**Concrete Analysis**

**Methodology:**

To start the project, one must first have a basic understanding of the constituents of concrete. Concrete strength is highly contingent on several components. These components include “Water, aggregate, and Portland cement” and should all be analyzed for maximum efficiency (Badole, 2021). In addition, the quality of these 3 components is important to understand how the cement will behave under certain conditions. To solve these real-world problems involving concrete, one can use machine learning techniques. Therefore, incorporating data in machine learning techniques will help us in determining the relationship between different components.

In this model, we imported several modules to make an effective model. The modules we made use of were numpy, pandas, statistics, matplotlib, scipy, seaborn and sklearn. We also imported specific functions from those modules to help us analyze and interpret the data. Next, we needed a spreadsheet containing the data relevant to our project. We imported said spreadsheet as a data frame and renamed the columns in order to be able to reference them more easily in our later code where we implement the machine learning technique. We can then extract information from the spreadsheet using Data.info() in order to obtain the relevant data types in this spreadsheet. Then using Data.describe() allows us to find any potential null values to clean the datasheet.We made use of Data=Data.dropna() to drop all null values in the data to make the model more accurate. The use of a pair plot produces a pairwise relationship within the dataset we have where “it will create a grid of axis where the y-axis belongs to row and the x-axis belongs to columns,” and this produces various graphs that allow for an effective model (Badole, 2021). When observing the pair plot, it becomes apparent that a high correlation between features appears. A feature correlation heatmap is a good way of visualizing how the different variables relate and how we can start building our model. Furthermore, we can further confirm the trend of the correlations by plotting the Pearson Correlation coefficients between individual variables. We also made a box plot to plot the mean value of each category in relation to the other categories to be able to relate them more precisely.

Next, the use of a scatter plot displays the relationship between two sets of data which allows us to get a more accurate version of the model, and it may appear that increasing or decreasing the amount of water or cement in concrete increases the quality of concrete. Three observations can be made. First, there appears to be a high correlation between cement and compressive strength further increasing strength. Compressive strength has a strong and high correlation between super plasticizer and age. Finally, water presents a high correlation with superplasticizer, and there is a positive correlation between fine aggregate and fly ash.

Next, below the Decision trees header, we split the data into train and test splits, using x and y variables to indicate different things, and proceeded to build the model using our linear regression process and the Decision Tree Regressor function from sklearn. We then plotted the prediction data against the true data on a scatter plot with a line of best fit to create a visual of the model.

In conclusion, we replicated the process using the Random Forest Regressor function from the sklearn library. Again, we plotted the predicted and true data to visualize and interpret the output. We decided to use the Random Forest Regressor because it has the lowest R^2 and RMSE values.

**Literature Research:**

Concrete is a basic component in development seeing that it can endure outrageous climate, demonstrating that substantial designs can be solid and versatile when made with the right combination. The technique for planning the substantial blend is trying because of numerous variables that add to its solidness and strength. A portion of the primary elements incorporate the proportion of fixings like water, concrete, added substances, and admixtures. These variables will influence the compressive strength which is basically a proportion of the substantial capacity to get through loads without cracking.

Compressive strength is a critical estimation on the grounds that the compressive strength makes an establishment that prompts no problem at all foundations. Observing the right extent of fixings is critical to deliver the expected strength, in any case, the testing can be a long-drawn-out system that can experience a few mistakes. An answer for these issues is making information models that will lessen the quantity of trials that should be tried. Developing information models is more customary for foreseeing the strength of substantial given that it will diminish how much time that it takes to test the strength while additionally working on the precision of the extent of fixings