**MACHINE LEARNING PROJECT**

***Project Name***

## Air Quality Index(AQI) prediction using Multi-layer perceptron, Decision Tree Regression and Linear Regression

***Submitting to*Vishwakarma Institute of Information Technology, Pune**

***Submitted by***

**Pradnya Dixit**

# Abstract of Project:

* Air Quality index is an important factor used for reporting daily air quality. It tells you how clean or polluted your air is, and what associated health effects might be a concern for you. The AQI focuses on health affects may experience within a few hours or days after breathing polluted air. Traditionally AQI is calculated with the contents of other pollutants as CO, NOx etc. whose concentration is air is measured through sensors. However, a need for use of Machine Learning techniques arises for a faster prediction of AQI and thus alerting the population for any further threat.
* The present project has many features (input parameters) to predict the AQI (Output parameter).
* The data is read and the characteristics are understood followed by developing model using Machine learning technique.
* After preparing the data apply training and testing to the model.
* In this project the machine learning techniques used are: Multi-layer perceptron, Decision Tree Regression and Linear Regression
* Using sklearn packages importing the machine learning algorithms and finding the Accuracy of the models.

# Project Summary:

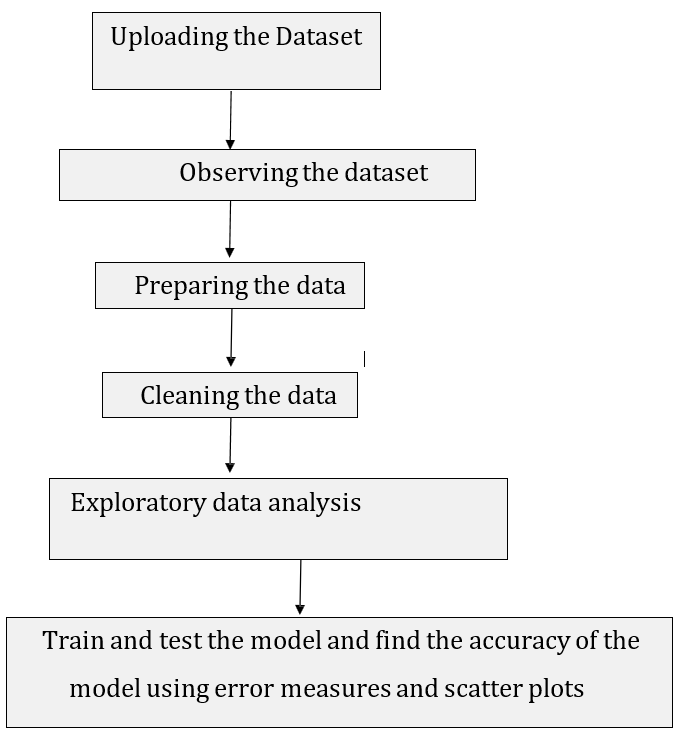
* From the AQI dataset we can analyze the different **input parameters** determining the AQI like: PM2.5, PM10, NO, NO2, NOX, NH3, CO, CO2, O3 and the **output parameter:** AQI along with category data.
* Using pandas we can visualize the dataset features.
* Prepare the dataset.
* Perform the exploratory data analysis using seaborn and matplotlib Modules.
* Analyze the insights of the dataset using different visualization techniques. Cleaning the dataset if there are any categorical values in the dataset.
* Find the independent and dependent variables after encoding the dataset.
* Splitting the dataset into training and testing from sklearn package import train\_test\_split
* Evaluate the model using Multi-layer perceptron, Decision Tree Regression and Linear Regression

# Objectives of Project:

* Predicting AQI Multi-layer perceptron, Decision Tree Regression and Linear Regression
* Judge the performance of the techniques to predict AQI and compare the same.

# Details of Project developed: (Refer fig.1)

* Firstly, imported all required packages and modules into your required format.
* Read data set file using pandas.
* Check duplicate values and remove those rows from the data set.
* Check null values in the dataset and using describe keyword get all required data.
* Perform the exploratory data analysis using seaborn and matplotlib.
* Clean the dataset using sklearn package import label encoder
* Determine the independent and dependent variables in the dataset.
* Using train\_test\_split, split the dataset into training and testing part. train the model 70% of the data and test the model 30% of the data.
* Evaluating the model using linear regression.
* Find the accuracy of the model using error measures as coefficient of determination and root mean square between Observed concrete strength and predicted concrete strength. Accuracy of model is also found through scatter plot.



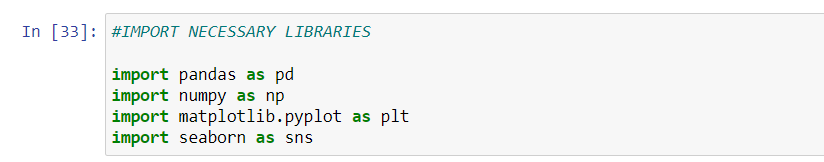
*Fig.1 Details of Project Developed*

## System Requirement Used:

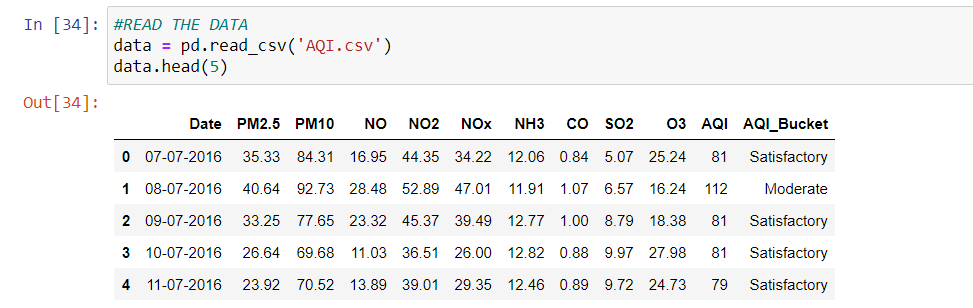
* + Windows 10
  + Python 3
  + Jupyter Notebook

