**Title: Project DeepCrete: Predicting Concrete Compressive Strength Using Artificial Neural Networks**

* **Author(s):** Your Name
* **Date:** September 22, 2025

**Abstract**

A brief, one-paragraph summary of the entire project. State the problem (predicting concrete strength), the method used (a neural network), the key results (e.g., "The final model achieved an R-squared value of 0.92 on the test set..."), and the conclusion.

**1. Introduction**

* **1.1. Background:** Briefly explain the importance of concrete in construction and why accurately predicting its compressive strength is crucial for safety, cost-efficiency, and quality control.
* **1.2. Problem Statement:** Clearly state the goal: to develop a machine learning model to predict the compressive strength of concrete based on its composition and age.
* **1.3. Objectives:** List the specific project goals (e.g., preprocess the dataset, build and train a neural network, evaluate its performance).

**2**Of course. Here is the text formatted for you to copy. The parts you need to fill in yourself are in **bold text**—simply delete the bolded instructions and insert your own content.

### 2.1. Dataset Description

The **UCI Concrete Compressive Strength dataset** is a collection of data used to predict the compressive strength of concrete. The dataset was donated to the UCI Machine Learning Repository by Professor I-Cheng Yeh. It contains **1030 samples** with 8 input features and 1 target variable.

The input features and the target variable are listed below:

* **Input Features**:
  + Cement (kg/m³)
  + Blast Furnace Slag (kg/m³)
  + Fly Ash (kg/m³)
  + Water (kg/m³)
  + Superplasticizer (kg/m³)
  + Coarse Aggregate (kg/m³)
  + Fine Aggregate (kg/m³)
  + Age (days)
* **Target Variable**:
  + Concrete Compressive Strength (MPa)

### 2.2. Data Analysis

## summary Statistics of Concrete Data

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | cement | slag | Fly ash | water | superplasticizer | Coarse  aggregate | Fine  aggregate | age | Cs  MPa |
| **count** | 1030 | 1030 | 1030 | 1030 | 1030 | 1030 | 1030 | 1030 | 1030 |
| **mean** | 281.17 | 73.90 | 54.19 | 181.57 | 6.20 | 972.92 | 773.58 | 45.66 | 35.82 |
| **std** | 104.51 | 86.28 | 64.00 | 21.35 | 5.97 | 77.75 | 80.18 | 63.17 | 16.71 |
| **min** | 102.00 | 0.00 | 0.00 | 121.80 | 0.00 | 801.00 | 594.00 | 1.00 | 2.33 |
| **25%** | 192.38 | 0.00 | 0.00 | 164.90 | 0.00 | 932.00 | 730.95 | 7.00 | 23.71 |
| **50%** | 272.90 | 22.00 | 0.00 | 185.00 | 6.40 | 968.00 | 779.50 | 28.00 | 34.45 |
| **75%** | 350.00 | 142.95 | 118.30 | 192.00 | 10.20 | 1029.40 | 824.00 | 56.00 | 46.14 |
| **max** | 540.00 | 359.40 | 200.10 | 247.00 | 32.20 | 1145.00 | 992.60 | 365.00 | 82.60 |

## Key Findings from Exploratory Data Analysis

The initial descriptive statistics reveal several important characteristics of the dataset, which consists of **1,030 complete observations** with no missing values across any of the features.

