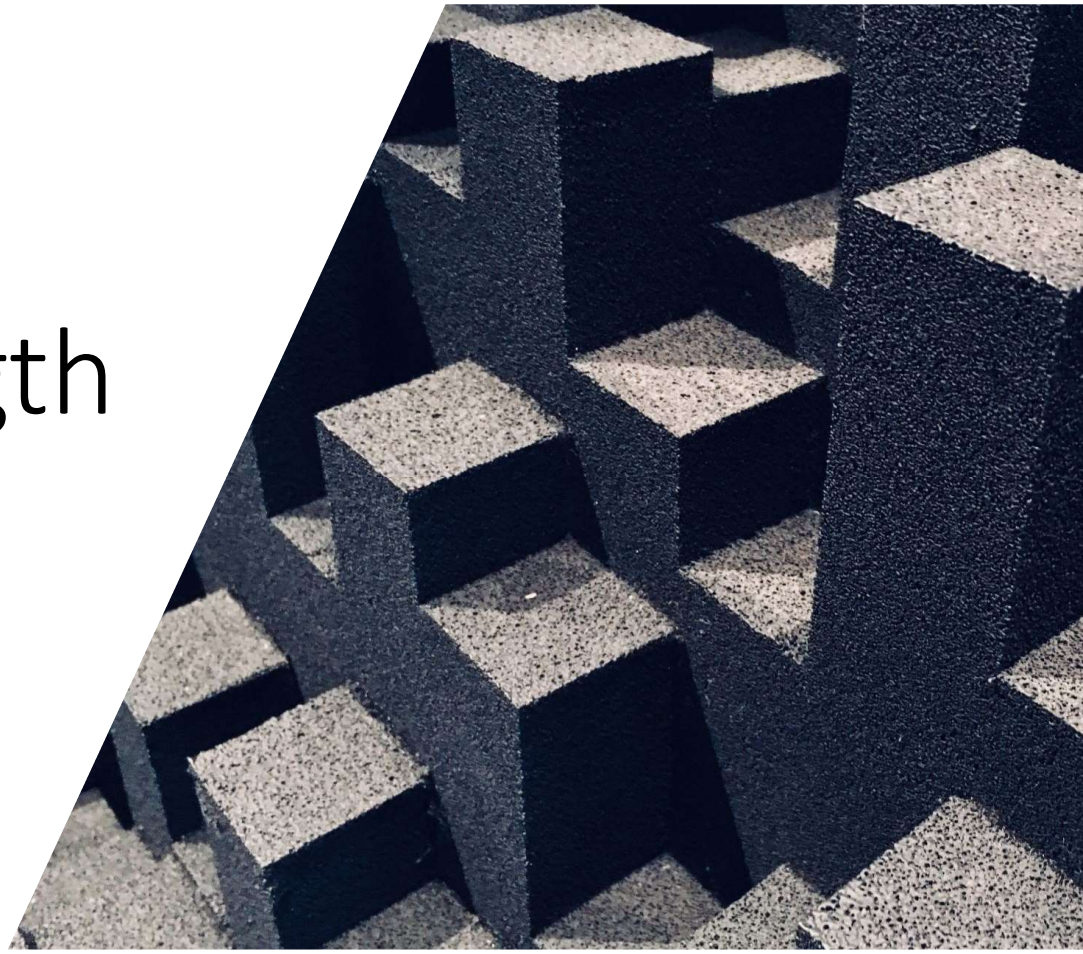
Concrete Compressive Strength Prediction

