Prediction of the Compressive Strength of High Performance Concrete Mix using Tree Based Modeling

C.Deepa
M.Phil ResearchScholar,
P.S.G.R. Krishnammal College for Women, Coimbatore, India

K.SathiyaKumari
Lecturer,GRGSACT,
P.S.G.R. Krishnammal College for Women, Coimbatore, India

V.Pream Sudha
Lecturer,GRGSACT,
P.S.G.R. Krishnammal College for Women, Coimbatore, India

ABSTRACT
Concrete is the safest and sustainable construction material which is most widely used in the world as it provides superior fire resistance, gains strength over time and gives an extremely long service life. Its annual consumption is estimated between 21 and 31 billion tones. Designing a concrete mix involves the process of selecting suitable ingredients of concrete and determining their relative amounts with the objective of producing a concrete of the required, strength, durability, and workability as economically as possible. According to the National Council for Cement and Building Materials (NCBM), New Delhi, the compressive strength of concrete is governed generally, by the water-cement ratio. The mineral admixtures like fly ash, ground granulated blast furnace, silica fume and fine aggregates also influence it. The main purpose of this paper is to predict the compressive strength of the high performance concrete by using classification algorithms like Multilayer Perceptron, M5P Tree models and Linear Regression. The result from this study suggests that tree based models perform remarkably well in predicting the compressive strength of the concrete mix.

General Terms
Concrete strength, Modeling Approaches, Experiments, Trend line analysis and Algorithms.

Keywords
Multilayer Perceptron, M5P Tree, Linear Regression.

1. INTRODUCTION
Concrete is an essential material in civil engineering. The property of concrete differs depending on various factors. The proportions of its constituents, the construction methods, the loading and environmental conditions to which it will be subjected over time are some of the issues that determine its strength. Therefore, the development of control methods to determine the condition and ascertain the quality of concrete is critical. The concrete compressive strength is a complex non-linear regression problem for construction engineering. It is highly difficult to predict the concrete strength due to non-linearity. Concrete testing is performed in order to determine whether specified strength requirements are met.

Concrete has high importance in civil engineering and is widely used for many kinds of structures. Its service is considered to be tremendously long exclusive of the disaster caused by earthquakes. It is the most important obstacle to be overcome and taken into account by the experts. In order to simulate the behavior of the structure, concrete compressive strength must be known. But it cannot be guessed easily due to its ingredients and processes.

Prediction of concrete strength is important for concrete construction as it gives an idea about the time for concrete form removal, project scheduling and quality control. Several approaches using regression functions have been proposed for predicting the concrete strength. For this purpose, concrete compressive dataset has been collected from a real time construction company. The concrete data sets have 300 instances that contain 8 attributes and a class attribute.

Data driven techniques like Linear Regression analysis, MLP (Multilayer Perceptron) and M5P modal trees are used in this study. These techniques have been applied to civil engineering problems in general and structural engineering, in particular. Also, comparative study between the different techniques has been carried out in different environments like strength of concrete mix.

A ready to use relationship between the strength of concrete and the properties of ingredients using linear regression analysis, multilayer perceptron and M5P was established. The approaches, capable of predicting reliably the compressive strength of hardened concrete based on the properties of the ingredients and wet concrete, were complementary to the existing workability tests routinely carried out during concreting.

It is found that there is a good correlation between these three algorithms. Time taken to build the model for MLP is high when compared to other algorithms. Linear regression had taken less time to build the models. In M5P model Error rate is less than other two algorithms. When compared to MLP and Linear Regression algorithm M5P tree algorithm is much better.

1.1 Concrete
Concrete is the only major building material that can be delivered to the job site in a plastic state. This unique quality makes concrete attractive as a building material because it can be casted to almost any form or shape. Concrete provides a wide latitude in surface textures and colors and can be used to construct a wide variety of structures, such as highways and streets, bridges, dams, large buildings, airport runways, irrigation

structures, breakwaters, piers and docks sidewalls, silos and farm buildings, homes, and even barges and ships.

The two major components of concrete are a cement paste and inert materials. The cement paste consists of Portland cement, water, and some air either in the form of naturally entrapped air voids or minute, intentionally entrained air bubbles. The inert materials are usually composed of fine aggregate, which is a material such as sand, and coarse aggregate, which is a material such as gravel, crushed stone, or slag.

