Big Data Analytics and Mining for Effective Visualization and Trends Forecasting of Crime Data

In this paper author is describing concept to forecast crime data by analysing crimes dataset which contains lakhs of records. Analysing such big data help us in obtaining useful information such as in which hours or which year or which type of crime is committing more crimes. We can visualize this data to the police or crime authorities which help them in deploying more force at the location/time which have more crime commits.

Using big data we are visualizing useful information and then using Deep Machine Learning Algorithms such as LSTM (Long short term memory) and Prophet, we are forecasting crime for next year. Using this prediction police peoples can take necessary decision on time. Here author is using Neural Network, LSTM and Prophet to predict crimes but LSTM and Prophet is giving better performance.

In this project I am implementing LSTM and Neural Network and then comparing RSME (root mean square error) between them. I am not implementing PROPHET as this library not getting installed.

In paper author has used 3 crimes datasets from San-Francisco, Chicago, and Philadelphia but on internet I am able to find San-Francisco dataset and I am using that one only.

Dataset contain following columns/features

1) IncidentNum - Case number of each incident;

2) Dates - Date and timestamp of the crime incident;

3) Category - Type of the crime. This is the

4) Descript - A brief note describing any pertinent details of the crime;

5) DayOfWeek - Day of the week that crime occurred;

6) PdDistrict - Police Department District ID where the crime is assigned;

7) Resolution - How the crime incident was resolved (with the perpetrator being, say, arrest or booked);

8) Address - The approximate street address of the crime incident;

9) X - Longitude of the location of a crime;

10) Y - Latitude of the location of a crime;

11) Coordinate - Pairs of Longitude and Latitude;

12) Dome - whether crime id domestic or not;

13) Arrest - Arrested or not;

This dataset saved inside dataset folder and you can see all columns and crimes details in that dataset. To implement this project we are following below pre-processing steps.

1. Replacing all missing values with mean of that column values
2. Converting time stamp into five various columns such as YEAR, MONTH, DAY, TIME, HOUR and Seconds which can help in forecasting based on time series and visualization.

NEURAL NETWORK MODEL Working Details

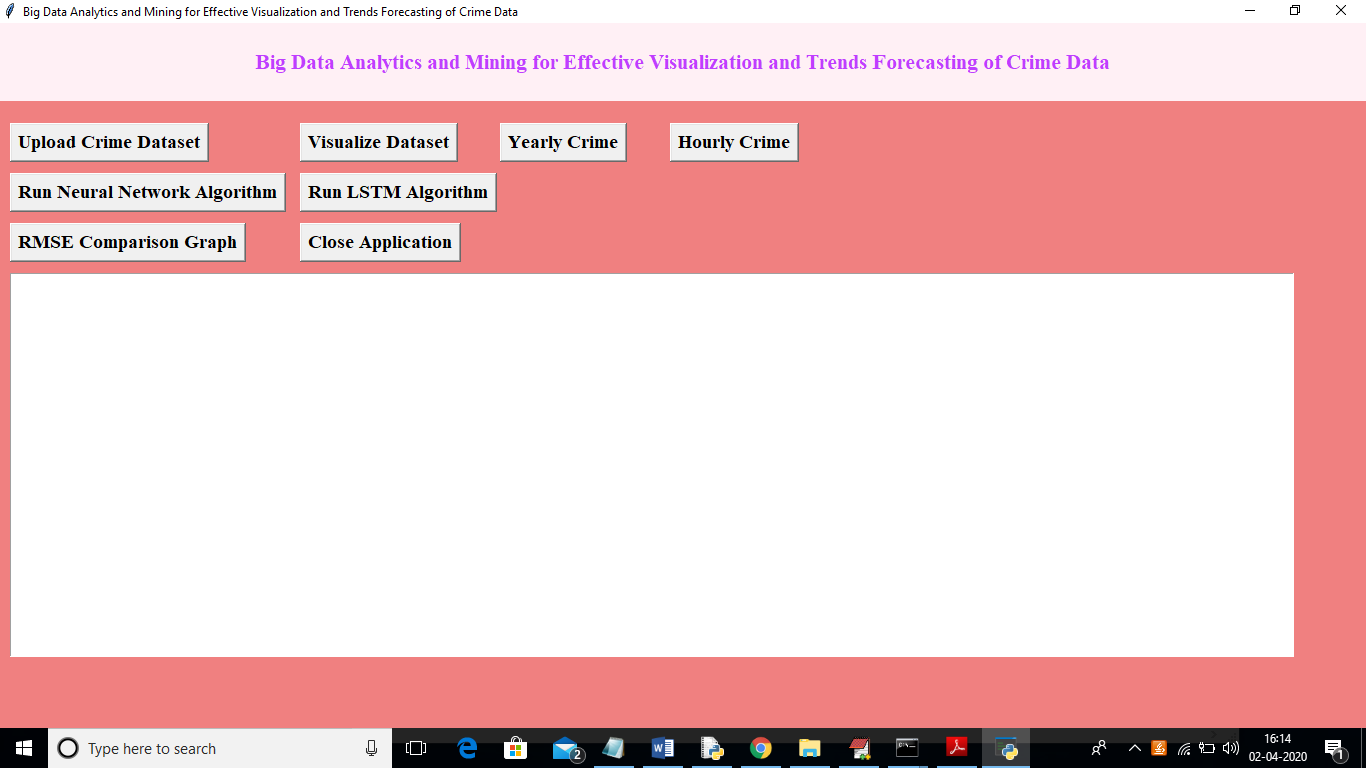
A neural network is composed of a certain numbers of neurons, namely nodes in the network, which are organized in several layers and connected to each other cross different layers. There are at least three layers in a neural network, i.e. the input layer of the observations, a non-observable hidden layer in the middle, and an output layer as the predicted results. In this paper we explored the multilayer feed-forward network, where each layer of nodes receives inputs from the previous layer. The outputs of the nodes in one layer will become the inputs to the next layer.

