NAME: **SWATHI R**

Batch: **DS Jan 2023 Batch**

Project Name**: Chemical Property Modelling in association with asktalos.com and UltraTech Cement Limited and Aditya Birla Group**

Problem Statement: Build a predictive model to find the compression strength.

This part of the project will include steps for building a predictive model to find the compression strength with highest efficiency possible.

With the dataset given, the most important aspect in building a project will be thorough analysis of the dataset.

**DATASET**

Dataset name : Material Compressive Strength Experimental Data

Number of columns : 7

Number of rows : 6139

Total number of attributes : 9

**SIMPLIFIED DETAILS OF DATASET**

| **Sl no** | **Attributes names** | **Data type** | **Type of data** | **Null values** |
| --- | --- | --- | --- | --- |
|  | Material Quantity (gm) | float | continuous | yes |
|  | Additive Catalyst (gm) | float | continuous | yes |
|  | Ash Component (gm) | float | continuous | yes |
|  | Water Mix (ml) | float | continuous | yes |
|  | Plasticizer (gm) | float | continuous | yes |
|  | Moderate Aggregator | float | continuous | yes |
|  | Refined Aggregator | float | continuous | yes |
|  | Formulation Duration (hrs) | float | continuous | yes |
|  | Compression Strength MPa | float | continuous | no |

**ATTRIBUTES**

1. Material Quantity (gm)
2. Additive Catalyst (gm)
3. Ash Component (gm)
4. Water Mix (ml)
5. Plasticizer (gm)
6. Moderate Aggregator
7. Refined Aggregator
8. Formulation Duration (hrs)
9. Compression Strength MPa

**STUDY OF ATTRIBUTES**

Material Quantity (gm)

Increasing cement content typically results in an **increase** in compressive strength.

Additive Catalyst (gm)

The use of the superplasticizer additive resulted in **higher** compressive strength.

Ash Component (gm)

Fly ash in concrete **reduces** the rate of hydration thus low early compressive strength is observed.

Water Mix (ml)

Concrete's durability and compressive strength are **decreased** due to increased dampness levels.

Plasticizer (gm)

The addition of plasticized admixture is able to reduce the water content of concrete to up to 25% whilst **doubling** the compressive strength compared to traditional concrete on day one.

Moderate Aggregator

In high-strength concretes, higher strength coarse aggregates typically yield **higher** compressive strengths, while in normal-strength concretes, coarse aggregate strength has little effect on compressive strength.

Refined Aggregator

The better the physical properties of the fine aggregate, the higher the workability of the concrete and specific gravity of the fine aggregate, and also the **higher** the compressive strength of the concrete.

Formulation Duration (hrs)

Longer durations can generally lead to **increased** compressive strength, there is a point of diminishing returns. Overly prolonged processes might not provide significant additional strength. Thus, requires threshold.

**DEPENDENT AND INDEPENDENT VARIABLE**

* The independent variable is what you change, the cause and Its value is independent of other variables in your study.
* The dependent variable is what changes because of that, the effect.

On examining the dataset attributes, we can conclude the following as Independent and dependent variables.

Independent varibales

1. Material Quantity (gm)
2. Additive Catalyst (gm)
3. Ash Component (gm)
4. Water Mix (ml)
5. Plasticizer (gm)
6. Moderate Aggregator
7. Refined Aggregator
8. Formulation Duration (hrs)

Dependent variables

1. Compression Strength MPa

**OBSERVATION - 1**

Almost all of the independent variables have missing values.

**SOLUTION**

The missing values present in the Independent variable columns can handled more than 1 way accordingly.

Dropping the missing value column

If there are way too many missing values in a column then we can drop that column.

Handling by mean, median and mode.

We can impute missing values with mean, median and mode.

Filling the values using Pandas function

Some functions that can be used in pandas for handling missing values are the

* Dropna
* Fillna
* Bfill
* Interpolate.

**OBSERVATION – 2**

The dataset is continuous ,its a case of regression problem.

**OBSERVATION – 3**

We are dealing with more than one Independent variable that has effect on our dependent variable. This is a Multilinear Regression Problem.

**MODEL PROPOSED**

Linear model.

**ALGORITHM**

Multilinear Regression Machine Learning Algorithm.