The following publication Wang, R., Asghari, V., Cheung, C. M., Hsu, S. C., & Lee, C. J. (2022). Assessing effects of economic factors on construction cost estimation using deep neural networks. Automation in Construction, 134, 104080 is available at https://doi.org/10.1016/j.autcon.2021.104080.

Investigating Effects of Economic Factors on Construction Cost Estimation Using Deep Neural Networks

Abstract: There is no consensus on the effects of the economic factors on construction cost estimation and little attention has been paid to incorporating the trend of economic factors into cost estimation. To explore those effects quantitatively, this study uses deep neural networks (DNN) as an estimator and SHapley Additive exPlanations (SHAP) as a model interpreter, adopting the data on 98 public school projects in Hong Kong. The analysis is also verified by a comparison analysis using several machine learning models popular in construction cost estimation. The results indicate that the economic factors do play an important role in reducing the construction cost estimation errors and are even more important than projects' characteristics. The findings would be helpful for stakeholders in the field of construction engineering and management to make appropriate decisions and for researchers to unveil the actual degree of the effects of other influential factors on construction cost estimation.

Keywords: external economic factors; construction cost estimation; public school projects; deep neural networks; SHapley Additive exPlanations

1. Introduction

A successful construction project needs to be delivered within three interrelated constraints, namely budget, time, and quality. Quality of projects can be monitored and improved during the construction phase whereas budget and time need to meet their agreed-upon estimations in the contracts. Both contractors and stakeholders are immensely engaged with these estimations [1]. Accurate cost estimation enables stakeholders and decision-makers to conduct more rational feasibility studies before commencement, determine the financial scale at the bidding stage, control and monitor cash flows during the construction phase of projects [2]. A project with an underestimated cost typically results in cost overrun and consequent financial losses for the stakeholders and/or contractors [3]. To mitigate the losses and to meet project profitability objectives, several approaches have been in practice and have been proposed in the literature to accurately estimate the construction costs.

Qualitative analysis and quantitative analyses are the main approaches for costs estimation. Qualitative approaches based on expert judgments could be simply biased and lead to inaccurate estimations [4]. Consequently, a growing body of literature has used traditional statistical methods (i.e., regression analysis [5,6]) and machine learning (ML) methods (e.g., support vector machine (SVM) [7,8], decision tree (DT) [9], and random forest (RF) [10,11]). But data is sometimes even more important than the methodology. In most cases, the development of construction cost estimation models merely relies on the projects' characteristics, for instance, project location, and project duration. However, external economic factors, in comparison to projects' characteristics, can influence the construction costs [12].

The literature analysis on the influence extent of economic factors in construction costs estimation reveals that there is no consensus among them. Some studies [13–15] insisted that external economic factors, such as inflation rate and consumer price index, exert less influence than projects' characteristics on the costs estimation. But not all studies share the same conclusion. Other studies [12,16,17] took an opposite view that economic factors such as price fluctuations are amongst the top influential factors. More importantly, past studies in this domain paid little attention to incorporating the trend and value of market indices, such as stock market index, construction indices, and daily wages, in construction cost estimation. This may be a reason behind the considerable deviation between the actual final cost and the original estimates for many construction projects even though advanced estimation approaches have been adopted by stakeholders and contractors [18]. All in all, there is a general lack of research in quantitative analysis of the influence and relative importance of economic factors in early costs estimation.

To fill the mentioned gap, this study aims to quantitatively examine the effects of external economic factors on early construction cost estimation using machine learning models. Although the novelty of machine learning approaches has been emphasized excessively in past studies, the role of other factors containing valuable information for estimation has been overlooked. Using deep neural networks (DNN) as the estimator and SHapley Additive exPlanations (SHAP) for explaining the trained model, this study analyzed the role of consumer price index (inflation rate), lending rate, building works tender price index, average daily wages, housing authority tender price index, and stock market index on early construction cost estimation of public-school projects in Hong Kong. Our results can be summarized as follows. Not only these factors and their trends can considerably reduce the estimation errors, but also, they can play a more important role than projects' characteristics in this regard. The main contribution of this study to the body of knowledge of construction engineering and management is the confirmation of the positive effect of economic factors in reducing early costs estimation errors through a quantitative approach. Thus, stakeholders, bidders, and contractors can make more informed decisions about initiating, bidding, and managing projects. During the construction phase, by monitoring the external economic factors, contractors and stakeholders could adjust the estimated construction cost in appropriate time, compare the adjusted estimations and the planned cash flows, and take necessary precautions to mitigate the possible shortage of funds. Our findings can be equally helpful to researchers in the field of construction cost estimation. With consideration of external economic factors, they can revisit the effects of the influential factors and uncover their actual degree of the effects on construction cost estimations.

The rest of this paper is organized as follows. The next section reviews the previous studies about the effects of external economic factors as well as the existing construction cost estimation methods. Then, Section 3 describes the dataset used in the present study. Next, Section 4 introduces the two methods DNN and SHAP as well as the metrics for evaluating the performance of the established construction cost estimation models. This is followed by the data pre-processing and feature engineering and the model development of DNN and several popular techniques selected as benchmarks, constituting Section 5. Section 6 provides the results and its discussion before the conclusion of the whole paper in Section 7.

2. Literature Review

This section, first, reviews the discussion of the controversial effects of external economic factors on early construction cost estimation as well as the methods used to explore that effects in prior research. Then the construction cost estimation methods proposed in previous literature are analyzed.

2.1 Effects of External Economic Factors

There is no consensus on the effects of external economic factors on construction cost estimation. On one hand, some researchers insisted that economic factors have limited impacts on the construction cost estimation. Based on a questionnaire survey, Elhag et al. [13] found that external factors and market conditions ranked the fourth, among six categories of cost-influencing factors. Similarly, Cheng [14] revealed that economic environmental and circumstantial factors pose the least influences on project cost. Among 27 factors, Hatamleh et al. [19] stated that market and economic conditions rank 21 and particularly, consultants have a notion that market and economic conditions have the least effect. On the other hand, economic factors have been reported to be disregarded. Aklnci and Flscher [20] described that economic factors have a relatively high correlation with the final cost. Similarly, Shane et al. [21] categorized 18 factors and verified that market conditions affect the construction cost, particularly for large projects. Then, a number of studies [17,22–24] revealed that market conditions are critical influential factors. For example, Zhao et al. [25] found that factors about market and industry conditions are the most significant among 30 influential factors. This controversy over the effects of economic factors on construction cost estimation causes researchers to hesitate to capture economic factors when estimating construction cost. More importantly, the trend and value of market indices, such as stock market index, construction indices, and daily wages, have less been incorporated into construction cost estimation. This may be an explanation of the considerable deviation between the actual final cost and the original estimates for many construction projects.

The above-mentioned inconsistent effects of external economic factors may be due to the lack of appropriate methodology [17]. To explore the effects of economic factors and other cost-influencing factors, many studies [14,15,19,26] have adopted some indexes, consisting of relative importance index, frequency cost adjusted importance index, Spearman's rank correlation, and other indexes. Based on these indexes, the important cost-influencing factors could be extracted. Besides, multiple linear regression [1,6], factor analysis [23], and structural equation modeling [17] have been used to examine the effects of influential factors on cost estimates. However, those indexes and techniques could be biased and subjective because the data is mainly collected from questionnaires. The traditional statistical methods assume the influential factors are independent of each other, irrespective of the interactive effects of factors. This may lead to underestimating the overall effects of the cost-influencing factors and thus missing some critical influential factors, such as economic factors.

To objectively test the effects of economic factors on construction cost estimation, this study embraced SHapley Additive exPlanations (SHAP), a recently developed method by

Lundberg and Lee [27]. This method could make use of the objective data of practical projects. More importantly, it is able to handle complex nonlinear interactions among input variables [28]. Due to its advantages, SHAP has been used in some studies. Mangalathu et al. [29] García and Aznarte [30] employed SHAP to rank the importance of input variables. Therefore, this study would use SHAP to explore the effects and relative importance of economic factors in a trained DNN model for the purpose of construction costs estimation.

