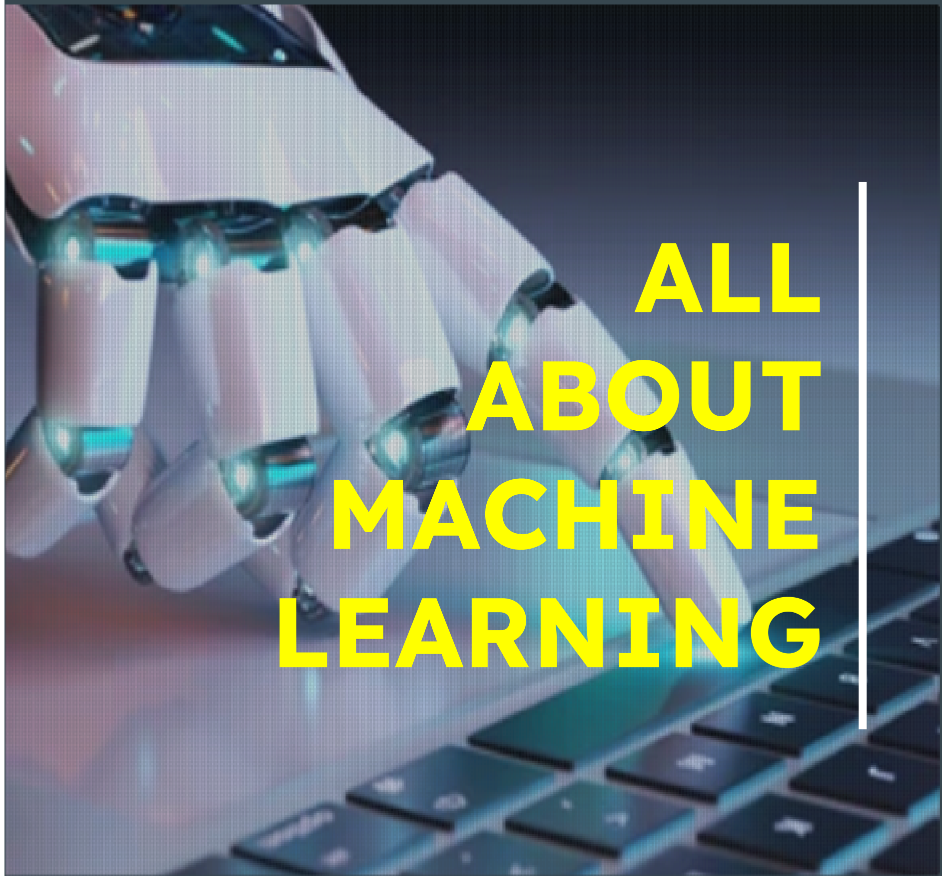
Compressive Strength of RCA Prediction

Leveraging AI for Sustainable Construction

Team Memeber: Shashi Ranjan
TeamId : 635011-U2PD388A



- Machine learning is a subset of artificial intelligence that allows computers to learn from data without being explicitly programmed.
- It has applications in various fields, including computer vision, natural language processing, and predictive analytics.
- In this presentation, we will explore the applications of machine learning in 'Predicting Compressive Strength of RCA' .

Problem Statement

- With increasing amount of Construction waste it's high time to adopt the sustainable development goals in construction also.
- It has been seen that lots of construction waste can be reused in new construction work.
- Although it is very difficult for normal people to know what amount of construction aggregate to use in fresh construction.
- Hence, to make this work easier we need to adopt the emerging technology 'Machine Learning'.

Data Description

Description of the dataset is provided below :
+ 236 rows and 19 colunms

```
data_copy.describe()
```

