FAKE ACCOUNT DETECTION USING MACHINE LEARNING AND DATA SCIENCE

V. SARALA,DASETTY DURGA BHAVANI

Assistant Professor MCA,DEPT, Dantuluri Narayana Raju college, Bhimavaram, AndhraPradesh
Email id:-vedalasarala21@gmail.com
PG Student of MSc Computer Science, Dantuluri Narayana Raju College, Bhimavaram, AndhraPradesh
Email id:-durgabhavanidasetti2000@gmail.com

ABSTRACT

In this paper author uses Artificial Neural Networks to identify whether given social network account details are from genuine or fake users. ANN algorithm will be trained with all previous users fake and genuine account dataset and then whenever we give 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.

1. INTRODUCTION

Artificial Neural Networks (ANNs) can be used to identify fake profiles on social media platforms. ANNs are a type of machine learning algorithm that are inspired by the structure and function of the human brain. ANNs can learn to identify patterns in data, making them well-suited for tasks like identifying fake profiles. To use ANNs for identifying fake profiles, a dataset of profiles is needed that contains both real and fake profiles. This dataset is used to train the ANN on features that are indicative of fake profiles, such as the number of friends, the frequency of posts, and the types of content shared. Once the ANN is trained, it can be used to classify new profiles as either real or fake. The ANN takes in the features of a profile as input and outputs a binary classification (real or fake) One of the key advantages of using ANNs for identifying fake profiles is that they can learn to identify patterns that are not immediately obvious to humans. For example, ANNs can learn to identify subtle differences in the language used by fake profiles that are not easily detected by humans. However, it's important to note that ANNs are not perfect and may still make errors in identifying fake profiles. It's also important to ensure that the training dataset is representative of the population of profiles on the social media platform in question, to avoid biases in the classification of profiles. Additionally, ethical considerations should be taken into account when developing such systems, such as ensuring user privacy and preventing discrimination.

2. LITERATURE SURVEY AND RELATED WORK

In 2018, Yeh-Cheng chen and Shystunfelix Wu [1] have presented Fake Buster: A Robust fake Account detection by Activity Analysis. They proposed an innovative method to detect fake account in OSNs(Online Social Networks). It is develop for accurately detecting fake account among social network users, based on various activity collection and analysis. In this research they have use Random forest, along with C$.5 and Adaptive Boosting, with decision stump as a second classifier that created behind it to focus on the instance in the training data , in case the accuracy of the first classifier is less effective. After finish training, a cluster of features for each testing account will input into models and output a prediction with rank score indicating the likelihood of being fake account.

In 2019, Faiza Masood, Ghana Ammad, Ahmad Almogren, Assad Abbas, Hasan Ali Khathak, Ikram Uddin, MohsenGuizani, and MansourZuair [2] have presented in their work Spammer detection and fake user identification on social network. A review of techniques used for detecting spammers on Twitter. Spammers can be identified based on: (i) fake content, (ii) URL based spam detection, (iii) detecting spam in trending topics, and (iv) fake user identification. The proposed taxonomy of spammer detection on twitter is categorized into four main classes, namely,(i) fake content, (ii) URL based spam detection, (iii) detecting spam in trending topics, and (iv) fake user identification. The first category (fake content) includes various techniques, such as regression prediction model, malware alerting system, and Lfun scheme approach. In the second category (URL based spam detection), the spammer is identified in URL through different machine learning algorithms.

The third category (spam in trending topics) is identified through Naïve Bayes classifier and language model divergence. The last category (fake user identification) is based on detecting fake usersthrough hybrid techniques. Farhan, Muhammad Ibrohim, Indra Budi [3] have presented in their work Malicious Account Detection on Twitter based on Tweet Account features using Machine Learning. In this research, build a malicious account detection that can distinguish genuine accounts from malicious accounts using only tweet features of the accounts. Also managed to build a multiclass classification for the two types of malicious accounts,fake followers and spam bots using only tweet features. Lastly, found the best combination of algorithms, features, and data transformation scenario that suits best of our problem.

3. EXISTING SYSTEM

There are several existing systems that use artificial neural networks (ANNs) to identify fake profiles on social media platforms. Here are a few examples:

- **Social Forensics by Indiana University:** Social Forensics is a web-based system that uses ANNs to identify fake profiles on social media platforms like Twitter and Facebook. The system analyzes a range of

features including profile information, network structure, and activity patterns to determine the likelihood that a profile is fake.

- **Fakey by the University of Georgia:** Fakey is a browser extension that uses ANNs to identify fake profiles on Twitter. The extension analyzes the account's profile, tweets, and network structure to determine the likelihood that the account is fake.
- **FRAUDAR by Stanford University:** FRAUDAR is a system that uses ANNs to identify fake profiles on social media platforms. The system analyzes the network structure of the social media platform to identify clusters of profiles that are likely to be fake.
- **Deep Social by the University of Vermont:** Deep Social is a system that uses ANNs to identify fake profiles on Twitter. The system analyzes a range of features including profile information, activity patterns, and network structure to determine the likelihood that a profile is fake.

