Machine Learning-Based Building Life-Cycle Cost Prediction: A Framework and Ontology

Xinghua Gao[1]; Pardis Pishdad-Bozorgi[2]; Dennis Shelden[3]; and Shu Tang[4]

[1]Myers-Lawson School of Construction, Virginia Polytechnic Institute and State Univ., Blacksburg, VA. E-mail: xinghua@vt.edu
[2]School of Building Construction, Georgia Institute of Technology, Atlanta, GA. E-mail: pardis.pishdad@design.gatech.edu
[3]Digital Building Laboratory, School of Architecture, Georgia Institute of Technology, Atlanta, GA. E-mail: dennis.shelden@design.gatech.edu
[4]Digital Building Laboratory, School of Architecture, Georgia Institute of Technology, Atlanta, GA. E-mail: shutang@gatech.edu

ABSTRACT

Numerous costs are associated with the design, construction, installation, operation, maintenance, and deconstruction of a building or building system. One of the challenges usually faced by an organization's capital planning department and/or facility management department is that they do not have an effective means to quickly estimate a new facility's whole life-cycle costs (LCC) during the programming phase when no building design is available. To provide facility managers and owners with an effective and reliable means to assess the total cost of the facility ownership, the authors are developing an approach that uses the historical data stored in multiple building systems and building information models (BIM) as basis to predict facilities' LCC—initial design and construction cost, utility cost, and operation and maintenance cost. In this paper, the authors propose a machine learning-enabled facility LCC analysis framework using data provided by building systems. The corresponding domain ontology—LCCA-Onto—is also presented. The proposed approach provides organizations who own multiple facilities with an innovative solution to the LCC prediction issue.

INTRODUCTION

Buildings are expensive to construct and can be demanding to maintain. Understanding the initial and future costs associated with the construction and maintenance of buildings is critical for an organization's resource planning. The life-cycle cost (LCC) analysis has become increasingly important in new building design and existing building retrofitting, refurbishment, and renovations. However, despite its importance, researchers and industry professionals are facing challenges when practicing LCC analysis in the Architecture, Engineering, Construction, and Owner-operated (AECO) industry. Two of the main barriers are the shortage of life-cycle cost (LCC) data (Cole and Sterner 2000; Bakis et al. 2003) and the complexity of predicting real future costs (Ferry and Flanagan 1991; Cole and Sterner 2000).

Currently, many building systems – such as Building Automation Systems (BAS), Computerized Maintenance Management Systems (CMMS), and Building Energy Management Systems (BEMS) – are constantly generating, recording, and storing data (Gao et al. 2018; Gao et al. 2019). Some of these data are related to building LCC, such as the building component price history, utility consumptions, and maintenance work orders. Utilizing these kinds of data housed in separate building systems is a potential solution for the data shortage issue in the facility LCC analysis field.

Machine learning is an automated process that extracts patterns from data (Kelleher et al.

2015). In the field of predictive data analytics, machine learning is a method used to devise complex prediction algorithms and models (Mitchell 1997; Kelleher et al. 2015). These analytical models enable data analysts to uncover hidden insights, predict future values, and produce reliable, repeatable decisions through learning from historical relationships and trends in the data (SAS 2018). With sufficient data, machine learning techniques can be used to estimate facility-related costs (Gao et al. 2019).

The authors' hypothesis is that the evolving building systems already contain many valuable data for LCC analysis but not being used because they are not connected, available to analysts in a consumable way. By extracting relevant data from these systems and implementing machine learning on the data, we can have a better understanding of the facility's LCC and overcome multiple barriers of current LCC analysis methods, and we can achieve more informed decisions in building design, construction, and facility management.

This research presents a machine learning-enabled facility LCC analysis framework using data provided by building systems. An LCC prediction system is being developed based on the proposed framework and corresponding domain ontology – LCCA-Onto. This system can be used in the programming phase to forecast the initial costs, utility costs, and operation and maintenance (O&M) costs of a new building.

