International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878 (Online), Volume-8 Issue-4, November 2019

Fake Profile Identification in Online Social Networks

Sk.Shama, K.Siva Nandini, P.Bhavya Anjali, K. Devi Manaswi,Sk.Wasim Akram

Abstract: ***There is a tremendous increase in technologies these days.. Mobiles are becoming smart. Technology is associated with online social networks which has become a part in every one's life in making new friends and keeping friends , their interests are known easier. But this increase in networking online make many problems like faking their profiles, online impersonation having become more and more in present days. Users are fed with more unnecessary knowledge during surfing which are posted by fake users. Researches have observed that 20% to 40% profiles in online social networks like facebook are fake profiles. Thus this detection of fake profiles in online social networks results into solution using frameworks.***

Keywords: Online Social Networks, Fake profiles, Classification, Neural Network.

I. INTRODUCTION

Online social media is the place each person has a outlook then be able to keep connecting their relations, transfer their updates, join with the people having same likes. Online Social Networks makes use of front end technologies, which permits permanency accounts in accordance with to know each other. Facebook, Twitter are developing along with humans to maintain consultation together with all others. The online accounts welcome people including identical hobbies collectively who makes users easier after perform current friends. Gaming and entertaining web sites which have extra followers unintentionally that means more fan base and supreme ratings.

Manuscript published on November 30, 2019.
* Correspondence Author
Sk.Shama*, Assistant Professor of Computer Science and Engineering Department Koneru Lakshmaiah Education Foundation situated Vaddeswaram, Guntur District Andhra Pradesh, India.
Sk. Wasim Akram, Assistant Professor of Computer Science and Engineering Department at Vasireddy Venkatadri Institute of Technology Guntur,
K. Siva Nandini, Computer Science And Engineering Department, Koneru Lakshmaiah Education Foundation Vaddeswaram, Guntur District Andhra Pradesh, India.
P.Bhavya Anjali, Computer Science and Engineering Department, Koneru Lakshmaiah Education Foundation Vaddeswaram, Guntur District Andhra Pradesh, India.
K.Devi Manaswi, from Computer Science and Engineering Department, Koneru Lakshmaiah Education Foundation Vaddeswaram, Guntur District Andhra Pradesh, India.

Ratings drives online account holders to understand newer approaches not naturally or manually to compete more with their neighbours. By these analogies , the maximum famous candidate in an election commonly get more number of votes. Happening of fake social media accounts and interests may be known. Instance is fake online account being sold on-line at a online market places for minimum price , brought from collaborative working offerings.

More often feasible to have Twitter fans and Facebook media likes in online. Fake user accounts may be created by humans or computers like bots , cyborgs. Cyborg is half bot and half human account. These accounts are usually opened by human, but their actions are made by bots. The another reason for people to create fake profiles for defaming accounts they dislike. This type of users create accounts with the username of the people they hate and post irrelevant stories and snap shots on their accounts to redirect everybody so that they assume that particular person is awful and make their reputation low.

Most attackers are in it to make money. They make money by distributing unwanted ads (spam) or capturing accounts they can reuse or resale (phishing). Spammers gather resources to know fake and real users, email ids ,ip locations and computing knowledge power. Every one of these advantages can have a huge expense related with them, and an assault, similar to any business adventure, needs benefit to continue onward. Attackers more often use facebook logins, applications, Events, Group users to gather login credentials, spam users, and ultimately gain profits. They need email records, treats, and a wide scope of IP delivers to go around notoriety based protections. Moreover, they use telephone numbers, taken charge cards, and CAPTCHA arrangements trying to go around validation checks.

Facebook security privileges its system to gather users to prevent spams and fishing accounts. Facebook Immune System does continuous minds all gather and each its activity made by it. Social bot is a known that stops and controls social online accounts. Bots socially is an auto generated software. Precised way a social account duplicates relies upon at the social media, also in contrast to general bot, a social bot interacting more in different customers that the social bot is a actual man or woman.

More auto generated programs or semi generated computer programs that duplicate the human behaviour in Social media. So to use them hackers attack online social networks. Thus primarily used for campaign, advertise and also thieve user non-public in more large scales. The bot online master gather inputs because of attackers.Cyborg bots appear like accounts of human from random calls of human, often selected human users image and user records published more often from collective accounts to be prepare from before online account attackers.

Cyborg bots ship gather random users. If a person acknowledge the request from user, ship to get request of the account who agree request, will increase popularity price because of lifestyles of mutual friends.

