

## Cost Optimisation Methodology

→ The complete implementation of the model is done in google colab and the dataset is acquired from the internet.

Data set link :-

→ Import the required packages such as numpy, pandas, scikit learn. NumPy is a Python library used for working with arrays. It also has functions for working in the domain of linear algebra, fourier transform, and matrices. Pandas is used for playing with the data such as manipulating the data, analyzing the data, storing the data and many more. Scikit learn is a library which stores all the machine learning algorithms.

→ Read the data set which is in csv format with the help of pandas. The pandas package here converts the dataset into the dataframe.

→ Now, go on exploring the dataset with the help small functions such as

Shape, which gives the dimensions of the dataset

Head, this gives first five rows of the dataset by default

Tail, this gives the last five rows of the dataset by default

Columns, this gives all the column names in the dataset.

→ The columns in the provided dataset are 'SERIAL NO', 'HEIGHT', 'LENGTH OF SHEER WALL', 'COST OF SHEAR WALL', 'CONCRETE COST', 'STEEL COST'

→ Preprocessing :- This is one of the important techniques in model building. In this step the data is cleaned. The major thing which is done in this step is to remove the null values. The values which are empty are directly dropped from the dataset using the function dropna()

→ Now, the data is error free and insightful. refined data from the original columns, so the original columns can be dropped.

→ Now all the columns except the scour depth are considered in the  $x_1, x_2, x_3, \dots$  Variables. And the scour depth in y variable which is the predicted variable.

→ Here we have used linear regression whose math intuition is based on

$$y = mx + c$$

the x parameters keep increasing based on the columns available. The equation might seem this way.

$$y = m(x_1 + x_2 + x_3 + x_4 + x_5 + \dots) + c$$

→ The data split into 80% and 20%, 80 percent of the data is used in the training phase and 20 percent for the testing phase.

→ Linear regression is imported from the scikit learn library and the model is built, and also accuracy of the model is also measured using the metrics such as  $r^2$ , root mean squared error and mean squared error.

→ The individual values of the  $r^2 = 0.999$

$$\text{Mean squared error} = 4597.74$$

$$\text{Root mean squared error} = 67.80$$