BACKGROUND

Accurate estimation in the early design stage is vital for the successful execution of a construction project. Using machine learning techniques, research studies have provided practitioners with decision-support tools for estimating construction duration and costs before the completion of a project's design stage, or even during the programming phase (Koo et al. 2010; Hong et al. 2011; Jin et al. 2016). Moreover, understanding the underlying dynamics of building utility consumption (energy, water, and gas) and predicting the consumption are essential for building resource planning, management, and conservation (Amasyali and El-Gohary 2018; Zhang et al. 2018). Energy (electricity) consumption prediction is the most extensively studied topic in the facility LCC prediction field (Gao et al. 2019). The most commonly used machine learning methods for energy forecasting involve: 1) Artificial Neural Network (ANN) (Mocanu et al. 2016; Park et al. 2018; Sala-Cardoso et al. 2018), 2) Support Vector Machines (SVM) Regression (Jain et al. 2014; Chou and Ngo 2016), and 3) Case-based Reasoning (An et al. 2007; Ji et al. 2014). Studies on using machine learning to predict O&M costs are relatively rare. This is probably because obtaining accurate maintenance data is challenging (Neely and Neathammer 1991). The most commonly used machine learning methods in O&M cost forecasting are multiple regression (Li and Guo 2012; Au-Yong et al. 2014; Weerasinghe et al. 2016; Krstić and Marenjak 2017) and ANN (Li and Guo 2012; Tu and Huang 2013).

Although machine learning techniques have been implemented in forecasting construction costs, utility consumption, and O&M costs, respectively, its application in predicting a building's whole LCC is rarely found in the literature. More studies that utilize machine learning to predict a building's overall LCC and shed light on the underlying relationships between each cost components are needed. It is possible to establish generalizable frameworks for developing facility LCC analysis machine learning models. This research presents a machine learning-enabled facility LCC analysis framework and the corresponding domain ontology that can be used to develop a machine learning-enabled LCC prediction system.

DEVELOPING MACHINE LEARNING MODELS FOR FACILITY LCC ANALYSIS

Figure 1 shows the proposed framework for developing machine learning models for facility LCC analysis, which consists of four major modules: 1) obtaining the descriptive attributes, 2) obtaining the target attributes, 3) training machine learning models, and 4) evaluating the models and selecting the most suitable one.

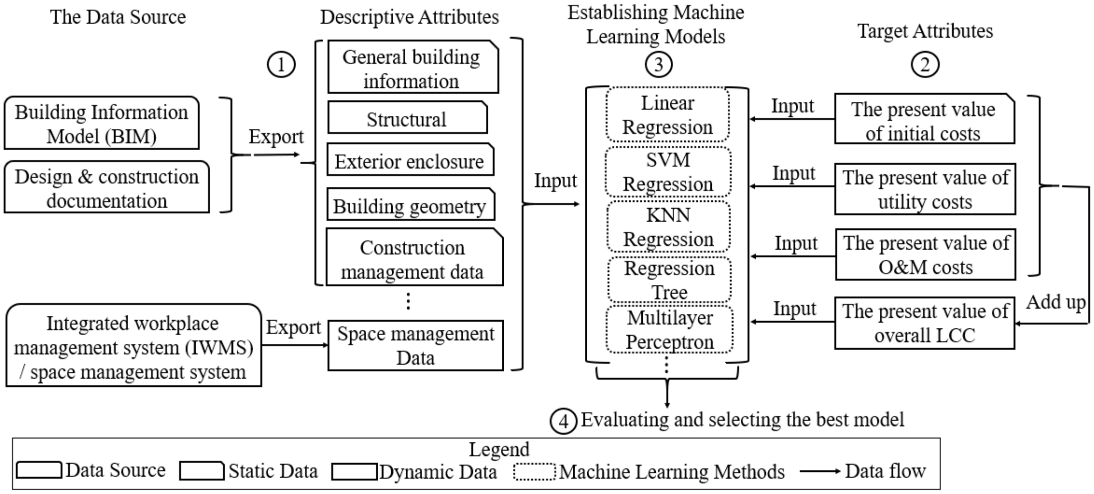
The Data Source
Descriptive Attributes
Establishing Machine Learning Models
Target Attributes
①
②
③
Building Information Model (BIM)
Design & construction documentation
Export
General building information
Structural
Exterior enclosure
Building geometry
Construction management data
Integrated workplace management system (IWMS) / space management system
Export
Space management Data
Input
Linear Regression
SVM Regression
KNN Regression
Regression Tree
Multilayer Perceptron
Input
Input
Input
Input
The present value of initial costs
The present value of utility costs
The present value of O&M costs
The present value of overall LCC
Add up
④ Evaluating and selecting the best model
Legend
Data Source
Static Data
Dynamic Data
Machine Learning Methods
Data flow