When Portland cement is mixed with water, the compounds of the cement react to form a cementing medium. In properly mixed concrete, each particle of sand and coarse aggregate is completely surrounded and coated by this paste, and all spaces between the particles are filled with it. As the cement paste sets and hardens, it binds the aggregates into a solid mass. Under normal conditions, concrete grows stronger as it grows older. The chemical reactions between cement and water that cause the paste to harden and bind the aggregates together require time. The reactions take place very rapidly at first and then more slowly over a long period of time.

1.2 Cement

Cement is a material that has adhesive and cohesive properties enabling it to bond mineral fragments into a solid mass. Cement consists of silicates and aluminates of lime made from limestone and clay (or shale) which is ground, blended, fused in a kiln and crushed to a powder. Cement chemically combines with water (hydration) to form a hardened mass. The usual hydraulic cement is known as Portland cement because of its resemblance when hardened to Portland stone found near Dorset, England. The name was originated in a patent obtained by Joseph ASP din of Leeds, England in 1824.

Typical Portland cements are mixtures of tricalcium silicate (3CaO • SiO2), tricalcium aluminates (3CaO • Al2O3), and dicalcium silicate (2CaO • SiO2), in varying proportions, together with small amounts of magnesium and iron compounds. Gypsum is often added to slow the hardening process.

1.3 Water

The water has two roles in concrete mixture: First is the chemical composition with cement and perform cement hydration and second is to make the concrete composition fluent and workable. The water which is used to make the concrete is drink water. The impurity of water can have undesirable effect on concrete strength.

1.4 Aggregates

Since aggregate usually occupies about 75% of the total volume of concrete, its properties have a definite influence on behavior of hardened concrete. Not only does the strength of the aggregate

Affect the strength of the concrete; its properties also greatly affect durability (resistance to deterioration under freeze-thaw cycles). Since aggregate is less expensive than cement it is logical to try to use the largest percentage feasible. Hence aggregates are usually graded by size and a proper mix has specified percentages of both fine and coarse aggregates. Fine aggregate (sand) is any material passing through a No. 4 sieve. Coarse aggregate (gravel) is any material of larger size.

Fine aggregate provides the fineness and cohesion of concrete. It is important that fine aggregate should not contain clay or any chemical pollution. Also, fine aggregate has the role of space filling between coarse aggregates. Coarse aggregate includes: fine gravel, gravel and coarse gravel In fact coarse aggregate comprises the strongest part of the concrete. It also has reverse effect on the concrete fineness. The more coarse aggregate, the higher is the density and the lower is the fineness.

1.5 Compressive strength of concrete

The strength of concrete is controlled by the proportioning of cement, coarse and fine aggregates, water, and various admixtures. The ratio of the water to cement is the chief factor for determining concrete strength. The lower the water-cement ratio, the higher is the compressive strength. A certain minimum amount of water is necessary for the proper chemical action in the hardening of concrete; extra water increases the workability (how easily the concrete will flow) but reduces strength. A measure of the workability is obtained by a slump test. Actual strength of concrete in place in the structure is also greatly affected by quality control procedures for placement and inspection.

2 MODELLING APPROACHES

The modeling approaches used for this study are briefly described below.

2.1 Multilayer Perceptron

An MLP is a network of simple *neurons* called *perceptrons*. The perceptron computes a single *output* from multiple real-valued *inputs* by forming a linear combination according to its input *weights* and then possibly putting the output through some nonlinear activation function. MLP networks are typically used in *supervised learning* problems. The supervised learning problem of the MLP can be solved with the *back-propagation algorithm*. The algorithm consists of two steps. In the *forward pass*, the predicted outputs corresponding to the given inputs are evaluated. In the *backward pass*, partial derivatives of the cost function with respect to the different parameters are propagated back through the network.

A multi-layer perceptron is especially useful for approximating a classification function that maps input vector (x_1,x_2, ... x_n) to one or more classes C_1,C_2,...C_m. By optimizing weights and thresholds for all nodes, the network can represent a wide range of classification functions. Optimizing the weights can be done by supervised learning, where the network learns from the large number of examples. Examples are usually provided one at a time. For each example the actual vector is computed and compared to the desired output. Then, weights and thresholds are adjusted, proportional to their contribution to the error made at the respective output. One of the most used methods is the back-propagation method, in which in the iterative manner, the errors are propagated (error = the difference between desired output and the output of actual ANN) into the lower layers, to be used for the adaptation of weights.

