DOI: https://doi.org/10.14256/JCE.4029.2024

Građevinar 1/2025

Primljen / Received: 26.4.2024.
Ispravljen / Corrected: 9.10.2024.
Prihvaćen / Accepted: 11.9.2024.
Dostupno online / Available online: 10.2.2025.

Novel meta-ensemble modelling approach and comparison of machine-learning models for rebar price estimation

Authors:

Assist.Prof. **Sahin Tolga Guvel**, PhD. CE
Osmaniye Korkut Ata University, Turkey
Department of Civil Engineering
sahintolgaguvel@osmaniye.edu.tr

Assist.Prof. **Abdulkadir Budak**, PhD. CE
Osmaniye Korkut Ata University, Turkey
Department of Civil Engineering
abudak@osmaniye.edu.tr

Ibrahim Karataş, PhD. CE
Osmaniye Korkut Ata University, Turkey
Department of Civil Engineering
ibrahimkaratas@osmaniye.edu.tr
Corresponding author

Research Paper

Sahin Tolga Guvel, Abdulkadir Budak, Ibrahim Karataş

Novel meta-ensemble modelling approach and comparison of machine-learning models for rebar price estimation

The early determination of costs in construction projects is crucial for the planning of expenses throughout each investment stage. Making realistic cost calculations is an effective way of preventing cost overruns that may occur in later stages. Rebar price prediction by considering economic indicators significantly affects investment costs and decisions. Therefore, in this study, using historical data for rebar construction material and economic indicators, nine machine-learning algorithms were used to determine the estimated rebar price for 1-, 3-, 6-, 9-, and 12-month lags. The voting meta-ensemble machine-learning algorithm exhibited the best performance for all lag periods investigated. The most successful estimate was obtained for a 3-month lag period. The mean absolute percentage error (MAPE) and coefficient of determination (R2) values for the rebar price estimation during this period were 3.79 % and 95.51 %, respectively.

Key words:

rebar price estimation, construction management, production planning, meta-ensemble, machine learning

Prethodno priopćenje

Sahin Tolga Guvel, Abdulkadir Budak, Ibrahim Karataş

Novi pristup modeliranju metaansambla i usporedba modela strojnog učenja za procjenu cijene armature

Rano utvrđivanje troškova u građevinskim projektima ključno je za planiranje troškova u svakoj fazi ulaganja. Realni izračun troškova učinkovit je način sprječavanja prekoračenja troškova do kojeg može doći u kasnijim fazama. Predviđanje cijene armaturnih šipki, uzimajući u obzir ekonomske pokazatelje, znatno utječe na troškove ulaganja i na odluke. Zato je u ovome istraživanju na temelju povijesnih podataka o materijalu za armiračke radove i ekonomskim pokazateljima upotrijebljeno devet algoritama strojnog učenja za dobivanje procijenjene cijene armaturne šipke za zakašnjenja od jedan, tri, šest, devet i dvanaest mjeseci. Algoritam za strojno učenje glasačkog metaansambla pokazao je najbolju izvedbu za sva istražena razdoblja zakašnjenja. Najuspješnija procjena dobivena je za zakašnjenje od tri mjeseca. Srednja apsolutna postotna pogreška (MAPE) i koeficijent determinacije (R2) za procjenu cijene armature tijekom tog razdoblja iznosile su 3,79 % odnosno 95,51 %.

Ključne riječi:

procjena cijene armature, upravljanje izgradnjom, planiranje proizvodnje, metaansambl, strojno učenje

1. Introduction

Feasibility studies are crucial in the construction sector for making investment decisions. However, determining whether investment decisions are made correctly remains challenging [1]. Therefore, investment cost is one of the most prominent parameters, particularly when considering the parameters that affect investment decisions [2]. However, investment costs change with technological developments and evolving employer expectations. Therefore, the most accurate possible investment cost estimation [3] affects the project's cost overruns, its timely completion, and the ability to achieve the quality desired by the employer [4].

Various tasks and procedures can affect the overall cost of building construction. One of the main parameters affecting production is the effect of the initial estimated cost of building materials on price fluctuation costs during a project [4]. Among construction materials, iron [4, 5] and cement [6] are the materials that most affect buildings in terms of cost [7]. This is because the change in their unit costs has a much more dynamic structure than other building materials, along with a high total cost.

Previous studies have been mainly based on different focal points related to the calculation of approximate construction costs [8]. Because many factors affect construction costs [1, 2], the use of economic indicators is now widely used. As economic indicators have a dynamic and changing structure, especially in developing countries, their impact on all resources constituting construction costs is inevitable. However, among the studies on the subject, those conducted considering structural features during the feasibility stage for calculating the upfront cost are extensive. For example, the Cost Index [8], unit price analysis [3] values, and project/material properties [1, 5] can be used to calculate construction costs. However, there are few studies on future building cost estimation based on building cost variability under the influence of economic indicators [4, 8–10]. Additionally, few studies have examined changes in the cost of building materials using economic indicators [4, 9, 11].

In this context, there is a knowledge gap in this field regarding changing the cost of building materials under the influence of economic indicators. Very few studies have been conducted on this subject, especially in the field of rebars, which significantly affect the cost of building materials. Therefore, this study examined the change in rebar prices under the influence of economic indicators.

In the construction sector, different analysis methods, such as basic statistical analysis, regression analysis, and artificial intelligence (AI) can be used for cost estimation [10, 12, 13]. However, in studies examining the effects of price changes over time using economic indicators, successful results have been obtained using time-series analysis, regression analysis, and AI applications [2, 7, 14]. Recently, not only in the construction sector but also in other sectors, AI applications have been shown to produce more successful results [6, 15-17]. Today, economic effects such as inflation cause uncertainties in construction sector costs. There is a great need for accurate cost forecasting of building materials such as iron, which are quickly affected by economic indicators, especially in the pre-investment and purchase stages. Accordingly, analysing the change in rebar prices under the influence of economic indicators with AI applications and estimating the closest cost during the feasibility stage would fill the aforementioned gap in knowledge. Therefore, the aims of this study were as follows:

- Estimating future iron prices using past rebar prices and economic indicator data of the relevant dates;
- Perform a comparison of machine-learning models used for iron price prediction, and select a model with a high predictive value; and
- Propose a new meta-ensemble model for iron price forecasting by combining existing computer-science-based models.

2. Literature review

Cost estimations are highly effective in the construction sector, especially in terms of investment decisions during the feasibility stage. In construction production, it is essential that future rebar price estimates, which are among the materials that significantly affect the total construction cost, be as close to the truth as possible. While analysing previous studies on the subject, it was found that the factors used in cost estimation varied [4]. For future building material predictions, Ou et al. [7] used time-series analysis to estimate the prices of iron ore and coking coal produced in a steel plant. In the models proposed by Faghih and Kashani [10] for future asphalt, steel, and cement price predictions for building materials in the United States (USA), building permits, consumer price index, construction spending, number of employees in the construction project, employment rate, gross domestic product, hourly earnings of construction labour, housing starts, industrial gas price, iron ore price, industrial producer price index, personnel income, and West Texas Intermediate parameters were used. Elfahham [8] carried out time-series analyses using the prices of brick, steel, cement, sand, and gravel to estimate the construction cost index in Egypt. Shiha et al. [9] also aimed to assess steel and cement prices in Egypt using macroeconomic indicators such as the consumer price Index, foreign reserves, gross domestic product, inflation rate, lending rate, money supply, producer price index, unemployment rate, and the US dollar exchange rate. Additionally, they performed estimations for 1-, 3-, and 6-month lags when selecting the lag time between macroeconomic indicators and output material prices. Mir et al. [4] developed a model to establish estimated ranges of asphalt and steel prices in the USA. The consumer price index, housing starts, and global iron ore price parameters were used to estimate steel prices.

In the construction sector, the use of AI applications in tasks requiring forecasting has increased in recent years [4, 15, 18, 19]. This is because the use of AI applications is easy and the level of success is higher than that of other estimation methods [4, 15]. Considering previous studies on cost estimation in building production, many different methods have been used for AI applications.

Ou et al. [7] applied the Gray extreme learning machine (GELM) forecasting model integrated with the Gray relation analysis (GRA) and extreme learning machine (ELM) methodologies. They also compared the proposed model with the autoregressive and integrated moving average (ARIMA) and generalised autoregressive conditional heteroskedasticity (GARCH) models. Using the proposed model, the estimated mean squared error (MSE) for iron price was 0.0097. The predictive success of the proposed combined model was higher than that of other models. Faghih and Kashani [10] presented a vector error correction (VEC) model to estimate short- and long-term construction material prices by characterising the relationship between economic parameters and material price estimations. For steel price estimation, the univariate VEC model has a higher accuracy than the other models, with a mean absolute percentage error (MAPE) value of 56.88 %, with 3-month-lag prediction values.

