Control Theory and Informatics
ISSN 2224-5774 (Paper) ISSN 2225-0492 (Online)
Vol.7, 2018
www.iiste.org

A Survey of Email Spam Filtering Methods

Madhvi Sharma
PG Scholar, CSE, VITS, Bhopal, India

Prof. Sumit Sharma,
HOD CSE, VITS, Bhopal, India

Abstract
E-mail is one of the most secure medium for online communication and transferring data or messages through the web. An overgrowing increase in popularity, the number of unsolicited data has also increased rapidly. To filtering data, different approaches exist which automatically detect and remove these untenable messages. There are several numbers of email spam filtering technique such as Knowledge-based technique, Clustering techniques, Learning based technique, Heuristic processes and so on. This paper illustrates a survey of different existing email spam filtering system regarding Machine Learning Technique (MLT) such as Naive Bayes, SVM, K-Nearest Neighbor, Bayes Additive Regression, KNN Tree, and rules. However, here we present the classification, evaluation and comparison of different email spam filtering system
Keywords: e-mail spam, spam filtering methods, machine learning technique, classification, SVM, ANN

I. INTRODUCTION

Nowadays, email is an individual most popular fastest and cheapest means of communication. It has grown to be a piece of daily life for millions of citizens for their information sharing [1]. Due to its ease email is exposed to a lot of threats. One of the most essential threats toward email is spam: at all unsolicited profitable communication. The expansion of spam traffic is appropriate a worrying problem given that it consumes the network bandwidth, time of users and wastes memory and causes economic loss together the users with the organizations [4]. Spam moreover stops up the email method through filling-up the server disk space while sent to numerous users from the identical organization [5].

The mainly worrying kind of spam is cruel spam, which aims to increase several emails through links leading to cruel websites. According to Symantec, a sharp increase in cruel URLs at the ending of 2014 in evaluation to 2013 was associated to modify in procedure and a flow in generally engineered spam emails. Unlike cybercrime to target 'low down volume towering value' fatalities such as banks other than frequently require superior hacking capacity, cruel spam enable cruel content to achieve 'towering high volume low down value' targets, which are fewer possible to have efficient anti-virus or additional offset measures in position [6]. In the case of learning institutes, cruel spam threatens the privacy and safety of huge amount of responsive data involving to staff as well as students.

According to [7], definite classes of user, such as executive or armed forces personnel, show to be targeted collectively in campaign of cruel spam. We can theorize that the cruel spam emails which are sending to staff in learning institutes allocate general features. These features contain to be exploring in order to recover detection of cruel spam in email. For the separation of such spams from important mails, spam filtering is important. Techniques for such spam filtering are classified in figure 1. Amongst these, Naïve Bayesian classification, Support Vector Machine, K Nearest Neighbor, Neural Networks [2, 3] are most used and appreciated by researchers. Also, number of freeware and paid tools are available for spam filtering, which makes use of these techniques.

