IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)
An International Open Access, Peer-reviewed, Refereed Journal

Detecting Fake Accounts on Social Media using Neural Network

Gokul Jadhav[1], Krupa Patel[2], Mr.Ranjit Gawande[3]
[1] *Department of Computer Engineering, Matoshri College of Engineering & R. C. Nashik, India.*
[2] *Programmer Analyst in Cognizant Technology Solutions India Private Limited, India*
[3] *Assistant Professor, Department of Computer Engineering, Matoshri College of Engineering & R. C. Nashik Maharashtra India.*
Savitribai Phule Pune University, Pune, India

Abstract—Social networking sites such as Facebook, Twitter, histogram, etc. are extremely famous among people. Users always interact with their friends via these social networks sites or media. They share their personal and public information using these social networks. an immense number of people use social networking sites due to their attractiveness. This fame causes problems to the websites due to the creation of fake accounts. The owners of fake accounts pull out personal information about other people and spread the fake data on social networks. In our proposed plan, we propose machine learning techniques such as Neural Networks and SVM for detecting fake accounts on Facebook or Twitter, or Twitter. Different data mining tools have been used for the simulation of the algorithm and the obtained results are presented by the proposed plan. Data mining tool which allows quick user interaction with a simple tool for the identification of fake accounts from available data. In this, we classify the data using the above machine learning techniques, which identify the fake accounts on the social sites.

***Keywords*:-** *Facebook, Fake Accounts, Feature Selection, Clustering, Classification*

I. INTRODUCTION

Social networking sites have been extensively used as a medium of communication between people in day-to-day life. Users using these sites always share their information and daily activities which attract several people to these sites. The increasing popularity of Facebook or Twitter or Twitter from the year 2006 to 2016 allows the users to add friends and share various kinds of information such as personal, social, economic, educational, political, business, etc. Moreover, they can also share photos, videos, and other day-to-day interaction. However, some people don't use these sites with a good objective. Therefore they create fake accounts on social sites. Fake accounts do not have any real identity so we can call them Attacker. This attacker uses incorrect information or statistics about some real-world person to create a fake account. Using these fake accounts, attackers spread fake information which affects other users. To protect such sensitive data of users is one of the major challenges of social sites.

There are several techniques in the field of machine learning that have been developed to detect fake accounts in social networking sites such as Neural Network (NN), Naive Bayes, Markov Model, and Bayesian Network. In recent researches, it has been found that these techniques make available enhanced results to detect fake accounts.

Neural Network consists of many interconnected processing elements. It takes decisions just like a human brain. Support vector machines (SVM) are supervised machine learning techniques used for classification. It finds the hyperplane to classify the data. Neural networks and SVM can accept a large amount of random data and are suitable to detect fake accounts on social networking sites based on various characteristics of accounts. The naive Bayes classifier is based on Bayes' theorem. It predicts the probability that a given variable belongs to a particular class.

II. PROBLEM STATEMENT

To detect and identify fake accounts from social media websites using machine learning algorithms.

III. LITERATURE SURVEY

A. M. Egele et al. (2015)

This paper presents a system named COMPA to detect compromised accounts on social networks. This system is based on the behaviour of users on social networks. The behaviours of normal users is stable and COMPA detects compromised accounts having behaviour more inconsistent. In COMPA behavioural profile is generated based on the previous message sent by the account.

B. D. Freeman et al. (2015)

This paper focused on detecting the clusters of fake accounts rather than an individual. This approach created a cluster, based on the features provided at the registration time such as registration IP address and registration date. Random forest, SVM and Logistic regression are used to train the model, and SVM is used to classify the cluster of accounts as fake or not. Arlington, USA published a paper dedicated to Reverse Engineering Mobile applications. It uses a technique to automatically Reverse Engineer Mobile Application User Interfaces (REMAUI). On a given input REMAUI identifies user interface elements such as images, texts, containers, and lists, via computer vision and optical character recognition (OCR) techniques. In this system 448 screenshots of Android and iOS applications were used, REMAUI generated user interfaces were similar to the originals, both pixel by pixel and in terms of their runtime.

C. Krishna B Kansara et al. (2016)

This paper proposed a Sybil node discovery method based on the social graph.

This approach overcomes the limitations of the previous graph-based approaches by adding user behavioral manners such as latent dealings and friendship refusal. The proposed design is divided into two parts, Sybil node identification (SNI) and Sybil node identification using behavioral analysis (SNI-B)

D. Ali M. Meligy (2017)

This paper presents a technique to detect fake accounts on social networking sites called fake profile recognizer. This technique is based on two methods i.e regular expression and deterministic finite automata. A regular expression is used to authenticate the profiles and deterministic automata recognize the identities in a trusted manner

IV. EXISTING SYSTEM

In online social network sites to detect fake accounts is a major challenge in research work. People using online social networks always experience various issues regarding security, which affects their personal as well as social life. The number of fake accounts on a social network is increasing day by day. Fake accounts spread fake news, fake rating, and increase spam. Our proposed plan is to detect fake accounts on Facebook or Twitter or Twitter. There are various techniques are available to detect fake accounts on online social networks. Each has its advantages, disadvantages, and purposes. According to the study it is observed that not all existing methods have a very high value of f-measure and recall value.