2.2 Construction Cost Estimation Methods

Traditionally, construction cost estimation is a process that highly relies on experts' experience [13,19]. Some large agencies or firms may have extensive facilities and sufficient information on past construction projects to construct an in-house cost database. Many others who lack expert knowledge or facilities may have to use a published construction cost index provided by commercial suppliers such as RSMeans [24] or governmental construction cost information. For example, Hong Kong Project Strategy and Governance Office would release construction expenditure forecasts and other information regularly [31]. Most of those databases or information consider past prices only while ignoring the trend and value of stock market index and other market indices. Then cost estimates are developed and adjusted based on experts' experience, which may lead to further low accurate results.

To enhance the accuracy and efficiency of cost estimation, scholars have adopted statistical and machine learning techniques. Statistical and machine learning techniques may facilitate decision-making by converting historical data into decision support systems. This can deal with the problem of insufficient information necessary for accurate estimation at an early stage of a project. Al-Momani [5] constructed a linear regression (LR) model for construction cost prediction using three project characteristics as explanatory variables. Erdis [32] adopted DT, SVM, and artificial neural networks to estimate the construction cost deviations using the data of 575 Turkey public construction projects. The input variable included the location and duration of the projects and the rate of price-cut. Fang et al. [8] developed a fusion method, which fused the Kalman filter with least-squares SVM and LR to predict the construction cost of projects in China. The input variables included building acreage, application, city level, and other features about the projects. Chakraborty et al. [11] used LR, RF, and other three methods for construction cost prediction considering structural assembly type, superimposed load, tributary area, the unit cost of concrete, and unit cost of formwork. Huang and Hsieh [10] combined RF and LR to improve the prediction accuracy on BIM labor cost in the construction phase. The input variable is composed of floor area, the number of stories, and the apartment complex.

As for estimation methods, statistical analysis especially regression analysis has been a traditional option [5,6] because it is intuitive and easily understood, which generates simple and concise predictions. However, it has an inherent disadvantage as it requires a defined mathematical form, and it is not appropriate for a dataset with the presence of high nonlinearity. As a result, the prediction power of traditional statistical methods is limited. With no requirement for specifying mathematical forms, machine learning tools could capture complex relationships between input and output. Among machine learning algorithms, SVM, DT, and RF are popular methods to do cost estimation in construction. SVM converts the raw data into a space with high dimensions by nonlinear mapping and then divides the data with a hyperplane. DT identifies an informative tree structure in a dataset by a recursive partitioning process and is easy-to-use and has no requirement for prior knowledge [33]. RF is an extension of DT, combing small trees into an ensemble via bagging or aggregating [34]. SVM, DT, and RF also have the problem of overfitting for regression problems [35–38].

DNN is a branch of deep learning. It can deal with datasets with high nonlinearity between outcome and predictors, and thus offer more accurate results. In addition, DNN is capable of resolving the overfitting issue via a number of methods. Mutis et al. [39] developed a DNN-based method to identify human activities and to estimate real-time space occupancy for indoor air quality control. Kassem et al. [40] proposed a DNN model to investigate the productivity of excavators in infrastructure projects. Though DNN has been applied to prevalent construction challenges, its application in construction cost estimation can be further explored [41].

3. Construction cost dataset: The case of public schools in Hong Kong

This study took public school projects in Hong Kong as a case study and employed their data. The reason behind this choice is three-fold. First, the overall budget of the Hong Kong government is tightening [42]. This exerts great pressure on the Government for optimally investing capital expenditures. Since public-school projects are authorized by the Hong Kong government, accurate cost estimates would be beneficial from all perspectives for the decision-making bodies. Second, we aimed to highlight the effect of the economic factors on rather smaller yet complex datasets. Public school projects have become more complex during the years [43,44], but there are not many similar projects built in countries. As a result, the construction of public schools may be more expensive than other private/public-owned projects underlining the importance of early construction costs estimation. Third, a single type of project is targeted mainly because project type could exert influences on the quality of cost estimate [2,45].

The collected data contain the information of 98 public school projects built in the Hong Kong Special Administration Region (HKSAR). All of these projects were handled by the Architectural Services Department (ArchSD) and endorsed by Public Works Subcommittee (PWSC) in Legislative Council Session 1999 to 2019. Among the 98 projects, 9 are special schools and the others are primary or secondary schools. The main scope of these projects was to build classrooms, assembly halls, and ancillary facilities. Special rooms were also built in 26 projects. Besides, 61 projects had some special requirements or features, such as high bedrock level, noise mitigation measures, and close proximity to slope.

The data includes 1 output variable and 13 input variables, summarized in Table 1. The output variable is the estimated final construction cost. Among the 13 input variables, 7 are related to the project characteristics selected mainly based on previous studies about construction cost estimation [6,46]. The other 6 variables are about the external economic factors. They are selected to show two levels of external economic environments, industry- and market-level. Industry-level economic factors have been reported to be associated with construction cost estimates via questionnaires [25] and are represented by building works tender price index (BWTPI), Housing Authority Tender Price Index for New Building Works (HATPI), and Average daily wages (ADW). The first two indexes are compiled by the Hong Kong government to indicate the tender price level for building works while the last one reflects the labor price level in the construction industry. Market-level economic factors mainly include interest rate on bank loans, stock market index, and inflation. They have been verified to affect the accuracy of construction cost estimates but merely via qualitative approaches such as expert judgments and questionnaires [12,14,15,20]. Interest rate on bank loans is operationalized as the best lending rate following Williams [47]. Stock market index is indicated as the Hang Seng index (HSI) when the market closed following Cao et al. [48]. Inflation is measured as the consumer price index (CPI) following Hwang [49].

Table 1 Summary of output and input variables

Var.	**Variable Name**	**Project Character**	**Description**
Y	Cost (m HK$)	Yes	Final estimated construction cost
X1	Special School	Yes	The type of project, a dummy variable indicating whether the school is for special education, 1=yes, 0=no
X2	Classrooms	Yes	Number of classrooms
X3	Special Rooms	Yes	Number of special rooms
X4	Location	Yes	The project locations based on their distance from mainland China and there are three categories, A, B, and C region
X5	CFA (m^{2})	Yes	Construction floor area
X6	Duration (year)	Yes	Duration of the overall construction works
X7	Special Feature	Yes	Special project features, 1=special features would increase the construction cost, 0= special features do not have any effect on construction cost, -1= special features would decrease the cost
X8	Lending Rate	No	Best lending rate
X9	BWTPI	No	Building works tender price index (BWTPI), which is quarterly compiled by the ArchSD to aid to adjust building cost data for the purposes of estimating. It also indicates the level of tender prices for new building works but not for building services works. It is calculated in a similar way that adopted by the Royal Institution of Chartered Surveyors' Building Cost Information Service in the United Kingdom.
X10	HSI-Close	No	Hang Seng index (HSI) when the market closed
X11	HATPI	No	Housing Authority Tender Price Index for New Building Works (HATPI), which is quarterly compiled to serve as an indication of the tender price level for new building works that are returned in that quarter and accepted by the housing authority. This index does not reveal the tender price level for building services works, site formation, piling, and alteration or fitting-out. It is calculated with the base index set at 100 for the 1st quarter in 1970. The accuracy relies on the number of tenders available in each quarter.
X12	CPI	No	Consumer price index (CPI), which is useful for analyzing the inflation that affects consumers
X13	ADW(HKD/day)	No	Average daily wages (ADW) workers engaged in public sector construction projects as reported by main contractors. It is issued by the Census and Statistics Department (C&SD).

4. Methods

4.1 Estimator: Deep Neural Networks

A typical artificial neural network model usually consists of three parts, i.e., one input layer, one hidden layer, and one output layer. DNN could be more complex. It could consist of several hidden layers. The number of hidden layers determines the depth of the architecture. Each hidden layer has an activation function and several nodes (i.e., neurons). Usually, within a layer, one activation function is used for all neurons. The activation functions in different layers may be different.