Overall, these systems demonstrate the potential of ANNs for identifying fake profiles on social media platforms. However, it's important to note that ANNs are not perfect and may still make errors in identifying fake profiles. Additionally, ethical considerations should be taken into account when developing and deploying such systems, including user privacy and preventing discrimination.

4. PROPOSED SYSTEM

A proposed system for identifying fake profiles on social media platforms using artificial neural networks (ANNs) could follow the following steps:

- **Data collection:** Collect a dataset of profiles from the social media platform in question. The dataset should contain both real and fake profiles.
- **Feature extraction:** Extract features from the profiles that are indicative of fake profiles, such as the number of friends, frequency of posts, and types of content shared.
- **Data preprocessing:** Normalize the extracted features to ensure they are on the same scale and remove any outliers or missing data.
- **Training the ANN:** Train an ANN on the extracted and preprocessed features using a supervised learning approach. The ANN should be designed to output a binary classification (real or fake) based on the input features.
- **Testing the ANN:** Test the ANN on a separate dataset of profiles to evaluate its performance in identifying fake profiles. This step can also be used to fine-tune the ANN parameters to improve its accuracy.
- **Deployment:** Deploy the trained ANN as a system to automatically identify fake profiles on the social media platform in question. The system should be regularly updated and maintained to ensure its continued accuracy.

Some additional considerations for the proposed system include ensuring user privacy and preventing discrimination. The system should be designed to protect user privacy and not collect any unnecessary personal information. Additionally, it's important to avoid any biases in the classification of profiles that could lead to discrimination against certain groups.

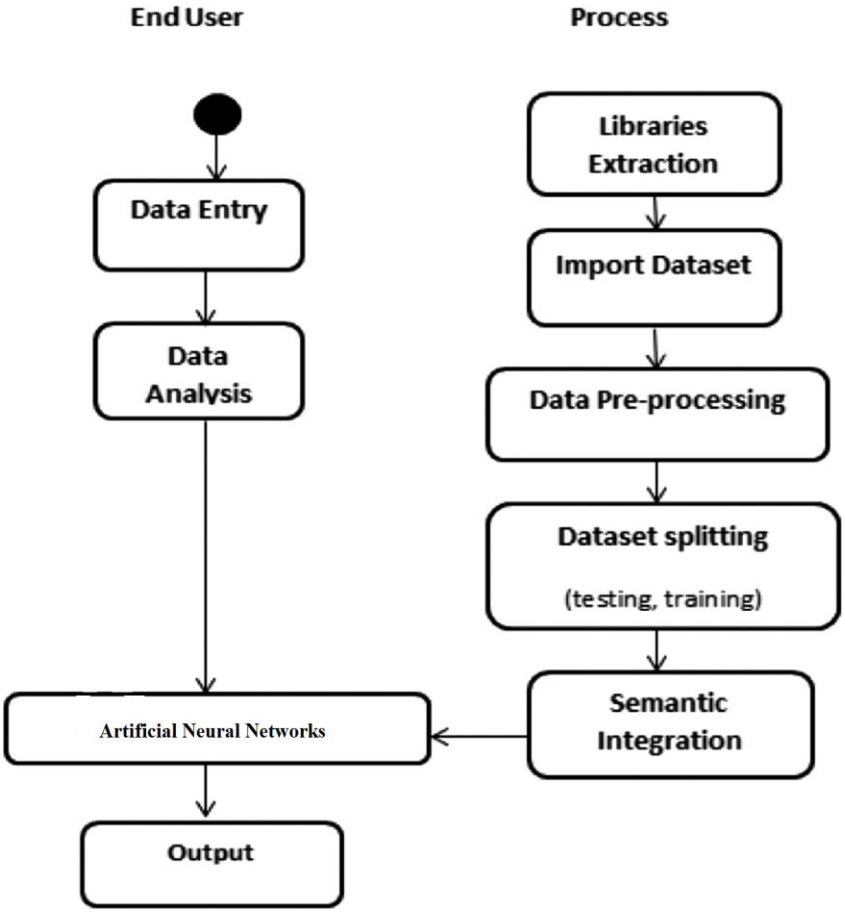

Fig 1: SYSTEM ARCHITECTURE

5. METHODOLOGIES

MODULES

1. Upload Social Network Profiles Dataset:

Using this module, we will upload dataset to application

2. Preprocess Dataset:

Using this module, we will apply processing technique such as removing missing values and then split dataset into

train and test where application use 80% dataset to train ANN and 20% dataset to test ANN prediction accuracy

3. Run ANN Algorithm:

Using this module, we will train ANN algorithm with train and test data and then train model will be generated and we can use this train model to predict fake accounts from new dataset.

