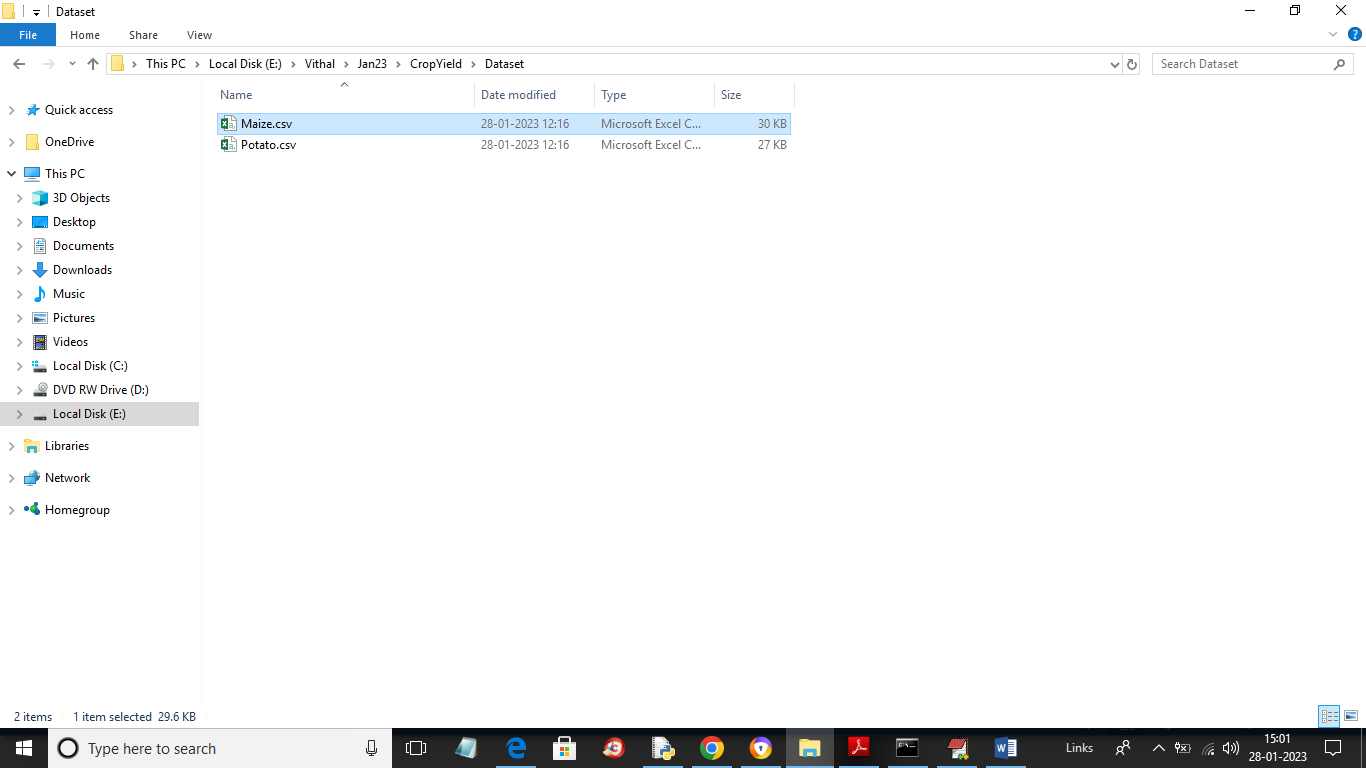
Crop Yield Prediction Using Machine Learning Models: Case of Irish Potato and Maize

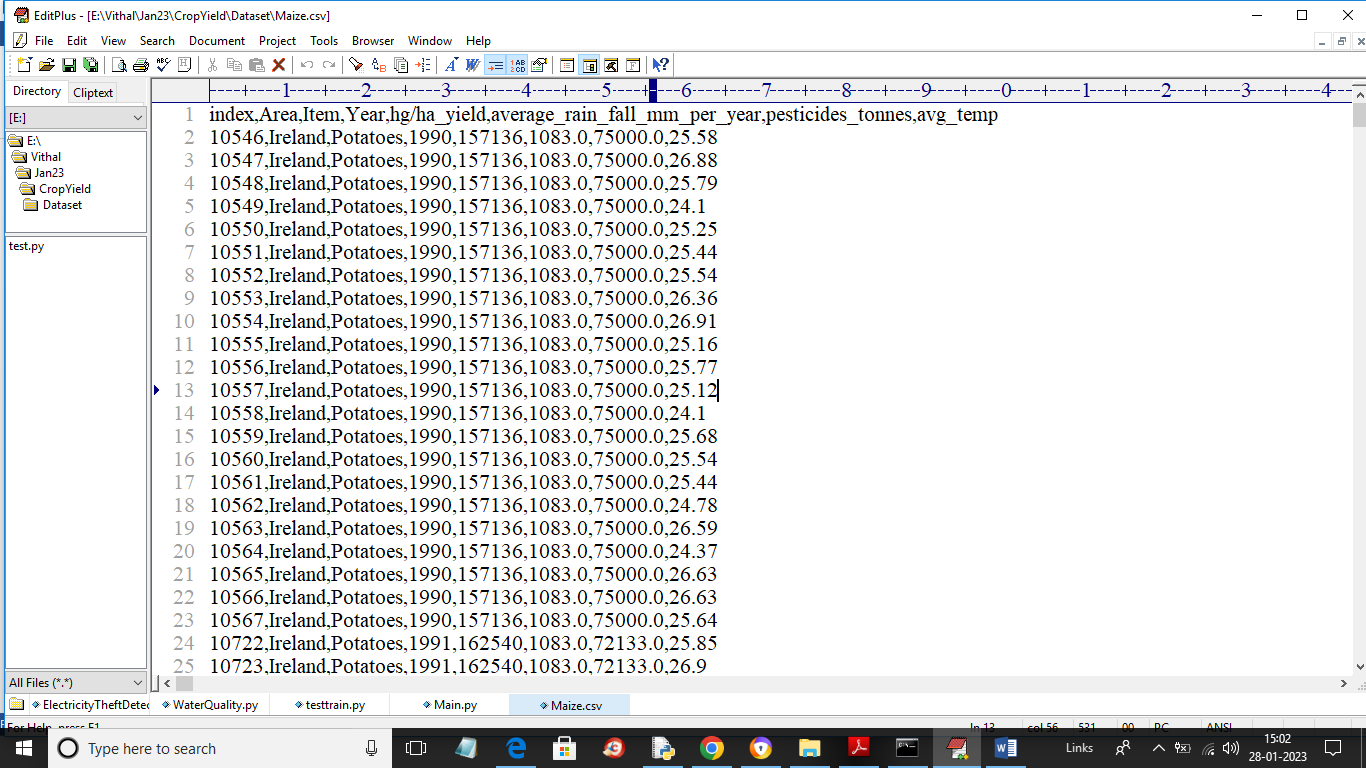
In this paper author employing Random Forest and SVR algorithm to predict future crop yield based on weather data such as temperature and rainfall. If farmers know the crop yield before sowing based on historical weather data then he may take better decision. So by employing machine learning algorithms we can inform farmers about future crop yield. In propose paper author using Irish Maize and Potato yield dataset to train all machine learning models and then this models can be used to predict future crop yield.

In propose paper author using Random Forest and SVR algorithm but you asked us to implement SVR, DNN, CNN, ANN and LSTM so we have implemented all 5 algorithms on both datasets. To evaluate performance of each algorithm we are calculating MSE and R2 Score where MSE refers to mean square error (difference between TEST crop yield and predicted yield). R2 refers to correct prediction rate. So for any algorithm MSE must be lower and R2 must be higher.

