**PROFESSIONAL TRAINING REPORT**

**at**

**Sathyabama Institute of Science and Technology (Deemed to be University)**

Submitted in partial fulfillment of the requirements for the award of Bachelor of Engineering in Computer Science and Engineering

By

**MOUNIKA.VAJRALA**

**REG.NO:39111050**

****

**DEPARTMENT OF COMPUITER SCIENCE AND ENGINEERING**

**SCHOOL OF COMPUTING**

**SATHYABAMA INSTITUTE OF SCIENCE AND TECHNOLOGY**

**JEPPIAAR NAGAR, RAJIV GANDHI SALAI,**

**CHENNAI – 600119, TAMIL NADU**

**NOVEMBER 2021**

|  |  |  |
| --- | --- | --- |
|  | **SATHYABAMA**  **INSTITUTE OF SCIENCE AND TECHNOLOGY**  **(DEEMED TO BE UNIVERSITY)**  **Accredited with Grade “A” by NAAC**  (Established under Section 3 of UGC Act, 1956)  JEPPIAAR NAGAR, RAJIV GANDHI SALAI  CHENNAI– 600119  [**www.sathyabama.ac.in**](http://www.sathyabama.ac.in) |  |

**DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING**

**BONAFIDE CERTIFICATE**

This is to certify that this Project Report is the bonafide work of **MOUNIKA.VAJRALA (Reg. No: 39111050)** who carried out the project entitled “**CONCRETE STRENGTH PREDICTION**” under my supervision from June 2021 to November 2021.

**Internal Guide**

**DR.M.Kanipriya M.E., Ph.D.**

**Head of the Department**

**Dr.S.VIGNESHWARI, M.E., Ph.D.**

**Dr.L.LAKSHMANAN, M.E., Ph.D.**

**.**



**Submitted for Viva voce Examination held on**

**Internal Examiner External Examiner**

**DECLARATION**

I, **MOUNIKA.VAJRALA** hereby declare that the project report entitled **Concrete Strength Prediction** done by me under the guidance of **Dr.M.KANIPRIYA M.E.,Ph.D** is submitted in partial fulfillment of the requirements for the award of Bachelor of Engineering Degree in Computer Science and Engineering.

**DATE: Vajrala.Mounika**

**PLACE: SIGNATURE OF THE CANDIDATE**

**ACKNOWLEDGEMENT**

I am pleased to acknowledge my sincere thanks to the **Board of Management** of **SATHYABAMA** for their kind encouragement in doing this project and for completing it successfully. I am grateful to them.

I convey my thanks to **Dr. T. Sasikala M.E., Ph.D**, **Dean**, School of Computing, **Dr.S. Vigneshwari,M.E.,Ph.D. and Dr.L.Lakshmanan,M.E.,Ph.D. Heads of the Department** of **Computer Science and Engineering** for providing me necessary support and details at the right time during the progressive reviews.

I would like to express my sincere and deep sense of gratitude to my Project Guide **Dr.M.KaniPriya,M.E., Ph.D.** for her valuable guidance, suggestions and constant encouragement paved the way for the successful completion of my project work.

I wish to express my thanks to all Teaching and Non-teaching staff members of the **Department of Computer Science and Engineering** who were helpful in many ways for the completion of the project.

**TRAINING CERTIFICATE**



**ABSTRACT**

Compressive Strength of Concrete determines the quality of Concrete. This is generally determined by a standard crushing test on a concrete cylinder. This requires engineers to build small concrete cylinders with different combinations of raw materials and test these cylinders for strength variations with a change in each raw material. The recommended wait time for testing the cylinder is 28 days to ensure correct results. This consumes a lot of time and requires a lot of labour to prepare different prototypes and test them. Also, this method is prone to human error and one small mistake can cause the wait time to drastically increase.

One of the most critical parameters in concrete design is compressive strength. As the compressive strength of concrete is correctly measured, time and cost can be decreased. Concrete strength is relatively resilient to impacts on the environment. The production of concrete compressive strength is greatly influenced by severe weather conditions and increases in humidity rates. In this research, a model has been developed to predict concrete compressive strength utilizing a detailed dataset obtained from previously published studies based on a deep learning method, namely, long short-term memory (LSTM), and a conventional machine learning (ML) algorithm, namely, support vector machine (SVM). The input variables of the model include cement, blast furnace slag, fly ash, water, superplasticizer, coarse aggregate, fine aggregate, and age of specimens. To demonstrate the efficiency of the proposed models, three statistical indices, namely, the coefficient of determination (*R*2), mean absolute error (MAE), and root mean square error (RMSE), were used. Findings shows that LSTM outperformed SVM with *R*2=0.98, *R*2= 0.78, MAE=1.861, MAE=6.152, and RMSE=2.36, RMSE=7.93, respectively. The results of this study suggest that high-performance concrete (HPC) compressive strength can be reliably measured using the proposed LSTM model.It was found that the accuracy of the random forest regression was considerable as per the result after applying the various regression techniques.

**Keywords-Machine learning, JupyterNotebook , Python language.**

i

|  |  |  |
| --- | --- | --- |
|  | **TABLE OF CONTENTS** |  |
| **CHAPTER No.** | **TITLE** | **PAGE No** |
|  | **ABSTRACT** | **i** |
|  | **LIST OF FIGURES** | **ii** |
|  |  |  |
| **iii** |
|  | **LIST OF ABBREVATIONS** |
| **1** | **INTRODUCTION** | **1** |
|  | 1.1 About Concrete Strength | 1 |
|  | 1.2 History and Features of Python  1.3 Features of Machine Learning  1.4 Summary of Project | 1  2  2 |
| **2.** | **AIM AND SCOPE** | **3** |
|  | 2.1 Aim of the project | 3 |
|  | 2.2 Concrete Strength Prediction | 3 |
|  | 2.3 Description of the dataset  2.4 Machine Learning Models | 4  6 |

**3. PROPOSED METHODOLOGY 8**

|  |  |  |
| --- | --- | --- |
|  | 3.1 Experimental Procedure | 8 |
|  | 3.2 Random Forest Regression (USED) | 9 |
|  | 3.3 Code Data  3.4 Research Significance | 10  19 |
|  |  |  |
| **4.** | **RESULTS AND DISCUSSION, PERFORMANCE ANALYSIS** | **20** |
|  | 4.1 Performance of Random Forest | 20 |
|  | 4.2 Feature importance analysis  4.3 Result | 22  24 |
| **5** | **SUMMARY AND CONCLUSIONS** | **25** |
|  |  |  |
|  | **APPENDIX** | **26** |
|  | * 1. SCREENSHOTS | **31** |
|  | * 1. SOURCE CODE   6.3 REFERENCES | **18**  **32** |

**LIST OF FIGURES**

**FIGURE NO. FIGURE NAME PAGE NO.**

2.1 Analysis of Strength 4

2.2 Example of Decision Tree 6

2.3 Example of Random Forest 8

3.3.1 - 3.3.3 Code Data 10

3.3.4 Code Data 11

3.3.5 - 3.3.6 Code Data 12

3.3.7 - 3.3.9 Code Data 13

3.3.10 - 3.3.11 Code Data 14

3.3.12 - 3.3.14 Code Data 15

3.3.15 - 3.3.16 Code Data 16

3.3.17 Code Data 17

3.3.18 - 3.3.19 Code Data 18

3.3.20 Code Data 19

4.1.1 Histogram of error for train 21

4.1.2 Histogram of error for test 22

4.2.1 Regression analysis for train 23

4.2.2 Regression analysis for test 23

4.3.1 Result 24

6.1-6.2 Screenshots 26

6.3-6.4 Screenshots 27

6.5 Screenshots 28

6.7-6.8 Screenshots 29

6.9-6.10 Screenshots 30

ii

**LIST OF ABBREVIATIONS**

**ABBREVATION EXPANSION**

LSTM **:** Long short-term memory

ML **:** Machine Learning

SVM **:** Support vector machine

R^2 **:** Coefficient of determination

MAE **:** Mean absolute error

RMSE **:** Root mean square error

HPC **:** High-performance concrete

ANN **:** Artificial neural network

AI **:** Artificial intelligence

LR **:** Linear Regression

MRA **:** Multiple regression analysis

MLR **:** Multiple Linear regression

RNN **:** Recurrent neural network

SRM **:** Structural risk minimization

RBF **:** Radial basis function

iii

**INTRODUCTION**

**1.1 ABOUT CONCRETE STRENGTH**

Quality of the concrete is generally determined by the compressive strength test of the concrete cylinder. Concrete strength is relatively resilient to impacts on the environment. The production of concrete compressive strength is greatly influenced by severe weather conditions and increases in humidity rates. The results of this study suggest that high-performance concrete (HPC) compressive strength can be reliably measured using the proposed LSTM model. Concrete is the mixture of Cement(c), Blast Furnace Slag, Water, Super Plasticizer, Coarse and Fine Aggregate, which is widely used construction material which is used in every kind of structures mainly in buildings. The procedure of determining the proportion of appropriate ingredients for producing concrete of required strength, workability, and durability is termed as mix design. The compressive strength of concrete is one of the widely used performance measure by the engineer in designing buildings. This study primarily focuses on finding the compressive strength of concrete.

.

**1.2 HISTORY AND FEATURES OF PYTHON**

Python was conceived in the late 1980s, and its implementation began in

December 1989 by Guido van Rossum at Centrum Wickenden & Informatica

(CWI) in the Netherlands as a successor to the ABC language (itself inspired by

SETL) capable of exception handling and interfacing with the Amoeba operating

system.

Python 3.0 (initially called Python 3000 or py3k) was released on 3 December

2008 after a long testing period. It is a major revision of the language that is not

completely backward-compatible with previous versions. Python is a multi-paradigm programming language. Object-oriented programming and structured programming are fully supported, and many of its features support functional programming and aspect-oriented programming (including by meta programming. Many other paradigms are supported via extensions, including design by contract and logic programming.

Python uses dynamic typing, and a combination of reference counting and a

cycle-detecting garbage collector for memory management. It also features

dynamic name resolution (late binding), which binds method and variable names

during program execution.

Python's design offers some support for functional programming in the Lisp

tradition. It has filter(), map(), and reduce() functions; list comprehensions,

dictionaries, and sets; and generator expressions. The standard library has two modules that implement functional tools borrowed from Haskell and Standard ML. Rather than having all of its functionality built into its core, Python was designed to be highly extensible.

