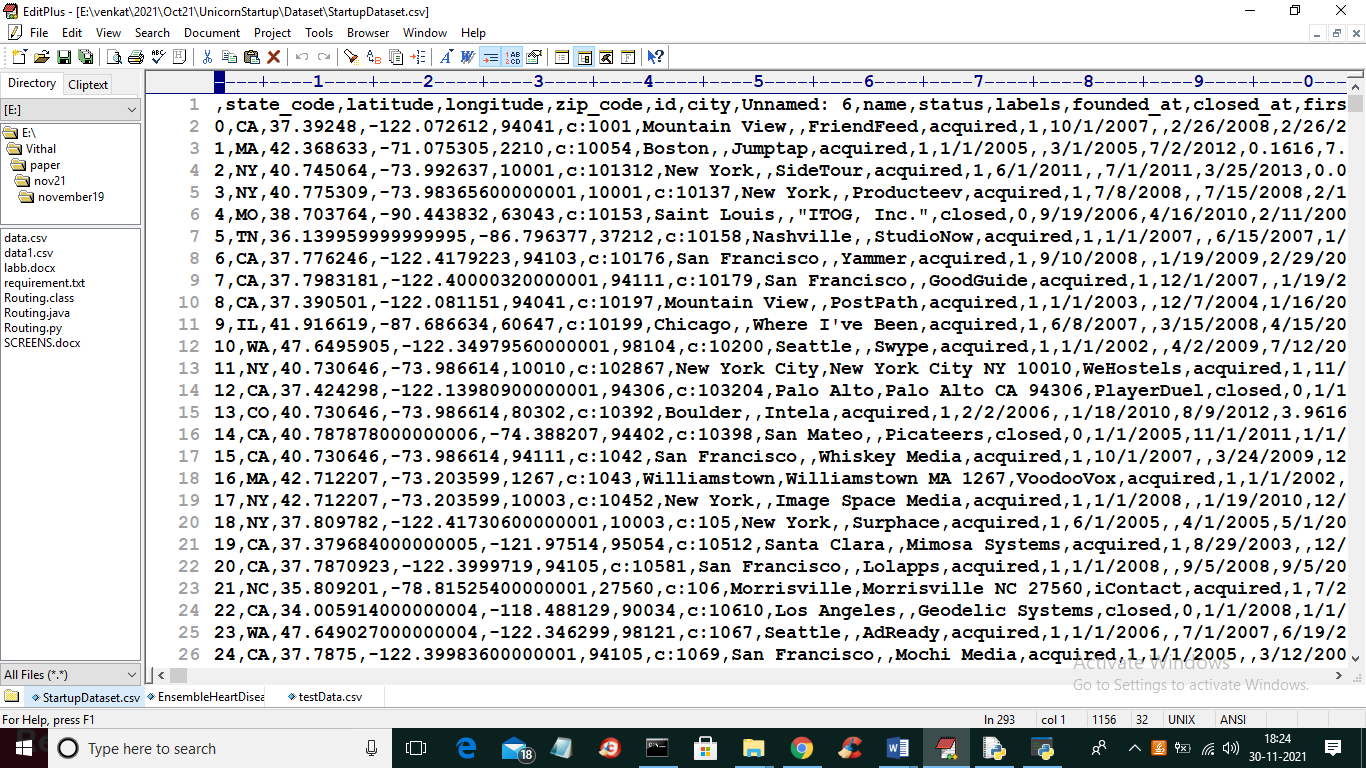
Startup Unicorn Prediction Using Advanced Machine Learning Algorithms

Startups play a major role in economic growth. They bring new ideas, spur innovation and create employment thereby moving the economy. There has been an exponential growth in startups over the past few years. Predicting the success of a startup allows investors to find companies that have the potential for rapid growth, thereby allowing them to be one step ahead of competition.

The objective of the project is to predict whether a startup which is currently operating turn into a success or a failure. The success of a company is defined as the event that gives the company's founders a large sum of money through the process of M&A (Merger and Acquisition) or an IPO (Initial Public Offering). A company would be considered as failed if it had to be shutdown.

To implement this project and predict success or failure of startup we have used advanced machine learning algorithms such as Gradient Boosting, SVM, Random Forest and Decision Tree. All ML algorithms will be trained on past performance of STARTUP dataset and then this trained model can be used to predict success or failure of new STARTUP TEST DATA.

Below is the dataset used to train all ML algorithms



In above dataset screen shots first row represents dataset column names and remaining rows represents dataset values and in all columns we column called ‘label’ which represents values as 0 or 1 where 0 means failure and 1 means success. In above dataset we have company location latitude, longitude, company founded date and closed date and many other parameters such as funding age, company in top 500 etc.

