ISSN No. 0976-5697

Volume 11, Special Issue I, May 2020

International Journal of Advanced Research in Computer Science

CONFERENCE PAPER

Available Online at www.ijarcs.info

HETEROGENEOUS ENSEMBLE STRUCTURE BASED UNIVERSAL SPAM PROFILE DETECTION SYSTEM FOR ONLINE SOCIAL MEDIA NETWORKS

Vinod A. M.
17TCS20, M.Tech, (CSE)
School of Computing and Information Technology
Reva University, Bangalore
viutaurus@gmail.com

Prof. Sathish. S. C.
Associate Professor,
School of Computing and Information Technology
Reva University, Bengaluru
sathish_gc@reva.edu.in

***Abstract*-The exponential rise in internet technology and online social media networks have revitalized human-being to connect and socialize globally irrespective of geographical and any demographic boundaries. Additionally, it has revitalized business communities to reach target audiences through social media networks. However, as parallel adverse up-surge the ever-increasing presence of malicious users or spam has altered predominant intend of such social media network by propagating biased contents, malicious contents and fraud acts. Avoiding and neutralizing such malefic users on social media network has remained a critical challenge due to gigantically large size and user's diversity such as Facebook, Twitter, and LinkedIn etc. Though exploiting certain user's behavior and content types can help identifying malicious users, majority of the existing methods are limited due to confined parametric assessment, and inferior classification approaches. With intend to provide spam profile detection system in this paper a novel heterogeneous ensemble-based method is developed. The proposed model exploits user profile features, user's activity features, location features and content features to perform spam user profile detection. To ensure optimality of computational significances, we applied multi-phased feature selection method employing WilcoxonRank Sum test, Significant Predictor test, and Pearson Correlation test, which assured retaining optimal feature sets for further classification. Subsequently, applying an array of machine learning methods, including Logistic regression, decision tree, Support Vector Machine variants with Linear, Polynomial and RBF kernels, Least Square SVM with linear, polynomial and RBF kernels, ANN with different kernels, etc we constituted a robust ensemble model for spam user profile classification. Simulations revealed that the proposed ensemble classification model achieves accuracy and F-score higher than 98%, which is the highest amongst major works done so far. It affirms suitability and robustness of the proposed model for real time spam profile detection and classification on social media platforms.**

Keywords-Social Media Network, Spam User Profile Detection, Heterogeneous Ensemble Learning, multi-phased feature selection.

I. INTRODUCTION

In the last few years, the exponential rise in internet technologies has revitalized socio-economic world to connect peers for making optimal decision and explore knowledge for optimistic potential generation. Amongst the major innovations, Online Social Networking platforms like Facebook, Twitter, Instagram, LinkedIn, MySpaceetc have emerged as the most influential tools across web-horizon [1]. Wish such gigantically huge online presence of users, peer-communication amongst known-unknown users has reached the pinnacle ever. On the other hand, from economic and business perspectives, such social media platforms facilitate private stakeholders or commercial entities as well as governmental agencies an optimal way to reach the audiences, do marketing of the product or services, assess customer behavior and performing opinion mining [2]. The predominant use of social media played a vital role in information sharing. With such dense user presence online social media enables information or targeted content to spread swiftly and easily across the network making it more visible. Unfortunately, being gigantic in nature and flexible authentication such online social media platforms turn out to be more susceptible to social spam attacks and corresponding fake content propagation [3]. A recent study revealed that approximately 83% of the legitimate social media users receive one or more unwanted or fake profile's (friend) request [4]. Moreover, connecting malicious users forces an individual user to see irrelevant posts, contents causing different socio-psychological stress [5]. A recent study also found that malicious links being propagated by such span users have been causing legitimate user's profile clone and taking over access right, which has caused many misdeeds across network. These facts alarm industry to address malicious user presence issue on social media platform by detecting and neutralizing such spam profiles.

One of such aspects is the ever-increasing rate of spamming and spammer profile on online social media platforms [6]. In practice, spam can be in varied forms such as fake profile, fraudulent reviews, click-jaking, advertisement, malicious links, and malicious files. Though, spam emails are the dominant challenge which has been increasing continuously since the 90s, the online fraud or malicious users' profile too has been increased alarmingly[7]. Though, in majority of the existing literatures, spam profile and spam messages are considered distinctly, they possess certain definite inherent associations. For example, spam profile used to have a highly likelihood to propagate spam messages, social-media posts etc [8]. Recent studies reveal that in major cases, spammers often cause or intend to incite nuisance, fake-news, vandalism, and anarchy by generating or posting malicious contents, which as a result causes indignation from users. Undeniably, in the last few years the events of unethical fake posts giving rise to the hatred across-society and violence has increased alarmingly. Additionally, the false advertisements, misguiding news or information too have been found being propagated by spam users or fake users on online social media networks. Factually, spam does not even distinguish other users as adult or young, male or female which as a result increases risk of online exploitation and crimes [9].

Considering the significance of malicious user identification, though numerous research efforts have been made; however, majority of the at hand systems either focuses on spam email detection or user identification with limited public features. On the other hand, the approaches developed so far address public information and content types for a specific social media platform, which turns out to be limited for the other platform. On contrary, for a real-world application a malicious or spam profile detection method is supposed to be robust enough to perform intended task irrespective of the platform (for instance, Facebook, Twitter, Instagram, LinkedIn, etc). In other words, there is the need of a robust common (i.e., applicable to all) approach to perform spam profile detection and classification in online social media networks. Doing so requires applying multiple features including user profile features, user activity features, location features and content features. Unfortunately, no significant research so far exploits above stated features altogether to perform spam user profile detection. Additionally, majority of the existing approaches applies classical machine learning methods such as Naïve Bayes, K-Nearest Neighbor (KNN), Support Vector Machine, Artificial Neural Network (ANN), Extreme Learning Machine etc; whose performance often remains suspicious towards a generalizing outcome demands. Though, ensemble-based methods have performed relatively better than the classical machine learning based spam user profile classification systems. However, the ensemble methods applied so far applied ELM ensemble or deep ensemble concepts to perform spam user classification. The inclusion of more classifiers pertaining to the different methods, such as regression, pattern mining, neuro-computing, etc could have exhibited better performance. Considering it as motivation, in this proposed model a highly robust heterogeneous ensemble structure is proposed for spam profile detection and classification in online social media network. Unlike classical approaches, our proposed method exploits maximum possible common feature traits including user profile features, user's activity features, location features and content features to perform spam user identification in online social media networks. Though, as case study we considered the data of Facebook publically available features, we managed to retain maximum possible features to perform optimal classification for reliable spam user identification for further neutralization measures. Some of our key contributions are:

1. We designed a novel multiple constraint assisted (i.e., user's profile features, user's activity features, location features and content features) spam user identification system, which employs different but common public information from major online social media networks. It enables the suitability of our proposed method to be applied over each social media network.
2. Our proposed system applies advanced multi-phased feature selection approach using Wilcoxon Significant Rank Sum test, Pearson Correlation test and Principle Component Analysis (PCA) to enable optimal feature sets for further classification. It strengthens the proposed model to yield optimal malicious profile identification and classification. Additionally, it reduces unwanted redundant computation which makes overall process highly efficient.
3. As classifier we have applied a highly robust and first of its kind ensemble structure with 10s of machine learning methods including Logistic Regression, SVM algorithms with Linear, Polynomial, and Radial Basis Function (RBF) kernels, Least Square SVM with Linear, Polynomial, and RBF kernels, ELM with different kernels, ANN with Gradient Descent (GD), GDx (adaptive learning), ANN with Levenberg Marquardt (ANN-LM) algorithms etc. Such inclusion of different machine learning methods from different principle(s) give rise to a heterogeneous ensemble structure which performs optimal classification over a large set of input user profile features.
4. The performance assessment exhibits robustness of the proposed model over major at hand approaches which recommends our proposed approach for realistic application environment.

The remaining sections of the presented manuscript are given as follows. Section II discusses some of the key related work pertaining to social media spam user identification and classification, followed by research questionnaire in Section III. Section IV presents proposed

method and its implementation, which is followed by the simulated results and its inferences in Section V. Conclusion and future scopes are discussed in Section VI. The references used in this study are presented at the end of the paper.

II. Related Work

Ruan et al [10] exploited the social behaviors of Online Social Network users including their use-patterns, subscribed services and active application to perform malicious user identification. Authors amalgamated user's social behavioral patterns and corresponding behavioral feature metrics to perform malicious user identification and classification. To perform spam profile detection on online social media, Ahmad et al [11] developed Markov Clustering (MCL) concept. To achieve it, authors applied Facebook user's profiles including benign as well as spam profiles. Based on the features of each user, they derived a weighted graph in which profiles were represented as nodes and their interactions as edges. Authors estimated the weight of an edge, connecting a pair of user profiles as the function of their real social interactions. More precisely, authors applied user's active friends, page likes and shared URLs. Thus, obtaining aforesaid features, authors applied Markov clustering method with majority voting to perform two class classifications. A similar work was done by Setiawan et al [12] who applied Markov clustering concept for spam profile detection on Japanese social media platform. Gheewala et al [13] recommended using machine learning methods for spam profile detection on social media platforms. Soman et al [14] too recommended using machine learning methods, especially clustering and classification methods for online social media span profile detection. Considering significant feature selection, authors suggested using user activity features, location features and text and content features. As solution, authors applied Jenson-Shannon Divergence (JSD) measure as feature extraction tool for Twitter data, and applied Fuzzy K-means (FKM) algorithm to cluster similar user profiles, which were later processed for Extreme learning machine (ELM) based classification to detect malicious users. Authors exploited user profile features, user activity features, location-based features and text and content features to perform malicious tweet identification on online social media platform. To extract key features, authors applied Jenson-Shannon Divergence (JSD) measure to characterize each labeled tweet using natural language models. Obtaining the aforesaid features, authors at first perform Fuzzy K-means (FKM) based similar-user profile clustering. Subsequently, they applied extreme learning machine (ELM) algorithm to perform two-class classification for malicious profile detection. Meda et al [15] applied random forest and non-uniform feature sampling method to perform spam profile detection and classification. Authors applied Twitter datasets with user-related 54 features to perform classification. Shahabadkar et al [16] performed compromised profile identification on online social media, Twitter and Facebook. To achieve it, authors applied sudden change in the social behavioral patterns for two-class classification. Alghamdi et al [17] focused on detecting malicious URLs on social media platforms. Kantepe et al [18] performed social-bot detection on Twitter. As feature, authors applied posted tweets, profile information and temporal behavior information to perform bot-detection. Chen et al [19] performed stream spam profile detection on Tweeter. Authors extracted 12 lightweight features for tweet representation and made two-class classification using machine learning methods. Realizing the fact that the statistical properties of spam tweets vary over time, and thus, the performance of existing machine learning-based classifiers decreases, Chen et al [20] derived a concept called "Twitter Spam Drift". In address such problems authors learnt over the large statistical properties of spam tweets and performed two-class classification. Madisetty et al [21] applied neural network assisted ensemble model for spam detection in Twitter. Noticeably, authors performed spam detection on tweet level information. Authors applied Word2Vec embedding concept to obtain tweet level features, which was subsequently processed for classification using convolutional neural networks (CNNs) for spam profile classification. Zhang et al [22] performed spam post identification on Instagram, where supervised learning methods were applied to perform two-class classification using K-fold cross validation. Realizing the need of significant features for online social media malicious used identification, Rajamohana et al [23] applied heuristic algorithms like Cuckoo search with Harmony search, which retained suitable set of user's features to assist Naïve Bayes classification for spam user detection.

Considering the significance of Social Spam Profiles (SSPs) detection on social media, especially Twitter, Hua et al [24] developed a swift and scalable spam profile detection model using behavioral and graph-based information. Retrieving aforesaid features, authors designed a threshold and association-based classification model to classify each use as non-spam or spam. In comparison to the classical machine learning methods, such as support vector machine, their proposed model exhibited better accuracy. A similar effort was made by Vuong et al [25] who exploited user behavior and content-based profile classification on Facebook. To perform profile classification, authors applied the information pertaining to the user's comment (Vietnamese Facebook Pages) and social media behavior information. However, to perform classification authors applied maximum entropy information. A more robust solution was proposed by Al-Zoubi et al [26] who exploited public feature information to perform spam profile classification on Twitter Online Social Media. Authors extracted publically available features of the users which were processed for feature selection using Relief and Information Gain. Obtaining the suitable feature sets, authors applied machine leaning algorithms including Decision Trees, Multilayer Perceptron, k-Nearest neighbors and Naive Bayes to perform two-class

