World Journal of Advanced Research and Reviews
eISSN: 2581-9615 CODEN (USA): WJARAI
Cross Ref DOI: 10.30574/wjarr
Journal homepage: https://wjarr.com/

(RESEARCH ARTICLE)

Check for updates

Road construction analysis using regression technique

JHERIMAE C. ANCHETA, PRIMO IVAN D. LAZO *, LLOVELYN B. MEDINA, CRISELLE J. CENTENO, ARIEL ANTWAUN ROLANDO C. SISON and MARK ANTHONY S. MERCADO

Information Technology Department, Pamantasan Ng Lungsod Ng Maynila, Manila, Philippines.

World Journal of Advanced Research and Reviews, 2023, 18(03), 658–664

Publication history: Received on 03 May 2023; revised on 09 June 2023; accepted on 12 June 2023

Article DOI: https://doi.org/10.30574/wjarr.2023.18.3.1125

Abstract

Estimating cost in construction is important to the city's design and planning management hence, the construction cost estimate must not be overpriced which may cause corruption or underpricing that leads to unreliable or low-quality road projects. The total estimated cost is only valid in the same year it was proposed because of the inflation rate the costs may change. The researchers applied Multiple Linear Regression technique in predicting total estimated cost for road construction analysis. The model is evaluated by the means of R-squared to determine the variables if they are correlated or overfitting. The calculated R-squared is equals to 0.696598 with the predictor variables (x1 & x2) Roadbed width and Net length and it means that the predictors (Xi) explain 69.7% of the variance of Y. The higher the R-squared result, the better fit it is for the Multiple Linear Regression model. It also shows that X1 and X2 are significant predictor variables. The coefficient of multiple correlation (R) is equals to 0.834624 and it means that there is a very strong correlation between the predicted data and the observed data whereas the dependent variable (y) is the Estimated cost.

CCS CONCEPTS: Multiple Linear Regression • Construction Estimation • Engineering

Keywords: Machine Learning; Supervised Learning; Prediction; Forecasting

1. Introduction

Department of Public Works and Highways, also known as DPWH, is the executive department of the Philippine government and it is the State's engineering and construction arm. They are responsible for planning, designing, construction and maintenance of infrastructure facilities, and other public works. Estimating cost in construction is important because it allows developers to make important choices. The initial estimate will not be the same as the total price of the project's final price tag, but the accurate estimate keeps all the parties focused on delivering the project on time and under budget. According to (DPWH NCR, 2022), there are ongoing construction projects that range from P93M - P140.2 million. One of the other general requirements needed in Construction Projects includes the Construction of Field Office for the Engineer that costs P488, 881.14.

Delays and unexpected occurrences may increase the cost estimation budget of an infrastructure project of DPWH. The analyzed data gathered from DPWH ADEO Projects in Aurora District shows that 317 projects are completed while 188 projects have encountered delays due to unexpected circumstances. There are nine (9) reasons for suspensions (Right of way problem, obstruction of public utilities, pandemic, waiting for the completion of other Phase, Scheduled Power Interruption, Christmas Season, non-availability of materials in the Locality, & obstruction of Existing Building) of the projects with a total of 236 suspensions for each reason. The reason that has the greatest number of suspensions is due to Weather occurrences. According to the surveys conducted by (Pantalunan, C et al. 2021), there are five (5) main causes of project delays which include Financial-related causes, Design-related causes, Natural/External causes, Management-related causes, and Construction-related causes. Among the five mentioned, the majority cause of delayed

* Corresponding author: PRIMO IVAN D. LAZO

projects is the Natural/External causes which includes national calamities and due to pandemic. With this related study conducted from Aurora District, the researchers aim to apply it for the construction estimation and cost prediction of DPWH in NCR amongst unexpected circumstances that may agitate the cost estimation.