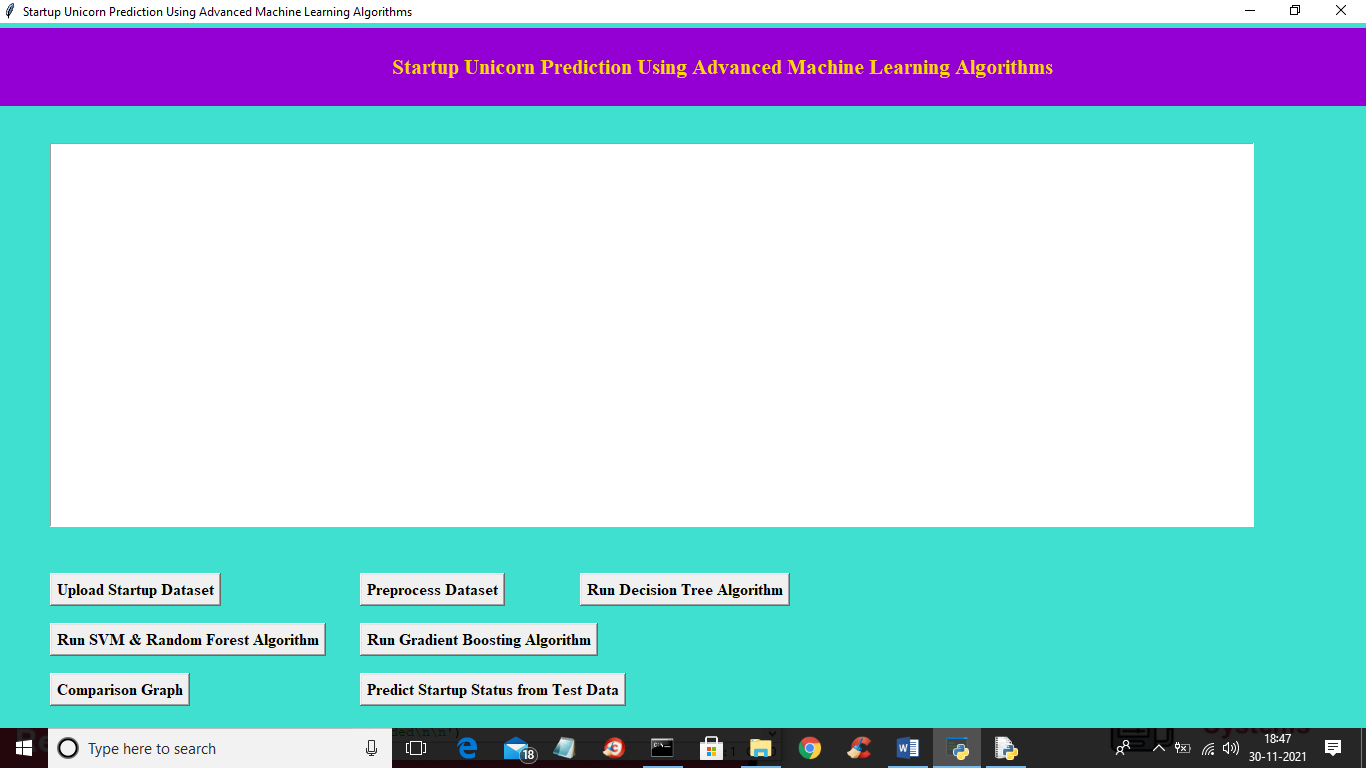
Will used above dataset to train ML algorithms

To implement this project we have designed following modules

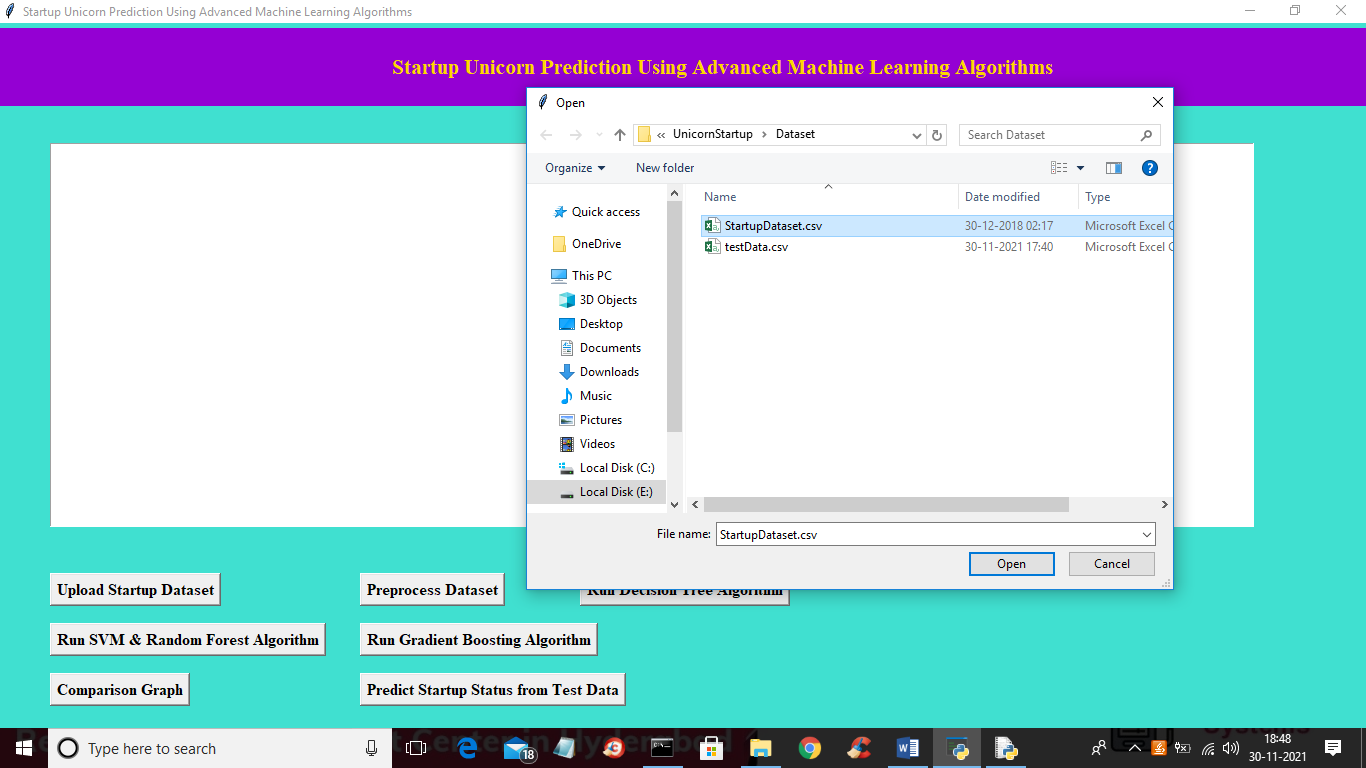
1. Upload Startup Dataset: using this module we will upload dataset to application
2. Preprocess Dataset: using this module we will process data to remove missing values and then split dataset into train and test where ML algorithms will take 80% dataset records to trained themselves and 20% records will be used to perform prediction and calculate accuracy
3. Run Decision Tree Algorithm: using this module we will trained decision tree algorithm
4. Run SVM & Random Forest Algorithm: using this module we will trained Random Forest and SVM algorithms
5. Run Gradient Boosting Algorithm: using this module we will trained Gradient Boosting algorithm
6. Comparison Graph: using this module we will evaluate performance of each algorithms and plot comparison graph of accuracy, precision, recall and FSCORE
7. Predict Startup Status from Test Data: using this module we will upload test data and then ML algorithms will analyse test data and predict whether that company or startup data show signs of success of failure

