

## APPLICATION OF MACHINE LEARNING APPROACH FOR LOGISTICS COST ESTIMATION IN PANELIZED CONSTRUCTION

SangJun Ahn\*, Mohammed Sadiq Altaf<sup>a</sup>, SangUk Han<sup>b</sup>, and Mohamed Al-Hussein<sup>c</sup>

\* University of Alberta, Department of Civil and Environmental Engineering, Canada  
sangjun@ualberta.ca

<sup>a</sup> University of Alberta, Department of Civil and Environmental Engineering, Canada  
msaltaf@ualberta.ca

<sup>b</sup> Hanyang University, Department of Civil and Environmental Engineering, South Korea  
sanguk@hanyang.ac.kr

<sup>c</sup> University of Alberta, Department of Civil and Environmental Engineering, Canada  
mohameda@ualberta.ca

**Abstract.** Logistics operations in panelized construction are vital daily tasks that connect the panel manufacturing facility to the job site. Although logistics operations are both important and prevalent in panelized construction, the cost of logistics has yet to be fully understood by either industry or academia due to the complicated relationship between multiple factors in logistics demands and operations. In practice, logistics is considered as an overhead cost that consists of various indirect or fixed costs in the panelized construction operation. As a result, logistics cost estimates are rendered inaccurate when subjected to project changes. Considering the number of construction projects over the course of a year, inaccurate logistics cost estimates are significant. Previous studies have shown that a machine learning approach could be used to predict costs that are influenced by multiple factors. To fill knowledge gaps in both research and practice, in this study machine learning based on historical logistics data is used to accurately predict logistics costs for a given project. The results from this study indicate that machine learning can be a reliable tool to predict logistics costs.

**Keywords:** Logistics cost, Machine learning, Panelized construction, SVM (Support Vector Machine)

### 1 INTRODUCTION

The demand for panelized construction in the residential sector has increased due to its cost and time advantages [1] over traditional stick-built construction; however, logistics costs have yet to be well understood despite the substantial influences on the overall cost of offsite construction projects [2,3,4]. Offsite construction utilizes prefabricated components (e.g., panel, module, or precast) that are manufactured in a controlled environment to minimize on-site construction activities. As a result, the offsite construction method often deals with heavy items that have large dimensions, and this directly affects the cost of delivery. If no load regulations exist, then the cost may not be affected by volume and weight, but strict road safety regulations exist in both Canada and the U.S., which limit the dimension and weight of a load in order to minimize any potential safety risks during a road delivery. Due to its heavy weight and large dimension, the delivery of precast modules must satisfy both weight and dimension regulations; as a result, a limited number of precast panels can be loaded onto transport trailers at one time [5,6]. Additionally, a large volume of a lightweight material such as plasterboard also reveals a high logistics cost, where up to 27% of the total logistics cost accounts for the cost of materials [7]. Panelized

construction, which is another type of offsite construction, prefabricates a panel in a factory using engineered woods. Panels can be considered as lightweight relative to precast concrete, but they require a large amount of space in the transport trailer during delivery due to their dimensions. In practice, frequent deliveries occur throughout daily logistics operations in the panelized construction process where drivers load panels from a factory onto transport trailers and then deliver them to multiple construction sites for on-site assembly. Considering the high volume of delivery cycles, each of which involves drivers, trucks, and trailers, the accumulated logistics cost in a single day of operation can be significant.

Previously, various methods have been introduced and researched to estimate accurate logistic costs in the construction industry. The proper estimation of logistics costs in the construction industry is imperative, because it provides several advantages, such as: (1) preparing a bid for a new construction project; (2) measuring potential savings from an improved logistics system; and (3) optimizing logistics management and decision making. The activity-based costing (ABC) method is one that has been used in several studies of construction logistics costs [8]. The ABC method identifies any construction activities that can potentially affect logistics cost and assigns estimated cost to each different activity in order to estimate a total logistics cost for a construction project. A simple logistics process can easily adopt the ABC method to estimate logistics cost, but logistics systems in panelized construction consist of multiple factors with complicated relationships in constantly changing environments (e.g., weather). Thus, estimating using ABC is not an ideal approach in panelized construction and previous study indicate that its application would be costly and time-consuming [9] due to the lack of a clear functional relationship between factors that are difficult to identify [10]. Therefore, a different approach has been introduced to overcome these issues, namely machine learning. The rapid developments of technologies in machine learning with low-cost computation [11] allow researchers and industries to forecast values by using an existing data set without having clear understanding of the relationships between multiple factors in the data.

