Prediction of Building Construction Project Cost Using Multiple Linear Regression and Artificial Neural Network

Viren B. Chandanshive[1] and Dr. Ajay R. Kambekar[2]

Ph.D. Student, Department of Civil Engineering, Sardar Patel College of Engineering, Andheri, Mumbai 400058, Maharashtra, India[1].
Associate Professor, Department of Civil Engineering, Sardar Patel College of Engineering, Andheri, Mumbai 400058, Maharashtra, India[2].

virenchandanshive@gmail.com1, arkambekar@gmail.com2

Abstract. The prediction of building construction cost becomes very crucial task at the early stage of construction due to lack of design data and limited available information. The prediction of project cost with higher degree of accuracy plays an important role in the success of every construction projects. The objective of this study is to develop Artificial Neural Network (ANN) and Multiple Linear Regression (MLR) model to predict the project cost at the early stage. Based on a literature survey and expert advice from the design professionals of Indian construction industry, the most influential eleven cost parameters are applied as input parameters while the project cost is the output parameter. A dataset of 78 building construction projects was collected from Mumbai (India) and its nearby region. The results obtained from the developed ANN and MLR model shown that, it is able to predict the cost of building construction projects. The coefficient of correlation *(R)* was about 0.9886 and the error criteria, the Mean Squared Error (MSE) 0.00081 of ANN model indicates that the ANN has better prediction and performance over the MLR model. This study contributes to construction management and provides a general idea about the project cost which will be helpful to the investors.

Keywords: Artificial Neural Network (ANN); Multiple Linear Regression (MLR); Cost Predictions; Early stage.

1 Introduction

At the early stage of construction, the prediction of project cost along with the higher accuracy is a very crucial task to the quantity surveyor as well as the project manager. The accuracy in the cost estimates plays a vital role in the success of the project as well as provides a general idea about the budget. At the pre-design stage of construction, due to a lack of design data and limited available information, it is very difficult to

predict the project cost and becomes a complex problem. At such conditions, an appropriate cost prediction performs the effective finance management entire every phase of construction. Also, useful to the project manager to the management of available capitals as well as for better decision making. Development of cost modeling methods is usually based on the historical data of the previous studies and construction experience as well as a knowledge of quantity surveyor. According to the literature study, Multiple Linear Regression (MLR) and Artificial Neural Network (ANN) tools are widely applied in such complex prediction problems.

Generally, the structural skeleton system, as well as interior and exterior walls, finishing works, etc., has a major impact on the total cost of residential building projects and hence cautious consideration must be reserved during the cost estimation. After a successful completion of literature study and expert interview with the Indian construction industry, the most significant cost parameters were identified. In this study, a dataset of 78 residential building construction projects was collected from the real estate sector of Mumbai (India) and its nearby regions. The basic objective of this study is to develop a statistical MLR based as well as an ANN model to predict the project cost at an early stage of construction.

2 Literature Review

The statistical approach MLR technique and ANN tool have numerous applications in several fields of civil engineering for prediction purpose. Some of the important literature of MLR and ANN are deliberated in this segment.

Gunaydın and Dogan developed an ANN model by using cost factors such as total are, ratio between typical floor area and total area, ratio between ground floor area and total area, number of floor, console direction of building, foundation, etc. for the early stage construction cost prediction. About 93% accuracy in prediction performance and low error criteria (MSE) shows ANN model has better good cost estimation. Adeli and Wu provides a complete overview of regularization of artificial neural network modeling with solved two examples. Smith and Mason examined the performance and stability of regression and neural networks cost estimation relationship (CER) models. They observed that the regression models are more advantageous over neural networks in terms of its accuracy, variability, development and validation. Attalla and Hegazi developed MLR as a statistical approach and ANN for predicting the cost deviation and concluded that both the developed model performed similar but ANN was more sensitive during the training of large dataset. Khamis et.al. compared MLR and ANN models for the estimation of house price in New York city. They observed that ANN model has 26.475% more regression criteria while lower in error criteria which indicates the better prediction than the MLR model. Wilmot and Mei developed an artificial neural network model for the prediction of highway construction cost. The level of significance of the predicted cost is about 95%. Al-Zwainy and Hadhal build a mathematical model for prediction of communication tower projects cost by using MLR technique. Total seven effective parameters were utilized for the development of MLR model. The results show that about 90.1% accuracy level with mean absolute percentage error 9.891% and

the correlation (R) was 98.6% between actual and predicted cost signifies a good prediction.

