A Comprehensive Survey of Spam Profile Detection Methods in Online Social Networks

R. Krithiga¹, Dr. E. Ilavarasan²

¹Assistant Professor, Department of Computer Applications, Perunthalaivar Kamaraj Arts College, Puducherry
²Professor, Department of Computer Science & Engineering, Pondicherry Engineering College, Puducherry

kriithiga@gmail.com¹, eilavarasan@pec.edu²

Abstract—Social networks have grown into a popular way for internet surfers to interact with friends in addition to family members, reading news, and also discuss events. Users spend more time on popular social platforms (e.g., Facebook, Twitter, etc.) storing and sharing their personal information. This fact together with the prospect of communicating thousands of users fascinates the concentration of malicious users. They exploit the implicit trust interactions concerning users with the purpose of accomplishing their malicious objectives, for instance, create malicious links inside the posts/tweets, spread fake news, send out unsolicited messages to genuine users and so on. In this paper, we reviewed various existing techniques on spam profile detection in online social networks.

Index Terms—Spam detection, Online Social Networks, Machine learning techniques, Social network security.

1. Introduction

Recently, Social networking sites (SNSs) have been changed into a crucial and prominent medium of communication and sharing of knowledge. Ultimately, SNS users are liable for sharing knowledge in the network as they are content providers in the network structure. The communities which include families, group of friends, and acquaintances are the primary component in the structure of a network. In SNSs, users can share information through posting links to their favorite web pages, files, photos, and videos.

Moreover, the SNS communities’ structure produces a network of credibility in addition to trust (Lee, 2015 and Singh, et.al, 2017). Facebook and Twitter are the most important SNSs. As per the report from Statista (Statista report, 2017), the total number of Facebook and Twitter users stood at 1,968 million and 319 million, respectively, as of April 2017. With the increasing number of users, an enormous amount of diverse knowledge is also being produced every day on these two SNSs (Sharma, et.al, 2017). Mainly, multimedia knowledge (in a text, audio files, videos, and images form) is produced, stored, and transferred in vast amounts. The multimedia knowledge posted on SNSs is often accompanied by user likes, comments, tags, hashtags, and so on.

Generally, spammers on Facebook as well as Twitter pose as legitimate users. Consequently, recognizing them in addition to distinguishing from legitimate users is a challenging task. In the past, spammers on these SNSs were usually simple, with a clear appearance that assisted in distinguishing them from legitimate users. However, spammers can still utilize inexpensive automated approaches on behalf of acquisition of credibility in addition to trust and making them hard to perceive in the huge population of SNS users. Detection of a spammer in SNSs is a classification issue in which legitimate users are well-known from spammers based on their corresponding features. Note, however, that these classification problems are associated with several unique challenges when attempting to identify spammers. These challenges includes high dimensionality of features, biased and less number of training sets, great computational classification difficulties, and public
inaccessibility of training sets. To overcome these challenges and achieve spammer classification accuracy, a number of traditional machine-learning techniques, like Support Vector Machine (Benevenuto, et.al, 2010 and Zheng, et.al, 2015), Decision Tree (Amleshwaram, et.al, 2013), Jrip (Ahmed, et.al, 2013), Bayesian Network (Rathore, et.al, 2017), Random Forest (Liu, et.al, 2016), k-Nearest Neighbors (Herzallah, et.al, 2017), etc. were used in the previous researches. Due to the high dimensionality of features, however, determining the optimum parameters for parametric supervised procedures are very slow and challenging (Zheng, et.al, 2016). Hence, existing models suffer from numerous problems, like poor generalization performance, extended training time, and greater false positive rates. Additionally, several methods, similar to (Miller, et.al, 2014), utilized biased dataset enclosing a much smaller number of spam profiles than legitimate ones in addition to offering imprecise classification outcomes.

2. Literature Review

This section deals with the previous studies related to spam profile detection in Online Social Networks (OSN). Some of the noteworthy studies include the following.

Aswani, et.al, in 2018 proposed a novel hybrid approach by combining analytics from social media and bio-inspired computing intended with the view of identifying spam profiles in twitter. The K-Means algorithm was combined with Levy flight Firefly Algorithm (LFA) along with chaotic maps. For this study, a total of 18,44,701 tweets were examined from a total of 14,235 Twitter profiles based on 13 statistically significant attributes from OSN analytics. The outcomes exhibited an accuracy of around 97.98% by examining the tweets based on those 13 attributes. One of the major disadvantages observed from the proposed system was the lack of consideration of content and semantics aspects, as most of current tweets include parody besides usage of non-English terminologies of millennial language (Aswani, et.al, 2018).