This paper aims to provide an accurate forecast of logistics costs in panelized construction where previously there has been little understanding of cost. Three different machine learning regression methods are applied to estimate delivery demands for over 100 residential construction projects based on historical data, and total logistics costs are calculated using verified logistics unit costs and the forecasted demands using the proposed approaches. For the purpose of validation, the final results are compared with the actual logistics cost data from panelized construction projects in Canada.

## **2 RESEARCH BACKGROUND**

The issue of transparency in logistics costs for offsite construction is considered one of the major hurdles that needs to be resolved in order to increase market shares in the construction industry [2]. Housing market data in Canada indicates that 12.5% of all single detached houses were built using panelized construction in 2011 [1]. The market share has been consistently increasing by small margins, but further rapid adaptations of panelized construction are necessary to achieve a reduction in housing prices while improving the quality of residential buildings. Considering the substantial influences of logistics cost in panelized construction, an accurate understanding of the cost would improve the deployment of panelized construction. Previously, logistics cost estimation methods, such as ABC, have been used to estimate logistics costs, but the study by Stapleton et al. indicates the drawbacks of the difficulties to determine a clear functional relationship between multiple factors [9]. Newly introduced estimation approaches, such as machine learning, have shown reliable cost forecasting of logistics costs in the construction industry [10,11].

### **2.1 Logistics cost estimation issues in panelized construction**

Logistics costs often involve multiple elements. They can vary depending on the characteristics of logistics in different industries. In the construction industry, the logistics costs can be broadly divided into five elements: labour, equipment, materials, facilities, and capital [12,13]. Labour could include any

person who participates in logistics activities. Truck drivers, workers in loading areas, and logistics managers are typical examples of this. Typical logistics equipment includes trucks, trailers, cranes, and forklifts. For the materials, any tools utilized for packaging and load holding are common items. Next, an example of the facilities are storage areas for goods and equipment. Finally, the capital can include any opportunity costs of unutilized equipment and goods in inventory. These are typical examples in the construction industry and may vary in other industries.

The logistics cost estimation in the ABC method [8] breaks down the logistics-related activities into detailed process levels, and each process includes multiple logistics elements. For example, a delivery process typically includes four different elements, such as driver salary, truck ownership, operation cost, and packaging costs, and each element has a unit cost which includes all involved expenditures (e.g., fixed cost, overhead, or variable cost) of activities. In logistics, unit costs are often expressed as cost per hour or cost per distance.

In total logistics cost, transportation processes often comprise the largest portion and have been a popular topic of study [8] due to its significant influence on the total cost. To calculate the costs, three different unit costs are applied with logistics demands (e.g., operation hours and distances) to determine the final transportation cost [14]. The unit costs include ownership costs per hour, operation costs per hour, and maintenance cost per hour, and the logistics demands are total operation hours and distances. The unit costs of ownership, operation, and maintenance can be calculated using historical logistics data; however, the total operation time and distance may not be accurately estimated from the historical data due to unique project environment conditions. Therefore, it is important to ascertain reliable and accurate logistics demand estimation, which has not been possible through the ABC method.

As indicated by previous research, complicated logistics systems are more costly and time-consuming when applying ABC [9], and more importantly, logistics demands may not be reliable, thus leading to inaccurate estimation. Panelized construction comprises complex logistics processes involving travel between the factory and various sites using multiple trucks, drivers, and trailers. In current practice, multiple construction projects are carried out concurrently in various locations and each project has unique design features that have an impact on both loading capacity as well as loading and unloading times. Thus, the historical data may not provide accurate information due to large deviations, such as truck demands, and the previous approach, ABC, is an inefficient method to apply. As a result, one of the leading panelized construction companies in Canada considers the logistics cost as a fixed cost or as overhead where a fixed percentage of the total house price is estimated as the logistics cost. However, this approach creates cost errors between estimated and actual costs due to a large variation of logistics demands among projects because of differences in design and location. Considering the large number of construction projects in a given year, the accumulated errors in logistics cost estimation are significant. Therefore, a different approach is necessary to ascertain accurate forecasting of logistics costs.