# Input Output Datasets / screenshots

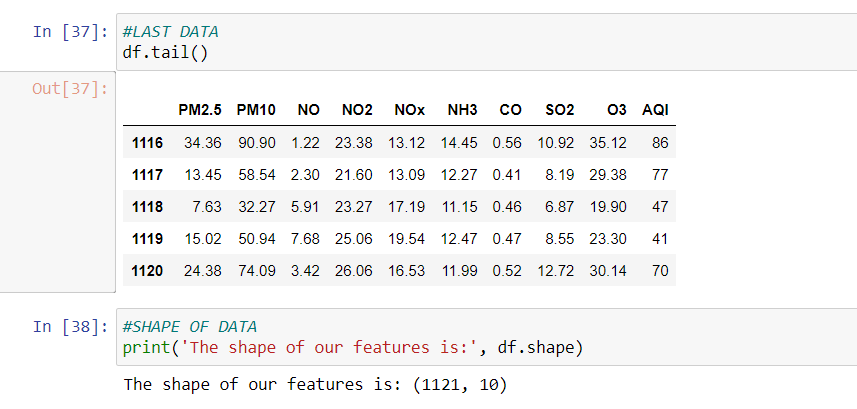
* *Import all required packages*

**

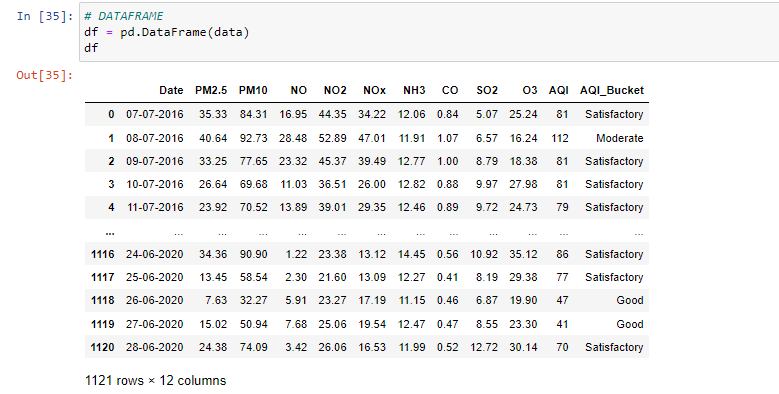
* *Uploading the dataset*

**

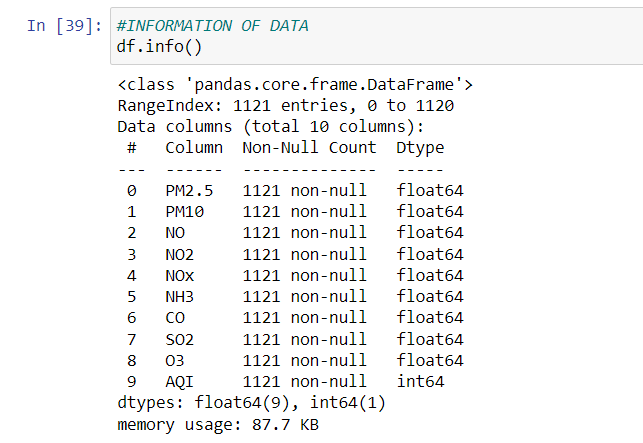
* *Details of the dataset*

