USING DEEP LEARNING MODEL AND ARTIFICIAL INTELLIGENCE (AI) FOR FAKE PROFILE IDENTIFICATION IN OSN NETWORK

B. SIVA JYOTHI [#1] , G.V.GAYATHRI [#2]

[#1] Assistant Professor, [#2] Assistant Professor,

Department of Computer Science & Engineering,

Anil Neerukonda Institute of Technology & Sciences,

Sanghivalasa, Visakhapatnam, AP 531162.

ABSTRACT

Now a days for identifying the fake profiles in online social network, it is very complicated task and one should keep more effort in order to identify them in accurate manner. There is a colossal expansion in innovations nowadays. Mobiles are getting more and more popular and there was a lot of innovation among online interpersonal organizations which has gotten a section in each one's life in making new companions and keeping companions, their inclinations are known simpler. As due to continuous usage of recent trends and technologies, every one is using these services for their day to day tasks and some are facing issues with some fake users who try to create fake profiles. Clients are taken care of with more pointless information during riding which are posted by counterfeit clients. Investigates have seen that 20% to 40% profiles in online social networks like facebook are fake profiles. Consequently this detection of fake profiles in online social networks results into arrangement utilizing structures.

Keywords:

Fake Profiles, Online Social Network, Facebook, Online Interpersonal Organizations, Innovation, Classification, Neural Network.

1. INTRODUCTION

Online social media is the spot where each and every individual has a standpoint at that point have the option to continue to associate their relations, move their updates, get together with individuals having same preferences. Online Social Networks utilizes front end advances, which grants permanency accounts as per to know one another. Facebook, Twitter are creating alongside people to keep up discussion along with all others. The online records invite individuals including indistinguishable interests all things considered who makes clients simpler after perform current companions. Gaming and engaging sites which have additional adherents accidentally that imply more fan base and incomparable evaluations. An appraisal drives online record holders to comprehend more up to date approaches not normally or physically to

contend more with their neighbors. By these analogies, the most extreme renowned up-and-comer in a political race usually gets more number of votes. Occurring of fake web-based media records and interests might be known. Example is fake online record being sold on-line at an online commercial centers for least cost, brought from cooperative working contributions.

As we all know that major people will try to use Twitter and Facebook media likes in on the web. Among several user profiles which are created in these famous networks we can see a lot of client records might be made by people or PCs like bots, cyborgs. Cyborg is half bot and half human record. These records are normally opened by human, yet their activities are made by bots. The another justification individuals to make counterfeit profiles for stigmatizing accounts they despise. This kind of clients make accounts with the username of individuals they disdain and post immaterial stories and previews on their records to divert everyone so they expect that specific individual is dreadful and make their standing low

Most assailants are in trying to create system crash or PC crash when we try to access such fake profiles in OSN networks. In some cases these bots who create fake profiles try to gain illegal profit in terms of collecting cash by appropriating undesirable promotions (spam) or catching records they can reuse or resale (phishing). Spammers assemble assets to know who are genuine clients and who can do genuine transactions and they try to gather their email ids, ip areas and figuring information power. All of these benefits can have an enormous cost related with them, and an attack, like any business experience, needs advantage to forge ahead. Aggressors all the more regularly use facebook logins, applications, Events, Group clients to assemble login accreditations, spam clients, and eventually acquire benefits. They need email records, treats, and a wide extent of IP conveys to circumvent reputation based insurances. In addition, they use phone numbers, assumed responsibility cards, and CAPTCHA game plans attempting to circumvent approval checks

Now a days OSN servers such as facebook and twitter are taking security advantages to accumulate clients to overcome spams and fishing accounts. Facebook Immune System does ceaseless personalities all assemble and every its action made by it. Social bot is a realized that stops and controls social online records. Bots socially is an auto produced programming. Précised way a social record copies depends upon at the online media, additionally as opposed to general bot, a social bot cooperating more in various clients that the social bot is a genuine man or lady. There was an seamless effort for designing several useful projects or semi created PC programs that copy the human data in Social media. So to utilize them programmers assault online informal communities. In this manner fundamentally utilized for crusade, promote and furthermore steal client non-public in more enormous scopes. The bot online expert accumulate inputs in light of aggressors

2. LITERATURE SURVEY

Literature survey is that the most vital step in the software development process. Before developing the new application or model, it's necessary to work out the time factor, economy, and company strength. Once all these factors are confirmed and got approval then we can start building the application. The literature survey is one that mainly deals with all the previous work which is done by several users and what are the advantages and limitations of those previous models. This literature survey is mainly used for identifying the list of resources to construct this proposed application.

