Fake Account Detection Using Machine Learning and Data Science

In this paper author using 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 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.

To train ANN algorithm we are using below details from social networks

**Account\_Age, Gender, User\_Age, Link\_Desc, Status\_Count, Friend\_Count, Location, Location\_IP, Status**

All fake users main intention is to send friend request to normal users to hack their machine or to steal their data and never they will have many number of posts or have many following friends and their account age also will have less number of years. By analysing this features Facebook will mark whether user profile is fake or genuine. This Facebook profile data we downloaded from Facebook website and using this data to train ANN model. Below are some values from profile dataset.

**Account\_Age, Gender, User\_Age, Link\_Desc, Status\_Count, Friend\_Count, Location, Location\_IP, Status**

10,1,22,0,1073,237,0,0,0

10,0,33,0,127,152,0,0,0

10,1,46,0,1601,405,0,0,0

10,0,25,0,704,380,0,0,0

7,1,34,1,64,721,1,1,1

7,1,30,1,69,587,1,1,1

7,1,36,1,61,782,1,1,1

7,1,52,1,96,827,1,1,1

In above dataset all bold names are the dataset column names and all integer values are the dataset values. As ANN will not take string value so we convert gender values to 0 or 1, if male value is 1 and if female value is 0. In above dataset last column give us information of fake or genuine account if last column contains value 0 then account is genuine otherwise fake. All fake account will have less number of posts as their main intention is to send friend requests not posts, so by analysing this features Facebook mark that record with value 1 which means it’s a fake account. We are using above dataset to train ANN model and this dataset saved inside code ‘dataset’ folder. After building train model we input test data with account details and ANN will give result as fake or genuine. Below are some values from test data

**Account\_Age, Gender, User\_Age, Link\_Desc, Status\_Count, Friend\_Count, Location, Location\_IP**

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

In above test data STATUS column and its value is not there and ANN will predict status and give us result whether above test data is fake or genuine. In output we can see result of above test data as genuine or fake.

ANN algorithms Details

To demonstrate how to build a ANN neural network based image classifier, we shall build a 6 layer neural network that will identify and separate one image from other. This network that we shall build is a very small network that we can run on a CPU as well. Traditional neural networks that are very good at doing image classification have many more parameters and take a lot of time if trained on normal CPU. However, our objective is to show how to build a real-world convolutional neural network using TENSORFLOW.

Neural Networks are essentially mathematical models to solve an optimization problem. They are made of neurons, the basic computation unit of neural networks. A neuron takes an input (say x), do some computation on it (say: multiply it with a variable w and adds another variable b) to produce a value (say; z= wx+b). This value is passed to a non-linear function called activation function (f) to produce the final output(activation) of a neuron. There are many kinds of activation functions. One of the popular activation function is Sigmoid. The neuron which uses sigmoid function as an activation function will be called sigmoid neuron. Depending on the activation functions, neurons are named and there are many kinds of them like RELU, TanH.

If you stack neurons in a single line, it’s called a layer; which is the next building block of neural networks. See below image with layers



To predict class label multiple layers operate on each other to get best match layer and this process continues till no more improvement left.

Module Details:

Upload Social Network Profiles Dataset: Using this module we will upload dataset to application