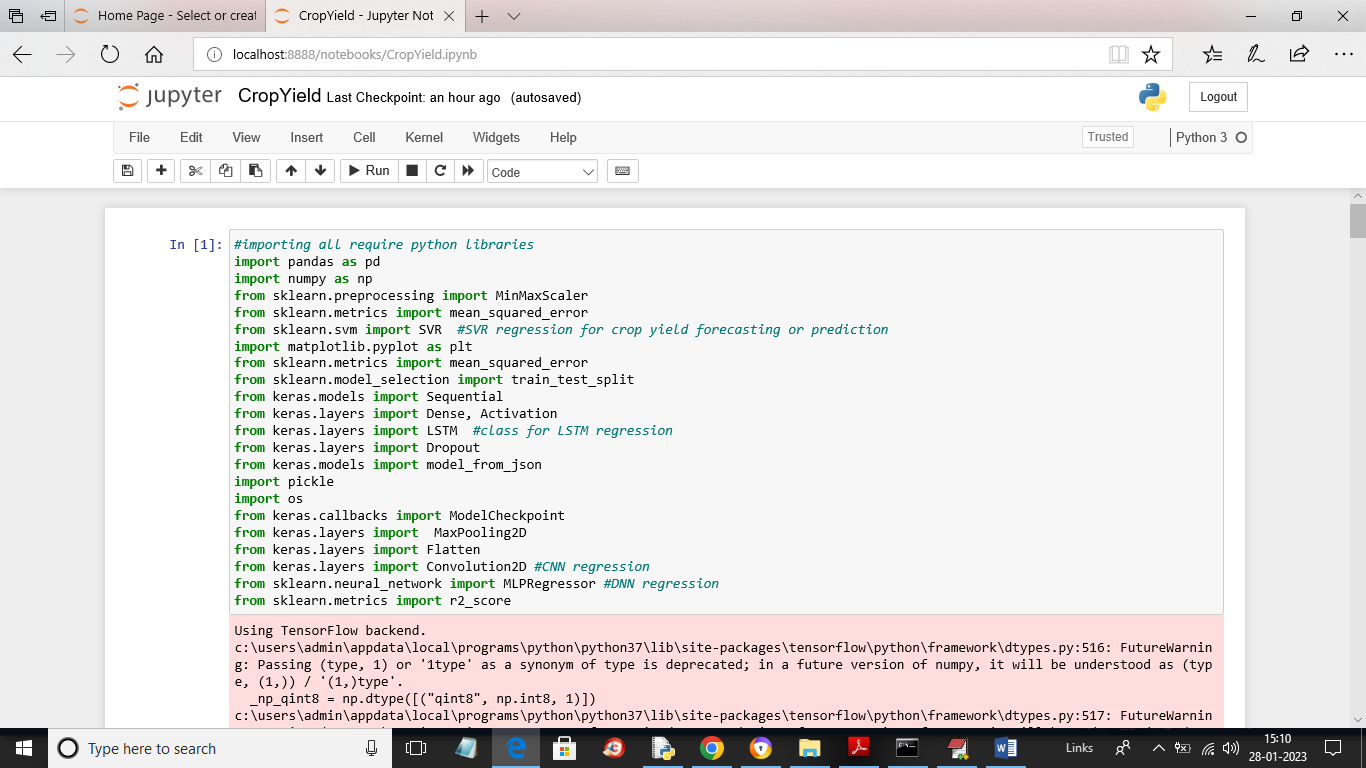
In below screen we are showing dataset details used in this project



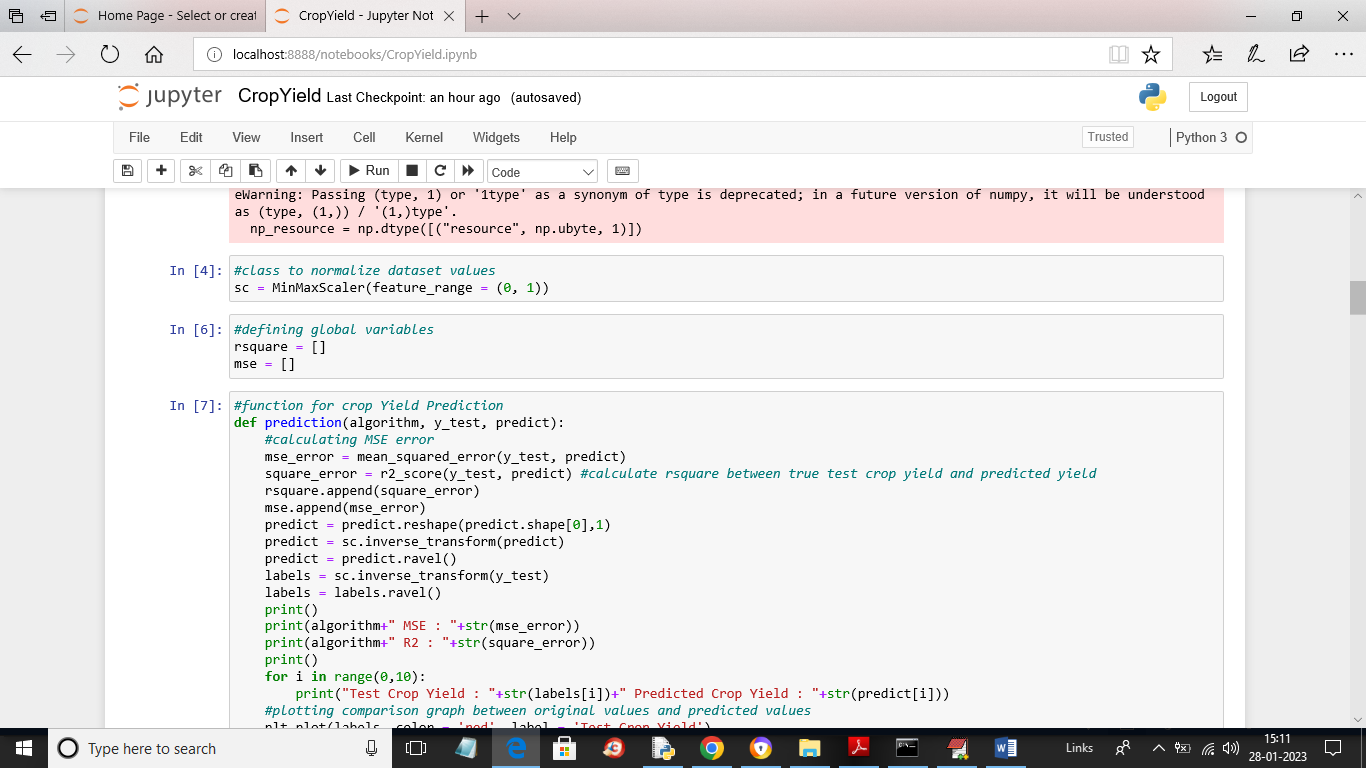
In above dataset folder we have 2 folders called as Maize and Potato and just open those files to view below data



