**Predicting Compressive Strength**

**Of Concrete**

**Introduction**

The development of cement compressive strength as a consequence of clinker characteristics and of physical and chemical properties of cement constitutes one of the most critical issues in the field of product design and its quality control. The traditional 28 days strength of mortar is considered as a sufficient indicator of overall quality of the product. All the norms and especially the norm applied in Europe, EN 197-1, apply specifications on this cement characteristic, concerning low and high strength limits. The manufacturer is obliged to respect these limits. The stability of cement quality is mainly characterized by the variance of 28 days strength around a predefined target.

The main purpose of the present study is to develop various models predicting 28 days compressive strength. The resulting tool is constantly applied in Halyps Cement Plant in the long-term. This article is structured as follows: a brief description of the cement production process is provided in the first section. The basic chemical moduli characterizing clinker quality are also referred to. The second section contains a summary of existing models relating cement strength with its chemical and physical characteristics. The proposed predicting models are developed in the next two sections. The implementation of the models is analyzed in the last section.

1

**Literature Survey**

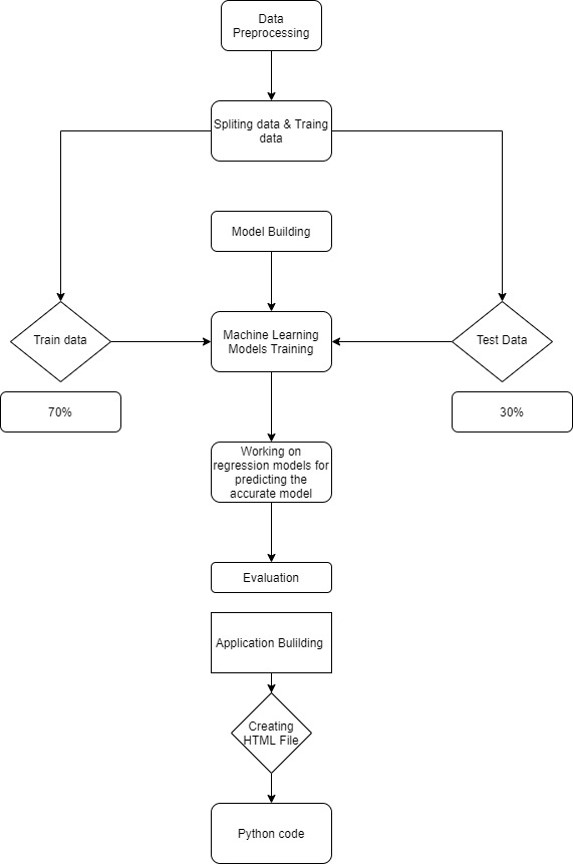
**PROBLEM STATEMENT-EXISTING SOCIETAL ISSUE:**

* In earlier days, the concrete strength is measure through other traditional methods like using drill holes, weight spring, or using sensors. But that requires a significant destruction of test sample and thereby increasing the cost.
* And the accuracy was also hardly 70%.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.
* Generally, the one-factor-at-a-time method is used in experimental designs to determine the concrete properties. The major disadvantage of this approach is that it does not consider the interaction between the factors (interaction terms). The higher the number of the controlled and uncontrolled effect variables that influence the concrete properties, the lesser the predicted accuracy*.*

**PROPOSED SOLUTION:**

The focus of this project is the application of machine learning process,Artificial nueral networks and their suitability to model concrete compressive strength compared with early models obtained from the literature and compared with some conventional approaches and also a recoomendation system is developed by applying various ML methods,Deep nueral network methods to predict the concrete strength from its components accurately and then looking for the optimal combination of components which increases the strength.

