**MACHINE LEARNING PROJECT**

***Project Name***

## Concrete Strength Prediction using Random Forest, Linear Regression and Support Vector Regression

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

***Submitted by***

**Preeti Kulkarni**

# Abstract of Project:

* Concrete strength determination is of vital importance. It is traditionally been done by casting the concrete and curing the same for 28 days. The traditional method utilizes manpower and consumes time. A need for use of Machine Learning techniques thus arises which can help prediction of concrete strength.
* The present project has many features (input parameters) to predict the concrete strength (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: linear regression, Random Forest and Support vector regression
* Using sklearn packages importing the machine learning algorithms and finding the Accuracy of the models.

# Project Summary:

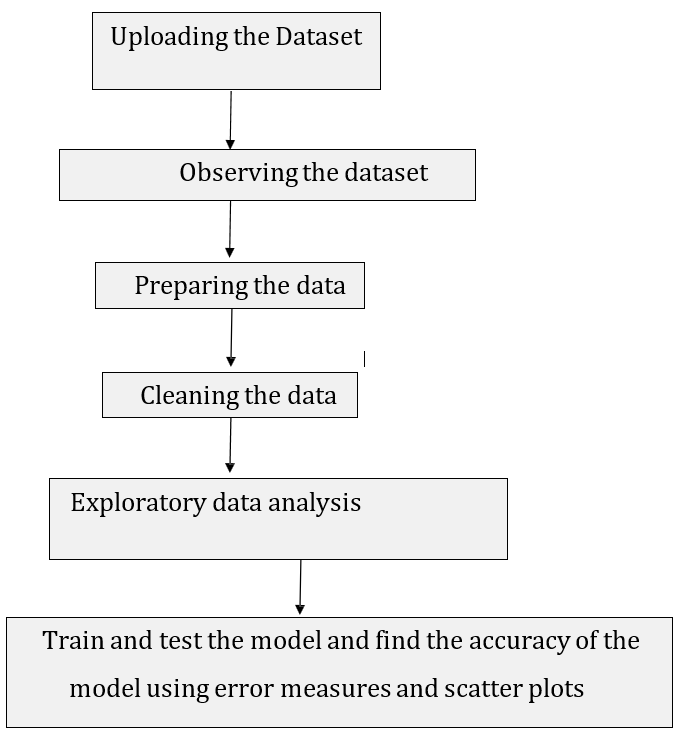
* From the concrete dataset we can analyze the different **input parameters** determining the concrete strength like: Cement (Kg/m3), Blast furnace slag (Kg/m3), Fly ash (Kg/m3), Water (Kg/m3), superplasticizer (Kg/m3), Coarse aggregate content (Kg/m3), Fine aggregate content (Kg/m3), Age (days) and the **output parameter:** Strength of Concrete(N/mm2)
* 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 linear regression, Random Forest and Support vector regression

# Objectives of Project:

* Predicting strength of concrete using Linear Regression, Random Forest and Support Vector Regression.
* Judge the performance of the techniques to predict concrete strength and compre 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:

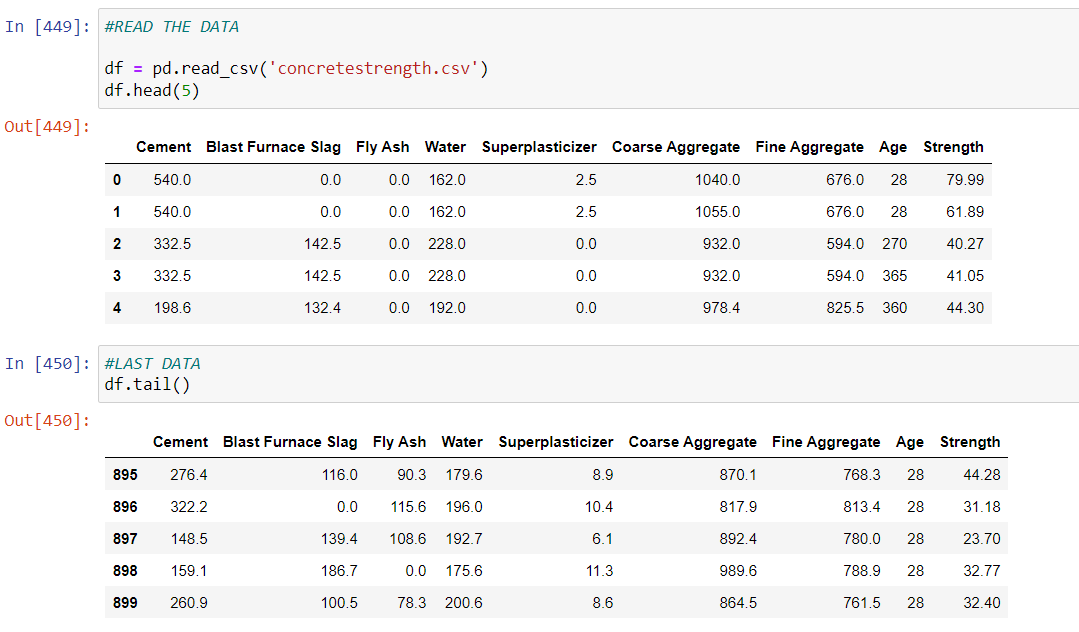
* + Windows 10
  + Python 3
  + Jupyter Notebook