MOTIVATION

Almost there are several records in online web-based media have loads of info information like name, sexual direction, partners, fans, inclinations, region numbers. Half pieces of this information are both of public and private. We need to utilize input that is public to know profiles which are fake for relational association as information from private is inaccessible. Regardless, in case our proposed plan is used by the relational association associations it, by then they can use the private information of the clients to know not from manhandling from security issues. Considered information is features for profiles to order of fake and authentic profiles. For identifying counterfeit profiles we followed these means:

1. Capacities are to be chosen after selection of characteristics, the dataset of profiles which are as of now delegated genuine or fake are needed for the tutoring rationale of the characterization calculation. We have utilized an openly accessible dataset of 1337 fake clients and 1481 real clients which incorporate various properties comprising of call, status tally, number of companions, fans depend, top picks, dialects respected, etc.

2. We try to choose the credit history factor which was extracted from the from user profile with the end goal of type.

3. After this the dataset of fake and real prepared documents are ready. From this dataset, 80% of both prepared documents (valid and imagine) are utilized to set up a tutoring dataset and 20% of the two profiles are utilized to assemble a testing dataset.

4. The tutoring dataset is then taken care of to the characterization set of rules. It gains from the training dataset and is anticipated to offer right style marks for the testing dataset.

5. The names from the testing dataset are killed and are left for assurance by the informed classifier.

6. The aftereffect of order calculation is appeared in 4.4. we have utilized two order calculations and have thought about the effectiveness of these calculations.

7. The proposed structure in the figure 1 shows the progression of methodology that ought to be sought after for tireless area of fake profiles with dynamic acquiring from the contribution of the result given by the game plan estimation

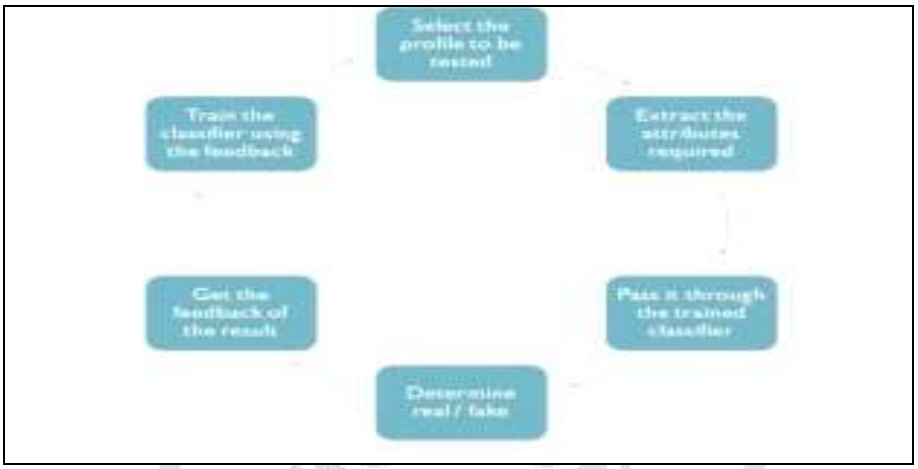


Figure 1. Represent the Fake Profile Detection Life Cycle

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 fake profiles.

3. EXISTING SYSTEM AND ITS LIMITATIONS

In the existing system, there was no concept for fake profile identification from OSN networks. Most aggressors are in it to bring in cash. They bring in cash by disseminating undesirable promotions (spam) or catching records they can reuse or resale (phishing). Spammers assemble assets to know phony and genuine clients, email ids, ip areas and processing information power. All of these benefits can have an enormous cost related with them, and an attack, like any business experience, needs advantage to progress forward. Aggressors all the more frequently use facebook logins, applications, Events, Group clients to assemble

login certifications, spam clients, and eventually acquire benefits. They need email records, treats, and a wide extent of IP conveys to circumvent reputation based securities. Also, they use phone numbers, assumed responsibility cards, and CAPTCHA game plans attempting to circumvent approval checks