classification. Liu et al [27] applied ELM method to perform online spam profile detection and classification system. In their proposed model, authors at first retrieved messages crawling from Sina Weibo social media, which was performed for feature extractions (based on social interactions, and profile properties). Authors found that ELM based classification yields better accuracy than major classical machine learning methods. Savyan et al [28] too focused on anomaly detection on Online Social Network site, Facebook. Authors applied unsupervised clustering algorithm to explore and learn over the user's reaction as "Smileys" and performing similarity measures followed by clustering, authors performed user's maliciousness classification. Hudli et al [29] proposed a machine learning classification-based model for candidate vacant position. To achieve it, authors applied user's profile information on organizational online platform to enable candidate screening candidates for a vacant position in an organization. As classifier authors applied Naive Bayes and k-Nearest Neighbors (kNN) classification methods. Unlike above stated researches, an enhanced approach was proposed by Bhat et al [30] who designed an ensemble-based method to perform spammer classification. Authors applied community-based structural features to perform classification. Vishagini et al [31] focused on spam classification using weighted SVM and Fuzzy K-Means clustering. In their proposed model, authors applied email as feature, which was processed by weighted SVM for spam filtering using weight variables obtained by KFCM algorithm. The weight variables reflect the importance of different classes.

III. Research Questions

This study or research the key focus is made on achieving answers for the following:

RQ1: Can the use of multiple user traits including profile features, activity features, local features and content features be efficient to perform spam profile detection and classification?

RQ2: Can the use of multi-phased cascaded feature selection method applying Wilcoxon Rank Sum test, significant prediction test, Pearson Correlation test can yield optimal feature set for spam profile detection and classification on social media networks?

RQ3: Can the strategic implementation of different base classifiers including Logarithmic regression, decision tree, ANN-GD, ANN-LM, ANN-GDX (adaptive learning weight), SVM-Linear, SVM-Polynomial, SVM-RBF, LS-SVM Linear, LS-SVM-Polynomial, LS-SVM RBF be efficient to constitute ensemble structure for more efficient and reliable spam user profile detection and classification on social media network?

IV. System Model

This section primarily discusses the proposed spam-profile detection and classification system and encompassing algorithmic implementation.

A. Data Preparation

As already discussed, considering the goal to develop a novel and robust spam profile detection and classification system for online social media network, in this paper the prime focus is made on exploiting maximum possible user's behavioral, personal traits to characterize its proneness towards genuine (i.e., non-spam) and spam profile types. Though, there are a number of online social media platforms available globally, some of the key platforms are, Facebook, Twitter, Instagram, LinkedIn, etc.

In our considered dataset, a complete set of 28 different features pertaining to the total of 1337 users were taken into consideration. Noticeably, the considered dataset was having both spam-profiles as well as non-spam profile.

TABLE I. User's Online Features or traits for spam proneness characterization

Variable Type	Specific User features/traits
User's Profile Features	– *ProfileID* – *Name* – *ScreenName* – *Time Zone* – *Language* – *Date of Registration*
User's Activity Features	– *Status uploaded* – *Number of followers* – *Total number of friends* – *Default Profile Picture* – *Profile Protected Status* – *Profile Verification Status* – *Number of Favorites*
User's Location Features	– *Profile Description* – *URL sharing* – *Location Enable Status* – *Location at the date of Registration*
Content Features	– *Profile Sidebar color* – *Profile Background Title* – *Profile Sidebar fills color* – *Profile Background color* – *Profile link color* – *Number of page listed* – *Profile Image* – *Profile banner (Cover Image)* – *Background Image* – *Background Image URL* – *Profile Description Text Color* – *URL sharing information*

Now, observing the above stated features it can be found that the considered data is of heterogeneous types encompassing integer variables, character, string, which at first require to be converted into uniform numerical form. To achieve it, we converted input features into equivalent numerical values directly, which makes further computation more efficient. Thus, obtaining the equivalent numerical outputs of each input feature variable, we performed normalization over each data element. The detailed discussion of the data normalization

algorithm applied in this study is given in the subsequent section.

B. Data Normalization

This is the matter of fact that in major classification or prediction systems, especially in large features-based models data imbalance is the key problem, which hinders the overall performance of the system. Considering the fact that in the considered dataset, there can be the probability that the dataset can have very small features signifying spam-proneness probability which can cause bias in classification and hence can affect overall prediction accuracy.

In practical scenario, for example Facebook, Twitter, Instagram, LinkedIn etc where there can be gigantically huge number of users and their details, the corresponding data can be of different size and range and hence computing over such unstructured and broad-scaled data can force learning model to undergo pre-mature convergence. Consequently, it can affect overall accuracy of the proposed model. Considering such data imbalance problem, we have performed data normalization using Min-Max algorithm. Functionally, our proposed Min-Max normalization model that normalizes input data in the range of 0 to 1. Our proposed normalization method linearly transforms and maps the input data-elements in the range of [0, 1]. Functionally, each data element x_i of the user's feature X is mapped to the corresponding normalized value x_i' in the range of [0, 1]. Mathematically, we used (4) to estimate normalized value(s) of the input datax_i .

$$Norm(x_i) = x_i' = \frac{x_i - min\ (X)}{max(X) - \ min\ (X)} \tag{1}$$

In (1), the data elements (user's feature) $\min(X)$ and $\max(X)$ state the minimum and maximum values of X, respectively.

C. Multi-Phased Feature Selection

This is the matter of fact that spam profile detection and classification is a complex problem where categorizing a profile as malicious merely based on one parameter or profile value is not fair. On the other hand, it is not mandatory that all features can have decisive impact on labeling a user as malicious or normal. Considering this fact, identifying a suitable set of features for spam user identification is must. As stated, in this study we have considered multiple user-traits or features such as user's profile features, activity features, location features and content features. However, to ensure computationally efficient architecture retaining most suitable features would be of great significance. With this motive, in this research we applied multi-phased feature selection model using three varied types of feature selection methods. We have applied the following three feature selection methods, which are implemented in sequence to accomplish eventual goal.

