Predicting the Compressive Strength of Concrete using ML Methods

Pavan Teja Pusala

K Rahul Shashank

JPNCE

Mahbubnagar,India

Rahulshashank21@gmail.com

Pavantejappt@gmailcom

.

I INTRODUCTION

Conventional concrete is a mixture of cement, water, and coarse and fine aggregates. Supplementary components such as chemical and mineral admixtures may be added to the basic concrete ingredients to enhance its properties in fresh or hardened state. The procedure of selecting appropriate ingredients for concrete and its relative amount with the aim of producing concrete of obligatory strength, workability, and durability as cost-spinning as possible is termed mix design. In recent years, many researchers have been working on developing accurate concrete compressive strength prediction models .The prediction of compressive strength of concrete has great connotation, if it is brisk and consistent because it offers an option to do the essential modification on the mix proportion used to avoid circumstances where concrete does not attain the mandatory design strength or by avoiding concrete that is gratuitously sturdy and also for more economic use of raw material and fewer construction failures, hence reducing construction cost The conventional process of testing the compressive strength of concrete involves casting several cubes for the respective grade (such as M5, M10, M15 etc.) and observing the strength of the concrete over a period of time ranging from 7 to 28 days. This is a time consuming and rather tedious process. Employing machine learning approaches, Artificial neural networks instead of traditional models makes it possible to develop a better understanding of the compressive strength of concrete. Hence, the focus of this paper is the application of machine learning process, Artificial neural networks and their suitability to model concrete compressive strength compared with early models obtained from the literature and compared with some conventional approaches and also a recommendation system is developed by applying various ML methods, Artificial neural network[15]methods to predict the concrete strength from its components accurately and then looking for the optimal combination of components which increases the strength.

II DATASET AND LITERATURE SURVEY

A) 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).

	Cement (component 1)(kg in a m^3 mixture)	Blast Furnace Slag (component 2)(kg in a m^3 mixture)	Fly Ash (component 3)(kg in a m^3 mixture)	Water (component 4)(kg in a m^3 mixture)	Superplasticizer (component 5) (kg in a m^3 mixture)	Coarse Aggregate (component 6)(kg in a m^3 mixture)	Fine Aggregate (component 7)(kg in a m^3 mixture)	Age (day)	Concrete compressi strength(M megapasca
0	540.0	0.0	0.0	162.0	2.5	1040.0	676.0	28	79.99
1	540.0	0.0	0.0	162.0	2.5	1055.0	676.0	28	61.89
2	332.5	142.5	0.0	228.0	0.0	932.0	594.0	270	40.27
3	332.5	142.5	0.0	228.0	0.0	932.0	594.0	365	41.05
4	198.6	132.4	0.0	192.0	0.0	978.4	825.5	360	44.30

B) Literature survey:

1)Title:Comparison of Machine Learning Techniques for the Prediction of Compressive Strength of Concrete.

Author:PalikaChopra,RajendraKumarSharma,Maneek Kumar,andTanujChopra.

Year: *12 Apr 2018*

Techniques: Three models have been used for the prediction of compressive strength of concrete:

(I) decision tree (DT) model,

(ii) random forest (RF) model, and

(iii) neural network (NN) model.

Limitations: The problem with the DT model is that it unable to predict beyond the range (minimum and maximum limits) given in the training dataset. therefore, the RMSE is high as compared to RF.

2) Title: Predictive modeling of high-performance concrete with regression analysis

Author: S.S. Wu ; B.Z. Li ; J.G. Yang ; S. K. Shukla

Year: 23 December 2016

Techniques: This paper deals with building a regression model for predicting concrete's compressive strength. First of all, eight process variables are identified as determinants of Concrete Compressive Strength (CCS). These variables are Cement, Blast Furnace Slag, Fly Ash, Water, Superplasticizer, Coarse Aggregate, Fine Aggregate, and Age.

Limitations: Appropriateness of the model is investigated by testing it against unseen data points. Further, correlation among these variables is computed and it is found that a few of them are highly correlated.

3) Title: Machine Learning Techniques in Concrete Mix Design

Author:PatrykZiolkowskiOrcID,Maciej Niedostatkiewicz

Year: 15 December 2019

Techniques: These methods are the Bukowski method, the Eyman and Klaus method, and the Paszkowski methods,Artificial Neural Networks (ANN).

Limitations: That the presented mathematical expression does not adequately reflect all the relationships between the components and have
certain boundary conditions It is likely to analyse the effect of admixtures and concrete durability by effective measures.

4)**Title**: Can the compressive strength of concrete be estimated from knowledge of the mixture proportions?: New insights from statistical analysis and machine learning methods.

Author:BenjaminYoung,Puneet Gupta,Gaurav Sant

Year: 14 June 2018

Techniques: Machinelearning and datamining algorithms, Artificial neural networks (ANNs),Decision trees,Support vector machines (SVMs).

Limitations: This study presents the first analysis of a large data set (>10,000 observations) of measured compressive strengths from actual (jobsite) mixtures and their corresponding actual mixture proportions.

*5)***Title**: Prediction of concrete strength using artificial neural networks

Author:seung-chang lee

Year: 19 April 2017

Techniques: Artificial neural network (ANN)

Limitations: This concept proposes that concrete of the same mix at the same maturity has approximately the same strength.

6) **Title**: Machine learning in concrete strength simulations: Multi-nation data analytics

Author:Jui-ShengChouaChih-FongTsaibAnh-DucPhamacYu-HsinLud

Year: 24 September 2019

Techniques: The individual learning classifiers are constructed from four different base learners, including multilayer perceptron (MLP) neural network, support vector machine (SVM), classification and regression tree (CART),and linear regression.

Limitations: This study validates the applicability of ML, voting, bagging, and stacking techniques for simple and efficient simulations of concrete compressive strength.

The further studies are needed to explore how the parameters in these models can be optimized automatically.

III MODEL BUILDING,ALGORITHM AND ARCHITECTURE DIAGRAM:

A) MODEL BUILDING:

ML methods:

The ML regression method estimates the output value using the input samples of the dataset. Such a procedure is also termed as the training set. The purpose of the regression method is to minimize the error between the predicted and actual outputs. The values were predicted using the ML regression models, namely, the regression tree, RF, support vector machines, artificial neural network [20].

Datasets were randomly split into 70% for the training set and 30% for the independent test set. The training data were used to train the ML model.[10] The independent test data were applied for the evaluation of the model's performance. The 10-fold cross-validation procedure helped in the estimation of the ML model skills.

In this study, four different normalization methods (i.e., min-max, decimal, sigmoid, and z-score) were applied to derive the most successful normalization method for the raw dataset. Then, the K-nearest neighbour (KNN) regression method was applied to the normalized datasets. The prediction results were compared to determine the most suitable normalization method. Later, the raw datasets were normalized with the determined normalization technique.

After preparing the data, we can fit different models on the training data and compare their performance to choose the algorithm with good performance. As this is a regression problem, we can use RMSE (Root Mean Square Error) and R^2 score as evaluation metrics.[11]

Evaluation Metrics:

To evaluate the predicted values of the regression methods, the actual and predicted values were compared. In this study, the R, RMSE, and MAE metrics were used to evaluate the prediction accuracy.

The model parameters were optimized for the highest R, lowest RMSE, and lowest MAE. All of them were calculated according to the following equations:

$$R = \frac{\sum_i^N \left(\text{actual}_i - \overline{\text{actual}}\right)\left(\text{predicted}_i - \overline{\text{predicted}}\right)}{\sqrt{\sum_{i=1}^N \left(\text{actual}_i - \overline{\text{actual}}\right)^2 \sum_{i=1}^N \left(\text{predicted}_i - \overline{\text{predicted}}\right)^2}},$$

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\text{predicted}_i - \text{actual}_i\right)^2},$$

$$\text{MAE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left|\text{predicted}_i - \text{actual}_i\right|}.$$

, N is the number of data points.

Linear Regression:

We will start with Linear Regression since this is the go-to algorithm for any regression problem [12]. 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,

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

Linear Regression

Where w_i corresponds to the coefficient of feature x_i.

The magnitude of these coefficients can be further controlled by using regularization terms to the cost functions. Adding the sum of the magnitudes of the coefficients will result in the coefficients being close to zero, this variation of linear regression is called **Lasso** Regression.

Adding the sum of squares of the coefficients to the cost function will make the coefficients be in the same range and this variation is called **Ridge** Regression. Both these variations help in reducing the model complexity and therefore reducing the chances of overfitting on the data.

Decision Trees:

A Decision Tree Algorithm represents the data with a tree-like structure, where each node represents a decision taken on a feature. This algorithm would give better performance in this case, since we have a lot of zeros in some of the input features as seen from their distributions in the pair plot above. This would help the decision trees build trees based on some conditions on features which can further improve performance. so, the Decision Tree Regressor has improved the performance by a significant amount.

Decision tree (DT) 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.

There are three types of nodes in DT: (i) root node (RN) (ii) decision node (DN), and (iii) leaf node (LN). The topmost node is RN. At DN, a conditional test is conducted, and further subtrees are decided according to the outcome of the test. LN is the target node (output) of the tree. DT algorithms calculate two types of entropies (a measure of homogeneity of the dataset).

Random Forests*:*

Using a Decision Tree Regressor has improved our performance, we can further improve the performance by ensembling more trees. Random Forest Regressor trains randomly initialized trees with random subsets of data sampled from the training data, this will make our model more robust.

The RMSE has further reduced by ensembling multiple trees. We can plot the feature importance's for tree-based models. The feature importance's show how important a feature is for a model when making a prediction.[6] Cement and Age are treated as the most important features by tree-based models. Fly ash, Coarse and Fine aggregates are the least important factors when predicting the Strength of Concrete.

Random forest (RF) is an ensemble method that combines many DTs. It can be used for both regression and classification. Each DT in the forest is created by the selection of different samples from the original dataset by the bootstrap technique [2]. These samples are then trained using a set of attributes selected by the bagging mechanism.

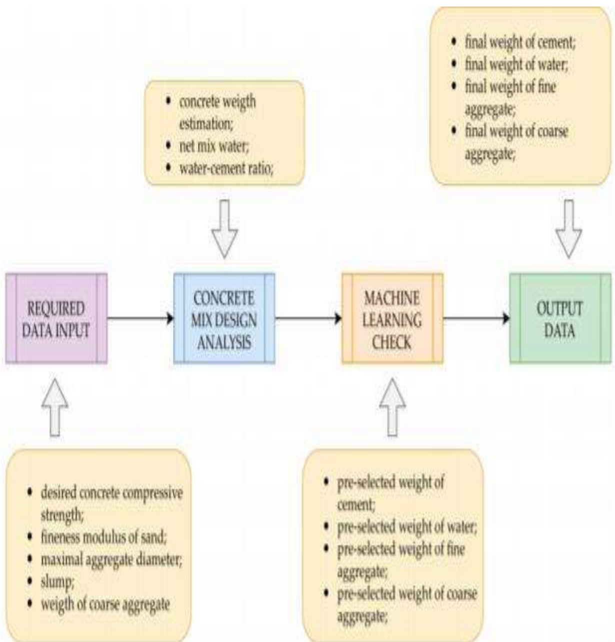
Each tree is grown with the following steps:

(i)The first step is to select the data frame. The data frame is 2/3 of the total training data selected randomly for each tree. This is known as bagging. Predictor variables are selected randomly, and the best split on these variables is used to split the node.

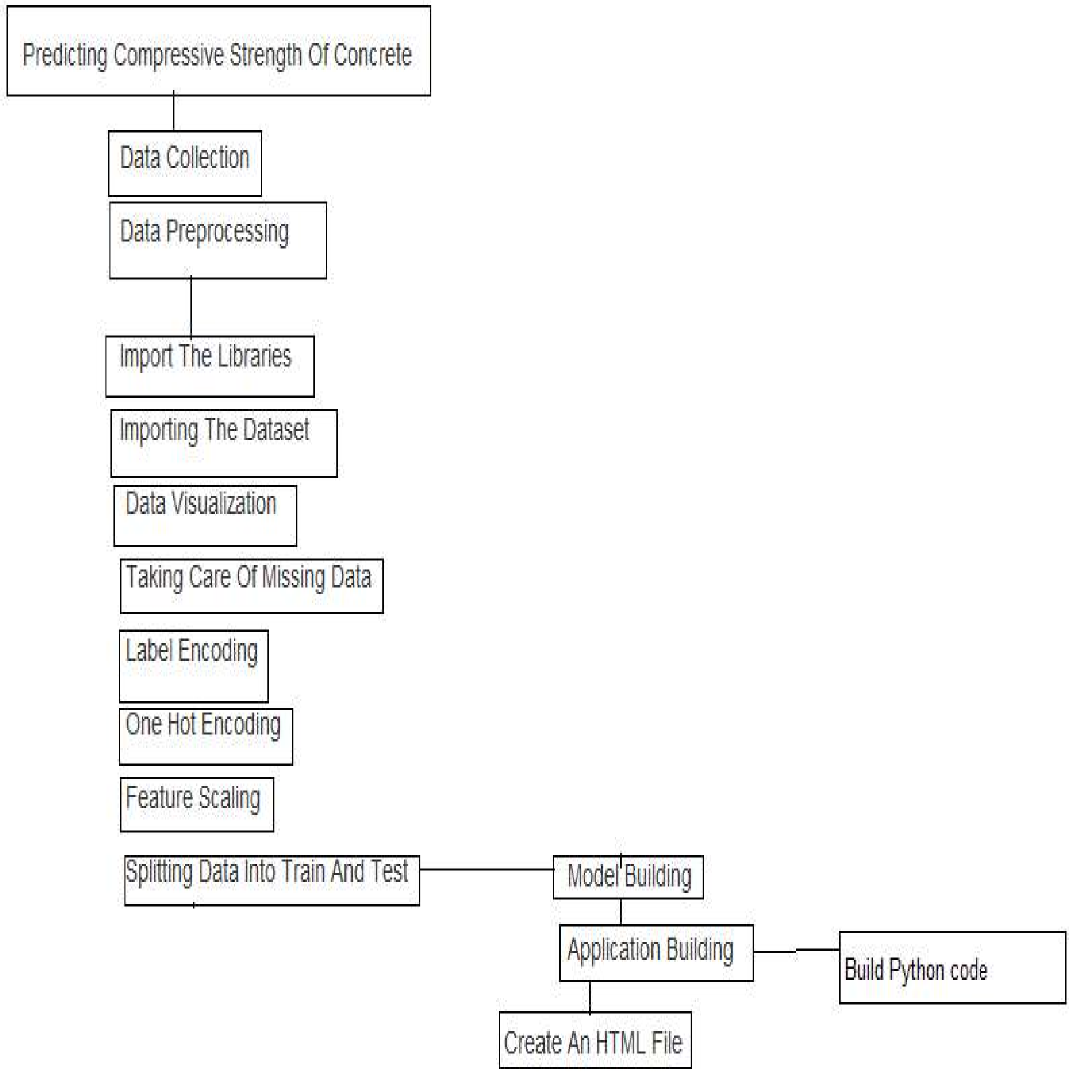
(ii)For each tree, calculate the out-of-bag error (validation or testing) using the rest of the data. Then, errors from all the trees are aggregated to find the overall out-of-bag error rate.

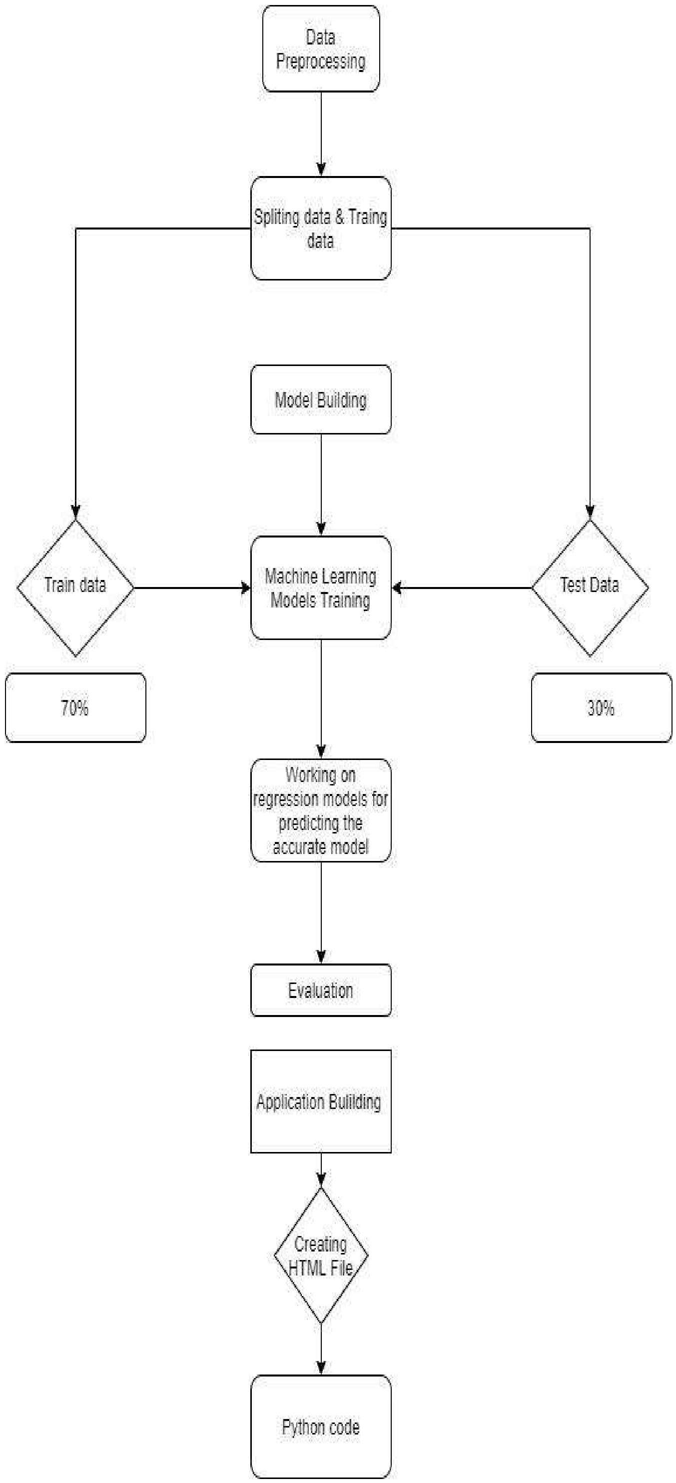
(iii)Each tree gives a classification, and the model chooses from the forest having most of the votes over all the trees in the forest. The votes could be 0's or 1's. The percentage of 1's received is defined as the prediction probability.

B) ARCHITECTURE DIAGRAM:



FLOWCHART:


Predicting Compressive Strength Of Concrete
Data Collection
Data Preprocessing
Import The Libraries
Importing The Dataset
Data Visualization
Taking Care Of Missing Data
Label Encoding
One Hot Encoding
Feature Scaling
Splitting Data Into Train And Test
Model Building
Application Building
Build Python code
Create An HTML File


Data Preprocessing
Spliting data & Traing data
Model Building
Train data
Machine Learning Models Training
Test Data
70%
30%
Working on regression models for predicting the accurate model
Evaluation
Application Bulilding
Creating HTML File
Python code

IV. RESULTS AND DISCUSSION:

FIG1: AIRPLOT OF DATA

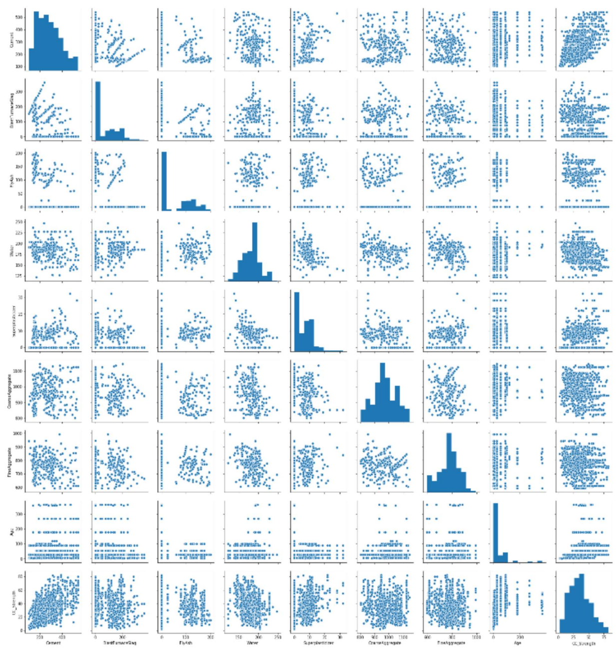


Fig 2.4 cement

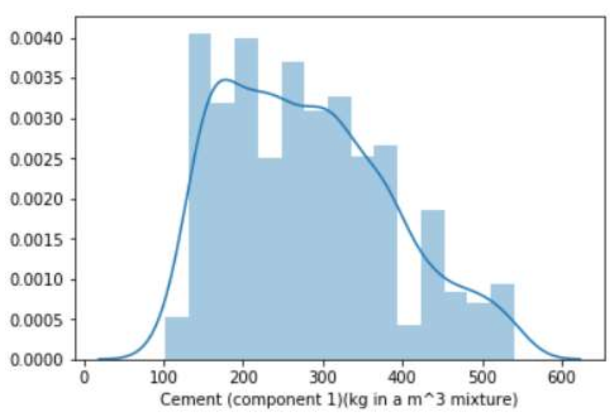


FIG 2: GRAPHS OF COMPONENTS

Fig 2.1 Blast furnace

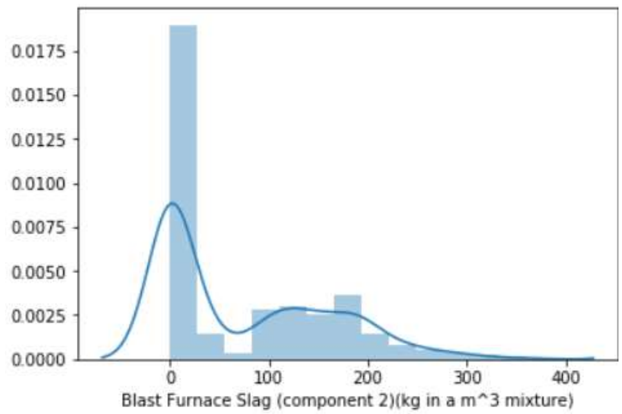


Fig2.2 Fly ash

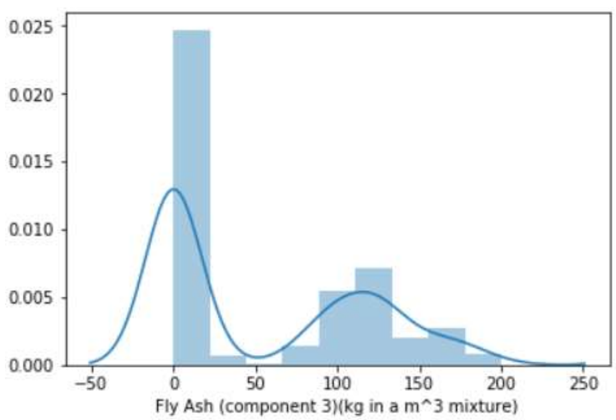


Fig2.3 Water

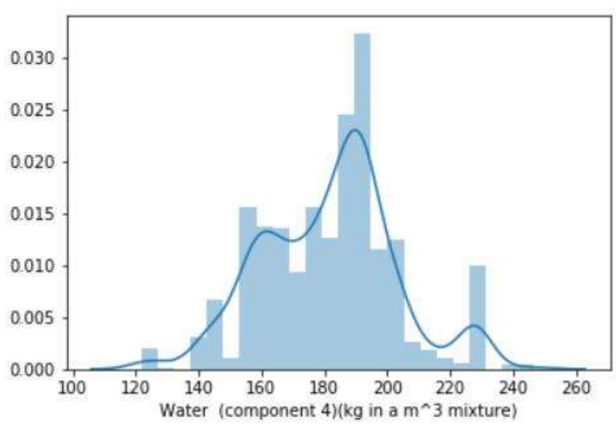


Fig 2.5 Coarse aggregate

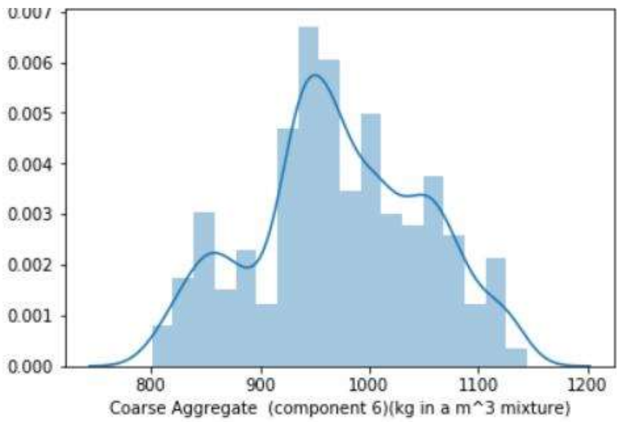
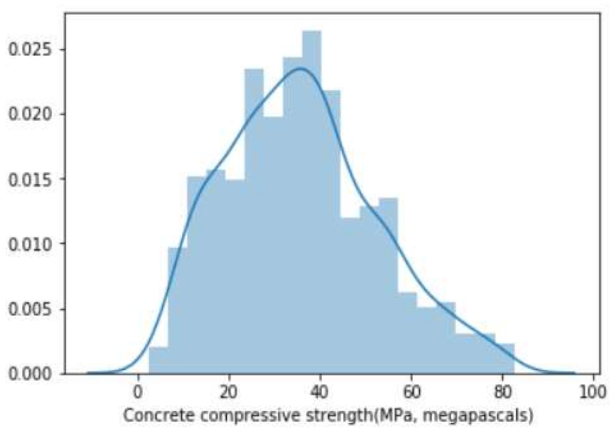
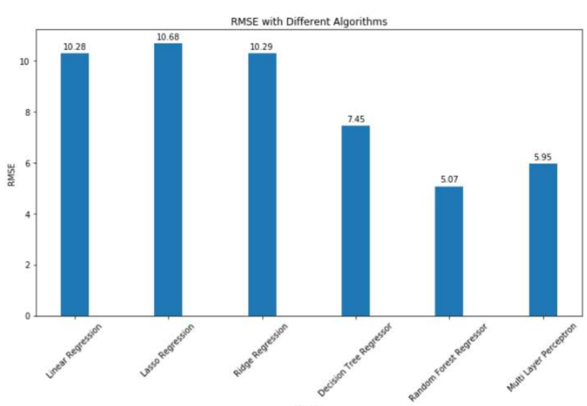
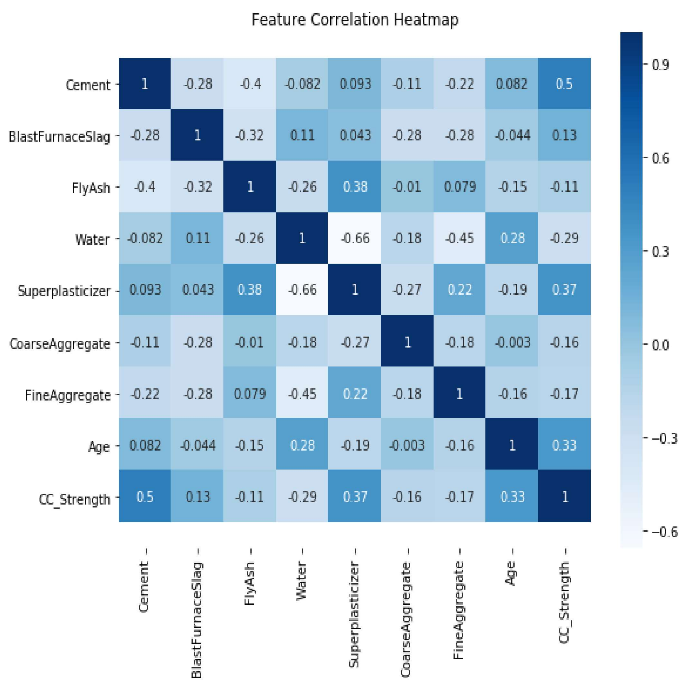


Fig 2.6 concrete compressive strength

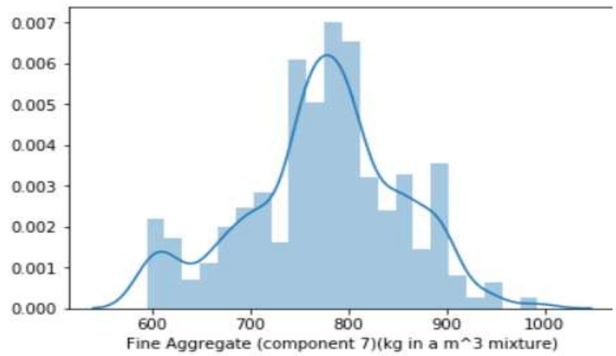


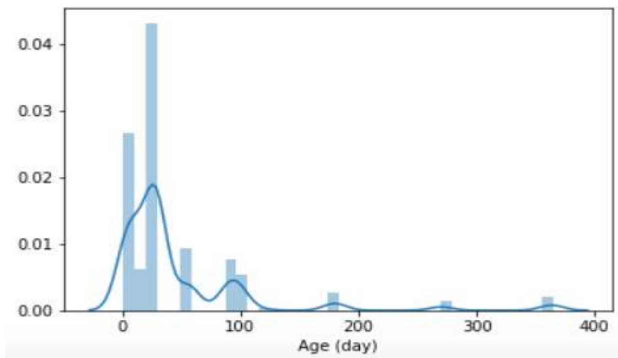


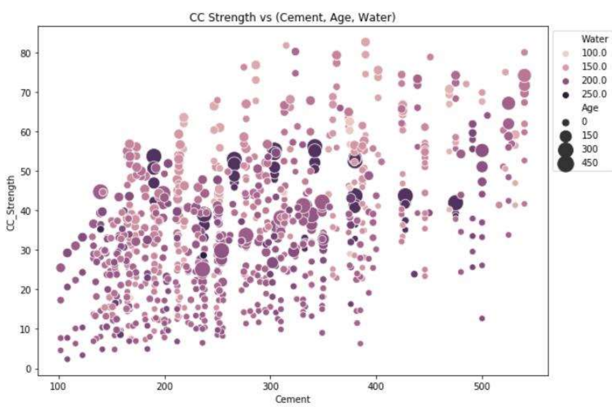


FEATURE CORRELATION HEAT MAP

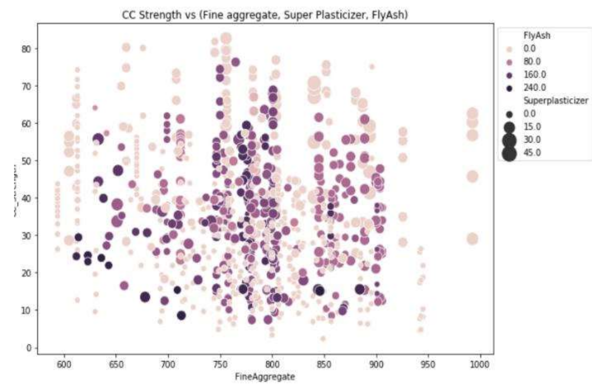
Fig 2.7 Fine aggregate and age







Compressive strength increases as the amount of cement increases, as the dots move up when we move towards right on the x-axis.



Compressive strength decreases Fly ash increases, as darker dots are concentrated in the region representing low compressive strength.

Compressive strength increases with Superplasticizer since larger the dot the higher they are in the plot.

FEATURE COEFFICIENTS

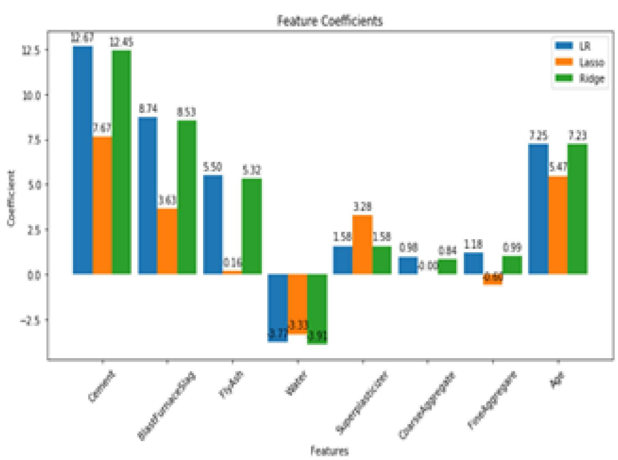
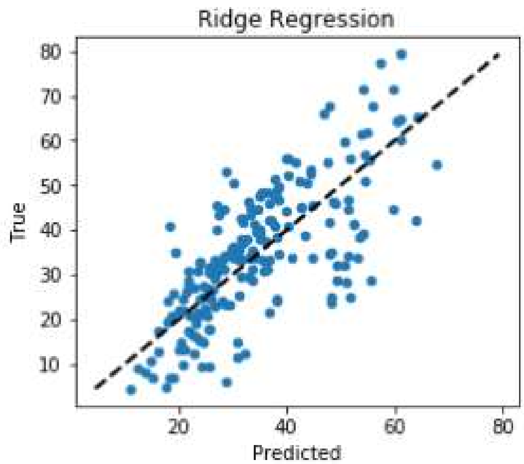
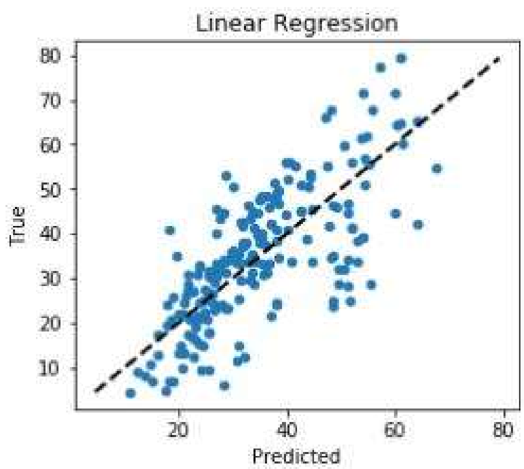
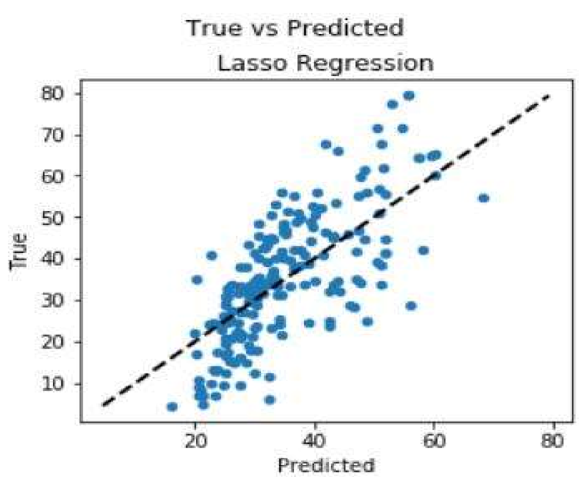


Fig 2.8 True vs Predicted


Ridge Regression
True
Predicted


Linear Regression
True
Predicted


True vs Predicted
Lasso Regression
True
Predicted

V.CONCLUSION:

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 and Random Forests to make predictions and compared their performance. Random Forest Regressor has the lowest RMSE and is a good choice for this problem. Also, we can further improve the performance of the algorithm by tuning the hyperparameters by performing a grid search or random search.

NN approaches combine the complexity of many statistical techniques with machine learning techniques and attributed as a black box which allows NN to be applied in all engineering disciplines. It comes out as the best possible model for the prediction of compressive strength of concrete. It has predicted with high accuracy for all the curing ages, that is, 28, 56, and 91 days.

The RF model has been concluded as the second possible alternative for 56- and 91-day compressive strength of concrete, but it consumes longer time as compared to NN [18]. The DT model has not been able to predict for the present dataset of compressive strength of concrete.

As an outcome, the NN model may serve as the best feasible prediction tool for predicting the compressive strength of concrete.

We then compared the performance of our intelligent prediction system to previously published works with a similar mission. Our comparison criteria was based on vital factors in computational intelligence and machine learning, [19]such as; the correct prediction rate, the minimum error, the size of the dataset, the number of input attributes for training the system, the data preprocessing and coding, and finally the learning schemes and also recommendation system helped to Predict concrete strength from its components accurately and then looking for the optimal combination of components which increases the strength. Further work involves investigating more learning schemes, exploring more ML algorithms like Genetic programming and improving the correct prediction rate.

VI. REFERENCES:

[1] Yaseen, Z. M., Deo, R. C., Hilal, A., Abd, A. M., Bueno, L. C., Salcedo-Sanz, S., & Nehdi, M. L. (2018). Predicting compressive strength of lightweight foamed concrete using extreme learning machine model. Advances in Engineering Software, 115, 112-125.

[2] Pham, A. D., Hoang, N. D., & Nguyen, Q. T. (2016). Predicting compressive strength of high-performance concrete using metaheuristic-optimized least squares support vector regression. Journal of Computing in Civil Engineering, 30(3), 06015002.

[3] Słoński, M. (2010). A comparison of model selection methods for compressive strength prediction of high-performance concrete using neural networks. Computers & structures, 88(21-22), 1248-1253.

[4] Asteris, P. G., Apostolopoulou, M., Skentou, A. D., & Moropoulou, A. (2019). Application of artificial neural networks for the prediction of the compressive strength of cement-based mortars. Computers and Concrete, 24(4), 329-345.

[5] Bui, D. K., Nguyen, T., Chou, J. S., Nguyen-Xuan, H., & Ngo, T. D. (2018). A modified firefly algorithm-artificial neural network expert system for predicting compressive and tensile strength of high-performance concrete. Construction and Building Materials, 180, 320-333.

[6] Zhang, J., Ma, G., Huang, Y., Aslani, F., & Nener, B. (2019). Modelling uniaxial compressive strength of lightweight self-compacting concrete using random forest regression. Construction and Building Materials, 210, 713-719.

[7] Omran, B. A., Chen, Q., & Jin, R. (2014). Prediction of compressive strength of "green" concrete using artificial neural networks. Sulbaran, T.

[8] Khashman, A., & Akpinar, P. (2017). Non-destructive prediction of concrete compressive strength using neural networks. Procedia Computer Science, 108, 2358-2362.

[9] Poorarbabi, A., Ghasemi, M., & Moghaddam, M. A. (2020). Concrete Compressive Strength Prediction Using Neural Networks Based on Non-destructive Tests and a Self-calibrated Response Surface Methodology. Journal of Nondestructive Evaluation, 39(4), 1-11.

[10] Khademi, F., & Behfarnia, K. (2016). Evaluation of concrete compressive strength using artificial neural network and multiple linear regression models.

[11] https://www.researchgate.net/profile/Susom_Dutta/publication/324459144_Comparison_of_machine_learning_techniques_to_predict_compressive_strength_of_concrete/links/5aceb35fa6f

dcc87840f2483/Comparison-of-machine-learning-techniques-to-predict-compressive-strength-of-concrete.pdf

[12] Dutta, D., & Barai, S. V. (2019). Prediction of compressive strength of concrete: machine learning approaches. In Recent Advances in Structural Engineering, Volume 1 (pp. 503-513). Springer, Singapore.

[13] Abuodeh, O., Abdalla, J. A., & Hawileh, R. A. (2019, April). Prediction of Compressive Strength of Ultra-High-Performance Concrete using SFS and ANN. In 2019 8th International Conference on Modeling Simulation and Applied Optimization (ICMSAO) (pp. 1-5). IEEE.

[14] Nguyen, T. A., Ly, H. B., Mai, H. V. T., & Tran, V. Q. (2020). Prediction of Later-Age Concrete Compressive Strength Using Feedforward Neural Network. Advances in Materials Science and Engineering, 2020.

[15] Shariati, M., Mafipour, M. S., Mehrabi, P., Ahmadi, M., Wakil, K., Trung, N. T., & Toghroli, A. (2020). Prediction of concrete strength in presence of furnace slag and fly ash using Hybrid ANN-GA (Artificial Neural Network-Genetic Algorithm). Smart Structures and Systems, 25(2), 183-195.

[16] Moretti, J. F., Minussi, C. R., Akasaki, J. L., Fioriti, C. F., Melges, J. L. P., & Tashima, M. M. (2016). < b> Prediction of modulus of elasticity and compressive strength of concrete specimens by means of artificial neural networks. Acta Scientiarum. Technology, 38(1), 65-70.

[17] Chou, J. S., & Tsai, C. F. (2012). Concrete compressive strength analysis using a combined classification and regression technique. Automation in Construction, 24, 52-60.

[18] Armaghani, D. J., & Asteris, P. G. (2020). A comparative study of ANN and ANFIS models for the prediction of cement-based mortar materials compressive strength. Neural Computing and Applications, 1-32.

[19] Kurpinska, M., & Kułak, L. (2019). Predicting performance of lightweight concrete with granulated expanded Glass and Ash aggregate by means of using Artificial Neural Networks. Materials, 12(12), 2002.