Feature Selection

Feature selection is a very important task for the SMS Spam filtering. Selected features should be correlated to the message type such that accuracy for detection of spam message can be increased. There is a length limit for SMS message and it contains only text (i.e. no file attachments, graphics, etc.) while in the email, there is no text limit and

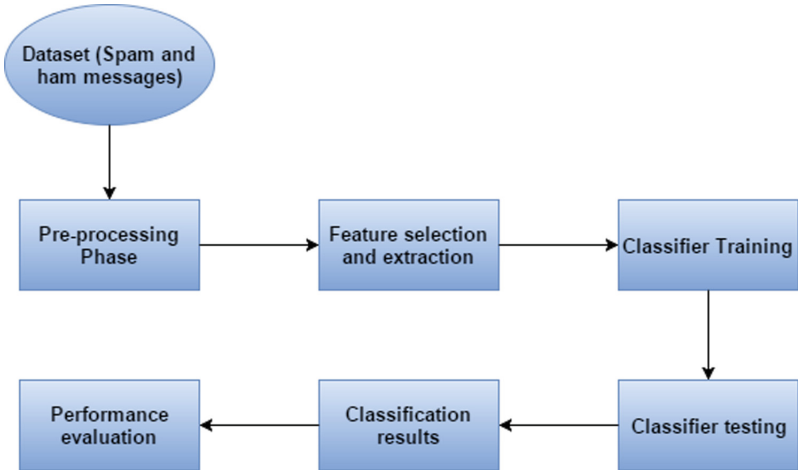

Fig. 2. System architecture

it contains attachments, graphics, etc. SMS message is usually of two types i.e. ham (legitimate) message and spam message. Spam and ham messages can be differentiated using various features. Identification of good feature that can efficiently filter spam SMS messages is a challenging task. Moreover, we study the characteristics of spam messages in depth and find some features, which are useful in the efficient detection of spam SMS. The features that we have extracted and evaluated for our proposed approach are summarized as follows:-

- *Presence of Mathematical Symbols* - Spammers usually uses mathematical symbols for creating spam messages. For example, symbol + can be used for free services messages. Mathematical symbols that we have considered in our experiment are +, −, < , > , / and ^. The first feature is defined as S_1 which could be 1 if any mathematical symbol is present in the message.

$$S_1 = \begin{cases} 1 & \textit{Mathematical symbol} \\ 0 & \textit{No mathematical symbol} \end{cases} \tag{1}$$

- *Presence of URLs* - We consider the presence of URLs as a feature since harmful spam SMS contains URLs and asks the user to visit those URLs to provide their personal details, debit/credit card information, password or to download some file (file containing the virus). The second feature S_2 which could be 1 if any URL (http or www) is present in the message.

$$S_2 = \begin{cases} 1 & \textit{URL is present} \\ 0 & \textit{No URL} \end{cases} \tag{2}$$

- *Presence of Dots* - The presence of dots seems to be good indicator for legitimate messages because people use dots while chatting. Moreover, People often use dots to separate the sentence, or words so that it becomes easy for the receiver to read the

message. We define the presence of dots as feature S_3, which will be 1 if the message contains dots.

$$S_3 = \begin{cases} 1 & Dot\ is\ present \\ 0 & No\ Dot \end{cases} \tag{3}$$

- *Presence of special symbols* - The presence of special symbols usually refers to spam messages because spammers use special symbols for various reasons. E.g. special symbol "$" is being used to represent money in the dollar in fake award messages, similarly symbol "!" is used to the special attention of user like CONGRATULATIONS! WINNER!, etc. Special symbols that we have used in our approach are !, *, &, # and ∼. The fourth feature is defined as S_4 which would be 1 if any special symbol is present.

$$S_4 = \begin{cases} 1 & Special\ symbol \\ 0 & No\ Special\ symbol \end{cases} \tag{4}$$

- *Presence of emotions* - The presence of emotion symbols seems to be a good indicator for legitimate messages because a person usually uses emotions while chatting. For example emotion :) is used for happy face, emotion :(is used for sad face, emotion -_- is used for angry face, etc. Emotion symbols that we have considered for our experiment are :), :(, -_-, :p, :v, :*, :o, B-) and :'(. We define the presence of emotions as feature S_5.

