Cybercrime: Suspicious Viber Messages Detection Model

Dr. Hanaa Mohsin Ahmed, Noor Basim Bethoon

Abstract—The rapid advancements in mobile phone systems and programs that provide free instant messaging (IM), short message service (SMS), and the accommodation of conveying millions of messages with virtually no delay and zero cost through Wi-Fi or 3G(third generation) has led to the increasing popularity of IM and SMS.The requisite for these advancements is an automatic classification system for expeditious relegation of the received messages to detect the suspicious message. This work proposes the use of a detection model in which social media messages are classified as predefined classes labeled"suspicious" and "not suspicious."The proposed system attemptsto solve this problem through three classifiers:Level Based Feature Content(LBFC) classifier, Naive Bayesian(NB) classifier, and IterativeDichotomiser 3(ID3) classifier.This system works offline, after collecting the messages online, saving them and then inputting them into the proposed system.In the LBFC classifier, the content feature is divided into four levels to detect suspicious classes.The second classifier presents an NB classifier and the the third is ID3 classifier is capable of identifying Viber messages as suspicious or non-suspicious, predicated on the content of these messages.From the experimental work, good results are achieved from the first classifer t features (Accuracy=0.882143%) and second testing using term frequency (TF)-based NB classifier (Accuracy=0.942857%), while results are achieved using TF-based ID3 (Accuracy= 0.957143%).

Index Terms— feature selection, ID3 Classification, level of features; Naive Bayesian; non-suspicious, suspicious, Viber instant message

1 INTRODUCTION

The amount of information stored in modern databases makes manual analysis intractable, using many tools such as data mining tools and machine learning tools. Data mining provides tools to reveal previously unknown information in large databases [1].

Lately, there is an urgent need to cope with huge textual data that became easier to gain and obtain through massive storage devices and high-speed Internet connections. Textual data can be more useful when extracting and employing them for important tasks, such as classifications and search engines that enable users to use a large text of data to acquire important information they need. Thus, fields that handle textual data attracted researchers to study and develop methods to use textual data [2].

Messaging is a general form of asynchronous near real-time communication between two or more users using different peripherals in their respective networks. Different types of messaging techniques exist in various communication applications, ranging from email to SMS. Exchanged messages are mostly text-based with the possibilityof carrying audio and video data. Instant messengers are one of the most popular types of interactive applications that communicate packet data directly between known (and unknown) people.

The procedure is commonlyreferred to as chatting. However, chatting covers a more general concept than instant messaging (IM).

Chatting often refers to communication between known and unknown people in a multiparty platform while IM is mainly performed between known people from their contact lists [3].

Viber is a communication application for the mobile phone. The current version works on both the Android and iPhone. Generally, users can call and send text messages or photos to their current Global Position System (GPS) location [4].

This paper is organized as follows. Section I provides the introduction. Section 2 summarizes the related work using various methods for classification. Section 3 gives general theoretical descriptionsabout text mining, text classification, text representation, preprocessing text, and feature selection. Section 4 describes the Viber application. Section 5 presents detailed steps of the proposed detection model and performance. Section 6 provides the experimentalresults. Section 7 concludes this paper.

2 RELATED WORK

Unlike the growing number of papers on email spam classifiers, there are still few studies on short message service (SMS) and IM spam filtering. However, the amount of junk SMS, IM, and email increases every day. Many people this domain (suspicious detection) on email massage, SMS, and IM. Below arethe relevant works related to this topic:

1. Almeida etal. (2011) displayed a new real, public, and non-encoded SMS spam collection that is the biggest one so far. In addition, they compare the performance achieved by several established machine learning methods. They found the support vector machine gives good results compared to other classifiers and, hence, it can be used as a good baseline for further comparison [5].
2. Patel and Bhatnagar (2011) propose a model that first uses the entropy term weighting scheme and then PCA is used for the re-parameterization. Next, they used the Artificial Neural Network for classification. Their mod-

el may be successful in efficiently classifying SMS text document implementation of this model will be further in future [6].