2

**ARCHITECTURE-MACHINE LEARNING WORKFLOW:**

3

**DATA SET COLLECTION:**

Data set is obtained from the website called Kaggle.It can be accessed using the link Kaggle.com

**DATA DESCRIPTION:**

* Number of Instances – 1030
* Number of Attributes – 9
* Attribute breakdown – 8 quantitative inputs,1 quantitative output

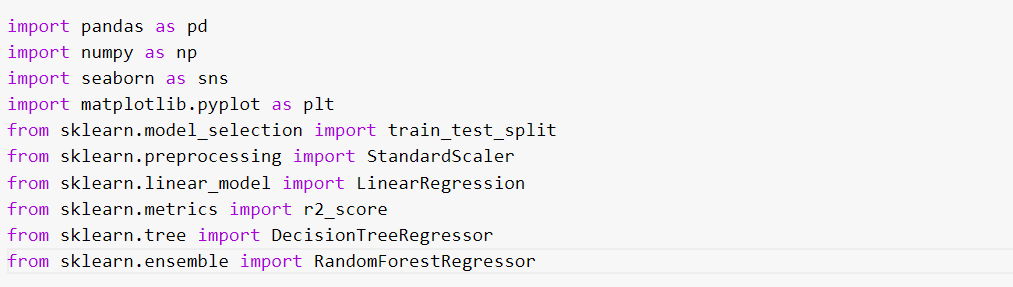
**ATTRIBUTE INFORMATION:**

* Inputs
* Cement (kg in a m^3 mixture)
* Blast Furnace Slag (kg in a m^3 mixture)
* Fly Ash (kg in a m^3 mixture)
* Water (kg in a m^3 mixture)
* Superplasticizer (kg in a m^3 mixture)
* Coarse Aggregate (kg in a m^3 mixture)
* Age (days)
* Output
* Concrete Compressive Strength (MPa)

4

**Data Pre-Processing**

**IMPORTING LIBRARIES:**

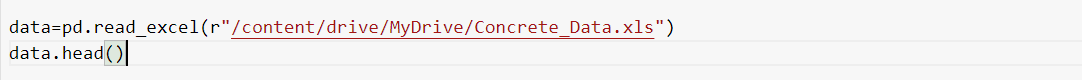


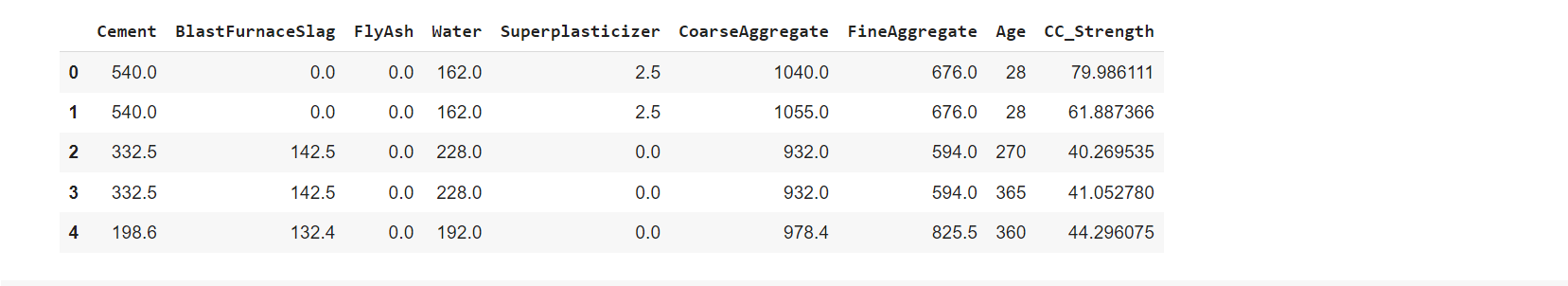
* **pandas:** pandas is a software library written for the Python programming language for data manipulation and analysis.
* **numpy:** NumPy is a python library used for working with arrays. It also has functions for working in domain of linear algebra, fourier transform, and matrices.
* **seaborn:** It is used for data visualization library used for plotting graph which will help us for understanding the data.
* **matplotlib:** Itisalso used for plotting graph.
* **train\_test\_split:** used for splitting data arrays into training data and for testing data.
* **StandardScaler:** standardizes a feature by subtracting the mean and then scaling to unit variance.
* **LinearRegression:** used to perform linear and polynomial regression and make predictions accordingly.
* **r2\_score:** Coefficient of Determination or R² is another metric used for evaluating the performance of a regression model. The metric helps us to compare our current model with a constant baseline and tells us how much our model is better.
* **DecisionTreeRegressor:** Decision tree builds regression or classification models in the form of a tree structure. It breaks down a dataset into smaller and smaller subsets while at the same time an associated decision tree is incrementally developed. The final result is a tree with decision nodes and leaf nodes.