In the same year, Fu, et.al, developed a dynamic metric for measuring the alterations in activities of users through the influence of progressive evolution patterns of OSN users. This was made accomplished by combining unsupervised as well as supervised learning with the view of distinguishing the variances among spammers and legitimate users. A real-world dataset comprising of a huge sum of users were tested under this methodology. The proposed technique effectively classified spammers and legitimate users based on their patterns of temporal evolution. Similarly, it demonstrated an extended level of resemblances in temporal evolution patterns of spammers (Fu, et.al, 2018).

Similarly, Dutta et.al developed an attribute selection methodology for enhancing the classification accuracy of spams, which provided better classification results by selecting a smaller subset of attributes. This attribute assortment algorithm was developed by applying the conceptions of rough set principles. Experiments were conducted over five different spam classification datasets and were validated with the performance of the proposed algorithm. The proposed technique selected a much smaller subset of attributes than the standard techniques with promising classification results than the other techniques in the literature (Dutta, et.al, 2018).

Eshraqi et.al, in 2015, reviewed various works related to spam detection in social networks. Most of the reviews in this study concentrated on social networking sites like Twitter and Facebook. According to the author, the study has been divided into five sections in which the collection of various data regarding spam profiles in social networks was made as the first phase. The examination of features of spam user account detection was carried out in the second phase. The third phase comprised of detection in features of a spam post while the fourth stage was devoted to the examination of different classification and clustering techniques used for detecting spam. And finally, the fifth stage summarised the results obtained through all the reviews (Eshraqi, et.al, 2015).

Gao, et.al, in 2012 developed a system for filtering online spams which could be organized in the OSN platform for examining the messages generated by users. Under this system, the spam messages rather than being inspected distinctly were reconstructed to their respective campaigns for classification. Almost 187 million wall posts from Facebook and 17 million tweets from Twitter
were collected and examined using this proposed structure. As a result of advanced examination through the proposed system, a message alerting “spam” was dropped before the spam message to assist the users. (Gao, et.al, 2012).

A novel and more sturdy attributes for identifying Twitter spammers were designed by Yang et.al, in 2013. This was based on an in-depth analysis strategy related to evasion which was generally utilized by spammers in Twitter. Under this study, at the initial stage, experimental analyses were made to have an in depth understanding of evasion tactics employed by spammers. Subsequently, 24 attributes for detection process were analysed for their robustness. It can be noted from the study that even under a low false positive rate, the detection rate of this proposed system was considerably higher than the other existing spam indicator techniques under four different prevalent machine learning classifiers (Yang, et.al, 2013).

In 2015, a performance evaluation was made by Chen, et.al, over three distinctive aspects of data, feature as well as a model. For this evaluation, a dataset comprising of almost 600 million public tweets was generated by means of a commercial URL-based security platform. Here, for detecting a tweet with spam message, 12 lightweight features were extracted from the tweet. As an extension, this detection was then transformed into a binary classification problem which could be solved using conventional machine learning algorithms. The performance evaluation was made over the attributes of spam to non-spam ratio, feature discretization, the training data size, data sampling, procedures related with machine learning and data regarding time (Chen, et.al, 2015).

In the year 2017, Liu, et.al suggested a fuzzy-based oversampling method that offered artificial data samples. These samples were obtained from limited observed samples depending on their idea of classification using fuzzy based information. Moreover, an ensemble technique was developed which examined certain classifiers based on their imbalanced data in three phases. It could be concluded from the study that by means of imbalanced class distribution, the rate of spam detection can greatly be influenced (Liu, et.al, 2017).

A set of characteristic attributes from Twitter, concerning network, activity, user, as well as tweet content were utilized in developing a supervised machine learning solution for recognizing cyberbullying on Twitter. This was suggested by Al-garadi et.al, in 2016 which resulted in an f-measure of 0.936 with an area under the receiver-operating characteristic curve of 0.943. These outcomes suggested that the model based on these features offered a reasonable solution for identifying Cyberbullying in online communication backgrounds. Lastly, the outcome acquired with the proposed features was proven the best by comparing with results from two baseline features for their performance (Al-garadi, et.al, 2016).

A novel methodology related to Markov Clustering (MCL) was presented by Ahmed & Abulaish in 2012 for determining the spam profiles over OSNs. A dataset comprising of Facebook profiles (both real and spam profiles) was utilized for the experiment. A weighed graph was applied which helped in modeling the social networks in which the profiles were characterized as nodes and their intersections as edges. Also, in case of a cluster comprising both real and spam profiles, the idea of majority voting was established. The outcomes have revealed that the introduction of majority voting not only reduces the sum of clusters but also increases the performance values of FP and FN measures (Ahmed & Abulaish, 2012).