## **2.2 Machine learning approach for logistics cost forecasting**

Machine learning is considered one of the fastest-growing technologies and has been expanding its application to broad areas such as health care, manufacturing, education, finance, policing, and marketing [11]. Any industries involving data-intensive issues, such as the logistics industry (e.g., DHL or UPS), have been actively utilizing this technology in order to gain understanding of a large amount of data and improve decision-making processes.

The machine learning techniques can be classified into two types: (1) supervised learning and (2) unsupervised learning. Supervised learning is applied when a user has both known output and input so that the trained model can predict outputs. On the other hand, unsupervised learning is employed when output data is unknown. Once the type of learning technique is determined based on the available data, an algorithm is selected to obtain optimum results. On a broad level, supervised learning utilizes two different algorithms, such as classification and regression, and unsupervised learning employs only clustering algorithms. Each type of algorithm offers various techniques that can be applied depending on the type of data and the goals to be met. For example, regression encompasses multiple methods,

including linear regression, support vector machine (SVM), ensemble methods, decision trees, and neural networks (NNs), among others.

Previously, the machine learning approach has been applied to forecast logistics cost in the construction industry to overcome difficulties in traditional forecasting methods such as ABC. Various studies describe the logistics system as an open complex system whose demands are non-linear and random [10,16]. Due to the nature of logistics systems, applying ABC to forecast logistics cost is time-consuming and costly. Furthermore, ABC is unable to accurately predict the demands of logistics. The same studies apply two different regression methods to predict cost, SVM and NN. SVM is based on the statistical learning theory, which is adapted from the structural risk minimization principle [17], and it is applied for nonlinear and multidimensional data to forecast output [18]. The NN is a highly nonlinear dynamic system that has an ability for self-learning [19]. The results of the machine learning application in logistics cost indicate that both regression models offer reliable results to predict the logistics cost, but the SVM presents a stronger predictive ability between the two, and, the NN reveals overfitting issues [10,16].

In this study, three different SVM methods, linear, quadratic, and cubic, are applied to determine an accurate prediction model. Logistics demands, such as the frequency of visits to a site and the duration of stay on a site, are forecasted using these three methods. Rather than directly estimating the total logistics cost, a verified unit cost for each activity is used with the forecasted values in order to compare with actual cost data.

3 RESEARCH METHOD

3.1 Data collection

GPS location data, such as longitude and latitude of transport trucks, are collected for a 6-month period by means of a GPS location transmitter installed in each truck. Using geocoding software, the data are analysed to determine the number of visits per project site, including the duration for unloading processes on site. Details regarding the GPS data analysis is outside the scope of this paper and thus will not be discussed. The collected data are matched with the logistics delivery schedule that includes the project name; floor area; wall area; availability of deck, stair, and veranda; roof built type; and house type. Table 1 presents an example of the collected data. A total of 10 attributes are used for training the SVM regression methods. The first eight attributes are selected as input data and the remaining two (e.g., duration of stay and number of visits) are used for outputs. The duration of stay is defined as the total unloading time for a truck at a project site, and the number of visits is defined as the frequency of visits to a site. Intuitively, a large house would require a longer duration of stay and greater number of visits.

Table 1. Sample data.

Inputs								Outputs	
Project Name	Floor Area (ft <sup>2</sup> )	Wall Area (ft <sup>2</sup> )	Deck	Stair	Veranda	Roof	House Type	Duration of Stay (hr)	Number of Visits
Rosenthal	811	3,936	0	1	0	Shop-built	RPL	9.73	8
DesRochers	2,126	3,936	0	1	0	Site-Built	RSL	10.54	11
...	...	...	...	...	...	...	...	...	...

3.2 Estimation of truck demands and duration of stay

The SVM regression method is applied to forecast the number of visits and on-site duration of stay of trucks for each project. Once the input data set is prepared, as presented in Table 1, the regression model is trained by using 10-folds cross-validation. This statistical technique is widely used in machine learning validation [20]. The 10-folds indicates that the data set is divided into 10 equal portions, then a single portion is used for testing while the others are used for training. Upon successful completion of training, new logistics operations data are entered into the trained model to determine the accuracy of forecasting durations and number of visits. Figure 1 represents user interfaces of the machine learning software.