LSTM MODEL Working Details

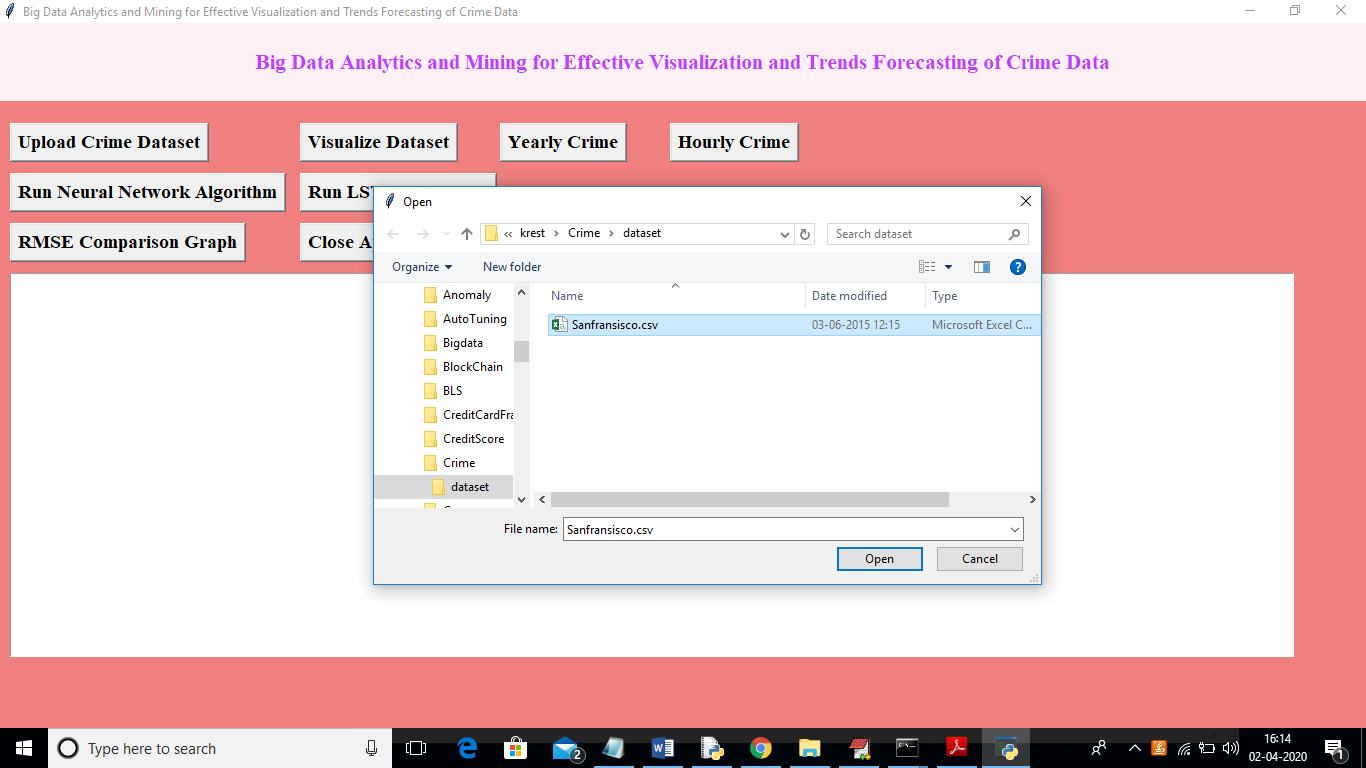
LSTM model is a powerful type of recurrent neural network (RNN), capable of learning long-term dependencies. For time series involves autocorrelation, i.e. the presence of correlation between the time series and lagged versions of itself, LSTMs are particular useful in prediction due to their capability of maintaining the state whilst recognizing patterns over the time series. The recurrent architecture enables the states to be persisted, or communicate between updated weights as each epoch progresses. Moreover, the LSTM cell architecture can enhance the RNN by enabling long term persistence in addition to short term

