**Problem Statement**

To build a regression model to predict the concrete compressive strength based on the different features in the training data.

**Data Description**

Given is the variable name, variable type, the measurement unit and a brief description.

The concrete compressive strength is the regression problem. The order of this listing corresponds to the order of numerals along the rows of the database.

|  |  |  |  |
| --- | --- | --- | --- |
| 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](https://en.wikipedia.org/wiki/Coal_combustion_product) that is composed of the [particulates](https://en.wikipedia.org/wiki/Particulates) (fine particles of burned fuel) that are driven out of coal-fired [boilers](https://en.wikipedia.org/wiki/Boiler) together with the [flue gases](https://en.wikipedia.org/wiki/Flue_gas). |
| 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](https://en.wikipedia.org/wiki/Concrete) or [mortar](https://en.wikipedia.org/wiki/Mortar_(masonry)) allows the reduction of the water to cement ratio without negatively affecting the [workability](https://en.wikipedia.org/wiki/Workability) of the mixture, and enables the production of [self-consolidating concrete](https://en.wikipedia.org/wiki/Self-consolidating_concrete) and high performance concrete |
| Coarse Aggregate (component 6) | quantitative | kg in a m3 mixture | Input Variable-- construction aggregate, or simply "[aggregate](https://en.wikipedia.org/wiki/Aggregate_(composite))", is a broad category of coarse to medium grained particulate material used in [construction](https://en.wikipedia.org/wiki/Construction), including [sand](https://en.wikipedia.org/wiki/Sand), [gravel](https://en.wikipedia.org/wiki/Gravel), [crushed stone](https://en.wikipedia.org/wiki/Crushed_stone), [slag](https://en.wikipedia.org/wiki/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 |

**Model Training**

1) Data Export from Db - The data in a stored database is exported as a CSV file to be used for model training.

2) Data Preprocessing

4) Model Selection - We are using two algorithms, "Random forest Regressor" and “Linear Regression”. For each cluster, both the algorithms are passed with the best parameters derived from Grid Search. We calculate the R squared scores for both models and select the model with the best score. Similarly, the model is selected for each cluster. All the models for every cluster are saved for use in prediction.

5) Once the prediction is made for all the models, the predictions along with the original names before label encoder are saved in a CSV file at a given location and the location is returned to the client.