Concrete Strength Prediction using Machine Learning Algorithms

Thrishal Vignesh Vajha[1], Prashanth Ragam[1], Megha Syam Bandlamudi[1], Jaya Kumar[1], and Dharanesh[1]

[1]School of Computer Science and Engineering, VIT-AP University, Near Vijayawada, 522237, Andhra Pradesh, India
thrishalvv@gmail.com, prashanth.rajam429@gmail.com, meghasyambandlamudi2355@gmail.com, jayakumar.s0134@gmail.com, dharaneesh

Abstr[2]act. Concrete is one of the most important materials required for the survival of a man in modern day. The strength of concrete is one of the most frequently utilized metric to asseess the strength of cement. Estimating the strength of concrete accurately is important for ensuring the safety of constructions in infrastructure.Some Machine learning algorithms are used for developing efficient ways to predict the concrete strength. Input parameters such as Cement, Blast Furnace Slag, Fly Ash, Water, Superplasticizer, Coarse Aggregate, Fine AggregateAge are used.This paper focuses mostly on estimating concrete strength by utilizing a variety of machine learning methods. The models utilized here are ensemble Linear Regression, SVR Kernel, Random forest, and Neural Networks out of which ensemble Neural Networks showed maximum accuracy with the R^2 of 0.872, and MAE of 4.27733.

Keywords: Concrete Strength, SVR Kernel, Random forest, Neural Networks, Linear Regression

1 Introduction

Concrete is very important Building material. In a country like India, where constructions are increasing rapidly ,Accurate Concrete Strength Prediction is essential to ensure the safety of Buildings. From a long time, The strength of concrete is simply predicted by using a formula based on material composition and conditions. But now with some Machine Learning Techniques, We can improve the accuracy of these prediction models. ML models can give idea about unusual patterns and exceptions. This helps Engineers to explore ML techniques and develop a more accurate prediction models. In this research paper, we explore machine learning techniques to predict concrete strength. Various Machine Learning techniques like Linear Regression, Neural Networks, SVR Kernel and Random forest to improve the efficiency of prediction.These algorithms help to predict the concrete strength which is way better than using emprical formulas to predict the strength. This work emphasizes concrete strength prediction with the help of machine learning (ML) algorithms. It is vital to predict concrete strength for longevity of buildings in infrastructure. So ML algorithms are used to predict the strength from previous data. Various ML algorithms such as Linear Regression, Neural Networks, SVR Kernel and Random forest are used to predict the strength based on various parameter like Cement, Blast Furnace Slag, Fly Ash, Water, Superplasticizer, Coarse Aggregate, Fine AggregateAge etc. By using the algorithms mentioned above, from the results, it can be concluded that the Neural Networks, a manchine learning model showed the maximum accuracy.

2 Concrete Strength

Determining concrete strength involves a comprehensive approach, starting with a precise mix design tailored to project specifications, including cement type and content, aggregate properties, and water-cement ratio. Testing methods, primarily compressive strength tests on cylindrical specimens, assess the concrete's strength under controlled conditions. Proper curing, through methods like water curing or curing compounds, is essential for optimal strength development. Quality control measures monitor materials, mixing processes, and concrete specimens to ensure desired strength levels are achieved. Environmental factors such as temperature and exposure to chemicals also impact strength. By managing these inputs effectively, engineers can ensure the concrete meets structural requirements for durability and safety.

Table 1. Minimum, maximum, and standard deviation of the parameters

Parameters	Minimum	Maximum	Standard deviation
Cement (in mm)	102	540	104.45562
Blast furnance slag (urea) (kg/acre)	0	359.4	86.237448

Fly ash(˚C)	0	200.1	63.96593
Water (N)	121.8	247	21.34385
Superplasticizer (P)	2.5	32.2	5.970940
Coarse Aggregate (K)	801	1145	77.7162
Fine Aggregate (Q/acre)	594	992.6	80.13705
Age	1	365	63.13923

3 Machine Learning Techniques

The data collection has become more important in every domain of human progress because of development of modern technologies. All organizations need large amounts of data to come to a conlusion be it Financial activities,Agriculture activities and Infrastructure development. At the same time, Data Engineers are looking on unusual patterns and trends that can only be found by Machine Learning Techniques but not by a classical formula or a trends. From analyzing previous data, it will become easy to predict the possible future outcomes. These algorithms analyses previous data and use what to make better conclusions for the future. Machine Learning algorithms are different from other techniques and can analyze from very large amounts of data and data with minimum structure. It facilitates Machine Learning algorithms to be effective in every domain compared to formukla-based appraoches.