4. ANN Accuracy & Loss Graph:

To train ANN model we are taking 200 epoch/iterations and then in graph we will plot /loss performance of ANN at each epoch/iteration.

5. Predict Fake/Genuine Profile using ANN:

using this module, we will upload new test data and then apply ANN train model to predict whether test data is genuine or fake.

6. RESULTS AND DISCUSSION SCREEN SHOTS

HOME SCREEN

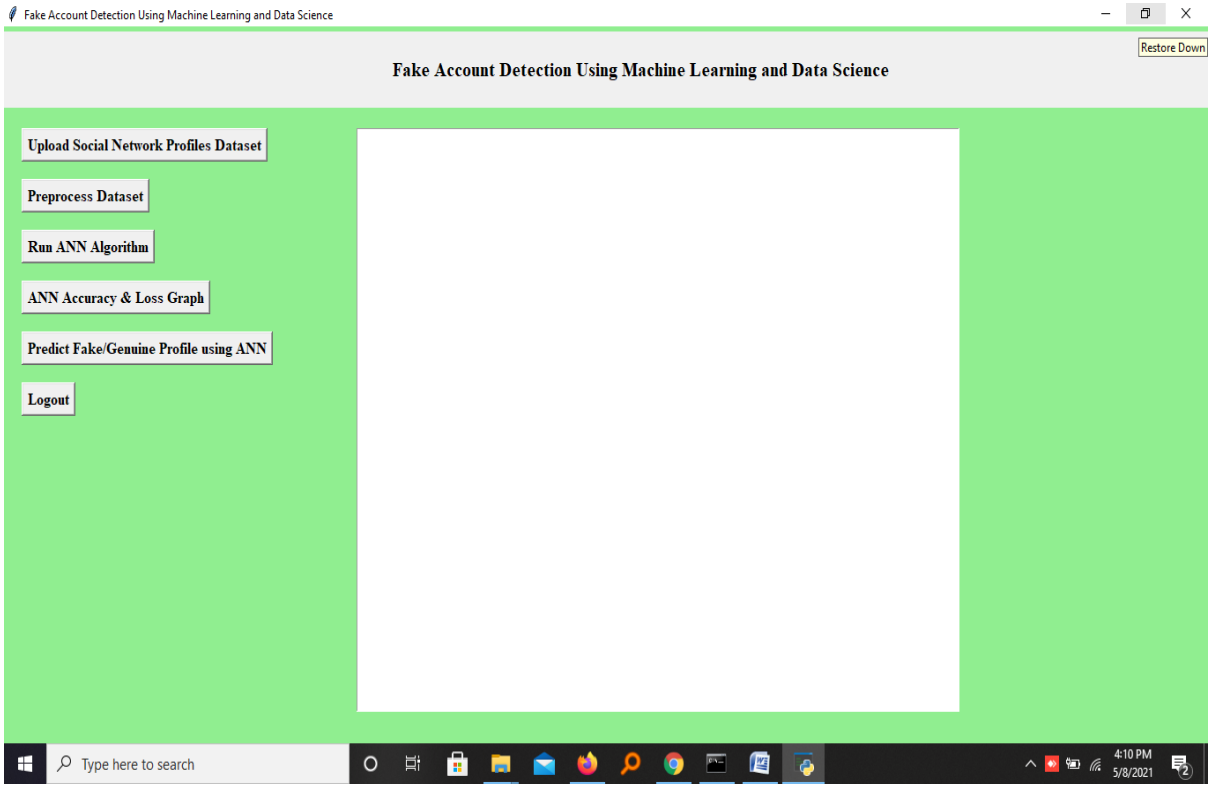

Fig 2:-In above screen click on 'Upload Social Network Profiles Dataset' button and upload dataset