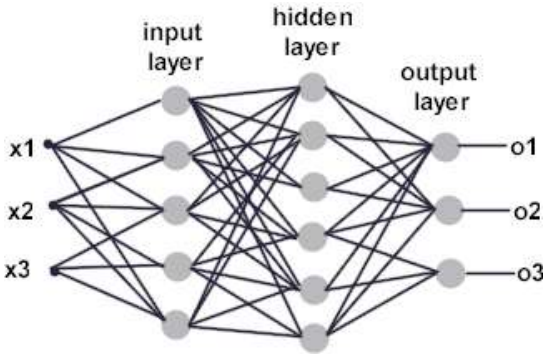

Figure 1: Multilayered Perceptron with 3 layers

This network has an **input layer** (on the left) with three neurons, one **hidden layer** (in the middle) with three neurons and an **output layer** (on the right) with three neurons.

There is one neuron in the input layer for each predictor variable. In the case of categorical variables, *N*-1 neurons are used to represent the *N* categories of the variable.

Input Layer — A vector of predictor variable values ($x_1 ... x_p$) is presented to the input layer. The input layer (or processing before the input layer) standardizes these values so that the range of each variable is -1 to 1. The input layer distributes the values to each of the neurons in the hidden layer. In addition to the predictor variables, there is a constant input of 1.0, called the *bias* that is fed to each of the hidden layers; the bias is multiplied by a weight and added to the sum going into the neuron.

Hidden Layer — Arriving at a neuron in the hidden layer, the value from each input neuron is multiplied by a weight (w_{ji}), and the resulting weighted values are added together producing a combined value u_j. The weighted sum (u_j) is fed into a transfer function, σ, which outputs a value h_j. The outputs from the hidden layer are distributed to the output layer.

Output Layer — Arriving at a neuron in the output layer, the value from each hidden layer neuron is multiplied by a weight (w_{kj}), and the resulting weighted values are added together producing a combined value v_j. The weighted sum (v_j) is fed into a transfer function, σ, which outputs a value y_k. The *y* values are the outputs of the network.

In MLP the computation layer-wise is done in the forward direction whereas weight adjustments are done in backward direction. It is easy to compute the error for nodes in the output layer as the actual outcome and desired results are known. The error computed in the last layer is propagated backwards as the desired result for the nodes in hidden layers are unknown. This process gives the change in the weight for the edges layer–wise. This standard method used in training MLPs is called the back propagation algorithm

2.2 M5P Tree model

M5P is a reconstruction of Quinlan's M5 algorithm for inducing trees of regression models. M5P combines a conventional decision tree with the possibility of linear regression functions at the nodes. First, a decision-tree induction algorithm is used to build a tree, but instead of maximizing the information gain at each inner node, a splitting criterion is used that minimizes the intra-subset variation in the class values down each branch. The splitting procedure in M5P stops if the class values of all instances that reach a node vary very slightly, or only a few instances remain. Second, the tree is pruned back from each leaf. When pruning an inner node is turned into a leaf with a regression plane.

Model trees are a sub-class of regression trees having linear models at the leaf node. In comparison with classical regression trees, model trees deliver better compactness and prediction accuracy. These advantages issue from the ability of model trees to leverage potential linearity at leaf nodes.

The model tree algorithm used in this work is based on M5P an optimized, open-source implementation of the classical M5P algorithm.

The algorithm known as the M5P algorithm is used for inducing a model tree, the aim is to construct a model that relates a target value of the training cases to the values of their input attributes. The quality of the model will generally be measured by the accuracy with which it predicts the target values of the unseen cases.

2.3 Linear Regression model

In general, the goal of linear regression is to find the line that best predicts Y from X. Linear regression does this by finding the line that minimizes the sum of the squares of the vertical distances of the points from the line.

Linear regression attempts to model the relationship between two variables by fitting a linear equation to observed data. One variable is considered to be an explanatory variable, and the other is considered to be a dependent variable. For example, a modeler might want to relate the weights of individuals to their heights using a linear regression model. Linear Regression is an excellent, simple scheme for numeric prediction. It is used for classification in domains with numeric attributes. The linear models serve very well as building blocks for more complex learning schemes. Linear regression analysis is carried out to establish a relationship between the parameters listed.