- Yatrik Shah



Introduction

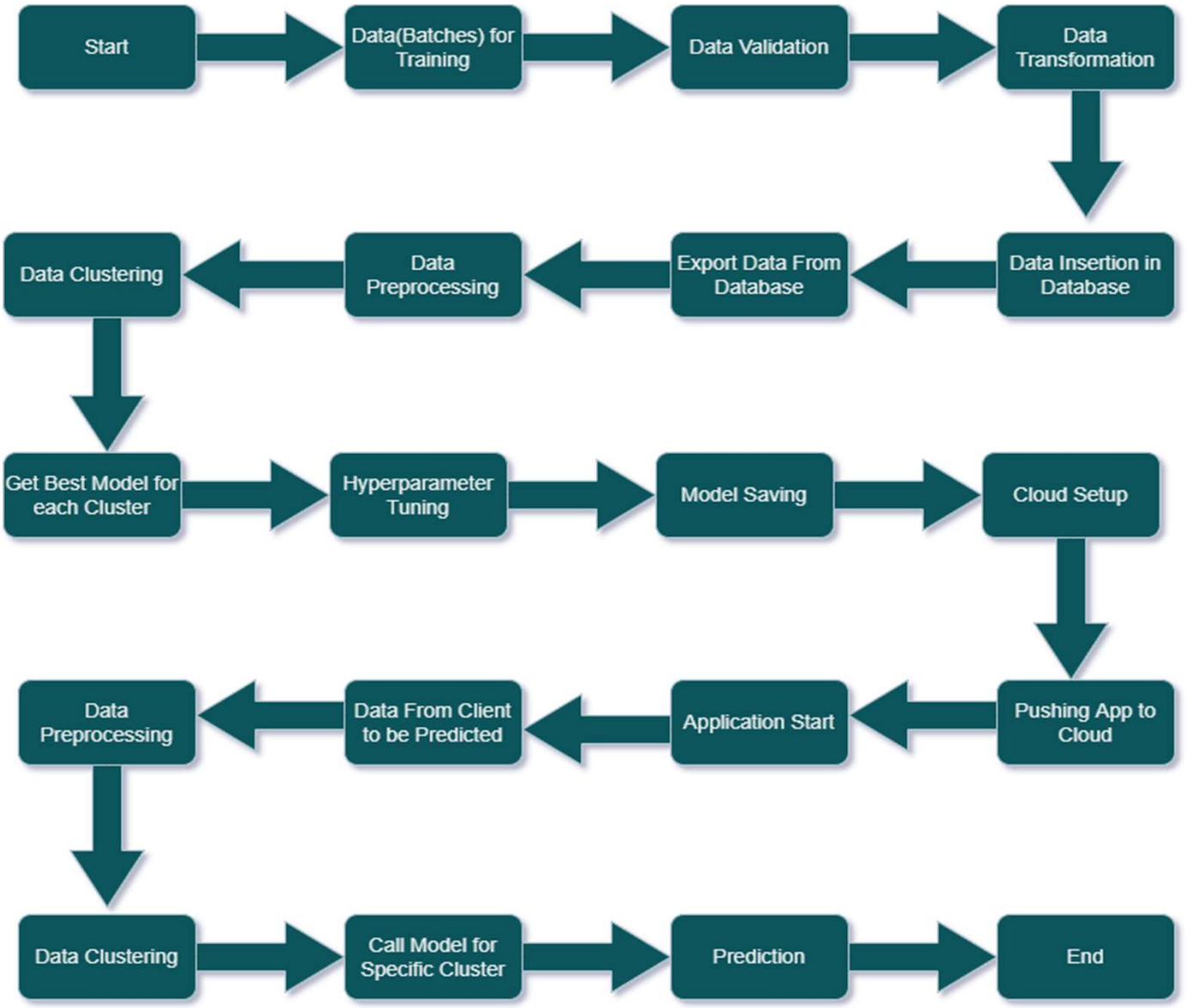
- **Concrete is most common and oldest material for construction worldwide. Infrastructure and construction are considered as extremely sensitive domains regarding the safety issue. So, since concrete being the main component or material for the construction it would be important to measure the strength or the power of the concrete using some reliable methods. Concrete is basically a composite material composed of various base materials like cement, water, Coarse Aggregate, Fine Aggregate, and some other components. Compressive strength of concrete is measured using a conventional crushing test on a concrete cylinder. Basically, it takes 28 days of time.**

Objective

The quality of concrete is determined by its compressive strength, which is measured using a conventional crushing test on a concrete cylinder. The strength of the concrete is also a vital aspect in achieving the requisite longevity. It will take 28 days to test strength, which is a long period.

We can save a lot of time and effort by using Data Science to estimate how much quantity of which raw material we need for acceptable compressive strength.

Workflow


Start
Data(Batches) for Training
Data Validation
Data Transformation
Data Insertion in Database
Export Data From Database
Data Preprocessing
Data Clustering
Get Best Model for each Cluster
Hyperparameter Tuning
Model Saving
Cloud Setup
Pushing App to Cloud
Application Start
Data From Client to be Predicted
Data Preprocessing
Data Clustering
Call Model for Specific Cluster
Prediction
End

Dataset Description

- The Following Dataset is used for the model training
- Which is Basically taken from UCI Public Repository:
- https://archive.ics.uci.edu/ml/datasets/Concrete+Compressive+Strength

