PREDICTING COMPRESSIVE STRENGTH OF CONCRETE USING IBM WATSON MACHINE LEARNING

A PROJECT REPORT

Submitted in partial fulfilment for the data science externship

with

SmartInternz

by

NAME	REGITRATION NO.
Rishika	18BEI0119
Simran Rajpal	18BEI0123
Kshitij Dubey	18BEI0128
Pragya Jain	18BEE0253

Under the guidance of

SmartInternz Team

VIT®
Vellore Institute of Technology
(Deemed to be University under section 3 of UGC Act, 1956)

INTRODUCTION

Overview

Considering that compressive strength (CS) is an important mechanical property parameter in many design codes, in order to ensure structural safety, concrete CS needs to be tested before application. However, conducting CS tests with multiple influencing variables is costly and time consuming. To address this issue, a machine learning-based modelling framework is put forward in this work to evaluate the concrete CS under complex conditions. The influential factors of this process are systematically categorized into five aspects: man, machine, material, method and environment.

The CS of concrete is obtained by testing specifically prepared and cured cubic or cylindrical specimens using a compression test instrument, which is cumbersome, time-consuming and costly in the entire experimental process. To improve this situation, empirical regression methods and numerical simulation have been developed to predict concrete CS based on the design recipe. The concrete production process is affected by several factors, which have strong nonlinear relationships with the product CS and are strongly interrelated. With the rapid development of machine learning, there is a trend to employ data-driven techniques for concrete CS prediction. Compared with the conventional regression methods, machine learning-based approaches adopt suitable algorithms to automatically “learn” from the process data, “distinguish” important influential factors from the interfering factors, approximate the intricate process mechanism with deterministic mathematical forms, and perform prediction with high accuracy over a specified confidence interval. To date, various machine learning algorithms have been applied to studying the correlations between the concrete recipe and the product CS.

Researchers have also proved that besides the type and basic properties of raw materials, environmental factors, such as temperature, and relative humidity, also significantly determine the concrete CS. To exclude the influence from these

environmental factors. Additionally, the recipe and environmental variables may not be sufficient in covering all the possible factors influencing concrete CS. It has been proved in many areas that to comprehensively evaluate the quality of an engineering product, the influences from man, material, machine, method and environment (shortened to 4M1E), should be considered. Each of these factors further represents the aggregation of various detailed influential factors.

Purpose

The combined effect of the above factors (4MIE) poses new challenges for the accurate prediction of concrete CS. Since the influence from the “material” aspect is already inherently complicated, when the joint influences from the other four aspects are included, the problem dimension becomes high, and it will be more difficult to manually identify the most relevant factors from the interfering ones.

Construction of both traditional regression models and machine learning models to predict the 28-day CS of no-slump concrete, based on the concrete ingredients.

Computer-aided feature selection techniques provide an alternative solution to this situation. Common feature selection algorithms can be categorized into two classes based on the search strategies—the filter methods and the wrapper methods. Applied principal component analysis (PCA) to reduce the noise in the input space and consequently improve the model predictive performance.

Principal component analysis is a typical filter method that performs feature selection based on the statistical performance of the original dataset and is independent of the subsequent learning algorithm. Different from the filter methods, the wrapper methods tightly couple the subsequent prediction algorithm with the feature selection process. In other words, the feature selection process is optimized based on the feedback from the subsequent algorithm performance. Commonly used wrapper methods include genetic algorithms (GAs) and particle swarm optimization (PSO).

LITERATURE SURVEY

Existing Problem

In the present study, a regression model was investigated as a performance prediction model for predicting the concrete compressive strength. Moreover, the effects of the changes of the coefficients of regression model of the performance curve were also examined. For this purpose, multiple regression analysis was carried out for predicting the compressive strength of concrete using four variables, namely, water-cementations ratio, fine aggregate cementitious ratio, coarse aggregate-cementitious ratio and cementitious content. Regression models were developed for concrete with medium and high workability at different curing ages (28, 56 and 91days). For models developed for compressive strength prediction at 56 days and 91 days, the compressive strengths at lower ages were also considered as a parameter.

For the experimental data obtained, the data sets on the dependent variable were normally distributed for each of the possible combinations of the level of the X variables, it was deemed suitable to use regression models for prediction of concrete strengths. Bayrak and Akgül predicted the lifetime performance and remaining service life of a bridge system using regression models, so that the best maintenance and repair strategies which kept the system safe can be obtained. Many attempts had earlier been made to obtain a suitable mathematical model that was capable of predicting the strength of concrete at various ages with good accuracy.

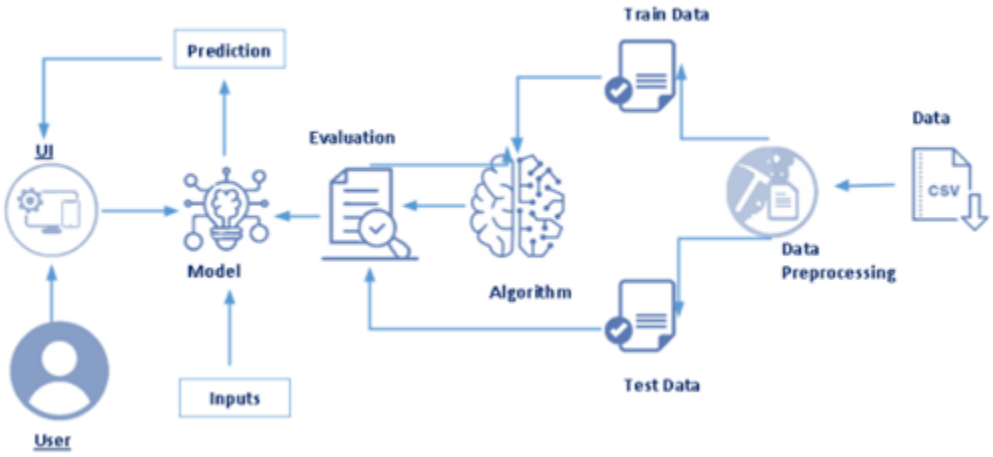
Proposed Solution

A mathematical model for prediction of compressive strength of concrete will be performed using statistical analysis for the concrete data obtained from experimental work done in this study. The multilinear regression yielded excellent correlation coefficient for prediction of compressive strength at different ages (3, 7, 14, 28, 91 days). The coefficient of correlation was 99.99% for each strength (at each age). Also, the model gives high correlation for strength prediction of concrete with different types of curing. Given the characteristics of the concrete

production process, multi linear regression is considered a very suitable modelling algorithm, due to its versatile merits, such as its good tolerance for outliers and noises, its ability to avoid overfitting, and its ability to deal with multicollinearity.

