


RESEARCH

Open Access

An efficient hybrid system for anomaly detection in social networks



Md. Shafiur Rahman^{1,2*} , Sajal Halder^{2,3}, Md. Ashraf Uddin^{2,4} and Uzzal Kumar Acharjee²

Abstract

Anomaly detection has been an essential and dynamic research area in the data mining. A wide range of applications including different social medias have adopted different state-of-the-art methods to identify anomaly for ensuring user's security and privacy. The social network refers to a forum used by different groups of people to express their thoughts, communicate with each other, and share the content needed. This social networks also facilitate abnormal activities, spread fake news, rumours, misinformation, unsolicited messages, and propaganda post malicious links. Therefore, detection of abnormalities is one of the important data analysis activities for the identification of normal or abnormal users on the social networks. In this paper, we have developed a hybrid anomaly detection method named DT-SVMNB that cascades several machine learning algorithms including decision tree (C5.0), Support Vector Machine (SVM) and Naïve Bayesian classifier (NBC) for classifying normal and abnormal users in social networks. We have extracted a list of unique features derived from users' profile and contents. Using two kinds of dataset with the selected features, the proposed machine learning model called DT-SVMNB is trained. Our model classifies users as depressed one or suicidal one in the social network. We have conducted an experiment of our model using synthetic and real datasets from social network. The performance analysis demonstrates around 98% accuracy which proves the effectiveness and efficiency of our proposed system.

Keywords: Anomaly Detection, Machine Learning, Hybrid Anomaly Detection, Social Networks

Introduction

Nowadays, social networks have become part and parcel of human lives in which people with similar interests, values and views communicate and interact on a large scale. Individuals visit social diverse kinds of network platforms including Facebook, MySpace and Twitter to create social and professional networks, collect information relevant to them, and share a significant amount of their sensitive data with others.

Although the popularity of online social networks (OSNs) is growing day by day, anomalies/spammers uploading messages with malicious content are increasingly targeting different social medias (Caruana and Li

2012). Online social network anomalies refer to unusual, and often illegal, user activity. Anomalies can have three different forms: a) point anomalies, b) contextual anomalies and c) collective anomalies. To detect malicious individuals including spammers, sexual predators, and online fraudsters, various machine learning techniques are used in the state-of-the-art research. Social network users are more likely to trust spam messages posted by their friends on online social networks in comparison to the number of spam emails (Zheng et al. 2015). Kayode et al. (2017) presented several existing works that detected the spammer on social networking sites. They mainly focused on the detection of phishing, spam or fake accounts and compromised accounts. Owing to the significantly growing popularity of social media, social networks have made it possible to gather a vast amount of information about user profile and content. This however, also increases the distribution of spam content and produces a number

*Correspondence: shafiurcse@gmail.com

¹Department of Computer Science and Engineering, Dhaka International University, Dhaka, Bangladesh

²Department of Computer Science and Engineering, Jagannath University, Dhaka, Bangladesh

Full list of author information is available at the end of the article

of new spammers who conduct and propagate different unusual and irregular activities (Wang et al. 2014). Spammers apply diverse techniques to spread spam messages on the social media. The anomalous messages on the social network has a great negative consequences in the society such as marketing advertisements of illegal drugs (Xu et al. 2016). In addition, spammers can follow and discover a large number of anonymous users particularly female and send them unwanted messages containing malicious URLs. Zephoria Digital Marketing's report (Zephoria Digital Marketing 2018) stated that currently there are approximately 2.20 billion monthly active Facebook users and the number of active users in Facebook is increasing by 13% in every year.

In modern times, social media is a significant cause of cyberbullying, which can influence many teenagers to commit suicide. Adolescent social media users are psychologically abused by anomalous individuals' messages, emails, tweets, and reactions. Therefore, misbehaving users such as Sybil accounts can be detected by analyzing data related to user's behaviour in the social networks (Yang et al. 2014). In order to differentiate between malicious and legitimate actors in the social network, many researchers have adopted the most widely used machine learning classifiers.

Belavagi et al. (2016) proposed a predictive model to detect intrusion using machine learning classifications that include SVM, Gaussian Naive Bayes and Random Forest Logistic Regression. The hybrid detection technique was found more efficient than a single classifier. Detection of intrusions is a potential field of security study, with the rapid growth of the Internet in daily life. Different classifier algorithms for classifying network traffic as regular or abnormal is used by many intrusion detection systems (IDS). Thaseen et al. (2017) adopted a multi class support vector machine with feature selection using chi-square to detect intrusion. NSL-KDD dataset was used to train the model and tuned the parameter of SVM. An ensemble of classifiers including SVM, MNB (modified Naive Bayes) and LPBoost were also developed by Thaseen et al. (2019) to detect network intrusion. To find out more important characteristics relevant to the class label, they graded the features using the Chi-square process.

However, legitimate users in the social networks are distinguished from the anomalous users of IDS (Intrusion Detection System) in the conventional network with respect to the dynamic characteristics of features. So, the classification algorithms for anomaly detection in the social network are associated with different significant challenges including the high dimensionality of features, high false-positive rate, biased and training sets, and high computational complexities (Rathore et al. 2018).

The downside of using a single classifier in the classification led to the concept of building a hybrid method

designed using Bagging and Boosting techniques. Islam et al. (2018) identified depressed users on the social networks using different machine learning algorithms including KNN, SVM, Decision Tree and ensemble method. They used different various psycho linguistic features to train the machine learning algorithms and found that decision tree produced the highest accuracy. Aljawarneh et al. (2018) proposed a hybrid anomaly detection system by cascading multiple classifiers such as J48, Naive Bayes, Meta Paggging, DecisionStump, RandomTree, REP-Tree and AdaBoostM1. We also focus on detection of the anomaly for one of the most popular OSN platforms, Facebook. Recent advances in machine learning show that cascading (Belavagi and Muniyal 2016) multiple machine learning methods yield a better performance than a single or hybrid use of classifiers. Our approach differs from above mentioned literature in the way we have used different machine learning algorithms and data features. We design a hybrid anomaly detection system in the social network. The model is developed by cascading three machine learning algorithms: 1) the Decision tree (C5.0) 2) Support Vector Machine (SVM) 3) Naive Bayesian Classifier (NBC). In the first phase, decision tree C5.0 is employed to classify social network's users into two classes: anomalous user and normal users. The decision tree C5.0 was chosen for the following reasons: 1) C5.0 algorithm requires comparatively less memory than other traditional decision trees and 2) C5.0 reduces error pruning and removes irrelevant or non associated attributes to the class labels on large datasets. In the second phase, SVM is applied to classify anomalous users to categorize individuals as happiest users or disappointed users. In the final phase, NBC is used to get the two groups of disappointed users into suicidal and not suicidal users. The result analysis shows that the proposed system provides a good performance to detect the anomaly and suicidal users in the social network.

