Using Deep Learning to Predict Plant Growth and Yield in Greenhouse Environments

In this paper author is predicting ficus plant growth/crop yield by evaluating performance of various machine learning algorithms such as SVR (Support Vector Regression), Random Forest Regression (RF) and LSTM (Long Short Term Memory) deep neural network algorithm. SVR and RF are the traditional old algorithms whose performance of prediction will be low due to unavailable of deep learning technique. To overcome from this problem author is using LSTM deep neural network algorithm to predict plant growth.

Deep Learning extends classical ML by adding more "depth" (complexity) into the model, as well as transforming the data using various functions that create data representations in a hierarchical way, through several levels of abstraction. A strong advantage of DL is feature learning, i.e., automatic feature extraction from raw data, with features in higher levels of the hierarchy being formed through composition of lower level features. DL can solve complex problems particularly well and fast, due to the more complex models used, which also allow massive parallelization. These complex models employed in DL can increase classification accuracy, or reduce error in regression problems, provided there are adequately large datasets available describing the problem. DL includes different components, such as convolutions, pooling layers, fully connected layers, gates, memory cells, activation functions, encoding/decoding schemes, depending on the network architecture used, e.g., Convolutional Neural Networks, Recurrent Neural Networks and Unsupervised Networks.

The LSTM model is introduce with the objective of modelling long term dependencies and determining the optimal time lag for time series problems. A LSTM network is composed of one input layer, one recurrent hidden layer, and one output layer. The basic unit in the hidden layer is the memory block, containing memory cells with self-connections memorizing the temporal state and a pair of adaptive, multiplicative gating units controlling information flow in the block. The memory cell is primarily a recurrently self-connected linear unit, called Constant Error Carousel (CEC), and the cell state is represented by the activation of the CEC. The multiplicative gates learn when to open and close. By keeping the network error constant, the vanishing gradient problem can be solved in LSTM. Moreover, a forget gate is added to the memory cell preventing the gradient from exploding when learning long time series.

This project consists of following modules

1. upload dataset: using this module we will upload FICUS plant dataset
2. Dataset cleaning: using this module we will find out empty values in the dataset and replace with mean or 0 values.
3. Train & Test Split: Using this module we will split dataset into two parts called and training and testing. All machine learning algorithms take 80% dataset to train classifier and 20% dataset is used to test classifier prediction accuracy. If classifier prediction accuracy high then Mean Square Error, Root Mean Square Error and Mean Absolute Error will be dropped.
4. Run SVR Classifier: Using this module we will train SVR classifier with splitted 80% data and used 20% data to calculate it performance
5. Run Random Forest Classifier: Using this module we will train Random Forest classifier with splitted 80% data and used 20% data to calculate it performance
6. Run LSTM Classifier: Using this module we will train LSTM classifier with splitted 80% data and used 20% data to calculate it performance
7. Predict Plant & Yield Growth: Using this module we will upload test data and then apply LSTM classifier to predict it growth value

Dataset information

To implement this project we are using FICUS plant dataset and this dataset saved inside ‘dataset’ folder. Below are some examples of dataset

CO2,Radiation,diameter,humidity,outside\_temperature,inside\_temperature,measurement,Yield

35.7, 20.85, 29.53, 0.91, 35.7, 27.48, 2.46, 35.7

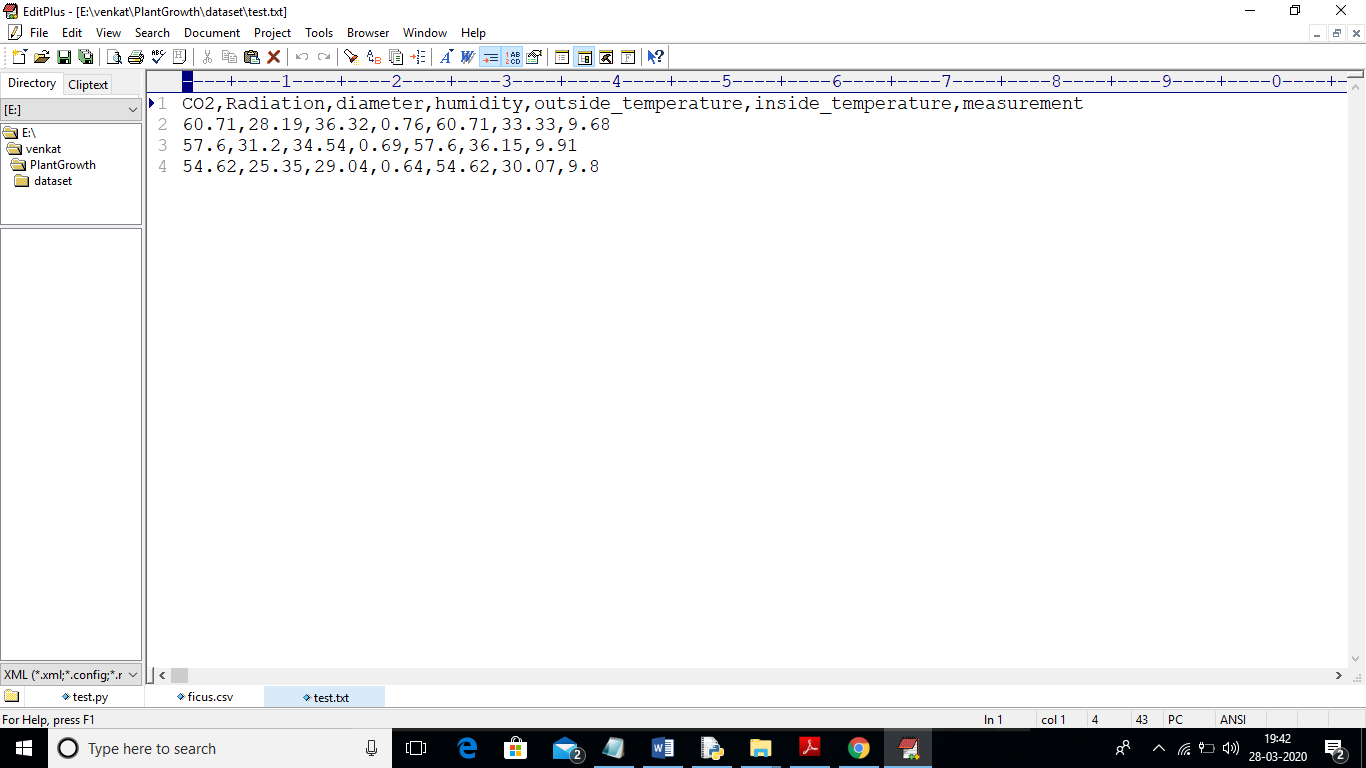
35.1, 26.92, 29.77, 0.93, 35.1, 26.92, 2.83, 35.7