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

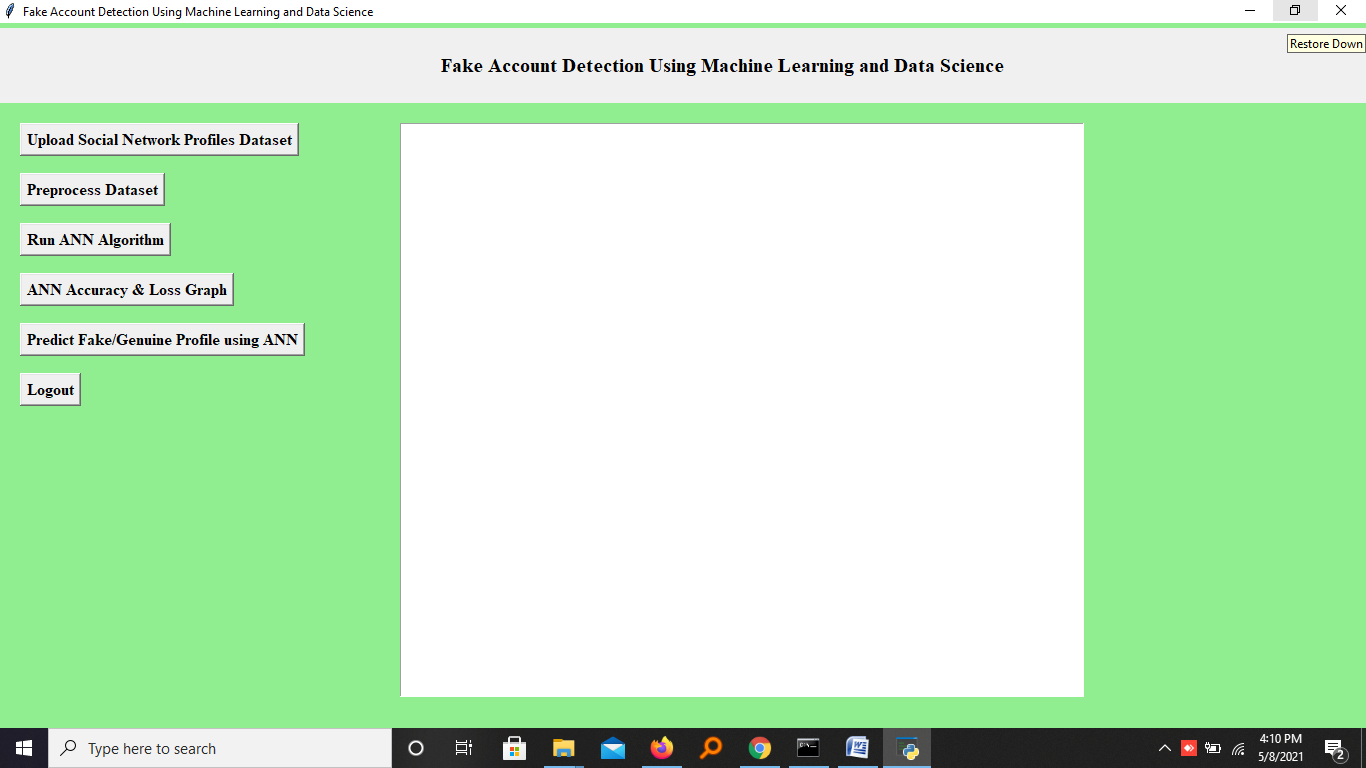
Run ANN Algorithm: Using this module we will train ANN algorithm with train and test data and then train model will be generated and we can use this train model to predict fake accounts from new dataset.

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

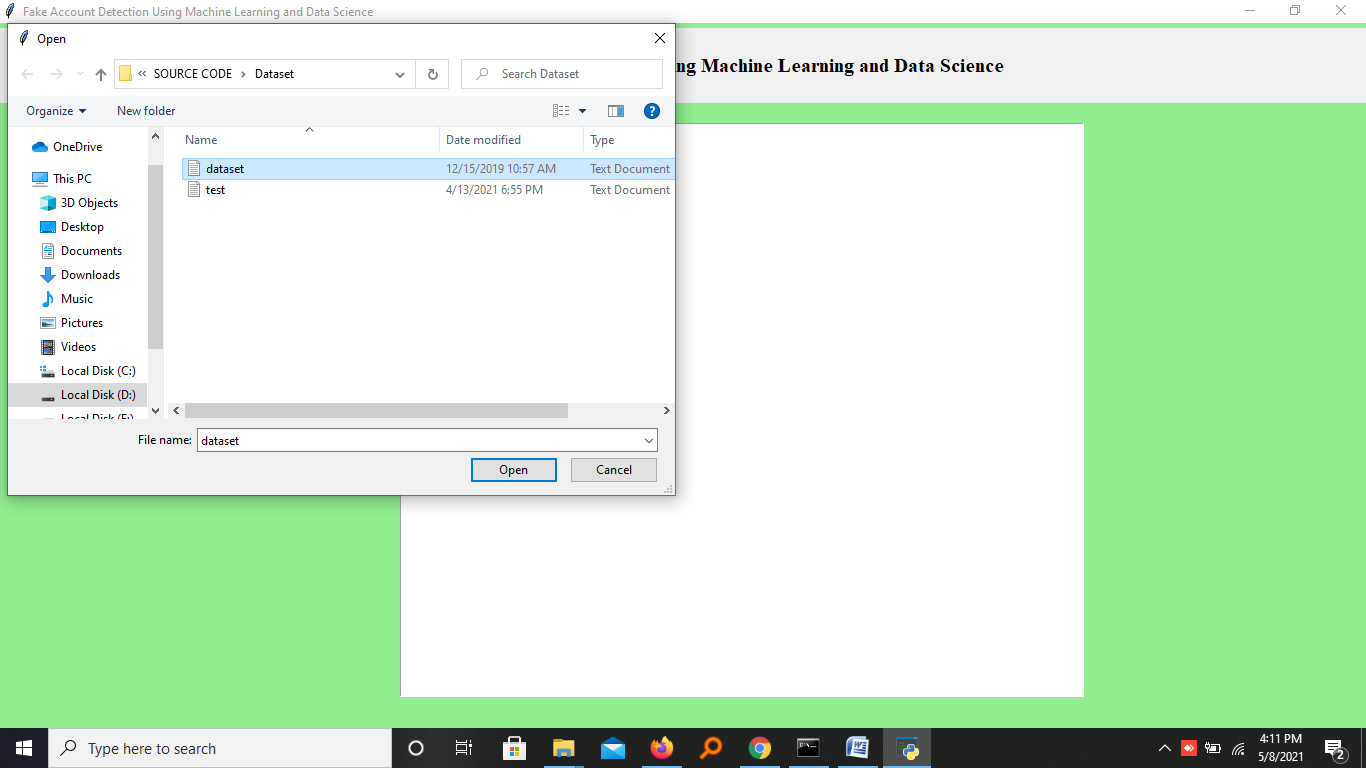
Predict Fake/Genuine Profile using ANN: using this module we will upload new test data and then apply ANN train model to predict whether test data is genuine or fake.