3 EXPERIMENTAL DESIGN

Experiments were performed using the machine learning tool with default parameters for the base learners. The data mining method used to build the model is classification. The WEKA, Open Source, Portable, GUI-based workbench is a collection of state-of-the-art machine learning algorithms and data pre processing tools. The real time data set consists of 300 instances with 9 different attributes. The instances in the dataset are pertaining to the attributes to represent the ingredient of concrete. To predict the compressive strength, in case of high strength concrete water to binder ratio, water content, fine aggregate ratio, fly ash replacement ratio, silica fume replacement ratio and super plasticizer were used as input parameters while compressive strength of concrete at all ages was used as output. The performance of the classifiers is evaluated and their results are analyzed.

Table 1: Concrete Compressive Strength Real Time Data Set

Parameter	Range	Avg	Std-Dev
Cement	55.8 – 491	190.58	88.61
Blast Furnace Slag	0 - 305.3	38.04	61.03
Fly ash	0 - 141.0	14.37	41.64
Water	32.4 - 930.0	121.99	80.74
Super plasticizer	0 - 12.3	1.3	2.9
Coarse aggregate	185.2- 1083.4	593.49	232.26
Fine Aggregate	118.8 - 942	469.81	209.20
Age	1 – 365	49.75	70.96
Compressive strength	3.32 – 79.99	33.53	15.58

In general, tenfold cross validation has been proved to be statistically good enough in evaluating the performance of the classifier. The 10-fold cross validation was performed to test the performance of the concrete compressive strength. The purpose of running multiple cross-validations is to obtain more reliable estimates of the risk measures. This technique consists of dividing the overall data in 10 disjoint subsets, or folds. Each algorithm is then trained using 9 of the subsets and evaluated using the tenth subset. The process is repeated 10 times and each time, a different subset is used for testing and the remaining 9 subsets are used to train the model. The algorithm is evaluated by averaging the prediction metrics from the 10 different models. Several prediction metrics can be employed to predict the accuracy for different algorithms.

4 RESULTS AND DISCUSSION

The performance of the MLP, Linear Regression and M5P tree techniques are given in Table.2. The predicted high strength values are not much more higher than the experimental compressive strength values in the real time data set. Comparitively, M5P algorithm has shown the lowest RMSE and MAE .It also has high correlation among the other algorithms.

4.1 Model assessment

To examine how close the predicted to the compressive strength of HSC mixtures, three indices, mean absolute error MAE, root mean square error RMSE and correlation coefficient R, were employed to evaluate the performance of the algorithms based on mixtures properties

$$\text{RMSE} = \sqrt{1/n\sum_{i=1}^{n}(Pi - Ai)^2}$$

$$\text{MAE} = 1/n\sum_{i=1}^{n}|Pi - Ai|$$

$$R = \frac{\sum_{i=1}^{n}(Pi - \overline{P})(Ai - \overline{A})}{\sqrt{\sum_{i=1}^{n}(Pi - \overline{P})^2 \sum_{i=1}^{n}(Ai - \overline{A})^2}}$$

Where Ai is the compressive strength of concrete mixtures, Pi is the predicted value, n is the total number of data points in validation, A is the mean value of observations, and P is the mean value of predictions.

Table 2: Prediction accuracy for different algorithms

Techniques	Correlat-ion	RMSE	MAE	Time taken In (sec)
Multilayer perceptron	0.7908	9.9054	7.678	2.06
Linear regression	0.7009	11.1066	8.8388	0.02
M5P model tree	0.8872	7.1874	5.008	0.41

Figure 2 shows that MLP takes more time(2.06 secs) to produce the result whereas the time taken by linear regression is the least (0.02 secs). M5P takes an average time of 0.41 seconds to build its model

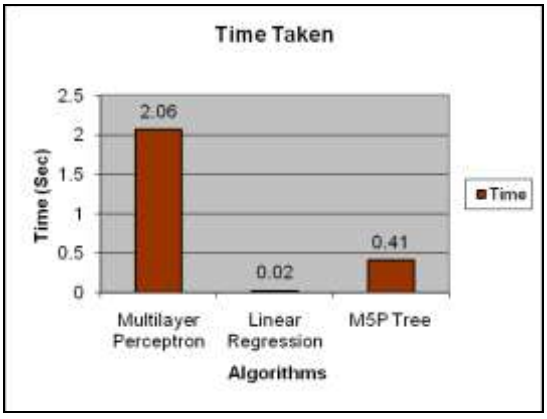

Figure 2: Time taken to build the model

M5P has the correlation coefficient closer to +1, indicating a strong correlation between variables. Figure 3, shows that correlation coefficient for M5P Tree is higher when compared to other algorithms.