$$S_5 = \begin{cases} 1 & Emotions \\ 0 & No\ Emotions \end{cases} \tag{5}$$

- *Lowercased words* - Checks if the message contains lowercased words or not as all lowercased words in a message can be used to seek user's attention. The presence of lowercased words is given by feature S_6 as:-

$$S_6 = \begin{cases} 1 & Lowercased\ words \\ 0 & No\ Lowercased\ words \end{cases} \tag{6}$$

- *Uppercased words* - We consider the presence of uppercased words as a feature as spammers usually use uppercased words to seek user's attention. For example, WON, PRIZE, FREE, RINGTONE, ATTENTION, etc. The seventh feature is S_7 given by rule:-

$$S_7 = \begin{cases} 1 & Uppercased\ words \\ 0 & No\ Uppercased\ words \end{cases} \tag{7}$$

- *Presences of Mobile Number* - We consider the presence of mobile number as a feature in order to identify spam messages because spammers usually give mobile number in a message. They ask the users to dial on the given number and when user dials on the given number, attacker on the other side ask for user's personal details,

bank details, etc. For example, "you have won a $2,000 price! To claim, call 09050000301". We define the presence of mobile number as feature S_8.

$$S_8 = \begin{cases} 1 & Mobile\ Number \\ 0 & No\ Mobile\ Number \end{cases} \tag{8}$$

- *Keyword specific* - The presence of suspicious keywords like send, ringtone, free, accident, awards, dating, won, service, lottery, mins, video, visit, delivery, cash, Congrats, Please, claim, Prize, delivery, etc. are considered as spam keywords because these keywords are generally used to attract users in spam messages. We define ninth feature as S_9 which will be 1 if message contains spam keywords otherwise it will be 0.

$$S_9 = \begin{cases} 1 & Presence\ of\ spam\ keywords \\ 0 & No\ spam\ keywords \end{cases} \tag{9}$$

- *Message Length* - It includes the total length of the message including space, symbols, special characters, smileys, etc. The text limit of SMS messages is 160 characters only. We define tenth feature as S_{10} which counts the total length of each message.

Table 1 shows that how each feature value is extracted from ham and spam messages.

Table 1. SMS message feature value for ham and spam messages

Feature type	Have you finished work yet?) (Ham message)	CONGRATULATIONS! Nokia 3650 video camera phone is your Call 09066382422 Calls cost 150 ppm Ave call 3 min vary from mobiles 16 + Close 300603 post BCM4284 Ldn WC1N3XX (Spam message)
Presence of mathematical symbols	0	1
Presence of URLs	0	0
Presence of dots	0	0
Presence of special symbols	0	0
Presence of emotions	1	0
Lowercased words	1	1
Uppercased words	0	1
Presence of mobile number	0	1
Keyword specific	0	1
Message length	30	157

4 System Design, Implementation and Results

Our aim is to construct a new classification model which can filter spam SMS efficiently. This section presents implementation details, dataset, the brief summary of machine learning algorithms and performance evaluation measures to judge the performance of our proposed approach. The detection of spam SMS is a binary classification problem where various features are used to train the classifier.

4.1 Dataset Collection

The SMS dataset that we have used for our experiment contains 2608 messages out of which 2408 collected from SMS Spam Corpus publically available [24] and 200 collected manually which consists of 25 spam messages and 175 ham messages. The SMS Spam Corpus v.0.1 consists of following two sets of messages:

- SMS Spam Corpus v.0.1 Small - It consists of 1002 ham messages and 82 spam messages. This corpus is useful and has been used in the research [23, 25].
- SMS Spam Corpus v.0.1 Big - It consists of 1002 ham messages and 322 spam messages. This corpus is useful and has been used by the researchers in their research work [26].

4.2 Machine Learning Algorithms

After extracting features, classification accuracy is being tested on WEKA tool using five machine learning algorithms: Naïve Bayes, Logistic Regression, J48, Decision Table, and Random Forest. These algorithms [27] are described in brief as:-

- *Naïve Bayes* - Naïve Bayes classification algorithm is based on Bayes theorem [28]. In Naïve Bayes assumptions between predictors is independent. It is simple and easy to use so it can be used for large datasets.
- *Logistic Regression* - In this machine learning algorithm the dependent variable is categorical and measures the relationship between the independent variable and categorical dependent variable using the logistic function.
- *J48* - J48 uses a training data o already classified samples and it is a java implementation of the C4.5 classification algorithm. This algorithm basically constructs a decision tree where is each feature is represented by the node.
- *Decision Table* - Decision table is a machine learning algorithm that represents a set of rules and in this result is good only for some continuous features.
- *Random Forest* - Random Forest algorithm is best machine algorithm for a large number of datasets. It basically constructs a set of decision trees at training phase and then each tree operates on randomly chosen attributes.