LIMITATIONS OF THE EXISTING SYSTEM

1. All the existing schemes are limited to the identify fake profiles under manual approach.
2. All the existing systems are failed to identify the fake profiles using CNN models.
3. The attacker who create some fake profile page in OSN can easily gather illegal access of genuine users by using some advertisement, promotions and so on.
4. There is no accurate model to classify the fake profiles and normal profiles easily in the primitive methods.
5. This is very time complexity approach to categorize the fake profiles and normal profiles.

4. PROPOSED SYSTEM AND ITS ADVANTAGES

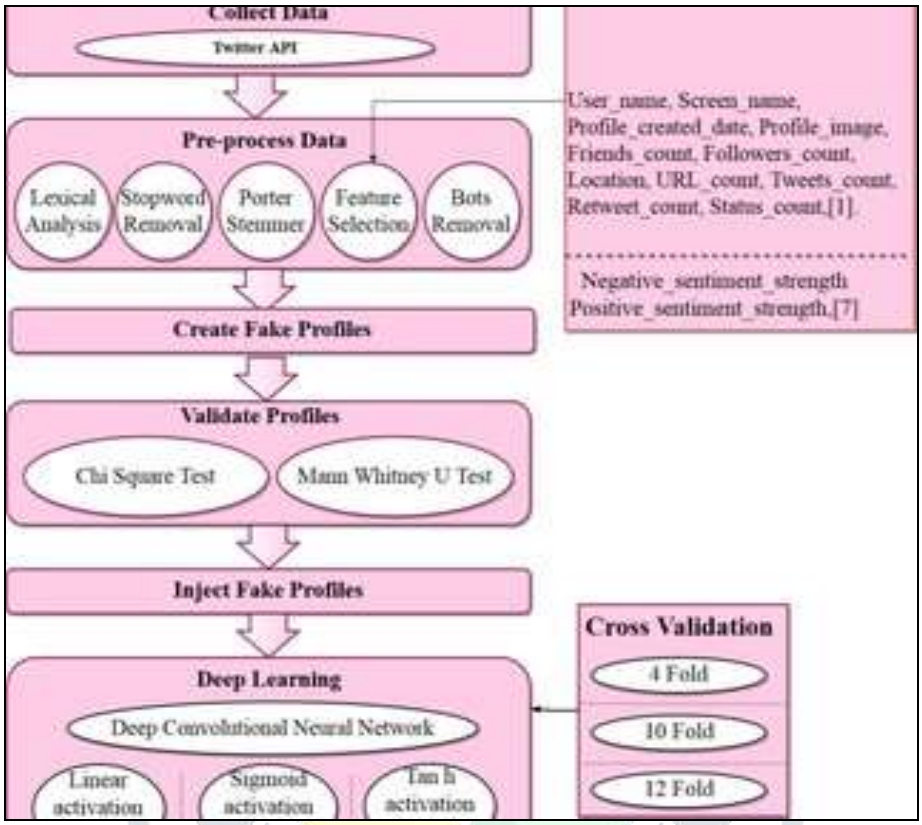
In this proposed work we try to design an application using AI and CNN model which can be used for prediction of fake profiles and normal profiles very easily from online social network. More auto produced projects or semi created PC programs that copy the human conduct in Social media. In the proposed system we try to figure out the assault who tries to utilize fake profiles in OSN network and steal the client non-public in more enormous scopes. The proposed CNN model can able to verify the identity of the user multiple times and then check if there is any assault program present in that profile creation. If there is any such program present immediately that will be identified as fake profile and corresponding user is identified as fake user. Here we develop a CNN model which can use to classify the profiles very accurately and design an automated method to predict the individual profiles. Cyborg bots transport assembles irregular clients. On the off chance that an individual recognize the solicitation from client, boat to get solicitation of the record who concur demand, will build prevalence value due to ways of life of common companions.

ADVANTAGES OF THE PROPOSED SYSTEM

1. The proposed scheme is very accurate in classification of Fake profiles and Normal profile from OSN network.
2. The proposed system gives accurate results.
3. The proposed system is capable of classification of bots who try to construct assault programs to gather the valuable information of genuine clients.