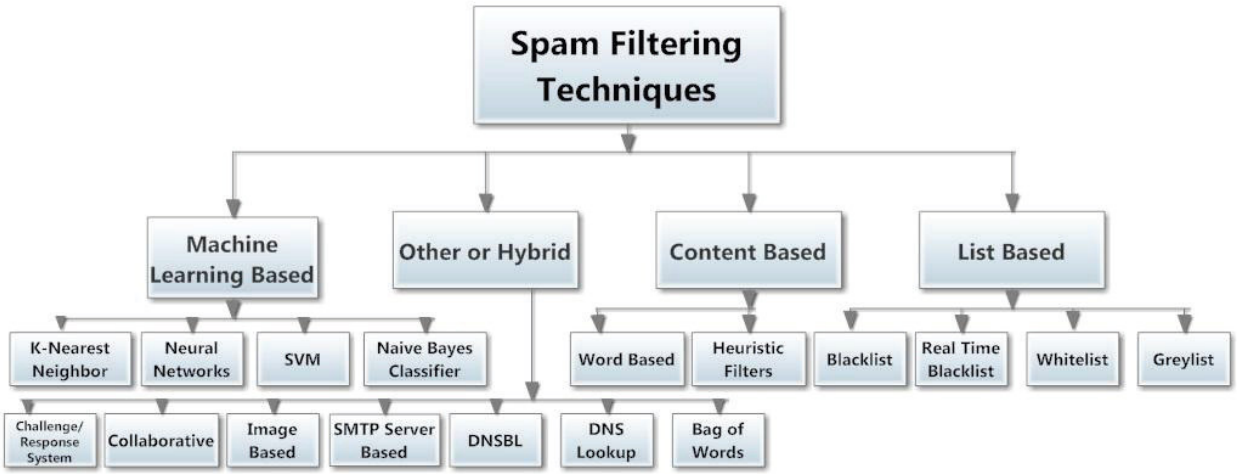

Fig. 1: Overall classification of spam filtering techniques

Several types of researches have been performed on email filtering, some acquired good accuracy and some are still going on. According to researcher's overview, Email filtering is a process to sort email according to some criteria. As there are various methods exist for email filtering, among them, inbound and outbound filtering is well known. Inbound filtering is the process to read a message from internet address and outbound filtering is to read the message from the local user. Moreover, the most effective and useful email filtering is Spam filtering which performs through anti-spam technique. As spammers are proactive natures and using dynamic spam structures which have been changing continuously for preventing the anti-spam procedures and thus making spam filtering is a challenging task [8].

Spam filtering is a process to detect unsolicited massage and prevent from entering into user's inbox. Now days, various systems have been existed to generate anti-spam technique for preventing unsolicited bulk email. Most of the anti-spam methods have some inconsistency between false negatives (missed spam) and false positives (rejecting good emails) which act as a barrier for most of the system to make successful anti-spam system. Therefore, an intelligent and effective spam-filtering system is the prime demand for web users.

II. EMAIL SPAM FILTERING METHODS

At present, number of spam email has increased for several criteria such as an advertisement, multi-level marketing, chain letter, political email, stock market advice and so forth. For restricting spam email, several methods or spam filtering system has been constructed by using various concept and algorithms. This section concluded by describing few of spam filtering methods to understand the process of spam filtering and its effectiveness.

a) Standard Spam Filtering Method

Email Spam filtering process works through a set of protocols to determine either the message is spam or not. At present, a large number of spam filtering process have existed. Among them, Standard spam filtering process follows some rules and acts as a classifier with sets of protocols. Figure.1 shows that, a standard spam filtering process performed the analysis by following some steps [9]. First one is content filters which determine the spam message by applying several Machines learning techniques [10]. Second, header filters act by extracting information from email header. Then, backlist filters determine the spam message and stop all emails which come from backlist file. Afterward, "Rules-based filters" recognize sender through subject line by using user defined criteria [19]. Next, "Permission filters" send the message by getting recipients pre-approvement. Finally, "Challenge response filter" performed by applying an algorithm for getting the permission from the sender to send the mail.