3. Shirani-Mehr (2012) used a real SMS database (UCIrvineMachine Learning Repository), beginning with preprocessing and feature extraction. After that, he applied different machine learning algorithms to the SMS spam classification problem and found the best result was in the use of Support Vector Machine(SVM) as the learning algorithm, which yielded overall accuracy of 97.64% [7].
4. Ahmed etal. (2014) proposed a hybrid system of SMS classification to detect spamor ham using aNaive Bayesian (NB) classifier and an Apriori algorithm. This system led to improved effective accuracy of 98.7% from the traditional NB approach of 97.4% while experimenting on the UCIrvine Data Repository [8].
5. Shahi and Yadav (2014) used the SVM-based and Naïve Bayes classification techniques to classify the Nepali SMS as spam and non-spam. The empirical analysis for different samples of text has been done to measure the accuracy of the classification methodologies used in this study. It is found to be 87.15% accurate in SVM and 92.74% accurate in the case of Naïve Bayes [9].

3. TEXT MINING

Text mining is an area of computer science thatlooks at strong links with Natural Language Processing, Data Mining, Machine Learning, Information Retrieval, and Knowledge Management. Text mining usually seeks to extract valuable information from unstructured textual data by identifying and exploringthose interesting patterns [10]. Commonly used methods for text classification are [11]: Decision Trees; Pattern (Rule)-based Classifiers; SVM classifiers; Neural Network Classifiers; Bayesian (Generative) Classifiers; and Other Classifiers.

3.1 Text Classification

Text classification refers to the problem of automatically assigning Zero, One, or more of a predefined set of labels to a given segment of free text. "The classification scheme will not be explicitly laid out according to human-designed rules – machine learning algorithms are instead designed to learn it from training examples. In other words, they extract the implicit knowledge contained in texts plus their labels" [12].

3.2 Text Representation

Free text information is unstructured. "Text documents vary in length and use different sets of words." As such, common classification algorithms cannot readily interpret them. Therefore, preprocessing procedures that map the free text into a structured representation are necessary before applying classification algorithms. "The most common way to represent text is based on the bag of words approach." It maps an input text (e.g., a document) to a vector of term weights, where terms can be words or phrases [13].

3.3 ThePreprocessing of Text

This process objective is to decrease the document spaces state and clean the text message from any additions. It consists of the following three steps [14]:

1. Tokenization: This "[i]s the process of breaking a stream of text into words, phrases, symbols, or other meaningful elements called tokens. The aim of the tokenization is the exploration of the words in a sentence. The list of tokens becomes input for further processing such as parsing or text mining."
2. Removing Stop Words: Stop words, such as "the," "a," "and," etc., frequently occur, so these insignificant words need to be removed. This process also reduces the text data and improves the system's performance. Every text document deals with these words, which are not necessary for text mining applications
3. Stemming Word: This technique transforms words into their stems, which generalizes the texts for similarity analysis. For this step, a process of conflating tokens reverts words to their root form. For example, both "computer" and "computers" are normalized to "compute"; "product," "produce," and "produced" to "produce"; "connection" to "connect"; and "computing" to "compute."

3.4 Feature Selection

Commonly used to feature selection, an algorithm of feature selection was used in this work: term frequency (TF), which aimsto reducethe dimensionality of the feature [15].

4. VIBER APPLICATION

Every person in the world can joinViber. The application has more than 664 million users worldwide, who can send messages and makehigh-definition (HD)quality phone and video calls over Wi-Fi or 3G—all for free [16]. A user can create group messages with up to 200 friends, share photos, video, stickers, links, and more. There is no need to register because users' phone numbersare their ID and Viber syncs with their mobile contact list automatically [17].

The properties of Viber are as follow [18]:

1. Message friends (texts can be up to 7,000 characters long).
2. Make free phone and video calls with HD sound quality.
3. Share "photos, videos, voice messages, locations, contact info, rich links, stickers and emoticons."
4. Attach files by "sending messages with DOC, PPT, PDF and utmost other files via Viber."
5. Delete a message "from all conversation members, even after it's been sent."
6. Support for the "Viber desktop application on Windows, Mac, Linux and Windows 8."

