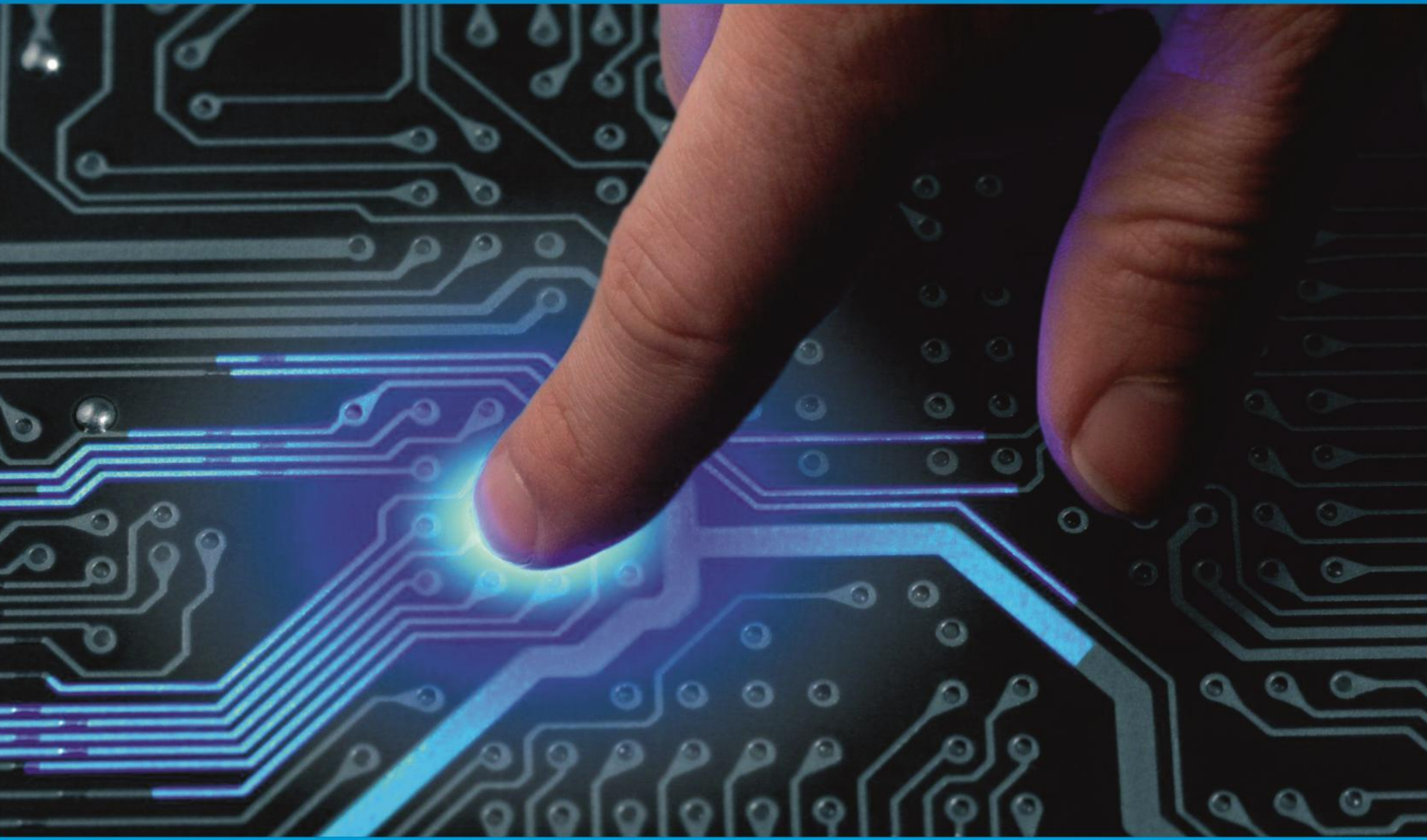
IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 8, Issue 11, November 2020

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 7.488

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | Impact Factor: 7.488 |

|| Volume 8, Issue 11, November 2020 ||

A Method of Fake Profile Detection using Random Forest with Feature Engineering

Priyanka Tiwari, Ankur Mudgal

PG Student, Dept. of CSE., SRIST Jabalpur, RGPV University, Bhopal, India

Assistant Professor, Dept. of CSE., SRIST Jabalpur, RGPV University, Bhopal, India

ABSTRACT: The social networking sites are making our social lives better but nevertheless there are a lot of issues with using these social networking sites. The issues are privacy, online bullying, potential for misuse, trolling, etc. These are done mostly by using fake profiles. In today’s online social networks there have been a lot of problems like fake profiles, online impersonation, etc. The users don’t care much about what they are sharing on social networks and the privacy of their post; it can be quite troublesome for them. On top of all benefits of online social networks, there are many disadvantages identity theft is one of the biggest concerns of online social networks. There are millions of fake profiles. Research work contains development of fake profile detection on social network using natural language processing applied in their features. In our proposed plan, we propose machine learning techniques such as Random Forest for detecting the fake accounts on Social Networks. In this, we classify the data as fake or real profiles using existing database using above techniques, which identifies the fake accounts on the social networking sites.

