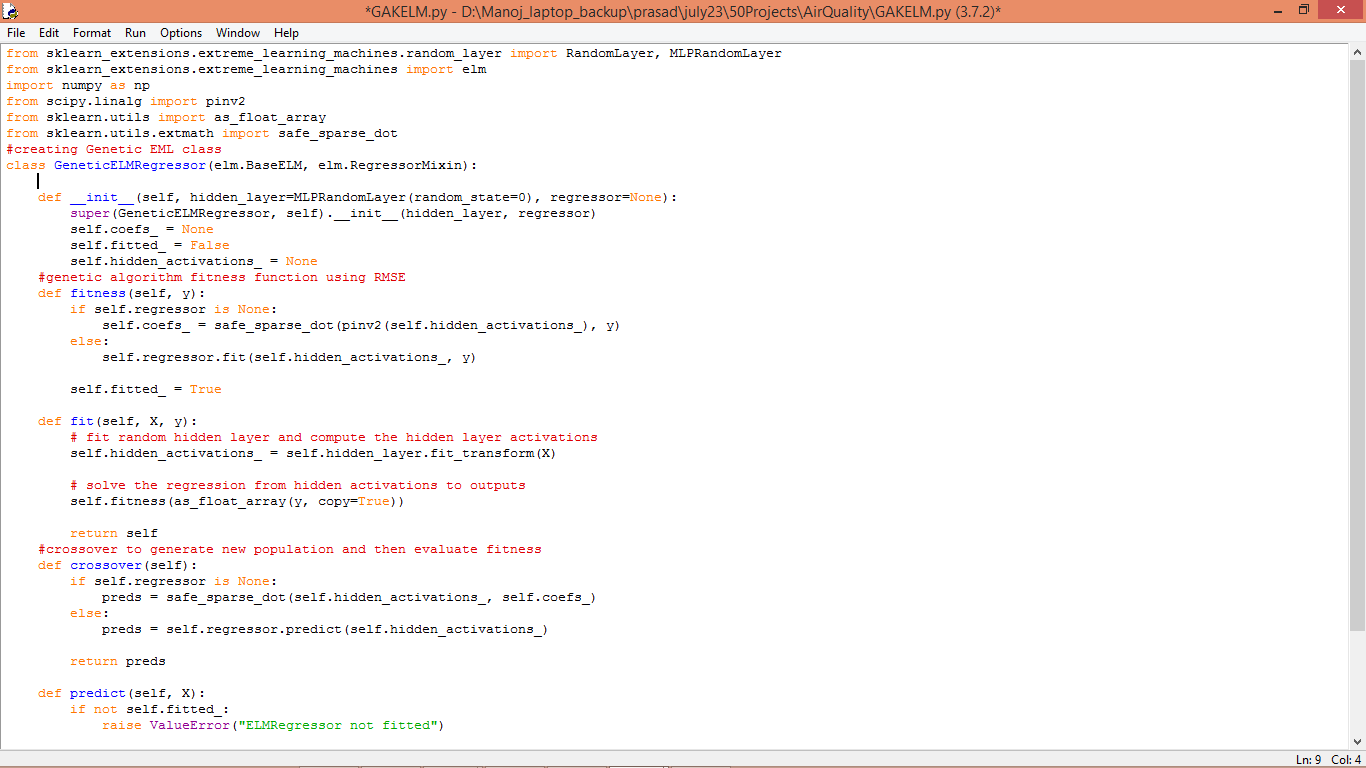
Air Quality Index Forecasting via Genetic Algorithm-Based Improved Extreme Learning Machine

For humans to live Air and Water is the basic requirement but because of pollutions both basic requirements are getting polluted and now it become mandatory to adjust pollution by monitoring Air Quality, if air quality is low then government can make necessary arrangement like reducing carbon emission and other dangerous chemicals.

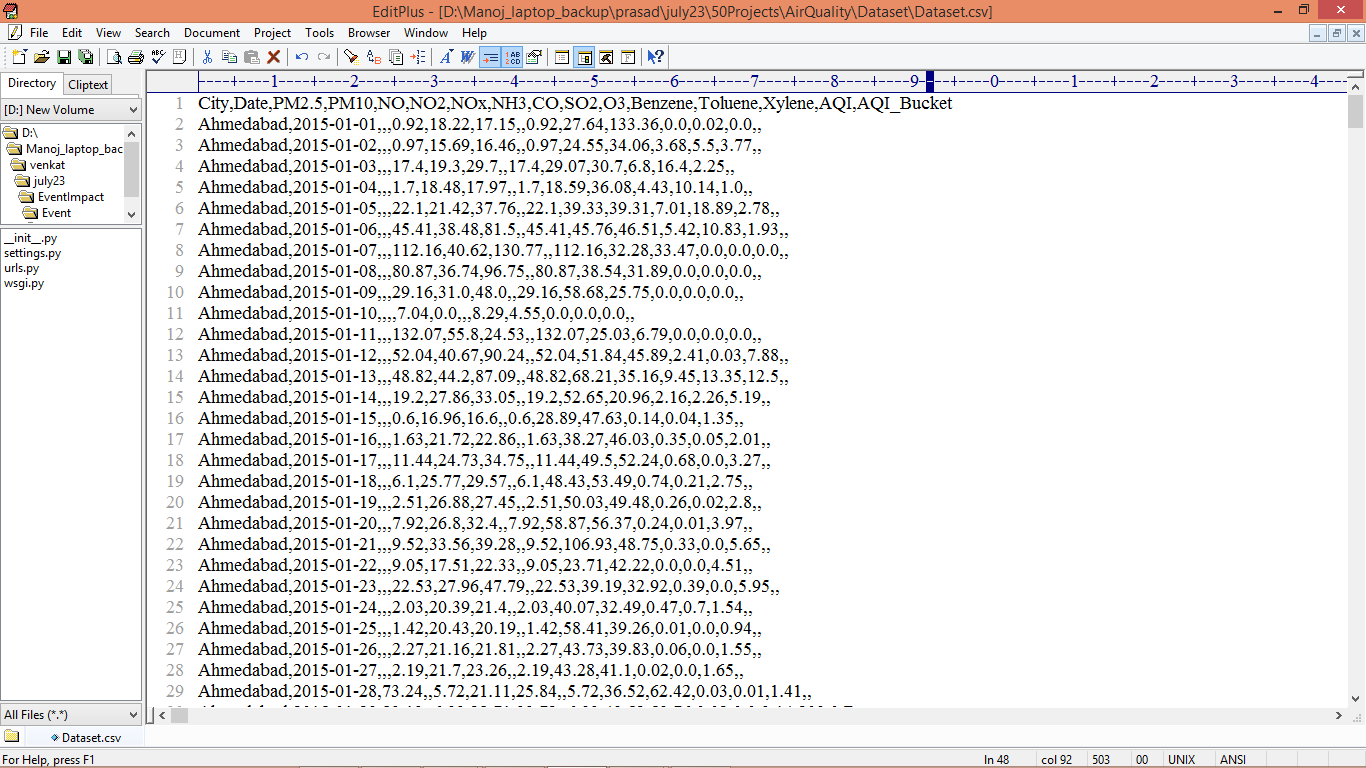
In the past many deep and machine learning algorithms are introduced but their performance is not accurate as training weights of those algorithms are not accurate enough to predict Air Quality with high accuracy and less error rate like MSE (mean square error) and RMSE (root mean square error). Both MSE and RMSE refers to difference between original and predicted values so the lower the MSE the better is the mode.

