A Prediction of Cement and Iron Product Price Index Based on Machine Learning Algorithm by Using Extreme Gradient Boosting (XGBoost)

Vichapon Sudsanguan[1] and Vachara Peansupap[2*]

[1,2] Department of Civil Engineering, Faculty of Engineering, Chulalongkorn University, Bangkok, THAILAND
**Corresponding author; E-mail address: vachara.p@chula.ac.th*

Abstract

Construction cost significantly impacts the analysis of feasibility, budget planning, and project success. For contractors, construction cost can be used to determine the bid price and profit. Construction Material Price Index (CMI) is a useful indicator for estimating costs and planning projects, as changes in material price can impact a contractor's ability to control construction costs. This research aims to apply an Extreme Gradient Boosting (XGBoost) for developing time series forecasting model of construction material prices in Thailand. The scope of this paper is focused on cement and iron products price index. The study collected eight influencing factors over a 276-month period from January 2000 to December 2022. The data was divided into three sections: model training, model validation, and model testing. The study evaluated the model using the walk-forward optimization technique. The evaluation of forecasting accuracy was done using Root Mean Squared Error (RMSE) and Mean Absolute Percentage Error (MAPE). The research results showed that the XGBoost models with multivariate and rolling window technique outperformed the univariate models and models without the rolling window technique in terms of RMSE and MAPE on both material price index. The developed models in this study offer an approach for forecasting construction material price index, providing an accurate short-term forecast of the material price index. The developed models can serve as a tool for stakeholders in the construction industry to forecast construction material price index.

Keywords: Construction Material Price Index (CMI), Machine learning models, Forecasting, Extreme Gradient Boosting

1. Introduction

Project construction cost is the total cost of building a construction project, including the cost of materials, labor, and other expenses. Therefore, it is an important factor to consider when planning and managing a construction project because it can have a significant impact on the project's feasibility, budget, and overall success [8]. Knowing the construction cost of a project is critical in determining whether the project is feasible and whether it makes financial sense to proceed [9]. If the construction costs are too high, the project may not be able to be completed within the budget, and the project may need to be redesigned or scaled back [16]. So, estimated project construction cost can help to determine how much money will be required to complete the project and how that money should be distributed among various tasks and materials [2]. Furthermore, Project cost estimation is useful in ensuring that projects stay on schedule and use resources efficiently [12]. Project managers can monitor cost overruns or areas for cost savings by comparing the actual construction cost to the budgeted cost [3].

In Thailand, the construction material price index (CMI) is an index that measures the change in the average cost of construction materials including lumber and wood products, cement, concrete ingredient, iron products, tiles, paints, sanitary ware, electrical and plumbing, and others in relation to a specific period (base year) which now uses the year 2015 as the base year for measuring the index. The construction material price index is a useful tool for understanding how the cost of materials used in the construction industry is changing, and it can be used to estimate the cost of materials for a construction project to support contractors in project planning by providing an indication of the expected cost of materials. For example, if a contractor is planning a construction project and needs to estimate the cost of materials, they can use the construction material price index to determine how much they can expect to pay for the materials they will need. So, the construction material price index can be used as cost estimation tool in the construction industry, as it can assist project manager in understanding current and projected material costs and making informed decisions about their construction projects.

One of the most important factors influencing the overall cost of construction is the cost of construction materials [18,20]. The uncertainty of material costs may have an impact on contractors' ability to control construction project costs. To control project material costs, consider the quantity and unit price of the material. Quantity may be related to material order and utilization, and several issues must be controlled, such as material quantity order accuracy, material utilization control, material waste, and so on. Whereas the price of material may be affected by price uncertainty. In order to manage and control material prices, Contractor may need to know the trend of material project prices so that construction companies can manage project costs. The goal of this research is to examine machine learning models for forecasting the Construction Materials Price Index (CMI) with additional influencing variables obtained from literature such as the producer price index, consumer price index, inflation rate, interest rate, and so on. So that contractor can use this forecasted information as a tool to analyze and evaluate construction budgets, or to analyze the trend of construction material prices, or through the bidding process, contractors can use this forecasted information as a tool to better estimate construction costs.