THEORITICAL ANALYSIS

Block Diagram



Hardware/ Software designing

https://github.com/rishika28/Applied-Data-Science/blob/main/Predicting%20Compressive%20Strength%20of%20Concrete%20(1).ipynb

EXPERIMENTAL INVESTIGATIONS

DATA PREPROCESSING

Importing Libraries

In [1]:
```
import pandas as pd #used for data manipulation
import numpy as np #used for numerical analysis
from collections import Counter as c # return counts
import seaborn as sns #used for data Visualization
import matplotlib.pyplot as plt
from sklearn.model_selection import train_test_split #splits data in random train and test array
from sklearn.metrics import accuracy_score,mean_squared_error,mean_absolute_error#model performance
import pickle #Python object hierarchy is converted into a byte stream,
from sklearn.linear_model import LinearRegression #Regresssion ML algorithm
```

Loading Dataset

In [2]:
```
data=pd.read_csv(r"D:\concrete.csv") #loading the csv dataset
```

In [4]:
```
data.head() #returns first 5 rows of data
```

Out[4]:

	Cement (component 1) (kg in a m^3 mixture)	Blast Furnace Slag (component 2)(kg in a m^3 mixture)	Fly Ash (component 3) (kg in a m^3 mixture)	Water (component 4) (kg in a m^3 mixture)	Superplasticizer (component 5)(kg in a m^3 mixture)	Coarse Aggregate (component 6) (kg in a m^3 mixture)	Fine Aggregate (component 7) (kg in a m^3 mixture)	Age (day)	Concrete compressive strength(MPa, megapascals)
0	540.0	0.0	0.0	162.0	2.5	1040.0	676.0	28	79.99
1	540.0	0.0	0.0	162.0	2.5	1055.0	676.0	28	61.89
2	332.5	142.5	0.0	228.0	0.0	932.0	594.0	270	40.27
3	332.5	142.5	0.0	228.0	0.0	932.0	594.0	365	41.05
4	198.6	132.4	0.0	192.0	0.0	978.4	825.5	360	44.30

Seeing Target, Nominal and Numerical Columns Count

In [14]:
```
np.unique(data.dtypes,return_counts=True)
```

Out[14]:
```
(array([dtype('int64'), dtype('float64')], dtype=object),
 array([1, 8], dtype=int64))
```

In [15]:
```
data.dtypes[data.dtypes!='O'].index.values
```

Out[15]:
```
array(['Cement', 'Blast Furnace Slag', 'Fly Ash', 'Water',
       'Superplasticizer', 'Coarse Aggregate', 'Fine Aggregate', 'Age',
       'Concrete compressive strength'], dtype=object)
```

Null Values

In [16]:
```
data.isnull().any() #it will return true if any columns is having null values
```

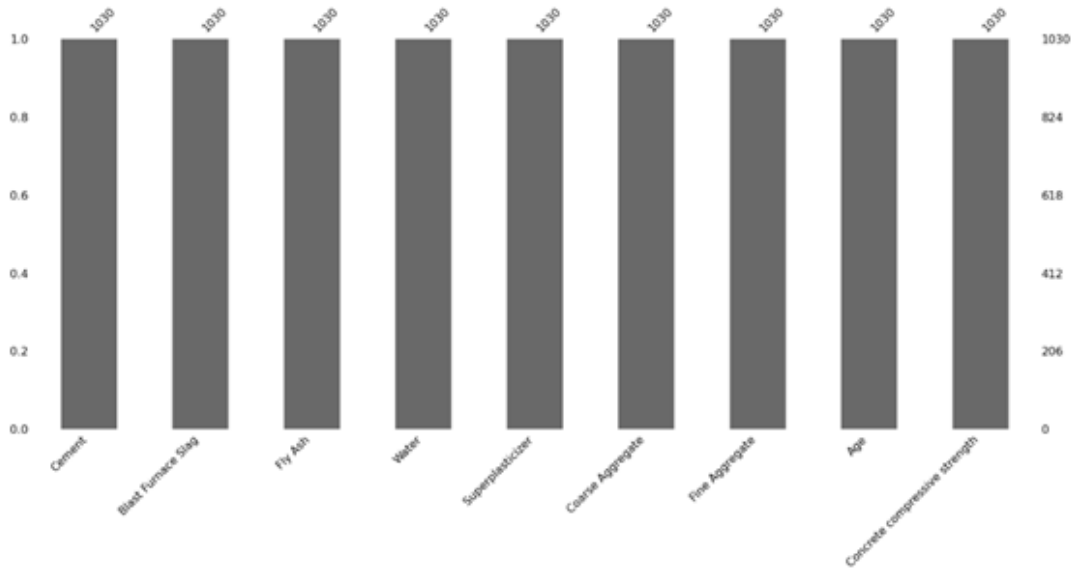
Out[16]:
```
Cement                          False
Blast Furnace Slag              False
Fly Ash                         False
Water                           False
Superplasticizer                False
Coarse Aggregate                False
Fine Aggregate                  False
Age                             False
Concrete compressive strength   False
dtype: bool
```

In [17]:
```
data.isnull().sum() #used for finding the null values
```

Out[17]:
```
Cement                          0
Blast Furnace Slag              0
Fly Ash                         0
Water                           0
Superplasticizer                0
Coarse Aggregate                0
Fine Aggregate                  0
Age                             0
Concrete compressive strength   0
dtype: int64
```

In [19]:
```
sns.heatmap(data.isnull(),cbar=False)
```

In [14]:
```
import missingno as msno #finding missing values
msno.bar(data)
plt.show()
```

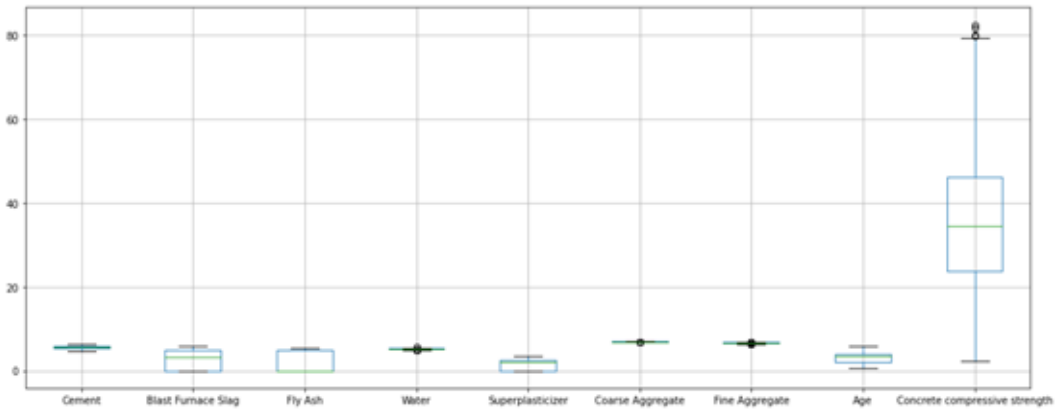


DATA VISUALISATION

Checking with Outliers

In [16]:
```
data.boxplot(figsize=(18,7))
```

Out[16]: <AxesSubplot:>



In [17]:
```
def outliers():
    for i in data.columns:
        print("\033[1m" + i + "\033[0m")
        Q1 = data[i].quantile(q=0.25) # 25% quartile value
        Q3 = data[i].quantile(q=0.75) # 75% quartile value
        print('1st Quartile (Q1) is: ', Q1)
        print('3st Quartile (Q3) is: ', Q3)
        IQR = Q3 - Q1 #finding Inter Quartile range
        print(IQR)
        UL = Q3 + (1.5 * IQR) #finding the upper level point

        print("upper limit : ",UL)
        LL = Q1 - (1.5 * IQR) #finding the lower level point
        print("lower limit : ",LL)
        print("Outlier present above UL",data[i][data[i]>UL].count())
        print("Outlier present above LL",data[i][data[i]<LL].count())
        print('---'*40)
outliers() # calling the function
```

Cement
1st Quartile (Q1) is: 5.26462568244329
3st Quartile (Q3) is: 5.860786223465865
0.5961605410225754
upper limit : 6.755027034999729
lower limit : 4.370384870909427
Outlier present above UL 0
Outlier present above LL 0