So to update weights accurately author of this paper enhancing Extreme Learning Machine with Genetic Algorithm (GA-KLEM). To solve the issue of the conventional limit learning machine where the number of hidden nodes and the random generation of thresholds and weights lead to the degradation of the network learning ability, a genetic algorithm is then used to optimize the number of hidden nodes and layers of the kernel limit learning machine. The thresholds, the weights, and the root mean square error are used to define the fitness function. Finally, the least squares method is applied to compute the output weights of the model. Genetic algorithms are able to find the optimal solution in the search space and gradually improve the performance of the model through an iterative optimization process.

To enhance ELM we have added genetic algorithm functions like Fitness (weight fitness will be check using RMSE function), crossover and mutation inside ELM to update weights. In below screen showing Enhance ELM code



In above screen read red colour comments to know about GA-KLEM algorithm. Propose algorithm is compare with traditional algorithm called SVM and each algorithm performance is evaluated using MSE and RMSE. To train all algorithm author has used Air Quality dataset and below screen showing dataset details.



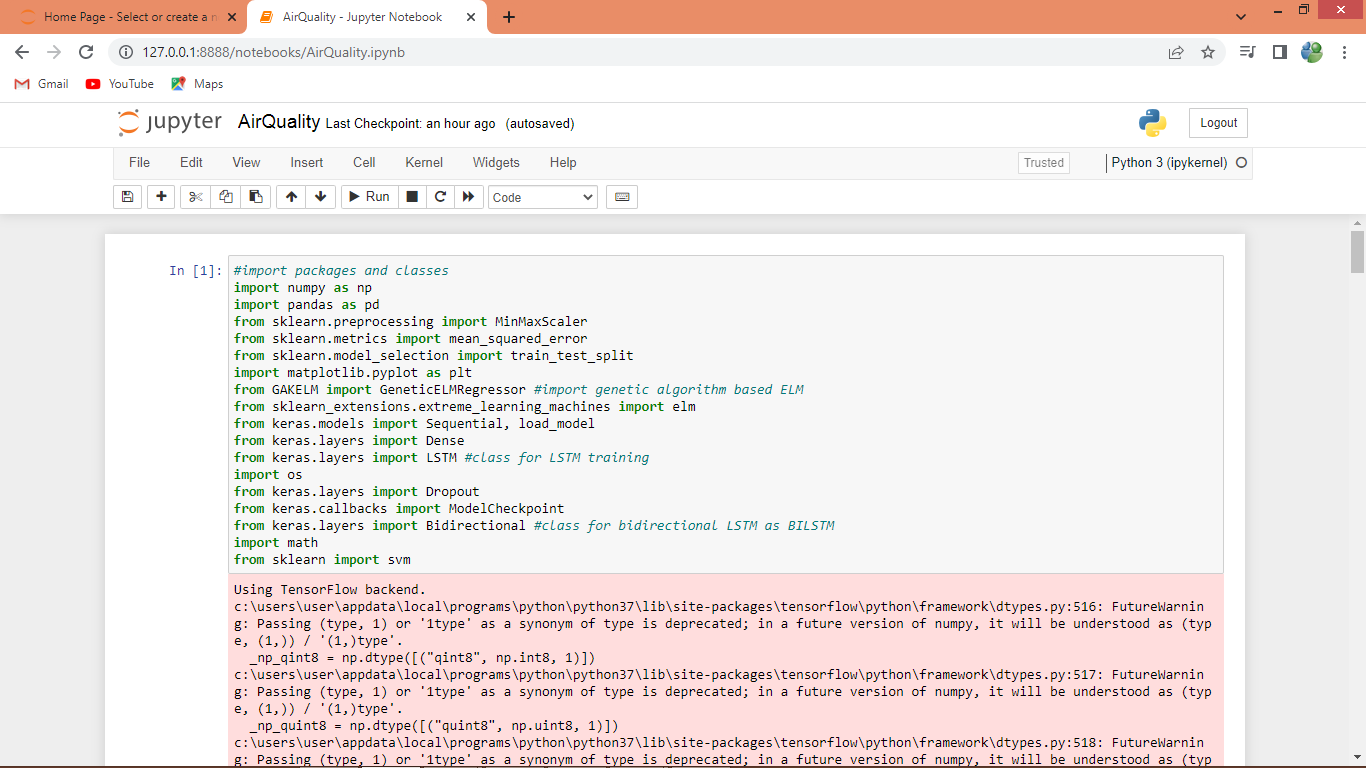
In above dataset screen first row represents dataset column names and remaining rows represents dataset values and we can test with each column for Air Quality prediction but we are using PM2.5 column for Air Quality prediction. So by using above dataset we will train and test each algorithm performance.

Extension Concept:

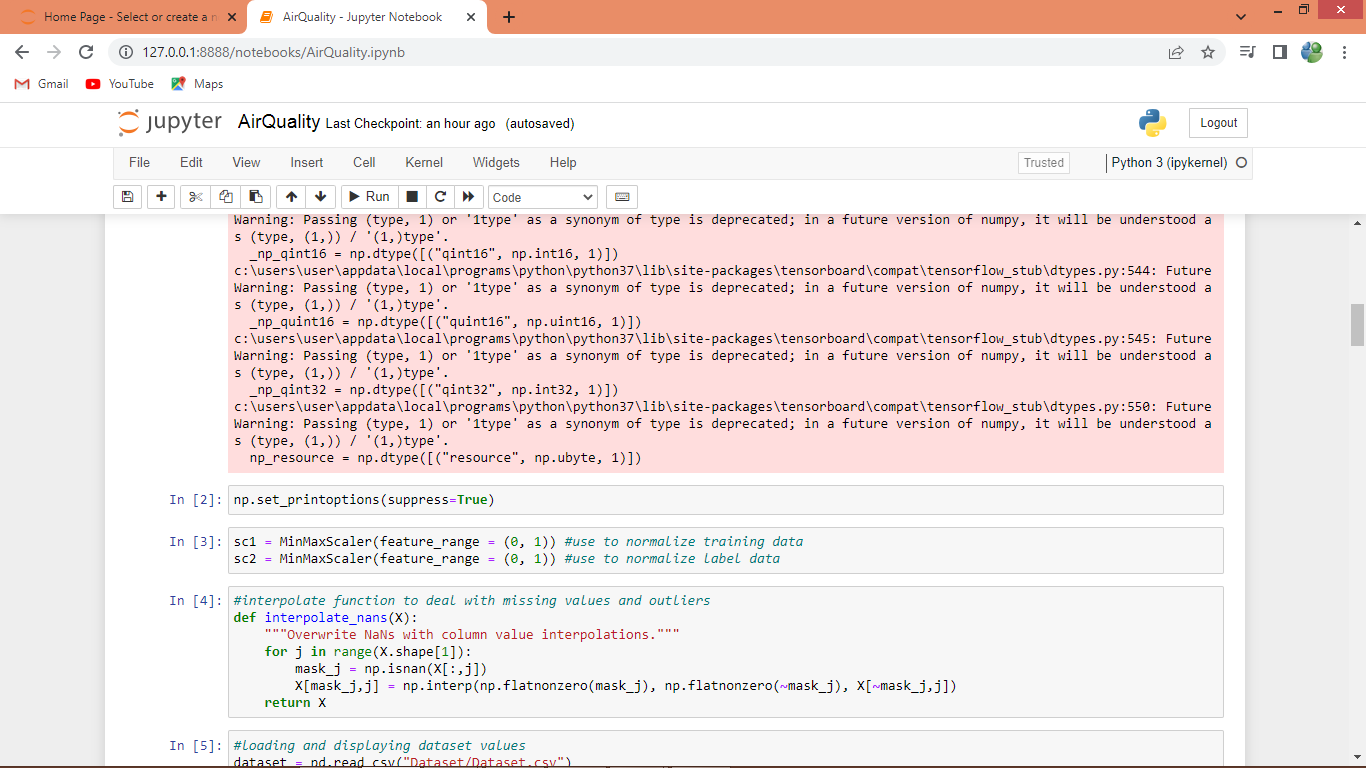
In propose paper author plan to enhance traditional ELM algorithm using GA optimization but as extension we have experimented with BI-LSTM algorithm which will optimize features weight in both forward and backward direction. BI-LSTM will optimized features till no more optimizations are possible so prediction accuracy will automatically get increased and error MSE rate will get decrease.

SCREEN SHOTS

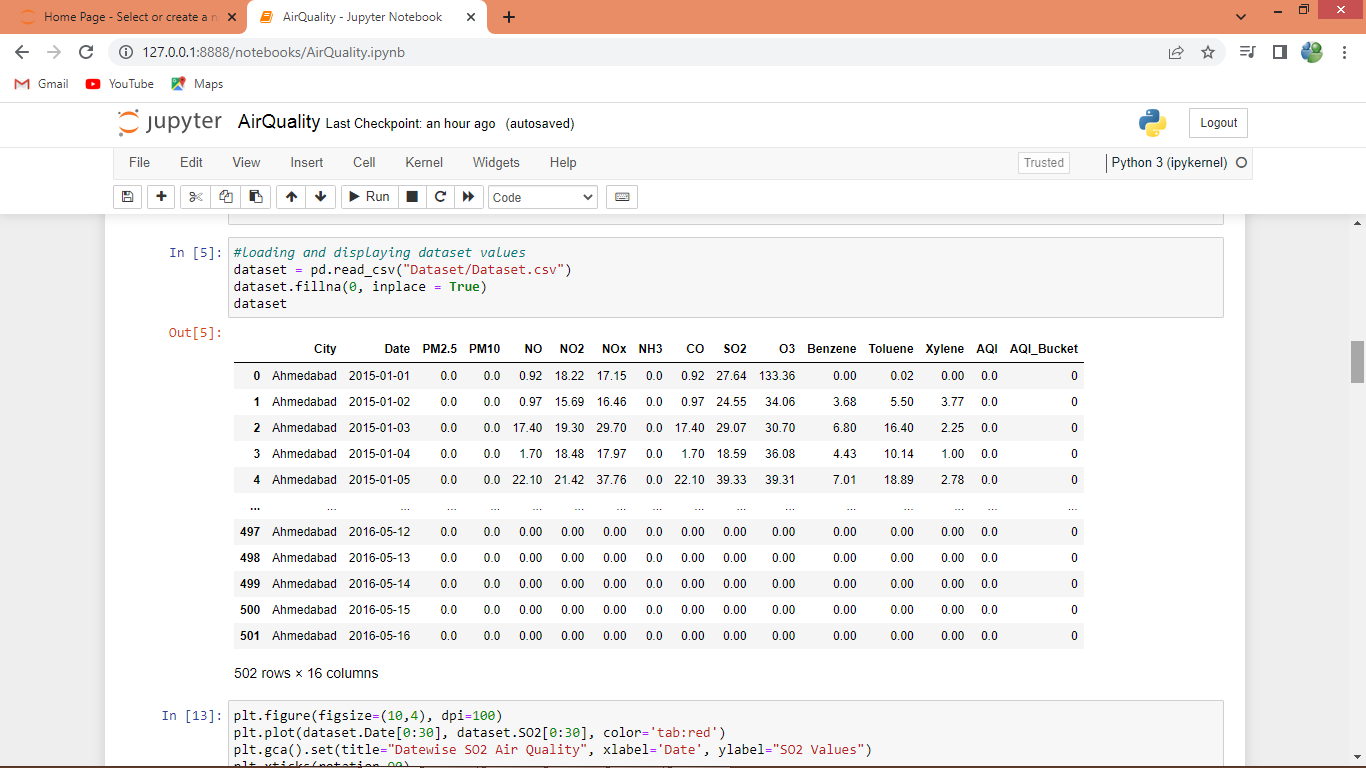
We have coded this project using JUPYTER notebook and below are the code and output screen with blue colour comments



