Nigerian Journal of Technology (NIJOTECH)
Vol. 37, No. 3, July 2018, pp. **770 – 778**

Print ISSN: 0331-8443, Electronic ISSN: 2467-8821
www.nijotech.com
http://dx.doi.org/10.4314/njt.v37i3.28

AN ARCHITECTURAL-BASED APPROACH TO DETECTING SPIM IN ELECTRONIC MEANS OF COMMUNICATION

O. H. Odukoya[1,*], O. B. Adedoyin[2], B. I. Akhigbe[3], T. A. Aladesanmi[4] and G. A. Aderounmu[5]
[1, 2, 3, 4, 5] DEPT. OF COMPUTER SCIENCE & ENGINEERING, OBAFEMI AWOLOWO UNIV., ILE-IFE, OSUN STATE, NIGERIA
E-mail addresses: [1] *oodukoya@oauife.edu.ng*, [2] *oyindamolaadedoyin2@gmail.com*,
[3] *biakhigbe@oauife.edu.ng* , [4] *taladesanmi@oauife.edu.ng*, [5] *gaderoun@oauife.edu.ng*

ABSTRACT
Spams are what users and developers should be aware of in all Internet-based communication tools (such as e-mail, websites, Social Networking Sites (SNS), instant messengers and so on). This is because spammers have not ceased from using these platforms to deceive and lure users into releasing vibrant and sensitive information (especially, financial details). This paper developed an architectural based technique for SPIM (Instant Message Spam or IM SPAM) detection using the classification method. The classification was done using the C4.5 classifier with a dataset of messages gotten from an instant messaging environment. The dataset served as the input to the classification algorithm method which was able to distinguish spam from non-spam messages. This classification method was depicted in a tree form to show its usefulness. The results show that its precision, recall and accuracy rate satisfied standard recommendation with a commendable error rate. The proposed technique will find implication in the reduction of the number of Internet users.

Keywords: *Social Networking sites, spammers, Instant message spam, C4.5 Classifiers, e-mails.*

1. INTRODUCTION
The Internet has revolutionized the way we communicate. Electronic mail (E-mail) has been the most rapid adopted form of communication ever known. Less than two decades ago, not many people had heard of it. Now, many of us use e-mail instead of writing letters or even calling people on the phone. People around the world send out billions of e-mail messages every day. Even though e-mails are still very much in use, there are other means of communicating electronically which even tends to be faster than e-mails. This technology is called Instant Messaging (IM). IM is a typical communication tool, which enables end-users to send messages in a one-to-one or one-to-many way in near real-time with awareness of each other's presence. Presence in this context refers to the realization that some principal thing is present or not on the network. IM allows effective and efficient communication. It allows the immediate receipt of acknowledgment or reply. Many IM services allow video calling features, voice over IP and Web conferencing services. Due to IM's unique features real-time nature and presence, the rate at which IM grows is extremely high. In a brave new smartphone and in an obsessed world that we live in, more and more people are sending messages through these IM applications such as Web Chat, AOL, Yahoo, WhatsApp, Snapchat, Viber, Skype, Facebook, Messenger, Live Chat, Hip Chat and so on across the world. IM apps are surging in popularity. In China alone, the number of mobile IM accounts approached 1.5 billion in 2013. As IM is more and more widely deployed, these and many more IM applications have fragmented IM user market. Not only does IM gain popularity among home or business users, it is also exposed to some security risks such as spamming. Spams are unsolicited messages ranging from advertisements and solicitations to jokes and chain letters. They are usually unwanted messages and users in general dislike being subjected to such messages. The most common form of spam that is recognized is e-mail spam (SPAM). As a matter of fact, spammers have started to expand their battle field from e-mail to IM. Instant Messaging Spam (SPIM) which is a type of spam targeting users of instant messaging services. Although, less ubiquitous than it's e-mail counterpart 500 million spam IMs were sent in 2003, which is twice the level in 2002 [1]. IMs are not usually blocked by firewalls. This makes them an especially useful channel. In order to solve this security risk, which is faced by IM, there is need to protect IMs

*** Corresponding author, tel: +234 – 813 – 948 – 4145**

against instant message spam (SPIM). This protection can be built into IM systems so that IM will not end up surrendering to spammers. The purpose of this paper is to develop a protection model, which is architectural-based and it is capable of detecting spam in an IM system. The remaining part of this paper is arranged as follows; Section 2 discussed the related work in SPIM detection, Section 3 described the methodology and simulation, while the result and evaluation are discussed in Section 4. Finally, the conclusion is presented in Section 5.

2. LITERATURE REVIEW

Information security research communities have proposed SPIM removal by handling botnet and worms since IM is mostly used to deliver their campaign messages and change point detection. Virus Throttling [2] are some of the known techniques. In another work [3] provided a characterization of spam traffic using work- load variation, density, inter-arrival time distribution, e-mail size distribution, temporal locality, which were compared with non-spam e-mails. The work showed that non-spam e-mail transmissions were driven by bilateral social relationship, while spam transmissions were usually unilateral actions that are based on the spammers' will to reach a large number of recipients.

In the work of [4], a group-based anti-spam framework was proposed. The work that investigated the clustering structures of spammers based on spam traffic that was collected at a domain mail server. The study showed that the relationship among spammers demonstrated a high clustering structure based on URL grouping.

