

Highway Project Clustering Using Unsupervised Machine Learning Approach

Hamed Alikhani¹ and H. David Jeong, Ph.D.²

¹ Ph.D. Student, Interdisciplinary Engineering, Texas A&M University, College Station, TX 77843. Email: hamedalikhani@tamu.edu

² Professor, Dept. of Construction Science, Texas A&M University, College Station, TX 77843. Email: djeong@tamu.edu

ABSTRACT

Many studies were conducted to predict project time, cost, and risks using historical data of similar past projects, assuming similar projects within the same group follow similar patterns. However, the accuracy of prediction models depends on the project grouping accuracy in categorizing similar projects together. Current practice in US highway agencies indicates categorizing projects based on similar activities into different work types, which can be challenging due to the lack of a solid definition of project work types and the variety of projects' activities. Besides, the accuracy of categorization with this method has not been determined. This paper develops a K-means clustering model to group similar highway projects together. The Euclidian distance was used as a metric to detect similarity and evaluate the performance of the method. The historical data of 730 projects from a highway agency in the US was used to validate the model and compare the performance of the model with the current categorization method used in US highway agencies based on project work type. The comparison showed the superiority of K-means clustering in grouping similar projects together. The proposed reliable clustering model can be used in future studies to increase the accuracy of prediction models that use the project type as an influential factor.