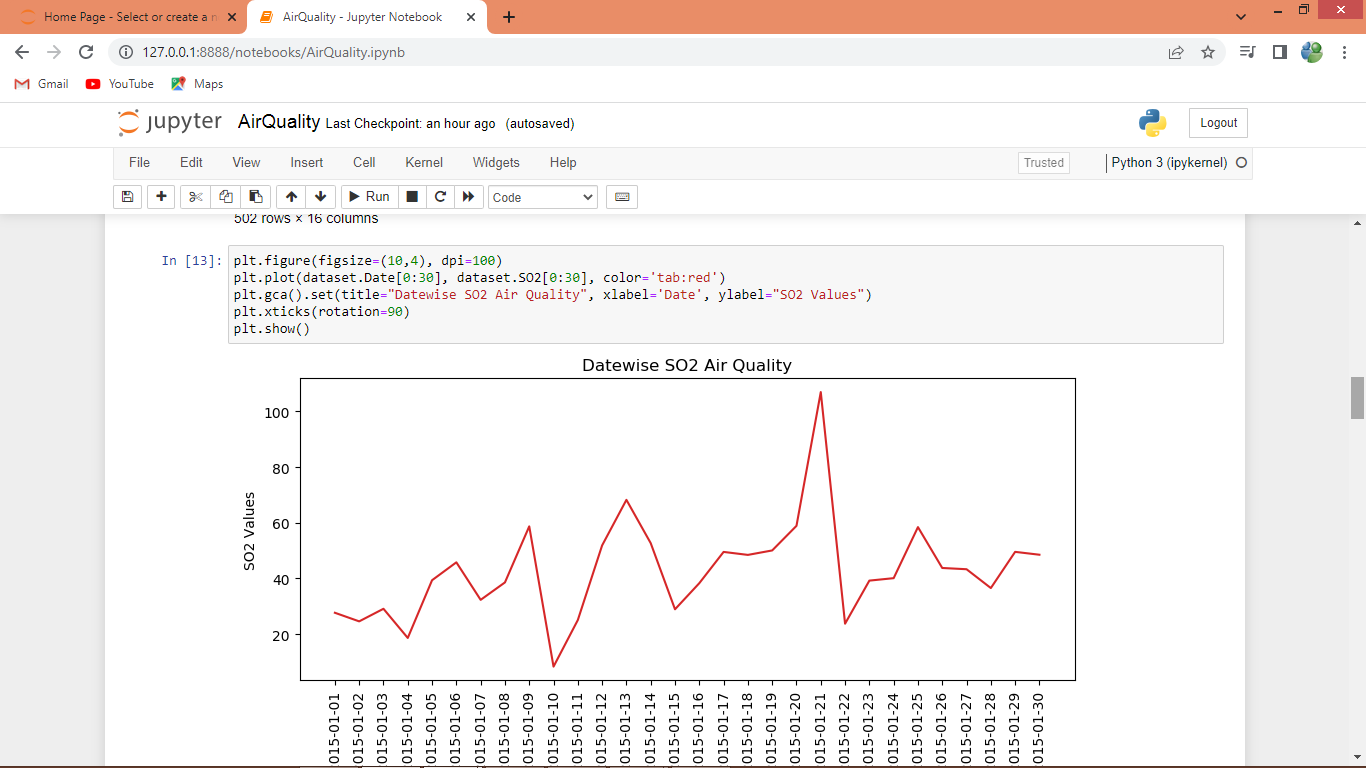
In above screen importing require python classes and packages



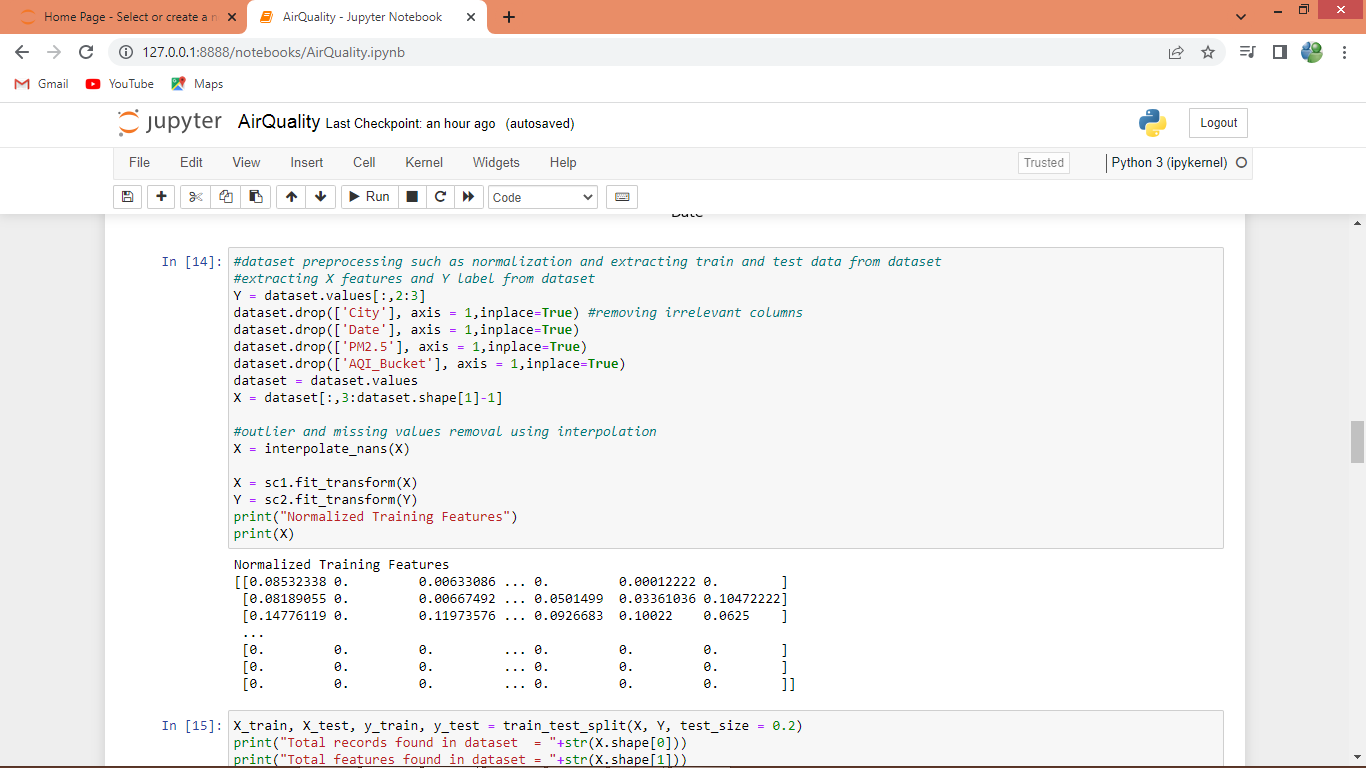
In above screen defining interpolate function to deal with missing values and outliers. Outliers refer to values which are not in normal range so such values will get replace with interpolate and standard deviation.