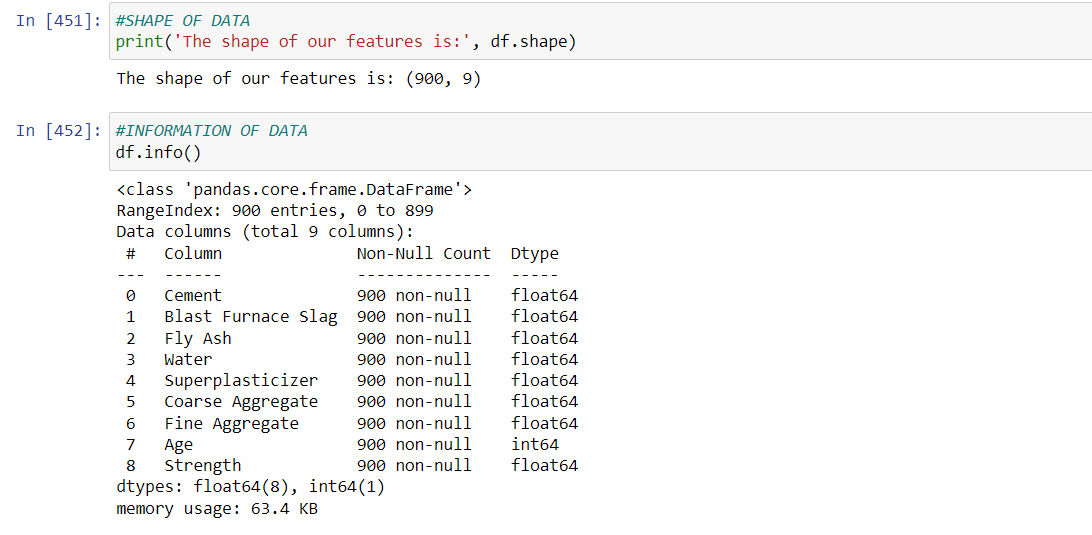
# Input Output Datasets / screenshots

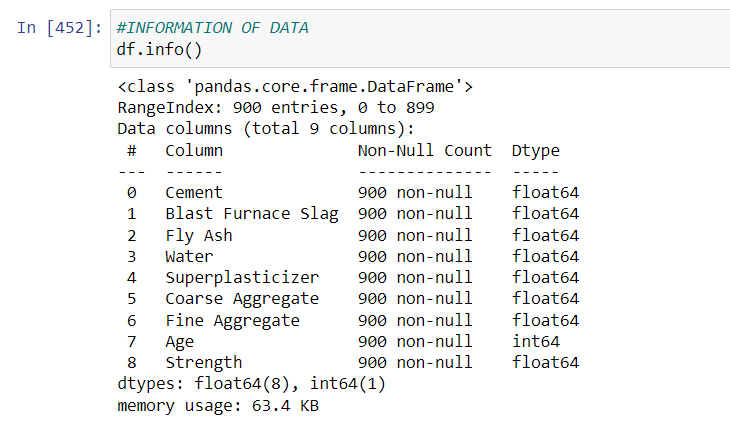
* *Import all required packages*

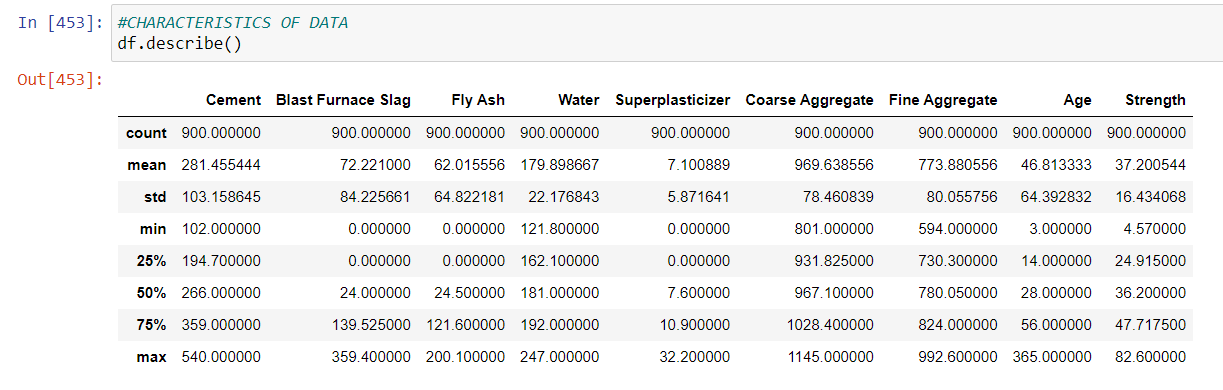
**

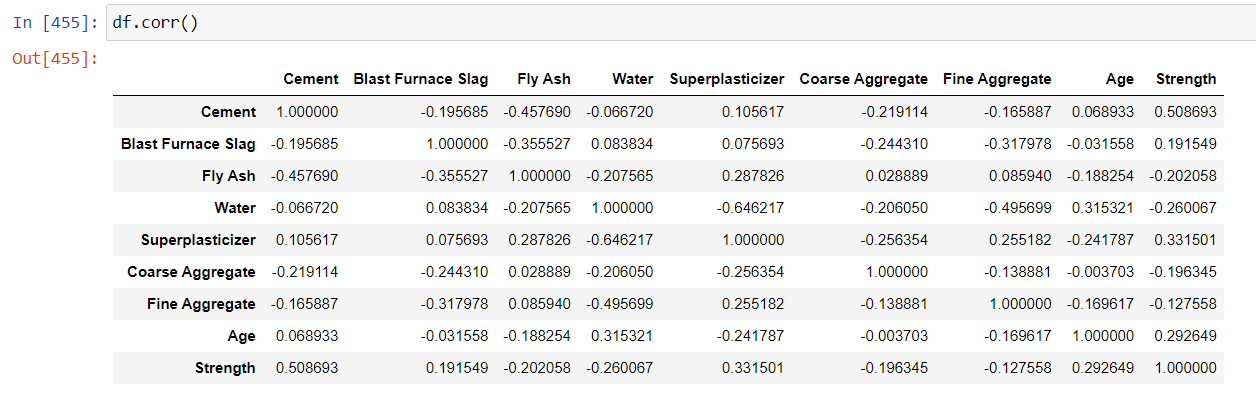
* *Uploading the dataset*

**

* *Details of the dataset*
* **
* *Observing the dataset features*

**

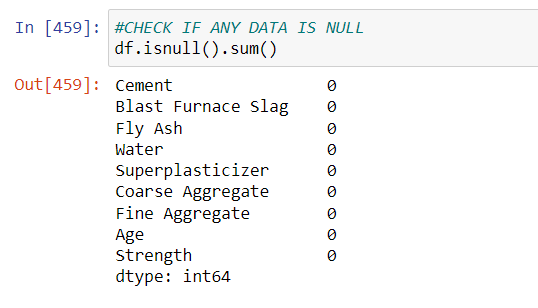
**

**

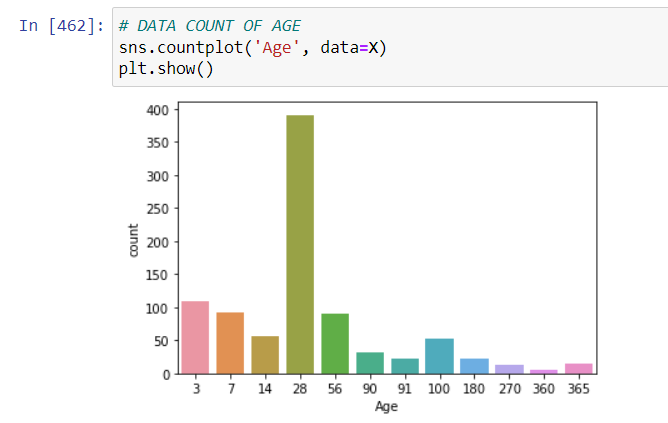
**

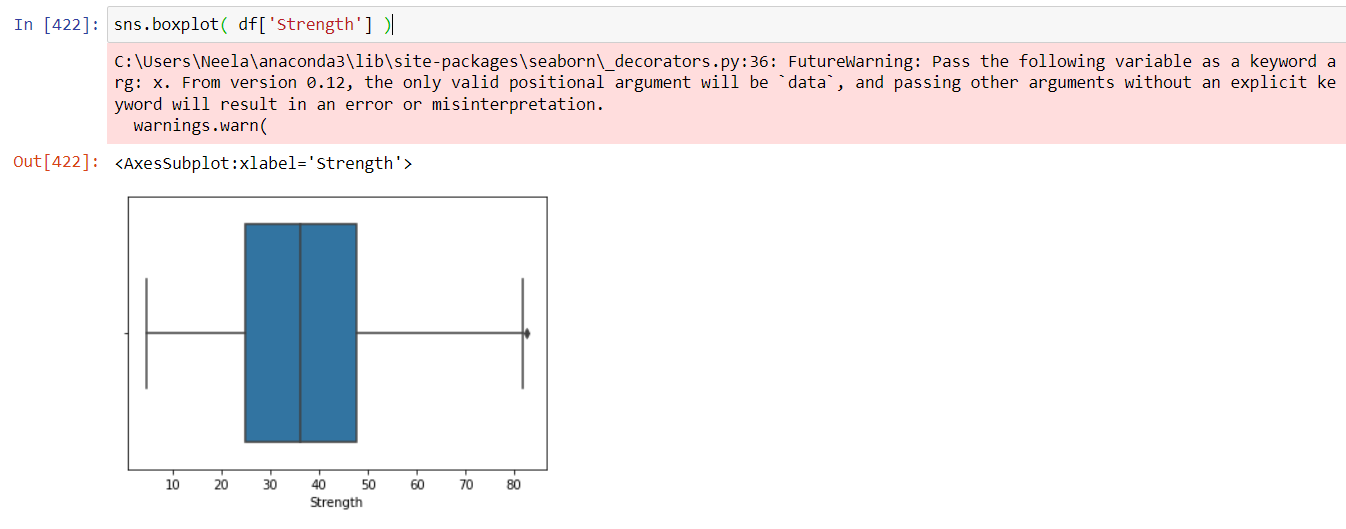
* *Drop any data if not required and check for null*