Combing the k-mean clustering approach with C5.0 decision tree and SVM can overcome two existing problems in k-mean clustering: 1) the problem of class dominance and 2) the problem forced assignment. The forced assignment problem arises when the k parameter in k-mean clustering is set to a value that is considerably less than the inherent number of natural groupings within the training data. The main contributions of this dissertation are described as follows:

- The first contribution is to detect anomalous users using Decision Tree (C5.0) classifier for mitigating the forced assignment and class dominance problem raised while classifying data as normal and anomalous behaviors in a social network.
- The second contribution includes a vulnerable users detection model called DT-SVMNB on the social

networks. The model discovers a cascade of machine learning algorithms that demonstrates a higher level of accuracy in detecting vulnerable social media users. From an anomaly detection perspective, the paper presents a high performance spammer with the suicidal detection system.

- The performance of the proposed DT-SVMNB classifier has been analyzed in this article. The model applies traditional decision tree (C4.5), Support Vector Machine (SVM) and Naive Bayesian classifier (NBC) methods in three different levels to evaluate its performances with respect to six metrics including accuracy. The result section demonstrated that our proposed hybrid approach outperforms over other existing methods. We conduct an extensive experiments on the proposed model using both synthetic and real datasets.

The remaining part of this paper is organized as follows. In “[Related work](#)” section, the related works are described while some preliminary concepts are discussed in “[Preliminaries about basic classifiers](#)” section. The proposed methodology has been described in “[The proposed vulnerable social media users prediction model](#)” section. The effectiveness, efficiency and scalability of the proposed method are shown in “[Experimental analysis and evaluation](#)” section. In “[Conclusions](#)” section, we concluded the paper with a direction of future work.

Related work

In this section, we discuss different existing machine learning models for detecting spammers in social networks. A social network is a platform for people sharing their activities, interests, background, and real-life connections via specific visual computer techniques. Online social networks (OSNs) have become more and more popular in nowadays society, and it would be hard to get rid of them from normal daily life.

The very first online social network is the email where people shared and transferred information via different email addresses. Benefiting from the flourish of smartphones, people have multiples choices of various social network applications or Apps including Facebook, Twitter, Snapchat, Tumblr, and Instagram. The social networks also create a lot of spammers every day. Various machine-learning techniques are used to detect spammers in social network. In such a technique, first a set of appropriate feature is selected. Next, the machine learning algorithm is trained using the dataset with the selected features. The training dataset contains behaviours of spam in the social network to classify the users into two categories: abnormal user and normal users. A number of machine learning techniques have been proposed to detect anomaly in data mining in literature. In the past

eight years, anomaly detection and filtering mechanisms have been widely implemented in a variety of social network-related applications such as email spam (Zhou et al. 2014), web (Erdélyi et al. 2011), and social networks (Yu et al. 2017) etc.

Chu et al. (2012) proposed an unsupervised technique to distinguish spam from the legitimate campaign using Random Forest algorithm based on content and behavioral features. Martinez et al. (2013) trained SVM classifier based on combined features of language and content. Benevenuto et al. (2012) used SVM classifier approach to classify spammers on Twitter. The authors identified the characteristics of spammers from tweet contents and user-behaviour to detect spammers and legitimate users. A large number of existing work focuses on the content-based model of machine learning approaches. Such content-based features are used to learn classification models to classify message and profiles as anomalous (Abulaish and Bhat 2015).

Gupta et al. (2015) proposed a hybrid spam detection mechanism for spammer detection in the social network taking three machine learning algorithms: Naive Bayes, Clustering and decision tree. Another spam detection (Sohrabi and Karimi 2018) technique is proposed to detect spam in Facebook cascading unsupervised machine algorithm clustering and supervised machine learning algorithm decision tree, SVM.

Manjunatha et al. (2018) suggested an anomaly detection model in the social network using Apache Kafka, Hbase database and Political Independence Index metrics. The risk of cyber attacks on social media can be reduced if anomalous activity is detected during data streaming. Mahmodi et al. (2020) recommended Online Fusion of Experts to forecast the drift in a data stream from a social media. Their proposed model used liner-order and Gaussian-order algorithm to identify anomalous changes in data stream. The design of anomaly detection in social media is challenging with the growth of data. Yasami et al. (2017) proposed a statistical infinite feature cascade approach to detect anomaly in social network where the model comprises two components 1) normal modeling 2) anomaly detection component. The first step is the birth, death and length of features, which are supposed to be distributed in this article for the first time in realistic statistics. The second approach is the development of the characteristics of nodes modelled by an Infinite Factorial Hidden Markov Model (IFHMM), considering the cascade of features.

Bindu et al. (2017) designed Anomaly Detection On Multilayer Social networks which is an unsupervised, parameter-free, and network feature-based technology. The proposed method can automatically identify anomalous users in a multilayer social network and rate them according to their anomalousness.

Current research aims to classify unusual activities using the methodology of behavior-based anomaly detection that reveals various patterns in the application of social media. Anomalous users are classified based on possible behavioural dissimilarity from others. Sudha et al. (2018) proposed a rich feature set using the K-means algorithm for identifying outliers in social media. An method was also proposed by the authors to provide a visual explanation of the results.

One of the methods to identifying malicious behaviours is social influence-based behavioural analysis. Savyan et al. (2017) proposed an unsupervised clustering algorithm to analyse the reactions given by users on the Facebook. The study of these reactions offers valuable information for detecting anomalous activity in Facebook accounts, as reactions are immediate.

To preserve confidentiality and security of social media users is paramount. Most conventional machine learning algorithm can not ensure privacy of the training data. To deal with this issue, Catak et al. (2018) suggested a privacy protection protocol approach for the extreme learning machine algorithm and included private classification protocols in order to avoid the possibility of sensitive data disclosure when outsourcing data analysis. Many research attempted to devise various methods for separating spam contents on the social media. However, predicting vulnerable users on the social media has not extensively done yet. According to the suicide prevention report of WHO, approximately 800000 people among 15–29 year old die by suicide every year. This rate increases due to social medias. In this article, we proposed a vulnerable user prediction model by cascading different machine learning algorithms. Basics of few machine learning algorithms are described in the next section.

Preliminaries about basic classifiers

Information theory

Information-theoretic measures are used to create and detect an appropriate anomaly detection model. Several Information-theoretic measures, such as entropy, conditional entropy, relative entropy, information gain and information cost, are used to explain the characteristics of a dataset. Definitions of these measures are given below:

- Entropy is a basic concept of information theory which detect the unexpected or anomaly of a collection of data items. For a dataset, D in which each data item belongs to a class $y \in C_D$, the entropy of D relative to the $|C_D|$ wise classification is defined as

$$H(D) = \sum_{y \in C_D} p(y) \log \frac{1}{p(y)} \tag{1}$$

Where $p(y)$ is the probability of y in D.