According to Bengio [50], starting with the input $\boldsymbol{x} = \boldsymbol{h}^0$, the output vector $\boldsymbol{h}^k$ of layer k is computed using the output $\boldsymbol{h}^{k-1}$ of the previous layer $k-1$, shown as Equation (1).

$$\boldsymbol{h}^k = \sigma^k(\boldsymbol{b}^k + \boldsymbol{W}^k \boldsymbol{h}^{k-1}) \tag{1}$$

where σ^k denotes the activation function, $\boldsymbol{b}^k$ denotes a vector of bias values, and $\boldsymbol{W}^k$ denotes a matrix of weights. The output of the final layer $\boldsymbol{h}^l$ is used to predict the target variable y.

Activation functions facilitate mapping the non-linearity relations between input and output. There are several types of activation functions, such as sigmoid, tanh, relu, and linear. This study uses tanh activation function for the input layer. Relu activation function is selected for the hidden layer because not all the neurons would be activated all the time and thus it facilitates model training. Linear activation function is used for the output layer because the ultimate goal of this study is to predict the construction cost as a regression problem. These choices were inspired by previous studies (e.g., in [51,52]) The formulas of these three activation functions are as follows:

tanh: $$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \tag{2}$$

relu: $$y(x) = \max(o, x) \tag{3}$$

linear: $$y(x) = x \tag{4}$$

The values of the bias vector are set to be constant (usually 1) and the weights in all layers are initialized randomly at the beginning of training. Then the weights will be updated to minimize the dissimilarity between the predicted and the actual output variables by optimization algorithms. This study uses Adam optimizer because its performance is slightly better than others such as Adagrad, RMSprop, and SGDNesterov [53]. DNN models are well-known algorithms and further details regarding their structures and training are beyond the scope of this study. For further details, please refer to [54].

4.2 Model Interpreter: SHapley Additive exPlanations

Although DNN models are commonly believed to be black boxes, some researchers [27,55] have developed rather novel techniques for interpreting the model behavior. These techniques fall into the class of additive feature attribution methods. The additive feature attribution methods have an explanation model, defined as the interpretable approximation of the original model. Particularly, Lundberg and Lee's work [27] has gained massive attention during the past few years. Their proposed method, SHapley Additive exPlanations (SHAP), incorporate Game Theory fundamentals to interpret machine learning models' output. The explanation model g is a linear function of binary variables:

$$g(z') = \emptyset_0 + \sum_{i=1}^{M} \emptyset_i z'_i \tag{5}$$

where $z' \in \{0,1\}^M$, M is the number of simplified input features, and $\emptyset_i \in \mathbb{R}$. z'_i denote a feature being unknown ($z'_i = 0$) or observed ($z'_i = 1$) and $\emptyset_i$ refer to the feature attribution values.

The class of additive feature attribution methods has an important property that there is a single unique solution in this class with three desirable properties, i.e., local accuracy, missingness, and consistency. Where these three properties are met and Equation (5) is followed, there is only one possible explanation model g:

$$\emptyset_i(f, x) = \sum_{z' \subseteq x'} \frac{|z'|!\,(M - |z'| - 1)!}{M!}[f_x(z') - f_x(z' \backslash i)] \tag{6}$$

where f denotes the original prediction model, $|z'|$ represents the number of non-zero entries in z', $z' \subseteq x'$ refers to all z' vectors where the non-zero entries are a subset of the non-zero entries in x', $z' \backslash i$ is the setting $z'_i = 0$. The SHAP values are the solutions of Equation (6) and indicate a unified measure of variable importance in DNN regression models. Further details behind these ideas can be found in [27,55,56].

4.3 Evaluation Metrics

To evaluate the performance of DNN in construction cost estimation, this study considers commonly used metrics such as coefficient of determination (R^2), the root mean square error (RMSE), the mean of absolute error (MAE), and the mean average percentage error (MAPE):

$$R^2 = 1 - \frac{\sum_{i=1}^{m}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{m}(y_i - \bar{y})^2} \tag{7}$$

$$RMSE = \sqrt{\frac{1}{m} \times \sum_{i=1}^{m}(\hat{y}_i - y_i)^2} \tag{8}$$

$$MAE = \frac{1}{m} \times \sum_{i=1}^{m}|\hat{y}_i - y_i| \tag{9}$$

$$MAPE = \frac{1}{m} \times \sum_{i=1}^{m}\left|\frac{\hat{y}_i - y_i}{y_i}\right| \tag{10}$$

where y_i is the observed target variable; $\hat{y}_i$ is the predicted target variable; $\bar{y}$ is the mean of y_i.

5. Model development

Given the nature of this study, feature engineering and data pre-processing were essential before training DNN and other machine learning models. This study calculated the exponential moving average and the linear extrapolation of the economic factors. One-hot encoding and feature scaling were also adopted based on the dataset mentioned in Section 3. Then, two datasets were created, one with economic factors (D1) and the other without economic factors (D2). Both datasets were used to train DNN and four ML models (i.e., LR, SVM, DT, and RF). The last four models were used as baselines to assess the performance of DNN. Their performances were evaluated by several metrics and compared using the two datasets to explore the effects of economic factors. To further explain the results of DNN, SHAP was employed to calculate the importance of economic factors relative to project characteristics based on the DNN model using the dataset including economic factors. The whole model development process is shown in Figure 1.

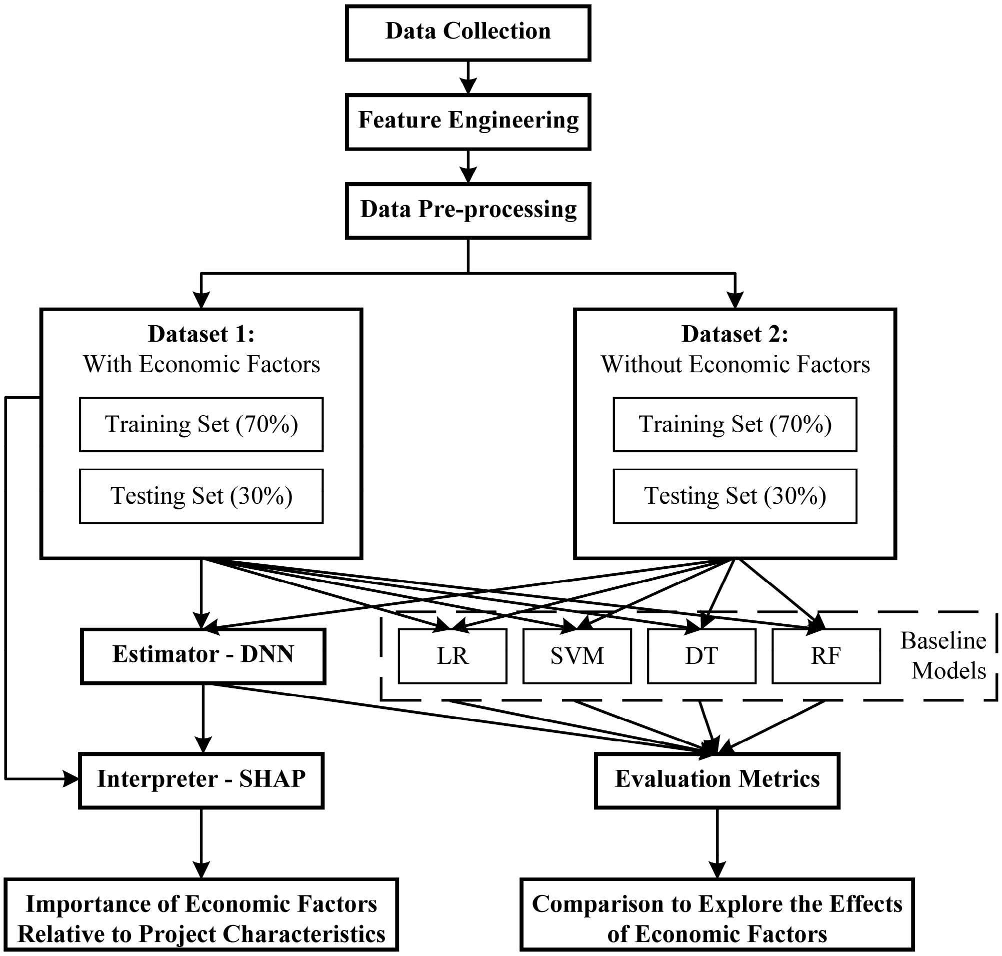


Figure 1 Research process

5.1 Feature engineering

To smooth out the short-term fluctuations of the economic indexes including BWTPI, HSI-Close, HATPI, CPI, and ADW, their exponential moving average were calculated and considered as input variables. Also, to capture one-step-ahead (one year after the date of cost estimation), this study simply included the linear extrapolation of these variables into the dataset. Transportation is extremely expensive in Hong Kong and could massively affect construction costs. Since most of the construction materials are imported from the border of Shenzhen city in mainland China, this study encoded the location of projects into three categories based on their distance from the border. Then, one-hot encoding was used to convert the location variable into binary variables before feeding to neural networks. Table 2 lists these newly generated variables with their explanations.

