Compressive Strength of Concrete Prediction Using ML model

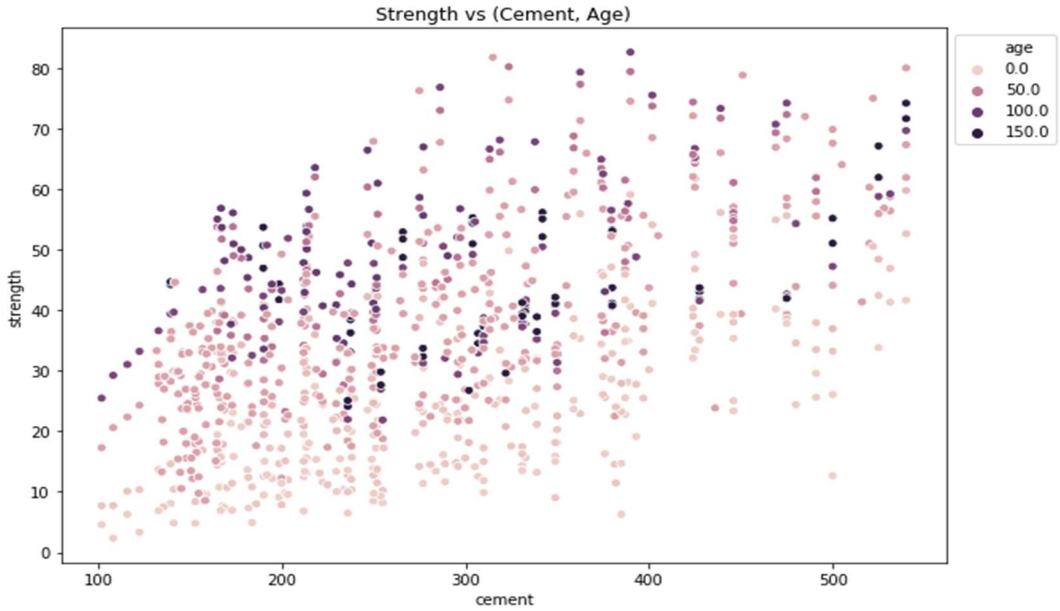
Abstract:

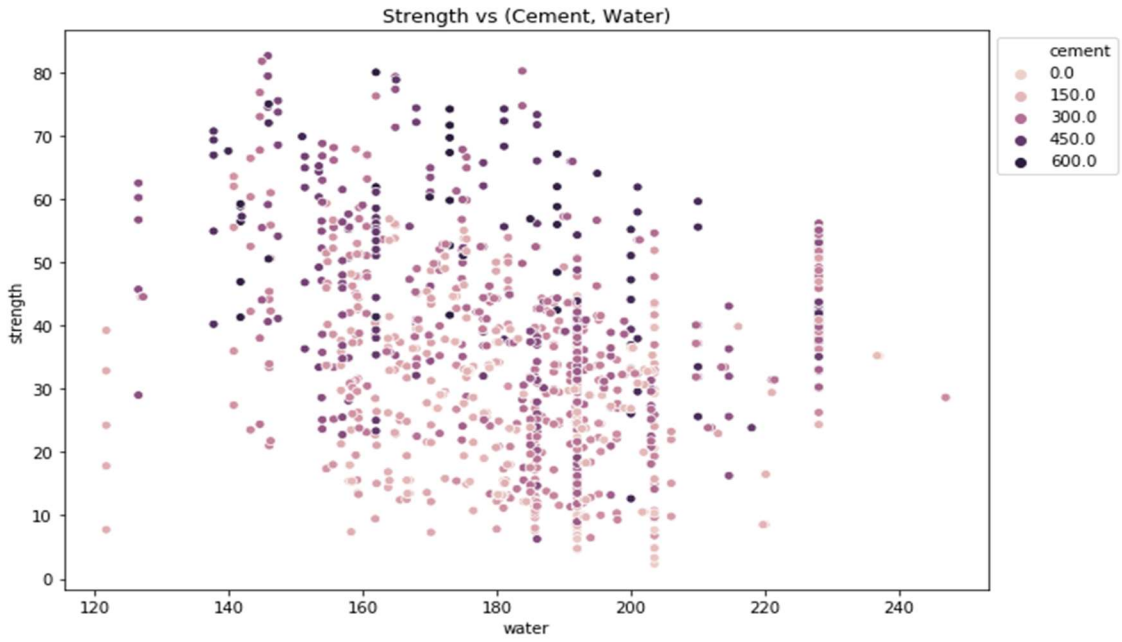
To predict the concrete strength of using the composition of its mixture and age of the concrete. Concrete has been widely used in recent years because its production compliments environmental conservation. It is a standard industrial practice that the concrete is classified based on grades. This grade is nothing but the Compressive Strength of the concrete cube or cylinder. Cube or Cylinder samples are usually tested under a compression testing machine to obtain the strength of concrete. The test requisites differ country to country based on the design code. ML methods, have been used to increase the prediction accuracy of concrete properties, and the data derived from the literature sources were used. Regression models tend to be used for the prediction of the compressive strength of high strength concrete. These models also demonstrate how the concrete compressive strength depends on the mixing ratios. With the use of Machine Learning Model, there will be no limitation of the complexity increasing number of variables. This Model and train and test the given population of concrete and with the best performing machine learning model it can effortlessly predict the strength of the concrete with much higher accuracy than traditional methods.

About the dataset:

The experimental data used for the prediction of concrete compressive strength in the present study have been taken from publicly available online dataset (UCI repository). The dataset consists of 1030 instances with 9 attributes and There are no missing values reported by the dataset providers. There are 8 input variables and 1 output variable. Seven input variables represent the amount of raw material (measured in kg/m³) and one represents Age (in Days). The target variable is Concrete Compressive Strength measured in (MPa — Mega Pascal). In this dataset we detect 25 duplicate value and we remove it. And in column we did outlier treatment by usinng IQR methodand we imputed upper fence value in the place of outlier.

Exploratory Data Analysis:





- The quantity of cement increases then the strength of the concrete increases.
- Strength of the concrete increases with the age.
- When the water decreases strength of the concrete is decreases.
- Water is the plays a crucial role in strength of concrete.
- The older cement is the more water require.

Data Modeling:

1. Linear Regression-

We will start with Linear Regression since this is the go-to algorithm for any regression problem. The algorithm tries to form a linear relationship between the input features and the

target variable i.e. it fits a straight line given by,

$$y = W * X + b = \sum_{i=1}^{n} w_i * x_i + b$$

Where w_i corresponds to the coefficient of feature x_i.

S.No	Metrics	Linear Regression
0	RMSE	8.506108
1	MSE	72.353867
2	MAE	6.563576
3	R2_SCORE	0.707303

2.Decision Tree:

Decision tree is a supervised ML algorithm. It can be used for both regression and classification. The aim of the DT algorithm is to divide the dataset into smaller, meaningful pieces, where each input has its own class label value. Different measurements are used for the DT splitting, such as Gini and information gain. Regression tree is a type of a DT and a hierarchical model for the supervised learning.

Classification and regression trees (CART), methods are the most important learning algorithms mentioned in the literature

S.No	Metrics	Decision Tree Regression
0	RMSE	7.189785
1	MSE	51.693011
2	MAE	4.898168
3	R2_SCORE	0.790883

3. Random Forest:

Random Forest Regressor trains randomly initialized trees with random subsets of data sampled from the training data, this will make our model more robust. Random forest (RF) is an ensemble method that combines many Decision trees. It can be used for both regression and classification. Each Decision Tree in the forest is created by the selection of different samples from the original dataset by the bootstrap technique. These samples are then trained using a set of attributes selected by the bagging mechanism.

S.No	Metrics	Random Forest Regression
0	RMSE	5.618542
1	MSE	31.568013

2	MAE	4.000963
3	R2_SCORE	0.872296

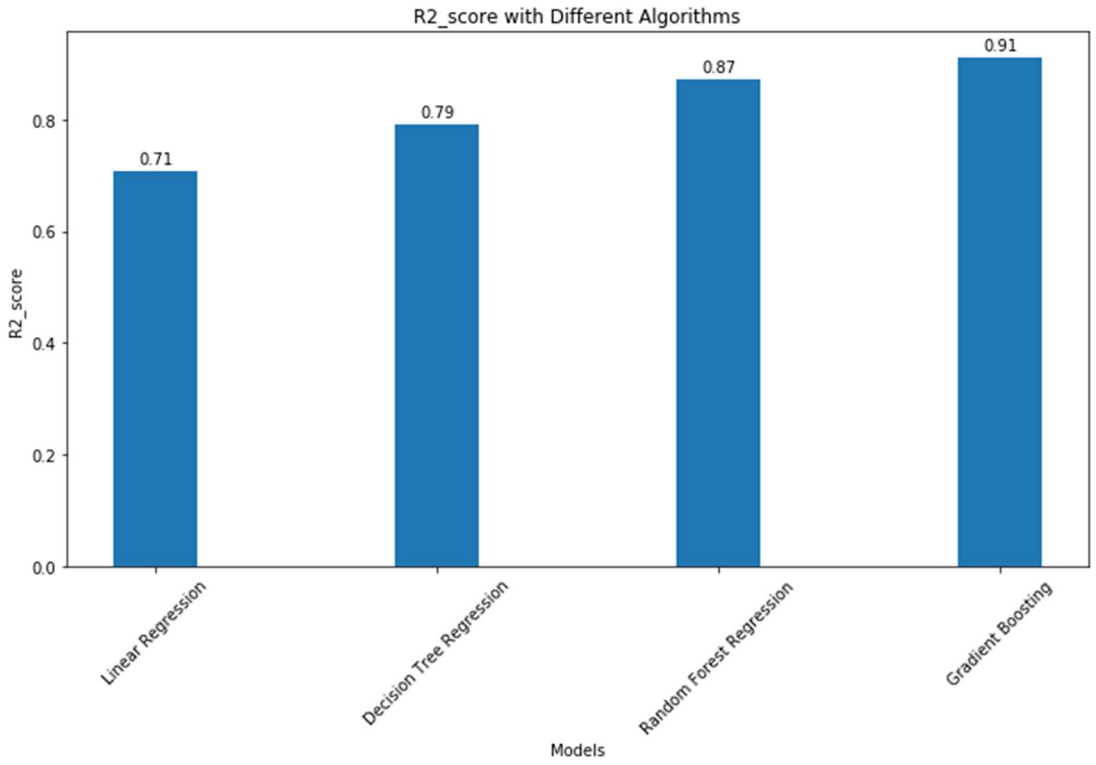
4. Gradient Boosting:

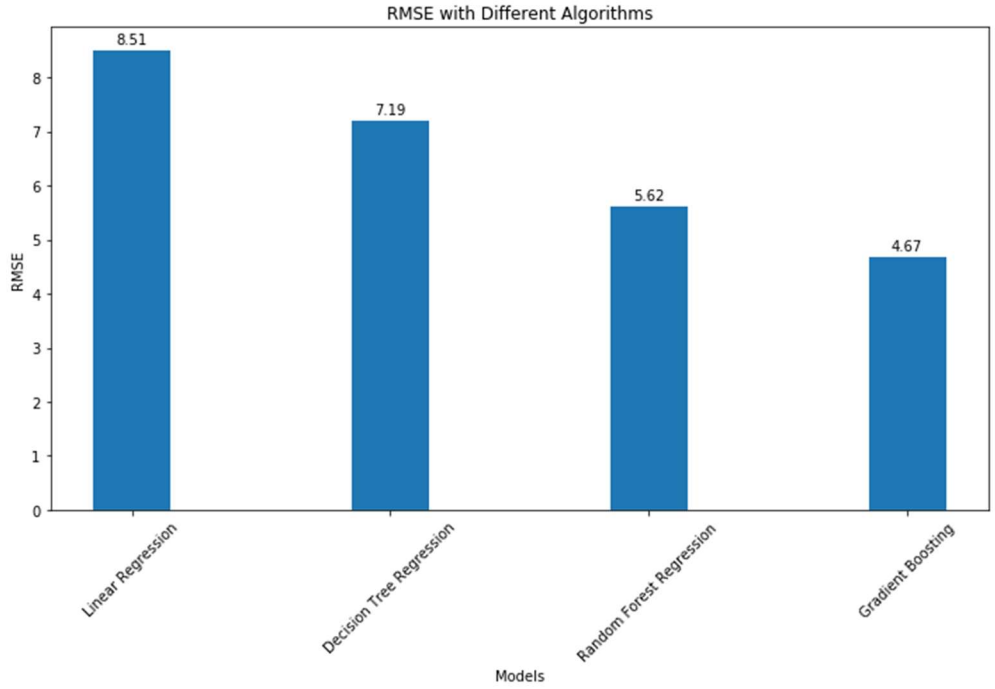
Gradient Boosting is a Machine Learning algorithm, uses for both classification and regression problem. It works on the principle that many weak learners can together make a more accurate predictor [Shallow Trees]. Gradient Boosting works by building simpler prediction model sequentially where is model try to predict the error leftover by the previous model.

Gradient Boosting works on principle of improving mistake of the previous learner through the next learner.

S.No	Metrics	Gradient Boosting Regression
0	RMSE	4.665450
1	MSE	21.766421
2	MAE	3.423395
3	R2_SCORE	0.911947

Conclusion:


R2_score with Different Algorithms
R2_score
0.0
0.2
0.4
0.6
0.8
0.71
0.79
0.87
0.91
Linear Regression
Decision Tree Regression
Random Forest Regression
Gradient Boosting
Models



- We have analysed the Compressive Strength Data and used Machine Learning to Predict the Compressive Strength of Concrete. We have used Linear Regression and its variations, Decision Trees, Random Forests and Gradient Boosting to make predictions and compared their performance.
- Gradient Boosting Regressor has the lowest RMSE and is good choice for problem.
- we can further improve the performance of the algorithm by tuning the hyperparameters by performing a grid search or random search.
- From the above plots, we concluded that Gradient Boosting has less RMSE value and good r2 score value.

Team Member:

1. Kovuru Harun Rasheed
2. Laxman Singh
3. Pulugu Vamsi
4. Mogallapalli Manibhargav