KEYWORDS: Facebook, Fake Accounts, Feature Selection, Clustering, Classification, Random Forest.

I. Introduction

Mobile Online social networks (OSNs), such as Facebook, Twitter, RenRen, LinkedIn, Google+, and Tuenti, have become increasingly popular over the last few years. People use OSNs to keep in touch with each others, share news, organize events, and even run their own e-business. Online Social Networks (OSNs) have also attracted the interest of researchers for mining and analyzing their massive amount of data, exploring and studying user’s behaviours as well as detecting their abnormal activities [1]. In [2] researchers have made a study to predict, analyze and explain customer’s loyalty towards a social media-based online brand community, by identifying the most effective cognitive features that predict their customers attitude. The implications of researchers attempt may helps an OSN operator detecting fake accounts efficiently and effectively, hence, improve the experience of their users by preventing annoying spam messages and other abusive content. The OSN operator can also increase the credibility of its user metrics and enable third parties to consider its user accounts [3]. Information security and privacy are among the primary requirements of social network users, maintaining and providing those requirements increases network credibility and subsequently its revenues. As recently, banks and financial institutions in U.S. have started to analyze Twitter and Facebook accounts of loan applicants before actually granting the loan [4].

The open nature of OSNs and the massive amount of personal data for its subscribers have made them vulnerable to Sybil attacks. In 2012, Facebook noticed an abuse on their platform including publishing false news, hate speech, sensational and polarizing, and others [5]. These phenomena raised the flag for the need of new techniques to detect such actions and avoid them. In 2015 Facebook estimated that nearly 14 million of its monthly active users are in fact undesirable, representing malicious fake accounts that have been created in violation of the websites terms of service. Facebook, for the first time, shared a report in the first quarter of 2018 that shows their internal guidelines used to enforce community standards covering their efforts between October 2017 to March 2019, this report illustrates the amount of undesirable content that has been removed by Facebook [6], and it covers six categories:

- ❖ graphic violence
- ❖ adult nudity
- ❖ sexual activity
- ❖ terrorist propaganda
- ❖ hate speech
- ❖ spam

In addition, 837 million posts containing spam have been taken down, and about 583 million fake accounts have been disabled, Facebook also has removed around 81 million undesirable content in terms of the rest violating content types. However, even after preventing millions of fake accounts from Facebook, it was estimated that around 88 million accounts are still fake [6]. Statistics show that 40% of parents in the United States and 18% of teens have a great concern about the use of fake accounts and bots on social media to sell or influence products. Another example, during the 2012 US election campaign, the Twitter account of challenger “Romney” experienced a sudden jump in the number of followers. The great majority of them were later claimed to be fake followers [7]. In December 2015, Adrian Chen, a reporter for the New Yorker, noted that he had seen a lot of the Russian accounts that he was tracking switch to pro-Trump efforts, but many of those were accounts that were better described as troll’s accounts managed by real people that were meant to mimic American social media users [8]. Similarly, before the general Italian elections of February 2013, online blogs and newspapers reported statistical data over a supposed percentage of fake followers of major candidates [9].

In 2017, fake posts have shared a roamer on social media that the actor Clint Eastwood has been dead, however, the claims were proven to be false In general, attackers follow the concept of having OSNs user accounts are “keys to walled gardens” [10], so they deceive themselves off as somebody else, by using photos and profiles that are either snatched from a real person without his/her knowledge, or are generated artificially, to spread fake news, and steal personal information. These fake accounts are generally called imposters [11]. To enhance their effectiveness, these malicious accounts are often armed with stealthy automated tweeting programs, to mimic real users, known as bots [12]. OSNs are employing different detecting algorithms and mitigation approaches to address the growing threat of fake/malicious accounts. Though Sybil accounts find a way to cloak their behavior with patterns resembling real accounts [12], [13], they manifest numerous profile features and activity patterns. Thus, automated Sybil detection is not always robust against adversarial attacks, and does not yield desirable accuracy. In all cases, such fake accounts have a harmful effect on users, and their motives would be anything other than good intentions as they usually flood spam messages, or steal private data [14], [15]. Inspired by the importance of this problem, researchers focus on identifying fake accounts through analyzing user level activity by extracting features from recent user’s e.g number of posts, number of followers, profiles. They apply trained machine learning technique for real/fake accounts classification. Another approach is using graph level structure where the OSN is modeled as a graph essentially presented as a collection of nodes and edges. Each node represents an entity (e.g. account), and each edge represents as a relationship (e.g. friendship). Though Sybil accounts find a way to cloak their behaviour with patterns resembling real accounts, they manifest numerous profile features and activity patterns. Thus, automated Fake account detection is not always robust against adversarial attacks, and does not yield desirable accuracy.