5. PROPOSED SYSTEM

The system focuses on the detection of suspiciousViber messages, based on anumber of steps. Step one is collecting the message (input), saving it as a text file, and reading the text. Step two is preprocessing, which aims to clean the text, reduce the document space, and eliminate redundancy based on the proposed system. This step uses data mining algorithms ((first classifier: Level Based Feature Content [LBFC] consisting of

four levels), (secondclassifier (NB) classifier), third classifier (IterativeDichotomiser 3 [ID3] classifier)) and uses feature selection methods to select the best subset of features usingTF. Figure1 is the block diagram of these steps.

The functionality of the suspicious detection system is based on the following phases, as shown in Figures 2, 3, and 4.

1. Prepare and preprocess suspicious dataset to be used in training and classification.
2. The algorithms of the suspicious detection are the LBFC classifier, NB classifier, and ID3 classifier, which train and classify the suspicious dataset.
   a. The first classifier LBFC uses a four level classifier with all of the features of the suspicious dataset applied as depicted in Figure 3.
   b. The second classifier uses TF feature selection with NB classifier. NB classifier is applied for a subset of features as depicted in Figure 4 that are selected using TF.
   c. The third classifier uses the algorithm TF feature selection with ID3 classifier. ID3 classifier is applied on a subset of features as depicted in Figure 5, which are selected using TF.