SCREEN SHOTS

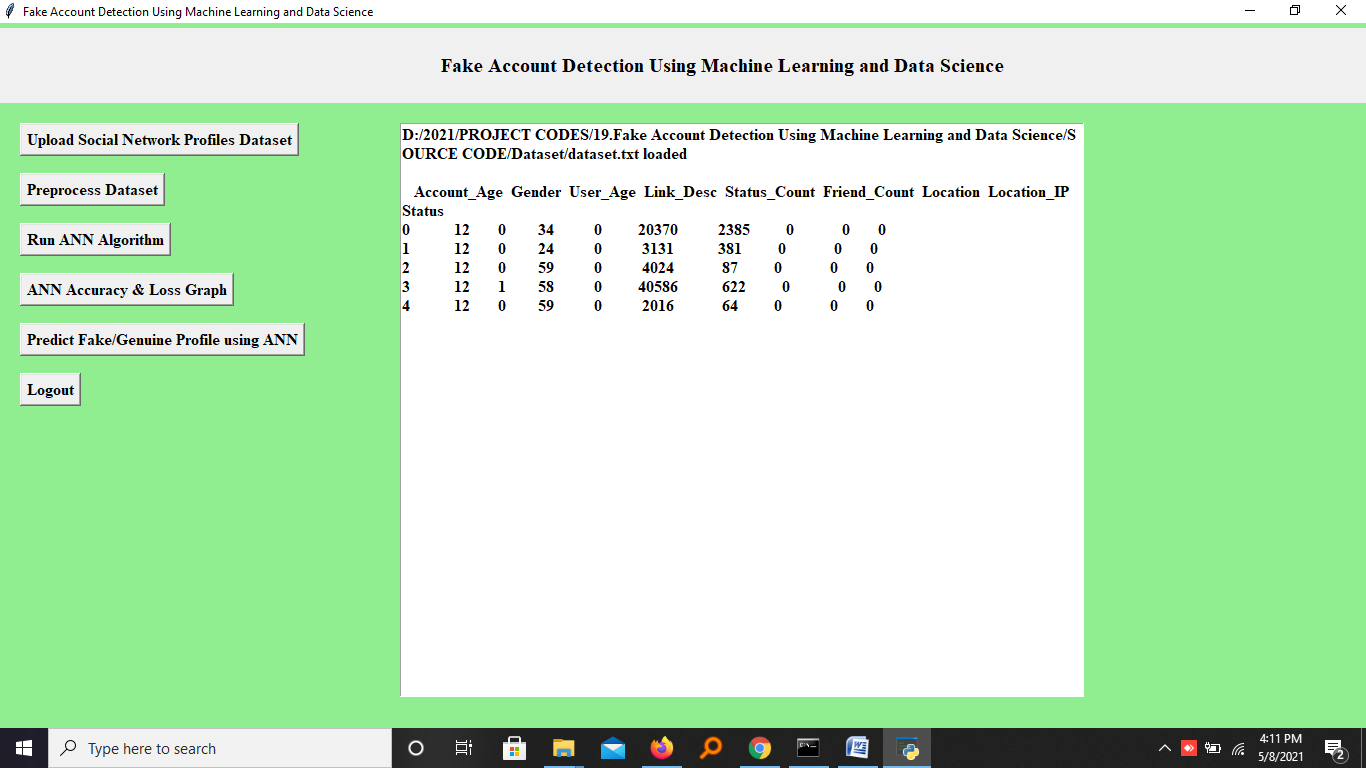
To run project double click on ‘run.bat’ file to get below screen