Screen shots

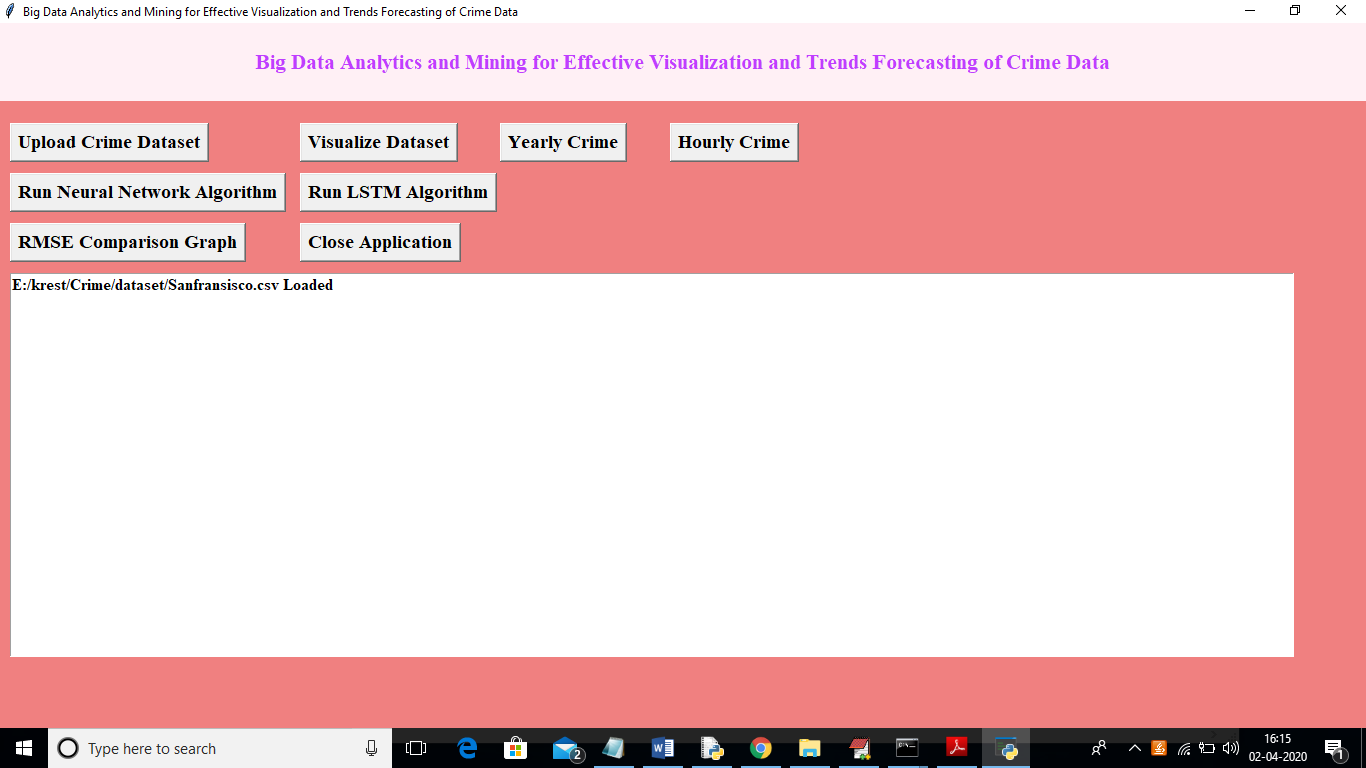
To run project double click on ‘run.bat’ file to get below screen