II. RELATED WORK

Accounts in online social media have heaps of input data like name, sexual orientation, companions, devotees, preferences, area numbers. Half part of this input data are both of public and private. We have to use input that are public to know profiles which are phony for interpersonal organization as data from private is unavailable.

In any case, on the off chance that our proposed plan is utilized by the interpersonal interaction organizations itself, at that point they can utilize the private data of the users to know not from abusing from security issues. Considered data is highlights for profiles to classify of phony and genuine profiles.

For detecting fake profiles we followed these steps:

1. Functions are to be selected after choice of attributes, the ataset of profiles which are already classified as fake or real are wanted for the schooling motive of the classification algorithm. We have used a publicly available dataset of 1337 fake customers and 1481 actual users which includes numerous attributes consisting of call, status count, number of friends, fans depend, favourites, languages regarded and so forth.
2. The selected attributes are extracted from profile for the purpose of type.
3. After this the dataset of fake and actual seasoned files are prepared. From this dataset, 80% of both seasoned files (authentic and pretend) are used to prepare a schooling dataset and 20% of both profiles are used to put together a testing dataset.
4. The schooling dataset is then fed to the classification set of rules. It learns from the education dataset and is predicted to offer correct elegance labels for the testing dataset.
5. The labels from the testing dataset are eliminated and are left for determination by the educated classifier.
6. The result of classification algorithm is shown in 4.4. we've got used two classification algorithms and have compared the efficiency of these algorithms.
7. The proposed structure in the figure 1 shows the succession of procedures that should be pursued for persistent location of phony profiles with dynamic gaining from the input of the outcome given by the arrangement calculation.

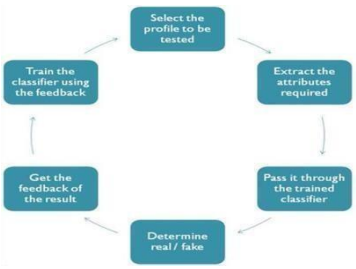


Fig 1. Cycle for Detection

The structure that can without much of a stretch be executed by long range informal communication organizations as they approach client data.

1. Order begins from the determination of profile that should be characterized.
2. When the profile is chosen, the helpful highlights are separated for the reason for order.
3. The separated highlights are then encouraged to prepared classifier.
4. Classifier is prepared routinely as new information is nourished into the classifier.
5. Classifier at that point decides if the profile is veritable or counterfeit.
6. The consequence of order calculation is then checked and input is sustained over into the classifier.
7. As the quantity of preparing information builds the classifier turns out to be more and increasingly precise in foreseeing the phony profiles.

III. METHODOLOGY

Implementation is a technique of categorizing an object into a particular class based on the training data set that was used to train the classifier. We feed the classifier with data set so that we can train it to identify related objects with as best accuracy as possible. Classifier is an algorithm used for classification. In this project we have used two classifiers namely Neural Networks and Support Vector Machines and have thereby compared their efficiencies.

Neural Network:

The conventional method by which a computer works is that you provide instructions or algorithms to the computer and it generates output based on it. But what if you do not know the algorithm to solve a problem? Will your computer still be able to provide solutions. If we use conventional techniques, then the computer will not be able to solve the problem unless you provide some instructions. Here comes the concept of Neural Networks.

We can still solve such a problem by training a network as such our program will learn on its own and will provide solution close to a certain accuracy. The term Neural Networks was coined in 1943 but could not be implemented then due to lack of technology. Neural Networks learn by example. Neural Networks are based on biological neurons i.e. brain cells and the way information is processed inside the brain. There are mainly two types of neural networks :

1. Single layer.
2. Multi layer.

Random Forest Classification Technique:

This classifier classifies collection of decision trees to subset of randomly generated training set. Then it augments the likes from decision sub trees to know subclass of handling object for tests. Random forest will generate NA missing values for attributes increase accuracy for larger sets of data. If more number of tress, it doesn't allow to trees to fit model.

Table.1 Comparison of accuracy for different algorithms

Algorithm	Precision	Recall	Accuracy
Decision Tree Network(Twitter and face book)	0.999	0.991	99.9%
Neural Networks Network (Twitter)	1	0.417	-
Naïve Bayes Network(Email and Twitter)	0.778	0.444	94.5%

IV. IMPLEMENTATION

1.Collect Data and pre-process the data 2.Generate fake accounts.

3.Data Validation to find fake and real . 4.Create new features.

5. Apply neural networks, random forest.
6. Evaluate results of accuracy, recall etc parameters.

Thus these steps are implemented for detecting fake profiles.