**1.3 FEATURES OF MACHINE LEARNING**

Machine language is the language understood by a computer. It is very difficult to understand, but it is the only thing that the computer can work with. All programs and programming languages eventually generate or run programs in machine language. Machine language is made up of instructions and data that are all binary numbers.

In machine learning and pattern recognition, a feature is an individual measurable property or characteristic of a phenomenon. Choosing informative, discriminating and independent features is a crucial element of effective algorithms in pattern recognition, classification and regression. Features are usually numeric, but structural features such as strings and graphs are used in syntactic pattern.

Features can be raw data which are very straightforward and can be derived from real-life as it is. However, not all problems can be solved using raw data or data in its original form. Many times, they need to be represented or encoded in different form. For example, a color can be represented in RGB format or HSV format. Thus, a color can have two different representations or encodings. And, both of these representations or encodings can be used to solve different kind of problems. Some tasks that may be difficult with one representation can become easy with another. For example, the task “select all red pixels in the image” is simpler in the RGB format, whereas “make the image less.

Machine-learning models are all about finding appropriate representations / features for their input data—transformations of the data that make it more amenable to the task at hand, such as a classification task.

**1.4 SUMMARY OF PROJECT**

The compressive strength of concrete determines the quality of the concrete. To Predict Concrete strength of concrete with maximum accuracy, for various quantities of constituent components. Using random forest regression we achieve to predict the Strength of the Concrete.

Using random forest regression, we achieve to predict the Strength of the Concrete. To evaluate nature of concrete strength gain pattern with time for a particular type of mix. To develop a simple relation which has the potential to predict the compressive strength of the concrete. The objective of this project is trying to predict the concrete compressive strength based important predictors.

The challenge in predicting concrete strength by machine learning was first described by Yeh et al. in 1998. Using seven input variables, they performed Random Forest regression to predict the strength of high-strength concrete. Yeh et al. trained their algorithm on many concrete samples, but they were not filtered in terms of content.

**AIM AND SCOPE OF PRESENT INVESTIGATION**

**2.1 AIM OF THE PROJECT**

The Objective of our project is to design a program that can predict the strength of the concrete. This helps to measure the accuracy of the concrete to sustain, it can run in any python environment.

**2.2 CONCRETE STRENGTH PREDICTION**

Compressive strength of concrete has been considered as an index of quality control for many years. Quality of the concrete is generally determined by the compressive strength test of the concrete cylinder. When engineers design a concrete structure, they basically rely on the characteristic strength of the concrete.

Concrete is the mixture of Cement(c), Blast Furnace Slag, Water, Super Plasticizer, Coarse and Fine Aggregate, which is widely used construction material which is used in every kind of structures mainly in buildings. The procedure of determining the proportion of appropriate ingredients for producing concrete of required strength, workability, and durability is termed as mix design. The compressive strength of concrete is one of the widely used performance measure by the engineer in designing buildings. This study primarily focuses on finding the compressive strength of concrete. One of the limitations with the current Mix Design methodology is,it provides a mechanism to find out the mix ratio for a given strength values but the reverse is not possible.

Strength gaining process of concrete is a multi-factor dependent complex process. There are numerous studies in this regard. Even at present the researchers have shown keen interest in it. Knowing concrete strength gain patterns enables one to predict the concrete characteristic strength at an early age and gives an idea about the quality of the concrete compliance with the design requirement.

The proposed model proved to be a major tool in predicting compressive strength of different concrete from a ready mix plant in spite of variations in the results.

The strengthening of concrete is a complex process involving many factors.

Statistical analysis can also provide insight into the key factors influencing comprehensive strength through correlation analysis.

**2.3 DESCRIPTION OF THE DATASET**

The business meaning of each column in

the data is as below

• Cement Component: How much cement is mixed

• Blast Furnace Slag: How much Blast Furnace Slag is mixed

• Fly Ash Component: How much Fly Ash is mixed

• Water Component: How much water is mixed

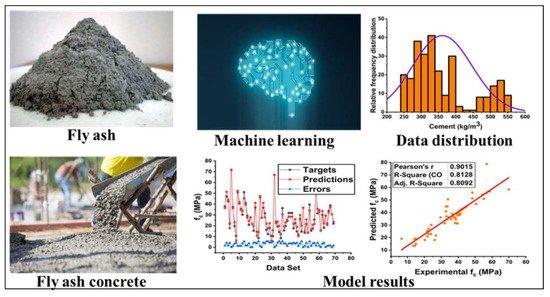
• Superplasticizer Component: How much Super plasticizer is mixed

• Coarse Aggregate Component: How much Coarse Aggregate is mixed

• Fine Aggregate Component: How much Coarse Aggregate is mixed

• Age in Days: How many days it was left dry

• Strength: What was the final strength of concrete



**FIGURE 2.1: ANALYSIS OF STRENGTH**

Cement is manufactured through a closely controlled chemical combination of calcium, silicon, aluminium, iron and other ingredients. Common materials used to manufacture cement include limestone, shells, and chalk or marl combined with shale, clay, slate, blast furnace slag, silica sand, and iron ore.

Blast furnace slag is a non-metallic coproduct produced in the process. It consists primarily of silicates, aluminosilicates, and calcium-alumina-silicates. The molten slag, which absorbs much of the sulphur from the charge, comprises about 20 percent by mass of iron production.

Fly ash is predominantly composed of silica, aluminum, iron, calcium, and oxygen, but the particles may also contain heavy metals such as arsenic and lead at trace levels. Most nations throughout the world do not consider fly ash a hazardous waste and therefore regulations on its disposal and storage are lacking.

This basically tells us that the water molecule is composed of two elements: hydrogen and oxygen or, more precisely, two hydrogen atoms (H2) and one oxygen atom (O). Hydrogen and oxygen are gases at room temperature. ... Splitting water into its two components is much easier to do and is called water electrolysis.

Superplasticizers (SP's), also known as high range water reducers, are additives used in making high strength concrete. Plasticizers are chemical compounds that enable the production of concrete with approximately 15% less water content. Superplasticizers allow reduction in water content by 30% or more. These additives are employed at the level of a few weight percent. Plasticizers and superplasticizers retard the curing of concrete.

Generally, superplasticizer can be classified into such types: purified lignosulfonates, carboxylate synthetic polymers, sulfonated synthetic polymers and synthetic polymers with mixed functionality cementitious materials.

SPs are used where well-dispersed particle suspension is required to improve the flow characteristics (rheology) of suspensions such as in concrete applications. Their addition to concrete or mortar allows the reduction of the water to cement ratio without negatively affecting the workability of the mixture, and enables the production of self-consolidating concrete and high-performance concrete. They greatly improve the performance of the hardening fresh paste. The strength of concrete increases when the water to cement ratio decreases.

Coarse aggregates have a wide variety of construction applications because they resemble standard rock particles, as opposed to fine aggregate, which more closely resembles sand.

Coarse aggregates are an integral part of many construction applications, sometimes used on their own, such as a granular base placed under a slab or pavement, or as a component in a mixture, such as asphalt or concrete mixtures. Coarse aggregates are generally categorized as rock larger than a standard No. 4 sieve (3/16 inches) and less than 2 inches.

Fine aggregate, It assists in producing workability and uniformly in mixture.

It assists the cement paste to hard the coarse aggregate particles.

It helps to prevent possible segregation of paste and coarse aggregate particularly during the transport operation of concrete for a long distance.

Fine aggregate reduces the shrinkage of binding material.

It prevents the development of a crack in the concrete.

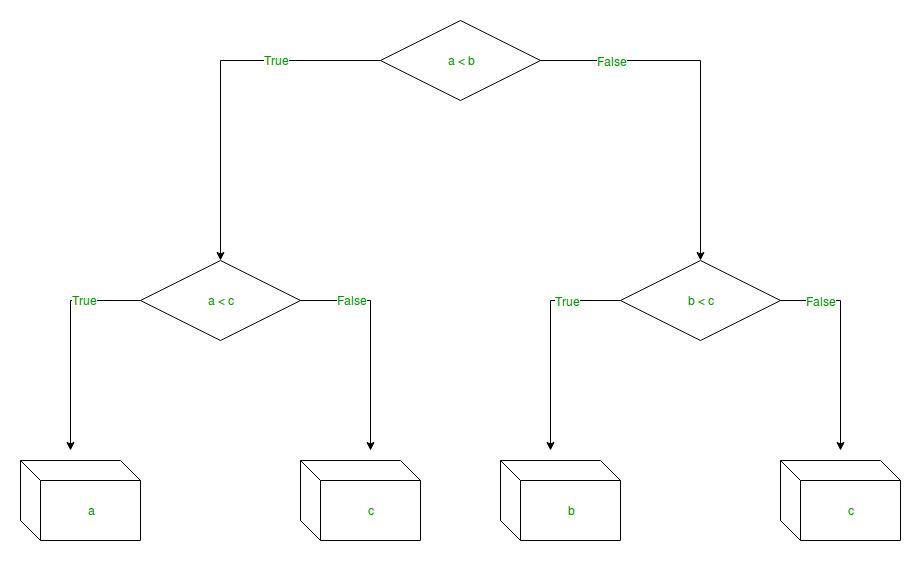
It fills the voids existing in the coarse aggregate. Thus, it helps in increasing the density of concrete.

**2.4 MACHINE LEARNING MODELS**

**Decision Tree**

A Decision Tree Algorithm represents the data with a tree-like structure, where each node represents a decision taken on a feature.

Decision tree regression observes features of an object and trains a model in the structure of a tree to predict data in the future to produce meaningful continuous output. Continuous output means that the output/result is not discrete, i.e., it is not represented just by a discrete, known set of numbers or values.



**FIGURE 2.2: EXAMPLE OF DECISION TREE**

**Linear Regression**

The linear regression algorithm tries to form a linear relationship between the input features and the target variable. Statistical approach for modelling the relationship between a scalar dependent variable and one or more explanatory variables in statistics, linear regression is a linear approach for modelling the relationship between a scalar response and one or more explanatory variables (also known as dependent and independent variables). The case of one explanatory variable is called simple linear regression; for more than one, the process is called multiple linear regression.

Linear regression is the simplest traditional method of regression models. The prediction equation consists of a polynomial of linear parameters associated with predictor variables.