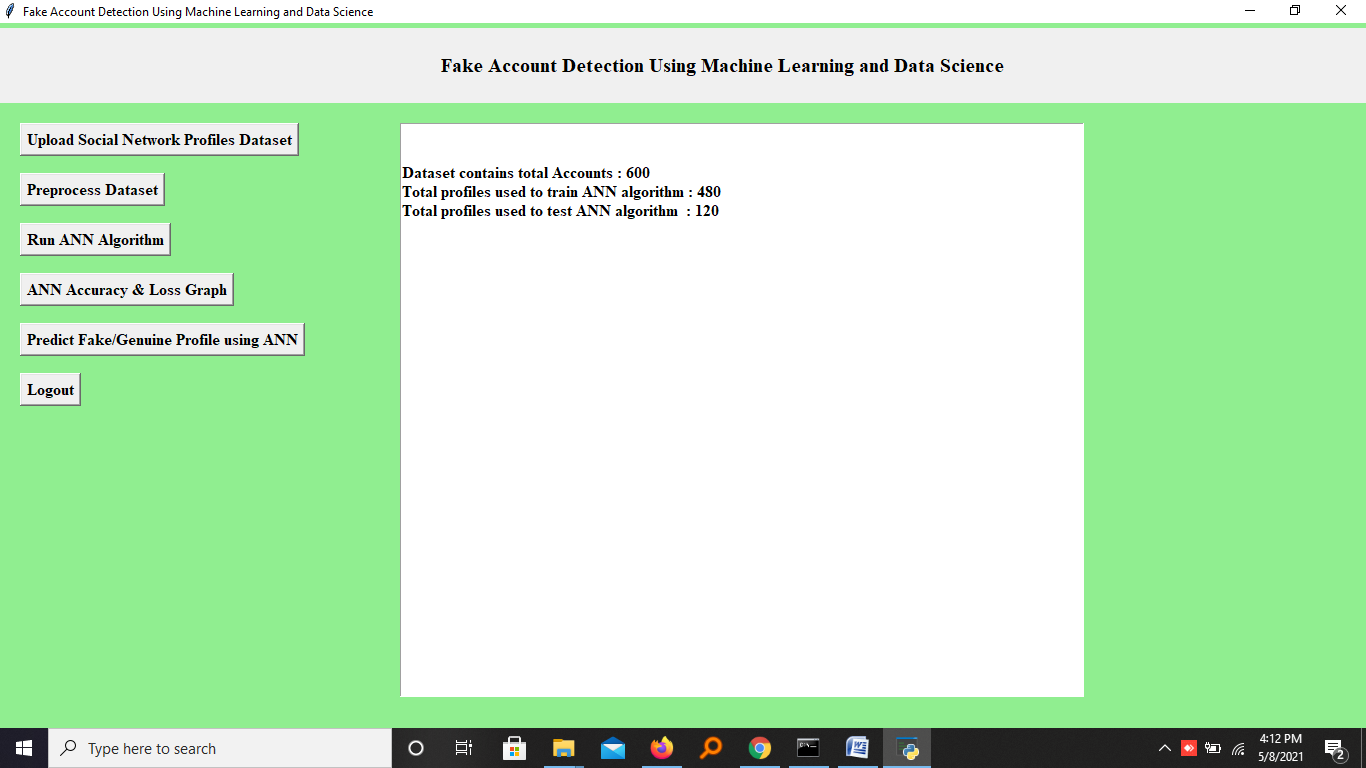
In above screen click on ‘Upload Social Network Profiles Dataset’ button and upload dataset



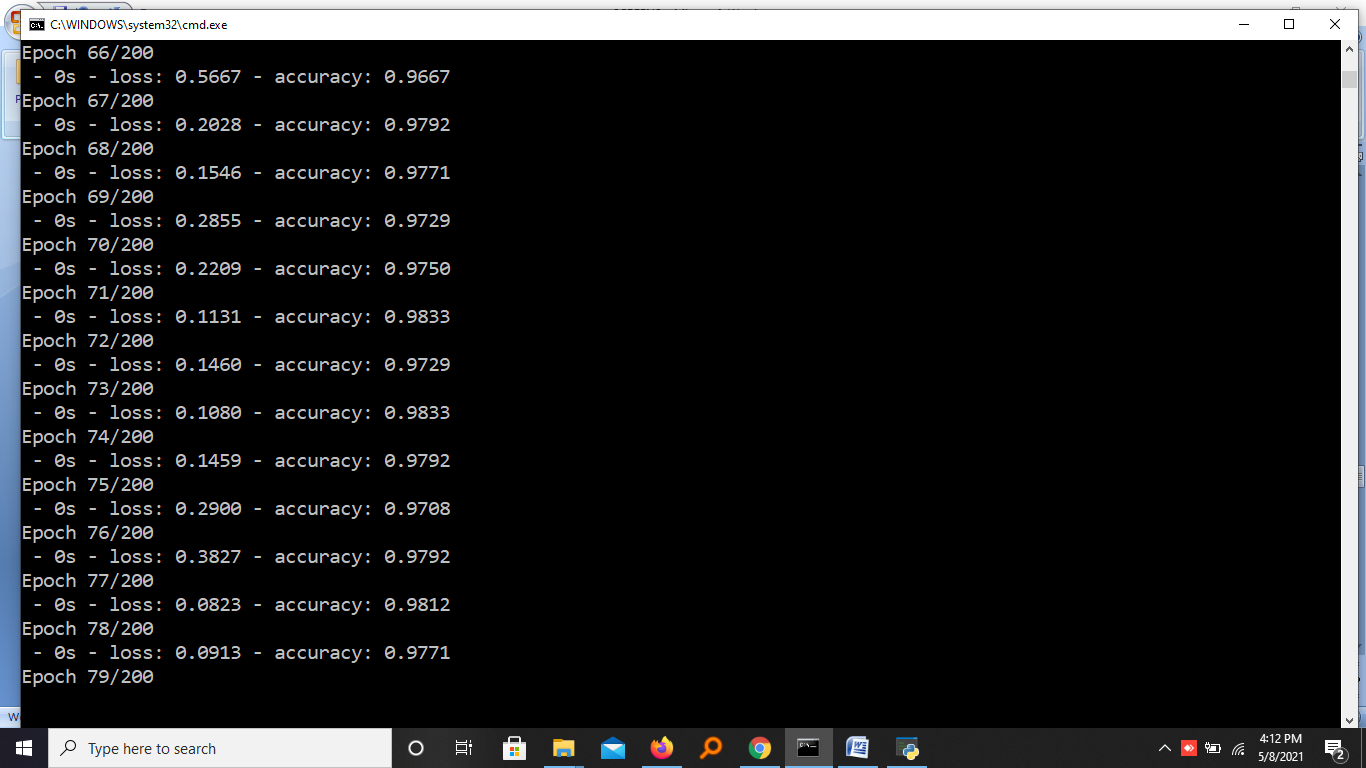
In above screen selecting and uploading ‘dataset.txt’ file and then click on ‘Open’ button to load dataset and to get below screen