Figure 1. A framework for developing machine learning models for facility LCC analysis

Module 1: Obtaining the Descriptive Attributes

The potential descriptive attributes in the LCC analysis machine learning model can be provided by BIM (Eastman et al. 2011; Pishdad-Bozorgi et al. 2018; Gao and Pishdad-Bozorgi 2019). For the organizations that do not have well-developed BIM (e.g. BIM with a level of development 400) for all facilities, the required data can be found in the design and construction documentation. For example, design drawings contain building geometry, structural, foundation, and general building information, such as building age and function, while the construction documents contain construction management related information, such as the delivery method and construction duration (which may influence the initial cost). After operation, the buildings' space allocations may change over time and this kind of change may not be timely reflected in the BIM. In this case, the up-to-date space allocation data can be found in the integrated workplace management system (IWMS) or other space management system.

Module 2: Obtaining the Target Attributes

The derivation of target attributes – the present value of initial costs, utility costs, and O&M costs – is shown in Figure 2. The raw data used for deriving the cost components are extracted from multiple building systems. After the data are stored in one database, machine learning techniques can be implemented on them to forecast each LCC component of a building. In contrast to the descriptive attributes, which are relatively static in a certain time period (such as three months), the target attributes are dynamic and can vary with the real-time utility consumption and O&M costs. The data indexed in time order are analyzed by time series methods, and projections are made when necessary, such as when there are missing values because sensors were not deployed in the past. The public statistics, such as the historical inflation rate, utility price, and labor rate, are incorporated into the analysis to calculate the monetary costs and to convert the costs to their present values. These present values of the LCC components are the target attributes of the LCC analysis machine learning models.

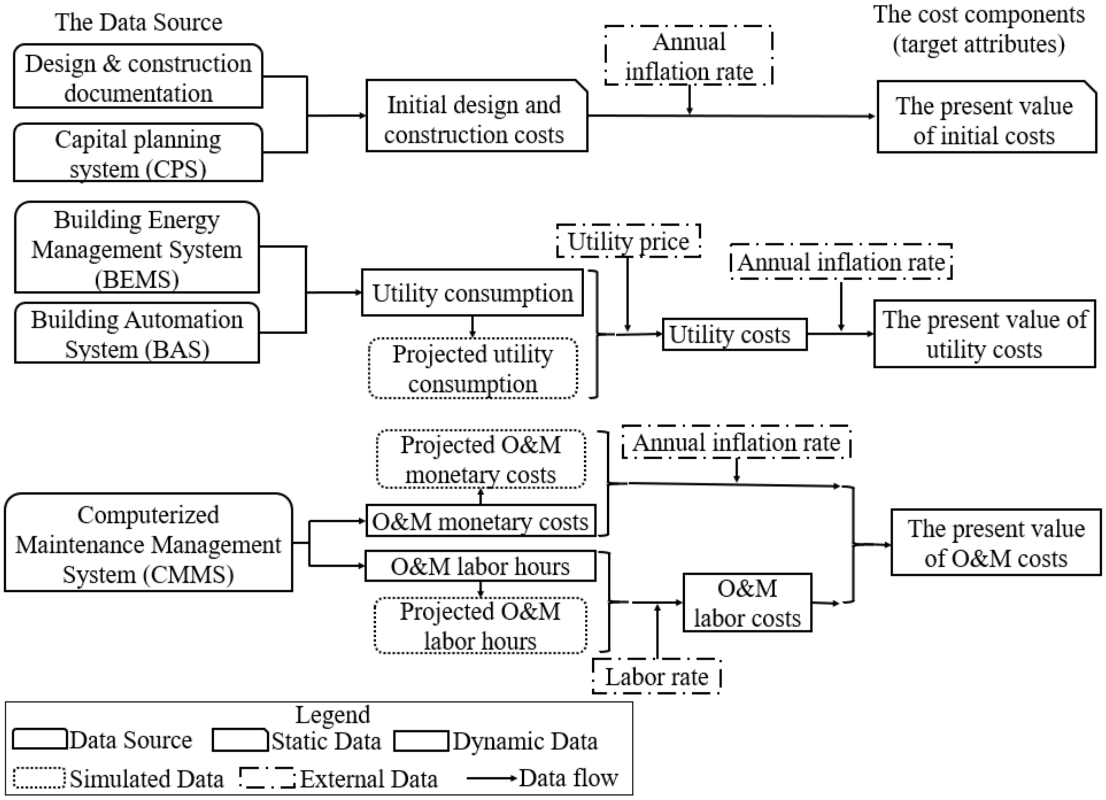
The Data Source
The cost components (target attributes)
Design & construction documentation
Capital planning system (CPS)
Initial design and construction costs
Annual inflation rate
The present value of initial costs
Building Energy Management System (BEMS)
Building Automation System (BAS)
Utility consumption
Projected utility consumption
Utility price
Utility costs
Annual inflation rate
The present value of utility costs
Computerized Maintenance Management System (CMMS)
Projected O&M monetary costs
O&M monetary costs
O&M labor hours
Projected O&M labor hours
Annual inflation rate
O&M labor costs
Labor rate
The present value of O&M costs
Legend
Data Source
Static Data
Dynamic Data
Simulated Data
External Data
Data flow

