Transactions on Computational and Scientific Methods | Vol. 5, No. 1, 2025
ISSN: 2998-8780
https://pspress.org/index.php/tcsm
Pinnacle Science Press

A CatBoost-Driven Model for Predicting and Analyzing Market Price Trends

Arjun Sharma
Syracuse University, Syracuse, USA
arjunSharma29@syr.edu

Abstract: Accurate prediction of building material prices is crucial for mitigating the impacts of cost fluctuations on project budgets, schedules, and quality. With rising competition and price volatility in the construction industry, traditional prediction models often face challenges such as slow training speeds and high sensitivity to parameter tuning. This study proposes a building material price prediction model utilizing the CatBoost algorithm, a gradient boosting framework optimized for categorical features and robust prediction performance. Using sample data of Φ16-25mm HRB400 type bar steel material prices and relevant influencing factors, the model is evaluated against common machine learning algorithms, including XGBoost and LightGBM. Experimental results demonstrate that the CatBoost-based model outperforms these alternatives in terms of mean squared error (MSE), root mean squared error (RMSE), mean absolute error (MAE), mean absolute percentage error (MAPE), and R^2. The findings highlight the superior prediction accuracy and generalization ability of the proposed model, offering valuable insights for construction cost control and resource planning. Future research will explore broader applications and refine the model to adapt to diverse market dynamics.

Keywords: Price of Building Materials; CatBoost Algorithm; Price Prediction.

1. Introduction

With the continuous development of our construction industry, the competition of building materials is increasingly fierce, while the price of building materials is also rising. The fluctuation of building materials price has an important effect on the result of project cost. When the price of construction materials rises, it will directly increase the project cost level. The improvement of the project cost level will often affect the project duration and construction quality, resulting in safety risks, serious will also be required to reverse work, resulting in the project cost exceeding the standard. At the same time, the price fluctuation of construction materials will also lead to contract disputes, and the owners and contractors will evade their responsibilities due to the difference between the project cost and the expected cost, which will increase the time cost and money cost of the project, so that the project can not be implemented smoothly. Therefore, it is necessary to establish a mathematical model that can be actually tested to predict the price of building materials, so as to reduce the impact of price fluctuations of building materials on the smooth implementation of the project.

At present, the academic circle has a more comprehensive and in-depth study and understanding of the impact of the price fluctuation of building materials on the cost of construction projects. It is recognized that the price fluctuation of building materials will cause a series of negative effects on the smooth implementation of the project. At the same time, the research method is reasonable, the results are scientific and credible, and there are real cases to support it. However, the academic research on the prediction of building materials price is not enough, only a few articles use relatively new technology to predict the price of building materials. For example, the grey system theory

analyzes and predicts the steel bar price by establishing the GM (1,1) grey dynamic model [5]. And BP neural network [8], by training the model with the characteristic value of historical building engineering cost, a model with certain accuracy can be obtained, which can quickly predict the project cost. And the grey neural network PGNN model [7] to predict the price of building materials. Many prediction methods have good prediction accuracy, but the training speed is easy to be reduced and the prediction effect is affected because there are too many super parameter tuning and the processing way of category variables is complicated in the process of data preprocessing. In recent years, with the rapid development of artificial intelligence technology, machine learning algorithm has been widely used in various fields of data analysis and prediction modeling. CatBoost is a GBDT framework based on symmetric decision tree algorithm [9], which can efficiently and reasonably solve category-type features and deal with gradient deviation and prediction migration, and improve the accuracy and generalization ability of the algorithm.

This paper establishes a building material price prediction model based on CatBoost algorithm, takes 30% of the sample data as the prediction set as an example to verify the prediction effect of this model, and compares it with common machine learning algorithms (xgboost regression algorithm and LightGBM regression algorithm). Verify the prediction effect of the construction material price prediction model based on this algorithm.

2. Prediction Algorithm Theory

2.1 Overview of CatBoost Algorithm

Among many machine learning methods, Catboost, developed by Yandex, a Russian search engine company, is a new algorithm based on decision tree [4], which has strong accuracy and scalability.

The main feature of CatBoost algorithm is that it can automatically process category features and missing values. Different from traditional gradient lifting algorithms, CatBoost uses a symmetric tree structure in the training process, and also utilizes weighted loss function, random sorting and statiscy-based category feature processing techniques to improve the accuracy and generalization ability of the model [10][11].