UPLOAD DATASET

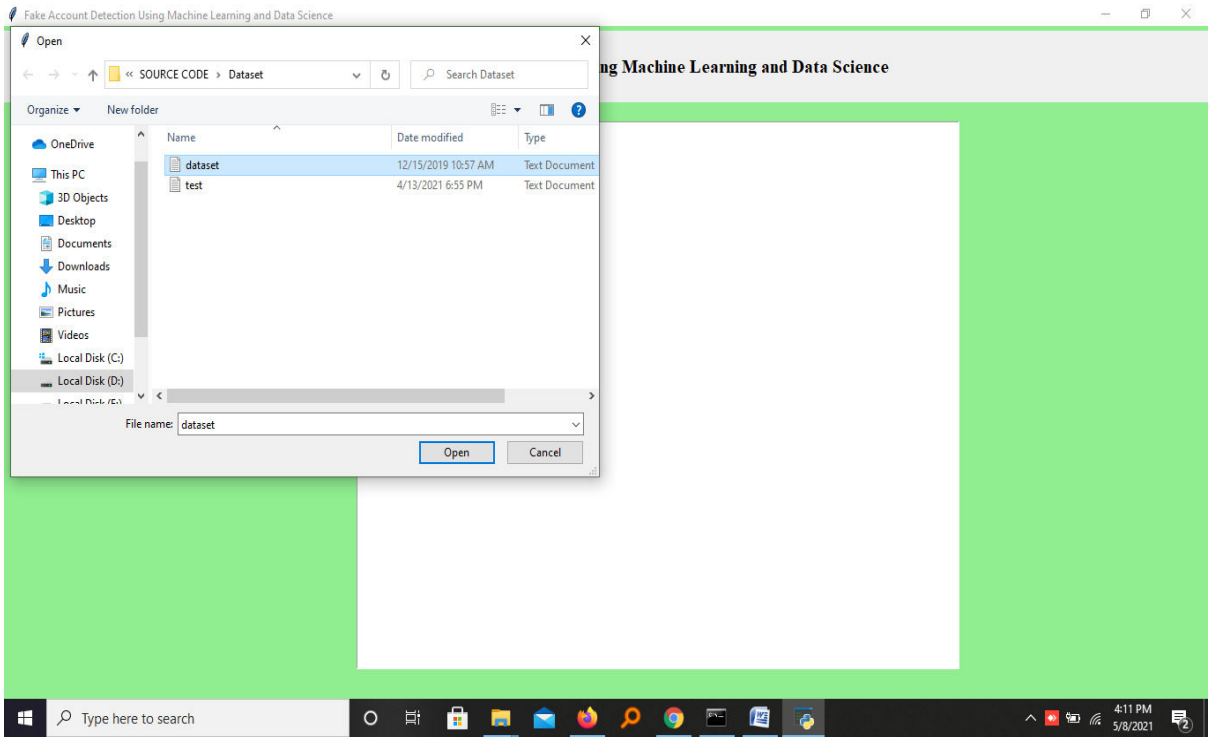

Fig 3:-In above screen selecting and uploading 'dataset.txt' file and then click on 'Open' button to load dataset and get below screen.

DATASET LOADED

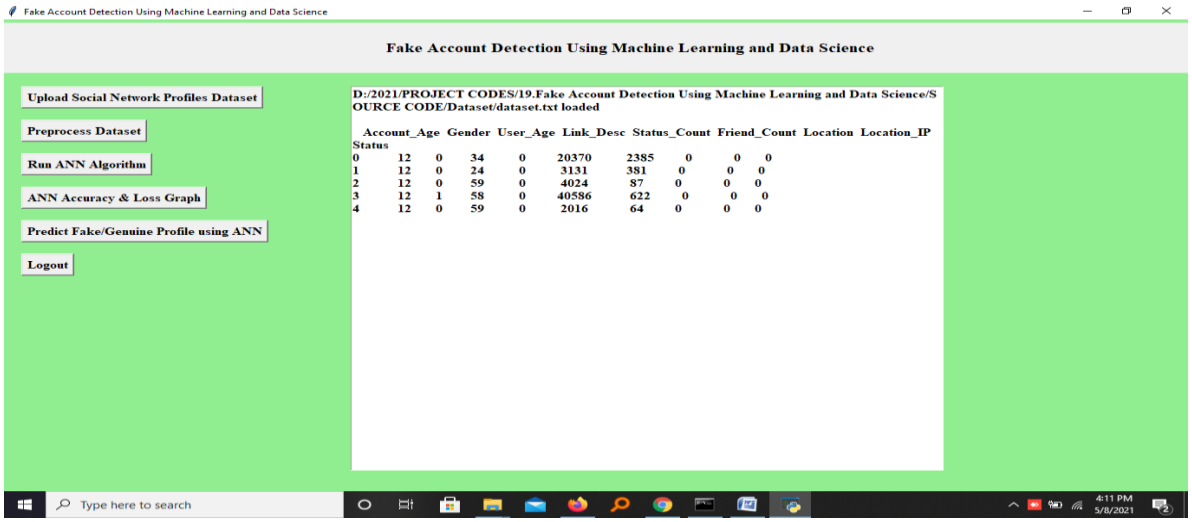

Fig 4:-In above screen dataset loaded and displaying few records from dataset and now click on 'Preprocess Dataset' button to remove missing values and to split dataset into train and test part.