Blast Furnace Slag
1st Quartile (Q1) is: 0.0
3st Quartile (Q3) is: 4.969465836003737
4.969465836003737
upper limit : 12.423664590009341
lower limit : -7.454198754005605
Outlier present above UL 0
Outlier present above LL 0

Fly Ash
1st Quartile (Q1) is: 0.0
3st Quartile (Q3) is: 4.78164132910387
4.78164132910387
upper limit : 11.954103322759675
lower limit : -7.172461993655805
Outlier present above UL 0
Outlier present above LL 0

Water

Water
1st Quartile (Q1) is: 5.111385197196399
3st Quartile (Q3) is: 5.262690188904886
0.15130499170848655
upper limit : 5.489647676467616
lower limit : 4.88442770963367
Outlier present above UL 2
Outlier present above LL 12

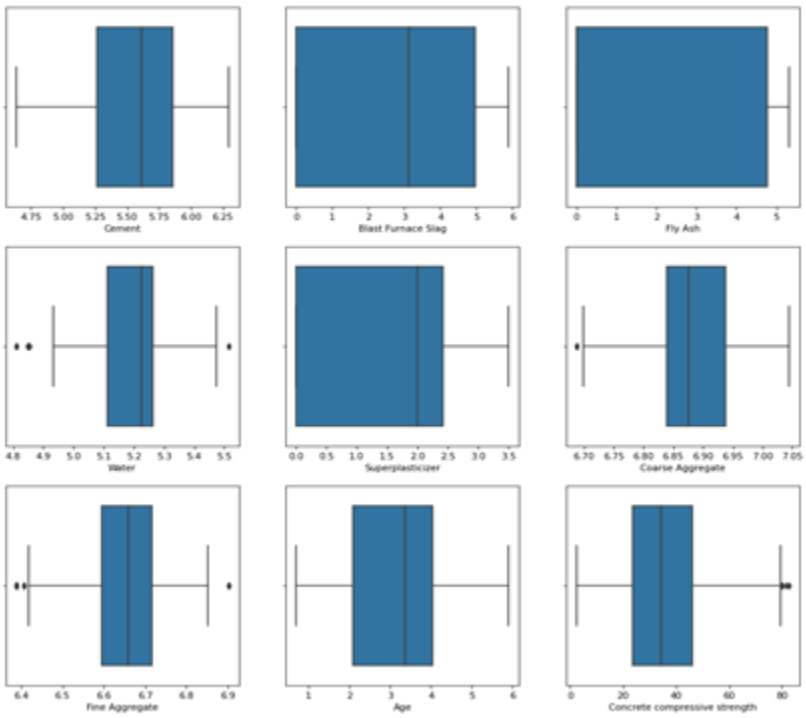
Superplasticizer
1st Quartile (Q1) is: 0.0
3st Quartile (Q3) is: 2.4159137783010487
2.4159137783010487
upper limit : 6.0397844457526215
lower limit : -3.6238706674515733
Outlier present above UL 0
Outlier present above LL 0

Coarse Aggregate
1st Quartile (Q1) is: 6.838405200847344
3st Quartile (Q3) is: 6.93770235535009
0.09929715450274568
upper limit : 7.086648087104209
lower limit : 6.689459469093226
Outlier present above UL 0
Outlier present above LL 6

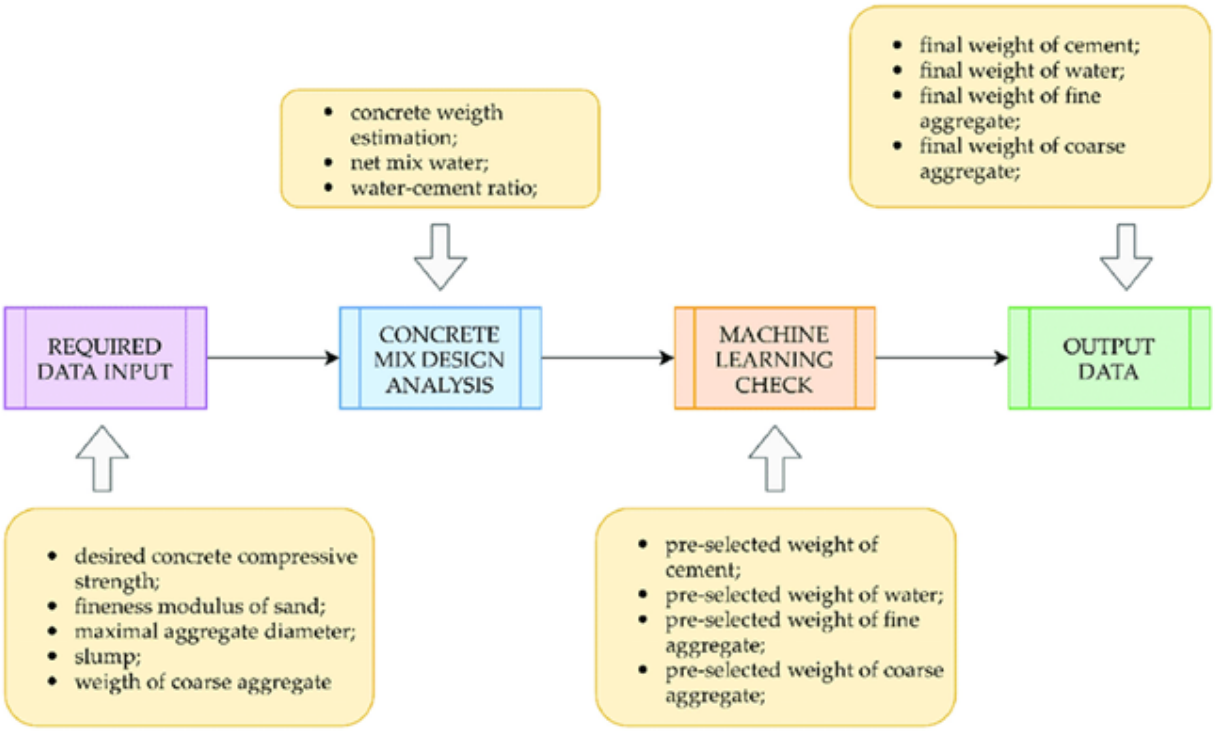
Fine Aggregate
1st Quartile (Q1) is: 6.595711359522024
3st Quartile (Q3) is: 6.715383386334681
0.11967202681265654
upper limit : 6.894891426553666
lower limit : 6.416203319303039
Outlier present above UL 5
Outlier present above LL 35

Age
1st Quartile (Q1) is: 2.0794415416798357
3st Quartile (Q3) is: 4.04305126783455
1.9636097261547145
upper limit : 6.988465857066622
lower limit : -0.8659730475522358
Outlier present above UL 0
Outlier present above LL 0

Concrete compressive strength
1st Quartile (Q1) is: 23.709999999999997



FLOWCHART


REQUIRED DATA INPUT
CONCRETE MIX DESIGN ANALYSIS
MACHINE LEARNING CHECK
OUTPUT DATA
concrete weigth estimation;
net mix water;
water-cement ratio;
final weight of cement;
final weight of water;
final weight of fine aggregate;
final weight of coarse aggregate;
desired concrete compressive strength;
fineness modulus of sand;
maximal aggregate diameter;
slump;
weigth of coarse aggregate
pre-selected weight of cement;
pre-selected weight of water;
pre-selected weight of fine aggregate;
pre-selected weight of coarse aggregate;