Table 3 also presents the descriptive information of all the output and input variables.

Table 2 New generated variables after feature engineering

Variable	Variable Name	Description
X4-1	Location-A	A dummy variable indicating the locations based on their distance from the border of Shenzhen City in mainland China, including Sham Shui Po, Kowloon City, Wong Tai Sin, Kwun Tong; 1=yes, 0=no
X4-2	Location-B	A dummy variable indicating the locations based on their distance from the border of Shenzhen City in mainland China, including Tsuen Wan, Tuen Mun, Yuen Long, North; 1=yes, 0=no
X4-3	Location-C	A dummy variable indicating the locations based on their distance from the border of Shenzhen City in mainland China, including Islands, Sai Kung, Eastern, Southern; 1=yes, 0=no
X9-1	BWTPI-EMA	Exponential moving average of BWTPI with a period of 6 samples
X9-2	BWTPI-Trend	Simple linear extrapolation of BWTPI with 6 samples
X10-1	HSI-Close-EMA	Exponential moving average of HIS-close with a period of 6 samples
X10-2	HSI-Close-Trend	Simple linear extrapolation of HIS-close with 6 samples
X11-1	HATPI-EMA	Exponential moving average of HATPI with a period of 6 samples
X11-2	HATPI-Trend	Simple linear extrapolation of HATPI with 6 samples
X12-1	CPI-EMA	Exponential moving average of CPI with a period of 6 samples
X12-2	CPI-Trend	Simple linear extrapolation of CPI with 6 samples
X13-1	ADW-EMA	Exponential moving average of ADW with a period of 6 samples
X13-2	ADW-Trend	Simple linear extrapolation of ADW with 6 samples

Table 3 Descriptive statistics of all variables

Variable	Mean ± St. Dev.	Variable	Mean ± St. Dev.
Y	121.87±93.01	X9-2	686.77±229.76
X1	0.09±0.29	X10	78.99±7.58
X2	27.59±5.99	X10-1	931.26±108.84
X3	4.29±7.28	X10-2	923.41±237.16
X4-1	0.4±0.49	X11	14433.86±3483.59
X4-2	0.42±0.5	X11-1	706.96±196.29
X4-3	0.18±0.39	X11-2	79.62±6.47
X5	11126.09±2011.97	X12	916.59±96.88
X6	1.8±0.29	X12-1	950.18±304.86
X7	0.34±0.72	X12-2	15531.62±4067.61
X8	6.97±1.68	X13	725.22±248.14
X9	908.63±282.36	X13-1	79.97±8.55
X9-1	14674.62±5023.14	X13-2	966.46±123.93

5.2 Data pre-processing

Before training the DNN model as well as the other ML models, this study used feature scaling. The input variables were transformed into the interval of [0,1], respectively:

$$y(x_i) = \frac{x_i - x_{min}}{x_{max} - x_{min}} \tag{11}$$

Then, to detect the possible overfitting issue and to create an optimal DNN model that demands testing and comparing numerous models, all the 98 observations were randomly divided into two parts. 70% of observations were employed as the training set and the other 30% as the testing set. The performance of the selected model using the testing set would be presented as the final results. Apart from the dataset (D1) with all the output and input variables including economic factors, this study created another dataset (D2) excluding economic factors to compare and thus explore the effects of economic factors. Both two datasets, D1 and D2, follow the above data pre-processing.

5.3 DNN training

During the DNN implementation, several hyperparameters need to be tuned: types of layers, number of layers, and number of nodes in each layer. Hyperparameter tuning and DNN structure design procedures have been the focus of multiple studies and are well documented (e.g., in [52,57–60]). Inspired by previous studies and following their recommendations, this study used fully connected layers, which is more applicable to engineering issues. Then the number of layers and the number of nodes in each layer were determined by the process of trial and error based on learning curve analysis. When there are 3 hidden layers and each containing 40 neurons, the developed DNN model has the lowest training and cross-validation errors. DNN is able to solve the overfitting problems via dropout layers and weight regularization. Considering the limited number of inputs and the simple DNN model structure, this study adopted weight regularization, ***L2*** regularization. When the cross-validation error is the lowest, the value of ***L2***-regularization weight is optimal, 10^{-6}. Apart from the developing process of DNN, there are several parameters in the optimization algorithm, Adam optimizer. Usually, an optimizer experiences a certain number of epochs' iteration or until the decrease of loss functions is lower than a certain tolerance. The period is called patience and the values are called minimum delta. In this study, the number of epochs was set to a large number, 1000, so that the termination of the training process would not be affected by the number of epochs. The optimization stops when the decrease in the loss function (MAPE) is lower than 2% in the previous 50 steps. That is, the minimum delta is 2% and the patience is 50. In one iteration, the mini-batches size, the number of training observations included was designated to 32. This study used the default value of the learning rate, 0.0001, in the KERAS library, a deep learning application programming interface [61]. The selected values of hyperparameters are summarized and presented in Table 4.

Table 4 Summary of the selected hyperparameter values of DNN

Hyperparameter	**Value**
Types of layers	Fully connected layers
Number of layers	3
Number of nodes in each layer	40
L2-regularization weight	10^{-6}
Number of epochs	1000
Minimum delta	2%
Patience	50
Mini-batches size	32
Learning rate	0.0001
Optimizer	Adam
Loss function	MAPE

5.4 Training other ML models as baselines

To assess the predictive performance of DNN and provide baselines, DNN prediction results were compared with that of LR, SVM, DT, and RF (four popular methods in prior literature mentioned in Section 2.2). LR depicts the process of estimating the linear relationship between a dependent variable and one or more independent variables. A general model of LR is as follows:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + +\beta_p X_p + \varepsilon \tag{12}$$

where Y is the estimated final cost, and $X_1,\ X_2, \ldots, X_p$ are the features of the projects and the external economic factors, $\beta_0, \beta_1, \beta_2, \ldots, \beta_p$ are parameters to be estimated using the least squares criterion and ε is an error term with a random normal distribution. After the parameters $\beta_0, \beta_1, \beta_2, \ldots, \beta_p$ are known, the LR model could be established to predict the construction cost.

SVM [62] maps the input vector nonlinearly to high dimensional feature space and then considers the linear correlation between the inputs and the outputs. During the implementation of SVM with a radial basis function (RBF) kernel, there are two parameters: the regularized constant C and the RBF kernel parameter γ. C was determined by grid search with 5-fold cross-validation. The value of γ is the reciprocal of the number of used features. The optimal C and γ were selected as 100000 and 0.04, respectively.

DT is a binary recursive partitioning [63]. Beginning from the root node, the data is split into two parts or say two child nodes. At each node, all the possible binary split is searched and computed, and the best split is chosen. Then the splitting process is repeated recursively until there is no further possible split and the tree is complete. RF is an extension of DT [64]. It is an ensemble of a number of small trees built on a part of training samples selected randomly. When all the trees are developed, they would be combined into an ensemble and their average predictions would be adopted as the final results. The parameters of DT and RF were determined by grid search with 5-fold cross-validation. That is, the sample size for each terminal node is 1, the minimum sample size required to split an internal node is 2, the number of trees is 100, the maximum depth of a tree is 5, and the minimum number of samples on a node for branching is 2.