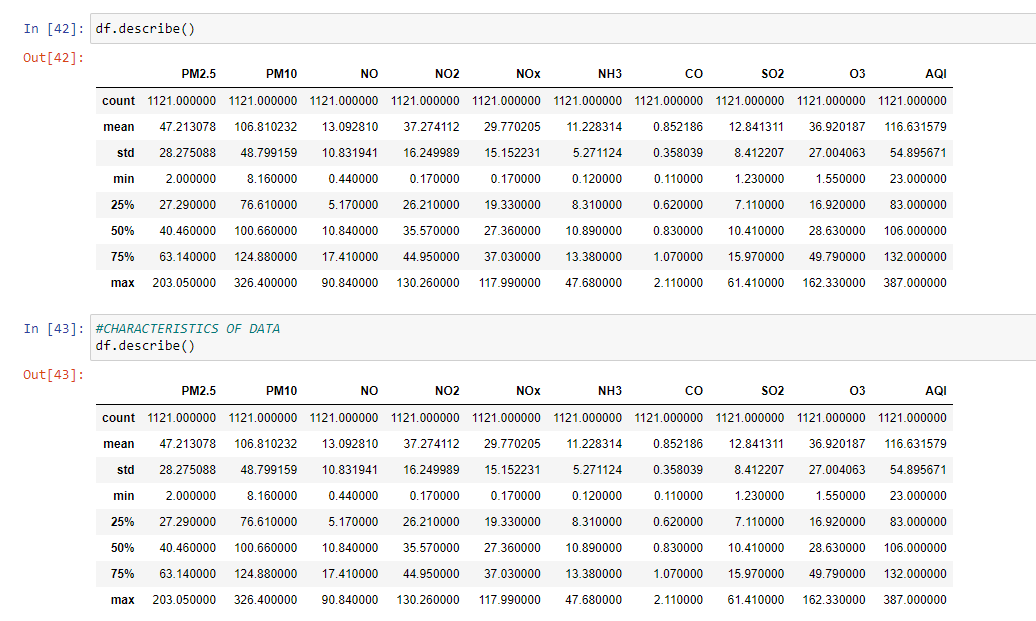
**

* *Dataframe*

**

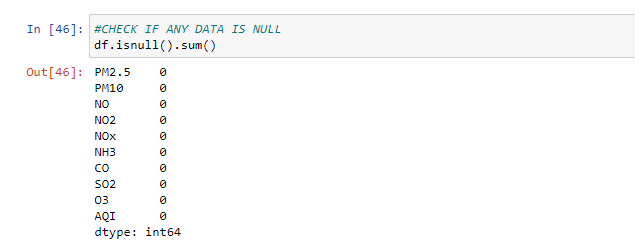
* *Observing the dataset features*

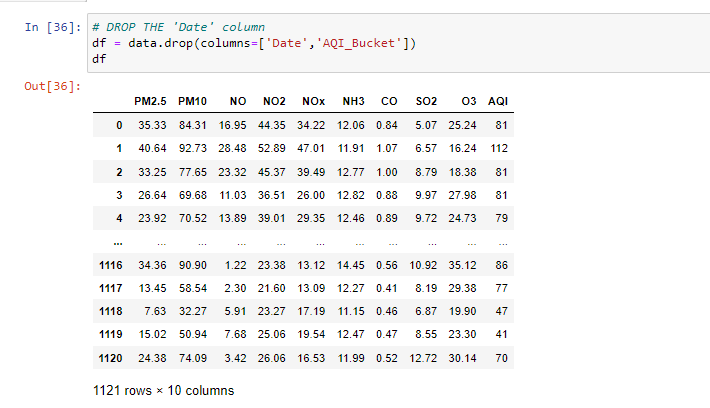
**

**

**

* *Drop any data if not required and check for null*