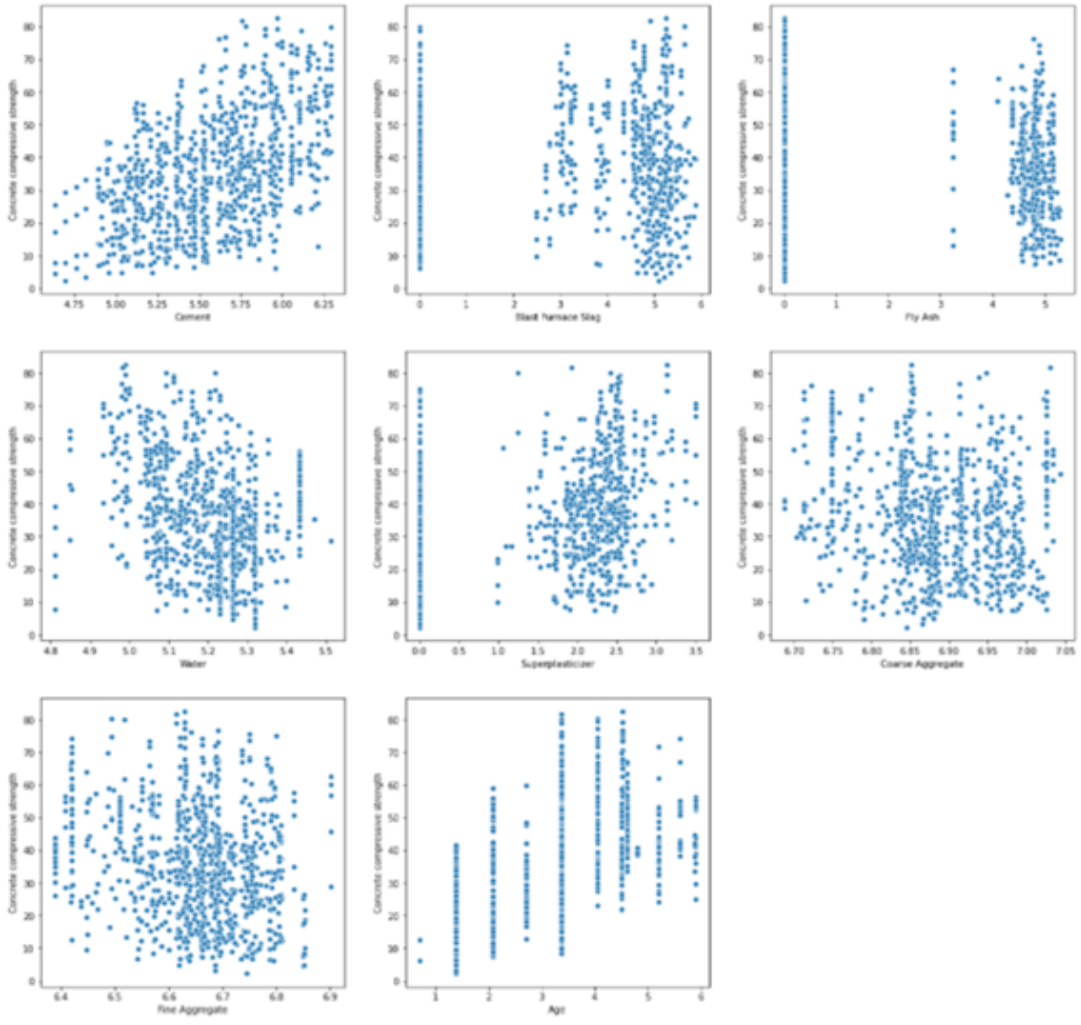
RESULT

ANACONDA (JUPYTER)

Scatterplot with the Target Column

In [30]:
```
y1=data["Concrete compressive strength"] #target column
plt.figure(figsize=(20,20),facecolor='white') #setting the figure size
plotnumber = 1

for column in data.columns[:8]: #iterating through each columns
    ax = plt.subplot(3,3,plotnumber) #as we have 9 columns that why we are 3X3
    sns.scatterplot(data[column],y1) #plottng scatterplot
    plt.xlabel(column,fontsize=10) #labeling the column name at x-axis
    plotnumber+=1
plt.show()
```



Correlation between the independent columns

In [31]:
```
plt.figure(figsize=(8,8)) #figure size of width-8 and height-8
sns.heatmap(data[:8].corr().abs(), vmin = -0.5,vmax = 1,annot=True)
```

Out[31]: <matplotlib.axes._subplots.AxesSubplot at 0x261bf172220>

Creating the Independent and Dependent Columns

In [21]:
```
x=pd.DataFrame(data,columns=data.columns[:8]) #independet columns
y=pd.DataFrame(data,columns=data.columns[8:]) #dependent column
```

Splitting the data into train and split

In [22]:
```
x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.2,random_state=1)
print(x_train.shape)
print(y_train.shape)
```

```
(824, 8)
(824, 1)
```

Multi Linear Regressor

In [51]:
```
from sklearn.linear_model import LinearRegression
mr=LinearRegression() #object of GradientBoostingRegressor
mr.fit(x_train,y_train) # training our data using fit method
```

Out[51]:
```
LinearRegression()
```

Checking the score with both the model

In [52]:
```
score=mr.score(x_test,y_test)

print("Accuracy of the model:",score)
```

```
Accuracy of the model: 0.7588745544697437
```

Predicting the output with our build models

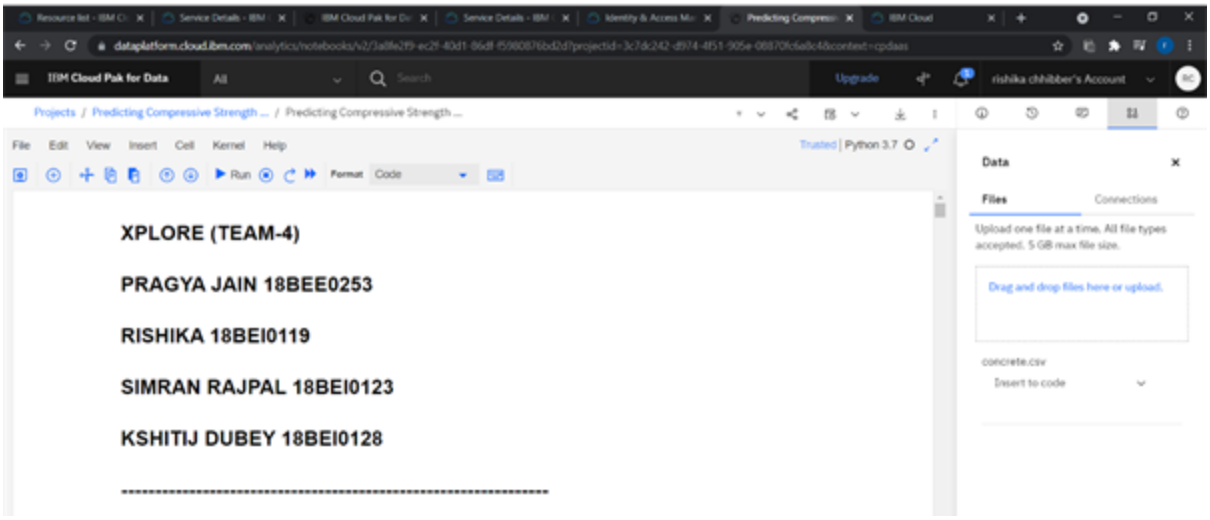
In [53]:
```
y_pred mr.predict(x_test) #predicting using predict method
print("Prediction made by Gradient Boosting model:",y_pred)
```