For [5] a new architecture for real-time SPIM defense and filtering is needed in a personalized setting and for various IM gateways [2]. In the work, a number of filtering methods that include collaborative feedback based filtering, content-based technique, challenge-response based filtering, IM sending rate, content-based SPIM defending techniques and so on were tested. After a number of experiments, they reported that the blacklisting spammers based on user feedback produced the most efficient blocking technique with least error rate.

In another related work by [6], a neural network-based system for automated email classification was formulated. The work also presented a linger technique with a neural network approach that automatically categorized emails and filters them into mailboxes. However, the work was not oriented towards real SPIM detection.

The work in [7] published a technique to distinguish BOTS from human users by mining user characteristics. The features used in the dataset were word-length, message length, URLs, Capital and small letters. However, the work implemented SPIM detection at the user end while ignoring vital architectural considerations. In a recent work by [8], data mining approaches were used to classify SMS spam and their performance was compared to gain insight and further explore available weaknesses. The simulation results showed that a multinomial naive Bayes; C4.5 and SVM with linear kernel were among the best classifiers for SMS spam detection. The work provided sufficient insight on how to use the C4.5 technique for classification.

In [9] the research work used a neural network method with a data set of email messages that captured single users. A descriptive characteristic of words and messages that are similar to those required to identify spam messages were used as a feature for defining spam messages. A total corpus of 1654 emails was used and they were received over undisclosed number of months. Results of comparisons between their technique and that of Naïve Bayesian technique showed that their technique only needed additional features to achieve a better result than the result from the neural network approach.

In [10] an anti-spam filtering model for agglutinative languages specified for Turkish was developed. They used a dynamic technique based on Artificial Neural Networks and Bayesian Networks with a user-specific algorithm. The algorithm adjusts itself to the characteristics of incoming e-mail messages in order to detect the spam. A total of 750 emails including 410 spams and 340 hams were used in the experiments with a success rate of approximately 90%.

SPIM classification is a challenging task for IM application providers, especially in telecommunication systems. This is due to its architecture, which is peer to peer and the use of heterogeneous communication protocols. Attempts at solving the problem have been done by trying to take advantage of interoperability, usability, Quality of Service (QoS) concerns and other secured solutions as suggested in Swagata [2]. This paper approached its solution-based technique provision through a multilayered architecture based technique. The contribution made in this paper are unlike the ones reviewed so far; aside being an architecture based approach it is useful for many scenarios. The implication of this is that SPIM will be detected in any of the one to one or one to many and in near real-time scenarios.

3. METHODOLOGY

The C4.5 classifier is a standard data mining classifier that was simulated in the WEKA (Waikato Environment for Knowledge Analysis) environment. In this paper the set of data used is a standard data set. The classification testing was done using the C4.5 classifier model within the WEKA simulation environment following the practice of [11] which proposed the C4.5 algorithm and was leveraged in this paper in an architecture based manner as shown in Figure.1 following the postulation in [11].

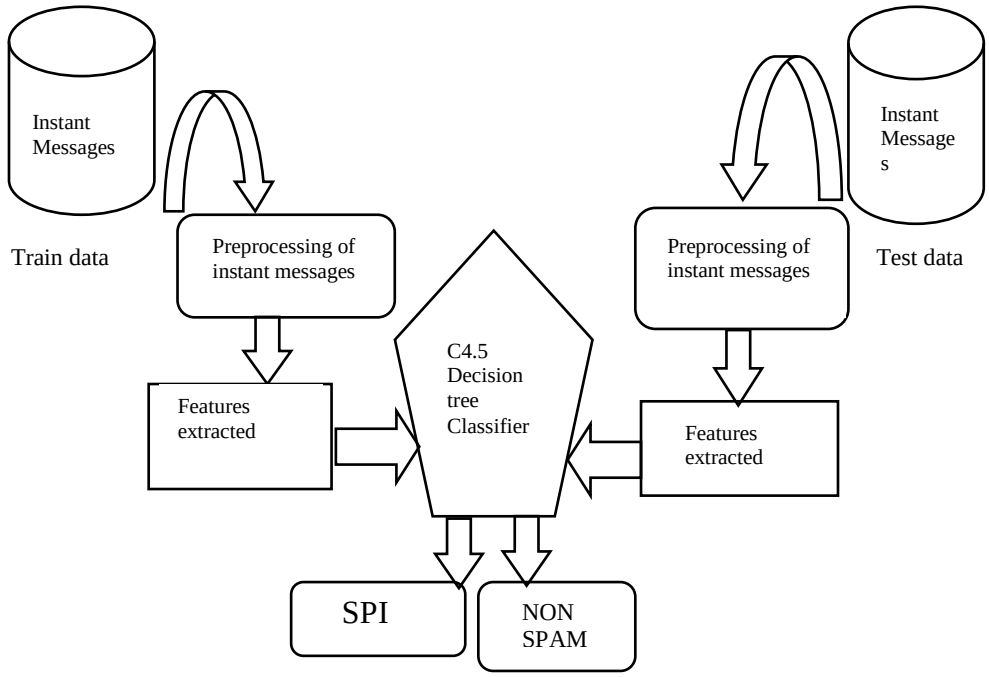

Figure 1: The standard architectural based SPIM detection method using the C4.5 algorithm