According to (COA, 2020), the significant delays of implementation of projects were caused by the following thirteen (13) factors issued by the DPWH Offices. The causes of these delays are unfavorable condition, peace and order situation, road right-of-way, pending issuance of excavation permit, modification of plans, design, and program of works, insufficient workforce, lack of equipment, unavailability & difficulty of hauling construction materials, inaccessible project site for heavy equipment, modification and realignment of location, scarcity of materials, inadequate project supervisions and monitoring, and late release of funds. There are fourteen factors that are affecting construction cost estimation according to (Mishra, 2014). The 14th factor is the construction contingency where it is stated that it is advisable to add at least 10% contingency towards the total projects costs for unforeseen costs and inflation. It's important that this budget is not viewed as extra money or money that needs to be spent. It is better to not use the construction contingency because it means that the project was properly planned, estimated, and managed properly. According to the Department Order 99 Series of 2018 (DPWH, 2018), The contingency amount must not exceed five percent (5%) of the amount of the contract excluding value added taxes. The researchers aim to apply these measures according to the related studies for the problem statements that are created.

Statement of the Problem

- The original estimated cost is not aligned with the actual project cost exceeding in the target budget plan of the client.
- Bidding Process is often biased and fixed by Assigned District Engineers which impacts the proper bidding winner.
- Inflation of construction materials may change the current cost estimation making it inconsistent.

Objectives of the Study

- To develop a system that provides construction estimation of materials with improved accuracy and applied construction contingency for additional or unexpected costs.
- To design a system that compares the bidder's cost estimation and the system's predicted outcomes.
- To produce a system for prediction of construction estimation and costs of the materials of how the inflation rate moves.

2. Review of related literature

A group of researchers developed a prediction model using the multiple linear regression technique in Iraq. They combined the Multiple Linear Regression technique (MLR) with the Weighted Least Squared (WLS). 501 sets of historical data were used as a data set within a time period of 10 years in Iraq (2005-2015). The cost of twenty five items are used for predicting the cost using MLR model and they involved the cost of (excavation the foundation works, Landfill works, filling with sub-base works, Construction works under moisture proof layer, Construction works above moisture proof layer, Construction works of sections, ordinary concrete for walkways, reinforced concrete foundation, reinforced concrete column, reinforced concrete lintel, reinforced concrete slabs, reinforced concrete beams, reinforced concrete stair, reinforced concrete for the sun bumper, plaster finishing works, cement finishing works, Plastic Paints, Pentellite paints, Stone packaging, Works of placing marble, Ceramic works for floor, Ceramic works for walls, Flattening (two opposite layers of lime), Flattening (Tiling). After conducting the study, they have concluded that the regression analysis technique proved its purpose and showed that it is very accurate with the degree of accuracy of 98.97%. MLR with WLS is promising to be used only at the initial stage of the project when limited data and incomplete information are present. (Abd et al., 2019).

According to (Gülden, 2013), Regression analysis is a statistical technique for estimating the relationship among variables which have reason and result relation. Main focus of univariate regression is to analyse the relationship between a dependent variable and one independent variable and formulate the linear relation equation between dependent and independent variable. Regression models with one dependent variable and more than one independent variable are called multiple linear regression. In this study, data for multilinear regression analysis occurred from Sakarya University Education Faculty student's lesson (measurement and evaluation, educational psychology, program development, counselling and instructional techniques) scores and their 2012-KPSS score. Assumptions of multilinear regression analysis-normality, linearity, no extreme-values and missing value analysis were examined. The data that

verify the assumptions were analyzed with multiple regression and lessons measurement and evaluation, instruction techniques, counselling, program development and educational psychology were estimated in the KPSS respectively.

3. Material and methods

3.1. Research Design

The research employs the quantitative - correlational research design utilizing a predictive method that involves regression analysis. It is used in those cases when there is an interest to identify the predictive relationship between the predictor and the outcome variable. This design is used to forecast outcomes, costs, consequences, or effects. It is designed to gather information regarding the materials to be used in construction projects in Manila. The researchers will use the Machine learning technique utilizing multiple linear regression analysis that needs statistical analysis skills. Multiple regression is a type of regression where the dependent variable shows a linear relationship with two or more independent variables. It can also be non-linear, where the dependent and independent variables do not follow a straight line. (CFI, 2022)

3.2. Research Development

The researchers used the Waterfall Model Software Development Life Cycle for designing, developing, and testing of software.