Name	Datatype	Measurement	Description
Cement	Quantitative	kg in a m3 mixture	Input Variable
Blast Furnace Slag	Quantitative	kg in a m3 mixture	Input Variable
Fly Ash	Quantitative	kg in a m3 mixture	Input Variable
Water	Quantitative	kg in a m3 mixture	Input Variable
Superplasticizer	Quantitative	kg in a m3 mixture	Input Variable
Coarse Aggregate	Quantitative	kg in a m3 mixture	Input Variable
Fine Aggregate	Quantitative	kg in a m3 mixture	Input Variable
Age	Quantitative	kg in a m3 mixture	Input Variable
Concrete compressive strength	Quantitative	Megapascal	Output Variable

Data Preprocessing

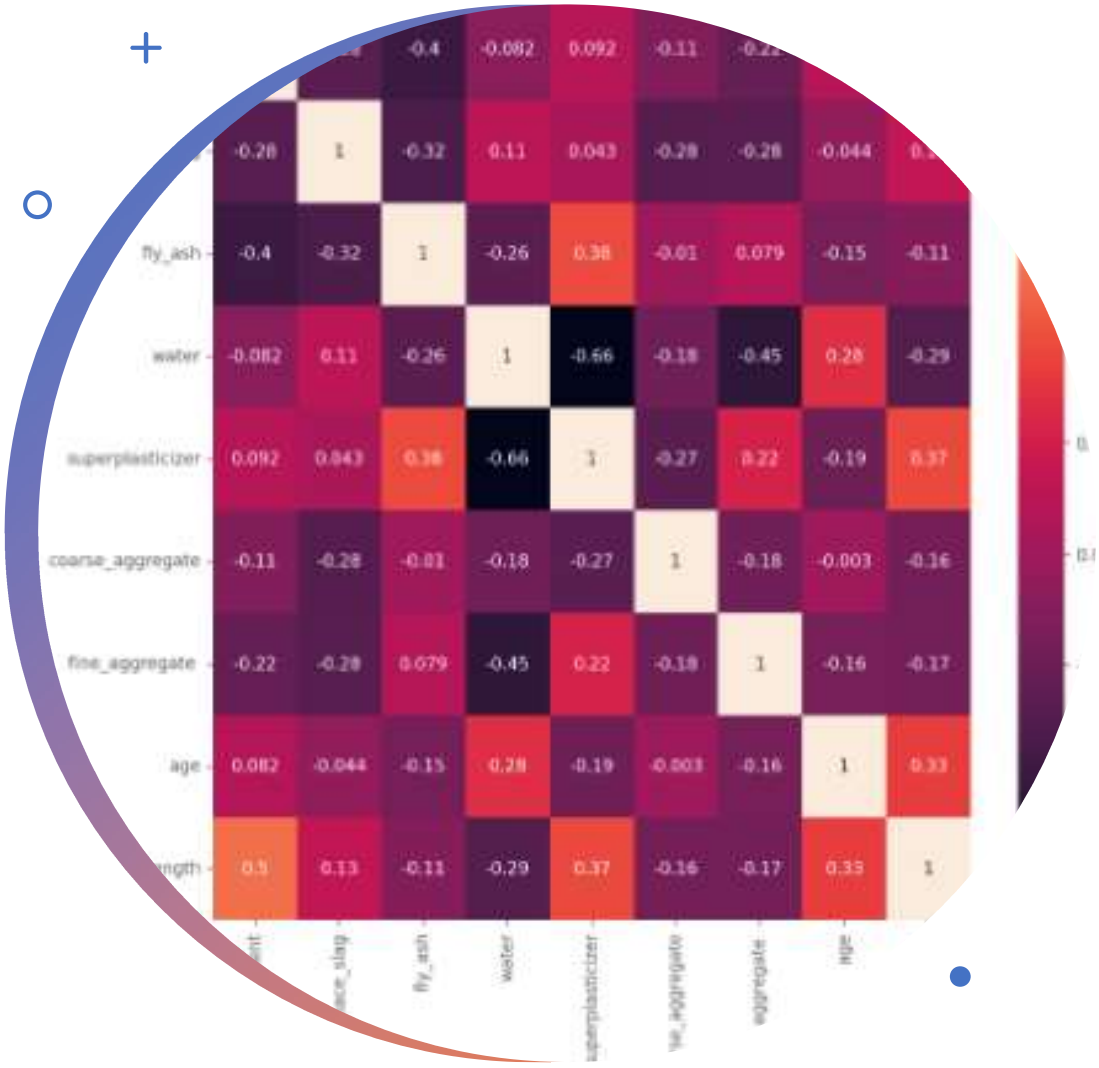
- This data is having 9 columns:
- 8: Independent columns
- 1: Dependent column