55.15, 25.42, 31.27, 0.67, 55.15, 31.8, 9.98, 45.6

54.87, 28.86, 32.39, 0.67, 54.87, 35.73, 9.97, 45.6

66.45, 34.7, 43.11, 0.75, 66.45, 39.12, 9.75, 13.1

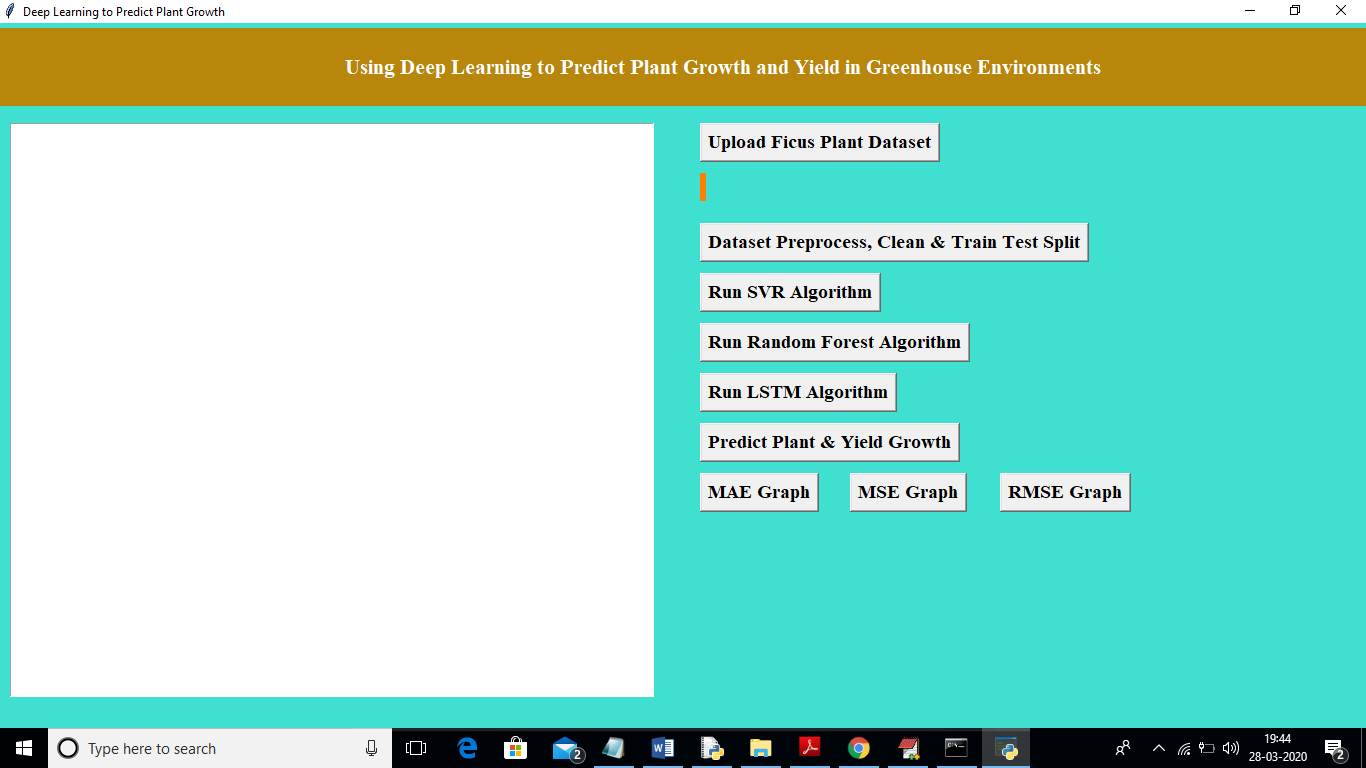
In above dataset we have columns as CO2, RADIATION, DIAMETER etc and last value is the YIELD of the crop under above environment values. By using above values we will train classifier and then upload test data to predict future growth or yield. Below are some test environment values but YIELD column is missing and classifier will predict



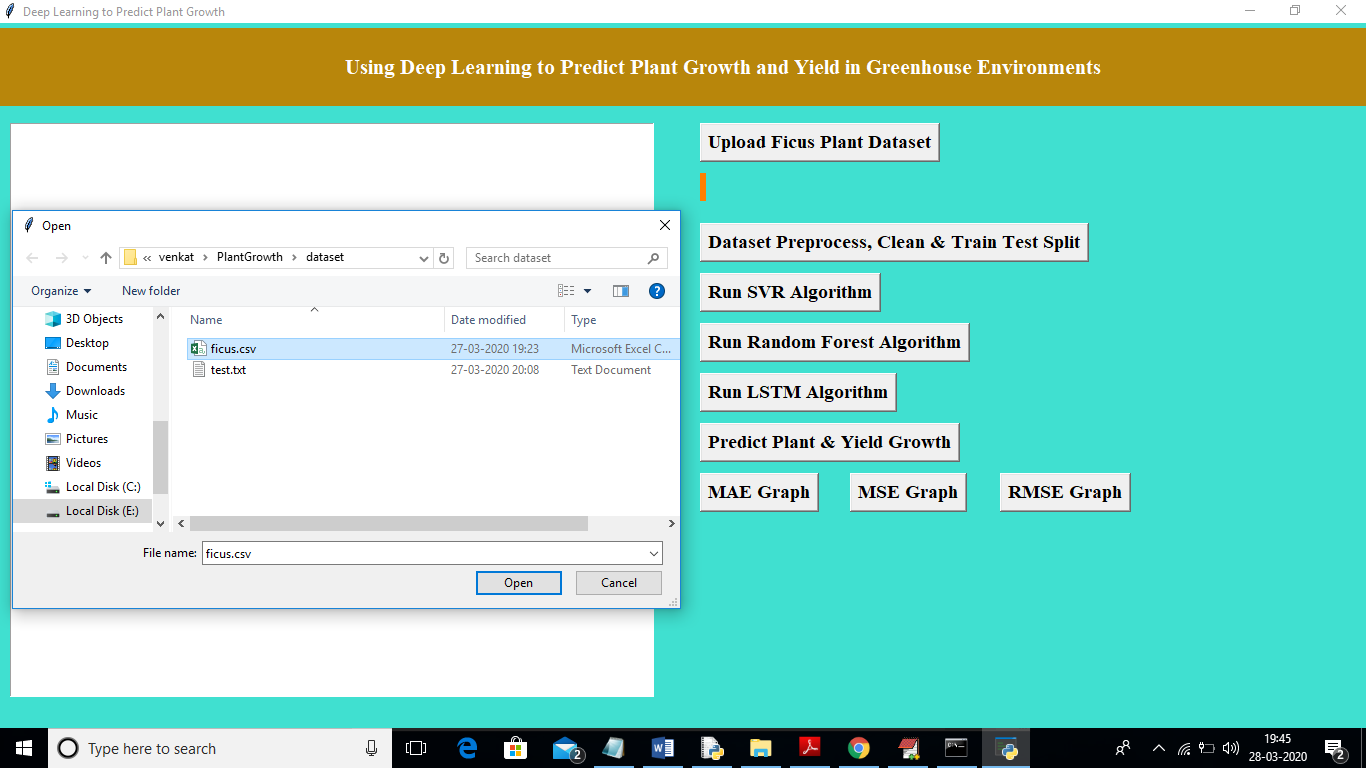
In above test data set we can see we have environment values but yield/growth value is missing and when we apply LSTM classifier on above test data then it will predict future growth for above test data.

