iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 9 **Issue:** XII **Month of publication:** December 2021

DOI: https://doi.org/10.22214/ijraset.2021.39629

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 9 Issue XII Dec 2021- Available at www.ijraset.com

Review and Analysis on Filtering of Unwanted Multimedia Messages from Online Social Network User Walls

Martand Ratnam[1], Dr. Sachin Singh[2]
[1]*M.Tech Scholar, BBDIT Ghaziabad*
[2]*Assistant Professor, BBDIT Ghaziabad*

Abstract: ***When it comes to sharing and exchanging various types of information, online social networks (OSNs) have become an increasingly popular and interactive medium in today's world. People who are connected to blogs and social networks see all of the publicly shared information, and it has a profound effect on the human mind. Messages or comments posted on a wall, a public or private area, may include unnecessary information or sensitive data. Thus, online social networks can benefit from information filtering, which can be used to help users organise messages written in public areas by removing unnecessary words. An information filtering system proposed in this paper may allow OSN users to control the posting and commenting on their walls directly. Every time a user posts a message, the message is intercepted by the filtered wall, which then applies Filtering and Black List Rules to it. The message will appear on the user's wall if it is not filtered or blacklisted.***
Keywords: Content Based Message Filtering, Demographic Filtering, Collaborative Filtering.

I. INTRODUCTION

For example, a social networking service can be used to connect people who share interests, activities, and a lot of human life information. The exchange of various types of content, including free text, image, audio, and video data, takes place on a daily basis. Users, particularly adolescents, are spending a significant amount of time on various social networking sites to connect with others, to share information, and to pursue common interests as social media has grown rapidly. No support is provided for preventing unwanted messages from being posted on user walls by social networks (OSNs). The majority of social network content is made up of short text, such as the messages that OSN users leave on the walls of specific public or private areas. Due to a lack of classification or filtering tools, a user receives all messages posted by the users he or she is following. Most of the time, the user is bombarded with a constant stream of notifications.

Various communication technologies, notably online social networks, necessitate the development of additional security mechanisms. Consequently, today's Online Social Networks (OSN) are tasked with filtering information. For textual documents and more recently, web content, information filtering has been extensively studied.

The ability to automatically control the messages written on one's wall by filtering out unwanted messages is one of the many benefits of message filtering. Filtered wall is an idea put forth by OSN users who want more control over the content that appears on their personal digital bulletin boards.

Machine Learning is used to assign each message to a category, and Filtering rules are used to allow the user to specify which content should not appear on their walls. There are also Black List Rules that can be used to block a user from accessing the wall at any given time. Social networking sites will be protected by the proposed system.

A. Overview

In today's world, electronic communication is a necessity. OSNs (online social networks) are used primarily for personal communication, while email is used primarily for official communication. Spam is the biggest problem that internet users are currently dealing with. Spam is a term used to describe unsolicited or unrequested electronic messages sent for the purpose of advertising, disseminating malicious code, phishing, or simply annoying the recipients. As a result of this, spam messages can spread very quickly because people have a tendency to believe and share content that has been shared by others. Spam is not only a waste of time for the users, but it can also result in financial losses. Spam is a difficult problem for internet users to deal with, but there are a variety of ways to combat it.

II. LITERATURE REVIEW

A. Content and Identity Based Filtering

Several studies in the field of spam detection have identified two broad categories of spam filtering techniques: content-based filtering and identity-based filtering. E-mails in the content-based category are parsed and graded based on spam-specific keywords and patterns. It is extremely vulnerable to poison attacks when using content-based filtering Spam e-mails contain a large number of legitimate words, making it less likely to be flagged as spam in a poison attack. The user maintains a whitelist and a blacklist of e-mail addresses with identity-based filtering. These are vulnerable to impersonation attacks because they are based on the sender's e-mail ID, which can be forged or hacked into to impersonate ordinary users. Spam filters must be able to withstand both of these types of attacks in order to work. Spam filtering is now more effective thanks to the addition of social network data to the email system. Some studies have also taken into account the trust, interests, and closeness of social network users in order to filter spam. However, because these factors are derived solely from social networks, they cannot be used to identify and classify spam emails. The e-mails were not classified based on any email-specific factors. Active learning-based spam message classification was proposed by Lizhou et al. (2016) to reduce classification time without sacrificing accuracy. Their work does not include the classification of e-mails that include images.