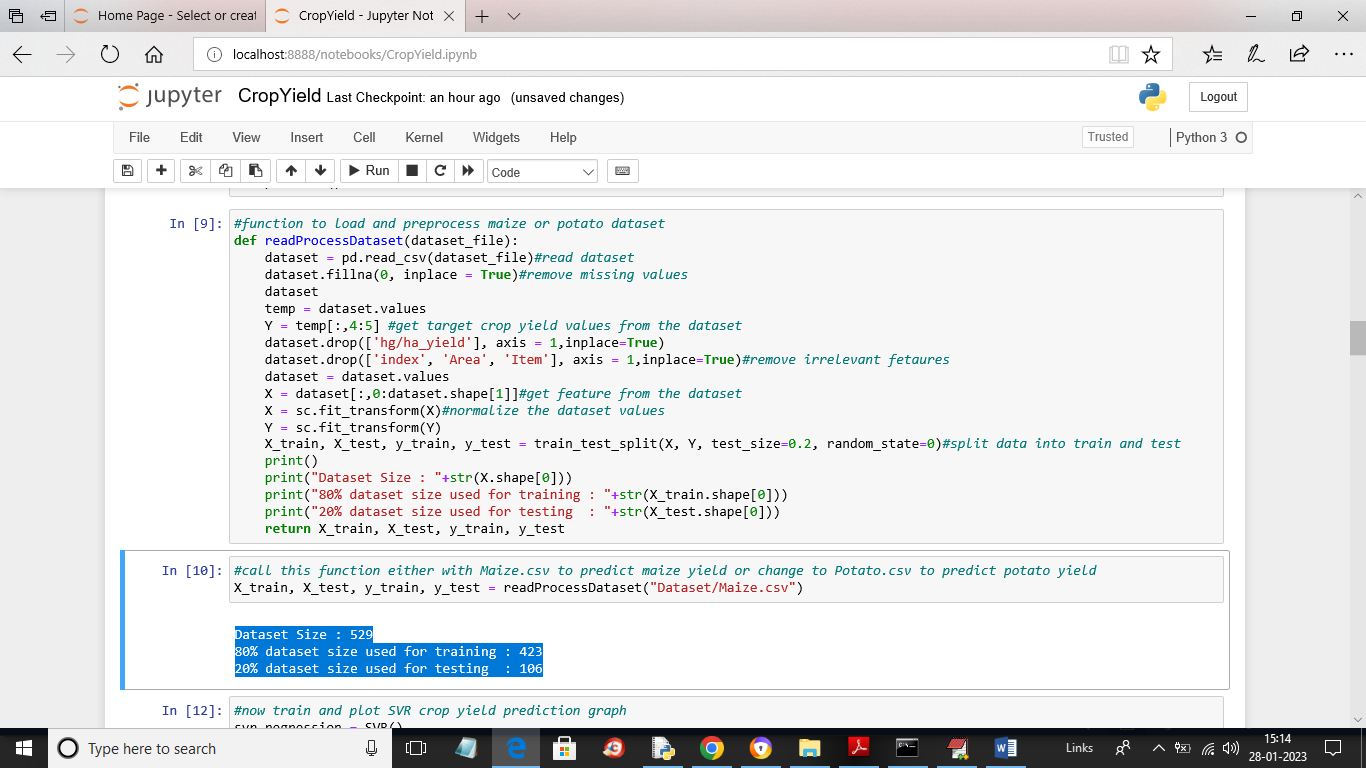
In above dataset screen first row contains dataset column names and remaining rows contains dataset values. We have coded this project using JUPYTER notebook and below are the output screens with code and below colour comments



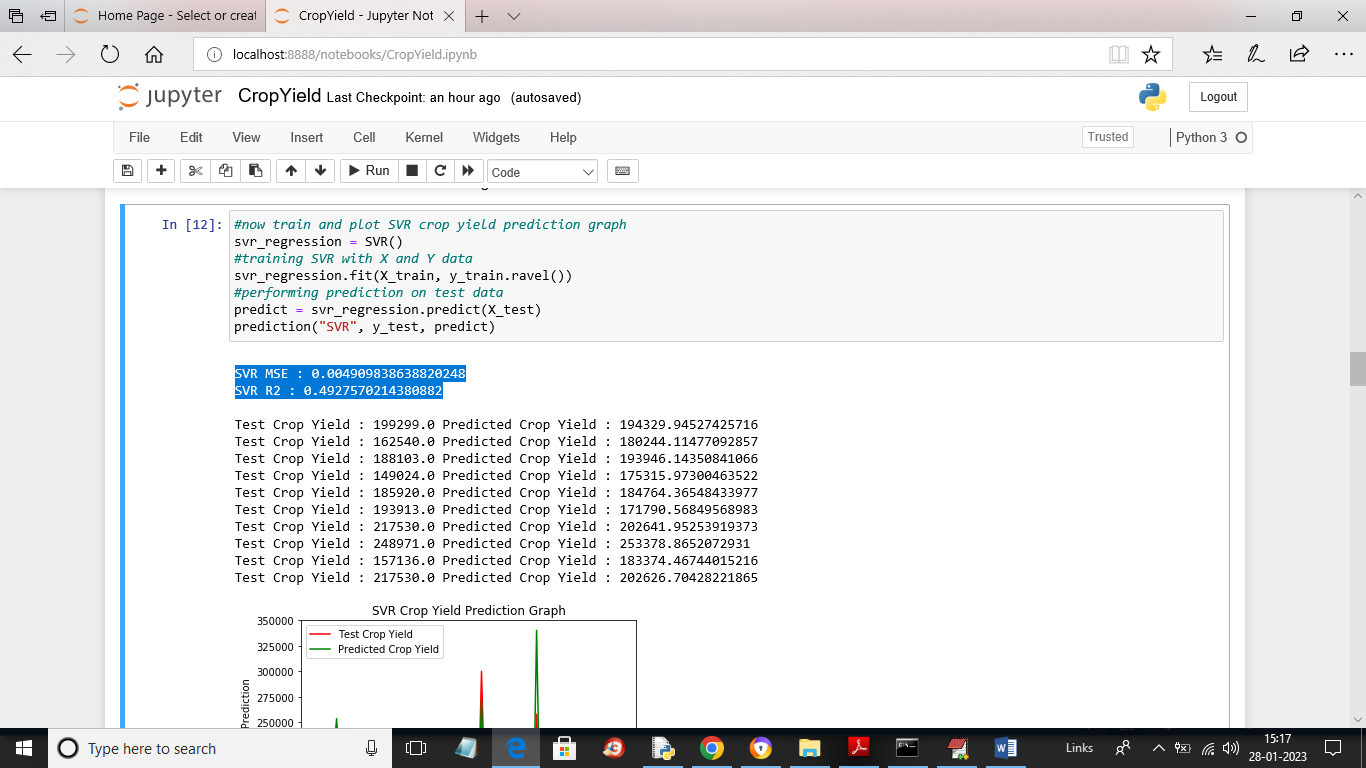
In above screen we are importing all require python packages



