International Journal of Psychosocial Rehabilitation, Vol. 24, Issue 06, 2020
ISSN: 1475-7192

PREDICTION OF FAKE PROFILES USING CLASSIFICATION ALGORITHM

[1]Bhavini Menariya, [2]Mrs. R Vidhya

ABSTRACT*— Online Social Networks are progressively affecting the manner in which individuals speak with one another and share data. The expanding protection dangers in social networking sites is drawing in security scientists attempting to recognize and relieve dangers to singular clients. With numerous online social networking sites having tens or many million clients all things considered creating billions of individual information content that can be abused, identifying and forestalling assaults on singular client security is a significant test. The greater part of the flow explore has concentrated on securing the protection of a current online profile in a given online social networking site.The proposed system aims to find the fake profiles using machine learning classification algorithm such as Random Forest that is highly employed. Algorithm has been developed in order to detect the fake profiles based on the dataset. Here the fake profiles are identified before anyone suffer by the suspicious account.*

Keywords*--Online Social Networks, Classification algorithm, Random Forest*

I. INTRODUCTION

In the present period, online social networks are the most famous and fast data spread applications on the Internet. Individuals of any age invest the greater part of their energy on social networking sites. Removing fake accounts from online social networks is a difficult task and researches have been done to detect fake profiles in different social networking sites. The proposed work aims to develop a framework based upon machine learning techniques that may assist with finding the fake profiles in social networking sites. Different social networking sites are improving our public activities however but there are a ton of issues with utilizing these social networking sites. Some of these issues are security, web-based harassing, potential for abuse, trolling, and so on. These are done mostly by using fake profiles. In this project, the main objective is to detect fake profiles using machine learning algorithms from the given dataset.

II. RELATED WORK

In this paper [1] unauthorized users abuse the authorized users and the algorithm used here is K means algorithm. It involves User based opinion datasets and handles large data.

[1]*Department of Computer Science Engineering,SRM Institute of Science and Technology, Kattankulathur Chennai, Tamil Nadu,India. bm8942@srmist.edu.in*

[2]*Assistant Professor ,Department of Computer Science Engineering, SRM Institute of Science and Technology,Kattankulathur, Chennai,Tamil Nadu, India, vidhyar@srmist.edu.in*

Received: 22 Feb 2020 | Revised: 13 Mar 2020 | Accepted: 05 Apr 2020

This paper[2] involves Naïve bayes algorithm and graph dataset. User behaviour can be predicted and performance measure used here is facebook fake user ground truth.

In this paper[3] automatic detection methods regularly need versatility and capacity to deduce forgery. It involves K-means algorithm and multimedia datasets. Dynamicity of trust and authenticity are used as performance measure.

In this research [4] fake or unauthorized users abuse the authorized user's details.The algorithm used here is K-nearest neighbor and the dataset is collected from the user. Real time implementation is used as performance measure.

This paper [5] tells that there are many fake accounts in facebook that people are using everyday.The algorithm used here is K-means clustering and user based opinion datasets are used.Similarity is used as performance measure.

In this paper[6] the nature of news is viewed as lower than customary news outlets bringing about a lot of fake news.The technique used here is node matching and it involves construction of two datasets with news data and social setting information.Trust analysis is used as performance measure.

This paper [7] involves the quantity of people groups via web-based networking media stages are increasing at a more prominent level for mailicious webpage. The algorithm utilized here is Naïve Bayes and dataset is taken from the user. Eliminate fake accounts made and cyborgs can't be utilized for separating fake accountsUsual practices, for example, spam which are found in sends and online networking stage is utilized as execution measure.

In this paper [8] it is noticed that there is a danger if you don't have an account in extravagant informal community. Decision tree algorithm is used and a facebook detecting application is made which can gather the necessary factual data from a profile. Execution measure is considered by discovering initial ones to dissect informal community diagrams from a unique perspective inside the setting of protection dangers.

This paper [9] includes interpersonal organization which is where social exercises, business arranged exercises, diversion and data are exchanged. Random forest algorithm and twitter dataset is used. Classifying each example into spam or non-spam class is utilized as execution measure.

This paper [10] defines the triats that give the probability of a client having a dangerous profile. The algorithm used here is K means clustering and user item rating matrix is used as dataset. Similarity among users and abnormal rating behavior are used as performance measure.