Figure 2. The derivation of cost components (target attributes)

Module 3: Training Machine Learning Models

With the descriptive attributes and target attributes ready, the next step is to train the machine learning models based on these data. The machine learning methods involve linear regression, SVM regression, KNN regression, regression tree, and multilayer perceptron. These methods are proved to be effective in building-related costs prediction (Gao et al. 2019). In this framework, the method pool of training regression models for facility LCC analysis is expandable. With the development of machine learning techniques in predictive data analytics, more methods can be adopted and implemented within the framework.

Module 4: Evaluating the Models

Evaluating the models and selecting the outperformed machine learning algorithm for facility LCC prediction can be done by repeated random sampling and cross-validation, and then comparing their performance (Hu and Castro-Lacouture 2018). The most suitable machine learning method would possibly be different from case to case, depending on the length of studied time span, attributes used, and data size and quality.

THE ONTOLOGY OF THE LCC ANALYSIS SYSTEM

The authors are developing an LCC analysis system based on the proposed framework. This system can be used by an organization's facility management and capital planning department to have a preliminary estimate of a building's LCC, new or existing. Without any input of professional estimators nor the detailed building design, the system is expected to yield a reasonable prediction (e.g. with an error of 20%) based on limited input parameters, such as gross square footage, number of floors, structural type, and space allocation (such as 30% residential, 20% office, 20% general usage, etc.). This section presents the system's ontology.

An ontology defines a common vocabulary for sharing information in a domain (Studer et al. 1998; Noy and McGuinness 2001). It includes both human-understandable and machine-interpretable definitions of concepts in the domain and relations among them (Studer et al. 1998; Noy and McGuinness 2001). Conceptual analysis and knowledge representation often require ontological support, therefore, developing a domain ontology is one of the fundamental steps when developing a shared model of knowledge (Cristani and Cuel 2004). Typically, an ontology involves the formal, explicit description of 1) concepts in a domain of discourse – usually referred as "classes" or "concepts", 2) properties of each concept, describing various features and attributes of the concept – usually referred as "slots" or "properties", and 3) restrictions on slots – usually referred as "facets" or "role restrictions" (Noy and McGuinness 2001).