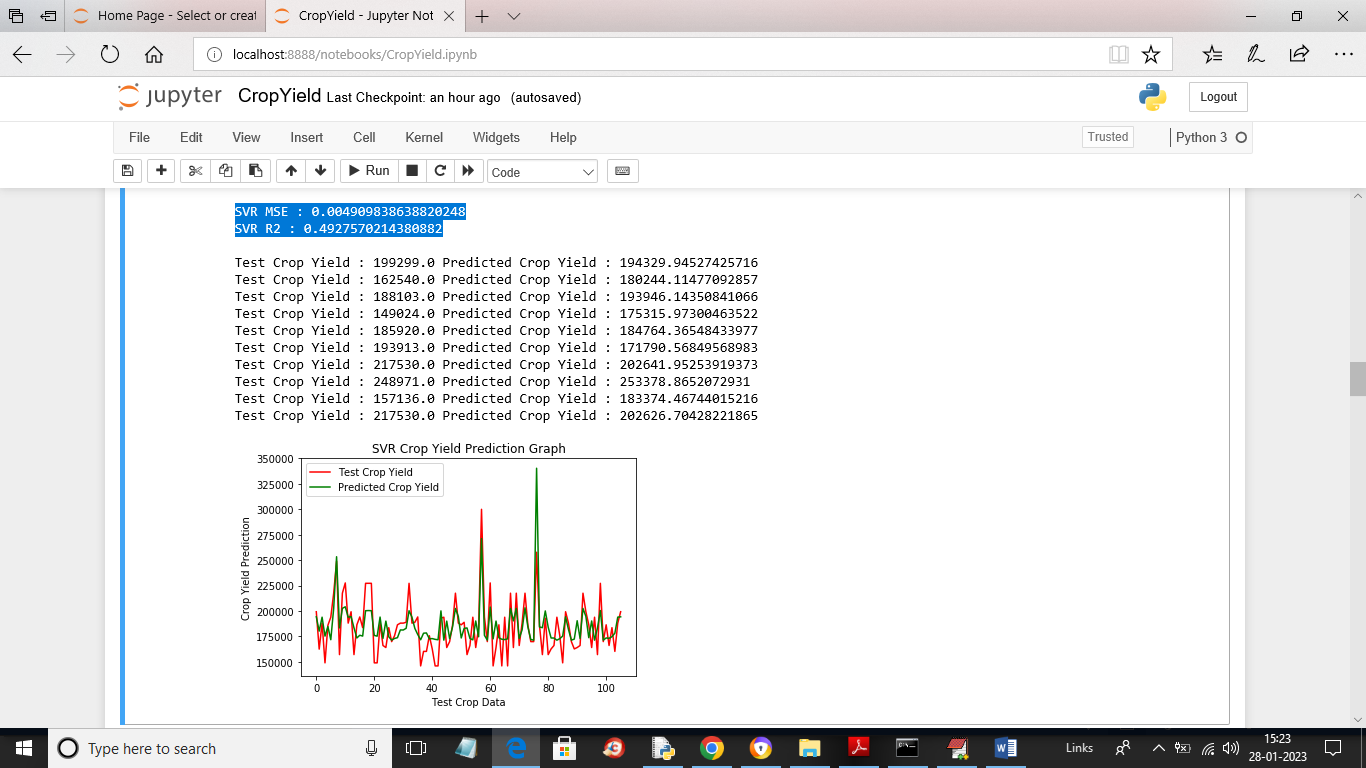
In above screen in first block defining MINMAX class to normalize dataset values and then defining MSE and square variables and then defining function predict and plot Crop Yield graph



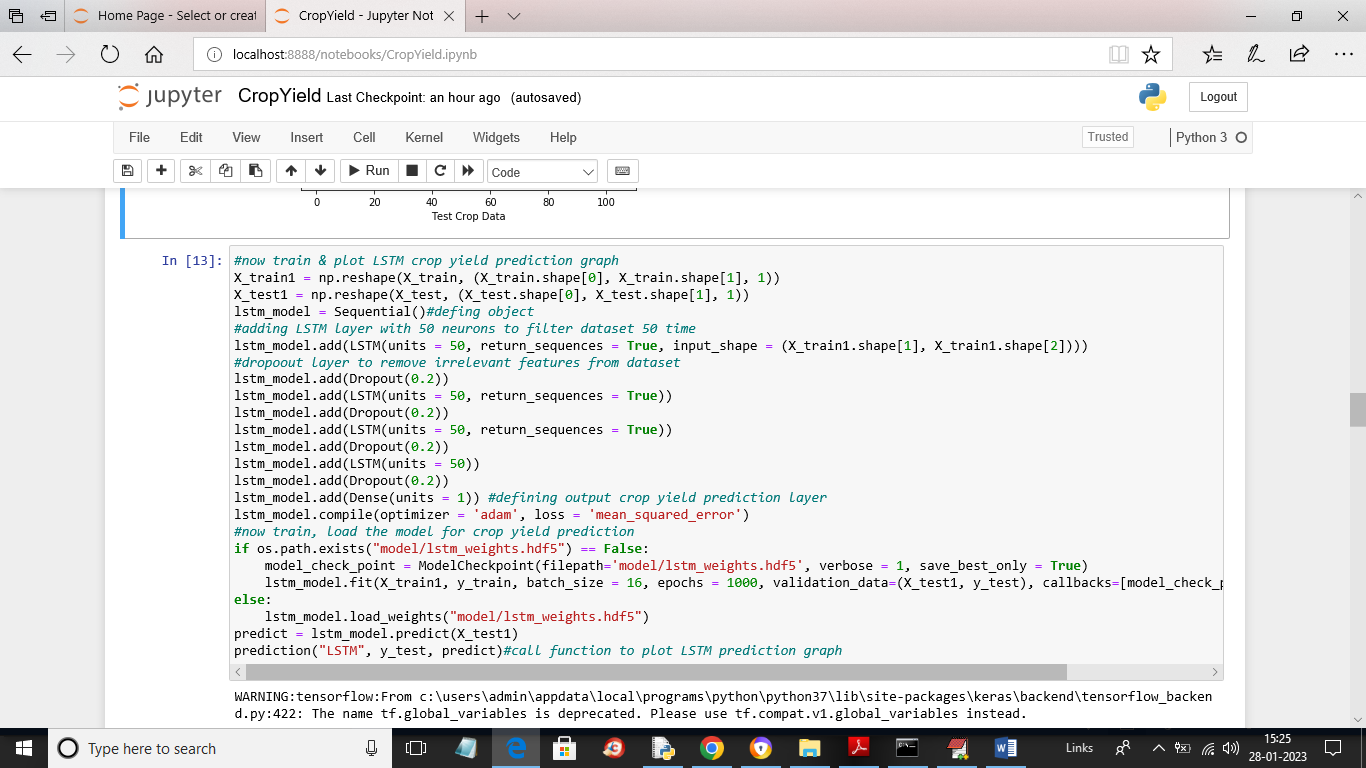
In above screen we are calling Read dataset function to read and process dataset and then get train and test data and now using train and test data we will train and test performance of all algorithms and in blue colour we can see total dataset size and training and testing size



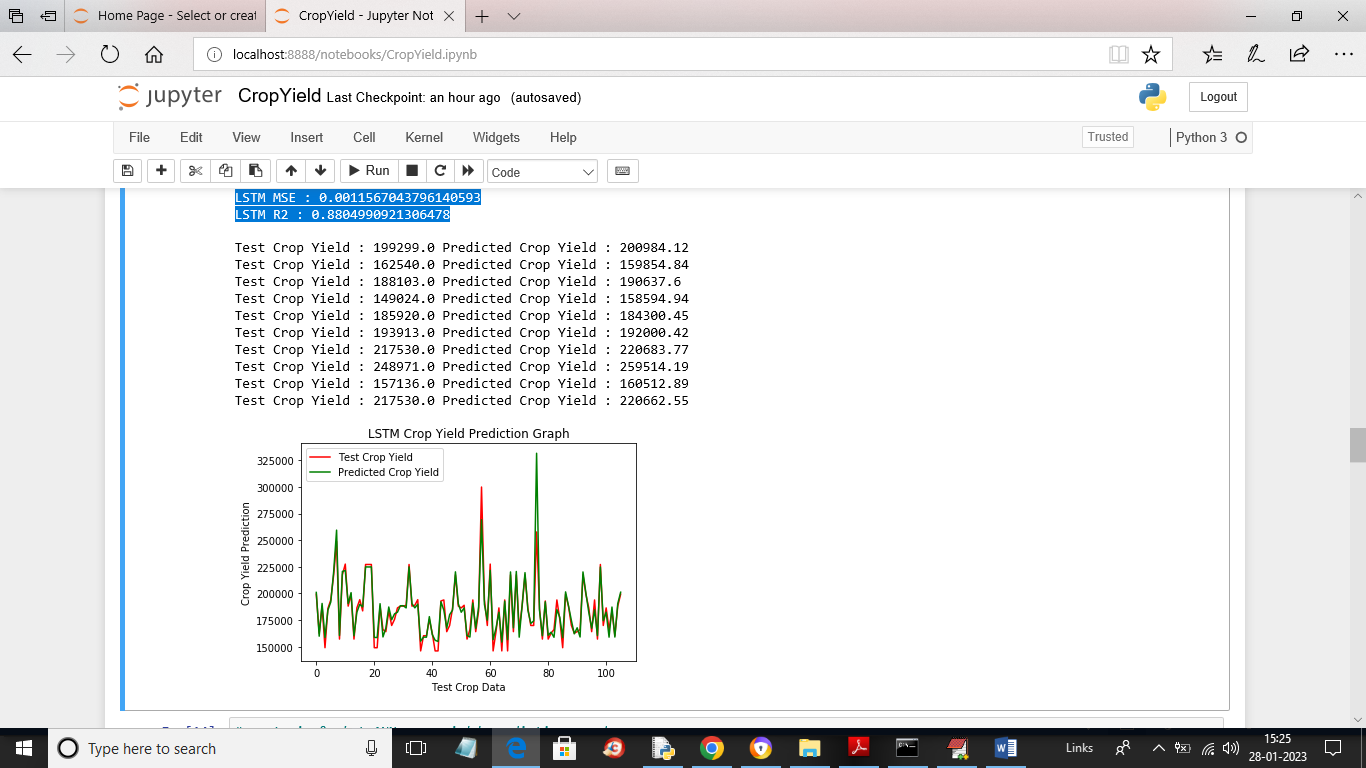
In above screen we are training with SVR algorithm and then in blue colour text we can see SVR MSE and R2 values and then in next lines we can see TEST crop yield values and SVR predicted crop yield values and in below screen we can see Test yield and predicted yield graph