5

* **RandomForestRegressor:** It is a supervised learning algorithm that uses **ensemble learning** method for regression. Ensemble learning method is a technique that combines predictions from multiple machine learning algorithms to make a more accurate prediction than a single model.

**IMPORTING DATA SET:**

****



* You might have your data in .csv files, .excel files or .tsv files or something else. But the goal is the same in all cases. If you want to analyse that data using pandas, the first step will be to read it into a data structure that’s compatible with pandas.
* Let’s load a .csv data file into pandas. There is a function for it, called read\_csv().We will need to locate the directory of the CSV file at first (it’s more efficient to keep the dataset in the same directory as your program).
* Path names on Windows tend to have backslashes in them. But we want them to mean actual backslashes, not special characters.

6

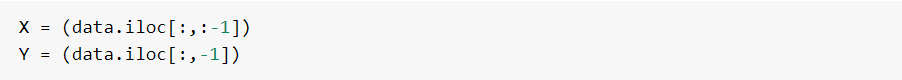
**FEATURE CORRELATION:**

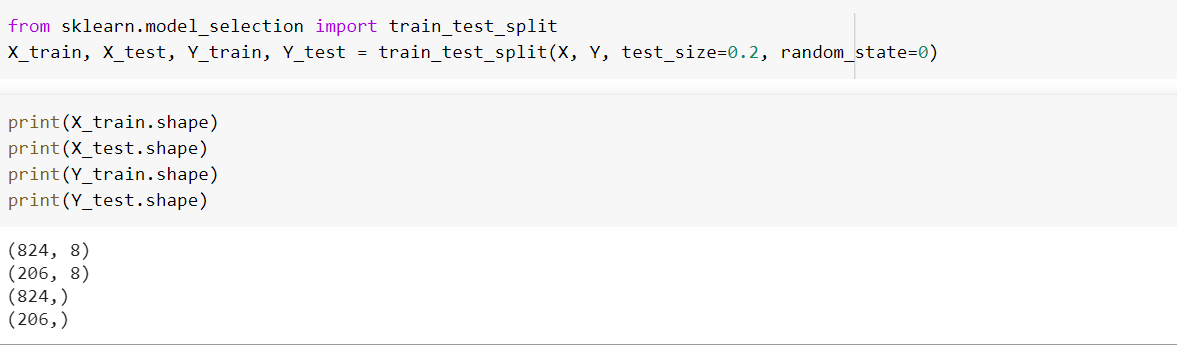
****

7

**SPLITTING DATA INTO TRAIN AND TEST:**

* When you are working on a model and you want to train it, you obviously have a dataset. But after training, we have to test the model on some test dataset. For this, you will a dataset which is different from the training set you used earlier. But it might not always be possible to have so much data during the development phase. In such cases, the solution is to split the dataset into two sets, one for training and the other for testing.
* Now split our dataset into train set and test using train\_test\_split class from scikit learn library.

****

****

8

**MODEL BUILDING:**

* Training and testing the model:
* There are several Machine learning algorithms to be used depending on the data you are going to process such as images, sound, text, and numerical values. The algorithms that you can choose according to the objective that you might have it may be Classification algorithms or Regression algorithms.

