Classification of SMS in Mobile System Using Multi Class Classifier and Randomizable Filtered Classifier

SMS Analysis for Future Security and User Convenience

[1]Neetu Gupta, [2]Ms. Sonal Arora

[1]Student, [2]Assistant Professor,
[1] Department of Computer Science and Engineering,
[1]DPGITM (MDU University), Gurgaon, India

Abstract: The intense boom of smart cell telephones and customers has contributed to the enlargement of online or offline Instant Messaging and SMS usage as an alternate way of Transaction and communication. Along with the faith they instinctively have of their devices makes this sort of messages a congenital environment for spammer.

In fact, reports distinctly suggests that extent of junk mail over Instant Messaging and SMS is rapidly increasing 12 months through yr. This represents a challenging problem for classical filtering methods these days. Smishing this term represents a phishing in SMS/Messages referred to as SMS-phishing is a cyber-protection attack, which utilizes Short Message Service (SMS) to scouse borrow personal statistics/credentials of cellular customers. The religion degree of cell users on their smartphones has attracted attackers to carry out various cellular protection assaults like SMS-Phishing. In this paper, we implement the SMS-Case-based statistics mining classification approach to classify them subpart of SMS category by detecting of legitimate, Illegitimate/Smishing messages and these classified messages will similarly categorize in 3 parts Primary, Other, Fake. In this research paper Multi Class Classifier and Randomizable Filtered Classifier algorithms are used to classify the SMS and furthermore, we will analyses the multiple algorithms using experimenter in Weka. During the lockdown period because of Pandemic suffering from COVID19 SMS- Phishing will become more energetic and the attackers send fraud messages to the mobile users intensively.

Keywords:
SMS, Message Analysis, Smishing, Illegitimate, SMS Classification, Data Mining, Cases, Phishing, Short Messages,

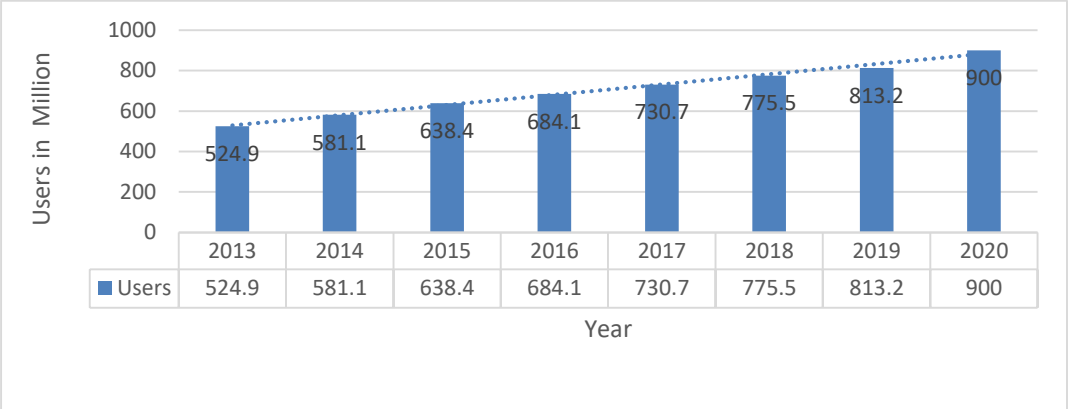
Benefits:
Using this proposed solution giant organization can enhance their existing mobile Message/SMS application to the user security and convenience.

Why this approach/ need of this solution:
As SMS/Instant messages has become alternative way of transaction and communication, so there is a need to make SMS services more secure and classified.

I. INTRODUCTION

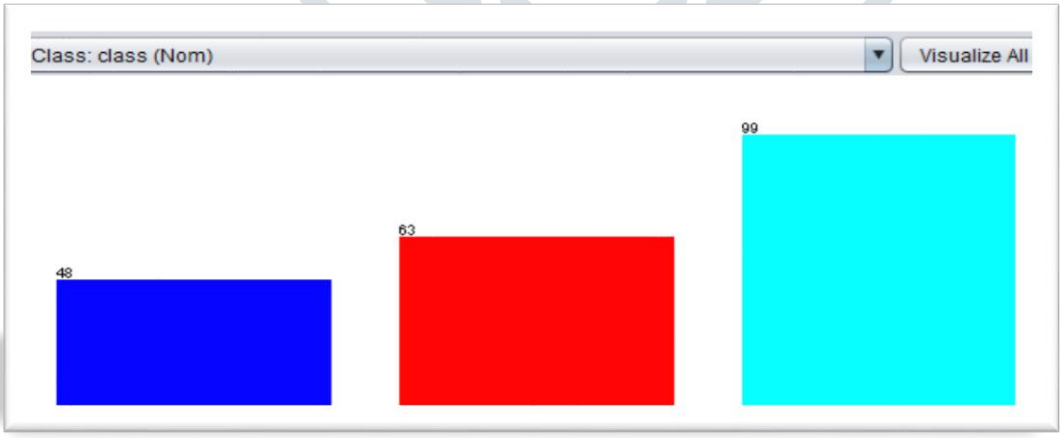
During the lockdown period because of Pandemic laid low with COVID19 SMS- Phishing turns into more active and the attackers send fraud messages to the mobile customers intensively. Not only attackers in truth the cash frauds lively a lot. As we all recognize Short text messaging has end up an average of verbal exchange those days. IM/SMS are truly the leading manner of communication. In fact, estimation says that near about 1.5 billion messages are sent a day by using considering just SMS. The popularity of (SMS) has been developing over the past decade. The intense growth of smart mobile telephones has contributed to the boom of online or offline SMS usage as an alternative way of Transaction and conversation. For organizations textual content messages are less difficult than even emails this is regularly because while 96% of cell users read their SMS via the top of the day about 85% of the emails stay unopened SMS assessment 2018 [3-4]. Example of a fraud/fake unsolicited mail text message is like ‘‘Hurray!! CONGRATULATIONS!!: YOUR MOBILE NO HAVE WON 1000,000 IN YOUR ACCOUNT— MOBILE DRAW USA, TO CLAIM PRIZE SEND BANK DETAILS, NAME, AGE, etc. Many greater messages patterns we can see on our cell system. So, this paper is in addition extension of my assessment paper, basically some studies work the usage of some algorithms and their comparative analysis might be discussed. The technique will be data mining technique is classification and algorithm used Multi Class Classifier and comparative analysis with and Randomizable Filtered Classifier algorithms a Meta approach through Weka.