**

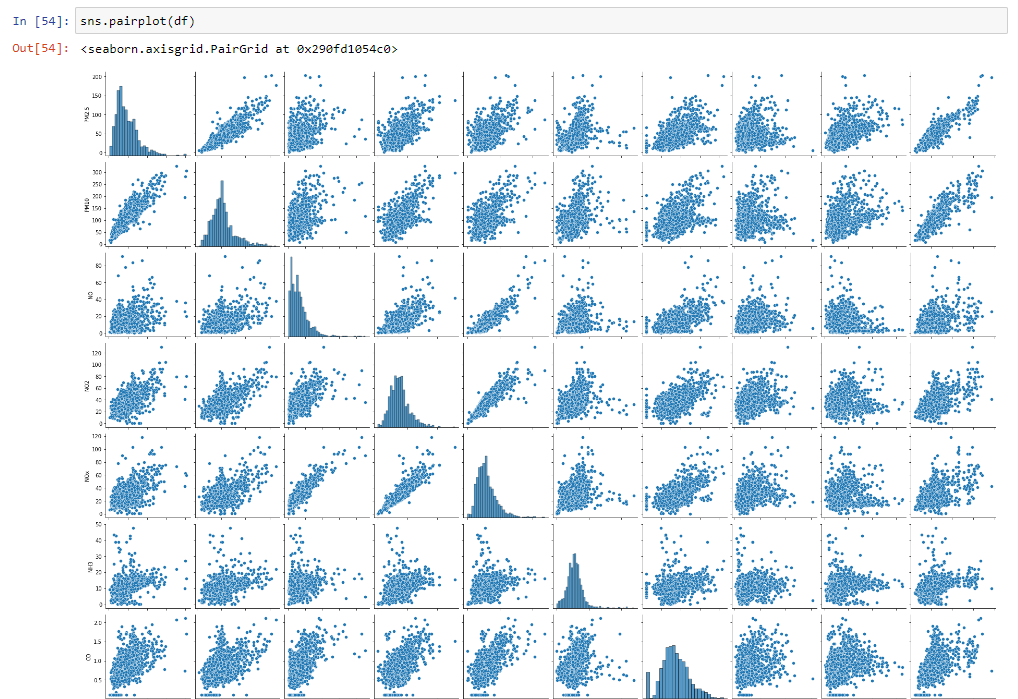
**

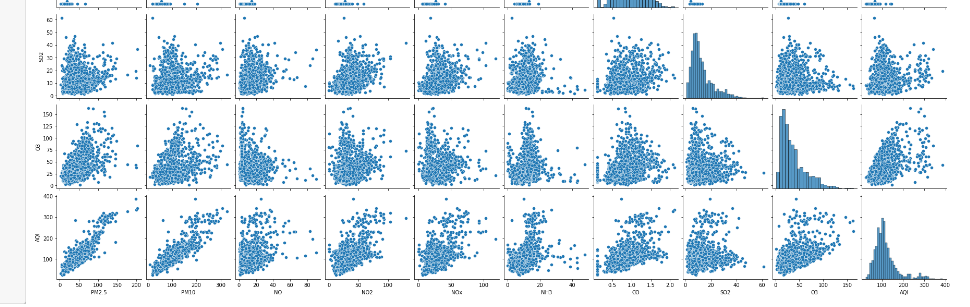
### EXPLORATORY DATA ANALYSIS

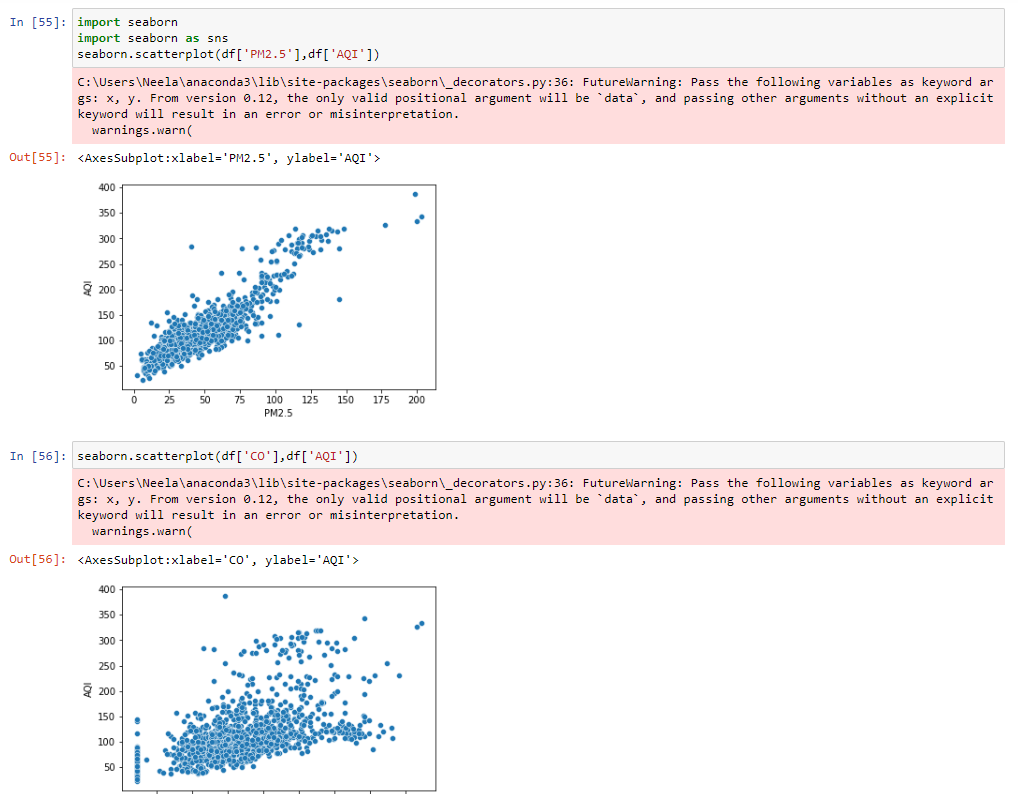
**

**

**

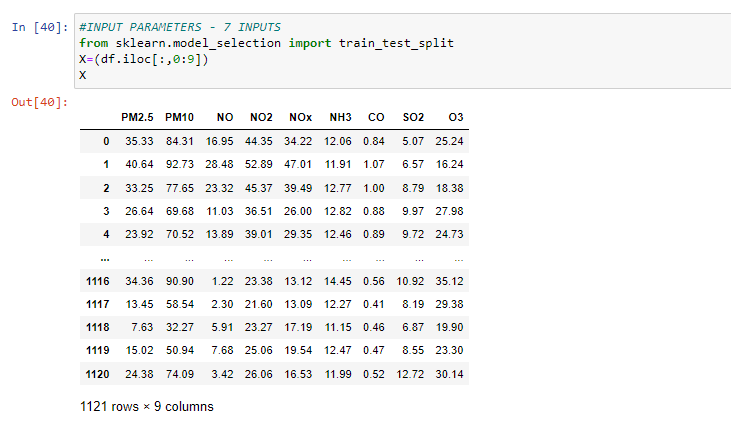
**

**

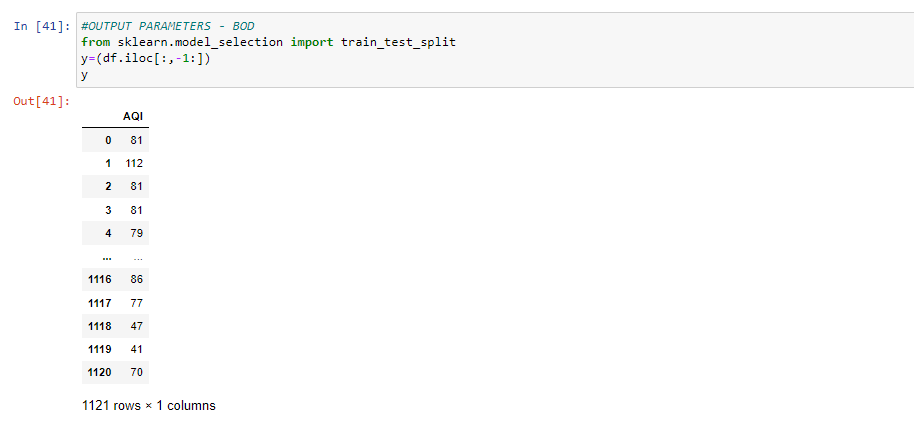