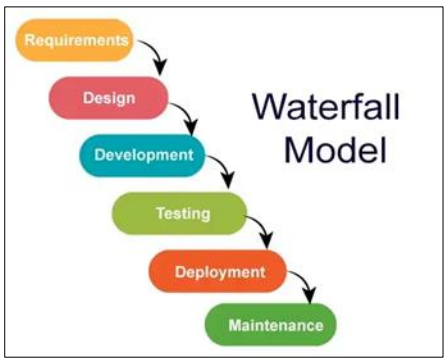

Figure 1 Waterfall Model

Upon interviewing the possible users and experts in the field related to the study the researchers analyzed all the necessary requirements for the system. The researchers will visualize the system through requirements analysis and come up with a conceptual model of the study then moves to the next phase, the implementation and coding. To identify the faults of the system the researchers will conduct software testing. Once the system is fully functional and has met the requirements then it may now proceed to deployment all the way to maintenance.

3.2.1. Hardware and Software Requirements

A Laptop will be used for developing the model using an ASUS Notebook with (Intel® core™ i3 7th gen) with a clock speed of 2.3 GHz, 8.0 gigabytes (GB) random access memory (RAM), and a Windows 10 operating system. The R Language will be used as the statistical computing programming Language and RStudio as the compiler of the Multiple Linear Regression Model.

3.2.2. Research Population and Sample

The population and sample that are used for data gathering are the datasets available about Road Projects within the vicinity of Manila City. Convenience Sampling is used as a technique for gathering samples according to the convenience of the researchers and respondents in Manila City Hall. This is used for any available dataset that the respondents have to offer for training it for the Multiple Linear Regression Model.

3.3. Research Evaluation\ Building Phase

In the article of (Cote, 2022), he mentioned that using a confusion matrix for regression problems is a very nice thing as a powerful visual tool to interpret the overall fit of the model and help spot problems in specific regions of the continuous data, even for multivariate problems where it is not possible to plot X vs Y.

A confusion Matrix is a classification table that is used to describe a certain model like the Multiple Linear Regression Model of this study. It describes its performance on a set of data for which it will know the true values of data. It is also called an error matrix for a reason, by giving a better idea of what is correctly predicted and what types of errors your algorithm is making. In the researchers' study, to verify the data that is imported into the system for the model, the Confusion Matrix shall classify the true values of the data of the Road Construction Projects that they have gathered so that it may determine its accuracy, misclassification of error, specificity, and sensitivity.

3.3.1. Statistical Tool and Instrument Use

- y = the predicted value of the dependent variable
- B_0 = the y-intercept (value of y when all other parameters are set to 0)
- B_1X_1 = the regression coefficient (B_1) of the first independent variable (X_1) (a.k.a. the effect that increasing the value of the independent variable has on the predicted y value)
- ... = do the same for however many independent variables you are testing
- B_nX_n = the regression coefficient of the last independent variable
- ϵ = model error (a.k.a. how much variation there is in our estimate of y)

$$y = \beta_0 + \beta_1 X_1 + \ldots + \beta_n X_n + \epsilon$$

Figure 2 Multiple Linear Regression Formula

According to (Hayes, 2022), A number of explanatory variables are combined in a statistical process called multiple linear regression (MLR), also referred to as multiple regression. It is a development of linear (OLS) regression that only takes into account one explanatory factor.

The researchers intended to use machine learning specifically the multiple linear regression to predict the construction cost materials. Based on Lani, 2021 to perform the multiple linear regression the researchers must go through its three stages: 1) analyzing the correlation and directionality of the data, 2) estimating the model, i.e., fitting the line, and 3) evaluating the validity and usefulness of the model.