Fig1: Mobile Phone User analysis in India [2]



Now a day's mobile security may be a major concern because attackers have diverted their mind from Computers to smartphones due to technology growth. Moreover, people are more attracted towards smartphones because it may be a portable and multi-functioning device, Smartphones are more popular now a days as compared to laptops due to their small screen size, lower cost, and portability. consistent with Dimensional Enterprise Mobile Security Survey report and it shows that Smishing attack stands at the second position altogether quite mobile devices attacks [14]. There's two sort of security methods are wont to identify Illegitimate/fake mobile SMS. The primary method is that the Blacklist based method that forestalls the incoming SMS from the fake sources [17]. However, blacklist-based techniques don't cover all the fake sources, as an attacker can buy any mobile number to send the Illegitimate/fake/bogus SMS. The second sort of solution is predicated on the machine learning algorithm during which various features are extracted and compute from the SMS to require appropriate decision. The advantage of the machine learning based technique is that it can detect the fake message coming from any source. Data processing methods help within the feature extraction and finding the relation between them [16]. These approaches identifying hidden knowledge from datasets in terms of Cases and make the choice supported extracted Cases. Human easily understands these Cases. In this paper, some fraud terms data processing classification approach within the prediction of useful/illegitimate/promotional/offers SMS. We've used WEKA tool to classify SMS/Messages for data processing. We study the varied characteristics of text messages thorough then found more than fifteen terms/Cases which may efficiently classify SMS to the subcategory. We then use classification algorithm namely Multi Class Classifier and Randomizable Filtered Classifier algorithms. In this, we've also identified the illegitimate/Smishing messages. The performance of the proposed approach is evaluated, and it achieved quite 98% of true negative rate and 99.0% true positive rate.

Fig 2: SMS frequency analysis



(Primary/Useful, Other/ Promotional/Offers, Illegitimate/Smishing)

II. REVIEW OF LITERATURE

Over the years, data scientists have proposed several ML models to spot Spam and not Spam. These aren't only for mobile text messages but also email spam and on social network platforms like Facebook Twitter delany et al [23] provided a survey of existing works for filtering spam SMS. They mostly covered articles that relied on conventional machine learning approaches but not deep learning for instance, [24] compared Bayesian classifier with other classification algorithms and located that the

previous was better to classify Spam text Messages. Androulidakis et al. [25] proposed another version to clear out Spam messages. Their version became supported the Android OS during which the user's cell manage was wont to filter out the Spam. The model checked the knowledge of message senders against a previously defined spammer list so when a message came from the users present within the list of spammers it had been treated as spam else not spam zainal et al.

Related Work

This section discusses the various existing mobile Text SMS classification detection techniques. The existing mobile classification and detection techniques divide into following section.

a) User Knowledge/Education Based Scheme

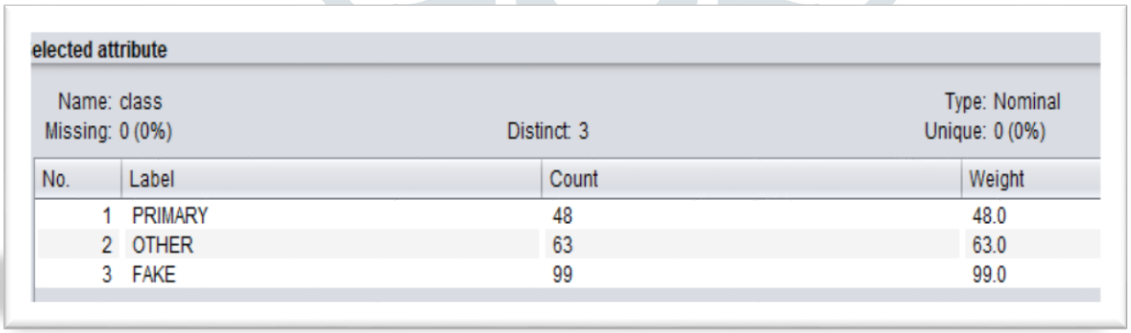
The educational based solutions focus on educating the mobile users about the characteristics of phishing message through training, workshop and awareness programs so that they correctly identify the phishing attack [8]. However, the phishing attack becomes successful due to human flaws and ignorance. This conceptual knowledge may help the users in avoiding phishing attacks.

b) Technical solutions to mitigate mobile phishing attack

The technical solutions are cost-effective and simple to enforce as compare to educational based answers. In this, Amrutkar et al. [9] cautioned mechanism named KAYO, which differentiates among the malicious and genuine cell webpages. It detects mobile malicious pages by measuring 44 mobile features from webpages. Among 44 features, 11 are newly identified mobile specific features. KAYO's 44 feature set is split into four classes namely HTML, mobile specific, URL and JavaScript features. Joo et al. [6] proposed a model 'S-Detector' for detecting Smishing attack. They used Naïve Bayesian Classifier in their system to filter Smishing messages by finding the words used more often in these messages. S-Detector consists of SMS monitor, SMS analyzer, SMS determinant, and Database. Foozy et al. [7] proposed a Rule-based methodology to filter Illegitimate/Smishing messages from spam messages. Authors applied two Rule namely 'winner announcement' and 'marketing advertisement'. They need applied the Bayesian technique in WEKA tool to see the accuracy of Smishing, spam and ham messages. Alfy et al. [15] proposed a spam filtering model for both email and SMS. The proposed technique used 11 features namely presence of URLs, likely spam words, emotion symbols, special characters, gappy words, message metadata, JavaScript code, function words, recipient address, discipline and spam domain. they need evaluated their proposed model with five email and SMS datasets. Within the literature, we will conclude that no single technique exists which will detect illegitimate/Smishing attacks efficiently. Therefore, we'd like a way which will protect the user against Illegitimate/fake/Smishing attacks.

I. RESEARCH METHODOLOGY

In this we have discuss our proposed methodology of classifying messages into different categories by using some terms/cases and accordingly we will classify them using WEKA tool by classification algorithms like a lazy approach i.e. Multi Class Classifier and Randomizable Filtered Classifier algorithms and will identify the accuracy, further will also use some detection cases to detect illegitimate/Smishing SMS/ Messages.



elected attribute

Name: class Type: Nominal
Missing: 0 (0%) Distinct: 3 Unique: 0 (0%)

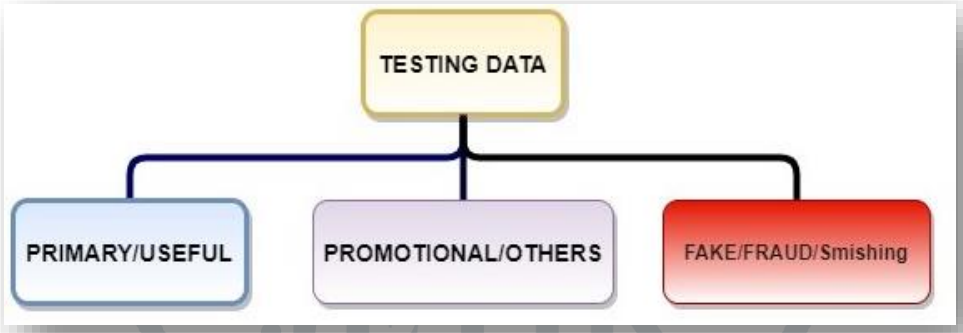
No.	Label	Count	Weight
1	PRIMARY	48	48.0
2	OTHER	63	63.0
3	FAKE	99	99.0

