Survey on Fake Profile Detection in Social Media

Jyothis Joseph, Assistant Professor
Computer Science and Engineering
College of Engineering Kidangoor
Kottayam, India
jyothis@ce-kgr.org

Angeetha Raju
Computer Science and Engineering
College of Engineering Kidangoor
Kottayam, India
angeetharaju27@gmail.com

Aparna Santhosh
Computer Science and Engineering
College of Engineering Kidangoor
Kottayam, India
aparnasanthoshpty@gmail.com

Ashitha Jenish
Computer Science and Engineering
College of Engineering Kidanggor
Kottayam, India
ashithajenish23@gmail.com

Minu KS
Computer Science and Engineering
College of Engineering Kidangoor
Kottayam, India
minuks1999@gmail.com

***Abstract*— In the present generation, online social networks (OSNs) have become increasingly popular, people's social lives ave become more associated with these sites. They use OSNs to keep in touch with each other, share news, organize events, and even run their own e-business. The rapid growth of OSNs and the massive amount of personal data of its users have attracted attackers, and imposters to steal personal data, share false news, and spread malicious activities. On the other hand researchers have started to investigate efficient technique to detect abnormal activities and fake accounts using machine learning algorithms. The various machine learning models widely used for fake profile detection are Support Vector Machine, Decision Tree, Neural Networks, Random Forest, Naive Bayes and K-nearest Neighbor.**

Keywords—Accounts, Fake profiles, Accuracy, machine learning algorithms.

I. INTRODUCTION

Social Media Platforms (SMPs) such as Twitter, Facebook, LinkedIn, Reddit etc. provide space for people around the globe to share their personal or career interest. They provide a place to share their ideas, interests, photos, and videos with other people. A recent study done by USC and Indiana University found that between 29M and 48M accounts on Twitter are fake (or bots). Detecting these fake identities, thus becomes important to protect genuine users from malicious intents. Various machine learning models have been proposed to detect bot accounts. This paper is a survey on fake profile detection methods using machine learning technologies.

II. SYSTEM ARCHITECTURE

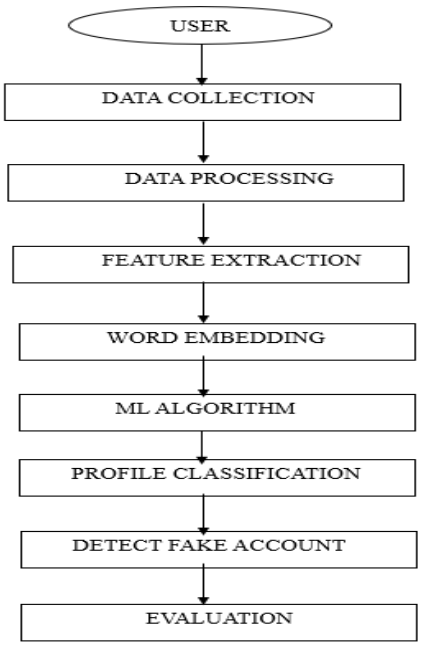

Fig.1. System architecture

DOI: 10.5281/zenodo.8013688

The user is logged in to the web interface, some data's had been already collected which includes users behavior, how the user manage these accounts etc. Then choose an account which is to be tested. By using machine learning algorithms it will classify whether an account is fake or not. It needs some data for the classification with the help of data pre-processing it takes the raw data into a text format which can be understood and analyze by machine learning algorithms and it eliminates redundancy. The next Feature Extraction, it is a process which decided which variables are most important to analysis. Word embedding is a method of extracting feature set of text. It will convert text data in to vector form. After all these procedures the profile is classified by using suitable machine learning algorithms.

III. RELATED WORKS

By using selected features Jyoti Kaubiyal et.al [1] aimed to use a feature-based approach to identify fake profiles. They collected real data from different profiles using the twitter API and they used 24 features from the profiles based on account, tweet, ownership URL etc. Then they preprocessed the data and three machine learning algorithms namely Logistic regression, SVM and Random Forest were applied to the same data for the classification of accounts. They used publicly available datasets for the development of classification model. The dataset was divided into 80:20 ratios for training and testing purpose. Also they created a table showing that their model is able to detect account as 'bot' and 'human' efficiently. Logistic regression and Random Forest were able to detect the fake accounts more efficiently with 95.3% and 97.9% accuracy respectively whereas SVM was having 80.8% accuracy. They showed the comparative results between the three different models. A curve is plotted between the True Positive rate and False Positive to see the ability of the model to classify the classes. Among all the three models Random Forest performed best.