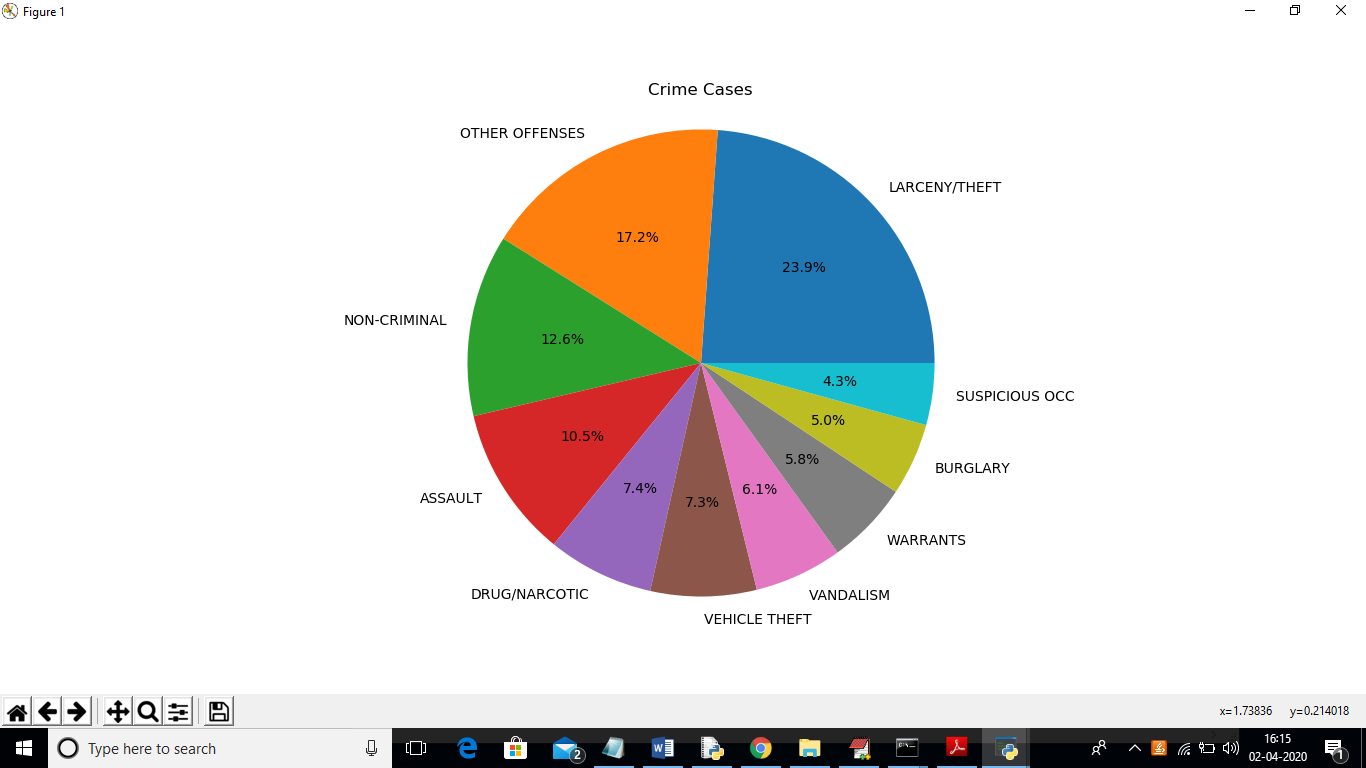
In above screen click on ‘Upload Crime Dataset’ button and upload dataset