1) Category feature

The basis of Catboost is still the lift tree. Different from the traditional gradient lift tree, Catboost creatively considers the prior distribution term in the calculation of node gain when processing category features, effectively eliminating the influence of low frequency features and noise in category variables on the generation of decision trees.

$$x_{i,k} = \frac{\sum_{j=1}^{p-1}\left[x_{\sigma_{j,k}=x_{\sigma_{p,k}}}\right]\cdot Y_j + a\cdot p}{\sum_{j=1}^{p-1}\left[x_{\sigma_{j,k}=x_{\sigma_{p,k}}}\right] + a}$$

In the formula, σj is the JTH data; xi and k represent the k-th column discrete features of the I-th row data in the training set; a is a prior weight; p is the prior distribution term (for regression problems, the prior term generally takes the mean value of the prediction label in the training set; Here is an indicator function that prints 1 when the internal condition is full and 0 otherwise. With the improved TS method, Catboost is able to convert class features into numerical values with minimal information loss. , effectively eliminate the influence of low frequency features and noise in class variables on the generation of decision trees.

2) boosting by ordering

The traditional GBDT model adopts the way of sampling without rows and rows. All base learners and Cart decision trees carry out gradient lifting on a complete data set, and each iteration uses the negative gradient of the last round of tree for training. This will lead to the accumulation of prediction bias and over fitting. Xgboost and the LGBM developed by Microsoft use row and column sampling and regularization to reduce the effect of overfitting. Catboost further proposed Ordered boosting. The pseudo-code of the algorithm is as follows:

\

Algorithm: Ordered boosting

Input $\{x_k, y_k\}_{k=1}^{n}$ *ordered according to* σ *, the numbers of trees I*

$\sigma \leftarrow$ *randompermutationof* $[1, n]$

$M_i \leftarrow 0 for i = 1 to n$

for iter $\leftarrow 1$ *to I do*

for $i \leftarrow 1$ *to n do*

for $j \leftarrow 1$ *to* $i-1$ *do*

$$g_j \leftarrow \frac{d}{da} Loss(y_j, a) \Big|_{a = M_i(X_j)}$$

$M \leftarrow$ *learn a tree* (X_j, g_j)

$M_i \leftarrow M_i + M$

return $M_1, M_2, \cdots, M_n$

Figure 1. Algorithm pseudocode

Where, σ is the number of random ordering of the training set; I is the number of symmetric decision trees that need to be generated which is the number of learners. Mi is initialized to 0 for all n samples. Then, by sampling the random sequence and obtaining the gradient based on it, the purpose of performing σ-order permutations is to enhance the robustness of the algorithm and effectively avoid overfitting. These permutations are the same as those used to calculate the improved TS. For each random permutation of sigma, n different model Mi as shown above are trained. Then, the gradient gj of the first i-1 data Loss function (Loss) is calculated successively, and i-1 gj is passed into the symmetric tree to establish a residual tree. For each permutation in s permutations, n model Mi is built, and the overall complexity is about O(s×n2). log2(n)), j < 2i+1). Mi'(Xj) is an approximation of the same j based on the first 2i samples. Finally, the predicted complexity of Mi'(Xj) will not be greater than $\sum_{0 \ll i \ll \log_2(n)} 2^{i+1} < 4n$.

1) Quick score

Catboost uses fully symmetric decision trees (ODT) as the base learning device. Its structure is shown in the figure below. Unlike normal decision trees, fully symmetric trees choose exactly the same features and feature thresholds when splitting internal nodes of the same depth. So a completely symmetric tree can also be transformed into a decision table with 2^d entries, where d represents the number of levels in the decision tree. The structure of decision tree is more balanced and the feature processing speed is much faster than the ordinary decision tree. In addition, floating-point features, statistical information (user id, etc.) and unique thermal coding features are uniformly processed in binary [12], which greatly reduces the requirement of parameter adjustment. The following is a completely symmetric tree structure:

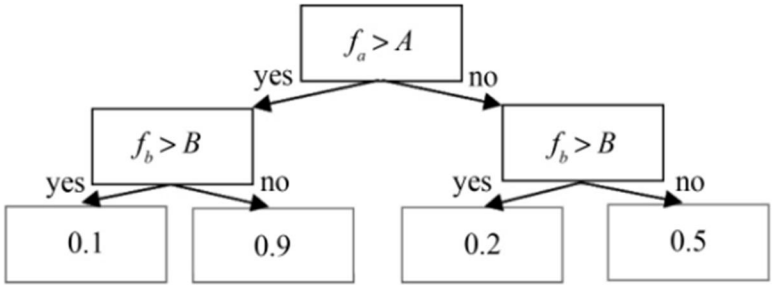


Figure 2. Completely symmetric tree structure

2) Feature importance ranking

CatBoost not only has a high prediction accuracy, but also can screen the relative contribution of different influence factors (i.e. features used in the prediction) to the prediction results. The relative contribution of a feature in a single decision tree can be measured by the following formula.

$$J_j^2 = \frac{1}{M\sum_{m=1}^{M} J_j^2(T_m)}$$

Where, L is the number of leaf nodes of the tree; L-1 is the number of non-leaf nodes in the tree; vt is the feature associated with node t; it2 is the reduction value of the square loss after node t splitting. The more it2 is reduced, the greater the benefit of the splitting, which means that this feature is more important to the feature of the owning node.

2.2 Other Algorithm Theory

1. XGBoost regression algorithm

XGBoost is an efficient implementation of GBDT. Unlike GBDT, xgboost adds regularization terms to the loss function; Moreover, since some loss functions are difficult to calculate derivatives, xgboost uses second-order Taylor expansion of loss functions as the fitting of loss functions [10].

2. LightGBM algorithm

LightGBM is an efficient implementation of XGBoost. Its idea isto discretize continuous floating-point features into k discrete values, and construct Histogram with width of k. Then the training data is traversed to calculate the cumulative statistics of each discrete value in the histogram. In feature selection, we only need to find the optimal segmentation points by traversing according to the discrete values of the histogram. In addition, the leaf-wise strategy with depth limitation saves a lot of time and space overhead [11].

3. Data Introduction and Processing

3.1 Index Selection

Considering the main factors affecting the price of rebar, this paper selects a total of 34 pieces of monthly data from March 2022 to December 2022, and sets 11 indicators such as iron ore output, producer price index, PMI, and national steel production. The specific indicators and contents are shown in Table 1.

Table 1. Index setting of influencing factors of steel bar price

Serial number	Time	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11
1	December-22	3992.6	7861.2	99.5	47	11193.4	540	7360611	276253.35	1564518.19	54691015	21512.38
2	November-22	3823.5	7953.6	100.1	48	10918.6	559	7353231	276253.35	1564518.19	56273188	20336.03
3	October-22	3841.2	7267.5	100.2	49.2	11484.7	518.4	6769961	276253.35	1564518.19	36293757	13189.88
4	September-22	3999.6	7897.6	99.9	50.1	11618.8	498.4	6976623	182976.01	1369479.32	39905355	14930.02
5	August-22	4089.4	8113.3	98.8	49.4	10832.9	615.3	9672933	182976.01	1369479.32	57782037	23112.61