6. Results and Discussion

6.1 Effects of Economic Factors

DNN and the other four ML models were trained on both two datasets, namely D1 including the economic factors while D2 excluding the economic factors. Their performances are reported in Table 5. DNN and the other four techniques have different performances using different evaluation metrics. Thus, several metrics rather than a single one need to be used when assessing the performance of ML tools. Compared with the results using D2, all the values of R^2 increase significantly and all the values of the error metrics (RMSE, MSE, MAE, and MAPE) decrease considerably when using D1 which includes the economic factors. In other words, economic factors play an important role in reducing the errors of construction cost estimation. This intuitively shows the significant effects of economic factors on construction cost estimation.

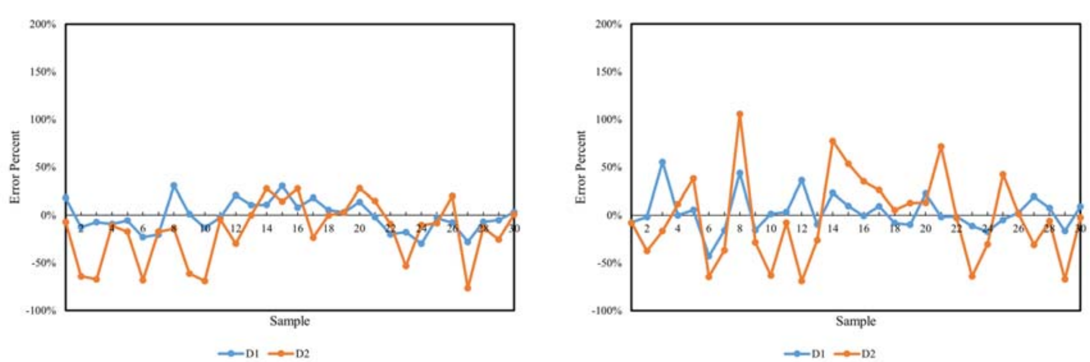
Then the performances of DNN and the other four ML techniques were compared and analyzed regarding the results using D1. As for R^2, all the established models fit the data well. DNN has the highest value of R^2 with a score of 0.95, followed by RF with a score of 0.91. Then, DNN has the lowest value of RMSE with a score of 33.52, followed by RF. That is, the standard deviation of the residuals between the calculated and actual target values for DNN and RF is the lowest, indicating is the construction cost predicted by DNN or RF is the closest to the observed construction cost estimates. Similarly, the MSE value of DNN is the lowest, with a score of 1123.90. As for MAE and MAPR, RF has the lowest values with scores of 19.02 and 10.75%, followed by DNN with scores of 20.79 and 12.91%. Overall, the performance of DNN is slightly more feasible and reliable than the other four ML models (LR, SVM, DT, and RF). Among the four baseline models, RF has the best performance, indicating the potential of RF for estimating construction cost.

Table 5 Model performance of DNN, LR, SVM, and DT using dataset with or without economic factors

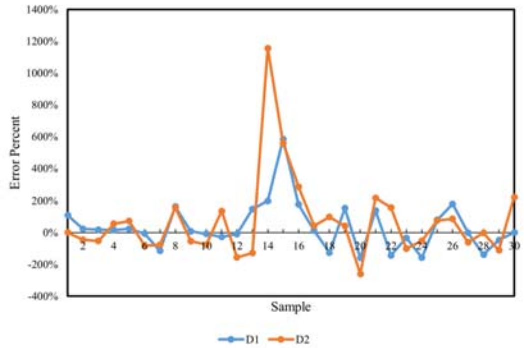
Metrics	Dataset	R^2	RMSE	MSE	MAE	MAPE
DNN	D1	0.95	33.52	1123.90	20.79	12.91%
	D2	0.55	115.11	13250.77	62.53	26.22%
LR	D1	0.87	49.69	2469.06	25.30	13.79%
	D2	0.53	80.58	6493.44	56.03	35.18%
SVM	D1	0.86	118.38	14013.89	63.85	26.42%
	D2	0.12	126.51	16004.17	68.03	27.75%
DT	D1	0.88	75.22	5657.43	35.29	16.80%
	D2	0.05	118.56	14057.39	66.71	31.20%
RF	D1	0.91	34.58	1195.63	19.02	10.75%
	D2	0.22	104.60	10940.17	58.54	26.22%

Note: D1 is the dataset with all the output and input variables including economic factors; D2 is the dataset with all the output and input variables excluding economic factors.

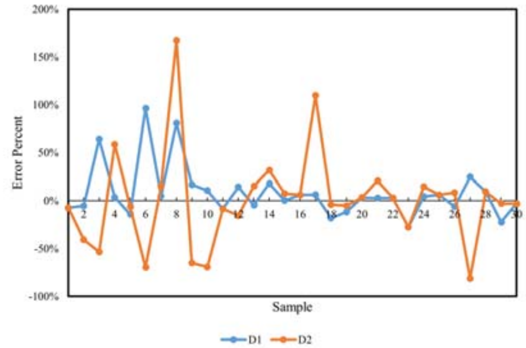
To amply show the effects of economic factors on construction cost estimation, this study drew several figures. Figure 2 provides the percent of prediction errors to the actual values on the testing set for DNN and the other four ML tools. Each part compares the prediction error percent using D1 (the dataset with economic factors) and D2 (the dataset without economic factors). When the dataset included economic factors, the prediction error percent decreased considerably for DNN, LR, DT, and RF models while slightly for the SVM model compared to that when the dataset excluded economic factors. Specifically, for all the five models, there are high errors for some of the samples. This may be partially because the data of some influential factors are not available and thus, they are not included in this study. Besides, there may be some inherent randomness for construction cost estimation. However, the aim of this study was not to train a perfect model but to explore the effects of economic factors. In each figure, almost all the samples with high errors using D2 would have lower errors when using D1. That is, the inclusion of the economic factors could reduce the prediction errors when estimating construction costs regardless of the estimation techniques.



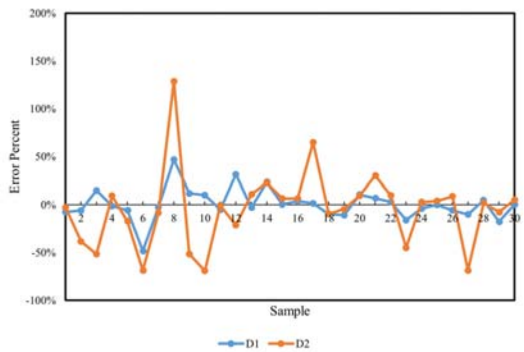
(a) DNN



(b) LR



(c) SVM



(d) DT

(e) RF

Figure 2 Prediction error percent on the testing set for DNN, LR, SVM, DT, and RF

6.2 Importance of external economic factors relative to project characteristics

Figure 3 provides the average SHAP values of all the input variables. The first six important variables are associated with the external economic factors rather than the project characteristics. To decide the number of top important variables that need to be given top priority, this study computed the change of DNN's prediction performance as the number of top important variables increases, presented in Figure 4. As the number of top variables input increases, R^2 increases, and MAPE decreases roughly. The jagged points in Figure 4 are inevitable because of the nature of the optimization method (stochastic gradient descent). However, this does not nullify the big picture that is depicted in Figure 4. When the first four important variables are considered, the DNN's R^2 is 0.93 and its MAPE is 13.79%. When more than four variables are input, the change of R^2 and MAPE becomes smoother and even R^2 decreases a bit while MAPE increases a bit. Thus, the first four variables are regarded as the most influential factors. All are relevant to the external economic environments, verifying their non-negligible influence on construction cost estimation. Once the external economic situation changes, the estimation needs to adjust timely. This is consistent with many previous studies [20–22] and governmental reports [65], which presented that macroeconomic dynamics are one of the top factors influencing construction costs. In particular, economic factors may affect the cost of equipment, materials, labor and loans and thus have significant indirect effects on the final construction cost [66]. If only the direct effects are emphasized, economic factors may be less influential. This may be the reason why some studies reported that economic factors were less important than project and site characteristics and then were neglected.