LCCA-Onto: Scope and Language

The domain ontology developed in this research, LCCA-Onto, is focused on machine learning-enabled facility LCC analysis using the data provided by BIM and building systems. Currently, there is no ontology that can fulfil this purpose in the facility LCC analysis domain. Existing ontologies, such as IFC (buildingSMART 2017), UniFormat (Charette and Marshall 1999), and MasterFormat (Construction Specifications Institute (CSI) 2016), are adopted to represent the classes (concepts) in the corresponding domain (e.g. IFC represents the building components, UniFormat and MasterFormat represent the construction work breakdown structure).

LCCA-Onto is developed with the W3C Web Ontology Language (OWL). OWL is a Semantic Web language designed to represent ontologies (OWL Working Group 2019). The tool used to develop LCCA-Onto is protégé (version 5.5) (Stanford Center for Biomedical Informatics Research 2019). The major classes involved in the LCCA-Onto are *Community*, *Building*, *Building_system*, *Data*, *Data_standard*, and *MACH_LRN_tool* (machine learning tool).

LCCA-Onto: definition of class

In LCCA-Onto, the classes *Community* and *Building* do not have any subclasses. The class *Building_system* contains five subclasses: *BAS* (Building Automation System), *BEMS* (Building Energy Management System), *CMMS* (Computerized Maintenance Management System), *IWMS* (Integrated Workplace Management System), and *BIM* (Building Information Model).

The subclasses of *MACH_LRN_tool* represent the machine learning tools for predictive data analytics (non-exhaustive), which involve *Linear_regression*, *SVM_regression*, *KNN_regression*, *Regression_tree*, *Time_series_regression*, and *MLP_regression*. The list can be further expanded with other machine learning methods that are proved to be effective in facility LCC analysis.

Data_standard contains subclasses: *City_data_standard*, *Building_data_standard*, and

IoT_data_protocol. An instance of *City_data_standard* is CityGML, a city-level open data standard and exchange format to store digital 3D models of cities and landscapes (Open Geospatial Consortium 2018). Instances of *Building_data_standard* can be IFC and gbXML (gbXML Schema Inc. 2018), which are commonly used building data standard. Instances of *IoT_data_protocol* can be the data protocols of building automation and control, such as BACnet, Modbus, and Zigbee.

LCCA-Onto: definition of property

In OWL, a property is a characteristic of a class – “a directed binary relation that specifies some attribute which is true for instances of that class” (OWL Working Group 2019). Properties may have domains and ranges. The domain is the subject of a relation while the range is the object of that relation. There are two types of OWL property: Object property (*owl:ObjectProperty*) and Datatype property (*owl:DatatypeProperty*) (OWL Working Group 2019). The object property is used to represent relations between instances of two classes. Datatype property (owl: DatatypeProperty) is used to represent relations between object and data values, such as numerical values (e.g. integer, double and, float), boolean, and string. Object properties of the LCCA-Onto are described in Table 1.

Table 1. Definition of Object Properties in LCCA-Onto

Object property	Domain	Range	Description
contains	Community	Building	Community contains buildings
isEquippedWith	Building	Building_ system	Buildings are equipped with building systems
exports	Building_ system	Data_raw	Building systems export the raw data
isProcessedBy	Data_raw	MACH_LRN_to ol	The raw data is processed by machine learning tools
derives	MACH_LRN_to ol	Data_ processed	Machine learning tools are used to derive the processed data
predicts	Data_ processed	Data_ predicted	The processed data is used to predict the facility LCC and its components
describes	Data_standard	Building Community Data	Data standards are used to describe buildings, communities, and data.

LCCA-Onto: the overall framework

The LCCA-Onto is expandable to include any facility cost related building system, data standard, data, and machine learning tools. Figure 3 shows the overall framework of LCCA-Onto, which summarizes the main classes and their relationships (object properties). The LCCA-Onto provides a foundation for the machine learning-based facility LCC analysis domain knowledge developed in this research.

