Spam Filtering With Fuzzy Categorization In Intelligent Email Responder

Najma Hanif [1], Mukaram Khan [2] , Sami Ullah Javaid [3], Babar Abbas [4], Amna Altaf [5], Malik Abdul Sami [6]

Abstract-

Communication through emails is the simplest and most consistent way of communication. Emails are used for fast and reliable communication at both personal and organizational levels, including academic institutions. Some organizations have deployed auto email responders to deal with a heavy volume of emails by auto responding to relevant routine mails while filtering out spams. Spammers send spam emails for hacking, phishing, denial of service or broadcasting marketing emails. There are various ways to identify spam emails. We propose a fuzzy logic based intelligent spam filtering technique as part of an auto email responder. Spam dictionary is created with ranked spam words, phrases, hyperlinks etc. Fuzzy rules are applied to categorize emails into spams and hams. The level of threat is identified by the matching words and phrases contained in email with the help of spam data dictionary. Our Model has been trained and tested on two data sets - CSDMC2010_SPAM (publicly available) and Strafford256 (a set of 256 real emails provided by Stratford University for this research).

Keywords: auto email responder, spam filtering, fuzzy logic, ranked dictionary

1. INTRODUCTION

Email is one of the most effective means of communication [1]. Emails are used to communicate all around the world for being a reliable and free means of communication. Depending of the social circle around a person, an individual may receive tens of emails every day. The number may rise to hundreds or thousands if it comes to an organizational service desk email address. A university teacher, teaching 2-3 classes of over 50 students in a semester, along with his present and past research students, is expected to receive over 100 mails a day, besides his other routine mails from friends, colleagues, admin staff, newsletters, conference intimation mails, mails from various literary forums, and from so many other groups. More spices are added to this by broadcast advertisements and other unwanted emails which we collectively may categorize as spam emails. It is humanly impossible to go through such a volume of emails or even scan through them on a daily basis. Email providers use certain filtering tools to categorize spams as “Junk Mail” in order to prevent readers from getting distracted from other relevant emails. These filters mostly work on email subject and sender’s address while a few provide sophisticated features to filter on the basis of contents as well. The fast and growing communication with emails requires an auto email responder for large corporate businesses and universities to deal with millions of emails on a monthly basis, that responds to routine emails with a typical stereotype responses. To take on this gigantic job in the humanly manner, a cognitive machine learning mechanism is necessary.

This research is part of a large project that involves developing an intelligent email response system to facilitate university lecturers in dealing with hundreds of routine query emails on daily basis. Intelligent response systems based on artificial intelligence have been developed for this purpose. It is challenging for an intelligent system to respond correctly in the presence of spam emails. Spam emails contain attachments, links and images full of malwares. These also flood the inbox of the receiver [2, 3]. Annual statistics on spam reports that the average user gets more than 50% spam emails. It is also reported that the digit goes from 50 to 150 billion emails sent as spam on daily basis [4, 5].

However, there are limitations with these techniques which hold us from obtaining satisfactory results in our automatic response systems.

Fuzzy logic deals with fuzzy sets that allow a degree of membership in terms of imprecisely defined values characterized by the degree of ranks [6]. Fuzzy logic is one of the flexible designs that is used in uncertain systems where is datasets are based on hazy values, such as spam filtering of emails [7], as it behaves exactly like human beings in its decision making process.

The proposed system is built on categorization of emails by applying fuzzy logic. The fuzzy rule-based system is used to distinguish an email into spam or ham. Features used in this work are extracted from emails that include sender's address, subject text, contents text and hyperlinks. Extracted features are compared against the spam data dictionary, which contains the spam sender's address, words, phrases and hyperlinks with ranks assigned with the level of threats. Email features are treated as input for fuzzification and rules are applied to predict the email as weak, moderate, strong or highly strong spam.

This paper is divided into following sections: Section 2 explains the review of literature, Section 3 explains the proposed work, Section 4 describes the results and Section 5 concludes the paper.

2. LITERATURE REVIEW

Conventionally Naïve Bayesian spam filters are considered as the simplest to implement and the most effective [8] as these work on mathematical rules and find the probability of the words to detect the spam. These are used mostly for content-based filtering, however, spammers make intentional spelling mistakes to fool the filtering process easily [9]. Moreover, it is computationally heavy and entails slow processing. A technique used by Varghese et al. [10] filters spam emails based on feature selection in four categories and eliminating the rare features on Naïve Bayesian score. They get information gain by feature selection method, construct TF-IDE weighted feature occurrence matrix, decompose it to singular value and finally generate a prediction model which provides better results as compared to [8].