SCREEN SHOTS

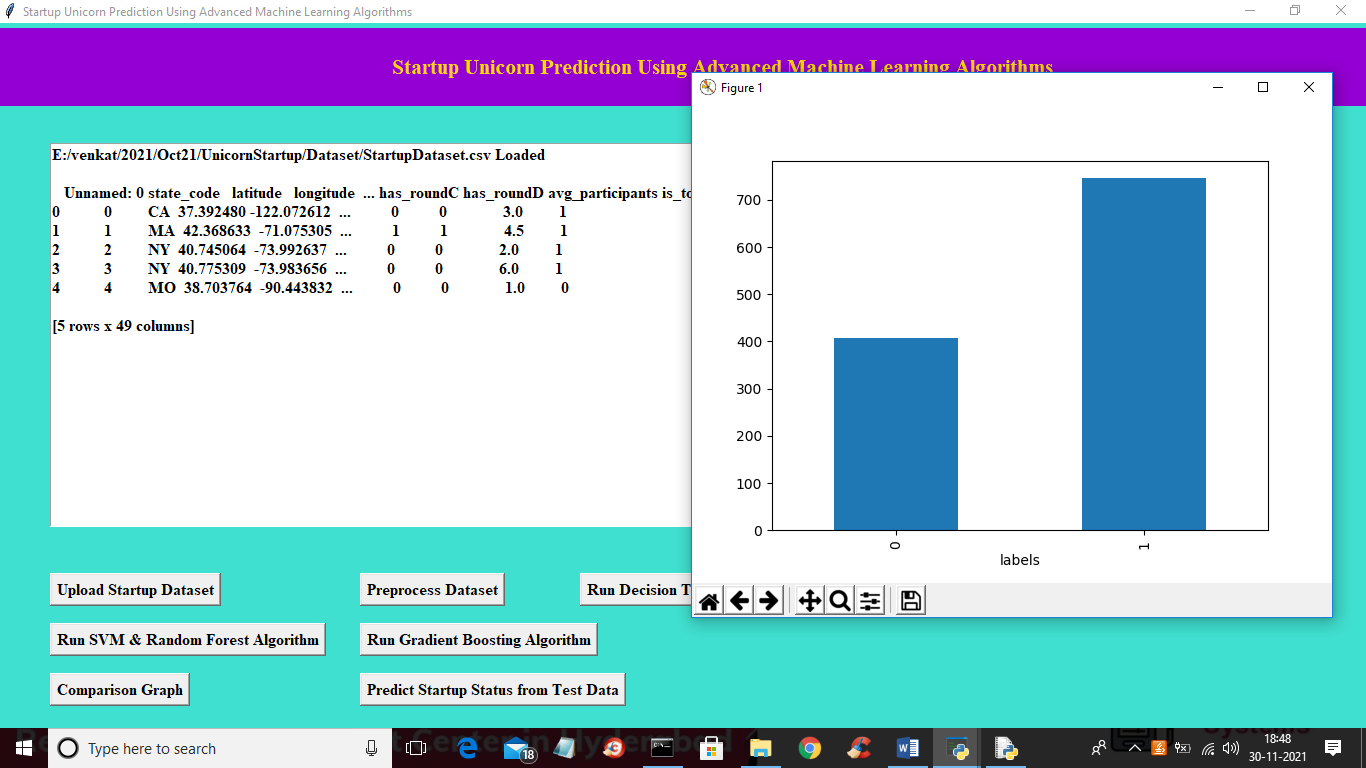
To run project double click on ‘run.bat’ file to get below screen