2. Literature Review

2.1 Material Price Influencing Factors

The cost of building materials plays a critical role in the overall cost of a construction project [7,20], and various factors can influence the price of materials. Several studies have been conducted to investigate the factors that affect the prices of construction materials in different contexts and locations. A review of the literature on this topic shows that various factors can influence the prices of construction materials.

Several studies found that inflation, exchange rate, interest rate, money supply and importation, as well as factors like crude oil prices, energy costs, local taxes and charges, fuel and power supply costs, high running costs, high raw material prices, transportation costs, and labor costs have a major impact on construction material prices [6,11,19].

Factors influencing construction materials prices can vary depending on location, such as in Vietnam, where the Consumer Price Index (CPI) is the most important factor influencing the construction price index, followed by Gross Domestic Product (GDP) and basic interest rate, foreign exchange rate, total export and import, and so on [17]. Other studies found a correlation between construction material prices and the inflation rate [15] and also showed that exchange rate, monetary policy, importation, interest rate, inflation rate, demand, and energy cost were found to be the major factors influencing construction material prices in some regions, like northern Malaysia [11]

This literature review demonstrates that a wide range of factors, including economic indicators, government regulations, logistics, and transportation costs, can influence construction material prices. To effectively manage costs, project managers should consider a range of factors when estimating and managing construction material costs.

2.2 Forecasting of Construction Material Price Index

Forecasting construction material price index is a critical task for the construction industry. Several studies have been conducted to predict these indices using a variety of methods and models, such as data from material suppliers [5], artificial neural network (ANN), and statistical models such as the Simple Average, Exponential Smoothing, and Simple Regression Analysis [13].

One of the common approaches used in the past research is the use of Time Series analysis, which is a statistical method used to analyze time series data, such as prices of construction materials over time. The time series models like ARIMA [25,26] have been found to be useful in predicting the future values of construction material price index. ANN have also been found to be useful in predicting the construction material price index, specifically in term of accuracy [13].

Multiple linear regression analysis [4] is also another approach that have been used in some studies, by including various variable attributes such as year, month, material prices, moving averages, secular trend, seasonal index, irregular movement, cyclical movement, and next month's material prices, these attributes can help to predict the price of construction materials.

However, previous research has some limitations. For example, some previous studies forecasted construction material price indices only in one category, all commodities, which may not accurately reflect changes in construction material price indices in other categories. Furthermore, some models, such as ARIMA, are designed to be used with univariate time series data, which means that the data was predicted using only one variable, the construction material price index, and no external variables, such as macroeconomic variables, were used in the model.

2.3 Forecasting with Machine Learning Algorithm

Forecasting with machine learning algorithms has gained a significant interest in recent years [14], as these algorithms have the ability to automatically learn patterns in historical data and make predictions about future events, making it a powerful tool for many fields such as businesses, finance, and economics.

Machine learning algorithms, such as XGBoost, ANN, and LSSVM, have been shown in previous research to be powerful tools for forecasting various types of data, including building material prices [20], housing prices [21], electricity load [1], sales volume [27], and exchange rates [10]. These researches revealed that these algorithms can analyze patterns in historical data and predict future events, providing useful insights for businesses, finance, and economics [24].

The XGBoost algorithm has been found to be particularly effective for forecasting, as it has been shown to perform well in a number of different forecasting tasks. [27] and [10] used XGBoost for forecasting sales volume time series and exchange rate respectively. They found that XGBoost performed the best in both datasets, and it was concluded that XGBoost outperforms the other models because it explicitly adds a regularization term to the objective function, which can help control the model's complexity, prevent overfitting, and improve the model's generalization ability. [21] also used the XGBoost algorithm to forecast second-hand house prices and compared the results to multivariate linear regression and decision tree models. They found that the XGBoost model was the most effective in predicting second-hand housing prices in Chengdu. Moreover, [1] predicted electricity load by calculating the importance of each feature dataset using the XGBoost feature selection technique and found that the XGBoost performs extremely well for time series prediction with efficient computing time and memory resource usage.

Artificial Neural Networks (ANNs) have also been found to be useful in forecasting building material prices [20], in terms of accuracy. Least Squares Support Vector Machine (LSSVM) with Improved Particle Swarm Optimization (IPSO) has been found to produce the most accurate predicted results when compared to other models such as neural networks and multiple linear regression [23].

