Construction Cost Estimation for Government Building Using Artificial Neural Network Technique

Sitthikorn Sitthikankun¹,², Damrongsak Rinchumphu²*, Chinnapat Buachart², Eakasit Pacharawongsakda³

¹ Graduate Program in Construction Engineering and Management, Faculty of Engineering, Chiang Mai University, THAILAND.
² Department of Civil Engineering, Faculty of Engineering, Chiang Mai University, THAILAND.
³ Big Data Engineering Program, College of Innovative Technology and Engineering, Dhurakij Pundit University, THAILAND.

*Corresponding Author (Tel +66-959959519, Email: damrongsak.r@cmu.ac.th).

Abstract

The construction bidding competition required effective precision to prevent losses in the bidding process, especially in the public sector. The bidders must have an estimate of the construction cost before the bidding. There are two widely used methods for construction cost estimation: 1) a rough estimation with an advantage of quick construction estimation cost and a disadvantage of a high price tolerance, and 2) a detailed estimation with an advantage of more accurate estimation of construction costs, and disadvantages of the need for a complete construction plan and time-consuming. Considering these disadvantages, research on the government construction cost estimation model was conducted by using the Artificial Neural Network (ANN) technique of forecasting modeling. The study’s results showed that the model consisted of two hidden layers which each layer has ten and eight nodes, respectively, with the best Root Mean Square Error (RMSE) value ± 0.331 million Baht. When the new data set was tested for validity, the R² equal to 0.914 proving the accuracy of the forecasting model as an alternative for government bidding participants to reduce the tolerances and to spend less time to estimate construction costs more efficiently.

Disciplinary: Civil & Construction Engineering and Management.

Cite This Article:

1 Introduction

The construction industry in Thailand is of great importance to the domestic economy, both public and private sectors, according to an assessment by the Economic Intelligence Center (EIC), Siam Commercial Bank Public Company Limited (2020). Currently, the construction industry market has contracted by 1% year-on-year (YOY), amounting to 1.29 trillion baht in which the private construction market contracted by 7.8% YOY, amounting to approximately 5.28 billion baht; in contrast, the public construction industry market has still been growing by 4.5% YOY, amounts to approximately 7.62 billion baht. The overall contraction of the construction industry in the country has caused the bidding competition to need effective precision in bidding to prevent losses from too-low price bidding (Tochaiwat et al., 2020). Wangniwetkul (2009) has mentioned that when there is a construction project of the private sector, the project owner will invite a few potential contractors to participate in the bidding process. Nowadays, the bid for the governmental sector can be succeeded through e-Bidding. At this stage, the tender documents of some projects may be able to be downloaded without any charge. However, the bidders must have an estimate of the construction cost. Currently, the construction cost estimation can be categorized into two widely used methods as a rough estimation and a detailed estimation. For the rough estimation, the construction cost estimation can be performed quickly, but there is a high tolerance in price. In contrast, the detailed cost estimation can provide a construction cost more accurately, but the plan must be complete and the duration for the construction cost estimation is taken longer.

2 Cost Estimation Methods

2.1 Rough Estimation

Rough Estimation is an estimate of the construction cost with an incomplete plan. Besides, the estimate of the construction cost is also based on the experiences of the estimator himself or is based on the data of previously completed projects. The tolerance is approximately 10-25% (Wangniwetkul, 2009) as shown in Figure 1. While Wangniwetkul (2009) mentioned that the tolerance could be as high as 50% which can cause a serious risk to the construction. Therefore, the avoidance of rough estimation should be considered if possible.

![Figure 1: The Tolerance on the Estimation Time (adapted from Wangniwetkul (2009))](http://TuEngr.com)
2.2 Detailed Estimation

This method could be conducted when the plan is completed by calculating the building materials in quantity and then estimating the cost of construction materials, construction wages, machine costs, operating costs, profits, taxes, also interest, etc.

While the rough estimation has a high tolerance range and the detailed estimation takes time to estimate the construction cost with the requirement of a complete plan and complete assembly lists, the construction cost estimation using forecast modeling techniques can reduce the tolerances and takes less time to estimate the construction cost with an incomplete plan and incomplete assembly lists. Therefore, it can be an alternative to help in the construction costs estimate for projects that have a limited time, or governmental bidders who received an unclear plan. According to the literature review, several new techniques for construction cost estimation are currently found. Nevertheless, construction cost estimation aid is to create a model for forecasting with modern techniques.