Elfahham [8] proposed a formula using neural networks, linear regression, and autoregressive time series to estimate the construction cost index for concrete structures based on historical construction cost records. The historical prices of brick, steel, cement, sand, and gravel were used as the basis for calculating the construction cost index. The average of the calculated absolute errors was as follows: 8.3 for the neural network, 17.5 for the regression method, which was the least accurate, and 3.5 for the time-series method. Strengthening the model using the inflation rate would be more effective.

Shiha et. al. [9] developed three rebar forecasting models. Model 1 was created using an Excel spreadsheet that utilised a genetic algorithm to minimise errors between neural network predictions and actual prices. Model 2 was developed using an Excel add-in called NeuralTools, and Model 3 was developed using the Python programming language in Spyder version 3.6. The proposed model for steel reinforcement price prediction was identified as Model 3, with MAPE values of 7.0 % and 4.3 % for the training and test sets, respectively. This model was chosen because it captured monthly fluctuations more effectively than Models 1 and 2.

By contrast, Mir et al. [4] proposed an artificial neural network (ANN)-based method for measuring uncertainties by creating prediction intervals. The optimal lower-upper bound estimation (optimal LUBE) method was adopted to train the ANN to generate intervals directly. The proposed method was used to estimate the construction material prices for asphalt and steel in the USA. Based on the results, they concluded that the optimal LUBE method yielded more realistic results than other methods for material price estimation.

Chan et al. [15] emphasised that machine-learning applications are effective in building material estimation and design; however, professionalism should be used in their application. Xu and Zhang [12] use Gaussian process regression models to forecast the daily price indices of steel products in the Chinese market. They employed cross-validation and Bayesian optimisation on various kernels and basic functions. Their goal was to provide forecasts of steel product prices, which hold significant economic importance for China as the largest steel consumer and producer worldwide.

Mi et al. [14] gathered rebar futures data from 2009 to 2020 and developed a VMD-EEMD-LSTM model to forecast rebar futures prices for the next 14 trading days. Dai et al. [13] collected and analysed rebar prices in the Guangdong Province of China during the first half of 2023 to understand the time-series characteristics of the rebar price composition. According to the literature, forecasting rebar prices in advance is crucial for the corresponding country's economy. Consequently, this study aimed to utilise machine-learning algorithms to predict rebar prices based on economic indicators.

3. Research methodology

This study aimed to estimate future rebar prices using past rebar prices and economic indicator data of relevant dates to compare the machine-learning models used to estimate and create a new meta-ensemble model. The flowchart shown in Figure 1 was developed for this purpose. First, we examined the economic indicators used in the literature.

Considering these economic indicators, the economic indicators that are/can be effective in the price policy of the Turkish iron and steel industry were determined. Data on the construction rebar prices between 2002–2022 were collected. The data prepared for the model were then divided into 80 % for the training dataset and 20 % for the test dataset. The model accuracies were calculated by comparing the results obtained by training the training data with the specified machine-learning algorithms with the test data. The model accuracies were evaluated using the MAPE, MAE, RMSE, and R^2 evaluation metrics. Subsequently, the target diagram was used to determine the model with the best estimation. Finally, for the model that provided the best estimation, the most important features affecting the construction rebar price estimation were determined.

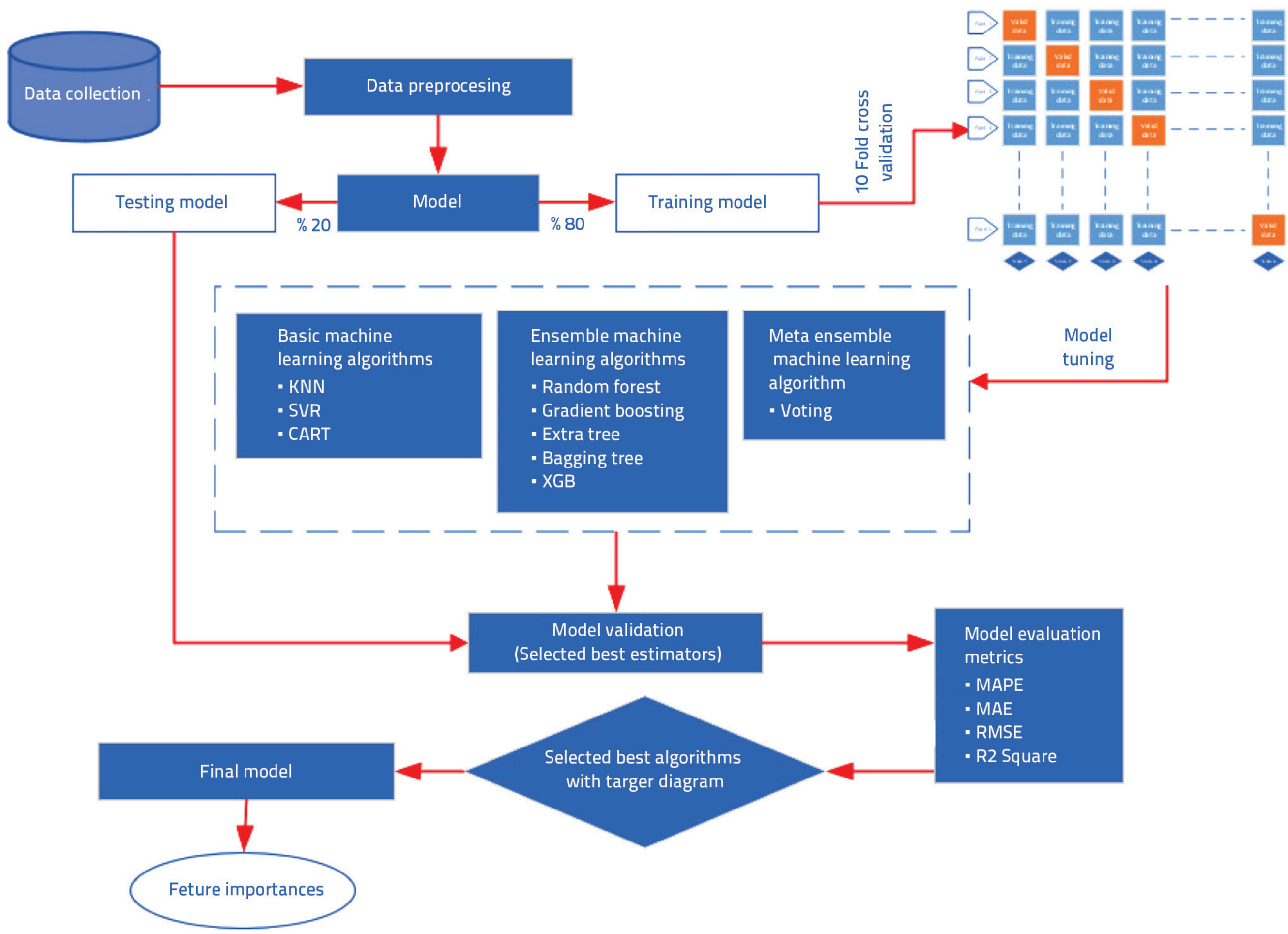

Figure 1. Flowchart of this study

3.1. Data collection

According to the literature, many economic indicators influence the calculation of construction costs. However, because each country has its own economic conditions, the economic indicators in the construction cost calculations differ according to the country. Because this study deals with the cost estimation of the rebar produced in Turkey, considering the economic indicators that will affect the costs of coal, gas, iron, petroleum, ferroalloys, electricity, producer price index, USD/TRY rate, and interest ratio in Turkey were determined. The prices of these indicators calculated by the Turkish Statistical Institute (TUIK) are price index values, whose changes are measured by comparing them over time. Therefore, the units of these values are expressed as 2003=100. The economic indicators used in this study were collected from the TUIK database [20] between January and March 2022. The descriptive statistics of these economic indicators are listed in Table 1. All the collected indicators are expressed as factors affecting the rebar price. Additionally, the prices of rebar exported from Turkey during this period were obtained from the SteelData company [21]. The changes in the construction rebar prices between January and March 2022 are shown in Figure 2. Fig. 3 shows the correlation matrix of the collected data. This matrix helps determine the relationships between different data points. The data indicate a strong correlation between coal, gas, iron, petroleum, ferroalloys, electricity, PPI, and USD/TRY. The interest variable exhibited a negative correlation with the other variables, whereas the iron cost variable exhibited a positive, albeit weak, correlation with the other variables.

Rebar prices are affected by the economic indicators calculated for a particular month in the following months. An accurate rebar price estimation, which is one of the most important components of the construction cost, is essential for the completion of the investment within the projected budget. Future rebar prices should be estimated and vary from project to project. In a project for which an investment decision has not yet been made, it is essential to accurately predict the prices 9 and 12 months ahead. By contrast, it is necessary to accurately predict prices one, three, and six months earlier in a project with a signed contract. Therefore, in this study, the effects of economic indicators on rebar prices were investigated after 1, 3, 6, 9, and 12 months. Five datasets were created for this purpose. A total of 243, 241, 238, 235, and 232 points were determined for the 1-, 3-, 6-, 9-, and 12-month lags.