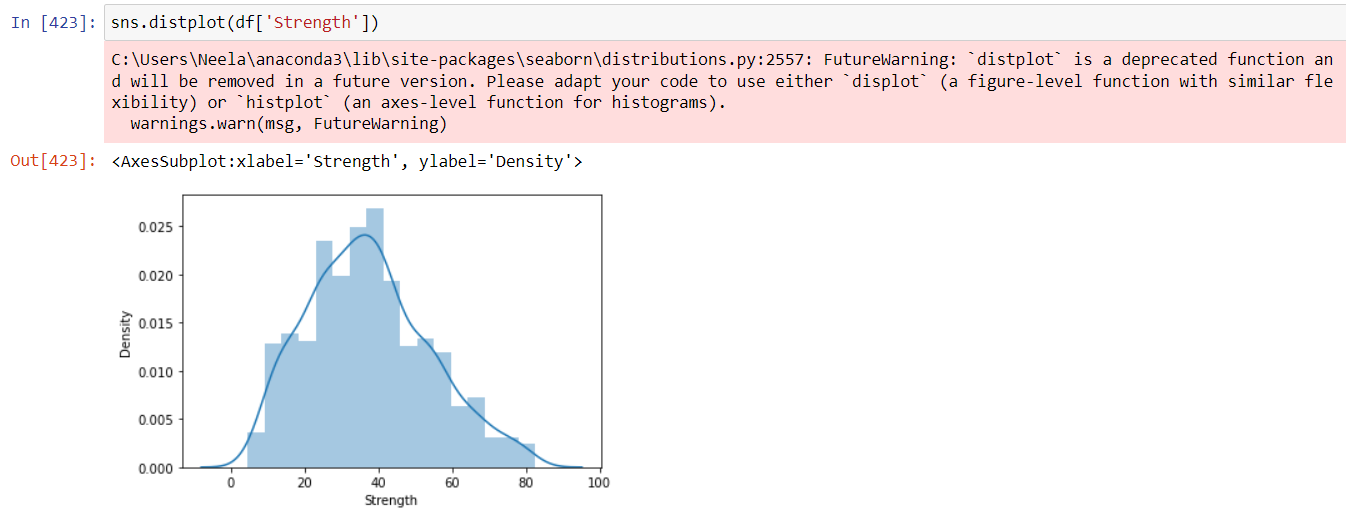
**

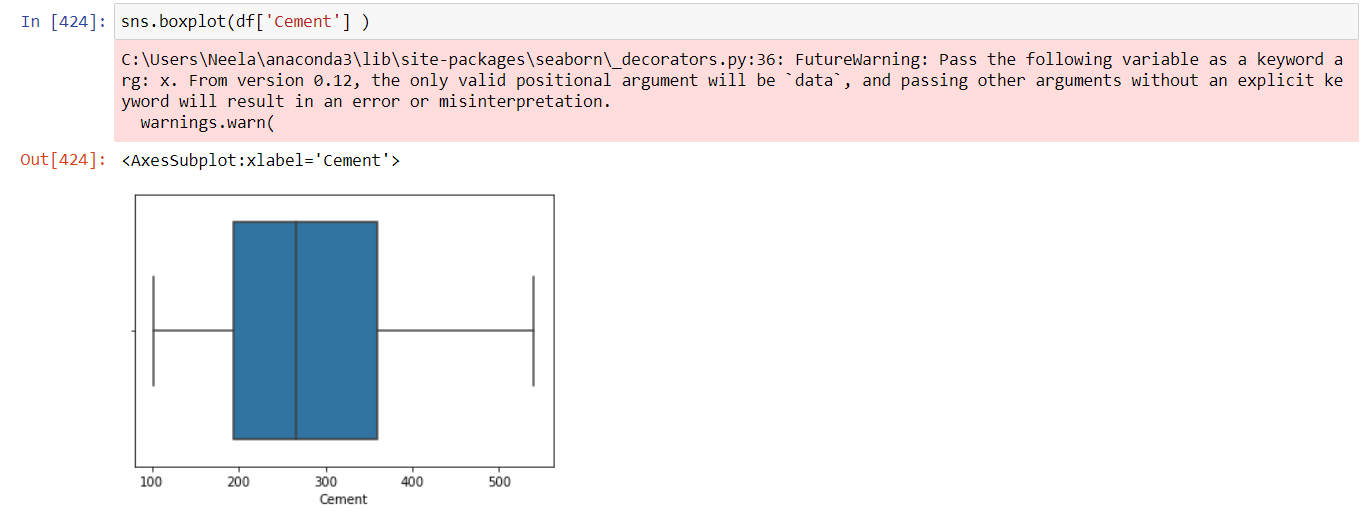
**

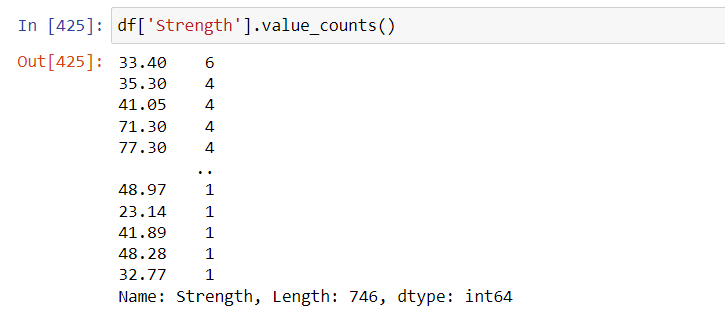
### EXPLORATORY DATA ANALYSIS

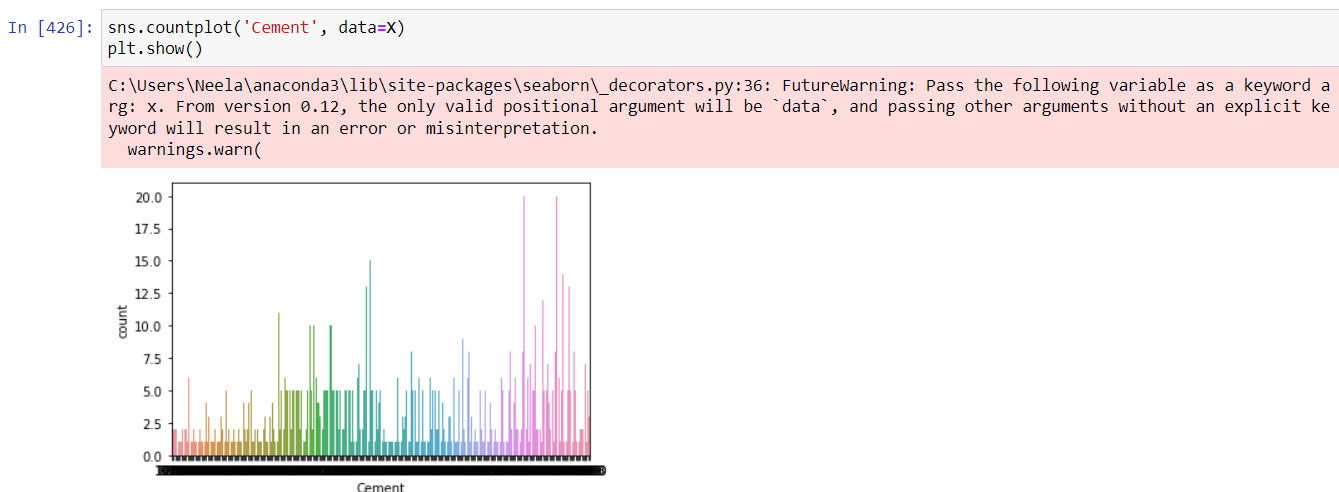
**

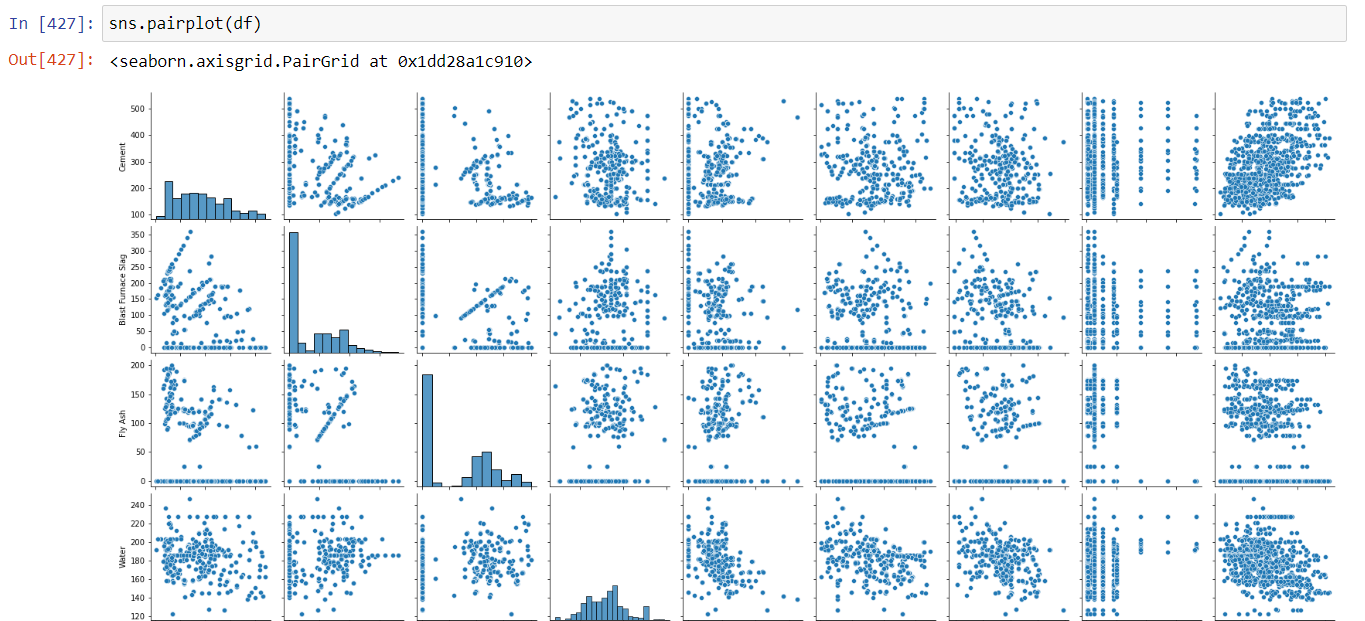
**

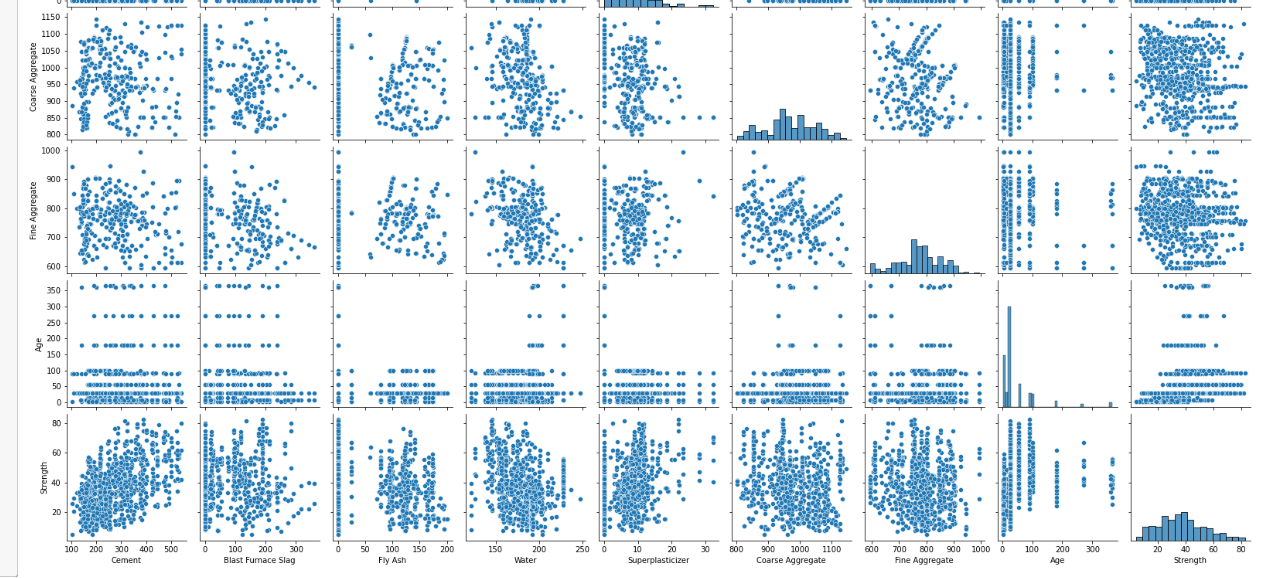