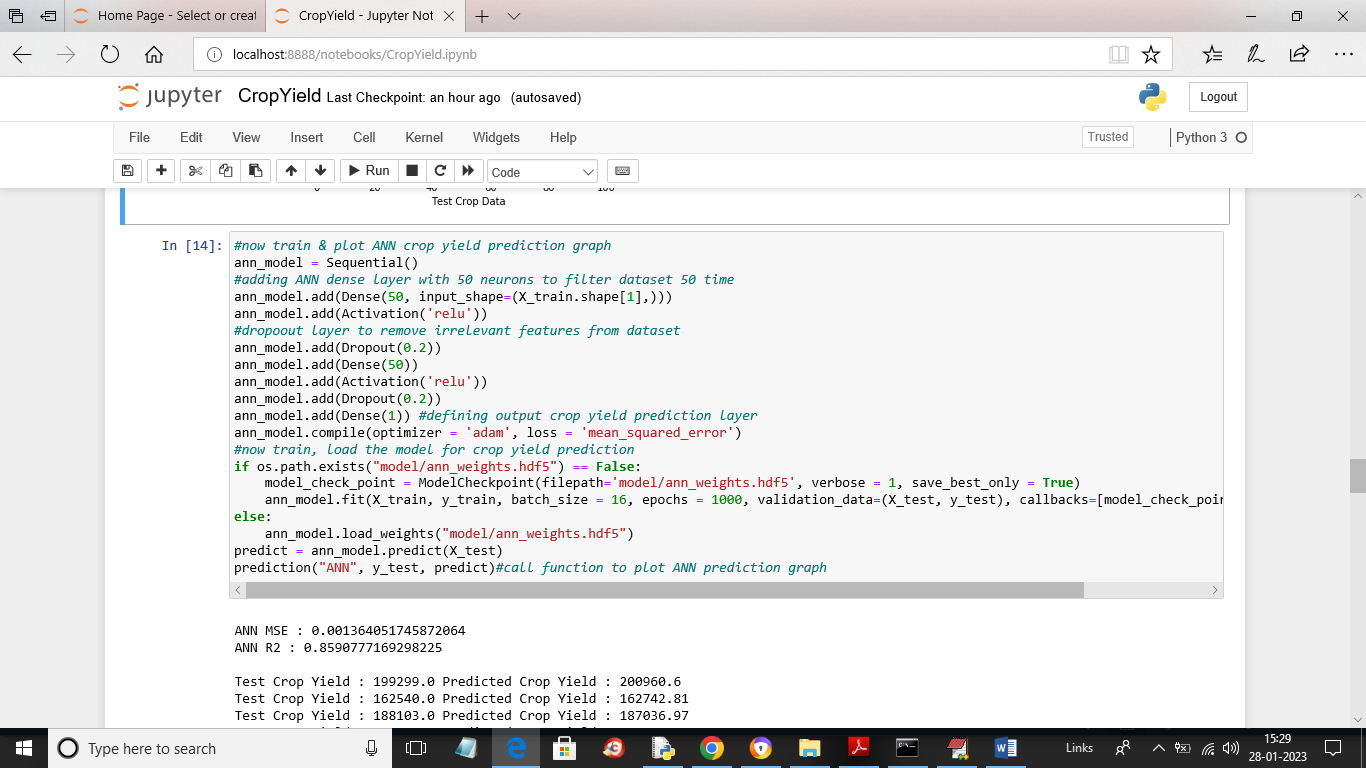
In above screen x-axis represents number of test records and y-axis represents Yield values where red line represents TEST yield and green line represents predicted yield and in above graph we can see there lots of gap between red and green line so SVR prediction is not accurate



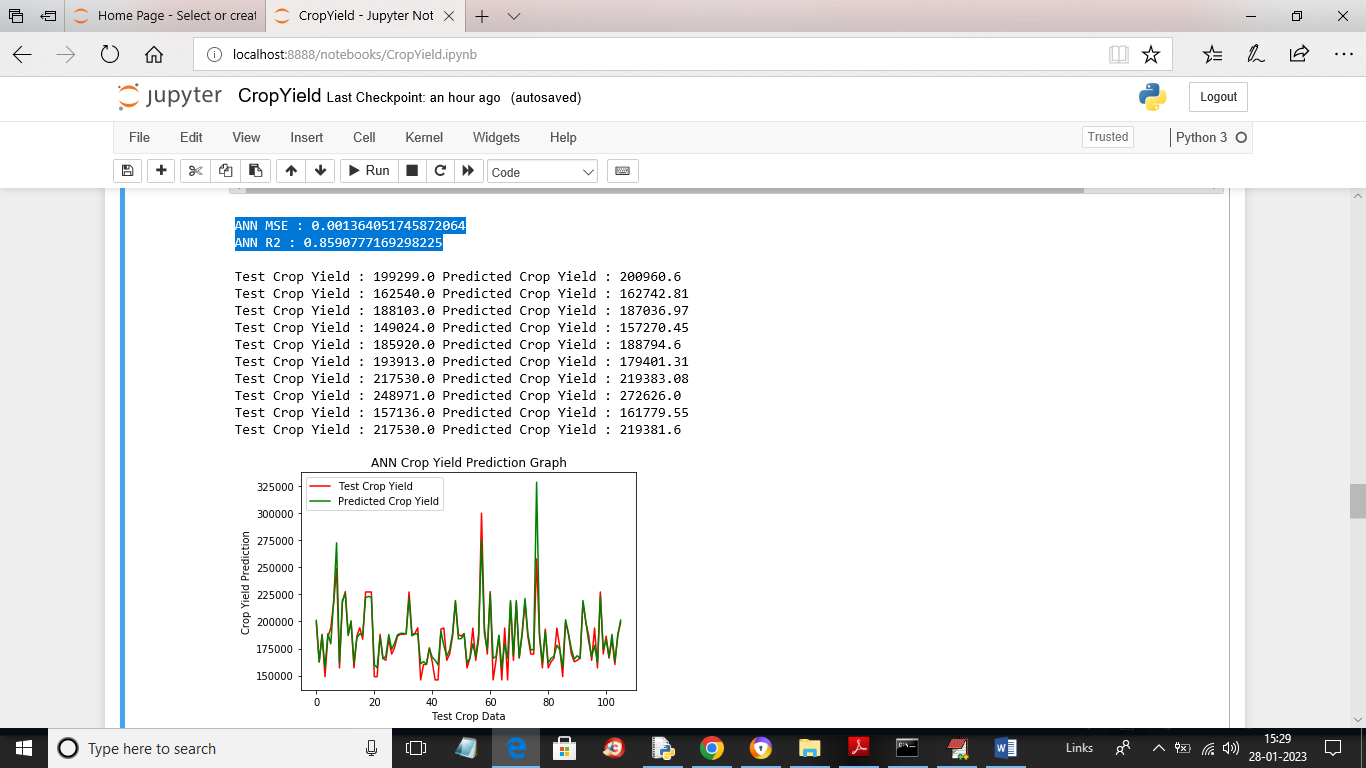
In above screen we are defining and training with LSTM and after executing above block will get below output