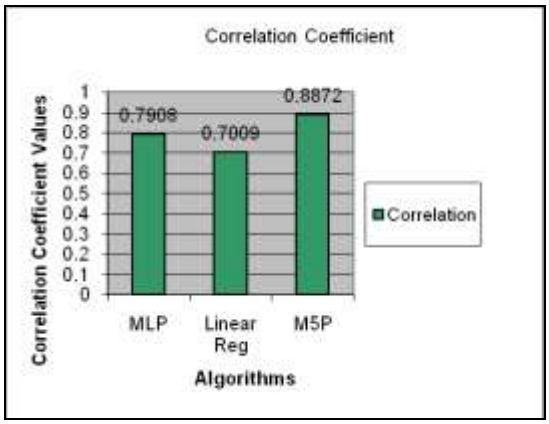

Figure 3: Correlation coefficent for different algorithms

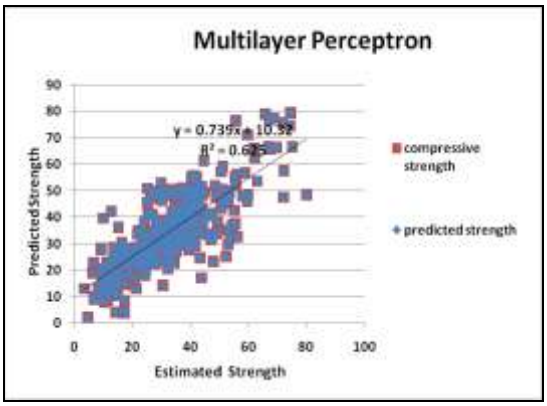

Figure 4: Plot between actual strength Vs Predicted strength by MLP

4.2 Trend line Analysis

Trend lines are used to graphically display trends in data and to help analyze problems of prediction. In a chart trend lines can be extended beyond the actual data to predict future values. A linear trend line is a best-fit straight line that is used with simple linear data sets. A linear trend line usually shows that something is increasing or decreasing at a steady rate. A linear trend line uses the following equation to calculate the least squares fit for a line: y= mx +b, where *m* is the slope and *b* is the intercept.

Consider a dataset having (x_i, y_i) .If there exists a **linear relationship** between the variables **x** and **y**, the data can be plotted and a "best-fit" *straight line* can be generated through the data. Of course, this relationship is governed by the familiar equation $y = mx + b$ with the **slope, *m*,** and **y-intercept, *b*,** for the data. The Trend lines generated by three algorithms are discussed below which are shown in the figures below.

A difficult task with the MLP method is choosing the number of hidden nodes. There is no theory yet to tell how many hidden units are needed to approximate any given function. The network geometry is problem dependent. Here, the three-layer MLP with one hidden layer is used and the common trial-and-error method is used to select the number of hidden nodes. Before applying the MLP method, the input data were normalized to fall in the range [0, 1]. The sediment concentration data were also standardized in a similar way. These normalized data were used to train each of the ANN models. After training was over, the weights were saved and used to test (validate) each network performance on test data. In Figure 4,

The MLP algorithm predict the equation as

$$Y = 0.739 x + 10.32$$

and $R^2 = 0.625$

In Figure 5, the M5P input space is recursively partitioned until the data at the leaf nodes constitute relatively homogeneous subsets such that a linear model can explain the remaining variability. This divide-and-conquer approach partitions the training data and provides rules for reaching the models at the leaf nodes. The linear models are then used to quantify, in a statistically rigorous way, the contribution of each attribute to the overall predicted value. A powerful aspect of the prediction model arrived at in this way is that it is interpretable, in contrast with other machine learning approaches, such as neural networks.

