Grenze International Journal of Engineering and Technology, January Issue

Prediction of Concrete Compressive Strength using Machine Learning

V Kathiresan[1], Esther Daniel[2], Premavathi T[3] and Damodharan Palaniappan[4]
[1]GITAM University (Deemed to be University), Karnataka, India
Email. xyzkathir@gmail.com
[2]Karunya Institute of Technology and Sciences, Coimbatore, India
Email. estherdaniell@gmail.com
[3-4]Marwadi University,Rajkot, India
Email. prema.cse05@gmail.com, damomtpcse@gmail.com

***Abstract*— Concrete Compressive strength is the attribute that is used to measure the consistency of the concrete. The concrete's compressive strength is measured subject to the size and the quality of cement, slag, fly ash, etc. Measuring the compressive strength in real-time scenarios takes a lot of effort. Linear regression is a kind of Machine learning methods methods used to predict the compressive strength of the concrete based on the attributes of supervised learning methods. Linear regression of machine learning can be appropriate for predicting the compressive strength of the concrete. The Linear regression based model has been created and trained using the dataset from UCI Machine Learning Repository. The model's performance was also tested by using the metrics RMSE, and MSE. The model can predict the concrete compressive strength for any unknown value of the given attributes.**

***Index Terms*—compressive strength, Machine Learning, Linear regression, RMSE, MSE.**

I. INTRODUCTION

This paper deals with measuring the strength of the concrete. The input parameter or attributes cement usage, blast furnace, fly ash, water mixture, super plasticizer mixture, coarse aggregate mixture, fine aggregate mixture and concrete age the compressive strength of the concrete is getting predicted [1][2]. The dataset used for training purpose is downloaded from UCI machine learning repository. Machine learning based linear regression model is used in the prediction of the compressive strength of the concrete [3][4]. Dataset used for experiment is donated by I- Cheng Yeh from Department of Information Management, Chung-Hua University (Republic of China). Independent variables are consider as an input for the prediction. Independent variables plays a vital role in testing as well as prediction. Independent variables is represented with the letter X. $X = \{\text{cement, blastfurnace, flyash, water, super plasticizer, coarseaggregate, fineaggregate , age}\}$
The dependent variable is the one that is going to be predicted based on the independent variable X. Dependent variable is denoted by the letter Y. The dependent(Y) variable given in the data set is concrete compressive strength. The dataset contains 103 tuples. The attributes are 8, those are considered as input or independent variables and 1 dependent or output variable in the data set.

Grenze ID: 01.GIJET.11.1.22

II. PROBLEM STATEMENT:

Predicting slump used to make the concrete is very difficult, it consumes a lot of effort and time when we do it manually. It involves civil engineering methods to solve. Incorporating machine learning techniques to solve the problem easily, it predicts the slump of the concrete based on the trained model with existing data.

III. MACHINE LEARNING IN PREDICTING THE COMPRESSIVE STRENGTH OF CONCRETE

Linear regression-based techniques can help in prediction of concrete compressive strength. Machine learning techniques are classified into two categories one is supervised and other is unsupervised learning. Supervised machine learning algorithms work based on training data. It helps to solve classification and prediction-based problems. In this paper supervised learning-based linear regression technique is used to train the model using the attributes cement usage, blast furnace, fly ash, water mixture, super plasticizer mixture, coarse aggregate mixture, fine aggregate mixture and concrete age. Concrete compressive strength is predicted based on the trained model.

A. Linear Regression's role in concrete compressive strength prediction:

Linear regression is widely used in machine learning. A simple and powerful approach of machine learning is linear regression. Predictive analysis can easily done using this statis tical methods and technique. Linear regression supports numerical prediction like houseprice, share value, temperature.,etc [5][6].

In linear regression the relationship among independent variable and dependent variable is getting established. The main objective is trying to fit the straight line in the data distribution of X and Y. once the line fitted properly in the data distribution Y can be predicted for any X value [7][8].slope of the line and Y_intercept supports the prediction of Y for any X.