An effective spammer detection outline which distinguished the legitimate users of Facebook from the spammers was proposed by Rathore, et.al, in 2018. In accordance with recent activities of Facebook, the outline presented the urgency of setting out a new feature set for detecting spams. A baseline dataset was used from Facebook which involved 300 spammers in addition to 700 legitimate user profiles. The baseline dataset encompassed feature set for an individual profile, which were extracted through a novel dataset construction mechanism. Furthermore, an intelligent decision support scheme was employed to differentiate spammers from genuine users. The evaluation outcomes proved that the suggested outline was precise and effective to deliver excellent performance with a maximum accuracy of 0.984 and Mathew correlation coefficient of 0.977 (Rathore, et.al, 2018).
A study related to spot the fake profiles created by people in contrast to those created by bots was reported by Van, et.al, in 2018. Initially, a series of early works regarding spam detection techniques were studied for any resemblance with their report. A number of supervised machine learning models were used with features exclusively engineered for spam profile detection without trusting behavioural data. The results of the study witnessed an F1 score of 49.75% upon detection of fake accounts (Van, et.al, 2018).

An integrated technique with added categories of features like metadata, interaction based as well as content-based attributes to previously existing community-based features for recognizing auto spammers was discussed by Fazil, et.al, in 2018. The novelty of this technique is that the users were categorized based on their exchanges with their followers as ignoring follower’s contents and metadata features were literally difficult. The experiment was made over a real-world dataset comprising of legitimate users as well as the spammers whose profiles were characterized through 19 predefined attributes along with 6 newly defined and 2 redefined features. The results showed that the classifications based on interaction as well as community proved to be effective while metadata classification was the least effective (Fazil, et.al, 2018).

With the view of recognizing spammers in Twitter, a hybrid machine learning model was proposed by hybridizing an existing meta-heuristic technique called Whale Optimization Algorithm (WOA) with Support Vector Machines (SVM). The idea behind the utilization of WOA in this hybrid technique was to enhance the parameters of SVM along with the task of selecting suitable features for spam recognition. Arabic, English, Spanish, as well as Korean lingual contexts, were recognized under this technique. The results have shown a considerable efficiency over other previously developed spam profile detection techniques. (Ala’M, et.al, 2018).

An alternative ontology-based methodology was proposed to detect spam tweets exclusively by means of content analysis. This was carried out as the accuracy of classification was much lower than the expected. The proposed ontology disabled the dependency on private as well as user-relationship information which most of the existing spam detection procedures use. Investigational outcomes demonstrated that the proposed method outperformed existing message to message spam detection methods by approximately 200% (Halawi, et.al, 2018).

A study based on spammers account observed over the OSNs was conducted using behavioural features by Almaatouq, et.al, in 2016. This study was carried out over a huge sum of 100 million tweets from Twitter collected over a span of a month. It identified two behaviourally dissimilar groups of spammers while they utilize different approaches of spamming. Similarly, it illustrated the way in which the legitimate, as well as the spammers in these groups, validate each individual property along with their patterns of social interaction. Finally, these groups were checked for their content attributes, social communication, as well as features related to profile for detecting spam messages (Almaatouq, 2016).

An innovative methodology was proposed that uses Open-source Intelligence (OSINT) along with machine learning techniques for localizing and outlining the movement of a social network user. This structure identified the location of users through their posts thereby raising the privacy issues involved. The outcomes have revealed an estimated accuracy of 77.72% stating that the users can be geo-located through this proposed technique without the use of GPS/GNSS related data (Pellet, et.al, 2019).

A real-time anti-phishing structure comprising of seven distinct procedures of classification along with Natural Language Processing (NLP) based attributes was suggested by Sahingoz, et.al, in 2019. Based on the following properties of language independence, usage of a large amount of phishing as well as authentic data, implementations in real-time, identification of new websites, independent from third-party services and usage of feature-rich classifiers, studies were carried out for recognizing their distinctive features. The proposed system was then tested with a novel dataset comprising of the above mentioned properties. As per the investigational and comparative outcomes from the implemented classification procedures, Random Forest algorithm with only NLP based features provides the improved performance with the 97.98% accuracy rate for recognition of phishing URLs (Sahingoz, et.al, 2019).
Based on the level of focused interest patterns of users, an advanced spammer detection system was proposed by Alghamdi, et al., in 2018. For recognizing the variation in interests of users, quantitative techniques were proposed and defined for the type of interest a user possess i.e. either focussed type or diverse-interest type. Based on the interest level, the users were recognized and a framework was developed by incorporating both unsupervised as well as supervised learning for categorizing between legitimate and spam users (Alghamdi, et al. 2018).

Crowd-retweeting of spam in Sina Weibo, the counterpart of Twitter in China was examined by Liu et al., in 2018. The characteristics of crowd retweeting spammers were carefully analyzed concerning their profile features, social relationships, and retweeting behaviours. The authors found that even though these spammers were probably as close as to legitimate users, the process of basic social connections of crowd retweeting campaigns were totally different from those of prevailing spam campaigns. This was recognized through the unique features of retweets that were spread in the cascade. Based on these outcomes, retweeting-aware link-based ranking algorithms were proposed to conclude more suspicious accounts by preferring recognized spammers as seeds. The evaluation outcomes presented that the procedures were more efficient compared to other link-based strategies (Liu, et al., 2018).