**

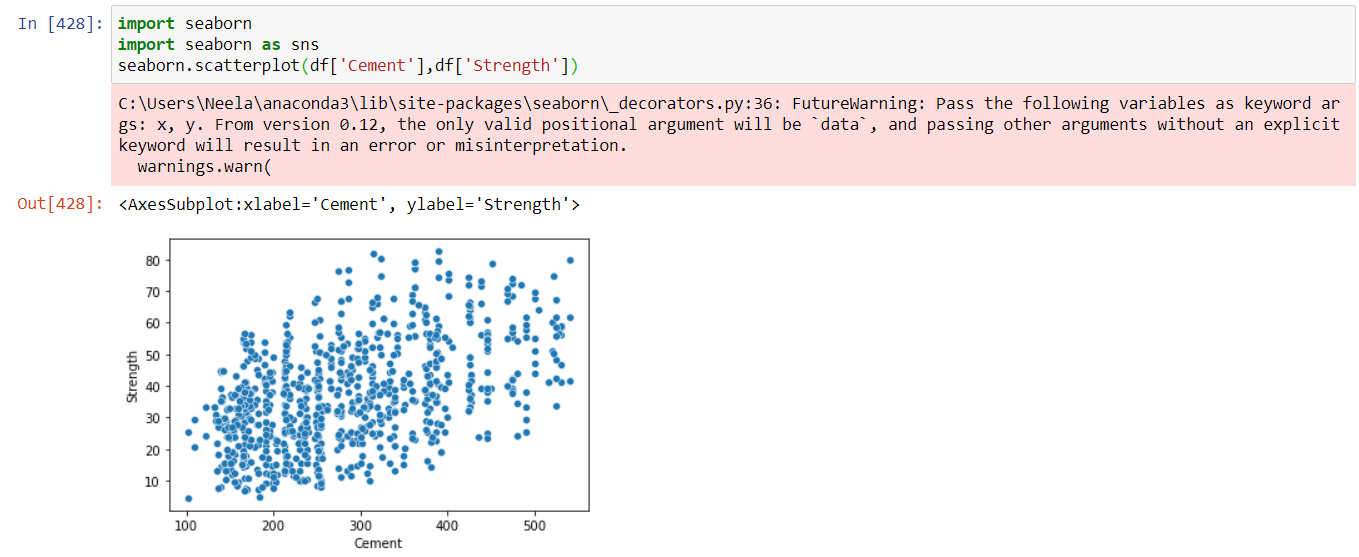
**

**

**

**

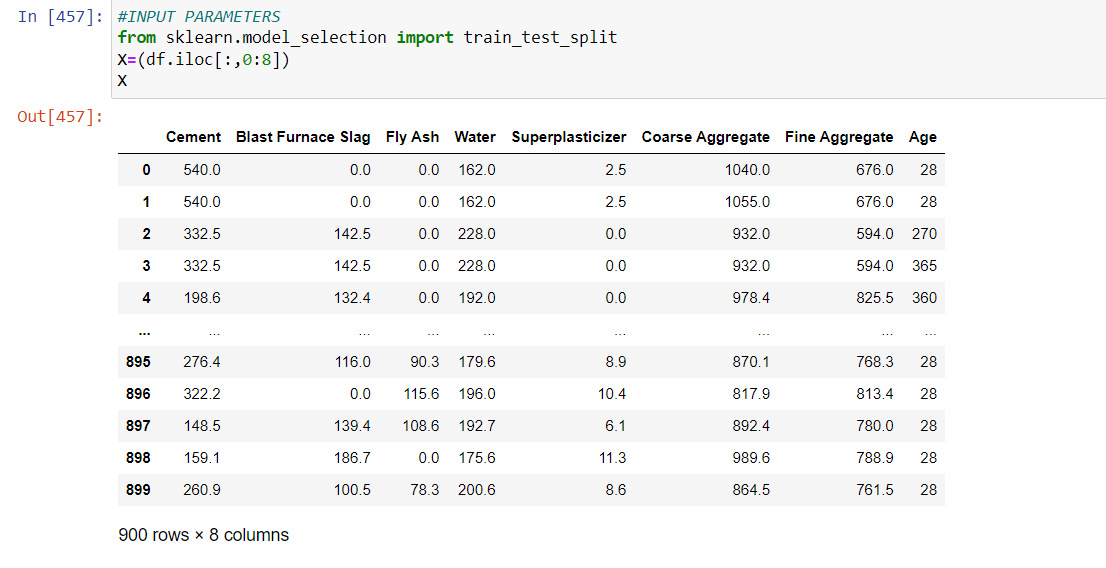
**

**

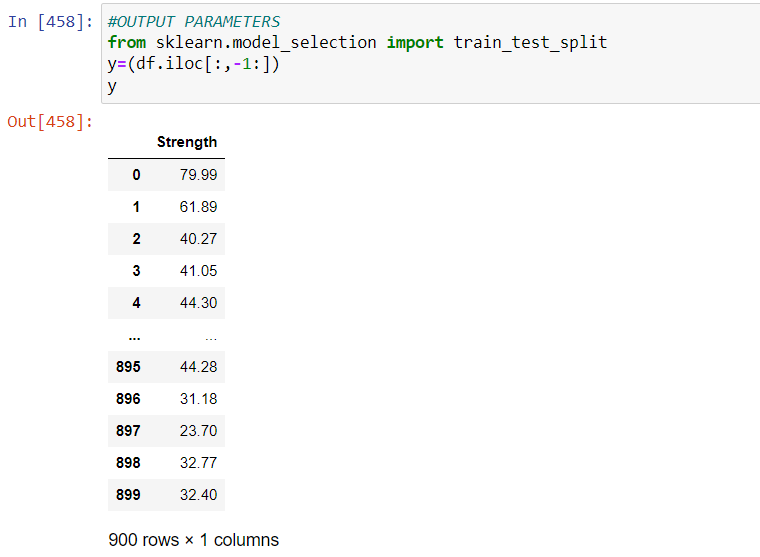
**

***PREDICTION MODEL DEVELOPMENT***

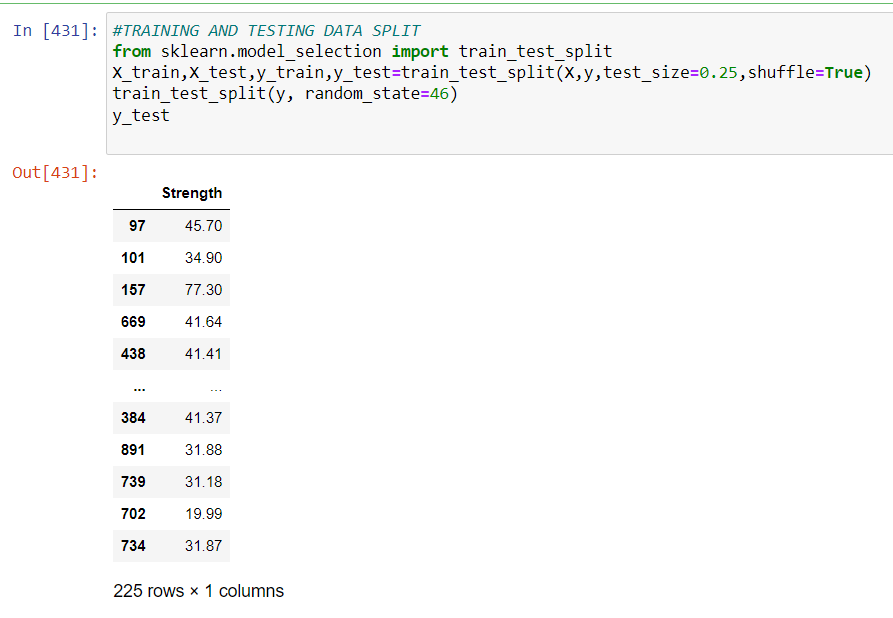
* *Input data*
* *O*