### Distribution of the Target Variable (csMPa)

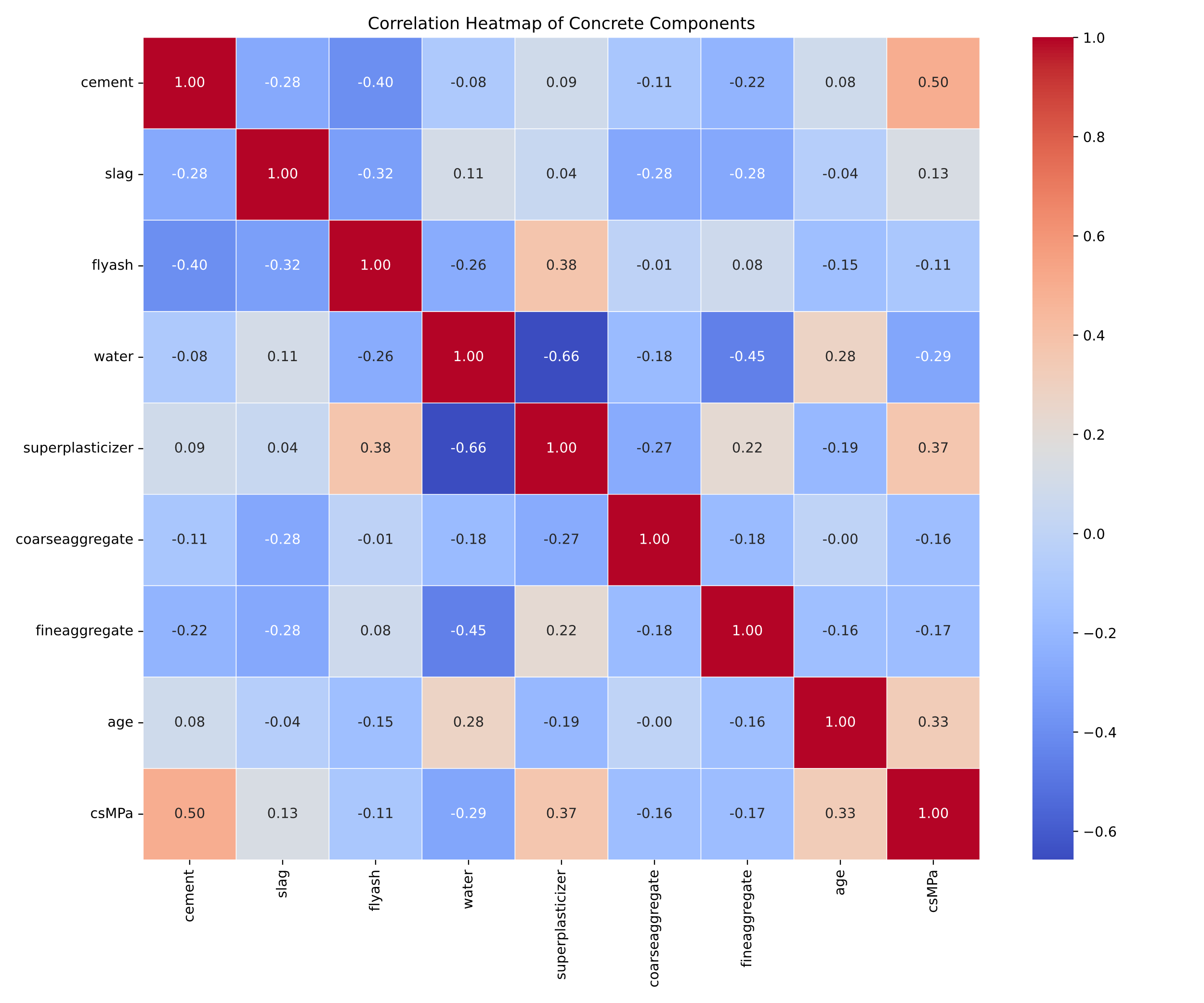
The target variable, concrete compressive strength (csMPa), shows a wide distribution, ranging from a low of **2.33 MPa** to a high of **82.60 MPa**. The **mean strength is 35.82 MPa**, which is very close to the **median of 34.45 MPa**, suggesting that the distribution of concrete strength is nearly symmetrical with a slight right skew. The high standard deviation (16.71 MPa) confirms the significant variability in the strength outcomes, indicating a diverse mix of concrete formulas and curing times.

### Characteristics of Predictor Variables

* **Prevalence of Additives:** A key finding is that several ingredients are not used in all mixtures. **Flyash**, **slag**, and **superplasticizer** all have minimum and 25th percentile values of 0. In fact, the median for flyash is also 0, indicating that **over half of the concrete samples in this dataset do not contain flyash**. This suggests that the presence or absence of these additives will be a critical factor in determining strength.
* **Age Skewness:** The age of the concrete, a critical factor for strength, is heavily right-skewed. While the median age is **28 days** (a standard curing time for testing), the mean is significantly higher at **45.66 days**, pulled by a long tail of older samples up to **365 days**. This wide range and skewness suggest that age will be a highly influential predictor.
* **Core Ingredients:** Core components like cement, water, coarseaggregate, and fineaggregate are present in all samples and exhibit distributions that are more centered than the additives. Cement content shows the most significant variation among these core ingredients, with a standard deviation of 104.51.

In summary, the primary drivers of variability in concrete strength are likely to be the **age** of the concrete and the **use (or non-use) of supplementary materials** like slag, flyash, and superplasticizer. The wide range and skewed nature of these features will be important considerations for feature engineering and model selection.