In 2018, a structured classification approach for spammer detection problem on OSNs was suggested by El-Mawass et al. This approach influenced the similarities amidst a range of users for transmitting the beliefs about their labels. A Markov Random Field model was generated in the form of a similarity graph by associating various analogous profiles based on their shared applications. It not only allowed the detection system to be more sustainable but utilized in designing adaptive classifiers. It could be concluded from the study that the detection accuracy of any technique can be influenced through similarities between user profiles (El-Mawass, et al., 2018).

A qualitative examination was made over the advantages and disadvantages of various previously developed spam detection techniques. Every detection technique was comprehensively reviewed and as a result, a discussion comprising of several gaps prevalent in existing approaches and the corresponding actions needed to address them were clearly stated (Kaur, et al., 2018).

The automatic discovery of cyber security related social network based on three diverse sets of a feature in addition to three unique machine learning approaches was proposed by Aslan et al., in 2018. Investigational outcomes presented a favorable performance with an extraordinary accuracy of 95%. The uppermost score (over 97%) was attained by combining the random forests technique with behavioural features. Maintaining a list of cyber security linked accounts manually requires domain-specific knowledge and consumes human efforts. As an outcome of this study, it was stated that a list can automatically be maintained which would help in complicated analysis like cyber security attacks, human behaviours of cyber criminals and cybersecurity experts by automatically checking the accounts in the list (Ashlan, et al., 2018).

A set of features were adopted for recognizing spammers on Twitter along with some additional features for enhancing the performance of classifiers by Hanif, et al., in 2018. WEKA and RapidMiner were used for estimating the performance of four machine learning algorithms namely Random forest (RF), Support vector machine (SVM), K nearest neighbor (KNN), and Multilayer Perceptron (MLP). The experiment showed that SVM, KNN, and MLP on WEKA outperformed those techniques on RapidMiner. However, in the case of RF, RapidMiner attained greater accuracy compared to RF on WEKA. Based on the 32 features in the dataset, MLP in addition to RF on both WEKA and RapidMiner outperformed compared to other classifiers with an accuracy of 95.42% and 95.44% respectively (Hanif, et al., 2018).

In 2018, machine learning based malicious account detection scheme called DeepScan for Location-Based Social Networks (LBSNs) was proposed by Gong et al. By separating every user’s activity information into many continuous time intervals, the users’ activities were measured. Through deep learning technologies, DeepScan makes application of time series features, which were more discriminative and informative, compared to conventional features, specifically statistical or demographic features. Through the real data collected from Dianping, the authors found that the DeepScan achieved an outstanding performance with an F1-score of 0.964 (Gong, et al., 2018).
A novel dataset based on various features collected from numerous latest research works was created for developing a more robust as well as accurate spammer detection model. This dataset was then tested with various standard classifiers for their prediction performance. Furthermore, an extended analysis was made for identifying the features that possess a higher impact of accuracy in detecting spam. For analysis, three different methodologies were chosen: change of mean square error (CoM), information gain (IG) and Relief-F method and compared for their performance effectiveness. The attributes such as the reputation of a profile, average words in a tweet, average mention per tweet, the active period of an account and average time within posts contributes more in detecting spammers in OSNs (Herzallah, et.al, 2018).

A Bagging ELM-based spammer detection framework for SNSs was proposed by Rathore et al., in 2018. The proposed framework had three major contributions in the area. Initially, it identified account and object-specific features to enable spammer detection in SNSs. Next, it built a novel dataset of the two most popular SNSs, i.e., Twitter and Facebook. Lastly, it presented a Bagging ELM classifier and applied this classifier to the dataset that was constructed from Twitter and Facebook. The experiments and assessment with other classifiers showed that the proposed structure was capable of attaining improved generalization performance than other frameworks. The proposed system attained an average rate of 99.01% accuracy for the Twitter dataset and 99.02% for the Facebook dataset while requiring shorter training time of 1.17s and 1.10s, individually (Rathore, et.al, 2018).

Perez et al., (2018) considered Twitter as a case study to measure the distinctiveness of the relationship concerning metadata and user uniqueness and comprehend the effects of potential obfuscation approaches. More precisely, atomic fields in the metadata were examined and scientifically integrated with a view of classifying new tweets belonging to an account through various machine learning algorithms. The authors demonstrated that, with the application of supervised learning procedure, they were capable of recognizing any user from a pool of 10,000 profiles with around 96.7% accuracy. (Perez, et.al, 2018).