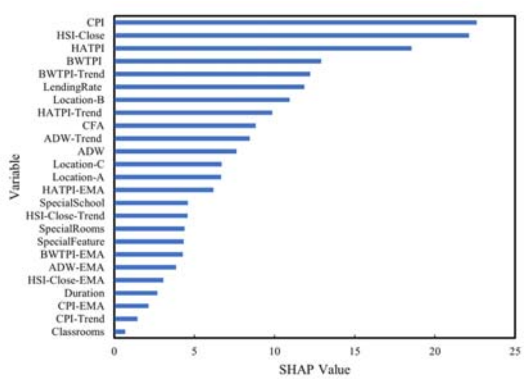


Figure 3 The relative importance of input variables

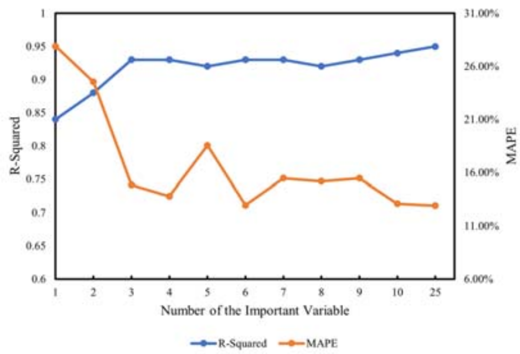


Figure 4 Selection of number of the top important variable

Among the top four variables, the most important one is CPI, commonly used as a measure of inflation [67]. Inflation directly decreases the real value of the currency and increases materials and equipment costs. Numerous financial and economic decision-makings especially the estimation of construction cost need to depend on inflation expectations [68]. This is because the construction industry is a project-based and capital-intensive sector and is strongly associated with national economic health [69]. This finding is consistent with the results of previous studies [15,70].

HSI-Close is the second important factor. It has been used as a proxy for macroeconomic impact to reflect the stock market conditions. Zhang [71] also provides evidence that stock market conditions make difference to the construction cost. This is because stock market conditions could mirror the extent of risks. If the market sentiment is risk-averse, the same sentiment would be applied in construction pricing. Therefore, stock market conditions need to be considered when estimating the construction cost for public school projects.

The third and fourth important factors are HATPI and BWTPI, respectively. The tender price indexes are the predicted pattern of movements of tender price, reflecting the prevailing market conditions [72]. They have been used as indications of construction cost levels for clients [73]. It is not surprising that HATPI and BWTPI contribute greatly to construction cost estimation.

7. Conclusion

Previous studies have developed construction cost estimation models without the consideration of economic factors mainly because of a lack of consensus among researchers. More importantly, little attention has been paid to combining the trend and value of market indexes into construction cost estimation. This study used DNN as an estimator and SHAP as a model interpreter to explore the effects of economic factors on construction cost estimation. This exploration is based on the data on 98 public school projects in HKSAR. The results show that economic factors and their trends exert much influence on reducing the errors of early construction cost estimation, indicating the effect of economic factors could not be overlooked on rather smaller yet complex datasets. This study also found that the economic factors can play a more important role than projects' characteristics in construction cost estimation.

This study has the following theoretical contributions. First, this study supplements the understanding of the effects of external economic factors on construction cost estimation. There has been a controversy over the effects of economic factors. Some researchers (e.g., in [14]) reported that economic factors pose the least influences on project cost. On the other hand, a number of studies [17,22–24] based on questionnaire surveys stated that economic factors could not be neglected. Using a quantitative approach, this study confirmed the external economic factors play a huge role in reducing the early costs estimation errors even in a limited number of samples for complex projects, contributing to the body of knowledge of construction engineering and management. Then, this paper facilitates the relevant literature on construction cost estimation. Many prior studies used projects' characteristics only to estimate construction cost, emphasizing excessively the importance of projects' characteristics. However, projects' characteristics were found to be less important than the economic factors in the present study. With consideration of the economic factors, researchers in the field of construction cost estimation may revisit the effects of influential factors including other projects' characteristics, and uncover their actual degree of effects on construction cost estimation.

For the practical implications, the finding that the external economic factors could reduce early construction cost estimation errors would be helpful for stakeholders, bidders, and contractors. Their decisions about initiating, bidding, and managing projects can be made more appropriately. This finding could help stakeholders and contractors, particularly during the construction phase. By monitoring the changes of the external economic factors, stakeholders and contractors could update the estimated construction cost in time and compare the updated cost with the planned cash flows so that some necessary precautionary measures can be taken to alleviate the possible shortage of funds. The finding that economic factors need to be considered even in small datasets implies that some stakeholders and contractors could employ their historical data only to estimate early construction cost when more external projects' data is difficult to access.

There are also some limitations in the present study. Apart from the external economic factors, construction cost estimation may be affected by other variables, e.g., the characteristics of clients, contractors, and other stakeholders as well as the development of construction technologies. Besides, the external social conditions including the occurrence of pandemic disease (e.g., COVID-19) may lead to an increase of force majeure cost and thus a low accuracy of construction cost estimation. Considering more variables may reduce the construction cost estimation errors. The effects of economic factors on construction cost estimation were explored using the data on Hong Kong public school projects. The findings in the present study may serve as a valuable reference in other small open economies due to the similarities in capital structure. Future studies are recommended to replicate the present study in other contexts with different project types and different regions.

Data Availability Statement

All data (including the project estimate cost, project duration, Hang Seng index-close, and others), models, or code that support the findings of this study are available from the corresponding author upon reasonable request.

References

[1] M.A. Azman, Z. Abdul-Samad, S. Ismail, The accuracy of preliminary cost estimates in Public Works Department (PWD) of Peninsular Malaysia, International Journal of Project Management. 31 (2013) pp. 994–1005. https://doi.org/10.1016/j.ijproman.2012.11.008.

[2] Z. Shehu, I.R. Endut, A. Akintoye, G.D. Holt, Cost overrun in the Malaysian construction industry projects: A deeper insight, International Journal of Project Management. 32 (2014) pp. 1471–1480. https://doi.org/10.1016/j.ijproman.2014.04.004.

[3] A. Akintoye, Analysis of factors influencing project cost estimating practice, Construction Management and Economics. 18 (2000) pp. 77–89. https://doi.org/10.1080/014461900370979.

[4] D.P. Alex, M. Al Hussein, A. Bouferguene, S. Fernando, Artificial neural network model for cost estimation: City of Edmonton's water and sewer installation services, Journal of Construction Engineering and Management. 136 (2010) pp. 745–756. https://doi.org/10.1061/(ASCE)CO.1943-

7862.0000184.

[5] A.H. Al-Momani, Construction cost prediction for public school buildings in Jordan, Construction Management and Economics. 14 (1996) pp. 311–317. https://doi.org/10.1080/014461996373386.

[6] D.J. Lowe, M.W. Emsley, A. Harding, Predicting construction cost using multiple regression techniques, Journal of Construction Engineering and Management. 132 (2006) pp. 750–758. https://doi.org/10.1061/(ASCE)0733-9364(2006)132:7(750).

[7] H. Son, C. Kim, C. Kim, Hybrid principal component analysis and support vector machine model for predicting the cost performance of commercial building projects using pre-project planning variables, Automation in Construction. 27 (2012) pp. 60–66. https://doi.org/10.1016/j.autcon.2012.05.013.

[8] S. Fang, T. Zhao, Y. Zhang, Prediction of construction projects' costs based on fusion method, Engineering Computations. 34 (2017) pp. 2396–2408. https://doi.org/10.1108/EC-02-2017-0065.

[9] S.Z. Doğan, D. Arditi, H. Murat Günaydin, Using decision trees for determining attribute weights in a case-based model of early cost prediction, Journal of Construction Engineering and Management. 134 (2008) pp. 146–152. https://doi.org/10.1061/(ASCE)0733-9364(2008)134:2(146).