**

***PREDICTION MODEL DEVELOPMENT***

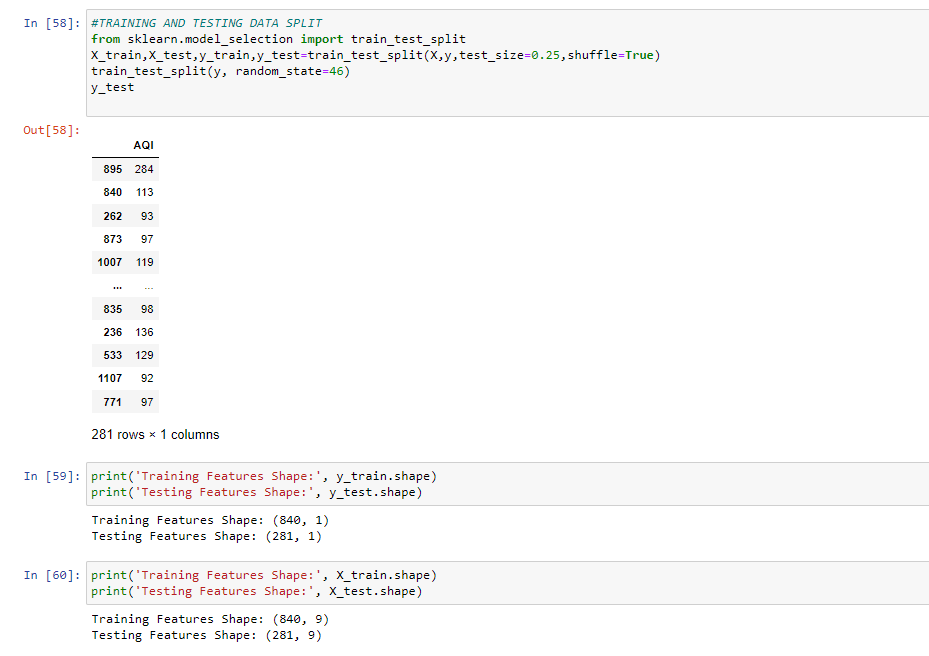
* *Input data*

**

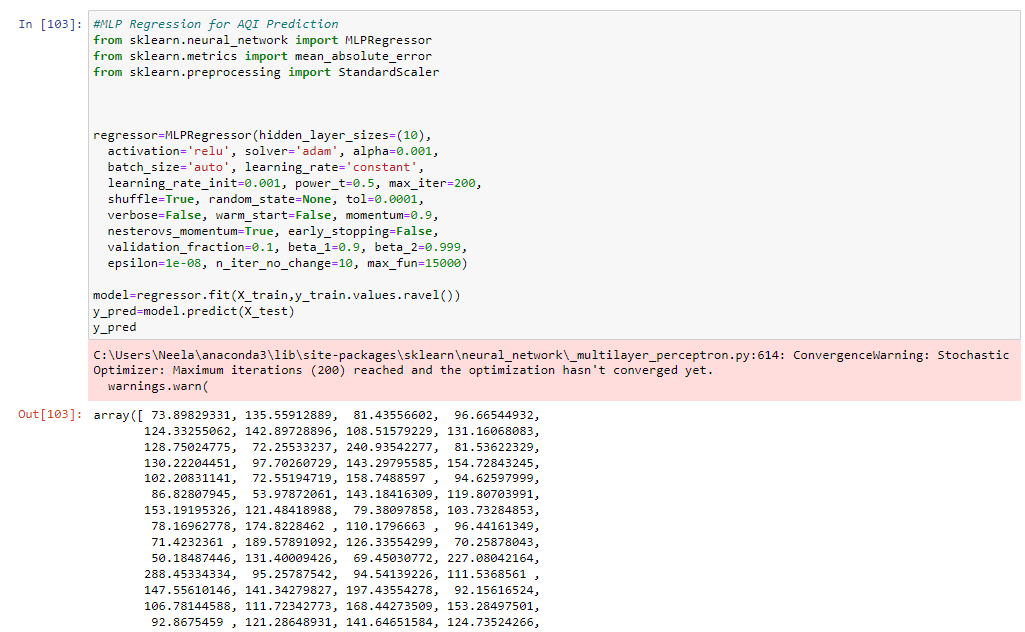
* *Output Parameter*

**

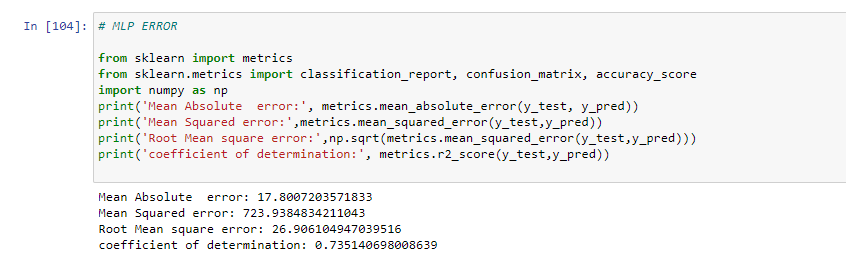
* *Split data*

**

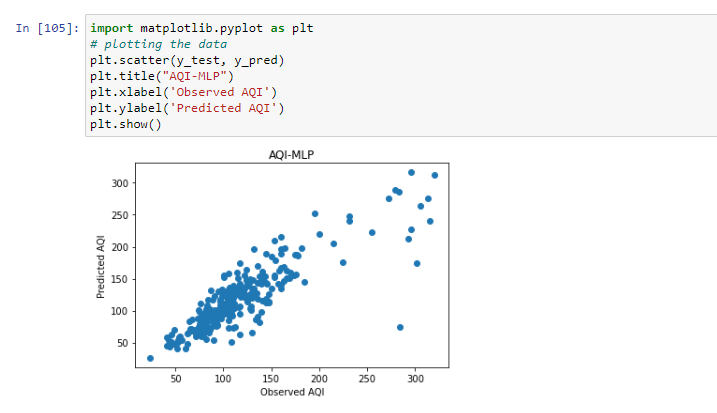