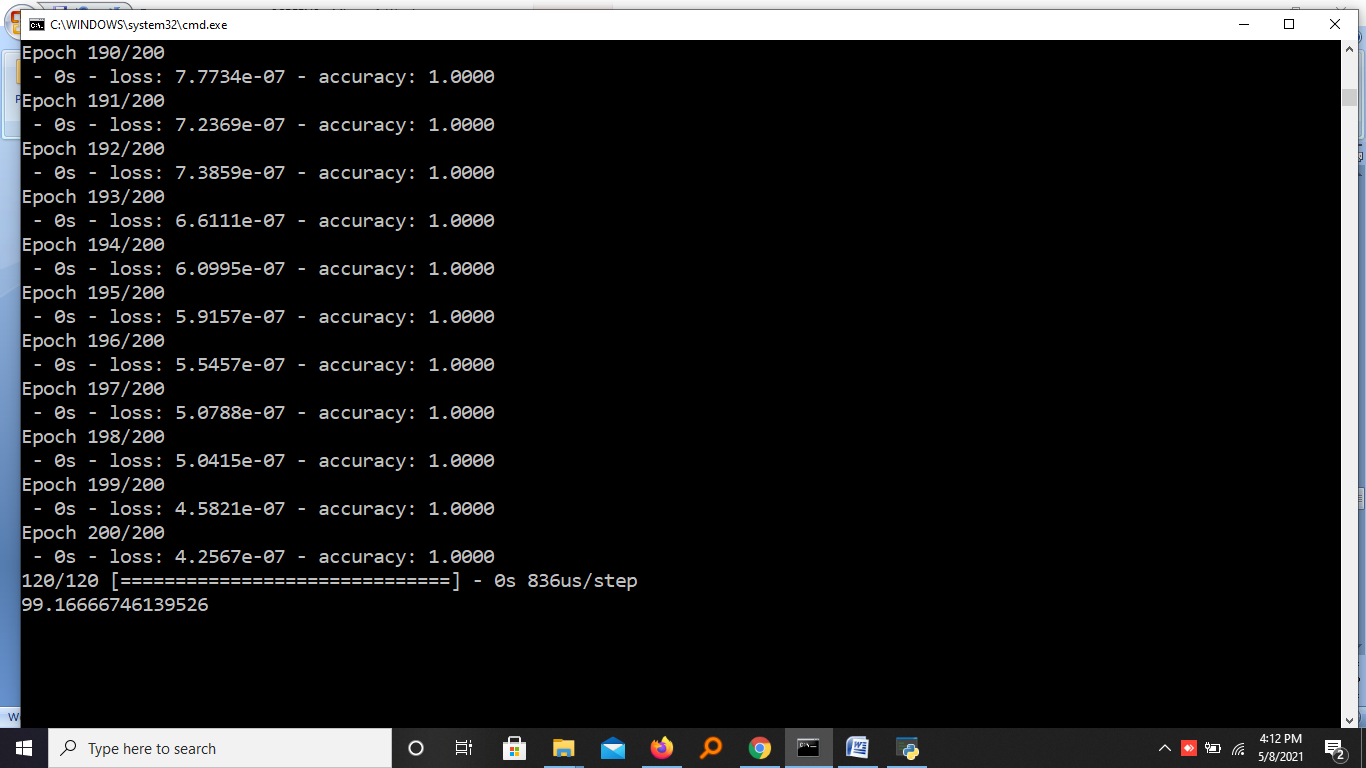
In above screen dataset loaded and displaying few records from dataset and now click on ‘Preprocess Dataset’ button to remove missing values and to split dataset into train and test part