[10] C.-H. Huang, S.-H. Hsieh, Predicting BIM labor cost with random forest and simple linear regression, Automation in Construction. 118 (2020) pp. 103280. https://doi.org/10.1016/j.autcon.2020.103280.

[11] D. Chakraborty, H. Elhegazy, H. Elzarka, L. Gutierrez, A novel construction cost prediction model using hybrid natural and light gradient boosting, Advanced Engineering Informatics. 46 (2020) pp. 101201. https://doi.org/10.1016/j.aei.2020.101201.

[12] D. Baloi, A.D.F. Price, Modelling global risk factors affecting construction cost performance, International Journal of Project Management. 21 (2003) pp. 261–269. https://doi.org/10.1016/S0263-

7863(02)00017-0.

[13] T.M.S. Elhag, A.H. Boussabaine, T.M.A. Ballal, Critical determinants of construction tendering costs: Quantity surveyors' standpoint, International Journal of Project Management. 23 (2005) pp. 538–545. https://doi.org/10.1016/j.ijproman.2005.04.002.

[14] Y.-M. Cheng, An exploration into cost-influencing factors on construction projects, International Journal of Project Management. 32 (2014) pp. 850–860. https://doi.org/10.1016/j.ijproman.2013.10.003.

[15] M. Gunduz, O.L. Maki, Assessing the risk perception of cost overrun through importance rating, Technological and Economic Development of Economy. 24 (2018) pp. 1829–1844. https://doi.org/10.3846/20294913.2017.1321053.

[16] D.C. Okpala, A.N. Aniekwu, Causes of high costs of construction in Nigeria, Journal of Construction Engineering and Management. 114 (1988) pp. 233–244. https://doi.org/10.1061/(ASCE)0733-9364(1988)114:2(233).

[17] L. Zhao, J. Mbachu, Z. Liu, Identifying significant cost-influencing factors for sustainable development in construction industry using structural equation modelling, Mathematical Problems in Engineering. 2020 (2020) pp. 4810136. https://doi.org/10.1155/2020/4810136.

[18] S. Bayram, S. Al-Jibouri, Efficacy of estimation methods in forecasting building projects' costs, Journal of Construction Engineering and Management. 142 (2016) pp. 05016012. https://doi.org/10.1061/(ASCE)CO.1943-7862.0001183.

[19] M.T. Hatamleh, M. Hiyassat, G.J. Sweis, R.J. Sweis, Factors affecting the accuracy of cost estimate: case of Jordan, Engineering, Construction and Architectural Management. 25 (2018) pp. 113–131.

https://doi.org/10.1108/ECAM-10-2016-0232.

[20] B. Akinci, M. Fischer, Factors affecting contractors' risk of cost overburden, Journal of Management in Engineering. 14 (1998) pp. 67–76. https://doi.org/10.1061/(ASCE)0742-597X(1998)14:1(67).

[21] J.S. Shane, K.R. Molenaar, S. Anderson, C. Schexnayder, Construction project cost escalation factors, Journal of Management in Engineering. 25 (2009) pp. 221–229. https://doi.org/10.1061/(ASCE)0742-597X(2009)25:4(221).

[22] H. Doloi, Cost overruns and failure in project management: Understanding the roles of key stakeholders in construction projects, Journal of Construction Engineering and Management. 139 (2013) pp. 267–279. https://doi.org/10.1061/(ASCE)CO.1943-7862.0000621.

[23] K.C. Iyer, K.N. Jha, Factors affecting cost performance: evidence from Indian construction projects, International Journal of Project Management. 23 (2005) pp. 283–295. https://doi.org/10.1016/j.ijproman.2004.10.003.

[24] S. Zhang, S.M. Bogus, C.D. Lippitt, G.C. Migliaccio, Estimating location-adjustment factors for conceptual cost estimating based on nighttime light satellite imagery, Journal of Construction Engineering and Management. 143 (2017) pp. 04016087. https://doi.org/10.1061/(ASCE)CO.1943-7862.0001216.

[25] L. Zhao, B. Wang, J. Mbachu, Z. Liu, New Zealand building project cost and its influential factors: A structural equation modelling approach, Advances in Civil Engineering. 2019 (2019) pp. 1362730. https://doi.org/10.1155/2019/1362730.

[26] J.K. Larsen, G.Q. Shen, S.M. Lindhard, T.D. Brunoe, Factors affecting schedule delay, cost overrun, and quality level in public construction projects, Journal of Management in Engineering. 32 (2016) pp.

04015032. https://doi.org/10.1061/(ASCE)ME.1943-5479.0000391.

[27] S.M. Lundberg, S.-I. Lee, A unified approach to interpreting model predictions, in: Proceedings of the 31st International Conference on Neural Information Processing Systems, Curran Associates Inc., Red Hook, NY, USA, 2017: pp. 4768–4777. https://dl.acm.org/doi/abs/10.5555/3295222.3295230 (accessed November 11, 2021).

[28] S.M. Lundberg, G.G. Erion, S.-I. Lee, Consistent individualized feature attribution for tree ensembles, ArXiv:1802.03888. (2019). http://arxiv.org/abs/1802.03888 (accessed November 1, 2020).

[29] S. Mangalathu, S.-H. Hwang, J.-S. Jeon, Failure mode and effects analysis of RC members based on machine-learning-based SHapley Additive exPlanations (SHAP) approach, Engineering Structures. 219 (2020) pp. 110927. https://doi.org/10.1016/j.engstruct.2020.110927.

[30] V.M. García, J.L. Aznarte, Shapley additive explanations for NO2 forecasting, Ecological Informatics. 56 (2020) pp. 101039. https://doi.org/10.1016/j.ecoinf.2019.101039.

[31] Project Strategy and Governance Office (PSGO), Construction Cost Information. https://www.psgo.gov.hk (accessed November 10, 2021).

[32] E. Erdis, The effect of current public procurement law on duration and cost of construction projects in Turkey, Journal of Civil Engineering and Management. 19 (2013) pp. 121–135. https://doi.org/10.3846/13923730.2012.746238.

[33] I. Dutta, S. Dutta, B. Raahemi, Detecting financial restatements using data mining techniques, Expert Systems with Applications. 90 (2017) pp. 374–393. https://doi.org/10.1016/j.eswa.2017.08.030.

[34] L. Breiman, Bagging predictors, Machine Learning. 24 (1996) pp. 123–140. https://doi.org/10.1007/BF00058655.

[35] S. Bhattacharyya, S. Jha, K. Tharakunnel, J.C. Westland, Data mining for credit card fraud: A comparative study, Decision Support Systems. 50 (2011) pp. 602–613. https://doi.org/10.1016/j.dss.2010.08.008.

[36] P.F. Pai, M.F. Hsu, M.C. Wang, A support vector machine-based model for detecting top management fraud, Knowledge-Based Systems. 24 (2011) pp. 314–321. https://doi.org/10.1016/j.knosys.2010.10.003.

[37] R. Wang, V. Asghari, S.-C. Hsu, C.-J. Lee, J.-H. Chen, Detecting corporate misconduct through random forest in China's construction industry, Journal of Cleaner Production. 268 (2020) pp. 122266. https://doi.org/10.1016/j.jclepro.2020.122266.

[38] M. Belkin, D. Hsu, P.P. Mitra, Overfitting or perfect fitting? Risk bounds for classification and regression rules that interpolate, in: Proceedings of the 32nd International Conference on Neural Information Processing Systems, Curran Associates Inc., Red Hook, NY, USA, 2018: pp. 2306–2317. https://dl.acm.org/doi/abs/10.5555/3327144.3327157 (accessed November 11, 2021).

[39] I. Mutis, A. Ambekar, V. Joshi, Real-time space occupancy sensing and human motion analysis using deep learning for indoor air quality control, Automation in Construction. 116 (2020) pp. 103237. https://doi.org/10.1016/j.autcon.2020.103237.