In this thesis, a classification algorithm has been used by running the Random Forest (RF) classification algorithm on the decision values. In addition, we also validated the detection performance of our classifiers over two other sets of real and fake accounts, disjoint from the original training dataset.

1.4 Characteristics of fake facebook friends

With all the personal information users entrust to Facebook, it's no surprise that scammers are also on Facebook, often posing as friends. When using Facebook, you will need to be aware of this scam in order to keep your personal information safe. Here are four characteristics of a fake Facebook profile that you should consider before confirming a friendship.

While it's flattering to receive a friend request, it's important to make sure you know who the person, and that they are someone you trust. If you accept every friend request you receive, even if you don't know them, then you will make your personal information available to hackers. Whenever someone befriends you on Facebook, they are no longer subject to your account security settings and they can view all of your posted information. In an attempt to gain this level of clearance, a hacker will make a fake Facebook profile and send out thousands of friend requests to random people. The fake profile may even include a normal-looking profile picture, along with other deceptive features to make a fake account appear legit. This is why it is important to ensure your account is secure. Before you confirm the friendship, you will want to investigate the profile and judge for yourself if the account is actively being used by a real person, or if it's something that was quickly slapped together by a hacker. Here are four clues that will tell you if a Facebook account is real or fake.

A) Pictures

Check out the user's pictures. A hacker may use a legitimate-looking profile image to draw you in (likely ripped off from someone else's profile) [16], and they may even add a few more generic pictures in their photo album to make the account appear more credible. Often times, the hacker will not take the time to use the same person for all of their pictures, this is a dead giveaway that the account is not real.

B) Joined Facebook Date

A hacker will make fake Facebook accounts all day long, and immediately send friend requests to random people as soon as the account is made. If the account sending you the friend request was started within the last day or two, and there are no mutual friends, then it's likely a fake account.

C) Account Inactivity

With a normal Facebook account, the user will post regularly, even if the user posts to their profile only once a month. You should be able to view a few of their posts, giving you insight into their life. A fake profile that's thrown together by a hacker will often not have any posts on their timeline [17].

D) Send a Message

You don't have to be Facebook friends with someone to send them a message through Facebook. You can easily send a person a message that's sending you a friend request. It's okay to ask them a basic question like, "Where do I know you from?" A hacker will often not write you back, and if they do, you can tell from their response if they are a hacker or not. If you discover that they're a hacker, you can block them from contacting you and even report them to Facebook so they won't harass anybody else.

1.5 Evolution of Fake Profiles in OSNs

OSNs provide huge amount of user generated content easily, so it's always under attack of spammers. Mostly the aim of these cyber criminals is to steal the user's personal, professional, political, social or financial information by exposing the users with unwanted information on the web likes, pornography etc. in order to deceive them. There are number of methods by which the users' data can be hacked by these adversaries, and creating fake profiles to perform malicious activities on OSNs is one of the mostly employed methods. From the users' point of view, personal, professional and even financial data is no more secure. Figure 1 provides a quick view of various kinds of fake profiles and several other kinds of profiles found in different online social networks.

Profiles which follow the rules and regulations provided by particular OSN service are real. Here rules and regulations in the context of OSN may mean the owner should not have more than one personal account, it should not spread any unlawful, misleading, malicious, or discriminatory content, and it should not collect the user's information or access Facebook by automated means (such as bots and spiders)[16,17]. A person handling more than one account; i.e. an account other than his principal account is categorized as fake1. Facebook has provided several options to enhance the privacy of user accounts like protecting the password and sending location specific login alerts and location alerts. Users can also use the extra security features of the network like how to logout from another device, how to keep Facebook password safe using app passwords etc. [18]. Moreover an approach called "Safebook" [19] is proposed in order to protect user's personal data from both the malicious users as well as service providers who violate privacy rules. In order to avoid cyber attacks, one should take proper care while using online social account. Also at the time of account creation, the terms and conditions should not be violated.

II. Related work

One of the most powerful and relatively cost effective techniques to solve problems involving big data is machine learning. Machine learning is defined by Arthur Samuel as the "field of study that gives computers the ability to learn without being explicitly programmed" Buczak and Guven (2015). Methods that use machine learning extract and gain knowledge from experience and analytical observations. Because of these advantages of machine learning, the applications of its techniques are expansive, including business, education, biological and medical studies and our scope of interest; cybersecurity.

The three main categories of this field are supervised, semi-supervised and unsupervised approaches.