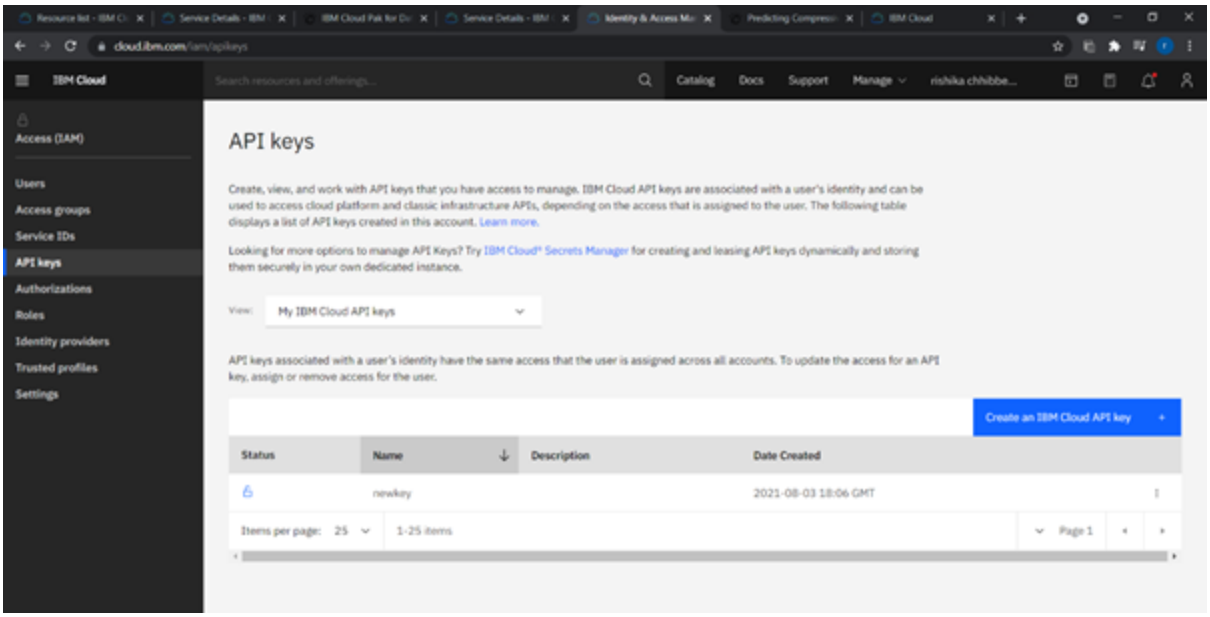
```
Prediction made by Gradient Boosting model: [[20.74680498]
 [10.7154792 ]
 [25.48286318]
 [19.49453633]
 [30.23373575]
 [60.90631683]
 [21.02945591]
 [50.93654851]
 [39.07451541]
 [23.81577986]
 [57.65533106]
 [17.30509287]
 [36.22802004]
 [25.26412982]
 [11.76260783]
 [28.34373385]
 [51.4278758 ]
 [55.11960116]
 [55.29194473]
 [32.60957553]
 [21.71693202]
 [40.61328039]
 [18.0597374 ]
 [57.38246876]
```

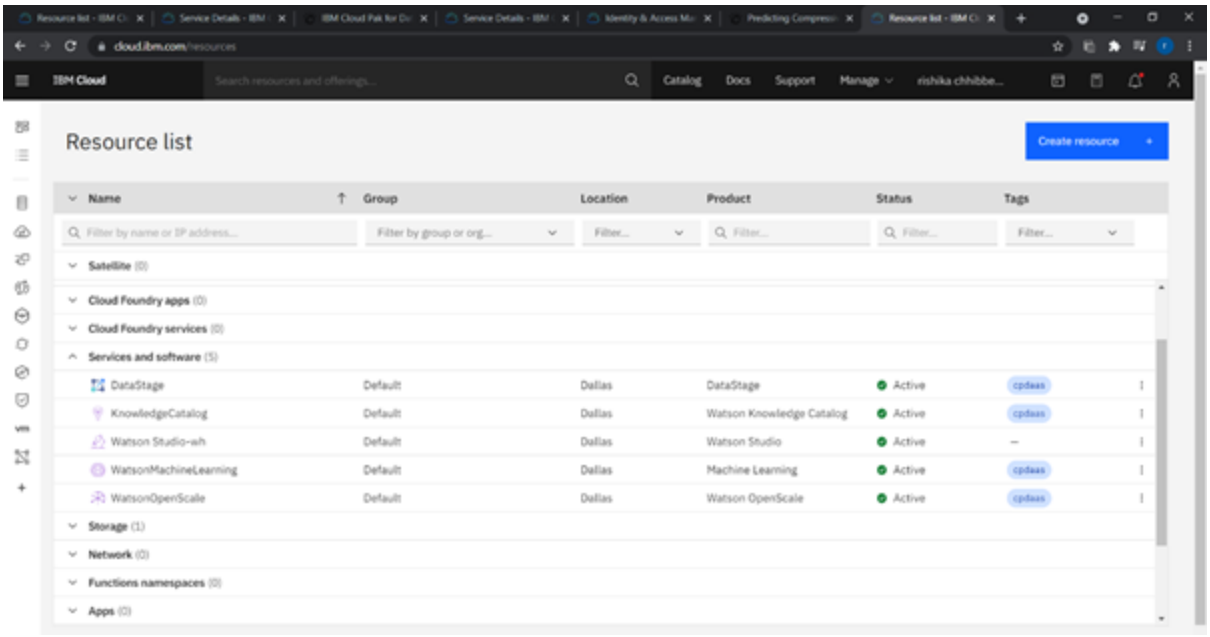
jupyter Predicting Compressive Strength of Concrete Last Checkpoint: Yesterday at 3:51 PM (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Not Trusted | Python 3
Run Markdown