This paper [11] involves a model that incorporates attributes that involve sentiments to separate genuine and fake profiles. The technique involved here is logistic regression and emotions dataset is used. Based on user emotions we can find whether the profile is fake or not. Emotions are measured and used as performance measure.

This paper [12] includes profile cloning which is the data fraud of existing client's profile certifications and makes a fake profile utilizing a few qualifications. The algorithm used here is Decision tree and facebook dataset are used.

This paper [13] includes perceiving the personalities of fake profiles which is one of the basic security issues in social networking platform. The algorithm used here is K nearest neighbor and user behavior dataset is used.

This paper [14] involves ensuring communication between users or groups or organizations.

III. MODULE DESCRIPTION

The automatic detection of fake profiles in general involves the following steps:

1. Data Collection
2. Data Pre-processing
   - Data cleaning
   - Data transformation
   - Data selection
3. Train and Test Data
4. Prediction and Result

3.1 DATA COLLECTION

Data Collection is a significant task in building any model. It involves gathering of task related information based on some targeted variables to analyse and produce some valuable outcome. The input dataset involves two datasets named as normal users and fake users. Both the datasets are combined here to detect the number of fake users. To get quality outcome which will be helpful for information generation and decision making the raw data should be pre-processed.

3.2 DATA PRE-PROCESSING

The procedure which includes transforming raw data into an understandable form for further processing is called data pre-processing. In real world the data are most often incomplete, uncertain, missing, and inconsistent and contains many errors. Hence, it is must to process the data before analyzing it and coming to the results. Data pre-processing can be done by data cleaning, data transformation and data selection.

DATA CLEANING

Data cleaning involves fill in missing values, preparing data for analysis, removing or modifying data that is irrelevant.

DATA TRANSFORMATION

Data transformation may include smoothing, aggregation, generalization, transformation which improves the quality of the data.

DATA SELECTION

Data selection includes some methods or functions which allow us to select the useful data for our system.

3.3 TRAIN AND TEST DATA

Separating the input dataset into training set and testing set is a significant piece of assessing models.Generally, the greater part of the dataset is utilized as training dataset and the little segment of the dataset is utilized as testing dataset.Now when algorithm is trained using the training set it is easy to predict the result for the test set.

3.4 PREDICTION AND RESULT

After the algorithm has gained some information about the dataset using the training set we can predict the result for the testing set. After the training set is used to train the algorithm we can predict the result using the remaining testing set. The result is displayed in the form of confusion matrix which shows the number of fake users from the dataset. Confusion matrix is used here to measure the performance of a classification problem .The algorithm also shows the accuracy rate of prediction.

IV. LEARNING ALGORITHM

In this proposed system we are using machine learning algorithm named as Random Forest Algorithm.

RANDOM FOREST ALGORITHM

It is used here because it gives better accuracy when compared to other algorithms. It can easily overcome the problem of overfitting by combining the results. The main advantage of Random Forest Algorithm is that scaling of data is not required. But more computation resources are required to implement the algorithm. When a large part of data is missing then also it maintains good accuracy. The accuracy of Random Forest Algorithm to identify the number of fake profiles is shown as result in the form of Receiver Operating Characteristic (ROC) curve. Accuracy is measured by the area under the ROC curve and is shown below.

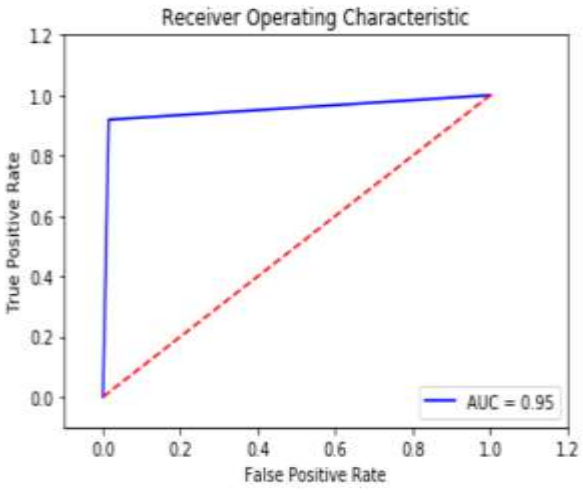

FIGURE 1: THE ROC CURVE AND IS SHOWN BELOW

V. CONCLUSION

Different social networking sites are improving our public activities however but there are a ton of issues with utilizing these social networking sites. Some of these issues are security, web-based harassing, potential for abuse, trolling, and so on. These are done mostly by using fake profiles. Hence the framework is designed to find fake profiles from the dataset. In this paper, we have used machine learning algorithm named as Random Forest Algorithm using which we can detect the number of fake users and display the result in the form of confusion matrix. This algorithm has improved the prediction accuracy rate in this paper.