A) Supervised approach: Supervised learning-based detection approaches are the most common among all detection methods categories. The goal of it is to build a model of the distribution of class labels in terms of predictor features. Then, given values of the predictor features, the resulting classifier predicts class labels for unlabeled instances. In other words, a classifier learns how to identify accounts (as bots or not) based on a known set of accounts features, by training this classifier on a set of labeled accounts. Therefore, a dataset of labeled instances along with their extracted features is needed to train the model and build the classifier. Models performance is dependent on the significance of the set of discriminating features, and the efficiency of the training set.

Mainly, there are two types of bots features exploited. First, behavioral features; such features describe users information (metadata), their actions, interactions, timestamps and may include text counting without deeply analysing textual content. Second, content features, in which users textual content is analysed to discriminate bots from real users. Articles that employ supervised methods can be mainly categorized into methods that exploit behavioral features only, referred to as behavior-based techniques. And methods that mainly use content features, sometimes combined with behavioral features, referred to as content-based techniques. A brief summary about each article can be found below.

i) Behaviour-based: The well-known BotOrNot [20] is an off-the-shelf system that leverages more than one thousand features to discriminate bots. It measures the 'botness' of a Twitter account; the likelihood that a given account is bot. It is available for public use through a website10. The authors extended their work in Varol et al. (2017) [21] by training the model on a new dataset (Lee et al., 2011) and released information about the features leveraged in their technique. Similarly, Alarifi et al. [22] created a Chrome browser plug-ins that can indicate if a Twitter account is human, Sybil or Cyborg. The authors created a dataset that is publicly available and proposed a set of extracted features. As well, Kantepe and Ganiz [23] extracted a set of the most efficient features inspired by the given features in DARPA competition. Then, they exploited these features to classify a set of Twitter accounts to normal and bot accounts. David et al. [24] identified a set of features to detect Sybils on Twitter platform. They measured the importance of each feature and applied five different classification approaches; Random Forest Classifier performed the best. Also, Khaled et al. [25] proposed detecting fake accounts and bots on Twitter using SVM-NN algorithm, which combines Supported Vector Machine and Neural Networks. The authors reduced the set of features addressed in Yang, to reduce the cost. Similarly, Velayutham and Tiwari [26] identified a list of ten user profile attributes and tweet pattern to calculate its 'botScore' of Twitter accounts, which is similar to the botness score in BotOrNot. The authors proposed BotClassifier; a supervised classification algorithm that shows better classification results compared to Naive Bayes classifier.

Besides, Ji et al. [27] empirically evaluating the evasion mechanism of existing Social bots. Then, they introduced a list of Social bots features as well as a list of some evasion mechanisms detected. The authors used nine new features along with other nine features to detect Social bots on a variety of platforms. By combining human inelegance and machine learning, Teljstedt et al. proposed a semiautomatic approach to detect bots on Twitter. The authors suggested using this method to label large ground truth sets, because optimized precision is the priority of this approach. In a different strategy that focuses on user's social status, Gilani et al. [28] partitioned the accounts in a Twitter dataset based on the popularity of the accounts, and identified a group of features that are efficient in classifying the accounts in each group. The authors classify the Twitter accounts into humans or automated agents (which can be non-malicious agents), rather than detecting malicious bots. Yet, their method, their set of features and their dataset might be useful for future detection methods.

B) Unsupervised approach

In unsupervised machine learning, the algorithm finds its way to cluster the input data. In other words, there is no need for labeled data to detect bots using this approach, and rather than depending on distinct features' values to classify each account, the unsupervised approaches focuses more on what is common between groups of accounts and cluster accounts based on similarity between accounts in a single cluster. In the proposed taxonomy, the articles which use unsupervised learning methods are divided in a similar way to articles that use supervised learning methods, namely, behavior-based and content-based. The following shows a review of approaches that fall under this category.

i) Behavior-based: A well-known model that detects SMBs by using unsupervised machine learning techniques is DeBot. The authors assume that human users are not expected to have many correlated activities for a long period of time. Therefore, they built DeBot which detects bots on Twitter based on their correlated activities. By focusing on accounts that tweet at least forty tweets in an hour, the highly correlated activities within groups of accounts will be the reason to flag them as bots by DeBot. A slightly different model was designed by Cresci et al [29]. The authors proposed the Digital DNA model, which encodes the sequence of online behavior for an account on social media. After comparing a digital DNA fingerprint of Twitter accounts, the authors suggested that groups which share similar sequence of actions (Longest Common String) are Spam bot campaigns. In their later work Cresci [29], applied the DNA-inspired model in both supervised and unsupervised approaches, where the unsupervised approach was more convenient to their model. Another approach was done by Ahmad and Abulaish. Throughout their work, the authors leveraged the fact that spammers use multiple accounts, which means that the features' values of the accounts in Spam campaigns can be an indicator to identify them. The work uses seven features to model profiles using a weighted graph. Spam profiles will be inter-linked based on the information contained in the identified set of features. Also, Minnich et al. [30] used a similar approach in their model BotWalk. By creating a vector representation for each user's features, BotWalk uses seed bots and this vector representation to detect Social bots on Twitter. The bot seeds are found by using DeBot, then the model crawls to connected users and detect anomalous users. The authors keep their dataset publicly available.