Yuanchun et al. (2011) proposed a method based on local concentration that uses content to classify spam email. Lourdes et al. discuss content and identity-based filtering (2010). Congfu et al. have proposed a content-based fusion algorithm for spam e-mail detection (2014). E-mails are classified based on the voting strategies used by the different classifiers. Salehi et al. (2017) proposed the use of a fuzzy-based method for the classification of spam emails. Spam messages are analysed for structural patterns and a fuzzy-based approach is used to reduce the problem of noise points. Sang et al. propose an approach for spam message identification based on optimising parameters and selecting features (2011). According to Zhenhai et al., spam messages can be identified by examining the departing messages (2012).

Gopi et al. (2018) discussed Features for spam message filtering in emails are selected using terms and category ratios based on term frequency and sample ratios. An economic metric method to improve the accuracy of spam detectors in e-mails was discussed by Fida et al. (2016). Amany et al. (2018) present a spam detection boosting method.

If the content of the benign email matches that of unrelated spam, no false positives are produced. All messages exchanged between nodes in the system must be forwarded without revealing the identity of the sender. Every detail about previously visited nodes must be removed before the information can be sent on to the next node in order to be shared across a community. Peer-to-peer communication is the norm in the system. In their approach, bandwidth costs are reduced while spam detection rates are raised at the same time. The detection of malicious attachments in e-mails was discussed by Yehonatan et al. (2018).

Ismaila et al. (2015) Also proposed an improved swarm optimization spam detection model for e-mail. In the Negative Selection Algorithm (NSA), a detector is generated using Particle Swarm Optimization (PSO), which uses a stochastic distribution to model the data (NSA). Particle swarm optimization and negative selection algorithm were used in conjunction. When compared to the NSA and PSO models, the performance and accuracy of the e-mail spam detection model.

III. SOCIAL NETWORKS FOR SPAM FILTERING

Haiying et al. (2014) put forth the idea of using social network data to filter the spam messages in e-mails. In addition to the social network factors, it employed a content-based and identity-based spam filtering technique. E-mails in the content-based category are analysed for spam-related keywords and patterns. Impersonation attacks, in which the identities of ordinary users are impersonated by forging their IDs or compromising their computers, are very common in the content-based category. A whitelist and blacklist of e-mail addresses are maintained by the users in identity-based filtering systems. This is because they are based solely on the e-mail addresses of the senders, which are more susceptible to poison attacks. Spam e-mails that have a lot of legitimate words added to them are less likely to be flagged as spam. E-mail systems used only a limited number of social network factors to distinguish between legitimate e-mails and spam, such as closeness, interest, and trust. The use of Bayesian spam filters resulted in better accuracy, protection against attacks, and a more efficient method of identifying spam. The trust and interest factors in email networks were not taken into account in this study. Chao et al. (2013) introduced The use of machine learning to identify spammers on Twitter. Based on a large dataset of millions of tweets, the empirical analysis was conducted. In order to identify spammers on Twitter, a neighbor-based detection feature was employed. Twitter spammers were identified through the collection and validation of evasion tactics based on profile features. The number of followers, the number of tweets, etc., were used to identify spam accounts on Twitter. The spammers' evasion tactics, including 24 detection features, were studied extensively and several detection features were examined. On the other hand, the datasets crawled from Twitter and classified as benign only included those accounts that had never posted malicious URLs, even though some of those URLs may be malicious.

Xianghan et al. (2015) proposed Spam detection method for social networks. A supervised machine learning approach to spam detection was proposed. The message's content and the user's behaviour were taken into consideration when implementing some key features. To identify spammers, the SVM classification algorithm was used with feature selection algorithms to identify the most important features and their weight. In addition, the feature extraction method was based on statistical analysis and human selection. However, the solution performs very well when the true positive rate of spammers and non-spammers is taken into consideration.

A. E-Mail Categorization

Mostafa et al. introduced attentive learning for e-mail categorization (2016). There are a lot of emails being sent out every day, and e-mail categorization is a hot topic. The dynamic behaviour of users is simulated while their new e-mails are categorised by an attentive learning approach for automatic e-mail categorization. Based on the actions of the users, researchers examined the different ways emails can be classified. E-mail categorization can be achieved by classifying the various aspects and structure format of an email as a feature set, which is a subset of the entire feature space, for an incoming message. After that, a sequence of the feature set was chosen for further study.

