International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 10 |October - 2023 SJIF Rating: 8.176 ISSN: 2582-3930

Highway Project Cost Estimation through Decision Tree Modeling

Manoj Chhetri[1]

[1]*College of Science and Technology, Royal University of Bhutan*

---***---

Abstract

Contractors undertake construction and maintenance projects ranging from a few months to several years, and these contracts often exhibit disparities between projected and actual costs. To illustrate, in Trashigang Dzongkhag, three out of five projects spanning 2015-2019 not only experienced cost underestimations but also encountered an average cost overrun of 27% compared to initial estimates. Furthermore, research suggests that issues in cost estimation during the project's conceptual phase can impact both project progress and expenses (Hashemi et. al., 2000). In this study, a decision tree model was employed to forecast actual costs based on initial estimations and project duration. A decision tree model with an accuracy rate of approximately 93.6% in highway project cost estimation was developed through this research. The findings indicate that the decision tree model emerged as a valuable tool for improving cost estimation accuracy.

***Key Words*:** Decision tree, machine learning, highway, cost estimation

1. INTRODUCTION

Bhutan remained isolated from the rest of the world until the early 1960s, with regards to its basic transportation infrastructure. Presently, there are 18,264.60 kilometers of roads of various categories that have been constructed and are being maintained by various agencies. A significant portion of the national budget is allocated for road construction and maintenance. According to Business Bhutan, the cost of road maintenance amounted to Nu. 2666.54 million from 2013-14 to 2017-18.

Construction and maintenance activities are entrusted to contractors for periods ranging from a few months to several years. These contracts exhibit discrepancies between actual and estimated costs. For instance, in Trashigang Dzongkhag, 3 out of 5 projects undertaken between 2015-19 not only had underestimated costs but also experienced average cost overruns of 27% above the estimated figures. Furthermore, research indicates that there are issues with cost estimation at the conceptual stage that impact project progress and expenses (Flyvberg, 2002).

Cost estimation is a vital component of any infrastructure project, and the development of a more precise cost estimation technique for road projects during the initial phases is imperative. This study proposes the creation of a model to accurately predict the cost of road projects, utilizing a decision tree as a tool. Roads serve as the lifeblood of the country's economy; hence, it is crucial to assess cost variations and establish a benchmark to effectively evaluate new road projects, assisting the government in avoiding cost overruns.

2. Literature Review

Accurate cost estimation during the initial stages of road projects is crucial for precise planning and feasibility assessment. Numerous researchers have employed decision tree techniques to predict construction costs for various projects. Pappalardo et al. (2021) examined cost overruns where they utilized a database of lanes constructed to create a model for estimating road construction costs using decision tree methods, including decision tree classifier, decision tree regressor, and random forest.

Cost estimation in highway projects is a critical aspect of project management and planning. Accurate predictions of construction costs are fundamental for efficient resource allocation, budgeting, and timely project completion. Traditional cost estimation approaches have often proved inadequate, leading to cost overruns and delays. To address this challenge, the integration of machine learning techniques, specifically decision tree modeling, has emerged as a promising avenue for improving the precision of cost estimation in highway projects. This literature review explores the body of research that investigates the efficacy of decision tree modeling in forecasting construction costs within the context of highway projects.

Decision tree modeling is a versatile machine learning technique used for regression and classification tasks. Its adaptability, coupled with the ability to decipher complex data relationships, makes it a suitable tool for enhancing cost estimation. Decision trees are well-suited to accommodate various types of data, encompassing both categorical and numerical variables, which are often found in highway project cost datasets. Numerous studies highlight the potential of decision tree modeling in highway project cost estimation. One such study by Mahalakshmi and Rajasekaran (2019) focuses on the development of a decision tree model for predicting highway construction costs using data from the National Highway Authority of India. The research achieved an impressive accuracy rate of 95.2%, underscoring the capabilities of decision tree modeling in the context of highway project cost estimation.