ii) Content-based In order to spread a message, bots must be open and they have to duplicate content. Chew (Chew, 2018) leveraged these assumptions to detect patterns of similarities and detect automated Twitter accounts, which appear to be Influence bots. The author used tweets to detect emergent patterns for groups of accounts, and claims that no further ground truth is needed to indicate that such accounts are automated. In a similar concept, Chen et al. exploited the used URL shortening services and content similarity to detect Spam bot campaigns on Twitter. In a later work, Chen and Subramanian [31] developed a system that detects Spam bot campaigns on Twitter. By monitoring the top trending URLs in tweets on Twitter's real-time streaming, the system collects the groups of accounts that share the same tweets' text. It flags the accounts as bots if the accounts share similar recent tweets, then a classifier is used to find Spam bot campaigns that share similar tweet text, and the system maps each campaign to the registrant email of the URL that the campaign shared. Abu-El-Rub & Mueen (2019) [32] leveraged content to detect SMBs in BotCamp. The model exploits trending topics to detect Social bot campaigns interested in political discussions.

7.2. Features

When trying to discriminate bot accounts from real human accounts, the first question to come to mind is: what are the discriminating features, and how does a bot account differ from a human account? Some researches rely on one feature to uniquely identify bot accounts. Such as screen names in Beskow and Carley (2019) [33] and posts locations in Echeverria and Zhou (2017). But it is easy for botmasters to create bots that can evade detection by such models, because overcoming one vulnerability is not an impossible task. To solve this issue, many researchers depended on a predefined list of features together with a labeled set of accounts fed to the machine. The machine's role is to find thresholds the features' values that help making decision (if an account is bot or not) or calculating the botness of an account. This is the main scenario for methods that use supervised machine learning. For instance, a table of 1150 features used in this way can be found in Varol et al. (2017) [55]. (Qi, AlKulaib, & Broniatowski, 2018) [69] Suggested a development of how features are used for detecting different types of bot accounts using features that are relevant to each type. Their motive was that each type of bots must have different features based on the objectives of each type. That is, Spam bots are more likely to have more posts than Social bots, whereas Social bots will have more connections with legitimate users than Spam bots. The authors developed the set of Twitter Bots features in Nimmo (2017) and composed a set of specialized bot type related features that is publicly available in their paper. Other researchers employed pattern recognition component to detect patterns in features instead of finding thresholds. The motive behind doing so is that algorithms used for managing bot accounts will eventually leave a pattern that can be exploited to identify bots. For example, patterns in temporal information for an account or in its posts textual content might be a result of an algorithm.

Profile information is sometimes called account's properties, e.g. profile's age, profile picture, username and so on. Whereas, posts content is widely used by researchers, because they are the carriers of the objectives of many types of bots. For example, Spam bots that promote a certain website are very likely to post URLs, while political bots will share political sentiments. Some works involved Spam detection component which compares the words in a post with known Spam datasets (Main & Shekokhar, 2015), others focus on emotions and sentiment analysis. Countless information can be extracted from this category of features. Posting behavior information is closely related to the previous category. Features of this category include temporal information and posts frequency. Some of the well-known detection models depend on detecting regular or periodic timing or comparing temporal information between accounts and detect synchronized accounts. Lastly, network structure category involves information about the community nature of a user, such as, interactions among users. Some authors assume that bots absence from the real world will prevent them from emerging in virtual communities. Moreover, many graph-based methods depend on the network structure in modeling users. Privacy is one of the obstacles that prevent researchers from getting some of the needed accounts features information. It causes less information exploited to classify and judge an account. This is one of the reasons why about 85% of the real datasets in the reviewed literature are of Twitter datasets, as Twitter users tend to keep their profiles public. It is obvious how critical the features chosen to detect SMBs are. As described before, one of the main challenges in this research area is the fast evolving abilities of SMBs, which causes changes in the values of the discriminating features and failing to detect them. Therefore, finding robust features is of much interest. An example of robust features is finding vulnerable victims as their features are relatively stable. Devakunchari Ramalingam, Valliyammai Chinnaiah [34] proposed a survey of the existing and latest technical work on fake profile detection.