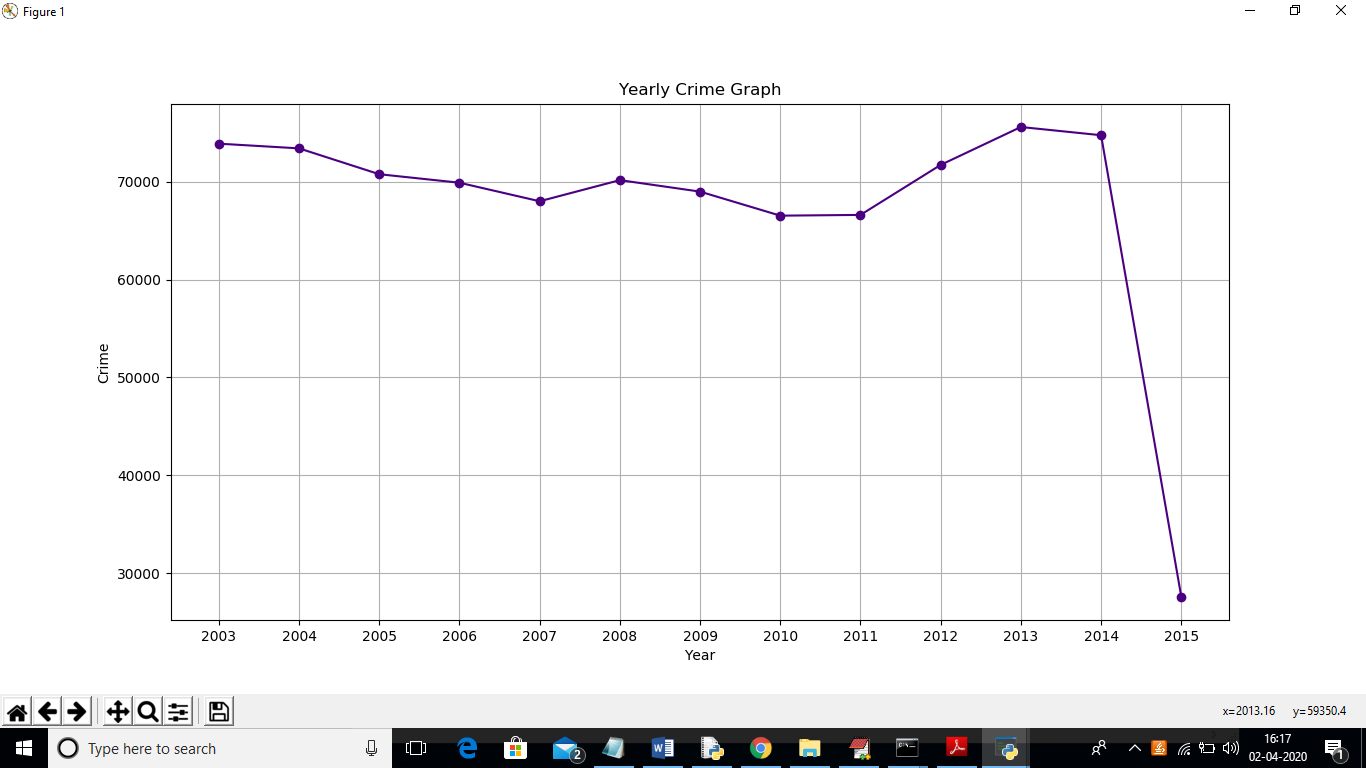
In above screen I am uploading San-Francisco dataset and uploading dataset will get below screen



Now click on ‘Visualize Dataset’ button to view top 10 crimes



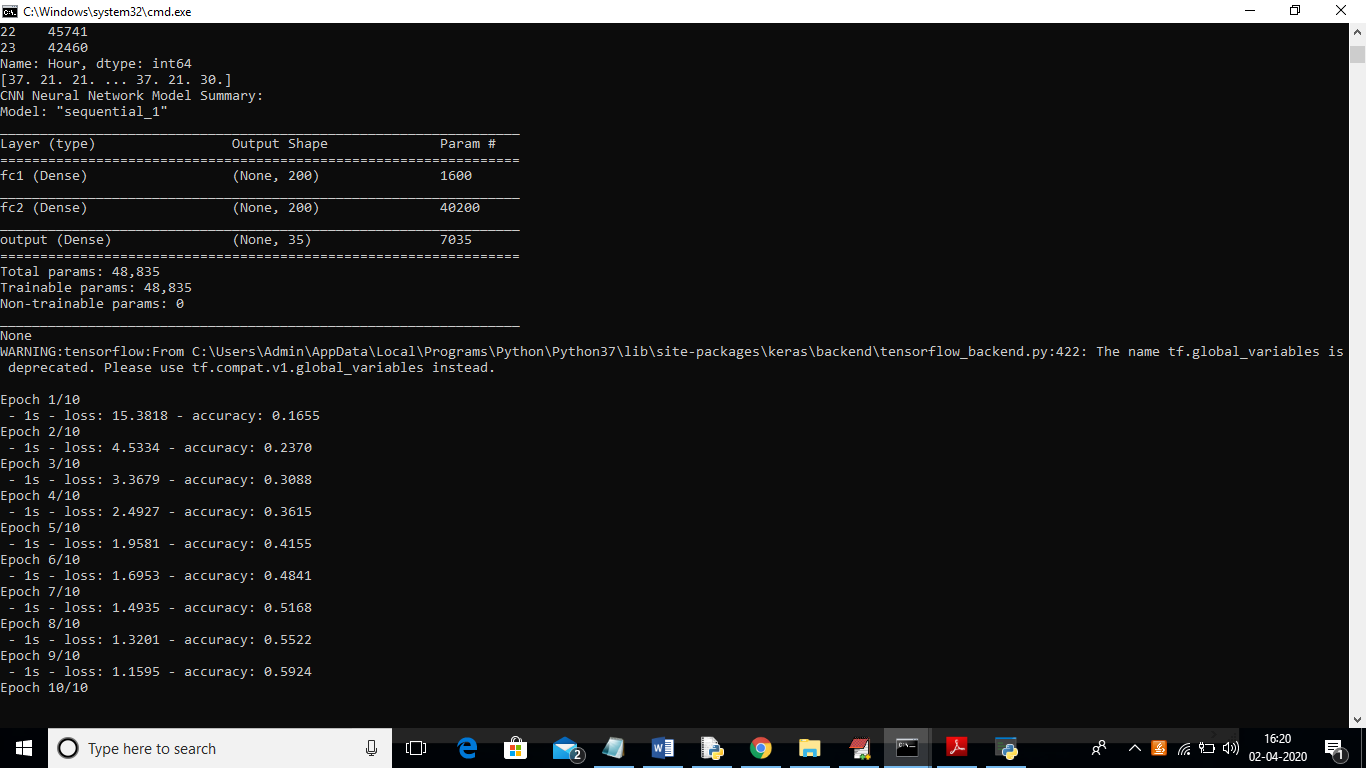
In above screen we can see top 10 crimes and which crime is committing more and now click on ‘Yearly Crime’ to visualize below graph