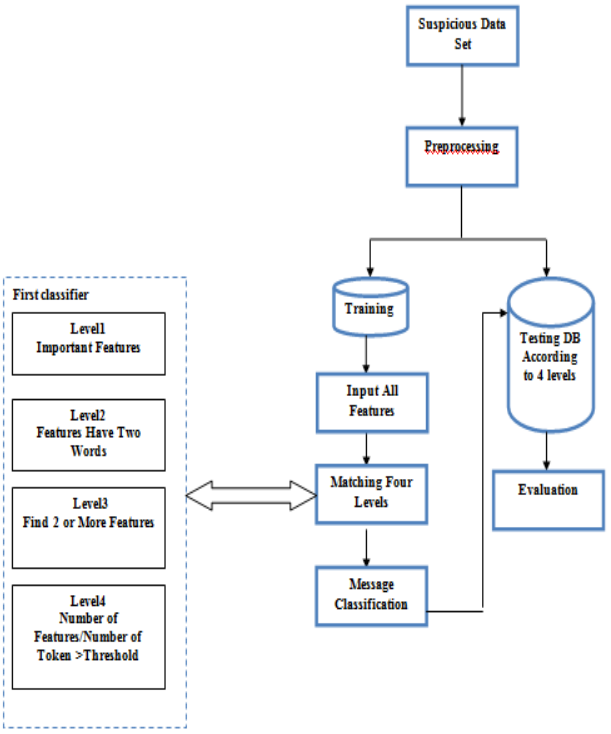

FIG 1. A BLOCK DIAGRAM OF THE STEPS OF CLASSIFICATION

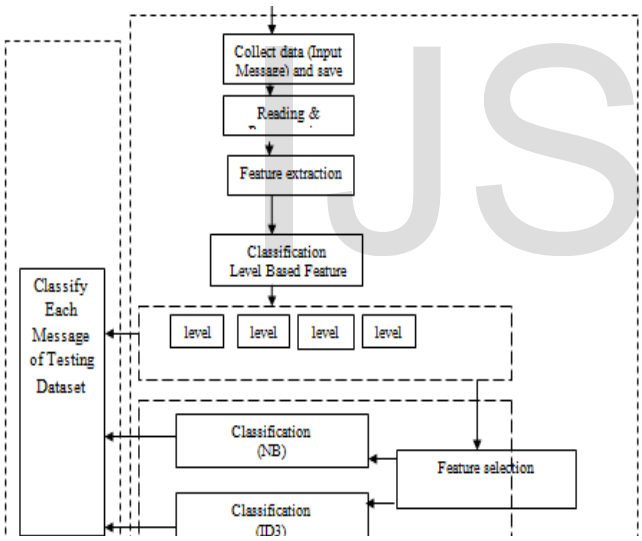

FIG 2. FUNCTIONALITY OF THE SUSPICIOUS DETECTION SYSTEM. (A) FIRST CLASSIFIER WITH ALL FEATURES

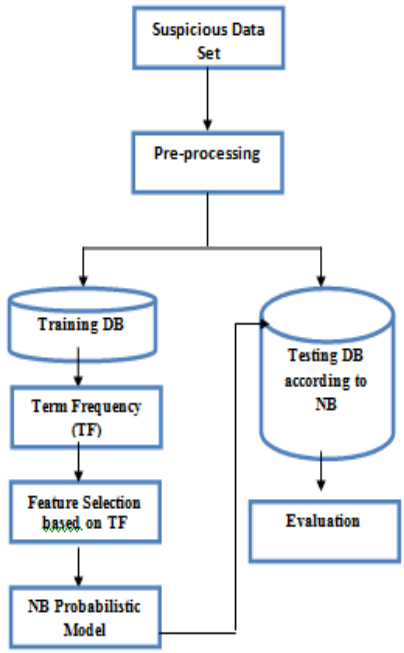

FIG 3. FUNCTIONALITY OF THE SUSPICIOUS DETECTION SYSTEM. (4): NB CLASSIFIER IS APPLIED ON A SUBSET OF FEATURES.

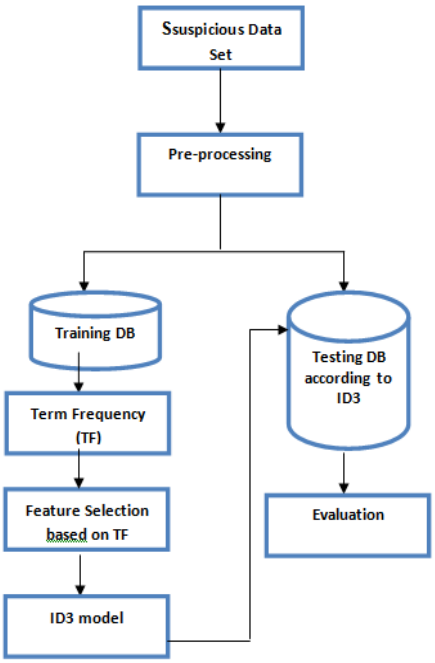

FIG4.ID3 CLASSIFIER IS APPLIED ON A SUBSET OF FEATURES.

a. Input File Reading the Message

The main method for collecting instant message data (Viber message) is to request mobile users to voluntarily give text messages because there is no such benchmark instant message dataset available in the counterterrorism domain. Primarily; this method has been used to collect real SMS text messages for research study. All messagetexts for training and testing are monolingual English written. Table 1 provides a sample of non-suspicious messaging and Table 2 provides a sample of suspicious message.

TABLE 1
NON-SUSPICIOUS MESSAGE

	Non - suspicious messages
1	Baghdad - Iraqi Prime Minister Haider al-Abadi called on Saturday for legal action over allegations that senior officials took millions of dollars in bribes to help major firms secure lucrative oil sector contracts.
2	Obama said tell us the person who made the statements doesn't know much about foreign policy or nuclear policy or the Korean peninsula or the world generally".
3	I'm gonna be home soon and i don't want to talk about this stuff anymore tonight I've cried enough today.
4	Health insurance protects you and your family from a financial hardship due to medical expenses in the untimely event that you are injured or become ill.