Since there is no research that uses XGBoost to forecast the construction material price index, and based on a review of the literature, XGBoost outperforms the other model in forecasting time-series in different areas of study. Its ability to handle multivariate time-series problems and utilizing built-in regularization techniques that can help prevent overfitting and improve model performance. As a result, the researcher aims to develop an XGBoost model to forecast Thailand's Construction Material Price Index.

3. Research Methodology

This section will be discussed about the research methodology with the following topic: data collection, XGBoost, data preparation, hyperparameters optimization, and evaluation of forecasting accuracy, The main objective of this section is to establish a robust modelling process that can accurately predict the price index of construction material, considering the forecast for 1, 3, 6, 9, and 12 months in advance. So that this forecasted information can provide contractors with greater flexibility and choice in selecting the model that best suits their needs. This can help contractors to obtain more accurate forecasts that are better aligned with their specific project timelines. For example, if a contractor is planning a project that will start in 6 months, they can use the 6-month forecasting model to gain the expected prices of the different materials required for their project. This can help them to plan their procurement and pricing strategies more effectively.

3.1 Data Collection

This study will collect all eight data, construction material price index, producer price index, consumer price index, inflation rate, interest rate, exchange rate, gross domestic product, and crude oil price, over a 276-month period beginning in January 2000 and ending in December 2022. These factors have been selected from the literature review based on their relevance to the construction industry and their potential impact on material prices. All factors were collected using monthly data, except for

GDP which was collected quarterly. As a result, the GDP data will contain the same value for each quarter, which was used for each month within that quarter to represent the average value for that period. The proportion of data can vary depending on the size of the dataset. The training set should contain enough data points for the model to learn the underlying patterns in the data. The validation set size can affect the model's ability to generalize to new data. A larger validation set may help identify overfitting, but it may also reduce the amount of data available for training the model and the testing set size should be large enough to provide a reliable estimate of the model's performance on new, unseen data. So, the data in this research is divided into three sections: 204 months for model training (74 percent training set), 36 months for model validation (13 percent validation set), and 36 months for model testing (13 percent testing set). The training set is used to train the model, while the validation set is used to tune the hyperparameters of the model, and the testing set is used to evaluate the final performance of the model. Table 1 depicts the time periods and sizes of the training, validation, and testing datasets used in this study, with data collected between January 2000 and December 2022.

Table 1 Dataset Periods and Sizes for Training, Validation, and Testing Sets

Type of Dataset	Periods		Number of data
	Start	End	
Training	Jan-2000	Dec-2016	204
Validation	Jan-2017	Dec-2019	36
Testing	Jan-2020	Dec-2022	36

3.2 XGBoost

XGBoost, also known as Extreme Gradient Boosting, is a popular machine learning algorithm that has gained wide usage across various fields. This algorithm offers unique features that improve the tree boosting approach, making it possible to process almost all data types rapidly and with high accuracy. Because of its exceptional capabilities, XGBoost is an effective tool for creating forecasting models such as regression and classification processes that target specific datasets. It is particularly useful for large datasets with many attributes and classifications [22].

3.3 Data Preparation

Because XGBoost develops gradient boosting for classification and regression problems. To use XGBoost for time series forecasting, the input dataset must first be turned into a supervised learning problem.

A supervised learning problem consists of input data and output data, with the purpose of teaching an algorithm how to predict the output data based on the input data.

The "sliding window technique" is used to convert time series into supervised learning problems. This approach uses previous time step data as input variables and following time step data as output variables.

After the dataset has been prepared, the dataset is then divided into training, validation and testing set and using walk-forward optimization so called “rolling window technique” to fit and evaluate the model in order to have an accurate forecasting result. The rolling window technique is used to determine the window width, which is the desired length in each sub-sequence. So, in each sub-sequence, there will be data for the model to train or learn on, as well as output data, which is the data that the model predicted.