Similarly, Weighted Naïve Bayesian (WBN) classifier is used in [11, 12] for subject-based spam filtering, which checks the subject of the email only for spam filtering and uses natural language expressions.

The model proposed in [13] focuses on detecting spams created by text modifications using Naïve Bayesian classifier to detect spam emails. System gathers keywords based on machine learning and semantic based algorithms to increase the detection rate and accuracy. Relationship between the spam score and email length has also been used to handle the Bayesian Poisoning, e.g. in [14], which discovers that intelligent systems with Naïve Bayes are strong on precision and categorization but weak on self-learning and self-adaptability, as compared to Artificial Immune System.

Mixing of different approaches gains value as it helps to detect the spam at different levels [15]. [16, 17] demonstrate a more effective mechanism by mixing Bayesian and SVM to filter spam. IP address black and white listing technique mixes various models from email address black and white lists, Real-time Blackhole Lists and Open Relay Database Lists, email DNS check, MIME header blocking, content filtering using words, phrases, wildcards, and regular expressions etc. [18] Youn et al.

[19] compare various email classification models, such as Neural Network, SVM, Naive Bayesian and J48 etc. They found J48 and NB classifiers showing better results as compared to SVM and NN classifiers.

In fuzzy logic based spam filtering, [20] proposes a model built upon fuzzy logic to detect spam mail while [21] demonstrates the classification of spam from ham on word ranking and fuzzy rules by building a database of words with ranks by which it distinguish content of spam emails by degree of threat. Similarly, [22, 23, 25] present adaptive fuzzy logic based model for spam detection, the functionality of model is improved with machine learning. In [24] fuzzy logic is used to reduce the uncertainty by identifying vague and ambiguous terms such as near, far, more, less etc. In [25] interactive human like inferencing and control systems is developed using fuzzy logic rules.

[27, 28] developed fuzzy logic-based spam detection models which use rules against five parameters i.e. sender address , sender IP , subject text , content text and attachments to be compared against the white lists. VSM and fuzzy logic based classifier has been developed in [29], while [30] proposes spam identification model called Disclosed Herein. Similarity measured hash and a sender hash are produced to detect spam for each email. Two or more previous emails are compared with a newly received email and these rules applied to specific match for a possible spam email.

All the above discussed methods are not built for intelligent response systems which require instant and precise decisions, to respond like human being. In our proposed spam filtering model for intelligent email response system, we developed a model using fuzzy logic, which uses various parameters for filtering based on user address, email subject, contents of emails, and hyperlinks. We also use phrase-based filtering using word. It also categorizes users in white and black lists for future decision making.

3. PROPOSED MODEL

The proposed model is highly suitable for real-time intelligent email responder as it counts on the sender's address, email subject and email contents as input for categorization using fuzzy logic based approach to filter out spam emails as output. Figure 1 shows the email categorized as spam after fuzzification.

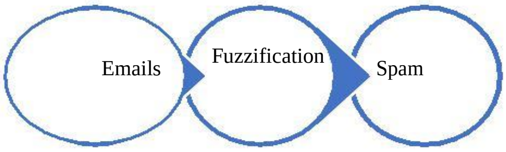

Figure 1: Email processing

Three major parameters are extracted from email as shown in Figure 2.

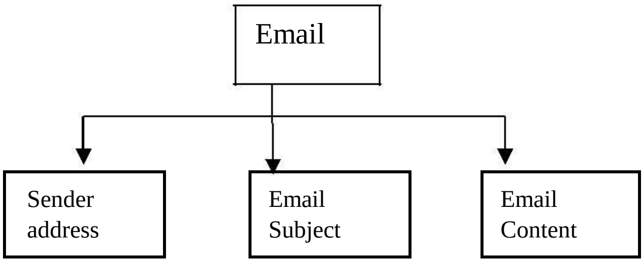

Figure 2: Feature of email for proposed model

Email categorization is divided in two main phases. In the first phase, the spam emails are filtered out based on sender's address included in the black list. Black list is initially populated from well-known spamming lists worldwide. White list includes the legitimate users that are registered students and faculty of university. New users are treated as spam or ham based on fuzzification process on subject and content of the email, as shown in Figure 4.