1. *Wilcoxon Signed Rank Test (WRS),*
2. *Significant Test, and*
3. *Cross-Correlation Test.*

A brief of these algorithms is given as follows.

a). Wilcoxon Signed Rank Test

Wilcoxon Signed Rank Test (WRS), often known as rank test is a non-parametric test with independent samples assesses the correlation amongst the different variables and their distinct impact on classification accuracy. With this motive, we applied this method to perform estimate correlation amongst the different feature's values and its corresponding significance towards spam-profile proneness prediction. In other words, the input vectors are characterized whether they signify spam-profile proneness or probability of being spam of a user. It shows how each user feature is related to the spam-profile proneness or malefic nature. It employs two different kinds of variables; independent variable and dependent variable, amongst which it estimates the correlation to identify the most significant variable having strong relation to the classification output. We defined user details as the independent variable while its spam-proneness was hypothesized to be the dependent variable. Implementing this method, we retrieve p-value of each user profile with reference to the spam-proneness probability and shows how closely the spam-profile proneness probability are related to those traits or features. WRS helps handling the uncertainty amongst the all extracted features and identifies significant features by removing insignificant elements.

b Significant Predictor Test

Similar to the rank test, ULR typically estimates inter-relation between the independent and dependent variables. In this work, it assess whether (user's profile features, activity features, location features and content features) features of each users on social media network is significant predictor for its spam-proneness. We applied ULR on the selected features from previous selection phase (i.e., rank-sum selected features). ULR for the selected features assessed whether the selected metrics is significant for user-level spam-profile proneness identification or characterization. It estimated the extent of variance (change percentage) in the dependent variable (spam-profile proneness) as inferred by the independent variable (i.e., user's profile features, activity features, location features and content features). Mathematically (2)

$$logit[\pi(x)] = \alpha_0 + \alpha_1 X \tag{2}$$

In (2), $\text{logit}[\pi(x)]$ and X state the dependent (i.e., spam-proneness) and the independent (user's online social media features) variables, respectively. Here, π signifies probability factor of significance of each category. Mathematically,

$$\pi(x) = \frac{e^{\alpha_0 + \alpha_1 X}}{1 + e^{\alpha_0 + \alpha_1 X}} \tag{3}$$

In our proposed model, the significance-level of each features is obtained based on the value of regression coefficient, p-value. Any metrics having the p-value more than 0.05 has been considered as significant feature for

the spam proneness prediction. Metrics having p-value less than 0.05 have been removed from the final selected feature set.

c). Cross-Correlation Test

In this method, once retrieving ULR filtered feature-set characterizing user's profile features, activity features, location features, and content features, we performed cross-correlation test using Pearson correlation estimation algorithm. User's trait or feature with correlation coefficient higher than 0.5 ($p > 0.5$) were considered as the final feature vector for further spam-proneness classification. Once obtaining the eventual user's social media features, we executed data normalization and augmentation, where we focus on enhancing input data for better computation. The detail of the pre-processing methods applied is given in subsequent section.

D. Heterogeneous Ensemble Learning based User level Spam-prone profile Detection and Classification in Online Social Media Network

Considering the implementation of machine learning methods for spam-profile identification and classification, in majority of the existing works, authors have applied different machine learning methods, where those algorithms are applied as standalone classifier. However, for the same dataset different algorithms give different results, characterizing diversity in classification performance. Considering this fact, in this research a highly robust ensemble learning model is developed that strategically amalgamates classifiers from the different categories including pattern mining SVM, decision tree and neural network including extreme learning machine. Thus, the strategic amalgamation of the different machine learning algorithms constitutes a heterogeneous ensemble model to perform spam-proneness prediction of a user on online social media network. A snippet of the different machine learning methods or classifiers used is given in the subsequent sections. To be noted, as base classifiers we applied different machine learning algorithms. The detailed discussion of the aforesaid base classifiers and ensemble models is given in the following sections.

1. Logistic Regression

Logistic regression is one of the most used regression methods for data classification. It exploits regression concept to categories outputs of a dependent variable on the basis of multiple inputs (say, code metrics). In our proposed spam-user detection or classification problem, the dependent variable (i.e., user's spam proneness probability) can have the two categories or values; Spam user or non-Spam user (or, normal user). Considering this fact, in logistic regression as base classifier examines spam proneness of each user on the basis of the relation amongst the feature traits of the individual user. Mathematically, logistic regression is presented as (4).

$$\text{logit}[\pi(x)] = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots \ldots . + \beta_m X_m \quad (4)$$

In (4), the component at the left side $\text{logit}[\pi(x)]$ states the dependent variable while x_i presents the independent variable. Logistic regression method applies linear regression concept to convert the dichotomous outputs by logit function and therefore makes $\pi(x)$ varying in the range of 0 to 1 to $-\infty$ to$+\infty$. In (4) m represents the total count of the independent variables (here, the code metrics). The other variable π states the likelihood of the spam proneness of the user during validation. Thus, the dependent variable $\pi(x)$predicted by LR is given by (5).

$$\pi(x) = \frac{e^{\beta_0+\beta_1 X_1+\beta_2 X_2+\cdots\ldots+\beta_m X_m}}{1 + e^{\beta_0+\beta_1 X_1+\beta_2 X_2+\cdots\ldots+\beta_m X_m}} \quad (5)$$

2. Decision Tree (DT) Algorithm

Decision Tree algorithm has been used as a classical machine learning approach to perform pattern classification [49,50]; however, its efficiency has made numerous value addition over time to achieve better accuracy. For example, DT evolves in the different forms including IDE, CART, DT C4.5 and DT C5.0, which has been used mainly for data mining and classification purposes. We applied the most recent DT variant, C5.0 algorithm as a base classifier to perform user-level spam-proneness prediction over the input data. Noticeably, C5.0 DT algorithm classified or labeled each user as spam user or non-spam user. Originating at the root node, employing association rule in between the split criteria, the input data metrics are split into multiple branches at each node of the DT. The C5.0 algorithm applies in this research employed Information Gain Ratio (IGR) information to perform two-class classification, characterizing each use as Spam-user or Non-spam user.

3. Support Vector Machine (SVM)

SVM is one of the most used supervised machine learning algorithms used for pattern classification. It learns over the data patterns and acts as a non-probabilistic binary linear classifier. Functionally, SVM minimizes the generalization error on unobserved instances by means of structural risk reduction paradigm. In this method, the support vectors signifies the subset of the training set which obtains the value of the boundary also called hyper-plane in between the two classes. SVM based prediction applies the following function to perform pattern-based classification.

4. Least Squares Support Vector Machine (LSSVM)

LSSVM is a statistical learning theory that adopts a least squares linear system as a loss function. LSSVM is closely related to regularization networks. With the quadratic cost function, the optimization problem reduces to find the solution of a set of linear equations. We applied LSSVM algorithm as per the reference [39], where three variants LSSVM with linear, polynomial and RBF kernel functions were applied. Due to space constraints, the detailed discussion of LSSVM is not given in this manuscript, though the detail of implemented model can be found in [39].

5. Artificial Neural Network

Being a heterogeneous ensemble structure, we applied ANN as neuro-computing (base-classifier) model for classification. In this research different ANN variants with different learning or weight estimation methods are applied as base leaner. The detailed discussion of the ANN models applied in this research is given in the subsequent sections.