3.4 Hyperparameters Optimization

The objective of hyperparameter optimization is to minimize the difference between the predicted and known values so that the machine learning algorithm can provide the best forecasting results for the machine learning problem. Therefore, in order to optimize a hyperparameter, find a combination of each hyperparameter that provides the best result and returns accurate predicted data to the model. The hyperparameters include max_depth, which represents the maximum number of nodes in a tree in the model, learning_rate, which represents the rate at which the model updates the weights of the features in each iteration, n_estimators, which represents the number of trees used in the model, min_child_weight, which represents the minimum number of samples required to be at a leaf node, subsample, which represents the fraction of the training instances used for each tree, colsample_bytree, which represents the fraction of the features used for each tree, gamma, which represents the minimum reduction in the loss required to split a node, reg_alpha, which represents the L1 regularization term used to control the complexity of the model, and reg_lambda, which represents the L2 regularization term used to control the complexity of the model. Each hyperparameter in the dictionary has a set of possible values that the model can choose from during the hyperparameter tuning process. Fig. 1 illustrates the hyperparameter space that will be searched to tune the XGBoost model for optimal performance. The search space includes a range of values for hyperparameters. These hyperparameters are crucial in determining the model's performance and can greatly impact its ability to accurately forecast the price of construction materials. By exploring a wide range of hyperparameters, the model can be optimized to achieve the best possible performance.

```
# Define the hyperparameter space to search
param_dist = {    'max_depth': randint(1, 10),
    'learning_rate': uniform(0.01, 0.99),
    'n_estimators': randint(50, 250),
    'min_child_weight': randint(1, 10),
    'gamma': uniform(0.0, 1.0),
    'subsample': uniform(0.1, 0.9),
    'colsample_bytree': uniform(0.01, 0.99),
    'reg_alpha': uniform(0.0, 1.0),
    'reg_lambda': uniform(0.0, 1.0)
}
```

Fig. 1 Hyperparameter Space for XGBoost Model

Then this research will perform random search on the XGBoost model. The objective of the random search is to find the best set of hyperparameters that results in the best model performance. The range of possible window width values for the rolling window will then define and the rolling window function will take the data and splits it into overlapping windows of the specified width. This is used to transform the data into a format that can be used for time series cross-validation. The results of each combination of hyperparameters and window width will then be stored in the results list. The time series cross-validation is performed using the TimeSeriesSplit function and the random search is performed using the RandomizedSearchCV function. To find the optimal set of hyperparameters for the XGBoost model, the RandomizedSearchCV function is applied on the training data. This function searches through a hyperparameter space and returns the best set of hyperparameters that yield the highest performance on the validation set. In this case, the goal is to minimize the RMSE, as this indicates the model's accuracy in predicting the future price of construction materials. The set of hyperparameters that produces the lowest RMSE on the validation set is then selected as the best combination. By selecting the optimal set of hyperparameters, the model can be tuned to provide the most accurate forecasts of construction material prices. The best combination of hyperparameters and window width will be stored and printed to determine which set of hyperparameters will generate the best results.

3.5 Evaluation of Forecasting Accuracy

In this part, the best hyperparameters found in the previous step are used to train the XGBoost model on the entire training set, which includes both the original training set and the validation set. The trained model is then used to make predictions on the training set, validation set, and test. Finally, the performance of the model is evaluated on the test set using two evaluation metrics: Root Mean Squared Error (RMSE) and Mean Absolute Percentage Error (MAPE). RMSE is a measure of the difference between the predicted values and the actual values. MAPE is a measure of the relative error between the predicted values and the actual values and then print the RMSE and MAPE values for the test sets to assess the model's performance on each dataset.

4. Research Findings and Results

This section will present the forecasting results of the construction material price index. We evaluate the models' performance for cement and iron products price index over various time horizons (e.g., 1, 3, 6, 9, and 12 months in advance). The models were developed using the best combinations of hyperparameters and size of window width and then report on the resulting root mean squared error (RMSE) and mean absolute percentage error (MAPE). RMSE is the square root of the mean of the square of all errors and MAPE is a percentage-based measure of the average of forecast errors.

4.1 Forecasting Result for Cement

Table 2 provides the best set of hyperparameters and window width size for the forecasting of cement for 1 month in advance. These hyperparameters were selected based on their performance on the validation set during the model development process, and the window width specifies the number of past observations used to predict future values.