1	Go until jurong point, crazy.. Available only in bugis n great world la e buffet... Cine there got amore wat...	0
2	Ok lar... Joking wif u oni...	0
3	Free entry in 2 a wkly comp to win FA Cup final tkts 21st May 2005. Text FA to 87121 to receive entry question(std txt rate)T&C's apply 08452810075over18's	1
4	U dun say so early hor... U c already then say...	0
5	Nah I don't think he goes to usf, he lives around here though	0
6	FreeMsg Hey there darling it's been 3 week's now and no word back! I'd like some fun you up for it still? Tb ok! XxX std chgs to send, Â£1.50 to rcv	1
7	Even my brother is not like to speak with me. They treat me like aids patent.	0
8	As per your request 'Melle Melle (Oru Minnaminunginte Nurungu Vettam)' has been set as your callertune for all Callers. Press *9 to copy your friends Callertune	0
9	WINNER!! As a valued network customer you have been selected to receivea Â£900 prize reward! To claim call 09061701461. Claim code KL341. Valid 12 hours only.	1
10	Had your mobile 11 months or more? U R entitled to Update to the latest colour mobiles with camera for Free! Call The Mobile Update Co FREE on 08002986030	1
11	I'm gonna be home soon and i don't want to talk about this stuff anymore tonight, k? I've cried enough today.	0
12	SIX chances to win CASH! From 100 to 20,000 pounds txt) CSH11 and send to 87575. Cost 150p/day, 6days, 16+ TsandCs apply Reply HL 4 info	1
13	URGENT! You have won a 1 week FREE membership in our Â£100,000 Prize Jackpot! Txt the word: CLAIM to No: 81010 T&C www.dbuk.net LCCLTD POBOX 4403LDNW1A7RW18	1
14	I've been searching for the right words to thank you for this breather. I promise i wont take your help for granted and will fulfil my promise. You have been wonderful and a blessing at all times.	0
15	I HAVE A DATE ON SUNDAY WITH WILL!!	0
16	XXXMobileMovieClub: To use your credit, click the WAP link in the next txt message or click here)) http://wap. xxxmobilemovieclub.com?n=QJKGIGHJJGCBL	1

Figure 2. The raw instant message training dataset downloaded from http://dcomp.sor.ufscar.br/talmuda/smsspamcollection

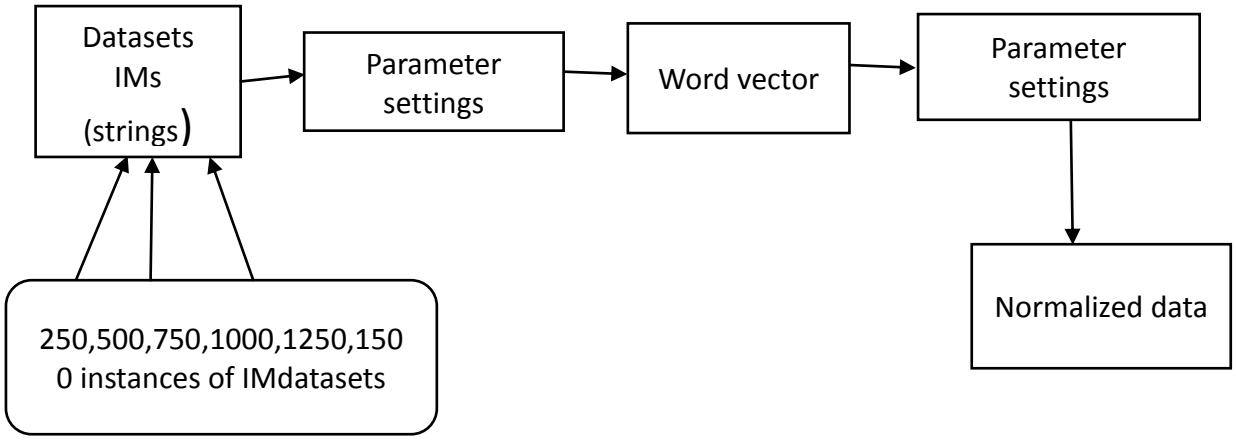

Figure 3: The workflow diagram to convert the training data from string in its normalized form

The dataset used in this work is a standard dataset. It is known as the SMS Spam Collection. It is a public set of data of 5574 SMS (Text) labeled messages. It has a dataset that is composed of real and non-encoded messages that are tagged as legitimate (ham) or spam.

WEKA was used to simulate the detection and protection model. The model was formulated using C4.5 data mining technique. In the Weka environment, the model was trained by 60% of the classified dataset, while the remaining part was used for testing. A sample (or snap shot) of the datasets that were used is as shown in Figure 2. In order to classify the message instances as spam and/or non-spam contents, the decision tree aspect of the C4.5 classifier was applied in the WEKA 3.8 environment on the dataset.

The dataset that was used was in the form of a string (text), which could not be analyzed directly using the C4.5. For the dataset to be analyzed using the model, the dataset was first preprocessed into an analyzable format. In order to simplify the process that was applied to the datasets, the workflow presented in Figure 3 was adopted. A screenshot of the output of the workflow diagram which is the normalized data (nominal dataset) is as shown in Figure 4.

Figure 5 shows the 'pre-processed' panel with the data set loaded into it. The panel showed the instances of the data, the class and attributes. The pre-process panel is an environment that allows for the editing of the dataset, to save the dataset in different formats as required and undo any process when necessary.

Figure 6shows the attributes of the training dataset. A result, each word in a message or sentence in the dataset that was sent was taken to discover the degree of spamminess, which is between 0 and 1 as the standard threshold.