Data set:

We needed dataset of fake and genuine profiles. Various attributes to include in the dataset are number of friends, followers, status count. Dataset is resulting to training and testing data. Classification algorithms are trained using training dataset and testing dataset is used to determine efficiency of algorithm. From the dataset used, More than 80 percent of accounts are used to train the data, 20 percent of accounts to test the data.

Attribute	Explanation
Post Count	The average number of posts created by users are expected to have a low count when the account is fake.
Comment Count	Fake accounts share and post unwanted links and advertisements which make a lower count.
Followers Count	Usually, fake profiles have low count but there is high follower count then they may belong to the same group.
Events	They won't add or share any event, live locations frequently.
Location	Fake profiles have irrelevant study and work locations.
Tagged Post	The number of tagged posts is comparatively less for fake users.
Created at	From the creation date, they use the timeline for less period of time.
Description	They make a description to advertise and connect with more number of people.
URL	The display name and URL don't match mostly.

Table.1
Description of attributes in data sets. V PERFOMANCE MEASURE

Efficiency = Count of correct predictions to that of total count of predictions.

Percent Error = (1-Efficiency)*100

Confusion Matrix : It is a way for summarizing the overall performance of a classification algorithm. Calculating a confusion matrix can come up with a better concept of what your category version is getting proper and what kinds of mistakes it is making.

TPR-True Positive Rate TPR=TP/(TP+FN)

FPR- False Positive Rate FPR=FP/(FP+TN)

TNR-True Negative Rate TNR=TN/(FP+TN)

FNR- False Negative Rate FNR=1-TPR

Recall – Number of the true positives were done,i.e. what number of the right hits were likewise found.

Recall = TP /(TP+FN)

Precision- Precision is how many hits are returned to true positive i.e. what number of the found were right hits.

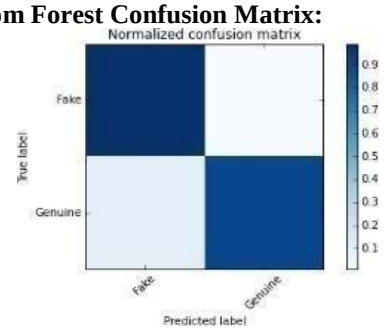
Precision - TP/(TP+FP)

F1 score measure of accuracy for tests. It accept exactness the review p,r of the test scoring the figure. ROC Curve is the plot of FPR versus TPR.

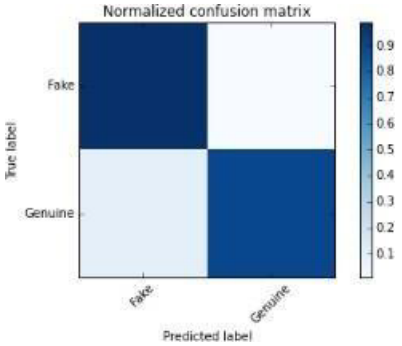
ROC used to differentiate the performance measurement of different classifying techniques.

Neural Networks Confusion Matrix:

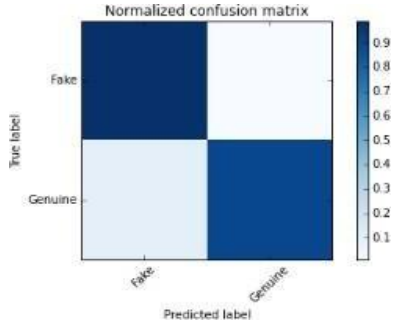
Random Forest Confusion Matrix:



Normalized Confusion Matrix for Neural Networks



Normalized Confusion Matrix for Random Forest

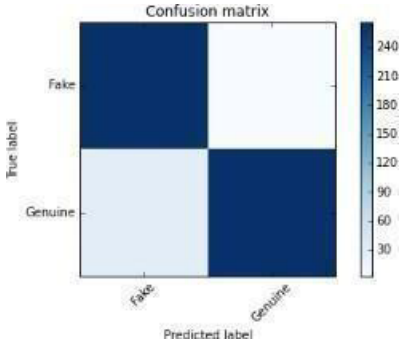


Confusion Matrix:

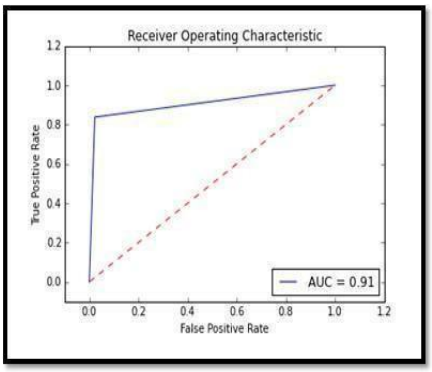
A confusion matrix is a summary of prediction outcomes on a classification problem. The number of accurate and incorrect predictions are summarized with depend values and damaged down by each elegance. that is the key to the confusion matrix.