3 Identification of cost influential factor

The most influential cost factors are identified from the previous associated studies and by directing practiced consultation and expert advice from the real estate sector of Indian construction industry. The most important eleven parameters that were applied in this study for the development of MLR and ANN models are; Ground Floor Area, Typical Floor Area, Structural Parking Area, Quantity of Flooring, No of Floors, Quantity of Elevator Wall, Quantity of Exterior Wall, Total Quantity of Exterior Plaster, Number of Columns, Types of Foundation and Number of Households.

4 Data Processing & Analysis

The database of small and medium scaled residential housing project which are recently constructed from year 2017-2019 is collected. Total 78 building construction project's database was collected from Mumbai (India) and its nearby region. To avoid overfitting and smooth performance the MLR and ANN model the collected database is normalized with the help of following equation 1.

$$P_i^{norm} = \left[2\left(\frac{P_{i-}\ P^{min}}{P^{max}{}_{-}\ P^{min}}\right)\right] - 1 \qquad (1)$$

5 Development of Regression analysis and neural networks

5.1 Regression Analysis

The regression techniques have precise mathematical background and have been widely used for cost estimation from last five decades. For the prediction of project cost, regression techniques are applied to observe the influence of various parameters to the project cost. The finest ability of regression methods is the dexterity to understand the associations between the cost and the variables measured.

In this study, the Stream 2* - IBM SPSS Modeler software is used to develop multiple linear regression model. For assessing the regression approach, the adjusted R2 value indicates the percentage changeability in the cost that can be enlightened by the parameters associated in the model. The p-values provide an indication of the significance level of the distinct parameters. According to T-test (at the P-values $P \leq 0.05$), the respective slope coefficients of independent variables like; Structural Parking Area, Quantity of Flooring, No of Floors and Quantity of Elevator Wall indicates the importance level from a statistical perspective. While, according to T-test But the independent variables Ground Floor Area, Typical Floor Area, Quantity of Exterior Wall, Total Quantity of Exterior Plaster, No. of Columns, Types of Foundation and No. of Households did not have a substantial influence on the developed MLR model.

Table 1. Regression Coefficients Values of MLR Model – Enter Method

Sr. No.	Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-0.144	0.039		-3.713	0.000
2	Ground Floor Area (X_1)	0.121	0.117	0.196	1.037	0.304
3	Typical Floor Area (X_2)	0.140	0.122	0.189	1.144	0.257
4	Structural Parking Area (X_3)	-0.252	0.085	-0.368	-2.981	0.004
5	Quantity of Flooring (X_4)	0.413	0.154	0.530	2.676	0.009
6	No of Floors (X_5)	0.122	0.051	0.169	2.406	0.019
7	Quantity of Elevator Wall (X_6)	0.307	0.090	0.328	3.411	0.001
8	Quantity of Exterior Wall (X_7)	0.098	0.231	0.127	0.422	0.675
9	Total Quantity of Exterior Plaster (X_8)	-0.181	0.248	-0.236	-0.730	0.468
10	No. of Columns (X_9)	-0.022	.079	-0.027	-0.274	0.785
11	Types of Foundation (X_{10})	-0.016	.014	-0.041	-1.163	0.249
12	No. of Households (X_{11})	0.159	.081	0.205	1.963	0.054

The values of unstandardized coefficients can be used to develop the multivariable linear regression equation;

$$\text{Total Cost} = (-0.144 + 0.121\ X_1 + 0.140\ X_2 - 0.252\ X_3 + 0.413\ X_4 + 0.122\ X_5 + 0.307\ X_6 + 0.098 \quad X_7 - 0.181\ X_8 - 0.022\ X_9 - 0.016\ X_{10} + 0.159\ X_{11}) \tag{2}$$

The equation 2 represents the mathematical relationship between the predictor variables and also the response variable and regression Coefficients Values of developed MLR Model.

5.2 Artificial Neural Network

An artificial neural networks are used to predict the project cost as the alternative for regression techniques. ANN has ability to learn from previous studies and works as a robust tool in many areas such as classification, pattern recognition, time series and prediction problems. The training process in the neural network consist the presenting set of examples (input parameters) with defined outputs (target output). The neural network system provides arrangement of weights to reduce the errors between network output and target output. Some of the well-known neural network systems consist of back propagation network, perception, ADALINE (Adaptive Linear Element), associative memory, etc. Fig. 1 shows the architecture of an artificial neuron.

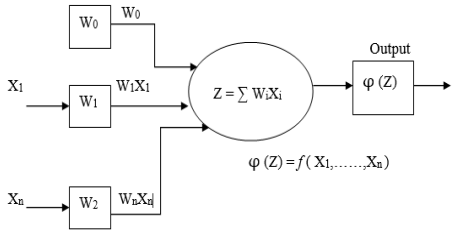

Fig. 1. Mathematical Model of Neuron.