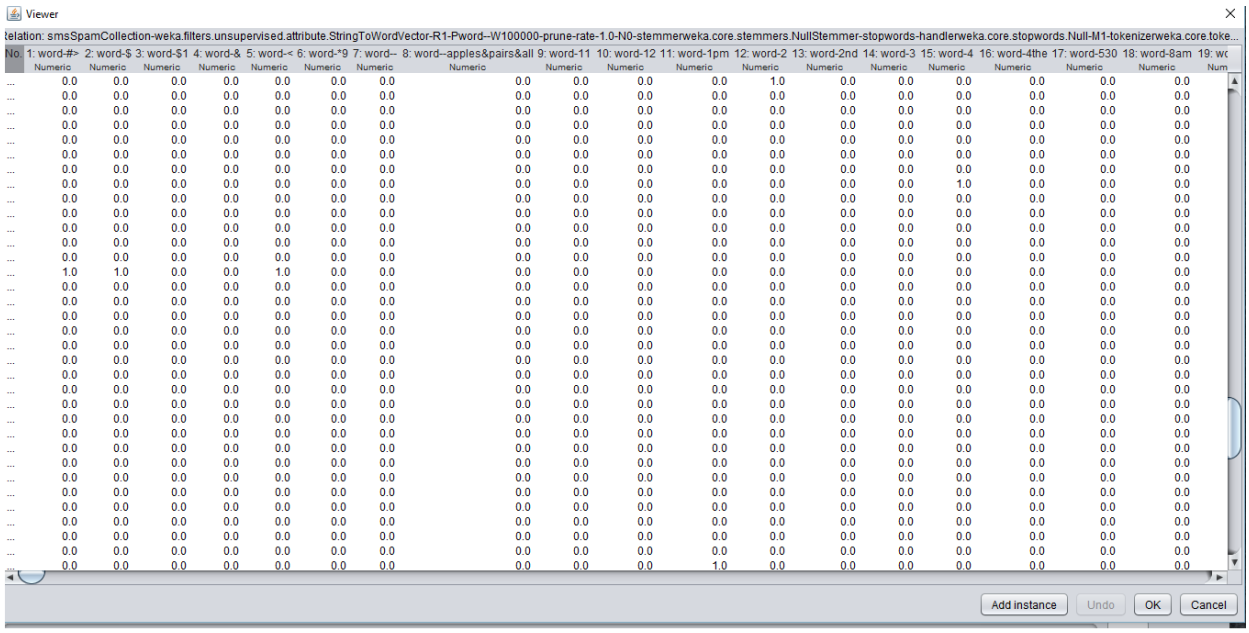

Figure 4: The normalized instant message dataset

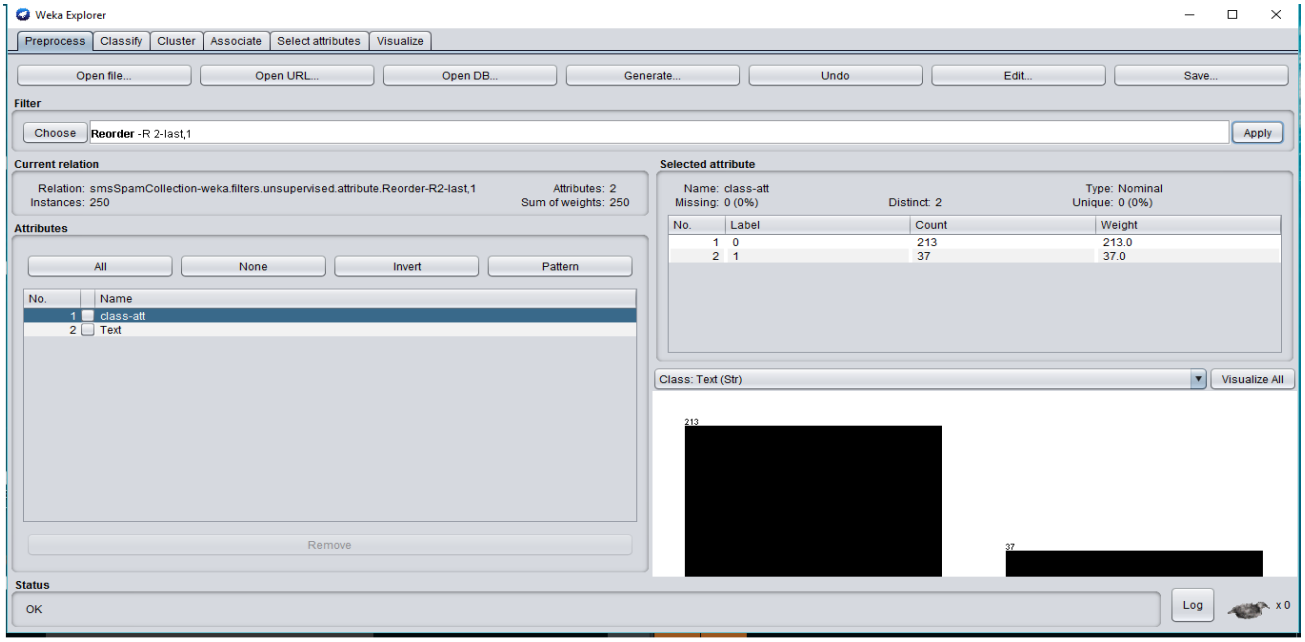

Figure.5 Pre-process panel interface with attributes and classes

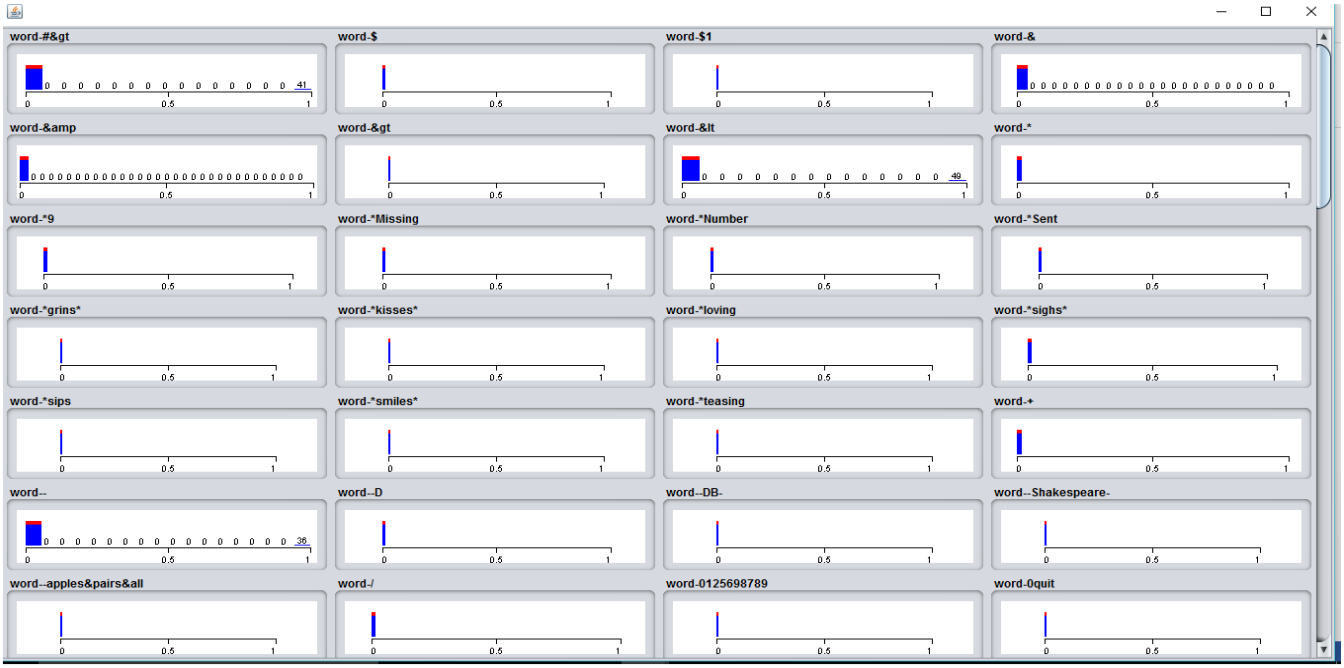

Figure 6 WEKA Interface showing the Attributes of the training dataset

Using the C4.5 Classifier, an algorithm for building, decision trees were developed. The texts used for the classification, training, and testing did not need to be changed to obtain a good performance. The evaluation of the performance was carried out using the training data, which was loaded at the preprocessing stage. Then, the 'Use a training set' from the test options part of the "Classify panel" as shown in Figure 6 was activated. The classifier was therefore built and evaluated. This processed the training set using the C4.5. Then it classified all the instances in the training data. The corresponding output performance statistics are shown in Figure 7.

The result of training the proposed model is shown in Figure 7 in the form of a decision tree as presented in Figure 8.

The same process of loading the training dataset and classification was repeated based on standard practice to test the data as shown in Figure 9 and the output is tabulated as presented in Table 1. This was also done for the testing dataset, but at the 'classify' panel, here the 'supplied test set' was chosen instead.