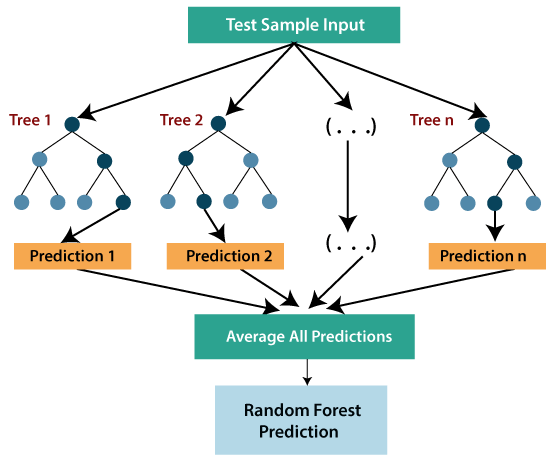
It attempts to model the relationship between two variables by fitting a linear equation to observed data. One variable is considered to be an explanatory variable, and the other is considered to be a dependent variable. For example, a modeler might want to relate the weights of individuals to their heights using a linear regression model.

**RANDOM FOREST**

Random forests can be used for classification and regression. Random Forests can evaluate variables importance by quantifying how the prediction error increases when data for a variable is permuted while keeping other variables constant. A randomly selected subset of explanatory variables is used for building single trees.

A Random Forest is an ensemble technique capable of performing both regression and classification tasks with the use of multiple decision trees and a technique called Bootstrap Aggregation, commonly known as bagging. What is bagging you may ask? Bagging, in the Random Forest method, involves training each decision tree on a different data sample where sampling is done with replacement.

They are easy to use without many pre-processing strategies. The idea is to build a collection of trees with a controlled variation. Random Forest is a clustering technique that can perform regression and classification tasks using multiple decision trees and bootstrap aggregation, commonly known as bagging. Bagging, in the Random Forest method, involves training each decision tree with a different data sample, where sampling is done with a substitution. The biggest problem with decision trees is that they tend to over-fit training data. Error pruning is the most common technique for avoiding this type of problem. In this project, the Random Forest model is defined with the help of Grid search CV.



**FIGURE 2.3: EXAMPLE OF RANDOM FOREST**

**3. EXPERIMENTAL OR MATERIALS AND METHODS, ALGORITHMS USED**

**3.1 EXPERIMENTAL PROCEDURE**

Recently, random forests have received a lot of attention in various fields because they can handle large numbers of variables with relatively small numbers of data. Also, Random Forests provide an assessment of variable importance. Random Forests approach (RF) is grown by a procedure called bagging, which is short for bootstrap aggregating, where trees are independently built by using a bootstrap sample (with replacement) of the entire data set.

In Random Forest a Decision Tree is built by randomly sampling a feature subset, and/or by the random sampling of a training data subset for each Decision Tree (the concept of Bagging). Each node of the trees is split using a subset of the explanatory variables chosen randomly for each tree. Nodes are homogeneous groups of data. Constructed trees are random, and overall prediction of the forest is the average of predictions from the individual trees.

The parameters that must be tuned for growing a random forest are the number of attributes chosen at each split and the number of trees to be grown. Once a tree has been built, the response for any observation can be forecasted by tracing the path from the root node down to the appropriate terminal node of the tree, based on the observed values for the splitting variables, and the predicted response value simply is the average response in that terminal node.

For regression analysis, data are partitioned to more homogeneous groups called nodes, each split is based on data of splitting variables. After a tree is split, the response variable can be predicted by following the path from the root node.

**3.2 RANDOM FOREST USING REGRESSION(USED)**

Random Forest Regressor trains randomly initialized trees with random subsets of data sampled from the training data, this will make our model more robust.

Random Forests model is a novel algorithm used in regression where it can assess the importance of input variables in a robust manner for a large number of variables. Adding Silica to concrete mix produces high early strength material which is a desirable feature. We analyze the results of 90 experimental samples that tested the effect of adding Silica on the compressive strength of concrete.

In this study, Random Forests model shows good performance and capability of predicting cement compressive strength-based Silicate percentage and other variables. The R2 for prediction model is high at 89%. Also, Random Forest model evaluated variables importance and indicated the curing time and milling time has higher impact than silicate percentage. The developed model can be used to predict compressive strength as a response to silicate percent with adequate accuracy.

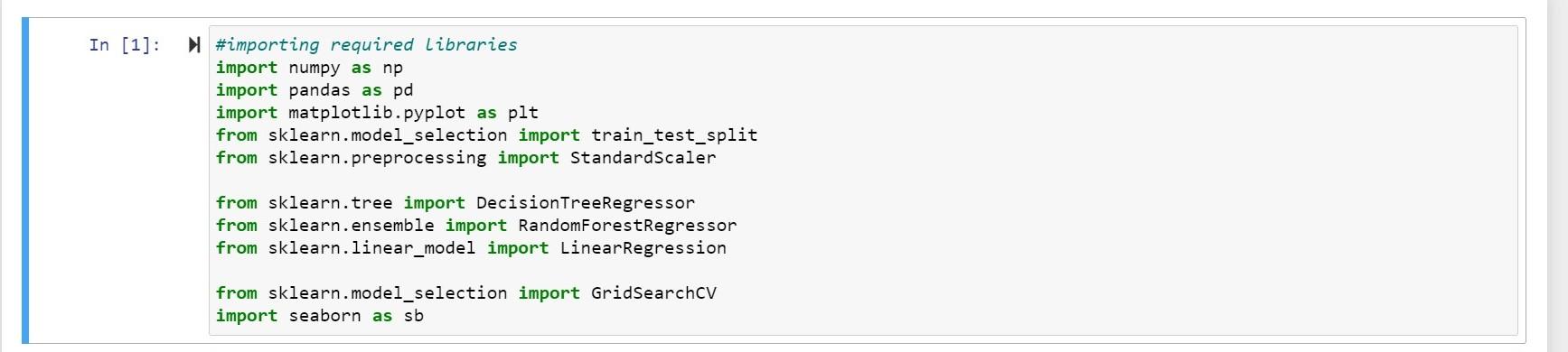
Random Forest models are constructed from a collection of decision trees.

Random forest (RF) is an ensemble method that combines many DTs. It can be used for both regression and classification. Each DT in the forest is created by the selection of different samples from the original dataset by the bootstrap technique. These samples are then trained using a set of attributes selected by the bagging mechanism. Subsequently, the decisions made by a large number of individual trees are subjected to voting. As such, the most voted class is presented as the class estimate of the community.

**3.3 CODE DATA**

**IMPORTING REQUIRED LIBRARIES**

The most common way to import pandas into your Python environment is to use the following syntax: import pandas as pd. The import pandas portion of the code tells Python to bring the pandas data analysis library into your current environment. The as pd portion of the code then tells Python to give pandas the alias of pd.

****

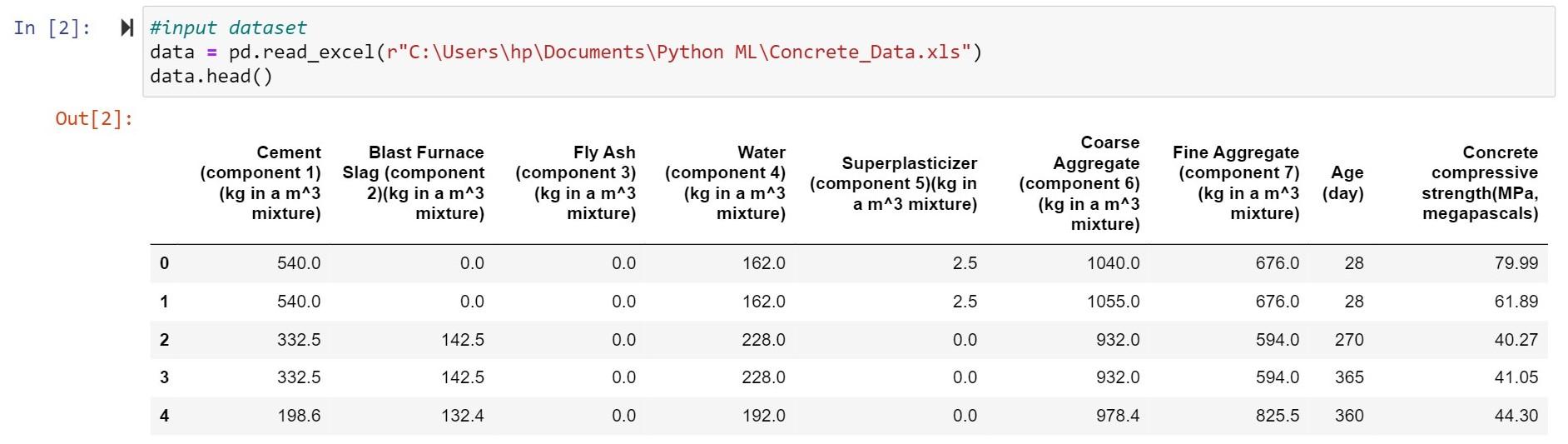
**FIGURE 3.3.1**

We import **pandas** for data analysis, NumPy contains a multi-dimensional array and matrix data structures, NumPy is an extension of Numeric and Num array, **seaborn**, and **matplotlib** to visualize the data.

Sklearn is for model selection and for importing various regression or classification models and to import train test split to train the models and to test them.