In this current paper:Four Machine Learning algoriths which includes Random forest, Linear Regression, SVR Kernel and Neural Networks are used and developed to predict the concrete strength.

3.1 Random Forest

This supervised machine learning technique is the most popular and effective for problems involving classification and regression. This method generates a large number of decision trees during training and in the process of producing class outputs. A bagging approach called random forest (RF) uses many decision trees on subsets of an observation set and averages the results to increase the estimated accuracy of the dataset.

Random Forest gathers the forecasts from every tree and uses the popular vote of guesses to determine the final result. The accuracy increases and the chance of mistakes decreases with the number of trees in the forest. A Random Forest has two random components. They are as follows:

1. An arbitrary subset of attributes.
2. Bootstrap data samples.

A random forest [3] is a collection of trees, each of which makes a forecast. The trees then pool their predictions, utilising the mean, mode, and median of the collection to determine the forest's prediction based on continuous or categorical data. This seems reasonable to a larger degree. However, since each tree had a unique set of circumstances, it's possible that the majority of them produced forecasts based on purely random possibilities.

3.2 SVR Kernel

SVR, or Support Vector Regression, uses different kernel functions to find relationships in data. The kernel function decides how the data is transformed to make predictions. Common kernel functions include:

1. Linear Kernel: Creates straight-line predictions.

2. Polynomial Kernel: Allows for curved predictions of different degrees.

3. Radial Basis Function (RBF) Kernel: Useful for complex, non-linear relationships.

4. Sigmoid Kernel: Less commonly used, but still has its applications**.**

Choosing the right kernel depends on the data and problem you're dealing with. Experimentation helps find the best one.

3.3 Linear Regression.

Linear regression is a fundamental technique in machine learning used to model the relationship between a dependent variable (often denoted as yy) and one or more independent variables (often denoted as xx). It's called "linear" because the relationship between the independent variables and the dependent variable is assumed to be linear, meaning it can be represented by a straight line.

The basic idea behind linear regression is to find the best-fitting straight line through the data points. This line is determined by estimating the coefficients (slope and intercept) that minimize the difference between the actual observed values of the dependent variable and the values predicted by the linear model.

The most common method for estimating the coefficients in linear regression is called Ordinary Least Squares (OLS). This method finds the coefficients that minimize the SSE, typically using optimization techniques like gradient descent.

Once the coefficients are estimated, the linear regression model can be used to make predictions on new data by plugging in the values of the independent variables into the equation.

3.4 Neural Networks Neural networks are a type of machine learning model inspired by the structure and function of the human brain. They consist of interconnected nodes called neurons, organized into layers. Each neuron receives input, performs a computation, and passes the result to neurons in the next layer. Neural networks learn complex patterns in data and are used for tasks like image recognition and natural language processing.

1. Input Layer: Receives raw input data, where each neuron represents a feature.

2. Hidden Layers: Between input and output layers, these perform computations.

3. Activation Function: Applies non-linearity to weighted sum of inputs.

4. Weights and Biases: Adjusted during training to minimize prediction error.

5. Output Layer: Produces final predictions or outputs.

6. Loss Function: Measures prediction error.

7. Training: Involves feeding data, computing loss, and updating parameters using backpropagation.

Neural networks can vary in architecture, including feedforward, convolutional, recurrent, and advanced structures like transformers and GANs, chosen based on data and task characteristics.

3.5 Evalution Metrics

The performance of developed algorithms are measured by mean absolute error (MAE), and coefficient of determination (R^2). The equations are as follows [7] :

$$MAE = \frac{1}{n}\sum_{i=1}^{n} |Y_i - \hat{Y}_i| \quad (2)$$

$$R^2 = 1 - \frac{\sum_{x=1}^{y}(b^x - f(a^x))^2}{\sum_{x=1}^{y}(b^x - b^*)^2} \quad (3)$$

where y refers to the number of target values; $b = (b^1, b^2, \ldots\ldots, b^y)^T$; b^* is the prediction value, and f (a^x) denotes the regression function for the feature vector a^x.

4 Results and Discussions

The gathering and processing of the data is the fundamental step in a machine learning model. For the data we have gathered 1030 values of data under different parameters. The parameters in the dataset are cement, blast furnance slag, fly ash, water, super plasticizer, coarse aggregate, fine aggregate and age. After the collection of data we had performed 4 machine learning algorithms and had verified the accuracy of each algorithm. In this project, we had performed 4 machine learning algorithms linear regression, SVR kernel, random forest, neural networks are implemented using python in google colaboratary.Pandas, Numpy, matplot and skmetrics are the python libraries used. The data is divided into two parts one to train the model and another to test the model.