Screen shots

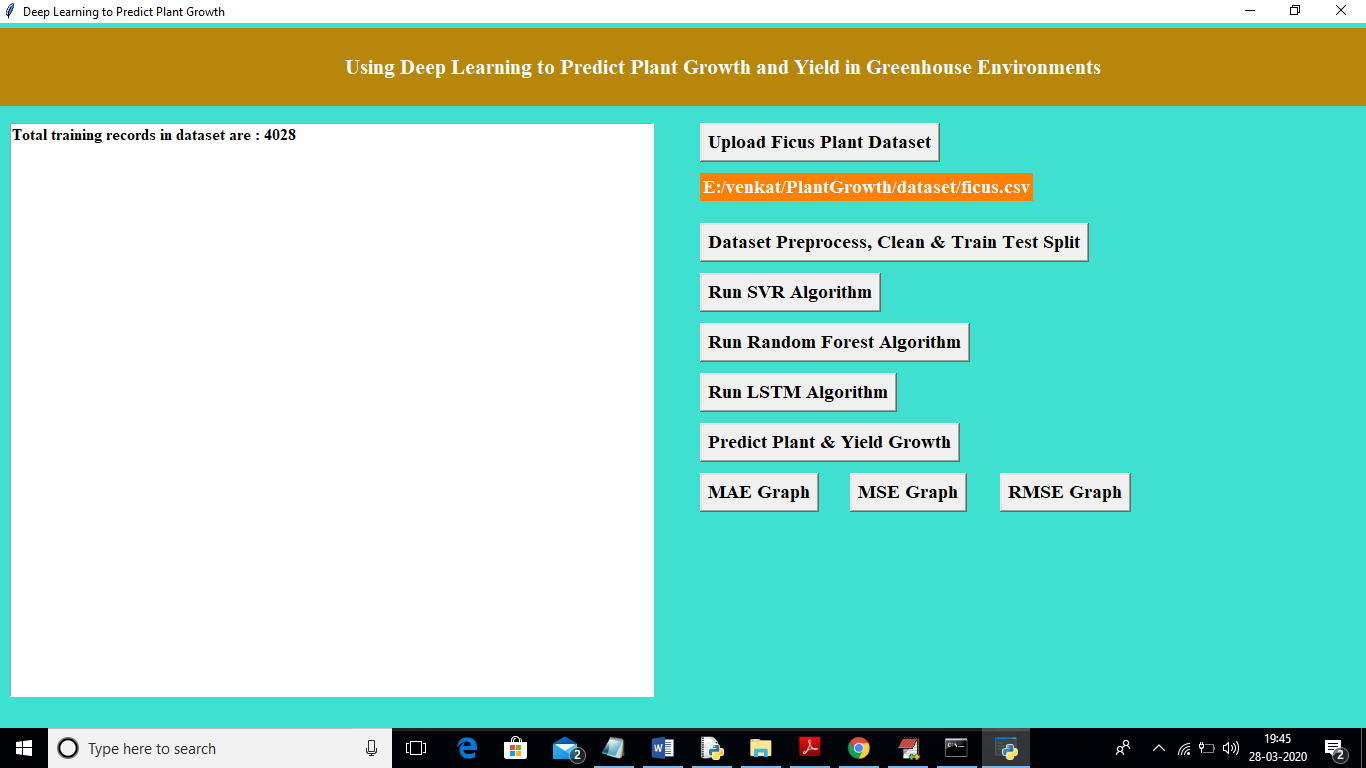
Double click on ‘run.bat’ file to get below screen



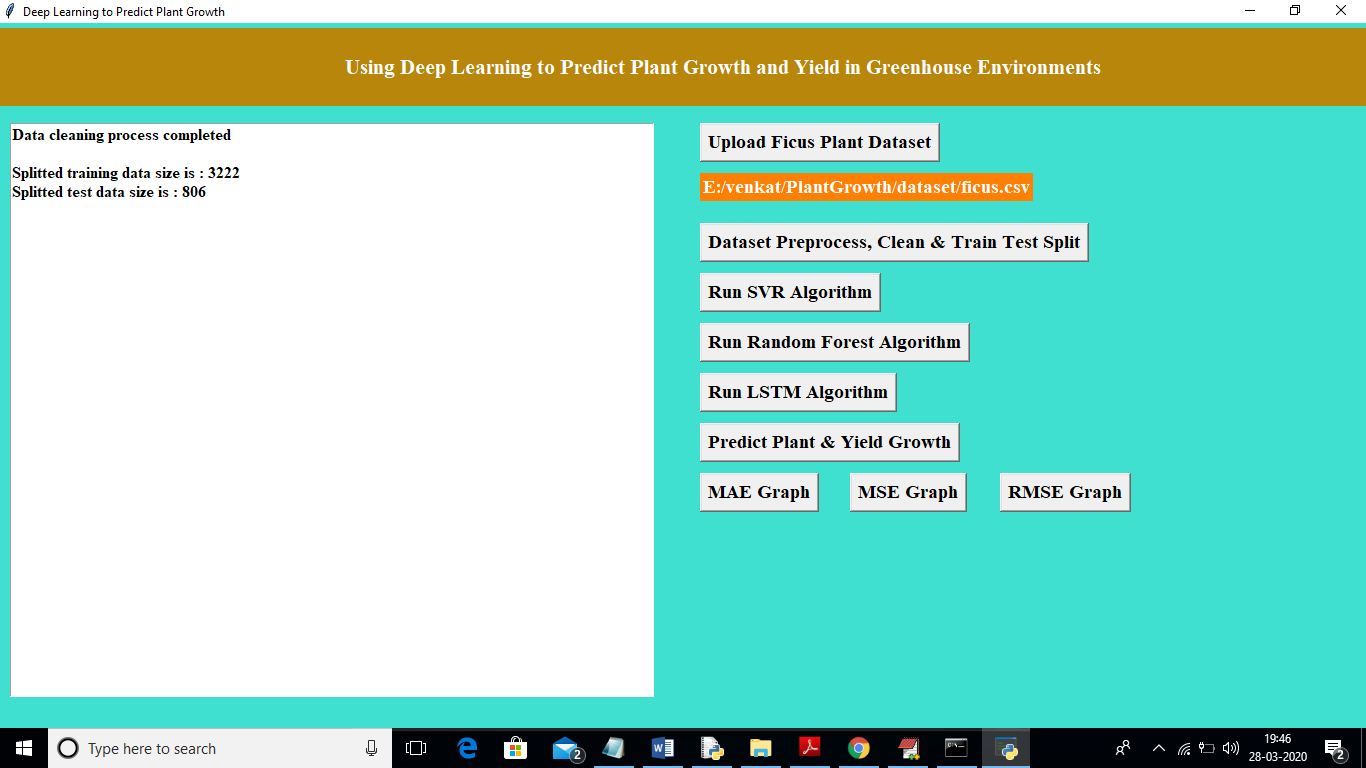
In above screen click on ‘Upload Ficus Plant Dataset’ button and upload dataset