TABLE 2
SUSPICIOUS MESSAGE

	suspicious messages
1	For the first time since deadly terrorist bombings rocked Brussels, The city main airport on Sunday will reopen to passengers albeit to a very limited number of them.
2	Three American contractors who were kidnapped last month by an armed group in Baghdad have been freed, two Iraqi security officials and a U.S. official said Tuesday. Elements of the Iraqi intelligence apparatus freed the three; the officials said Further details about their rescue weren't immediately available.
3	ISIS has claimed responsibility for a suicide attack in a city south of Baghdad that killed at least 35 people, Iraqi police officials said. At least 105 others were wounded in the attack, which targeted one of the busiest checkpoints in Hilla, according to police officials in the city.
4	AThe nuclear security summit comes in the wake of attacks in Paris and Brussels that have killed dozens and exposed Europe's inability to thwart destabilizing attacks or track Islamic State operatives returning from Iraq and Syria.

b. Preprocessing of the Message

To classify the text documents, first, the authors determine the suspicious features and store them in a database. Then, preprocessing is applied as shown in Figure (5). This process objective is to decrease the document space and clean the text message. The steps are: (1) tokenization and normalization, (2) stop word removal (see Table 3), and (3) stemming. There are many types of the stemming algorithms, several of which can output incomplete stems that do not have any meaning.

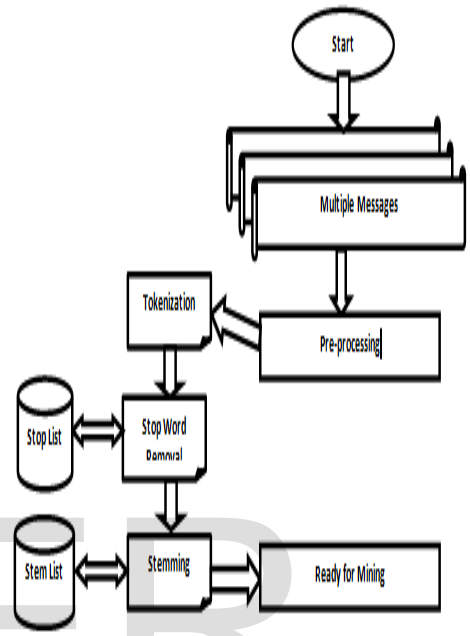

FIG5.PREPROCESSING OF MESSAGE

TABLE 3
SAMPLE OF STOP WORD

able	amid	asked	as	both
about	amidst	asking	bill	bottom
above	among	asks	consider	brief
abroad	amongst	associated	already	briefly
abst	amoungst	at	around	but
abstract	amount	aug	big	by

c. Evaluation Metrics

1. Accuracy = (TP + TN)/N: "is the ratio between the number of text documents which were correctly categorized and the total number of text documents" [19].
2. Error = (FP + FN)/N: "is the ratio between the number of text documents which were not correctly categorized and the total number of text documents" [19].
3. Precision = TP / (TP + FP): "is the percentage of correctly categorized text documents among all text documents that were assigned to the category by the classifier" [19].

4. Recall = TP / (TP + FN): "is the percentage of correctly categorized text documents among all text documents belonging to that category" [19].
5. F1 = ((2*Recall*Precision)/(Recall + Precision)): is anaccuracy test's measurededucedby way ofcomputing the weighted average of precision and recall. F1 score ranges its top value on 1 and worst on 0 [20].
6. CTrue Positive Rate (TPR)= TP/ (TP+FN): the fraction of positive target that are classified as positive [21].
7. False Positive Rate(FPR) = FP/(FP+TN): the fraction of negative examples (No, False, -) classified as positive (Yes, True, +) [21].