4.3 Evaluation Metrics

In order to evaluate the effectiveness of our proposed approach, we will consider eight possible outcomes i.e. true positive rate, false positive rate, true negative rate, false negative rate, f1 score, accuracy, precision, and recall. These are the standard metrics to judge any spam detection system. These evaluation metrics are described in brief as follows:-

- True Positive Rate (TP) - It denotes the percentage of spam messages that were accurately classified by the machine learning algorithm. If we denote spam messages as S and spam messages that were accurately categorized as P, then

$$\text{TP} = \frac{P}{S} \tag{10}$$

- True Negative Rate (TN) - It denotes the percentage of ham messages that were accurately categorized as ham messages by the machine learning algorithm. If we denote ham message as H and ham messages that were accurately categorized as ham by Q, then

$$\text{TN} = \frac{Q}{H} \tag{11}$$

- False Positive Rate (FP) - It denotes the percentage of ham messages that were wrongly categorized as spam by the machine learning algorithm. If we denote ham messages as H and ham messages that were wrongly classified as spam by R, then

$$\text{FP} = \frac{R}{H} \tag{12}$$

- False Negative Rate (FN) - It denotes the percentage of spam messages that were incorrectly classified as ham message by the machine learning algorithm. If we denote spam messages as S and number of SMS spam messages that were incorrectly classified as ham by T, then

$$\text{FN} = \frac{T}{S} \tag{13}$$

- Precision - It denotes the percentage of messages that were spam and actually classified as spam by the classification algorithm. It shows the exact correctness. It is given as:-

$$\text{Precision} = \frac{TP}{TP + FP} \tag{14}$$

- Recall - It denotes the percentage of messages that were spam and classified as spam. It shows the completeness. It is given as:-

$$\text{Recall} = \frac{TP}{TP + FN} \tag{15}$$

- F-measure - It is defined as the harmonic mean of Precision and Recall. It is given as:-

$$\text{F-measure} = \frac{2 * Precision * Recall}{Precision + Recall} \tag{16}$$

- Receiver Operating Characteristics (ROC) area - In this an area is plotted between True Positive Rate and False Positive Rate for different threshold values.

4.4 Results and Discussions

Various experiments are performed to evaluate the performance of our proposed SMS Spam detection system. Initially we have selected features on the basis of behavior of spam and ham messages and then extracted these features from the dataset to get the feature vector. After extracting features from the dataset, various classification algorithms like Naïve Bayes, Logistic Regression, J48, Decision Table and Support Vector Machine (SVM) is being applied to get the performance metrics. We have used WEKA tool for classification and have used cross validation of 10-fold in which 90% of data is used for training purpose and remaining 10% data for testing purpose. Table 2 presents the results of our proposed approach on various classification algorithms i.e. Naïve Bayes, Logistic Regression, J48, Decision Table and Support Vector Machine (SVM) algorithm. Receiver Operating Characteristics (ROC) curve for our proposed approach is shown in Fig. 3.

Table 2. Results of proposed approach on various machine learning algorithms

Algorithm	TP rate	FP rate	Precision	ROC area	F-measure	Recall
Naïve Bayes	0.941	0.077	0.948	0.985	0.943	0.941
Logistic regression	0.959	0.135	0.958	0.989	0.958	0.959
J48	0.961	0.143	0.960	0.953	0.960	0.961
Decision table	0.960	0.133	0.959	0.984	0.960	0.960
Random forest	0.965	0.102	0.965	0.983	0.965	0.965

After comparing the performance metrics for various machine learning algorithms we have analyzed that best results were achieved by Random Forest Classification Algorithm. The Random Forest machine learning algorithm achieved the best classification results with the high accuracy. Moreover, we achieved 96.5% true positive rate and 1.02% false positive rate with Random Forest machine learning algorithm. Random Forest classification algorithm basically constructs a set of decision trees at training phase and then each tree operates on randomly chosen attributes.