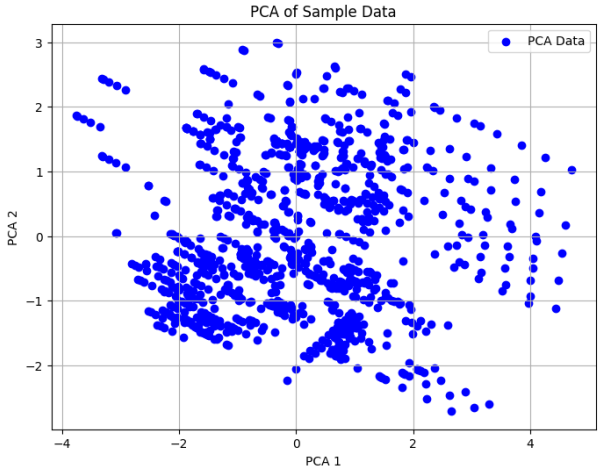

Fig. 1. The scattered plot after reducing the feautures into 2 dimensions using PCA

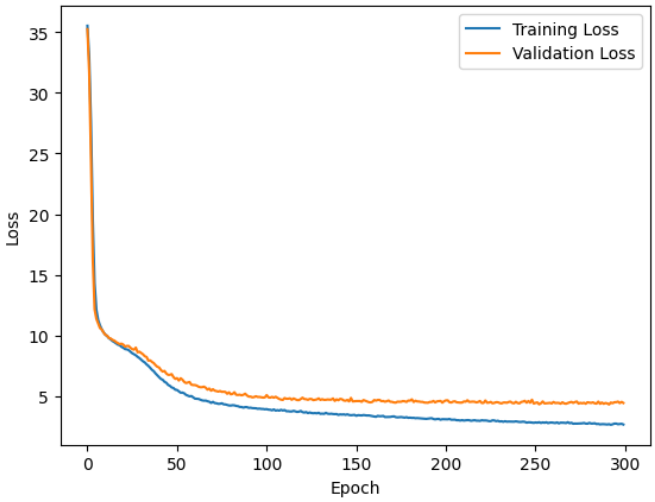

Fig. 2. Graph showing the plot of loss vs epoch of the data

The values of R^2 and MAE of the performed four machine learning algorithms were shown in the below table 2. The R^2 value of Linear Regression is 0.59, which represents that the accuracy level of the algorithm is 59% Similarly, the R^2 for SVR Kernel is 60.02%, for random forest the accuracy is 73.57%, for neural networks the accuracy is 87.2%. Higher the R^2 score, accurate will be the algorithm. From the observations we can say that neural networks gives more accuracy. The MAE values for Linear regression is 8.382. Similarly, the MAE for SVR kernel is 8.36861, the MAE for random forest is 6.6822, the MAE for neural networks is 4.27733. The algorithm will be more efficient if the MAE value of the algorithm will be less. By the above observations we can conclude that neural networks is the best algorithm for prediction model.

Table 2. Comparison of R^2 and MAE

Model	R^2	MAE
Linear Regression	0.596985	8.382613
SVR Kernel	0.600295	8.368612
Random Forest	0.735759	6.682209
Neural Networks	0.872009	4.277332

5. Conclusion

From the analysis on concrete strength prediction model, The main observations were made below

- The parameters taken for the prediction model are cement, blast furnance slag, fly ash, water, superplasticizer, coarse aggregrate, fine aggregrate, age.
- Four machine learning algorithms Linear regression, random forest, SVR Kernel, neural networks were used, to predict the best algorithm values of R^2 and MAE were considered.
- In the results, neural networks had given an R^2 score of 0.872 and MAE 4.27733 making it the best algorithm for the prediction model out of four algorithms.
- From the Results we can say that the neural networks is prefered over the linear regression, SVR Kernel, Random Forest, making it best algorithm for concrete strength prediction.

References

1. Ahmad, S., & Warid, M. N. (2018). Predicting the compressive strength of concrete using multiple regression analysis and artificial neural networks. Journal of Engineering Research and Application, 8(1), 64-70.

2. Kheder, G. M., & Muhamed, J. J. (2020). Prediction of Compressive Strength of Normal Concrete Using Artificial Neural Network. International Journal of Engineering Research & Technology, 9(2), 179-184..

3. ASTM International. (n.d.). ASTM C 42/C42M-20 Standard Test Method for Obtaining and Testing Drilled Cores and Sawed Beams of Concrete. ASTM International