PREPROCESSING DATA:

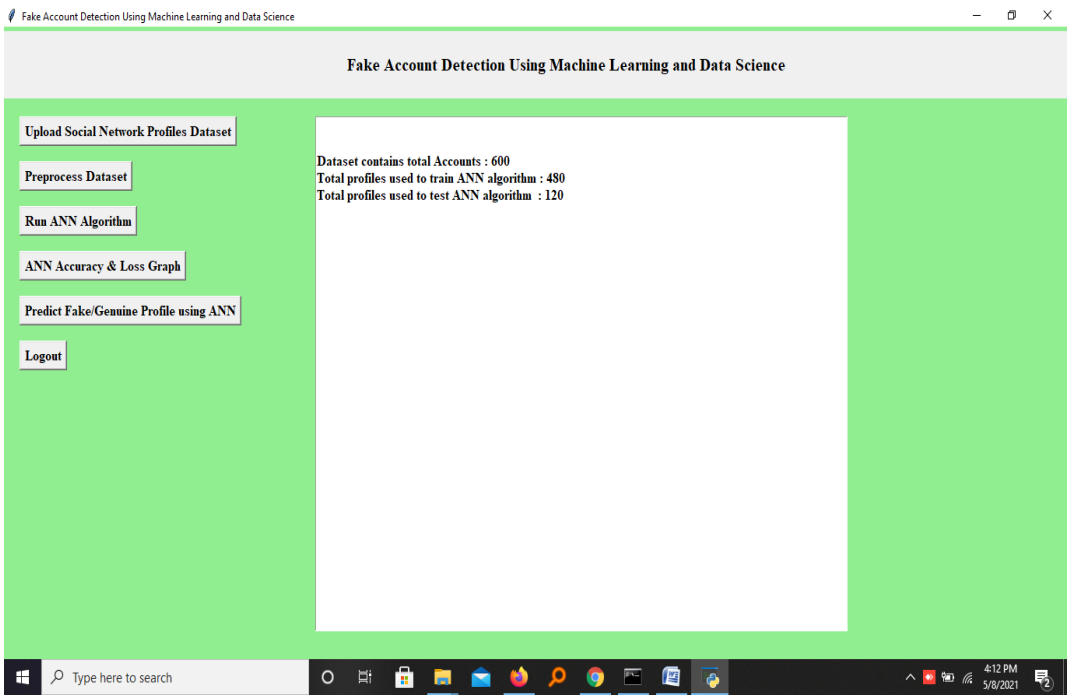

Fig 5:-In above screen we can see dataset contains total 600 records and application using 480 records for training and 120 records to test ANN and now dataset is ready and now click on 'Run ANN Algorithm' button to ANN algorithm.

5.3**.5 RUN ANN ALGORITHM:**

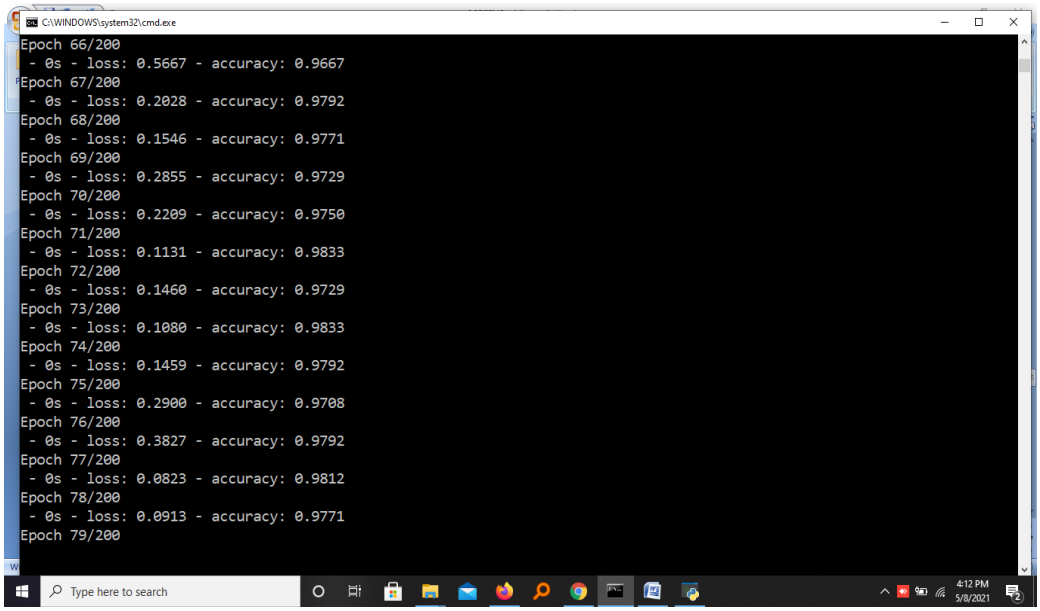

Fig 6:-In above screen we can see ANN start iterating model generation and at each increasing epoch we can see accuracy is getting increase and loss getting decrease.