Table 2 The Hyperparameters and Window Width for 1 Month in Advance of Cement

Hyperparameters	Value
Max_depth	7
Learning_rate	0.03093
N_estimators	209
Min_child_weight	4

Gamma	0.89035
subsample	0.62486
Colsample_bytree	0.55046
Reg_alpha	0.00001
Reg_lambda	0.05798
Window_width	5

The performance of the forecasting model for cement at 1 month in advance was evaluated using XGBoost. The model incorporated the cement price index as well as additional macro-economic factors and utilized a rolling window technique. Fig. 2 displays the comparison of the predicted and actual values of the index in the testing dataset. The model achieved an RMSE of 0.884 and a MAPE of 0.789%.

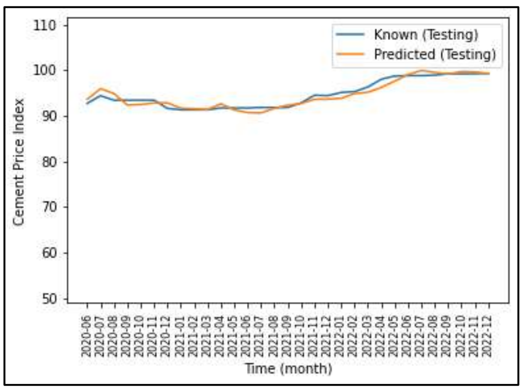

Fig. 2 Known VS Predicted Data for 1 Month in Advance of Cement

To evaluate the impact of the rolling window technique and the inclusion of additional variables, two additional models were tested. The first model used only the cement price index as a univariate variable, while the second model included additional variables but did not use the rolling window technique. The RMSE and MAPE results for the testing dataset were 1.637 and 1.372% for the univariate model and 2.282 and 1.858% for the model without rolling window technique, respectively. Fig. 3 and Fig. 4 depict the known versus predicted values for the univariate model and the model without rolling window technique, respectively.

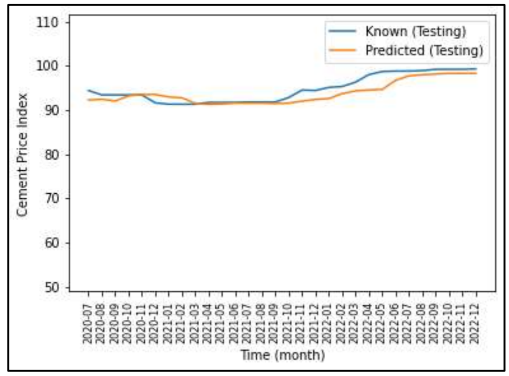

Fig. 3 Known vs. Predicted Values for Cement Univariate Model

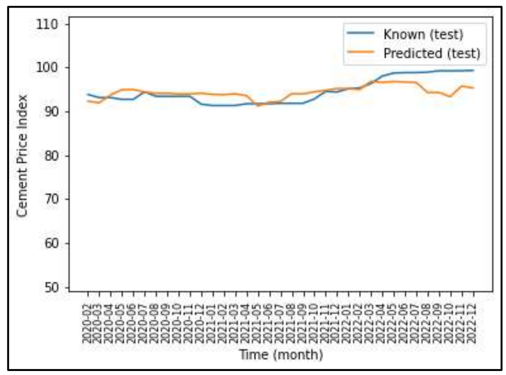

Fig. 4 Known vs. Predicted Values for Cement Model Without Rolling Window Technique

Then this research employed a multivariate model with a rolling window technique to predict the cement price index for the remaining 3, 6, 9, and 12 months in advance. Table 3 summarizes the performance of the forecasting model for the cement price index.

Table 3 Summary of Forecasting Model Performance for Cement Price Index

Cement Price Index		
Month in advance	RMSE	MAPE
1	0.884	0.789%
3	1.572	1.434%
6	1.906	1.754%
9	1.833	1.612%
12	2.719	2.578%

4.2 Forecasting Result for Iron Products

Table 4 provides the best set of hyperparameters and window width size for the forecasting of iron products for 1 month in advance. And Fig. 5 illustrates the performance of the forecasting model for iron products at 1 month in advance. The root mean squared error (RMSE) and mean absolute percentage error (MAPE) of this model is 5.031 and 2.914% respectively. Fig. 6 and Fig.7 compare the known and predicted values for the univariate model and the model without the rolling window technique applied to Iron products. The results show that the univariate model had RMSE and MAPE scores of 9.206 and 5.256%, respectively, on the testing dataset, while the model without the rolling window technique had scores of 6.774 and 4.253%, respectively.