2.3 Cost Estimation Using Artificial Neural Network (ANN)

There are many methods of modeling for forecasting using modern techniques. ANN is considered an accurate and popular method. ANN is an imitation of the nervous system of the organisms that are connected by learning from the basics first and taking the experience from the preliminary learning to predict further information. Additionally, Matel et al. (2019) said that the ANN method is inspired by the study of human brain processes. Furthermore, Polat (2012) said that ANN is started as a correlation in the Input Layer and the Output Layer to find the relationship between the two and set the correlation weight in the Hidden Layer, where the Input Layer is forwarded to the Hidden Layer, and then the Hidden Layer will calculate the result according to the specified weight. After that, those calculated results will be sent to the Output Layer.

ANN is a technology that simulates the human brain and nervous system (Boussabaine, 1996). It learns from the experiences in previous examples and does new things. It also learns key characteristics from the data that are imported into the input layer. Therefore, it can be interpreted that the artificial neural network consists of three layers as the Input Layer, the Hidden Layer, and the Output Layer. Moreover, Geetha (2014) said that ANN is a connection to internal systems. The system consists of 3 layers, the first part is the Input Layer Part that receives the input and forwards

![Figure 2: Artificial Neural Network.](http://TuEngr.com)
it to the next part which is the Hidden Layer that will calculate the result according to the specified weight, and the last part is the Output Layer, which will present the result of the model.

In the civil engineering aspect, there is the utilization of ANN to forecast or assist in some processes. For instance, Abd et al. (2019) used ANN to estimate the final cost of the Iraqi construction projects, from 501 data sets since 2005-2015. There were 25 Input Layers, and a Hidden Layer was created by using 2 Nodes obtained through several experiments. The result value obtained from this model was $R^2 = 0.987$ which made ANN has been proved for its accuracy of the least Root Mean Square Error (RMSE) from the trial-and-error process. Additionally, Kaviya (2019) also used ANN for forecasting the compressive strength of high-performance concrete from 446 concrete data samples, using 326 data. The model had 8 Input Layers and 1 Hidden Layer and 30 nodes were used. The results presented that ANN was suitable compared to the multiple linear regression model, which can reduce the errors in the concrete industry in any safety issue.

However, before inventing the forecast modeling, it is necessary to consider the important independent variables to be used as the input layer for the forecasting model as the independent variables are the variables that are used for forecasting the dependent variables. From the literature review, it was found that the independent variables used in the building construction cost forecasting consisted of 11 variables, which the definition and measurement of each variable could be explained as the following:

2.3.1 Usage Area of the Building (X1)

It is the total area of the building that can be used. The area can be calculated by multiply the width of the building by the length of the building on every floor and get the total sum in square meters for the measurement. Also, Chakan (2010) used the total usable area of the building, excluding the rooftop, as variables to create a forecasting model.

2.3.2 Average Perimeter (X2)

It is the total sum of each floor perimeter length divided by the number of the floors and to be measured in meters. Rujirayanyong (2012) used the average perimeter variable as one of the variables to create a predictive model, saying that the average perimeter was calculated by the sum of the perimeter of all layers divided by the number of layers, measured in meters.

2.3.3 Average Inter-Floor Height (X3)

The height of the floor can be measured from ground level to floor level of the next floor. To get the average inter-floor height is to have a total sum of each floor height and divided by the number of floors. If the last floor is covered by a roof, the distance from the last floor to the roof beam level should be used. The measurement is in meters. The average inter-floor height was also used by Rujirayanyong (2012), who mentioned about this variable that the average floor height was obtained by calculating the distance from the floor level to the floor level of the next floor of all floors together and dividing it by the number of floors of the building, including the rooftop. If
there is a covered roof, the calculation must include the roof beam level. The measurement is in meters. Additionally, Chakan (2010) also used the mean height between the floors as a variable in forecast modeling.

2.3.4 Building Height (X4)

It is the height from the original soil level or + 0.00 of the building to the highest end of the roof or deck level which is measured in meters. Rujirayanyong (2012) used this variable and said that the building’s height is the same as the distance from the bottom floor to the roof level or up to the roof ridge if the building is covered with a roof. The unit of measurement is in meters. Additionally, Chakan (2010) used the height variable of a residential building as a forecasting model.

2.3.5 Number of Floors (X5)

It is the total number of floors of the building, including the rooftop and basement, measured in units. Rujirayanyong (2012) also used as a variable to create a forecasting model and mentioned that the number of floors means all floors including the basement and the rooftop. Also, Chakan (2010) also used the number of floors above the ground as a variable but separated the number of basements as another variable.