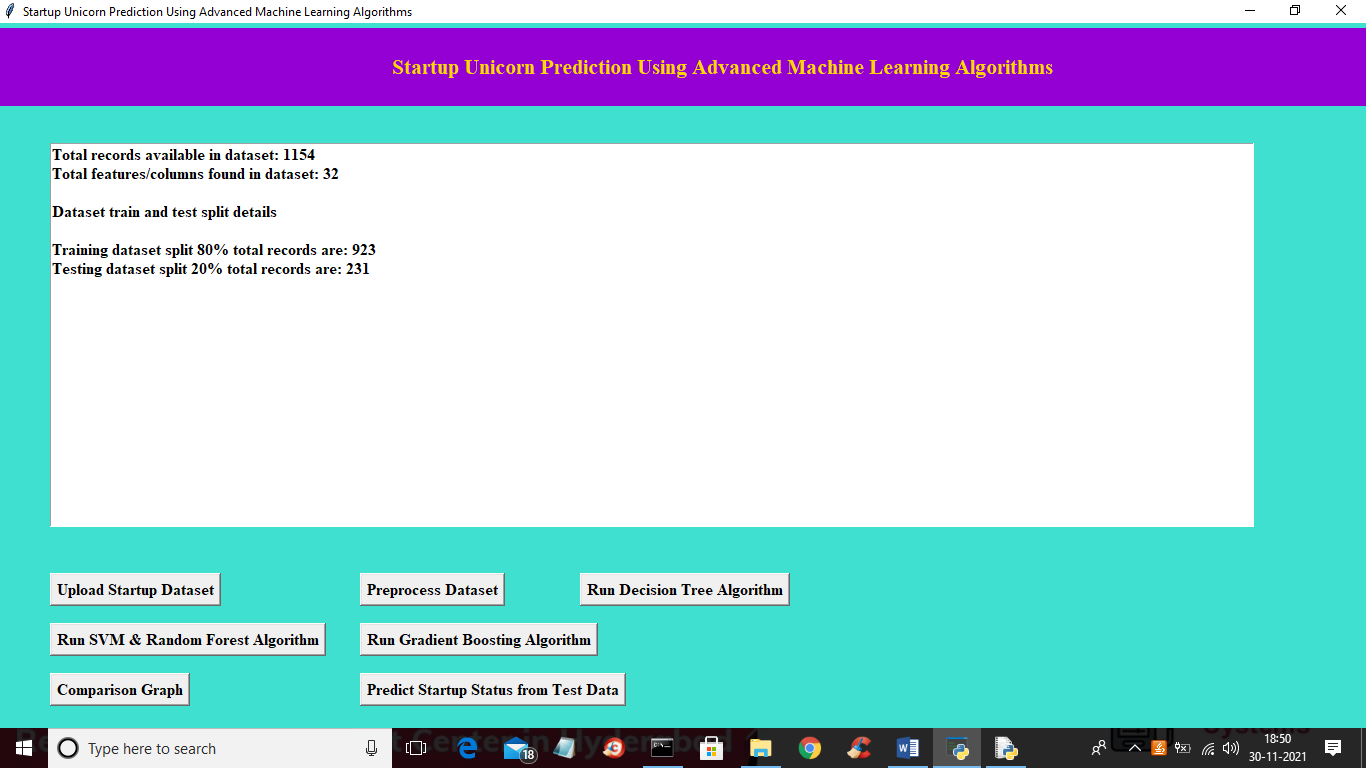
In above screen click on ‘Upload Startup Dataset’ button to upload dataset and to get below screen



