International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 07 Issue: 05 | May 2020 www.irjet.net p-ISSN: 2395-0072

Identifying Fake Twitter Account using Neural Network in Machine Learning

Bommisetty Ganesh[1], Dhanush R[2], Gurram Akhil[3], Gnanasekar [4]

[1] Student, Department of Computer Science, R.M.D. Engineering College
[2]Student, Department of Computer Science, R.M.D. Engineering College
[3]Student, Department of Computer Science, R.M.D. Engineering College
[4]Assistant professor, R.M.D. Engineering College

Abstract –

In day today life Social media became a part of everyone's life. Twitter and Facebook become the major platform for using social media. Nowadays social media is used for spreading real time information, and we have to make sure that the information published is real. But some people who tend to spread the fake news of a person or a product uses the fake Id's called as bots. So it is difficult for a common user to know which news is genuine and which is false news. These fake ID's if used in large scale can create a huge damage to the society. The main objective of our project is to identify these fake ID's. In this project to find the fake twitter account we presented a classification method. The paper determines the minimized set of attributes that influence for finding the fake Twitter account, and then the determined factors are applied using different classification techniques. The most accurate algorithm has been determined by comparing the techniques results. By using minimum set of attributes the amount of data to be analyzed is reduced and we can get faster results.

***Key Words*: Identifying fake twitter account, classification algorithm, neural network**

1.INTRODUCTION

SOCIAL MEDIA HAS GROWN ENORMOUSLY FROM THE PAST FEW YEARS. DURING THIS RISE, DIFFERENT SOCIAL MEDIA HAVE CREATED MANY ONLINE ACTIVITIES WHICH ATTRACT LARGE NUMBER OF USERS WHERE USERS INCREASES DEPENDS UPON THE INFORMATION PUBLISHED IN THE ONLINE SOCIAL NETWORKS(OSNS)[1]. WHERE ON THE OTHER HAND OSNS ARE SUFFERING WITH THE INCREASING IN THE NUMBER OF FAKE ACCOUNTS THAT HAS BEEN CREATED . FAKE ACCOUNTS MEANS THAT ARE NOT REAL .THESE FAKE ACCOUNTS PUBLISH FAKE NEWS AND SPAM. OSNS OPERATOR ARE NOW EXPEND AND DETERMINED RESOURCES TO DETECT THE FAKE ACCOUNTS.

Twitter is widely used by so many clients (almost 46%) [2] for sharing messages pictures post or some other type of data. Twitter allows the user to send the information to a large group of users who are active in real time. In Twitter the fake accounts are called as bots .A Bot is a simple software which autonomously processes repitative tasks. [3] Bots (short for software robots) were merely known to the world since the early days of computers. One of the captivating example of bots is chatbots, algorithms were designed to maintain conversation with humans, that was visualized by Alan Turing in the 1950s.[4] The designing of computer algorithm which passes the Turing test driven artificial intelligence research for decades, that was witnessed by the previous resarchers like the Loebner Prize, awarding progress in Natural Language Processing(NLP)g. Since the early days of AI, we have identified drastic changes that are evolving. When bots like Joseph Weizenbaum's ELIZA,[5] mimicking a Rogerian psychotherapist, were developed as demonstrations for delight.Spam bots collect email addresses to send unwanted spam mails. Social bots are misused by political parties and state to distort the public opinion. They are used for spreading fake news at a faster pace. Reference [6] has presented a study that Fake accounts are used in US election for spreading rumors in Ukraine conflict to mislead the public about the news.

2. EXISTING SYSTEM

In previous researchers, uses some factors to conclude whether a twitter account is fake or not. These factors may largely affect the way of making decision towards fake Id. When number of factors are low we will not obtain the correct result significantly. Now using advanced technology they have great improvement in creating fake accounts which cannot be matched by the software that are used to detect. Because of the advancement in fake account creation, the existing methods have turned obsolete. The commonly used method to detect fake account is Random forest algorithm [7]. These methods have few downsides such as inefficiency to handle the variables used in this algorithm in different number of levels. Also there is an increase in the time efficiency that is taken as hit. It also uses more no of trees. Clock activity is also used to detect whether an account is used by a bot[3]. At this stage we intensely look for likes, comments and shares for this particular account from the time of creation. If this account has enormous no of likes, shares and comments then it will be concluded as fake on used by a bot. This rate cannot be achieved by a normal social user. Also, the total amount of time it was online will be looked before concluding. On other factors we consider the information provided by the user like phone number, email address etc.