2.3.6 Roof Area (X6)

It is the material used for covering the building, which is measurable by using the area of the roofing material according to the actual slope. The unit of measurement is square meters. Rujirayanyong (2012) used the roof area as a variable for the forecast modeling which stating that the roof area is equal to the area roofed by any material and measured according to the slope of the roof. The measurement unit is square meters.

2.3.7 Bathroom Area (X7)

This is the area of the entire bathrooms inside the building, which can be measured by using the width multiplied by the length of each bathroom and combine all bathroom areas altogether which to be measured in square meters. For this bathroom size variable, Rujirayanyong (2012) used the bathroom space as the forecast modeling variable and said that the bathroom area is equal to the sum of all bathrooms inside the building. The unit of measurement is square meters. Besides, Duangsangthong (2016) also used the same variable for the forecast modeling.

2.3.8 Ground Slab (X8)

It is the area of the building that is adjacent to the ground or the area on the bottom floor of the building. If the building is a raised building, it is considered that there is no ground slab. The measurement unit is square meters. This variable was also utilized by Rujirayanyong (2012) as a variable for the forecast modeling which stating that the value of the ground slab is equal to the total area of the slab on the ground which is mostly the bottom floor adjacent to the ground. The unit of measurement is square meters.
2.3.9 Open Space (X9)

It is the area that has space with the niches in and around the building, whether it is light, doors, windows, inside or around the building, or combined altogether. The measurement unit is square meters. Rujirayanyong (2012) used it as a variable for the forecast modeling and stated that the value of the open space is the sum of the various opening areas, whether it is a light niche, door, or window. The unit of measurement is square meters.

2.3.10 Types of Roof (X10)

It is how the types of roofing material are determined whether it is metal sheets or other roofing materials.

2.3.11 Types of Floor Structure (X11)

It is the floor structure inside the building, which is divided into 2 types as a precast, regardless of what type of precast concrete floor and a cast-in-place, regardless of what form it is as well.

From all the literature reviews, the modeling for forecasting with modern techniques by the artificial neural network (ANN) technique is suitable and meets the demands of promptness and accuracy. The forecasting price unit in this study is Thai Baht. There are 11 forecasting variables in total. All variables which have been used for the construction cost forecast modeling according to the literature review can be summarized in Table 1.

<table>
<thead>
<tr>
<th>Table 1: The Summary of Variables from the Literature Review.</th>
</tr>
</thead>
<tbody>
<tr>
<td>---------------------</td>
</tr>
<tr>
<td>X1</td>
</tr>
<tr>
<td>X2</td>
</tr>
<tr>
<td>X3</td>
</tr>
<tr>
<td>X4</td>
</tr>
<tr>
<td>X5</td>
</tr>
<tr>
<td>X6</td>
</tr>
<tr>
<td>X7</td>
</tr>
<tr>
<td>X8</td>
</tr>
<tr>
<td>X9</td>
</tr>
<tr>
<td>X10</td>
</tr>
<tr>
<td>X11</td>
</tr>
</tbody>
</table>

In conclusion, as per the literature review, it is found that in the process of project proposal preparation, a faster and more accurate estimation is needed. Currently, ANN has been proved as a usable method. However, there have been no trials in the government context. Also, further studies are still needed via the methodology which is presented in the next section.

3 Methodology

The data for modeling were collected from the public procurement system. The information was downloaded via the Internet which was free of charge, and no tender envelopes were sold.
Therefore, 50 data sets were collected. The collecting data were the building construction data for the fiscal year of 2020 based on the period of the cost of materials, labor, the Bureau of Public Procurement Standards, and the Comptroller General’s Department. The prices of the information have been changed into the figure of $1 \times 10^6$ to be noted at once.

For forecast modeling, the knowledge of computer programming with source code writing is required. For any non-coding staff who wants to create the computer program, it may be difficult for them to comprehend and learn to write the source code. Therefore, it would be a good idea to have a forecast modeling wizard for non-coding staff as the background of the source code writing knowledge is not much required as mentioned in the following sections.

3.1 Forecast Modeling Creating

The forecast modeling creating requires a computer program with source code writing to generate the forecasting model. However, source-code writing is difficult for non-coding staff. Therefore, the semi-finished program is an alternative for non-coding staff to create a very simple forecasting model by only applying a little knowledge of source code writing. Nowadays, many semi-finished programs are available. One of them is RapidMiner Studio which is a free program with limited functionality. The function is a free version that has been verified via the literature review (László & Ghous, 2020) that is sufficiently precise for an accurate analysis. Also, the program has been utilized in Thailand (Chaysiri & Ngauv, 2020).