Here, X1, X2, X3,…Xn are the n inputs to the artificial neuron and W1, W2, W3,……Wn are the weights attached to the input links. The total input *I* received by the soma of the artificial neuron is given by equation 3;

$$I = w_1x_1 + w_2x_2 + \ldots + w_nx_n = \sum_{i=1}^{n} w_ix_i \quad (3)$$

In this study, Matlab R2013a version Software is used for the development ANN model. The process was carried out in three phases; Training phase, validation phase and Testing phase. The database was used for the processing of training, validation and testing sets was 70%, 15%, and 15% respectively. The Scaled Conjugate Gradient (trainscg) training function is utilized to modify the weights and biases for minimization of error. The error on the validation set is monitoring during training process. A reasonable level of performance error function was adopted.

6 Comparison between Neural Network and Regression Model

In this study, the comparison between ANN and MLR model is carried out. The comparison was based on the two criteria: Regression and Error criteria. Regression criteria includes the coefficient of correlation R, coefficient of determination R2, and Adjusted R Square (R-Adjacent) to explain the overall effectiveness of the model as well as to identify the correlation between the target cost verses the predicted cost. The error criteria measure the overall performance of developed models.

Table 2. Comparison between the Models

Comparison Factor	Model	
	Artificial Neural Network	Regression Analysis
	ANN (trainscg)	MLR (Enter Method)
R	0.9886	0.973
R2	0.9774	0.946
R-Adjacent	0.9771	0.937
SSE	0.1315	5.736
MSE	0.00081	0.521

7 Conclusion

This paper investigated the various application of regression analysis as well as neural networks in the estimation of construction cost of project. Based on literature study and an expert interview as well as advice from the various agencies involved in real estate sector the most important cost factors that influence the project cost are identified. Two different models were developed to predict the project cost at the early stage of construction. A statistical regression analysis approach is used to develop MLR model and other one was developed using artificial neural networks. About Seventy-eight cases from the Mumbai and its near-by region were used for the development of both model. The data sets are divided into three sets; training (70%), testing (15%) and validation (15%), while the enter method is used to develop MLR model and the Scaled Conjugate Gradient (trainscg) training function for ANN model.

The results shown that both the MLR and ANN predictive models are suitable for such cost prediction problems. The coefficient of correlation R is about 0.9886 between the target cost and predicted cost have proven that ANN has slightly more effective than the MLR model. Also the error criteria of ANN model shown that the performance of ANN model is better than the statistical regression approach i.e. MLR model. Such cost prediction models provide the general idea about the budget of the project hence better decisions can be made.

8 Reference

1. Chester G. Wilmot, "Neural Network Modeling of Highway Construction Costs," Journal of Construction Engineering and Management, vol. 131, no. 7, pp. 765-771, July 2005.
2. E. S. Alice and K. M. Anthony, "Cost Estimation Predictive Modeling: Regression Versus Neural Network," The Engineering Economist: A Journal Devoted to the Problems of Capital Investment, vol. 42, no. 2, pp. 137-161, 2010.
3. H. Adelil and M. Wu, "Regularization Neural Network for Construction Cost Estimation," Journal of Construction Engineering and Management, vol. 124, no. 1, January 1998.
4. J. E. David, D. H. Gary and C. H. Frank, "A comparative analysis between the multilayer perceptron ``neural network" and multiple regression analysis for predicting construction plant maintenance costs," Journal of Quality in Maintenance Engineering, vol. 6, no. 1, pp. 45-61, 2000.
5. K. Azme Bin, N. K. and K. Khalilah Binti, "Comparative Study On Estimate House Price Using Statistical And Neural Network Model," International Journal of Scientific & Technology Research, vol. 3, no. 12, pp. 126-131, 2014.
6. K. Gwang-Hee, A. Sung-Hoon and K. Kyung-In, "Comparison of construction cost estimating models based on regression analysis, neural networks, and case-based reasoning," Building and Environment, vol. 39, no. 10, pp. 1235-1242, 2004.
7. M. Attalla and T. Hegazy, "Predicting Cost Deviation in Reconstruction Projects: Artificial Neural Networks versus Regression," Journal of Construction Engineering and Management, vol. 129, no. 4, pp. 405-411, August 2003.
8. M. G. H and Z. D. S, "A neural network approach for early cost estimation of structural systems of buildings," International Journal of Project Management, vol. 22, no. 7, pp. 595-602, October 2004.