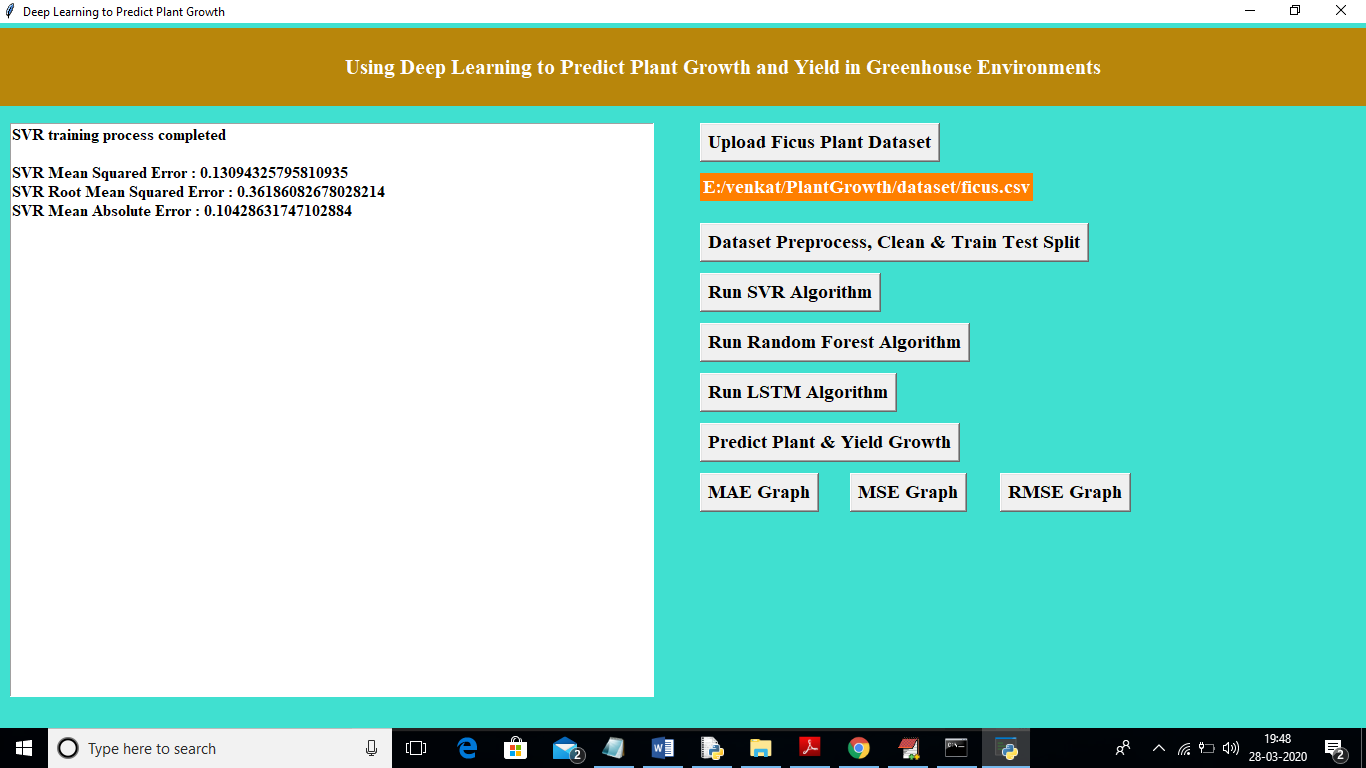
In above screen I am uploading ‘ficus.csv’ dataset file and after uploading dataset will get below screen



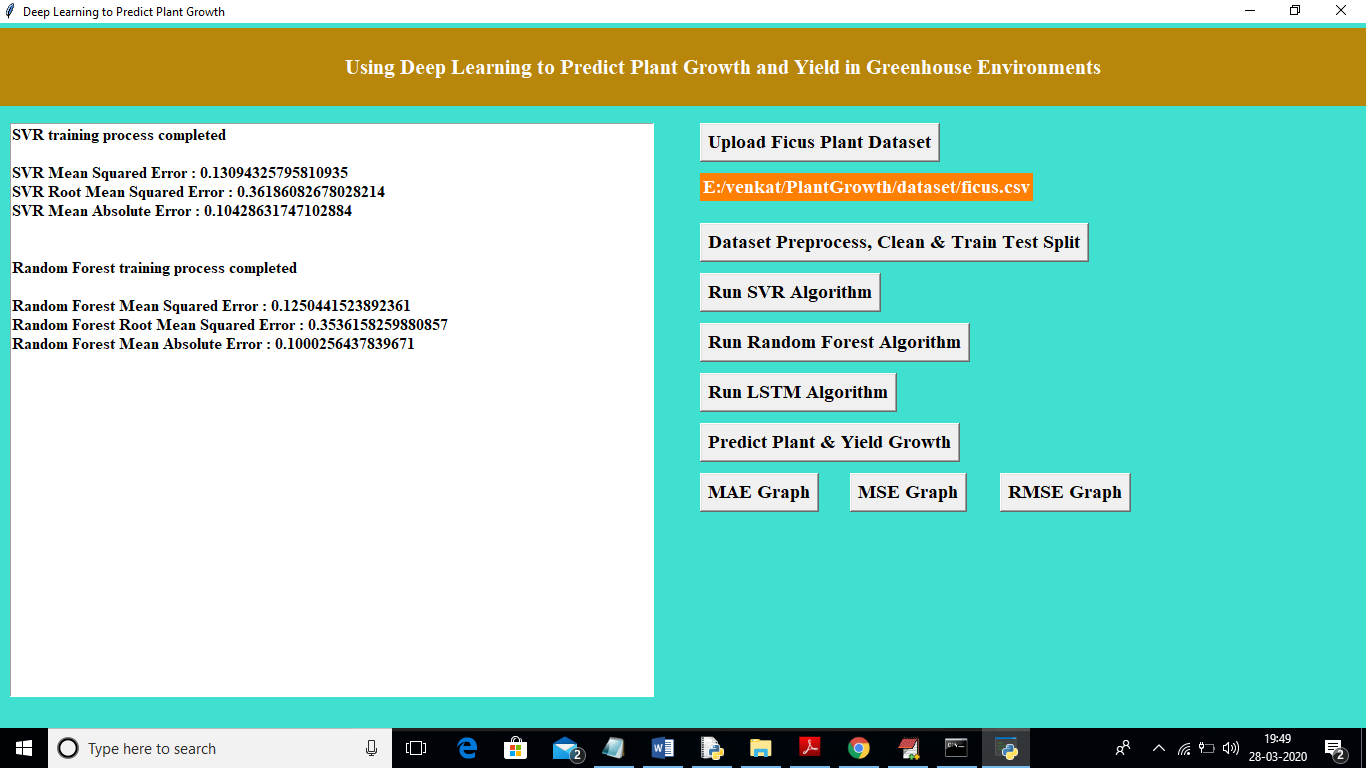
In above screen we can see dataset loaded and dataset contains total 4028 records. Now click on ‘Dataset Preprocess, Clean & Train Test Split’ button to clean dataset and to split dataset into train and test part