**READING DATASET**

Generally, we use a dataset in the form of a CSV file, for reading this CSV file we will use the **panda’s** library

****

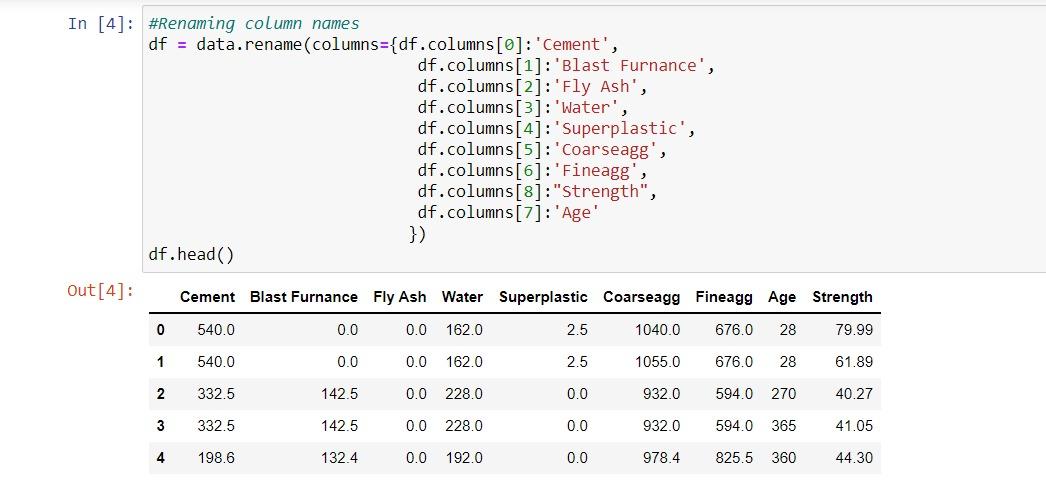
**FIGURE 3.3.2**

Copying the dataset into another:



**FIGURE 3.3.3**

**Changing the column name in the dataset**

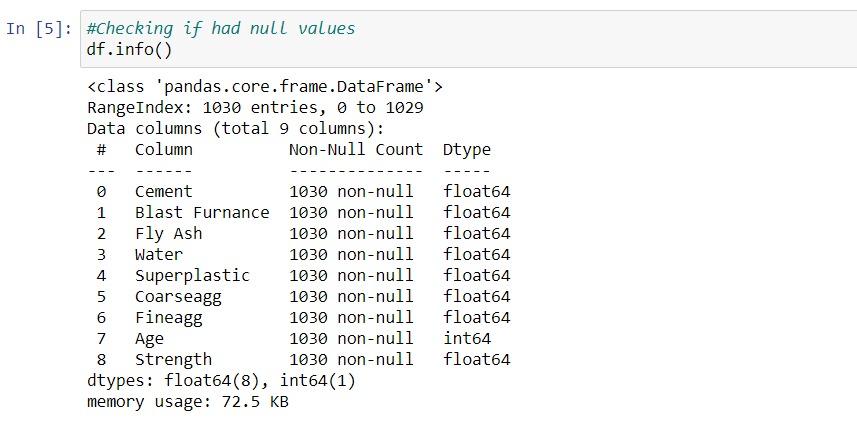
****

**FIGURE 3.3.4**

Changing the dataset column name using the rename function.

**Dataset info**

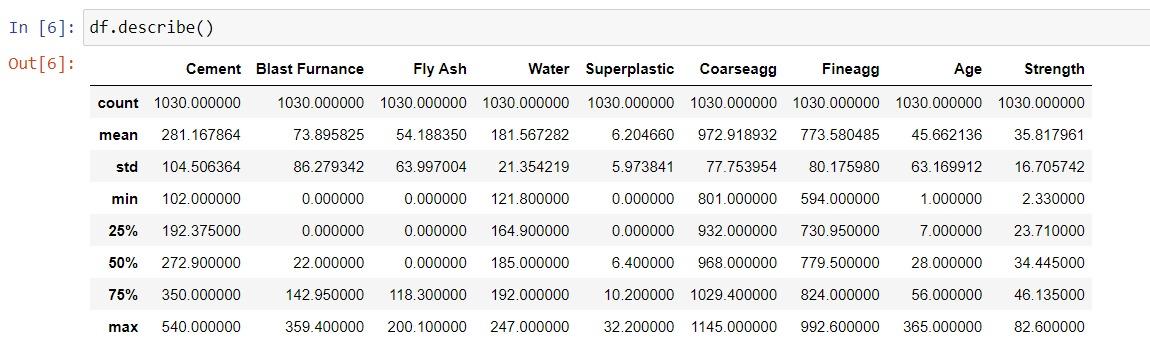
After reading the dataset we have to extract information from the data, for that we use the certain function:



**FIGURE 3.3.5**

Here we notice the count of null values in each feature and see what is the data type of features present in the dataset.

**Describing the dataset**



**FIGURE 3.3.6**

**describe()**method calculate the various calculation of each data point in the feature.

**Dividing Dependent and Independent Variables**

Before starting model building we have to divide the dataset into two parts,

1. **Independent** variables contain a list of those variables in which concrete quality is dependent.
2. The**dependent** variable is that variable that is dependent on other variables’ values.



**FIGURE 3.3.7**

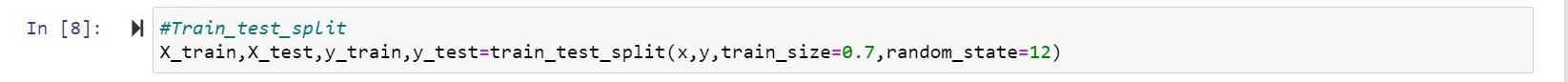
In this program, x contains the list of independent variables, and y contains the dependent variable in this case:

1. Independent variables are **cement, fly ash, water, superplasticizer, coarse aggregate, fine aggregate, age**.

2. dependent variable is the only **Strength.**

**Splitting the data**

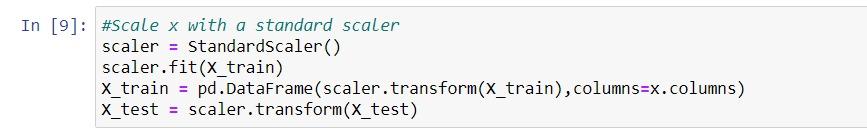
Now we use the scikit-learn module train\_test\_split, which is used for splitting the training and testing parts.



**FIGURE 3.3.8**

**Feature Scaling**

We do scaling of data for balancing the data points.



**FIGURE 3.3.9**

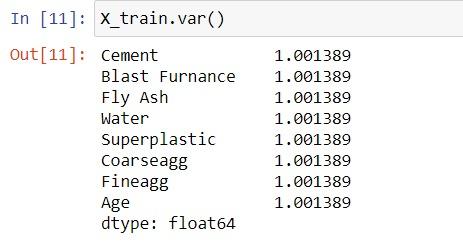
In this program first, we import train\_test\_split from scikit-learn then create **StandardScaler()** class object, after creating the object we fit train data into StandardScaler for scaling the data and then we transform the train and test data into an array.

**Printing the X\_train**



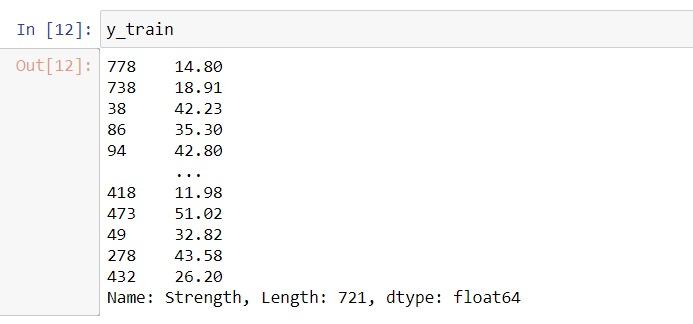
**FIGURE 3.3.10**

**Printing the variance of X\_train**

****

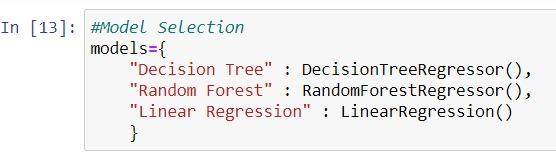
**FIGURE 3.3.11**

**Printing the y\_train result**

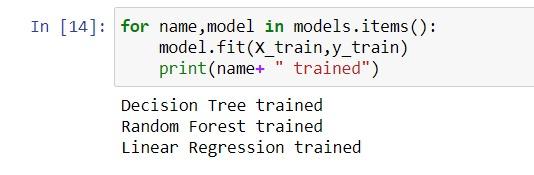
****

**FIGURE 3.3.12**

**Creating models into dictionary**

****

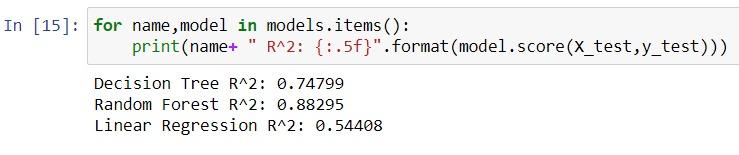
**FIGURE 3.3.13**

**Training the models **

**FIGURE 3.3.14**

Training the Decision tree, Random Forest and Linear Regression to get the best result from them.

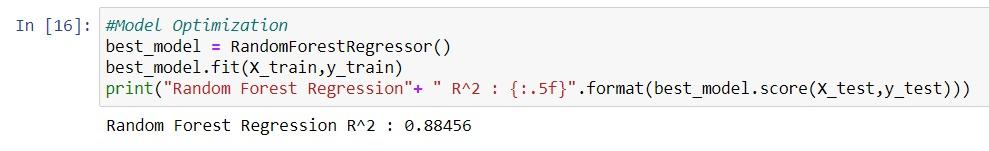
**Score of the model**



**FIGURE 3.3.15**

Getting the score of the models which we trained from the above figure, we get the Random Forest as the best model we done above.

**Optimizing the model**



**FIGURE 3.3.16**

By getting the best model as a random forest we train the same model again to get the best result.

**Correlation plot**

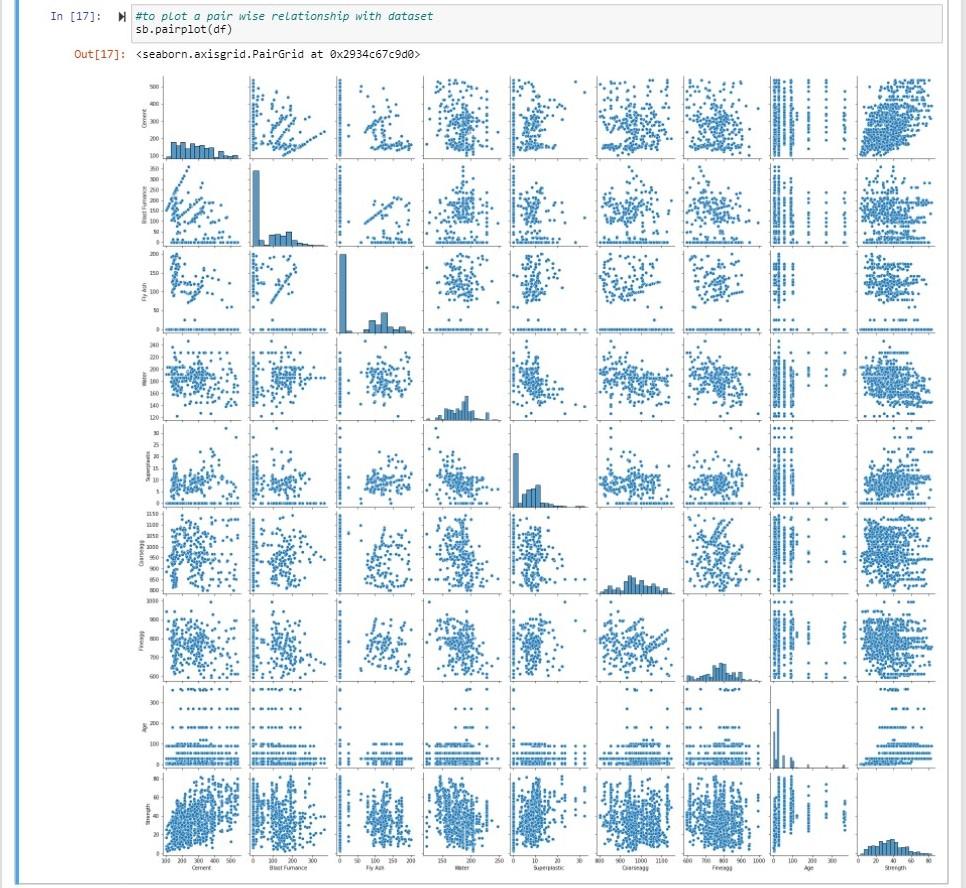
The correlation plot shows the correlation coefficient between variables. This plot contains the correlation matrix-like table.