Amongst the major neuro-computing concepts, ANN has emerged as one of the most applied and sought algorithms. Functionally it mimics the human-brain characteristics to learn over the different data or patterns so as to make future classification in unknown dataset(s). The learning efficiency, dept information learning and allied classification capacity makes ANN a potential solution for major Artificial Intelligence (AI) purposes or decision-making purposes. Structurally, it encompasses multiple neurons possessing input data to be processed at the different layers such as input layer, hidden layer and eventually outputs classification output at the output layer. Functionally, it employs error reduction concept to learn over the input data, where it calculates the disparity in between the expected value and the observed value called error. It intends to reduce the error iteratively to attain least or the zero error signifying convergence to result final output at the output layer. Noticeably, at the output layer, ANN classifies input data into expected categories, for example in this paper classifies each user as Spam-user and Non-spam user.

Practically, achieving optimal classification result requires suitable weight selection, swift computing etc; else it undergoes local minima and pre-mature convergence, which affects overall computational efficiency. To alleviate such problems and to gain zero-error condition ANN requires optimal weight estimation and respective learning efficiency. To achieve it, ANN has undergone numerous phased evolutions with enhancement in weight estimation and learning efficiency. In this paper, we applied ANN with different variants to perform spam-proneness prediction or classification, where these algorithms (ANN-GD, ANN-GDX, ANN-LM and ANN-RBF) have been applied as base learners to constitute ensemble.

A. ANN-GD

Gradient descent is one of the optimization technique used to minimize error function iteratively for all training sets. Performing GD based weight estimation and corresponding learning ANN-GD classifies each user on the social media platforms or online social media platform as spam-prone or non-spam prone and labels it as "1" for spam-prone profile and "0" for non-spam prone profile.

B. ANN-RBF

ANN-RBF(Radial Basis Function neural network has an input layer, hidden layer and an output layer. The neurons in hidden layer consist of Gaussian Transform Function whose output will be inversely proportanl to the distance from the center of the neuron.

C. ANN-GDX/LM

Though, ANN-GD and ANN-RBF algorithms have been applied for many classification problems; however adaptive weight assignment and learning remained challenge. Unlike classical neuro-computing models, ANN-LM and ANN-GDX iteratively performs localization of the minimum value of the multivariate function, which is often called as the Sum of Squares (SoS) of the non-linear real-valued functions. This ability strengthens ANN-GDX to perform swift weight update which not only makes learning faster but also avoids the problem of local minima and convergence. Additionally, ANN-LM model amalgamates efficacy of both SD-ANN as well as ANN-GD models, where selecting the learning rate, it achieves error minimization swiftly.

6. Heterogeneous Ensemble Learning Model

A snippet of the ensemble model applied in this research is given as follows:

A. Maximum Voting Ensemble (MVE) Model

In this paper, we have applied above stated base classifiers as the base classifier to constitute a novel heterogeneous ensemble learning model. To form ensemble structure all classifiers are run over the same dataset and predict each class as spam prone profile (label as "1") or non-spam profile (label as "0"). Thus, obtaining class outputs (i.e., label) of each software code/class "Maximum Voting is obtained for each class and a class with maximum votes (i.e., either 1 or 0), is categorized as final category (spam prone profile or non- spam prone profile).

B. Best Trained Ensemble Model

Unlike MVE ensemble, the Best Trained Ensemble (BTE) model at first identifies the best performing base classifier. In our proposed model, BTE was designed in such manner that for each feature set (rank sum test, significant predictor test, and correlation test), it identifies the best performing base classifier, and again trains the input data by means of the identified classifier overall all feature sets. Thus, the maximum accuracy or performance obtained by that classifier throughout the three different feature sets is predicted as the final output. Thus, applying BTE ensemble over the extracted user's features, user-level spam proneness classification has been performed.
The detailed discussion of the simulation results and their corresponding inferences are given in the subsequent section.

V. RESULTS AND DISCUSSION

Considering significance of a robust spam or malicious profile identification and classification system, in this research the predominant emphasis was made on exploiting maximum possible user's features and map their corresponding associations to signify a user as spam prone user or non-spam (prone) user. Though, a few researches have made efforts to use classical machine learning methods for spam-profile classification, in this

paper at first, we emphasized on enhancing multiple level processes, including large scale user's public profile variables, activity features, location and content features. Subsequently, to achieve a better trade-off between computational overhead and efficiency we applied multi-phased feature selection method applying Wilcoxon Rank Sum test, significant predictor test and Pearson correlation test in sequence. Noticeably, in significant prediction and correlation test we assigned significant coefficient or the level of significance as 0.5 (i.e., 50% significance level). This mechanism ensured that only those features or the user's trait(s) having certain significant towards its spam-proneness prediction or relation would be retained and remaining or relatively insignificant feature element(s) would be dropped before classification. As already stated, to design a universally applicable spam profile detection and classification system, we retained those all publically accessible or API assisted retrieved user's information which is common in major online social media networks. For example, date of registration, location, language, profile photo, URL, comments or posts and their corresponding frequency, followers, following user's etc are the common public access-level information of a user. With this motive, in this motive, in this research we considered a total of 28 features pertaining to a user on online social media network [40]. The detail of the features considered in this work is given in Table I. Additionally, unlike major classical efforts where authors merely applied 100 or lower number of users to train their model, we applied a total of 1337 user details, which comprised both spam profile(s) as well as non-spam profile(s). Considering data heterogeneity, we at first converted all data types in equivalent numerical form, which were subsequently processed for Min-Max normalization followed by cross validation-based training. To be noted, as a pre-processing step, realizing characters diversity such as lower case, upper case, special characters, numerical or floating values, we transferred upper case into lower case using MATLAB functions, which was subsequently converted into readable or executable numerical form. Performing Min-Max normalization in the range of [0-1], we performed multi-phased feature selection which enabled it to retain optimal feature sets for further classification.