Shivangi Gheewala et.al [2] aimed to develop a model that analyze, detects and recovers from defamatory actions in twitter. Machine learning gives any application model which has the capability to learn and make prediction, as they can carried out classification, clustering, regression, visualization, data processing and feature selection tasks. Spam Detection framework is a binary classification problem. This paper is structured which gives an overview of previous related work on spam. The continuous increase in the volume of social network has contributed to the growth in spamming accounts. To reduce the effect of spam accounts on real user, spammer evasion tactics are to be analyzed and effort is required to develop a productive system that detects spam accounts and prevent real users from getting attacked by spammers. A profound survey has been carried out in this paper. Thus, nowadays research has started to work on solving the above-mentioned issues in order to achieve satisfactory results.

"Spam profile detection in social networks based on public features" by Ala M-Zoubi et.al [3] developed it based on a set of publicly available features. It is extracted from the profile information which include languages used by the users. Also, they have used four machine learning algorithms to develop the detection model and two feature selection methods. In feature Engineering they convinced the information's they gathered were able to decide which feature is suitable to classify the profile and which is not. In this work they follow 10 features for identifying the spam profiles from legitimate ones - like suspicious works which include deist check here, Health, make money, give me, Vote, Free, also other similar terms in Arabic languages also were checked. The threshold to the feature is set to 80%. Data collection is done by extracting features from 82 profiles. This process is carried out in 3 stages, first stage is to collect the authentication key from the twitter API's and then the TwitterR and R script are used to extract the exact profile data and finally they save all information in CSV file for labeling. They have marked 0 for genuine account and 1 for spam. In order to evaluate the spam detection model developed in this work, the experiment in this work is carried out in two stages. In first stage the basic classifier including decision tree,K-NN ,NB and MLP are applied. In second stage features in the data set which are already constructed and are selected the most representative features. This work was a preliminary study for future work aimed to collect much larger data set for different languages using the same methodology. Now a days detection of fake accounts are more challenging, therefor efficiency of more scalable models will be investigated and applied.

Sarah Khaled et.al [4] started to investigate an efficient technique to detect abnormal activities and fake accounts based on the account features and used to solve this problem by using supervised machine learning algorithms. In this paper they introduced a new algorithm SVM-NN that should provide an efficient detection on fake profile and fake bots .This is a combination of neural network and support vector machine. The SVM-NN algorithm used less amount of account features and the algorithm gives 98% of relevant information about the accounts. The researchers identify the fake accounts through the followers list and profile details. Then the information are applied in some techniques for account classification. The dataset from fake followers collected from fastfollowers.com and inerttwitter.com. Using this dataset built 3 dataset such as fast followers' dataset, inert twitter dataset and technology.

Kumud Patel et.al [5] had reviewed on Fake Profile Detection on social site by using Machine Learning . Here the authors focused on different machine learning algorithms to identify the spam accounts. The first process is to select the target profile for the account details such as follower, likes, friend list, comments and so on. The information was applied

into some machine learning classifiers and identify the fake accounts. Most of the machine learning algorithm give an 50%-96% of accuracy. Find the account genuine by analyzing the followers count. In psychology studies ,many users are lied on their name ,age ,location and other details. Check the email ids that are linked in social media ,that is the another way to identify the fake profiles and analyze the location of users .

Preethi Harris et.al[6]experimented classification algorithms to train the dataset and compare each of them. The dataset for detection of fake profiles in Instagram, a Kaggle dataset with has been used is divided into 80% training data and 20% testing data. Classification algorithm are used to determine the efficiency of algorithm. It enables to learn assignment of class labels to problem domain. Here the proposed work is to classify the data as fake or genuine Instagram profile. This paper conclude that the Machine Learning algorithm can easily detect fake profile from social networking sites. The algorithms such as SVM, KNN, Random Forest along with Naive Bayes and XGBoost are used to classify the Instagram profiles. The results of classification based on profile IDs are written in a data dictionary to identify the fraudulent IDs for action by the concerned authorities.