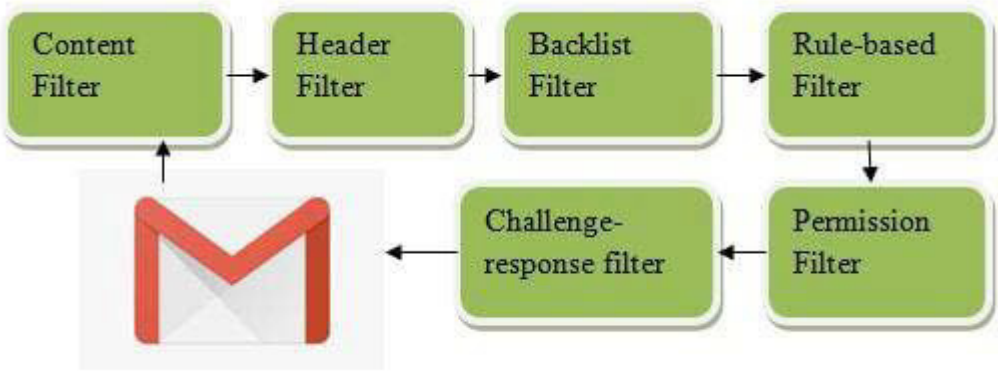

Figure 1: A standard process of Email spam filtering system

b) Client Side and Enterprise Level Spam Filtering Methods

A client can send or receive an email by just one clicking through an ISP. Client level spam filtering provides some frameworks for the individual client to secure mail transmission. A client can easily filter spam through these several existing frameworks by installing on PC. This framework can interact with MUA (Mail user agent) and filtering the client inbox by composing, accepting and managing the messages [11].

Enterprise level spam filtering is a process where provided frameworks are installing on mail server which interacts with the MTA for classifying the received messages or mail in order to categorize the spam message on the network. By this system, a user on that network can filter the spam by installing appropriate system [12] more efficiently. By far most; current spam filtering frameworks use principle based scoring procedures. An arrangement of guidelines is connected to a message and calculates a score based principles that are valid for the message. The message will consider as spam message when it exceeds the threshold value. As spammers are using various strategies, so all functions are redesigned routinely by applying a list-based technique to automatically block the messages. Figure 2 represents the method of client side and enterprise level spam filtering [13].

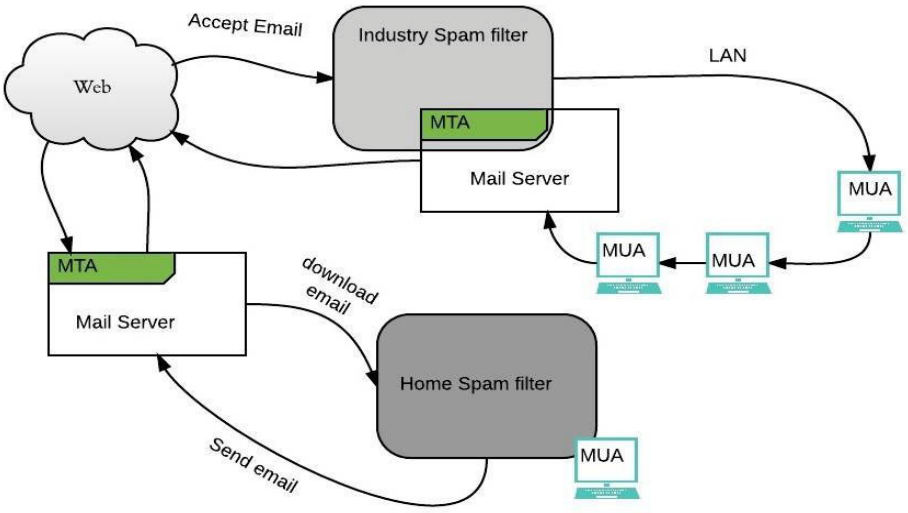

Figure 2: Client Side and Enterprise level Email spam filtering system

c) *Case Base Spam Filtering Method*

Among several spam filtering methods; case base or sample base filtering is one of the prominent method for Machine Learning Technique.

