Annals of R.S.C.B., ISSN: 1583-6258, Vol. 25, Issue 5, 2021, Pages. 4170 - 4174
Received 25 April 2021; Accepted 08 May 2021.

Fraudulent Users Detection on Social Media Platforms Using Machine Learning Techniques

1Avik Kumar Ghosh, 2P. Visalakshi, 3Hiren J Doshi

SRM Institute of Science and Technology

1ag2989@srmist.edu.in, 2visalakp@srmist.edu.in, 3hj2865@srmist.edu.in

ABSTRACT

The social network, a crucial part of our life is plagued by online impersonation and fake accounts. In this project, a model is trained such that it could be used to classify a social media account as fake or genuine. This model uses Support Vector Machine and Deep Neural Network as a classification technique and can deal with a huge dataset of records immediately, taking out the need to assess each record physically. The primary concern to us here is Fake Accounts which are to be identified and so our problem can be said to be a classification or a clustering problem. As, this is an automatic detection method, it can be applied easily by online social networks which has millions of profiles, whose profiles cannot be examined manually.

Keywords **–** Classification, Fake User Detection

Introduction

- In the current period, online interpersonal organizations are the most well-known and quick data propagation applications.
- Eliminating counterfeit records has pulled in the consideration of numerous explores; consequently, broad investigates have been done on the ID of not normal records.
- The proposed work means to build up a framework dependent on information mining strategies that may assist with finding the phony record in web-based media networks.

As the process of detection of fake human accounts isvery difficult, these loop holes are widely abused. We believe that such forging of identity may be used for other purposes:

- We are not expecting that people will give their accurate details in the privacy policies of social media. An example of cyber-bullying is when someone blackmails or harasses a person for some specific reason.
- Peoples and groups who forge their identities on social media arestriving to spread chaos in our society.
- The process is developed to increase the popularity by making websites more gamified which helps in increasing the social rating sand popularity with more likes, followers and remarks.
- The creation of fake accounts is very easy now adays. In the present day, the creation of fake accounts is very easy and can be created easily by anyone..Now a day it is easier to buy Twitter and Instagram followers and likes online.

Researches have been conducted to find fake accounts created by bots. ML is used to identify botson social media but ML is also used to identify the purpose of the bot. We can recognize counterfeit accounts by different various methodologies by the likes neural network and support vector machine.

The paper presents a fake profile detection method using algorithms like deep neural networks and also support vector machine. The data is collected from Facebook Network. The feature set consists of

Annals of R.S.C.B., ISSN: 1583-6258, Vol. 25, Issue 5, 2021, Pages. 4170 - 4174
Received 25 April 2021; Accepted 08 May 2021.

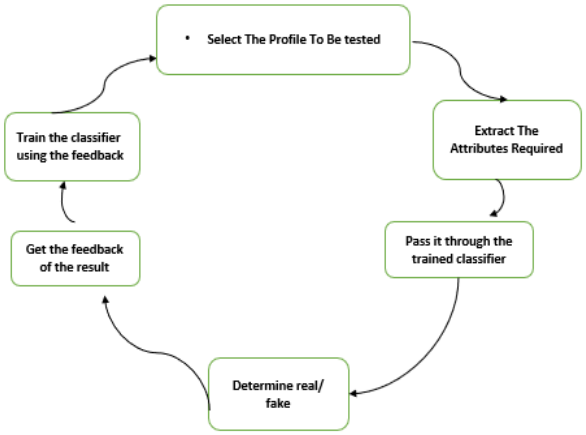
attributes like follower count, status count, friends count, language code and also gender and a few others.

I. RELATEDWORK

So we have done related work regarding the survey in which we found that one paper has done the fake profile detection model using the Naïve Baye's algorithm in which the major drawback was the accuracy which they got. and one more paper used the algorithm K means algorithm to detect the fake profiles. So the major problem with these algorithms is that it cannot handle large datasets and with large datasets the accuracy and effectiveness of the model gets reduced.

