### **Programming for AI**

### **(AL2001)**

### **Project Proposal**

### 

### **Concrete Slump Dataset**

### 

### **Group Members:**

### Alisha Zaidi [23K-0025]

Laiba Bint-e-Zia[23K-0006]

### 

### 

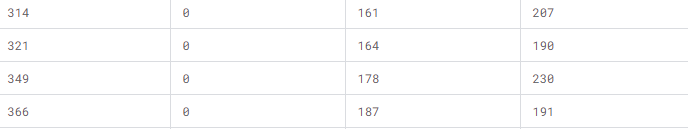
### 

### 

### **Dataset Overview**

The dataset that is been working on is the *Concrete Slump* dataset from Kaggle. It contains **103 rows** and **10 columns** of data. Each row represents a different concrete mixture and slump test result. The columns include inputs like cement amount, water, fine and coarse aggregates, and outputs like compressive strength and slump values.

### **Data Cleaning and Preparation**

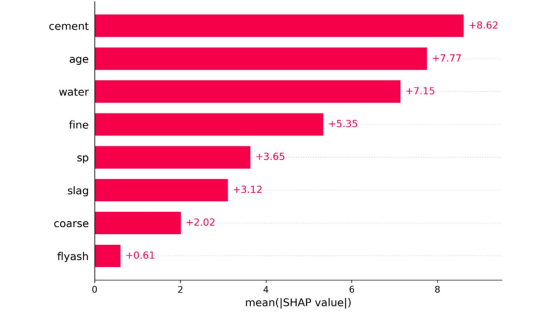
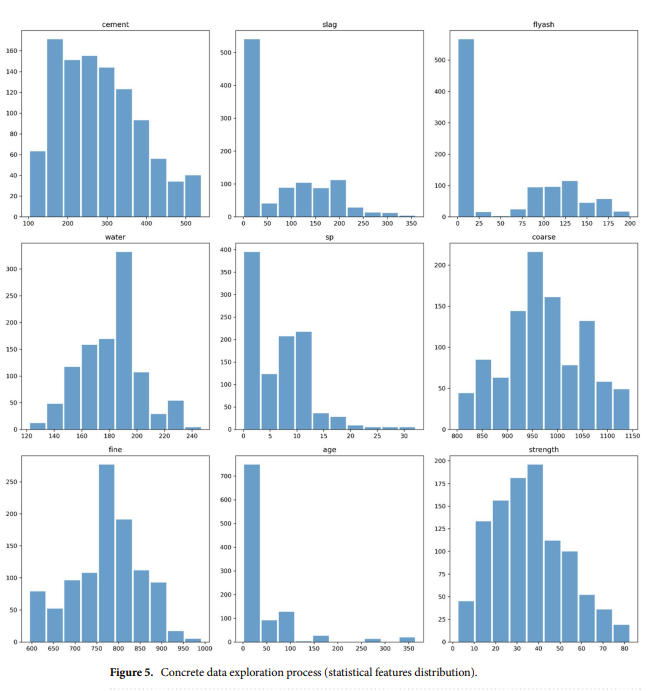


To ensure the dataset is clean, Pandas will be used to:

* Fill in any missing values.
* Convert string data to numeric types where necessary.
* Ensure no invalid values remain (e.g., negative values where they don’t make sense).

In case the data was already clean, some intentional missing values will be introduced and inconsistencies to demonstrate data cleaning techniques.

### **Statistical Insights and Visualization**



Using **Numpy**, **Pandas,** **Matplotlib** and **Seaborn**, to extract key statistical information, including:

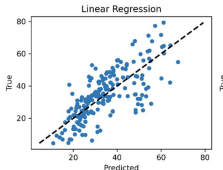
* **Mean, Median, Mode**: These will provide insight into the average trends in materials used.
* **Standard Deviation and Variance**: These measures help understand the variability in the dataset.
* **Correlation Matrix**: This will show how different ingredients correlate with the slump value and compressive strength.

Visualizations:

* **Histograms** for distribution of materials like cement, water, etc.
* **Box plots** to observe outliers in slump values.
* **Heatmaps** for correlations between different variables.

*Graph Example*: A heatmap of the correlation matrix will clearly show relationships between input materials and the concrete slump value. This visualization helps in identifying which materials have the most impact on the final concrete mix properties.

### **Machine Learning Model**



A **Linear Regression** model will be applied using the **Scikit-learn** library. The model will predict the slump value based on input features like cement, water, and aggregate amounts. the model will be evaluated using metrics like Mean Squared Error (MSE) and R-squared (R²).

**User Interface**

A simple UI will be developed using **Tkinter**, where users can input different material proportions and predict the slump value. Alternatively, a web version can be built using **Flask** or **Django**.