Equation developed using M5P Modal Trees is:

Concrete compressive strength = 0.1324 * Cement + 0.0288 * Blast Furnace Slag + 0.1239 * Fly Ash - 0.2619 * Water + 0.0517 * Superplasticizer - 0.0081 * Coarse Aggregate - 0.0127 * Fine Aggregate + 0.0585 * Age (day) + 48.9924

The M5P Tree model predict the equation as

$$Y = 0.777x + 7.073$$

and $R^2 = 0.787$

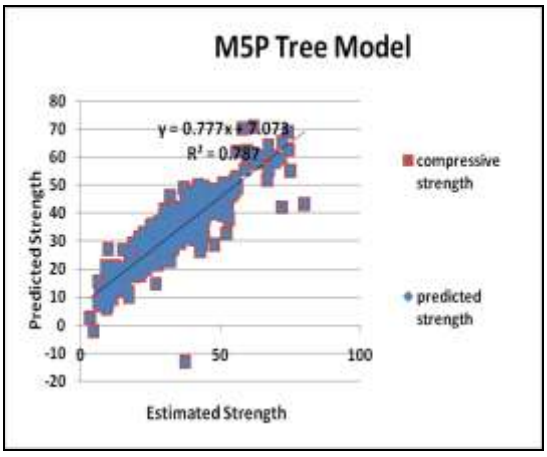

Figure 5: Plot between actual strength Vs predicted strength by M5P Tree

In Figure 6, the goal of linear regression is to adjust the values of slope and intercept to find the line that best predicts Y from X. Linear regression analyzes the relationship between two variables, X and Y. For each subject, have both X and Y and need to find the best straight line through the data. In some situations, the slope and/or intercept have a scientific meaning. In other cases to use the linear regression line as a standard curve to find new values of X from Y, or Y from X.

Equation developed using Linear Regression is:

Concrete compressive strength = 0.083 * Cement + 0.0601 * Blast Furnace Slag + 0.0826 * Fly Ash + 0.6622 * Superplasticizer + (-0.0147 * Coarse Aggregate) + (-0.0439 * Fine Aggregate) + 0.0883 * Age + 38.3411

Linear Regression model predict the equation as

$$Y = 0.514x + 16.30$$

and $R^2 = 0.491$

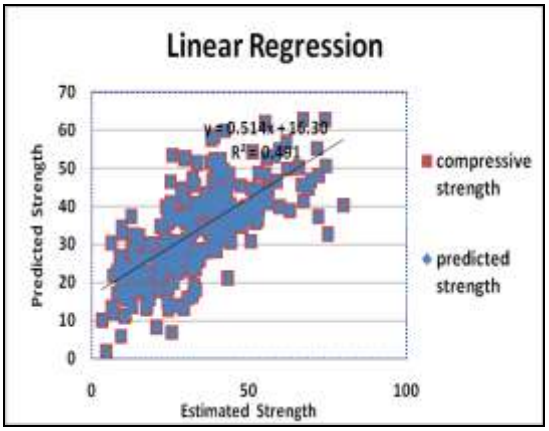

Figure 6: Plot between Actual strength Vs Predicted Strength by linear regression model

R^2 is the square of the correlation coefficient R. The R^2 value describes how much of the variation in the data is accounted for by the trend line. The closer the R value is to +1 or –1, the better the line fits the data. MLP and Linear regression has an R value of 0.625 and 0.491 respectively whereas M5P has its R value as 0.787 which is closer to +1 showing that the line generated by M5P best fits the data.

Figure 7, Shows the results obtained by the M5P Classifier Tree Model. Successful analysis and prediction should be always based on the use of various types of models. Different models, although in close accuracy, offer various advantages over each other.

- MLP approach gives results with good prediction and has an inbuilt flexibility for choosing any number of independent variables without assuming an explicit equation. But it requires non-linear optimization with the possibility of converging only in local minima.