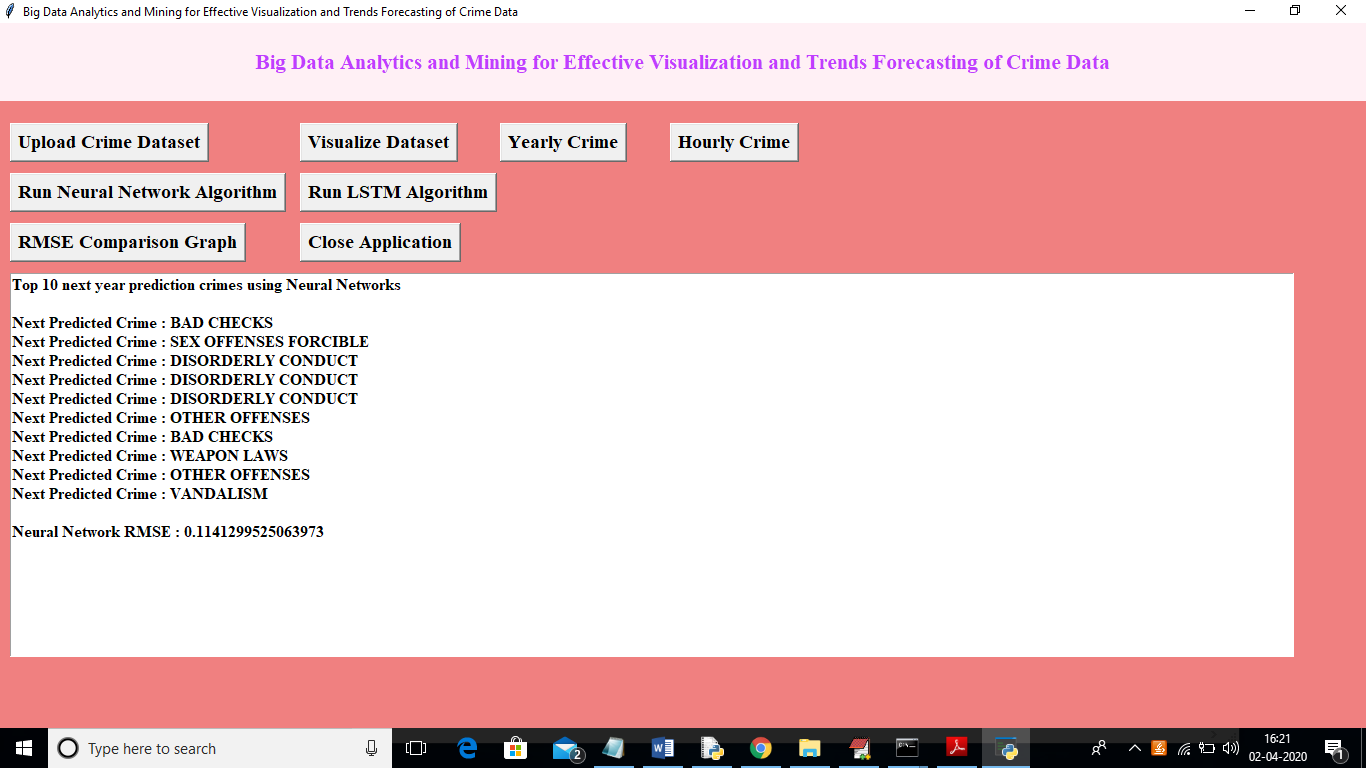
In above graph x-axis represents year and y-axis represents crimes count and from above graph from 2003 to 2014 more number of crimes was committed. Now click on ‘Hourly Graph’ button to visualize below graph