6	July-22	3966.8	8022.7	98.7	49	10624.18	667.1	10717609	182976.01	1369479.32	57485187	22712.9
7	June-22	4203.2	9871.1	100	50.2	11841.71	755.7	11702041	114347.34	1207367.21	44203283	19653.84
8	May-22	4673.1	9780.5	100.1	49.6	12261.1	775.9	10698338	114347.34	1207367.21	38238052	17654.57
9	April-22	4972.9	8579.4	100.6	47.4	11482.7	497.7	7123436	114347.34	1207367.21	37269339	18567.9
10	March-22	4898.6	9476.2	101.1	49.5	11688.72	494.5	6728494	45330.74	1002800.25	44973011	21968.78
11	February-22	4764.9		100.5	50.2		362	4980396	45330.74	1002800.25	30423132	14465.79
12	January-22	4723.8		99.8	50.1		461	6901942	45330.74	1002800.25	27640801	12695.5
13	December-21	4707.1	7851.2	98.8	50.3	11354.8	502.6	8483507	259205.39	1575495.26	52348380	22977.57
14	November-21	4667.1	7839.6	100	50.1	10102.9	436.1	7542200	259205.39	1575495.26	92231255	38910.75
15	October-21	5371.7	8012.7	102.5	49.2	10173.7	450	6872983	259205.39	1575495.26	55631298	28448.75
16	September-21	5684.9	8469.8	101.2	49.6	10195.09	492	7296771	169298.61	1350005.42	53073347	29237.73
17	August-21	5167.9	8391.5	100.7	50.1	10880.4	505.3	7295573	169298.61	1350005.42	58081620	30772.01
18	July-21	5322.8	7992.9	100.5	50.4	11099.7	567	7592665	169298.61	1350005.42	44128200	24212.48
19	June-21	4883.9	8786.9	100.3	50.9	12072.3	646	8287866	105935.77	1193094.65	50563487	25543.1
20	May-21	5058.6	8761.2	101.6	51	12469.4	527.1	6016971	105935.77	1193094.65	58938503	31744.31
21	April-21	5119.2	8689.4	100.9	51.1	12127.5	797	7401178	105935.77	1193094.65	56994979	29475.67
22	March-21	4775.4	8175	101.6	51.9	11987.2	754	6320464	41405.4	973881.9	66165072	31482.2
23	February-21	4616.8		100.8	50.6		490	4169320	41405.4	973881.9	25561138	11317.79
24	January-21	4294.6		101	51.3		524	4676502	41405.4	973881.9	42269431	18339.42
25	December-20	4358.4	7702.1	101.1	51.9	12033.8	485	4588448	232717.11	1494743.36	58747857	24545.32
26	November-20	4058.8	7521.2	100.5	52.1	11734.4	440.2	4069718	232717.11	1494743.36	22781367	8699.77
27	October-20	3758.3	7842.1	100	51.4	11848.3	403.9	3627306	232717.11	1494743.36	16780351	6069.18
28	September-20	3699.7	7348.1	100.1	51.5	11806.3	382.8	3479369	148671.23	1245358.95	26405491	9584.51
29	August-20	3759	7699.9	100.3	51	11913.3	367.8	3323232	148671.23	1245358.95	24742608	9335.52
30	July-20	3720	7302.7	100.4	51.1	11688.5	417.6	3644371	148671.23	1245358.95	28696298	10572.39
31	June-20	3701.6	7831.7	100.4	50.9	11585.1	370.13	3373904.41	88905.04	1120565.45	29602017	10637.94
32	May-20	3699.6	7464.2	99.6	50.6	11452.7	440.1	3663415	88905.04	1120565.45	27464942	9555.63
33	April-20	3550.7	7437.9	98.7	50.8	10701.2	632	4753602	88905.04	1120565.45	35933957	11903.79
34	March-20	3577	7475.7	99	52	9888	647.6	4859733	31379.79	893586.26	46859153	16099.82

Note: A1 is the average market price of rebar (Φ16-25mm,HRB400) -10 days (yuan/ton), A2 is the current value of iron ore production (10,000 tons), A3 is the PPI index (%), A4 is the PMI index _ current value (%), A5 is the current value of national steel production (10,000 tons). A6 is the national steel exports _ current value (10,000 tons), A7 is the national steel exports _ current value (thousands of dollars), A8 is the cumulative value of construction engineering output value (hundreds of millions of yuan), A9 is the construction area of construction enterprises (tens of thousands of square meters), A10 is the spiral steel futures trading volume (hand), A11 is the spiral steel futures trading volume (hundreds of millions of yuan).

3.2 Data Preprocessing

In order to eliminate the differences between different variables in nature, dimension, dimension and other characteristic attributes, the collected data will be Z-score standardized processing. Convert it to a dimensionless relative value. To prevent the loss of the accuracy of the result because the dimension of a single variable is too large. The specific transformation formula of variables is

$$z = \frac{(x - \mu)}{\sigma}$$

Where, z is the standardized data, x is the original data, the mean value of the original data, and the standard deviation of the original data.

4. Learning and Prediction

4.1 Model Construction

With the market price of Φ16-25mm, HRB400 type rebar as the dependent variable Y, and 10 indexes of rebar futures trading volume, national steel export volume, iron ore raw ore output as the independent variable X, CatBoost prediction model was constructed. 70% of the 34 data sets were divided into training sets, and the remaining 30% were used as prediction sets to evaluate the prediction ability of the model. At the same time, the ten-fold cross-validation method was used to avoid the unreasonable division of data sets overfitting problem on training set.