	Effective water- to-cement ratio	Aggregate-to-cement ratio (a/c)	RCA replacement ratio (RCA %)	Parent concrete strength(MPa)	Nominal maximum RCA size(mm)	Nominal maximum NA size(mm)	Bulk density of RCA (kg/m3)	Bulk density of NA (kg/m3)	Water absorption of RCA(WARCA) (%)	Water absorption of NA	Los Angeles abrasion of RCA	Los Angeles abrasion of NA	Density of hardened concrete AD (qad) (kg/m3)	Density of hardened concrete SSD (qSSD) (kg/m3)	Compressive strength (f'c) (MPa)
count	232.000000	210.000000	232.000000	22.000000	205.000000	205.000000	185.000000	155.000000	173.000000	149.000000	54.000000	43.000000	41.000000	30.000000	232.000000
mean	0.482716	3.163619	54.301724	42.718182	21.560976	22.302439	2418.708108	2682.451613	5.158960	1.155369	36.540741	25.139535	2365.390244	2368.366667	40.159052
std	0.116830	0.940518	38.928110	10.806254	6.053879	5.854966	139.598731	105.381015	1.724429	0.710237	9.302536	4.347037	61.415746	113.168621	15.539161
min	0.240000	1.700000	0.000000	28.250000	7.000000	7.000000	1946.000000	2387.000000	1.500000	0.200000	15.100000	11.900000	2210.000000	2115.000000	13.000000
25%	0.410000	2.600000	20.000000	37.300000	19.000000	20.000000	2330.000000	2600.000000	4.000000	0.600000	30.800000	22.500000	2340.000000	2320.000000	30.175000
50%	0.470000	3.000000	50.000000	38.325000	20.000000	20.000000	2400.000000	2670.000000	5.100000	1.100000	34.000000	24.800000	2365.000000	2360.000000	39.300000
75%	0.540000	3.500000	100.000000	42.425000	25.000000	25.000000	2489.000000	2730.000000	6.100000	1.480000	40.200000	29.100000	2390.000000	2425.000000	46.305000
max	0.840000	6.400000	100.000000	66.000000	32.000000	38.000000	2880.000000	2970.000000	11.900000	3.000000	59.800000	32.000000	2530.000000	2602.000000	102.500000

Data Cleaning

- The dataset that were provided was not cleaned data, so were expected to clean it then use in our model to predict the output.

Data Cleaning steps that were used :

- Removing columns in which target variabled were missing.

  I deleted that columns from dataset because if imputaions would have been done withot removing those columns it would have affected the accuracy of the result.

- Removing outlieres from the data by imputing with lower_limit and upper_limit values.

  Where, lower_limit=1st Quartile - 1.5*Inter Quartile Range

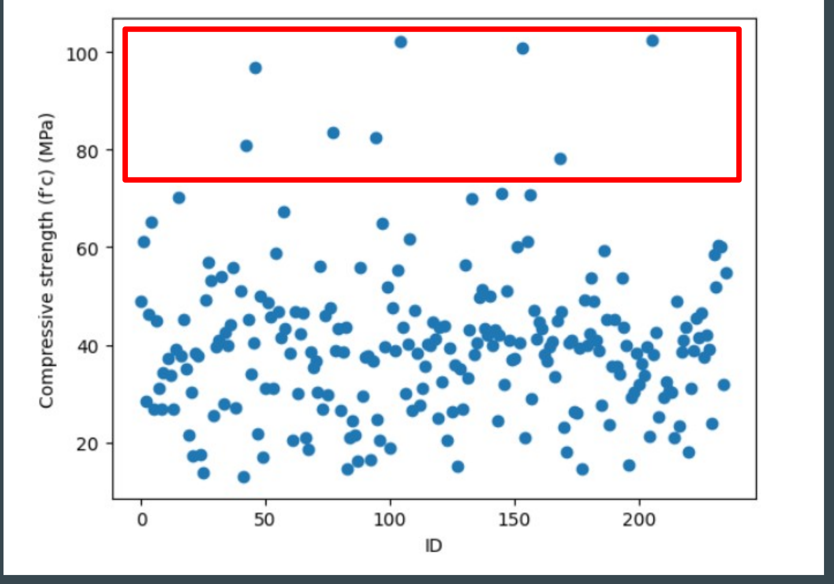
  upper_limit=3rd Quartile + 1.5*Inter Quartile Range

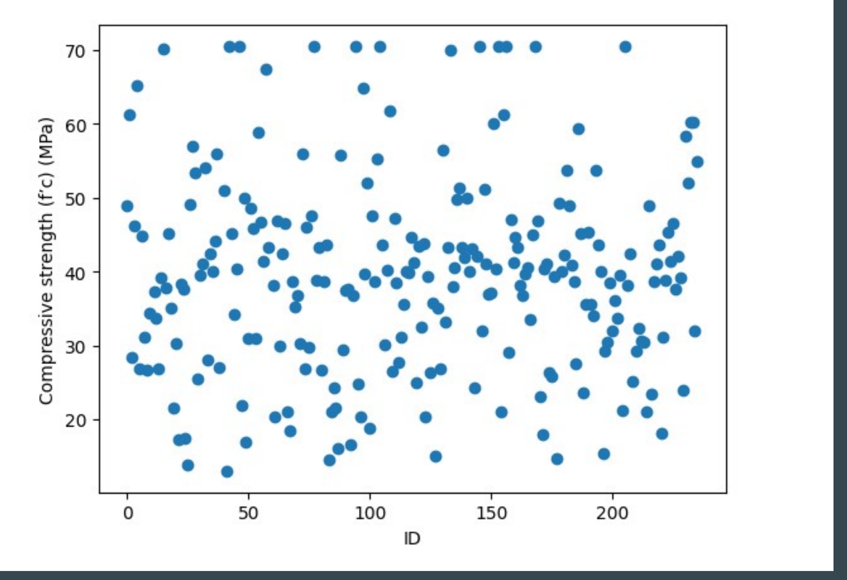
  The outliers greater then upper_limit is repleced with upper_limit value and the outliers lesser then lower_limit is replaced with lower_limit.