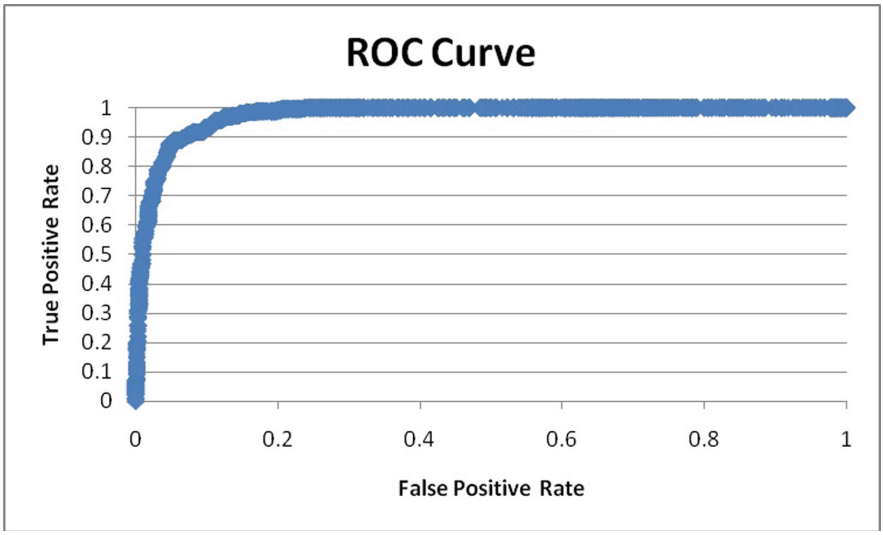

Fig. 3. Random forest ROC area

After getting the results a comparative study is done such that to compare our results with the previous results. Table 3 shows the comparative study of our proposed approach with the previous approach available in the literature.

Table 3. Comparative study of our approach with previous approach

Paper	Machine learning algorithm	TP rate
Content based spam detection [15]	SVM	95.5
Proposed approach	Random forest	96.5

5 Conclusion and Future Work

The SMS Spam problem is increasing nowadays with the increase in the use of text messaging. SMS Spam filtering is the big challenge these days. In this paper, we propose a technique for SMS Spam filtering based on 10 feature using five machine learning algorithms namely Naïve Bayes, Logistic Regression, J48, Decision Table and Random Forest. The dataset that we have used in our work consists of 2608 messages out of which 2408 messages were collected from the SMS Spam Corpus v.0.1 publically available and 200 messages collected manually. Out of all classification algorithms, Random Forest Classification Algorithm gives best results with 96.1% true positive rate.

In our future work, we will try to add more features as best spam features help in detecting spam messages more accurately. We will also try to collect more and more datasets from the real world.