Table 1. Descriptive statistics of the composed dataset for this study

Economic indicator	Max	Min	Mean	Median	Standard deviation	Skewness	Kurtosis
Coal	2319.26	95.33	421.49	377.79	347.60	2.71	10.22
Gas	2060.51	56.28	251.67	215.98	257.21	4.81	29.75
Iron	3622.76	78.35	509.26	483.63	446.03	2.85	12.66
Petroleum	4257.26	77.43	476.91	342.42	530.26	4.10	22.39
Ferroalloys	3005.04	58.95	379.99	247.18	449.27	3.44	13.68
Electricity	1610.03	85.46	230.87	202.29	183.87	4.29	24.80
Producer price index (ppi)	1423.27	71.11	259.95	202.08	205.17	2.75	9.82
USD/TRY rate	14.67	1.17	2.99	1.78	2.57	2.27	5.72
Interest ratio	0.38	0.04	0.18	0.17	0.08	0.85	0.50
Rebar Price (USD/Ton)	1445.00	190.00	518.95	498.25	176.93	1.58	6.32

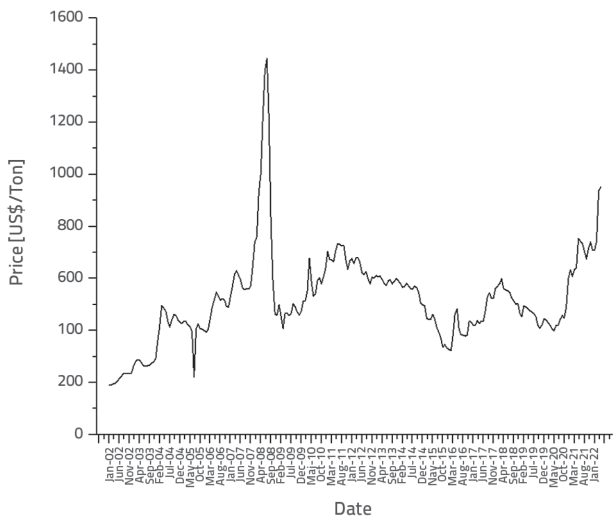

Figure 2. Change in construction rebar prices between January and March 2022

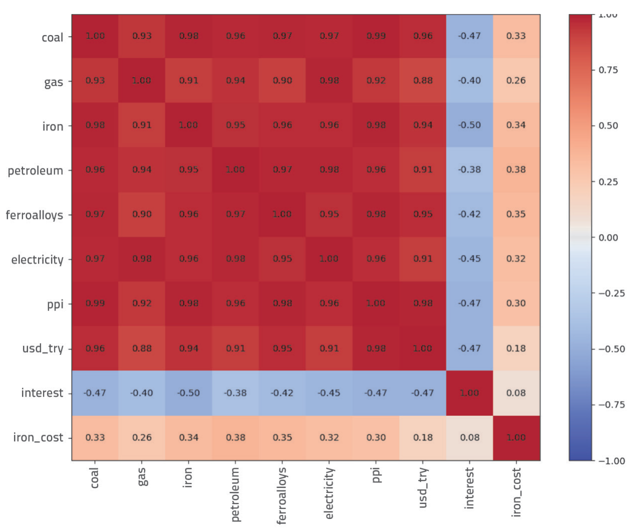

Figure 3. Correlation matrix among the different variables

3.2. Data pre-processing

Pre-processing was performed using the obtained data to further increase the predictive ability of the machine-learning models. First, missing data control and normalisation processes were performed. No missing data were found in the collected data. However, the input values appeared to differ. Some values range from zero to unity, whereas others range from 100 to 500. These values required normalisation. The normalisation process was performed as follows [22]:

$$z_i = \frac{x_i - x_{min}}{x_{max} - x_{min}} \tag{1}$$

3.3. 10-fold cross validation

After the data prepared for the model were separated in an 80:20 ratio, a 10-fold cross-validation process was performed to reduce the training data bias and determine the model performance randomness [23]. This process analysed 10 % of the training data as validation data and 90 % as training data. This process was repeated ten times, and the final result was determined by averaging all scores.

3.4. Machine-learning algorithms

In this study, ensemble and basic machine-learning algorithms were compared for construction rebar price estimation.

3.4.1. K-Nearest neighbour regression

The K-nearest neighbour (KNN) algorithm estimates the output value by training the input features, stocking them in the feature space, and comparing new incoming inputs with this feature space based on the closest distance [24]. The principal parameter to be determined for this model is the integer value *k* [25]. In the KNN algorithm, the best prediction is determined by optimising the *k*-neighbour value [25]. The optimal distance (*D*) is measured by calculating the distances of the *k*-nearest observations for each observation, as follows:

$$D = \sqrt{\sum_{j=1}^{k} \left(x_i - y_i\right)^2} \tag{2}$$

where x_i and y_i are the coordinate values of each observation and *D* is the distance between vectors.

3.4.2. Support vector regression

The aim of the support vector regression (SVR) model is to find a function that is as straight as possible, which is the closest to the features towards the Є maximum deviation determined between the predictions obtained from the trained data and the actual values. In other words, it ignores the errors as long as the deviation is small and evaluates the model according to the errors larger than the Є deviation value [11, 26, 27]. Theoretically, the SVR function can be formulated as follows [28]:

$$f(x) = w^t \times \varphi(x) + b \tag{3}$$

where *f*(*x*) is the estimated value obtained with the SVR function and *w* and *b* are coefficients determined by minimising the adjusted risk function, which is calculated as follows:

$$R = \frac{1}{2} w^2 + C \cdot \frac{1}{n} \sum_{i=1}^{n} \left| y_i - f\left(x_i\right) \right|_{\epsilon} \tag{4}$$

where *R* is the regularised risk function, *C* denotes the Euclidean norm and a cost parameter that quantifies the empirical risk, and $|y_i - f(x_i)|_{\epsilon}$ is an ε-insensitive loss function that controls the bias and makes the estimate robust. The kernel function is utilised to transform the data into a higher-dimensional space, which helps linearise nonlinear data and establish a stronger relationship between model and data. In this study, the radial basis function (RBF) kernel was used as the default kernel model in the SVR model used. The RBF kernel computes the similarity between two data points based on the distance between them, using a Gaussian distribution [29].

3.4.3. Classification and regression tree

The classification and regression tree (CART) algorithm, which has been applied to both classification and regression problems, was chosen for the regression problems in this study. In other words, the decision tree algorithm consists of decisions and leaf nodes [30]. The purpose of this algorithm is to boost prediction success by separating complex heterogeneous structures in the data into simpler homogeneous subbranches. Starting from the top node, the splitting process is repeated until the leaves are pure. To split the nodes in the most informative manner, we must define an objective function to be optimised using the tree-learning algorithm. Here, the objective function should maximise the information gain (IG) in each partition, as follows [31]:

$$IG\left(D_P, F\right) = I\left(D_P\right) - \left(\frac{N_{left}}{N_P} I\left(D_{left}\right) + \frac{N_{right}}{N_P} I\left(D_{right}\right)\right) \tag{5}$$

where *F* is the property of performing the split; D_p, D_{left} and D_{right} are the datasets of the root and child nodes; *I* is the impurity measure; N_p is the total number of samples at the parent node; and N_{left} and N_{right} are the number of samples at the child nodes.

3.4.4 Random forest regression

The random forest algorithm, which evaluates and combines predictions obtained using multiple decision tree algorithms, combines the bagging and random subspace methods [32]. In other words, the observation values for trees forming the Random Forest are selected by bagging the random sample selection, and the variables are selected by random subspace. The predicted values are calculated by weighting the error values from each tree. The basic parameters of this method are determined as *m* (number of parameters) and *k* (number of trees) [32, 33].

3.4.5. Gradient boosting machine

The gradient boosting machine (GBM) is a powerful machine-learning algorithm used for both regression and classification tasks. It builds models in a stepwise manner and generalises them, allowing optimisation of the differentiable loss function. In other words, it is a flexible and powerful tool for predictive modelling that exploits the gradient descent and boosting principles to iteratively improve model accuracy. It is particularly effective when combined with decision trees and provides a balance between model complexity and interpretability [34]. A series of model sets with GBM was created by fitting the errors of the previous model. This process was iterated until the maximum number of repetitions was reached [35].

3.4.6. Extra tree regression

The extra tree algorithm, which has emerged as an extension of the random forest algorithm, trains each base estimator using a random subset of features [36]. All training data were used to train each decision tree [37].