Decision Tree, Support Vector Machine (SVM), and NB classifiers were all compared to the Random Forest algorithm (RF) to see how well they performed. Only the user's classification of files and e-mails were successful in determining whether or not the automatic system was successful. E-mail folders are a major problem because the subject matter changes over time. Thus, the effectiveness of co-training for spam filtering was evaluated. As Iryna et al. (2013) explained, evolutionary algorithms can be used to improve the performance of spam filters. According to Faeze and colleagues (2019), the NB classifier can be used to identify web spam.

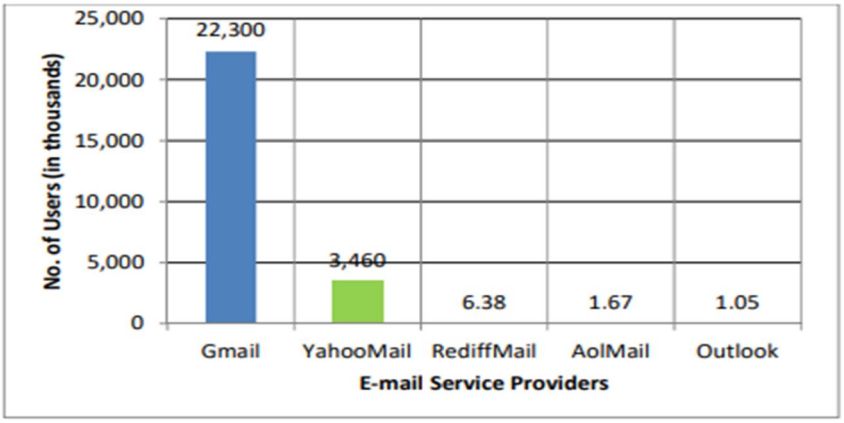


Figure 3.1 Usage of e-mail domain by social network users

Szde Yu (2015) proposed Digital forensics faces a major hurdle here. Thousands and hundreds of artefacts were used in the digital investigation of the crime. Only e-mails are capable of generating a large amount of data and information. There have been numerous scenarios drawn from real e-mail forensics investigations. Criminals can use e-mail spamming as a means of reaching potential victims and facilitating their schemes. There are no forensic tools that can reliably detect spam e-mails, so digital investigators must have the necessary knowledge and patience to find out how crucial data is being sent out through content analysis. Detecting e-mail spam while minimising false positives and minimising false negatives is the primary concern in the field of e-mail spam detection.

IV. EFFECTIVE FILTERING OF UNSOLICITED MESSAGES FROM ONLINE SOCIAL NETWORKS

A. Introduction

One-on-one networking (ONN) has become an essential part of modern communication. Everyone has a good time on social media, regardless of their age. In addition to their many advantages, OSNs have a number of drawbacks, the most prominent of which is spam. Sending unwanted or unwelcome emails, microblogs, instant messengers and so on is known as spam. It's a nuisance and a waste of time for everyone who has to deal with it. It's currently impossible to tell whether an email is spam or not based solely on the content of the message. Some machine learning techniques are used to assist in spam detection, but the unit of measurement is different for each type of spam and has different effects on web users.

The advantages of the BF and SF over other approaches have made them popular. Faster convergence with the BF than with other models based on the probability values. In addition, it only requires a small amount of training data and is simple to implement. Both classes are further away from the SF than they are from each other. This is the case for feature vectors located near the separating plane, where disruptions are less likely to have an impact on them This means that even though there may be an interruption, the feature vector will not be affected. The SF performs well in high-dimensional space, has a high classification accuracy rate, and a wide variety of kernel functions. Each one is designed for a specific type of data. For non-linear datasets, the appropriate kernel functions are used to separate them. Overfitting can be avoided by incorporating soft margins. Slack values are added to soft margins, and error rates are calculated by taking into account the slack values. Soft margins and acceptable slack values can help reduce error rates. For text classification, the two classifiers mentioned above are excellent, but they are not sufficient for effectively classifying messages from the OSN.