The confusion matrix suggests the methods in which your classification model is confused while it makes predictions. It gives us perception now not only into the mistakes being made by a classifier but extra importantly the forms of mistakes which can be being made.

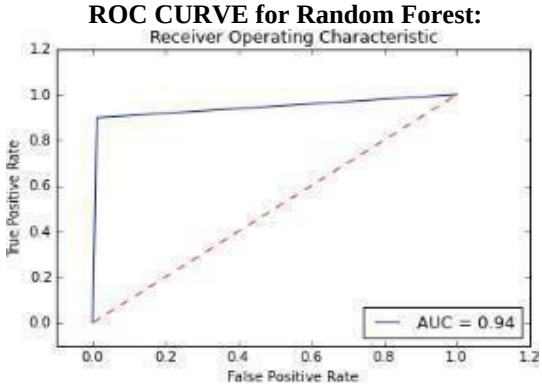
Random Forest Confusion Matrix:



ROC CURVE For Neural Networks:



ROC CURVE for Random Forest:



ROC CURVES:

Efficiency of neural neural network in classifying data is 91%. We have taken 80% of data for training neural network and 20% for classification.

Efficiency of random forest in classifying data is 91%. We have taken 80% of data for training random forest and 20% for classification.

V. CONCLUSION

Fake profiles are created in social networks for various reasons by individuals or groups. The results are about detecting the account is fake or genuine by using engineered features and trained using machine learning models like neural networks and random forest. The predictions indicate that the algorithm neural network produced 93% accuracy. In the future, there is a hope that new features make to detect and identify easily like implementing skin detection can be done by using natural language processing techniques more accurate. When Facebook introduces new features then it will be easy to identify fake accounts easily.

VI. FUTURE WORK

Main problem is that a person can have multiple Facebook accounts which makes them an advantage of creating fake profiles and accounts in online social networks.

The idea is of attaching Aadhar card number when signing up an account so that we can restrict to create a single account and there is no chance of fake profiles at any moment.

