IDENTIFICATION OF FAKE ACCOUNTS ACROSS SOCIAL NETWORKS

[1]MS. AFREEN SUBUHI, [2]B. SWARANI, [3]B. SAI TEJA

[1]Assistant Professor, Department of Information Technology, CMR College of Engineering & Technology

2, 3B-Tech, Department of Information Technology, CMR College of Engineering & Technology

Abstract:

Online Social Networks (OSNs) presently engage the majority of people, from a child to an adult and even old age people as they spend a good amount of time on these platforms exchanging their information and creating interaction with other people of the world. On one hand, these social networks provide the advantage of direct connectivity between people, information sharing, ways to create a large audience, etc. on the other hand people also misuse them in many ways. Social networking sites are suffering from people who own bulk of fake accounts to take advantage of vulnerabilities for their immoral benefits such as intriguing, targeted accounts to click on malicious links or to attempt any other cybercrimes. These actions motivate researchers to develop a system that can detect fake accounts on these OSNs. Several attempts have been made by the researchers to detect the accounts on social networking sites as fake or real, relying on account's features (user-based, graph-based, contentbased, time-based) and various classification algorithms. In this paper, we provide an overview of various studies done in this direction and a survey of all the techniques already used and can be used in the future.

INTRODUCTION :

Throughout the human history people have worked on developing better ways of communication. One such attempt led to the birth of social networking sites in the 1990s. Though it took some time for people to catch on with these sites but by the early 2000s the social media platform began to flourish. At least 3 social networking sites were launched every year. The world witnessed one of the biggest revolutions of all time. Currently everyone feels the need to have an online presence. The social media has become the number 1 activity on the internet. In 2010 the number of active users of social media was 970 million, within a decade this number rose up to 3.6 billion in 2020. In the coming 5 years it is expected that the social media will continue to bloom, eventually attracting more than 4.41 billion users by the year 2025. These platforms have transformed the web into a social web where it has become easy for people to find old friends via Facebook, get latest news updates on Twitter, job hunt on

LinkedIn and watching trending videos with a single click on YouTube. These platforms have become our go to space for entertainment and keeping abreast. Not only this social media has also given us a space to voice our opinion without any fear. As a result, we have witnessed hundreds of revolutions on the internet like the popular me-too movement. These platforms have also evolved into a new marketing platform which is free of cost unlike the traditional forms of publicity like TV, billboards, radio etc. Social media has become so entangled with our everyday activities that it has become impossible to imagine our lives without these platforms. The growing popularity of social media platforms has not only benefitted the people but also caught the attention of scammers. On one hand social media is bringing people together and on the other hand it has created a guarded space for fraudsters to carry out a number of illegal activities. The absence of any authentication process has made it easy for anyone to make a fake account. This serves as an advantage for the scammers encouraging them to use fake account for illegal activities as there is a good chance that the account holder will not get caught. Owing to this the popularity of fake accounts has increased. The use of these phantom accounts to impersonate someone in hope of defaming them has become a common issue. At times these accounts serve the bigger purpose of acting as a trusted acquaintance to get personal information from a person. This obtained information can be used to carry out phishing attacks. People often use these dummy accounts to spread fake news which in the worst case can cause riot like conditions. Some people make use of fake accounts to spread hate which can be directed at certain race, religion, country or often at a particular person. This has increased the cases of cyber bullying leading to rise in the cases of depression and anxiety in teenagers. The social media platforms have also seen an increase in the number of accounts which provide services or products in exchange of money. But most of these accounts are fake, as a result thousands of people are sold fake products and are promised fake services by these accounts. Sometimes these fake accounts are used by companies to build hype for their bad products and services. Not only scammers but also a lot of influencers also use fake bot followers to appear popular, which helps them in gaining more offers from companies asking to publicise their products. At times the fake accounts can also be used to befriend a person in order to stalk them.

OBJECTIVE

As we see the number of people using OSNs is increasing, the creation of fake

social media accounts is also increasing. The main motivating factor for the identification of these fake accounts is that these accounts are mostly created to carry out cyber extortion or to commit cybercrimes anonymously or with an untraceable identity as a result of this, the rate of cybercrime has increased noticeably from the last one year. Also, the owner of fake accounts sometimes aims to take advantage of the kindness of people by making false announcements or by spreading fake news through these accounts to usurp money from innocent people. Moreover, people are creating multiple accounts that do not belong to someone and only created to get a hike of votes in online voting systems and so as in online gaming in the greed of getting referral incentives.