A graph-based technique for identifying tourist movement patterns from Twitter information was presented by Hu et al. Initially, tweets collected using geo-tags were gutted to filter those not issued by tourists. Next, a DBSCAN-based clustering technique was modified to construct tourist graphs containing the tourist attraction vertices and edges. Third, network analytical approaches were used to identify tourist movement patterns, including popular and centric attractions, and popular tour routes. The New York city in US was considered to validate the proposed methodology utility. The detected tourist movement patterns support business in addition to government activities such as tour planning, transportation, and enlargement of both shopping and accommodation centers (Hu, et.al, 2018).

Wani et al recommended a model that can be opted by the OSN service providers to alert their users with a list of suspicious connection (links) from their corresponding friend lists so that users can themselves authenticate the recommended links and filter their friend list based on the alerts. Data from 1000 interconnected Facebook profiles were collected for designing the classifier proposed using mutual clustering coefficient metric (Wani, et.al, 2018).

Sohrabi et al. recommended a novel technique regarding the recognition of spam comments on Facebook. In this study, an online spam filtering scheme was considered through revising the posts in addition to comments and reviewing their features. The suggested filtering system was capable of exploiting different techniques and optimization procedures such as simulated annealing, PSO, ACO, and differential evolution to perceive and remove the impish contents and avoids issuing spam comments to offer a secure environment for users of the standard social network. Moreover, supervised machine learning techniques, clustering methods, and decision trees were used for the suggested filtering scheme (Sohrabi, et.al, 2018).

In 2010, an exploration regarding the limits of classifying social spam profiles on MySpace was presented by Irani, et.al. The study focussed on zero-minute fields or commonly known as static fields which were presented during the creation of a profile in order to try and determine whether this sort of technique would be feasible in preventing spammers. The results displayed a standard C4.5
A feature-based technique was developed along with a supervised learning method for detecting spam posts from Instagram. K-fold cross-validation was utilized for determining the optimum pair of supervised learning model as well as the parameters for the model. Many profiles from Instagram were collected and for marking media posts quickly a two-pass clustering method (i.e. Minhash clustering and K-medoids clustering) was utilized. This was primarily formed in order to group the near-duplicate posts into a similar form of clusters. It can be witnessed that the accuracy of this proposed system was around 96.27% (Zhang & Sun, 2017).

Adikari & Dutta studied the minimal set of profile data that were required for determining the fake profiles in LinkedIn and identified the appropriate data mining technique for carrying out such a task. In this study, two predominant techniques of data mining namely RBF kernel and Polynomial kernel were compared for their accuracy and a better technique was chosen among these. Although the former technique remains as the most commonly utilized method in data mining context, the latter provides higher accuracy. The accuracy of the proposed polynomial kernel was found to be 84.04% (Adikari & Dutta, 2014).

A new set of features through EdgeRank algorithm was proposed for detecting the video spammers on YouTube. This study comprised of experiments with nine classifiers of various learning, decision tree, function-based and Bayesian techniques. Through the study, it was witnessed that a maximum average of 98 percent was observed over Bayes Network and Naïve Bayesian network while the minimum was found to be 94 % percent with LibLINEAR technique (Yusof & Sadoon, 2017).