The proposed approach is a model to filter SMS, protects the user from the phishing SMSs by blocking these messages and also implemented system can classify them into different category and delivering only Normal ones to the mobile user instead of making all into a single category it will further filtered into different category like- Primary/Useful, illegitimate/fake, Other/Promotional. The SMS detection is a type of ternary classification problem where a message can be the divide in three categories (i.e., Primary, Illegitimate/fake, Other/Promotional). Illegitimate message is a dangerous spam message that steals personal data/credentials. As per our research and observation, we find the followings characteristics of fraud message:

- ✓ *It can have .exe message content in the link form.*
- ✓ *SMS having URLs.*
- ✓ *Now a days a different format seen like SMS includes.txt files.*
- ✓ *It can have any honey coated audio/video content that can trap the user.*

- ✓ *Advertising for offers/ Promotions.*
- ✓ *It can have the bogus fake links. Advertising something like providing free minutes, etc.*
- ✓ *It can have email address or a phone number.*
- ✓ *Links can have harmful viruses.*
- ✓ *Machine Recorded voice/Self-answering SMS asking the user to subscribe or unsubscribe any service.*
- ✓ *Announcing to users as a winner for fake contest and attract him using the prize money.*
- ✓ *Intended to spread some fake news.*
- ✓ *Message can have link and link have steganography.*
- ✓ *Java script contents.*
- ✓ *Long SMS can have fake. Etc.*

Fig 4: Testing Data Classification Hierarchy



III. TOOL AND TECHNIQUES

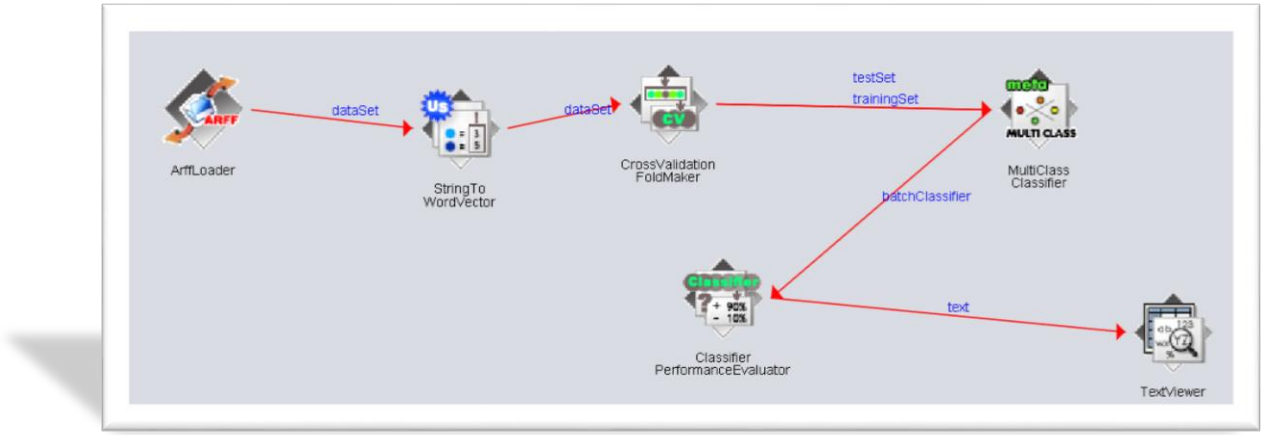
In our research work we used data mining classification techniques to classify the data and accordingly we found accuracy, and result set data, to do so we used WEKA tool, PC configuration having windows 10 core i5 processor 8 GB RAM etc.

Basic Steps will follow the different stages of data mining-

- ❖ Text Data collecting.
- ❖ Text Data Pre-processing.
- ❖ Classifier Training, Testing and Validation data sets
- ❖ Fraud Terms/Case- Extraction.
- ❖ Data Analysis (Attribute generation & selection)
- ❖ Visualization (Applying Text Mining algorithms)
- ❖ Evaluation

In Weka data mining tool, there is a different way we can analyse data like: Explorer, Experimenter, Knowledge Flow Environment Workbench, simple CLI. Below is the snapshot using Weka Knowledge Flow. Weka Knowledge Flow is the way to analyse data by graphical design approach [30]. Designs are as ArffLoader, String to word vector, Cross Validation fold maker, Multi Class Classifier and Randomizable Filtered Classifier algorithms, Classifier Performance Evaluator TextViewer etc. In ArffLoader we will load the data set which we need to analyse, here in my case my data set is in Text format so we need to convert this to the Word vector format, then here in fig we used Cross Validation Fold maker and then we pass the train set and test set data to the used algorithm and then a Classifier is used and finally Text Viewer is used to display the result.

Fig 5: Knowledge Flow Environment link using Multi Class Classifier (Meta Approach):

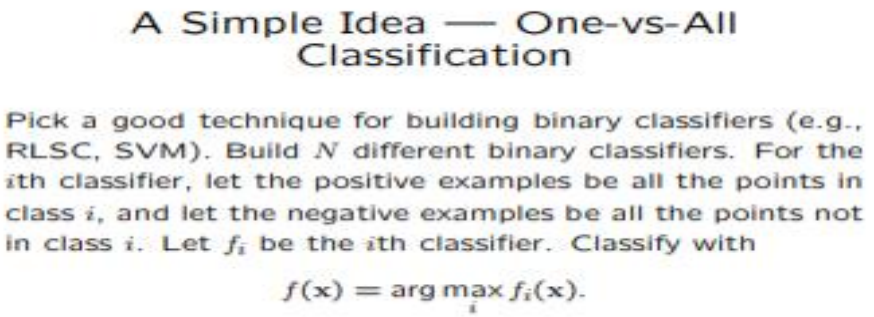


IV. Algorithm Analysis and Details:

Multi Class Classifier (Meta Approach)

Classification tasks more than two classes. Classification is the act of deciding the category of a given object based on several attributes related to that object. Despite the long history of classification, the research on this topic was limited in theory before 1960.[38]. Alongside the progress of computers and due to new interest, automatic pattern classification has gained more attention. Automatic pattern classification employs a machine learning algorithm to induce a classier given a training data set. The induced classier than should be able to assign a predefined class label for new data from the same domain.[39]. In the subsequent we provide a quick review of two usually used decomposition techniques accompanied by brief abstracts of the well known alternative decomposition techniques proposed inside the literature.

1. One-Against-All: Perhaps the most standard method for decomposition of a multiclass classification problem into binary subproblems is the One-Against-All (OAA) strategy. In this strategy, k different binary classifiers are trained to classify k different classes, each of which separate a single class from the remaining. That is, the samples in one class are considered positive examples and the remaining samples belonging to the other classes are considered negative examples. Using the highest output value for an unknown sample, OAA reveals the corresponding class of the sample. The main disadvantage of OAA is that it may induce an inaccurate binary classifier for given classes when the data is unbalanced,30 i.e. the number of positive examples is too low compared to the number of negative examples and vice versa.