So that's why it is not the perfect model to be used. One more related work was with the paper which has done or created a privacy system for the vulnerable users but in that case too the drawbacks are majorly involving weak authentication system as there is no effective algorithm is used here to determine the vulnerable users.

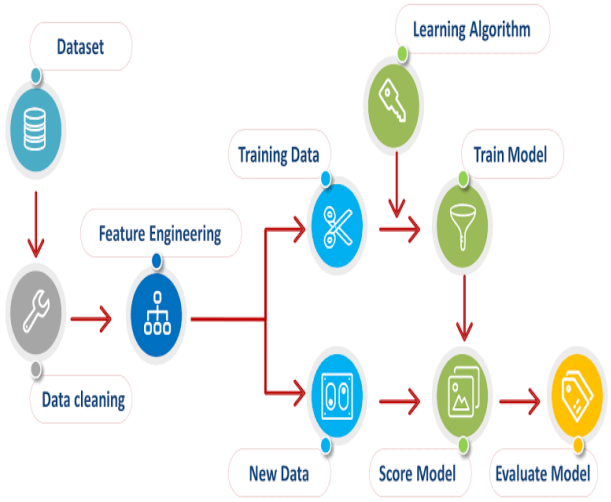
Machine Learning Model FlowDiagram



This is the major flow diagram which defines our model which we will be using in making to determine the fake profiles.

So the dataset mainly consists of all the Facebook data which is needed for the features extraction. These are basically the user id, the actual name of user, the screen name, status count, followers count, the account creation date, time zone ,favorite count and friend list count. So these are the main features and attributes which will be used to train the model.

Annals of R.S.C.B., ISSN: 1583-6258, Vol. 25, Issue 5, 2021, Pages. 4170 - 4174
Received 25 April 2021; Accepted 08 May 2021.



So as we can see the dataset is made and then the features we got earlier and all of them will be fed to the data which will be trained and then after the process is done , the new data will be fed to the model to get the results and the accuracy of the model as whether its working well or not working well. The algorithm which we will mainly use will be deep neural networks and support vector machine and then we will compare both of them to determine which is the most accurate.

The dataset which is being used contains 2818 rows and 15 columns or the attributes which are required to train the model and then test the model by providing relevant data.

The dataset primarily consists of various attributes which helps to determine whether the profile is fake or not. Some of the attributes which are present in the dataset are id, name, followers count, statuses count, friends count, favorites count, profile image, date of creation of the account, time-zone, location, Verified, language and many more.

Dataset statistics

Number of variables	9
Number of observations	2818
Missing cells	1749
Missing cells (%)	6.9%
Duplicate rows	26
Duplicate rows (%)	0.9%
Total size in memory	187.3 KiB
Average record size in memory	68.0 B

The above table gives more data and insights about the dataset that is being used to train and test the model. It includes information of the number of observations, number of duplicate rows, the percentage of duplicate rows, the total size it occupies in the memory and so on.

The dataset which is being used has gone through various processes before using like Data Cleansing. In this the data which is present in the dataset is cleaned by trying to

Annals of R.S.C.B., ISSN: 1583-6258, Vol. 25, Issue 5, 2021, Pages. 4170 - 4174
Received 25 April 2021; Accepted 08 May 2021.

remove the redundant values or the duplicate values and also fixing the fields which have null values.

Now when there is a much more clean dataset and it is ready to go through the training process.

II. FINDING FAKEACCOUNTS

In order to find a fake account created by human, first we should provide our machine learning model with fake accountcreatedbyussothatthealgorithmmayknowwhat a fake accountis.

1Order begins from the choice of profile that should be characterized.

2. When the profile is chosen, the helpful highlights are separated with the end goal of grouping.

3. The separated highlights are then taken care of to prepared classifier.

4. Classifier is prepared consistently as new information is taken care of into the classifier.

5. Classifier at that point decides if the profile is authentic or counterfeit.

6. The aftereffect of characterization calculation which will be profound neural organization and furthermore support vector machine is then confirmed and criticism is taken care of once again into the classifier.