[40] M. Kassem, E. Mahamedi, K. Rogage, K. Duffy, J. Huntingdon, Measuring and benchmarking the productivity of excavators in infrastructure projects: A deep neural network approach, Automation in Construction. 124 (2021) pp. 103532. https://doi.org/10.1016/j.autcon.2020.103532.

[41] T.D. Akinosho, L.O. Oyedele, M. Bilal, A.O. Ajayi, M.D. Delgado, O.O. Akinade, A.A. Ahmed, Deep learning in the construction industry: A review of present status and future innovations, Journal of

Building Engineering. 32 (2020) pp. 101827. https://doi.org/10.1016/j.jobe.2020.101827.

[42] T. Siu, C. Jim, Hong Kong plans lower budget deficit as economy expected to recover, Reuters. (2021). https://www.reuters.com/article/us-hongkong-economy-budget-idUSKBN2AO0GB (accessed July 29, 2021).

[43] J.M. Vincent, D. McKoy, The complex and multi-faceted nature of school construction costs: Factors affecting California, American Institute of Architects, Sacramento, CA, 2008. https://escholarship.org/uc/item/7hp8820v (accessed July 29, 2021).

[44] J.M. Vincent, P. Monkkonen, The impact of state regulations on the costs of public school construction, Journal of Education Finance. 35 (2010) pp. 313–330. http://dx.doi.org/10.1353/jef.0.0020.

[45] B.-G. Hwang, S.R. Thomas, C.T. Haas, C.H. Caldas, Measuring the impact of rework on construction cost performance, Journal of Construction Engineering and Management. 135 (2009) pp. 187–198. https://doi.org/10.1061/(ASCE)0733-9364(2009)135:3(187).

[46] M. Juszczyk, A. Leśniak, Modelling construction site cost index based on neural network ensembles, Symmetry. 11 (2019) pp. 411. https://doi.org/10.3390/sym11030411.

[47] T.P. Williams, Predicting changes in construction cost indexes using neural networks, Journal of Construction Engineering and Management. 120 (1994) pp. 306–320. https://doi.org/10.1061/(ASCE)0733-9364(1994)120:2(306).

[48] M.T. Cao, M.Y. Cheng, Y.W. Wu, Hybrid computational model for forecasting Taiwan construction cost index, Journal of Construction Engineering and Management. 141 (2014) pp. 04014089. https://doi.org/10.1061/(ASCE)CO.1943-7862.0000948.

[49] S. Hwang, Dynamic regression models for prediction of construction costs, Journal of Construction

Engineering and Management. 135 (2009) pp. 360–367. https://doi.org/10.1061/(ASCE)CO.1943-7862.0000006.

[50] Y. Bengio, Learning deep architectures for AI, Machine Learning. 2 (2009) pp. 1–127. https://doi.org/10.1561/2200000006.

[51] V. Nair, G.E. Hinton, Rectified linear units improve restricted boltzmann machines, in: Proceedings of the 27th International Conference on Machine Learning, Omnipress, Wisconsin, 2010. https://openreview.net/forum?id=rkb15iZdZB (accessed October 6, 2020).

[52] V. Asghari, Y.F. Leung, S.-C. Hsu, Deep neural network based framework for complex correlations in engineering metrics, Advanced Engineering Informatics. 44 (2020) pp. 101058. https://doi.org/10.1016/j.aei.2020.101058.

[53] D.P. Kingma, J. Ba, Adam: A method for stochastic optimization, in: Proceedings of the 3rd International Conference on Learning Representations, 2015. https://wtf-deeplearning.github.io/optimization-regularization/1412.6980.pdf (accessed November 11, 2021).

[54] I. Goodfellow, Y. Bengio, A. Courville, Deep learning, MIT press, Cambridge, MA, 2016. ISBN 9780262337373, https://www.deeplearningbook.org. (accessed November 11, 2021)

[55] M.T. Ribeiro, S. Singh, C. Guestrin, “Why Should I Trust You?”: Explaining the Predictions of Any Classifier, in: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Association for Computing Machinery, New York, 2016: pp. 1135–1144. https://doi.org/10.1145/2939672.2939778.

[56] E. Štrumbelj, I. Kononenko, Explaining prediction models and individual predictions with feature contributions, Knowledge and Information Systems. 41 (2014) pp. 647–665.

https://doi.org/10.1007/s10115-013-0679-x.

[57] J. Brownlee, Better Deep Learning: Train Faster, Reduce Overfitting, and Make Better Predictions, Machine Learning Mastery, Melbourne, Australia, 2018. https://machinelearningmastery.com/better-deep-learning. (accessed November 11, 2021)

[58] N.S. Keskar, J. Nocedal, P.T.P. Tang, D. Mudigere, M. Smelyanskiy, On large-batch training for deep learning: Generalization gap and sharp minima, in: Proceedings of the Workshop of the 5th International Conference on Learning Representations, 2017. http://arxiv.org/abs/1609.04836 (accessed October 10, 2020).

[59] A.Y. Ng, Feature selection, L 1 vs. L 2 regularization, and rotational invariance, in: Proceedings of the 21st International Conference on Machine Learning, Association for Computing Machinery, New York, 2004: pp. 78. https://doi.org/10. 1145/1015330.1015435.

[60] N.G. Andrew, Machine learning yearning: Technical strategy for AI engineers, in the era of deep learning, deeplearning.ai, 2018. https://www.deeplearning.ai/machine-learning-yearning. (accessed November 11, 2021)

[61] F. Chollet, KERAS, GitHub Repos, (2015). https://github.com/keras-team/keras (accessed October 10, 2020).

[62] C. Cortes, V. Vapnik, Support-vector networks, Machine Learning. 20 (1995) pp. 273–297. https://doi.org/10.1023/A:1022627411411.

[63] L. Breiman, J. Friedman, C.J. Stone, R.A. Olshen, Classification and regression trees, Wadsworth International Group, Belmont, CA, 1984. ISBN 9780412048418.

[64] L. Breiman, Random forests, Machine Learning. 45 (2001) pp. 5–32.

https://doi.org/10.1023/A:1010933404324.

[65] Ministry of Business, Innovation & Employment (MBIE), Housing Pressures in Christchurch, Wellington, New Zealand, 2013. https://www.mbie.govt.nz/dmsdocument/1088-housing-pressures-christchurch-2013-pdf (accessed July 18, 2021).

[66] W. Jung, S.H. Han, Which risk management is most crucial for controlling project cost?, Journal of Management in Engineering. 33 (2017) pp. 04017029. https://doi.org/10.1061/(ASCE)ME.1943-5479.0000547.

[67] M. Aiken, Using a neural network to forecast inflation, Industrial Management & Data Systems. 99 (1999) pp. 296–301. https://doi.org/10.1108/02635579910291984.

[68] Y. Cao, A. Shapiro, Will inflation remain low?, FRBSF Economic Letter. 19 (2014). https://www.frbsf.org/economic-research/files/el2014-19.pdf (accessed October 10, 2020).

[69] H. Ren, S.S. Lin, The UK construction industry under cyclical high inflation, high interest rates and recession, International Journal of Project Management. 14 (1996) pp. 301–305. https://doi.org/10.1016/0263-7863(96)84513-3.

[70] C. Kaliba, M. Muya, K. Mumba, Cost escalation and schedule delays in road construction projects in Zambia, International Journal of Project Management. 27 (2009) pp. 522–531. https://doi.org/10.1016/j.ijproman.2008.07.003.

[71] Y.Y. Zhang, Forecasting the trend of construction cost indices for Taiwan employing support vector machine, Chao Yang University of Technology, Taiwan. 2007. http://ir.lib.cyut.edu.tw:8080/handle/310901800/28310. (accessed November 11, 2021)

[72] S.T. Ng, S.O. Cheung, R.M. Skitmore, K.C. Lam, L.Y. Wong, Prediction of tender price index

directional changes, Construction Management and Economics. 18 (2000) pp. 843–852. https://doi.org/10.1080/014461900433122.

[73] S.M. Rowlinson, A. Walker, The construction industry in Hong Kong, Longman, Hong Kong, 1995. ISBN 9620007794.