Scatter Plot

The below Scatter plot shows the outlies :

The points circled with red are the outliers .

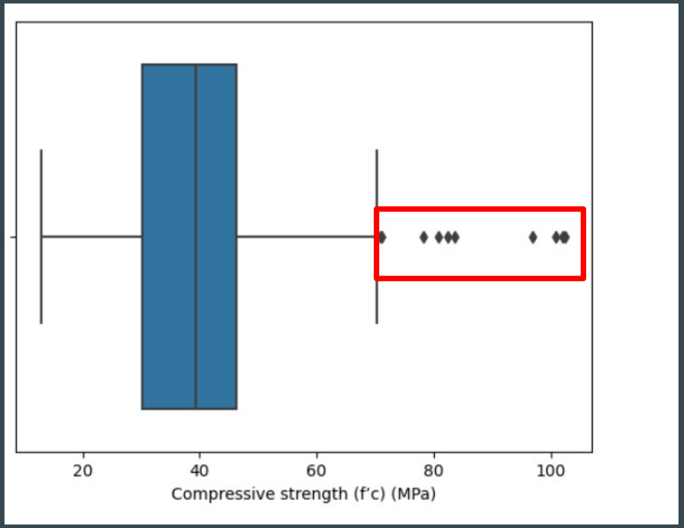


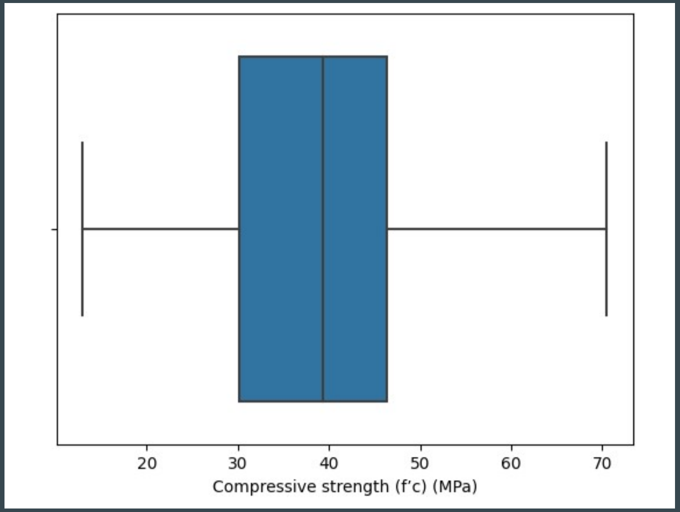


Box Plot

The below Box plot shows the outlies :

The points circled with red are the outliers .





Imputaion

- Mean imputaion have been applied on the columns that has integer or flaoting point values to fill the missing values.
- Columns that have been imputed using mean imputaion are :
    1. ‘Effective water- to-cement ratio'
    2. 'Aggregate-to-cement ratio (a/c)'
    3. 'RCA replacement ratio (RCA %)'
    4. 'Aggregate-to-cement ratio (a/c)'
    5. 'Parent concrete strength(MPa)'
    6. 'Nominal maximum RCA size(mm)'
    7. 'Nominal maximum NA size(mm)'
    8. 'Water absorption of RCA(WARCA) (%)'
    9. 'Water absorption of NA'
    10. 'Los Angeles abrasion of RCA'
    11. 'Los Angeles abrasion of NA'

These imputed values are saved to a .csv file ‘MeanImputedData.csv’. In the further process this new file is used.