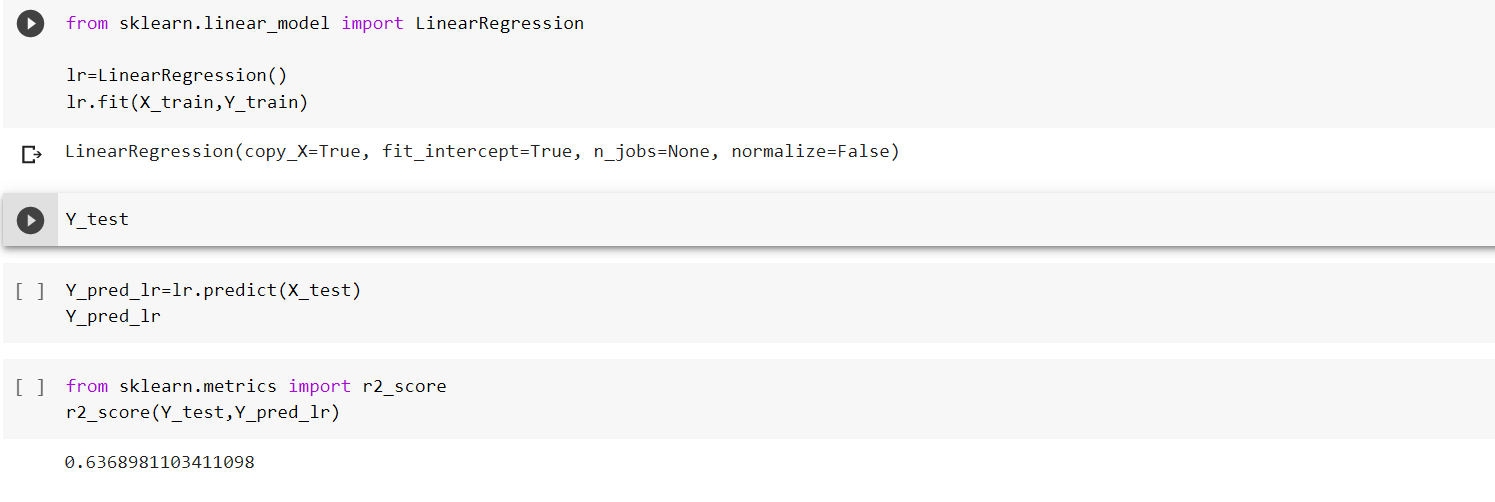
Example: 1. Linear Regression.

2. Random Forest Regression / Classification.

3. Decision Tree Regression / Classification.

* You will need to train the datasets to run smoothly and see an incremental improvement in the prediction rate.
* Now we apply Linear Regression algorithm on our data set.

**Linear Regression** is a machine learning algorithm based on **supervised learning**. Linear regression performs the task to predict a dependent variable value (y) based on a given independent variable (x). So, this regression technique finds out a linear relationship between x (input) and y(output). Hence, the name is Linear Regression.