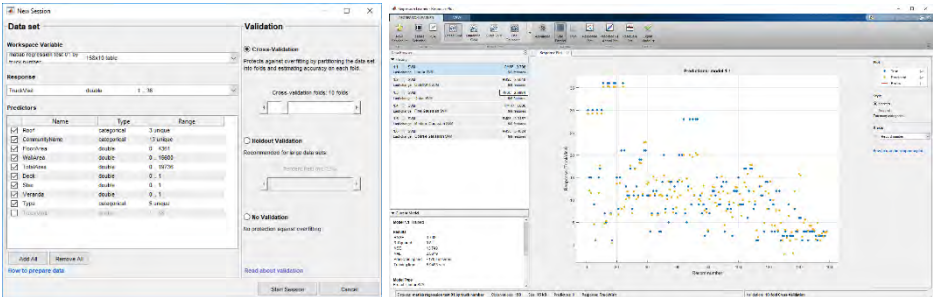


Figure 1. User interfaces of machine learning software.

3.3 Calculation of logistics cost

The final logistics cost is calculated using Equation (1). The number of site visits ( $N$ ) and the duration of stay ( $T$ ) are the only two parameters that are estimated by the SVM regressions. Distances ( $D$ ) are determined by measuring the distance between a factory and a site using Google maps. Drivers occasionally take different routes in the case of traffic or weather conditions, but the historical data indicates that deviations are negligible. The two unit costs are also used to calculate the logistics cost and they consider two different operation conditions: (1) moving and (2) idling. Although trucks are not moving when idle, this condition still influences fuel consumption, driver salary, maintenance cost, and truck depreciation. Therefore, unit idling cost is used for accurate calculation of the logistics cost.

$$C_T = \sum(N \times D \times C_{op}) + \sum(T \times C_{id})$$

(1)

Table 2. Description of parameters of Equation (1).

	Description
$C_T$	Total logistics cost, in CAD
$N$	Forecasted number of truck visits, in each
$D$	Distance between a factory to a site, in km
$C_{op}$	Unit operation cost, in CAD/km
$T$	Forecasted duration of stay, in hr
$C_{id}$	Unit idling cost, in CAD/hr

4 RESULTS AND DISCUSSION

4.1 Truck demand and duration of stay estimation

The three SVM methods are trained to compare training performances. Linear, quadratic, and cubic SVM regressions are the selected models. The results indicate that the cubic models in both truck visits and duration of stay feature the least root-mean-square error (RMSE) and the highest  $R^2$ , as presented in Table 3. These results are expected due to the nonlinear characteristics in logistics data as described in the previous studies [10,16].

Table 3. Results of the trained models (10-folds).

Results	Number of Truck Visits			Duration of Stay (hr)		
	SVM Linear	SVM Quadratic	SVM Cubic	SVM Linear	SVM Quadratic	SVM Cubic
RMSE	3.7111	3.5019	2.9087	3.0322	2.9729	2.9133
$R^2$	0.81	0.83	0.89	0.83	0.84	0.84

To validate the accuracy of the trained models, actual logistics operation data are entered and the outcomes are presented in Figure 2. As expected, both cubic SVM models result in the best accuracy among the models. In truck visit forecasting, the linear SVM results in the lowest prediction accuracy with 2.29 times mean error while the cubic model results in 1.15 times. In the duration of stay, the linear model also indicates the highest error with 1.75 hr and the cubic reveals the lowest with 0.88 hr.

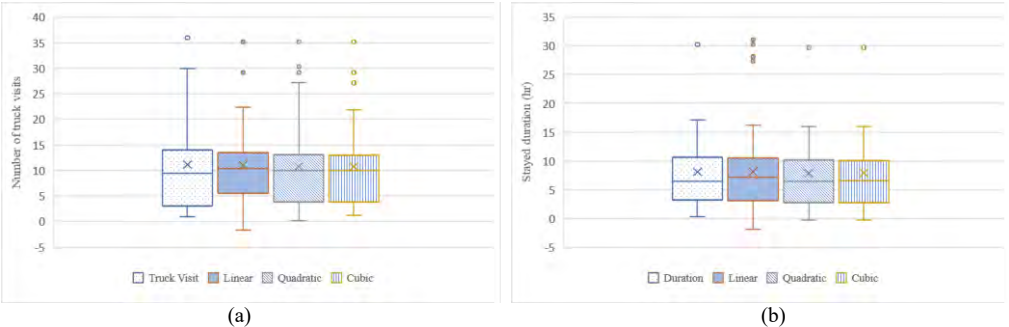


Figure 2. Validation results using sample data: (a) truck visits; (b) duration of stay

