Machine Learning-based Life-cycle Cost Analysis for Educational Facilities: A Case Study

Xinghua Gao, Pardis Pishdad-Bozorgi, Ph.D., and Dennis R. Shelden, Ph.D., AIA
Georgia Institute of Technology
Atlanta, GA

Because of the long life spans of buildings, robust decisions regarding the economic efficiency of alternative materials, components, and systems demand a full lifecycle perspective that goes beyond the initial cost and regular maintenance and repair. The Life Cycle Cost Analysis (LCCA) has become increasingly important in new building design and existing building retrofitting, refurbishment, and renovations. However, unlike commercial and residential buildings, the real service lives and costs of many complex educational buildings are difficult to predict because 1) there is always a mismatch between the predicted energy performance of buildings and actual measured performance, 2) building systems and components, with proper maintenance and repair, can function beyond the warranty, which makes their true costs difficult to predict, and 3) even the same type of systems used in different buildings may have different LCCs because the monetary and labor costs vary depending on each university's operational profile on building systems. Machine learning is an automated process that extracts patterns from data and devises complex prediction algorithms and models. The developments of machine learning techniques and more advanced building systems provide building experts with new opportunities to achieve more accurate predictions of facility-related costs.

This study aims to investigate the feasibility of utilizing the historical data housed in heterogeneous building systems of a multi-facility entity to predict the LCCs of its facilities through machine learning. An LCCA approach is proposed to collect the data generated by different building systems and analyze the data with machine learning techniques to predict the future costs of the organization's new and existing facilities, and thus to achieve better decision makings in building design, retrofitting, refurbishment, and renovations.

A case study is conducted on multiple facilities of a university to demonstrate the proposed machine learning-based LCCA process. Multiple machine learning methods are tested to determine the best ones for predicting each LCC component and to develop the LCC regression model, including linear regression, SVM, multilayer perceptron, k-Nearest Neighbors (kNN), decision trees, multi-output support vector regression, and multi-output regression trees. In a 20-year time span, the ratio of the utility cost and O&M cost to the initial cost showed an approximately normal distribution. On average, the overall utility cost is equal to 30.40% of the initial cost (standard deviation = 24.48%), while the O&M cost 71.19% (standard deviation = 66.08%), both with inflation considered. Athletic facilities and computing centers consume much higher energy than other types of facilities. In addition, the building age has a negative correlation (-0.454) with the initial cost, which indicates that the design and construction of buildings become increasingly expensive over time. On the other hand, the older buildings tend to cost more on O&M and less energy efficient.

This research contributes to the body of knowledge by investigating the feasibility of obtaining an accurate estimation of facilities' life-cycle costs (LCCs) by implementing machine learning on historical data. By exploring the new possibility for better prediction of a facility' LCC through leveraging historical data housed in heterogeneous building systems across a continuous network of buildings, this research has a greater impact than simply studying the LCC of an individual project in the programming or design phase. The impact involves data-based LCC inputs in future facility cost benchmarking and informed project developments by incorporating the data pertaining to the total cost of ownership. Using existing available data to benchmark facility costs can assist decision making, and new data can be incorporated as they become available.

Keywords: Machine learning, Life-cycle Cost Analysis (LCCA), data mining, cost prediction