In the training part of LR model, the model is trying to fit the straight line in the training data distribution.

$$\mathrm{Y} = \mathrm{MX} + C$$

$X - Independent\ Variable$
$Y - Dependent\ Variable$
$\mathrm{C} =$ intercept
$\mathrm{M} =$ Slope

Model is getting trained with the help of gradient descent algorithm. The X and Y are the input value for training. The gradient descent algorithm finds the best fitting line in the data distribution. Slope and Y_intercept of the best fitting line is used to predict Y for any X.

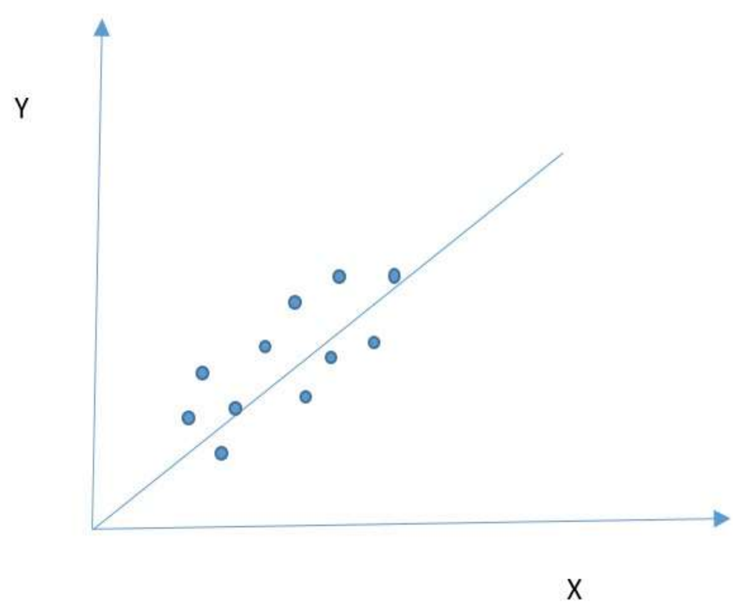

Fig.1 Linear regression

B. Linear regression based numerical prediction:

Linear regression plays a vital role in numerical prediction since it is a well-known technique for predicting numerical values. When the data is linearly distributed the regression method shows significant performance concerning the prediction

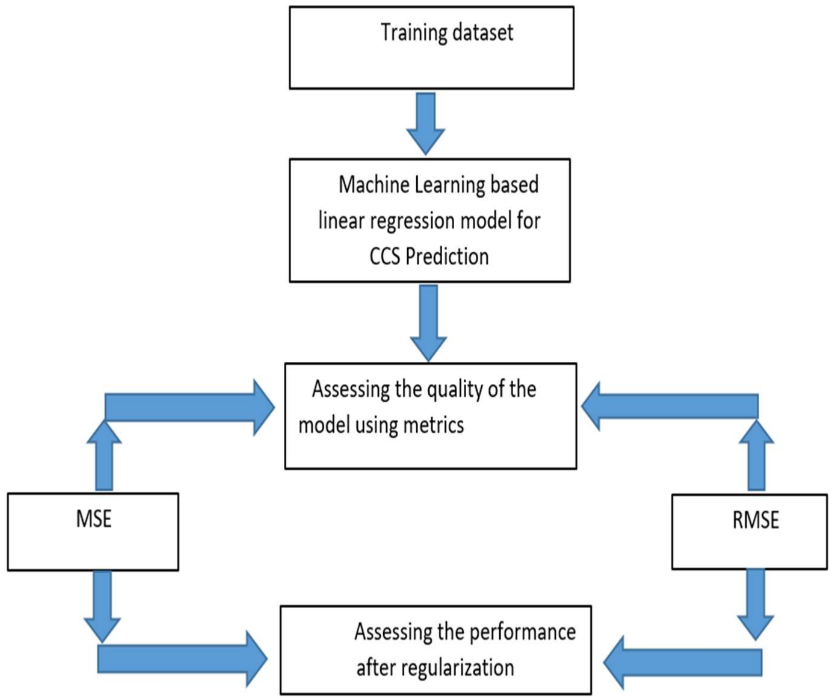

Fig. 2 Flow Diagram of the CCS (Concrete Compressive Strength) Prediction Model.