Spammer detection in online social network is challenging and high task. This section evaluates the performance of some ensemble learning approaches for task of spam detection by Sajid Yosuf Bhat et.al[7]. The evaluation results state that bagging ensemble learning approach using J48 which means decision tree, base classifier performs better than its model. Mostly used three ensemble methods are bagging, boosting, stacking dataset with artificially planned spammers. They also compare the performance of multiple classifier including decision tree ,Navie Bayes, K-NN and their ensemble variants implemented in WEKA. They have used real world datasets. Also they look forward the accuracy of two classifiers which where 0.963.How ever in the case of naive bayes classifier the ensemble approach showed low performance compared to their individual performance for spammer detection task using structural features. When it came to three classifier J48, IBk and naive bayes the performance is lower than the best case of the other two ensemble. They have concluded that the observed bagging ensemble learning approach using J48 classifier performs better than individual performance of IBk and naive bayes classifier also better than ensemble approaches ,boosting and stacking for the task of spammer detection in social network.

Yasyn Elyusufi et.al.[8] described social networks fake profiles detection based on account setting and activity in 2019.In this paper they assessed the impact of using Decision Tree (DT), and Naıve Bayes (NB) to classify the user profiles into fake and genuine. This paper consists of three sections. They have assessed the impact of using Naive Bayes classifiers and Decision Trees classifiers in the prediction of fake or genuine profiles in social networks. In feature selection phase initially 33 profile features are used. Then they decided to use only features which will affect directly the results. The features on the final dataset were: statuses count, followers count, friends count, favorites count. Before the training phase ,they split the dataset. In this work the training test set is defined with 80% while the test set is defined with 20%. At last they compared the results of two machine learning algorithms (Decision Tree and Naive Bayes) to determine the most appropriate approach to differentiate the legitimate profiles from fake profiles in Facebook dataset. In addition, they showed the accuracy calculation for each algorithm based on the result tests of the two algorithms.

IV. CONCLUSION

Fake profiles have become a major concern in social networking sites and are difficult to detect. The main objective of our model is to detect fake profiles using selected features efficiently. In the classification phase different machine learning algorithms were used. This paper provides a survey of different machine learning algorithms to detect fake accounts.

References

[1] Jyoti Kaubiyal, and Ankit Kumar Jain. "A feature based approach to detect fake profiles in Twitter." In Proceedings of the 3rd international conference on big data and internet of things, pp. 135-139. 2019.

[2] Shivangi Gheewala, and Rakesh Patel. "Machine learning based Twitter Spam account detection: a review." In 2018 Second International Conference on Computing Methodologies and Communication (ICCMC), pp. 79-84. IEEE, 2018.

[3] Ala'M, Al-Zoubi, Ja'far Alqatawna, and Hossam Paris. "Spam profile detection in social networks based on public features." In 2017 8th International Conference on information and Communication Systems (ICICS), pp. 130-135. IEEE, 2017.

[4] Sarah Khaled, Neamat El-Tazi, and Hoda MO Mokhtar. "Detecting fake accounts on social media." In 2018 IEEE international conference on big data (big data), pp. 3672-3681. IEEE, 2018.

[5] Kumud Patel, Sudhanshu Agrahari, and Saijshree Srivastava. "Survey on fake profile detection on social sites by using machine learning algorithm." In 2020 8th international conference on reliability, infocom technologies and optimization (trends and future directions)(ICRITO), pp. 1236-1240. IEEE, 2020.

[6] Preethi Harris, J. Gojal, R. Chitra, and S. Anithra. "Fake Instagram Profile Identification and Classification using Machine Learning." In 2021 2nd Global Conference for Advancement in Technology (GCAT), pp. 1-5. IEEE, 2021.

[7] Sajid Yousuf Bhat, Muhammad Abulaish, and Abdulrahman A. Mirza. "Spammer classification using ensemble methods over structural social network features." In 2014 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT), vol. 2, pp. 454-458. IEEE, 2014.

[8] Yasyn Elyusufi, Zakaria Elyusufi, and M'hamed Ait Kbir. "Social networks fake profiles detection based on account setting and activity." In Proceedings of the 4th International Conference on Smart City Applications, pp. 1-5. 2019.