Now we visualize the correlation between variables by plotting plot:



**FIGURE 3.3.17**

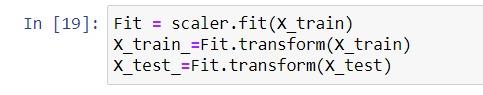
**Seaborn Diagram**



**FIGURE 3.3.18**

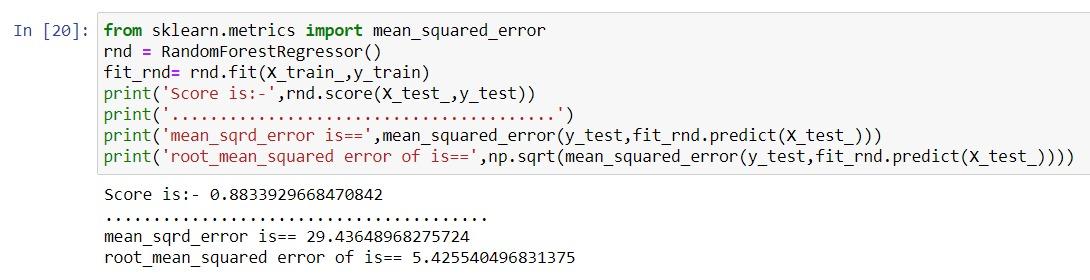
We use the seaborn library to plot the correlation plot between variables, here you see that there is one to one relationship between variables. Every variable is showing a relationship with another variable.

**Transforming the data**

****

**FIGURE 3.3.19**

**Random Forest Regression**

****

**FIGURE 3.3.20**

The accuracy score of Random Forest Regressor is highest among linear, lasso, and ridge regression, so we use the Random Forest Regressor model, Here the highest accuracy means it predicts the quality of concert by using training, which contains independent variables, and also it gives less error rate.

**3.4 Research Significance**

The significance and novelty of this research are  conducting experimental work for fly ash-based concrete (FAC) through ASTM standards, and  making FAC model using the Machine learning (ML) algorithms. Moreover, the focus of this study is based on the prediction and comparison of concrete (fly ash based) compressive strength via supervised machine learning approaches. The DT, ANN, and boosting ML approaches were employed to predict and compare the outcome with the actual results. The boosting approach was used for optimization by developing twenty sub-models to have a more accurate value of the coefficient of correlation (R2). These applications were used to compare the prediction performance of each technique. The importance of this research is to understand the role of input parameters and the accuracy level of the outcomes from the various ML algorithms. This study also presents the comparison approach of both ensemble and individual ML approaches with the results obtained for the experimental work. Each model performance was also evaluated from the statistical checks and k-fold cross-validation.

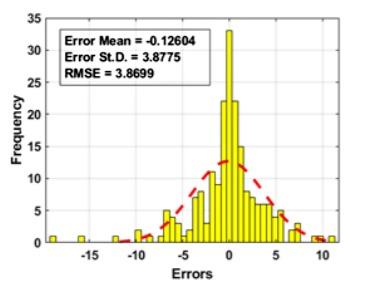
**4. RESULTS AND PERFORMANCE ANALYSIS**

**4.1 PERFORMANCE OF RANDOM FOREST**

This section demonstrates the performance of RF algorithms in predicting the compressive strength of manufactured sand concrete. In the RF algorithm, the number of trees is an important parameter that directly affects the prediction accuracy. An appropriate number of decision trees could ensure the quality in predicting the output of RF. Fig 4.1.1 shows the out-of-bag error in function of the number of trees in RF. As can be seen, a number of 50 trees could guarantee the low error values and the rate of convergence decreased when increasing the number of trees. It could be concluded that, in this study, the use of 500 grown trees was sufficient to achieve reliable prediction results

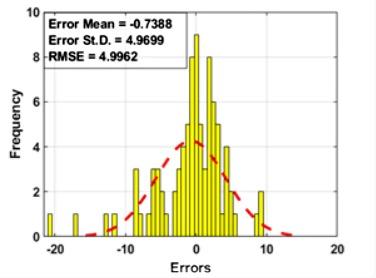
In this study, the possibility of using Random Forest algorithm to predict the compressive strength of manufactured sand concrete was investigated. In order to construct the database for the use of Random Forest algorithm, reliable experimental results were gathered from the available literature. The dataset was next divided into the training (70% of data) and testing datasets (30% of the remaining data). Several well-known statistical criteria were used to evaluate the performance of the Random Forest Algorithm.

In the following parts, the prediction results of RF are presented. Fig. 4.1.2 shows the histogram of error between the actual values of compressive strength of manufactured sand concrete versus those obtained from experiments. As can be seen, a high concentration of error around 0 was obtained, showing the efficiency of RF. Several higher values of error were found (i.e., -20, -15) but only 2 cases were observed. This clearly showed that, for the training part, the RF algorithm performed well in solving the problem.

****

**FIGURE 4.1.1: HISTOGRAM OF ERROR FOR TRAINING DATASET**

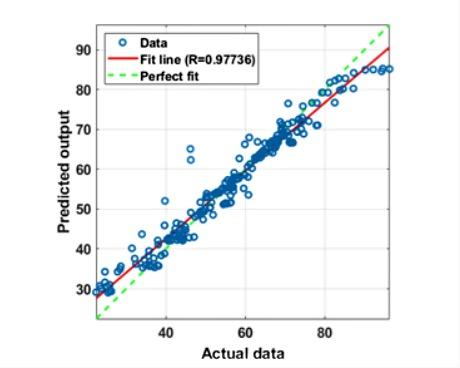
Next, the performance of the testing dataset is presented. It shows the histogram of error between the actual and predicted output. Again, only 4 values of error superior to 10 were observed. The remaining errors were highly concentrated between 0 and ±5 (MPa). The values of RMSE were 3.8699 and 4.9962 for the training and testing datasets, respectively. The values of mean error were computed as - 0.12604 and -0.7388 for the training and testing parts, respectively. Idem, values of standard deviation were 3.8775 and 4.9699 for the training part and testing one, respectively. It can be seen that the accuracy (i.e. error mean, standard deviation error and RMSE) of the training data was superior to the testing one, thus helpful in preventing selection as to avoid overfitting. Thus, the results were reliable and the RF model could be used for further investigation.



**FIGURE 4.1.2: HISTOGRAM OF ERROR FOR TESTING DATASET**

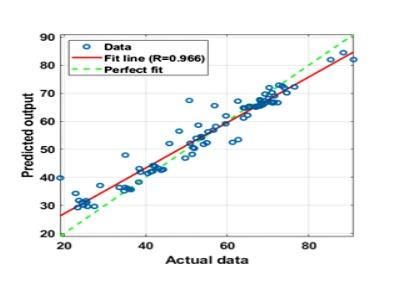
**4.2 FEATURE IMPORTANCE ANALYSIS**

Feature importance analysis is performed in this section. Basically, while constructing and developing the RF model, this algorithm could provide information related to the sensitivity of the features used to predict the output. The importance of features is computed as the sum of changes in the risk made by the splitting process then divided by the sum of the number of branches. In general, the out-of-bag is the criterion to deduce the importance of the feature. The higher the value of out-of-bag, the more important the input. Fig. 4.2.1 shows the sensitivity analysis of 11 input parameters in this study in predicting the compressive strength of manufactured sand concrete. It can be seen that the curing age and the water to binder ratio were the two most affecting features to the prediction of compressive strength. As the range of the curing age in this study was large, it was understandable to conclude that this input exhibited an important role in the compressive strength of manufactured concrete.



**FIGURE 4.2.1: Regression analysis for the training dataset.**

Overall, the sensitivity analysis using RF feature importance analysis is helpful in finding several high impact parameters to the prediction process. Based on the results, further prediction of manufactured concrete compressive strength might be improved by selecting the appropriate features.

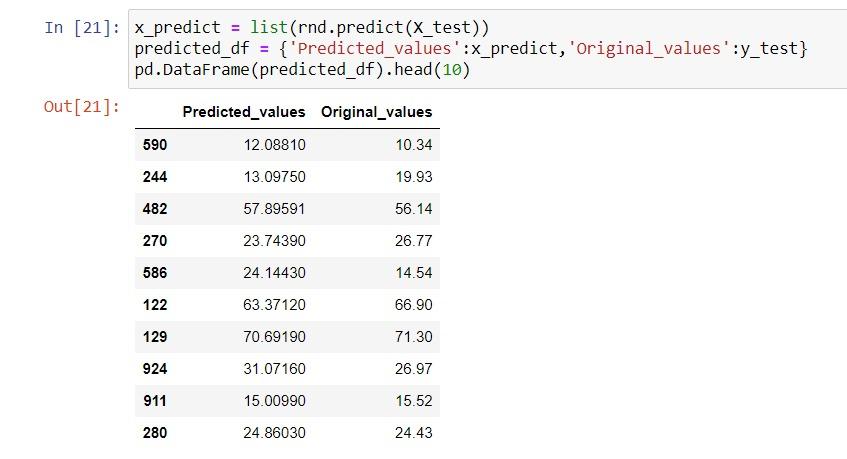
****

**FIGURE 4.2.2: Regression analysis for the testing dataset.**

**4.3 RESULT**

**Predicting Values Vs Original Values**

Now, we take a compression between the predicted values of the dependent variable **csMPa** and the original values of variable csMPa.



**FIGURE 4.3.1 : Result**

You can see that by applying the Random Forest Regressor model, the predicting values are quite similar to our original values.