Fig 1 represents the fitness of the straight line in linearly distributed data. The proposed training dataset is given to the machine learning based linear regression model for CCS prediction. The model quality is assessed by MSE,RMSE ,MAE and R2 metrics[9][10][11]. After the assessment, the model quality is improved by regularization techniques. MSE (Mean squared Error) is one of the metrics used to assess the quality of the linear regression model.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - {y'}_i)^2$$

$$n - number\ of\ data\ points\ in\ dataset$$

$$y_i\ - Actual\ Value$$

$$y_i' - Predicted\ Value$$

Root Mean squared Error is the metric to check the quality of the prediction model . It is square root of MSE.

IV. EXPERIMENTAL ANALYSIS:

The dataset that is used is donated by I Cheng Yeh, Department of Information Management, Chung-Hua University (Republic of China) [15]. It is a multivariate data. It contains 103 instances . it was donated in the year of 2009. Initially, the number of data points was 78. After several years 25 new data points are added. The dataset contains 8 independent variables cement_usage, blast_furnace, fly_ash, water_mixture, super_plasticizer_mixture, coarse_aggregate_mixture, fine_aggregate_mixture and concrete_age.The dependent variable is ccs. The sample data is represented in Fig 3 and Fig 4 represents the sample data and its type respectively.

$$X = \{cement\_usage, blast\_furnace, fly\_ash, water\_mixture,$$
$$super\_plasticizer\_mix, coarse\_aggregate\_mixtu, fine\_aggregate\_mixture\ and\ concrete\_age\ \}$$
$$Y = \{ccs\}$$

	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)	CCS
0	540.000	0.000	0.000	162.000	2.500	1040.000	676.000	28	79.990
1	540.000	0.000	0.000	162.000	2.500	1055.000	676.000	28	61.890
2	332.500	142.500	0.000	228.000	0.000	932.000	594.000	270	40.270
3	332.500	142.500	0.000	228.000	0.000	932.000	594.000	365	41.050
4	198.600	132.400	0.000	192.000	0.000	978.400	825.500	360	44.300