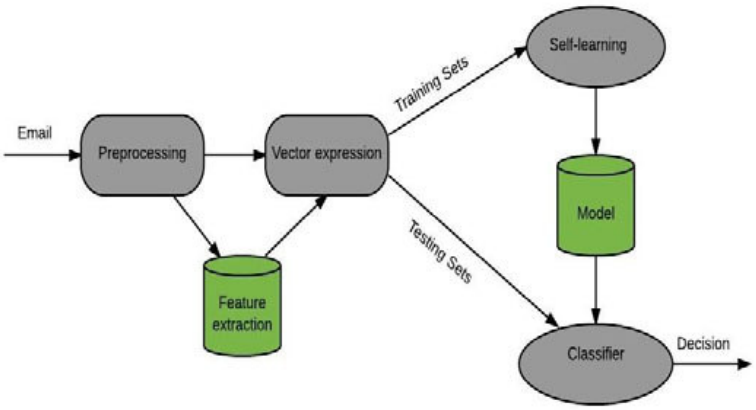

Figure 3: Case Base Spam Filtering System

Here, describes a sample of case base spam filtering architecture by applying Machine learning techniques [Fig. 3] in detail. The full process perform through several steps which followed by the figure 3. At the first step, extracted all email (spam email and legitimate email) from individual users email through collection model. Then, the initial transformation starts with the pre-processing steps through client interface, highlight extraction and choice, email data classification, analyzing the process and by using vector expression classifies the data into two sets.

Finally, machine learning technique is applied on training sets and testing sets to determine email whether it is spam or legitimate. The final decision makes through two steps; through self observation and classifier's result to make decision whether the email is spam or legitimate.

III. SPAM DETECTION TECHNIQUES

There are lots of existing techniques which try to prevent or reduce the expansion of huge amount of spam or junk email. The available techniques usually move around using of spam filters. Generally, spam detection techniques or Spam filters inspect different sections of an email message to determine if it is spam or not.

On the basis of different sections of email messages; Spam detection techniques can be classified as Origin based spam detection techniques and Content based spam detection techniques [14]. Generally, most of the techniques applied to the issue of spam detection are effective but the important role in minimizing spam email is the content based filtering. Its positive outcome has forced spammers to regularly change their methods, behaviors, and to scam their messages, in order to avoid these kinds of filters. Spam detection techniques are discussed below:

Origin-Based Technique:

Origin or address based filters are techniques which based on using network information to detect whether a email message is spam or not. The email address and the IP address are the most important parts of network information used. There are few main categories of origin-Based filters like Blacklists; Whitelists based systems [14].

1) ***Blaclists:*** Blacklists are records of email addresses or IP addresses that have been earlier used to send spam [15]. In creating a filter; if the sender of mail has its entry in the black list then that mail is undesirable and will be considered as spam [16]. For example those websites can be put in blacklist which have a past record of fraudulent or which exploits browser's vulnerabilities. The main problem of a blacklist is maintaining its content to be accurate and up-to-date.
2) ***WhiteLists:*** It is opposite to the black list concept. It consists of the list of entries which can penetrate through and are authorized. These mails are considered as ham mails and can be accepted by the user. It has a set of URLs and domain names that are legitimate [16]. Spam is blocked by a white list with a system which is exactly opposite to existing blacklist. Rather than define which senders to block mail from, a white list define which senders to permit mail from; these addresses are placed on a trusted-users list [15].

   The main difficulty of white listings is the assumption that trustworthy contacts do not send junk, for a while this theory could be invalid. Great number of spammers uses PCs that have been harmed using viruses and Trojans for sending spam, to every single one contacts of address book, thus we could receive a spam message from a recognized sender if a virus has infected his computer. Seeing as these contacts are present in the white list, all messages arriving from them are labeled as secure [17].
3) ***Realtime Blackhole List (RBL):*** This spam-filtering method acts something like the same to a accepted blacklist on the converse less hands-on maintenance is required, and the Mail Abuse Prevention System

and System administrators (third-party) operate it using spam detection tools [17]. This filter basically needs to connect to the third-party system whenever an email comes in, to authenticate the sender's IP address against the list. As the list is probably to be preserved by a third party, we don't have as much of control on what addresses are there on the list [15].

Content Based Spam Detection Techniques:

Content based filters are based on examining the content of emails. These content based filters are based on manually made rules, also called as heuristic filters, or these filters are learned by machine learning algorithms [17].These filters try to interpret the text in respect of examine its content and make decisions on that basis have spread among the Internet users, ranging from individual users at their personal computers, to big commercial networks. The success of content-based filters for spam detection is so large that spammers have performed more and more complex attacks intended to avoid them and to reach the users mailbox.