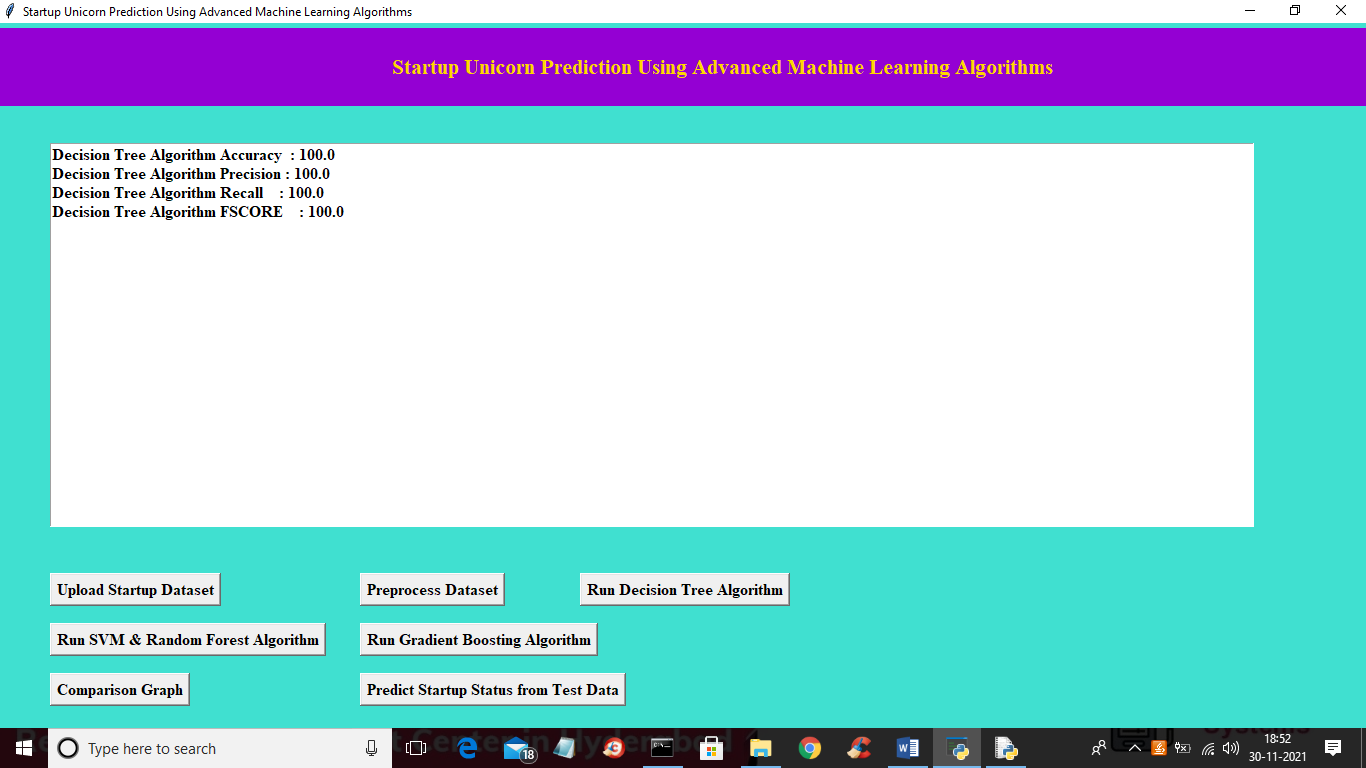
In above screen selecting and uploading ‘StartupDataset.csv’ file and then click on “Open” button to load dataset and to get below screen



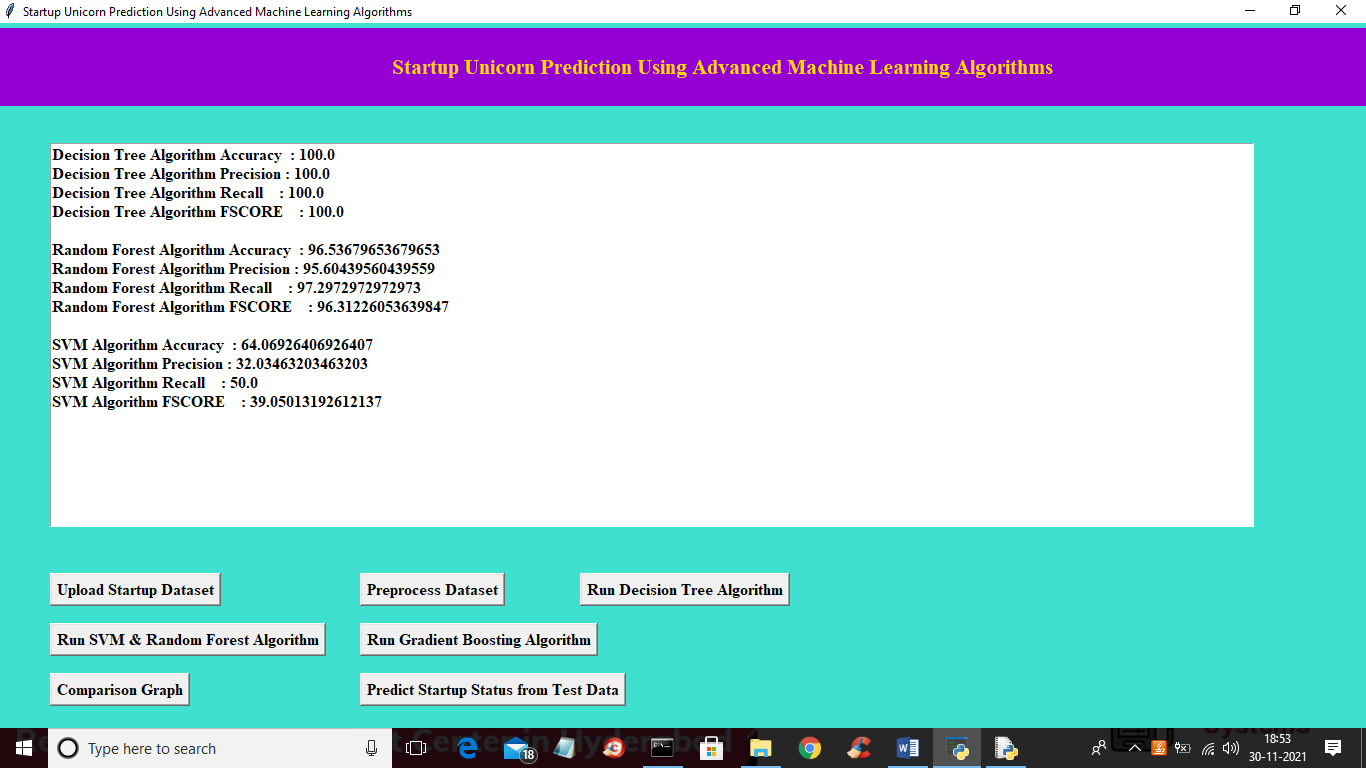
In above screen dataset loaded and I am displaying few values from dataset and in graph x-axis represents values 0 (failure) and 1 (success) and y-axis represents number of records available under that label. Now close above graph and then click on ‘Preprocess Dataset’ button to remove missing values and then split dataset into train and test part