5. USING DEEP LEARNING MODEL AND ARTIFICIAL INTELLIGENCE (AI) FOR FAKE PROFILE IDENTIFICATION

In this section we try to discuss about proposed CNN and AI combine called as ANN model which is used to detect fake profiles which are present in the OSN network.



Execution is a strategy of ordering an item into a specific class dependent on the preparation informational index that was utilized to prepare the classifier. We feed the classifier with informational index so we can prepare it to recognize related items with as best precision as could be expected. Classifier is a calculation utilized for arrangement. In this undertaking we have utilized two classifiers in particular Neural Networks and Support Vector Machines and have consequently thought about their efficiencies.

Neural Network: The ordinary strategy by which a PC works is that you give directions or calculations to the PC and it creates yield dependent on it. Yet, imagine a scenario in which you don't have the foggiest idea about the calculation to tackle an issue. Can your PC actually give arrangements? On the off chance that we utilize ordinary methods, the PC won't tackle the issue unless you give a few directions. Here comes the idea of Neural Networks. We can in any case tackle such an issue via preparing an organization as such our program will learn all alone and will give arrangement near a specific precision. The term Neural Networks was existed in 1943 yet couldn't be carried out then because of absence of innovation. Neural Networks learn as a visual cue. Neural Networks depend on natural neurons for example synapses and the manner in which data is prepared inside the mind. There are essentially two sorts of neural organizations:

1. Single layer.

2. Multi-facet

Random Forest Classification Technique:

This classifier orders assortment of choice trees to subset of arbitrarily created preparing set. At that point it expands the preferences from choice sub trees to know subclass of taking care of item for tests. Arbitrary timberland will produce NA missing qualities for ascribes increment precision for bigger arrangements of information. On the off chance that more number of braid, it doesn't permit 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%

6. IMPLEMENTATION

It is a stage where the theoretical design is converted into programatical manner and here we divide the application into number of modules. The application is designed in Python programming language and we try to gather OSN dataset which contains both fake and normal profiles from Kaggle repository. The following are the modules:

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.

Table.1 Description of attributes in data sets.

7. 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.

8. EXPERIMENTAL RESULTS

Implementation is a stage where the theoretical design is converted into a programmatic manner. In this proposed application we try to use PYTHON as a programming language in which Google Collaboratory or Jupiter Notebook as a working platform to process the current application.