- Conditional entropy is the entropy of D given that Y is the entropy of the probability distribution ($P(x, y)$) as

$$H(D|Y) = \sum_{x,y \in C_D, C_Y} p(x, y) \log \frac{1}{p(x|y)} \tag{2}$$

where $P(x, y)$ is the joint probability of x and y and $P(x|y)$ the conditional probability of x given y.

- Relative entropy is the entropy between two probability distributions $p(x)$ and $q(x)$ defined over the same $y \in C_D$ as

$$reEn(p|q) = \sum_{y \in C_D} p(y) \log \frac{p(y)}{q(y)} \tag{3}$$

- Relative conditional entropy is the entropy between two probability distributions ($p(x,y)$ and $q(x,y)$) defined over the same $x \in C_D$ and $y \in C_D$ as

$$relConEn(p|q) = \sum_{x,y \in C_D, C_Y} p(x, y) \log \frac{p(x|y)}{q(x|y)} \tag{4}$$

- Information gain is a measure of the information gain of an attribute or feature B in a dataset D and is

$$Gain(D|B) = H(D) - \sum_{v \in Values_B} \frac{|D_v|}{|D|} \tag{5}$$

where values B is the set of possible values of B and D_v the subset of D where B has the value v.

Decision tree(C5.0)

Decision Tree (C5) is one of the most popular classification techniques. C5.0 produces smaller decision trees or fewer rules set in comparison with C4.5. The accuracy of C5.0 is good comparing with C4.5. C5.0 automatically allows removing unhelpful attributes. C5.0 algorithm provides less and efficient memory usage, Cross-validation, knowledge discovery, reduced error pruning and Feature selection facilities than C4.5.

Support vector machine (SVM)

SVM is a supervised learning model with associated learning algorithms used for classification and regression analysis. SVM is used in classification function to distinguish between two classes in the training data. Aims of the SVM is to find best line to separate data showed in Fig. 1. Support vector is used to construct model and not sensitive other data. We used Gaussian kernel in SVM for linear and non-linear space.

Naive bayesian classifier (NBC)

In Naive Bayesian approach, users in the social network were classified by calculating the posterior probability. Naive Bayesian is used to classify the disappointed users as

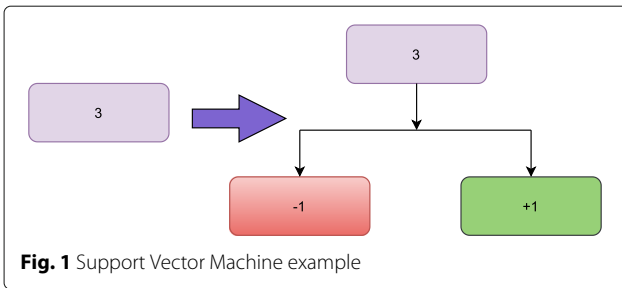


Fig. 1 Support Vector Machine example

suicidal and not suicidal based on the suicidal dictionary feature. Mathematically, Naive Bayes and Naive Bayesian theorems can be expressed into the following simple form:

$$Gain(D|B) = H(D) - \sum_{v \in Values_B} \frac{|D_v|}{|D|} \tag{6}$$

Where, $P(c|x)$ is the posterior probability of target class given predictor. $P(c)$ is the prior class probability. $P(x|c)$ is the likelihood which is the probability of predictor given class. $P(x)$ is the prior probability of predictor.

The proposed vulnerable social media users prediction model

In this proposed system, the three algorithms are combined for the development of anomaly detection system in the social network. Figure 2 shows the model of the proposed anomaly detection. We filter the dataset at multiple levels cascading different classifiers based on the characteristics of the machine learning algorithms. As, decision tree C5.0 algorithm can show better accuracy on large volume of dataset. So, Fig. 3 illustrates that the original