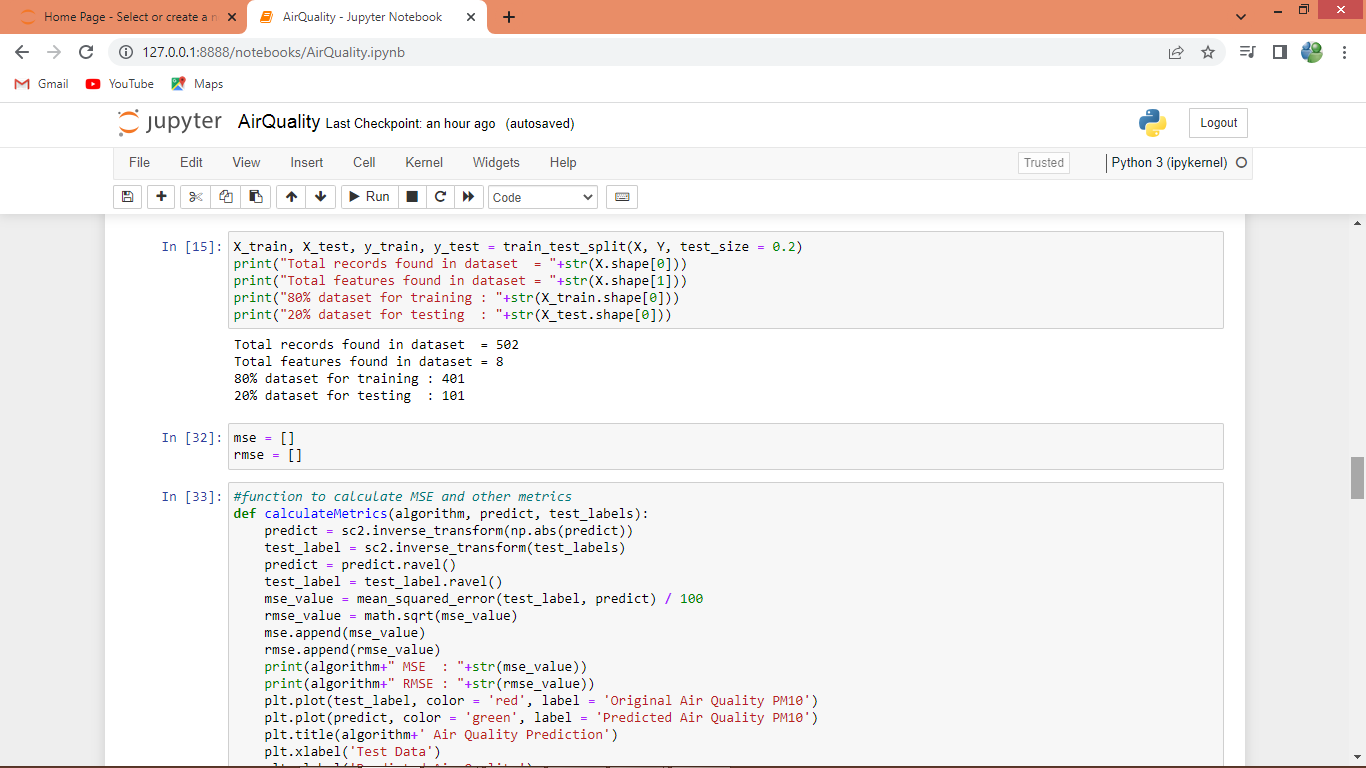
In above screen reading and displaying dataset values



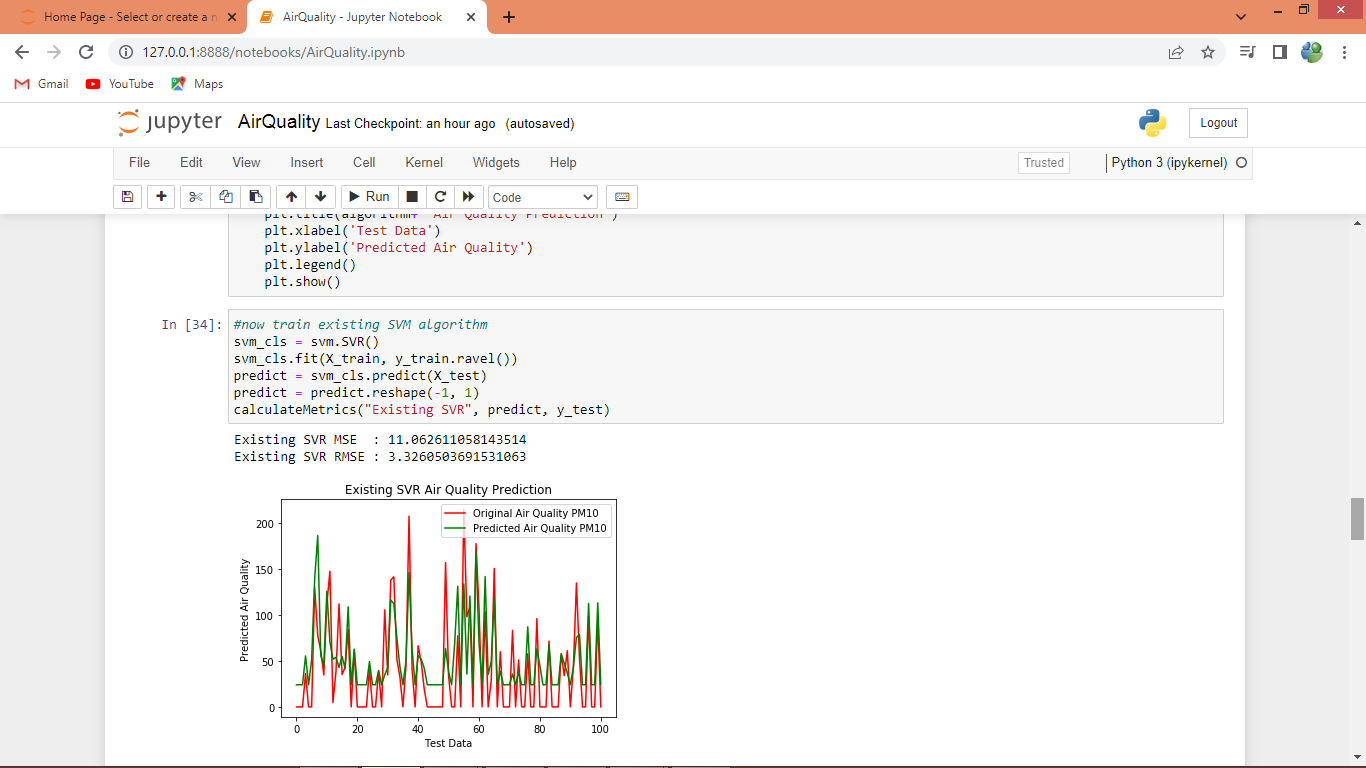
In above graph displaying Air Quality where x-axis represents Date and y-axis represents air quality