3.4.7. Bagging tree regression

The bagging algorithm proposed in [38] was created by combining multiple decision tree algorithms. A novel prediction was achieved by averaging the predictions of decision tree algorithms. This helps reduce the high variance in the decision tree algorithm.

3.4.8. eXtreme gradient boosting (XGBoost) regression

XGBoost was developed to improve the speed and prediction performance of the GBM algorithm [39]. In addition to the high predictive performance of this algorithm, another feature is that it prevents overfitting [40]. XGBoost consists of a loss function and normalisation term in the learning process, as in Equation 6. The normalisation term controls the model complexity, avoiding overlearning, whereas the loss function calculates the difference between each predicted and true value [41].

$$Aim^p = \sum_{i=1}^{n} l\left(\overline{y_i}, y_i\right) + \sum_{i=1}^{p} \sigma\left(f_i\right) \tag{6}$$

where l is the loss function, n is the number of observations used, σ is the normalisation term, and f_i is the estimate at step i.

3.4.9. Average voting ensemble model

The average voting ensemble method is based on a voting scheme that combines the machine-learning algorithms described above to achieve superior performance. The purpose of average voting is to average the predictions of multiple models, thereby obtaining more accurate overall predictions. It combines the predictions assuming that the model with the most votes is the winner, as shown in Figure 4 [42]. In this approach, the predictions of each constituent model are assigned equal weights. The model predictions are combined and averaged. This method prevents machine-learning algorithms from generating various errors and prevents overfitting. It also improves the overall performance by combining the advantages of different models.

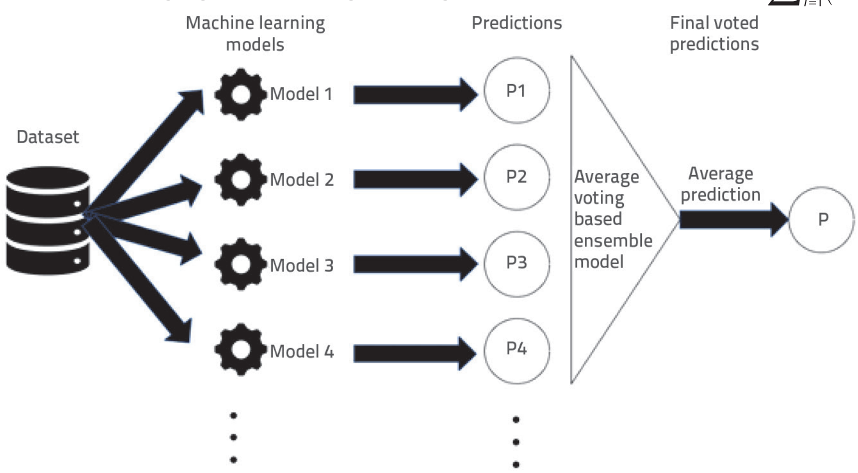

Figure 4. Voting ensemble model structure [34]

3.5. Model evaluation metrics

To calculate and compare the estimation accuracy and error rates of the models analysed for rebar price estimation, the MAPE, mean absolute error (MAE), root mean squared error (RMSE), and coefficient of determination (R^2) metrics were used to evaluate the models. MAPE is an evaluation metric that provides the model error rate as a percentage and is calculated as described in Equation 7. The MAE was calculated by taking the absolute value of the difference between each predicted value and the actual value, as in Equation 8. The closer the calculated MAE value is to zero, the better the model prediction performance. RMSE, which is a widely used metric for regression problems, was calculated using Equation 9. For the RMSE, the minimum calculated values, such as MAE, indicate a preferable prediction. Additionally, as it has the same units as the dependent variable, the RMSE is more often employed than MSE and MAE to evaluate the performance of regression models compared with other random models. However, the regression evaluation metric, which varies between zero and unity (providing a superior model fit as it approaches unity), is the R^2 value, calculated using Equation 10. The R^2 result can also be evaluated as the variance ratio explained by the model.

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \cdot 100 \tag{7}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|\hat{y}_i - y_i\right| \tag{8}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\hat{y}_i - y_i\right)^2} \tag{9}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2}{\sum_{i=1}^{n}\left(y_i - \overline{y}_i\right)^2} \tag{10}$$

where:
- n - number of observations in the dataset
- y_i - actual value
- $\hat{y}_i$ - predicted value
- $\overline{y}_i$ - mean of the actual values

3.6. Target diagram

In this study, a target diagram [43] was used to compare the prediction performance of the basic machine-learning, ensemble machine-learning, and meta-ensemble algorithms used for construction rebar price prediction, as it is difficult to compare a total of nine algorithms. Using these diagrams, a graphical representation can be used to determine which algorithm has the best predictive performance. The target diagram emerges from the relationship between the statistical metric of bias (B) and the unbiased root mean square difference (uRMSD). The relationship between B and uRMSD returns the root mean square difference (RMSD), as follows:

$$RMSD^2 = B^2 + uRMSD^2 \quad (11)$$

Algorithms with better predictive performance are expected to have the lowest RMSD values.

3.7. Feature importance

Feature importance is a concept that measures the influence of input features on a model's prediction results, thereby enhancing the model explainability [44]. It has been used in numerous previous studies [16, 45]. The feature importance values of the meta-ensemble model were calculated by averaging the machine-learning models used in this study. In other words, the mean feature importance values were obtained based on the average coefficients of each feature obtained from the algorithm results in this study. These values show the degree of influence of these elements on the iron price forecasts.

4. Results and discussion

Within the scope of this study, the rebar price estimations, which is among the building materials that more significantly affect the building production cost, for the next 1, 3, 6, 9, and 12 months were examined in two stages using machine-learning models. First, for each month, the MAPE, MAE, RMSE, and R^2 values for each model were determined via basic machine-learning models (KNN, SVR, and CART), and ensemble machine-learning models (Random Forest, Gradient Boosting Regression, Extra Tree Regression, Bagging Tree Regression, XGB Regression) were carried out comparatively within themselves. While analysing these models, each requires specific parameters that must be adjusted during the training process. The hyperparameters used in the models are listed in Table 2.

Subsequently, to increase the prediction accuracy, different variations of the basic and ensemble models were tested, and a voting regression meta-ensemble model was created by combining them, as shown in Figure 5. This model was then compared with other estimation models. Second, the relationship between the price estimates of the 1-, 3-, 6-, 9- and 12-month lags and the actual values and the effect of each input factor on model performance were graphically expressed and examined.

The evaluation metrics obtained from the analysis performed according to the effects of economic indicators for the 1-month-lag rebar price estimation case are presented in Table 3. As can be seen, when the basic and ensemble models are compared, the model with the highest prediction accuracy according to the MAPE and MAE metrics is the Random Forest model, while the model with the highest prediction accuracy according to the RMSE and R^2 metrics is the Bagging Tree model. Based on the comparison, the Voting Regression meta-ensemble model was determined to be the most accurate. The MAPE, MAE, RMSE, and R^2 values for this model were 3.90 %, 19.2721, 28.59717, and 0.953522, respectively.

The evaluation metrics obtained from the analysis performed according to the effects of economic indicators for the 3-month-lag rebar price estimation case are presented in Table 4. As can be seen, when the basic and ensemble models are compared, the model with the highest prediction accuracy according to all evaluation metrics is the Bagging tree model. When all models are compared, the Voting Regression meta-ensemble model was determined to be the most accurate. The MAPE, MAE, RMSE, and R^2 values for this model were 3.80 %, 19.06886, 28.10399, and 0.955111, respectively.

The evaluation metrics obtained from the analysis performed according to the effects of economic indicators for the 6-month-lag rebar price estimation case are presented in Table 5.As can be seen, when the basic and ensemble models are compared, the model with the highest prediction accuracy according to all evaluation metrics is the Extra Tree model. When all models are compared, the voting meta-ensemble model was determined to be the most accurate estimation model, except for the R^2 metric. The MAPE, MAE, and RMSE values for this model were 3.92 %, 19.62414, and 28.95394, respectively. In terms of the R^2 metric, the values of the Extra Tree and Voting Regression models are very close to each other, they were determined as 0.952496 and 0.952355, respectively.