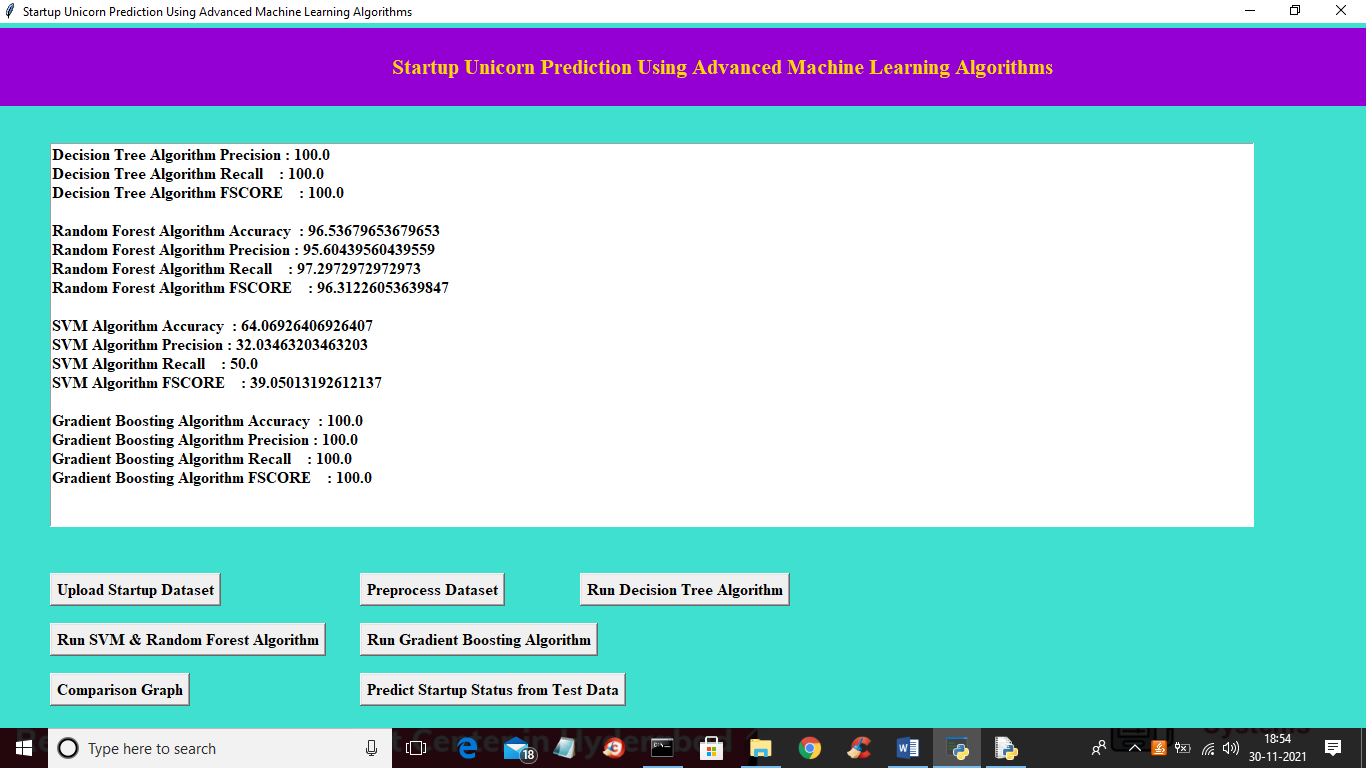
In above screen we can see dataset contains 1154 total records and each record contains 32 features/columns and then application split dataset into 80% (923 records) for training and remaining 231 records will be used to test trained algorithms prediction performance in terms of accuracy. Now train and test data is ready and now click on ‘Run Decision Tree Algorithm’ button to train Decision Tree with above dataset and get below output