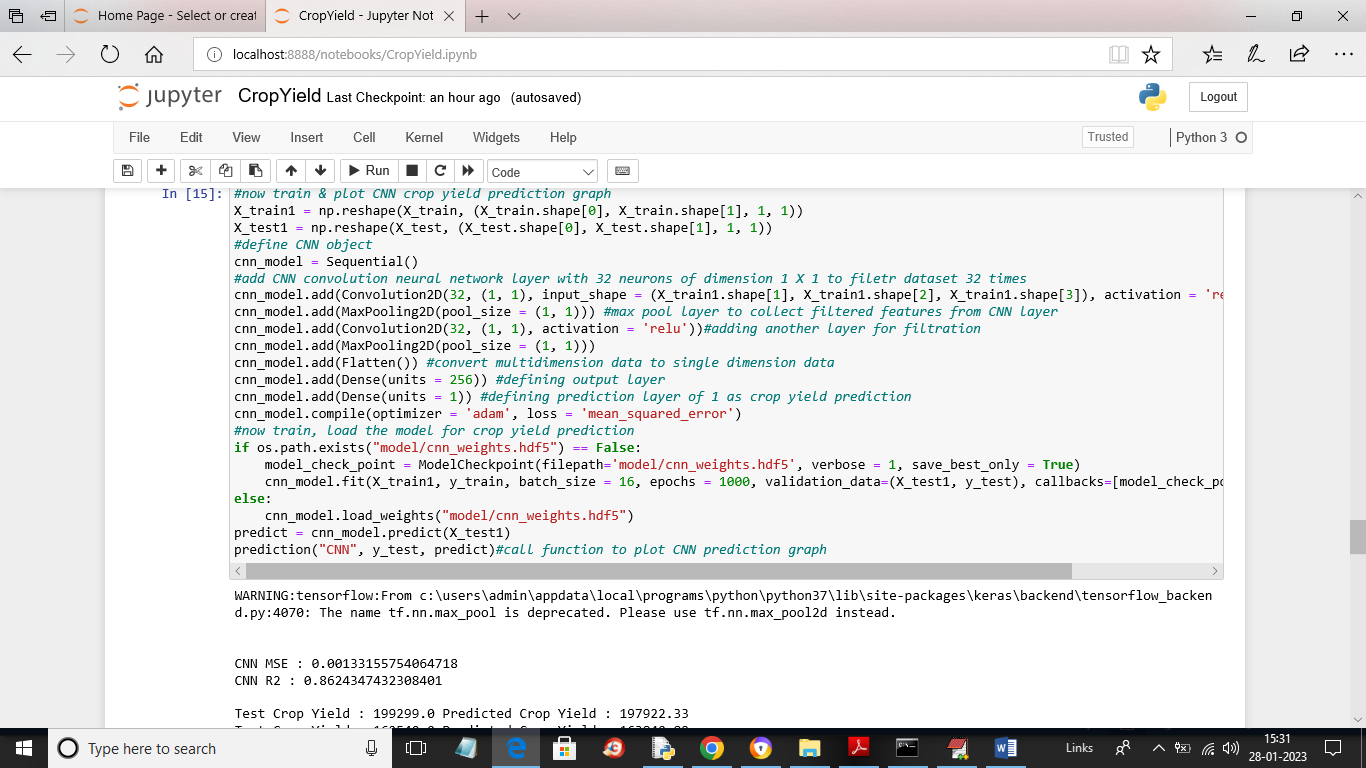
In above screen in blue colour we can see LSTM MSE and R2 values and then in next lines we can see TEST crop yield and LSTM predicted crop yield and in LSTM graph we can see both green line and red is fully overlapping so LSTM prediction is accurate



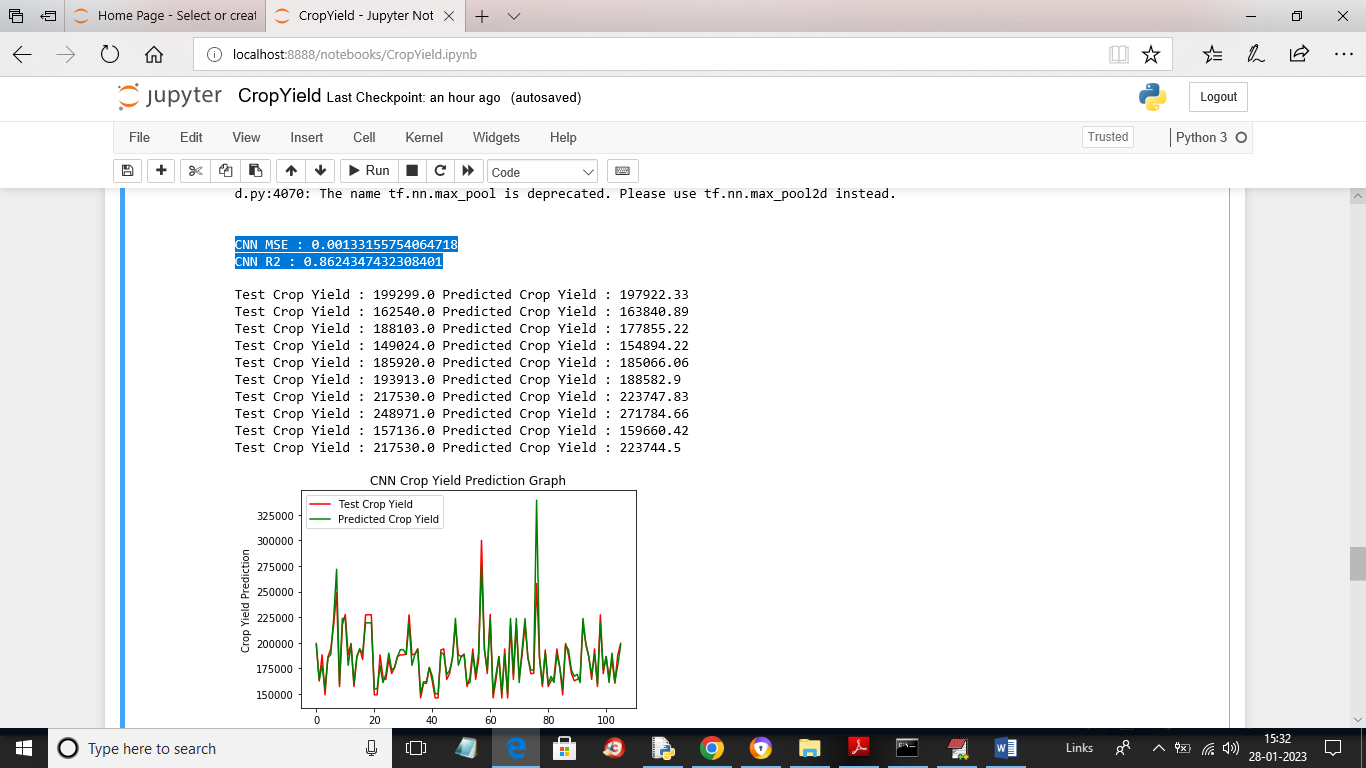
In above screen we are training with ANN model and after executing above block will get below output