B. Identification and Filtering of OSN Spam

Spammers use the internet to distribute a variety of spam, including text spam, URL spam, review spam, comment spam, and more. So, researchers came up with a variety of methods for detecting spam. In De Wang et al., URL spam can be detected by mapping URLs to destination URLs (2015). The invalid or inactive URLs were then filtered out, and the URL analysis was completed. In order to categorise URLs, a number of trends were taken into consideration, which included statuses with unique URLs. URL spam can be detected using a Markov chain model that is then converted into a classifier. For the most part, spam URLs are of a higher quality than legitimate URLs. Spam URLs can be identified and filtered out based on these values. The spam URLs can be sent via text message or tweeted. URL spam can be identified and then filtered out by matching specific patterns of URL spam. The features of tweets are first extracted, and then the classifiers are trained on the dataset. Tweets that were captured at a specific time are also captured and sent to the classification model for analysis. When it comes to spam tweet detection, Chao et al. compare the performance of a variety of machine learning algorithms (2015).

Spammers can disseminate their spam through social media by posting or commenting on irrelevant texts. Different machine learning approaches to combating spam comments were discussed by Mansour et al. (2015). Some of the features like post-comment similarity, inter-comment similarity, length of comments, phone and email information, number of words in comments, interval between posts and comments, URL link, black words list, stop words ratio, and word duplication ratio were extracted to detect spam.... Spammers use a variety of methods to disseminate their spam, including OSN. With the rise of online shopping, customer reviews are critical for product sales, but there is no way to monitor who is writing them. It's possible that this type of review spam will have an impact on a product's sales. According to Siddu et al. (2015), spam detection techniques have been reviewed in the literature. Review spam was discovered using sentiment analysis, which looked for a review's rating as a function of the review's content. When a review's content exceeded a predetermined threshold, it was flagged as spam and removed from the site.

V. SPAM DETECTION TECHNIQUES

When it comes to catching spam, data mining techniques play a critical role. Mohammad Noor et al. provided an overview of various social media data mining techniques (2016). Data quality was assessed using quality assessment rules. If the quality score is greater than some threshold, the data is reliable. This quality score was maintained as a cutoff point. Some questions could only be answered by using the data extraction method. In addition, we used the reciprocal translation technique. Data mining was also used to find the anomalies.

For anomaly detection, a variety of techniques were used: behavior-based techniques, structural methods using graphs, and spectral methods, all of which were discussed in Ravneet and colleagues (2016). The internal content of send and receive messages was analysed to detect unusual behaviour using the behavior-based technique's content-based filtering. A graph metric was determined for each node using the structure-based technique, and the nodes had different values when compared to those of the anomalous users. A simplified algorithm known as SPCTRA was used to partition the network using a spectral-based technique that eliminated links between nodes. Subgraphs with dense subgraphs were created for attackers and anomalous users. Tingmin et al. carry out the word-to-vector identification of spam messages in twitter (2017b). Spam detection in OSNs is done using the usual methods. Manajit et al. conducted a study on the current state of spam detection in social networks (2016). It was used to detect web spam and spammers using a co-classification framework, as well as least-square SVM classifiers (both linear and non-linear). To determine whether a review is spam or not, features such as similarity to other product reviews, similarity to reviews of alternative products, the frequency with which a reviewer or a product is reviewed, and repeatability measures are used.

In Ruxi et al., the classifiers for detecting comment spam in social networks were described (2015). Sample selection, extraction of words from a sample and generation of the results are the four stages of classifier construction. An internet search engine, or "web crawler," is used to gather information about a web page's content and then weed out words that are irrelevant. This was followed by the creation of a classifier and its subsequent validation testing.