III.PROPOSED ALGORITHM

Proposed processing sequence is as follows:

1. The detection process starts with the selection of the profile that needs to be tested.
2. After the selection of the profile, the suitable attributes (i.e. features) is selected on which the classification algorithm is implemented.
3. The attributes extracted is passed to the trained classifier. The classifier gets trained regularly as new training data is feed into the classifier.
4. The classifier determines whether the profile is spam or real.
5. The classifier may not be 100% accurate in classifying the profile so; the feedback of the result is given back to the classifier. For example, if the profile is identified as spam, social networking site can send a notification to the profile to submit identification. If the valid identification is given, feedback is sent to the classifier that the profile was not fake.
6. This process repeats and as the time proceeds, the no. of training data increases and the classifier becomes more and more accurate in predicting the fake profiles.

3.2 Proposed Method

A) Feature extraction:

We extracted many features from Twitter like TextData, TweetCreatedAt, RetweetCount, TweetFavouriteCount, TweetSource, UserID, UserScreenName, UserName, UserCreatedAt, UserDescription, UserDescriptionLength, UserFollowersCount, UserFriendsCount, UserLocation, HttpCount, HashtagCount, MentionCount, and TweetCount. That features wecollected from Twitter and a few more we will generate and add by using these features.

B) Cross k-Validation for splitting training and test data:

Cross-validation is a technique used to evaluate predictive models. In this technique, the original samples are divided into two categories: training set for model training and test set for evaluation. The original sample is randomly divided into k subsamples with equal size. One of these subsamples is considered as evaluative data in order to test the model, and the rest of them, k-1 subsamples, are considered as training data. The cross-validation process is repeated k times for k subsamples, each time for one of them as evaluative data. The first advantage of this method is that all samples are used for both training and validation process, and the second one is that each sample is used for validation just once.

C) Classification:

Proposed classifier is Random forests, it is a supervised learning algorithm. A forest is comprised of trees. It is said that the more trees it has, the more robust a forest is. Random forests create decision trees on randomly selected data samples, get a prediction from each tree and select the best solution by means of voting. It also provides a pretty good indicator of the feature importance.

It works in four steps:

1. Select random samples from a given dataset.
2. Construct a decision tree for each sample and get a prediction result from each decision tree.
3. Perform a vote for each predicted result.
4. Select the prediction result with the most votes as the final prediction.

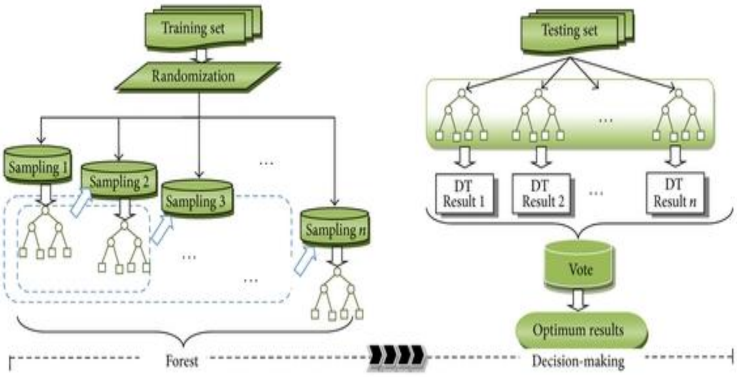


Figure 3.1: Proposed classifier.

3.3 Flow-Chart of Proposed Method

Here T: number of features,
D: number of trees to be constructed,
V: output (class with high votes)

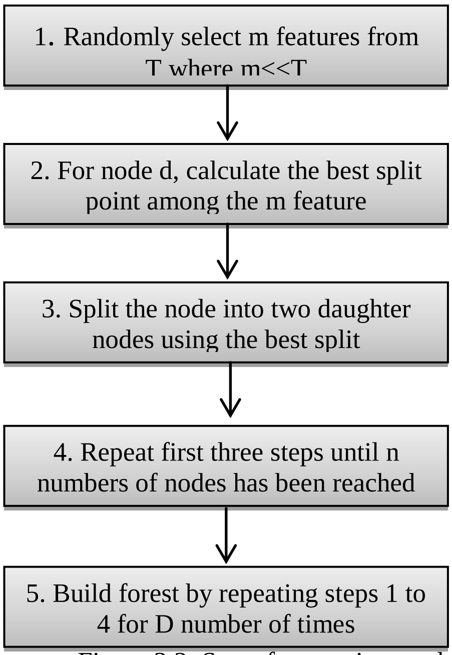

1. Randomly select m features from T where m<<T
2. For node d, calculate the best split point among the m feature
3. Split the node into two daughter nodes using the best split
4. Repeat first three steps until n numbers of nodes has been reached
5. Build forest by repeating steps 1 to 4 for D number of times

Figure 3.2: Steps for creating random forest.