| S.No | Author | Advantages | Limitations | Social Networks examined | Metrics | Size of Dataset |
|------|--------|------------|-------------|-------------------------|---------|-----------------|
| 1.   | Gao et al., (2012) | Extraordinary accuracy, latency at lower levels and high throughput were some of the advantages of this system. Also, they incur low cost for maintenance after deployment. | Although this system detects fake accounts in Facebook and Twitter, it cannot be greatly used as they are not generic. They focus only on detecting accounts that spread URLs through messages. | Facebook and Twitter. | High accuracy, low latency and high throughput. | 187 million wall posts from Facebook and 17 million tweets from twitter. |
| 2.   | Yang et al., (2013) | While keeping an even lower false positive rate, the detection rate using the new proposed feature set is significantly higher than that of existing features. | It uses a crawled dataset which may still possess a sampling bias. This deteriorates the accuracy of spam profile detection technique. | Twitter | High detection rate, low false positive rate. | 20000 twitter accounts |
|   | Authors, Year   | Description                                                                                                                                                                                                 | Dataset/Performance Metrics                                                                 | Type of Data/Features                                                                 |
|---|-----------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------|-----------------------------------------------|
| 3. | Chen et al., (2015) | In the evaluation, it was found that the classifiers’ ability to detect Twitter spam reduced when in a near real-world scenario since the imbalanced data brings bias. The proposed method overcome this scenario effectively. | Performance decreases because of the distribution of features changes of later days’ dataset, whereas the distribution of training dataset stays the same. This problem will exist in streaming spam tweets detection, as the new tweets are coming in the forms of streams, but the training dataset is not updated. | Twitter detection ability, feature discretization, spam detection rate. 600 million tweets |
| 4. | Liu et al., (2017) | Improves the spam detection performance on imbalanced Twitter datasets with a range of imbalance degrees.                                                                                                   | The synthetic data generation scheme to incorporate correlations among features was not done. | Twitter spam detection performance 600 million tweets |
| 5. | Al-garadi et al., (2018) | The proposed model can be used by the organization’s members as well as non-government organizations (NGOs), including crime-prevention foundations, social chamber organizations, psychiatric associations, policymakers, and enforcement bodies. | Not utilized the data from other social media to investigate cyber bullying behaviours.        | Twitter Precision, F-measure, recall and AUC. 2.5 million geotagged tweets |
| 6. | Jantan et al., (2017) | FFNN-EBAT has better quality results than the other training algorithms.                                                                                                                                      | Not verified the performance of the EBAT algorithm with other neural network types. (e-mails) | Mean, median and standard deviation of Mean Square Error (MSE). 8467 e-mail instances |
|   | Author(s)                        | Description                                                                 | Platform(s)          | Efficiency Criteria | Dataset Size |
|---|----------------------------------|-----------------------------------------------------------------------------|----------------------|---------------------|---------------|
| 7. | Rathore et al., (2018)           | More accurate and efficient to deliver excellent performance. Fast response and high accuracy rate for spammer detection on Facebook. | Facebook             | Efficiency and accuracy. | 1000 Facebook user profiles. |
| 8. | Ala'M et al., (2018)             | This model can be helpful in designing more accurate and insightful spam detection models for online social networks. The model can also be helpful in identifying the most influencing features. | Twitter              | Accuracy, Precision, F-measure and AUC | 800 instances |
| 9. | Pellet et al., (2019)            | Localization of user was made possible through the developed social graph technique. | Twitter, Instagram and Facebook | Prediction accuracy. | 5447 instances |
| 10. | Singh & Batra (2018)             | The model can be successfully deployed for spam detection in massive data sets. The framework was tested for tweets generated by Twitter alone which demands further enhancement to | Twitter              | Performance, Precision, F-measure and recall. | - |
3. Conclusion
Due to the increasing popularity and intensive use of social networks such as Twitter, Facebook, LinkedIn, etc. the number of spammers are rapidly growing. This has resulted in the development of several spam detection techniques. From the papers reviewed, it can be concluded that most of the work have been developed using classification approaches like SVM, Decision Tree, Naïve Bayesian, and Random Forest, KNN. The effectiveness of methods has been determined based on the metrics such as detection accuracy, efficiency, false prediction rate, F-Measure, Precision and recall. Detection rate is measured from user-based or content based features or a combination of both and network level features. It is to be noted that spam detection methods have been proposed considering only any one of the of social network type which is not effective in the current scenario as an individual has accounts across various social networks. There is a huge scope of improvements in terms of unified system for spam identification, introduction and exploration of new features which would account the activities of users at various time frames and autonomous system that updates itself at periodic intervals so as to defend against the evolving and invading spammers.

4. REFERENCES
[1]. Adikari, S., & Dutta, K. (2014), “Identifying Fake Profiles in LinkedIn”, In PACIS (p. 278).
[2]. Ahmed, F., & Abulaish, M. (2012, June), “An mcl-based approach for spam profile detection in online social networks”. In Trust, Security and Privacy in Computing and Communications (TrustCom), 2012 IEEE 11th International Conference on (pp. 602-608), IEEE.
[3]. Ahmed, F., & Abulaish, M. (2013), “A generic statistical approach for spam detection in Online Social Networks”, Computer Communications, 36(10-11), 1120-1129.
[4]. Ala’M, A. Z., Faris, H., & Hassonah, M. A. (2018), “Evolving Support Vector Machines using Whale Optimization Algorithm for spam profiles detection on online social networks in different lingual contexts”, Knowledge-Based Systems, 153, 91-104.
[5]. Algaradi, M. A., Varathan, K. D., & Ravana, S. D. (2016), “Cybercrime detection in online communications: The experimental case of cyberbullying detection in the Twitter network”, Computers in Human Behavior, 63, 433-443.
[6]. Alghamdi, B., Xu, Y., & Watson, J. (2018, November), “A Hybrid Approach for Detecting Spammers in Online Social Networks”, In International Conference on Web Information Systems Engineering (pp. 189-198), Springer, Cham.
[7]. Almaatouq, A., Shmueli, E., Nout, M., Alabdulkareem, A., Singh, V. K., Alsaleh, M., & Pentland, A. S. (2016), “If It Looks Like a Spammer and Behaves Like a Spammer, It Must Be a Spammer. Analysis and Detection of Microblogging Spam Accounts”, International Journal of Information Security 15.5 (2016): 475–491.
[8]. Amleshwaram, A. A., Reddy, N., Yadav, S., Gu, G., & Yang, C. (2013, January), “Cats: Characterizing automation of twitter spammers”, In Communication Systems and Networks (COMSNETS), 2013 Fifth International Conference on (pp. 1-10). IEEE.
[9]. Aslan, Ç. B., Sağlam, R. B., & Li, S. (2018), “Automatic Detection of Cyber Security Related Accounts on Online Social Networks: Twitter as an example”, In: 9th International Conference on Social Media & Society (SM&Society 2017), 18-20 July 2018, Copenhagen, Denmark. (doi:https://doi.org/10.1145/3217804.3217919)
[10]. Aswani, R., Kar, A. K., & Ilavarasan, P. V. (2018), “Detection of spammers in twitter marketing: A hybrid approach using social media analytics and bio-inspired computing”, Information Systems Frontiers, 20(3), 515-530.
[11]. Benevenuto, F., Magno, G., Rodrigues, T., & Almeida, V. (2010, July), “Detecting spammers on twitter” In Collaboration, electronic messaging, anti-abuse and spam conference (CEAS) (Vol. 6, No. 2010, p. 12).