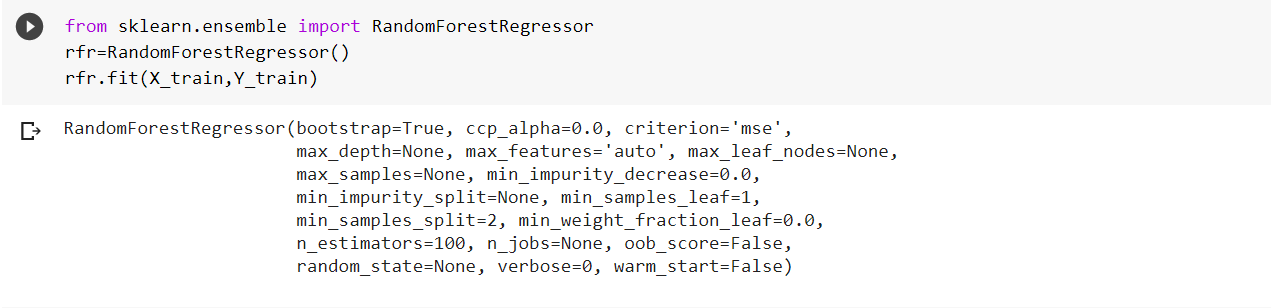


9

**DECISION TREE REGRESSION:**

**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.

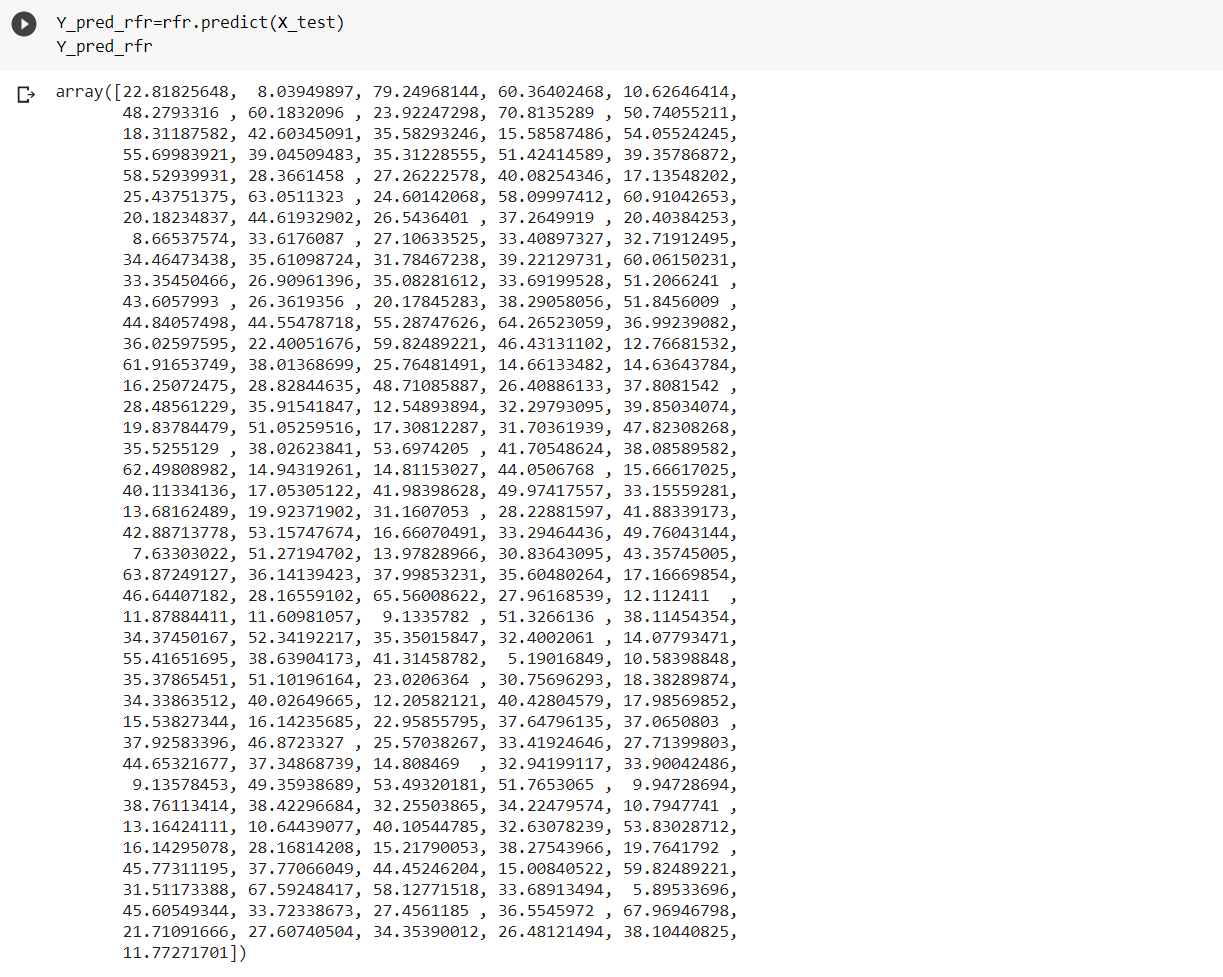
**RANDOM FOREST REGRESSOR:**

**Random Forest Regression** is a supervised learning algorithm that uses **ensemble learning** method for regression. Ensemble learning method is a technique that combines predictions from multiple machine learning algorithms to make a more accurate prediction than a single model.