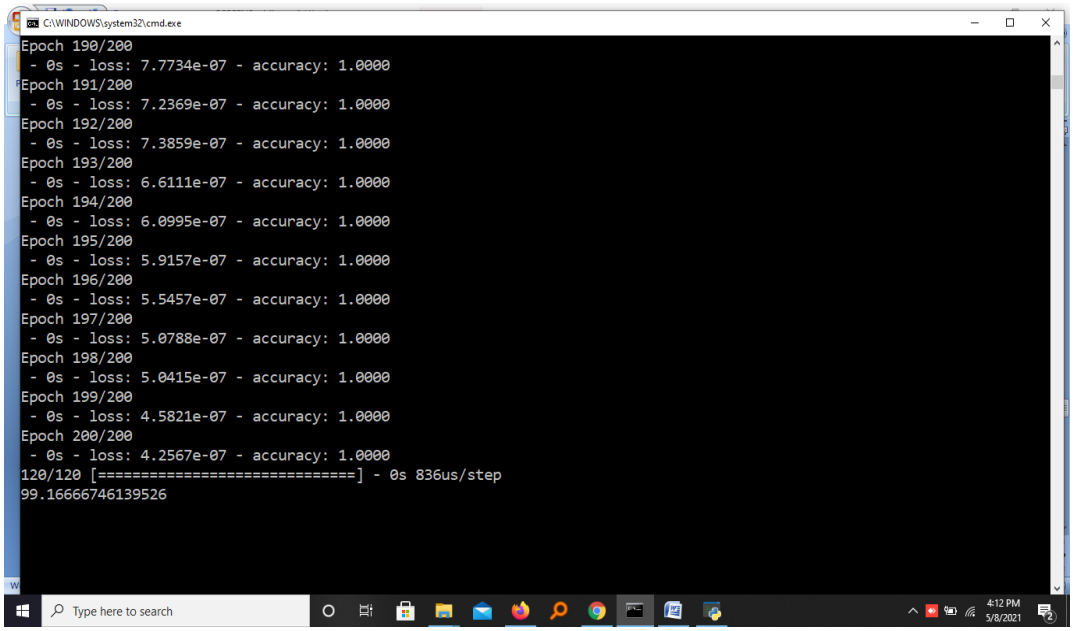

Fig 7:-EPOCH/ITERATION:

In above screen we can see after 200 epoch ANN got 100% accuracy and in below screen we can see final ANN accuracy.

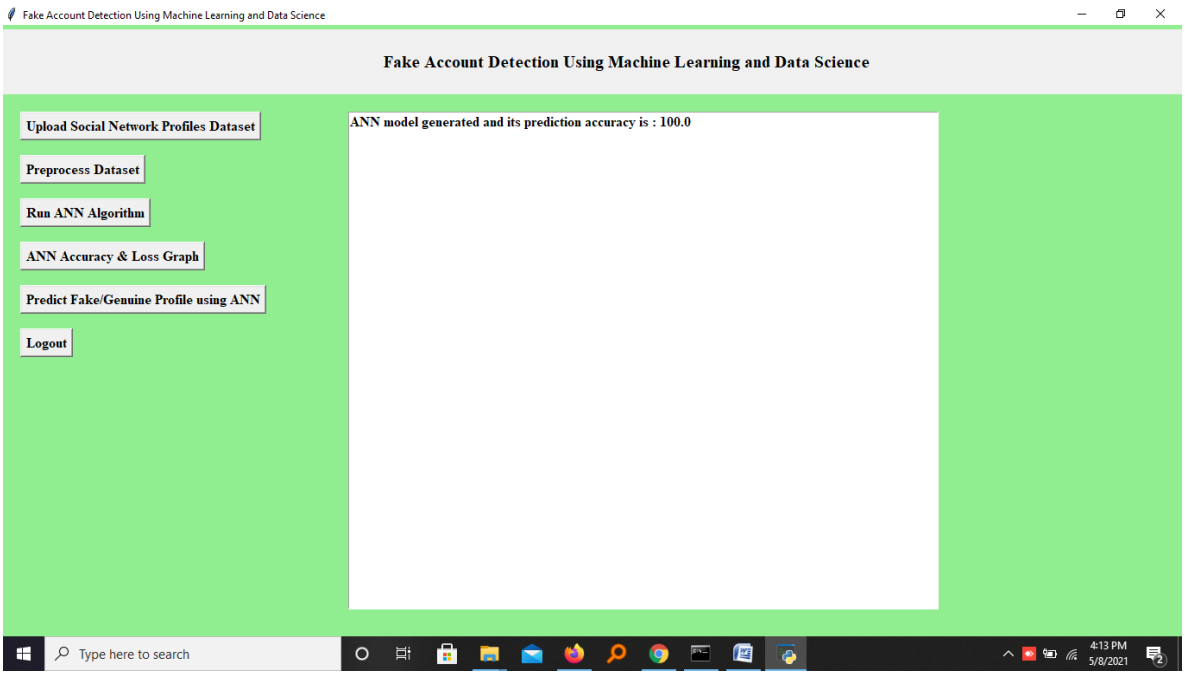

Fig 8:-PREDICTING ACCURACY:

in above screen ANN model generated and now click on 'ANN Accuracy & Loss Graph' button to get below graph.