V. PROPOSED SYSTEM

This proposed work uses the techniques like neural networks i.e. NN and a support vector machine called SVM for the classification of real and fake accounts. The feature set that influences the detection of fake accounts detection of the fake on Facebook or Twitter or Twitter will be used. This proposed work is expected to generate the higher value of f-measure and recall required for detection of fake accounts in Facebook or Twitter or Twitter. The machine learning techniques are neural network and support vector machine that provides accurate results. Neural networks and Support vector machines give better results in data classification.

VI. RESEARCH METHODOLOGY

The different machine learning algorithms can mock-up a problem with the knowledge or background for the model preparation process that helps in the selection of the best algorithm which accepts given input data to get the best result.

1. Support Vector Machine (SVM): Support vector machines (SVMs, also support vector networks) are the supervised learning models with associated learning algorithms that analyse data used for classification and regression analysis. For the given labelled training data (supervised learning), the algorithm outputs an optimal hyperplane that categorizes new examples.

2. Neural Networks sense, an artificial neural network, composed of artificial**:** A neural network is a network or circuit of neurons, or in a modern neurons or nodes. A neural network (NN), in the case of artificial neurons, is an interconnected group of natural or artificial neurons that uses a mathematical model for information

The methodology consists of the following steps.

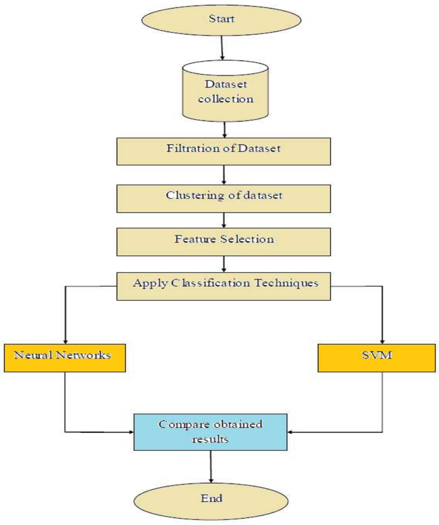

Figure 1. Methodology of Proposed Technique

1. Dataset Collection

The first step of detecting the fake account on Facebook or Twitter is collecting the data set of any of these two as per the availability. For the proposed work the data set of Facebook or Twitter is collected by survey method. By using the survey method we collect the data set of Facebook or Twitter from Facebook or Twitter users. For this purpose, we create a Google form. Google form consists of various types of questions that help us to accurately classify data-free accounts and fake accounts.

2. Filtration of Dataset

In the second step with the filter of the collected data set for filtration, we apply a randomized filtration technique. Randomization randomly changes the position of accounts in the dataset. The filtration is also used to filter the wrong values filled in the dataset and the wrong value is replaced with the average value of its upper and lower column value.

3. Clustering of Dataset

After the filtration clustering technique is applied to the data set. K-media clustering technique is applied to dataset set witch assign the data set to clusters. A cluster of fake accounts is identified by using the clustering technique. The k-Medoid clustering algorithm is better than the K-mean clustering algorithm because K-method is robust to outliers.

4. Feature Selection

In this step feature selection technique is applied to the feature set. For feature selection principal component analysis is used for feature selection. The principal component analysis also combines the related features and assigns a value to them. The feature selection technique is used to get higher accuracy with the minimum feature set. Because by using the feature selection technique we can eliminate the features having the lowest weight zero weight.

5. Classification of Dataset

The classification technique is applied after the clustering and feature selection technique. Neural networks and SVM are used for the classification of data. Neural networks and SVM are machine learning techniques that efficiently work on different kinds of data sets. The execution time of SVM is less because it takes less time to train the SVM.

6. Compare the Results

Two different results are produced by neural network and SVM in the existing technique and proposed technique. The results of both techniques are compared with existing techniques. The parameters used and the time taken by the existing technique and proposed hybrid technique are compared. The results of Neural Network and SVM in the proposed hybrid technique are also compared and the results with higher accuracy are considered.

VII. CONCLUSION

This proposed hybrid technique is used to most successful classifier neural network and SVM. K-Medoid clustering is also used to improve the accuracy and reduce the time complexity of the algorithm. In proposed work collected real-time data set of Facebook or Twitter from Facebook or Twitter users.

VII. FUTURE SCOPE

i. This technique can also be used for other social networking sites such as Facebook or Twitter and LinkedIn with minor changes.

ii. The accuracy of the proposed technique can also be improved using different feature selection techniques.

REFERENCES

[1] M. Egele, G. Stringhini, G. Stringhini, and G. Vigna, “Towards Detecting Compromised Accounts on Social Networks,” IEEE, vol. 5971, no. c, 2015.

[2] D. M. Freeman and T. Hwa, “Detecting Clusters of Fake Accounts in Online Social Networks Categories and Subject Descriptors,” AISec, 2015.

[3] K. B. Kansara, “Security against Sybil attack in social network,” ICICES, no. Icices, 2016.

[4] A. M. Meligy, “Identity Verification Mechanism for Detecting Fake Profiles in Online Social Networks,” IJCNIS, no. January, pp. 31–39, 2017.

[5] Ashraf Khalil, Hassan Hajjdiab, and Nabeel Al-Qirim , Detecting Fake Followers in Twitter: A Machine Learning Approach, International Journal of Machine Learning and Computing, Vol. 7, No. 6, December 2017

[6] S. Khaled, N. El-Tazi and H. M. O. Mokhtar, "Detecting Fake Accounts on Social Media," 2018 IEEE International Conference on Big Data (Big Data), Seattle, WA, USA, 2018, pp. 3672-3681.