2.1 DISADVANTAGES

1.Accuracy is less
2.Less attributes for evaluation

3. PROPOSED SYSTEM

This concept is based on the confidence that humans usually behave differently than the fakes, therefore, detecting this behavior will lead to the revealing of the fake accounts. In this section, we will demonstrate some of the works that have been presented in this area. It has reached an accuracy of 84.5% to detect spammers by identifying 23 attributes, most of these attributes. (17 attributes) are demonstrated. However, in our research, we have reached more accuracy with smaller set of attributes as will be discussed. In the set of attributes has been minimized by identifying ten attributes for detected. However, in the previous research, the result was not promising for identifying fake accounts with more optimistic perspective that it is able to identify fake tweets with higher accuracy by using the support of graphical method. Although has presented a minimized set of attributes which contained six attributes, however, it is mentioned that it could only detects determined types of spammers, they are bagger, and poster spammers. In our approach, we propose minimized set of attributes for detecting all types of false news. In addition, one of these attributes requires text analysis procedure for finding the similarities among messages which is not required for our proposed approach. Moreover, it is mentioned in that Random Forest algorithm is the best results for detection for Twitter.

3.1 ADVANTAGES

1. High Accuracy
2. Efficient result prediction

4. APPROACH

4.1 Data collection

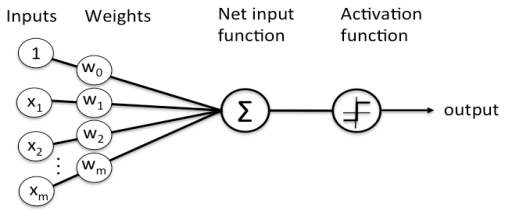
At first we collect dataset of fake and genuine profiles. Various attributes included in the dataset are a number of friends, followers, status count. Dataset is divided into training and testing data. Classification algorithms are trained using a training dataset and the testing dataset is used to determine the efficiency of the algorithm. From the dataset used, 80% of both profiles (genuine and fake) are used to prepare a training dataset and 20% of both profiles are used to prepare a testing dataset.

4.2 Pre-processing

In training dataset, it contain status count, follower count, favourties count, friends count, listed count, sex code, lang code is obtained for numerous individual users. Features are extracted from it.

4.3 Training

Neural networks are a group of algorithms are used to design for recognizing patterns. They interpret sensory data to sort machine perception, and also used for labeling and clustering raw input. Then these patterns are recognized as numerical vectors, real world data, images, sounds, text are translated.

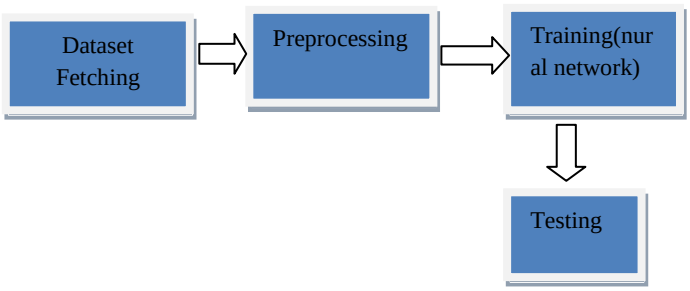


4.4 Testing

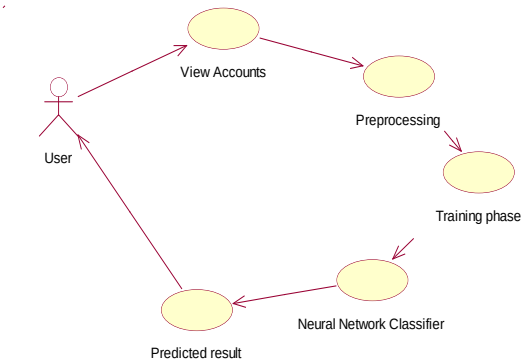
Based on the prediction value the account is classified into fake or normal account.