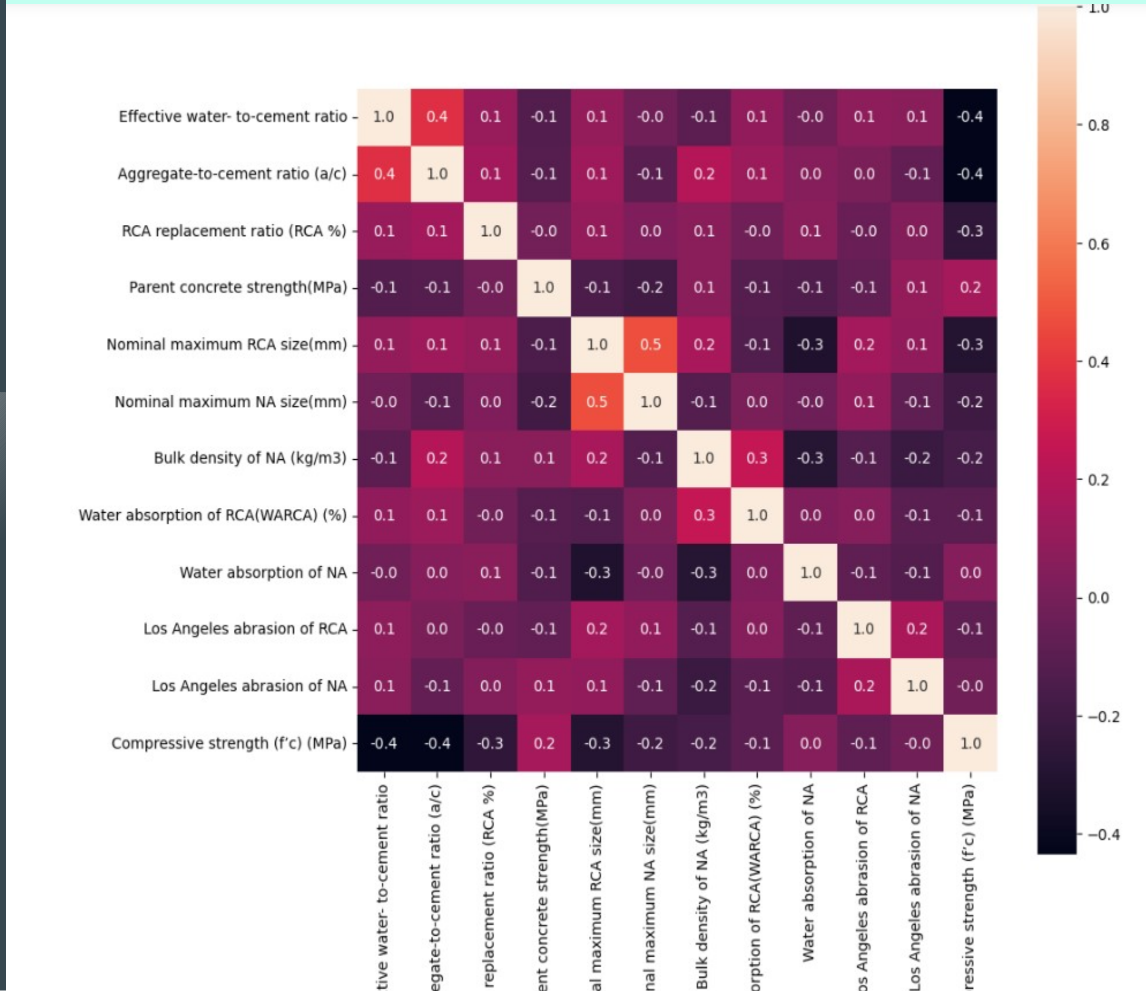
Deleting Unwanted Columns

- In ‘MeanImputedData.csv’ file there were certain columns whose datatypes was ‘Object’ and they were not of much use in the prediction. Hence i deleted them.
- Columns that have been deleated using drop function :
  1. 'Compressive strength tests'
  2. 'Elastic modulus tests'
  3. 'Flexural strength tests'
  4. 'Splitting tensile strength tests'
  5. 'Bulk density of RCA (kg/m3)'
  6. 'Bulk density of RCA (kg/m3)'
  7. 'Density of hardened concrete AD (qad)(kg/m3)'
  8. 'Density of hardened concrete SSD (qSSD)(kg/m3)'
  9. Elastic modulus(Ec)(MPa)
  10. Flexural strength(f’r)(MPa)
  11. Splitting tensile strength(MPa)

After deleting the columns, ‘MeanImputedData.csv’ is updated so that it can be used in to preapare the model.

The hetamap in the figure shows the correlation between the elements.

