World Academy of Science, Engineering and Technology
International Journal of Civil and Architectural Engineering
Vol:18, No:07, 2024

Predicting Costs in Construction Projects with Machine Learning: A Detailed Study Based on Activity-Level Data

Authors : Soheila Sadeghi

Abstract : Construction projects are complex and often subject to significant cost overruns due to the multifaceted nature of the activities involved. Accurate cost estimation is crucial for effective budget planning and resource allocation. Traditional methods for predicting overruns often rely on expert judgment or analysis of historical data, which can be time-consuming, subjective, and may fail to consider important factors. However, with the increasing availability of data from construction projects, machine learning techniques can be leveraged to improve the accuracy of overrun predictions. This study applied machine learning algorithms to enhance the prediction of cost overruns in a case study of a construction project. The methodology involved the development and evaluation of two machine learning models: Random Forest and Neural Networks. Random Forest can handle high-dimensional data, capture complex relationships, and provide feature importance estimates. Neural Networks, particularly Deep Neural Networks (DNNs), are capable of automatically learning and modeling complex, non-linear relationships between input features and the target variable. These models can adapt to new data, reduce human bias, and uncover hidden patterns in the dataset. The findings of this study demonstrate that both Random Forest and Neural Networks can significantly improve the accuracy of cost overrun predictions compared to traditional methods. The Random Forest model also identified key cost drivers and risk factors, such as changes in the scope of work and delays in material delivery, which can inform better project risk management. However, the study acknowledges several limitations. First, the findings are based on a single construction project, which may limit the generalizability of the results to other projects or contexts. Second, the dataset, although comprehensive, may not capture all relevant factors influencing cost overruns, such as external economic conditions or political factors. Third, the study focuses primarily on cost overruns, while schedule overruns are not explicitly addressed. Future research should explore the application of machine learning techniques to a broader range of projects, incorporate additional data sources, and investigate the prediction of both cost and schedule overruns simultaneously.

Keywords : cost prediction, machine learning, project management, random forest, neural networks
Conference Title : ICCPM 2024 : International Conference on Construction and Project Management
Conference Location : Santorini, Greece
Conference Dates : July 11-12, 2024