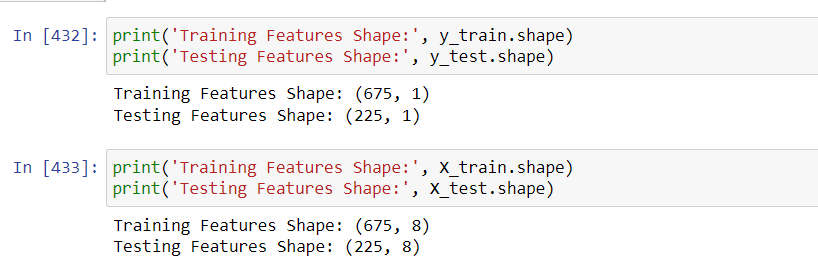
**

* *Output Parameter*

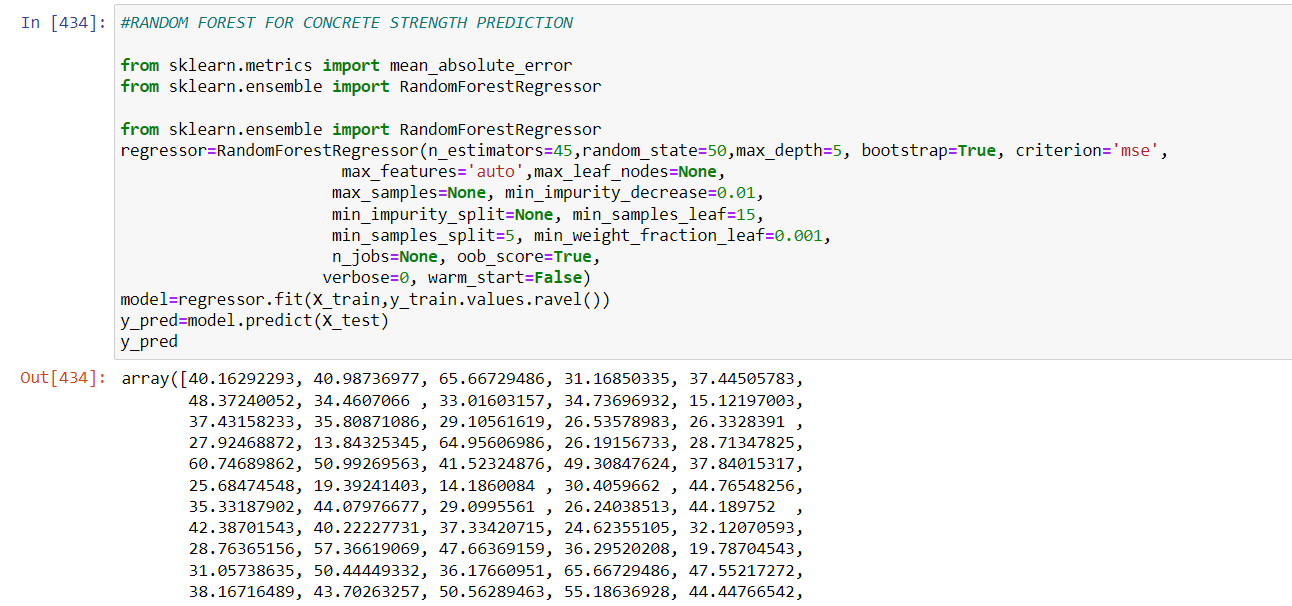
**

* *Split data*

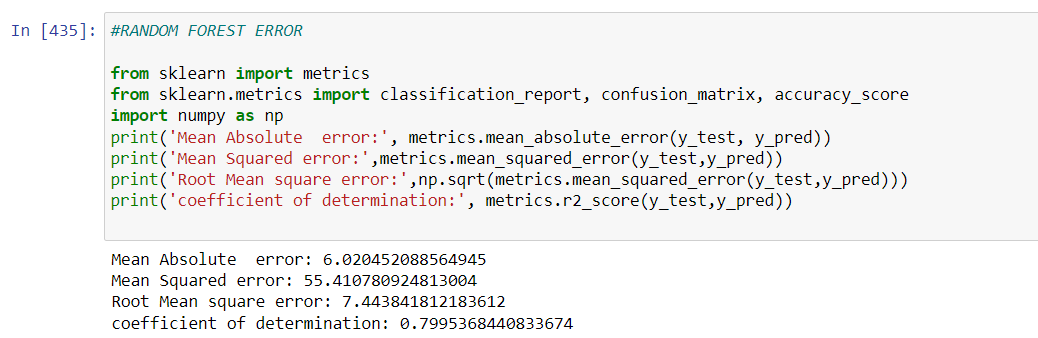
**

**

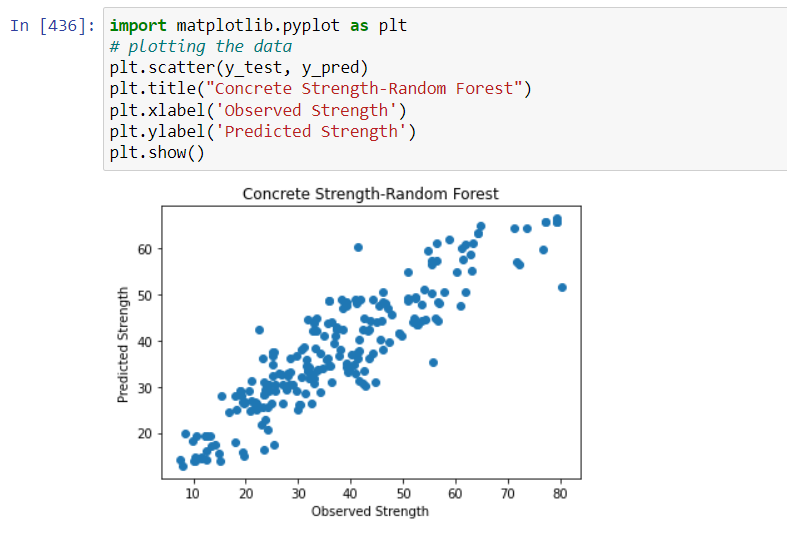
* *Concrete Strength prediction using Random Forest*