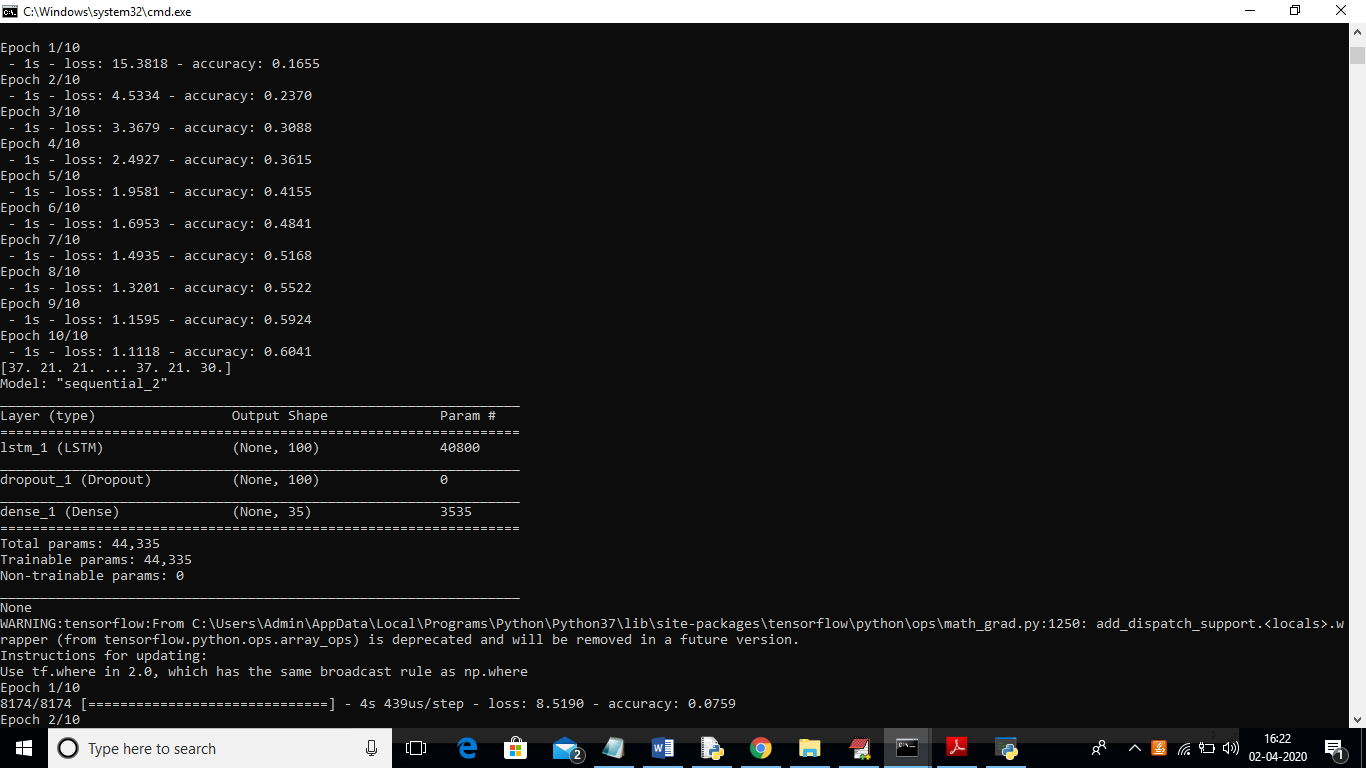
In above graph x-axis represents Hour and y-axis represents crime count and in above graph we can see more crimes are committing at night hours. Now we visualize all possible data and now to forecast/predict crime we need to train LSTM and Neural Network model. So click on ‘Run Neural Network Algorithm’ button to generate neural network model