3.1.1 RapidMiner Studio

This program is a tool for data analysis such as statistics, data correlation, or forecast modeling to predict prospects, whether the forecasting of sales, customer service usage, or numerical forecasting from the available data. Furthermore, RapidMiner Studio has been utilized for various forecast modeling by many users. For instance, Geetha (2014) used RapidMiner Studio for ANN creating to forecast the weather, which its accuracy was 81.78%. This is proved that the artificial neural network for the weather forecasting model has been enough accurate and could be used for other further weather forecasts. Moreover, Çelik (2017) used ANN to forecast the price of precious metals such as gold, silver, platinum, and palladium via RapidMiner Studio Software. The results also indicated that the prices of gold and palladium were highly predictable while silver and platinum were less effective. In summary, artificial neural networks were useful for the price of precious metals forecasting.

With the easy-to-use functionality of the RapidMiner Studio Software, it is not necessary to be an expert in source code writing in creating an efficiently forecasting model, but only basic knowledge is sufficient. Therefore, in creating a construction cost forecast modeling, the ANN has been generated in the RapidMiner Studio Software as described in the next section.
3.1.2 Forecast Modeling

For modeling, the tools which have been provided by RapidMiner Studio must be selected following the user’s needs. The construction cost forecasting model which has been using the ANN is shown in Figure 3.

![Image](http://TuEngr.com)

**Figure 3:** The Process of Forecast Modeling.

According to Figure 3, it is the process of a forecasting model which has been using the ANN. In this process, ‘Read Excel’ is a tool for reading the Excel format data set. The tool can change the type of reading according to the type of available data files. Also, ‘Select Attributes’ is a tool to define the attributes of the information, and ‘Set Role’ is a tool for forecasting variables or defining the kind of the desired variable. For this forecast, variable Y has been defined as the Label for the forecast variable. Also, ‘Split Data’ has been used for determining the proportion, Train and Test of the forecast model which the proportion of Train and Test is shared at 0.8 and 0.2, respectively. ‘Neural Net’ is the process of the neural network forecasting modeling for a defined Train proportion. Normally, the Hidden Layer would define the Activation Function as the Sigmoid function, while the Output Layer is the Sigmoid function itself if the Learning Data describes a classification task. Furthermore, it would also become the Linear Function if the Learning Data described a Numerical Regression Task. ‘Apply Model’ performs a forecast using ANN and ‘Performance’ is the calculation of the error rate of the forecast model. Moreover, Çelik (2017) used the same forecast modeling process to create a precious metal forecasting model.

Regarding the number of layers of the hidden layer in the ANN modeling process, the appropriate number of layers is still questionable. According to the rules-of-thumb for ANN modeling, Ranjan (2019) presented there should be 2 Hidden Layers, the number of nodes in the first Hidden Layer must be half of the number of the independent variables, half of the number of nodes in the next Hidden Layer must decrease from the first Hidden Layer, and the number of nodes in the Output should be equal to the number of the independent variables.

From the modeling process, the results were satisfactorily forecasted by the lowest value of the Root Mean Square Error described in the next section.

4 Results of the Model

After obtaining a forecasting model using ANN, the least RMSE must be achieved by an experiment that changed the number of nodes of each layer in the Hidden Layer. It was starting
with the reference of the rules-of-thumb, it could be concluded that using two layers of the Hidden Layer results is the best model that presents the least RMSE value. Therefore, only the two-layer test method will be shown for the testing of the node number of each layer.

### 4.1 The Test of the Numbers of Nodes in the Hidden Layer

With the suitable number of nodes, the result would be the least RMSE of the total number in the test. It was started with the rules-of-thumb mentioned that the number of nodes in the first Hidden Layer was approximately half the number of independent variables. This meant that the number of nodes in the first layer of the 11 independent variables must be 5 and 6, and the number of nodes in the next Hidden Layer should be halved from the hidden layer, 2 and 3 Nodes, respectively.

Nevertheless, other forms of node testing were not abandoned. The test of the first layer was started from 1 node to 12 nodes and in the next layer, the test was started from 1 node to 10 nodes by 120 times. The value of RMSE has been presented in the next section of the test results.

### 4.2 Test Results

Regarding the test results, the RMSE values of each node cannot describe the relevance of the information as presented in Table 2.