**5. CONCLUSIONS**

In this study, the possibility of using Random Forest algorithm to predict the compressive strength of manufactured sand concrete was investigated. In order to construct the database for the use of Random Forest algorithm, reliable experimental results were gathered from the available literature. The dataset was next divided into the training (70% of data) and testing datasets (30% of the remaining data). Several well-known statistical criteria were used to evaluate the performance of the Random Forest algorithm.

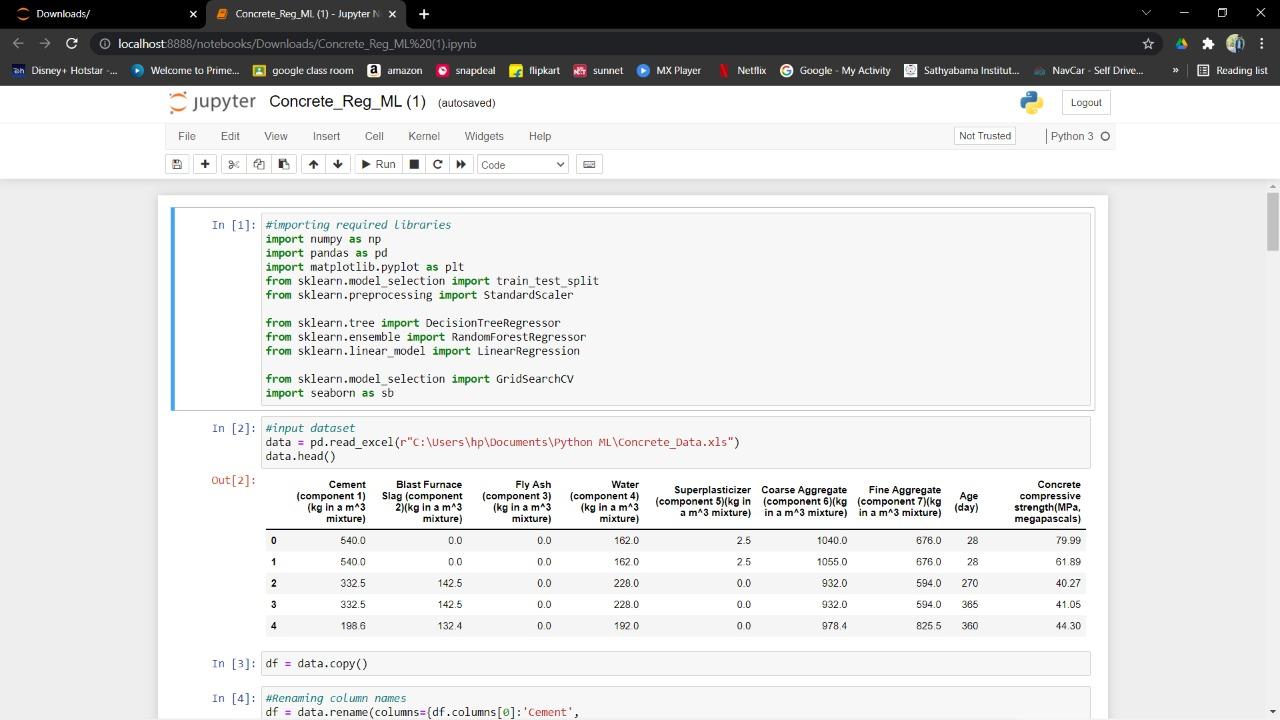
One of the most critical issues in the concrete structure manufacturing technological process is to obtain predictable properties of both fresh concrete mix and hardened concrete. Manufacturers of concrete mix are obliged to guarantee the appropriate properties of concrete delivered to the construction site. However, obtaining the appropriate concrete properties at all stages of the production process is difficult, especially when chemically complex additives and admixtures for concrete are present.

Predictive analytics is an area of knowledge dealing with predicting all kinds of phenomena, properties, and trends.

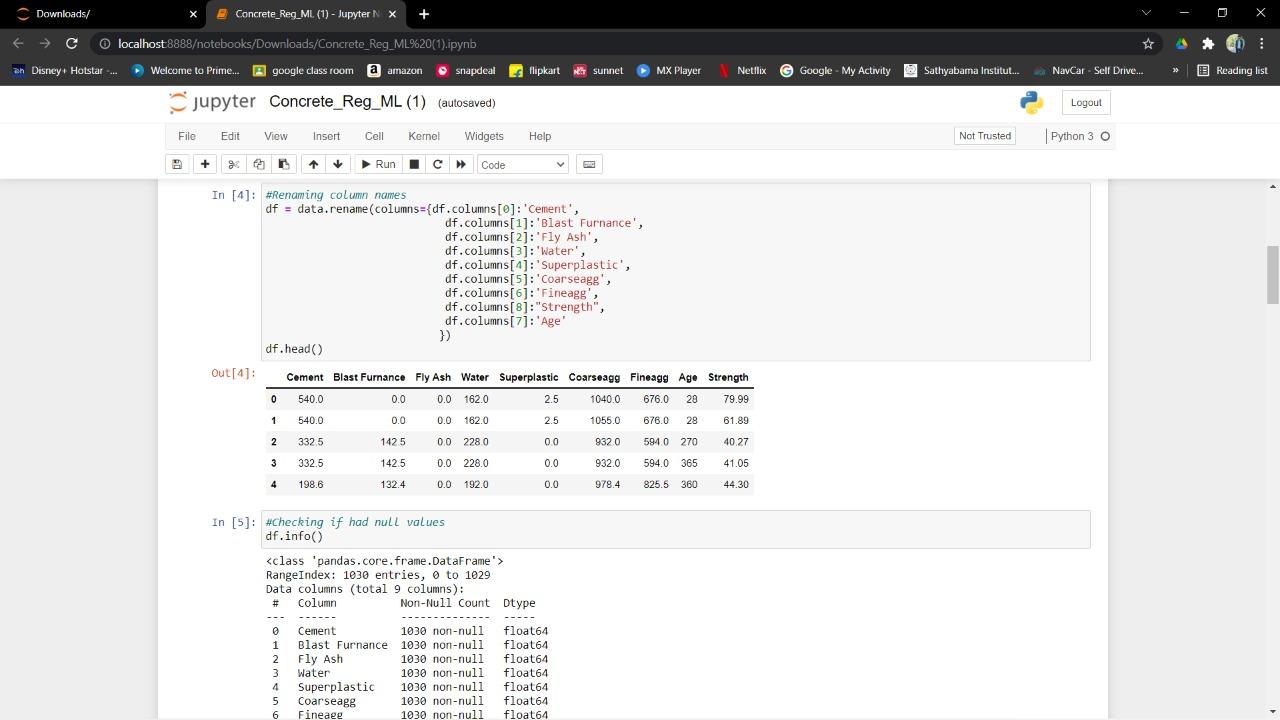
One of the most promising branches in predictive analytics is machine learning, which has acquired particular attention in recent years due to the significant development of technology. It could be used to significantly improve the process of concrete mix design.