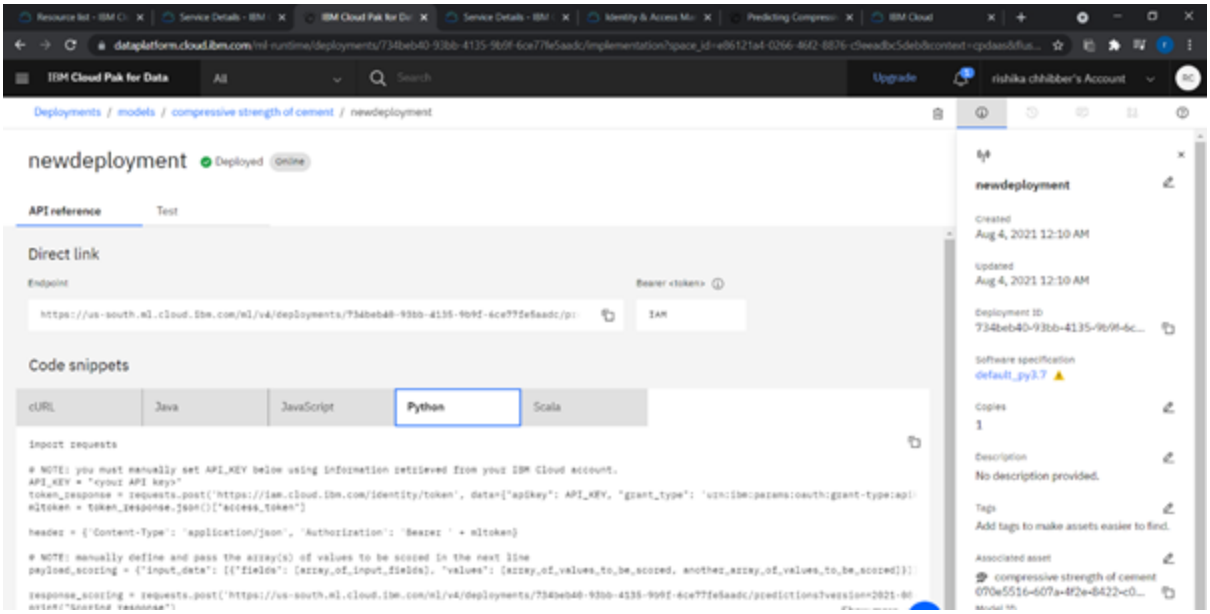
```
 [42.05886032]
 [34.15083446]
 [29.09776951]
 [41.22731239]
 [12.58416631]
 [63.41095022]
 [55.42795058]
 [31.05300015]
 [23.88761991]
 [43.18784356]
 [40.81992552]
 [22.51663109]
 [29.28228007]
 [33.6830338 ]
 [16.37132846]
 [43.23771114]
 [13.29237951]
 [56.23626758]
 [35.88185111]
 [31.52618463]
 [31.73064379]
 [32.76252501]
 [30.53397513]
 [45.31164158]
 [41.39914922]
 [24.81062003]
 [38.20024311]
 [61.45589018]
 [57.9601133 ]
 [64.77993313]
 [ 7.09111489]
 [41.61535275]
 [23.5902705 ]
 [25.75721171]
 [38.05831671]
 [46.64588311]
 [64.73753838]
 [50.88197961]
 [46.61663661]
 [31.89482732]
 [36.93548382]
 [45.02084293]
 [36.57075226]
 [33.47791281]
 [25.72600744]
 [38.46297641]
 [25.86603423]
 [40.91607439]
 [17.30221504]
 [61.35685586]
```

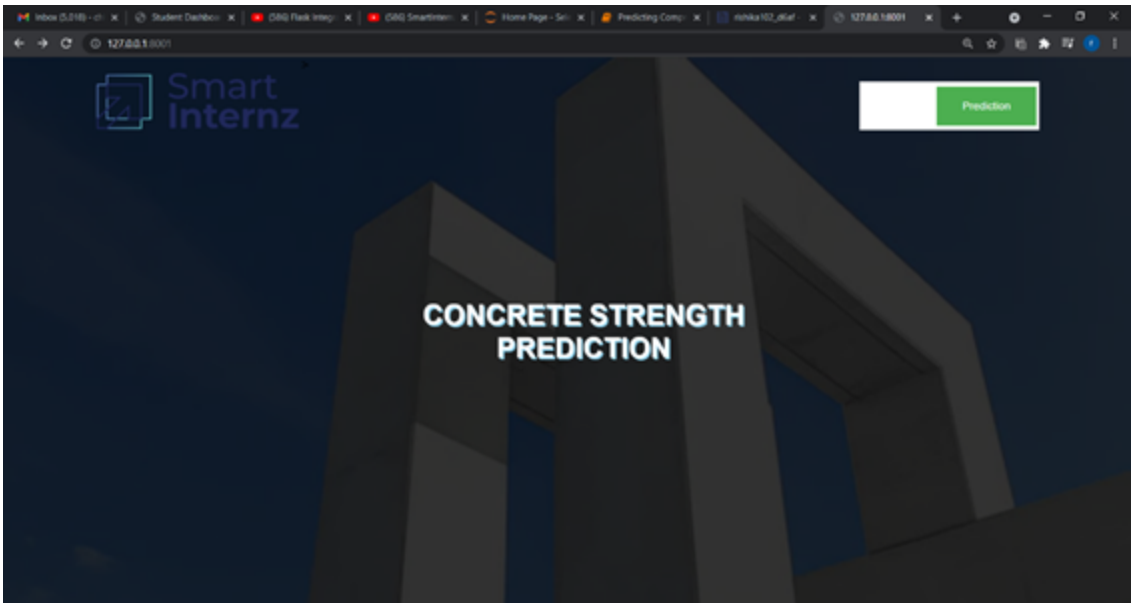
IBM WATSON STUDIO



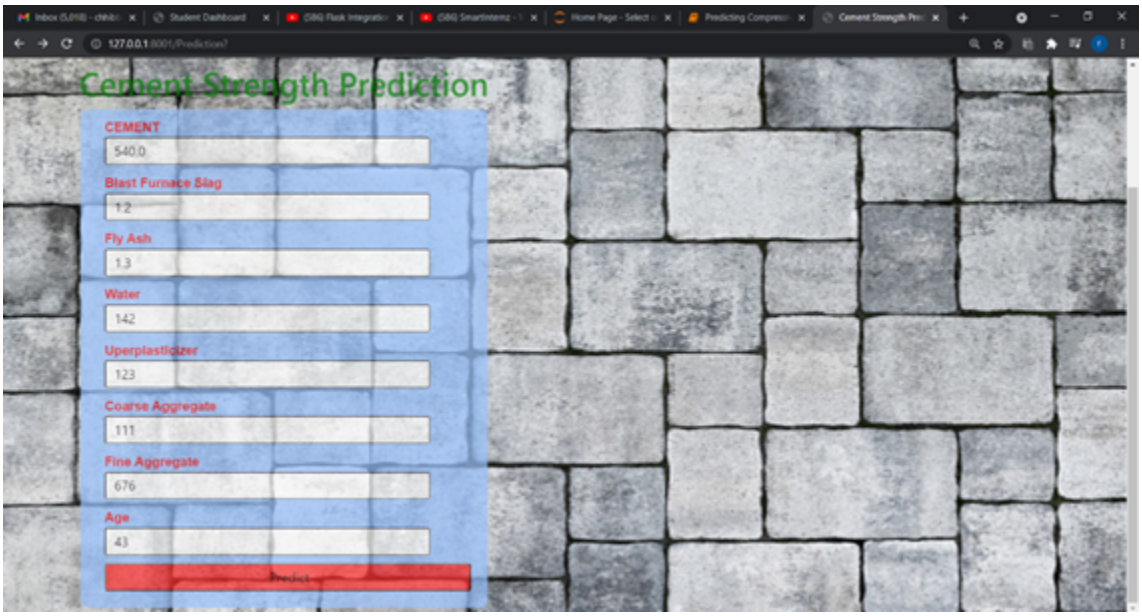



IBM Cloud
Catalog
Docs
Support
Manage
Resource list
Create resource
Name
Group
Location
Product
Status
Tags
Satellite (0)
Cloud Foundry apps (0)
Cloud Foundry services (0)
Services and software (5)
DataStage
Default
Dallas
DataStage
Active
KnowledgeCatalog
Watson Knowledge Catalog
Watson Studio-wh
Watson Studio
WatsonMachineLearning
Machine Learning
WatsonOpenScale
Watson OpenScale
Storage (1)
Network (0)
Functions namespaces (0)
Apps (0)