The evaluation metrics obtained from the analysis performed according to the effects of economic indicators for the 9-month-lag rebar price estimation are presented in Table 6. As can be seen, when the basic and ensemble models are compared, the model with

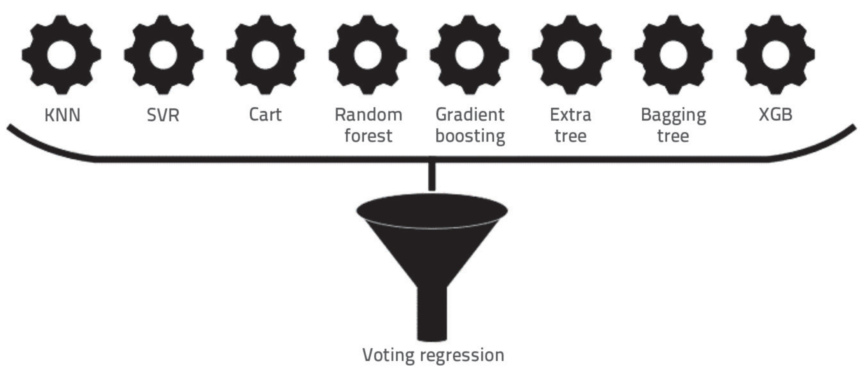

Figure 5. Voting for the meta-ensemble model

Table 2. Hyperparameters used for each machine-learning model

	1 month delay	3 month delay	6 month delay	9 month delay	12 month delay
KNN	k:3, weights: uniform	k:3, weights: uniform	k:2, weights: uniform	k:4, weights: uniform	k:3, weights: uniform
SVR	C:7200, kernel: RBF	C:7400, kernel: RBF	C:9500, kernel: RBF	C:51000, kernel: RBF	C:54400, kernel: RBF
CART	max_leaf_nodes:30, min_samples_split: 3	max_leaf_nodes: 38, min_samples_split: 5	max_leaf_nodes: 18, min_samples_split: 2	max_leaf_nodes: 5, min_samples_split: 8	max_leaf_nodes: 33, min_samples_split: 2
Random forest	bootstrap: False, max_depth: 200, max_features: sqrt, min_samples_leaf: 1, min_samples_split: 2, n_estimators: 600	bootstrap: False, max_depth: 750, max_features: sqrt, min_samples_leaf: 1, min_samples_split: 2, n_estimators: 300	bootstrap: False, max_depth: 200, max_features: sqrt, min_samples_leaf: 2, min_samples_split: 3, n_estimators: 200	bootstrap: False, max_depth: 300, max_features: sqrt, min_samples_leaf: 2, min_samples_split: 4, n_estimators: 20	bootstrap: False, max_depth: 1100, max_features: sqrt, min_samples_leaf: 4, min_samples_split: 5, n_estimators: 9
Gradient boosting	learning_rate: 0,05, max_depth: 50, max_features: 0,1, min_samples_leaf: 2, n_estimators: 2000, subsample: 1,0	learning_rate: 0,05, max_depth: 10, max_features: 0,01, min_samples_leaf: 2, n_estimators: 400, subsample: 0,5	learning_rate: 0,05, max_depth: 20, max_features: 0,3, min_samples_leaf: 2, n_estimators: 500, subsample: 0,5	learning_rate: 0,05, max_depth: 90, max_features: 0,5, min_samples_leaf: 5, n_estimators: 700, subsample: 0,5	learning_rate: 0,01, max_depth: 30, max_features: 0,5, min_samples_leaf: 4, n_estimators: 800, subsample: 1,0
Extra tree	bootstrap: False, max_depth: 100, max_features: 0,75, min_samples_leaf: 1, min_samples_split: 2, n_estimators: 100	bootstrap: False, max_depth: 50, max_features: 0,25, min_samples_leaf: 1, min_samples_split: 2, n_estimators: 400	bootstrap: False, max_depth: 30, max_features: 0,8, min_samples_leaf: 1, min_samples_split: 2, n_estimators: 800	bootstrap: False, max_depth: 30, max_features: 0,6, min_samples_leaf: 2, min_samples_split: 2, n_estimators: 10	bootstrap: False, max_depth: 40, max_features: 1,0, min_samples_leaf: 1, min_samples_split: 2, n_estimators: 8
Bagging tree	base_estimator: None, bootstrap: False, bootstrap_features: True, max_features: 0,5, max_samples: 1,0, n_estimators: 800	base_estimator: None, bootstrap: False, bootstrap_features: netočno, max_features: 0,6, max_samples: 1,0, n_estimators: 800	base_estimator: None, bootstrap: False, bootstrap_features: False, max_features: 0,5, max_samples: 1,0, n_estimators: 800	base_estimator: None, bootstrap: Istina, bootstrap_features: True, max_features: 0,5, max_samples: 1,0, n_estimators: 60	base_estimator: None, bootstrap: False, bootstrap_features: False, max_features: 5, max_samples: 100, n_estimators: 200
XGB	colsample_bytree: 0,5, learning_rate: 0,1, max_depth: 5, min_child_weight: 2, n_estimators: 1500, subsample: 1,0	colsample_bytree: 0,3, learning_rate: 0,05, max_depth: 20, min_child_weight: 1, n_estimators: 1000, subsample: 0,1	colsample_bytree: 0,5, learning_rate: 0,05, max_depth: 10, min_child_weight: 3, n_estimators: 3000, subsample: 1,0	colsample_bytree: 0,7, learning_rate: 0,1, max_depth: 12, min_child_weight: 11, n_estimators: 200, subsample: 1,0	colsample_bytree: 0,7, learning_rate: 0,1, max_depth: 20, min_child_weight: 1, n_estimators: 50, subsample: 1,0

Table 3. Machine-learning model results for the 1-month-lag construction rebar price estimation

Model	MAPE [%]	MAE	RMSE	R^2
KNN	5.97	28.44813	39.48362	0.9114
SVR	8.51	39.51059	52.64472	0.842489
Cart	7.83	37.09775	57.3994	0.812753
Random forest	**4.37**	**21.40749**	31.50025	0.943607
Gradient boosting regression	4.79	23.0923	30.915	0.945683
Extra tree regression	4.78	23.44043	32.84048	0.938706
Bagging tree regression	4.49	21.64282	**30.83886**	**0.94595**
XGB regression	4.69	23.3149	34.86586	0.930912
Voting regression	**3.90**	**19.2721**	**28.59717**	**0.953522**

Table 4. Machine-learning model results for the 3-month-lag construction rebar price estimation

Model	MAPE [%]	MAE	RMSE	R^2
KNN	10.90	52.51305	74.39734	0.805022
SVR	12.87	58.86498	89.51642	0.717722
Cart	9.24	41.96255	69.92181	0.827775
Random forest	8.45	37.0673	52.3793	0.903352
Gradient boosting regression	9.82	44.61394	61.82911	0.865334
Extra tree regression	8.63	36.52421	52.80427	0.901778
Bagging tree regression	**8.12**	**35.40851**	**51.50008**	**0.90657**
XGB regression	11.38	51.66069	67.52914	0.83936
Voting regression	**3.80**	**19.06886**	**28.10399**	**0.955111**

Table 5. Machine-learning model results for the 6-month-lag construction rebar price estimation

Model	MAPE [%]	MAE	RMSE	R^2
KNN	7.82	51.57516	100.8423	0.758166
SVR	13.58	80.882	155.4422	0.425395
Cart	10.99	52.47877	72.66854	0.874419
Random forest	8.33	45.35044	66.37932	0.895215
Gradient boosting regression	7.37	40.54662	57.1967	0.922201
Extra tree regression	**6.62**	**34.20766**	**44.69392**	**0.952496**
Bagging tree regression	8.37	41.76773	53.85572	0.931024
XGB regression	8.17	42.14981	54.58297	0.929149
Voting regression	**3.92**	**19.62414**	**28.95394**	**0.952355**

Table 6. Machine-learning model results for the 9-month-lag construction rebar price estimation

Model	MAPE [%]	MAE	RMSE	R^2
KNN	9.28	56.20286	120.9789	0.719113
SVR	13.60	83.25784	184.2311	0.348614
Cart	17.57	104.1335	193.7997	0.279194
Random forest	9.23	55.1738	119.7518	0.724782
Gradient boosting regression	**8.52**	**44.34587**	**75.30403**	**0.89117**
Extra tree regression	8.95	49.73613	95.94524	0.823332
Bagging tree regression	10.11	56.2993	115.2702	0.744996
XGB regression	9.93	50.05226	81.11086	0.873739
Voting regression	**4.42**	**22.19465**	**31.59304**	**0.943274**

the highest prediction accuracy according to all evaluation metrics is the Gradient Boosting model. When all models are compared, the Voting Regression model is determined to be the most accurate. The MAPE, MAE, RMSE, and R^2 values for this model were 4.42 %, 22.19465, 31.59304, and 0.943274, respectively.

The evaluation metrics obtained from the analysis performed according to the effects of economic indicators for the 12-month-lag rebar price estimation are presented in Table 7. As can be seen, when the basic and ensemble models are compared, the model with the highest prediction accuracy according to all evaluation metrics is the XGBoost model. When all models are compared, the Voting Regression model is determined to be the most accurate. The MAPE, MAE, RMSE, and R^2 values for this model were 4.11 %, 20.39988, 30.71019, and 0.9464, respectively.