Table 4 The Hyperparameters and Window Width for 1 Month in Advance of Iron Products

Hyperparameters	Value
Max_depth	3
Learning_rate	0.03531
N_estimators	175
Min_child_weight	6
Gamma	0
subsample	0.22804
Colsample_bytree	0.86849
Reg_alpha	0.00001
Reg_lambda	0
Window_width	6

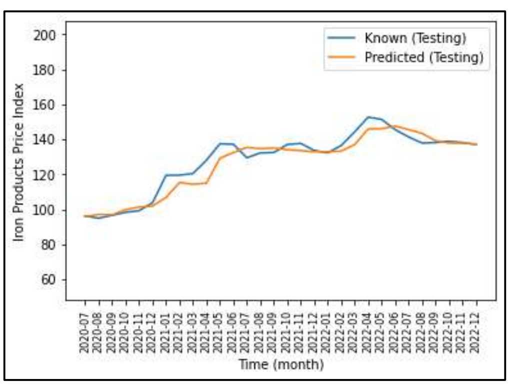

Fig. 5 Known VS Predicted Data for 1 Month in Advance of Iron Products

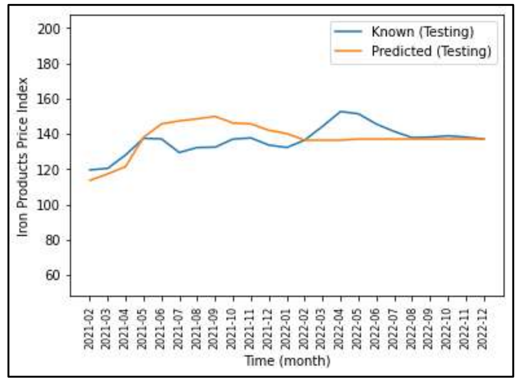

Fig. 6 Known vs. Predicted Values for Iron Products Univariate Model

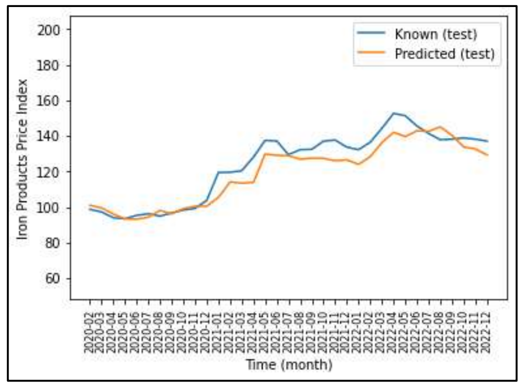

Fig. 7 Known vs. Predicted Values for Iron Products Model Without Rolling Window Technique

The results show that the model with a multivariate variable and a rolling window technique generated the most accurate predicted results for the iron products price index one month in advance. This technique was further developed to forecast the iron products price index for the next 3, 6, 9, and 12 months in advance. Table 5 summarizes the performance of the forecasting model for the iron products price index.

Table 5 Summary of Forecasting Model Performance for Iron Products Price Index

Iron Products		
Month in advance	RMSE	MAPE
1	5.031	2.914%
3	5.813	2.997%
6	5.690	3.253%
9	6.627	4.277%
12	8.16	4.602%

5. Research Discussion and Conclusion

The construction industry is a crucial contributor to the global economy. One of the most important factors influencing the overall cost of construction is the cost of construction materials such as cement and iron products, which can be affected by price fluctuations. Forecasting the prices of construction materials can help stakeholders make informed decisions and mitigate the risks associated with price volatility. This study aimed to forecast the price index of cement and iron products using machine learning algorithm. The study employed a dataset that included the price index of cement and iron products, along with additional macro-economic variables. The dataset was preprocessed and split into training, validation, and testing sets. XGBoost were developed to forecast the price index for cement and iron products for different time horizons, ranging from one month to twelve months in advance.