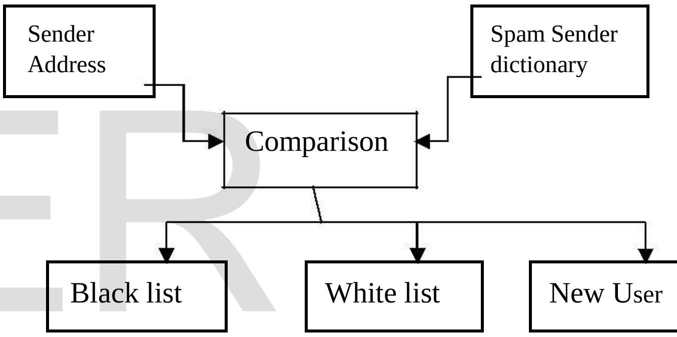

Figure 3: List assignment in proposed model

Second part of model focuses on textual categorization of subject and contents of email by using fuzzy rules. The text of email contains words, phrases and hyperlinks. Unsolicited mail contains spam words, spam phrases and spam hyperlinks which are assigned diverse values by using fuzzy rules. The values assigned to spam words with level of threat associated to that word or phrase in spam dictionary of system. Fuzzy rules are applied to detect spam emails as shown in Figure 4.

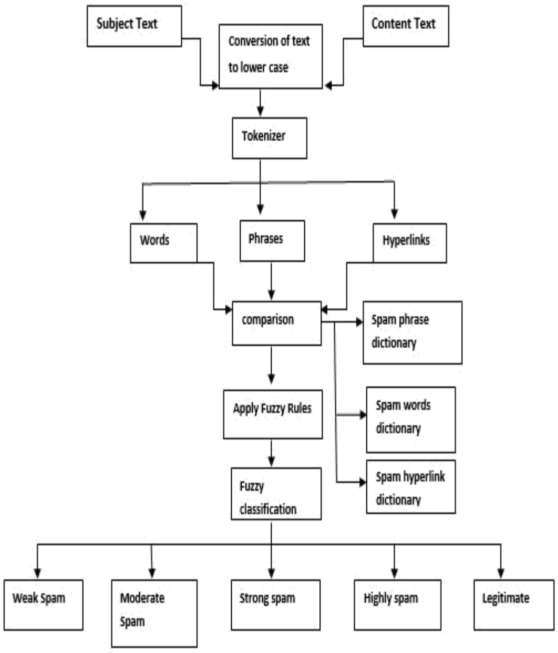

Figure 4: Subject and contents text filtering by Fuzzy Logic Based Intelligent Response System

Spammers usually use upper and lower cases to deceive the anti-spam systems, so the proposed model converts all the text into lower case before processing text. Words, phrases and hyperlinks are tokenization to feed the system as input before comparing against the spam dictionary. The matched inputs are further applied with fuzzy rules to classify the mail as weak spam, moderate spam, strong spam or legitimate email.

3.1 CATEGORIZATION MODEL BASED ON FUZZY LOGIC

Fuzzification process used in this model consist of rules based on partial logic [22]. The system contains a dictionary of users with ranked values to black list and white list. It also contains a dictionary having words, phrases and hyperlinks with assigned spam ranking. Ranks are the mathematical values assigned to spam dictionary data. On the basis of these, emails are placed into different categories. The system has four spam categories:-

- □ weak spam
- □ moderate spam
- □ strong spam
- □ highly strong spam

The implemented system is tested on two data sets as:-

- □ A publically available dataset (CSDMC2010) containing 4000 spam emails.
- □ A custom-build dataset consisting of 256 real emails provided by the University of Stratford for this purpose (Stratford256).

The spam dictionary has been built on the basis of spam words in CSDMC2010 dataset. Values assigned to the words and phrases are based on the degree of threat e.g. the words and phrases like ‘Congratulation’, ‘You won cash price’, ‘you won’, ‘lottery’, ‘Get free tour’, ‘Award’ etc are likely contents of spam emails catching our eyes quickly and luring us into an unwanted situation. These types of words are placed in the highly-strong spam category. Hyperlinks are also included in the emails contents text asking the recipient to subscribe to the websites in different emails. Subscription links are provided to get notifications if clicked. Hyperlinks, such as ‘Get now’, ’don’t miss a chance’, ‘click me’, Subscribe now’, ’here’ etc are also ranked as strong spam. Unlike strong spam and highly strong spams, there are less dangerous spam words and phrases used in daily life and are difficult to distinguish as spam, e.g. ‘Dear Sir ‘ , ‘money’, ‘information’, ‘free trial’ etc. So the system counts the spam words and add up values assigned to the words. Based on final value, fuzzy rules categorize the mail into the weak, moderate, strong and highly strong spam. Rank assignment in word/phrase dictionary are as per the following rules:-