GRAPH:

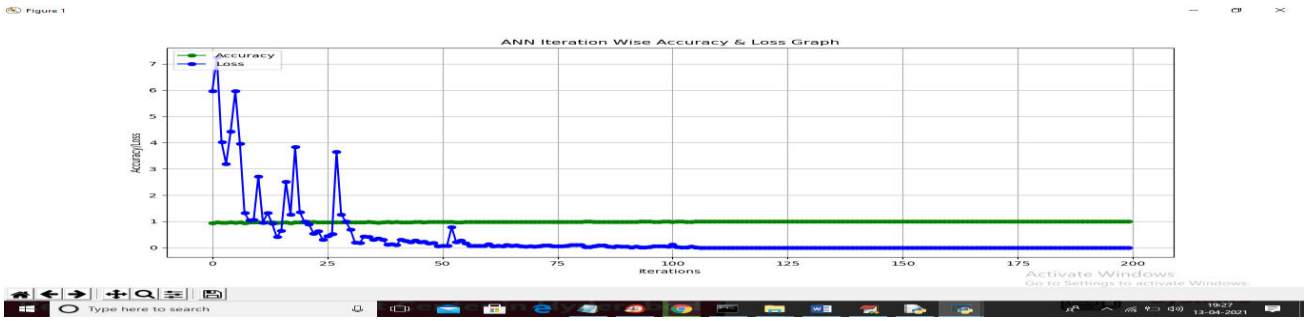

Fig 9:- In above graph x-axis represents epoch and y-axis represents accuracy/loss value and in above graph green line represents accuracy and blue line represents loss value and we can see accuracy was increase from 0.90 to 1 and loss value decrease from 7 to 0.1. Now model is ready and now click on 'Predict Fake/Genuine Profile using ANN' button to upload test data and then ANN will predict below result.

5.3.9 TEST DATA:

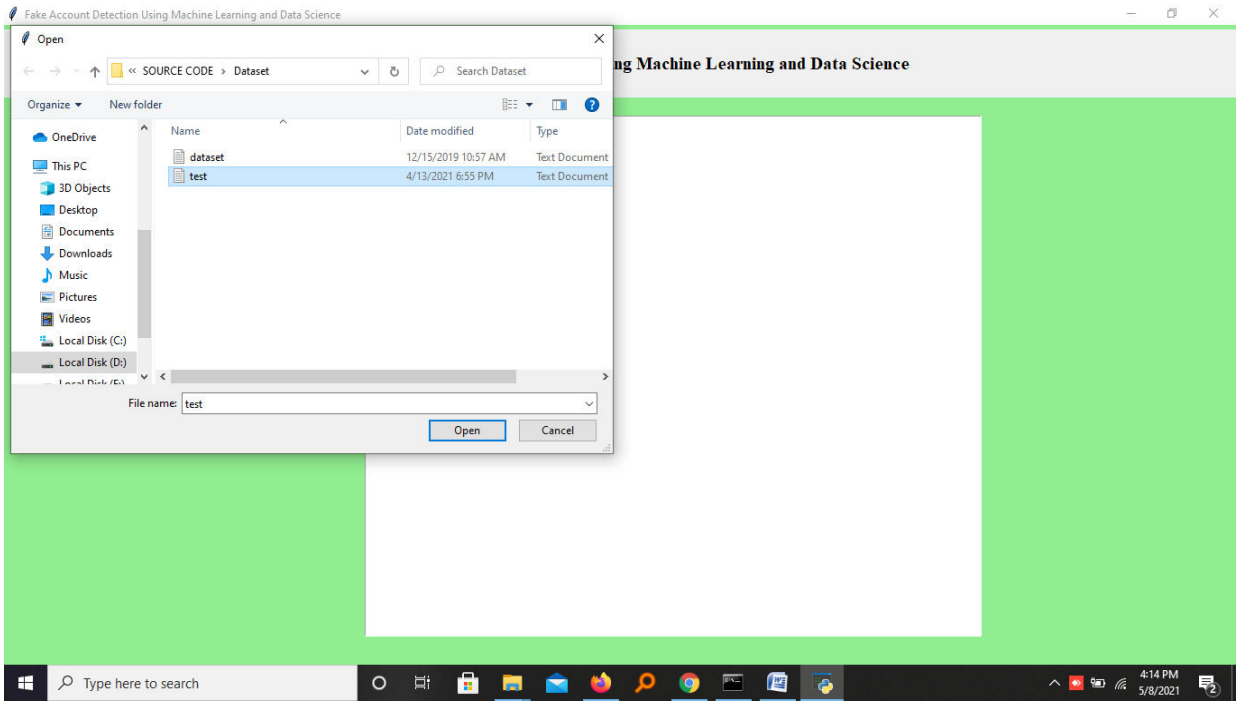

Fig 10 :-In above screen we are selecting and uploading 'test.txt' file and then click on 'Open' button to load test data and to get elow prediction result.

