

Predicting Compressive Strength of Concrete using Machine learning

Authors

- **Andrew Bushnell**

-  [andrewb7777](#)

Department of CEE, University of Illinois

- **Kanchan Kulhalli**

-  [Kanchan-uiuc](#)

Department of CEE, University of Illinois

- **Vikram Gadge**

-  [vgadge2](#)

Department of CEE, University of Illinois

Abstract

This project aims to create a machine learning model that will be able to predict which components and the combinations of these components will work best based on strength requirements of certain structures and predict maximum allowable loads that can be achieved based on the mixtures. We also plan to create a model to predict the most cost effective concrete mixture.

The dataset on “Concrete Compressive Strength” was obtained using Kaggle.com [1]. The data comes in a CSV format and contains 9 columns and a total of 1032 rows. The next section gives a detailed information on the dataset.

As the first step, we have done an extensive exploratory analysis on our dataset. We have studied the effects of all the different columns(the independent variables) on concrete compressive strength(the dependent variable). Most of the plots have been quite intuitive while some of the plot haven't provided much information on how exactly they impact the concrete compressive strength.

The final section contains an overview on how we plan on generating the machine learning models to predict the concrete compressive strengths based on different mixture composition.

Exploratory Data Analysis

In this section, we aim to explain the various components in our dataset and its effect on the concrete compressive strength.

The data set is composed of nine columns of data that state the following information: Fly Ash, Water, Superplasticizer, Coarse Aggregate, Age, and Concrete Compressive Strength. These columns have the following units of measurements: the first 7 columns have the units kg in m³ mixture, 8th column in days and the 9th column in MPa megapascals. The data set[1] is in a CSV format and has a total of 1030 rows. We found a few discrepancies in the data set and hence we had to perform some data cleaning tasks before doing any exploratory analysis. The below section describes our data cleaning process in detail.

Data Cleaning

The dataset that we selected from Kaggle comprised of rows that had rows that were repeated multiple times. To remove the redundancy, we used the `unique()` function to remove the duplicate rows. The number of rows reduced to 1005 after this operation.

The other issue we found out in our dataset was that the compressive strength was different despite all the factors affecting it i.e., all the 8 columns affecting it were the same. We combined those rows into a single row by taking a mean of the compressive strength. After this operation, the number of rows in our dataset reduced to 992.

We also observed that one of the columns i.e., fine aggregate had an extra space in its name and we had to remove the extra space using the `rename()` method to make the column access easier.

```
begin
df_orig = CSV.read("/Users/kanchankulhalli/Documents/CEE 492 - Data
                  Science/concrete_data.csv", DataFrame)
df_uniq = unique(df_orig)
rename!(df_uniq, "fine_aggregate " => :fine_aggregate)
df_group = groupby(df_uniq, [:cement, :blast_furnace_slag, :fly_ash, :water,
                           :superplasticizer, :coarse_aggregate, :fine_aggregate, :age])
df = combine(df_group, :concrete_compressive_strength => mean)
rename!(df, :concrete_compressive_strength_mean =>
        :concrete_compressive_strength)
end
```

Description of the Dataset

Now that we cleaned our dataset, we set out to ask some interesting questions by studying each column and its effect on concrete compressive strength but before we did that, we generated Table [1](#) to get a general understanding of the columns in our dataset.

Table 1: Ranges and statistics of the columns in our dataset.

| | variable | min | mean | median | max |
|---|-------------------------------------|-------|---------|---------|--------|
| | Symbol | Real | Float64 | Float64 | Real |
| 1 | :cement(kg per m3) | 102.0 | 276.873 | 259.95 | 540.0 |
| 2 | :blast_furnace_slag(kg per m3) | 1.0 | 73.0007 | 20.0 | 359.4 |
| 3 | :fly_ash(kg per m3) | 1.0 | 55.6028 | 1.0 | 200.1 |
| 4 | :water(kg per m3) | 121.8 | 182.368 | 185.7 | 247.0 |
| 5 | :superplasticizer(kg per m3) | 1.0 | 6.34415 | 6.0 | 32.2 |
| 6 | :coarse_aggregate(kg per m3) | 801.0 | 974.597 | 968.0 | 1145.0 |
| 7 | :fine_aggregate(kg per m3) | 594.0 | 773.081 | 780.0 | 992.6 |
| 8 | :age(days) | 1 | 46.1663 | 28.0 | 365 |
| 9 | :concrete_compressive_strength(MPa) | 2.33 | 35.119 | 33.73 | 82.6 |

Water and Cement are the two most basic ingredients of concrete. The strength of the concrete mixture is heavily influenced by the proportions of these ingredients. We decided to take a look at how the water/cement ratio affects the strength of the concrete.

Water and Cement

The Abram's water-to-cement ratio (w/c) pronouncement of 1918 has been described as the most useful and significant advancement in the history of concrete technology. The generally accepted Abrams rule is a formulation of the observation that an increase in the w/c decreases the strength [2].

We decided to check how the Abrams' law holds up for our dataset considering the fact that there are several other ingredients that affect the strength of the concrete. Figure 1 plots the water cement ratio versus the concrete compressive strength.

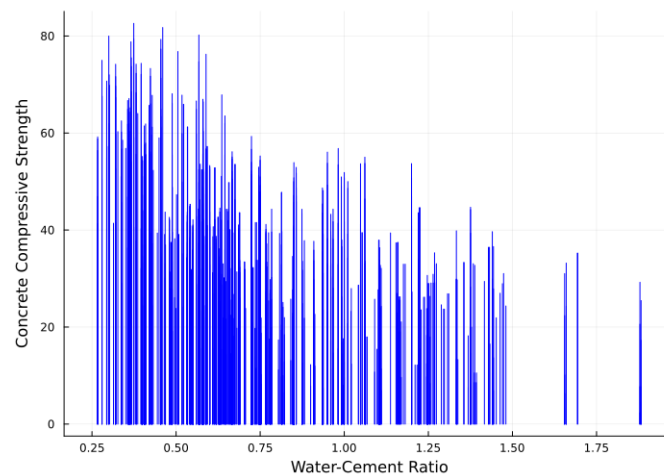


Figure 1: Water-Cement Ratio vs Compressive strength of Concrete

We can observe that the law holds quite good from a general perspective i.e., the compressive strength decreases with the increase in the w/c ratio but doesn't hold true on a case by case basis. This is understandable since there are several factors that also influence the strength.

In the below sub sections, we try to analyze the effects of various other ingredients on the compressive strength of concrete.

Superplasticizer

Superplasticizers are powerful water reducers that enable an increase in the ultimate stress of concrete by decreasing the w/c ratio, a decrease in the cement content while maintaining the same range of strength or workability, an increase in concrete compacity, and other effects [3].

An important component of most modern concrete mixes, water reducers improve the workability of wet concrete while decreasing the amount of water used in the mix. Superplasticizers decrease the water-cement ratio while providing such benefits as increased density and improved bond strength.

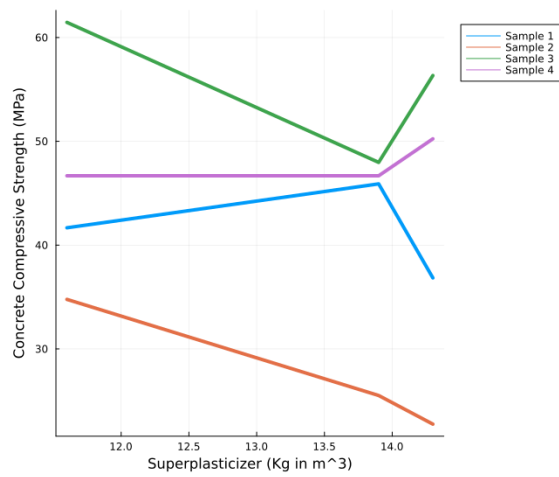


Figure 2: Superplasticizer vs Compressive strength of Concrete

To study the effect of superplasticizer in isolation, we kept all the other columns constant and compared the varying superplasticizer with the concrete compressive strength. Figure 2 shows the line plots for different samples. It is not clear how the superplasticizer is affecting the strength. We looked through a number of articles and research papers and we found out that the effect of superplasticizers is more pronounced when the ratio of superplasticizer to cement is between 0.5% to 3% [4]. To verify this, we plotted Figure 3 with two samples, the first one has a superplasticizer/cement(s/c) ratio equal to 0.2% and the second one has s/c ratio equal to 2%.

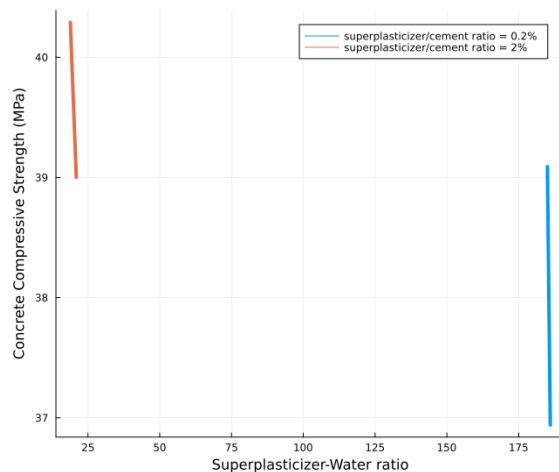


Figure 3: Superplasticizer to Water ratio vs Compressive strength of Concrete

We can see that a decrease in the superplasticizer to water ratio results in decrease in the concrete compressive strength. This suggests a positive correlation between the superplasticizer and the concrete strength. However, for ratios greater than 3%, we did not find any kind of correlation between these two suggesting that the effect of superplasticizer depends on the s/c ratio.

Fly Ash

In recent decades, fly ash has become an increasing common component used in concrete mixtures. Fly ash is used to increase the workability of plastic concretes along with increase the strength and durability of regular concretes [5]. Fly ash can also be used to replace a portion of the amount of the cement mixture needed. This in return reduces the cost while not decreasing the strength.

Using our data set, we first began by organizing the data set to only include unique values of fly ash since there were many rows that had no fly ash component in their concrete mixture. Then using this data set, we created a histogram to show the general trendline of the impact that the fly ash

component has on the compressive strength. We can see from Figure 4 that using fly ash does increase the strength of the concrete but peaks at approximately 20 kg per m³ mixture and then decreases as more fly ash is used. This is mostly due to fly ash not being suitable to replace fully or the most of the cement used in the mixture and is used more as a binding unit that helps with the workability and durability to a certain degree.

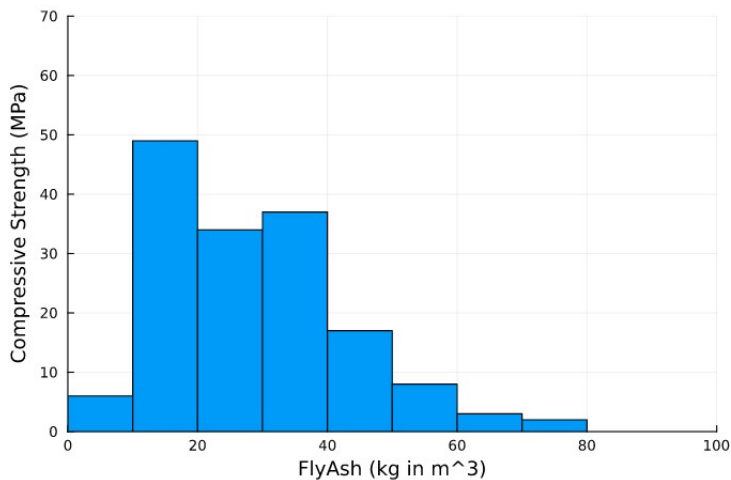


Figure 4: Fly Ash vs Concrete Compressive Strength

Coarse Aggregate and Fine aggregate

Aggregates are inert granular materials such as sand, gravel, or crushed stone that, along with water and Portland cement, are an essential ingredient in concrete. Fine aggregates generally consist of natural sand or crushed stone with most particles passing through a 3/8-inch sieve. Coarse aggregates are any particles greater than 0.19 inch, but generally range between 3/8 and 1.5 inches in diameter [6].

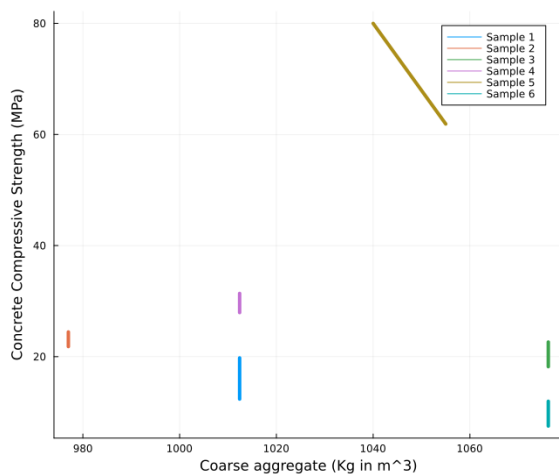


Figure 5: Coarse Aggregate vs Concrete Compressive Strength

We can observe from Figure 5 that there is no positive correlation between the coarse aggregate and the concrete compressive strength. We couldn't really understand how to interpret the graph. Its possible that it's effects weren't pronounced in the samples that we have taken. We need a more refined model to understand the effects of this property.

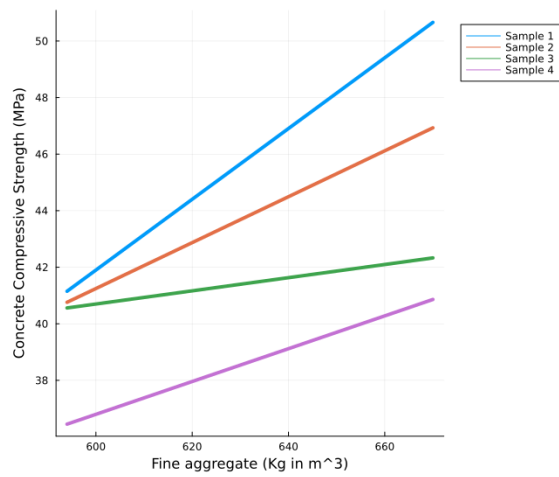


Figure 6: Fine Aggregate vs Concrete Compressive Strength

Figure 6 shows a positive correlation between the fine aggregate and the concrete compressive strength which is as expected. For both the graphs, we have kept all the columns constant except for the column that we are analyzing.

Age

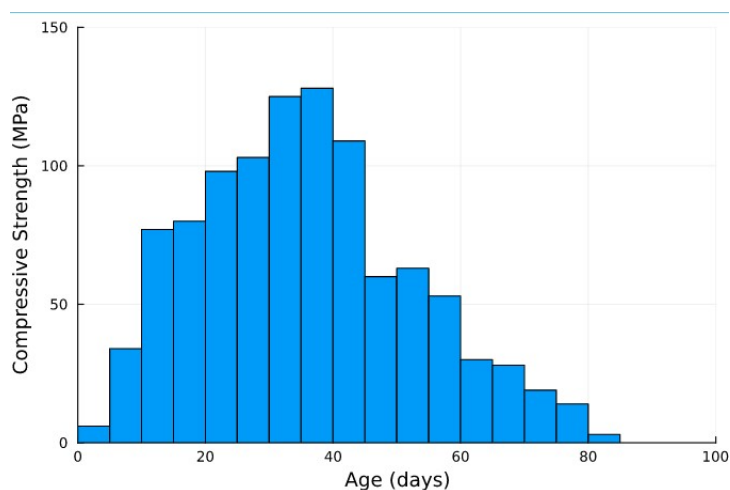


Figure 7: Age vs Concrete Compressive Strength

Figure 7 shows a histogram plot of Age vs Concrete Compressive Strength where age is the number of days after the concrete has been placed, we see that as concrete age increases the compressive strength increases until it reaches a peak at approximately 28 days and then gradually decreases in strength as age increases due to wear and tear of the concrete material. From [7], we understood that the concrete requires a curing time where once the concrete is placed it needs time to cure which is where the water content in the concrete mixture evaporates, leading to the concrete to settle and harden. This in return leads to increase in the concrete compressive strength. Based on this information and looking at the data set, to have the concrete mixture result in the strongest compressive strength, we want our age to be around the 28-day mark which is the ideal curing time.

Blast Furnace Slag

Concretes containing slag as a partial replacement of cement (up to 40%) had higher compressive and flexural strengths casting and curing at +42°C than those of concretes made with Portland cement alone[8].

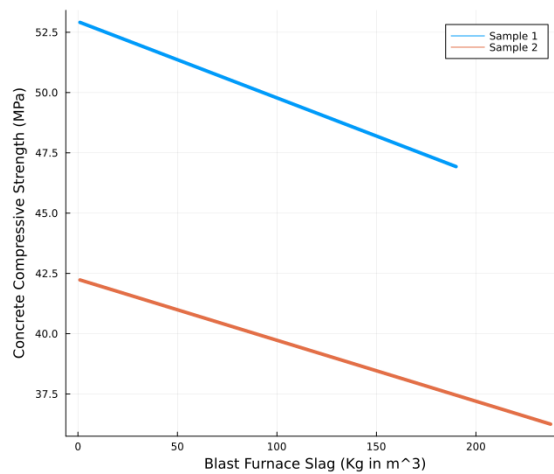


Figure 8: Blast Furnace Slag vs Compressive strength of Concrete

However, Figure 8 shows a negative correlation where the compressive strength of the concrete is decreasing with the increase in the Blast furnace slag content. We couldn't really figure out how to interpret this graph and we believe that this is similar to the case we explained for coarse aggregate.

Concrete Compressive Strength

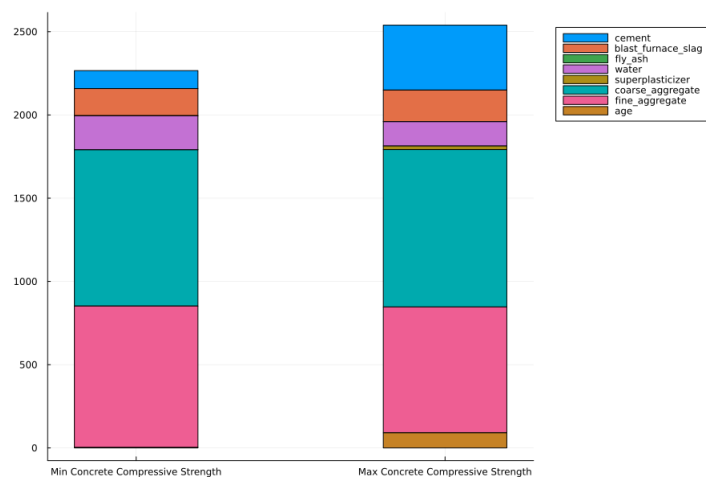


Figure 9: Proportions of the components at minimum and maximum concrete compressive strengths

The graph above explains how the proportions of the 8 components are affecting the concrete compressive strength. It clearly shows that the second bar graph has more quantity of superplasticizer and less quantity of water than the first bar graph. Also, the second bar graph has good aging time than the first one. Hence clearly the second plot has more concrete compressive strength than the first one.

Predictive Modeling

Using our data set we can create a machine learning model to produce a predictive modeling code for solving for which combinations of concrete mixtures would be ideal to meet a certain strength requirement, based on different construction projects, and then find the minimum optimal cost. To do this we intend to first use the data available in our dataset to find out the unit costs of each component in our concrete. We then will look up research papers over the different strength requirements set in place at the State and Federal level for different construction projects, such as bridges and highways where highway bridges require a minimum of 25 MPa to pass the strength requirements set in place in California [9].

As a first step towards create a feasible model, we plan to use regression analysis. The reason is that we have 8 independent variables (the 8 columns described in detail in the previous section) affecting the single dependent variable which is the concrete compressive strength. Depending on the results of the model, we plan to explore more advanced machine learning models. We will create rough estimations of how much of each concrete component would need to be used to create the optimal combination for each project. This would be done through the creation of a model of ideal solutions to meet the lowest price and still meet the strength requirements for each unique project. This would be very useful in the construction industry which would be able to use our machine learning models to evaluate which combination of concrete would best work to meet the requirements of their project while also saving the construction company the most capital.

Based on what we have discussed with the TA, this is a feasible idea since our data could be used to create combinations that are not currently in our dataset by using what we learned in class to create rough estimates of new combinations based on the current data. For example, we could find the new compressive strength of a certain new combination to be the average strength of two known combinations from our data set or use certain criteria found online, such as impact of water to cement ratio or fly ash to concrete ratio, to make a rough estimate of the new combinations results compressive strength. We intend to do this by creating something like the solver function in excel where we will have an objective function, such as minimize price and maximize strength, and set up constraints, such as have strength be greater than or equal to 25 MPa or have at least 20% cement. Then solving this would give us our result of a new matrix of the ideal values for the combinations in the concrete to be used for that specific project. # Preliminary Predictive Modeling

In this section, we will explore multiple methods for creating drafts for our predictive modeling analysis. After reviewing the potential examples that could be used, we decided to use clustering (k-means) and ... to get our initial drafts. With advice from the profesor we also will attempt to create a constrained optimization using "JuMP" in Julia.

K-Means

To begin, we tried using the k-means clustering method in Julia to create a predictive model of our data using unsupervised clustering of our data. To start we created a function to give us random initial centroids in our data set.... Since the size of our data set is a 1030x10 Matrix, we decided to get 10 initial clusters for our data set. Next we want to calculate the distance between our initial centroids and the vector we are looking for "x_obs" which in our case will be a 10 x 1 matrix and will have all of its values be equal to 20 which represents the theoretical minimum strength required for the highway construction project we are basing our predictive modeling analysis for which will be done using our function calc_distance we created... After we have the distances for our centroids, we want to assign them to a specific group based on the location they are at in the data set using the calc_group function we created which returns a vector assigning group numbers to our data set.... After we have our groups made, we now want to update our centroids so they become the actual centroids for the observations we assigned to them when we created the groups and we do this by using the function we created update_centroids!.... After we update our centroids to be located at the center of the groups they were assigned to, we now can do the actual k-mean algorithm to find our clusters of our data for our predictive modeling analysis. Using our function we created kmeans, we can find which cluster each of our observations should be located in our data set where we do 10000 steps for our data to get an accurate result for our k-mean data and we choose 10 for our k integers since we originally chose 10 initial clusters and want to be consistent...

References

1. **Concrete Compressive Strength** <https://www.kaggle.com/datasets/sinamhd9/concrete-comprehensive-strength>
2. https://www.researchgate.net/publication/222447231_Modeling_of_Strength_of_High-Performance_Concrete_Using_Artificial_Neural_Networks_Cement_and_Concrete_research_2812_1797-1808
3. **The effect of superplasticizers on the mechanical performance of concrete made with fine recycled concrete aggregates**
P Pereira, L Evangelista, J de Brito
Cement and Concrete Composites (2012-10-01)
<https://www.sciencedirect.com/science/article/pii/S0958946512001369>
DOI: [10.1016/j.cemconcomp.2012.06.009](https://doi.org/10.1016/j.cemconcomp.2012.06.009)
4. **Superplasticizers in Concrete** <http://www.buildingresearch.com.np/newfeatures.php>
5. **The Study of the Properties of Fly Ash Based Concrete Composites with Various Chemical Admixtures**
M Ondova, N Stevulova, A Estokova
Procedia Engineering (2012) <https://doi.org/gq5qj3>
DOI: [10.1016/j.proeng.2012.07.582](https://doi.org/10.1016/j.proeng.2012.07.582)
6. **Aggregates** <https://www.cement.org/cement-concrete/concrete-materials/aggregates>
7. **Compressive strength development of concrete with different curing time and temperature**
J-K Kim, Y-H Moon, S-H Eo
Cement and Concrete Research (1998-12-01)
<https://www.sciencedirect.com/science/article/pii/S0008884698001641>
DOI: [10.1016/s0008-8846\(98\)00164-1](https://doi.org/10.1016/s0008-8846(98)00164-1)
8. **Properties of concrete containing ground granulated blast furnace slag (GGBFS) at elevated temperatures**
Rafat Siddique, Deepinder Kaur
Journal of Advanced Research (2012-01) <https://doi.org/d2n4rj>
DOI: [10.1016/j.jare.2011.03.004](https://doi.org/10.1016/j.jare.2011.03.004)
9. **Statistical Analysis of Concrete Compressive Strengths for California Highway Bridges**
Christian Unanwa, Mark Mahan
Journal of Performance of Constructed Facilities (2014-02-01)
<https://ascelibrary.org/doi/10.1061/%28ASCE%29CF.1943-5509.0000404>
DOI: [10.1061/\(asce\)cf.1943-5509.0000404](https://doi.org/10.1061/(asce)cf.1943-5509.0000404)