2. One-Against-One: Another common and popular decomposition strategy is the One-Against-One (OAO). In this strategy, all possible pairs of different classes are taken into account and therefore kðk 1Þ=2 binary classifiers are induced, each of which separate a pair of classes. Then the ¯nil classifier is built by combining individual binary classifiers. The main drawback of OAO is as follows: if the number of training samples is not enough and the binary classifiers are not regularized carefully, the nil classifier will tend to over¯t.30 The training process of the binary classifiers in this approach is simplified and needs less time compared to OAA. This is due to the fact that in OAO for each binary classifier, only the samples of two classes are considered, while in OAA all the samples are used for training binary classifiers. In this strategy, however, the number of binary classifiers grows super linearly with the number of classes.

Another Simple Idea — All-vs-All Classification

Build $N(N-1)$ classifiers, one classifier to distinguish each pair of classes i and j. Let f_{ij} be the classifier where class i were positive examples and class j were negative. Note $f_{ji} = -f_{ij}$. Classify using

$$f(\mathbf{x}) = \arg\max_i \left(\sum_j f_{ij}(\mathbf{x}) \right).$$

Also called all-pairs or one-vs-one classification.

Summary:

Correctly Classified Instances	***200***	*95.2381 %*
Incorrectly Classified Instances	***10***	*4.7619 %*

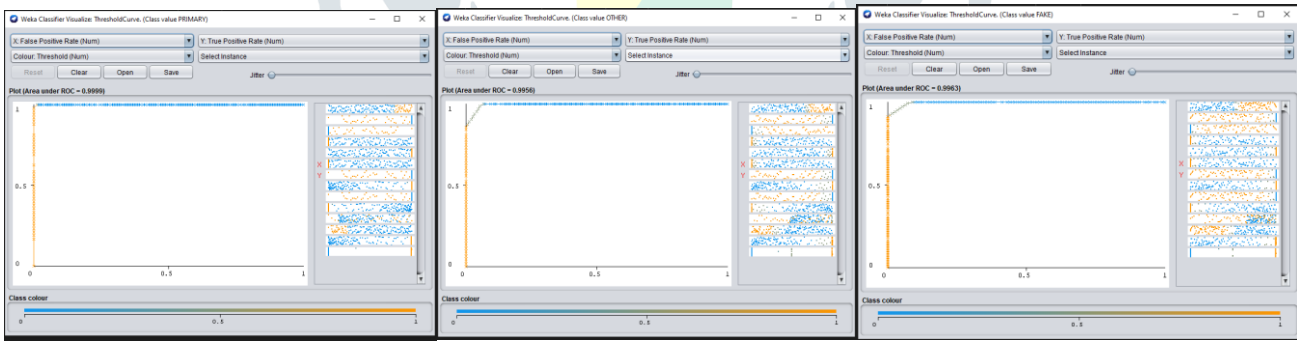
Current relation

Relation: SMSTEXTANALYSIS-weka.filters.unsupervis... Attributes: 2054
Instances: 210 Sum of weights: 210

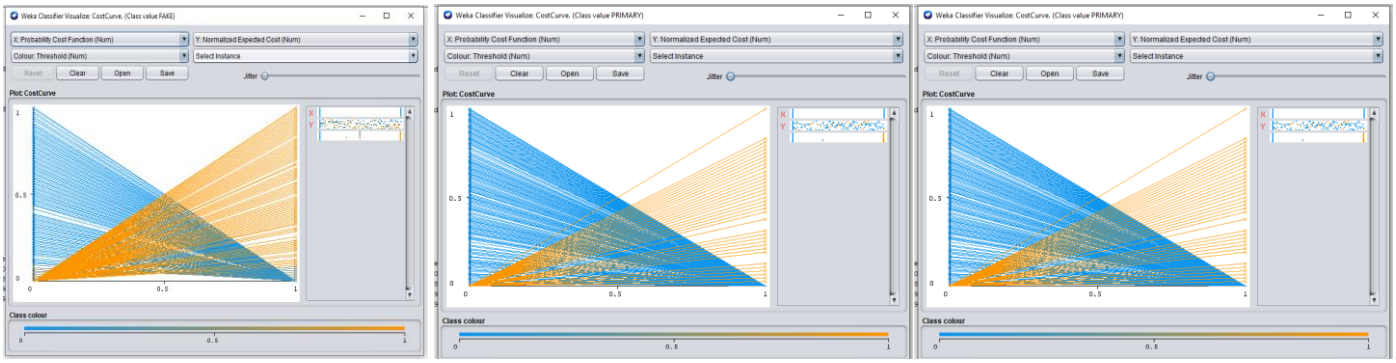
Detailed Accuracy by Class:

TP Rate	*FP Rate*	*Precision*	*Recall*	*F- Measure*	*MCC*	*ROC Area*	*PRC Area*	*Class*
0.979	0.000	1.000	0.979	0.989	0.986	1.000	0.999	Primary
0.968	0.048	0.897	0.968	0.931	0.902	0.996	0.989	Other
0.929	0.027	0.968	0.929	0.948	0.905	0.996	0.995	Fake
0.952	0.027	0.954	0.952	0.953	0.923	0.997	0.994	***Weighted Avg***

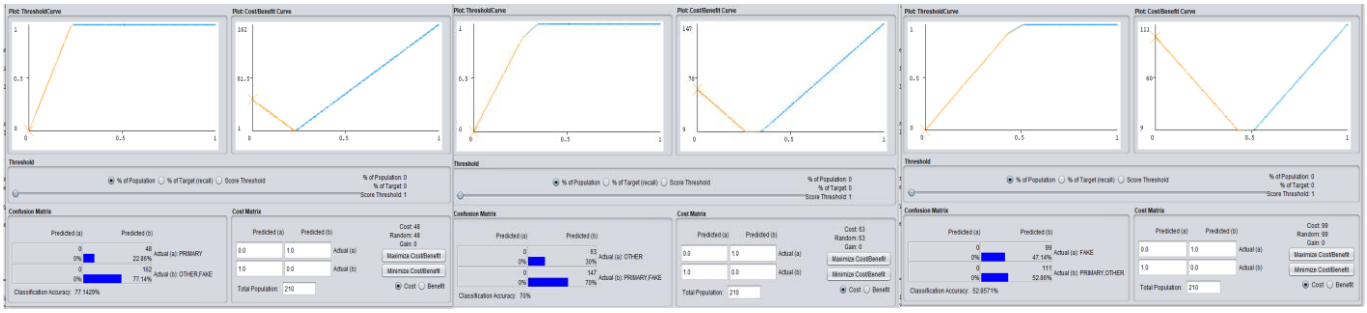
Other Measures and Graph Evaluations:



Cost Curve analysis: For Primary, Other, and Fake Classification



Cost Benefit Analysis: For Primary, Other, and Fake Classification.



