International Journal of Advance Research in Science and Engineering
Volume No. 12, Issue No. 04, April 2023
www.ijarse.com
IJARSE
ISSN 2319 - 8354

DETECTING FAKE PROFILES ON SOCIAL MEDIA USINGMACHINE LEARNING

M. Mounika,L. Mahesh ,N. Siva Jyothi, M. Sai Mohan

Mr. M. Chennakesava Rao, Assoc. Professor

Department of Computer Science and Engineering,

Tirumala Engineering College, Narasaraopet, Andhra Pradesh

ABSTRACT

In the present generation, On-Line social networks (OSNs) have become increasinglypopular, which impacts people's social lives and impel them to become associated withvarious social media sites [1]. Social Networks are the essential platforms through which many activities such as promotion, communications, agenda creation, advertisements, and news creation have started to be done. Adding new friends and keeping in contact with them and their updateshas become easier. Researchers have been studying these online social networks to see theimpact they make on the people. Some malicious accounts are used for purposes suchas misinformation and agenda creation. Detection of malicious account is significant. The methods based on machine learning-based were used to detect fake accounts that could mislead people. The dataset is pre-processed using various python libraries and a comparison model is obtained to get a feasible algorithm suitable for the given dataset [2]. An attempt to detect fake accounts on the social media platforms is determined by various Machine Learning algorithms. The classification performances of the algorithms Random Forest, Neural Network and Support Vector Machines are used for the detection of fake accounts.

1. INTRODUCTION

Online Social Networks (OSNs), such as Facebook, Twitter and LinkedIn, have become increasingly popular over the last few years. People use OSNs to keep in touch with each other's, share news, organize events, and evenrun their own e-business. Facebook communitycontinues to grow with more than 2.2 billion monthly active users and 1.4 billion daily active users, with an increase of 11% on a year-overyear basis. For the purpose to detect fake accounts on the social media platforms the dataset generated was pre-processed and fake accounts were determined by machine learningalgorithms.[3] The classification performancesof the algorithms Random Forest, Neural Network and Support Vector Machines are
used for the detection of fake accounts. The accuracy rates of detecting fake accounts using the mentioned algorithms are compared and thealgorithm with the best accuracy rate is noted.

2. LITERATURE SURVEY

Sarah Khaled et al. presented a new algorithm, SVM-NN, to provide efficient detection for fake Twitter accounts and bots, feature selection and dimension reduction techniques. This proposed algorithm (SVMNN) uses less number of features, while still being able to correctly classify about 98% of the accounts ofour training

dataset [1].

Sreenivas Kuncham et al. proposed a machine learning model to predict the student placements using various Machine Learning algorithms that include J48, Naïve Bayes, Random Forest etc., The model tries to obtain the results from various algorithms and these results are compared to predict the best algorithm for any given dataset.[2]

3. METHODOLOGY

Proposed system is equipped with various Machine Learning tasks and the architecture followed is as shown below. The proposed system collects the dataset which are pre- processed by providing a framework of algorithms using which we can detect fake profiles in Facebook by comparing the accuracy of three machine learning algorithms and the algorithm with very high efficiency is found for the given dataset. Fig 1. Proposed Methodology

The different ways in which an algorithm can model a problem is based on its interaction with the experience or environment for the model preparation process that helps in choosing the most appropriate algorithm for the given input data in order 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 analyze data used for classification and regression analysis. For the given labelled training data (supervised learning), the algorithm outputs an optimal hyper plane which categorizes new examples.

2. Neural Networks:

A neural network is a network or circuit of neurons, or in a modern sense, an artificial neural network, composed of artificial 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 processing based on connectionistic approach.

3. Random Forest:

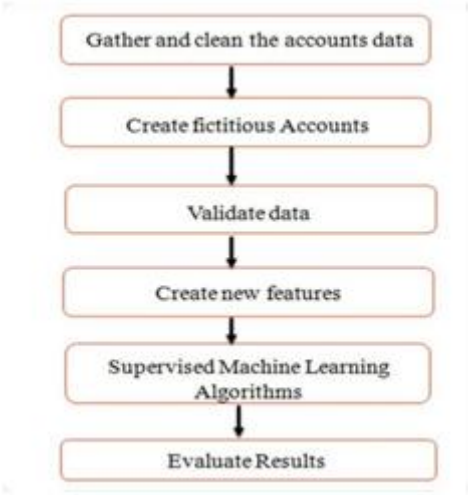
Random forest algorithm is a supervised classification algorithm. As the name suggest, this algorithm creates the forest with a number of trees. In general, the more trees in the forest the more robust the forest looks like. In the same way in the random forest classifier, the higher the number of trees in the forest gives the high accuracy results.

4. EXPLANATION OF ATTRIBUTES

Attribute importance is a supervised function that identifies and ranks the attributes that are most important in predicting a target attribute.[4] Raw machine learning data contains a mixture of attributes, some of which are relevant to making predictions.

Table 1: Attributes that define a Dataset

Attribute Name	Description
ID	The unique ID given to the account holder
NAME	The name given to the account holder
SCREEN_NAME	The pseudonym given to the account holder
CREATED_AT	The date when the account is created
FRIENDS_COUNT	The number of friends for the account
STATUSES_COUNT	The number of statuses posted from the account
FOLLOWRES_COUNT	The number of followers for the account
LISTED_COUNT	The number of groups the account belongs to
URL	The URL of the account
TIMEZONE	The time zone of the account holder
UTC_OFFSET	The UTC offset, given TIMEZONE
LOCATION	The location of the account holder
GEO_ENABLED	This field must be true for the current user to attach geographic data when using POST statuses / update.
VERIFIED	When true, indicates that the user has a verified account.