4.2 Model Prediction under Different Eigenvalue Numbers

Input 10 independent variables of standardized data into CatBoost model, and the importance of eigenvalues can be calculated, as shown in Figure 3.

It can be seen from Figure 3 that the current value of iron ore raw ore production occupies the highest proportion of characteristic importance in this model, accounting for nearly 1/4, which has a great influence on the accuracy and establishment of the model. The least influence on the model is the national steel export volume and PMI index, which account for less than 5%. In order to better measure the predictive effect of building material price of CatBoost model, the number of input features will be adjusted in the process of model training according to the weight of features analyzed by CatBoost algorithm, and the changes of various indicators of the model will be observed [15]. Before the next model training, the national steel export volume and PMI index were excluded from the original data.

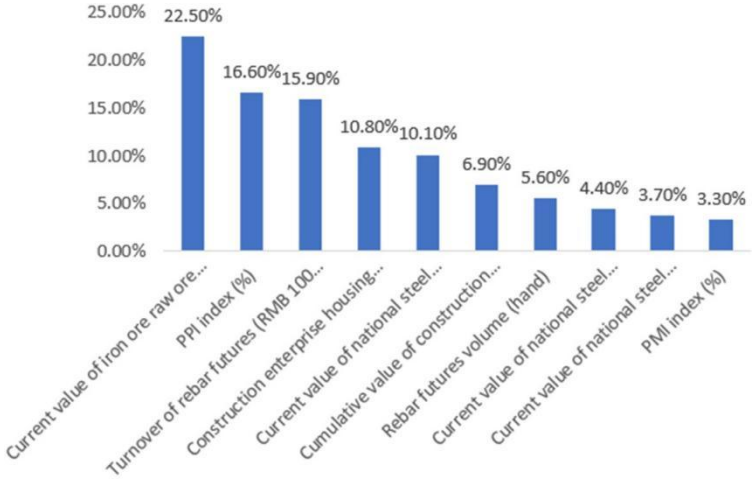


Figure 3. Importance of 10 eigenvalues in catboost model

FIG. 4 to FIG. 6 show the weight of each feature obtained by training under the condition of different number of input features in the adjustment process of the number of input features of CatBoost model.