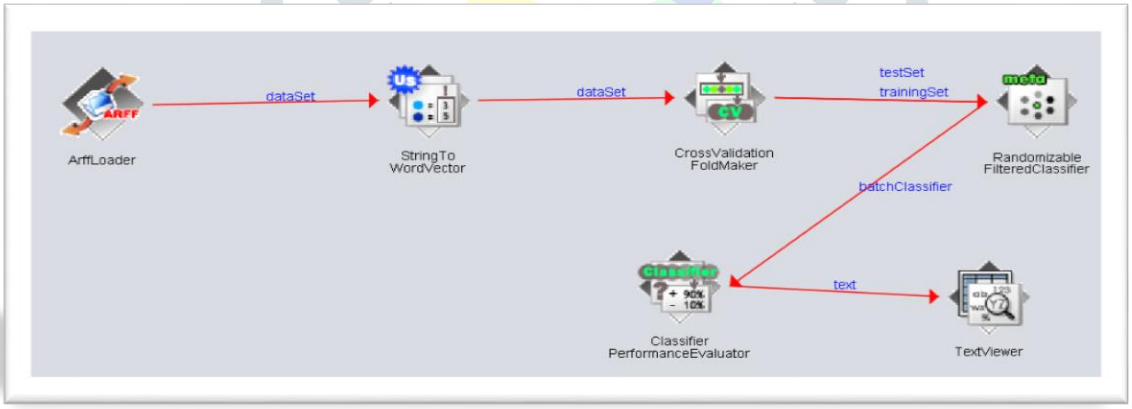
Time Taken:

Time taken to build model: 0.01 seconds
Time taken to test model on training data: 0.04 seconds

RANDOMIZABLE FILTERED CLASSIFIER:

A simple variant of the Filtered Classifier that instantiates the model with a randomizable filter, more specifically, Random Projection, and IBk as the base classifier. Other than this and checking that at least one of the two base schemes implements the Randomizable interface, it implements the same functionality as Filtered Classifier, which (now) also implements Randomizable. This approach employed an arbitrary classifier on statistics that has been passed through an arbitrary filter. Like the classifier, the structure of the filter worked exclusively on the training data and test instances will be processed by the filter without altering their structure (Hall et al., 2009). In using randomizable filter (RF) as an ensemble base classifier, each base classifier is built using a different random number of seed (but based on the same data). The final prediction is a straight average of the predictions generated by the individual base classifiers. Class for creating a committee of random classifiers. The base classifier (that forms the committee members) needs to implement the Randomizable interface.

Fig 6: Knowledge Flow Environment link using Randomizable Filtered Classifier:



Summary:

Correctly Classified Instances	***200***	*95.2381 %*
Incorrectly Classified Instances	***10***	*4.7619 %*

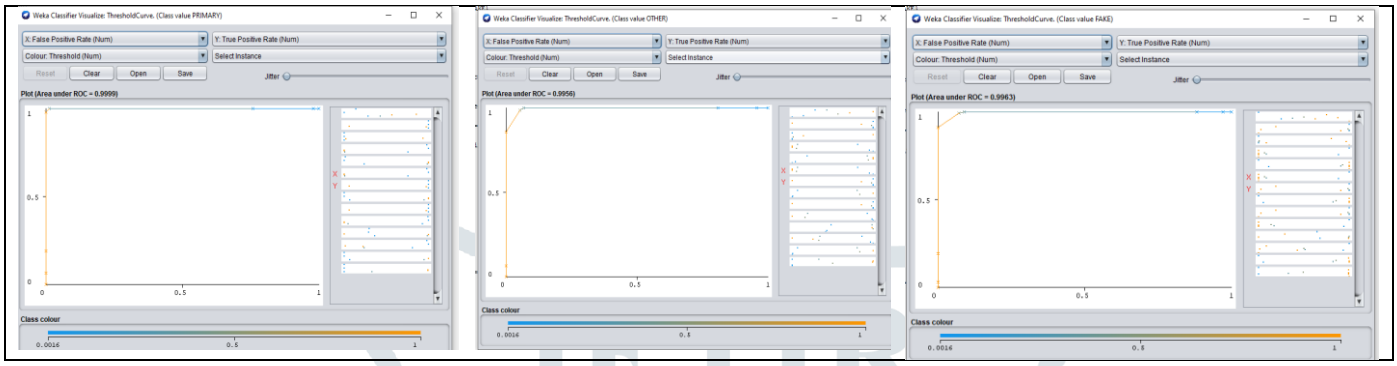
Time Taken:

Time taken to build model	0.03
Time taken to test model on training data: 0.13 seconds	0.08

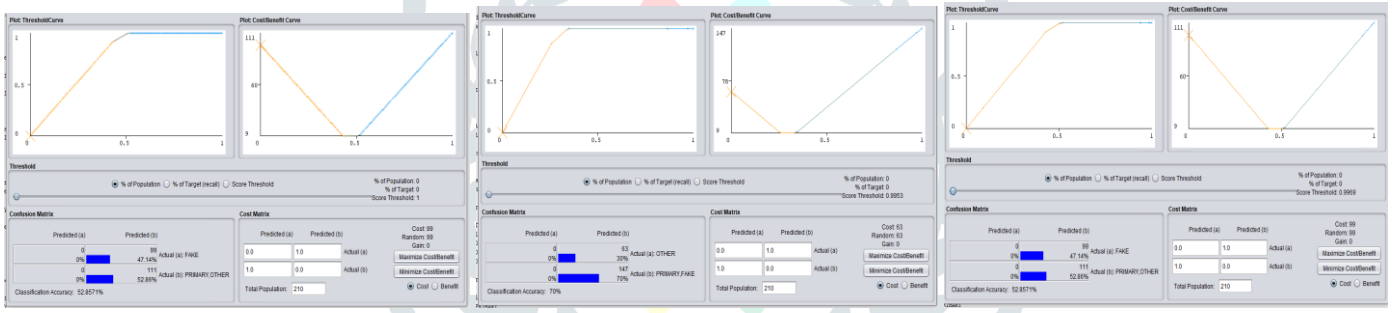
Detailed Accuracy output analysis by Class:
This classifier accuracy is better than Multi Class Classifier

TP Rate	*FP Rate*	*Precision*	*Recall*	*F- Measure*	*MCC*	*ROC Area*	*PRC Area*	*Class*
1.000	0.012	0.960	1.000	0.980	0.974	1.000	0.999	Primary
0.984	0.054	0.886	0.984	0.932	0.904	0.996	0.983	Other
0.909	0.000	1.000	0.909	0.952	0.917	0.996	0.993	Fake
0.952	0.019	0.957	0.952	0.953	0.926	0.997	0.991	***Weighted Avg***

Other Measures and Graph Evaluation: Threshold Curve Visualization for all the classified categories of Randomizable Filtered Classifier.