6. EXPERIMENTAL RESULTS

A. Stage 1: First Classifier (LBFC)

1. Level1: Find Important Feature on one Word

TABLE 4
LEVEL1 RESULT

Accuracy	0.957143
Error	0.042857
Precision	0.970588
Recall, TPR, Sensitivity	0.942857
FPR	0.028571
F1	0.956522

2. Level2: Feature Consists of Two Words

TABLE 5
LEVEL2 RESULT

Accuracy	0.7
Error	0.3
Precision	1
Recall, TPR, Sensitivity	0.4
FPR	0
F1	0.571428571

3. Level3: Test by Threshold

TABLE 6
LEVEL3 RESULT

Accuracy	0.942857
Error	0.057143
Precision	0.969697
Recall, TPR, Sensitivity	0.914286
FPR	0.028571
F1	0.941176

4. Level4: Find Two or More Features

TABLE 7
LEVEL4 RESULT

Accuracy	0.928571
Error	0.071429
Precision	0.941176
Recall, TPR, Sensitivity	0.914286
FPR	0.057143
F1	0.927536

The average accuracy of four lvels=0.957143+0.7+0.942857+0.928571= 0. 882143.The classification result after applying LBFC to the training sample is as follows. The maximum accuracy obtained is 0.957143that occurred in Level One. Important features achieved by classification counters results are (TP=33, FN=2, FP=1 and TN=34). The minimum accuracy obtained is 0.7, which is achieved by classification counters results are (TP=14, FN=21, FP=0 and TN=35) that occurred in Level Two: Feature consists of two words, as shown in Figure (6).

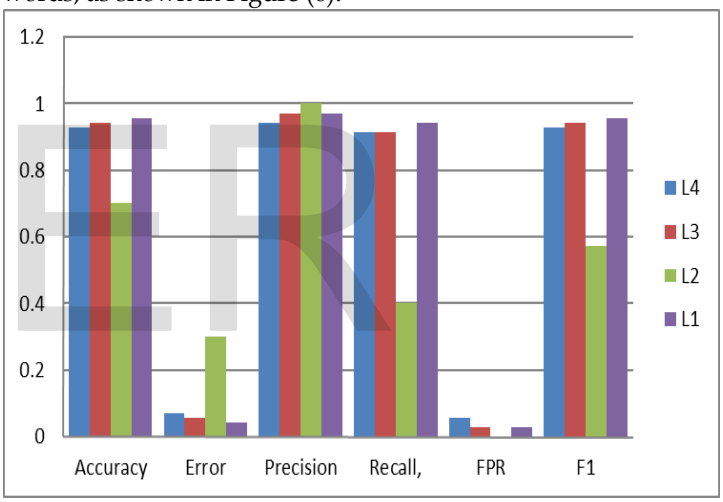

FIG6. Evaluation matrices level4.

B. Stage 2: second Classifier NB

Now we will introduce the second classifier (NB), which is trained to the sample of "training dataset." The results of testing NB3 classifier to "testing dataset" are displayed in Table 8

TABLE 8
NB CLASSIFIER MEASURES EVALUATION (TRAINING SAMPLE)

Accuracy	0.942857
Error	0.057143
Precision	0.969697
Recall, TPR, Sensitivity	0.914286
FPR	0.028571
F1	0.941176

C. stage3: third ID3 Classifier Training and Testing

Now we will introduce the third classifier, which is ID3 and is

trained to the sample of "training dataset." The results of testing ID3 classifier to "testing dataset" are displayed in Table IX.

TABLE 9
ID3 CLASSIFIER MEASURES (TRAINING SAMPLE)

Accuracy	0.957143
Error	0.042857
Precision	1
Recall, TPR, Sensitivity	0.914286
FPR	0
F1	0.955224

The total analysis of the proposed Suspicious Detection system after the three classifiers have been implemented is shown in Table X. The analysis is based on the accuracy of each used classifier.

TABLE 10
TRAINING SAMPLES AND ACCURACY RESULTS

No. of Training Samples	**No. of Testing Samples**	**Accuracy**			
		First Classifier		**Second Classifier NB**	**Third Classifier ID3**
130	70	Level 1	0.957143	**0.942857**	**0.957143**
		Level2	0.7		
		Level 3	0.942857		
		Level4	0.928571		
		Average of Accuracy	**0.882143**		

7. CONCLUSION

1. The first classifier is clear that the best value of accuracy derives from levelone and the average of accuracy is 0.882143.
2. For more accuracy, the proposed system uses NB and ID3.
3. Searching for features consisting of twowords, for example "Islamic state," that have suspicious meaning if they appear in a message. When one of these features appear in a message it is not necessarily wholesalesuspicious if find Islamic not mean message is suspicious or find state that not mean is suspicious but if find Islamic state that mean is message is suspicious.
4. When using semantic meaning, the problem of reducing dataset was solved.
5. The dimensions of the data and, consequently, the computation time to construct the classifier are considerably affected by feature selection techniques.
6. The feature set size has an evident effect on the performance of the suspicious detection classifier. It has been noticed that an increasing size of features set also leads to increased performance of the classifier.