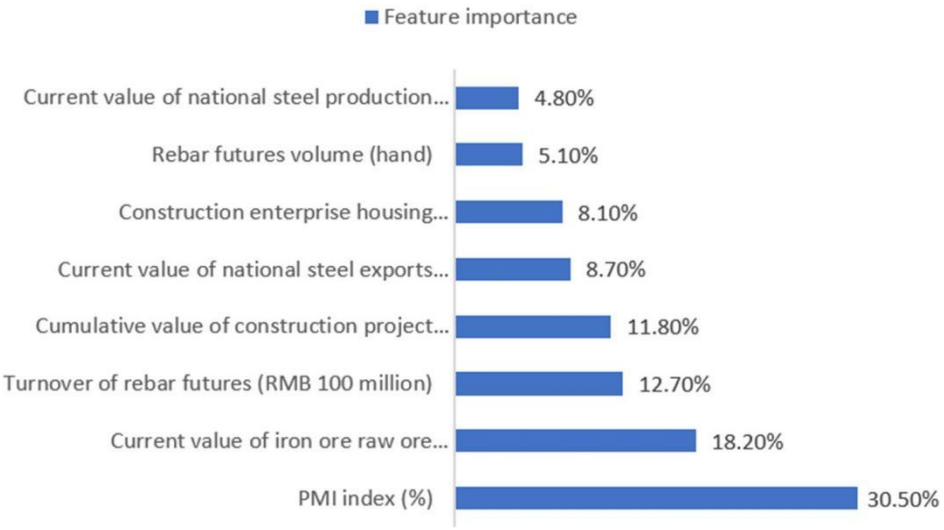


Figure 4. Importance of 8 eigenvalues in catboost model

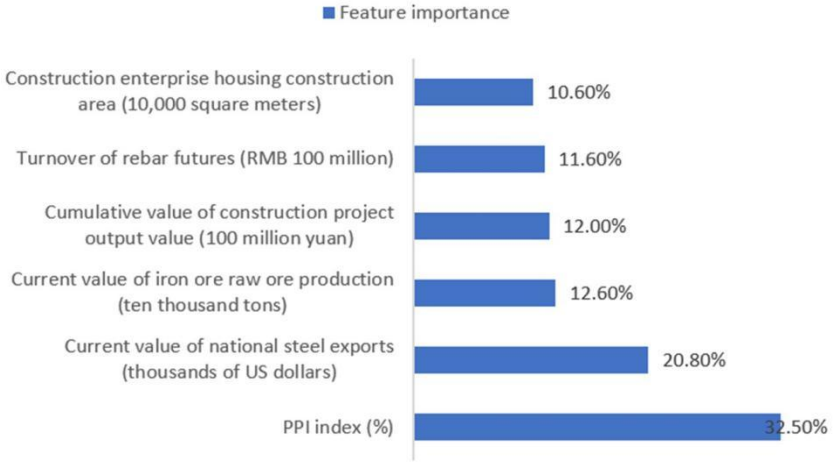


Figure 5. Importance of 4 eigenvalues in catboost model

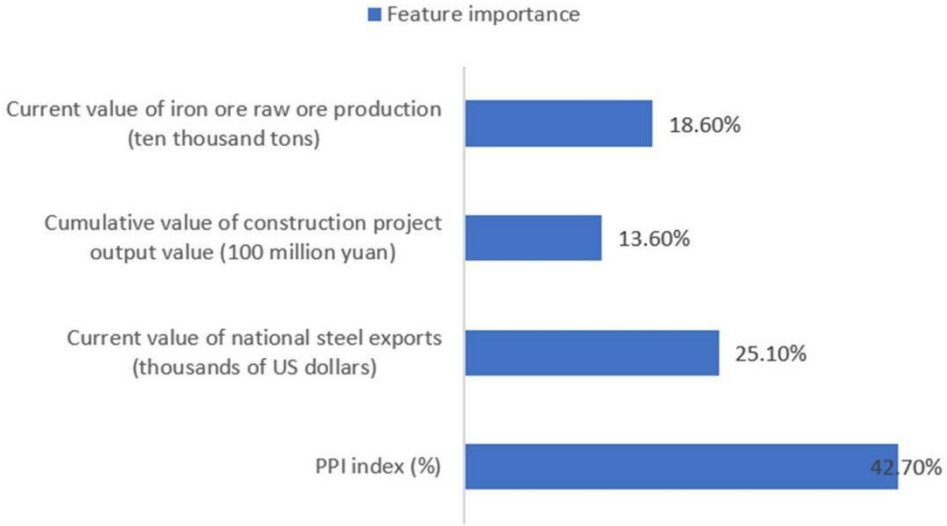


Figure 6. Importance of 6 eigenvalues in catboost model

It can be seen that among the input of 8 characteristic values, PPI index accounts for 30.5%, which has the greatest impact on the accuracy and establishment of the model. The national steel production volume and the steel bar futures trading volume account for 4.8% and 5.1% respectively, which have relatively little impact on the model. Therefore, the two characteristic values of the national steel

production volume and the steel bar futures trading volume are removed. In the case of input of 6 characteristic values, PPI index still accounts for the highest proportion and has the greatest impact on the model results. The housing construction area of construction enterprises and the turnover of steel rebar futures account for relatively small proportion, which is removed. When four eigenvalues were input, the importance of each eigenvalue had little difference, and the PPI index accounted for 42.7%, which was the highest.

The evaluation index results obtained by training CatBoost model when the number of input features is 14, 11, 8 and 5 are shown in Table 2.

Table 2. Comparison of evaluation indexes of price prediction errors with different numbers of characteristic values

	Number of eigenvalues		Evaluation index			
		MSE	RMSE	MAE	MAPE	R squared
Test set	10	0.014	0.12	0.103	27.534	0.784
	8	0.013	0.113	0.094	32.299	0.816
	6	0.007	0.086	0.07	22.932	0.883
	4	0.01	0.101	0.085	31.855	0.866

As shown in Table 2, when the number of eigenvalues is 6, the evaluation index of the test set reaches the best. Therefore, the CatBoost prediction model was established with six characteristic values such as PPI index and national steel exports as independent variables, and the market price of rebar as dependent variable. The results are shown in Figure 7.