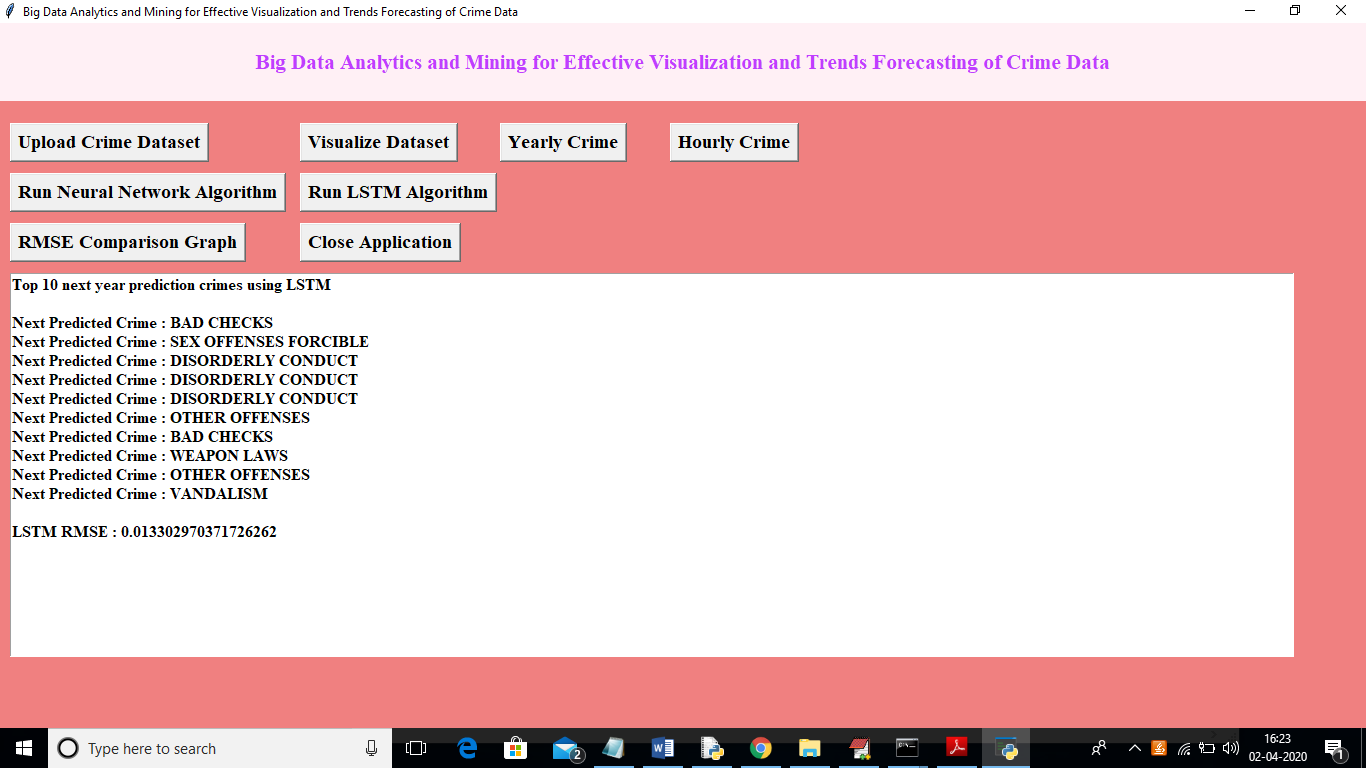
In above black console we can see neural network layers are created and its building model on different epochs to filter features and to choose best features to generate model. After training model we can get next year top 10 crime prediction in below screen



In above screen just I am displaying first top 10 crimes and below I am displaying RSME error for neural network. Now click on ‘Run LSTM Algorithm’ button to generate model using LSTM



In above console we can see LSTM model also start generating and now see below screen for prediction



In above screen we got prediction with LSTM and LSTM got less RMSE error compare to neural network so we can say LSTM is better than Neural Network. Now click on ‘RMSE Comparison Graph’ button to get RMSE comparison graph between LSTM and neural network



In above graph x-axis represents algorithm name and y-axis represents RMSE value. From above graph we can say LSTM got less RMSE so it performance is better. If any prediction model accuracy is high then it will consider as best prediction model and its RMSE error will reduce