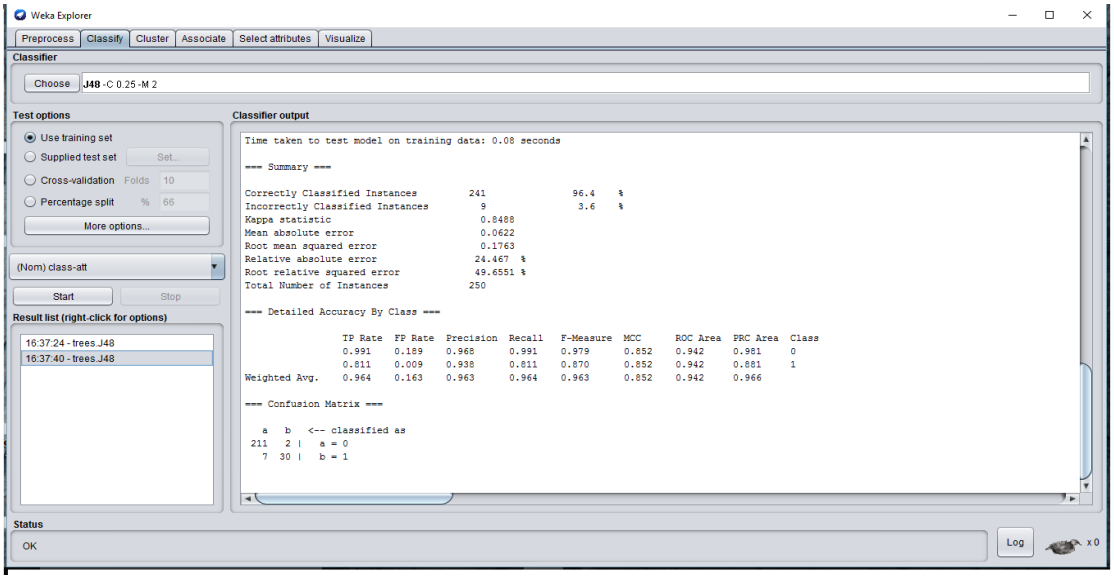

Figure.7 Output interface of the training dataset

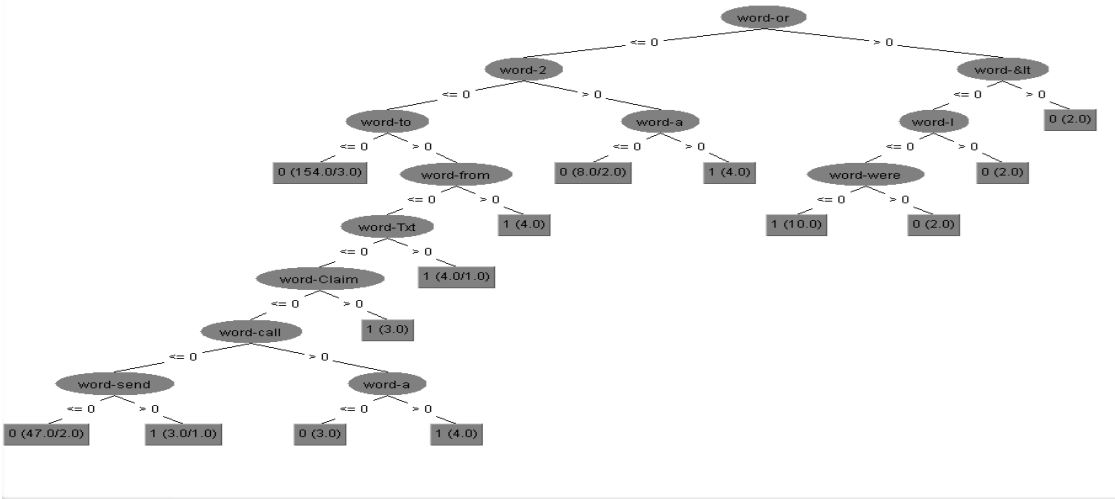

Figure.8 The decision tree output interface of the training dataset

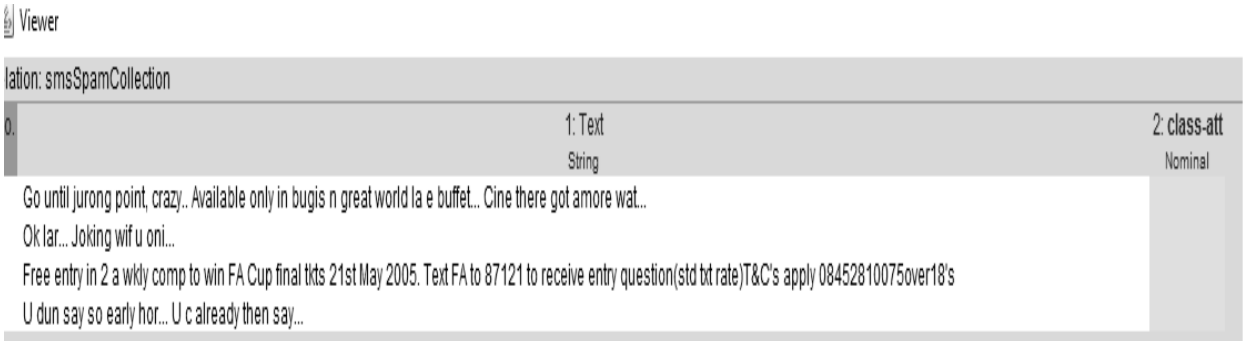

Figure 9: Sample Test data

Table.1 Classification of the test data

S/N	Text (string)	Classification as a spam
1	Go until jurong point, crazy.. Available only in bug is n great world la e buffet... Cine there got amore wat...	No
2	Ok lar... Joking wif u oni...	No
3	Free entry in 2 a wkly comp to win FA Cup final tkts 21st May 2005. Text FA to 87121 to receive entry question(std txt rate)T&C's apply 08452810075over18's	No
4	U dun say so early hor... U c already then say...	No

4. RESULT AND DISCUSSION

4.5 Results Evaluation

The simulation process that was followed in the work reported in this paper used estimators such as:

True Positive (TP) rate, which was used to establish instances of correctly classified data as a given class while,

False Positive (FP) rate is used to show (instances) of falsely classified data as a given class.

Precision is the proportion of instances that are truly of a class divided by the total instances of the classified dataset was leveraged. The equation for precision is given by;

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (1)$$

Recall is described as the proportion of instances that are classified as a given class divided by the actual total in that class (equivalent to TP rate). The formula for recall is given by;

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (2)$$

Accuracy was used to measure the degree of correctness of the classifier. The accuracy is easily derived from a confusion matrix, where the TP and the TN were summed up over all other parameters in the matrix as seen below;

$$\text{Accuracy} = \frac{TP + TN}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad (3)$$

Error rate was used to signify the degree of deviation from the correctness of a classified results/output. This was measured as error rate as shown in equation (4)

$$\text{Error Rate} = \frac{FP + FN}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad (4)$$

A Classification time was simply introduced to show how long it took the system to classify an inputted dataset. While training the C4.5 classifier, six (6) iterations of the datasets were loaded into WEKA. These iterations are seen in Table 2, which shows a summary of the results obtained based on some of the key model estimators.

The result from the training session revealed that the model has high accuracy ability with an accuracy measure between 95.80 and 97.00%. A graph of the generated accuracy was plotted against the number of instances as shown in Figure 10. The graph generates a line that looks so straight and very close to 100%.This accuracy measure is very reliable in binary classification, since it is very close to 100%. Likewise, the precision, true positives, that were generated for all six (6) instances were more than 95.00%. The error rate generated was very reliable for binary classification since it amounted to a very small percentage that is as low as 5.61% and 6.88%.