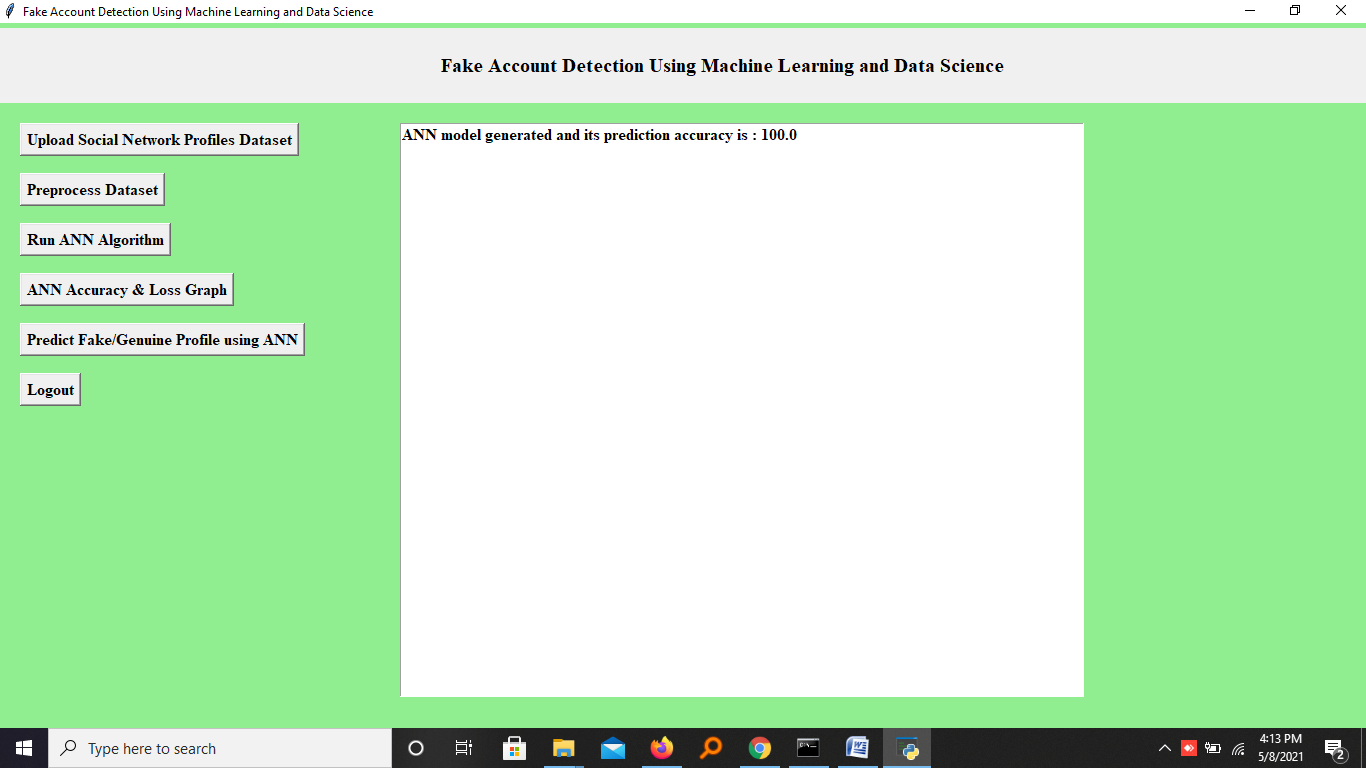
In above screen we can see dataset contains total 600 records and application using 480 records for training and 120 records to test ANN and now dataset is ready and now click on ‘Run ANN Algorithm’ button to ANN algorithm