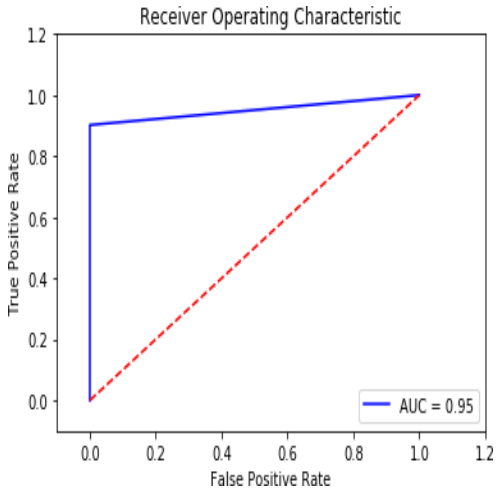
5. EXPERIMENTAL RESULTS ANDDISCUSSION

a. Performance of model usingRandom Forest Algorithm:

The random forest is a model made up of many decision trees. When trainingthe model using Random forest algorithm, each tree in a random forestlearns from a random sample of the data points and the samples drawn with replacement are known as bootstrapping in which some samples will be used multiple times in a singletree.

b. Performance of model using Support Vector Machine Algorithm:

In many supervised learning tasks, labelling instances to create a trainingset is time consuming and costly; thus, finding ways to minimize the number Fig : Accuracy using Random Algorithm



of labelled instances is beneficial. TheSupport Vector Machine algorithm is used to minimize the instances byimproving efficiency. In this algorithm, we plot each data item as apoint in n-dimensional space (where nis number of features you have) withthe value of each feature being thevalue of a particular coordinate. Wethen perform the detection of fakeaccounts through classificationtechnique by finding the hyper-planethat

differentiate the two classes verywell (look at the below snapshot).

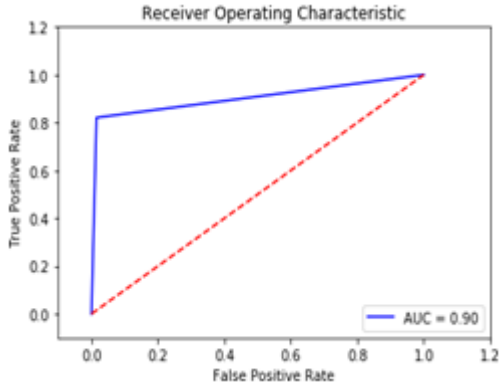


Fig : Accuracy Using Support Vector Machine

c. Performance of model using NeuralNetworks Algorithm:

Neural networks (NNs) can be defined as "The algorithms in machine learning are implemented by using the structure of neural networks. These neural networks model the data using artificial neurons. Neural networks thus mimic the functioning of the brain." The 'thinking' or processing that a brain carries out is the result of these neural networks in action.

The Neural networks algorithm tries to improve the performance of the model by using smart computational methodsto create new and better performing types of prediction and detection model.

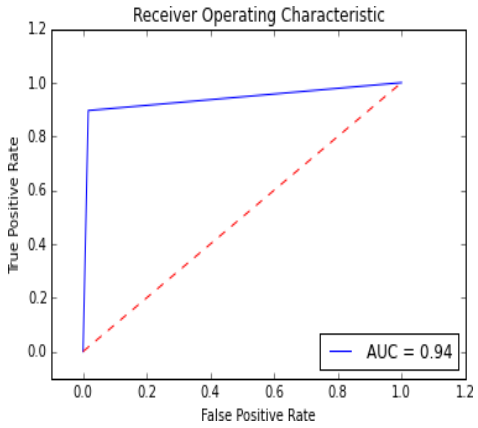


Fig : Accuracy Using Neural Network

6. CONCLUSION

Through utilization of different kinds ofMachine Learning Algorithms, this paper is aimed to exploit different aspects of dataset which has not been deeply considered in literature and to find a good way of detection of the fake and automated accounts. In this paper we have presented a Machine Learning pipeline for detecting fake accounts in online social networks. Rather than making a prediction using one single algorithm, our system uses three different classification algorithms to determine whether or not an account in the provided dataset is a fake account or not. Our evaluation using Support Vector Machine, Random Forest and Neural Networks

showed strong performance, and the comparison of the accuracy of prediction seemed to be higher using Support Vector Machine for the given dataset. The Accuracy of detecting fake accounts is found to be higher using Random Forest Algorithm followed by Neural Networks Algorithm for a given dataset. As a future work,[5] recurrent neural networks can be utilized for the time series user data for a better detection of fake accounts and the algorithms can be applied to various social online platforms such as Instagram, LinkedIn and Twitter to detect the fake accounts.

REFERENCES

[1] 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. 36723681.

[2] Rao, K. Sreenivasa, N. Swapna, and P.Praveen Kumar. "Educational data mining for student placement prediction using machine learning algorithms." *Int. J. Eng. Technol. Sci* 7.1.2 (2018): 43-46.

[3] Y. Boshmaf, D. Logothetis, G. Siganos, J. Lería, J. Lorenzo, M. Ripeanu, K. Beznosov, H. Halawa, "Íntegro: Leveraging victim prediction for robust fake account detection inlarge scale osns", *Computers & Security*, vol. 61, pp. 142-168, 2016.

[4]N. Singh, T. Sharma, A. Thakral and T. Choudhury, "Detection of Fake Profile in Online Social Networks Using Machine Learning," *2018 International Conference on Advances in Computing and Communication Engineering (ICACCE)*, Paris, 2018, pp. 231-234.

[5]D. M. Freeman, "Detecting clusters of fake accounts in online social networks", *8th ACM Workshop on Artificial Intelligence and Security*, pp. 91-101.