A histogram of the generated accuracy, precision, TP rate, FP rate and error rate of all the number of instances are also presented as depicted in Figure 11.

The classification time that was generated from all instances of the datasets shows that the more the number of instances, the longer time it took the proposed model to make classification into spam and non-spam. A graph showing the classification time against the number of instances is also shown in Figure 12. The graph shows a progressing line that increases as the number of instances increases.

5. CONCLUSION

In this paper, a model that uses an architectural based approach and a data mining technique (C4.5 algorithm) is presented, to show how the rate of SPIM invasion can be reduced in electronic communication. The architectural model presented in this paper, when implemented will go a long way in: (i) reducing the security risks that could be caused by password stealing, Trojans and other threats in a workplace where Instant Messaging applications are used; and the (ii) reduction of network bandwidth consumption due to the reduction of malicious content that consumes the available bandwidth.

The limitation of this work is with respect to the learning classifiers used which showed an ability to learn but none of these could show 100% predictive accuracy. In the future, we intend to look at deceptive suspicious messages in other formats other than text and also take care of cases with encrypted suspicious messages.

6. ACKNOWLEDGEMENT

This research is funded by Africa Center of Excellence in Software Engineering at the Obafemi Awolowo University, Ile-Ife. Nigeria.

Table 2: Iterated results on training dataset

Number of instances	Accuracy (%)	Precision (%)	True positive rate (%)	False positive rate (%)	Error rate (%)	Classification time(seconds)
250	96.40	96.30	96.40	16.30	6.22	0.03
500	97.00	96.90	97.00	13.40	5.61	0.10
750	96.53	96.50	96.53	16.80	6.45	0.17
1000	95.80	96.80	95.80	14.60	7.75	0.28
1250	96.30	96.30	96.30	18.60	6.88	0.49
1500	96.33	96.33	96.33	19.10	6.78	0.55

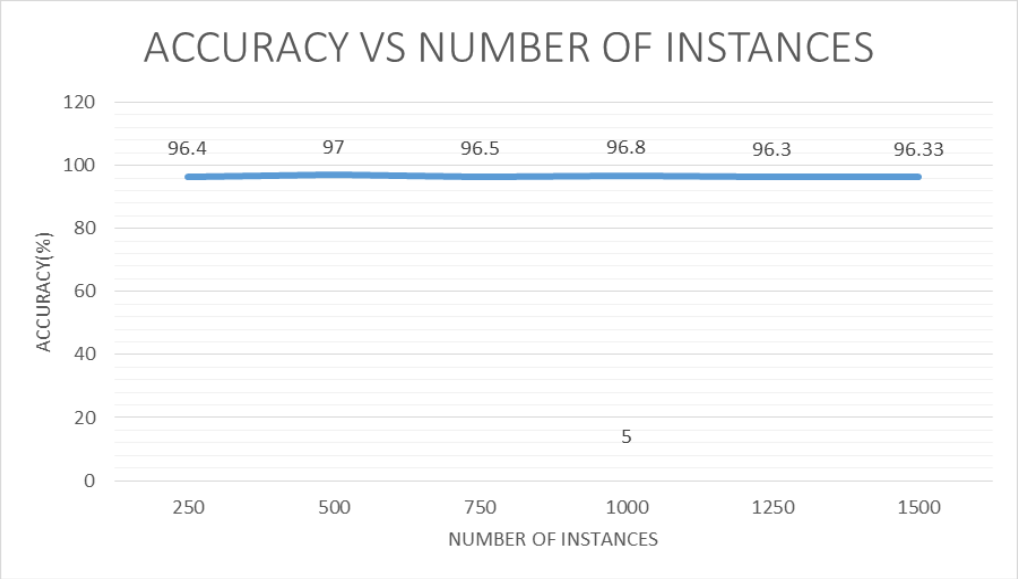

Figure 10: Graph showing the accuracy of the classifier against the number of instances of the training data