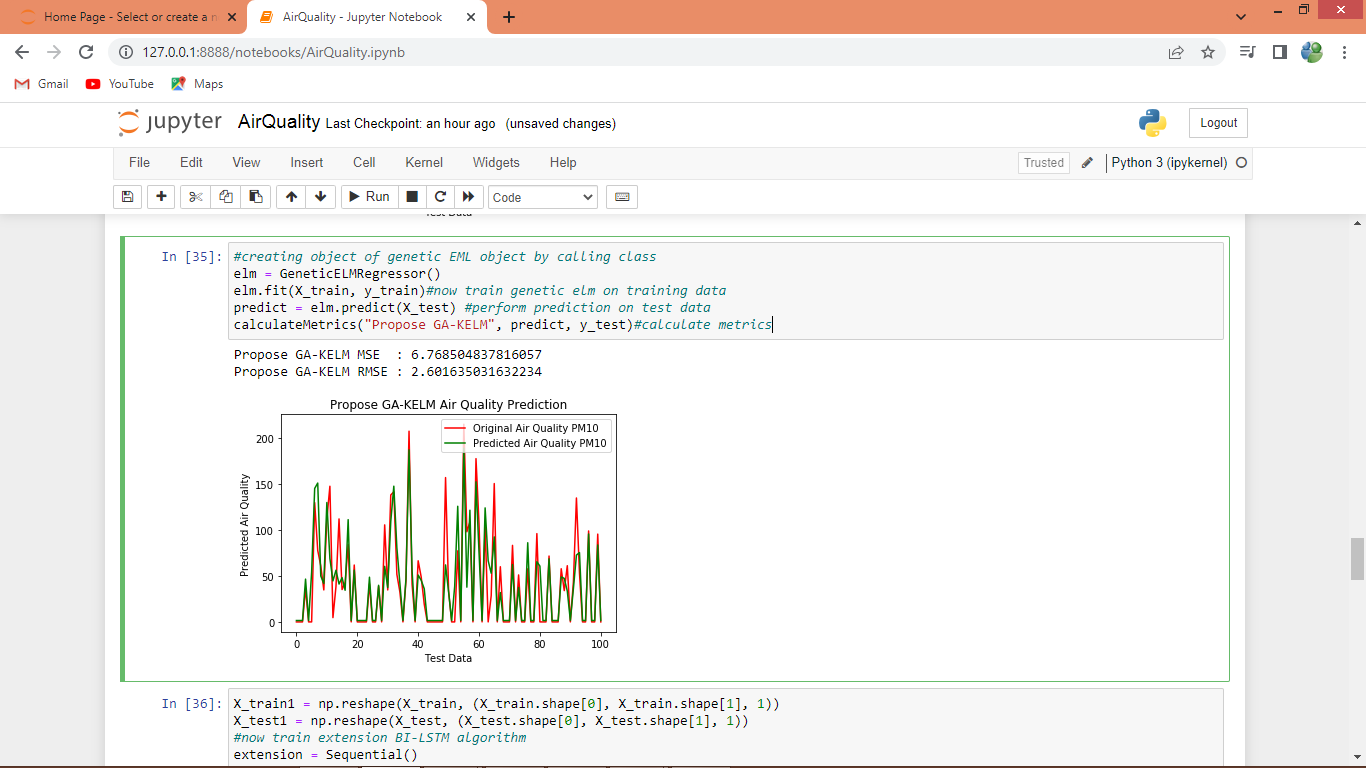
In above screen dataset Pre-processing such as extracting X and Y features, interpolation and normalization



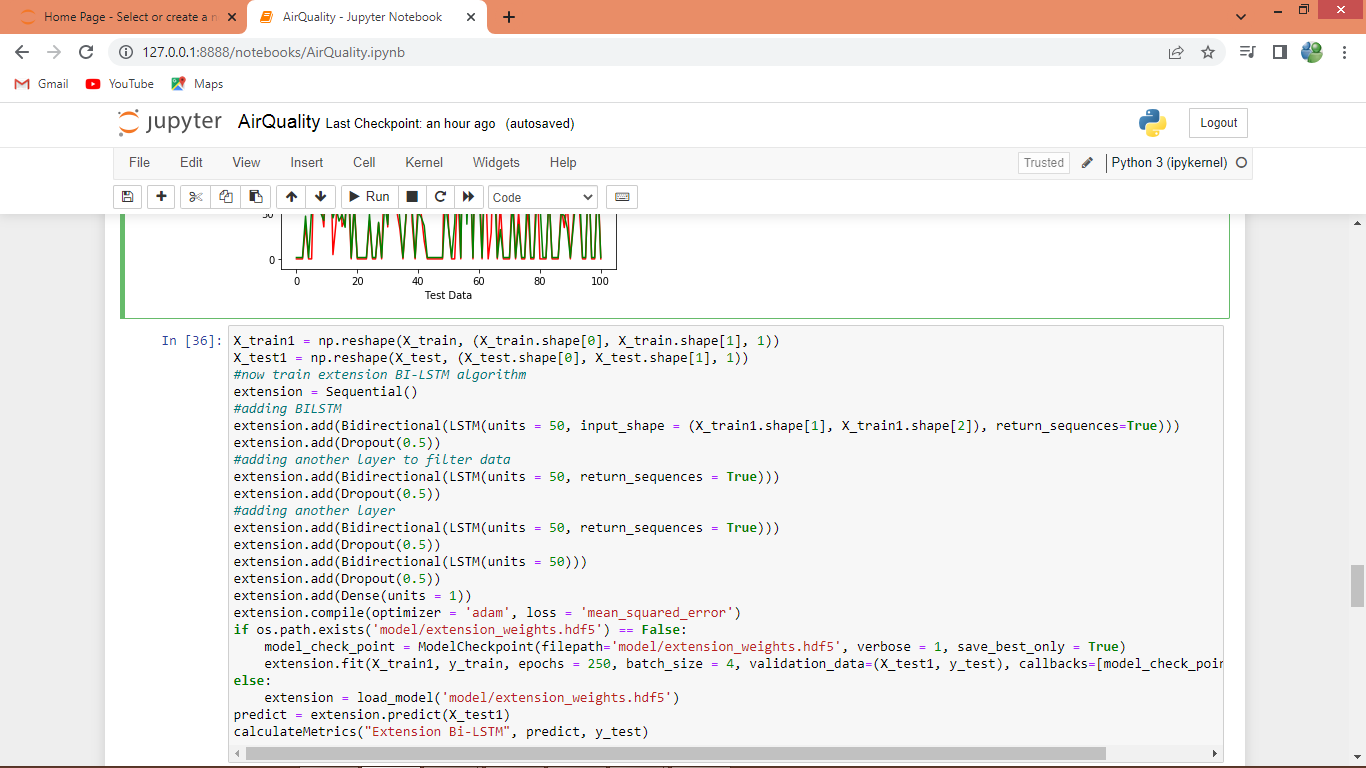
In above screen splitting dataset into train and test and then defining function to calculate MSE and RMSE.