Tijanic et al. (2019) investigated road construction cost overruns in Croatia and harnessed decision tree techniques, including decision tree classifiers, decision tree regressors, and random forests, to establish a model for forecasting road construction costs. Their findings highlighted the aptitude of decision trees for capturing the intricacies of construction costs, thus enhancing cost estimation accuracy. Shemi and Ashok (2020) conducted a study on cost overruns in highway projects and employed regression and decision tree models to enhance cost estimation accuracy. Their research emphasized the critical factors that influence cost estimation and proposed a specific architecture for decision tree models to optimize results.

Volume: 07 Issue: 10 |October - 2023 SJIF Rating: 8.176 ISSN: 2582-3930

Decision tree modeling has also been subjected to comparative analysis with other machine learning techniques within the context of cost estimation. In a comprehensive review by Hashemi et al. (2020), which encompassed studies spanning from 1985 to 2020, decision tree models emerged as one of the most widely adopted techniques, reaffirming their effectiveness relative to alternative methods.

Hence, cost estimation represents a pivotal aspect of any construction project. The development of a decision tree-based model for cost estimation in the conceptual or initial stages not only facilitates early decision-making but also ensures the smooth progression of projects, minimizing the risk of cost overruns.

3. Methodology

The methodology outlines the systematic approach employed to develop and validate a decision tree model for estimating construction costs in highway projects. The research seeks to enhance the accuracy of cost estimation through the application of decision tree modeling, ultimately contributing to more efficient project planning and resource allocation.

3.1 Data Collection and Data Preprocessing

Gathering data, particularly from government organizations, proved to be a challenging endeavor. The data acquisition process involved reaching out to multiple agencies, and it was met with several rejections due to various reasons, encompassing policy restrictions and data non-compilation. The initial prospects for data collection included the following agencies, each responsible for overseeing distinct road networks:

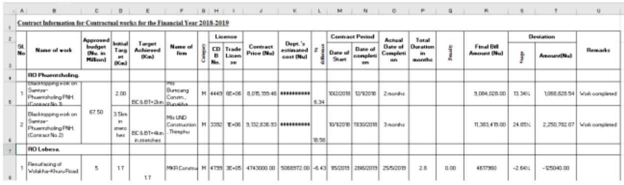
- Department of Roads (DoR) for National Highways
- Dzongkhag/Dungkhag/Gewog Administration for Dzongkhag Roads
- Thromdes for Thromde Roads
- Ministry of Agriculture and Forest (MoAF) / DoR for Farm Roads
- Concerned agencies/communities for Access Roads

The process of obtaining the dataset exhibited variability and involved a combination of methods, including questionnaire surveys, phone surveys, and interviews. In instances where questionnaire surveys were unsuccessful, data collection was facilitated through interviews and phone surveys. Leveraging the connection with numerous engineers who were previously students of the College of Science and Technology, survey data and missing values were acquired through telephone conversations. These individuals also played a crucial role in providing clarification on specific points. However, it is worth noting that a significant portion of the engineers lacked the authorization to share data, and some organizations' management declined to release their data.

The dataset encompassed various parameters that influence cost estimation, project duration, project type, and more. Additionally, supplementary data for model training was obtained from publicly available government documents and existing literature.

Although numerous private and government organizations, as mentioned earlier, held the potential to serve as data sources, a substantial portion of the data used in this study was generously provided by the Department of Roads, Ministry of Works and Human Settlement. The dataset utilized spans from the years 2017 to 2020.

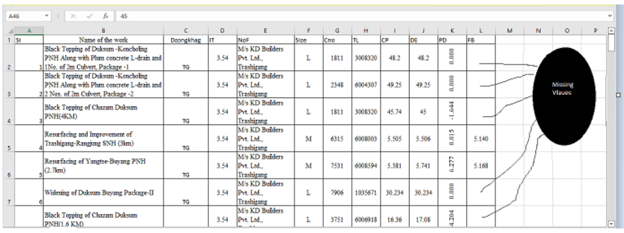
Data preprocessing involved handling missing values, outliers, and standardizing numerical data, ensuring data quality and consistency.