**6.APPENDIX**

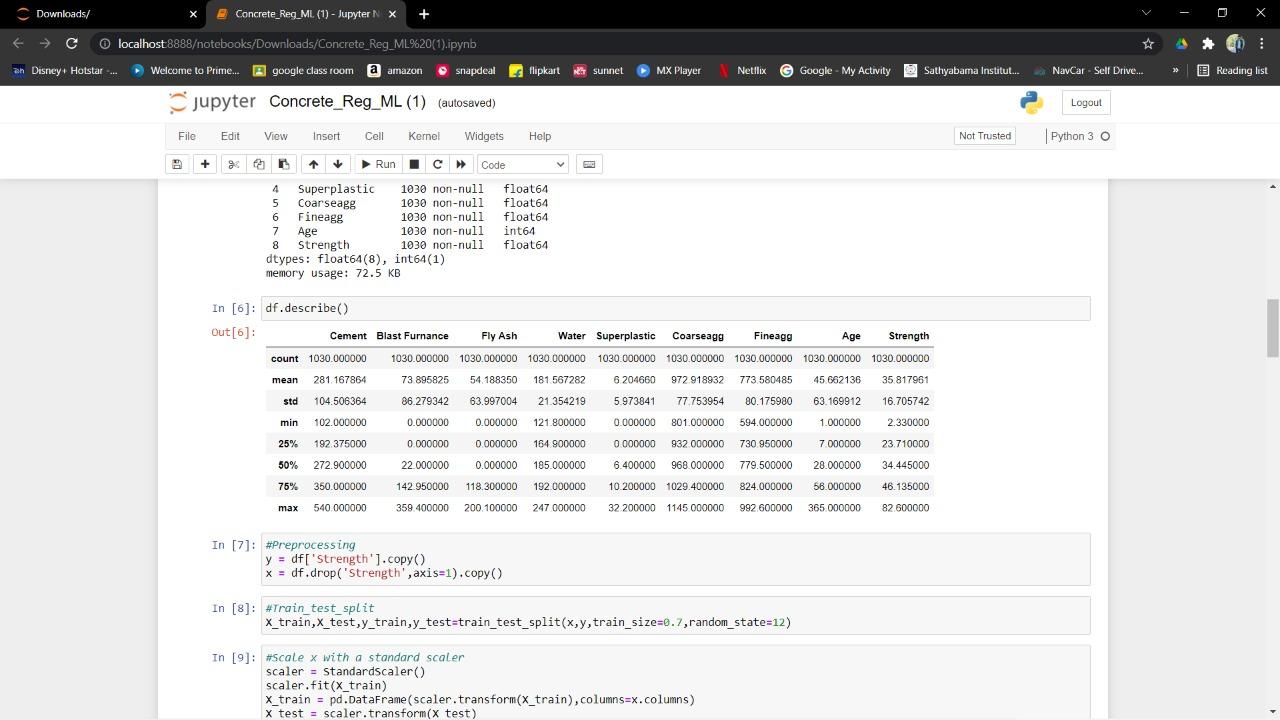
* 1. **Implemented Screenshots**

****

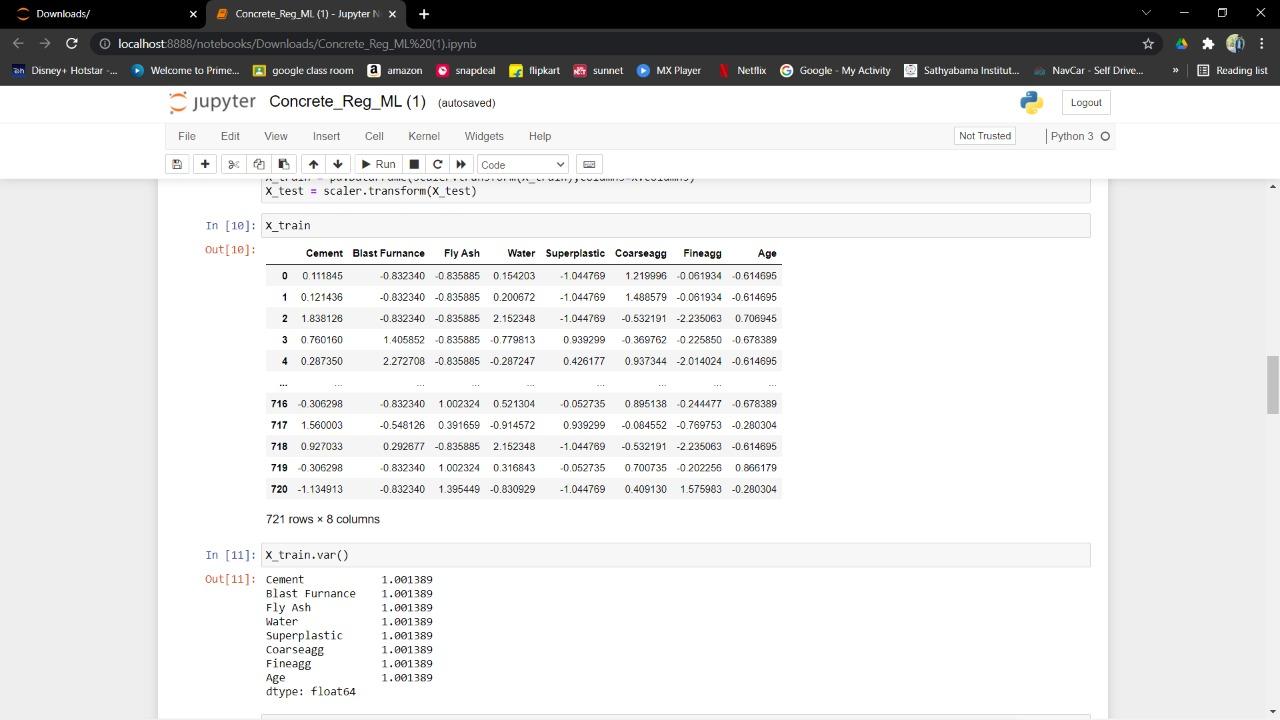
**FIGURE 6.1**

****

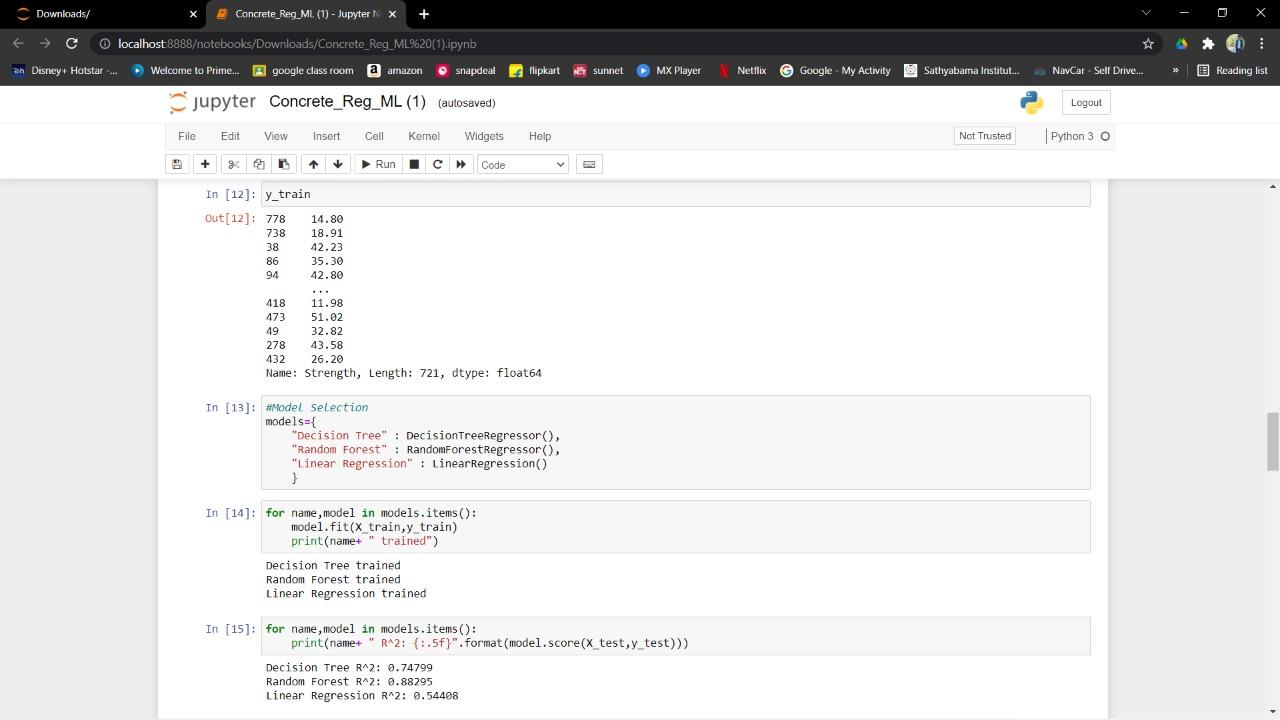
**FIGURE 6.2**

****

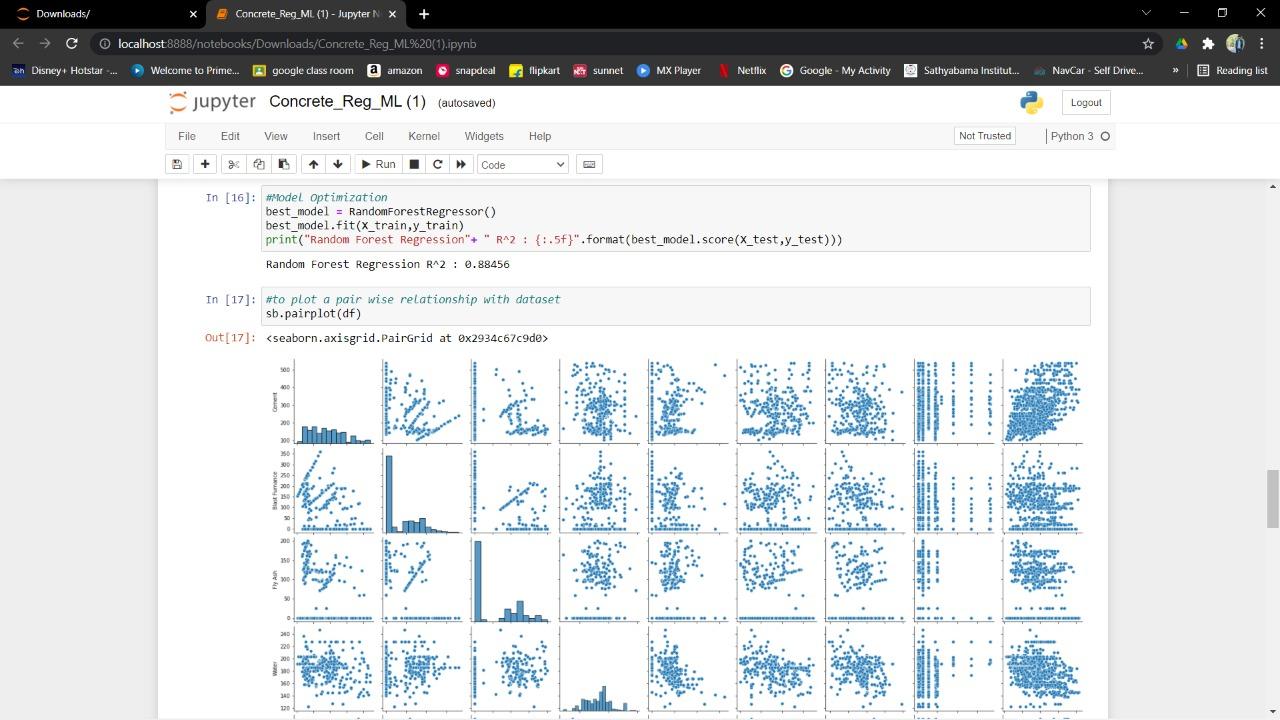
**FIGURE 6.3**

****

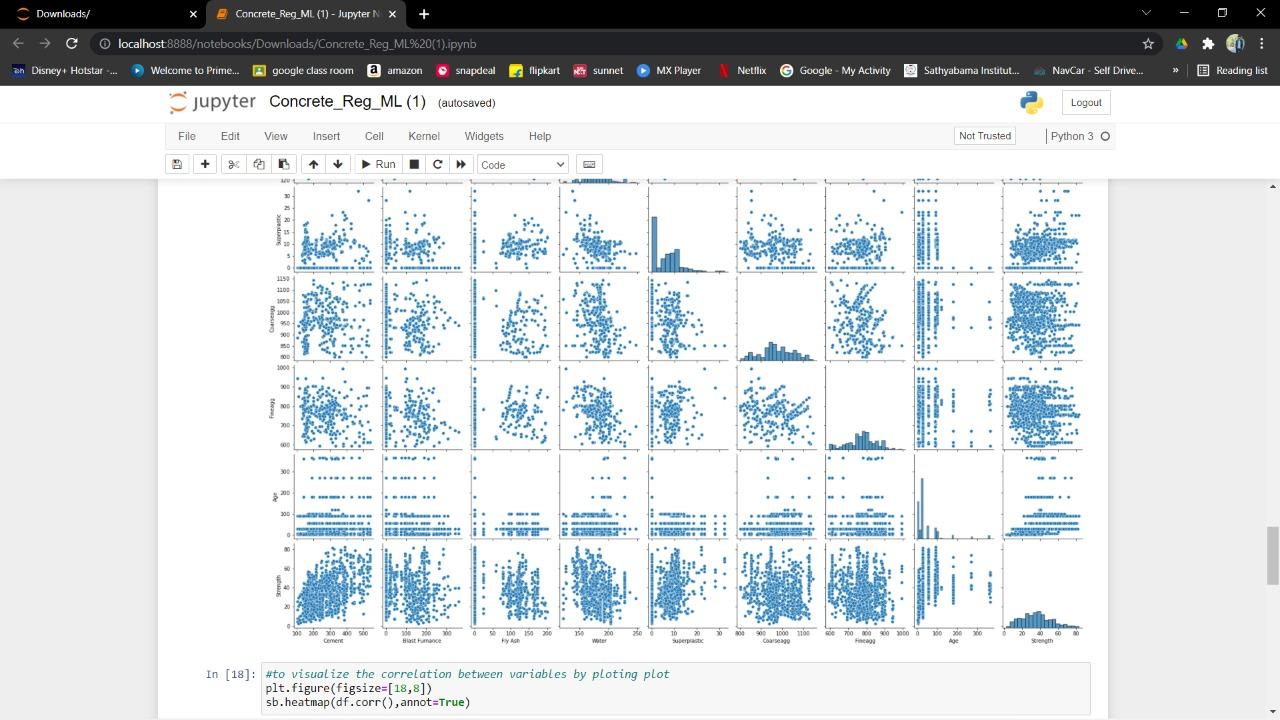
**FIGURE 6.4**

****

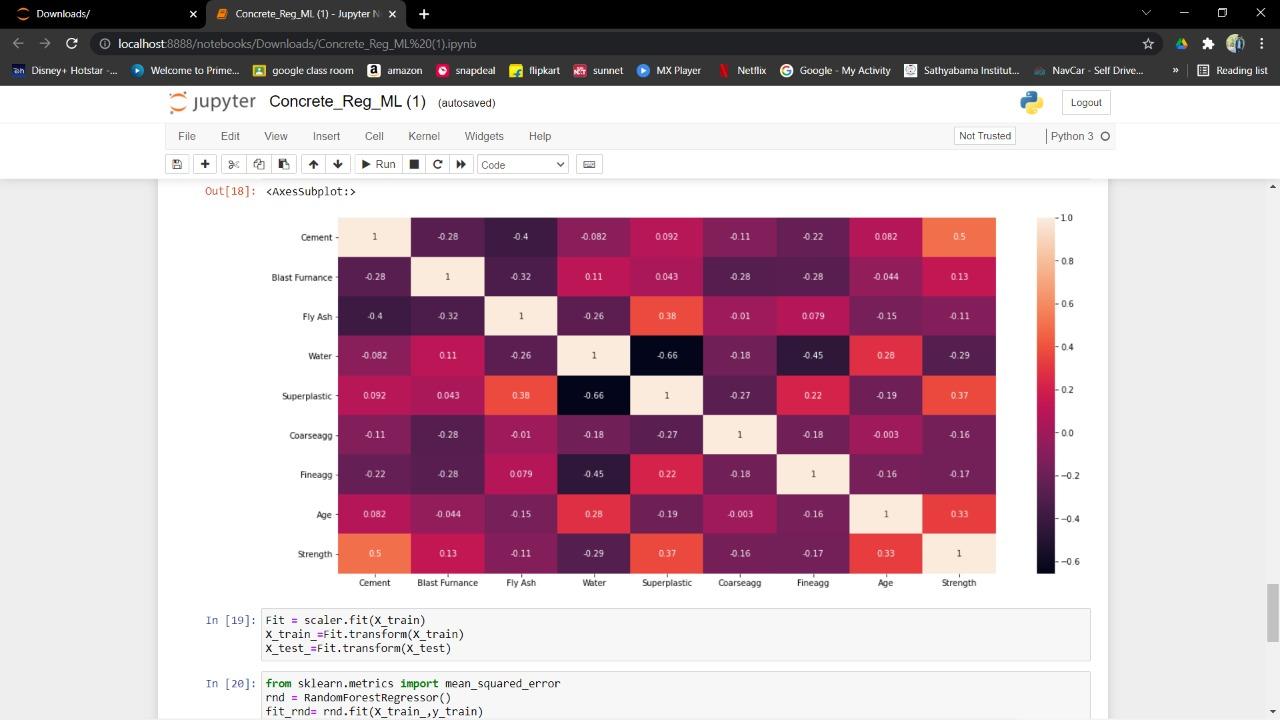
**FIGURE 6.5**

****

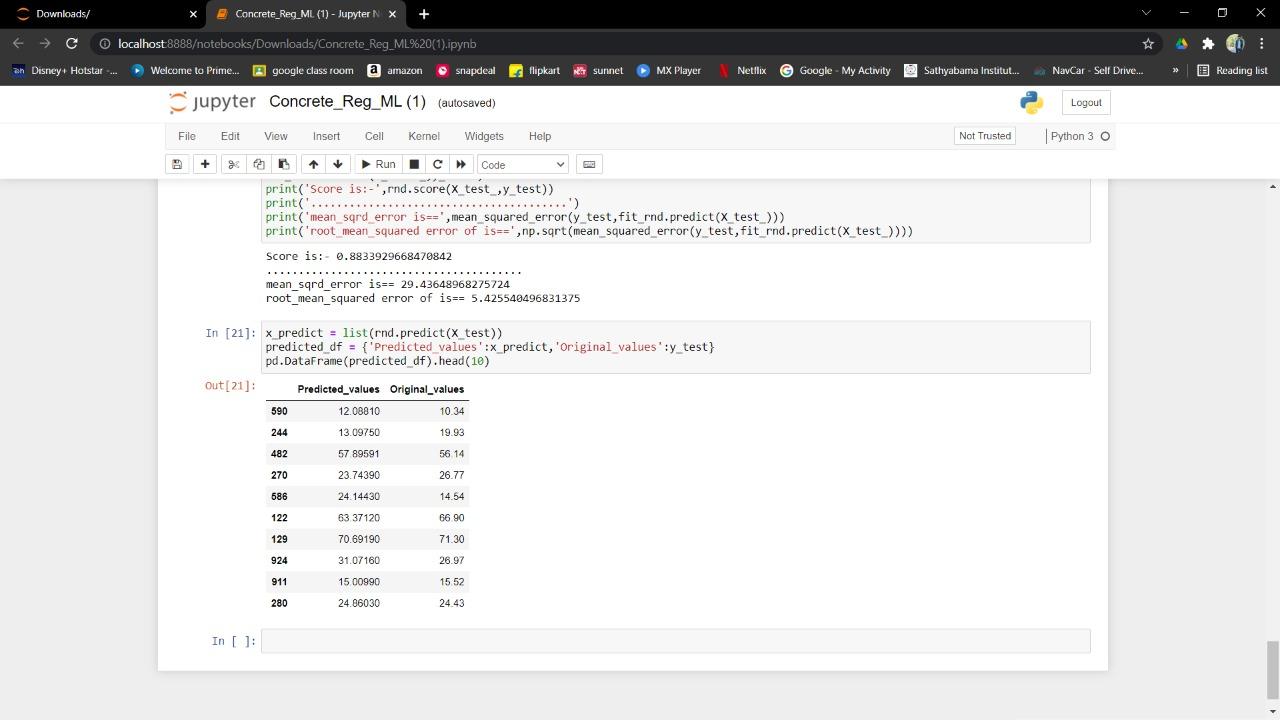
**FIGURE 6.7**

****

**FIGURE 6.8**

****

**FIGURE 6.9**



**FIGURE 6.10**

* 1. **Source Code**

#importing required libraries

import numpy as np

import pandas as pd

import matplotlib.pyplot as plt

from sklearn.model\_selection import train\_test\_split

from sklearn.preprocessing import StandardScaler

from sklearn.tree import DecisionTreeRegressor

from sklearn.ensemble import RandomForestRegressor

from sklearn.linear\_model import LinearRegression

from sklearn.model\_selection import GridSearchCV

import seaborn as sb

#input dataset

data = pd.read\_excel(r"C:\Users\hp\Documents\Python ML\Concrete\_Data.xls")

data.head()

df = data.copy()

df = data.rename(columns={df.columns[0]:'Cement',

df.columns[1]:'Blast Furnace',

df.columns[2]:'Fly Ash',

df.columns[3]:'Water',

df.columns[4]:'Superplastic',

df.columns[5]:'Coarse Agg',

df.columns[6]:'Fine Agg',

df.columns[8]:"Strength",

df.columns[7]:'Age’})

#Checking if had null values

df.info()

df.describe()

#Preprocessing

y = df['Strength'].copy()

x = df.drop('Strength',axis=1).copy()

#Train\_test\_split

X\_train,X\_test,y\_train,y\_test=train\_test\_split(x,y,train\_size=0.7,random\_state=12)

#Scale x with a standard scaler scaler = StandardScaler()

scaler.fit(X\_train)

X\_train = pd.DataFrame(scaler.transform(X\_train),columns=x.columns)