IMPLEMENTATION

Inspired by the significance of identification of fake accounts recently researchers have started to examine efficient mechanisms. In most of the previous works, user based and graphbased features were majorly used in prediction and classification of accounts. Various studies have been done in this area by using different datasets, feature reduction techniques, classification techniques. To maximize accuracy the support vector machine used, k-nearest neighbor and random forest classifier and achieved 90.3% accuracy in detecting fake users. And extracted the features from profile attributes user activity values and nodes edge weight as it is graph-based and content-based action. CMRCET B. Tech (IT) Page No 10 Identification of fake accounts across social networks Faurecia Benevento detected spammers by using SVM classification algorithm. They collected dataset of twitter and this dataset contains information of 54M users, 1.9B links and approximately 1.8B tweets. Their approach detected 70% of spammers and 96% of nonspammers correctly. They collected 16 features using Twitter API, 13 of which were directly collected from API and 3 were created by them using API. As a pre-processing method on the collected data mainly the focus is on the effect of discretization method. By using the discretization technique and using the Naïve Bayes classifier they increase their accuracy from 85 to 90% the shortcoming in that research is they can improve the result by using feature selection. And has reached a classification accuracy of around 98% by using the support vector machine and Neural Network classifier and a hybrid approach of both and compare the accuracy result of the three mentioned classifier. As far as we know from our study in most of the work as OSN platform Twitter is chosen because in Twitter information is available publicly by default

and this information can be easily accessed with the help of Twitter APIs. The different fake profile detection approaches for different OSNs are summarized along with the features used, datasets and techniques which are discussed above in this section expansively

PROPOSED SYSTEML:

Objective of Proposed Model Each social media profile has a lot of data associated with it like the user's name, account holder's name, number of friends of the user, date of birth of the user, phone number of the user, etc. We can make use of the associated data to comment on the genuineness of the account. In this research we have used deep neural network to recognise fake Instagram accounts. A six layered ANN model is used for the purpose. The designed model uses features like username, number of followers, profile description length, number of accounts followed by the user, number of posts etc to declare a given account as genuine or fake. Nowadays for security reasons the social media companies have started providing the facility of making account private to the users but if this method is deployed then they can even access the information of private accounts for examination without any violation of privacy. The dataset used for this case study is an open dataset which is available on Kaggle. The dataset has details of 96 Instagram accounts. These 696 data entries are divided into 2 folders train set and test set. he trains set has data of 576 accounts and the test set has data of 120 accounts. Both train and test set are balanced meaning they have a ratio of 1:1 between real and fake accounts, i.e., the train set has 288 real and 288 fake account details and test set has 60 real and 60 fake account entries. The target variable which expresses whether the given account is real or fake can take up 2 values 0 and 1. The value being 1 if the account is fake and 0 if the account is genuine. Other than the target variable the dataset has 10 features out of which 4 take binary value and the rest take integer value. Before the dataset can be used for the estimation process it is essential to process it. The dataset was checked in order to eliminate any missing values. Below figure shows the dataset used for the training purpose.

Algorithm used for Proposed Model ANNs are computational models that are designed to emulate the human brain. The working of the brain is used as a basis by the ANN for development of complex algorithms that can be used to solve a given problem. Like the human brain ANN can decipher and arrive to a conclusion from a given vague information. ANN is made up of thousands of artificial neurons which are simply called units or nodes. These nodes are like the natural neurons

both in structure and working. These nodes are connected to each other to form layers and the interconnection of these layers produces a web like structure which is called a network. A fundamental ANN comprises of input, hidden and output layers. The input layer accepts input from the outside world. The nodes present in the input layer are passive in nature, this means that they are incapable of making any changes in the data provided to them. The input provided to this layer can be in form of a pattern or a vector in case of visual data. This accepted input is then transformed into something meaningful by the hidden layers. The hidden layers refine the features of the input before passing them on to the output layer. Finally, the output layer responds to the given information and produces output for the system. Based on the number of hidden layers an ANN can either be shallow neural network or deep neural network.