There are various popular content based filters such as: Rule Based Filters, Bayesian filters, Support Vector Machines (SVM) and Artificial Neural Network (ANN).

1) Rule- Based Filters: The Rule-Based Filters use a set of rules on the words incorporated in the whole message to find out whether the message is spam or not. In this approach, a comparison is done between each email message and a set of rules to find out whether a message is spam or ham. A set of rules contains rules with a variety of weights assigned to each rule. In the beginning, each received email message has a zero score. Then email is parsed to detect the existence of any rule, if it exists. If the rule is found in the message, then the weight of the rule is added to the final score of the email. At the end, if the final score is found to be exceeding some threshold value, then the email is declared as spam [18].

The drawback of Rule-Based Spam Detection Technique is that it is a set of rules that is very huge and static that causes less performance [14]. The spammers can effortlessly surmount these filters by simple word obfuscation, for example, the word "SALE" could be changed to S*A*L*E so it will bypass the filters. The inflexibility of the rule-based approach is it's another major disadvantage. The rule based spam filter is not intelligent as there is no self-learning ability available in the filter.

2) Bayesian filters: The Bayesian filters are most advanced form of content-based filtering, these filters uses the laws of probability to find out which messages are legitimate and which are spam. Bayesian Filters are also the well known machine learning approaches [19]. In order to identifying each message as either junk or legitimate, initially the end user must "train" the Bayesian filter manually for efficiently block the spam messages.

Eventually, the filter takes words and phrases found in legitimate emails and adds them to a list; it also applied the same method with words found in spam. To decide which received messages are classified as spam emails, the content of the email are scan by the Bayesian filter and then compare the text against its two-word lists to calculate the probability that the message is spam.

For example, if the occurrences of word "free" is 62 times in a list of spam messages but only 3 times in ham(legitimate) emails, then there is a 95% possibility that an arriving email containing the word "free" is spam or junk email. Because a Bayesian filter is continuously building its list of word based on the messages that an individual user receives, it theoretically becomes more efficient the longer it's used.

However, since the Bayesian filter method requires a training before it starts working well, we will require exercising patience and will probably have to delete few junk messages manually, at least at first time [14].

3) Support Vector Machines: The Support Vector Machines (SVM) has successes at using as classifying text documents. SVM has encouraged important researches into applying them to spam filtering. SVMs are kernel methods whose vital idea is to embed the data indicating the text documents into a vector space where geometry and linear algebra can be performed [19]. SVMs try to create a linear separation between the two classes in the vector space [14].

Support Vector Machine An example is shown above. In this example, the objects belong either to BLUE (ham) class or PINK (spam) class. The separating line defines a boundary on the left side of which all objects are PINK and to the right of which all objects are BLUE. Any new object (white circle) falling to the right is labeled, i.e., classified, as BLUE (or classified as PINK should it fall to the left of the separating line).

4) Artificial Neural Network An artificial neural network is a group of interconnected nodes these nodes are called as neurons. The well known example of artificial neural network is the human brain. The term artificial neural network has moved around a huge class of models and machine learning methods. The central idea is to extract linear combinations of the inputs and derived features from input and then model the target as a nonlinear function of these features [14]. Neural Network as an interconnected collection of nodes

ANN is an adaptational system that changes its structure based on internal or external information that flows through the network during the learning phase. They are generally familiar with model complex relationships between inputs and outputs or to find patterns in data. The neural network must first be "trained" to categorize emails into spam or non spam starting from the particular data sets. This training includes a computational analysis of message content using huge representative samples of both spam and non-spam messages [18]. To generate training sets of spam and non-spam emails, each email is attentively reviewed

according to this simple, yet limited definition of spam.

IV. RELATED WORK

Many algorithms have been proposed for classifying spam and legitimate emails. N. Radha and R. Lakshmi [20] compared the performance of Naïve Bayesian (NB), Multi-layered Perceptron (MLP), J48, and Linear Discriminant Analysis (LDA) algorithms. Using WEKA software, they achieved a prediction accuracy of 93% for J48, slightly exceeding MLP's 92%, however at the expense of an increased computational time. Using RapidMiner, MLP accuracy surpassed that of LDA by 1% and that of NB by 3%.