4. Results and discussion

4.1. Analysis Results

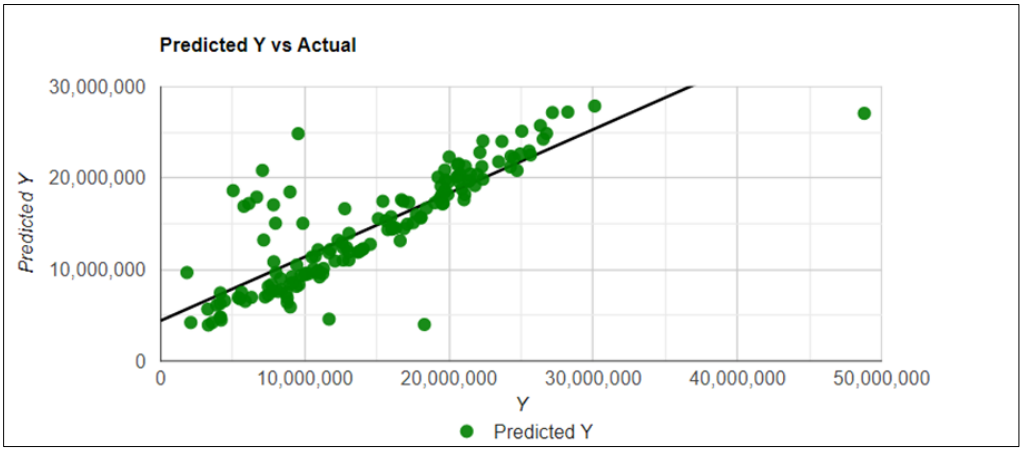

Figure 3 Predicted Y vs Actual

The researchers applied Multiple Linear Regression technique in predicting total estimated cost for road construction analysis. The model is evaluated by the means of R-squared to determine the variables if they are correlated or overfitting. The calculated R-squared is equals to 0.696598 with the predictor variables (x1 & x2) Roadbed width and Net length and it means that the predictors (Xi) explain 69.7% of the variance of Y. The higher the R-squared result, the better fit it is for the Multiple Linear Regression model. It also shows that X1 and X2 are significant predictor variables. The coefficient of multiple correlation (R) is equals to 0.834624 and it means that there is a very strong correlation between the predicted data and the observed data whereas the dependent variable (y) is the Estimated cost. Results of

the multiple linear regression indicated that there was a very strong collective significant effect between the X1, X2, and Y, ($F(2, 147) = 168.75$, $p < .001$, $R2 = 0.7$, $R2adj = 0.69$).

The individual predictors were examined further and indicated that X1 ($t = 3.496$, $p < .001$) and X2 ($t = 16.591$, $p < .001$) were significant predictors in the model.

5. Conclusion and Recommendations

The primary objective of this study is to provide an adequate and reliable prediction of construction estimation cost projects by using Multiple Linear Regression. Given by the data gathered from the detailed estimation of construction secondary road projects in Manila City Hall have similar items and descriptions to be used for the whole project but have different quantities and unit costs depending on the supplier of materials that the contractors will procure. These data are sufficient to train for the Multiple Linear Regression Model and Prediction of cost. In the detailed estimation of road projects, there are six (6) parts of items considered and it is indicated per item its unit cost, quantity, and amount.

The items are not specified which materials to be used for it has a more summarized approach. The sample items that were included are the Bunkhouse, Barricade, Billboard, Construction Safety and Health, Mobilization/Demobilization, aggregate subbase course, aggregate base course, tapping drainage, concrete approach, cutting of concrete pavement, curb and gutter type A, catch basin with C.I Steel grating, Pipe culverts 610 mm dia., Pipe Culverts 910mm dia., Foundation Fill, Removal of Existing Curb and Gutter, Sidewalk (100mm thk), Nawasa Leak/20m, Roadway Excavation, Removal of Sidewalk/Driveways, Hauling of Excavated Materials/Debris, and PCCP 200mm thick 3000psi pumpcrete des.

The researchers were also given the Construction Material Price Data for the region of the National Capital Region and it displays the price differentiation for each city of the region. There are over Seven Hundred and Forty-Four (744) Materials included in the data which have prices from the locations of Malabon, Metro Manila 1st DEO, Metro Manila 2nd DEO, Metro Manila 3rd DEO, North Manna DEO, Quezon City 1st DEO, Quezon City 2nd DEO, South Manila DEO, and Las Pinas-Muntinlupa DEO.

The Multiple Linear Regression Model has proven its purpose by having 0.981 accuracy according to the Confusion Matrix that was utilized as an evaluation tool to predict the cost estimation of each secondary road project given as training data. With thirty-five (35) of the Overpriced as true positive showing that it is accurately predicted while it is also true for the Under-priced which shows sixty-eight (68) for the true negative. With the functionality of uploading CSV files in RStudio, it is considered as time efficient and reliable for importing datasets with a large amount of data involved.