- Here the datatypes of the columns are perfect, No need to update or typecast

- Here there are no Null values

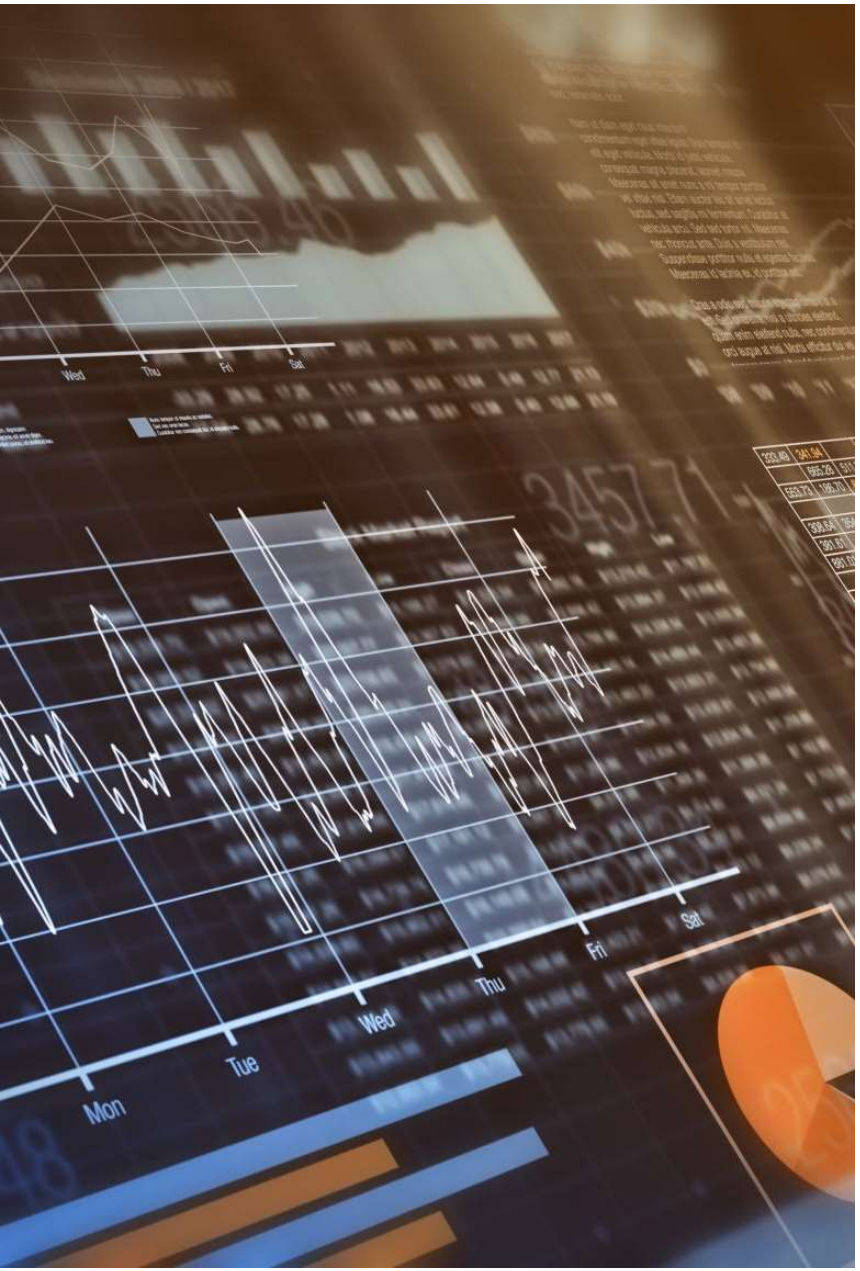
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1030 entries, 0 to 1029
Data columns (total 9 columns):
 #   Column                         Non-Null Count  Dtype  
---  ------                         --------------  -----  
 0   cement                         1030 non-null   float64
 1   blast_furnace_slag             1030 non-null   float64
 2   fly_ash                        1030 non-null   float64
 3   water                          1030 non-null   float64
 4   superplasticizer               1030 non-null   float64
 5   coarse_aggregate               1030 non-null   float64
 6   fine_aggregate                 1030 non-null   float64
 7   age                            1030 non-null   int64  
 8   concrete_compressive_strength  1030 non-null   float64
dtypes: float64(8), int64(1)
memory usage: 72.5 KB
```



EDA – Data Visualization

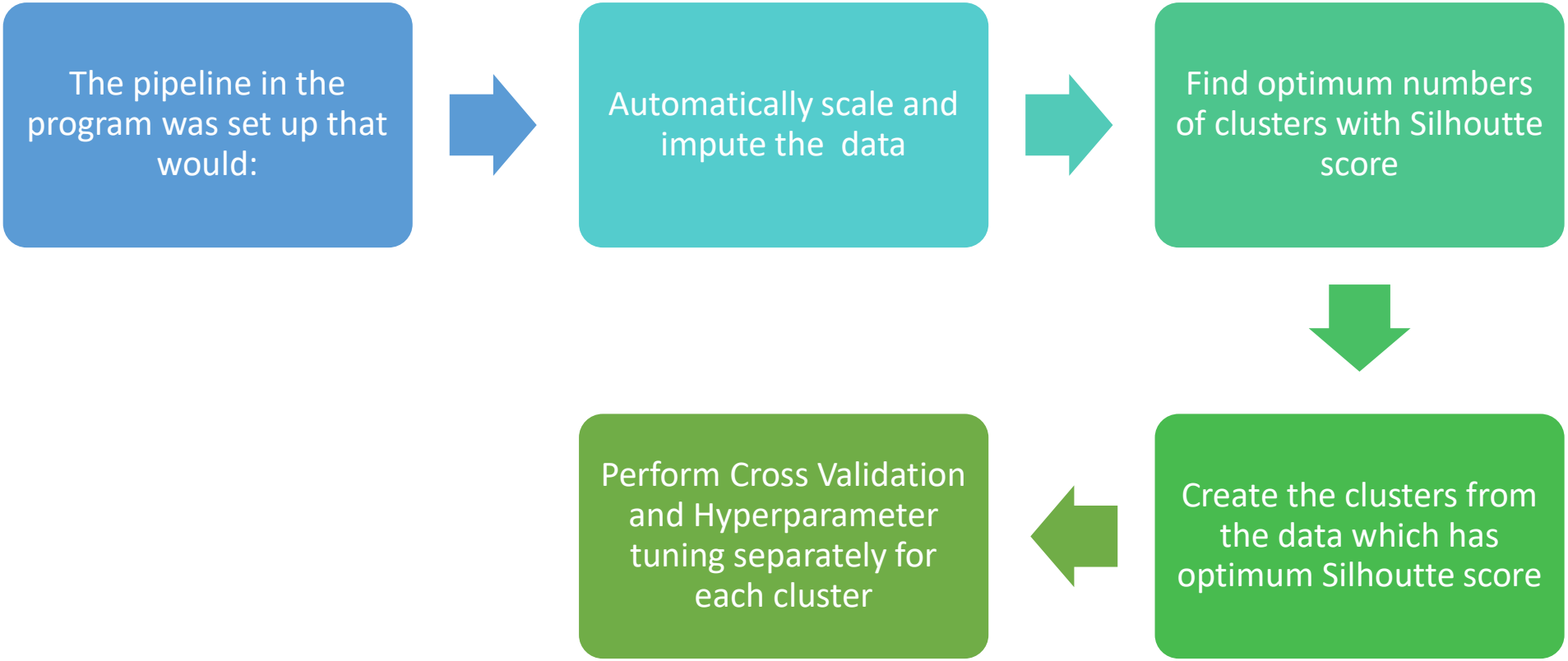
- Here Exploratory Data Analysis is done.
- Here no multicollinearity issue detected
- But by plotting the distribution plots of all the columns, it was detected that : not all columns are having normal or Gaussian distribution.
- So in the feature engineering step, feature scaling will be done.



Feature Engineering

- Data was clean but some of the feature engineering was needed.
- All the columns were scaled with Standardization to , which were required for clustering algorithm and other Machine Learning process.
- Data Imputation is done of there is any Null values in any new incoming data

Model Building

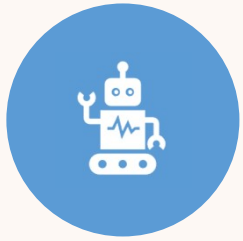


Model Building

AFTER PERFORMING HYPERPARAMETER TUNING GET BEST R_SQUARED OF EACH MACHINE LEARNING MODEL.

SAVE ALL THE PERFORMANCE METRICES AND SCORES IN THE RESPECTING LOG FILE.

THEN SAVE THE FINAL MODEL WHICH HAS THE MAXIMUM R_SQUARED BY REPLACING EXISTING MODELS.



APPLICATION IS READY FOR THE PREDICTION.

Model Deployment

Model deployment is done on Heroku server.

Link : https://ml-concrete-strength.herokuapp.com/

Thank You