10

**PREDICT THE VALUES:**

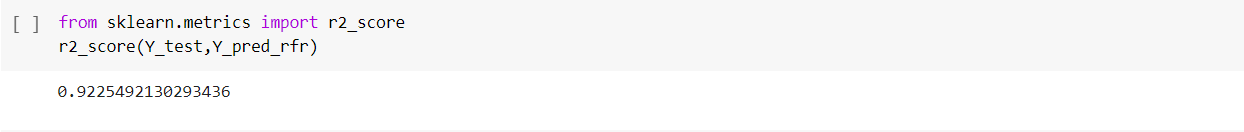
* Once the model is trained, it’s ready to make predictions. We can use the predict method on the model and pass x\_test as a parameter to get the output as pred.
* Notice that the prediction output is an array of real numbers corresponding to the input array.



11

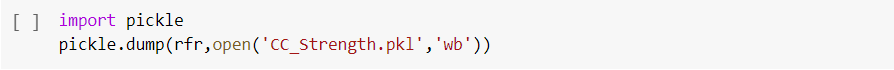
**EVALUATION:**

* Finally, we need to check to see how well our model is performing on the test data. There are many evaluation techniques are there. For this, we evaluate r2\_score produced by the model.

****

**SAVING THE MODEL:**

Model is saved so it can be used in future and no need to train it again.



**APPLLICATION BUILDING:**

Creating a HTML File, flask application.

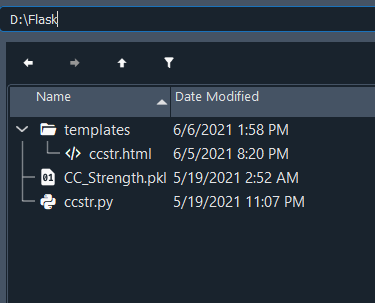
* Build python code
* Importing Libraries

* Routing to the html Page.
* Showcasing prediction on UI
* Run The app in local browser.

12

**PROJECT STRUCTURE:**

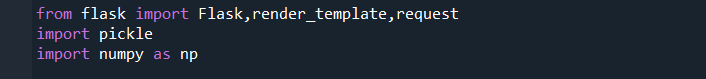
Create a project folder that contains files as shown below



We are building a Flask Application that needs HTML pages stored in the templates folder