The Multiple Linear Regression Model is plotted using a Scatter plot including a regression line to determine if it is going from zero (0) to one (1). If it is approaching one (1) or upwards, then the regression model appears to be accurate in predicting and if it is going to zero (0) or going downwards, the regression model may be having some lapses and appears to be inaccurate.

The primary objective of this research was to help the government to estimate easier and more accurately when doing different projects. The researchers recommend this study to the government to provide a more accurate estimation with construction contingency in order to be prepared for any unexpected situations resulting in the buyer being able to prepare the exact amount. With the contingency amount added, both the client and government will not have a problem with the budget getting inadequate. Also showing the prices of each material that will be used in every project will make them more trusted since they will be transparent with how they will spend the budget wisely.

The researchers also recommend that it can also compare the bidders' proposals in order to have a fair bidding. They can easily compare each proposal, they can see the materials that will be used, how much they estimate the construction project using the materials they would like to use, and the days they estimated to finish the project. It also indicates if the price that they proposed is overpriced or underpriced which can be easily compared to the predicted cost.

And lastly, the researchers would recommend the prediction of construction estimation and costs of the materials of how the inflation rate moves. It shows the predicted construction estimation with the added inflation rate in the cost of materials. With this feature, they can easily predict costs accurately with the inflation added in every year.

Compliance with ethical standards

Acknowledgments

The completion of this study would not be possible without the help of certain individuals that contribute and extend their help in gathering data. We give our gratitude to our families and loved ones that supported us along the way.

Disclosure of conflict of interest

No conflict of interest.

References

[1] Abd, A., Jasim, N.A., & Naseef, F.S. (2019, February). Developing a Prediction Model to Predict the Construction Project Cost by Using Multiple Linear Regression Technique. Indian Journal of Science and Technology. https://indjst.org/articles/developing-a-prediction-model-to-predict-the-construction-project-cost-by-using-multiple-linear-regression-technique

[2] Asian Development Bank. (2022). COST ESTIMATION IN SOVEREIGN OPERATIONS TECHNICAL GUIDANCE NOTE. https://www.adb.org/sites/default/files/cost-estimation-sovereign-operations.pdf

[3] Branchoux, C., Fang, L., & Tateno, Y. (1 August 2018). Estimating Infrastructure Financing Needs in the Asia-Pacific Least Developed Countries, Landlocked Developing Countries, and Small Island Developing States. Economies 2018.

[4] Bresnahan, B.W., Babigumira, J. B., Vodicka, E., Malik, A. M., Yego, F., Lokangaka, A., Chitah, B.M., Bauer, Z., Chavez, H., Moore, J.L., Garrison, P., Swanson, J.O., Swanson, D., McClure, E.M., Goldenberg, R.L., Esamai, F., Garces, A.L., Chomba, E., Saleem, S., … Nathan, R.O. (2021). Cost estimation alongside a multi-regional, multi-country randomized trial of antenatal ultrasound in five low-and-middle-income countries. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8513981/

[5] Burns, E. (2021, March 30). What is machine learning and why is it important? SearchEnterpriseAI. https://www.techtarget.com/searchenterpriseai/definition/machine-learning-ML

[6] CFI Team. (2022, May 5). Multiple Linear Regression. Corporate Finance Institute website. https://corporatefinanceinstitute.com/resources/knowledge/other/multiple-linear-regression/

[7] Chayalakshmi, C.L., Jangamshetti, D.S., & Sonoli, S. (April 2018). MULTIPLE LINEAR REGRESSION ANALYSIS FOR PREDICTION OF BOILER LOSSES AND BOILER EFFICIENCY. International Journal of Instrumentation and Control Systems.

[8] Conduct and interpret a multiple linear regression. (2021, August 11). Statistics Solutions. https://www.statisticssolutions.com/free-resources/directory-of-statistical-analyses/multiple-linear-regression/#:~:text=Multiple%20Linear%20Regression%20Analysis%20consists,and%20usefulness%20of%20the%20model

[9] DPWH GOV PH. (n.d.). DPWH Infrastructure Projects. DPWH. www.dpwh.gov.ph

[10] DPWH GOV.PH. (2018, June 14). Guidelines in the Preparation of Approved Budget for the Contract (ABC) for Consulting Services for Locally-Funded Infrastructure Projects. Department Order 99 Series of 2018. https://www.dpwh.gov.ph/dpwh/sites/default/files/issuances/DO_099_s2018.pdf