IBM Cloud Pak for Data
Deployments / models / compressive strength of cement / newdeployment
newdeployment
Deployed
Online
API reference
Test
Direct link
Endpoint
Bearer <token>
IAM
Code snippets
cURL
Java
JavaScript
Python
Scala
import requests
# NOTE: you must manually set API_KEY below using information retrieved from your IBM Cloud account.
API_KEY = "<your API key>"
# NOTE: manually define and pass the array(s) of values to be scored in the next line
Created
Aug 4, 2021 12:10 AM
Updated
Aug 4, 2021 12:10 AM
Deployment ID
Software specification
default_py3.7
Copies
1
Description
No description provided.
Tags
Add tags to make assets easier to find.
Associated asset
compressive strength of cement


Smart
Internz
Prediction
CONCRETE STRENGTH
PREDICTION


Cement Strength Prediction
Strength of the cement is [68.96999649] MPa


127.0.0.1:8001/Prediction?
Cement Strength Prediction
CEMENT
540.0
Blast Furnace Slag
1.2
Fly Ash
1.3
Water
142
Uperplasticizer
123
Coarse Aggregate
111
Fine Aggregate
676
Age
43
Predict

ADVANTAGES & DISADVANTAGES

Advantages:

Over the past five years, the implementations of the non-tuned machine learning model have shown a noticeable progress in multidisciplinary of science and engineering fields. This is due to its advantages (particularly over conventional artificial neural network algorithms) such as the randomly initiated hidden neurons without the need for iterative tuning process for free parameters or connections between hidden and output layer.

It is remarkably efficient to reach a global optimum, following universal approximation capability of single layer feed-forward network. Also, this model is featured by the efficiency and generalization performance over traditional learning algorithms (e.g., SVMs or ANNs) as revealed in the estimation problems in many different fields.

With the development of artificial intelligence, it is easier to forecast concrete compressive strength. When compared with other traditional regression methods, machine learning adopts specific algorithms that can learn from the input data and gives highly accurate results.

Disadvantages:

Concrete compressive strength requirements can vary from 2500 psi (17MPa) for residential concrete to 4000 psi (28 MPa) and higher in commercial structures. Higher strengths up to and exceeding 10000 psi (70 MPa) are specified for certain applications.

Machine Learning requires massive data sets to train on, and these should be inclusive/unbiased, and of good quality. There can also be times where they must wait for new data to be generated.

Any disadvantage of using a multiple regression model usually comes down to the data being used. Two examples of this are using incomplete data and falsely

concluding that a correlation is a causation.

APPLICATIONS

Concrete is considered by many to be a strong and durable material, and rightfully so. But there are different ways to assess concrete strength.

Water/cementitious ratio (w/cm)

This refers to the ratio of water to cement in the concrete mixture. A lower water-to-cement ratio makes for a stronger concrete, but it also makes the concrete more difficult to work with.

The right balance must be struck to achieve the desired strength while maintaining workability.

Proportioning

Traditional concrete is made of water, cement, air, and an aggregate mixture of sand, gravel, and stone. The right proportion of these ingredients is key for achieving a higher concrete strength.

A concrete mixture with too much cement paste may be easy to pour—but it will crack easily and not withstand the test of time. Conversely, too little cement paste will yield a concrete that is rough and porous.

Mixing

Optimal mixing time is important for strength. While strength does tend to increase with mixing time to a certain point, mixing for too long can actually cause excess water evaporation and the formation of fine particles within the mix. This ends up making the concrete harder to work with and less strong.

There is no golden rule for optimal mixing time, as it depends on many factors, such as: the type of mixer being used, the speed of the mixer rotation, and the specific components and materials within a given batch of concrete.

Curing methods

The longer the concrete is kept moist, the stronger it will become. To protect the concrete, precautions must be taken when curing concrete in extremely cold or hot temperatures.

CONCLUSION

Earlier and accurate estimation of concrete strength are valuable to the construction industry. The presence of such model would possibly obtain the hard balance and equality between controlling the quality (quality control process) and economics (saving time and expense, i.e., this model could be used in construction to make the necessary adjustments on mix proportion used, to avoid situations where, concrete does not reach the required design strength or by avoiding concrete that is unnecessarily strong.

This methodology allows a fast and accurate prediction of values for compressive strength on site. Common methods for estimation of in place strength requires extensive use of curing of mortar cubes at constant temperatures or the use of databases containing a large number of compressive strength values made at many ages and cured at different temperatures. These databases have to be fed with a statistical relevant amount of data before a reliable estimation of the strength can be made. Furthermore, all of these methods requires many hours of lab and field time for testing, collecting and analysing data.

Furthermore, the existing variables in the model yielded good reasonable results. Also, it is not preferred to load the prediction model with large number of variables, because it is preferred to use a model with lesser number of variables with higher possible accuracy to assure the rapid and easy use of the model.

FUTURE SCOPE

Compressive strength is one of the most important properties of concrete and mortar. The strength of the binder (cement) therefore has a significant effect on the performance characteristics of the mixture and ensures the overall quality of the finished product. The test for compressive strength is generally carried out by crushing cubes of hardened cement-sand mortar in a compression machine.

For now, we have created the web application of the process, which can be further converted to an android app which will be more useful and convenient for the user to use and will help in saving time.

App development will help in creating the project process oriented.

BIBLIOGRAPHY

1. Multiple Regression Model for Compressive Strength Prediction of High-Performance Concrete

   Article in Journal of Applied Sciences · January 2009

2. Comprehensive Machine Learning-Based Model for Predicting Compressive Strength of Ready-Mix Concrete

3. PREDICTING COMPRESSIVE STRENGTH OF CONCRETE FOR VARYING WORKABILITY USING REGRESSION MODELS

   Article in International Journal Of Engineering & Applied Sciences · December 2014

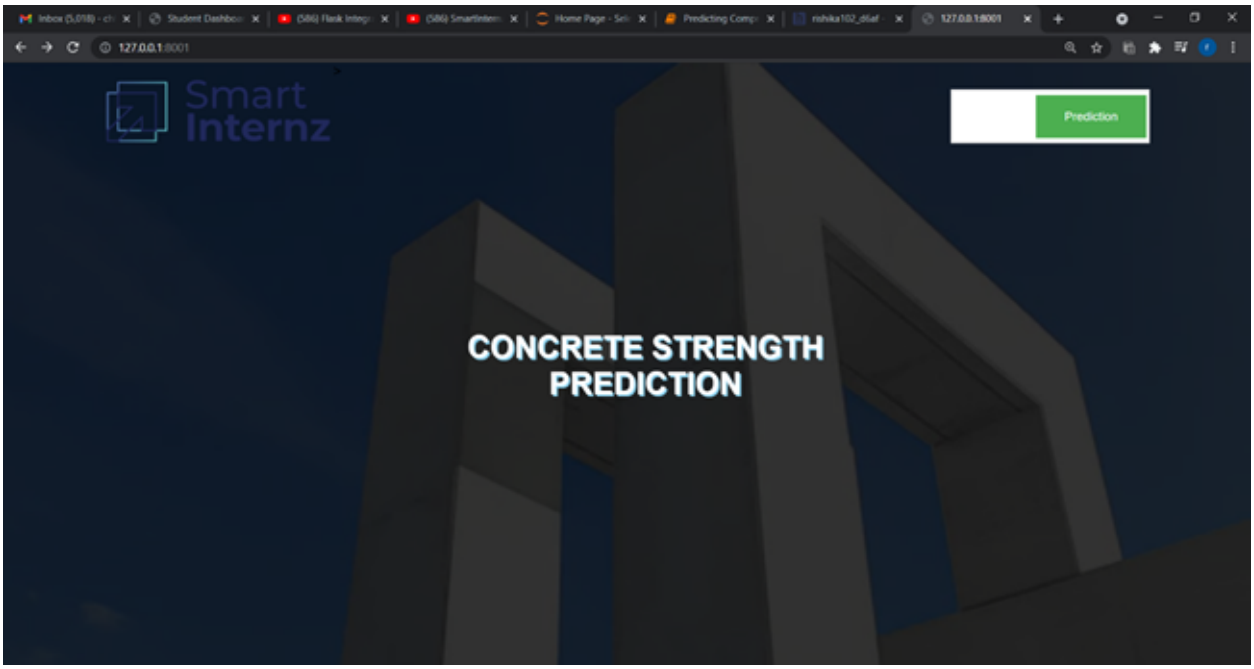
4. https://cor-tuf.com/everything-you-need-to-know-about-concrete-strength/