<table>
<thead>
<tr>
<th>Node</th>
<th>Hidden 1</th>
<th>Hidden 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.452</td>
<td>0.517</td>
</tr>
<tr>
<td>2</td>
<td>0.497</td>
<td>0.489</td>
</tr>
<tr>
<td>3</td>
<td>0.439</td>
<td>0.525</td>
</tr>
<tr>
<td>4</td>
<td>0.445</td>
<td>0.480</td>
</tr>
<tr>
<td>5</td>
<td>0.474</td>
<td>0.494</td>
</tr>
<tr>
<td>6</td>
<td>0.430</td>
<td>0.371</td>
</tr>
<tr>
<td>7</td>
<td>0.404</td>
<td>0.430</td>
</tr>
<tr>
<td>8</td>
<td>0.419</td>
<td>0.446</td>
</tr>
<tr>
<td>9</td>
<td>0.460</td>
<td>0.411</td>
</tr>
<tr>
<td>10</td>
<td>0.412</td>
<td>0.374</td>
</tr>
<tr>
<td>11</td>
<td>0.456</td>
<td>0.435</td>
</tr>
<tr>
<td>12</td>
<td>0.448</td>
<td>0.486</td>
</tr>
</tbody>
</table>

As per the results in Table 2, the test was started by applying the rules-of-thumb suggested by Ranjan (2019) as by 120 times of trial-and-error. It has been found that the number of nodes in the first Layer was 10 and 8 nodes in Layer 2 and the obtained value of the least RMSE was 0.331 million Baht, as the pattern of ANN presented in Figure 4.
To ensure the accuracy of the forecasting model, the new 10 sets of data were used for the test of the forecast model by adding the ANN model process as shown in Figure 5.

According to the test of the new 10 sets of data, the result showed that the value of $R^2$ was 0.914 that has been more accurate than the other rough estimation as presented in Figure 6.

5 Conclusion

According to the fact that the rough estimation can cause more tolerances, the estimation method can also result in damages by too-low bidding cost and eventual losses. Small companies
usually are small companies with no engineering staff or insufficient staff. They also have not enough time to have a detailed estimate of the price. Due to the above mentioned, the small companies often use the rough estimation of the construction cost to estimate the price to participate in the bidding process which is resulting in losses due to the tolerances of the estimation method. However, the control of tolerances in construction cost estimation is possible by using the ANN method as a tool. From the presented forecasting model, the value of the RMSE is ± 0.33 million Baht and was found the accuracy of the forecasting is 91.4%. Also, the comparison with the rough estimation by Wangniwetkul (2009), the value of tolerances may reach up to 50%. Therefore, the ANN method is a better accurate construction cost estimation technique. Furthermore, the selected tools for creating the ANN in this study do not require source-code writing skills or any complex computer languages. They are suitable for any civil engineers, architects, or those in charge of the pricing processes to use since they do not require much additional learning. This will be more widely useful in the type of behavior introduction than the programmatic modeling as is consistent with László & Ghous (2020). Moreover, it is also a testament to the potential in applying data analysis principles to Thai civil engineering and construction management. Besides, they can create work efficiency and further working opportunities in this field.

6 References
Chakan, C. (2010). Cost estimation for condominium construction using the Artificial Neural Network theory. (in Thai), Kasetsart University,
Economic Intelligence Center (EIC) Siam Commercial Bank Public Company Limited. (2020). EIC Evaluation of the Thai construction industry How should I be prepared during the COVID-19. Economic Intelligence Center (EIC), 7.


Sithikorn Sithikankun is a student at the Department of Civil Engineering Faculty of Engineering in Construction Engineering and Management, Chiang Mai University, Thailand. He got a Bachelor’s degree in Civil Engineering from Rajamangala University of Technology Lanna, Thailand. His research focuses on Construction Management.

Dr. Damrongsak Rinchumphu is a Researcher at the Center of Excellence for Natural Disaster Management (CENDIM) and an assistant professor at the Department of Civil Engineering, Faculty of Engineering, Chiang Mai University, Thailand. He has a Doctoral degree in Civil Engineering and Built Environment from Queensland University of Technology, Australia. His research focuses on Sustainability and Technology in Civil and Construction Management.

Dr. Chinnapat Buachart is a Researcher at the Center of Excellence for Natural Disaster Management (CENDIM), and an Assistant Professor at the Department of Civil Engineering, Faculty of Engineering, Chiang Mai University, Thailand. He received his Ph.D. in Structural Engineering from the Asian Institute of Technology (AIT), Thailand. His research focuses on Computational Methods in Structural Mechanics.

Dr. Eakasit Pacharawongsakda is a Lecturer at the Big Data Engineering Program, College of Innovative Technology and Engineering, Dhurakij Pundit University, Thailand. He received his Ph.D. in Computer Science from Sirindhorn International Institute of Technology, Thammasat University Thailand. His research focuses on Data Mining, Machine Learning, and Big Data.