Though, in numerous existing researches authors have applied classical machine learning methods such as Naïve Bayes, K-NN, SVM, ANN etc individually in this paper we focussed on achieving higher accuracy by designing a robust ensemble learning model. More specifically, we developed a heterogeneous ensemble learning model comprising pattern learning algorithms, decision tree, neuro-computing models etc. As name indicates, the inclusion of machine learning algorithms from the different pattern mining or learning ability enables proposed ensemble model to be called as heterogeneous. In our proposed ensemble model, we applied a total of 11 base classifiers, which constituted 2 different ensemble paradigms, MVE and BTE. As base classifier we applied the key machine learning methods like, including Logistic Regression, SVM algorithms with Linear, Polynomial, and Radial Basis Function (RBF) kernels, Least Square SVM with Linear, Polynomial, and RBF kernels, ELM with different kernels, ANN with Gradient Descent (GD), GDx (adaptive learning) algorithms. In order to enhance the reliability of classification to train the model, we applied 10-fold cross validation method. To be noted, being a two-class classification problem, it performed each user labeling as Spam-user profile (or spam-prone profile) or non-spam profile. Thus, our proposed model performed user level classification (as spam profile or non-spam profile) for all 1337 users by learning over their corresponding features. In addition to the cross-validation based performance assessment, to manually test the classification reliability and allied performance of the proposed model, we provide the facility to feed user's manual information to the system, which on the basis of pre-trained model classifies that user as spam-prone or non-spam prone profile.To perform statistical performance analysis, we obtained confusion metrics in terms of True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN). The definitions of these statistical variables and their combination to yield accuracy, precision, recall and F-Measure are given in Table II.

TABLE II. PERFORMANCE VARIABLES

Parameter	Mathematical Expression	Definition
Accuracy	$\frac{(TN+TP)}{(TN+FN+FP+TP)}$	Signifies the proportion of predicted spam-prone user(s) that are inspected out of all modules.
Precision	$\frac{TP}{(TP+FP)}$	States the degree to which the repeated measurements under unchanged conditions show the same results.
F-measure	$2.\frac{Recall.Precision}{Recall+Precision}$	It combines the precision and recall numeric value to give a single score, which is defined as the harmonic mean of the recall and precision.
Recall	$\frac{TP}{(TP+FN)}$	It indicates how many of the relevant items are to be identified.

The statistical performance has been obtained for the different base classifiers and ensemble models (Logistic regression, Decision tree, SVM-Linear, SVM-Poly, SVM-RBF, LSSVM, LSSVM-Lin, LSSVM-Poly, LSSVM-RBF, ANN-GD, ANN-GDX, ANN-RBF, MVE and BTE).

TABLE III. PERFORMANCE VALUES

Techniques	Accuracy (%)	Precision (%)	Recall (%)	F-Measure (%)
Logistic regression	52.70	99.00	51.10	67.40
Decision Tree	52.98	98.50	52.00	68.06
ANN-GD	97.01	98.01	99.40	98.70
ANN-GDX	97.52	97.60	99.40	98.49
ANN-RBF	59.61	98.92	56.20	71.67

SVM-Lin	52.00	98.70	56.20	71.61
SVM-Poly	52.71	98.72	56.20	71.62
SVM-RBF	89.94	97.20	88.41	92.59
LSSVM-Lin	59.20	98.60	56.30	71.67
LSSVM-Poly	58.99	98.60	56.30	71.67
LSSVM-RBF	61.42	98.60	56.30	71.67
BTE	98.80	98.20	99.67	98.91
MVE	62.80	99.01	57.00	72.34

The recall performance for BTE ensemble has been obtained as 99.67, which is significantly higher than MVE ensemble (57%) and other machine learning methods. It shows robustness of the BTE ensemble to learn and classify data under varied diversity and non-linear structure. F-Measure parameter, which employs both Recall and Precision too affirmed superiority of the proposed BTE model (98.91%) than the MVE (72.31%) and other machine learning models. Thus, observing overall statistical performance parameters and their corresponding significance it can be inferred that the proposed BTE ensemble model can be most suitable and effective towards spam profile detection and classification on online social media network. The box-plot presentation of the performance by different machine learning algorithms (as base classifiers) and ensemble learning methods is given in Fig. 3 to Fig. 6.

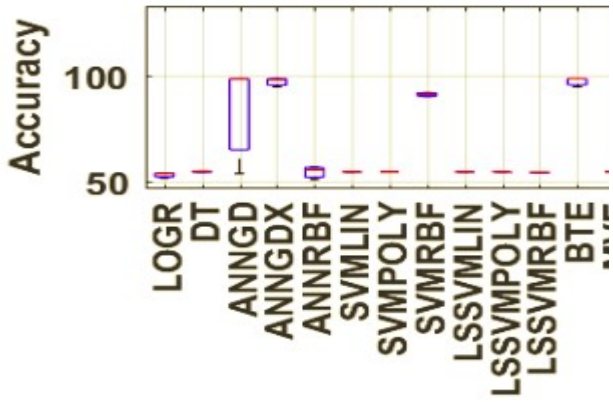

Fig. 3 Box-plot for accuracy by different machine learning and ensemble methods for spam profile detection and classification

Observing above stated results (Table II), it can be found that BTE ensemble model outperforms other base classifiers and even MVE ensemble model. The maximum accuracy has been achieved by BTE, enables followed by ANN-GDX (with adaptive learning ability). Interestingly, SVM and its advanced version named LS-SVM which is expected to perform better has perform very lower in comparison to the neuro-computing models and eventual our proposed BTE ensemble structure. Similar to accuracy, the proposed ensemble model has exhibited maximum precision of (BTE) 98.20%, though it is lower than the MVE, which exhibited 99.01%. However, in reference to the other parameters, accuracy, recall, F-measure, BTE can be more reliable and precise.

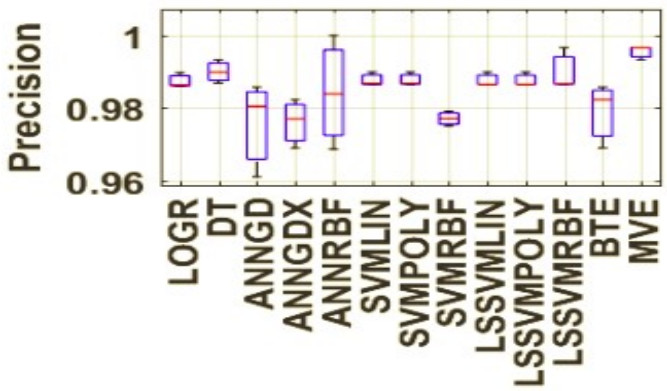

Fig. 4 Box-plot for precision by different machine learning and ensemble methods for spam profile detection and classification