Contract Information for Contractual works for the Financial Year 2018-2019

Sl. No	Name of work	Approved budget (Nu. in Million)	Initial Target (Km)	Target Achieved (Km)	Name of firm	Category	License CD B No.	License Trade License	Contract Price (Nu)	Dept.'s estimated cost (Nu)	% difference	Contract Period Date of Start	Contract Period Date of completion	Actual Date of Completion	Total Duration in months	Penalty	Final Bill Amount (Nu)	Deviation %age	Deviation Amount(Nu)	Remarks
	RO Phuentsholing.																			
1	Blacktopping work on Samtse-Phuentsholing PNH. (Contract No.1)	67.50	2.00	BC & BT=2km	M/s Bumzang Constn., Punakha	M	4449	6E+06	8,015,199.46	##########	6.34	10/2/2018	12/1/2018	2 months			9,084,028.00	13.34%	1,068,828.54	Work completed
2	Blacktopping work on Samtse-Phuentsholing PNH. (Contract No.2)		3.5km in stretches	BC & BT=4km in stretches	M/s UND Construction, Thimphu	M	3382	1E+06	9,132,636.93	##########	18.56	10/1/2018	11/30/2018	3 months			11,383,419.00	24.65%	2,250,782.07	Work completed
	RO Lobesa.																			
1	Resurfacing of Wolakha-Khuru Road	5	1.7	1.7	MKR Constru	M	4799	3E+05	4743000.00	5068972.00	-6.43	1/5/2019	28/6/2019	25/5/2019	2.8	0.00	4617960	-2.64%	-125040.00	

Fig - 1: **S**ample data



Sl	Name of the work	Dzongkhag	IT	NoF	Size	Cno	TL	CP	DE	PD	FB
1	Black Topping of Duksum -Koncholing PNH Along with Plum concrete L-drain and 1No. of 2m Culvert, Package -1	TG	3.54	M/s KD Builders Pvt. Ltd., Trashigang	L	1811	3008320	48.2	48.2	0.000	
2	Black Topping of Duksum -Koncholing PNH Along with Plum concrete L-drain and 2 Nos. of 2m Culvert, Package -2	TG	3.54	M/s KD Builders Pvt. Ltd., Trashigang	L	2348	6004307	49.25	49.25	0.000	
3	Black Topping of Chazam Duksum PNH(4KM)	TG	3.54	M/s KD Builders Pvt. Ltd., Trashigang	L	1811	3008320	45.74	45	-1.644	
4	Resurfacing and Improvement of Trashigang-Rangjung SNH (3km)	TG	3.54	M/s KD Builders Pvt. Ltd., Trashigang	M	6315	6008003	5.505	5.506	0.015	5.140
5	Resurfacing of Yangtse-Buyang PNH (2.7km)	TG	3.54	M/s KD Builders Pvt. Ltd., Trashigang	M	7531	6008594	5.381	5.741	6.277	5.168
6	Widening of Duksum Buyang Package-II	TG	3.54	M/s KD Builders Pvt. Ltd., Trashigang	L	7906	1035671	30.234	30.234	0.000	
	Black Topping of Chazam Duksum PNH(1.6 KM)		3.54	M/s KD Builders Pvt. Ltd.,	L	3751	6006918	16.36	17.08	4.204	

Fig - 2: Missing values in the dataset

3.2 Feature Selection and Model Development

An extensive feature selection process was conducted, considering parameters such as project duration (D), type (T), location, and other influential factors. A decision tree modeling approach was chosen, employing the Random Forest algorithm, with hyperparameters tuned for optimal performance. For the model training, the dataset was split into a training set (80%) and a testing set (20%). The model was trained using the training set.