References

1. Mobile Commons Blog. https://www.mobilecommons.com/blog/2016/01/how-text-messaging-will-change-for-the-better-in-2016/
2. SMS Blocker Award. https://play.google.com/store/apps/details?id=com.smsBlocker&hl=en
3. TextBlocker. https://play.google.com/store/apps/details?id=com.thesimpleandroidguy.app.messageclient&hl=en
4. Androidapp. https://play.google.com/store/apps/details?id=com.mrnumber.blocker&hl=en
5. Puniškis, D., Laurutis, R., Dirmeikis, R.: An artificial neural nets for spam e-mail recognition. Elektronika ir Elektrotechnika **69**, 73–76 (2006)
6. Jain, A.K., Gupta, B.B.: Phishing detection: analysis of visual similarity based approaches. Secur. Commun. Netw. **2017** (2017). Article ID 5421046. doi:10.1155/2017/5421046
7. Gupta, B.B., Tewari, A., Jain, A.K., Agrawal, D.P.: Fighting against phishing attacks: state of the art and future challenges. Neural Comput. Appl. 1–26 (2016). doi:10.1007/s00521-016-2275-y
8. Jain, A.K., Gupta, B.B.: A novel approach to protect against phishing attacks at client side using auto-updated white-list. EURASIP J. Inf. Secur. 1–11 (2016). doi:10.1186/s13635-016-0034-3
9. Choudhary, N., Jain, A.K.: Comparative Analysis of Mobile Phishing Detection and Prevention Approaches (Accepted)
10. Tatango Learning Center. https://www.tatango.com/blog/top-25-sms-spam-area-codes/
11. Adaptive Mobile Press Releases. https://www.adaptivemobile.com/press-centre/press-releases/five-top-spam-texts-for-2012-revealed-in-adaptivemobiles-ongoing-threat-ana
12. Cloudmark Report. https://www.tatango.com/blog/sms-spammers-exploit-twilio-send-385000-spam-text-messages/
13. Action Fraud News. http://www.actionfraud.police.uk/news/latest-scams-to-watch-out-for-apr16
14. ACMA Cybersecurity Blog. http://www.acma.gov.au/theACMA/engage-blogs/engage-blogs/Cybersecurity/Banks-targetted-by-SMS-phishing-scam
15. El-Alfy, E.S.M., AlHasan, A.A.: Spam filtering framework for multimodal mobile communication based on dendritic cell algorithm. Future Gen. Comput. Syst. **64**, 98–107 (2016). doi:10.1016/j.future.2016.02.018
16. Jialin, M., Zhang, Y., Liu, J., Yu, K., Wang, X.: Intelligent SMS spam filtering using topic model. In: International Conference on Intelligent Networking and Collaborative Systems (INCoS), pp. 380–383. IEEE (2016). doi:10.1109/INCoS.2016.47
17. Chan, P.P.K., Yang, C., Yeung, D.S., Ng, W.W.Y.: Spam filtering for short messages in adversarial environment. Neurocomputing **155**, 167–176 (2015). doi:10.1016/j.neucom.2014.12.034
18. Delany, S.J., Buckley, M., Greene, D.: SMS spam filtering: methods and data. Expert Syst. Appl. **39**, 9899–9908 (2012). doi:10.1016/j.eswa.2012.02.053
19. Xu, Q., Xiang, E.W., Yang, Q., Du, J., Zhong, J.: SMS spam detection using non-content features. IEEE Intell. Syst. **27**(6), 44–51 (2012)
20. Nuruzzaman, M.T., Lee, C., Abdullah, M., Choi, D.: Simple SMS spam filtering on independent mobile phone. Secur. Commun. Netw. 1209–1220 (2012). doi:10.1002/sec.577
21. Uysal, A.K., Gunal, S., Ergin, S., Gunal, E.S.: A novel framework for SMS spam filtering. In: International Symposium on Innovations in Intelligent Systems and Applications (INISTA), pp. 1–4. IEEE (2012). doi:10.1109/INISTA.2012.6246947

22. Yadav, K., Kumaraguru, P., Goyal, A., Gupta, A., Naik, V.: SMSAssassin: crowdsourcing driven mobile-based system for SMS spam filtering. In: 12th Workshop on Mobile Computing Systems and Applications, pp. 1–6. ACM (2011). doi:10.1145/2184489.2184491
23. Hidalgo, J.M.G., Bringas, G.C., Sánz, E.P., García, F.C.: Content based SMS spam filtering. In: ACM Symposium on Document Engineering, pp. 107–114. ACM (2006). doi:10.1145/1166160.1166191
24. SMS Spam Corpus. http://www.esp.uem.es/jmgomez/smsspamcorpus
25. Cormack, G.V., Hidalgo, J.M.G., Sánz, E.P.: Feature engineering for mobile (SMS) spam filtering. In: 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 871–872. ACM (2007). doi:10.1145/1277741.1277951
26. Cormack, G.V., Hidalgo, J.M.G., Sánz, E.P.: Spam filtering for short messages. In: 16th ACM Conference on Conference on Information and Knowledge Management, pp. 313–320. ACM (2007). doi:10.1145/1321440.1321486
27. Ayodele, T.O.: Types of machine learning algorithms. In: New Advances in Machine Learning. INTECH Publisher (2010)
28. Machine Algorithm Algorithms. http://machinelearningmastery.com/naive-bayes-for-machine-learning

http://www.springer.com/978-981-10-5779-3

Advanced Informatics for Computing Research
First International Conference, ICAICR 2017, Jalandhar, India, March 17–18, 2017, Revised Selected Papers
Singh, D.; Raman, B.; Luhach, A.K.; Lingras, P. (Eds.)

2017, XIII, 370 p. 172 illus., Softcover
ISBN: 978-981-10-5779-3