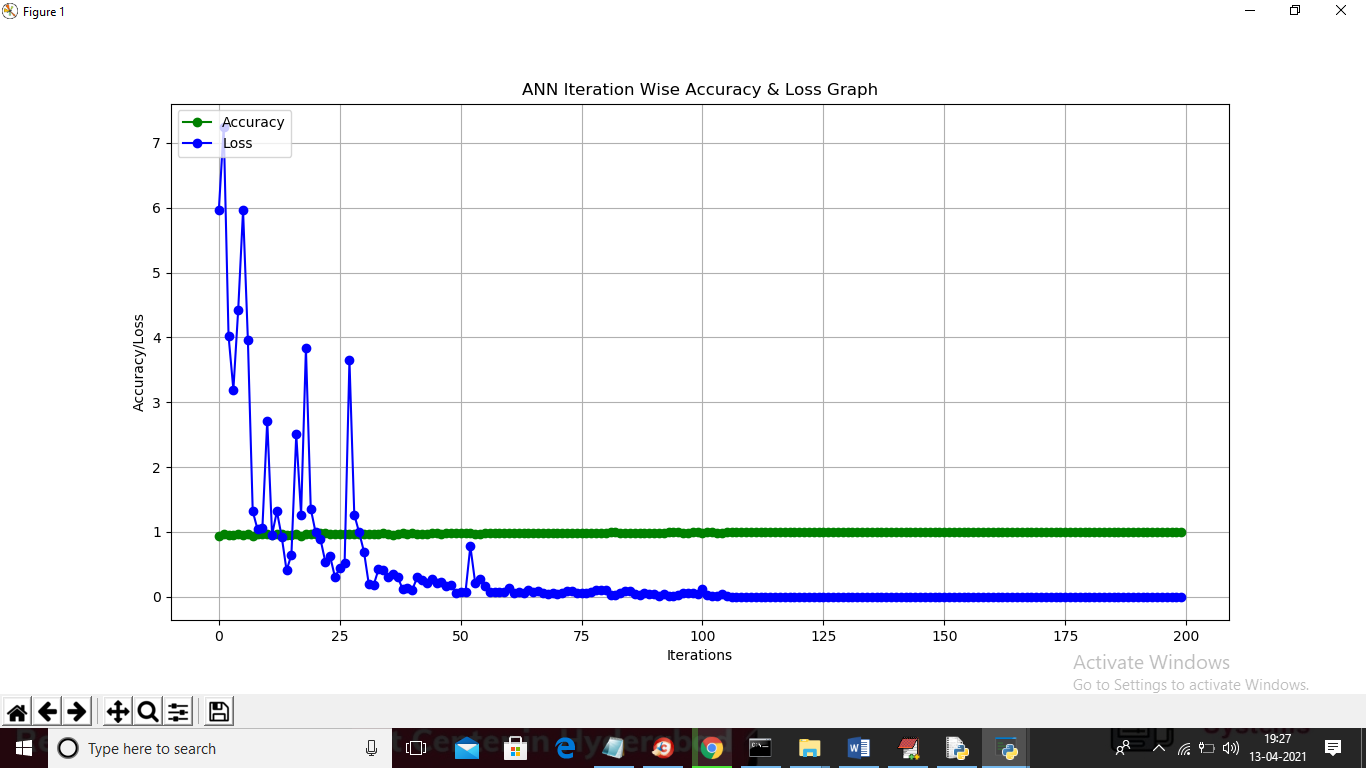
In above screen we can see ANN start iterating model generation and at each increasing epoch we can see accuracy is getting increase and loss getting decrease.



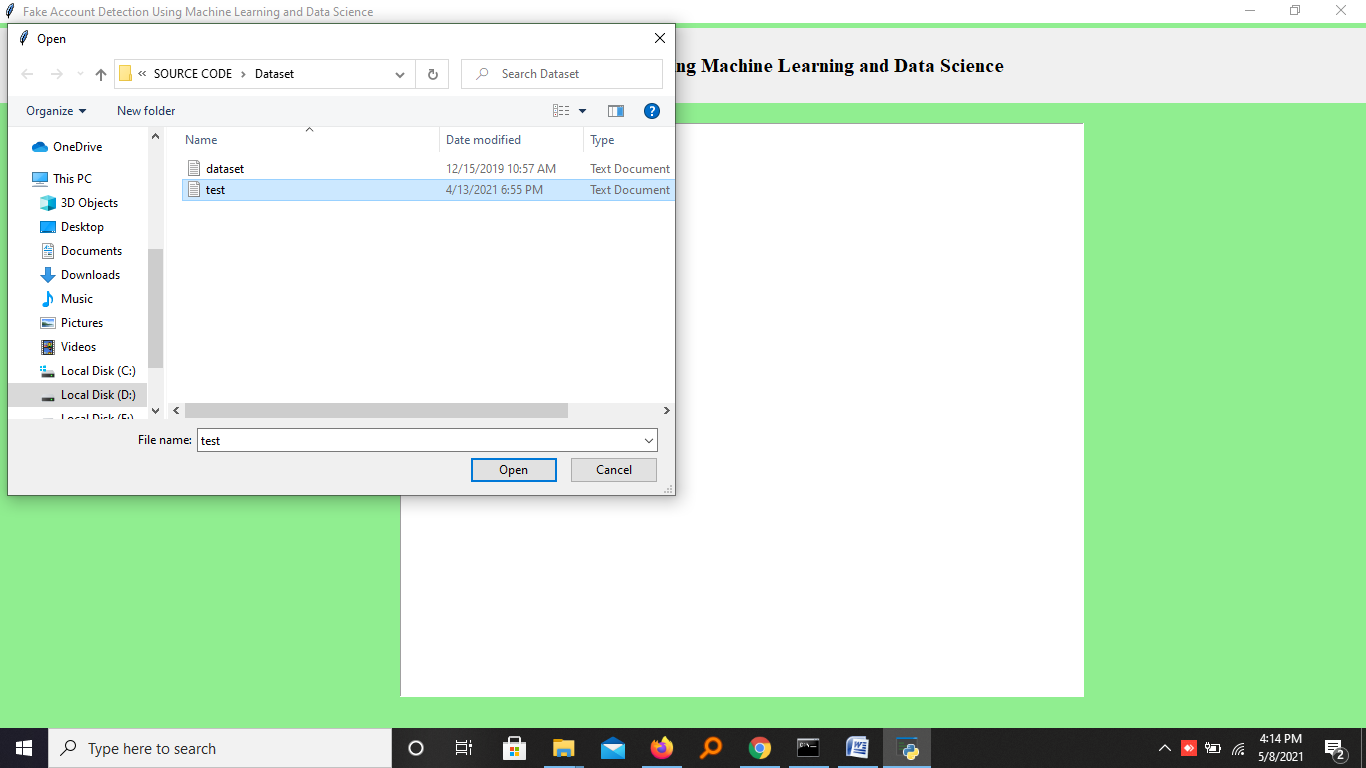
In above screen we can see after 200 epoch ANN got 100% accuracy and in below screen we can see final ANN accuracy