Fig.3 Sample Data Points of The Dataset

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1030 entries, 0 to 1029
Data columns (total 9 columns):
 #   Column                                                 Non-Null Count  Dtype  
---  ------                                                 --------------  -----  
 0   Cement (component 1)(kg in a m^3 mixture)              1030 non-null   float64
 1   Blast Furnace Slag (component 2)(kg in a m^3 mixture)  1030 non-null   float64
 2   Fly Ash (component 3)(kg in a m^3 mixture)             1030 non-null   float64
 3   Water  (component 4)(kg in a m^3 mixture)              1030 non-null   float64
 4   Superplasticizer (component 5)(kg in a m^3 mixture)    1030 non-null   float64
 5   Coarse Aggregate  (component 6)(kg in a m^3 mixture)   1030 non-null   float64
 6   Fine Aggregate (component 7)(kg in a m^3 mixture)      1030 non-null   float64
 7   Age (day)                                              1030 non-null   int64  
 8   CCS                                                    1030 non-null   float64
dtypes: float64(8), int64(1)
memory usage: 72.5 KB
```

Fig.4 Attribute Details and Data Types

	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)	CCS
Cement (component 1)(kg in a m^3 mixture)	1.000	-0.275	-0.397	-0.082	0.092	-0.109	-0.223	0.082	0.498
Blast Furnace Slag (component 2)(kg in a m^3 mixture)	-0.275	1.000	-0.324	0.107	0.043	-0.284	-0.282	-0.044	0.135
Fly Ash (component 3)(kg in a m^3 mixture)	-0.397	-0.324	1.000	-0.257	0.378	-0.010	0.079	-0.154	-0.106
Water (component 4)(kg in a m^3 mixture)	-0.082	0.107	-0.257	1.000	-0.658	-0.182	-0.451	0.278	-0.290
Superplasticizer (component 5)(kg in a m^3 mixture)	0.092	0.043	0.378	-0.658	1.000	-0.266	0.223	-0.193	0.366
Coarse Aggregate (component 6)(kg in a m^3 mixture)	-0.109	-0.284	-0.010	-0.182	-0.266	1.000	-0.178	-0.003	-0.165
Fine Aggregate (component 7)(kg in a m^3 mixture)	-0.223	-0.282	0.079	-0.451	0.223	-0.178	1.000	-0.156	-0.167
Age (day)	0.082	-0.044	-0.154	0.278	-0.193	-0.003	-0.156	1.000	0.329
CCS	0.498	0.135	-0.106	-0.290	0.366	-0.165	-0.167	0.329	1.000

Fig 5. Correlation Matrix

Fig 5 represents the co-relation matrix among the variables, it shows whether the variables have co-relation among themselves or not [12][13][14]. If correlation exist whether it is come under the category of positive correlation or negative correlation. It can be used as primary data for dimensionality reduction. Once the model has been created the model is ready to get unforeseen data and predict the ccs. Now we should assess the quality of the model based on the well-known metrics R2, MAE, MSE and RMSE. Fig 6 represents the training and testing data's value associated with metrics. Fig 7 represents the comparison of actual and prediction.

	linear_train	linear_test
R2	0.620	0.594
mae	8.175	8.299
mse	107.240	109.751
rmse	10.356	10.476

Fig 6. Quality of the Model Based on the Metrics R2, MAE, MSE and RMSE

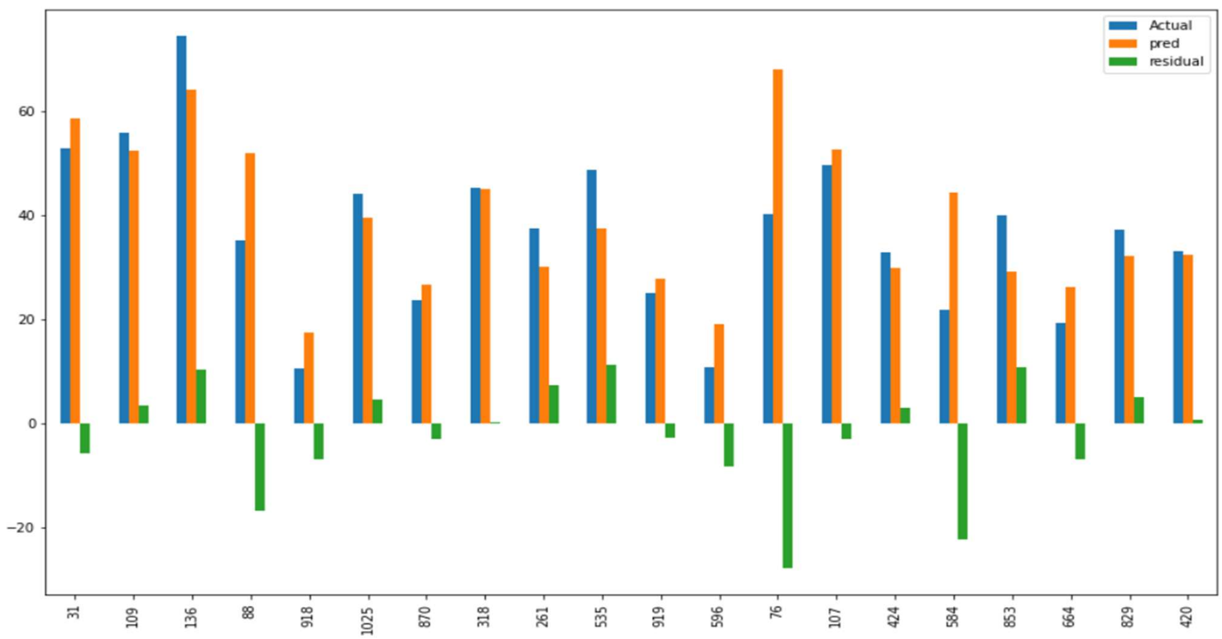

Fig 7. Comparison of Actual and Prediction

IV. Conclusion