5. System Design

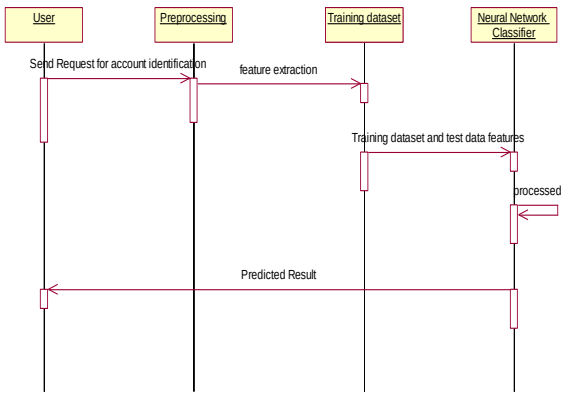
5.1 Architecture Diagram



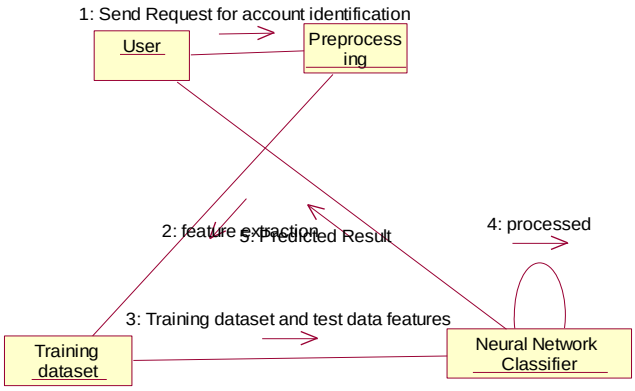
5.2 Use case Diagram



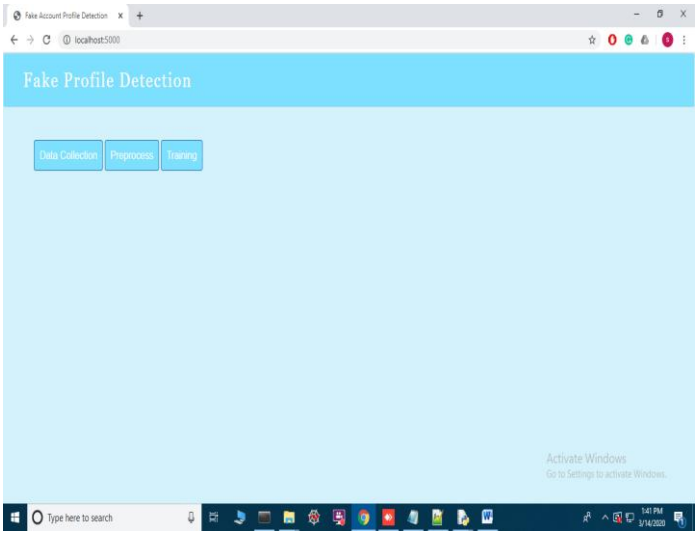
5.3 Sequence Diagram



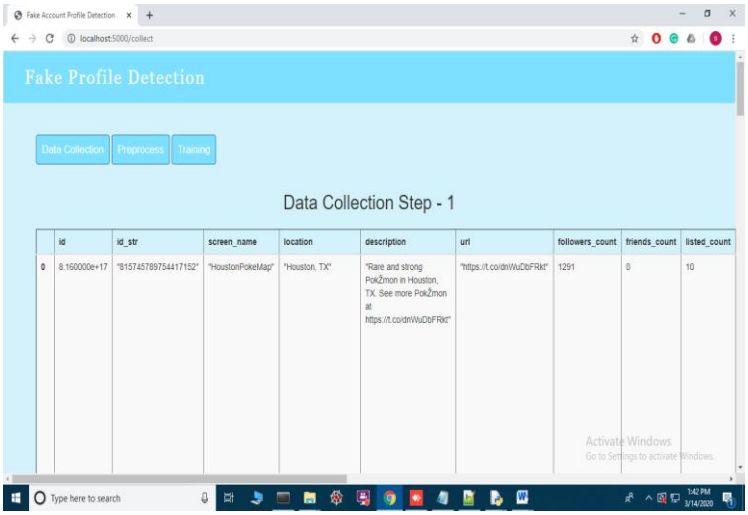
5.4 Collabration Diagram



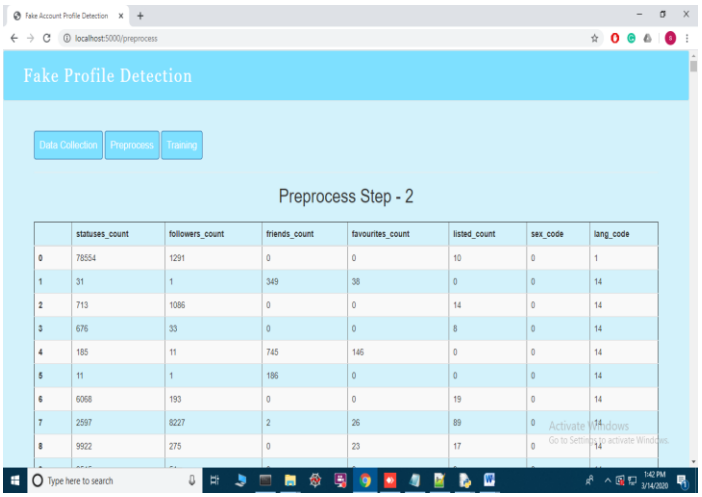
6.Implementation Results



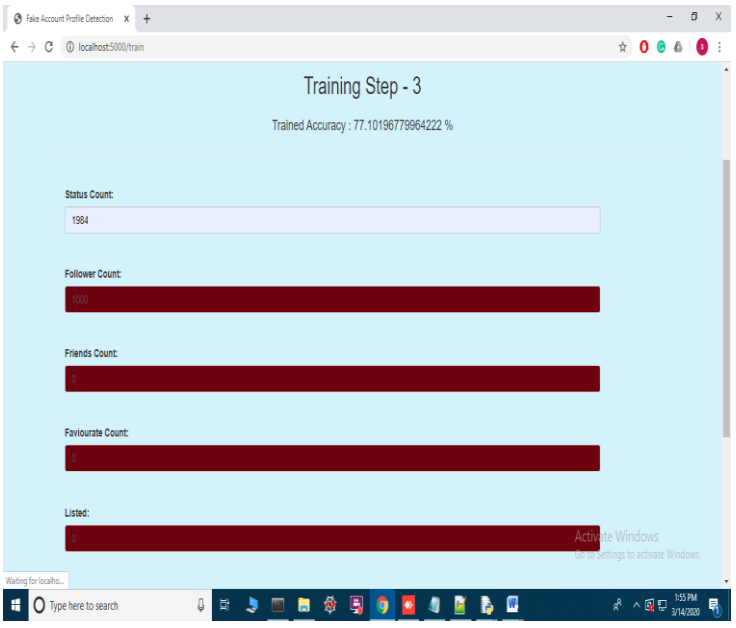
Home page



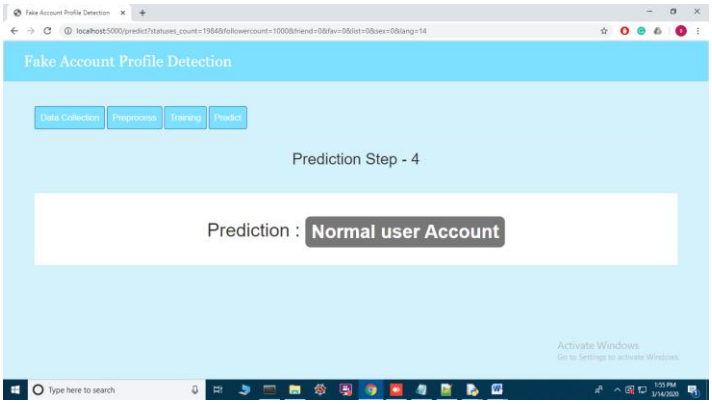
Data collection



Preprocessing



Training



Testing

7. Conclusion

We have given a framework using which we can identify fake profiles in any online social network by using Neural Network with a very high efficiency as high as around 77%.The model presented in this project demonstrates Neural Network (NN) is an elegant and robust method for classification in a large dataset. Regardless of the non-linearity of the decision boundary, NN is able to classify between fake and genuine profiles with a reasonable degree of accuracy. This method can be extended on any platform that needs classification to be deployed on public profiles for various purposes. This project uses only publicly available information which makes it convenient for users.

REFERENCES

[1] Qiang Cao, Michael Sirivianos, Xiaowei Yang, and Tiago Pregueiro, "Aiding the detection of fake accounts in large scale social online services," in Proceedings of the 9th USENIX conference on Networked Systems Design and Implementation, 2012.

[2] Carlos Castillo, Marcelo Mendoza, and Barbara Poblete, "Information credibility on twitter," in Proceedings of the 20th international conference on Worldwide web, 2011.

[3] B. Y.E. Ferrara, O.Varol, C. Davis, F.Menezer, and A. Flammini, “The rise of social bots,” Commum.ACM, vol.59 No.7,pp.96-104,2016.

[4] Turing, A.M. Computing machinery and intelligence. Mind 49, 236 (1950), 433–460

[5] Weizenbaum, J. ELIZA—A computer program for the study of natural language communication between man and machine. Commun. ACM 9, 1 (Sept. 1966), 36–45

[6] M. Camisani-Calzolari. (2012, August) Analysis of Twitter followers of the US Presidential Election candidates: Barack Obama and Mitt Romney. (Online). http://digitalevaluations.com/

[7] L. Breiman, "Random forests," Machine Learning, 2001.