* ***AQI Prediction using MLP***

**

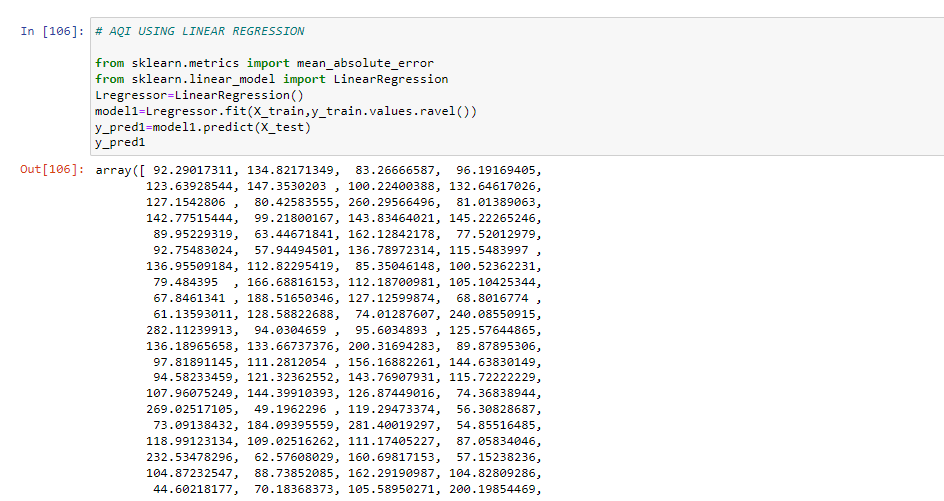
* *Error*

**

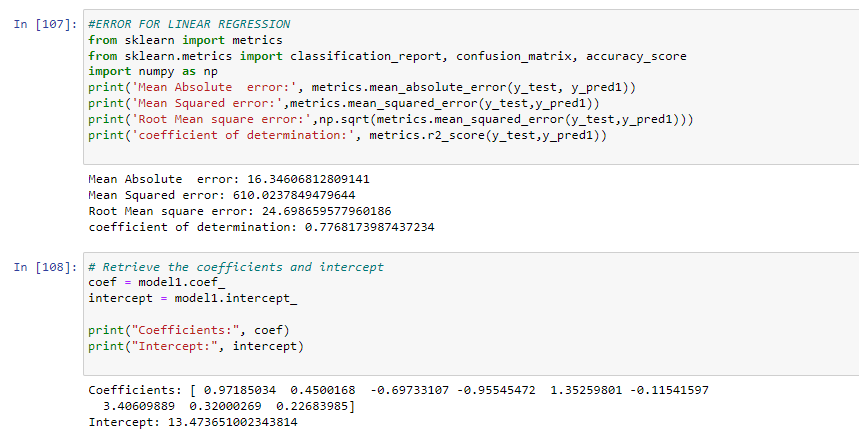
* *Scatter plot between Observed and Predicted*

**

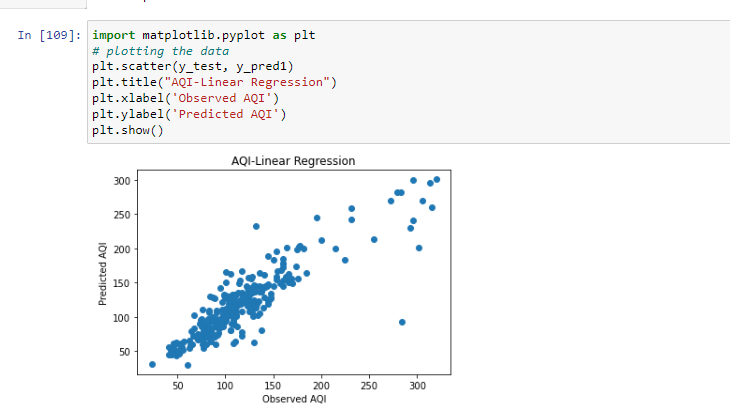
* ***AQI Prediction using Linear Regression***