4.2 Logistics cost comparison

The three different logistics costs resulting from the three SVM methods (linear, quadratic, and cubic) of over 100 projects are compared to determine the performance of the machine learning approach. Actual, estimated, and machine learning forecasted logistics costs are selected. The actual cost represents the actual expenditure for logistics costs, and the estimated costs are costs used for estimating purposes by the company’s estimation department. In Figure 3, four different comparisons are presented for detailed analysis of the logistics costs. Overall, the machine learning forecasting reveals better outputs than the estimation values. The estimation values have a tendency to underestimate by an average of  $-11.4\%$  than the actual costs, and the machine learning values also tend to underestimate by  $-1.21\%$ . In comparison among various house sizes, the estimation values are overestimated in small houses and underestimate in large houses, but the forecasting values by the machine learning maintain a similar trend to that of the estimation, but with far less deviation from the actual costs. In travel hours and distance comparison, both the estimation and machine learning values feature accurate predictions.

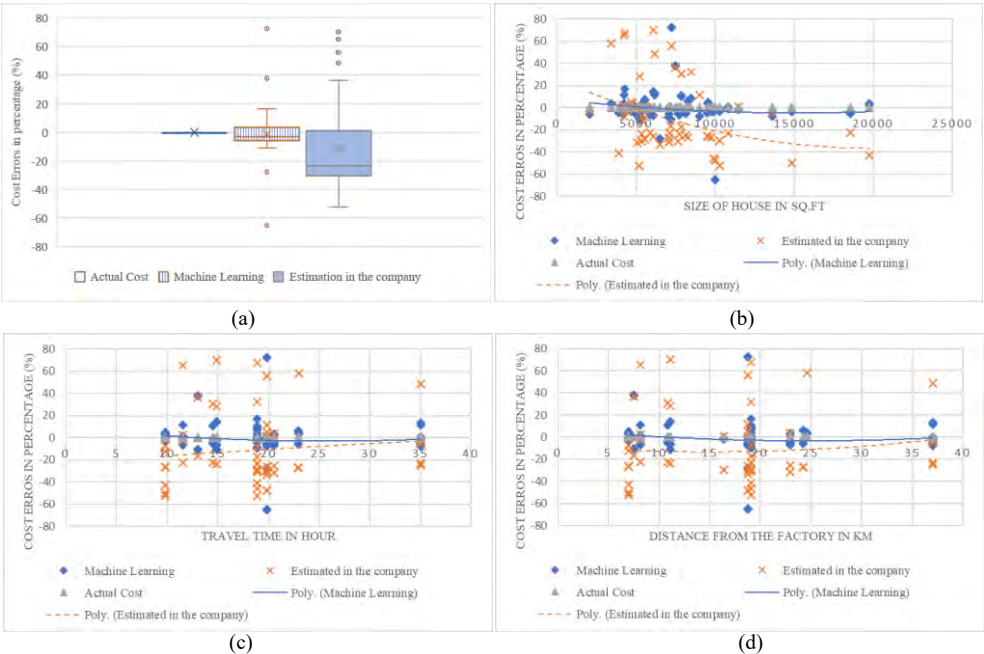


Figure 3. Results of cost comparison: (a) distribution of logistics cost errors; (b) logistics cost error vs. size of house; (c) logistics cost error vs. travel time; and (d) logistics cost error vs. distance)

In this paper, the machine learning approach is used to forecast logistics demands (e.g., number of trucks and duration of stay on site) in the panelized construction industry. The results indicate that the SVM regression methods can predict the demands with reliable accuracy. Unlike the ABC logistics cost estimating method, the machine learning method is able to provide accurate results without having an extensive understanding among multiple factors in logistics operations. Also, this approach can save a significant amount of time and labour compared to that which is required to perform ABC. The entire machine learning approach process can be carried out by a single person in a relatively short amount of time, thus, if accurate logistics data is available, the machine learning approach is more efficient.