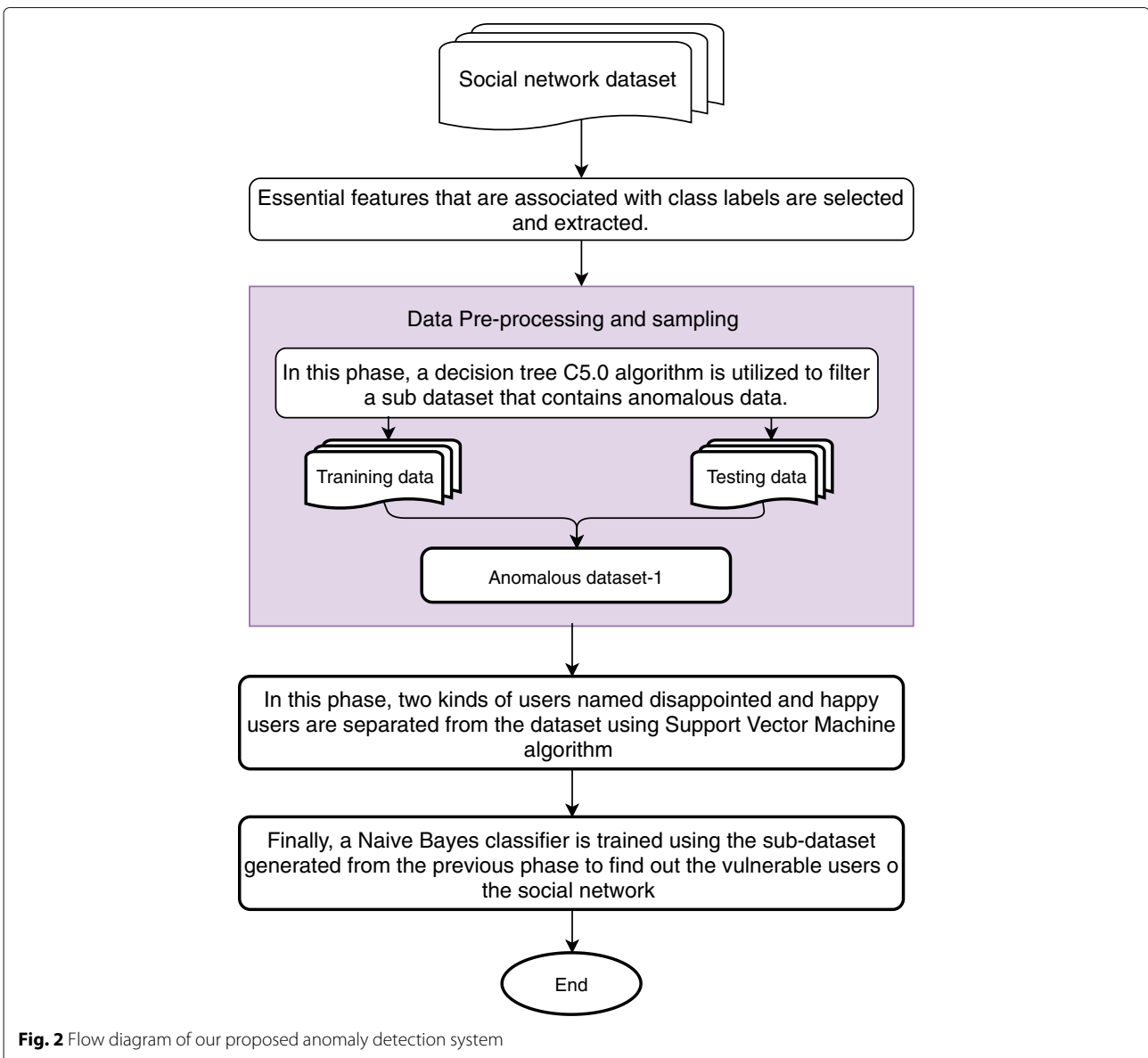


Fig. 2 Flow diagram of our proposed anomaly detection system

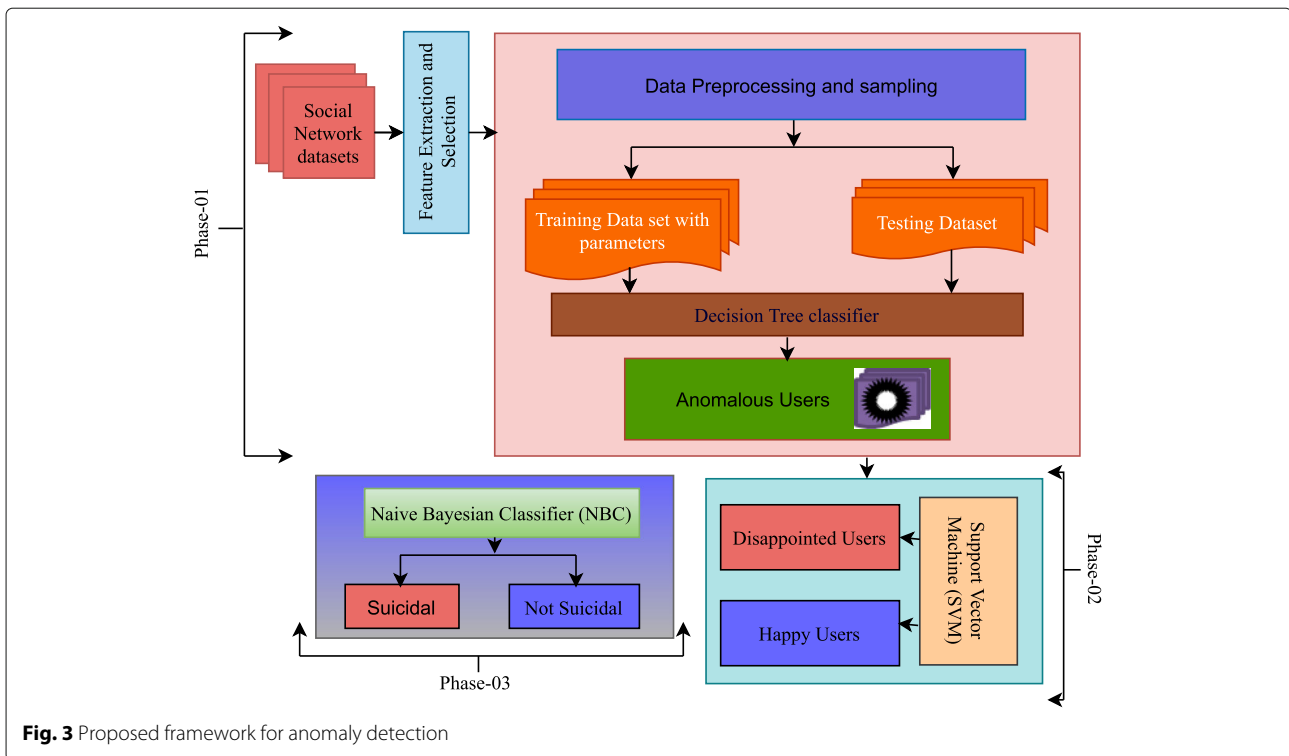


Fig. 3 Proposed framework for anomaly detection

size of dataset is first fed into C5.0 algorithm to filter all the anomalous entities. Next, the SVM is applied on the filtered data to find out only disappointed users as the SVM can result in higher accuracy when it is used as binary classifier. Finally, Naïve Bayesian classifier (NBC) is trained with the filtered dataset from the previous stage to recognize the potential suicidal users. Decision making process of the proposed method is depicted in Fig. 4. The reason of using NBC is that NBC shows higher accuracy on small volume of dataset. We describe the features of the proposed model below.

Features analysis

The use of machine learning approach depends on many factors to identify anomalous users in the social network. Among these, the most important factor is the selection of the most important features that distinguish anomalous users from legitimate class. Unlike normal users in the social network, abnormal users usually aim at commercial intent. In this section, we analyze an enormous set of features of users from both content and user’s behaviour point of view according to the synthetic dataset and real dataset extracted from the social network. We have collected two types of feature list:

- User’s Profile-based features: Features related to profile were used in Yang et al. (2013) to detect anomalous in the social network. The profile based features include the behavioural changes and the

basic information of a particular user. The examples of this kind of features are the number of followers, the number of wall post in specific time duration, the number of friends, and so on. We retrieved this type of features that depict the behavioral information of a user on his or her account of users from the Facebook.

- Content-based features: Content based features indicate unusual parts of the comments, status, wall post on the social networks. For Facebook, content based features can be tags, comments, URLs, likes, spam words, re-posts and hashtags contained in the profile post and message. We identified a significant numbers of content-based features by rating them.