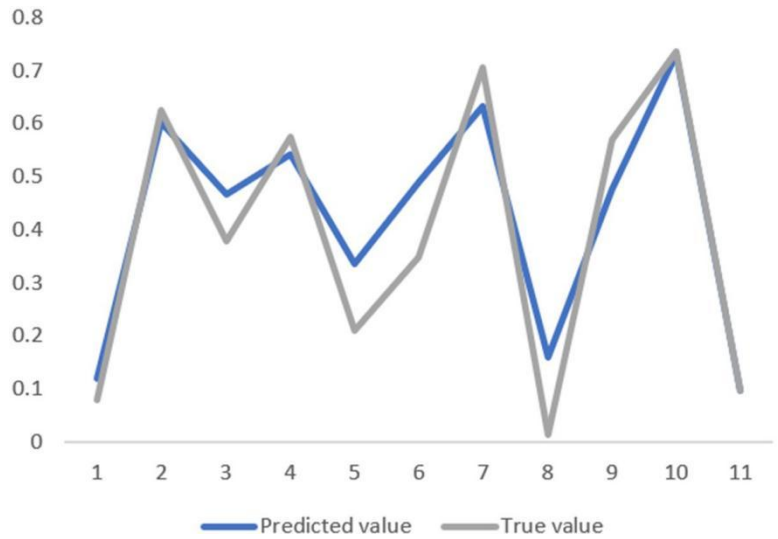


Figure 7. Comparison between the predicted value and the real value of the training data of 6 eigenvalues of CatBoost algorithm

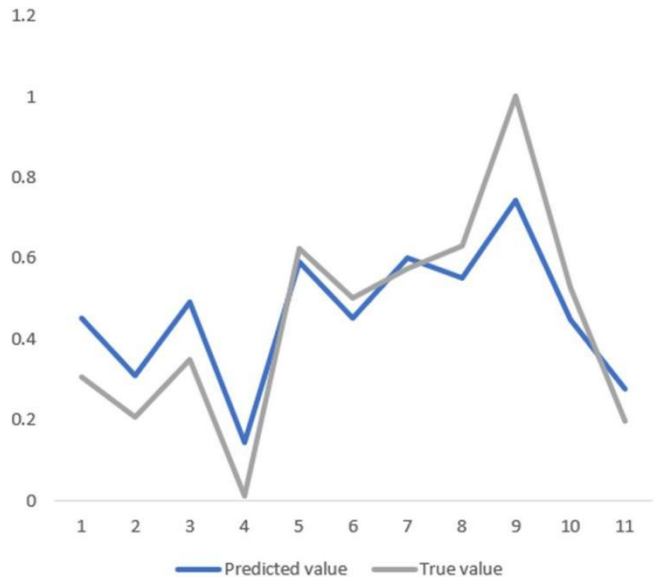


Figure 8. Comparison results between the predicted value and the real value of CatBoost algorithm data with all eigenvalues participating

FIG. 8 shows the comparison result between the predicted value of data and the real value when all 10 eigenvalues participate in model building. Compared with FIG. 7, the predicted result in FIG. 7 is closer to the real value and the model training result is more reliable.

4.3 Model Comparison

Based on the above analysis results, six characteristic values, including PPI index and national steel exports, were selected as independent variables, and the market price of rebar was selected as dependent variable. They were input into XGBoost and LightGBM models for training, and the evaluation indexes obtained by the two models were compared with those obtained by CatBoost model. The results are shown in Table 3.

As can be seen from the comparison results of evaluation indexes in Table 3, CatBoost model is superior to XGBoost model and LightGBM model in 5 evaluation indexes under the condition of input of 6 characteristic values. The MSE, RMSE, MAE, MAPE and R^2 of CatBoost model increased by 0.008, 0.038, 0.016, 18.226 and 0.088, respectively. Compared with LightGBM model, MSE,RMSE, MAE, MAPE, R^2 of CatBoost model increased by 0.006, 0.028, 0.019, 21.2, 0.052, respectively. It can be seen that the prediction accuracy and effect of construction material price based on CatBoost algorithm is higher, which can provide useful reference for construction engineering cost.