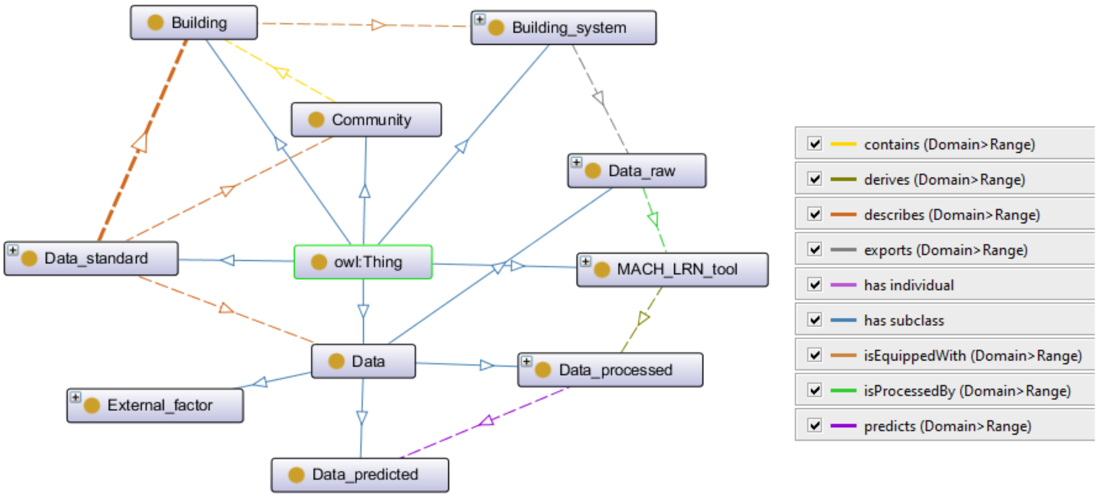

Figure 3. The overall framework of LCCA-Onto

CONCLUSION

This research investigates the approach of forecasting facilities' LCC by implementing machine learning on historical data provided by building systems. The authors propose a machine learning-enabled facility LCC analysis framework and the corresponding domain ontology – LCCA-Onto. The proposed framework minimizes human involvement to the greatest extent possible. People make mistakes – the more people involved in the data processing and analysis process, the higher is the risk exposure to human errors. In addition, some stakeholders tend to be very protective of the money-related data, which makes collecting historical data extremely difficult (Weerasinghe et al. 2016). This research bypassed some of the existing barriers in cost analysis and uses the data from building systems directly. The proposed transparent approach provides reliable insights into facility LCC patterns.

The authors are developing a facility LCC prediction system based on the proposed framework and ontology. Future research will be focusing on the automation of data collection, processing, and LCC analysis. After the system is developed, a usability assessment will be conducted to evaluate the system's usefulness.

REFERENCES

Amasyali, K., and El-Gohary, N. M. (2018). "A review of data-driven building energy consumption prediction studies." *Renewable & Sustainable Energy Reviews*, 81, 1192-1205. DOI: 10.1016/j.rser.2017.04.095.

An, S.-H., Kim, G.-H., and Kang, K.-I. (2007). "A case-based reasoning cost estimating model using experience by analytic hierarchy process." *Building and Environment*, 42(7), 2573-2579. DOI: doi.org/10.1016/j.buildenv.2006.06.007.

Au-Yong, C. P., Ali, A. S., and Ahmad, F. (2014). "Prediction cost maintenance model of office building based on condition-based maintenance." *Maintenance and Reliability*, - 16(- 2), - 324. DOI.

Bakis, N., Amaratunga, R., Kagioglou, M., and Aouad, G. (2003). "An integrated environment for life cycle costing in construction." DOI.

buildingSMART (2017). "IFC Overview summary." <http://www.buildingsmart-tech.org/specifications/ifc-overview>. (Internet, cited Sep 29th, 2019).

Charette, R. P., and Marshall, H. E. (1999). *UNIFORMAT II elemental classification for building specifications, cost estimating, and cost analysis*, US Department of Commerce, Technology Administration, National Institute of Standards and Technology.

Chou, J. S., and Ngo, N. T. (2016). "Time series analytics using sliding window metaheuristic optimization-based machine learning system for identifying building energy consumption patterns." *Applied Energy*, 177, 751-770. DOI: 10.1016/j.apenergy.2016.05.074.