Features in the training dataset

The important features of the datasets used in training machine learning algorithm in our model is described below:

1. **Total number of friends (Rathore et al. 2018; Ahmed and Abulaish 2013; Rahman et al. 2017)**(frnd) f_1 : The users who are virtually connected with an individual, are considered the friends of that individual on social networks. In general, regular users have a higher number of friends than spammers.
2. **Number of followers (Sohrabi and Karimi 2018) (followers) f_2** : The number of each user’s friends cannot exceed 5000 people in the social network

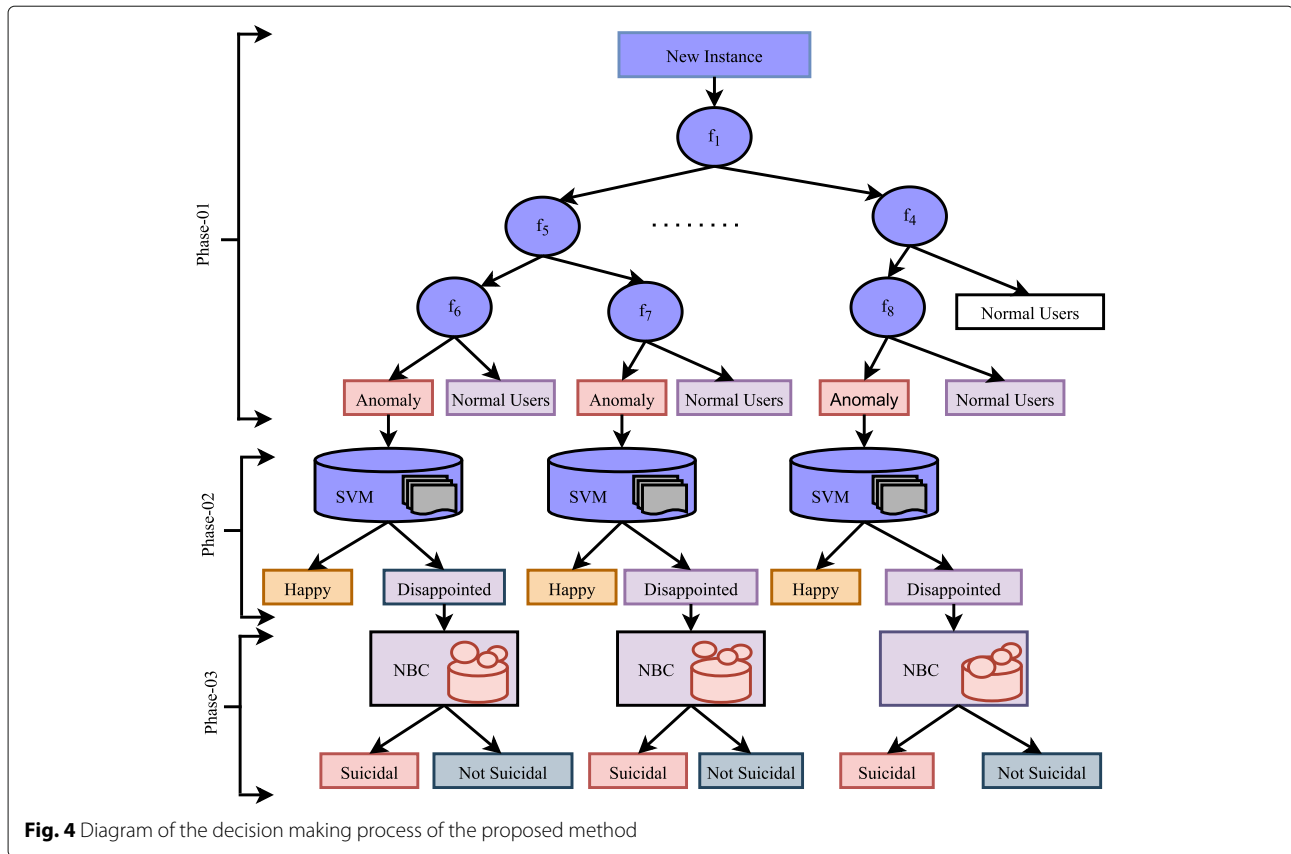


Fig. 4 Diagram of the decision making process of the proposed method

(Facebook) but the number of followers of a user can be unlimited. Followers are the users who follow a particular user. Spammers have a small number of followers. Therefore, a user with a small number of followers can potentially be considered as a spam account. Spammers do not usually have lots of followers.

3. **Number of negative emoticon symbol per day in message and post (f_3):** Normal users use emoticon symbols including sad, sigh, and happy in their messages and wall-posts to express their moods. A regular expression is used to find the number of emoticon symbol in the message.
4. **Number of links (URLs) (f_4) in the wall post:** Spammers can redirect their normal users (victims) to phishing website for collecting users' sensitive information for the malicious purpose (Chen et al. 2014). For this reason, the number of URLs that appear in each message and wall posts is computed using a regular expression.
5. **Frequency of phone number (f_5):** This feature represents the number of time a phone number appears in each message. Almeida et al. (2013) has shown that spammers use phone number

intentionally to lure their victim in their message. Another regular expression is used to extract this feature.

6. **Hash-Tag (Wall Posts, Comments) (f_6):** Putting hash-tags in message, comment and wall status is very popular in the social network. Hash-tags are more noticeable for spammers seen by more users.
7. **Likeness on wall post or status (f_7):** Users can use "like" button to give their positive opinions about wall status, photos or shared URLs links. The number of likes of wall posts and comments of anomalous users is much less than normal users. $f_7 = \frac{p_{like}}{n}$ where p_{like} is total number of like in n recent posts of a specific user.
8. **Frequency of money words (f_8):** In some situations, anomalous user tries to overpower normal users in the social network by sending unsolicited SMS that requests for money to perform illegal activities. So, we collected post containing the phrases related to money such as thousand, million, and trillion to compute the frequency of word " money" which is considered as one of the features.
9. **Frequency of money symbols or words (f_9):** Money symbol is used to perform crime activities.

This feature’s value is counted using a regular expression. Finally, we get the total number of time used the money symbol in the message.

10. **Message size (f_{10}):** Spam comments vary in sizes. Some of those include the URL(s). Some of them contain advertisements and are larger in sizes due to advertising. The explanation of the commodity will be in the form of comments, and the others contain both text and URL.
11. **Share (f_{11}):** Users in the social networks can share photos, videos, and blog entries with their friends. The number of sharing contents is an essential factor for detecting anomalous users.
12. **The Average Time Interval (f_{12}):** (BurstTime) Anomalous users usually send messages within short intervals. So, time taken for making a post is also an important feature to detect an abnormal user.
13. **Comments per wall post (f_{13}):** An anomalous user’s wall post has less unique comment than a legitimate user’s post in the social network.
 $f_{13} = \frac{p_{cmnt}}{n}$ Where p_{cmnt} is a total number of comments in n recent posts of a specific user.
14. **Post sharing (f_{14}):** Legitimate user’s post is shared more time than anomalous users post. So, post sharing can be also considered as another important feature to detect abnormal users.
15. **Disappointed word (f_{15}) and Suicidal word (f_{16}):** Legitimate users use less disappointed word and suicidal word in their post in social networks. f_{15} and f_{16} are calculated by averaging the counted word based on Tables 1 and 2 respectively.