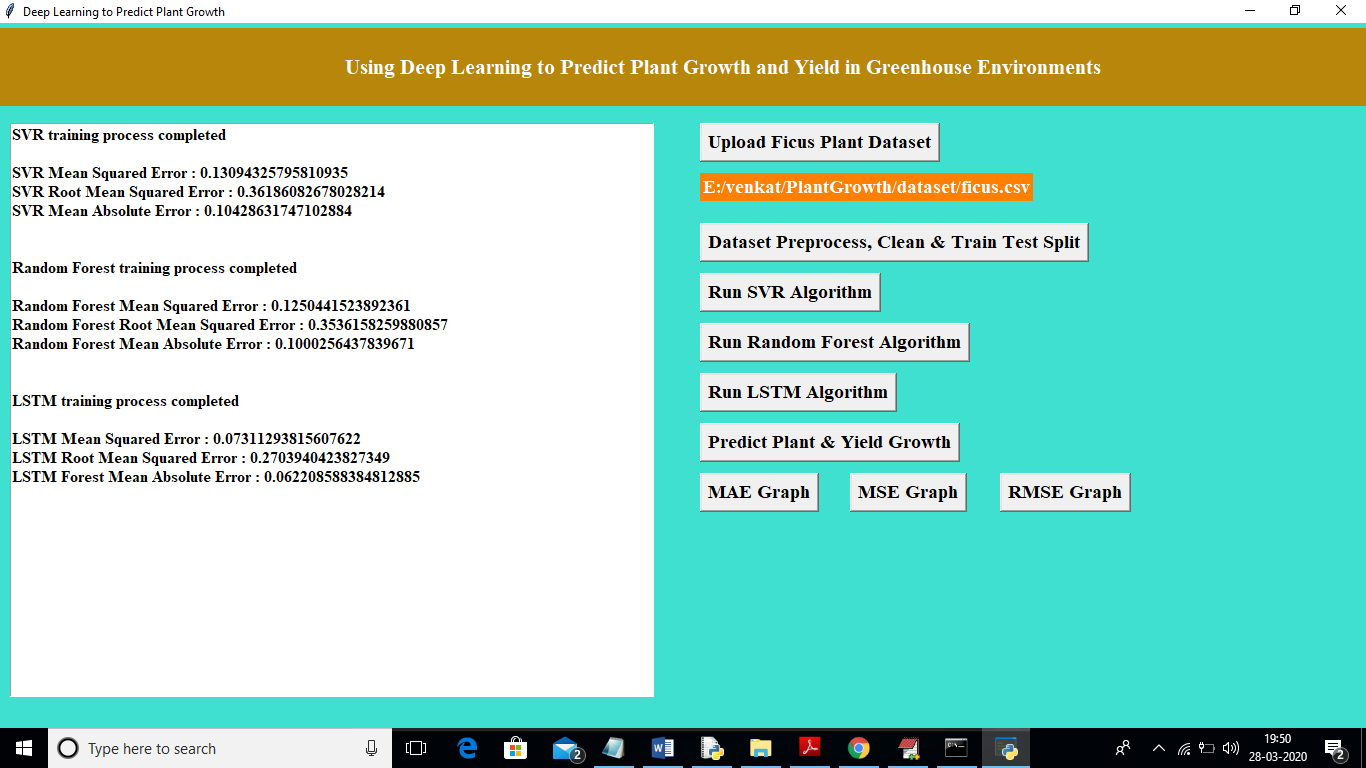
In above screen we can see application split dataset into 80 and 20% and application using 3222 records for training and 806 for testing. Now dataset loaded and splitted and now click on ‘Run SVR Algorithm’ button to train SVR algorithm