Table 7. Machine-learning model results for the 12-month-lag construction rebar price estimation

Model	MAPE [%]	MAE	RMSE	R^2
KNN	12.12	70.28362	149.2291	-0.09984
SVR	8.86	50.52747	86.82169	0.627712
Cart	08.65	47.66665	77.24122	0.70534
Random forest	7.98	45.74283	65.88371	0.785623
Gradient boosting regression	6.31	35.44273	53.35277	0.859416
Extra tree regression	7.20	41.23574	76.28565	0.712586
Bagging tree regression	7.63	40.80243	58.88146	0.82877
XGB regression	**5.81**	**30.42101**	**43.77327**	**0.905367**
Voting regression	**4.11**	**20.39988**	**30.71019**	**0.9464**

According to these results, the voting ensemble models exhibited the most accurate estimation values at all time lags. Additionally, except for the voting model, the ensemble models exhibited better estimation at each monthly lag. Conversely, the basic machine-learning models exhibited a lower estimation success than the ensemble models. To compare the machine-learning models used for rebar price estimation, target diagrams were formed for the 1-, 3-, 6-, 9-, and 12-month lags, as shown in Figure 6. Based on these results, the following conclusion can be drawn:

For the 1-month-lag case, among the models analysed, the ensemble and meta-ensemble estimation errors were close to each other, and the model with the lowest estimation error value was used as the voting meta-ensemble model (Figure 6-a).

There was a significant prediction error difference between the meta-ensemble and the ensemble and basic machine-learning models for the 3-, 9-, and 12-month lags. The model with the lowest estimation error was used as the voting ensemble model (Figures 6.b, 6.d, 6.e).

There was a significant prediction error difference between the meta-ensemble and the ensemble and basic machine-learning models for the 6-month-lag case. Although the model with the lowest estimated error value was the voting meta-ensemble model, the closest predictions to this model were determined by the KNN and SVR basic machine-learning models (Figure 6.c).

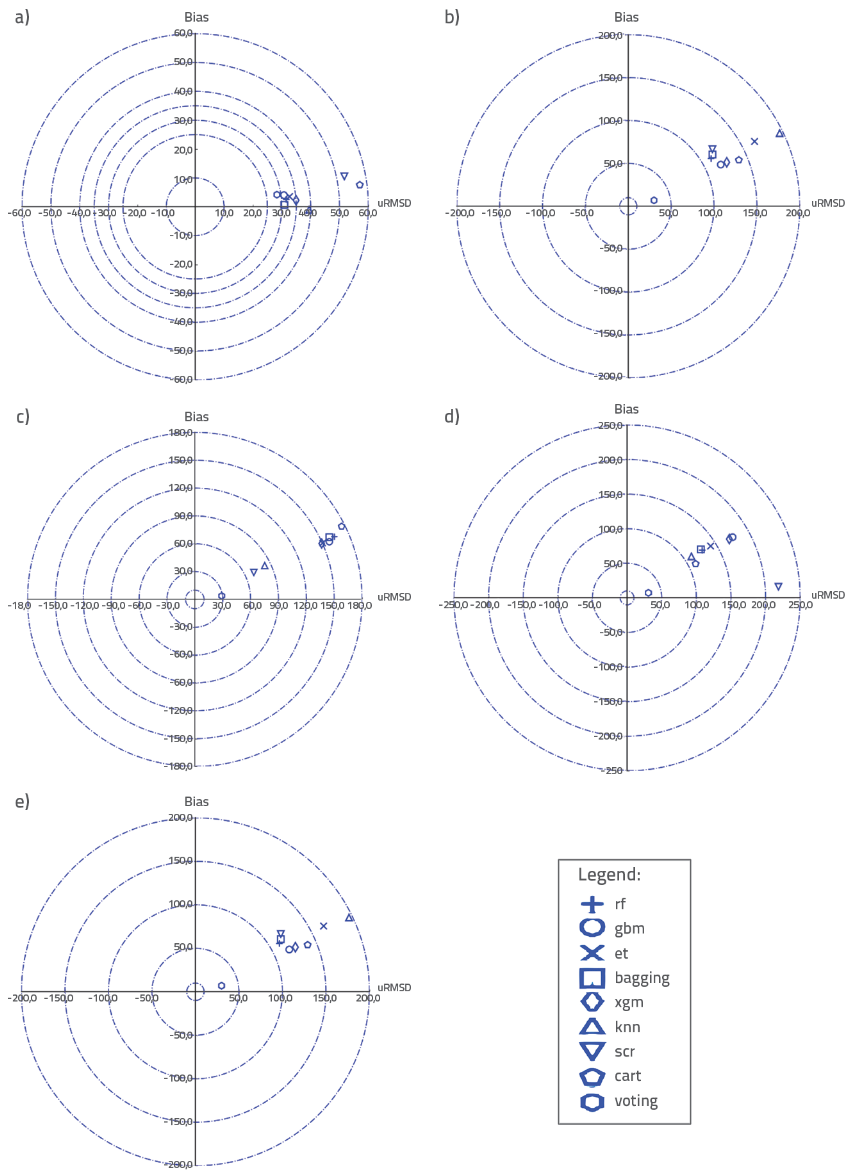

Figure 6. Target diagrams for all machine-learning models analysed in this study: a) 1 month late; d) 3 months late; c) 6 months late; d) 9 months late; e) 12 months late

The basic, ensemble, and meta-ensemble machine-learning models used in the construction rebar price estimation were compared based on the MAPE and R^2 evaluation metrics for the 1-, 3-, 6-, 9-, 12-month-lag periods. Based on these results, the following conclusion can be drawn:

The 1-month-lag period was determined to exhibit the most accurate predictive value among the basic machine-learning models. The basic machine-learning model with the most accurate predictive value during this period was the KNN model (Figure 7).

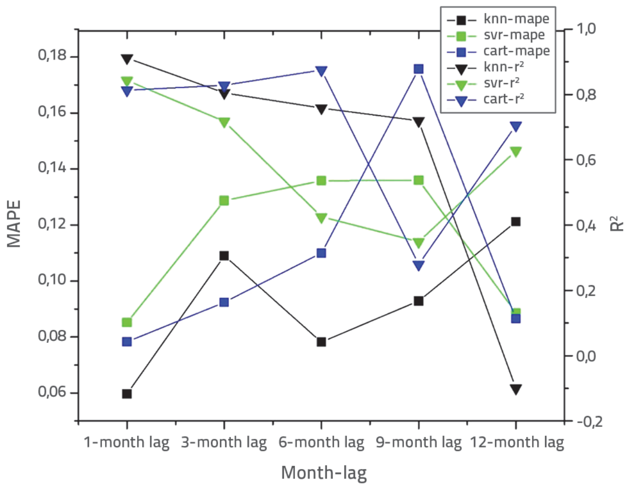

Figure 7. Comparison between the basic machine-learning models for the 1-, 3-, 6-, 9-, 12-month-lag periods

The 1-month-lag period was determined to exhibit the most accurate predictive value among the ensemble machine-learning models. Although all models in this period exhibited similar prediction values, the model with the lowest MAPE value was the Random Forest model, whereas the model with the highest R^2 value was the Bagging Tree model (Figure 8).

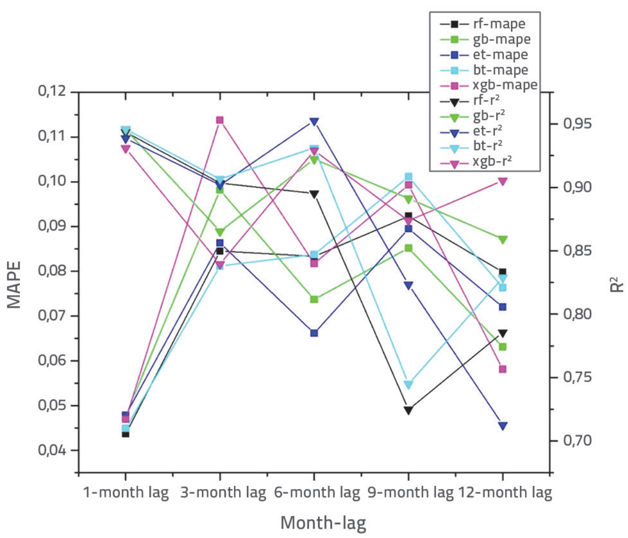

Figure 8. Comparison between the ensemble machine-learning models for the 1-, 3-, 6-, 9-, 12-month-lag periods

For the Voting meta-ensemble machine-learning model, the period with the highest predictive value was the 3-month-lag period. Additionally, the estimated values for the 1- and 6-month lags were relatively high (Figure 9).