#### Correlation Heatmap



### Discussion of Relationships

The correlation heatmap quantifies the linear relationships between the concrete components, its age, and its final compressive strength (csMPa). The results highlight several key relationships that are crucial for understanding and predicting the material's behavior.

#### Key Predictors of Compressive Strength

The most influential factors on concrete strength are clearly identifiable:

* **Cement:** As expected, **cement** has the strongest positive correlation with compressive strength, with a coefficient of **+0.50**. This confirms its role as the primary binding agent responsible for strength development.
* **Superplasticizer and Age:** Following cement, **superplasticizer** shows a notable positive correlation of **+0.37**, and **age** has a similar positive correlation of **+0.33**. The superplasticizer's role in reducing water content and the ongoing curing process over time are significant contributors to higher strength.
* **Water:** The amount of **water** has the most significant negative relationship with strength, with a correlation of **-0.29**. This underscores the critical importance of the water-to-cement ratio; less water leads to a denser, stronger concrete.

#### Weaker Relationships

The remaining components show a much weaker linear relationship with the final strength:

* The aggregates (**coarseaggregate** at **-0.16** and **fineaggregate** at **-0.17**) have a slight negative correlation.
* Interestingly, **slag** (+0.13) and **flyash** (-0.11) show very little correlation. The weak negative correlation for flyash may seem counter-intuitive but could be explained by its frequent use as a *substitute* for the more influential cement.

#### Inter-Feature Correlations

The heatmap also reveals important relationships between the input variables themselves:

* The strongest correlation in the entire matrix is between **superplasticizer** and **water**, with a coefficient of **-0.66**. This makes perfect sense, as the function of a superplasticizer is to increase workability, thus allowing for a significant reduction in water content.
* Moderate negative correlations exist between **cement** and **flyash (-0.40)** and between **cement** and **slag (-0.28)**, which supports the fact that these materials are often used to replace a portion of the cement in a mix.

## 3. Methodology ⚙️

### 3.1. Data Preprocessing

To prepare the data for training the neural network, the following steps were taken:

* **Train-Validation-Test Split**: The dataset was split into three sets:
  + **Training set**: 70%
  + **Validation set**: 15%
  + **Test set**: 15%
* **Feature Scaling**: **StandardScaler** was used to scale the input features. This is a necessary step because the features have different units and scales. StandardScaler transforms the data to have a mean of 0 and a standard deviation of 1, which helps the neural network learn more effectively and converge faster.

### 3.2. Model Architecture

A sequential neural network was designed with the following structure:

| Layer | Number of Neurons | Activation Function |
| --- | --- | --- |
| Input Layer | 8 | - |
| Hidden Layer 1 | 64 | ReLU |
| Hidden Layer 2 | 32 | ReLU |
| Output Layer | 1 | Linear |

### 3.3. Training Process

The model was trained using the following configuration:

* **Optimizer**: **Adam** was chosen as the optimizer due to its efficiency and adaptability.
* **Loss Function**: **Mean Squared Error (MSE)** was used as the loss function, which is a standard choice for regression problems.
* **Metrics**: **Mean Absolute Error (MAE)** was monitored during training to provide a more interpretable measure of the model's performance.
* **Batch Size**: **[Specify the batch size you used, e.g., 32]**
* **Number of Epochs**: **[Specify the number of epochs you trained for, e.g., 100]**

**4. Results and Discussion**

* **4.1. Model Performance:** Present the final performance of your model on the **unseen test data**. A clean table showing the final MSE, MAE, and R2 score is essential.
* **4.2. Analysis of Results:** Discuss what these results mean. Is an MAE of 4 MPa good or bad in a real-world context? How well did your model perform based on the R2 score?
* **4.3. Prediction Examples:** You can show a plot of the model's predicted values versus the actual true values for the test set. A perfect model would have all points lying on a 45-degree line.

**5. Conclusion**

* **5.1. Summary of Work:** Briefly summarize what you accomplished.
* **5.2. Final Conclusion:** State the main takeaway. For example, "This project successfully demonstrated that an artificial neural network can be an effective tool for predicting concrete compressive strength with a high degree of accuracy..."
* **5.3. Future Work:** Suggest potential improvements. Could you use more data? Try other models (like Gradient Boosting)? Build a simple web application for users to get predictions?

Good luck with your project! It's a fantastic way to learn and apply neural networks.

Data set site: https://www.kaggle.com/datasets/maajdl/yeh-concret-data?resource=download