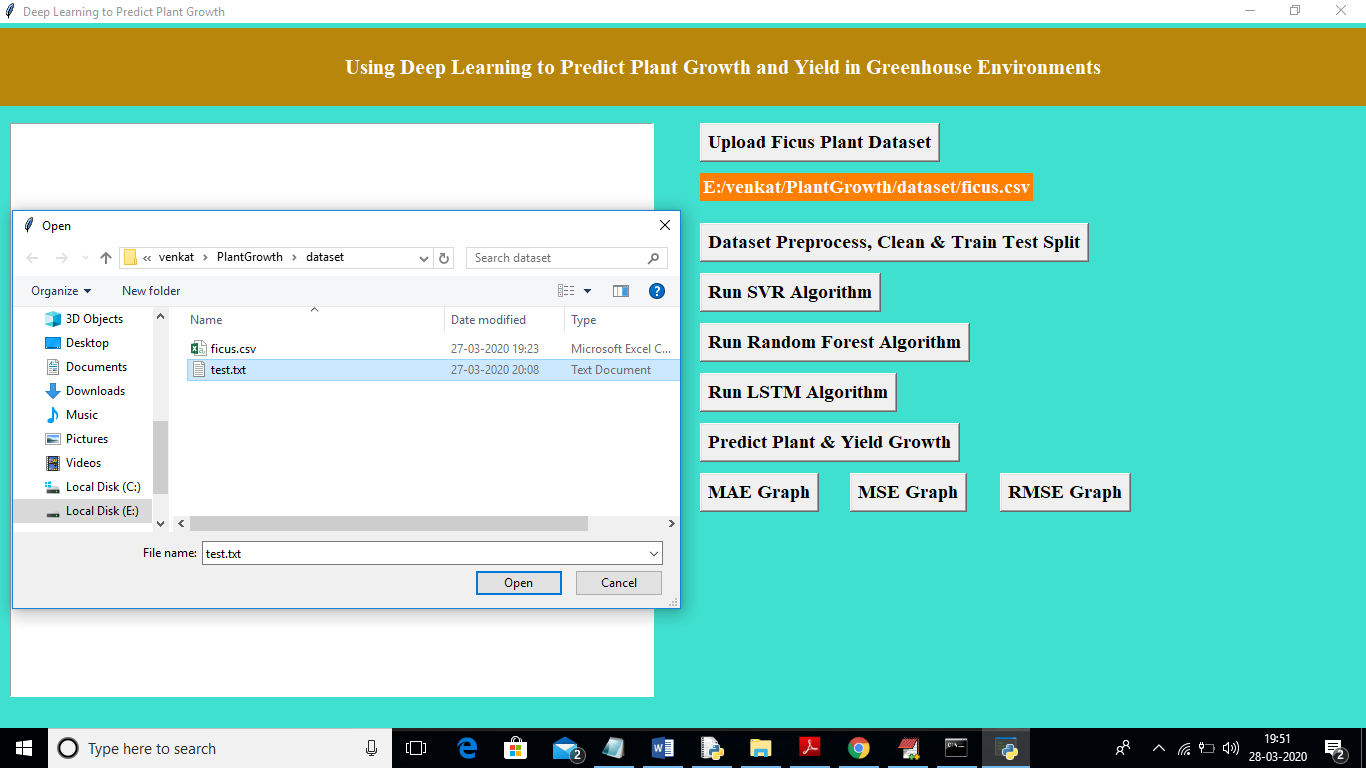
In above screen we got RMSE, MAE and MSE error for SVR algorithm and now click on ‘Run Random Forest Algorithm’ button to train random forest algorithm



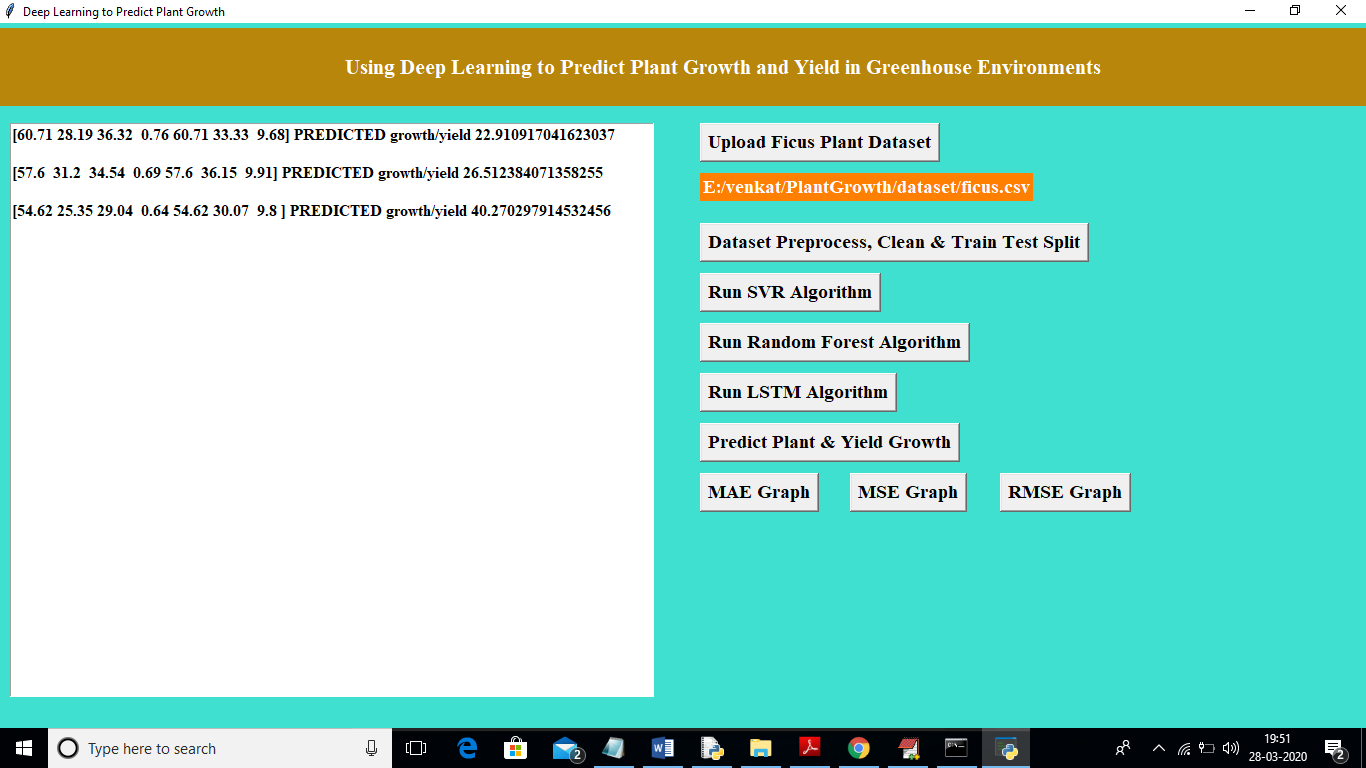
In above screen we got random forest MSE, RMSE, MAE error and now click on ‘Run LSTM Algorithm’ button to train dataset with LSTM algorithm