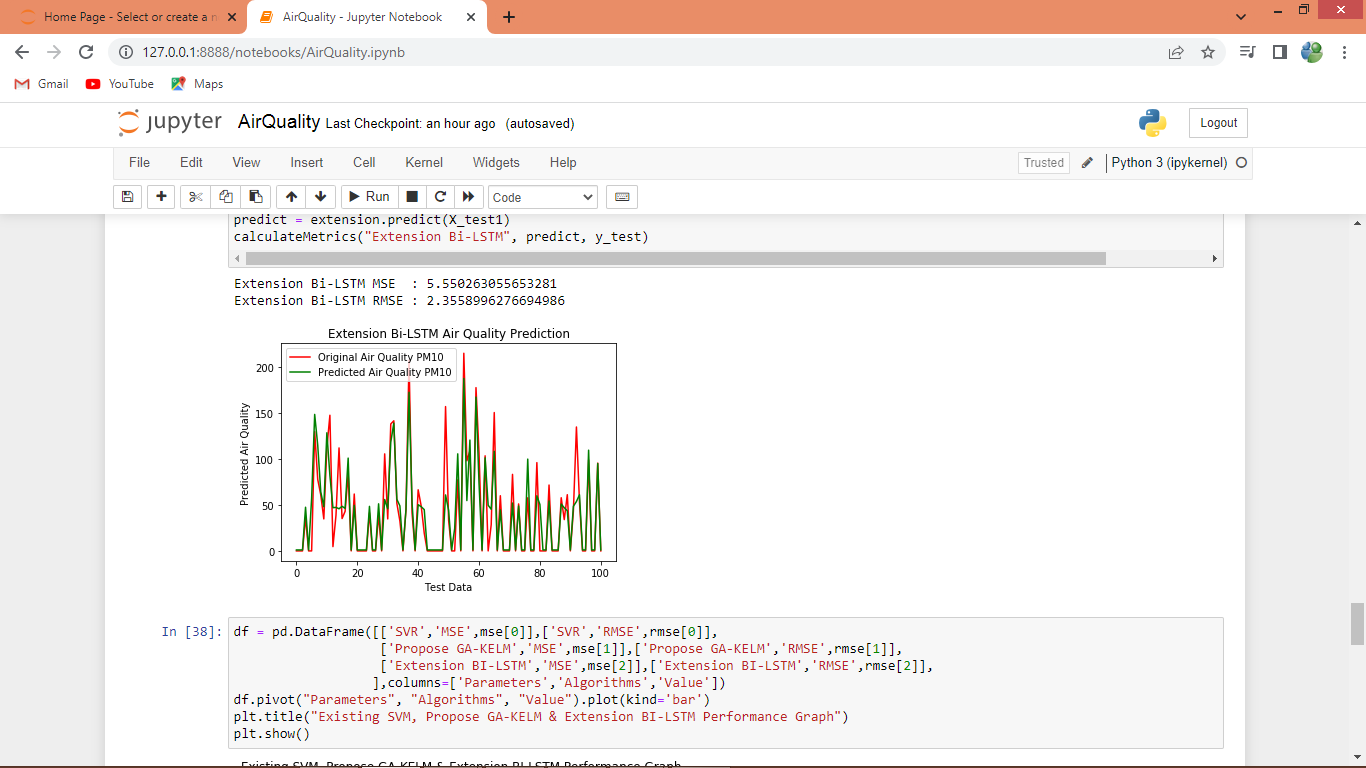
In above screen training SVM algorithm and then with SVM we got 11 as the MSE and in graph x-axis represents TEST COUNT and y-axis represents air quality. Red line represents Original Test Air Quality and green line represents Predicted Air Quality and we can see both lines are closed as they are overlapping with little gaps and this gap can reduce by applying propose algorithm



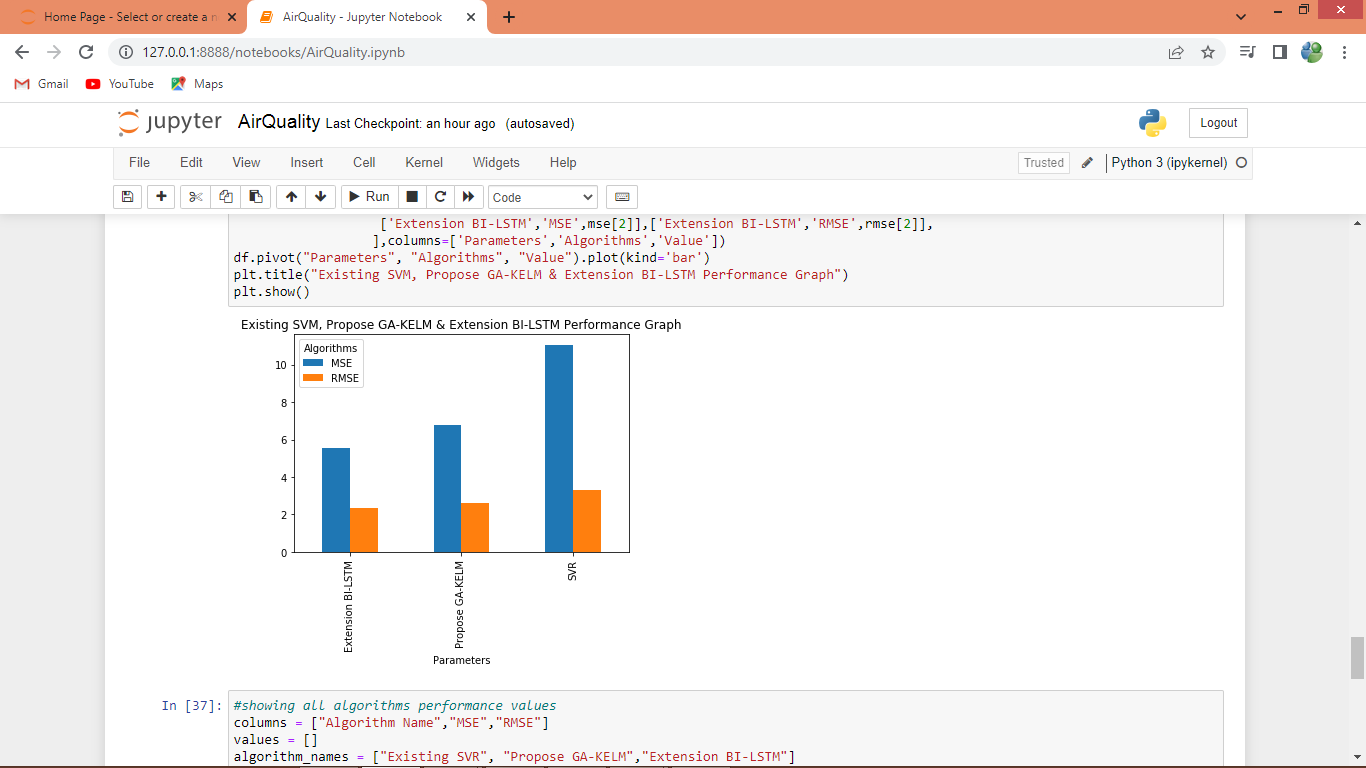
In above screen we are training propose genetic ELM called GA-KELM and we got its MSE as 6 and in graph we can see now both lines are overlap with too few gap.