Finding concrete compressive strength based on civil engineering methods is time time-consuming and resource-consuming process. In this work, ccs have been predicted for unforeseen data through machine learning prediction techniques. The dataset donated by I-cheg yeh from the Department of Information Management, cheng-Hua university has been used for training purposes. The prediction of ccs through machine learning reduces the complexity of the process. It can save resources too. Linear regression model implemented through sklearn library of python for prediction. The model's quality is also assessed through well-known quality metrics like R2, MAE, MSE and RMSE. The obtained results prove the performance of the model.

REFERENCE

[1]. Feng, D. C., Liu, Z. T., Wang, X. D., Chen, Y., Chang, J. Q., Wei, D. F., & Jiang, Z. M. (2020). Machine learning-based compressive strength prediction for concrete: An adaptive boosting approach. Construction and Building Materials, 230, 117000.

[2]. Naderpour, H., Rafiean, A. H., & Fakharian, P. (2018). Compressive strength prediction of environmentally friendly concrete using artificial neural networks. Journal of Building Engineering, 16, 213-219.

[3]. Deng, F., He, Y., Zhou, S., Yu, Y., Cheng, H., & Wu, X. (2018). Compressive strength prediction of recycled concrete based on deep learning. Construction and Building Materials, 175, 562-569.

[4]. Kewalramani, M. A., & Gupta, R. (2006). Concrete compressive strength prediction using ultrasonic pulse velocity through artificial neural networks. Automation in Construction, 15(3), 374-379.

[5]. James, G., Witten, D., Hastie, T., Tibshirani, R., & Taylor, J. (2023). Linear regression. In An introduction to statistical learning: With applications in python (pp. 69-134). Cham: Springer International Publishing.

[6]. Su, X., Yan, X., & Tsai, C. L. (2012). Linear regression. Wiley Interdisciplinary Reviews: Computational Statistics, 4(3), 275-294.

[7]. Uyanık, G. K., & Güler, N. (2013). A study on multiple linear regression analysis. Procedia-Social and Behavioral Sciences, 106, 234-240.

[8]. Aalen, O. O. (1989). A linear regression model for the analysis of life times. Statistics in medicine, 8(8), 907-925.

[9]. Spüler, M., Sarasola-Sanz, A., Birbaumer, N., Rosenstiel, W., & Ramos-Murguialday, A. (2015, August). Comparing metrics to evaluate performance of regression methods for decoding of neural signals. In 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) (pp. 1083-1086). IEEE.

[10]. Grömping, U. (2007). Relative importance for linear regression in R: the package relaimpo. Journal of statistical software, 17, 1-27.

[11]. Tatachar, A. V. (2021). Comparative Assessment of Regression Models Based On Model Evaluation Metrics. International Journal of Innovative Technology and Exploring Engineering, 8(9), 853-860.

[12]. Steiger, J. H. (1980). Tests for comparing elements of a correlation matrix. Psychological bulletin, 87(2), 245.

[13]. Dziuban, C. D., & Shirkey, E. C. (1974). When is a correlation matrix appropriate for factor analysis? Some decision rules. Psychological bulletin, 81(6), 358.

[14]. Kohonen, T. (1972). Correlation matrix memories. IEEE transactions on computers, 100(4), 353-359.

[15]. Santosa, S., & Santosa, Y. P. (2017). Evolutionary artificial neural networks for concrete mix design modelling. Int. J. Comput. Appl, 5(7).