REFERENCES

[1] Saxena, N., Bhargava, N. and Mahor, U. (2012) A Competent Technique on Cluster Based Master Slave Structural Design. International Journal,1, no. 1, pp.
[2] Giorgino, T. (2008) An introduction to Text Classification, Retrieved on October 13 (2004)
[3] Mehta, S., Eranna U.and Soundararajan, K. (2012) A Neural Technique for SMS Classification Using Keywords Search and Identification of Captured Messages, Using Hebbian Learning. International Journal of, pp.
[4] Sarl, V.M. (2016) Viber on the App Store. https://itunes.apple.com/mr/app/viber/id382617920?mt=8
[5] PatelD. and Bhatnagar M. (2011) Mobile SMS Classification an Application of Text Classification. International Journal of Soft Computing and Engineering, 2231-2307.
[6] Shirani-Mehr, H. (2013) SMS spam detection using machine learning approach. 1-4.
[7] Mahmoud, T.M. and Mahfouz, A.M. (2012) SMS Spam Filtering Technique Based on Artificial Immune System. International Journal of Computer Science Issues, 9, no.1, 589-597.
[8] Shahi, T.B. and Yadav, A. (2013) Mobile SMS Spam Filtering for Nepali Text Using Naïve Bayesian and Support Vector Machine. International Journal of Intelligence Science, 4, no. 1, 24.
[9] Merriam-Webster. (2015) Definition of SUSPICIOUS. http://www.merriam-webster.com/dictionary/suspicious
[10] S. Inc. (2016) Final report. https://www.scribd.com/document/156343006/Final-Repor
[11] Beal, V. (no date) What is IM (instant message)? Webopedia definition. http://www.webopedia.com/TERM/I/IM.html
[12] Patel, D. and Bhatnagar, M. (2011) Mobile SMS Classification an Application of Text Classification. International Journal of Soft Computing and Engineering,2231-2307.
[13] Shirani-Mehr, H. (2013) SMS Spam Detection Using Machine Learning Approach. 1-4.
[14] Mahmoud, T.M. and Mahfouz, A.M. (2012) SMS Spam Filtering Technique Based on Artificial Immune System. IJCSI International Journal of Computer Science Issues, 9, no. 1, 589-597.
[15] Awad, W.A. and ELseuofi, S.M. (2011) Machine Learning Methods for E-mail Classification. International Journal of Computer Applications (0975–8887), 16, no.1, pp.
[16] Guan, A.D. and Chung, T.C. (2014) SMSClassification Based on Naive Bayes Classifier and Apriori Algorithm Frequent Itemset. International Journal of Machine Learning and Computing, 4, no. 2, 183.
[17] Shahi, T.B. and Yadav, A. (2013) Mobile SMS Spam Filtering for Nepali Text Using Naïve Bayesian and Support Vector Machine. International Journal of Intelligence Science, 4, no. 1, 24.

[18] Merriam-Webster. (2015) Definition of SUSPICIOUS. http://www.merriam-webster.com/dictionary/suspicious
[19] Song, Y. (2009) Machine Learning for Text Mining: Classification. Dissertation, The Pennsylvania State University, Pennsylvania.
[20] Sasaki, Y. (2008) Automatic Text Classification. University of Manchester: presentation.
[21] Internet survey, Data (2014b) Simple Guide to Confusion Matrix Terminology. http://www.dataschool.io/simple-guide-to-confusion-matrix-terminology