[12]. Chen, C., Zhang, J., Xie, Y., Xiang, Y., Zhou, W., Hassan, M. M., & Alrubaian, M. (2015), “A performance evaluation of machine learning-based streaming spam tweets detection”, IEEE Transactions on Computational social systems, 2(3), 65-76.

[13]. Dutta, S., Ghatak, S., Dey, R., Das, A. K., & Ghosh, S. (2018), “Attribute selection for improving spam classification in online social networks: a rough set theory-based approach”, Social Network Analysis and Mining, 8(1), 7.

[14]. El-Mawass, N., Honeine, P., & Vercouter, L. (2018, July), “Supervised Classification of Social Spammers using a Similarity-based Markov Random Field Approach”, In Proceedings of the 5th Multidisciplinary International Social Networks Conference (p. 14). ACM.

[15]. Eshraqi, N., Jalali, M., & Moattar, M. H. (2015, November), “Spam detection in social networks: A review”, In Technology, Communication and Knowledge (ICTCK), 2015 International Congress on (pp. 148-152). IEEE.

[16]. Fazil, M., & Abulaish, M. (2018), “A Hybrid Approach for Detecting Automated Spammers in Twitter”, IEEE Transactions on Information Forensics and Security, 13(11), 2707-2719.

[17]. Fu, Q., Feng, B., Guo, D., & Li, Q. (2018), “Combating the evolving spammers in online social networks”, Computers & Security, 72, 60-73.

[18]. Gao, H., Chen, Y., Lee, K., Palsetia, D., & Choudhary, A. N. (2012, February), “Towards Online Spam Filtering in Social Networks”, In NDSS (Vol. 12, pp. 1-16).

[19]. Gong, Q., Chen, Y., He, X., Zhuang, Z., Wang, T., Huang, H., & Fu, X. (2018), “DeepScan: Exploiting Deep Learning for Malicious Account Detection in Location-Based Social Networks”, IEEE Communications Magazine, Feature Topic on Mobile Big Data for Urban Analytics, 56(1).

[20]. Halawi, B., Mourad, A., Otrok, H., & Damiani, E. (2018), “Few are as Good as Many: An Ontology-Based Tweet Spam Detection Approach”, IEEE Access.

[21]. Hanif, M. H. M., Adewole, K. S., Anuar, N. B., & Kamsin, A. (2018), “Performance Evaluation of Machine Learning Algorithms for Spam Profile Detection on Twitter Using WEKA and RapidMiner”, Advanced Science Letters, 24(2), 1043-1046.

[22]. Herzallah, W., Faris, H., & Adwan, O. (2018), “Feature engineering for detecting spammers on Twitter: Modelling and analysis”, Journal of Information Science, 44(2), 230-247.

[23]. Hu, F., Li, Z., Yang, C., & Jiang, Y. (2018), “A graph-based approach to detecting tourist movement patterns using social media data”, Cartography and Geographic Information Science, 1-15.

[24]. Irani, D., Webb, S., & Pu, C. (2010, May), “Study of Static Classification of Social Spam Profiles in MySpace” In Icwsm.

[25]. Jantan, A., Ghanem, W. A., & Ghaleb, S. A. (2017), “Using Modified Bat Algorithm To Train Neural Networks For Spam Detection”, Journal of Theoretical & Applied Information Technology, 95(24).

[26]. Kaur, R., Singh, S., & Kumar, H. (2018), “Rise of spam and compromised accounts in online social networks: A state-of-the-art review of different combating approaches”, Journal of Network and Computer Applications.

[27]. Lee, D. H. (2015), “Personalizing Information Using Users’ Online Social Networks: A Case Study of CiteULike”, JIPS, 11(1), 1-21.
[28]. Liu, B., Ni, Z., Luo, J., Cao, J., Ni, X., Liu, B., & Fu, X. (2018), “Analysis and defense against crowd-retweeting based spam in social networks”, World Wide Web, 1-23.

[29]. Liu, L., Lu, Y., Luo, Y., Zhang, R., Itti, L., & Lu, J. (2016), “Detecting Smart Spammers On Social Network: A Topic Model Approach”, arXiv preprint arXiv:1604.08504.