0 < word value <=0.25 → weak spam

0.25 < Word value <=0.50 → Moderate spam

0.50 < Word value <=0.75 → Strong spam

0.75 < Word value <=0.9) → Highly strong spam

Emails are also ranked after fuzzification into different spam categories by level of threat as:-

0 < mail-value <=0.25 → weak spam

0.25 < mail-value <=0.50 → Moderate spam

0.50 < mail-value <=0.75 → Strong spam

0.75 < mail-value <=0.9) → Highly strong spam

4. RESULTS AND DISCUSSION:

The proposed system is trained and tested with two datasets, as mentioned before. The first dataset, i.e. CSDMC2010 [2], is publicly available and consists of 4327 emails. In this research, we trained the system with 800 emails in three phases, as mentioned later in this section, and tested it against another 1000 emails. The results after initial training phase with 500 emails of dataset 1 are shown in Table 1.

TABLE 1: RESULT AFTER TRAINING WITH 500 EMAILS

Measures	***Value***
Sensitivity	0.9836
Specificity	0.8619
Precision	0.8414
Negative Predictive Value	0.9860
False Positive Rate	0.1381
False Discovery Rate	0.1586
False Negative Rate	0.0164
Accuracy	0.9138
F1 Score	0.9069
Matthews Correlational Coefficient	0.8364

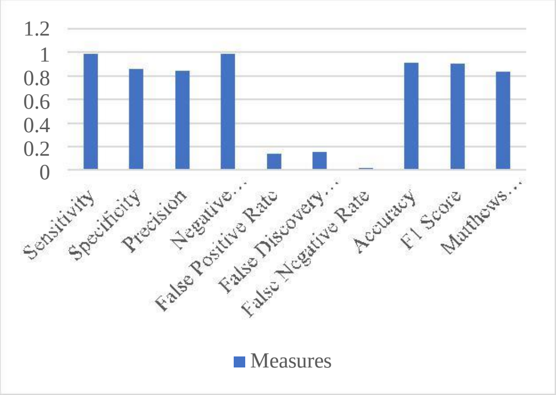

Figure 5: Graphical depiction of Table 1.

The results improved after training the system by 200 more mails and it showed better results as more spam data is added into spam dictionary, as shown in Table 2.

TABLE 2: PREDICTION RATE AFTER TRAINING 700 SPAM MAILS

Measures	**Value**
Sensitivity	0.9895
Specificity	0.9447
Precision	0.9420
Negative Predictive Value	0.9900
False Positive Rate	0.0553
False Discovery Rate	0.0580
False Negative Rate	0.0105
Accuracy	0.9660
F1 Score	0.9652
Matthews Correlational Coefficient	0.9331

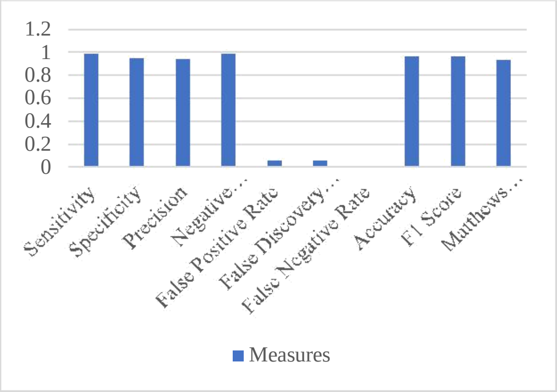

Figure 6: Graphical depiction of Table 2.

The accuracy of up to 98 % is achieved by training the model with 100 more mails, as shown in Table 3.

TABLE 3: PREDICTION RATE AFTER 800 EMAIL TRAINING

Measures	**Value**
Sensitivity	0.9959
Specificity	0.9689
Precision	0.9680
Negative Predictive Value	0.9960
False Positive Rate	0.0311
False Discovery Rate	0.0320
False Negative Rate	0.0041
Accuracy	0.9820
F1 Score	0.9817
Matthews Correlational Coefficient	0.9644

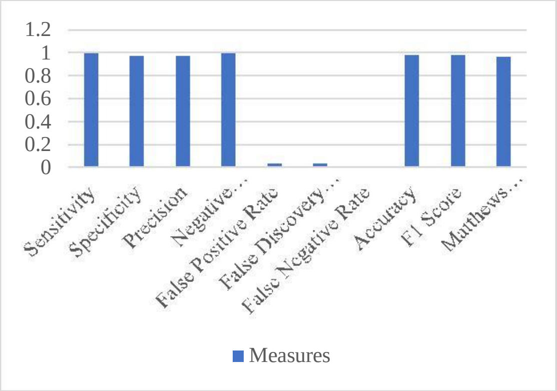

Figure 7: Graphical depiction of Table 3.