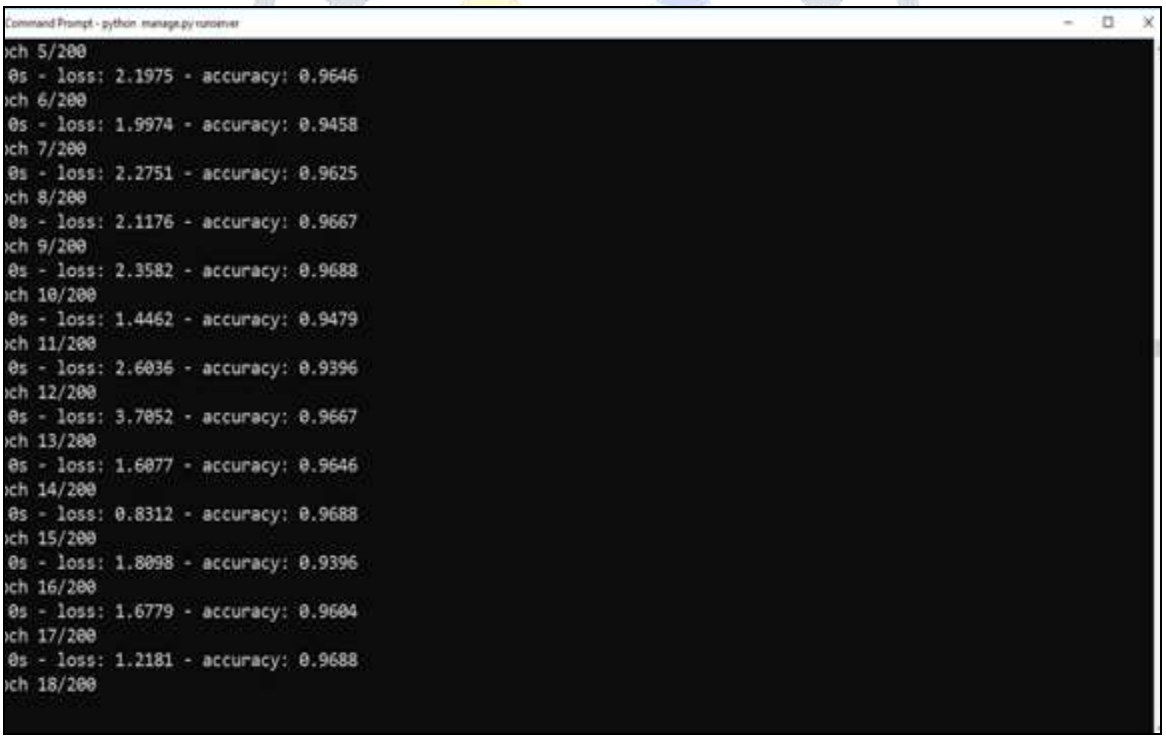
MAIN PAGE

In above screen click on 'ADMIN' link to get below login screen

In above screen enter admin and admin as username and password to login as admin. After login will get below screen.

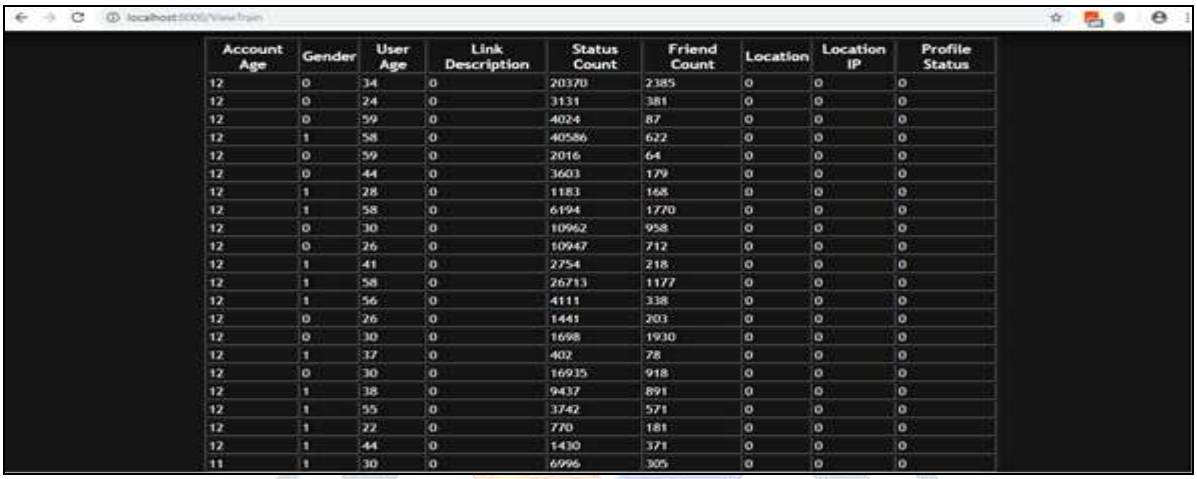
ADMIN MAIN PAGE

In above screen click on 'Generate ANN Train Model' to generate training model on dataset. After clicking on that link you can see server console to check ANN processing details with accuracy



In above black console we can see all ANN details.

In above screen we can see ANN got 98% accuracy to train all Facebook profile. Now click on 'View Ann Train Dataset' link to view all dataset details



Account Age	Gender	User Age	Link Description	Status Count	Friend Count	Location	Location IP	Profile Status
12	0	34	0	20370	2385	0	0	0
12	0	24	0	3131	381	0	0	0
12	0	59	0	4024	87	0	0	0
12	1	58	0	40586	622	0	0	0
12	0	59	0	2016	64	0	0	0
12	0	44	0	3603	179	0	0	0
12	1	28	0	1183	168	0	0	0
12	1	58	0	6194	1770	0	0	0
12	0	30	0	10962	958	0	0	0
12	0	26	0	10947	712	0	0	0
12	1	41	0	2754	218	0	0	0
12	1	58	0	26713	1177	0	0	0
12	1	56	0	4111	338	0	0	0
12	0	26	0	1441	203	0	0	0
12	0	30	0	1698	1930	0	0	0
12	1	37	0	402	78	0	0	0
12	0	30	0	16935	918	0	0	0
12	1	38	0	9437	891	0	0	0
12	1	55	0	3742	571	0	0	0
12	1	22	0	770	181	0	0	0
12	1	44	0	1430	371	0	0	0
11	1	30	0	6996	305	0	0	0