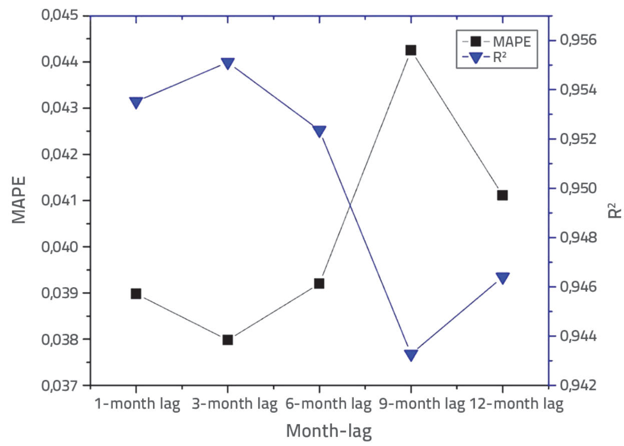

Figure 9. Comparison between the voting meta-ensemble machine-learning models for the 1-, 3-, 6-, 9-, 12-month-lag periods

In this study, machine-learning models for rebar price estimation were compared. The voting meta-ensemble model obtained by combining these models yielded the most accurate prediction results. This main contributions of this study to the field of construction rebar price estimation are discussed in the following sub-sections.

4.1. Estimation evaluation metrics

The limited number of studies on this subject clearly reveals the need to increase the success of iron price estimation. In previous studies, the price prediction performance was evaluated by considering error rates such as the MAE, RMSE, and MAPE [4, 9, 10, 23]. Additionally, one or more error rates were considered [4, 7, 9, 10]. In AI regression analyses, only R^2 [8] or R^2 with error rates [23] is used as a common evaluation method. In the literature on rebar price estimation, no study was found in which different error rates or R^2 values were evaluated. In contrast to previous studies on this subject, the R^2 value was considered along with the error rate in this study. Therefore, both the prediction errors and compatibility of the estimated values with the actual values were evaluated.

4.2. Featue importance

Shiha et al. [9] stated that the CPI, PPI, unemployment rate, GDP, foreign reserves, USD exchange rate, and lending rate parameters affect rebar price estimation. However, they did not indicate the parameters that affected the influence amount. The feature importance values of the parameters affecting the rebar price estimated using the voting ensemble model are shown in Figure 10. Accordingly,

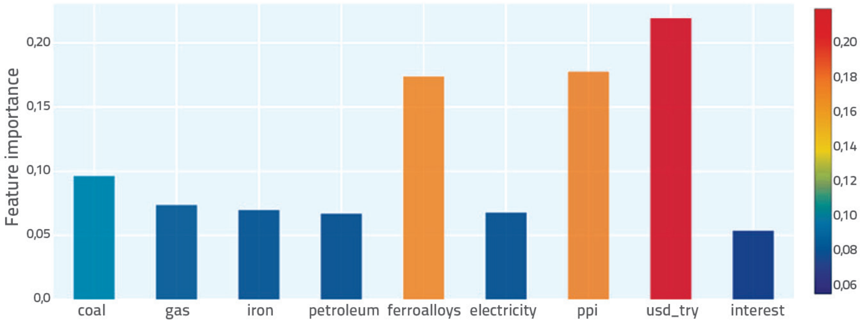

Figure 10. Feature importance of rebar prices estimated based on voting meta-ensemble model

differentiated via updating or adding/ removal. Furthermore, it can be applied to other subjects. A comparison between the actual and predicted values for the voting meta-ensemble machinelearning model with a 3-month-lag period is shown in Figure 11.

In this study, because the voting meta-ensemble model is more successful than the other models, it can be used by all parties in the construction sector for rebar price estimation. This model is expected to be useful for determining the lowest-cost time for investors while determining the investment start date. Additionally, we expect that this will provide an advantage in terms of cost when determining the time of rebar purchase in construction projects for which contracts have been signed and construction is ongoing. The newly developed model offers accurate price predictions not only for certain periods but also for periods covering a full year. In future work, the proposed model can be further developed and be used for the price estimation of other construction materials.

the most influential parameters were the USD/TRY, PPI, and ferroalloy. The global determination of the iron price and the widespread use of the US dollar in international trade confirm the effectiveness of the USD/TRY parameter. Furthermore, rebar prices from an industrial enterprise involved in iron and steel production are directly affected by PPI and ferroalloy prices, making these parameters essential for price forecasting.

4.3. Estimation performance

In previous studies on rebar price estimation, satisfactory results were obtained for different periods using different estimation models [9, 10]. However, in this study, successful results were obtained using a single model for 1-, 3-, 6-, 9-, and 12-month-lag periods. Obtaining a successful estimate using a single model contributes to practical application and saves time. In other words, although eight different models were used, using a new single model by combining these different models effectively reduced the computational complexity and shortened the analysis time.

4.4 New prediction model

Whereas previous studies combined one [4, 9, 10], two [10, 23], or three [10] estimation models, in this study, a voting prediction model was developed by combining eight machine-learning models, three of which were basic and five of which were ensembles. The eight machine-learning models used in this study are commonly used in computer science. In this study, these models were combined to create a new model for construction material price prediction. The newly developed voting prediction model achieved more accurate results than the other models. The parameters in this estimation model can be

5. Conclusion

Accurate estimates are crucial to determine future rebar prices in the construction sector. However, studies on this topic are limited, which further increases its importance. Accordingly, this study aimed to estimate future rebar prices in the construction sector using an AI-based method. In this context, while analysing the future rebar price estimates, the relationship with the change in economic parameters was considered.

In this study, machine learning was used as an AI method, and coal, gas, iron, petroleum, ferroalloys, electricity, producer

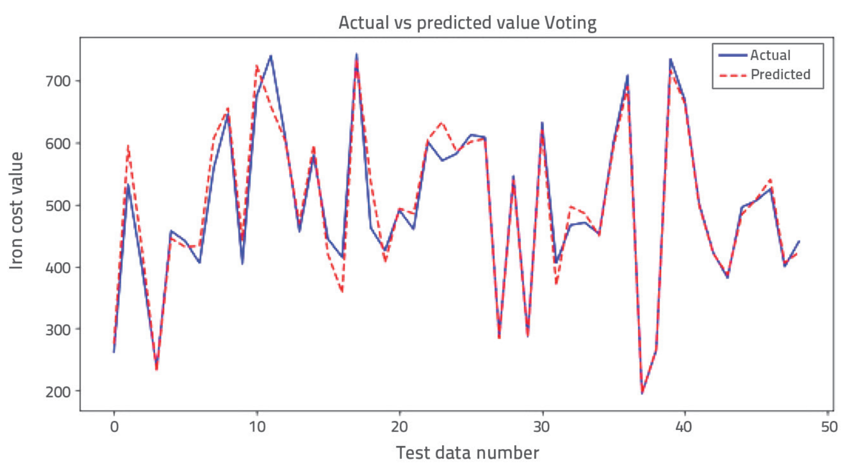

Figure 11. Comparison between the actual and predicted values according to the vote meta-ensemble model over a 3-month lag period

price index, USD/TRY rate, and interest ratio were determined as the economic parameters. Machine-learning models, basic machine-learning models (KNN, SVR, and Cart), and ensemble machine-learning models (Random Forest, Gradient Boosting Regression, Extra Tree Regression, Bagging Tree Regression, XGB Regression) were comparatively analysed. Based on the results, the Voting Regression meta-ensemble model, which was created by testing different variations of the basic and ensemble models, was developed to increase the estimation accuracy.

Unlike previous studies, in this study, for the rebar price estimation in the 1-, 3-, 6-, 9-, and 12-month-lag periods, the percentage error rate (MAPE) was below 4 %, and the model fit (R^2) was over 94 % for all periods. The best results for the basic and ensemble machine-learning models were obtained in the 1-month-lag period, while those for the meta ensemble machine-learning model were obtained in the 3-month-lag period. However, when the basic, ensemble, and meta-ensemble models were compared, the voting ensemble model exhibited the most accurate prediction results in the 1-month-lag period. Additionally, the voting model produced more accurate results in the 3-month-lag period than the 1-month-lag period. Therefore, in general, the voting meta-ensemble model was determined to be more successful. This model exhibits acceptable prediction accuracy for investment and rebar purchasing decisions in all periods examined. However, more accurate predictions can be obtained in the short term.

In summary, the estimated performance, economic parameter effect, and success percentage during different estimation periods were analysed, and a new price estimation model is proposed. In future work, based on these findings, different estimates can be made for different fields, in addition to Civil Engineering, by adding or removing new parameters. We expect that the findings of this study will contribute significantly to all parties involved in the construction industry and academia.

Acknowledgements

The authors would like to thank SteelData for helping with the rebar price dataset construction.

REFERENCES

[1] Yun, S.: Performance Analysis of Construction Cost Prediction Using Neural Network for Multioutput Regression, Applied Sciences, 12 (2022) 19, p. 9592.

[2] Aram, S., Eastman, C., Beetz, J.: Qualitative and Quantitative Cost Estimation: A Methodology Analysis, American Society of Civil Engineers, 2014.

[3] Baalousha, Y., Çelik, T.: An integrated web-based data warehouse and artificial neural networks system for unit price analysis with inflation adjustment, Journal of Civil Engineering and Management, 17 (2011) 2, pp. 157-167.