IV. Simulation Results

The evaluation is based on a confusion matrix associated metrics variables TP, FP, TN, and FN in the confusion matrix refer to the following:

- True positive (TP): number of fake nodes that are identified as fake nodes.
- False positive (FP): number of normal nodes that are identified as fake nodes.
- True negative (TN): number of normal nodes that are identified as normal nodes.
- False negative (FN): number of fake nodes that are identified as normal nodes.

To evaluate the classifier, accuracy, and area under the curve (AUC) are used. The AUC is performance metrics for binary classifiers; the closer this AUC is to one, the more favorable the final performance of the classification will be. By comparing the ROC curves with the AUC, it captures the extent to which the curve is up in the northwest corner. The metrics which are introduced below are used to calculate the ROC.

- True negative rate (TNR) =TN/ (TN+FP).
- False positive rate (FPR) =FP/ (FP+TN).
- True positive rate (TPR) = TP/ (TP+FN).
- False negative rate (FNR) =FN/ (FN+TP)

There is another measure which is used to evaluate the performance:

Accuracy= (TP+TN)/ (TP+FP+TN+FN).

Method	Training (%)	Accuracy
Naive Bayes	75.00	
KNN	90.10	
Decision Tree	90.60	
Proposed	94.34	

Table 5.1: Accuracy Comparison for existing and Proposed Classifiers.

It is observed that proposed classifier achieved best accuracy for detection of fake profiles from datasets.

V. Conclusion and Future Work

We have given a framework which collects data from Twitter and facebook using API and from every tweet; we extract features that we need to feed our classifiers. We develop our proposed method that is based on binary classification through the Random Forest algorithm. Results shows that it is more efficient than through any other classifier. Using our proposed method, we have achieved the highest efficiency in terms of detection of fake profiles based on user content features. Future work will investigate the enrichment of the feature set used in the research for this thesis by engineering features from the social sciences knowledge domain -especially psychology. The aim will be to enrich the corpus with new features engineered from the same attributes, as used in this study, found on SMPs.

References

[1] R. Kaur and S. Singh, “A survey of data mining and social network analysis based anomaly detection techniques,” Egyptian journal, vol. 17, no. 2, pp. 199–216, 2016.

[2] L. M. Potgieter and R. Naidoo, “Factors explaining user loyalty in a social media-based community,” African Journal of Information Management, vol. 19, no. 1, pp. 1–9, 2017.

[3] Y. Boshmaf, D. Logothetis, G. Siganos, J. Lerıa, J. Lorenzo, M. Ripeanu, K. Beznosov, and H. Halawa, “´Integro: Leveraging victim prediction for robust fake account detection in large scale osns,” Computers & Security, vol. 61, pp. 142–168, 2016.

[4] (2012) Cbc.facebook shares drop on news of fake accounts. Internet draft. [Online]. Available: http://www.cbc.ca/news/technology/facebook-shares-drop-onnews- of-fake-accounts-1.1177067

[5] (2018) Facebook publishes enforcement numbers for the first time. Internet draft. [Online]. Available: https://newsroom.fb.com/news/2018/05/enforcement-numbers.

[6] (2018) How concerned are you that there are fake accounts and bots on social media platforms that are used to try to sell you things or influence you? Internet draft. [Online]. Available: https://www.statista.com/statistics/881017/fakesocial- media-accounts-bots-influencing-selling-purchases-usa.

[7] (2012) Buying their way to twitter fame. Internet draft. [Online]. Available: www.nytimes.com/2012/08/23/fashion/twitterfollowers- for-sale.html?smid=pl-share

[8] (2017) Welcome to the era of the bot as political bogeyman. Internet draft. [Online]. Available: https://www.washingtonpost.com/news/politics/wp/2017/06/12/welcome to the-era-of-the-bot-as-political-bogeyman.

[9] (2018) Human or 'bot'? Doubts over Italian comic beppe grillo's twitter followers. Internet draft. Available: https://www.telegraph.co.uk/technology/twitter/9421072/Human-or bot- Doubts-over-Italian-comic-Beppe-Grillos-Twitter-followers.html.

[10] S.-T. Sun, Y. Boshmaf, K. Hawkey, and K. Beznosov, "A billion keys, but few locks: the crisis of web single sign-on," in Proceedings of the 2010 New Security Paradigms Workshop. ACM, 2010, pp. 61–72.

[11] S. Fong, Y. Zhuang, and J. He, "Not every friend on a social network can be trusted: Classifying imposters using decision trees," in Future Generation Communication Technology (FGCT), 2012 International Conference on. IEEE, 2012, pp. 58–63.

[12] Y. Boshmaf, I. Muslukhov, K. Beznosov, and M. Ripeanu, "The social bot network: when bots socialize for fame and money," in Proceedings of the 27th annual computer security applications conference. ACM, 2011.