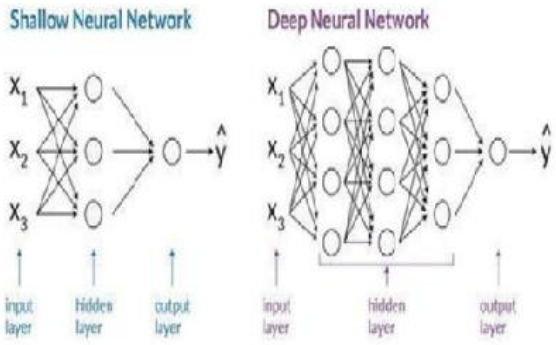


Fig:-1 Shallow and deep neural network architecture

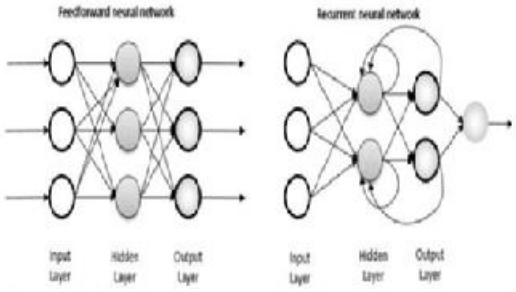


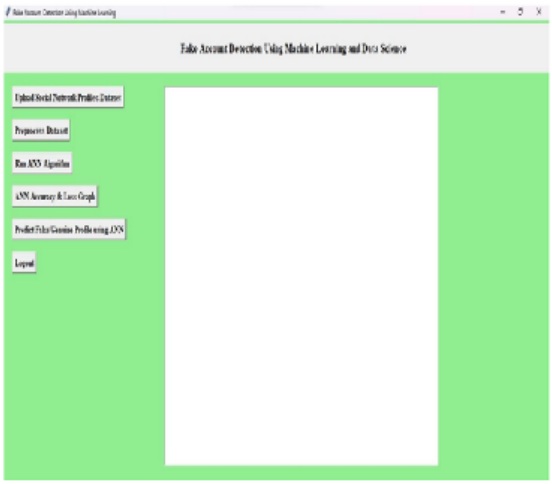
Fig:-2 Feed forward and feedback network

RESULTS AND DISCUSSION

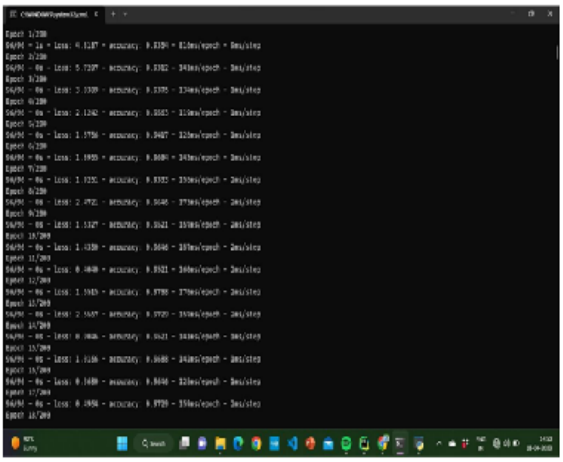
ANN algorithm will be trained with all previous users fake and genuine account dataset and then whenever we gave new test data then that ANN train model will be applied on new test data to identify whether given new account details are from genuine or fake users. Online social networks such as Facebook or Twitter contains users details and some malicious users will hack social network database to steal or breach users information, To protect users data we are using ANN 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 methods to create new and better performing types of

prediction and detection model. To train ANN algorithm we have used below details from social networks. Account_Age, Gender, User_Age, Link_Desc, Status_Count, Friend_Count, Location, Location_IP, Status.

4.2 Screenshots of output



4.2.1 Before uploading dataset



4.2.3 Running of ANN algorithm

CONCLUSION :

In this research we have recognized a serious issue haunting the social media platforms which is the ever-increasing number of fake accounts on them. To overcome this problem, we have proposed a deep learning model which can be used to identify the dummy accounts in matter of seconds which can be then removed before they cause any serious harm to the people. The suggestion of a deep learning has been done in this project keeping in mind the drawbacks of the currently existing methods. The model used studies the data associated with the accounts to derive a relation between it and the genuineness of the account. To represent the performance of the model we have used confusion matrix. and learning curves along with the accuracy of the model. The model has shown good performance in case of both training and testing set.