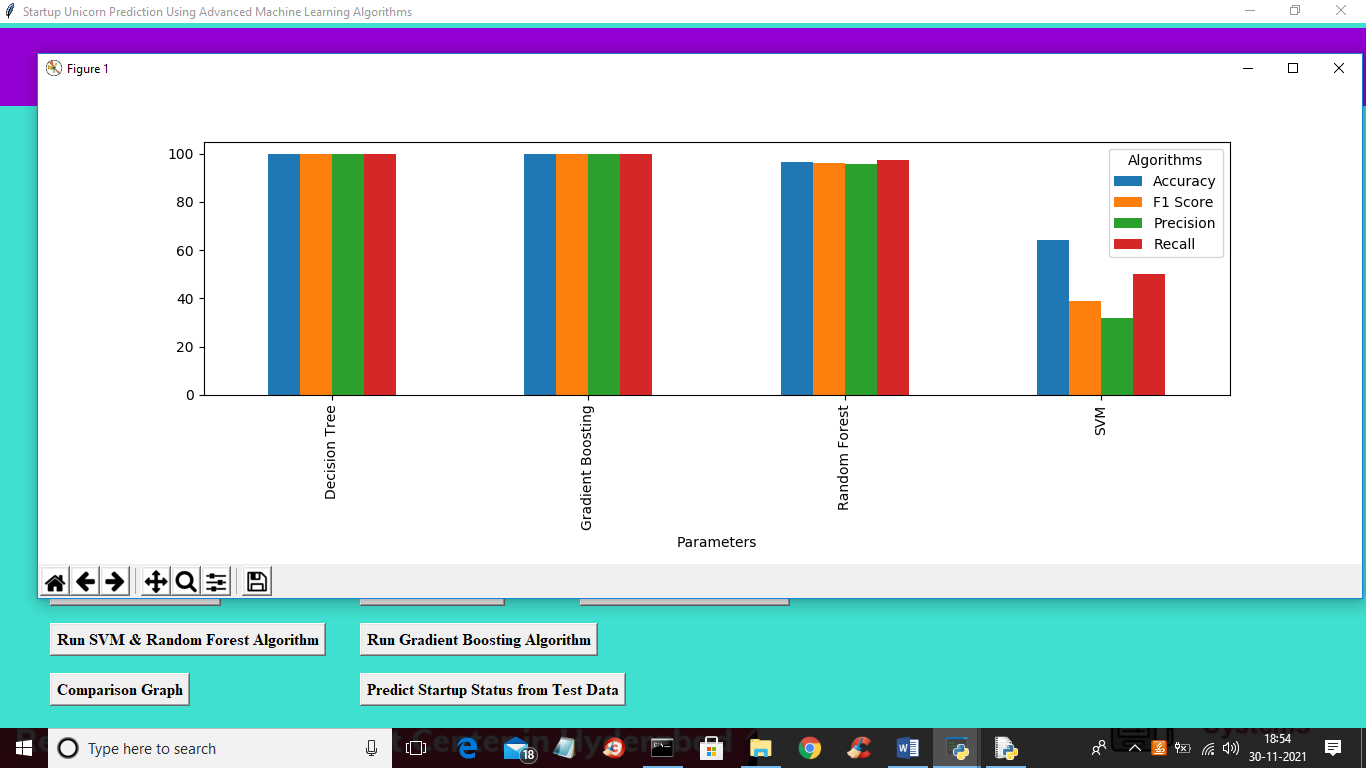
In above screen with decision tree we got 100% accuracy and now click on ‘Run SVM & Random Forest Algorithms’ button to get below output



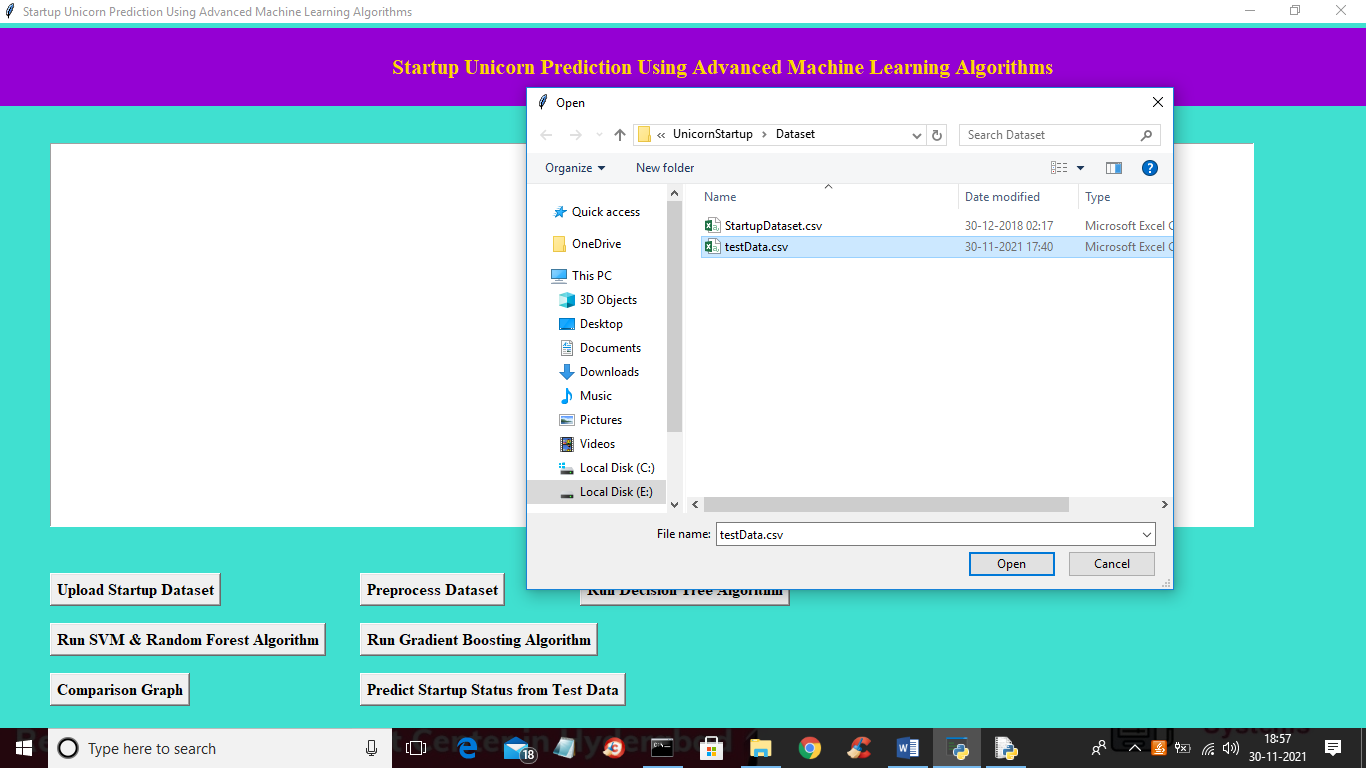
In above screen with random forest we got 96% accuracy and with SVM we got 64% accuracy and now click on ‘Run Gradient Boosting Algorithm’ button to train gradient boosting with above dataset and to get below output