* Templates folder contains ccstr.html

**Task 1 :- Importing Libraries**

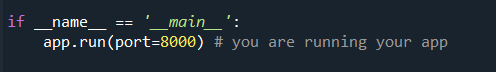
****

13

**Task 2 :- Routing to the html Page**

****

**Task 3 :- Main Function**

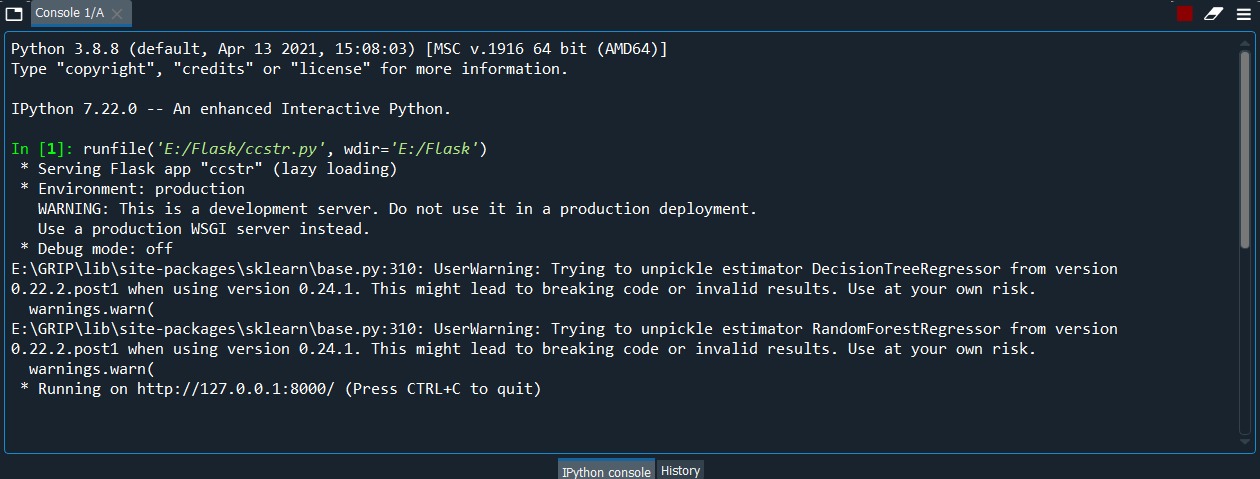
****

14

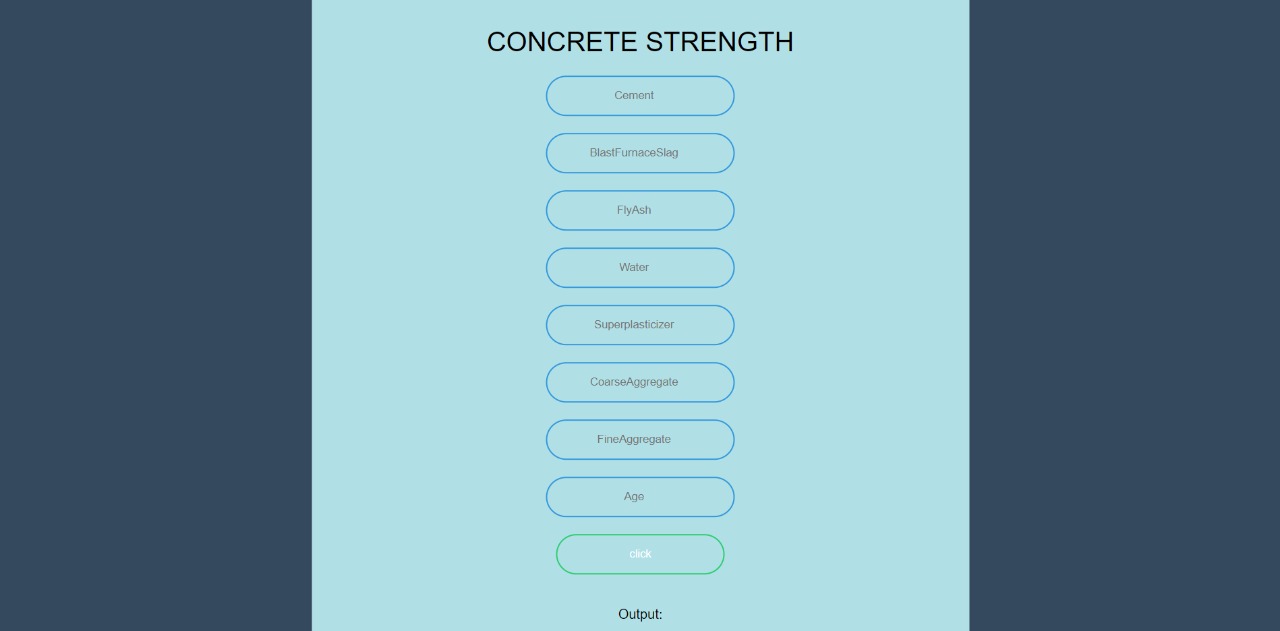
**ACTIVITY:**

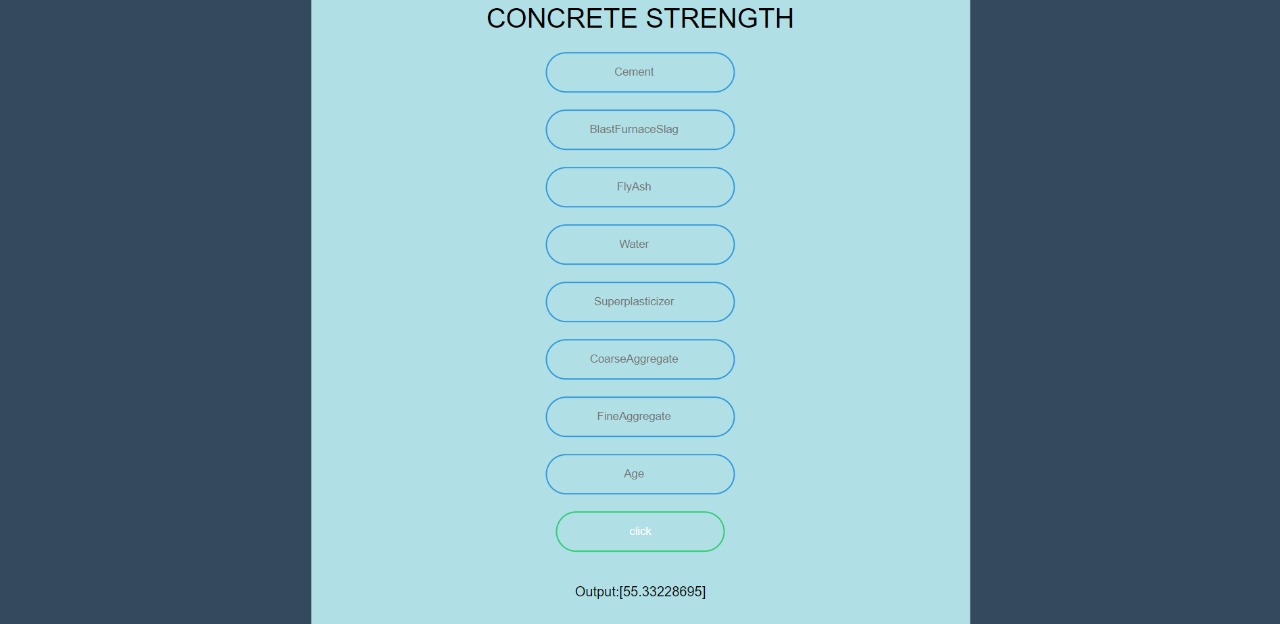
Open the anaconda prompt from the start menu.

* Navigate to the folder where your app.py resides.
* Now type “python app.py” command.
* It will show the local host where your app is running on <http://127.0.0.1:8000/>
* Copy that local host URL and open that URL in the browser. It does navigate me to where you can view your web page.
* Enter the values, click on the predict button and see the result/prediction on the web page.



15

**OUTPUT SCREEN:**



16

**FINDINGS AND SUGGESTIONS:**

Through Exploratory Data Analysis

* The Accuracy for Linear Regression is 63.6%
* The Accuracy for Decision Tree Regressor is 78.7%
* The Accuracy for Random Forest Regression is 92.2%

**CONCLUSION:**

Analysed the Compressive Strength and used Machine Learning to Predict the Compressive Strength of Concrete. We have used Linear Regression and its variations, Decision Trees and Random Forests to make predictions and compared their performance. Random Forest Regressor has the lowest RMSE and is a good choice for this problem. Also, we can further improve the performance of the algorithm by tuning the hyperparameters by performing a grid search or random search.

NN approaches combine the complexity of many statistical techniques with machine learning techniques and attributed as a black-box which allows NN to be applied in all engineering disciplines. It comes out as the best possible model for the prediction of compressive strength of concrete. It has predicted with high accuracy for all the curing ages, that is, 28, 56, and 91 days.

**REFERENCE:**

1.www.kaggle.com

2.www.quora.com

3.www.wikkipedia.com

M.Sahithi

*sahithisk9*@gmail.com