The results showed that the multivariate models with a rolling window technique outperformed the univariate models and models without the rolling window technique in terms of RMSE and MAPE. The performance of models for forecasting cement price index are better than the iron products price index. This may be because the raw materials required for iron production, such as iron ore, coal, and scrap metal, may be more volatile in price than the raw materials required for cement production, such as limestone and clay. So, changes in the availability and cost of these raw materials may have a significant impact on the price of iron products. In addition, iron production may require more processing steps than cement production, which can make it more expensive to produce. Thus, changes in the cost of energy, labor, and other production inputs can have a greater impact on the price of iron products than on the price of cement.

For cement, the model achieved an RMSE of 0.884 and a MAPE of 0.789% when predicting one month in advance. The RMSE and MAPE increased as the time horizon increased, with the model achieving a RMSE of 2.719 and a MAPE of 2.578% when predicting twelve months in advance. For iron products, the multivariate model with a rolling window technique achieved an RMSE of 5.031 and a MAPE of 2.914% when predicting one month in advance. Similar to cement, the RMSE and MAPE increased as the time horizon increased, with the model achieving a RMSE of 8.160 and a MAPE of 4.602% when predicting twelve months in advance. For longer time horizons, the accuracy of the models decreased, with higher RMSE and MAPE scores. However, the models still provided useful information for decision-makers in the construction industry, allowing them to make more informed decisions regarding purchasing and budgeting.

When comparing the cement price index predictions results with previous studies. We discovered that utilizing the xgboost model produced better RMSE values than [25]. Although the timeline of this study's testing dataset (January 2020 to December 2022) differs from prior research, which used data from July to December 2016. However, using the rolling window technique and additional macroeconomic factors resulted in accurate forecasting results. As a result, this study suggests that the xgboost model, when combined with these methodologies, provides good forecasting capabilities for the cement price index.

To sum up, the models developed in this study offer an approach for forecasting construction material price index. The stakeholders in the construction industry may incorporate these models into their planning processes to make informed decisions and improve project outcomes as the accurate forecasting of material prices can help reduce project costs and increase efficiency. Future research could continue research on development of forecasting models for construction material prices and explore the possibility of incorporating other relevant variables to improve the accuracy of the models further.

References

[1] Abbasi, R. A., Javaid, N., Ghuman, M. N. J., Khan, Z. A., Ur Rehman, S., & Amanullah. (2019). Short Term Load Forecasting Using XGBoost. In Web, *Artificial Intelligence and Network Applications* (pp. 1120-1131). https://doi.org/10.1007/978-3-030-15035-8_108

[2] Aibinu, A., & Jagboro, G. (2002). The effects of construction delays on project delivery in Nigerian construction industry. *International journal of project management*, 20(8), 593-599.

[3] Al-Jibouri, S. H. (2003). Monitoring systems and their effectiveness for project cost control in construction. *International journal of project management*, 21(2), 145-154.

[4] Bundasak, S. (2016). *Estimate Construction materials Price Using Regression Analysis with Variable Time Series*. Thesis, Kasetsart University, Thailand.

[5] Chooprasit, S. (2001). *A study of factors affecting construction material price index for the development of improvement guidelines*. Thesis, Chulalongkorn University, Thailand. http://cuir.car.chula.ac.th/handle/123456789/11644

[6] Danso, H., & Obeng-Ahenkora, N. K. (2018). Major determinants of prices increase of building materials on Ghanaian construction market. *Open Journal of Civil Engineering*, 8(2), 142-154.

[7] Donyavi, S., & Flanagan, R. (2009). The impact of effective material management on construction site performance for small and medium sized construction enterprises. Proceedings of the 25th *Annual ARCOM Conference*, Nottingham, UK,

[8] Elfahham, Y. (2019). Estimation and prediction of construction cost index using neural networks, time series, and regression. *Alexandria Engineering Journal*, 58. https://doi.org/10.1016/j.aej.2019.05.002

[9] Elfaki, A. O., & Alatawi, S. (2015). Representing the knowledge of public construction project cost estimator by using rule-based method. *Journal of Building Construction and Planning Research*, 3(04), 189.

[10] Islam, S. F. N., Sholahuddin, A., & Abdullah, A. S. (2021). Extreme gradient boosting (XGBoost) method in making forecasting application and analysis of USD exchange rates against rupiah. *Journal of Physics: Conference Series*, 1722(1), 012016. https://doi.org/10.1088/1742-6596/1722/1/012016

[11] Kamaruddeen, A. M. (2019). Factors Influencing the Price of Selected Building Materials in Northern Malaysia: Borneo. *Journal of Sciences and Technology*, 2(1), 7-12.