**

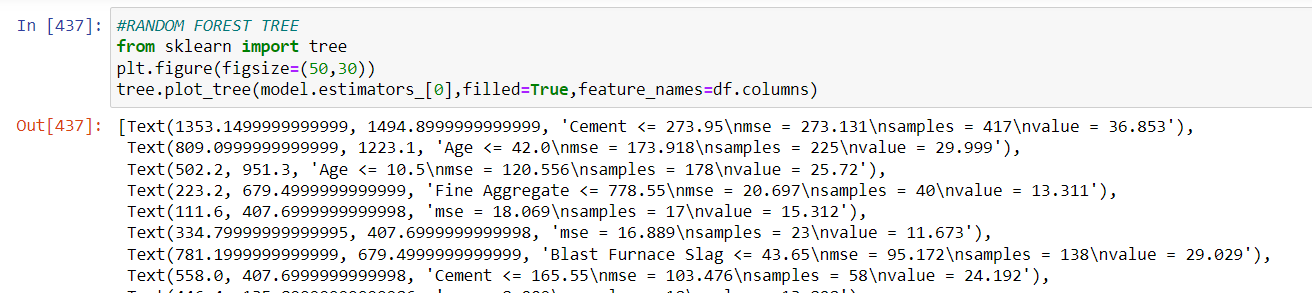
* *Error*

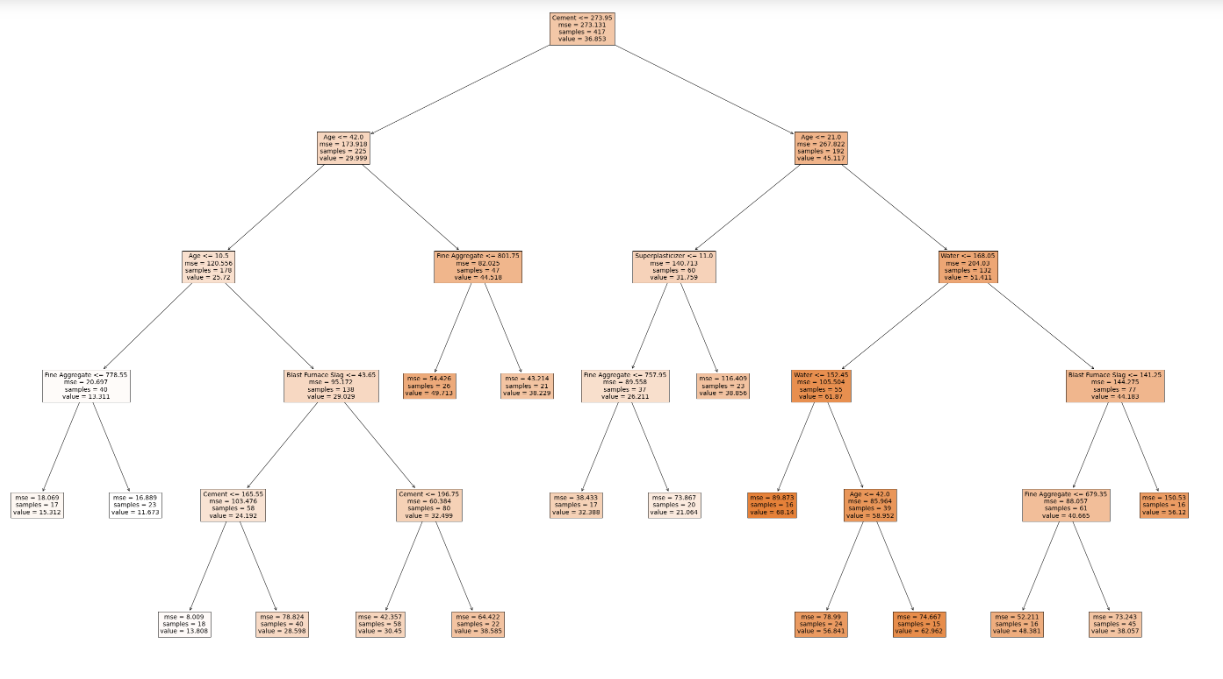
**

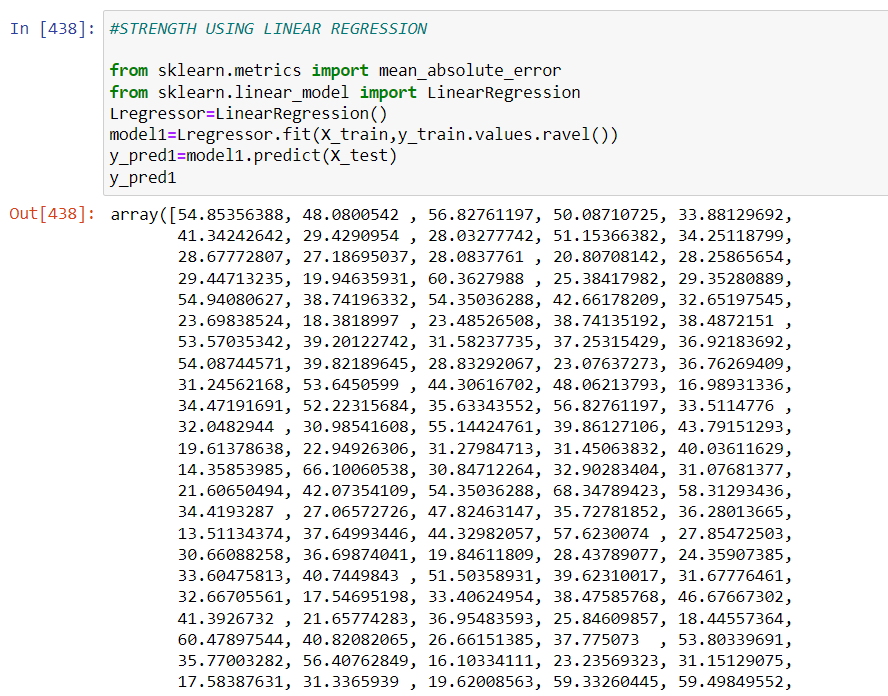
* *Scatter plot between Observed and Predicted*

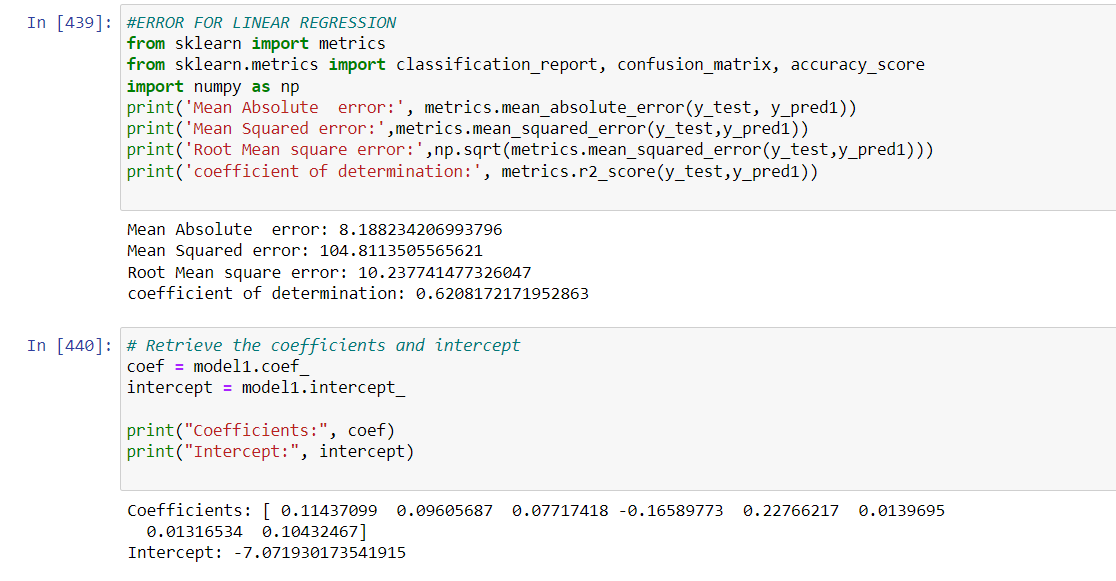