The Other dataset i.e. Stratford256 consist of 256 emails based on actual students queries at the University of Stratford. This dataset gave only weak spam indication. The fuzzy rules let the emails pass for further response. Table 4 shows results obtained after testing on this dataset.

TABLE 4: PREDICTION ON STRATFORD256 DATASET

Measures	**Value**
Sensitivity	0.9766
Specificity	
Precision	1.0000
Negative Predictive Value	0.0000
False Positive Rate	
False Discovery Rate	0.0000
False Negative Rate	0.0234
Accuracy	0.9766
F1 Score	0.9881
Matthews Correlational Coefficient	

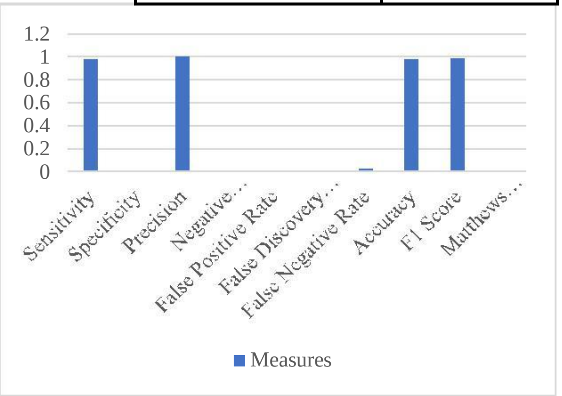

Figure 8: Graphical depiction of Table 4.

5. CONCLUSION AND FUTURE WORK:

The proposed spam filtering model, based on fuzzy logic, is found effective in detecting spam emails as part of an intelligent email response system developed jointly with the University of Stratford with more than 97% accuracy. The proposed model categorizes emails on the basis of sender address, email subject and its contents. Spam words and phrases are extracted from emails to rank the spam severity in a fuzzy manner. Fuzzy rules are then applied to categorize the emails with level of threat and identify emails as strong to weak spam to reply accordingly. In future we aim to considered IP address, URL, images and attachments to further improve our spam filtering.

REFERENCES

[1] Al-Alwani, Abdulkareem. "*A novel email response algorithm for email management systems."* Journal of Computer Science 10.4 (2014): 689.

[2] Santhi, G., S. MariaWenisch, and P. Sengutuvan. *"A Content Based Classification of Spam Mails with Fuzzy Word Ranking"* International Journal of Computer Science Issues *(IJCSI)* 10.3 (2013): 48.

[3] Rathi, Megha, and Vikas Pareek. *"Spam mail detection through data mining-A comparative performance analysis"* International Journal of Modern Education and Computer Science 5.12 (2013): 31.

[4] Lee, Chih-Ning, Yi-Ruei Chen, and Wen-Guey Tzeng. *"An online subject-based spam filter using natural language features"* IEEE Conference on Dependable and Secure Computing, 2017.

[5] Pomerol, Jean-Charles. "*Artificial intelligence and human decision making"* European Journal of Operational Research 99.1 (1997): 3-25.

[6] Kobersi, I.S., Finaev, V.I., Almasani, S.A., Kaid, W.A.A. *"Control of the Heating System with Fuzzy Logic"*, World Applied Sciences Journal 23 (11): 1441-1447, 2013ISSN 1818-4952, 2013.

[7] Almasan, Siham AM, et al. "*Filtering Spam Using Fuzzy Expert System"* Journal of Emerging Trends in Computing and Information Sciences 6.12 (2015).

[8] T. Sun, *"Spam Filtering based on Naive Bayes Classification"* Arch. Res. Pap. Babes Bolyai University, 2009.

[9] Roy, Kaushik, Sunil Keshari, and Surajit Giri. *"Enhanced Bayesian spam filter technique employing LCS"* International Conference *on* Computer, Electrical & Communication Engineering (ICCECE), 2016.