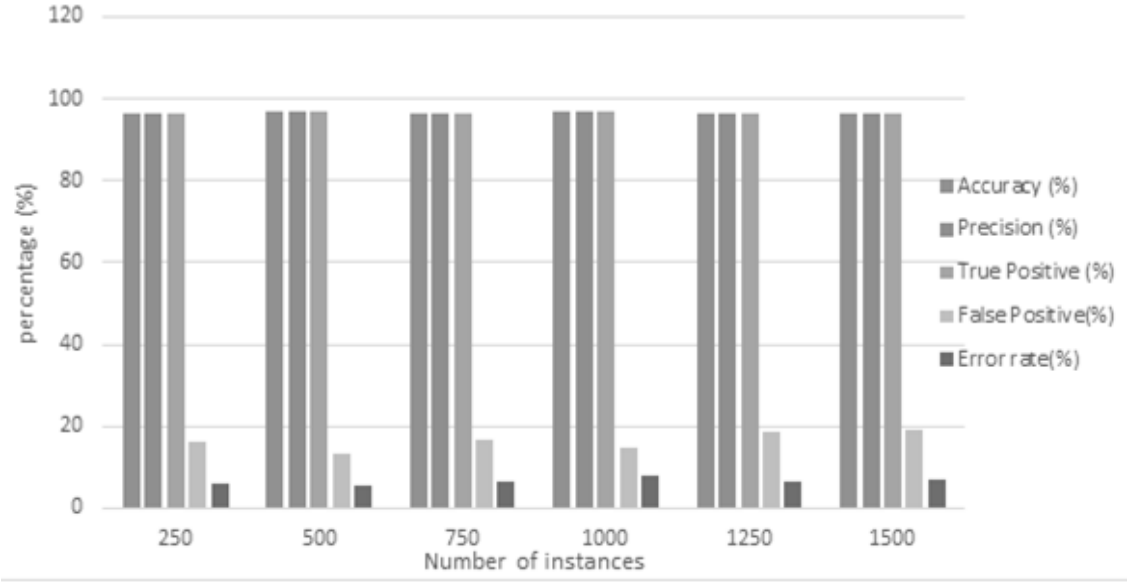

Figure 11 Graph showing the Accuracy, precision, TP rate and Error Rate of the training dataset.

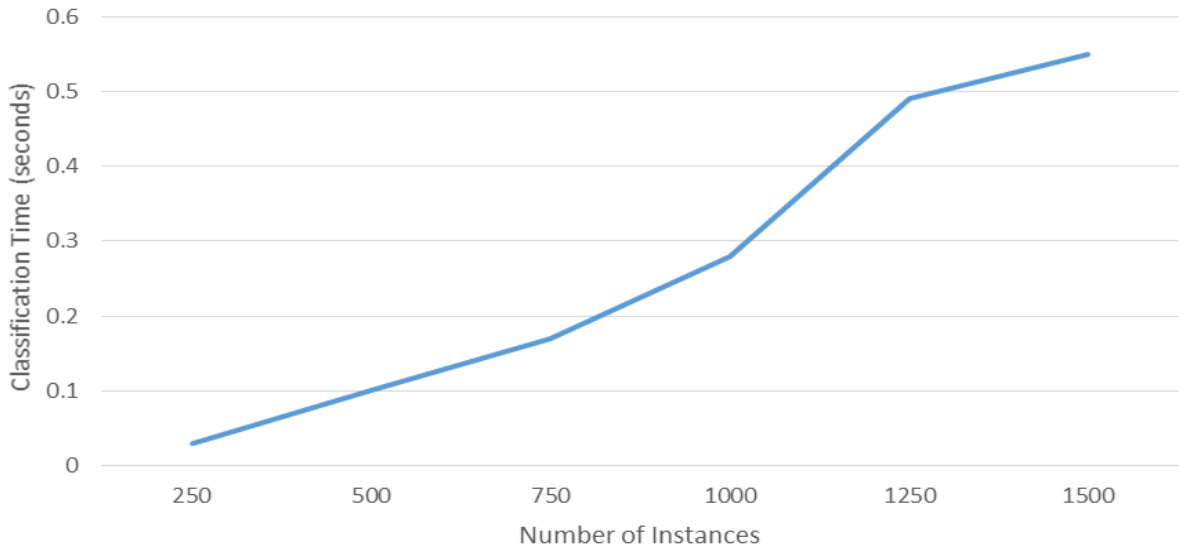

Figure.12: Graph showing the classification time against the number of instances.

7. REFERENCES

[1] Ferris Research Report, Spam Control: Problems & Opportunities'. http://www.ferris.com. Accessed on May 10,207, 2003.

[2] Swagata, D. 'Instant Messaging Spam Detection In Long Term Evolution Networks'. Published M.sc Thesis, Submitted to the Faculty of Engineering and Computer Science, Concordia University, Canada, 2013.

[3] Gomes, L. H., and Cazita, C. 'Characterizing a Spam Traffic. *In the proceeding of IMC' Internet Measurement Conference*, October 25-27, Taormina, Sicily, Italy.2004.

[4] Li, F. and Hsieh, M. 'An Empirical Study of Clustering Behavior of Spammers and Group-based Anti-Spam Strategies, CEAS 2006 *Third Conference on Email and Anti-Spam*.' Mountain View, California USA. 2006

[5] Liu, Z., Lin, W., Li, N. and Lee, D. "Detecting and filtering Instant Messaging Spam – a global and personalized approach. *International Conference on Secure Network Protocols,* Boston, Massachusetts, USA, November 6, 2005.

[6] James, C., Irena, K., and Josiah, P. A. 'Neural Network Based Approach to Automated Email Classification'. *Proceeding of the 2003 IEE/WIC International Conference on Web Intelligence* pg 702, October 13-17, Halifax, 2006, Canada.

[7] Maroof, U. " Analysis, and detection of Spin using Message Statistics", *International Conference on Emerging Technologies (ICET),* Islamabad, Pakistan, October 18-19, 2010.

[8] Shirani-Mehr, H. " Sms Spam Filtering detection using Machine learning approach. CS229

[9] Levent, O., Tunga, G., and Fickert, G."Adaptive anti-spam filtering for agglutinative languages": special case for Turkish, *Journal of Pattern Recognition Letters*. Volume 25, Issue 16, pg 1819-1831, 2004.

[10] Stuart, I., Sung-Hyuk, C. and Charles, T. " A Neural Network Classifier for Junk Email". *Proceedings of Student/ Faculty Research Day, CSIS, Pace University*, London. 2004.

[11] Erritali, M., Hasina, B., Merbouha, A. and Ezzikouri, H. "A comparative study of decision tree ID3 and C4.5". *International Journal of Advanced Computer Science and Applications*, Special Issue on Advances in Vehicular Ad Hoc Networking and applications, pp13-19, 2011.