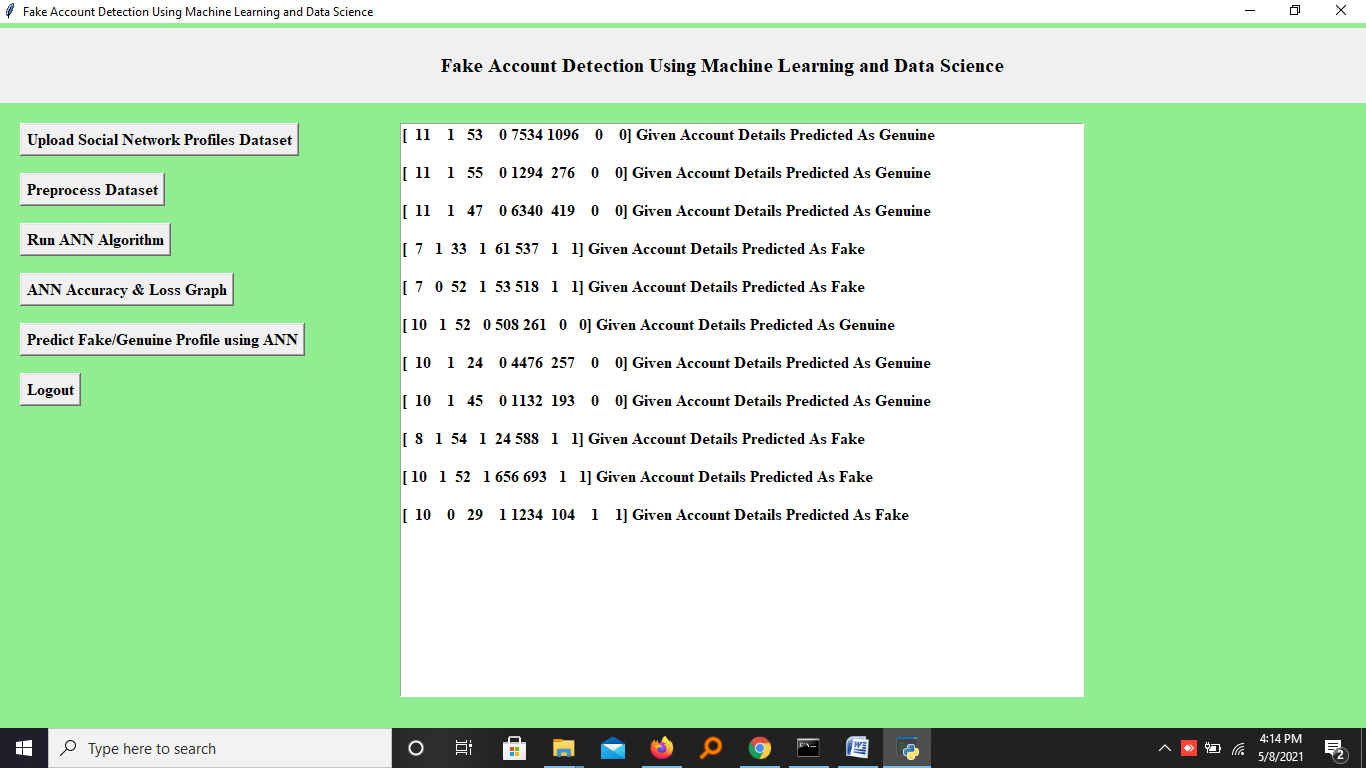
In above screen ANN model generated and now click on ‘ANN Accuracy & Loss Graph’ button to get below graph



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



In above screen we are selecting and uploading ‘test.txt’ file and then click on ‘Open’ button to load test data and to get below prediction result



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