[12] Kim, G.-H., An, S.-H., & Kang, K.-I. (2004). Comparison of construction cost estimating models based on regression analysis, neural networks, and case-based reasoning. *Building and Environment*, 39(10), 1235-1242.

[13] Kumsiritrakul, N. (2003). *A prediction of construction cost indexes using neural network models*. Thesis, Chulalongkorn University, Thailand. http://cuir.car.chula.ac.th/handle/123456789/1439

[14] Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2018). Statistical and Machine Learning forecasting methods: Concerns and ways forward. PloS one, 13(3), e0194889.

[15] Musarat, M. A., Alaloul, W. S., Liew, M., Maqsoom, A., & Qureshi, A. H. (2020). Investigating the impact of inflation on building materials prices in construction industry. *Journal of Building Engineering*, 32, 101485.

[16] Nega, F. (2008). *Causes and effects of cost overrun on public building construction projects in Ethiopia*. Unpublished doctoral dissertation, Addis Ababa University, Ethiopia, 34.

[17] Nguyen, P. T., & Nguyen, Q. L. H. T. T. (2020). Critical Factors Affecting Construction Price Index: An Integrated Fuzzy Logic and Analytical Hierarchy Process. *The Journal of Asian Finance, Economics and Business*, 7(8), 197-204. https://doi.org/10.13106/jafeb.2020.vol7.no8.197

[18] Oghenekevwe, O., Olusola, O., & Chukwudi, U. S. (2014). An assessment of the impact of inflation on construction material prices in Nigeria. *PM World Journal*, 3(4), 1-22.

[19] Oladipo, F., & Oni, O. (2012). Review of selected macroeconomic factors impacting building material prices in developing countries–A case of Nigeria. *Ethiopian Journal of Environmental Studies and Management*, 5(2), 131–137-131–137.

[20] OuYang, H., Zhang, X., & Hu, C. (2013). Application Research on the Artificial Neural Network in the Building Materials Price Prediction. *The 19th International Conference on Industrial Engineering and Engineering Management*, Berlin, Heidelberg.

[21] Peng, Z., Huang, Q., & Han, Y. (2019, 18-20 Oct. 2019). Model Research on Forecast of Second-Hand House Price in Chengdu Based on XGboost Algorithm. *2019 IEEE 11th International Conference on Advanced Infocomm Technology (ICAIT)*,

[22] Shehadeh, A., Alshboul, O., Al Mamlook, R. E., & Hamedat, O. (2021). Machine learning models for predicting the residual value of heavy construction equipment: An evaluation of modified decision tree, LightGBM, and XGBoost regression. Automation in Construction, 129. https://doi.org/10.1016/j.autcon.2021.103827

[23] Tang, B.-q., Han, J., Guo, G.-f., Chen, Y., & Zhang, S. (2019). Building material prices forecasting based on least square support vector machine and improved particle swarm

optimization. *Architectural Engineering and Design Management*, 15(3), 196-212. https://doi.org/10.1080/17452007.2018.1556577

[24] Tang, Y., Song, Z., Zhu, Y., Yuan, H., Hou, M., Ji, J., Tang, C., & Li, J. (2022). A survey on machine learning models for financial time series forecasting. *Neurocomputing*, 512, 363-380. https://doi.org/https://doi.org/10.1016/j.neucom.2022.09.003

[25] Trakulgoson. R. (2018). *Cement price index forecasting using ARIMA and ARIMA X model*. Thesis, Sukhothai Thammathirat University, Thailand. Bibliography: Sheet [57]-59.

[26] Wongkhrua, C. (2004). *Building-material Price Index Forecasting by ARIMA Method*. Thesis, Chiang Mai University, Thailand. 10.14457/CMU.res.2004.182

[27] Zhang, L., Bian, W., Qu, W., Tuo, L., & Wang, Y. (2021). Time series forecast of sales volume based on XGBoost. *Journal of Physics: Conference Series*, 1873(1), 012067. https://doi.org/10.1088/1742-6596/1873/1/012067