The proposed approach depicted in Fig. 3 for detecting anomalies in Social networks is divided into the following three logical phases:

1. Phase-1: Data Pre-processing and Sampling (Feature Identification)- Behavior of anomalous users differs

Table 1 Disappointed Words Dictionary

Disappointed Word Dictionary
Despair, disappoint, Rehabilitation, Revelation, Risk, Sad, Scared, gloom, dejected, misery, hapless, desperate, ignore, despised, worried, moody, gloomy, loss, lonely, sadness, failure, grief, worry, suck, pain, sorrow, Oscillation, Percentage, Prevention, Problem, Psychiatry, Puzzle, Raw emotion, Recovery, Sense, Sensitive, Solution, Sorrow, Stigma, Humankind, Suffering, Likelihood, Mystery, psychiatric problem, Obsessive-compulsive, Mental health, Tragedy, Stress, Struggles, Stunned, Teenagers, Tendency, Terminal, Mental, fatigue, despondency, prostration, crisis, stress, dread, sulking, miserable, trouble, weariness, divast, pessimistic, grieving, frustat, offended, bitted, bad, unhappy, alone, confused, blindfold, betrayed, guilty, vicious, wicked, empty, angry, heartbroken, unbearable, nervous, shy, tired, silly, Happening, Hard, Harm, Support, Suspicion, Illness, Impact, Injury, Legend, Treatable, Treatment, Troubled, Unresponsive, Unworthy, Haunting, Health, Help, Socially, Treatable, Treatment, Troubled, Unresponsive, Unworthy.

Table 2 Suicidal Words Dictionary

Suicidal Word Dictionary
Upset, Crying, Controversy, stress, Addiction, shocking, suffering, crisis, Depression, Emotion, shock, dissatisfaction, good, Pain, blame, Claim, Brood, Devastating, Fear, tragedy, hope, fear, Hurt, conflict, suspicion, anxiety, Pain, Failure, Finality, Anger, Conflicted, despair, Tears, Death, Disease, Brain, Despair, Feel, worry, Feelings, Heartbreak, heart, Die, Fight, against, Hopelessness, Unworthy, Incoherent, Frustration, sorry, Diagnosis, Mourn, Compassion, Love, Personal, Fret, Loss, Grief, Life, Misunder, tood.

from legitimate users in the social network. In this step, we have identified a list of feature or characteristics used to identify the anomalous users. The features used in the model are described in details in the above section. Data Pre-processing and sampling data preprocessing which is the first phase of the model receives dataset with the selected features. If training dataset holding much irrelevant, noisy, unreliable data is used in the data mining process, incorrect outcomes might be produced which reduces accuracy level of the learning algorithm. Normally, data preparation and filtering steps take a considerable amount of processing time. We perform the pre-processing following the steps described below:

- (a) A set of users who has friends more than 3050, is deleted.
- (b) Users in the dataset who has more than 15000 followers is excluded.
- (c) Normalization is applied to construct the dataset based on different sampling equation.

The dataset labelled as anomalous and legitimate, were used for training the learning algorithms. In the pre-processing step, all the continuous features were converted into discrete using different threshold valued sampling equations. After pre-processing, we get the actual dataset that is used in our proposed method, where 70% data is assigned as training data and 30% data is assigned as testing data.

After that, decision tree (C5.0) is used to classify users account. The feature which has the highest information gain is consider as root of the tree and the tree is split based on that feature. In this technique, a decision was made at each level of the constructed tree based on the feature value. Finally, the user accounts were classified as legitimate user and anomalous users.

2. From the above Phase-1, we can collect the anomalous users for the extracted dataset in the social network. Then, anomalous users was inserted as a new dataset for Support Vector Machine (SVM) to classify the user as disappointed and happy users.

- From the above phase-2, disappointed user list was generated for the social network datasets to use the machine learning algorithm Naive Bayesian classifier (NBC) based on the selected features to detect the suicidal user. To classify the disappointed users using the probability, NBC theorem was used to calculate the probability. Mathematically, NBC theorem can be expressed into the following simple form:

$$P(c|x) = P(x|c) \times P(c) \tag{7}$$

Experimental analysis and evaluation

We used a standard PC (Intel Core i5-6500 3.20 GHz, 8 GB RAM) to conduct experiments. The model was implemented using R language and Python machine learning packages. The python packages include numpy, and sklearn libraries. The model consists of three phases where the first phase, second phase and third phases involves training C5.0, SVM and Naive Bayesian classification algorithm respectively. We have used information gain method for C5.0 algorithm and adopted percentage method for analyzing performances (70% training dataset and 30% testing dataset). We applied poly kernel based SVM in the second phase. To analyze the performances of Bays algorithm in the third phase, 10-fold cross validation process was applied.

To evaluate our proposed approach, we calculate the true positive rate (TPR) and the false positive rate (FPR) for various thresholds values. We consider the following metrics for analyzing outcomes of the model.

- The classification results are listed in the confusion matrix (Uddin et al. 2020) shown in Table 3, also called the contingency table. The True Positive upper left corner is the number of individuals that were listed as true positive, while those were true. The False-positive lower right cell reflects the number of samples that, though false, were labelled as false negative. False-negative shows the number of individuals, while these were false, being counted as true. False-positive reflects, as these were true, the number of individuals that were listed as true.

$$Accuracy = \frac{\sum True\ Positive + \sum True\ Negative}{\sum Total\ Samples} \tag{8}$$

Table 3 The confusion matrix

	Condition Positive	Condition Negative
Predicted Condition Positive	True Positive	False Negative
Predicted Condition Negative	False Positive	True Negative

Table 4 Accuracy Comparison for Anomalous User Detection on Dataset1

Algorithms	Recall	Precision	F-Measure	Accuracy
KNN (K=10)	0.936120	0.983026	0.959000	0.951073
NBC	0.973680	0.947904	0.960619	0.956849
SVM	0.963492	0.965959	0.964724	0.959741
C4.5	0.931444	0.997881	0.963519	0.966161
Random Forest	0.959831	0.983749	0.971643	0.975035
DT-SVMNB	0.96609	0.985876	0.975883	0.978178

$$Precision = \frac{\sum True\ Positive}{\sum Predicted\ Condition\ Positive} \tag{9}$$