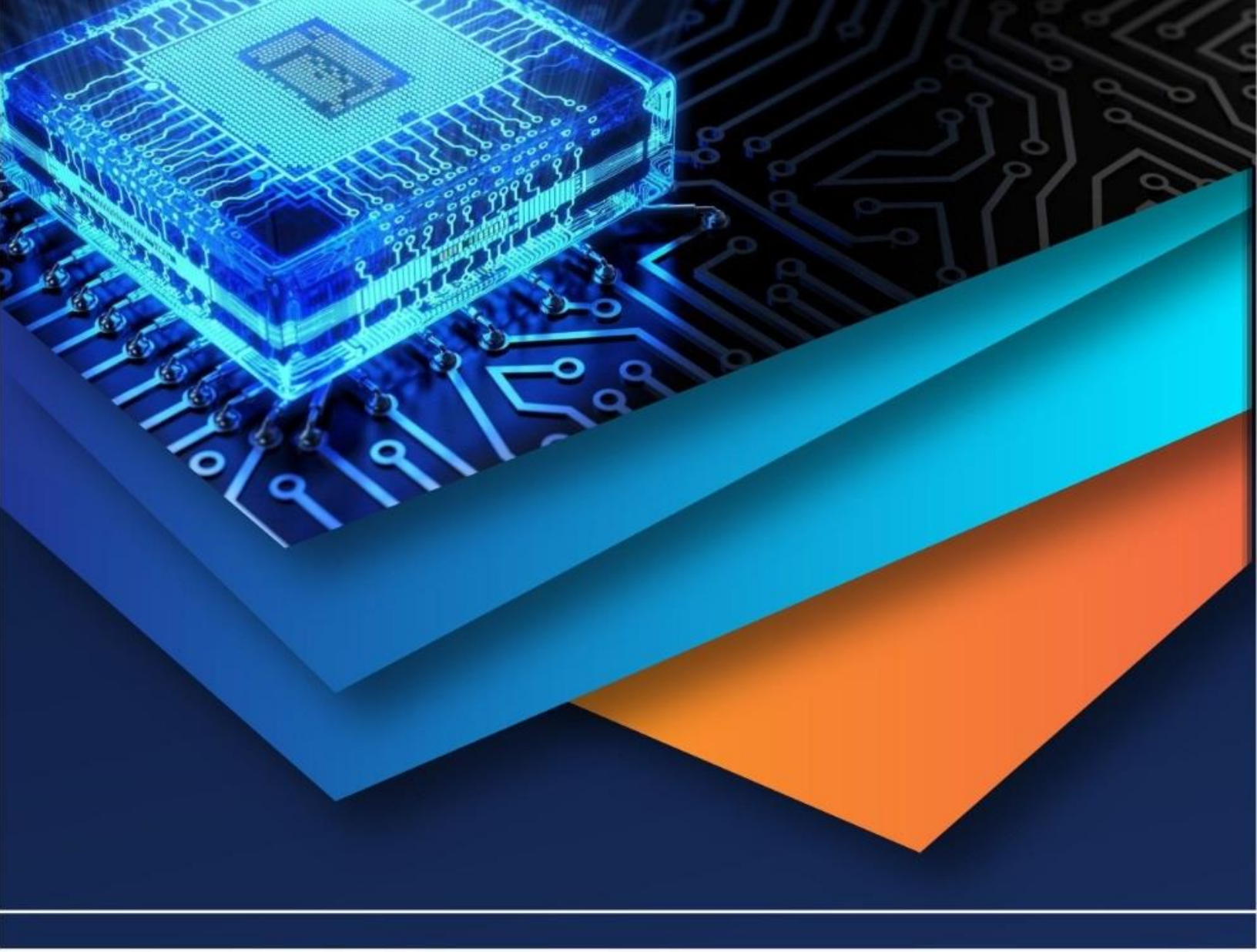
The manual annotation of comments and word segmentation were used to ensure the accuracy of the calculation of comments based on manual tags. Surendra et al. discussed content-based filtering in a semi-supervised manner (2018).

VI. CONCLUSION

E-mail and OSN spam is a problem for users, despite the fact that so many spam message filtering techniques are available in real time. In addition to irritating the users, spam messages can cause financial losses and a loss of trust between the users. In order to identify and filter spam messages without causing problems for the end users and the service providers, an effective method is needed. It is the focus of the proposed model to identify and filter out unwanted emails and messages from the OSN. In order to enhance the filter's performance, the proposed effective spam e-mail identification and filtering mechanism takes into account factors from social media and email datasets. In addition to e-mail features such as trust, reputation, and interest, social network factors such as the strength and degree of connection between the users are taken into account. Logistic regression is used to combine the independent variables. Using the OCR method, spam can be effectively categorised by finding the text in the images that appear in the incoming e-mails.

