Architecture

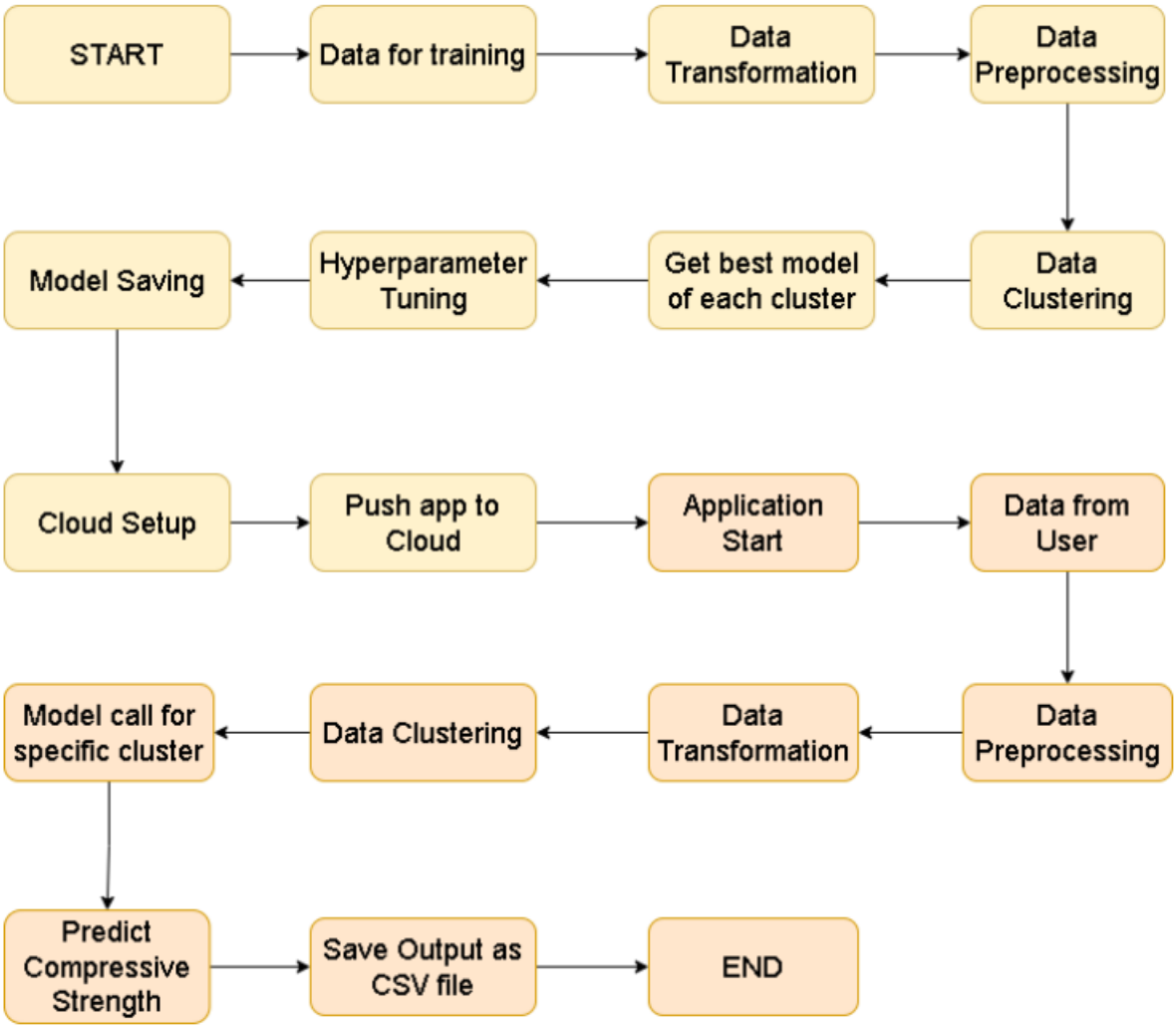
CONCRETE COMPRESSIVE STRENGTH PREDICTION
USING
RANDOM FOREST REGRESSOR AND LINEAR REGRESSION