7. As the quantity of preparing information builds the classifier turns out to be increasingly more precise in foreseeing the phony profiles.

Now we will get the results as we can see we get around 98% accuracy which is really efficient.

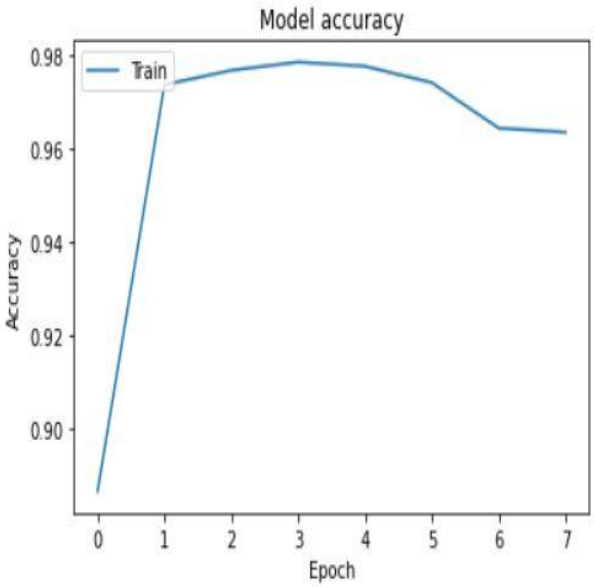
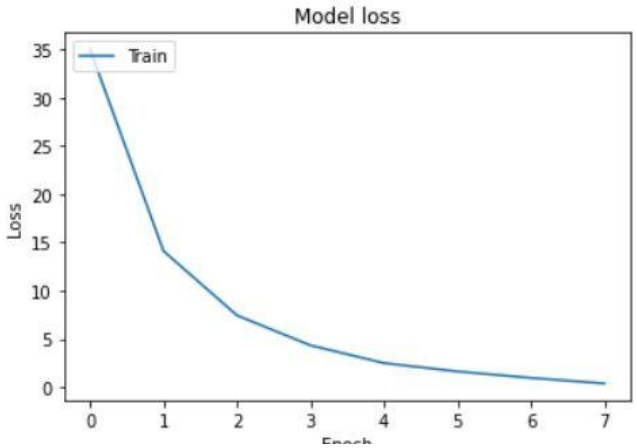


Fig: Model Accurate Output

Annals of R.S.C.B., ISSN: 1583-6258, Vol. 25, Issue 5, 2021, Pages. 4170 - 4174
Received 25 April 2021; Accepted 08 May 2021.



III. CONCLUSION

In the end, it can beconclude that fake profiles that are created on the social media either by humans or by bots for any purpose can be detected and identified. This is done by training the model by providing both the datasets which are both of the genuine users and the fake users. So, now when the actual dataset will be given to the model it can identify and detect the fake profiles from it very effectively.

IV. REFERENCES

[1] Matt. Social Media Comparison Infographic.2014.at https://leveragenewagemedia.com/blog/socialmediainfographic/.
[2]Social Media comparison infographic https:// leveragenewagemedia.com /blog /social-media-infographic/.
[3] Tumblr. URL:www.tumblr.com/
[4]Foursquare. URL:https://foursquare.com/.
[5] Al-Jarrah, O.Y.; Yoo, P.D.; Muhaidat, S.; Karagiannidis, G.K. and Taha, K. Efficient Machine Learning for Big Data: A Review, Big Data Research, 2 (2015), pp. 87–93.
[6] PengGao, Neil Zhenqiang Gong, SanjeevKulkarni, Kurt Thomas, and Prateek Mittal, “SybilFrame: A Defense-in-Depth Framework for Structure-Based Sybil Detection”. URL https://arxiv.org
[7] Wang, “Don’t follow me - spam detection in twitter,”in SECRYPT 2010, 2010. [15] G.
[8] G. Stringhini, C. Kruegel, and G. Vigna, “Detecting spammers on social networks,” in ACSAC, 2010.
[9] H. Yu, M. Kaminsky, P. B. Gibbons, and A. Flaxman.,