**

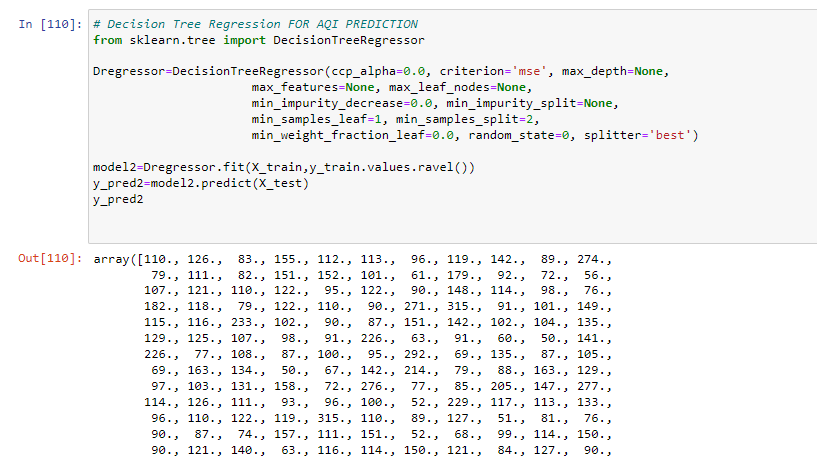
* *Error and Coefficients*

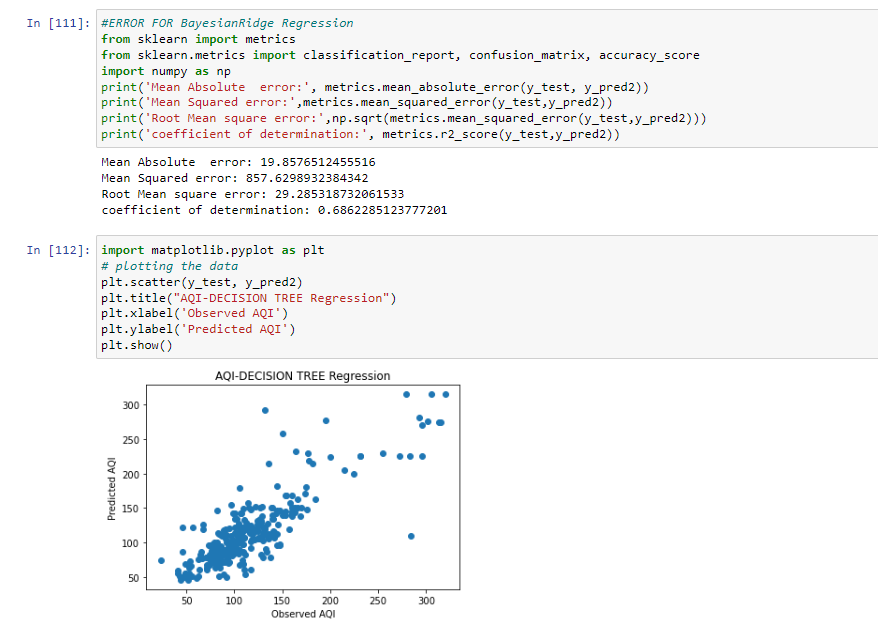
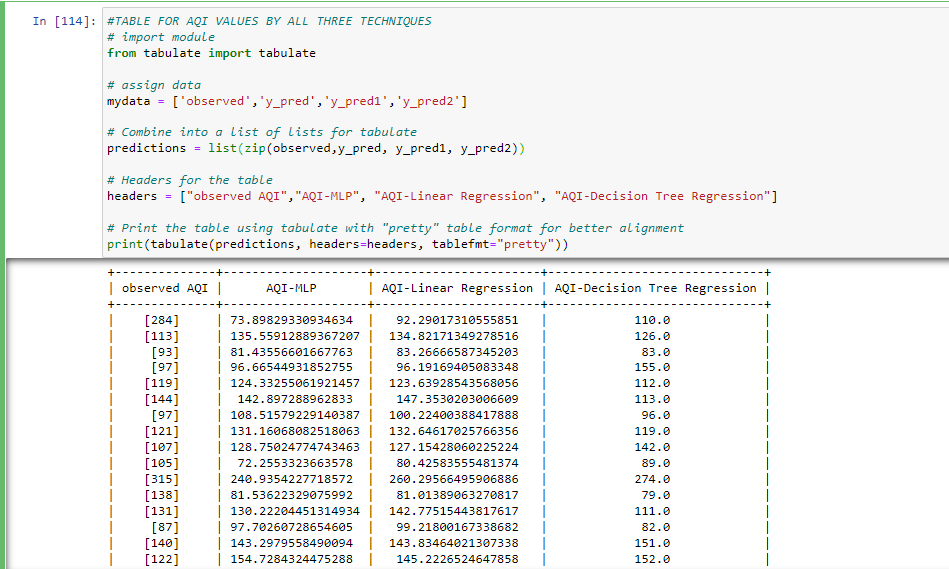
**