Cost Benefit Analysis: For Primary, Other, and Fake Classification.



Furthermore, a comparative different algorithms analysis using Experimenter for the same data set:

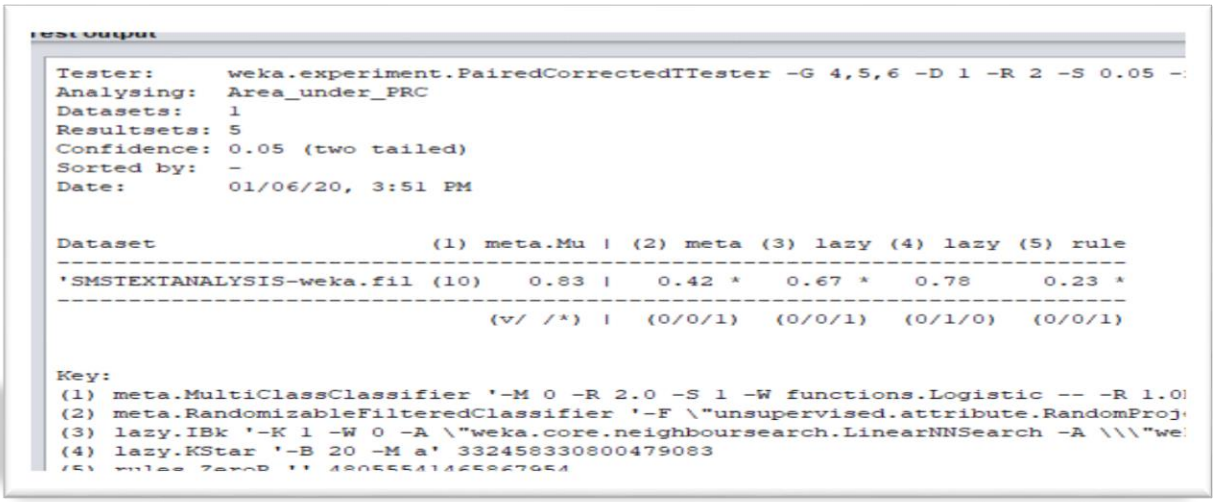
***Root Relative Square Errors of multiple algorithms together*:** Multiple algorithms analysis together using experimenter in Weka, here I used 5 different algorithms i.e. Multi Class Classifier, Randomizable Filtered Classifiers, IBK, K-Star and ZeroR.

```
Resultsets: 5
Confidence: 0.05 (two tailed)
Sorted by:  -
Date:       01/06/20, 3:44 PM

Dataset                    (1) meta.Mult | (2) meta.R (3) lazy.I (4) lazy.K (5) rules.
-----------------------------------------------------------------------------------------
'SMSTEXTANALYSIS-weka.fil (10)    110.00 |    130.61 v    109.27      107.00      100.00
-----------------------------------------------------------------------------------------
                                  (v/ /*) |     (1/0/0)     (0/1/0)     (0/1/0)     (0/1/0)


Key:
(1) meta.MultiClassClassifier '-M 0 -R 2.0 -S 1 -W functions.Logistic -- -R 1.0E-8 -M -1 -num-d
(2) meta.RandomizableFilteredClassifier '-F \"unsupervised.attribute.RandomProjection -N 10 -R
(3) lazy.IBk '-K 1 -W 0 -A \"weka.core.neighboursearch.LinearNNSearch -A \\\"weka.core.Euclidea
(4) lazy.KStar '-B 20 -M a' 332458330800479083
(5) rules.ZeroR '' 48055541465867954
```

Area Under PRC Multiple algorithms analysis:

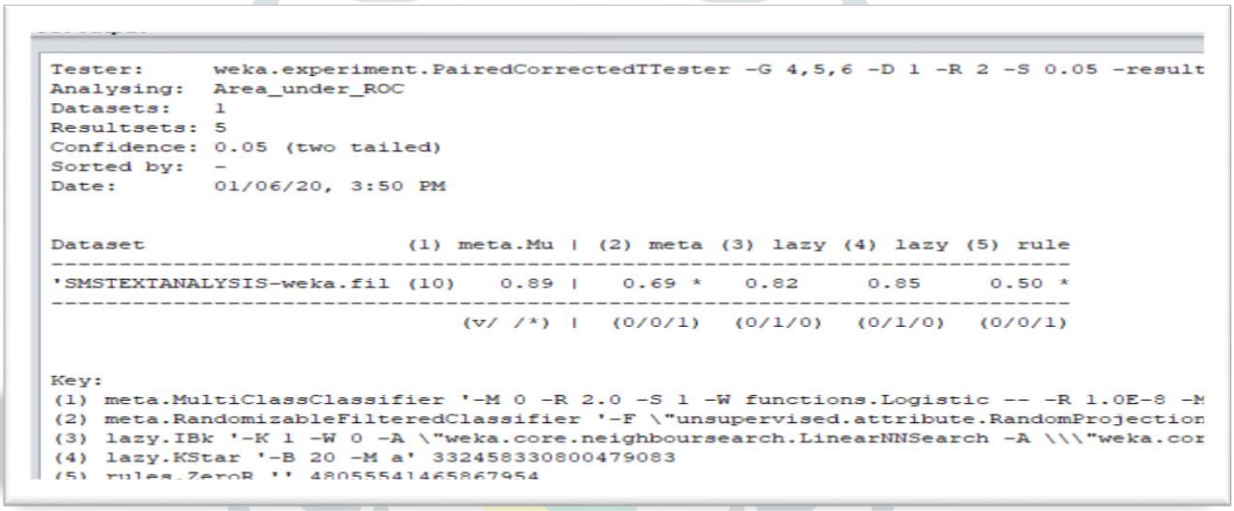


```
Tester:      weka.experiment.PairedCorrectedTTester -G 4,5,6 -D 1 -R 2 -S 0.05 -
Analysing:   Area_under_PRC
Datasets:    1
Resultsets: 5
Confidence: 0.05 (two tailed)
Sorted by:   -
Date:        01/06/20, 3:51 PM

Dataset                      (1) meta.Mu |  (2) meta  (3) lazy  (4) lazy  (5) rule
-----------------------------------------------------------------------------------
'SMSTEXTANALYSIS-weka.fil (10)    0.83 |    0.42 *    0.67 *    0.78      0.23 *
-----------------------------------------------------------------------------------
                                (v/ /*) |   (0/0/1)   (0/0/1)   (0/1/0)   (0/0/1)


Key:
(1) meta.MultiClassClassifier '-M 0 -R 2.0 -S 1 -W functions.Logistic -- -R 1.0
(2) meta.RandomizableFilteredClassifier '-F \"unsupervised.attribute.RandomProj
(3) lazy.IBk '-K 1 -W 0 -A \"weka.core.neighboursearch.LinearNNSearch -A \\\"we
(4) lazy.KStar '-B 20 -M a' 332458330800479083
(5) rules.ZeroR '' 48055541465867954
```