FUTURE ENHANCEMENT :

Currently only the data available for Instagram profiles has been used for the training and testing purpose but in future we can also train the model to identify fake accounts on other popular platforms like Facebook, LinkedIn, Twitter and many more by providing an efficient dataset for them.

REFERENCES :

• Boshmaf, Y., et al.: Integro: leveraging victim prediction for robust fake account detection in OSNs. In: NDSS, vol. 15, pp. 8–11, February 2015.

• Erşahin, B., Aktaş, Ö., Kılınç, D., Akyol, C.: Twitter fake account detection. In: 2017 International Conference on Computer Science and Engineering (UBMK), pp. 388–392. IEEE, October 2017.

• Mateen, M., Iqbal, M.A., Aleem, M., Islam, M.A.: A hybrid approach for spam detection for Twitter. In: 2017 14th International Bhurban Conference on Applied Sciences and Technology (IBCAST), pp. 466–471. IEEE, January 2017.

• Pekel, Engin & Kara, Selin. (2017). A Comprehensive Review For Artifical Neural Network Application To Public Transportation. Sigma Journal of Engineering and Natural Sciences. 35. 157-179.

• Raturi, Rohit. (2018). Machine Learning Implementation for Identifying Fake Accounts in Social Network.

• Samala Durga Prasad Reddy,"Fake Profile Identification using Machine Learning" in IRJET 2019.

• Bharat Sampatrao Borkar, Dr. Rajesh Purohit,"Recognition of fake profiles in social media : a literature review",Volume 5, Issue 2, 2019

• Fabiyi, Samson Damilola. (2019). A Review of Unsupervised Artificial Neural Networks with Applications. International Journal of Computer Applications. 181. 22-26. 10.5120/ijca2019918425

• Bala Anand, M., Karthikeyan, N., Karthik, S., Varatharajan, R., Manogaran, G., Sivaparthipan, C.B.: An enhanced graph-based semi-supervised learning algorithm to detect fake users on Twitter. J. Supercomput. 75(9), 6085–6105 (2019). https://doi.org/10.1007/s11227- 019-02948-w

•Reddy, b. V. R., dasari, n., & venkateswararao, k. (2021). A steganography system with gausian markov random fields and error detection codes.

• Revathy, G., Gurumoorthi, E., Sasikala, C., & Latha, T. M. (2023, June). Training superbot with learning automata and multi kernel SVM. In AIP Conference Proceedings (Vol. 2782, No. 1). AIP Publishing.

• Latha, Ch & Soujanya, K. & Amulya, C.. (2020). Remote Monitoring and Maintenance of Patients via IoT Healthcare Security and Interoperability Approach. 10.1007/978-981-15-1632-0_22.

Sujihelen, L., Boddu, R., Murugaveni, S., Arnika, M., Haldorai, A., Reddy, P.C.S., Feng, S., Qin, J., 2022, Node Replication Attack Detection in Distributed Wireless Sensor Networks, Wireless Communications and Mobile Computing, 10.1155/2022/7252791

Venkataiah, V., Nagaratna, M., Mohanty, R., 2022, Application of Chaotic Increasing Linear Inertia Weight and Diversity Improved Particle Swarm Optimization to Predict Accurate Software Cost Estimation, International Journal of Electrical and Electronics Research, 10.37391/IJEER.100218

Narasimha, V., Dhanalakshmi, M., 2022, Risk Factor of Diabetes with Comorbidity Using Machine Learning Techniques, Lecture Notes in Electrical Engineering, 10.1007/978-981-16-7985-8_37

Rashid, E., Ansari, M.D., Gunjan, V.K., 2022, Innovation and Entrepreneurship in the Technical Education, Lecture Notes in Electrical Engineering, 10.1007/978-981-16-7985-8_125

Sunitha, P., Ahmad, N., Barbhuiya, R.K., Gunjan, V.K., Ansari, M.D., 2022, Impact ofÂ Covid-19 onÂ Education, Lecture Notes in Electrical Engineering, 10.1007/978-981-16-7985-8_124