Cole, R. J., and Sterner, E. (2000). "Reconciling theory and practice of life-cycle costing." *Building Research & Information*, 28(5-6), 368-375. DOI.

Construction Specifications Institute (CSI) (2016). "MasterFormat 2016 Edition: Numbers and Titles." <www.edmca.com/media/35207/masterformat-2016.pdf>. (Internet, cited Sep 29th, 2019).

Cristani, M., and Cuel, R. (2004) "A comprehensive guideline for building a domain ontology from scratch." *Proc., proceeding of" International Conference on Knowledge Management (I-KNOW'04)", Graz, Austria*.

Eastman, C., Teicholz, P., Sacks, R., and Liston, K. (2011). *BIM Handbook: A Guide to Building Information Modeling for Owners, Managers, Designers, Engineers and Contractors (2nd Edition)*, John Wiley & Sons.

Ferry, D. J., and Flanagan, R. (1991). *Life cycle costing: A radical approach*, Construction Industry Research and Information Association London.

Gao, X., and Pishdad-Bozorgi, P. (2019). "BIM-enabled Facilities Operation and Maintenance: A Review." *Advanced Engineering Informatics*, 39, 227–247. DOI: doi.org/10.1016/j.aei.2019.01.005.

Gao, X., Pishdad-Bozorgi, P., Shelden, D., and Hu, Y. (2019). "Machine Learning Applications in Facility Life-cycle Cost Analysis: A Review." *The 2019 ASCE International Conference on Computing in Civil Engineering*, ASCE, Atlanta, GA.

Gao, X., Pishdad-Bozorgi, P., Shelden, D., and Tang, S. (2019). "A Scalable Cyber-Physical System Data Acquisition Framework for the Smart Built Environment." *The 2019 ASCE International Conference on Computing in Civil Engineering*, ASCE, Atlanta, GA.

Gao, X., Tang, S., Pishdad-Bozorgi, P., and Shelden, D. (2018). "Foundational Research in Integrated Building Internet of Things (IoT) Data Standards."Center for the Development and Application of Internet of Things Technologies (CDAIT).Available from: https://cdait.gatech.edu/sites/default/files/georgia_tech_cdait_research_report_on_integrated_building_-_iot_data_standards_september_2018_final.pdf, (2018) (Internet, cited Sep 29th, 2019).

gbXML Schema Inc. (2018). "gbXML." <www.gbxml.org>. (Internet, cited Sep 29th, 2019).

Hong, T., Hyun, C., and Moon, H. (2011). "CBR-based cost prediction model-II of the design phase for multi-family housing projects." *Expert Systems with Applications*, 38(3), 2797-2808. DOI: doi.org/10.1016/j.eswa.2010.08.071.

Hu, Y., and Castro-Lacouture, D. (2018). "Clash Relevance Prediction Based on Machine Learning." *Journal of Computing in Civil Engineering*, 33(2), 04018060. DOI.

Jain, R. K., Smith, K. M., Culligan, P. J., and Taylor, J. E. (2014). "Forecasting energy consumption of multi-family residential buildings using support vector regression: Investigating the impact of temporal and spatial monitoring granularity on performance accuracy." *Applied Energy*, 123, 168-178. DOI: 10.1016/j.apenergy.2014.02.057.

Ji, C., Hong, T., Jeong, K., and Leigh, S. B. (2014). "A model for evaluating the environmental benefits of elementary school facilities." *Journal of Environmental Management*, 132, 220-

229. DOI: 10.1016/j.jenvman.2013.11.022.

Jin, R., Han, S., Hyun, C., and Cha, Y. (2016). "Application of Case-Based Reasoning for Estimating Preliminary Duration of Building Projects." *Journal of Construction Engineering and Management*, 142(2). DOI: 10.1061/(asce)co.1943-7862.0001072.

Kelleher, J. D., Namee, B. M., and D'Arcy, A. (2015). *Fundamentals of Machine Learning for Predictive Data Analytics: Algorithms, Worked Examples, and Case Studies*, MIT Press.