X\_test = scaler.transform(X\_test)

X\_train

X\_train.var()

y\_train

#Model Selection models={ "Decision Tree" : DecisionTreeRegressor(),

"Random Forest" : RandomForestRegressor(), "Linear Regression" : LinearRegression() }

for name,model in models.items():

model.fit(X\_train,y\_train) print(name+ " trained")

for name,model in models.items():

print(name+ " R^2: {:.5f}".format(model.score(X\_test,y\_test)))

#Model Optimization best\_model = RandomForestRegressor() best\_model.fit(X\_train,y\_train) print("Random Forest Regression"+ " R^2 : {:.5f}".format(best\_model.score(X\_test,y\_test)))

#to plot a pairwise relationship with dataset

sb.pairplot(df)

#to visualize the correlation between variables by plotting plot plt.figure(figsize=[18,8]) sb.heatmap(df.corr(),annot=True)

Fit = scaler.fit(X\_train) X\_train\_=Fit.transform(X\_train) X\_test\_=Fit.transform(X\_test)

from sklearn.metrics import mean\_squared\_error

rnd = RandomForestRegressor() fit\_rnd= rnd.fit(X\_train\_,y\_train)

print('Score is:-',rnd.score(X\_test\_,y\_test))

print('........................................')

print('mean\_sqrd\_error is==',mean\_squared\_error(y\_test,fit\_rnd.predict(X\_test\_)))

print('root\_mean\_squared error of is==',np.sqrt(mean\_squared\_error(y\_test,fit\_rnd.predict(X\_test\_))))

x\_predict = list(rnd.predict(X\_test))

predicted\_df = {'Predicted\_values':x\_predict,'Original\_values':y\_test} pd.DataFrame(predicted\_df).head(10)

* 1. **References**

[**https://www.youtube.com/watch?v=S3Wn-RK7NnY**](https://www.youtube.com/watch?v=S3Wn-RK7NnY)

[**https://www.analyticsvidhya.com/blog/2021/04/concrete-strength-prediction-using-machine-learning-with-python-code/**](https://www.analyticsvidhya.com/blog/2021/04/concrete-strength-prediction-using-machine-learning-with-python-code/)