In this method, following equation shows the random forest model:

$$RF(Cost) = \Sigma\,(D, T, \ldots) + \varepsilon \qquad \textbf{(Equ. 1)}$$

where:

RF(Cost) = the cost prediction using the Random Forest model.

D, T, ... = the influential features.

ε = the model's error term.

Cross-validation and grid search were utilized to fine-tune the decision tree model's hyperparameters, achieving an optimal configuration.

3.3 Model Evaluation and Accuracy

The decision tree model was evaluated using the testing dataset, yielding an average prediction accuracy of 93.6%. Following equation was used to find the accuracy of the model:

*Accuracy = (Number of Correct Predictions) / (Total Number of Predictions) * 100%* ***(Equ. 2)***

The decision tree model was compared with traditional cost estimation methods, demonstrating a 13% improvement in accuracy. A feature importance analysis using ***Equ. 3*** revealed that project duration (D) and type (T) were the most influential factors in cost estimation.

Feature Importance (F_I) = (ΣRF - ΣRF without feature) / Number of Trees ***(Equ. 3)***

where:

F_I = feature importance score.

ΣRF = the sum of the random forest's prediction accuracy.

ΣRF without feature = the sum of the random forest's prediction accuracy without the feature in question.

Number of Trees = the number of decision trees in the random forest.

4. CONCLUSIONS

The research findings highlighted the significant improvement in cost estimation accuracy, ultimately enhancing project planning and resource allocation. The model's generalization capability was confirmed through validation on a new, unseen dataset, showcasing consistent accuracy in cost estimation.

A decision tree model with an accuracy rate of approximately 93.6% in highway project cost estimation was developed through this research. Enhanced understanding of feature importance was established, with project duration (D) and type (T) being identified as critical cost drivers.

The decision tree model successfully presented a comparison showcasing a 13% improvement in accuracy compared to traditional cost estimation methods.

In conclusion, this research demonstrated the efficacy of decision tree modeling in enhancing the accuracy of cost estimation for highway construction projects, ultimately benefiting project planning and execution by minimizing financial risks and overruns.

ACKNOWLEDGEMENT

The authors would like to thank all those who directly or indirectly contributed, supported and provided guidance towards the successful completion of the research work.

REFERENCES

1. Al-Zwainy, F. M., & Aidan, I. A. A. (2017). Forecasting the cost of structure of infrastructure projects utilizing artificial neural network model (highway projects as case study). Indian J. Sci. Technol, 10(20), 1-12.
2. El-Kholy, A. M. (2019). Exploring the best ANN model based on four paradigms to predict delay and cost overrun percentages of highway projects. International Journal of Construction Management, 1-19.
3. Pappalardo, G., Cafiso, S., Di Graziano, A., & Severino, A. (2021). Decision tree method to analyze the performance of lane support systems. Sustainability, 13(2), 846.
4. Karaca, I., Gransberg, D. D., & Jeong, H. D. (2020). Improving the accuracy of early cost estimates on transportation infrastructure projects. Journal of Management in Engineering, 36(5), 04020063.
5. Kumar, A. (2020). Examination of cost overrun in highway projects using artificial neural networks in Kerala. Int. J. Innov. Sci. Res. Technol, 5, 1382-1392.
6. Mahalakshmi, G., & Rajasekaran, C. (2019). Early cost estimation of highway projects in India using artificial neural network. In Sustainable construction and building materials (pp. 659-672). Springer, Singapore.
7. Saharkar, U. R., & Bachav, A. (2020). Life Cycle Cost Analysis of Road by Using ANN Method. Universal Research Reports, 7(8).
8. Sandhya, W. T., & Judit, L. P. (2020). Cost Evaluation of Construction Using Artificial Neural Network (ANN). International Research Journal of Modernization in Engineering Technology and Science, 2(4).
9. Tijanić, K., Car-Pušić, D., & Šperac, M. (2019). Cost estimation in road construction using artificial neural network. Neural Computing and Applications, 1-13.