S. Youn and D. McLoed [21] explored how the size of a dataset affects classification performance. For a dataset size of 1000, support vector machines (SVM), NB, and J48 achieved accuracies of 92.7%, 97.2%, and 95.8% respectively. Nevertheless, when the size increased to 5000, accuracy of SVM dropped by 1.8% and of NB by 0.7%, whereas that of J48 increased by 1.8%. Moreover, they deduced that accuracy increased with increasing feature size. The authors of [22] performed weighted SVM on spam filtering. Un-weighted SVM ignores the particular importance of every sample, which often leads to imbalance classification and less precise results. They tested their algorithm on 400 emails from the Chinese corpus ZH1 with half of them being spam. Results revealed that as the weight value of the legitimate emails class increased from 1 to 10 with the spam weight fixed to 1, the precision increased from 97.47% to 99.44%.

Furthermore, [23] modified SVM into a relaxed online SVM so that it trains only on actual errors. This resulted in less number of iterations and minimized the computational cost of the algorithm. The precision of the algorithm on the benchmark email datasets trec05p-1 and trec06p was very close to that of the online SVM while reducing the CPU execution time.

In [24], the authors combined the Best Stepwise feature selection with a classifier of Euclidean nearest neighbor and created a Naïve Euclidean approach. Each email was represented in D-dimensional Euclidean space. Using SpamBase from the UCI repository, and a 10-fold cross validation, they achieved an accuracy of 82.31% compared to 60.6% for the Zero rule.

R. Laurutis et al. [25] applied artificial neural network (ANN) to classify spam emails. Their main contribution was to replace the frequency of words in the content with descriptive properties of elusive patterns created by the spammers. Their data corpus contained around 1800 spam and 2800 legitimate emails. Their experiments revealed that ANN provided a maximum precision value of 90.57% after training it with 57 email parameters.

Authors of [26] compared SVM to Rocchio classifier. Rocchio classifier is based on the normalized TF-IDF modeling of training vectors. The dot product of the test and prototype vectors is obtained to classify the document as spam or non-spam. The threshold value of classification is obtained such that the training error is minimized after rank ordering the generated dot products of the prototype vector with the whole set of training. Compared to binary SVM with an error rate 0.213, Rocchio classifier reached a 0.327 error rate using a dictionary of mixed upper and lower cases.

J. Provost [27] evaluated the rule-learning RIPPER algorithm compared to the NB algorithm. RIPPER generates keyword-spotting rules to set and bag valued attributes. It performs its classification according to the predefined rules that determine the impact of having certain words in the header fields or content of the emails. The experiments performed on junk email provided by several users and on legitimate ones form the inbox of the author achieved 90% accuracy after training it with 400 emails whereas NB reached 95% after only 50 training emails.

In [28], B. Medlock proposed the smoothed n-gram language interpolation and modeling which assumes that the probability of a specific word in a sequence is solely dependent on the previous n-1 words. Separate language models were built for legitimate and spam email followed by computing the probability that the model generated this text message. Bayes rule was applied later to find the class with the highest probability for the provided message. An accuracy of 98.84% and 97.48% was obtained for the adaptive bigram and unigram language model classifier after applying it to the GenSpam corpus of 9072 legitimate and 32332 spam emails.

Finally, we discuss similar comparative studies specific to the Spambase corpus used herein. In [29], Kiran et al. analyzed the performance of several classifiers in identifying spam emails based on the Spambase corpus using the WEKA toolset. Ensemble classifiers were employed as well in a set of experiments measuring classification accuracy, precision, and recall. Different validation techniques including half splits, leave-one-out, and 10-fold cross validation were used while noting consistent results across all. Ensemble Decision Tree (25 trees into total) was shown to achieve the best accuracy rate with 96.4%. In [30], in the most recent relevant work (2013), Sharma et al. evaluated Spambase using 24 classifiers from the WEKA toolset as well. Ten fold cross validation was used and accuracy, precision, and recall are reported on for each algorithm. Random Committee achieved the best result with 94.28% accuracy.