This study explores the SVM regression models, but there may be a more suitable model. For future research, it is recommended to examine various regression models in the supervised machine learning approach. And, the current results are based on 3 months of data and over 100 construction projects, but a larger amount of data would provide more accurate results.

5 CONCLUSIONS

This study proposes a machine learning approach to forecast logistics costs in the panelized construction process. Collected logistics data from panel delivery trucks and logistics schedules are prepared to use as inputs for the SVM regression models. The outputs from the models are compared with actual and estimation costs from a panelized construction company for validation purposes. The results indicate that the machine learning approach provides more accurate results than the estimation costs without having in-depth understanding of multiple factors as in the ABC. Also, in comparison among distance, travel time, and house size, the machine learning approach outperforms the estimations overall.

Thus, the machine learning approach is recommended in forecasting logistics costs that are considered complex and difficult to determine reliable results.

## REFERENCES

- [1] CMHC, "Profile and Prospects of the Factory Built Housing Industry in Canada", *Canada Mortgage and Housing Corp*, 2006.
- [2] MBI, "Challenges and Opportunities for Modular Construction in Canada", *Workshop Summary Repot*, 2015.
- [3] Browne, M., "Chapter - 01: The challenge of construction logistics." *Supply Chain Management and Logistics in Construction*, KoganPage, 2015.
- [4] Franklin, M., "Chapter - 02: Aggregating global products for just-in-time delivery to construction sites." *Supply Chain Management and Logistics in Construction*, KoganPage, 2015.
- [5] Arditi D, Ergin U, Gunhan S., "Factors Affecting the use of Precast Concrete Systems", *Journal of Architectural Engineering*, 6(3), 2000.
- [6] Polat G., "Factors affecting the use of precast concrete systems in the United States.", *Journal of Construction Engineering and Management*, 134(3):169-178, 2008.
- [7] Wegelius-Lehtonen, T., "Performance measurement in construction logistics", *International Journal of Production Economics*, 69:107-116, 2001.
- [8] Fang Y, Ng ST., "Applying activity based costing approach for construction logistics cost analysis", *Construction Innovation*, 11:259-281, 2011
- [9] Stapleton D, Pati S, Beach E, Julmanichoti P., "Activity based costing for logistics and marketing", *Business Process Management Journal*, 10.5:584-597, 2004.
- [10] Tian J, Gao M., "The Research of Building Logistics Cost Forecast Based on Regression Support Vector Machine", *International Conference on Computational Intelligence and Security*, 1:648-652, 2009.
- [11] Jordan MI, Mitchell TM., "Machine leering: Trends, perspectives and prospects.", *Science*, 349(6245):255-260, 2015.
- [12] Kaplan RS, Copper R., "cost and effect", *Harvard Business School Press*, 1998.
- [13] Brimson JA., "Activity Accounting: An Activity based Costing Approach", *Wiley*, 1991.
- [14] Peurifoy RL, Ledbetter WB., "Construction Planning Equipment and Methods", *McGraw-Hill Inc.*, 2006.
- [15] Mathworks, "Machine Learning with MATLAB." *MathWorks*. Accessed 09 24, 2017. [https://www.mathworks.com/campaigns/products/offer/machine-learning-with-matlab-conf.html?elqsid=1506277080587&potential\\_use=Student](https://www.mathworks.com/campaigns/products/offer/machine-learning-with-matlab-conf.html?elqsid=1506277080587&potential_use=Student).
- [16] Tian J, Gao M., "Building Logistics Cost Forecast Based on Improved Simulated Annealing Neural Network", *Intelligent Computation Technology and Automation*, 3:914-917, 2009.
- [17] Weston J, Watkins C., "Multi-class support vector machines", *Department of Computer Science, Royal Holloway, University of London*, 1998.
- [18] Cristianini NM, Shawe-Taylor J., "An introduction to support vector machines and other kernel-based learning methods", *Cambridge university press*, 2000.
- [19] Kosko B., "Neural networks and fuzzy systems: a dynamical systems approach to machine intelligence/book and disk.", *Vol. 1 Prentice hall*, 1992
- [20] Refaailzadeh P, Tang L, Liu H., "Cross-validation.", *Encyclopedia of database systems*, 532-538, Springer US, 2009.