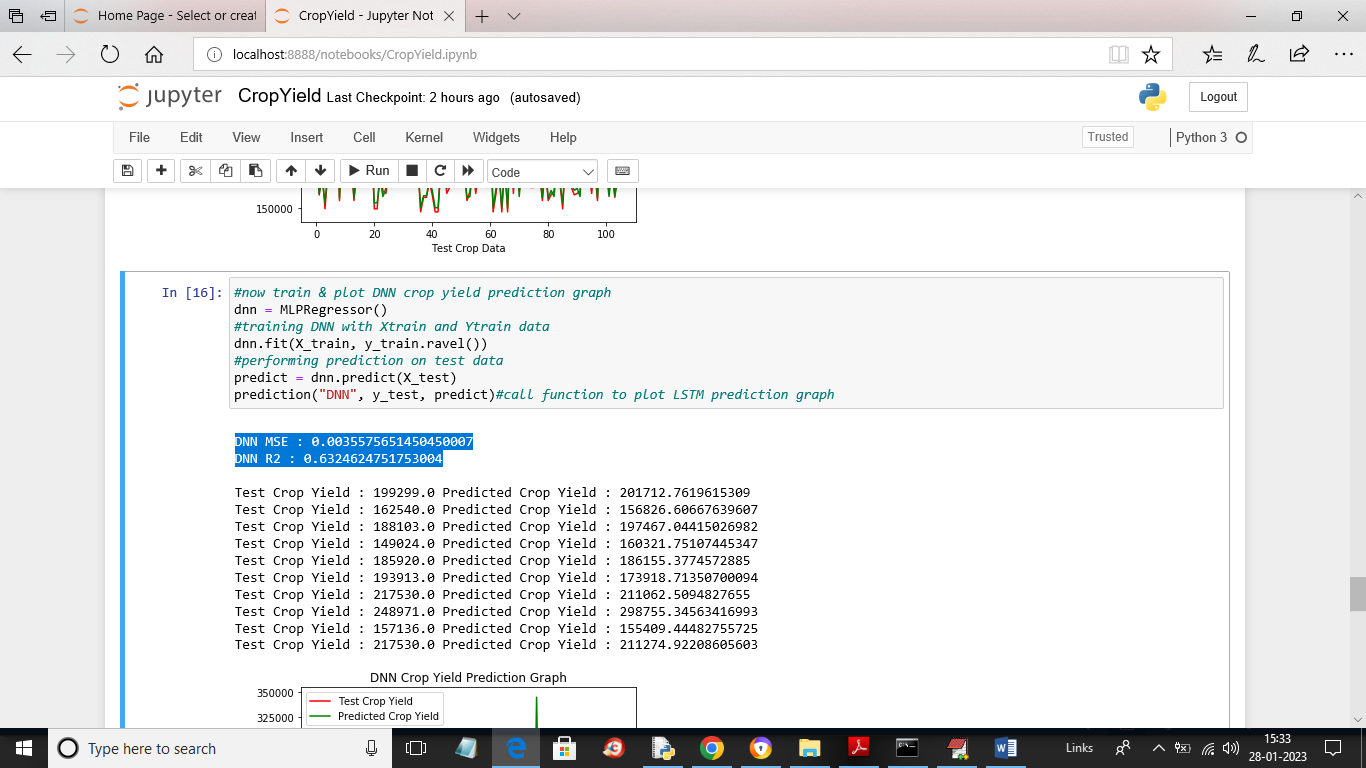
In above screen in blue colour text we can see ANN MSE and R2 values and then in next lines we can see TEST and predicted crop yield for ANN and then in ANN graph we can see both lines are fully overlapping so ANN prediction also accurate



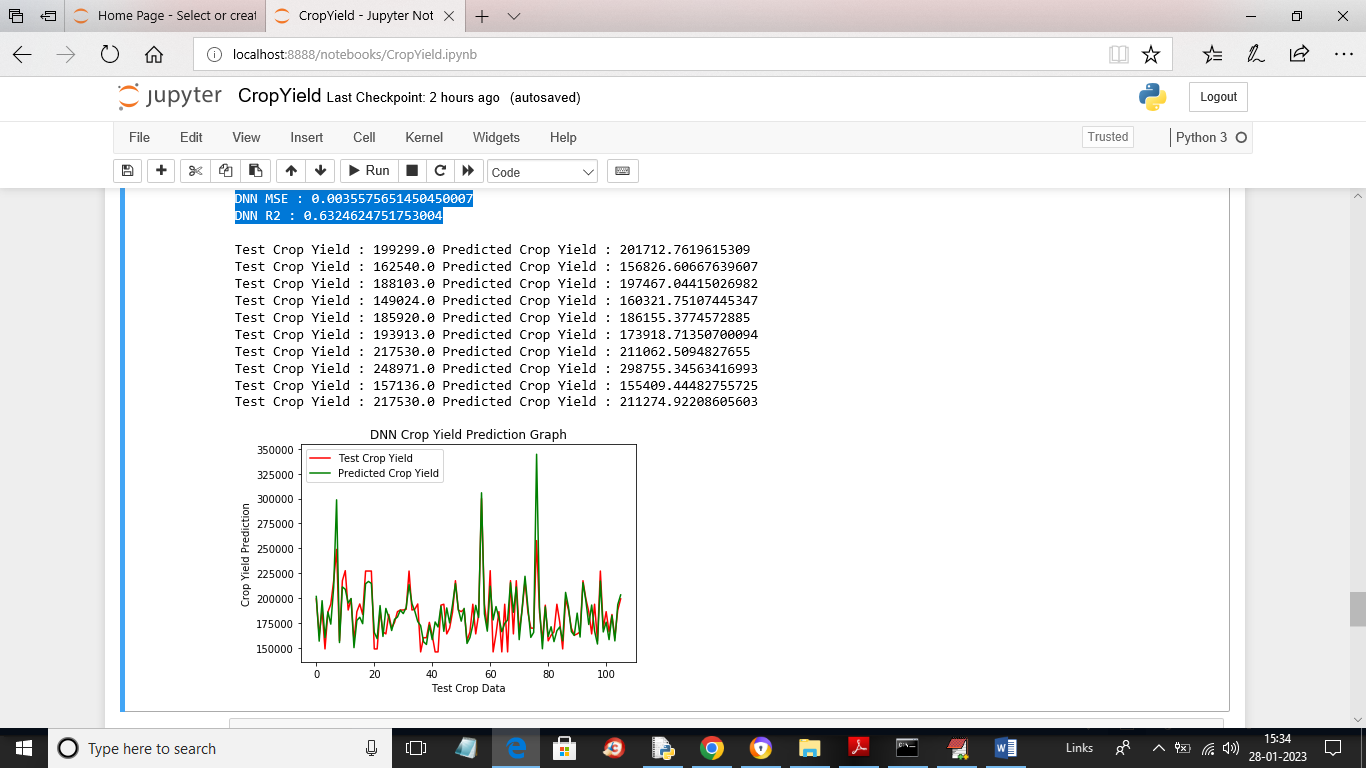
In above screen we are training with CNN and after executing above block will get below output