V. PROBLEM STATEMENT

Web spam which is a most important issue through today's web search tool; therefore it is essential for web crawlers to contain the capacity to identify web spam among creeping. The categorization Models are considered by machine learning classify algorithm. The one machine learning algorithm is Naïve Bayesian Classifier which is as well used in to part the spam as well as non-spam mails. Big Data analyze framework which is as well outline used for spam detection. Extract the feeling as of a message is a method for get the important data. In Machine learning innovation can get from the training datasets additionally anticipate the preference making framework hence they are broadly utilize as a fraction of feeling order through the exceptionally accuracy of framework.

VI. FUTURE RESEARCH SCOPE

This work proposes a model for improving recognition of cruel spam in email. Our model resolve employ a novel dataset intended for the process of feature choice, and then validate the set of chosen features using three classifiers identified in spam detection: Support Vector Machine, Naïve Bayes, and Multilayer Perception. Feature selection is projected to recover training time as well as accuracy for the classifiers

VII. CONCLUSIONS

To review the results of the hypothesis it can be said, that the design of a Meta spam filter make sense as well as has its ground. Although the notion deals with existing spam filters as well as e-mail corpus, the over describe methodology can as well be applied for extra filters also. Studies of Bayesian networks have provided a fine base for the creation of a Meta spam filter.

REFERENCES

[1.] S. Whittaker, V. Bellotti and P. Moody, "Introduction to this special issue on revisiting and reinventing e-mail", Human-Computer Interaction, 20(1), 1-9, 2005.
[2.] Blanzieri, Enrico, and Anton Bryl. "A survey of learning-based techniques of email spam filtering." Artificial Intelligence Review 29, no. 1 (2008): 63-92.
[3.] Zhang, Le, Jingbo Zhu, and Tianshun Yao. "An evaluation of statistical spam filtering techniques." ACM Transactions on Asian Language Information Processing (TALIP) 3, no. 4 (2004): 243-269.
[4.] G. Santhi, S. M. Wenisch and P. Sengutuvan, "A Content Based Classification of Spam Mails with Fuzzy Word Ranking", IJCSI International Journal of Computer Science Issues, 10, 48-58, 2013.
[5.] I. Koprinska, J. Poon, J. Clark and J. Chan, "Learning to classify e-mail", Information Sciences, 177(10), 2167-2187, 2007.
[6.] M. Alazab and R. Broadhurst, "Spam and Criminal Activity", Australian Institute of Criminology, 2014.
[7.] R. Amin, J. Ryan, and J. van Dorp, "Detecting Targeted Malicious Email Using Persistent Threat and Recipient Oriented Features", IEEE Secur, Priv. Mag, (99), 1-1, 2012.
[8.] Drucker, Harris, Donghui Wu, and Vladimir N. Vapnik. "Support vector machines for spam categorization." IEEE Transactions on Neural networks 10, no. 5 (1999): 1048-1054.
[9.] Christina, V., S. Karpagavalli, and G. Suganya. "A study on email spam filtering techniques." International Journal of Computer Applications 12, no. 1 (2010): 0975-8887.
[10.] Wang, Qiang, Yi Guan, and Xiaolong Wang. "SVM-Based Spam Filter with Active and Online Learning." In TREC. 2006.
[11.] Saad, Omar, Ashraf Darwish, and Ramadan Faraj. "A survey of machine learning techniques for Spam filtering." International Journal of Computer Science and Network Security (IJCSNS) 12, no. 2 (2012): 66.
[12.] Cormack, Gordon V., José María Gómez Hidalgo, and Enrique Puertas Sánz. "Spam filtering for short messages." In Proceedings of the sixteenth ACM conference on Conference on information and knowledge management, pp. 313-320. ACM, 2007.
[13.] Cunningham, Padraig, Niamh Nowlan, Sarah Jane Delany, and Mads Haahr. "A case-based approach to spam filtering that can track concept drift." In The ICCBR, vol. 3, pp. 03-2003. 2003.
[14.] Sabri, Alia Taha, Adel Hamdan Mohammads, Bassam Al-Shargabi, and Maher Abu Hamdeh. "Developing new continuous learning approach for spam detection using artificial neural network (CLA_ANN)." European Journal of Scientific Research 42, no. 3 (2010): 525-535.
[15.] Jaswal, Vandana, and Nidhi Sood. "Spam Detection System Using Hidden Markov Model." International Journal of Advanced Research in Computer Science and Software Engineering 3, no. 7 (2013): 304-308.
[16.] Puri, Sahil, Dishant Gosain, Mehak Ahuja, Ishita Kathuria, and Nishtha Jatana. "Comparison and analysis of spam detection algorithms." International Journal of Application or Innovation in Engineering & Management (IJAIEM) (2013).
[17.] Sanz, Enrique Puertas, José María Gómez Hidalgo, and José Carlos Cortizo Pérez. "Email spam filtering." Advances in computers 74 (2008): 45-114.