**

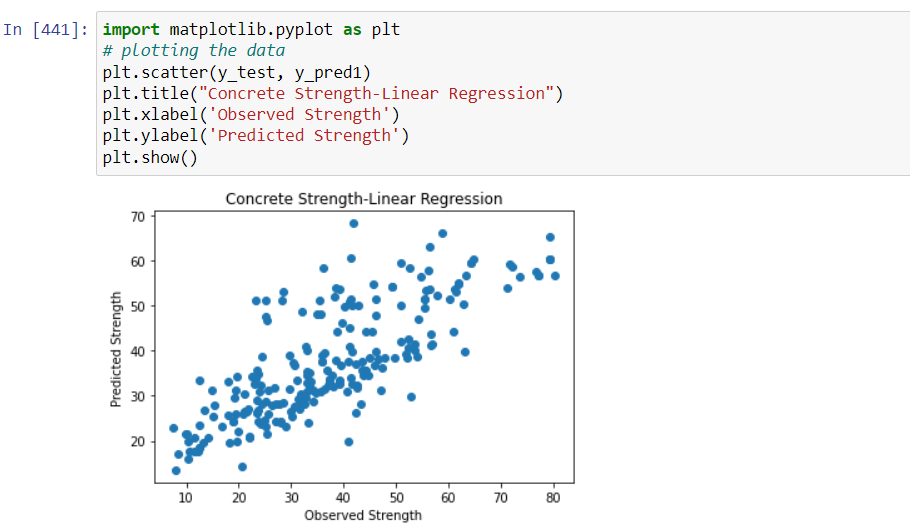
* *Random Forest Tree*



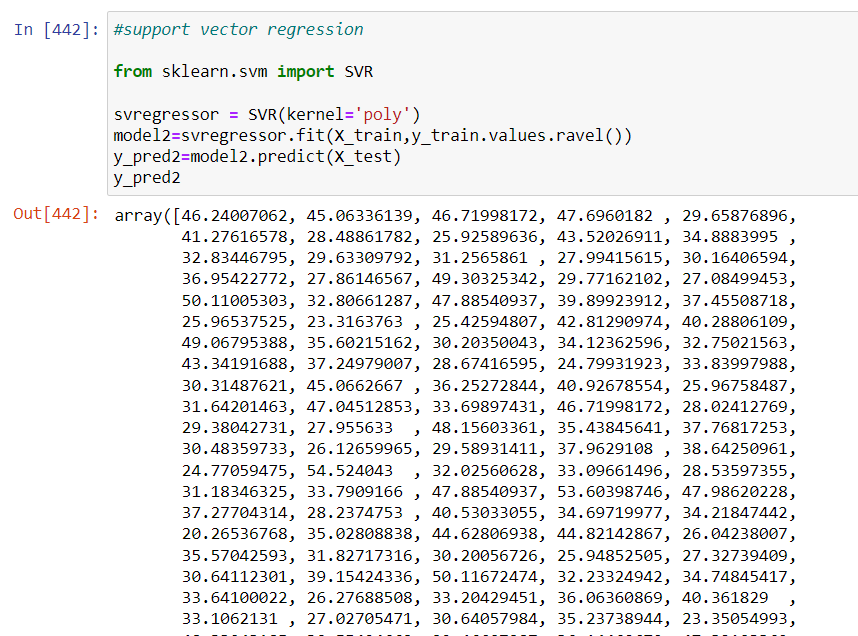
**

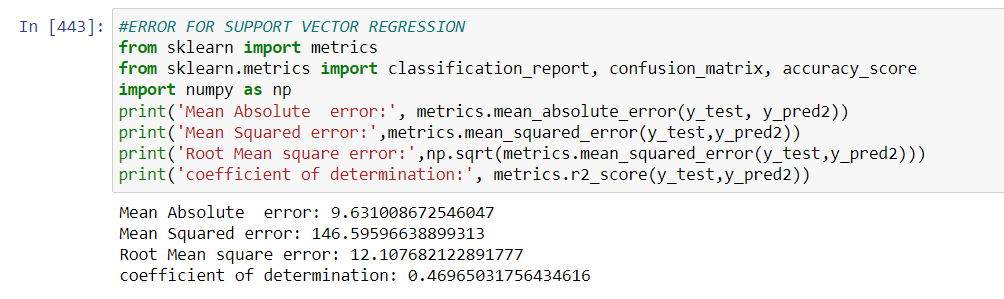
* *Concrete Strength Prediction using Linear Regression*
* *Error and Coefficients*