Area Under ROC Curve Multiple algorithms analysis:

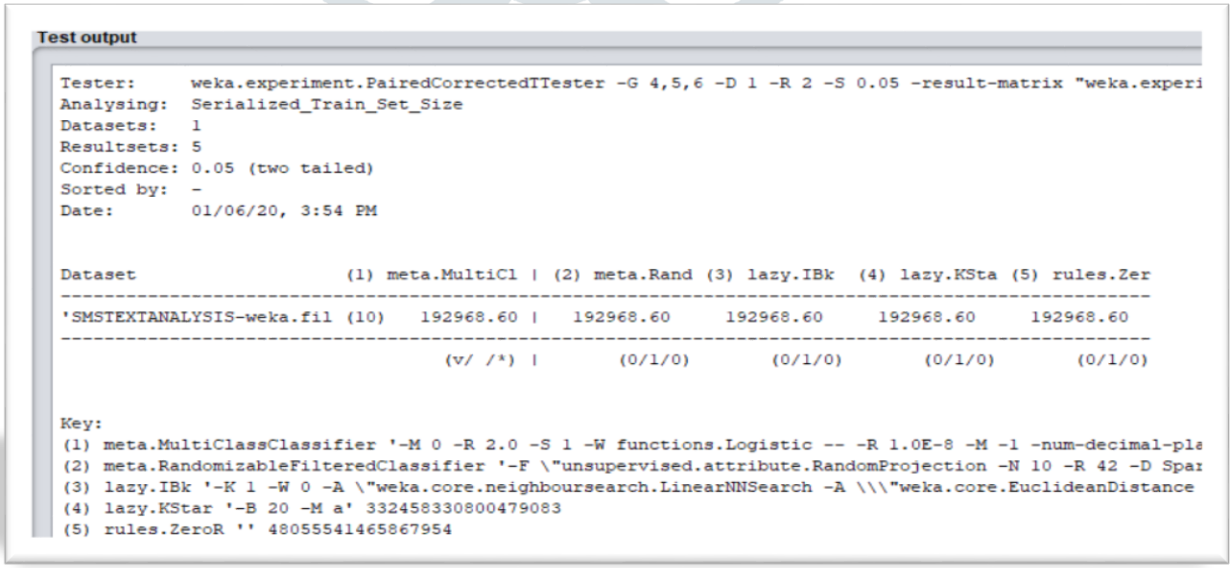


```
Tester:      weka.experiment.PairedCorrectedTTester -G 4,5,6 -D 1 -R 2 -S 0.05 -result
Analysing:   Area_under_ROC
Datasets:    1
Resultsets: 5
Confidence: 0.05 (two tailed)
Sorted by:   -
Date:        01/06/20, 3:50 PM

Dataset                      (1) meta.Mu |  (2) meta  (3) lazy  (4) lazy  (5) rule
-----------------------------------------------------------------------------------
'SMSTEXTANALYSIS-weka.fil (10)    0.89 |    0.69 *    0.82      0.85      0.50 *
-----------------------------------------------------------------------------------
                                (v/ /*) |   (0/0/1)   (0/1/0)   (0/1/0)   (0/0/1)


Key:
(1) meta.MultiClassClassifier '-M 0 -R 2.0 -S 1 -W functions.Logistic -- -R 1.0E-8 -M
(2) meta.RandomizableFilteredClassifier '-F \"unsupervised.attribute.RandomProjection
(3) lazy.IBk '-K 1 -W 0 -A \"weka.core.neighboursearch.LinearNNSearch -A \\\"weka.cor
(4) lazy.KStar '-B 20 -M a' 332458330800479083
(5) rules.ZeroR '' 48055541465867954
```

Serialized Training set size Multiple algorithms analysis using Experimenter in Weka:



```
Test output

Tester:      weka.experiment.PairedCorrectedTTester -G 4,5,6 -D 1 -R 2 -S 0.05 -result-matrix "weka.experi
Analysing:   Serialized_Train_Set_Size
Datasets:    1
Resultsets: 5
Confidence: 0.05 (two tailed)
Sorted by:   -
Date:        01/06/20, 3:54 PM

Dataset                    (1) meta.MultiCl |  (2) meta.Rand  (3) lazy.IBk  (4) lazy.KSta  (5) rules.Zer
-----------------------------------------------------------------------------------------------------------
'SMSTEXTANALYSIS-weka.fil (10)   192968.60 |    192968.60      192968.60      192968.60      192968.60
-----------------------------------------------------------------------------------------------------------
                                  (v/ /*) |        (0/1/0)        (0/1/0)        (0/1/0)        (0/1/0)


Key:
(1) meta.MultiClassClassifier '-M 0 -R 2.0 -S 1 -W functions.Logistic -- -R 1.0E-8 -M -1 -num-decimal-pla
(2) meta.RandomizableFilteredClassifier '-F \"unsupervised.attribute.RandomProjection -N 10 -R 42 -D Spar
(3) lazy.IBk '-K 1 -W 0 -A \"weka.core.neighboursearch.LinearNNSearch -A \\\"weka.core.EuclideanDistance
(4) lazy.KStar '-B 20 -M a' 332458330800479083
(5) rules.ZeroR '' 48055541465867954
```

I. CONCLUSION AND FUTURE WORK:

In our paper we have given an approach to secure SMS for now as well as in future. We had studied a lot with collected message data and made some cases to classify the messages into different category, so that it can be implemented into the existing application by the giant organizations like Google, Microsoft, and Apple not with different application but with the same existing application. Different inbuilt algorithms used by WEKA tool. True positive rate is 99.99%, and FPR is 0.01%.
As there is another application exist but giving the permission to the other application it's also a privacy breach so instead of securing system by other application it will be good to have a functionality in existing one application. Future research can have more data sets collected from different users; we will make more terms/cases through it so that we can have a better security to the users. Also, our planning to make more and more cases accordingly. We are exploring more algorithm and data set as well as cases and characteristics so that we can get better classification accuracy. Also, we will extend this research work using data mining tools like Rapid Miner and for sample and testing purpose we will develop a software, an android based application for the same.

II. ACKNOWLEDGMENTS:

The whole project of Weka tool was done as part of the WEKA project, at the University of Waikato. We would like to thank the members of the WEKA team, who made such a great Weka tool to analyze data sets. Really a great feature ahead to use for machine learning projects. Throughout my research work the support documentation by Weka team helped us a lot.