[30]. Liu, S., Wang, Y., Zhang, J., Chen, C., & Xiang, Y. (2017), “Addressing the class imbalance problem in Twitter spam detection using ensemble learning”, Computers & Security, 69, 35-49.

[31]. Miller, Z., Dickinson, B., Deitrick, W., Hu, W., & Wang, A. H. (2014), “Twitter spammer detection using data stream clustering”, Information Sciences, 260, 64-73.

[32]. Mv Ngo Tien HoA, High Speed And Reliable Double Edge Triggered D- Flip-Flop For Memory Applications”, Journal of VLSI Circuits And Systems, 1 (01), 13-17,2019

[33]. MN BORHAN,”Design Of The High Speed And Reliable Source Coupled Logic Multiplexer”, Journal of VLSI Circuits And Systems 1 (01), 18-22,2019

[34]. Metadata Information”, arXiv preprint arXiv:1803.10133.

[35]. Rathore, S., Loia, V., & Park, J. H. (2018), “SpamSpotter: an efficient spammer detection framework based on intelligent decision support system on facebook”, Applied Soft Computing, 67, 920-932.

[36]. Rathore, S., Sangaiah, A. K., & Park, J. H. (2018), “A novel framework for internet of knowledge protection in social networking services”, Journal of computational science, 26, 55-65.

[37]. Sahingoz, O. K., Buber, E., Demir, O., & Diri, B. (2019), “Machine learning based phishing detection from URLs”, Expert Systems with Applications, 117, 345-357.

[38]. Sharma, P. K., Rathore, S., & Park, J. H. (2017), “Multilevel learning based modeling for link prediction and users’ consumption preference in Online Social Networks”, Future Generation Computer Systems.

[39]. Singh, A., & Batra, S. (2018), “Ensemble-based spam detection in social IoT using probabilistic data structures”, Future Generation Computer Systems, 81, 359-371.

[40]. Sohrabi, M. K., & Karimi, F. (2018), “A feature selection approach to detect spam in the Facebook social network”, Arabian Journal for Science and Engineering, 43(2), 949-958.

[41]. Statista, Most famous social network sites worldwide as of April 2017, ranked by a number of active users (in millions). https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/ (accessed 13 September 2017).

[42]. Van Der Walt, E., & Eloff, J. (2018), “Using Machine Learning to Detect Fake Identities: Bots vs. Humans”, IEEE Access, 6, 6540-6549.

[43]. Wani, M. A., & Jabin, S. (2018), “Mutual Clustering Coefficient-based Suspicious-link Detection approach for Online Social Networks”, arXiv preprint arXiv:1805.00537.

[44]. Yang, C., Harkreader, R., & Gu, G. (2013), “Empirical evaluation and new design for fighting evolving twitter spammers”, IEEE Transactions on Information Forensics and Security, 8(8), 1280-1293.

[45]. Yusof, Y., & Sadoon, O. H. (2017), “Detecting video spammers in youtube social media”, in Zulikha, J. & N. H. Zakaria (Eds.), Proceedings of the 6th International Conference of Computing & Informatics (pp 228-234). Sintok: School of Computing.

[46]. Zhang, W., & Sun, H. M. (2017, January). “Instagram Spam Detection”. In Dependable Computing (PRDC), 2017 IEEE 22nd Pacific Rim International Symposium on (pp. 227-228). IEEE.
[47]. Zheng, X., Zeng, Z., Chen, Z., Yu, Y., & Rong, C. (2015), “Detecting spammers on social networks”, Neurocomputing, 159, 27-34.

[48]. Zheng, X., Zhang, X., Yu, Y., Kechadi, T., & Rong, C. (2016), “ELM-based spammer detection in social networks”, The Journal of Supercomputing, 72(8), 2991-3005

[49]. MN BORHAN,”Design Of The High Speed And Reliable Source Coupled Logic Multiplexer”, Journal of VLSI Circuits And Systems 1 (01), 18-22,2019

[50]. Mv Ngo Tien HoA, High Speed And Reliable Double Edge Triggered D- Flip-Flop For Memory Applications”, Journal of VLSI Circuits And Systems, 1 (01), 13-17,2019

Author Biography

First Author

Mrs. R. Krithiga, currently works as an Assistant Professor, Department of Computer Applications at Perunthalaivar Kamarajar Arts College, Madagadipet, Puducherry. She completed her Master of Computer Science from Pondicherry University. She is also a research scholar at Pondicherry Engineering College. Her area of interests includes Evolutionary algorithms, Data mining and Intelligent systems.

Second Author

Dr. E. Ilavarasan, is a Professor of Department of Computer Science & Engineering at Pondicherry Engineering College, Puducherry. He has an experience of more than 25 years in the teaching field. He is expertized in Web service computing.