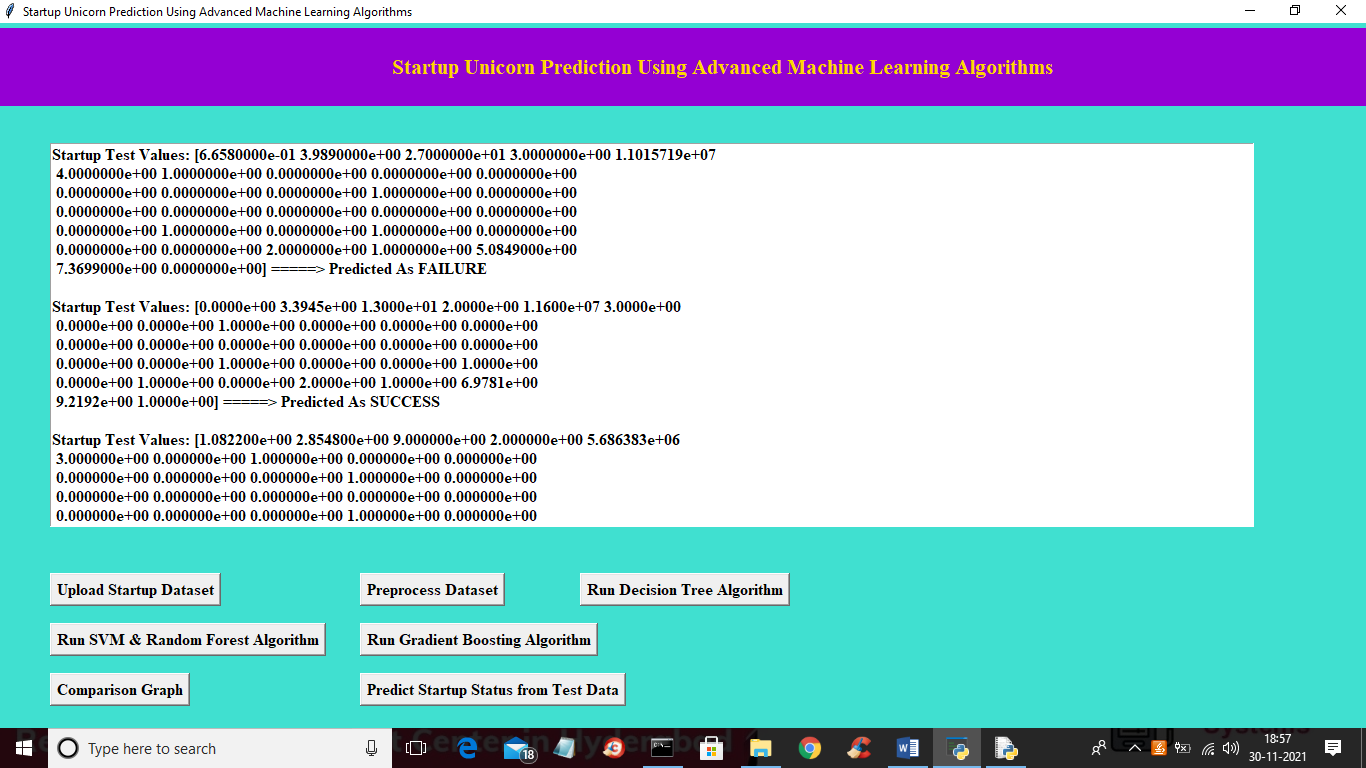
In above screen with Gradient Boosting also we got 100% accuracy and now click on ‘Comparison Graph’ button to get below graph



In above graph x-axis represents algorithms names and y-axis represents accuracy, precision, recall and FSCORE and in above graph different colour bar graph represents different metrics for each algorithm. In above graph we can see Decision Tree and Gradient Boosting giving high performance compare to other algorithms. Now close above graph and then click on ‘Predict Startup Status from Test Data’ button to upload test data and to get below output



In above screen selecting and uploading ‘testData.csv’ file and then click on “Open’ button to load test data and to get below prediction result



In above screen inside square bracket we can see Startup test values and based on test values ML algorithms perform prediction. We can see prediction result for each test record after ====🡺 arrow symbol as FAILURE or SUCCESS