References:

[1] A. K. Jain and B. B. Gupta, A novel approach to protect against phishing attacks at client side using auto-updated white-list. EURASIP Journal on Information Security, 2016(9), 2016.
[2] N. Gupta, S. Arora Review paper SMS Categorization for Future Security and User Convenience http://www.jetir.org/papers/JETIR1908A97.pdf
[3] SMS, C, The real value of sms to businesses, 2018, https://www. smscomparison.co.uk/sms-gateway-uk/2018-statistics/. (Accessed March 2019).
[4] T.A. Almeida, J.M.G. Hidalgo, A. Yamakami, Contributions to the study of sms spam filtering: new collection and results, in: Proceedings of the 11th ACM Symposium on Document Engineering, ACM, 2011, pp. 259–262.
[5] C. Wang, Y. Zhang, X. Chen, Z. Liu, L. Shi, G. Chen, F. Qiu, C. Ying, W. Lu, A behavior-based sms antispam system, IBM J. Res. Dev. 54 (2010) 3–1.
[6] T. Yamakami, Impact from mobile spam mail on mobile internet services, in: International Symposium on Parallel and Distributed Processing and Applications, Springer, 2003, pp. 179–184.
[7] V. Gupta, A. Mehta, A. Goel, U. Dixit, A.C. Pandey, Spam detection using ensemble learning, in: Harmony Search and Nature Inspired Optimization Algorithms, Springer, 2019, pp. 661–668.
[8] Z. Chen, Q. Yan, H. Han, S. Wang, L. Peng, L. Wang, B. Yang, Machine learning based mobile malware detection using highly imbalanced network traffic, Inform. Sci. 433 (2018) 346–364.
[9] C. Amrutkar, Y.S. Kim and P. Traynor, Detecting Mobile Malicious WebPages in Real Time, IEEE Transactions on Mobile Computing (2016)
[10] J.W. Joo, S.Y. Moon, S. Singh and J.H. Park, S-Detector: an enhanced security model for detecting Smishing attack for mobile computing, Telecommunication Systems vol. 66(1), 29–38 (2017).
[11] M. Foozy, C. Feresa, R. Ahmad and M.F. Abdollah, A practical Case based technique by splitting SMS phishing from SMS spam for better accuracy in mobile device, International Review on Computers and Software, vol. 9(10), pp. 1776-1782 (2014).
[12] E. M. El-Alfy and Ali A. AlHasan, Spam filtering framework for multimodal mobile communication based on dendritic cell algorithm, Future Generation Computer Systems, vol. 64, pp. 98-107, (2016).
[13] Symantec Internet Security Threat Report, Available at: http://www.symantec.com/content/en/us/enterprise/other_resources/bistr_main_report_v19_21291018.en-us.pdf. Accessed August 2017
[14]Mobile messaging fraud report, Available at: https://mobileecosystemforum.com/mobile-messaging-fraud-report-2016/.
[15] Smishing Report, Available at : http://resources.infosecinstitute.com/category/enterprise/phishing/phishing-variations/phishing-variationssmishing/, last accessed 2017/07/15.
[16] The Social Engineering Framework, Available at: https://www.social-engineer.org/framework/attack-vectors/smishing/.
[17] J.W. Joo, S.Y. Moon, S. Singh and J.H. Park, S-Detector: an enhanced security model for detecting Smishing attack for mobile computing, Telecommunication Systems vol. 66(1), 29–38 (2017).
[18] M. Foozy, C. Feresa, R. Ahmad and M.F. Abdollah, A practical rule based technique by splitting SMS phishing from SMS spam for better accuracy in mobile device, International Review on Computers and Software, vol. 9(10), pp. 1776-1782 (2014).
[19] A. Tewari, A. K. Jain and B. B. Gupta, Recent survey of various defense mechanisms against phishing attacks. Journal of Information
[20] Dimensional Enterprise Mobile Security Survey, Available at: http://blog.checkpoint.com/wpcontent/uploads/2017/04/Dimensional_Enterprise-Mobile-Security-Sury.pdf.
[21] N. Choudhary and A.K. Jain, Towards Filtering of SMS Spam Messages Using Machine Learning Based Technique. Advanced Informatics for Computing Research, 18-30, 2017.
[22] A. K. Jain and B. B. Gupta, A novel approach to protect against phishing attacks at client side using auto-updated white-list. EURASIP Journal on Information Security, 2016(9), 2016.
[23] Mobile User Statistics referred from here https://www.statista.com/statistics/274658/forecast-of-mobile-phone-users-in-india/.
[24] Walter Daelemans; Antal van den Bosch (2005). Memory-Based Language Processing. Cambridge University Press.
[25] Stuart Russell and Peter Norvig (2003). Artificial Intelligence: A Modern Approach, second edition, p. 733. Prentice Hall. ISBN 0-13-080302-2
[26] Tom Mitchell (1997). Machine Learning. McGraw-Hill.
[27] D. Randall Wilson; Tony R. Martinez (2000). "Reduction techniques for instance-based learning algorithms". Machine Learning.
[28] Gagliardi, F (2011). "Instance-based classifiers applied to medical databases: Diagnosis and knowledge extraction". Artificial Intelligence in Medicine. 52 (3): 123–139.
[29] John G. Cleary, Leonard E. Trigg: K*: An Instance-based Learner Using an Entropic Distance Measure. In: 12th International Conference on Machine Learning, 108-114, 1995.
[30] https://www.cs.waikato.ac.nz/ml/weka/
[31] D. Aha, D. Kibler (1991). Instance-based learning algorithms. Machine Learning. 6:37-66.
[32] John G. Cleary, Leonard E. Trigg: K*: An Instance-based Learner Using an Entropic Distance Measure. In: 12th International Conference on Machine Learning, 108-114, 1995.
[33] Dayana C. Tejera Hernández University of the Informatics Sciences/Department of Software Engineering and Management, La Habana, 10800, Cuba
[34] S.J. Delany, M. Buckley, D. Greene, Sms spam filtering: methods and data, Expert Syst. Appl. 39 (2012) 9899–9908.

[35] K. Mathew, B. Issac, Intelligent spam classification for mobile text message, in: Computer Science and Network Technology (ICCSNT), 2011 International Conference on, vol. 1, IEEE, 2011, pp. 101–105.
[36] I. Androulidakis, V. Vlachos, A. Papanikolaou, Fimess: filtering mobile external sms spam, in: Proceedings of the 6th Balkan Conference in Informatics, ACM, 2013, pp. 221–227.
[37] International Journal of Pattern Recognition and Arti¯cial Intelligence Vol. 25, No. 8 (2011) 12191241 #.c World Scienti¯c Publishing Company.
[38] S. Theodoridis and K. Koutroumbas, Pattern Recognition, 2nd edn. (Academic Press, 2003).
[39] A. C. Lorena, A. C. P. L. F. de Carvalho and J. M. P. Gama, A review on the combination of binary classi¯ers in multiclass problems, Artif. Intell. Rev. 30 (2009) 1937.