DOMAIN: INFRA

Submitted by: Piyush Sharma

Date: 22/10/2022

Architecture



2.1 Detailed Architecture

Data Description

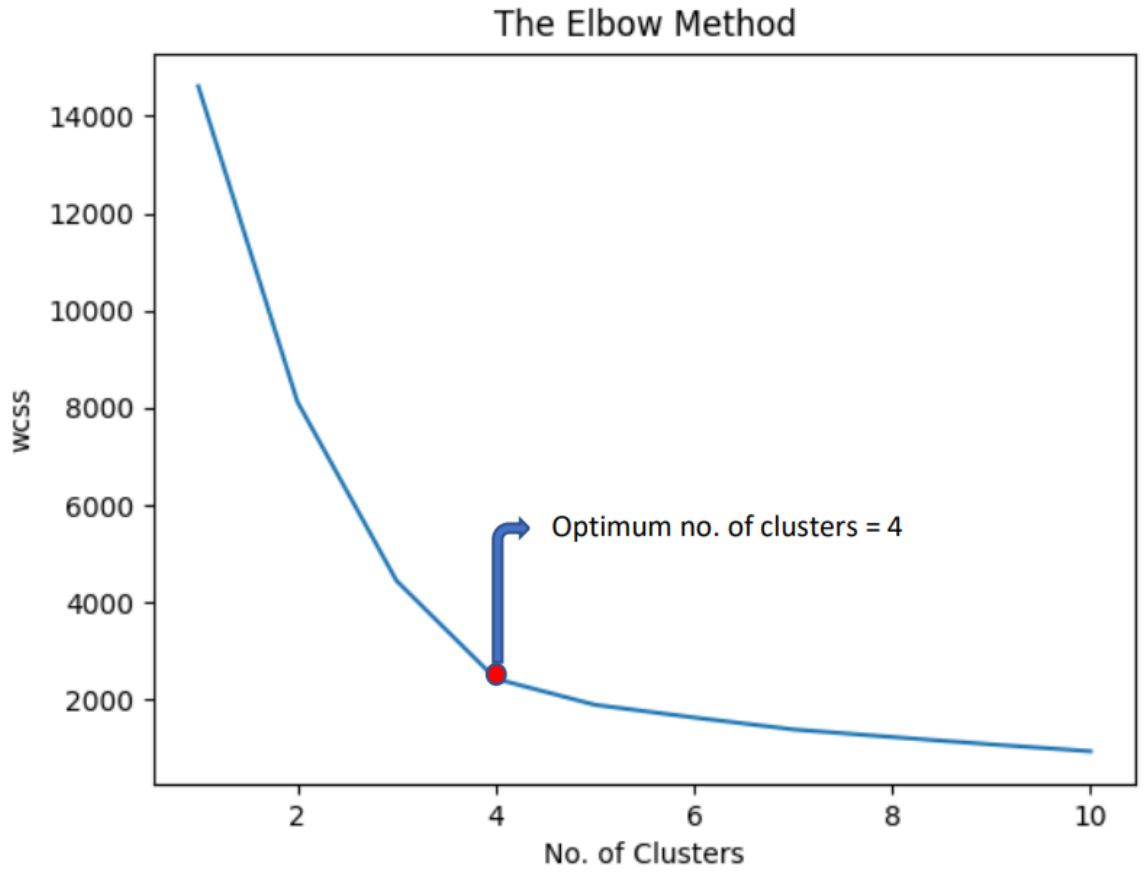
Name	Data Type	Measurement	Description
Cement (component 1)	quantitative	kg in a m3 mixture	Input Variable
Blast Furnace Slag (component 2)	quantitative	kg in a m3 mixture	Input Variable-- Blast furnace slag is a nonmetallic coproduct produced in the process. It consists primarily of silicates, aluminosilicates, and calcium-alumina-silicates
Fly Ash (component 3)	quantitative	kg in a m3 mixture	Input Variable- it is a coal combustion product that is composed of the particulates (fine particles of burned fuel) that are driven out of coal-fired boilers together with the flue gases.
Water (component 4)	quantitative	kg in a m3 mixture	Input Variable
Superplasticizer (component 5)	quantitative	kg in a m3 mixture	Input Variable-- Superplasticizers (SP's), also known as high range water reducers, are additives used in making high strength concrete. 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
Coarse Aggregate (component 6)	quantitative	kg in a m3 mixture	Input Variable-- construction aggregate, or simply "aggregate", is a broad category of coarse to medium grained particulate material used in construction, including sand, gravel, crushed stone, slag, recycled concrete and geosynthetic aggregates
Fine Aggregate (component 7)	quantitative	kg in a m3 mixture	Input Variable—Similar to coarse aggregate, the constitution is much finer.
Age	quantitative	Day (1~365)	Input Variable
Concrete compressive strength	quantitative	MPa	Output Variable

Data Preprocessing

- Check for null values
- Impute missing values using KNN Imputer
- Remove columns with zero standard deviation
- Transform features using log transformation
- Scale features using Standard Scaler

Data Clustering

- Kmeans clustering is used to create clusters of preprocessed data.
- The idea behind using clustering is to use different models on each cluster.
- The optimum number of clusters is determined by using the elbow plot's "Knee Locator" function
- After training Kmeans on preprocessed data the model is saved.



Model Building

- After clusters are created, we find the best model for each cluster.
- We are using two algorithms that are Random Forest Regressor and Linear Regression for each cluster.
- For each cluster, algorithms will be passed with the best parameters which are derived by GridSearchCV.
- We will calculate the r squared error score (r2 score) of each model and the model with the best r2 score will get selected.
- Similarly the models will be saved for each cluster that will be used in prediction.

Receiving data from users

- Users have to provide data in CSV format for training and prediction

Data Clustering

- The Kmeans model saved during training will be loaded again to predict the cluster of user's data.

Model call for specific cluster

- Based on cluster number, the respective model will be loaded and will be used to predict the data for that cluster.

Deployment

- We will deploy the model on Heroku cloud platform.