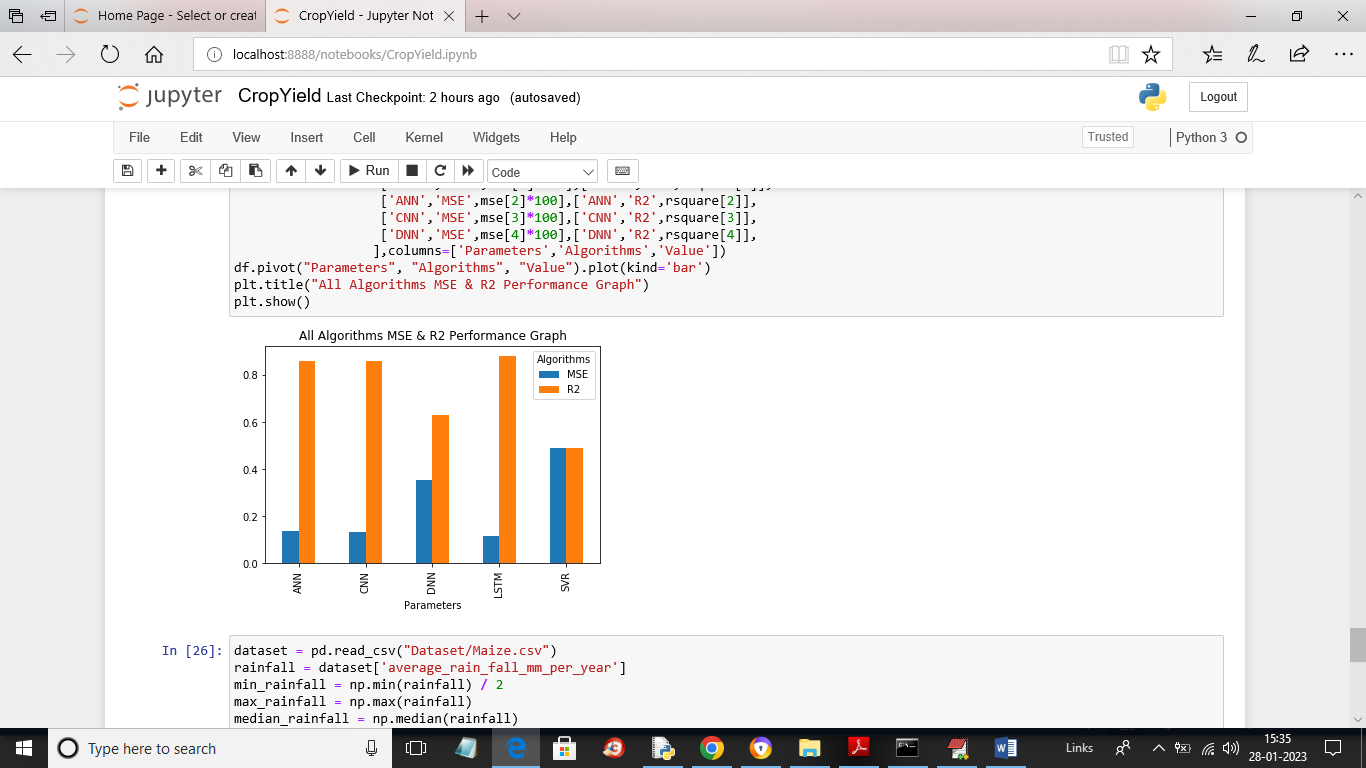
In above screen we can see CNN MSE and R2 values and then we can see TEST and predicted yield for CNN and then in graph we can see both lines are fully overlapping so CNN prediction is also accurate



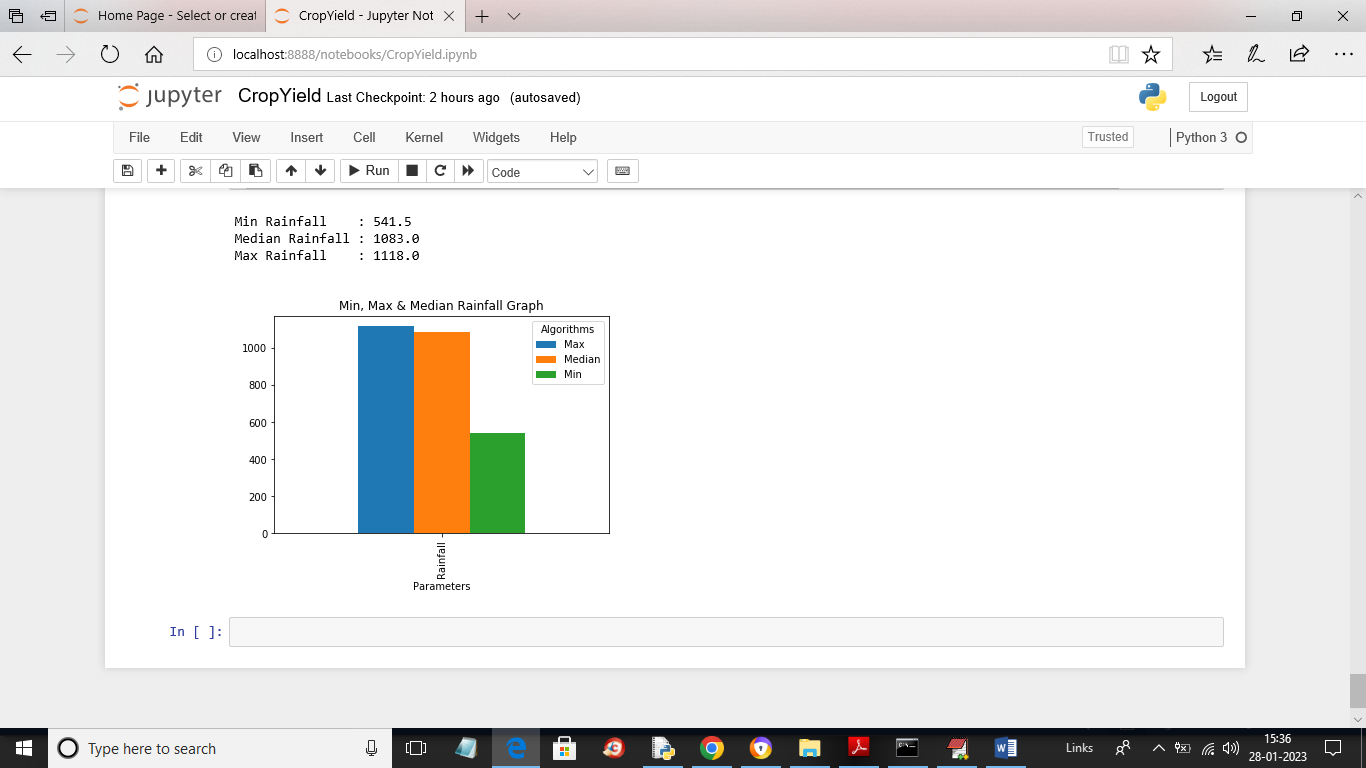
In above screen we are training with DNN and then we can see DNN MSE and R2 values and then we can see test and predicted crop yield for DNN



In above DNN graph there is little difference in red and green line as its contains some gap so DNN prediction is good but not accurate



In above graph blue bar represents MSE and orange bar represents R2 and x-axis represents algorithm names and y-axis represents values and in all algorithms we can see LSTM got high R2 and less MSE compare to all algorithms so we can say LSTM is good at crop yield prediction.



In above graph we are plotting MIN, MAX and median rainfall graph found in the dataset