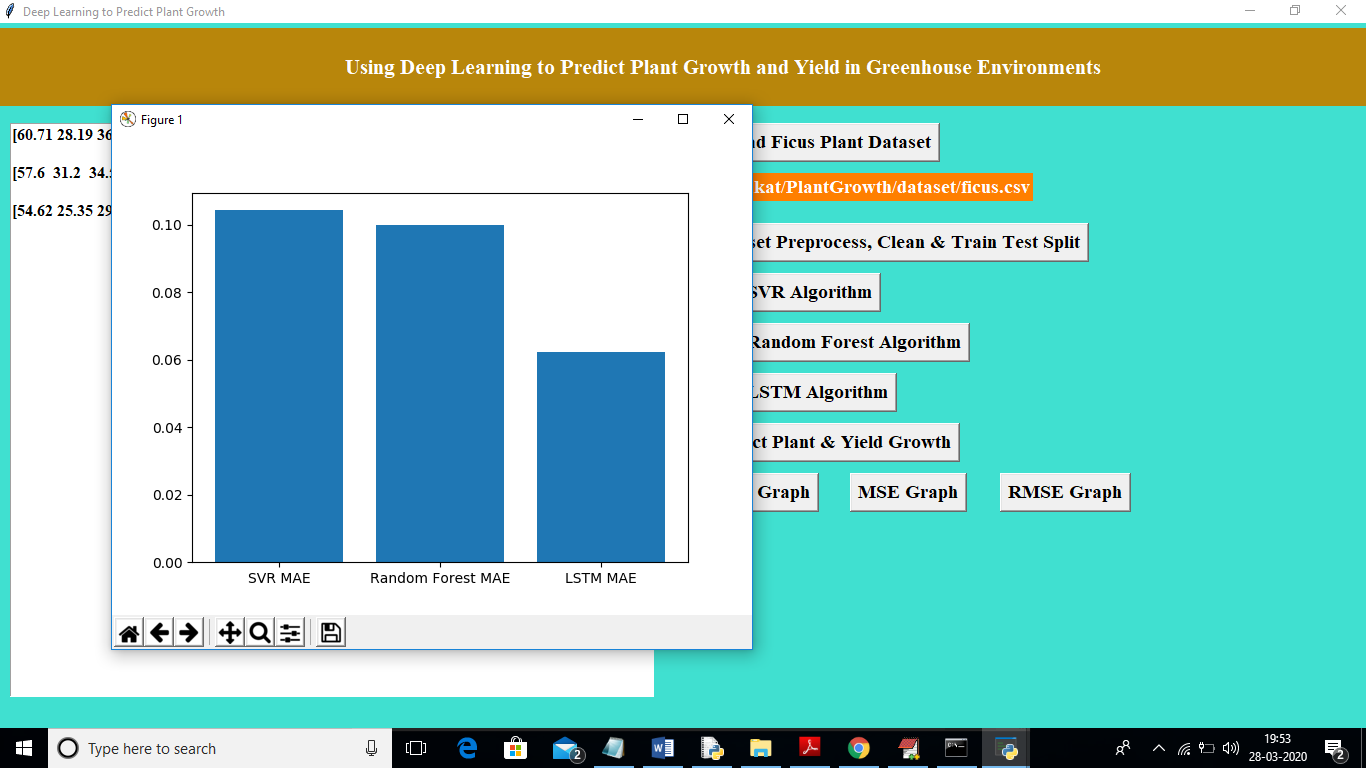
In above screen we can see LSTM got less MSE, RMSE and MAE error compare to traditional algorithm. Now all algorithms training process completed and now we can upload test file and predict its growth



In above screen I am uploading ‘test.txt’ file and now click on ‘Open’ button to predict growth for test data

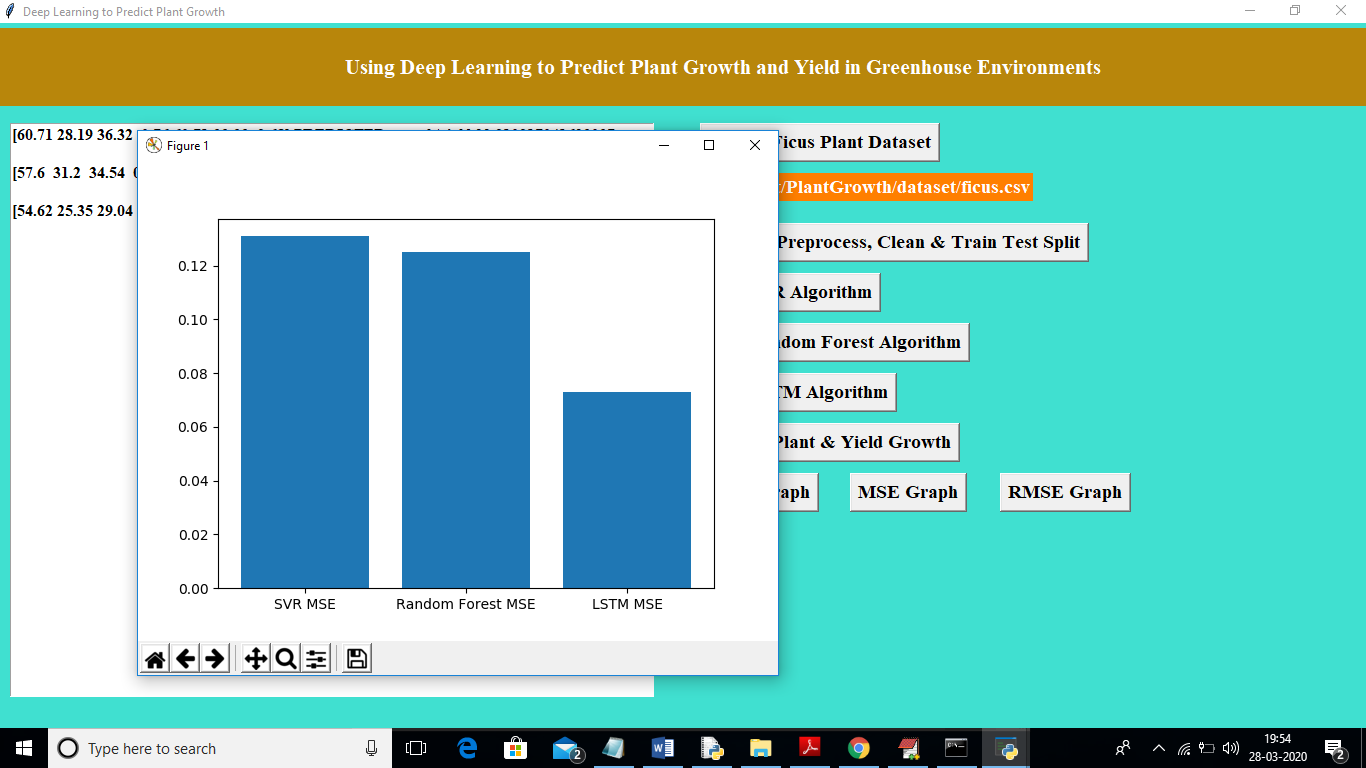


In above screen for first record growth prediction is 22% and second record 26% and third record having 40% growth prediction. Similarly u can add new records to test data and can predict its growth. Now click on ‘MAE Graph’ button to see MAE comparison graph between all algorithms