REFERENCES

1. Vijay Tiwari, "Analysis and detection of fake profile over social network",2017 International Conference on Computing, Communication and Automation (ICCCA) ,Greater Noida,India,5-6 May 2017
2. Aditi Gupta and Rishabh Kaushal, "Towards detecting fake user accounts in facebook" 2017 ISEA Asia Security and Privacy (ISEASP), Surat,India, 29 Jan.-1 Feb. 2017
3. Khaled A.N. Rashed ,Dominik Renzel , Ralf Klamma ,"Trust-aware media quality profiles in fake multimedia detection" Workshop on Multimedia on the Web, RWTH Aachen University,Ahornstr,55,D-52056 Aachen,Germany,2011
4. S.Revathi and Dr.M.Suriakala," Profile Similarity Communication Matching Approaches for Detection of Duplicate Profiles in Online Social Network",2018 3rd International Conference on Computational Systems and Information Technology for Sustainable Solutions (CSITSS), Bengaluru,India,20-22 Dec. 2018
5. P.V. Savyan and S.Mary Saira Bhanu, "Behaviour Profiling of Reactions in Facebook Posts for Anomaly Detection", 2017 Ninth International Conference on Advanced Computing (ICoAC),Chennai,India,14-16 Dec. 2017
6. Kai Shu, Suhang Wang, Huan Liu, "Understanding User Profiles on Social Media for Fake News Detection", 2018 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR),Miami,FL,USA,10-12 April 2018
7. Naman Singh,Tushar Sharma,Abha Thakral,Tanupriya Choudhury, "Detection of Fake Profile in Online Social Networks Using Machine Learning",2018 International Conference on Advances in Computing and Communication Engineering (ICACCE), Paris,France,22-23 June 2018
8. Mauro Conti,Radha Poovendran,Marco Secchiero, "FakeBook: Detecting Fake Profiles in On-Line Social Networks", 2012 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, Istanbul,Turkey,26-29 Aug. 2012
9. Shivangi Gheewala,Rakesh Patel,"Machine Learning Based Twitter Spam Account Detection:A Review",2018 Second International Conference on Computing Methodologies and Communication (ICCMC),February 2018

Received: 22 Feb 2020 | Revised: 13 Mar 2020 | Accepted: 05 Apr 2020

10. Anahita Davoudi and Mainak Chatterjee, "Detection of profile injection attacks in social recommender systems using outlier analysis", 2017 IEEE International Conference on Big Data (Big Data),Boston,MA,USA,11-14 Dec. 2017
11. Mudasir Ahmad wani,Nancy Agarwal,Suraiya Jabin,Syed Zeeshan Hussain,"Analyzing Real and Fake users in Facebook Network based on Emotions",2019 11th International Conference on Communication Systems & Networks (COMSNETS),Bengaluru,India,7-11 Jan. 2019
12. M.A. Devamane and N.K. Rana,"Detection and prevention of Profile Cloning in Online Social Networks", International Conference on Recent Advances and Innovations in Engineering (ICRAIE-2014),Jaipur,India,9-11 May 2014
13. Mohamed Torky,Ali Meligy,Hani Ibrahim,"Recognizing Fake identities in Online Social Networks based on a Finite Automaton approach",2016 12th International Computer Engineering Conference (ICENCO),Cairo,Egypt,28-29 Dec. 2016
14. Neha M.Yadav and P.N.Chatur,"Compromised Account Detection and Prevention by Profiling Social Behavior and FASS Key Concept", 2017 International Conference on Recent Trends in Electrical, Electronics and Computing Technologies (ICRTEECT),Warangal,India,30-31 July 2017