[4] Mir, M., Kabir, H.D., Nasirzadeh, F., Khosravi, A.: Neural network-based interval forecasting of construction material prices, Journal of Building Engineering, 39 (2021) 102288, pp. 1-13.

[5] Mashhadban, H., Kutanaei, S.S., Sayarinejad, M.A.: Prediction and modeling of mechanical properties in fiber reinforced self-compacting concrete using particle swarm optimization algorithm and artificial neural network, Construction and Building Materials, 119 (2016) 2016, pp. 277-287.

[6] Shi, L., Lin, S., Lu, Y., Ye, L., Zhang, Y.X.: Artificial neural network based mechanical and electrical property prediction of engineered cementitious composites, Construction and Building Materials, 174 (2018), pp. 667-674.

[7] Ou, T.Y., Cheng, C.Y., Chen, P.J., Perng, C.: Dynamic cost forecasting model based on extreme learning machine - A case study in steel plant, Computers & Industrial Engineering, 101 (2016), pp. 544-553.

[8] Elfahham, Y.: Estimation and prediction of construction cost index using neural networks, time series, and regression, Alexandria Engineering Journal, 58 (2019) 2, pp. 499-506.

[9] Shiha, A., Dorra, E.M., Nassar, K.: Neural Networks Model for Prediction of Construction Material Prices in Egypt Using Macroeconomic Indicators, Journal of Construction Engineering and Management, 146 (2020) 3, pp. 1-16.

[10] Faghih, S.A.M., Kashani, H.: Forecasting Construction Material Prices Using Vector Error Correction Model, Journal of Construction Engineering and Management, 144 (2018) 8, p. 4018075.

[11] Nguyen, H., Vu, T., Vo, T.P., Thai, H.T.: Efficient machine learning models for prediction of concrete strengths, Construction and Building Materials, 266 (2021) 120950, pp. 1-17.

[12] Xu, X., Zhang, Y.: Price forecasts often steel products using Gaussian process regressions, Engineering Applications of Artificial Intelligence, 126 (2023) 106870, pp. 1-13.

[13] Dai, X., Gao, P., Ma, S.: Cost Forecasting for Building Rebar under Uncertainty Conditions: Methodology and Practice, Buildings, 14 (2024) 7, p. 1900.

[14] Mi, J., Xie, X., Luo, Y., Zhang, Q., Wang, J.: Research on Rebar Futures Price Forecast Based on VMD—EEMD—LSTM Model, in Chen, C.H., Scapellato, A., Barbiero, A., Korzun, D.G. (Eds.): Applied Mathematics, Modeling and Computer Simulation, IOS Press, 2023.

[15] Chan, C.H., Sun, M., Huang, B.: Application of machine learning for advanced material prediction and design, EcoMat, 4 (2022) 4, 1-28.

[16] Karatas, I., Budak, A.: Development and comparative of a new meta-ensemble machine learning model in predicting construction labor productivity, Engineering, Construction and Architectural Management, 31 (2024) 3, pp. 969-9988.

[17] Karatas, I., Budak, A.: Prediction of Labor Activity Recognition in Construction with Machine Learning Algorithms, Icontech International Journal, 5 (2021) 3, pp. 38-47.

[18] Üneş, F., Kaya, Y.Z., Mamak, M.: Daily reference evapotranspiration prediction based on climatic conditions applying different data mining techniques and empirical equations, Theoretical and Applied Climatology, 141 (2020), 1-2, pp. 763-773.

[19] Kayadelen, C., Önal, Y., Altay, G., Öztürk, M., Serin, S.: Effects of maintenance, traffic and climate condition on International Roughness Index of flexible pavement, International Journal of Pavement Engineering (2022), pp. 1-13.

[20] TUIK: Turkish Statistical Institute, https://data.tuik.gov.tr/, accessed September 20, 2024.

[21] SteelData: SteelData Informatics and Information Technologies, https://www.steel-data.com/, accessed September 20, 2024.

[22] Jain, S., Shukla, S., Wadhvani, R.: Dynamic selection of normalization techniques using data complexity measures, Expert Systems with Applications, 106 (2018), pp. 252-262.

[23] Farooq, F., Ahmed, W., Akbar, A., Aslam, F., Alyousef, R.: Predictive modeling for sustainable high-performance concrete from industrial wastes: A comparison and optimization of models using ensemble learners, Journal of Cleaner Production, 292 (2021), p. 126032.

[24] Han, J., Micheline, K., Jian, P.: Data mining concepts and techniques third edition, The Morgan Kaufmann Series in Data Management Systems, 2011.

[25] Swetapadmaa, A., Yadav, A.: A novel single-ended fault location scheme for parallel transmission lines using k-nearest neighbor algorithm, Computers & Electrical Engineering, 69 (2018), pp. 41-53.

[26] Smola, A.J., Schölkopf, B.: A tutorial on support vector regression, Statistics and Computing, 14 (2004) 3, pp. 199-222.

[27] Drucker, H., Burges, C.J.C., Kaufman, L., Smola, A., Vapnik, V.: Support Vector Regression Machines, Advances in neural information processing systems, 9 (1997), 155-161.

[28] Chen, R., Liang, C.Y., Hong, W.C., Gu, D.X.: Forecasting holiday daily tourist flow based on seasonal support vector regression with adaptive genetic algorithm, Applied Soft Computing, 26 (2015), pp. 435-443.

[29] Kovacevic, M., Ivanisevic, N., Petronijević, P., Despotović, V.: Construction cost estimation of reinforced and prestressed concrete bridges using machine learning, Journal of the Croatian Association of Civil Engineers, 73 (2021) 01, pp. 1-13.

[30] James, G., Witten, D., Hastie, T., Tibshirani, R.: An Introduction to Statistical Learning, Springer New York, New York, NY, 2013.

[31] Breiman, L., Friedman, J.H., Stone, C.J., Olshen, R.A.: Classification and Regression Trees, Biometrics, 40 (1984) 3, p. 874.

[32] Breiman and Leo: Random Forests, Machine Learning, 45 (2001), pp. 5-32.

[33] Kayadelen, C., Altay, G., Önal, Y.: Numerical simulation and novel methodology on resilient modulus for traffic loading on road embankment, International Journal of Pavement Engineering, 23 (2021) 9, pp. 3212-3221.

[34] Friedman, J.H.: Greedy function approximation: a gradient boosting machine, Annals of statistics, (2001), pp. 1189-1232.

[35] Gong, M., Bai, Y., Qin, J., Wang, J., Yang, P., Wang, S.: Gradient boosting machine for predicting return temperature of district heating system: A case study for residential buildings in Tianjin, Journal of Building Engineering, 27 (2020), p. 100950.

[36] Geurts, P., Ernst, D., Wehenkel, L.: Extremely randomized trees, Machine Learning, 63 (2006) 1, pp. 3-42.

[37] Ahmad, M.W., Reynolds, J., Rezgui, Y.: Predictive modelling for solar thermal energy systems: A comparison of support vector regression, random forest, extra trees and regression trees, Journal of Cleaner Production, 203 (2018), pp. 810-821.

[38] Breiman, L.: Bagging Predictors, Machine Learning, 24 (1996) 2, pp. 123-140.

[39] Chen, T., He, T.: xgboost: eXtreme Gradient Boosting, R package version 0.4-2, 1(4), 2015, https://mran.microsoft.com/web/packages/xgboost/vignettes/xgboost.pdf, accessed April 26, 2021.

[40] Chen, T., Guestrin, C. (Eds.): XGBoost: A Scalable Tree Boosting System, 2016.

[41] Alshboul, O., Shehadeh, A., Almasabha, G., Almuflih, A.S.: Extreme Gradient Boosting-Based Machine Learning Approach for Green Building Cost Prediction, Sustainability, 14 (2022) 11, p. 6651.

[42] Géron, A.: Hands-on machine learning with Scikit-Learn and TensorFlow: Concepts, tools, and techniques to build intelligent systems / Aurélien Géron, First edition, O'Reilly, Beijing, 2017.

[43] Stow, C.A., Jolliff, J., McGillicuddy, D.J., Doney, S.C., Allen, J.I., Friedrichs, M.A.M., et al.: Skill Assessment for Coupled Biological/Physical Models of Marine Systems, Journal of Marine Systems, 76 (2009), 1-2, pp. 4-15.

[44] Zien, A., Kraemer, N., Sonnenburg, S., Raetsch, G. (Eds.): The Feature Importance Ranking Measure, Springer, Berlin Heidelberg, 2009,

[45] Kayadelen, C., Önal, Y., Altay, G., Öztürk, M., Serin, S.: Effects of maintenance, traffic and climate condition on International Roughness Index of flexible pavement, International Journal of Pavement Engineering, 24 (2023) 1.