PREDICTING ACCOUNT AS FAKE OR GENUINE:

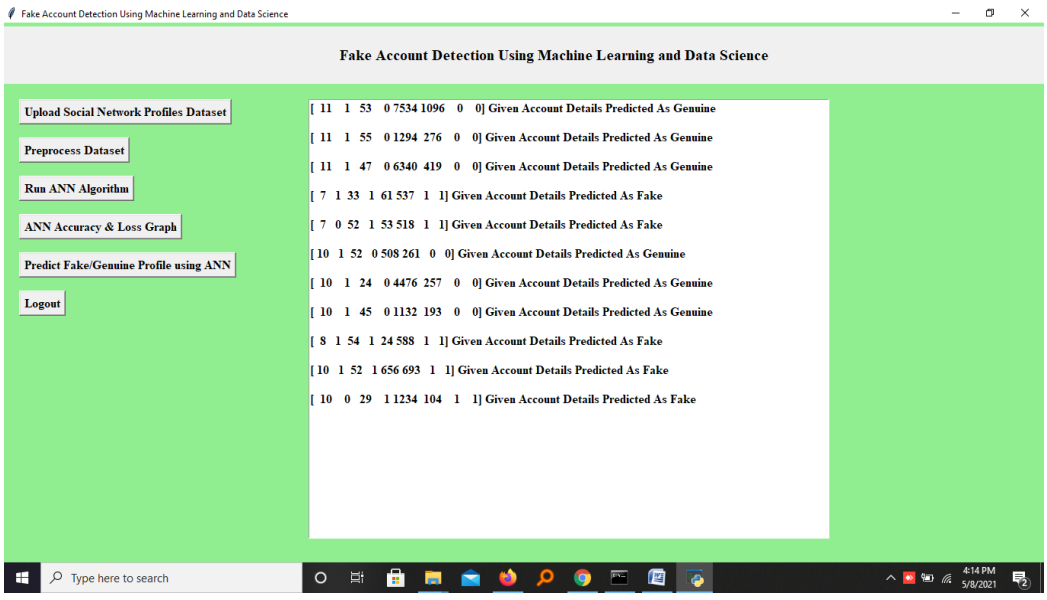

Fig 11:- In above screen in square bracket we can see uploaded test data and after square bracket we can see ANN prediction result as genuine or fake.

6. CONCLUSION AND FUTURE SCOPE

In this research, We have come up with an ingenious way to detect fake accounts on OSNs By using machine learning algorithms to its full extent, we have eliminated the need for manual prediction of a fake account, which needs a lot of human resources and is also a time-consuming process. Existing systems have become obsolete due to the advancement in the creation of fake accounts. The factors that the existing system relayed upon is unstable. In this research, we used stable factors such as engagement rate, artificial activity to increase the accuracy of the prediction

With ongoing research and development, these techniques can become more accurate and efficient, helping online platform to better protect users from fraudulent activities.

7. REFERENCES

1. ”Detection of Fake Twitter accounts with Machine Learning Algorithms” Ilhan aydin,Mehmet sevi, Mehmet umut salur.

2. ”Detection of fake profile in online social networks using Machine Learning” Naman singh, Tushar sharma, Abha Thakral, Tanupriya Choudhury.

3. ”Detecting Fake accounts on Social Media” Sarah Khaled, Neamat el tazi, Hoda M.O. Mokhtar.

4. ”Twitter fake account detection”, Buket Ersahin, Ozlem Aktas, Deniz kilinc, Ceyhun Akyol.

5. ” A new heuristic of the decision tree induction” ning li, li zhao, ai-xia chen, qing-wu meng, guo-fang zhang.

6. ” Statistical machine learning used in integrated anti-spam system” peng-fei zhang, yu-jie su, cong wang.

7. ” A study and application on machine learning of artificial intellligence” ming xue, changjun zhu.

8. ” learning-based road crack detection using gradient boost decision tree” peng sheng, li chen, jing tian.

9. ” Verifying the value and veracity of extreme gradient boosted decision trees on a variety of datasets” aditya gupta, kunal gusain, bhavya popli.

10. ” Fake account identification in social networks” loredana caruccio, domenico desiato, giuseppe polese.