Koo, C. W., Hong, T., Hyun, C. T., Park, S. H., and Seo, J. o. (2010). "A study on the development of a cost model based on the owner's decision making at the early stages of a construction project." *International Journal of Strategic Property Management*, 14(2), 121-137. DOI: doi.org/10.3846/ijspm.2010.10.

Krstić, H., and Marenjak, S. (2017). "Maintenance and operation costs model for university buildings." *Technical Gazette*, - 24, - 200. DOI: 10.17559/TV-20140606093626.

Li, C. S., and Guo, S. J. (2012). "Development of a Cost Predicting Model for Maintenance of University Buildings." *Proceedings of the 2011 2nd International Congress on Computer Applications and Computational Science, Vol 1*, F. L. Gaol, and Q. V. Nguyen, eds., 215-221.

Mitchell, T. M. (1997). *Machine Learning*, McGraw-Hill.

Mocanu, E., Nguyen, P. H., Kling, W. L., and Gibescu, M. (2016). "Unsupervised energy prediction in a Smart Grid context using reinforcement cross-building transfer learning." *Energy and Buildings*, 116, 646-655. DOI: 10.1016/j.enbuild.2016.01.030.

Neely, E. S., and Neathammer, R. (1991). "Life-cycle maintenance costs by facility use." *Journal of Construction Engineering and Management-Asce*, 117(2), 310-320. DOI: 10.1061/(asce)0733-9364(1991)117:2(310).

Noy, N. F., and McGuinness, D. L. (2001). "Ontology development 101: A guide to creating your first ontology." Stanford knowledge systems laboratory technical report KSL-01-05 and ….Available from: http://www.corais.org/sites/default/files/ontology_development_101_aguide_to_creating_your_first_ontology.pdf, (2001) (Internet, cited Sep 29th, 2019).

Open Geospatial Consortium (2018). "CityGML." <https://www.citygml.org/>. (Internet, cited Sep 29th, 2019).

OWL Working Group (2019). "Web Ontology Language (OWL)." <www.w3.org/OWL>. (Internet, cited Sep 29th, 2019).

Park, B. R., Choi, E. J., Hong, J., Lee, J. H., and Moon, J. W. (2018). "Development of an energy cost prediction model for a VRF heating system." *Applied Thermal Engineering*, 140, 476-486. DOI: 10.1016/j.applthermaleng.2018.05.068.

Pishdad-Bozorgi, P., Gao, X., Eastman, C., and Self, A. P. (2018). "Planning and developing facility management-enabled building information model (FM-enabled BIM)." *Automation in Construction*, 87, 22-38. DOI: doi.org/10.1016/j.autcon.2017.12.004.

Sala-Cardoso, E., Delgado-Prieto, M., Kampouropoulos, K., and Romeral, L. (2018). "Activity-aware HVAC power demand forecasting." *Energy and Buildings*, 170, 15-24. DOI: 10.1016/j.enbuild.2018.03.087.

SAS (2018). "Machine Learning: What it is & why it matters." <https://www.sas.com/it_it/insights/analytics/machine-learning.html>. (Internet, cited Sep 29th, 2019).

Stanford Center for Biomedical Informatics Research (2019). "protégé." <protege.stanford.edu>. (Internet, cited Sep 29th, 2019).

Studer, R., Benjamins, V. R., and Fensel, D. (1998). "Knowledge engineering: principles and methods." *Data and knowledge engineering*, 25(1), 161-198. DOI.

Tu, K. J., and Huang, Y. W. (2013). "Predicting the operation and maintenance costs of condominium properties in the project planning phase: An artificial neural network approach." *International Journal of Civil Engineering*, 11(4A), 242-250. DOI.

Weerasinghe, A., Ramachandra, T., and Rotimi, J. O. B. (2016). "A Simplified model for predicting running cost of office buildings in Sri Lanka." *Proceedings of the 32nd Annual ARCOM Conference*, - 340.

Zhang, C., Cao, L. W., and Romagnoli, A. (2018). "On the feature engineering of building energy data mining." *Sustainable Cities and Society*, 39, 508-518. DOI: 10.1016/j.scs.2018.02.016.