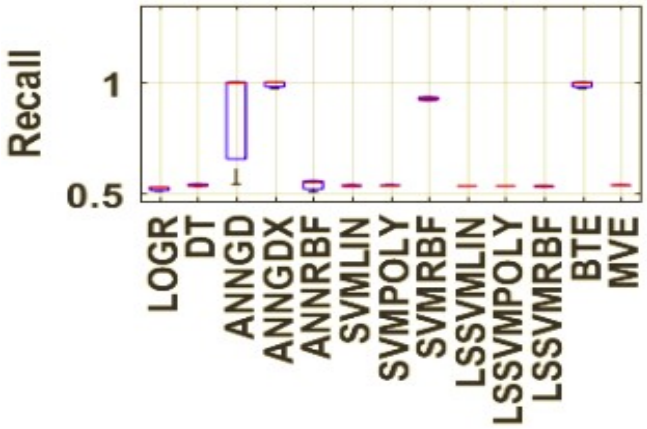

Fig. 5 Box-plot for recall by different machine learning and ensemble methods for spam profile detection and classification

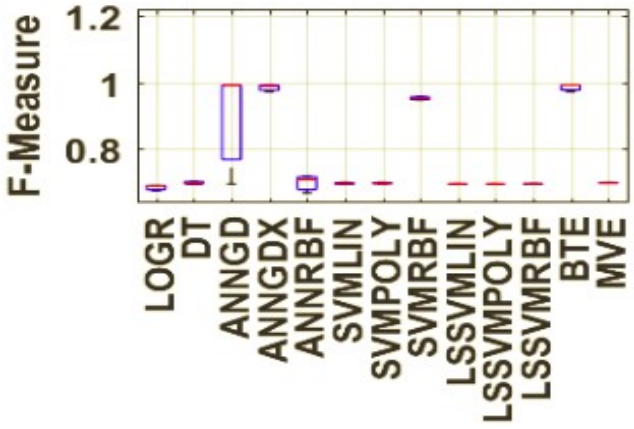

Fig. 6 Box-plot for F-Measure by different machine learning and ensemble methods for spam profile detection and classification

Though, the above discussed simulation results reveal and affirm the robustness of the proposed BTE heterogeneous ensemble model for spam profile detection and classification on social media platform, we have performed a qualitative assessment of the proposed model with reference to the other works done so far.

TABLE IV. PERFORMANCE COMPARISON

Methods	Accuracy (%)	Precision (%)	Recall (%)	F-Measure (%)
Hua et al [31]	79.26	66.18	68.32	-
Vuong et al [25]	92.00	92.32	92.12	91.24
Al-Zoubi et al [26]	95.70	94.00	96.00	-
Bhat et al [30]	96.90	96.90	96.90	96.90
*Proposed BTE Ensemble	98.80	98.20	99.67	98.91

The performance comparison of the proposed spam profile detection and classification system is given in Table III. Observing the comparative results it can be found that the proposed BTE ensemble model with Logistic regression, Decision tree, SVM-Linear, SVM-Poly, SVM-RBF, LSSVM, LSSVM-Lin, LSSVM-Poly, LSSVM-RBF, ANN-GD, ANN-GDX, ANN-RBF as base classifier with BTE ensemble concept outperforms all existing approaches. Though, authors in [31] applied SVM as classifier with different feature selection and threshold adaptive classification, its performance is far below our proposed model. Authors in [37] applied j48 and k-NN algorithms to constitute ensemble learning for spam classification however, our proposed BTE ensemble outperforms it. On the other hand, in [33] who applied j48 and Naïve Bayes classifier too performs inferior in comparison to our proposed heterogeneous ensemble model BTE. Observing overall performance, it can be confirmed that the proposed heterogeneous ensemble model with aforesaid base classifier combination with BTE structure or ensemble paradigm can achieve optimal performance to perform spam profile detection and classification over online social media networks. The overall results affirm acceptance of all research questions *(RQ1, RQ2 and RQ3)* defined in Section III. The overall research conclusion is given in the subsequent section.

VI. CONCLUSION

Considering the significance of multi-parametric learnt model for spam user profile detection in online social media network, in this paper a highly robust heterogeneous ensemble learning structure-based approach is developed. Unlike classical efforts the key contributions of this paper can be visualized in terms of the multiple features learning or training, multi-phased feature selection and finally heterogeneous ensemble learning based classification. The proposed method exploited diverse user's profile variables including profile features, activity features, location features and content features. Such multi-constructs amalgamation strengthens learning to make accurate spam profile identification or classification. Once obtaining the aforesaid feature-set the proposed method applied sequential implementation of multiple feature selection methods, including Wilcoxon Rank Sum test, significant predictor test, and Pearson Correlation test, which helped in retaining optimal features for further computation. This approach not only reduced bulkiness of large feature data, but also enhanced computational efficacy over large input space. Obtaining the optimal feature set, the proposed method applied a robust heterogeneous ensemble structure encompassing pattern analysis method, regression approaches, neuro-computing, decision tree methods etc, whose strategic amalgamation as ensemble learning can help achieving an optimal classification solution. More specifically, in this paper we applied Logarithmic regression, decision tree, ANN-GD, ANN-GDX (adaptive learning weight), ANN-RBF, SVM-Linear, SVM-Polynomial, SVM-RBF, LS-SVM Linear, LS-SVM-Polynomial, LS-SVM RBF and two key ensemble models Base Trained Ensemble (BTE) and Maximum Voting Ensemble (MVE). Performing extensive performance assessment over realistic Facebook user profile data with 1334 users containing both genuine or non-spam as well as spam profile revealed that amongst the all classifiers the BTE ensemble exhibits classification accuracy of 98.8%, precision of 98.2%, F-Measure and Recall of approximate 99.67% and 98.91%, respectively. On contrary, MVE ensemble exhibits accuracy of merely 62.8%, while precision of 97%, F-Measure 74.14% and recall of 60%. Observing overall performance, it can be found that the proposed heterogeneous ensemble with BTE structure outperforms major base classifiers and existing methods. Thus, the amalgamation of multiple features or user-specific traits, suitable feature selection and heterogeneous BTE ensemble model can be the best suited solution for spam or malicious user account identification and classification over social networking sites. In future, the efficacy of deep-ensemble mechanism can also be examined to perform spam user profile identification over large scale dataset.

VII. REFERENCE

[1] Statistic, Most famous social network sites — Global social media ranking 2016, 2016 (accessed November 15, 2016).

[2] J. Alqatawna, "An adaptive multimodal biometric framework for intrusion detection in online social networks," IJCSNS International Journal of Computer Science and Network Security, vol. 15, no. 4, pp. 19–25, 2015.