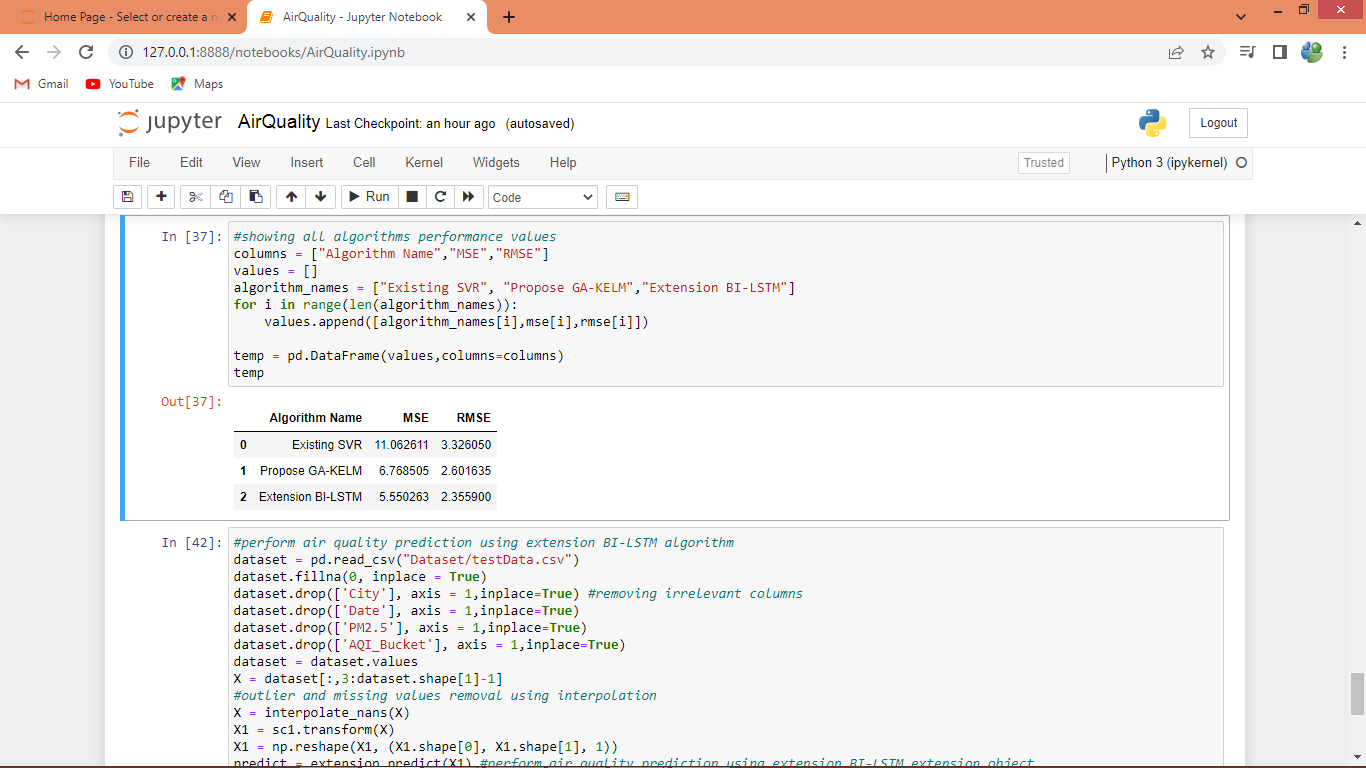
In above screen training extension BI-LSTM algorithm and after executing above block will get below output



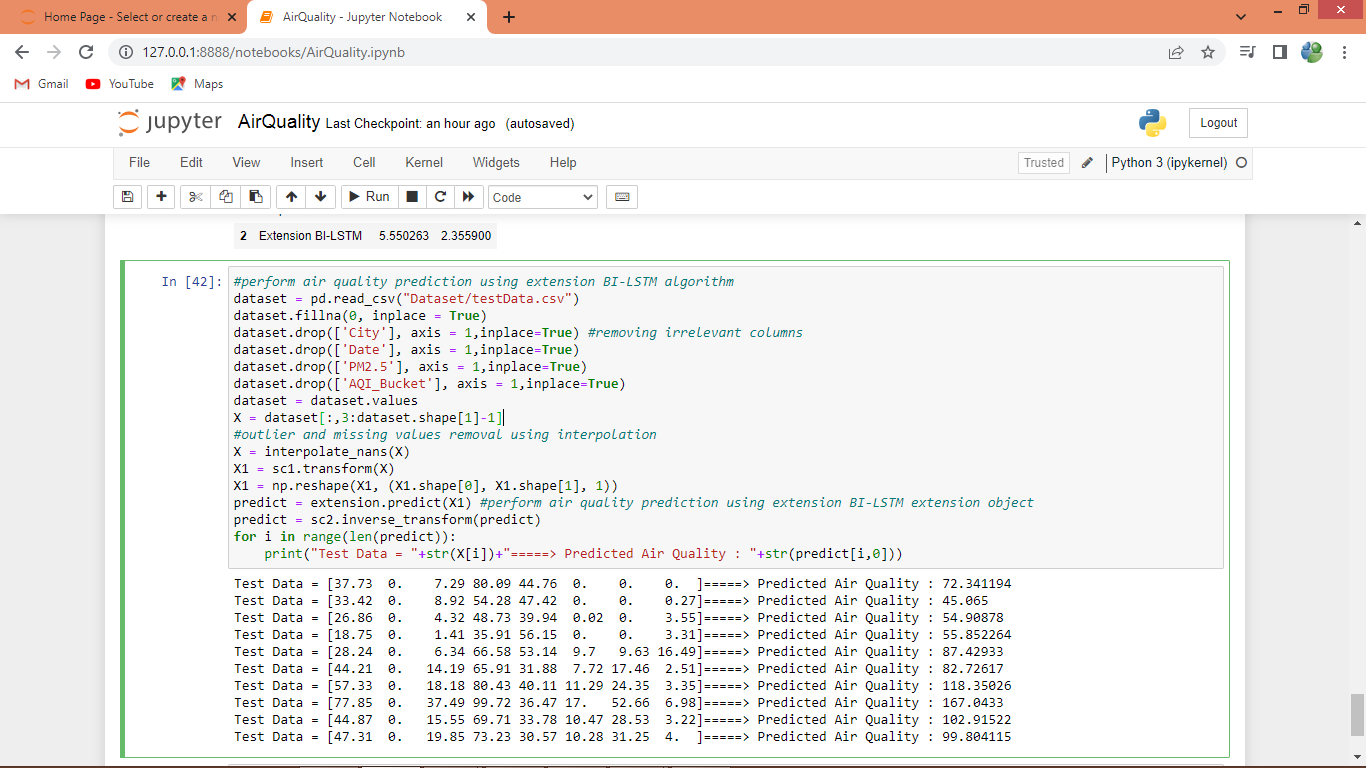
In above screen with extension BI-LSTM we got MSE as 5% and in graph also nearly 90% test and predicted air quality lines are overlapping and in all algorithms extension BI-LSTM has got less MSE and RMSE



In above graph x-axis represents algorithm names and y-axis represents MSE and RMSE in different colour bars and in all algorithms extension BI-LSTM got less RMSE and MSE



In above screen displaying all algorithms performance in tabular format



In above screen loading TEST data and then predicting air quality using extension object and then we can see TEST data and predicted air quality after =🡺 arrow symbol