[13] P. Patel, K. Kannoorpatti, B. Shanmugam, S. Azam, and K. C. Yeo, "A theoretical review of social media usage by cyber-criminals," in Computer Communication and Informatics (ICCCI), 2017 International Conference on. IEEE, 2017, pp. 1–6.

[14] K. Thomas, C. Grier, D. Song, and V. Paxson, "Suspended accounts in retrospect: an analysis of twitter spam," in Proceedings of the 2011 ACM SIGCOMM conference on Internet measurement conference. ACM, 2011.

[15] J. Ratkiewicz, M. Conover, M. Meiss, B. Gonc¸alves, S. Patil, A. Flammini, and F. Menczer, "Truthy: mapping the spread of astro turf in microblogs streams," in Proceedings of the 20th international conference companion on World Wide Web. ACM, 2011, pp. 249–252.

[16] Facebook: Statement of rights and responsibilities https://www.facebook.com/terms, January 30, 2015, doi: June 12, 2016.

[17] Conti, M., Poovendran, R., & Secchiero, M. (2012, August). Facebook: Detecting fake profiles in on-line social networks. In Proceedings of the 2012 International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2012) (pp. 1071-1078). IEEE Computer Society.

[18] Twitter: The Twitter Rules https://twitter.com/rules, doi: June 12, 2016.

[19] Facebook "Security Tips" https://www.facebook.com/help/379220725465972 [Accessed: 10- August-2016].

[20] Davis, C. A., Varol, O., Ferrara, E., Flammini, A., & Menczer, F. (2016). Botornot: A system to evaluate social bots. Proceedings of the 25th international conference companion on World Wide Web. International World Wide Web Conferences Steering Committee273 274.

[21] Varol, O., Ferrara, E., Davis, C. A., Menczer, F., & Flammini, A. (2017). Online human-bot interactions: Detection, estimation, and characterization. Eleventh international AAAI conference on web and social media280289.

[22] Alarifi, A., Alsaleh, M., & Al-Salman, A. (2016). Twitter turing test: Identifying social machines. Information Sciences, 372, 332–346.

[23] Kantepe, M., & Ganiz, M. C. (2017). Preprocessing framework for Twitter bot detection. 2017 International conference on computer science and engineering (ubmk). IEEE630–634.

[24] David, I., Siordia, O. S., & Moctezuma, D. (2016). Features combination for the detection of malicious Twitter accounts. 2016 IEEE international autumn meeting on power, electronics and computing (ROPEC). IEEE1–6.

[25] Khaled, S., El-Tazi, N., & Mokhtar, H. M. (2018). Detecting fake accounts on social media. 2018 IEEE international conference on big data (big data). IEEE3672–3681.

[26] Velayutham, T., & Tiwari, P. K. (2017). Bot identification: Helping analysts for right data in Twitter. 2017 3rd international conference on advances in computing, communication & automation (ICACCA) (fall). IEEE1–5.

[27] Ji, Y., He, Y., Jiang, X., Cao, J., & Li, Q. (2016). Combating the evasion mechanisms of social bots. Computers & Security, 58, 230–249.

[28] Gilani, Z., Kochmar, E., & Crowcroft, J. (2017). Classification of Twitter accounts into automated agents and human users. Proceedings of the 2017 IEEE/ACM international conference on advances in social networks analysis and mining 2017. ACM489–496.

[29] Cresci, S., Lillo, F., Regoli, D., Tardelli, S., & Tesconi, M. (2019). Cashtag piggybacking: Uncovering spam and bot activity in stock microblogs on Twitter. *ACM* Transactions on the Web (TWEB), 13(2), 11.

[30] Minnich, A., Chavoshi, N., Koutra, D., & Mueen, A. (2017). Botwalk: Efficient adaptive exploration of Twitter bot networks. Proceedings of the 2017 IEEE/ACM international conference on advances in social networks analysis and mining 2017. ACM467–474.

[31] Chen, Z., & Subramanian, D. (2018). An unsupervised approach to detect spam campaigns that use botnets on Twitter. ArXiv: 1804.05232.

[32] Abu-El-Rub, N., & Mueen, A. (2019). Botcamp: Bot-driven interactions in social campaigns. The World Wide Web conference. ACM2529–2535.

[33] Beskow, D. M., & Carley, K. M. (2019). It's all in a name: detecting and labeling bots by their name. Computational and Mathematical Organization Theory, 25(1), 24–35.

[34] D. Ramalingam, V. Chinnaiah, Fake profile detection techniques in large scale online social networks: A comprehensive review, Computers and Electrical Engineering (2017), http://dx.doi.org/10.1016/j.compeleceng.2017.05.020.

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

☎ 9940 572 462 ☎ 6381 907 438 ✉ ijircce@gmail.com

www.ijircce.com