$$Recall = \frac{\sum True\ Positive}{\sum Condition\ Positive} \tag{10}$$

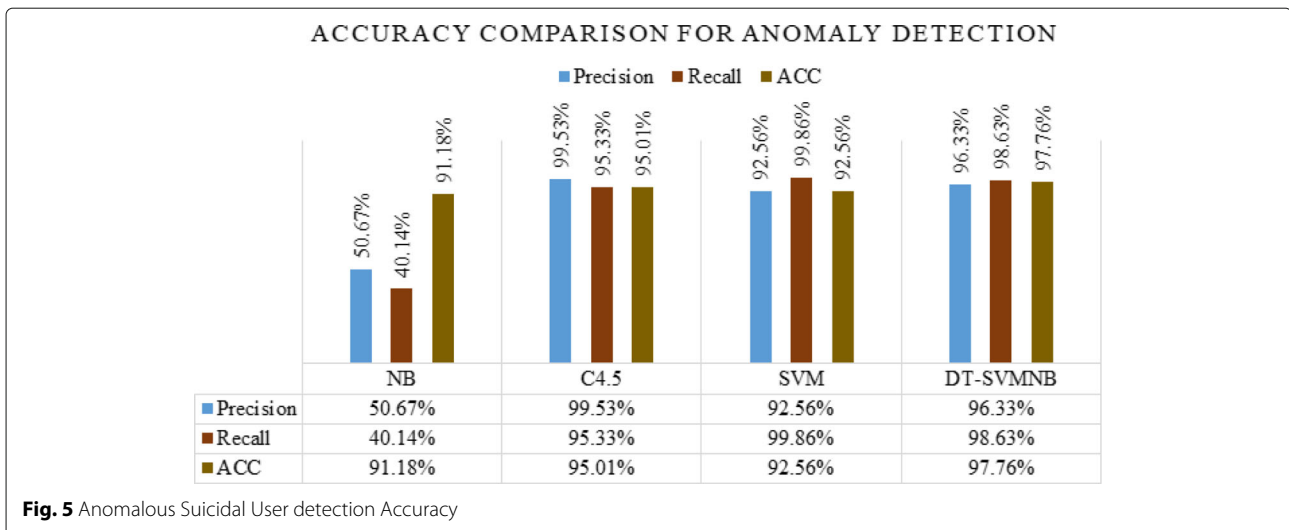
- Receiver The Operating Characteristic Curve (or ROC Curve) (Ashraf Uddin et al. 2020) is a plot of the true positive rate for the various potential diagnostic test cutpoints against the false-positive rate. The trade-off between sensitivity and specificity is exposed by ROC (a reduction in specificity would follow any increase in sensitivity). The more the curve follows the left border and the more closely the curve follows the ROC space’s top border, the more precise the test would be.

In the proposed hybrid algorithm, we use the C5.0 decision tree in the first phase of DT-SVMNB that out-comes higher accuracy than other algorithms to detect the anomalous users in the social network. Table 4 shows the Recall, Precision and ACC metrics values of the first phase for the synthetic dataset and Table 5 shows performance values of the first phase for real datasets. Tables 4 and 5 illustrates that Decision tree (C5.0) provides better accuracy than NBC, SVM, decision tree(C4.5)and KNN algorithms. Though SVM also shows better accuracy, SVM requires longer time than C5.0 for large dataset.

We have generated a sub dataset named dataset1 from the synthetic dataset. This dataset contains 18615 users

Table 5 Accuracy Comparison for Anomalous User Detection on Dataset2

Algorithms	Recall	Precision	F-Measure	Accuracy
KNN (K=10)	0.854452	0.853477	0.853964	0.838332
NBC	0.979729	0.620542	0.759825	0.835821
SVM	0.859259	0.661912	0.747784	0.813134
C4.5	0.998808	0.603312	0.752244	0.823811
Random Forest	0.829763	0.789699	0.809236	0.835623
DT-SVMNB	0.904228	0.754500	0.822606	0.855729



with features values. Among these, the number of anomalous user is 4893, and the number of normal users is 13722. To generate real dataset, the user profile list was created from the initial data library by crawling the followings and friends for each profile. For this process, we have used a web crawler. Proposed features were extracted using Facebook-Graph API (Facebook-tools, accessed May 20, 2020). Thus, we successfully obtained 10,513 Facebook profiles.

Anomalous suicidal user detection accuracy of the proposed method is illustrated in Fig. 5. Figure 5 shows the Accuracy, Precision and Recall of DT-SVMNB, SVM, C4.5 and NBC, respectively for the dataset1. The Accuracy of DT-SVMNB which is approximately 98% is higher than the other algorithms. This paper evaluates the performance of DT-SVMNB algorithms to detect the abnormal and suicidal user in social network and then compare the performance of DT-SVMNB with some studied anomaly detection algorithms. The result shows that DT-SVMNB outperforms other anomaly detection algorithms.

Conclusions

In recent year, many researchers have worked on anomaly detection in different fields including social network. In this paper, we discussed the importance of detecting anomaly in the social network. We have developed a machine learning-based anomaly detection solution for social networks. Sequentially, we define anomaly detection and presented the existing anomalous users detection approaches in social network. Feature extraction of users from the social network in our proposed solution is based on statistical analysis and manual selection. Importantly, our methodology is able to detect anomaly patterns unknown to the trained model, thus showing promising results for real-life cybersecurity applications in common OSNs. In this work, we have proposed an efficient hybrid

system named as DT-SVMNB for anomalous users detection and suicidal users detection simultaneously based on the proposed features list to solve some problem of existing works. The solution considers the user’s content and behaviour feature, and apply them into DT-SVMNB based algorithm for anomalous user and suicidal users classifications. Further, we set up an extensive comparison between our proposed approach and other existing techniques using a real and a synthetic set of data.

Our future work will be to explore user interest dynamically through different activities to characterize user interest patterns comprehensively. As our detection approach is online, our further future work is to devise an online real-time detection system.

Abbreviations

SVM: Support Vector Machine; NBC: Naive Bayesian classifier; OSN: Online Social Networks; DT: Decision Tree

Acknowledgements

The authors would like to thank all the anonymous reviewers for their rigorous review and comments in several revision rounds. The reviews are detailed and helpful to improve and finalize the manuscript. The authors are highly grateful to them.

Authors’ contributions

MS Rahman and S Halder conceived the study and participated in the proposed design. MS Rahman assisted in collecting the data and generating the result. All authors have read and approved the final manuscript.

Funding

Not applicable.

Availability of data and materials

The datasets generated during and analysed during the current study are available from the corresponding author on reasonable request.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Department of Computer Science and Engineering, Dhaka International University, Dhaka, Bangladesh. ²Department of Computer Science and

Engineering, Jagannath University, Dhaka, Bangladesh. ³Department Computer Science and Information Technology, RMIT University, Melbourne, Australia. ⁴Internet Commerce Security Laboratory, Federation University Australia, Ballarat VIC 3350, Australia.

Received: 27 October 2020 Accepted: 26 January 2021

Published online: 02 March 2021

References

- Abulaish M, Bhat SY (2015) Classifier ensembles using structural features for spammer detection in online social networks. *Found Comput Decis Sci* 40(2):89–105
- Adewole KS, Anuar NB, Kamsin A, Varathan KD, Razak SA (2017) Malicious accounts: dark of the social networks. *J Netw Comput Appl* 79:41–67
- Ahmed F, Abulaish M (2013) A generic statistical approach for spam detection in online social networks. *Comput Commun* 36(10–11):1120–1129
- Aljawarneh S, Aldwairi M, Yassein MB (2018) Anomaly-based intrusion detection system through feature selection analysis and building hybrid efficient model. *J Comput Sci* 25:152–160
- Almeida T, Hidalgo JMG, Silva TP (2013) Towards sms spam filtering: Results under a new dataset. *Int J Inf Secur Sci* 2(1):1–18
- Ashraf Uddin M, Stranieri A, Gondal I, Balasubramanian V (2020) Dynamically recommending repositories for health data: a machine learning model. In: *Proceedings of the Australasian Computer Science Week Multiconference*. ACM, pp 1–10. <https://dl.acm.org/doi/abs/10.1145/3373017.3373041>
- Belavagi MC, Muniyal B (2016) Performance evaluation of supervised machine learning algorithms for intrusion detection. *Procedia Comput Sci* 89:117–123
- Benevenuto F, Rodrigues T, Cha M, Almeida V (2012) Characterizing user navigation and interactions in online social networks. *Inf Sci* 195:1–24
- Bindu P, Thilagam PS, Ahuja D (2017) Discovering suspicious behavior in multilayer social networks. *Comput Hum Behav* 73:568–582
- Caruana G, Li M (2012) A survey of emerging approaches to spam filtering. *ACM Comput Surv (CSUR)* 44(2):9
- Çatak FÖ, Mustacoglu AF (2018) Cpp-elm: cryptographically privacy-preserving extreme learning machine for cloud systems. *Int J Comput Intell Syst* 11(1):33–44
- Chen C-M, Guan D, Su Q-K (2014) Feature set identification for detecting suspicious urls using bayesian classification in social networks. *Inf Sci* 289:133–147
- Chu Z, Widjaja I, Wang H (2012) Detecting social spam campaigns on twitter. In: *International Conference on Applied Cryptography and Network Security*. Springer, pp 455–472
- Erdélyi M, Garzó A, Benczúr AA (2011) Web spam classification: a few features worth more. In: *Proceedings of the 2011 Joint WICOW/AIRWeb Workshop on Web Quality*. ACM, pp 27–34. <https://dl.acm.org/>
- Gupta A, Kaushal R (2015) Improving spam detection in online social networks. In: *2015 International Conference on Cognitive Computing and Information Processing (CCIP)*. IEEE, pp 1–6. <https://ieeexplore.ieee.org/document/7100738>
- Islam MR, Kabir MA, Ahmed A, Kamal ARM, Wang H, Ulhaq A (2018) Depression detection from social network data using machine learning techniques. *Health Inf Sci Syst* 6(1):8
- Manjunatha H, Mohanasundaram R (2018) Brnads: Big data real-time node anomaly detection in social networks. In: *2018 2nd International Conference on Inventive Systems and Control (ICISC)*. IEEE, pp 929–932. <https://ieeexplore.ieee.org/abstract/document/8398937>
- Martinez-Romo J, Araujo L (2013) Detecting malicious tweets in trending topics using a statistical analysis of language. *Expert Syst Appl* 40(8):2992–3000
- Rahman MS, Dey LR, Haider S, Uddin MA, Islam M (2017) Link prediction by correlation on social network. In: *2017 20th International Conference of Computer and Information Technology (ICIT)*. IEEE, pp 1–6. <https://ieeexplore.ieee.org/abstract/document/8281812>
- Rathore S, Loia V, Park JH (2018) Spamspotter: An efficient spammer detection framework based on intelligent decision support system on facebook. *Appl Soft Comput* 67:920–932
- Rathore S, Sangaiah AK, Park JH (2018) A novel framework for internet of knowledge protection in social networking services. *J Comput Sci* 26:55–65
- Savvan P, Bhanu SMS (2017) Behaviour profiling of reactions in facebook posts for anomaly detection. In: *2017 Ninth International Conference on Advanced Computing (ICoAC)*. IEEE, pp 220–226. <https://ieeexplore.ieee.org/abstract/document/8441402>
- Sohrabi MK, Karimi F (2018) A feature selection approach to detect spam in the facebook social network. *Arab J Sci Eng* 43(2):949–958
- Sudha MS, Priya KA, Lakshmi AK, Kruthika A, Priya DL, Valarmathi K (2018) Data mining approach for anomaly detection in social network analysis. In: *2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT)*. IEEE, pp 1862–1866. <https://ieeexplore.ieee.org/abstract/document/8472985>
- Thaseen IS, Kumar CA (2017) Intrusion detection model using fusion of chi-square feature selection and multi class svm. *J King Saud Univ-Comput Inf Sci* 29(4):462–472
- Thaseen IS, Kumar CA, Ahmad A (2019) Integrated intrusion detection model using chi-square feature selection and ensemble of classifiers. *Arab J Sci Eng* 44(4):3357–3368
- Uddin MA, Stranieri A, Gondal I, Balasubramanian V (2020) Rapid health data repository allocation using predictive machine learning. *Health Inf J* 26(4):3009–3036. SAGE Publications Sage UK: London, England
- Wang D, Irani D, Pu C (2014) Spade: a social-spam analytics and detection framework. *Soc Netw Anal Min* 4(1):189
- Xu H, Sun W, Javadi A (2016) Efficient spam detection across online social networks. In: *2016 IEEE International Conference on Big Data Analysis (ICBDA)*. IEEE, pp 1–6. <https://ieeexplore.ieee.org/abstract/document/7509829>
- Yang C, Harkreader R, Gu G (2013) Empirical evaluation and new design for fighting evolving twitter spammers. *IEEE Trans Inf Forensic Secur* 8(8):1280–1293
- Yang Z, Wilson C, Wang X, Gao T, Zhao BY, Dai Y (2014) Uncovering social network sybils in the wild. *ACM Trans Knowl Discov Data (TKDD)* 8(1):2
- Yasami Y, Safaei F (2017) A statistical infinite feature cascade-based approach to anomaly detection for dynamic social networks. *Comput Commun* 100:52–64
- Yazdi HS, Bafghi AG, et al. (2020) A drift aware adaptive method based on minimum uncertainty for anomaly detection in social networking. *Expert Syst Appl* 162:113881
- Yu D, Chen N, Jiang F, Fu B, Qin A (2017) Constrained nmf-based semi-supervised learning for social media spammer detection. *Knowl-Based Syst* 125:64–73
- Zephoria Digital Marketing (2018) The Top 20 Valuable Facebook Statistics – Updated April 2018. <https://zephoria.com/top-15-valuable-facebook-statistics/>. Accessed 11 May 2018
- Zheng X, Zeng Z, Chen Z, Yu Y, Rong C (2015) Detecting spammers on social networks. *Neurocomputing* 159:27–34
- Zhou B, Yao Y, Luo J (2014) Cost-sensitive three-way email spam filtering. *J Intell Inf Syst* 42(1):19–45

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen® journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)