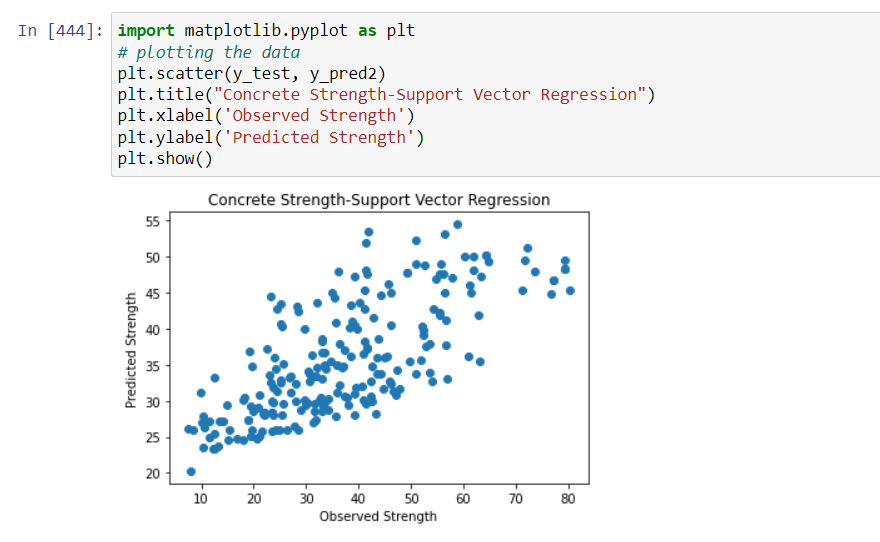
**



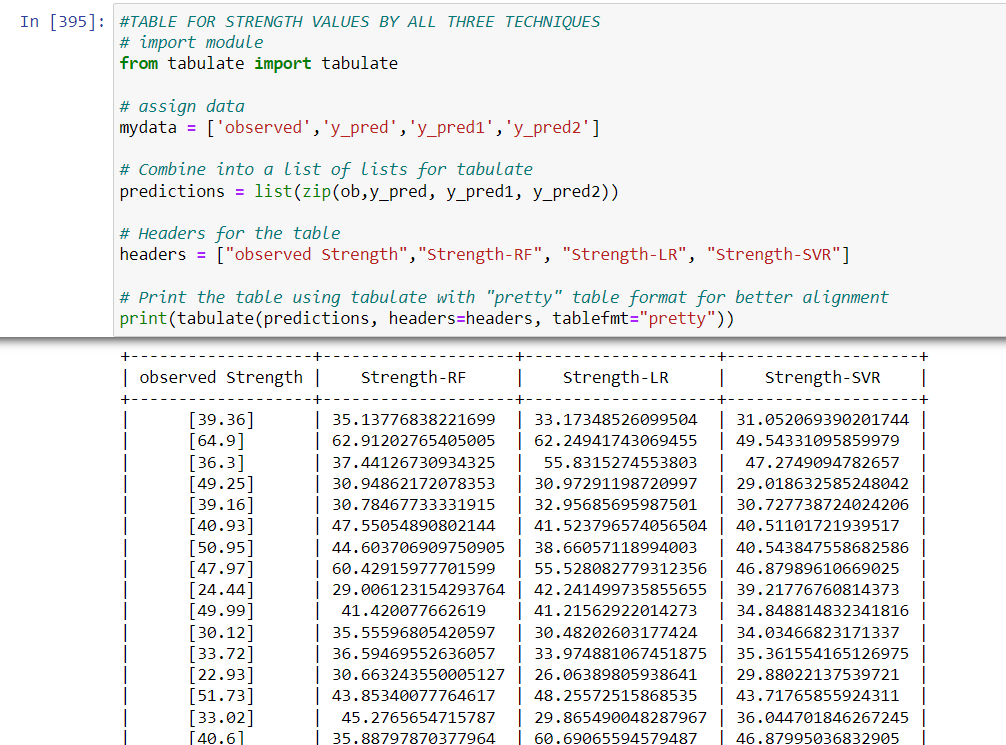
* *Concrete Strength Prediction using Support Vector Regression (Kernel-Polykernel)*

**

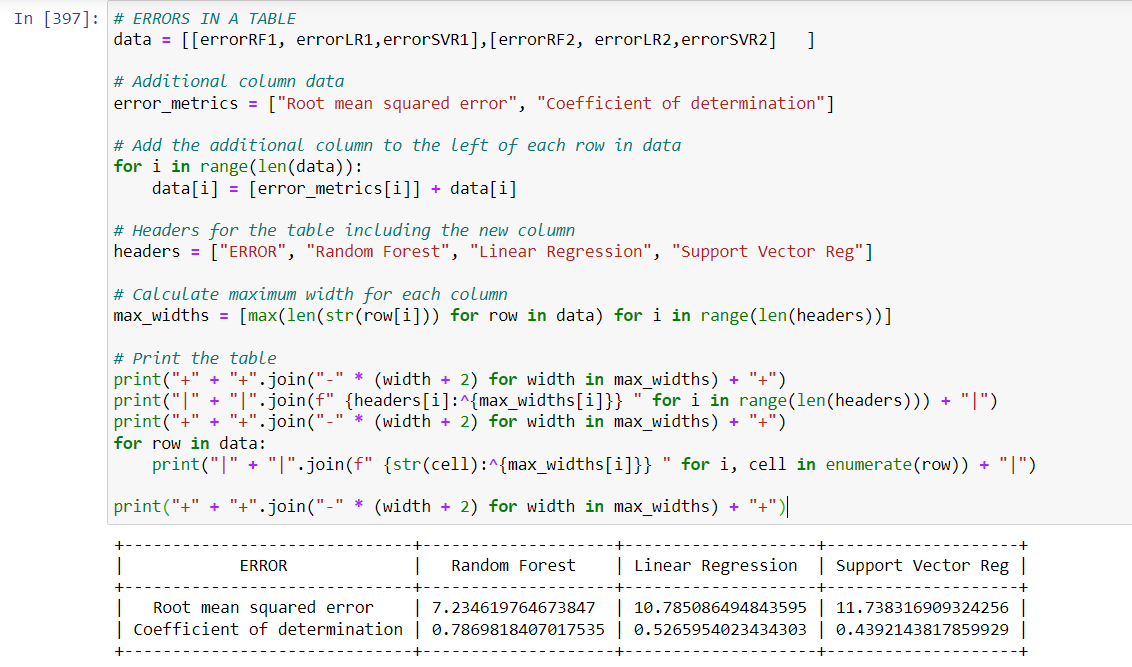
**

**

* *Predictions in Tabluar format*

**

* *Errors in table*

**

# Conclusion:

The present project’s aim is to predict strength of concrete using Linear regression, Random Forest and Support vector regression.

* The exploratory data analysis also shows a higher correlation of strength wit water and Age as input parameters.
* The error table shows that concrete strength prediction is done better by Random Forest with coefficient of determination 0.787 followed by Linear regression (coed=0.52) and Support vector regression (COD=0.439).
* The Root mean square by Random Forest is also lower followed by Linear Regression and Support Vector 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/>