REFERENCES

[1] Aboli, SV & Rupa, AF 2016, 'Automated content based short text classification for filtering undesired posts on facebook', Proceedings of the IEEE World Conference on Futuristic Trends in Research and Innovation for Social Welfare (Startup Conclave), Coimbatore, pp. 1-5.
[2] Adarsh, MJ & Ravikumar, P 2018, 'An effective method of predicting the polarity of airline tweets using sentimental analysis', Proceedings of the IEEE 4th International Conference on Electrical Energy Systems (ICEES), pp. 676-679.
[3] Adel, HM & Raed, AZ 2011, 'Application of genetic optimized artificial immune system and neural networks in spam detection', Elsevier Journal of Applied Soft Computing, vol. 11, no. 4, pp. 3827-3845.
[4] Ali, AA & El-Sayed, ME 2015, 'Dendritic cell algorithm for mobile phone spam filtering', Elsevier Journal of Procedia Computer Science, vol. 52, pp. 244-251.
[5] Aliaksandr, B & Petr, H 2018, 'Spam filtering using integrated distribution- based balancing approach and regularized deep neural networks', Springer Journal of Applied Intelligence, vol. 48, no. 10, pp. 3538-3556.
[6] Amany, AN, Neveen, IG & Afaf, AS 2018, 'Antlion optimization and boosting classifier for spam email detection', Elsevier Journal of Future Computing and Informatics, vol.3, no. 2, pp. 436-442.
[7] Ashraf, A, Zanaty, EA & Ghoniemy, S 2013, 'Improving the Classification Accuracy using SVM (SVMs) with New Kernel', Journal of Global Research in Computer Science, vol. 4, pp. 1-7.
[8] Bing, X, Mengjie, Z, Will, NB & Xin, Y 2016, 'A survey on evolutionary computation approaches to feature selection', IEEE Transactions on Evolutionary Computation, vol. 20, no. 4, pp. 606-626.
[9] Brian, W & Tong, L 2009, 'Channel e-mail: a sociotechnical response to spam', IEEE Transactions on Computer, vol. 42, no. 7, pp. 63-72.
[10] Chao, C, Jun, Z, Yi, X, Yang, X, Wanlei, Z, Mohammad, MH, Abdulhameed A & Majed, A 2015, 'A performance evaluation of machine learning based streaming spam tweets detection', IEEE Transactions on Computational Social Systems, vol. 2, no. 3, pp. 65-76.
[11] Chao, C, Yu, W, Jun, Z, Yang, X, Wanlei, Z & Geyong, M 2017, 'Statistical features-based real-time detection of drifted twitter spam', IEEE Transactions on Information Forensics and Security, vol. 12, no. 4, pp. 914-925.
[12] Chao, Y, Robert, H & Guofei, G 2013, 'Empirical evaluation and new design for fighting evolving twitter spammers', IEEE Transactions on Information Forensics and Security, vol. 8, no. 8, pp. 1280-1293.
[13] Chenwei, L, Jiawei, W & Kai, L 2016, 'Detecting spam comments posted in micro – blogs using self-extensible spam dictionary', Proceedings of the IEEE International Conference on Communications (ICC), Kuala Lumpur, pp. 1-7.
[14] Chi-Yao, T, Pin-Chieh, S, & Ming-Syan C 2011, 'Cosdes: a collaborative spam detection system with a novel e-mail abstraction scheme', IEEE transactions on Knowledge And Data Engineering, vol. 23, no. 5, pp. 669-682.
[15] Chunyao, S & Tingjian, G 2015, 'Window-chained longest common subsequence: common event matching in sequences', Proceedings of the IEEE 31st International Conference on Data Engineering, Seoul, pp. 759-770.
[16] Clotilde, L, Paulo, C, Pedro, S, Miguel, R & Miguel, R, 2011, 'Symbiotic filtering for spam email detection', Elsevier journal of Expert Systems with Applications, vol. 38, no. 8, pp. 9365-9372.
[17] Congfu, X, Baojun, S, Yunbiao, C, Weike, P & Li, C 2014, 'An adaptive fusion algorithm for spam detection', IEEE Intelligent Systems, vol. 29, no. 4, pp. 2-8.
[18] David, R, Florentino F & Jose, RM 2018, 'Concept drift in e-mail datasets: An empirical study with practical implications', Elsevier journal of Information Sciences, vol. 428, pp. 120-135.
[19] David, R, Florentino, F & Jose, RM 2018, 'Using evolutionary computation for discovering spam patterns from e-mail samples', Elsevier journal of Information processing and management, vol. 54, pp. 303-317.
[20] DeWang & Calton Pu 2015, 'BEAN: A behaviour analysis approach of url spam filtering in twitter', Proceedings of the IEEE International Conference on Information Reuse and Integration, San Francisco, pp. 403-410.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY