Compression Strength Prediction Using Machine Learning Techniques

Phijo J Cherickal[1*], Vidhya V Kumar[2], Ranjini Mariam Philipose[3], Shailesh S[4], Tiji Thomas[5]
[1] Department of Computer Applications, MACFAST, India, meetphijo@gmail.com
[2] Department of Computer Applications, MACFAST, India, vidhyaviswakumar@gmail.com
[3] Department of Computer Applications, MACFAST, India, ranjinimariam@gmail.com
[4] Department of Computer Science, SACRED HEART COLLEGE, India, shaileshsivan@gmail.comm
[5] Department of Computer Applications, MACFAST, India, tiji@macfast.org

ABSTRACT

The advanced computing techniques and its applications on other engineering disciplines accelerated the different aspects and phases in engineering process. Nowadays there are so many computer aided methods widely used in civil engineering domain. The mathematical relationship between ratios of different concrete components and other influencing factors with its compression strength need to be analyzed for different engineering needs. This paper aims to develop a mathematical relationship after analyzing the above factors and to foresee the compressive strength of concrete by applying various regression techniques such as linear regression, support vector regression, decision tree regression and random forest regression on assumed data set., It was found that the accuracy of the random forest regression was considerable as per the result after applying the various regression techniques.

Keywords: Compressive Strength, Regression, Prediction, Concrete, SVM, Random Forest, Decision Trees

1. INTRODUCTION

In the fields of Science, Engineering, Health sector, the information technology provides a means for access to a broad variety of methods of modeling complex systems. These methods can be applicable in the field of civil engineering especially in construction strategies. Concrete is the mixture of Cement(c), Blast Furnace Slag, Water, Super Plasticizer, Coarse and Fine Aggregate, which is widely used construction material which is used in every kind of structures mainly in buildings. The procedure of determining the proportion of appropriate ingredients for producing concrete of required strength, workability, and durability is termed as mix design. The compressive strength of concrete is one of the widely used performance measure by the engineer in designing buildings. This study primarily focuses on finding the compressive strength of concrete. One of the limitation with the current Mix Design methodology is, it provides mechanism to find out mix ratio for a given strength values but the reverse is not possible. That is for a given proportion there is no mathematical model which defines relationship between this mix components and

resultant compressive strength value. So this can be done only through physical experiments. This study primarily focused on filling this gap. This paper focuses on developing a prediction model by using machine learning techniques for predicting the compressive strength of concrete, if we are given with the mix ratio. For building this prediction model, a publically available compression strength dataset is utilized. This study also proposes a simple mathematical relationship between the coefficient p (strength at infinite time) with the strength values of concrete of a particular day. The developed mathematical model is validated for commonly used concrete data. The reliable prediction of concrete strength at different days (7, 14, 28 days etc.) found excellent after the analyzing the data set.

2. LITERATURE EREVIEW

The Compression Strength of Concrete generally determined by a standard compression test on a small concrete specimen like cube or cylinder. This requires engineers to build the specimen with different combinations of raw materials and test these cylinders for strength variations with a change in each raw material. To ensure 98% strength, it is recommended to test the specimen after 28 days. The above test takes a lot of time and labour to prepare different prototypes and test them. Also, this method is prone to human error.

Using the neural network and fuzzy logic methods Topcu et al. [1] developed models, to improve the accurate prediction of the compression strength of the fly ash concrete and heavy-weight concrete. It is concluded, that in both studies using the neural network and fuzzy logic methods the compression strength could be predicted without any further experiments.

Khademiet et al.[2] compared the methods using multiple linear regression, neural network, and adaptive neuro-fuzzy inference system to estimate the concrete compression strength for 28 days and reported that the neural network and adaptive neuro-fuzzy models gives the reliable results.

M.Timur Cihan et al.[3] used the machine learning regression methods to predict the concrete compressive strength. The samples were prepared in the laboratory by using seven variables. For predicting the concrete compression strength, M.Timur Cihan et al. [3] compared the decision tree (DT), random forest (RF), support vector machine (SVM), artificial neural networks and Fuzzy logic to find the efficient regression method.

An image processing technique is proposed by Celalettin Basyigit, et al. [4] to arrive at the compressive strength values of different concrete classes. Celalettin Basyigit, et al. [4] uses seven water/cement ratios obtained using the different sets of concrete. The correlations between the values of compression strength of the concrete specimen and the results obtained in image processing techniques were compared. A good correlation is identified during the mixture design.

3. PRELIMINARIES

3.1. Regression

Regression is one of the machine learning methods in artificial intelligence, it allows predicting a continuous outcome variables based on the value of predictor variable. This model uses a mathematical equation to predict unknown from the known variable. Regression analysis is used to solve the prediction problems in machine learning techniques, which is treated as a supervised learning method to find the correlation between the variables. Different types of regression in machine learning techniques are used for analysis are Linear Regression, Ridge Regression, Support Vector regression, lasso Regression, polynomial Regression, Bayesian Linear Regression, Decision Tree regression, Random forest regression.

In our problem for the analysis, we focused on

1. Linear Regression
2. Support Vector Regression
3. Decision Tree Regression
4. Random Forest Regression

3.2. Linear Regression

The above mentioned are the basic types of regression techniques used in machine learning. In linear regression the predictor variable and a dependent variable is inter relate linearly.

$$Y = mx + c + e \qquad (1)$$

Equation (1) is used to denote linear regression model. Where m is considered as the slope of the line, c is an intercept and e will be the error in given model.

Linear regression mainly focuses on the conditional probability of the. They are Simple linear regression and multiple linear regression models. The simple linear regression model is used in many applications such as for predicting students marks based on number of hours they ideally studied.

The multiple linear regression models help to find t h e relation between multiple input variables and corresponding output variable. Figure 1 shows an example for linear regression model.

3.3. Support Vector Regression

For classifications in machine learning, the supervised learning model support vector machines are used, which predicts the acceptable amount of error. It also helps to find a hyper plane to fit the data. By using training sample, support vector regression forecast a proper mapping of input to real numbers. The SVR and SVM use the same principle in classification, with few differences. A real number is produced as output.

In figure 2, consider the top and bottom lines as the decision boundaries and the middle line as the hyperplane. In support vector regression, we observe the points only within the decision boundary. Since the hyper plane has the maximum number of points and is treated as the best fit.
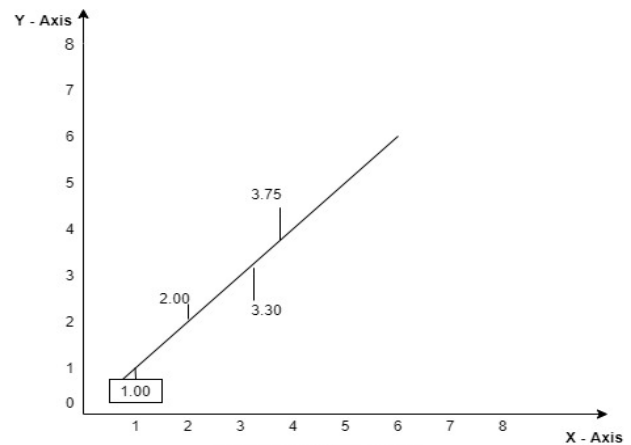

Figure 1 : Linear Regression

Consider the two decision boundary lines in the figure 2 as the distance say 'y', from the hyper plane.
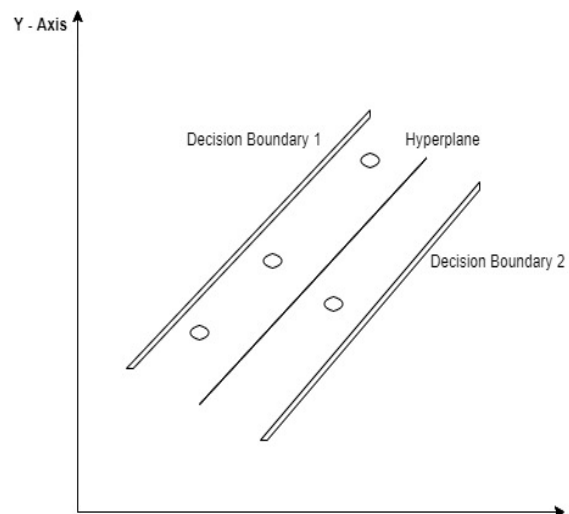

Figure 2: Support Vector Regression

We can draw the line from the hyperplane at a distance of '+y' and '-y'. Here 'a' in the text is basically referred to as epsilon.

$$Y = wx + b \qquad (2)$$

The equations of decision boundary become:

$$wx + b = +y \qquad (3)$$

$$wx + b = -y \qquad (4)$$

Thus, any hyperplane that satisfies our SVR should satisfy:

$$-y < wx + b < +y \qquad (5)$$

3.1.2 Decision Tree Regression

Decision Tree Regression is a supervised learning algorithm. This is used in solving classification and regression problems. In this type of regression, a tree like structure called decision tree is created. In this tree structure, all internal nodes denote the condition of the attribute, where branch indicates the result of the condition, and leaf node denotes the final decision. Here, from the assumed dataset the decision tree is start constructing as the root node. Next step is to divide the root node into left and right child nodes which acts as the subsets of the assumed dataset. These subsets are again divided into their subsets as shown in figure3.
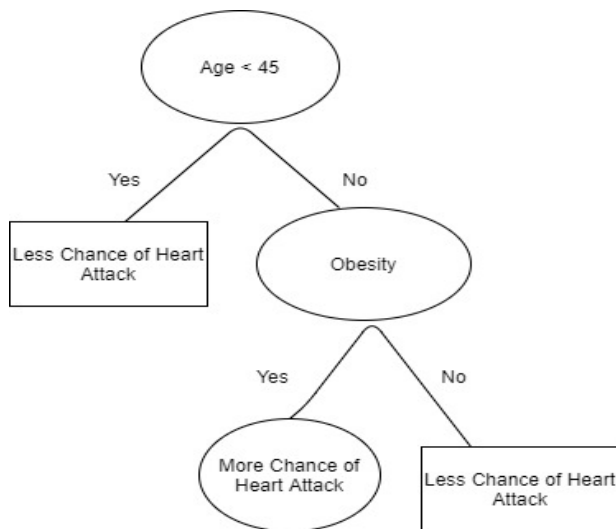

Figure 3 : Decision Tree Regression

3.4. Random Forest Regression

To solve the regression tasks random forest regression can be used as a powerful supervised learning algorithm. In this type of regression, multiple set of decision trees can be combined together. The final outcome of this regression method is calculated as the average of each combined decision tree output. Figure 4 shows a sample random forest regression. These participated decision trees are known as base models, and can be represented as:

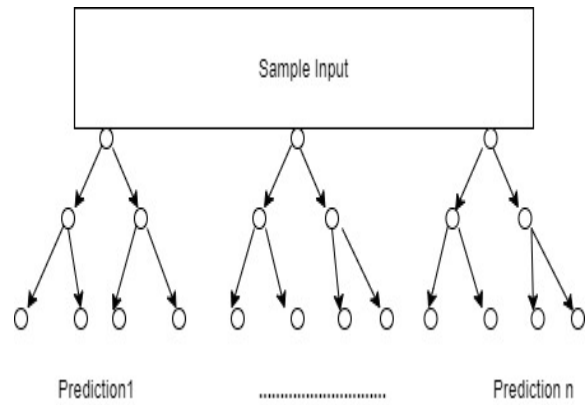$$g(x) = f_0(x) + f_1(x) + f_2(x) + \dots \qquad (6)$$

Figure 4: Random Forest Regression

4. METHODOLOGY

The purpose of this work is to create a mathematical relationship between the raw materials used to make a concert and its compression strength with the quantity, for describing the concept we introduced a machine learning pipeline

4.1. Regression Pipeline

The machine learning used in this work is shown in figure 5. The pipeline includes the dataset, preprocessing steps and model building.
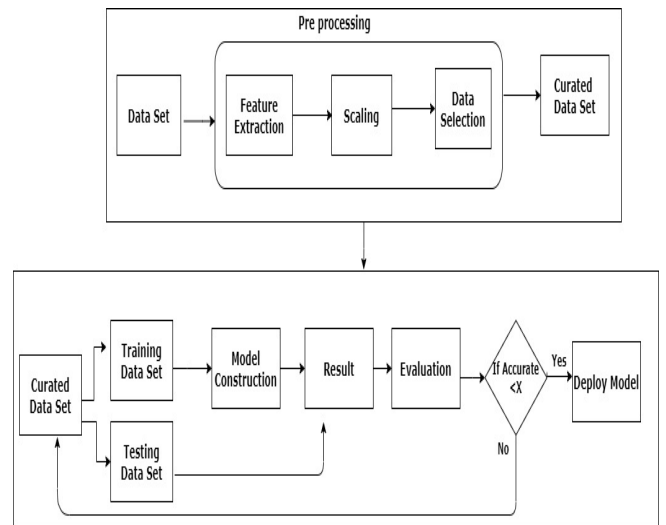

Figure 5: The Overall ML Pipeline

4.2. Dataset

We applied machine learning regression models to predict the concrete compression strength. The different techniques of regression considered in this study are linear regression models, random forest, support vector regression and decision tree regression model. The data set is collected from kaggle for the analysis. In this study the input variables are cement, fly ash, blast furnace slag, water, coarse Aggregate, super plasticizer, age and fine aggregate. The output variable is the concrete compressive strength which is measured in mega pascals.

4.3. Preprocessing

The machine learning preprocessing steps were applied to the raw datasets before they could be utilized for the regression

method. In preprocessing, the following steps need to be applied. They are feature selection, scaling and data selection. In feature selection, as per the expert opinion from the domain we need to consider only the six components from the dataset by omitting Super plasticizer and fly ash. The ML regression method estimates the output value using only the selected features from the dataset. Next step is scaling in which the component values are scaled to between zero and one by using minmax scaling procedure. Next step is data selection in which age of the cement in days is calculated. Here we consider three days which is 7, 14, and 28. After feature selection curated dataset is ready for further processing.

4.4. Regression Algorithms

The next phase in the pipeline was building the regression model. From the curated dataset 80% is selected as training set and the rest is kept testing set. On the training data the following regressions namely liner regression, support vector regression, decision tree regression and random forest regression are explained in the preliminary section. The model is build based on the type of regressions we applied.

5. IMPLEMENTATION AND RESULTS

The implementation of the work was done for evaluating the proposed model. The machine learning method for the compression strength prediction was modeled as a regression model. It was then implemented using python using the packages like numpy, sklearn, pandas, etc. The models were trained during a CPU based system with 8GB RAM.

For evaluating the model correctness, we have considered the RMSE value. The table below shows the RMSE value of the different regression models for the compression strength prediction for different curing time. We have considered 7, 14 and 28-days curing time in this study.

From the table V, it is clear that, for different curing days the random forest regression is giving least RMSE value which means better model. Also, the plots show the actual and predicted values for different regression models and for different curing period.

Table 1: RMSE values for different models Vs different curing time

Days	Linear	Support Vector	Decision Tree	Random Forest
7	13.1	15.6	30.7	3.1
14	14.5	24.5	33.1	4.5
28	12.7	35.6	50.3	2.7

The following graphs show the actual verses predicted compressive strength of the applied regression methods.
Figure 6, Figure7, and Figure 8 shows the result prediction graph of linear regression for 7, 14 and 28 days respectively. The following figure numbers Figure 9, Figure 10, and Figure 11represents compression strength prediction result of support

vector regression on different ages. Now the figure numbers Figure 12, Figure 13, and Figure 14 shows the result of random forest regression in different days. The compression strength prediction result of decision tree regression in 7, 14, 28 days is shown in Figure 15, Figure 16, Figure 17respectively.
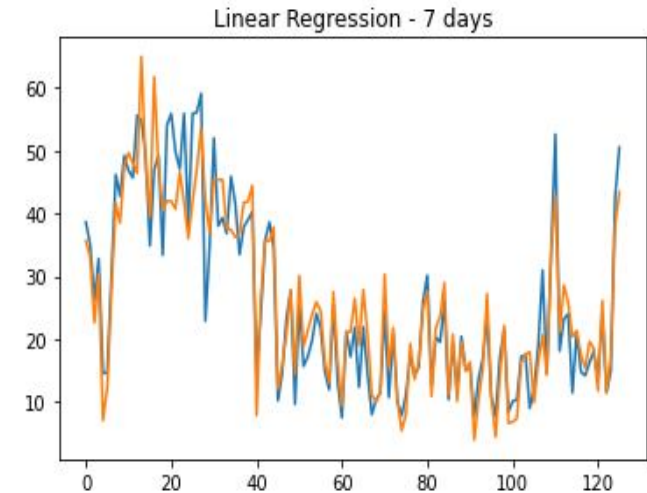

Figure 6: Compression Strength Prediction Graph - LR 7 Days
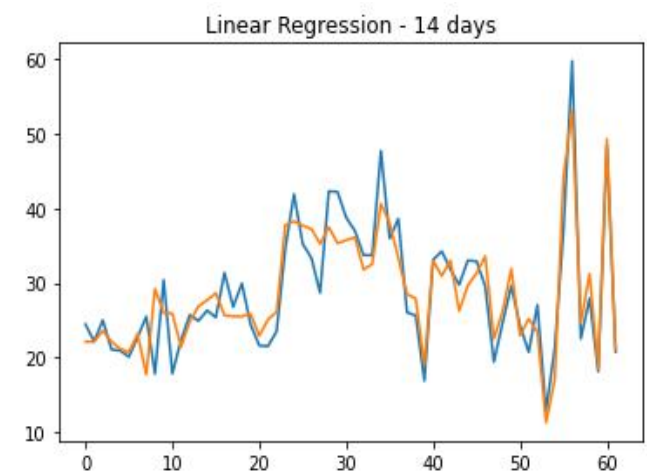

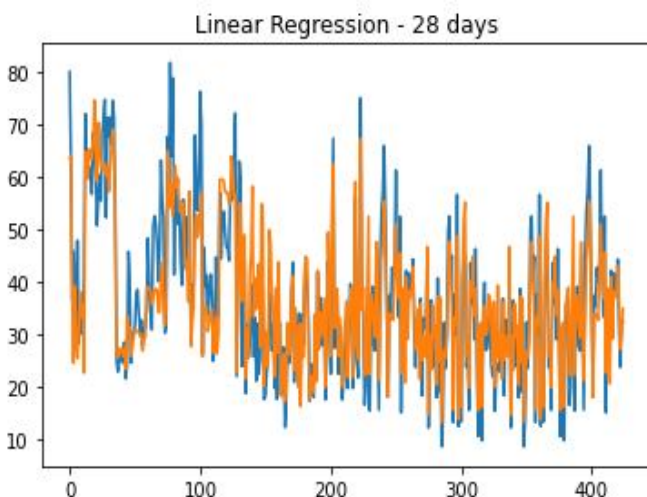Figure 7: Compression Strength Prediction - LR 14 Days

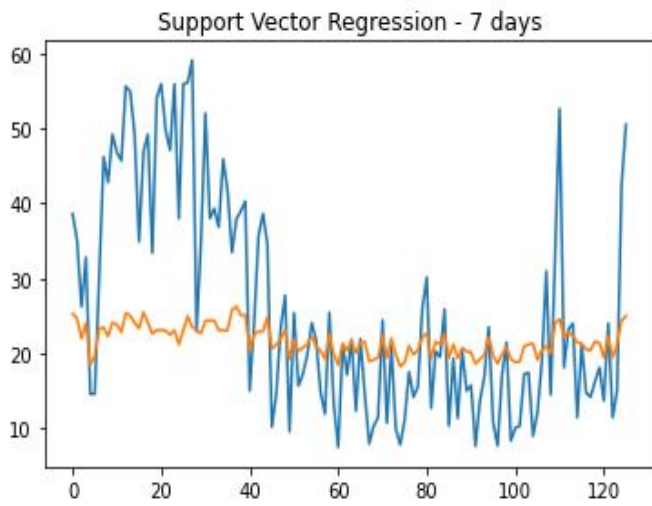Figure 8: Compression Strength Prediction - LR 28 Days

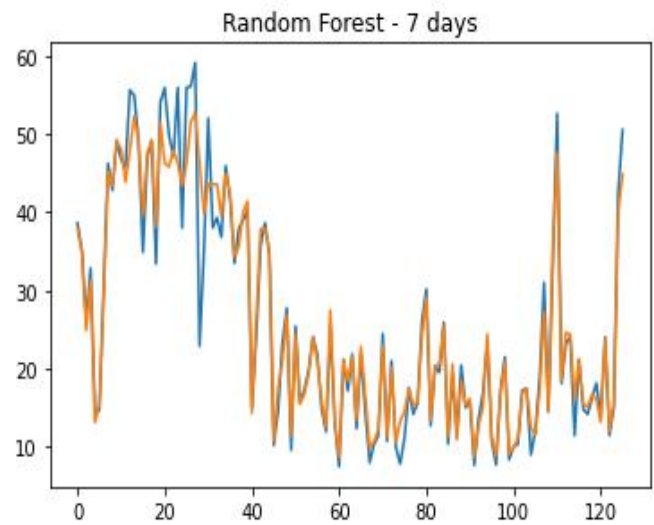Figure 9: Compression Strength Prediction Graph - SVR 7 Days

Figure 12: Compression Strength Prediction Graph - RF 7 Days

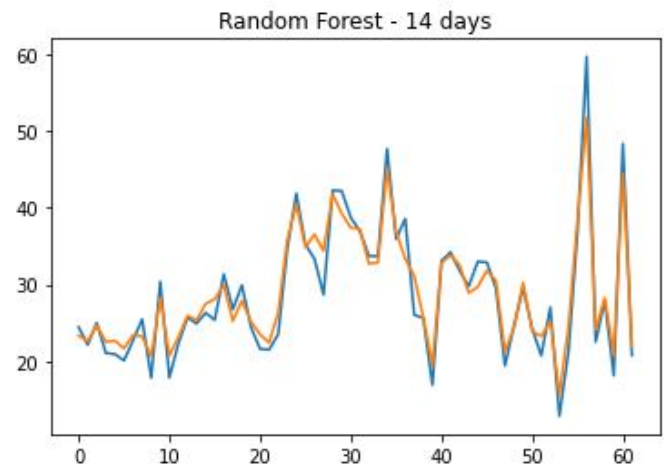Figure 10: Compression Strength Prediction Graph - SVR 28 Days

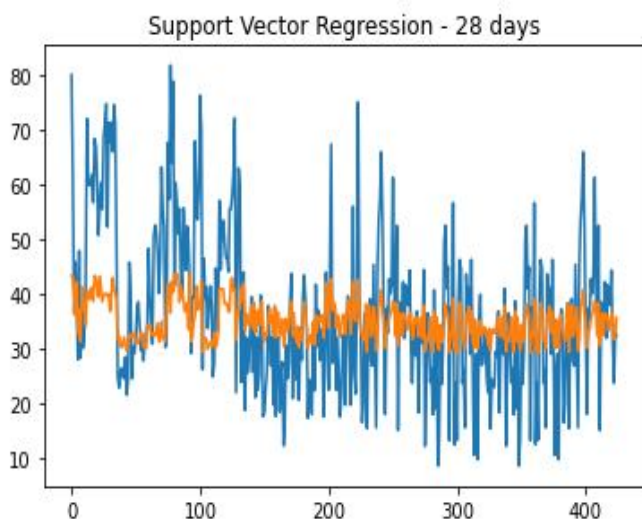FiFigure 13: Compression Strength Prediction Graph - RF 14 Days

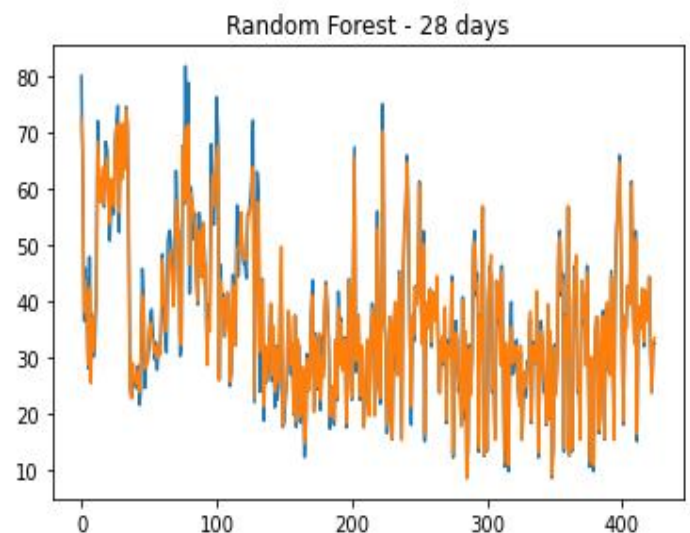Figure 11: Compression Strength Prediction Graph - SVR 14 Days

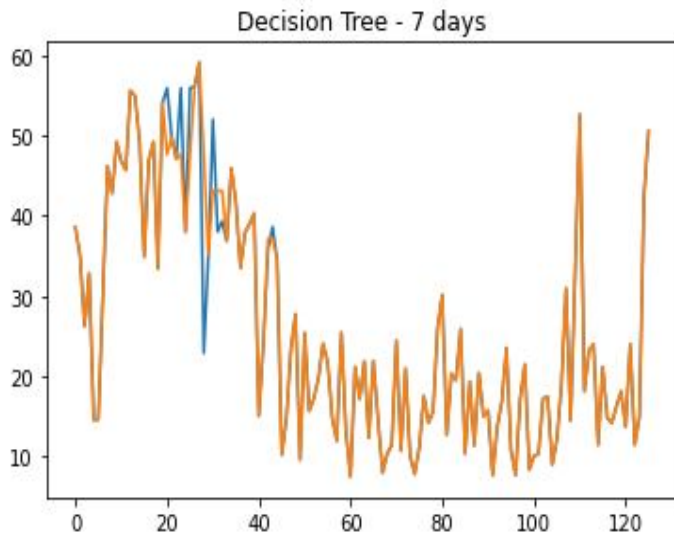Figure 14: Compression Strength Prediction Graph - RF 28 Days

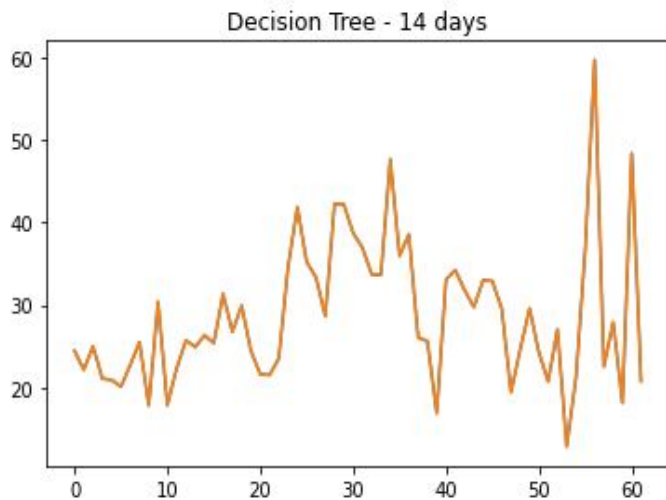Figure 13: Compression Strength Prediction Graph - DT 7 Days

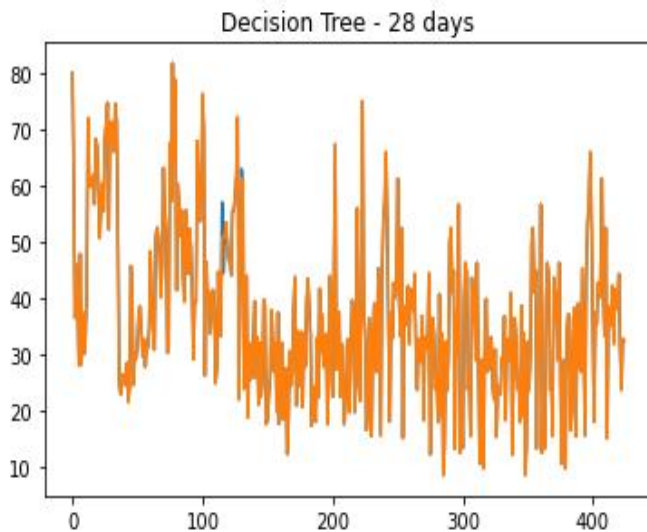Figure 14: Compression Strength Prediction Graph - DT 14 Days

Figure 15: Compression Strength Prediction Graph- DT 28 Days

6. CONCLUSION

The technological advancements in the field of machine learning pave a great way to reach interdisciplinary research application. One such problem in civil engineering, to find a mathematical relationship for the amount of different compo-nets used for creating a concrete mix and the compression strength was considered. In this paper we proposed a machine learning pipeline, using different regression techniques for predicting the compression strength of a particular concrete mix. From the raw dataset by apply feature selection 7 attributes where selected and pre-processed to create a curated dataset. Then different machine learning models were tried and Random Forest Regression gives less RMSE values and chosen as best regression model. Many such computationally solvable problems exist in civil engineering domain which has very high relevance in real time application.

7. REFERENCES

1.İ. B. Topçu and M. Sarıdemir, "Prediction of compressive strength of concrete containing fly ash using artificial neural networks and fuzzy logic," Computational Materials Science, vol. 41, no. 3, pp. 305–311, 2008.

2.F. Khademi, M. Akbari, S. M. Jamal, and M. Nikoo, "Multiple linear regression, artificial neural network, and fuzzy logic prediction of 28 days compressive strength of concrete," Frontiers of Structural and Civil Engineering, vol. 11, no. 1, pp. 90–99, 2017.

3.M. Timur Cihan, Tekirdag˘ Namık, C¸ orlu "Prediction of Concrete Compressive Strength and Slump by Machine Learning Methods" Vol. 2019 ,Article ID 3069046,2019

4.Celalettin Bas¸ yig˘it, Bekir C¸ omak, S¸ emsettinKılınc¸arslan, Ismail Serkan U¨ncu¨ "Assessment of concrete compressive strength by image processing technique"

5.H.-G. Ni and J.-Z. Wang, "Prediction of compressive strength of concrete by neural networks," Cement and Concrete Research, vol. 30, no. 8, pp. 1245–1250, 2000.

6.A. Baykasoglu, T. Dereli, and S. Tanis¸, "Prediction of cement strength using soft computing techniques," Cement and Concrete Research, vol. 34, no. 11, pp. 2083–2090, 2004.

7. S. Akkurt, G. Tayfur, and S. Can, "Fuzzy logic model for the prediction of cement compressive strength,"Cement and Concrete Research, vol. 34, no. 8, pp. 1429–1433, 2004.

8.A. Oztas ¨ ¸, M. Pala, E. Ozbay, E. Kanca, N. C ¨ ¸ aglar, and M. ˘ A. Bhatti, "Predicting the compressive strength and slump of high strength concrete using neural network," Construction and Building Materials, vol. 20, no. 9, pp. 769–775, 2006.

9.Palika Chopra,1 Rajendra Kumar Sharma,1 and Maneek Kumar2 "Prediction of Compressive Strength of Concrete Using Artificial Neural Network and Genetic Programming, "Hindawi Publishing Corporation Advances in Materials Science and Engineering Vol. 2016, Article ID 7648467, 10 pages

10. P. Muthupriyaa, K. Subramanianb, and B. Vishnuramc, "Prediction of compressive strength and durability of high preformance concrete by artifical neural networks " Int. J. Optim. Civil Eng, Vol. 1, pp. 189-209, 2011.

11. J. Sobhani, M. Najimi, A. R. Pourkhorshidi, and T. Parhizkar, "Prediction of the compressive strength of no-slump concrete: a comparative study of regression, neural network and ANFIS models," Construction and Building Materials, vol. 24, pp. 709-718, 2010

12. Y. J. Kim, J. Hu, S. J. Lee, and B. J. Broughton, "Prediction of Compressive Strength of Aerated Lightweight Aggregate Concrete by Artificial Neural Network," Applied Mechanics and Materials, vol. 84, pp. 177-182, 2011