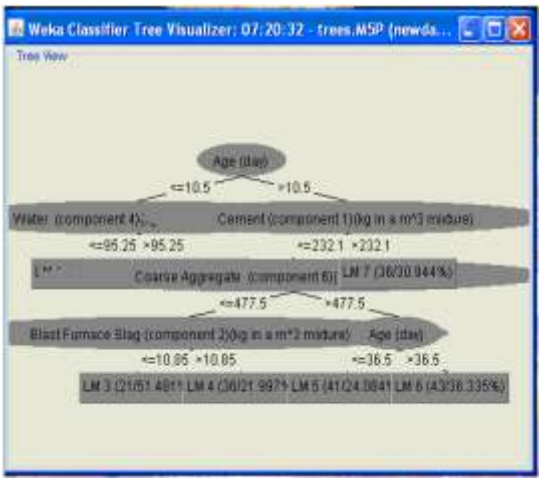

Figure 7: M5P Classifier Tree Model

- Prediction of strength by linear regression is found to be adequate and the approach can be easily adopted for ready use because of the explicit nature of the strength equation.
- Modal Trees have advantages in both compactness and prediction accuracy, attributable to the ability of modal trees to use the local linearity in the data. It is more understandable and allows one to build a family of models of varying complexity and accuracy.

5 CONCLUSION

In this paper, three data mining algorithms are applied on the task of classifying concrete compressive strength data set and the most accurate learning method is evaluated. With high correlation coefficient, lesser RMSE and MSE, the M5P tree is found to work well for this dataset in comparison to MLP and Linear regression. Trend line analysis also gives an approving result for M5P. Result with the data set suggests that tree based modeling approach can effectively be used in predicting the compressive strength of high performance concrete.

6 FUTURE WORK

There are a number of potential avenues for further work. In the future, the work can be extended by applying different data mining techniques. An in-depth evaluation of why different algorithms exhibit different classification accuracy and computational performance can be performed. Investigating the robustness of ML classification and a comparison between ML and non-ML techniques on an identical dataset would also be valuable.

7 REFERENCES

[1]. Yogesh Aggarwal, "Modeling of Reinforcement in Concrete Beams Using Machine Learning Tools", World Academy of Science, Engineering and Technology 32, 2007.

[2]. S.M. Gupta, "Support Vector Machines based Modelling of Concrete Strength", World Academy of Science, Engineering and Technology 36, 2007.

[3]. Vahid. K. Alilou & Mohammad. Teshnehlab, "Prediction of 28-day compressive strength of concrete on the third day using artificial neural networks" International journal of Engineering, Vol (3) , 565-575.

[4]. Serkan Subasi, "Prediction of mechanical properties of cement containing class C fly ash by using artificial neural network and regression technique", Academic Journals Vol.4 940 pp.289-297 April 2009.

[5]. Noorzaei J., Hakim S.J.S., Jaafar M.S., Thanoon W.A.M. "Predicting the compressive strength and slump of high strength concrete using neural network", *International Journal of Engineering and Technology*, Vol. 4, No. 2, 2007, pp. 141-153.

[6]. H. Witten and E. Frank, *Data Mining: ractical Machine Learning Tools and Techniques with Java Implementation*. Morgan Kaufmann Publisher, 2000.

[7]. L. Breiman, J. Friedman, R. Olshen, and C. Stone. *Classification and Regression Trees*. Wadsworth International Group, 1984.

[8]. Y.Wang and I.Witten. Inducing model trees for continuous classes. In *Proceedings of the 9th European Conf. on Machine Learning, Poster Papers*, 1997.

[9]. R. Quinlan. Learning with continuous classes. In *Proceedings of the 5th Australian Joint Conference on Artificial Intelligence (AI'92)*, 1992.

[10]. Jong In Kim, Doo Kie Kim, "Application of neural networks for Estimation of Concrete Strength', KSCE Journal of Civil Engineering, 6(4): 429-438, 2002.

[11]. Rishi. Garge. "Concrete Mix Design using Artificial Neural Network", m.sc Thesis, Thapar Institute of Engineering and Technology, June 2003.

[12]. I-Cheng Yeh. "Design of High-Performance Concrete Mixture Using Neural Networks and Nonlinear Programming". Journal of Computing in Civil Engineering, Vol. 13, No. 1, January, 1999.

[13]. Sergio Lai and Mauro Serra. " Concrete strength prediction by means of neural network", Constntction and Building Materials, Vol. 11, No. 2, 1997 pp. 93-98.

[14]. R. Kohavi. A study of cross-validation and bootstrap for accuracy estimation and model selection. In *Proceedings of 14th International Joint Conference on Artificial Intelligence*, 1995.

[15]. Rangwala, "Engineering Materials (Material Science), by Charotar Publishing 28th edition 2001.