[3] D. Wang, D. Irani, and C. Pu, "A Social-Spam Detection Framework," in Proceedings of the 8th Annual Collaboration, Electronic messaging, Anti-Abuse and Spam Conference, New York, NY, pp. 46-54, 2011.

[4] Harris Interactive Public Relations Research. A study of social networks scams. 2008.

[5] K. Beck, "Analyzing Tweets to Identify Malicious Messages," in Proceedings of the 2011 IEEE International Conference on Electro/Information Technology, Mankato, MN, pp. 1-5, 2011.

[6] A. Hai Wang, "Detecting Spam Bots in Online Social Networking Sites: a Machine Learning Approach," in Proceedings of the 24th annual IFIP WG 11.3 working conference on Data and applications security and privacy, Springer-Verlag Berlin, Heidelberg, pp. 335-342, 2010.

[7] G. Stringhini, C. Kruegel, and G. Vigna, "Detecting Spammers on Social Networks," in Proceedings of the 26th Annual Computer Security Applications Conference, New York, NY, pp. 1-9, 2010.

[8] K. Varnsen, "Types of Twitter Spam," (Ranker), [online] 2013, http://www.ranker.com/list/types-of-twitter-spam/kel-varnsen. [Accessed on 22 Feb 2020]

[9] A Fast and Minimal JSON Parser for Java, Eclipse Source Developer, [online] 2013, http://eclipsesource.com/blogs/2013/04/18/minimal-jsonparser-for-java/.[Accessed on 22 Feb 2020]

[10] X. Ruan, Z. Wu, H. Wang and S. Jajodia, "Profiling Online Social Behaviors for Compromised Account Detection," in IEEE Transactions on Information Forensics and Security, vol. 11, no. 1, pp. 176-187, Jan. 2016.

[11] F. Ahmed and M. Abulaish, "An MCL-Based Approach for Spam Profile Detection in Online Social Networks," 2012 IEEE 11th International Conference on Trust, Security and Privacy in Computing and Communications, Liverpool, 2012, pp. 602-608.

[12] E. I. Setiawan, C. P. Susanto, J. Santoso, S. Sumpeno and M. H. Purnomo, "Preliminary study of spam profile detection for social media using Markov Clustering: Case study on Javanese people," 2016 International Computer Science and Engineering Conference (ICSEC), Chiang Mai, 2016, pp. 1-4.

[13] S. Gheewala and R. Patel, "Machine Learning Based Twitter Spam Account Detection: A Review," 2018 Second International Conference on Computing Methodologies and Communication (ICCMC), Erode, 2018, pp. 79-84.

[14] S. J. Soman and S. Murugappan, "Detecting malicious tweets in trending topics using clustering and classification," 2014 International Conference on Recent Trends in Information Technology, Chennai, 2014.

[15] C. Meda et al., "Spam detection of Twitter traffic: A framework based on random forests and non-uniform feature sampling," 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, San Francisco, CA, 2016.

[16] R. Shahabadkar, Mukesh Kamath B and K. R. Shahabadkar, "Diagnosis of compromised accounts for online social performance profile network," 2016 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET), Chennai, 2016.

[17] B. Alghamdi, J. Watson and Y. Xu, "Toward Detecting Malicious Links in Online Social Networks through User Behavior," 2016 IEEE/WIC/ACM International Conf. on Web Intelligence Workshops, Omaha, NE, 2016.

[18] M. Kantepe and M. C. Ganiz, "Preprocessing framework for Twitter bot detection," 2017 International Conference on Computer Science and Engineering (UBMK), Antalya, 2017

[19] C. Chen et al., "A Performance Evaluation of Machine Learning-Based Streaming Spam Tweets Detection," in IEEE Transactions on Computational Social Sys. vol. 2, no. 3, pp. 65-76, Sept. 2015.

[20] C. Chen, Y. Wang, J. Zhang, Y. Xiang, W. Zhou and G. Min, "Statistical Features-Based Real-Time Detection of Drifted Twitter Spam," in IEEE Trans. on Infor. Forensics and Security, vol. 12, no. 4, pp. 914-925, 2017.

[21] S. Madisetty and M. S. Desarkar, "A Neural Network-Based Ensemble Approach for Spam Detection in Twitter," in IEEE Transactions on Computational Social Systems, vol. 5, no. 4, pp. 973-984, Dec. 2018.

[22] W. Zhang and H. Sun, "Instagram Spam Detection," 2017 IEEE 22nd Pacific Rim International Symposium on Dependable Computing (PRDC), Christchurch, 2017

[23] S. P. Rajamohana, K. Umamaheswari and S. V. Keerthana, "An effective hybrid Cuckoo Search with Harmony search for review spam detection," 2017 Third International Conference on Advances in Electrical, Electronics, Information, Comm. and Bio-Informatics, Chennai, 2017.

[24] W. Hua and Y. Zhang, "Threshold and Associative Based Classification for Social Spam Profile Detection on Twitter," 2013 Ninth International Conference on Semantics, Knowledge and Grids, Beijing, 2013.

[25] T. Vuong, V. Tran, M. Nguyen, C. Thi Nguyen, T. Pham and M. Tran, "Social-spam profile detection based on content classification and user behavior," 2016 Eighth International Conference on Knowledge and Systems Engineering, Hanoi, 2016, pp. 264-267.

[26] A. M. Al-Zoubi, J. Alqatawna and H. Faris, "Spam profile detection in social networks based on public features," 2017 8th International Conference on Information and Communication Systems (ICICS), Irbid, 2017, pp. 130-135.

[27] Chen Liu and Genying Wang, "Analysis and detection of spam accounts in social networks," 2016 2nd IEEE International Conference on Computer and Communications, Chengdu, 2016.

[28] P. V. Savyan and S. M. S. Bhanu, "Behaviour Profiling of Reactions in Facebook Posts for Anomaly Detection," 2017 Ninth International Conference on Advanced Computing (ICoAC), Chennai, 2017.

[29] S. A. Hudli, A. V. Hudli and A. A. Hudli, "Application of data mining to candidate screening," 2012 IEEE International Conference on Advanced Communication Control and Computing Technologies (ICACCCT), Ramanathapuram, 2012.

[30] S. Y. Bhat, M. Abulaish and A. A. Mirza, "Spammer Classification Using Ensemble Methods over Structural Social Network Features," 2014 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT), Warsaw, 2014

[31] V. Vishagini and A. K. Rajan, "An Improved Spam Detection Method with Weighted Support Vector Machine," 2018 International Conference on Data Science and Engineering (ICDSE), Kochi, 2018.