Heat-Map



Machine Learning Models Used

- Following are the model applied to predict the Compressive strength of RCA :

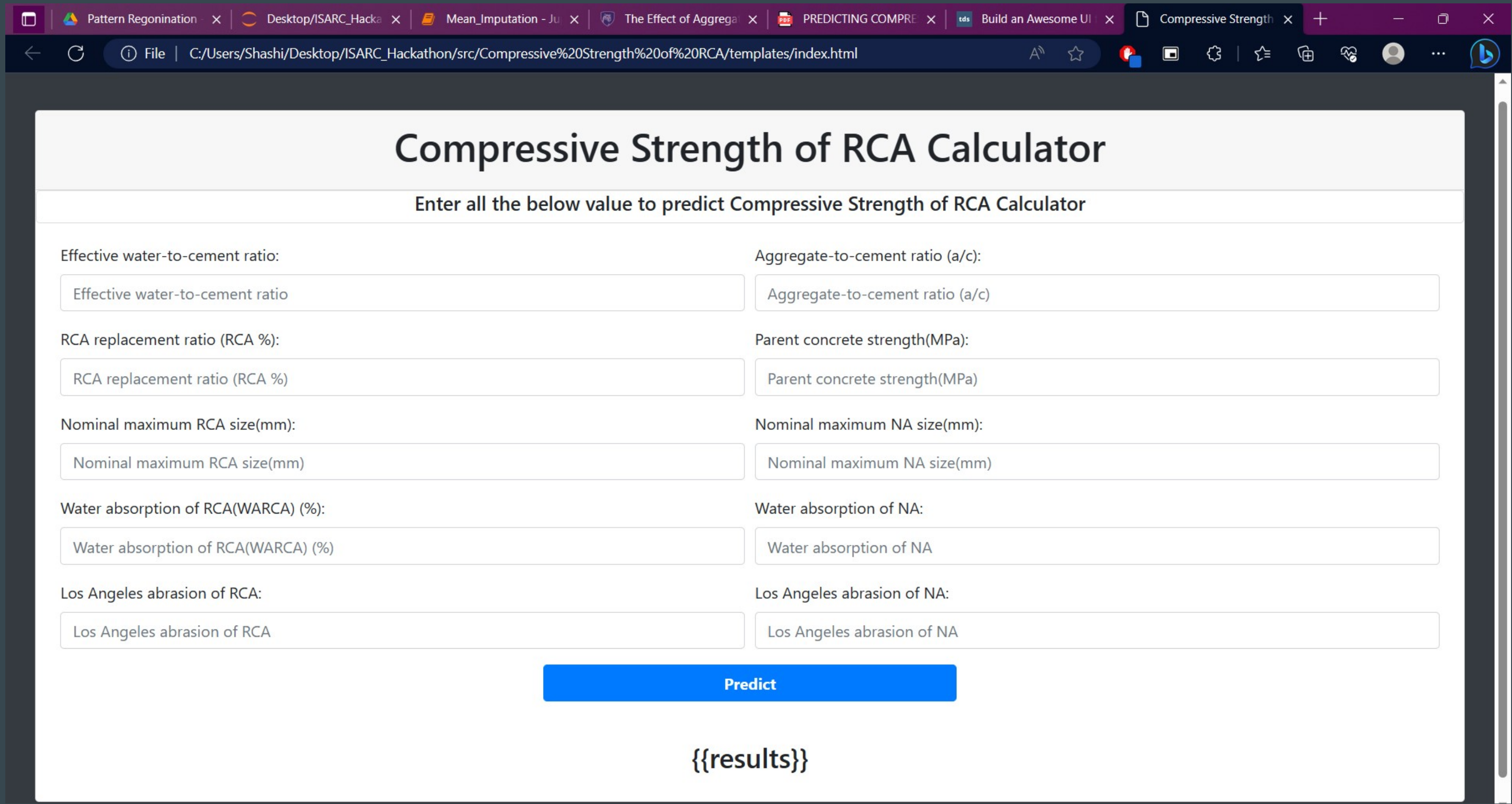
  - ✓ Multiple Linear Regression:
  - ✓ Decision Tree Regression
  - ✓ Random Forest Regression
  - ✓ Gradient Boosting Regression
  - ✓ AdaBoost Regressor
  - ✓ XGBoost Regressor

Results

Following are results obtained from those Machine Learning Algorithms :

S.No	Model	Mean Squared Error	Score
1.	Multiple Linear Regression	111.64	0.24369595103796715
2.	Decision Tree Regression	132.61	0.10163528858518911
3.	Random Forest Regression	132.61	0.5393021553791466
4.	Gradient Boosting Regression	74.45	0.50
5.	AdaBoost Regressor	76.58	0.4812255093140996
6.	XGBoost Regressor	67.36	0.5436879380510502

Web Interface



Conclusion

As it can be seen that most accurate result is provided by the XGBoost Regression with

Mean Squared Error .(MSE) 67.36 &

Score 0.5436879380510502

Hence, XGBoost Algorithms is used to predict the ‘Compressive Strength of RCA’.