In above screen we can see all train data and scroll down to view all records. Now ANN train model is ready and you can logout and click on 'User' link to get below screen

USER ACCOUNT CREATION

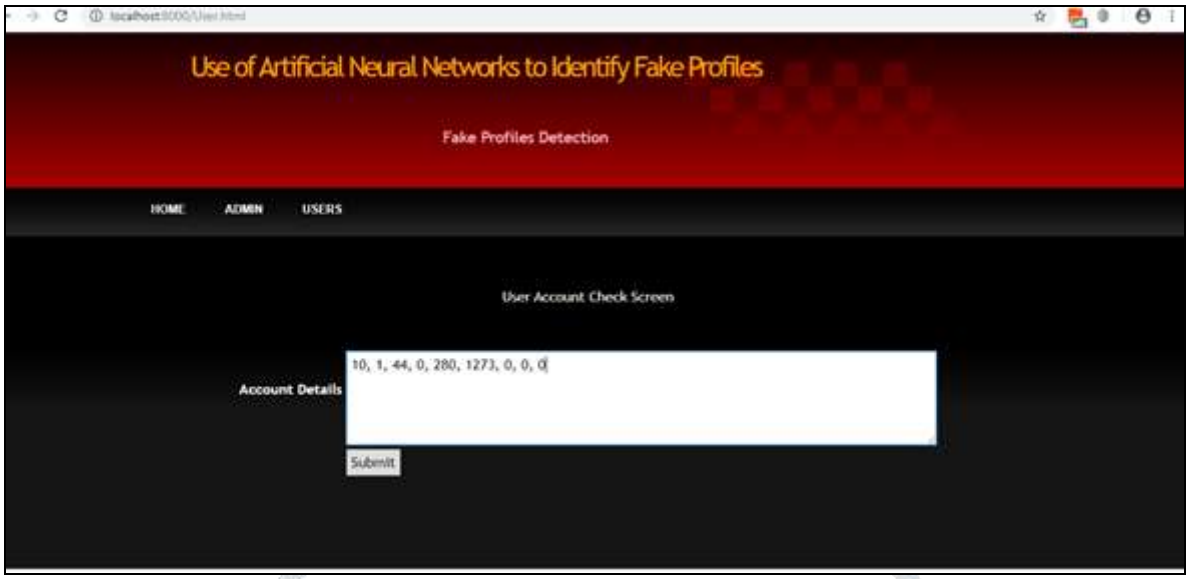
In above screen enter some test account details to get prediction/identification from ANN. You can use below records to check

10,1,44,0,280,1273,0, 0

10,0,54,0,5237,241,0,0

7,0,42,1,57,631,1,1

7,1,56,1,66,623,1,1



For above input will get below result

In above screen we can see the result predicted as genuine account

For above account details we got below result

In above screen we got result as fake for given account data

9. CONCLUSION

Counterfeit profiles are made in informal communities for different reasons by people or gatherings. The outcomes are tied in with identifying the record is phony or authentic by utilizing designed highlights and prepared utilizing AI models like neural organizations and arbitrary backwoods. The forecasts show that

the calculation neural organization created 93% exactness. Later on, there is an expectation that new highlights make to recognize and distinguish effectively like carrying out skin location should be possible by utilizing common language handling procedures more precise. At the point when Facebook presents new highlights then it will be not difficult to distinguish counterfeit records without any problem. VI. FUTURE WORK Main issue is that an individual can have various Facebook accounts which makes them a benefit of making counterfeit profiles and records in online informal communities. The thought is of connecting Aadhar card number when joining a record with the goal that we can limit to make a solitary record and there is zero chance of phony profiles at any second.

10. 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 SocialNetwork-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.