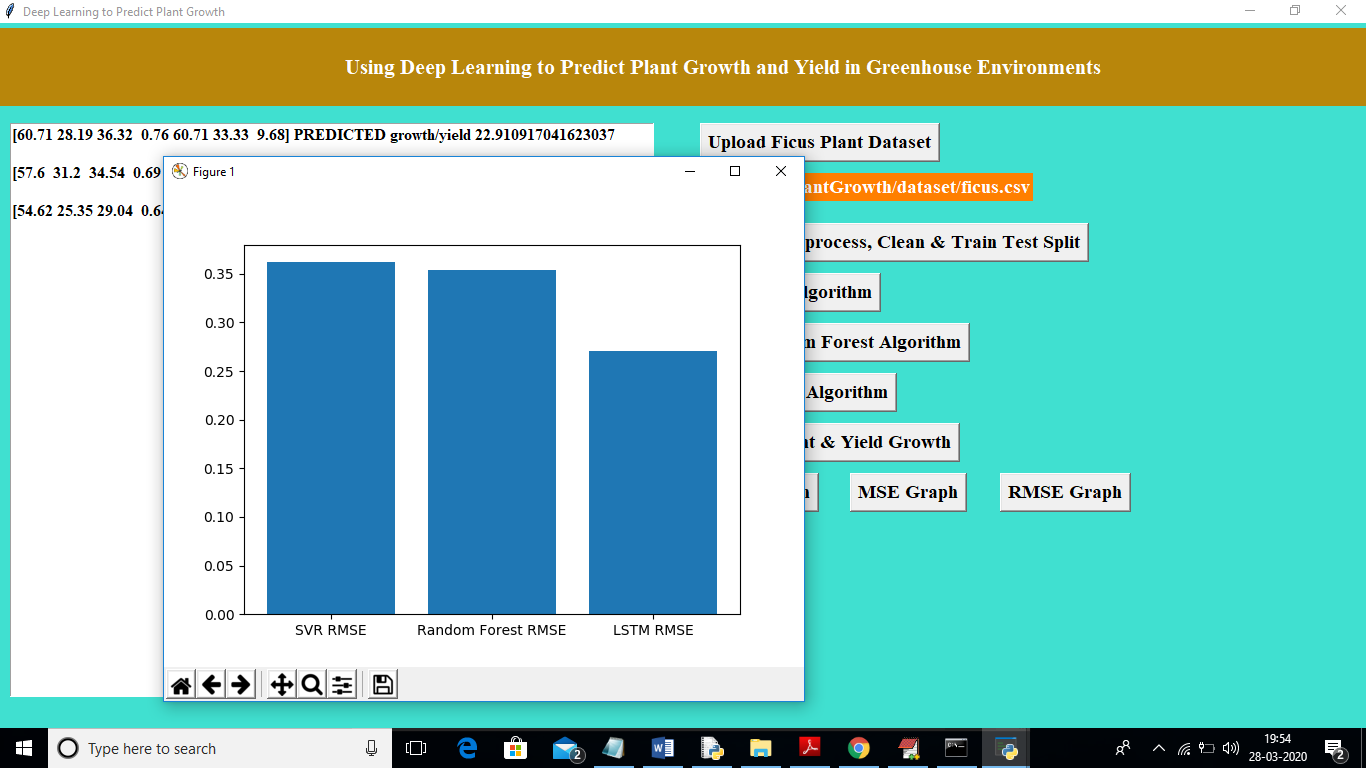


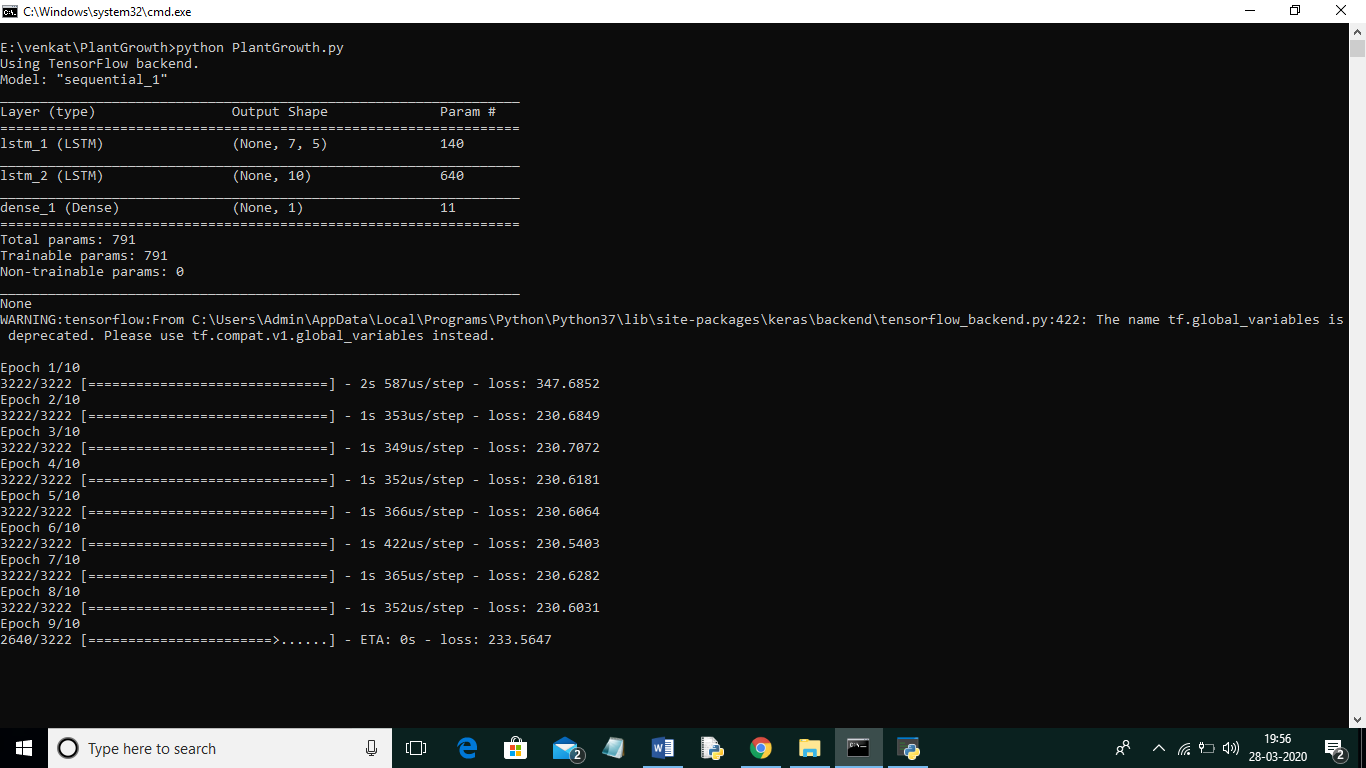
In above graph x-axis represents algorithm name and y-axis represents MAE error. From above graph we can conclude that LSTM got less error and its prediction performance will be best compare to other two.

Below MSG error graph



Below RMSE graph





In above black screen we can see training model generation for LSTM and to build this model I am using 10 epoch and in each epoch LSTM will use recent data to train model and forgot old data refrence