REFERENCES

1. Sai Pooja, G., Rajarajeswari, P., Yamini Radha, V ., Navya Krishna.G., Naga Sri Ram.B., Recognition of fake currency note using convolutional neural networks(2016). International Journal of Innovative Technology and Exploring Engineering, 58-63,8(5).
2. Mohammed Ali Al-Garadi,Mohammad Rashid Hussain, Henry Friday Nweke, Ihsanali,Ghulammujtaba1,Harunachiro Ma,Hasanalikhattak,Andabdullahgani "Predicti-Ngcyber Bullying On Social Networks.
3. Yadongzhou,Daewookkim,Junjiezhang,(Member,Ieee),Lili Liu1, Huanjin3, "(IEEE)ProGuard: Detecting Malicious Accounts in Social-Network-Based Online Promotions".
4. Mauro Conti University of Padua, Radha Poovendran University of Washington, Marco Secchiero University of Padua,"FakeBook: Detecting Fake Profiles in On- line Social Networks(2012)", ACM /IEEE International Conference on Advances in Social Networks Analysis and Mining.
5. ni .N., Smruthi.M.,"A Hybrid Scheme for Detecting FakeAccounts in Facebook" ISSN: 2277- 3878, (IJRTE)International Journal of Recent Technology and Engineering (2019) , Issue-5S3, Volume-7.
6. NarsimhaGugulothu,JayadevGyani, Srinivas Rao Pulluri "A Comprehensive Model for Detecting Fake Profiles in Online Social Networks(2016)".
7. Dr.Narsimha.G, Dr.JayadevGyani, P. Srinivas Rao ,"Fake Profiles Identification in Online Social Networks Using Machine Learning and NLP(2018)", International Journal of Applied Engineering Research ISSN 0973-4562, Number 6, Volume 13.
8. Reddy, A. V. N., &Phanikrishna, C. Contour tracking based knowledge extraction and object recognition using deep learning neural networks(2016). Paper presented at the Proceedings on 2nd International Conference on Next Generation Computing Technologies in 2016, NGCT 2016, 352-354. doi:10.1109/NGCT.2016.7877440.
9. V. Rama Krishna,& K.Kanaka Durga. Automatic detection of illegitimate websites with mutual clustering.(2016) International Journal of Electrical and Computer Engineering, 6(3), 995-1001. doi:10.11591/ijece.v6i3.9878
10. D.Rajeswara Rao & V.Pellakuri. Training and development of artificial neural network models: Single layer feedforward and multi layer feedforward neural network(2016). Journal of Theoretical and Applied Information Technology, 150-156,84(2).
11. Challa, N.,. Pasupuleti, S. K, & Chandra, J. V. A practical approach to E-mail spam filters to protect data from advanced persistent threat.(2016) Paper presented at the Proceedings of IEEE International Conference on Circuit, Power and Computing Technologies, ICCPCT 2016, doi:10.1109/ICCPCT.2016.7530239.
12. D.Rajeswara Rao , & P.Vidyullatha. Machine learning techniques on multidimensional curve fitting data based on r_ square and chi_square methods(2016). International Journal of Electrical and Computer Engineering, . doi:10.11591/ijece.v6i3.91556(3), 974- 979.
13. K.Anand., &J.Kumar,Anomaly detection in online social network: A survey. Paper presented at the Proceedings of the(2017) International Conference on Inventive Communication and Computational Technologies,ICICCT 2017, 456-459. doi:10.1109/ICICCT.2017.7975239
14. Pradeepini, G., Patil, S. T. ,& Bangare, S. L (2017). Brain tumor classification using mixed method approach. Paper presented at the International Conference on Information Communication and Embedded Systems, ICICES, doi:10.1109/ICICES.2017.8070748
15. Jaya Lakshmi, R., &Subba Rao, G. V. A re- constructive algorithm to improve image recovery in compressed sensing. Journal of Theoretical and Applied Information Technology(2017), 95(20), 5443- 5453
16. N.Jayanthi, B.V.Babu., & N.S.Rao,. Survey on clinical prediction models for diabetes prediction. Journal of Big Data, 4(1) 2017 doi:10.1186/s40537-017-0082-7
17. .Pradeepini, G., Pradeepa, G., Tejanagasri, B., &Gorrepati, S. H. Data classification and personal care management system by machine learning approach. International Journal of Engineering and Technology,2018 (UAE), 7(2.32 Special Issue 32), 219-223. doi:10.22444/IBVS.6227_old
18. .Praveena, M., Asha Deepika, R., & Sai Raghavendhar, C. Analysis on prediction of heart disease using data mining techniques. Journal of Advanced Research in Dynamical and Control Systems(2018), 10(2), 126-136.
19. Satish Babu, J., Niveditha, M., Bhavya, V., &Gowthami, K. Data mining techniques for herbs. International Journal of Engineering and Technology,2018(UAE), 7(1.1 Special Issue 1), 406- 410.
20. Shinde, S. A., &Rajeswari, P. R. Intelligent health risk prediction systems using machine learning: A review. International Journal of Engineering and Technology,2018(UAE), 7(3), 1019-1023. doi:10.14419/ijet.v7i3.12654
21. Sucharitha, G., &Senapati, R. K. Local extreme edge binary patterns for face recognition and image retrieval(2018). Journal of Advanced Research in Dynamical and Control Systems_10, 644-654.
22. V.N.Mandhala, D.Bhattacharyya, &T.Kim . Face detection using image morphology - A review. International Journalof Security and its Applications,2016, 10(4), 89-94. doi:10.14257/ijsia.2016.10.4.10

AUTHORS FORFILE

Sk.Shama, working as Assistant Professor of Computer Science and Engineering Department at the Koneru LakshmaiahEducation Foundation situated at Vaddeswarm, received Bachelors Degree from JNTUK , PG from Acharya Nagarjuna University..

Sk. Wasim Akram, working as Assistant Professor of Computer Science and Engineering Department at Vasireddy Venkatadri Institute of Technology Guntur , received Bachelor's Degree from JNTUK PG from JNTUK, pursuing Ph. D from JNTUA.

K. Siva Nandini, from Computer Science and Engineering Department, Koneru Lakshmaiah Education Foundation student located at Vaddeswaram,Guntur.

P.Bhavya Anjali, from Computer Science and Engineering Department, Koneru Lakshmaiah Education Foundation student located at Vaddeswaram,Guntur.

K.Devi Manaswi, from Computer Science and Engineering Department, Koneru Lakshmaiah Education Foundation student located at Vaddeswaram, Guntur.