# CTC-34 Concrete Compressive Strength prediction

## Context:

The concrete compressive strength of concrete is a highly nonlinear function of age and ingredients.

Concrete is the most important material in civil engineering. The concrete compressive strength is a highly nonlinear function of age and ingredients. These ingredients include cement, blast furnace slag, fly ash, water, superplasticizer, coarse aggregate, and fine aggregate.

The actual concrete compressive strength (MPa) for a given mixture under a specific age (days) was determined from laboratory. Data is in raw form (not scaled).The data has 8 quantitative input variables, and 1 quantitative outputvariable, and 1030 instances (observations).

## Domain:

Material manufacturing

**Data reference:** https://www.kaggle.com/c/ctc-34-concrete-compressive-strength-prediction/overview

**Citation** :
NOTE: Reuse of this database is unlimited with retention of copyright notice for
Prof. I-Cheng Yeh and the following published paper:

I-Cheng Yeh, "Modeling of strength of high performance concrete using artificial
neural networks," Cement and Concrete Research, Vol. 28, No. 12, pp. 1797-1808 (1998)

## Attributes:

| Name | Data Type | Measurement | Description |
|---|---|---|---|
| Cement | Quantitative | Kg in a m^3 mixture | Input variable |
| Blast Furnace Slag | Quantitative | Kg in a m^3 mixture | Input variable |
| Fly ash | Quantitative | Kg in a m^3 mixture | Input variable |
| Water | Quantitative | Kg in a m^3 mixture | Input variable |
| Superplasticizer | Quantitative | Kg in a m^3 mixture | Input variable |
| Coarse Aggregate | Quantitative | Kg in a m^3 mixture | Input variable |
| Fine Aggregate | Quantitative | Kg in a m^3 mixture | Input variable |
| Age | Quantitative | Days | Input variable |
| | | | |
| Concrete compressive strength | Quantitative | MPa | Output Variable |

## Key asks:

- Build a model that helps map the input features to arrive at an appropriate representation of

the inputs to the compressive strength

- Model performance range at 95% confidence level

## Objective:

Modeling of strength of high performance concrete using Machine Learning

## Steps and tasks:

1. Deliverable -1 (Exploratory data quality report reflecting the following)
    a. Univariate analysis
        i. Univariate analysis – data types and description of the independent attributes which should include (name, meaning, range of values observed, central values (mean andmedian), standard deviation and quartiles, analysis of the body of distributions / tails, missing values, outliers
    b. Multivariate analysis
        i. Bi-variate analysis between the predictor variables and between the predictor variables and target column. Commenton your findings in terms of their relationship and degree of relation if any. Presence of leverage points. Visualize the analysis using boxplots and pair plots, histograms or density curves. Select the most appropriate attributes
    c. Strategies to address the different data challenges such as datapollution, outliers and missing values
2. Deliverable -2 (Feature Engineering techniques)
    a. Identify opportunities (if any) to create a composite feature, drop afeature
    b. Decide on complexity of the model, should it be simple linear mode in terms of parameters or would a quadratic or higher degreehelp
    c. Explore for gaussians. If data is likely to be a mix of gaussians, explore individual clusters and present your findings in terms ofthe independent attributes and their suitability to predict strength
3. Deliverable -3 (create the model )
    a. Obtain feature importance for the individual features usingmultiple methods and present your findings
4. Deliverable -4 (Tuning the model)
    a. Algorithms that you think will be suitable for this project
    b. Techniques employed to squeeze that extra performance out of themodel without making it overfit or underfit
    c. Model performance range at 95% confidence level