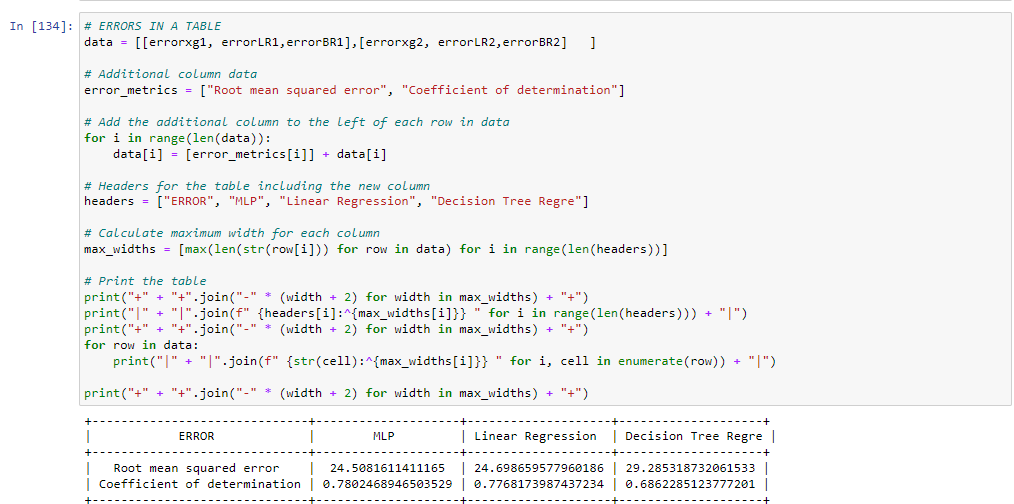
* *Scatter Plot*

**

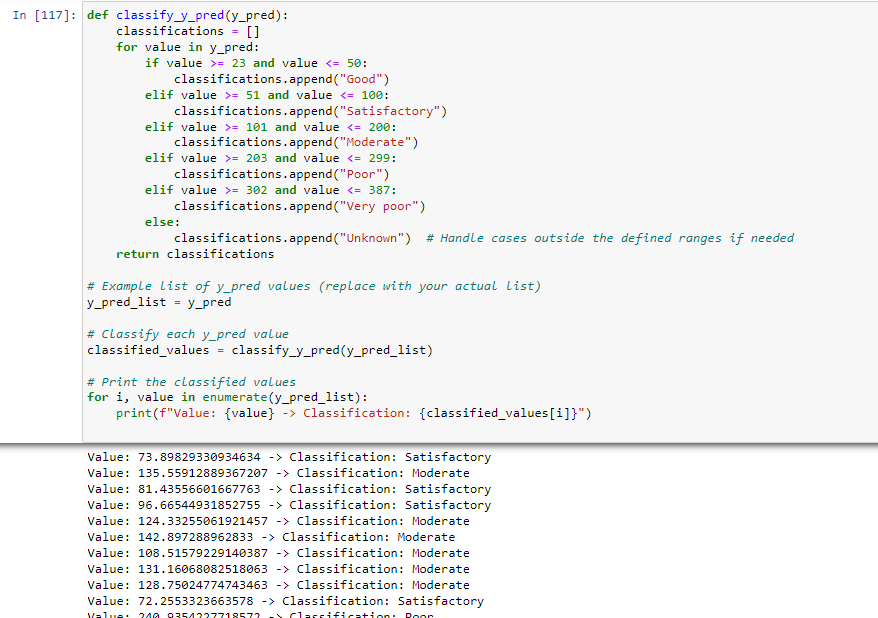
* *AQI using Decision Tree Regression*

**

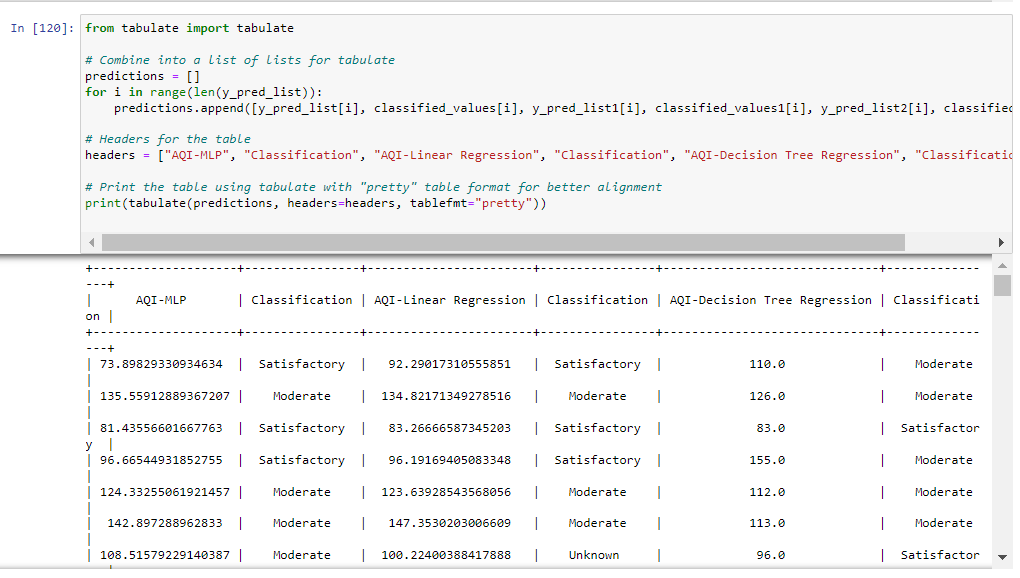
* Error and scatter plot
* 
* Data in Tabular format
* 



* **Classification of Data**



* All results and classification in tabular format



# Conclusion:

The present project’s aim is to predict AQI using Multi-layer perceptron, Decision Tree Regression and Linear Regression

The exploratory data analysis also shows a higher correlation of AQI with PM2.5 and PM10 as input parameters.

* The error table shows that concrete strength prediction is done better by MLP with coefficient of determination 0.78 followed by Linear regression (coed=0.77) and Decision Tree Regression (COD=0.68).
* The Root mean square by MLP is also lower followed by Linear Regression and Decision Tree Regression.
* The Scatter plot with random forest also shows a balanced scatter with no obvious under or over prediction.

# References:

* <https://www.w3schools.com/python/pandas/default.asp>
* <https://www.w3schools.com/python/numpy/numpy_random_seaborn.asp>
* <https://www.w3schools.com/python/matplotlib_pyplot.asp>
* Dataset: <https://www.kaggle.com/>