[18.] Shrivastava, Jitendra Nath, and Hima Bindu Maringanti. "E-mail spam filtering using adaptive genetic algorithm." International Journal of Intelligent Systems and Applications 6, no. 2 (2014): 54.
[19.] Nosseir, Ann, Khaled Nagati, and Islam Taj-Eddin. "Intelligent word-based spam filter detection using multi-neural networks." International Journal of Computer Science Issues (IJCSI) 10, no. 2 Part 1 (2013): 17.
[20.] N. Radha and R. Lakshmi . “Supervised learning approach for spam classification analysis using data mining tools.” (IJCSE) International Journal on Computer Science and Engineering Vol. 02, No. 09, 2010, pp. 2783-2789
[21.] Youn, Seongwook, and Dennis McLeod. "A comparative study for email classification." In Advances and innovations in systems, computing sciences and software engineering, pp. 387-391. Springer, Dordrecht, 2007.
[22.] Chen, Xiao-li, Pei-yu Liu, Zhen-fang Zhu, and Ye Qiu. "A method of spam filtering based on weighted support vector machines." In IT in Medicine & Education, 2009. ITIME'09. IEEE International Symposium on, vol. 1, pp. 947-950. IEEE, 2009.
[23.] Sculley, David, and Gabriel M. Wachman. "Relaxed online SVMs for spam filtering." In Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval, pp. 415-422. ACM, 2007.
[24.] Chan, T. Y., Jie Ji, and Qiangfu Zhao. "Learning to Detect Spam: Naive-Euclidean Approach." International Journal of Signal Processing 1 (2010): 31-38.
[25.] Puniškis, D., R. Laurutis, and R. Dirmeikis. "An artificial neural nets for spam e-mail recognition." Elektronika ir Elektrotechnika 69, no. 5 (2006): 73-76.
[26.] Drucker, Harris, Donghui Wu, and Vladimir N. Vapnik. "Support vector machines for spam categorization." IEEE Transactions on Neural networks 10, no. 5 (1999): 1048-1054.
[27.] Provost, J. "Naive-bayes vs. rule-learning in classification of email. The University of Texas at Austin." Artificial Intelligence Lab. Technical Report AI-TR-99-284 (1999).
[28.] Medlock, Ben. "An Adaptive, Semi-Structured Language Model Approach to Spam Filtering on a New Corpus." In CEAS. 2006.
[29.] Kiran, SS Ravi. "Spam or not spam--that is the question." (2009).
[30.] Sharma, Sumant, and Amit Arora. "Adaptive approach for spam detection." International Journal of Computer Science Issues (IJCSI) 10, no. 4 (2013): 23.