[11] Gonzales, C. (2020, October 19). COA: Over P101B worth of DPWH projects delayed, unimplemented. INQUIRER.net. https://newsinfo.inquirer.net/1349782/coa-over-p101b-worth-of-dpwh-projects-delayed-unimplemented

[12] Hayes, A. (2022, July 8). Multiple linear regression (MLR) definition. Investopedia. https://www.investopedia.com/terms/m/mlr.asp#:~:text=Key%20Takeaways-,Multiple%20linear%20regression%20(MLR)%2C%20also%20known%20simply%20as%20multiple,uses%20just%20one%20explanatory%20variable.

[13] Hussain Khan, T., & Qadir Khan, A. (2015). Effects of Lowest Bidding Bid Awarding System in Public Sector Construction Projects in Pakistan. https://globaljournals.org/GJMBR_Volume15/3-Effects-of-Lowest-Bidding.pdf

[14] Jung, S., Pyeon, J., Lee, H., Yoon, I., & Rho, J. (2020, September 24). Construction Cost Estimation Using a Case-Based Reasoning Hybrid Genetic Algorithm Based on Local Search Method. Sustainability.

[15] Koloğlu, Y., Birinci, H., Kanalmaz, S. I., & Özyılmaz, B. (2018). A Multiple Linear Regression Approach For Estimating the Market Value of Football Players in Forward Position. ABDULLAH GÜL UNIVERSITY INDUSTRIAL ENGINEERING DEPARTMENT.

[16] Ma, Z. (2014, December 3). BIM-based intelligent acquisition of construction information for cost estimation of building projects. Procedia Engineering. https://www.sciencedirect.com/science/article/pii/S1877705814019274

[17] Mishra, G. (2014). https://www.dpwh.gov.ph/dpwh/sites/default/files/issuances/DO_099_s2018.pdf. The Constructor Website. https://theconstructor.org/construction/cost-estimation-factors/9316/?amp=1

[18] Pantalunan, C., Renomeron, C., De Castro, P.S., Manzon, R.D., & Jocson, J. (2021). Elements of Delays in Construction Management of DPWH Aurora District Engineering Office. Engineering and Technology Journal.

[19] Peper, S. (2022). The High Cost of Low Bids: When Contractors Compete on Price, Everyone Loses. https://www.levelset.com/blog/high-cost-low-bids/

[20] Peterson, A. D., & Ziegler, L. (2021). Building a Multiple Linear Regression Model With LEGO Brick Data. Journal of Statistics and Data Science Education.

[21] Predicting Construction Cost Using Multiple Regression Techniques. (n.d.). Journal of Construction Engineering and Management, 132(7). https://www.researchgate.net/publication/245283472_Predicting_Construction_Cost_Using_Multiple_Regression_Techniques

[22] Predicting Construction Cost Using Regression Techniques For Residential Building. (2020, October 5). Journal of Physics: Conference Series. https://iopscience.iop.org/article/10.1088/1742-6596/1899/1/012120/pdf

[23] Robes, P. (2019, August 13). Software Quality Standards - How and Why we Applied ISO 25010. Monterail website. https://www.monterail.com/blog/software-qa-standards-iso-25010

[24] Tafeyeh Hashemi, S., Ebadati, O.M., & Kaur, H. (2020, September 5). Cost estimation and prediction in construction projects: A systematic review on machine learning techniques - SN Applied Sciences. SpringerLink. https://link.springer.com/article/10.1007/s42452-020-03497-1

[25] Uyanık, G. K., & Güler, N. (2013). A Study on Multiple Linear Regression Analysis. Procedia - Social and Behavioral Sciences, 106. https://www.sciencedirect.com/science/article/pii/S1877042813046429#bibl0005

[26] Writer, S. (2015, August 4). What Are Scope and Limitations in Research? Reference.com. Retrieved December 12, 2022, from https://www.reference.com/science-technology/scope-limitations-research-93aa9f6a2e9c469

[27] Goodwin, K. (2018, November 8). How to Calculate Future Cost of Goods Using Inflation Rates. https://pocketsense.com/recessions-effects-cpi-8135118.html