APPENDIX

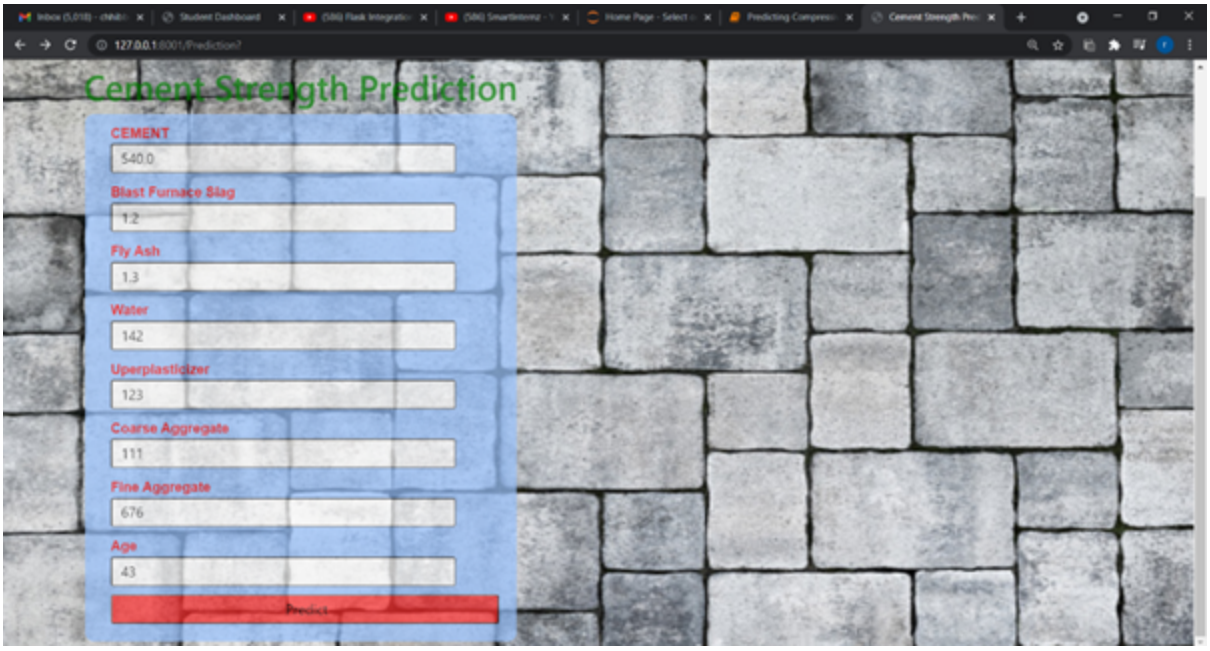
a. **Source Code**

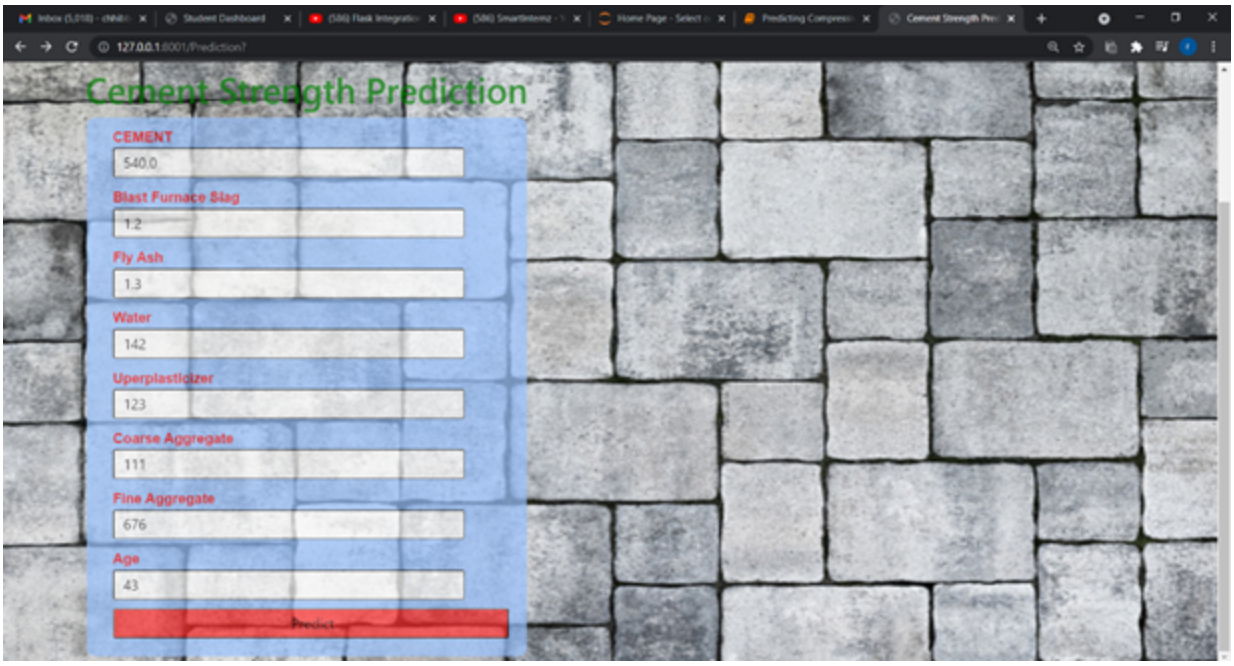
https://github.com/rishika28/Applied-Data-Science/blob/main/Predicting%20Compressive%20Strength%20of%20Concrete%20(1).ipynb

b. **UI output Screenshots**




127.0.0.1:8001/predict
Cement Strength Prediction
Strength of the cement is [68.96999649] MPa


127.0.0.1:8001/Prediction?
Cement Strength Prediction
CEMENT
540.0
Blast Furnace Slag
1.2
Fly Ash
1.3
Water
142
Uperplasticizer
123
Coarse Aggregate
111
Fine Aggregate
676
Age
43
Predict


127.0.0.1:8001/Prediction?
Cement Strength Prediction
CEMENT
540.0
Blast Furnace Slag
1.2
Fly Ash
1.3
Water
142
Uperplasticizer
123
Coarse Aggregate
111
Fine Aggregate
676
Age
43
Predict