[10] Varghese, Reshma, and K. A. Dhanya. *"Efficient Feature Set for Spam Email Filtering"* 7th International Conference on Advance Computing (IACC), 2017.

[11] Lee, Chih-Ning, Yi-Ruei Chen, and Wen-Guey Tzeng. *"An online subject-based spam filter using natural language features"* IEEE Conference on Dependable and Secure Computing, 2017.

[12] Al-Alwani, Abdulkareem. *"A novel email response algorithm for email management systems"* Journal of Computer Science 10.4 (2014): 689.

[13] Peng, Wuxu, et al. *"Enhancing the Naive Bayes Spam Filter Through Intelligent Text Modification Detection"* 17th IEEE International Conference On Trust, Security And Privacy In Computing And Communications, 2018.

[14] Luo, Qin, et al. "*Research of a spam filtering algorithm based on naive Bayes and AIS"* International Conference on Computational and Information Sciences. IEEE, 2010.

[15] Chawathe, Sudarshan. "Improving Email Security with Fuzzy Rules." *2018 17th IEEE International Conference On Trust, Security And Privacy In Computing And Communications/12th IEEE International Conference On Big Data Science And Engineering (TrustCom/BigDataSE)*. IEEE, 2018

[16] Zhonghui, Zhang, and Wu Bin. "2010 Seventh International Conference on Fuzzy Systems and Knowledge Discovery (FSKD)."

[17] COSOI, A.C. VLAD, M.S. AND SGARCIU, V. *"On neural networks and the future of spam"*, IEEE International Conference on Automation, Quality and Testing, Robotics (AQTR 2008), Piscataway, NJ, USA: (2008), pp. 230 - 233.

[18] Christina, V., S. Karpagavalli, and G. Suganya. *"A study on email spam filtering techniques."* International Journal of Computer Applications 12.1 (2010): 0975-8887.

[19] Youn, Seongwook, and Dennis McLeod. *"A comparative study for email classification*." Advances and innovations in systems, computing sciences and software engineering. Springer, Dordrecht, 2007. 387-391.

[20] Begol, Moslem, and Keivan Maghooli. *"Improving digital image edge detection by fuzzy systems."* World Academy of Science, Engineering and Technology 81 (2011): 76-79.

[21] Santhi, G., S. MariaWenisch, and P. Sengutuvan. *"A Content Based Classification of Spam Mails with Fuzzy Word Ranking."* International Journal of Computer Science Issues (IJCSI) 10.3 (2013): 48.

[22] Mehdi Samiei yeganeh, Li Bin and G. Praveen Babu, *"A Model for Fuzzy Logic Based Machine Learning Approach for Spam Filtering"*, IOSR Journal of Computer Engineering 2012, ISSN: 2278-0661 Vol.4, No.5, pp. 07-10

[23] Subhodini gupta, Parekh .B.S and Jaimine N.Undavia, *"A Fuzzy Approach for Spam Mail Detection Integrated with Wordnet Hypernyms*

Key term Extraction”, IJERT, 2012, Vol. 1, No.5, pp.1-5.

[24] Kanagavalli, V. R., and K. Raja. *"A fuzzy logic based method for efficient retrieval of vague and uncertain spatial expressions in text exploiting the granulation of the spatial event queries."* International journal of computer applications (0975-8887), national conference on future computing CoRR. 2013.

[25] Sun, Jiping, et al. *"Fuzzy logic-based natural language processing and its application to speech recognition"* 3rd WSES International Conference on Fuzzy Sets & Systems. 2002.

[26] Santhi, G., S. Maria Wenisch, and P. Sengutuvan. *"Fuzzy Rule based Novel Approach to Spam Filtering."* International Journal of Computer Applications 71.14 (2013).

[27] Sudhakar, P., et al. *"Fuzzy logic for e-mail spam deduction"* Proceedings of the 10th WSEAS international conference on Applied computer and applied computational science. 2011.

[28] Fuad, M. Muztaba, Debzani Deb, and M. Shahriar Hossain. *"A trainable fuzzy spam detection system."* Proc. of the 7th Int. Conf. on Computer and Information Technology. 2004.

[29] Sonia, Dr. *"Spam Filter: VSM based Intelligent Fuzzy Decision Maker"* International Journal of Computer Science and Technology 1.1 (2010): 48-52.

[30] Barber, Mark H., Carsten Hagemann, and Christopher J. Hockings. *"Similar email spam detection" U.S. Patent Application No. 15/270,237.*