Table 3. Comparison of price prediction error evaluation indexes of different algorithms

algorithm	MSE	RMSE	MAE	MAPE	R^2
CatBoost	0.007	0.086	0.07	22.932	0.883
XGBoost	0.015	0.124	0.086	41.158	0.795
LightGBM	0.013	0.114	0.089	44.123	0.831

5. Summary and Outlook

In the process of the development of the construction market, the price of building materials will change with time, external environment and market supply and demand relations, so the accurate prediction of the price of building materials and the trend of volatility is a key step to control the construction cost. In this paper, with the data of Φ16-25mm, HRB400 type bar steel material price and other relevant influencing factors as sample data, a construction material price prediction model

based on CatBoost algorithm is proposed. By optimizing and improving the feature engineering of the construction material price prediction model, the optimal feature number under the ideal prediction results of the model is compared. Compared with common machine learning algorithms (xgboost regression algorithm and LightGBM regression algorithm), the prediction results are better than XGboost regression algorithm and LightGBM regression algorithm in terms of evaluation indexes MSE, RMSE, MAE, MAPE and R^2. Therefore, the prediction ability of building materials price prediction model based on this algorithm is stronger, so it has certain reference significance for the research of building materials price prediction.

References

[1] Shi Baofeng, Li Aiwen, Wang Jing. Research on price discovery function of China Rebar futures Market [J]. Operations Research and Management,2018, 27(6): 162-171.

[2] Li Lintai. Study on Main and Secondary Influencing Factors of the Current Price Rise of Rebar [J]. Price Theory & Practice,2021(6): 85-89.

[3] PROKHORENKOVA L,GUSEV G,VOROBEV A,et al. CatBoost:unbiased boosting with categorical features[C]//proceedings of the Proceedings 32nd Conference on Neural Information Processing Systems(NIPS).Montreal,CANADA:[s.n.],2018.

[4] Chen Dian-dian, Chen Yun-zhi, FENG Xian-feng, Wu Shuang. Remote Sensing Retrieval of river suspended matter concentration based on CatBoost algorithm with Hyperparameter optimization [J]. Journal of Geoinformation Science, 2002,24(04):780-791.

[5] Wang Yanxia. Price prediction of main materials based on GM(1,1) grey dynamic model of tender offer [J]. Railway engineering cost Management,2013,28(05):39-43.

[6] Wang Jiaming, Fan Xuening. Construct the engineering material price forecast based on ARIMA study [J]. Journal of engineering cost management, 2020, No. 171 (01) : 65-75. The DOI: 10.19730 / j.carol carroll nki. 1008-2166.2020-01-065.

[7] Rozemin, Buyuyue. Research on Price Prediction Method of Building Materials Based on Grey Neural Network PGNN Model [J]. Building economy, 2020, 9 (10) : 115-120. The DOI: 10.14181 / j.carol carroll nki. 1002-851 - x., 202010115.

[8] Teng Lingyun. Research on Construction of construction cost prediction Model based on BP Neural network [J]. Housing Industry,2020,No.239(12):110-113.

[9] Gong Hong, Chen Yang, Zhou Chenhui et al. Prediction Model of Postgraduate employability Based on CatBoost Algorithm [J]. Journal of xi 'an university of posts and telecommunications, 2021, 26 (6) : 89-96. The DOI: 10.13682 / j.i SSN. 2095-6533.2021.06.012.

[10] Gu Chongyin, Xu Xiaoyuan, Wang Mengyuan, et al. Fault Diagnosis Method of photovoltaic Array Based on CatBoost Algorithm [J]. Automation of Power Systems: 1-13.

[11] GU Chongyin,XU Xiaoyuan,WANG Mengyuan,et al.Fault diagnosis method of photovoltaic array based on CatBoost algorithm [J].Power System Automation:1-13.

[12] Kong Rui-yao, Xie Tao, Ma Ming, Kong Rui-lin. CatBoost model application in depth inversion [J]. Bulletin of surveying and mapping, 2022 (7) : 33-37. DOI: 10.13474 / j.carol carroll nki. 11-2246. 2022. 0199.

[13] Xu Ling, Jing Xiangnan, Yang Ying et al. Classification and evaluation of national surface water quality based on SMOTE-GA-CatBoost algorithm [J/OL]. China environmental science: 1-11 [2023-04-05]. https://doi.org/10.19674/j.cnki.issn1000-6923.20230221.033.

[14] Wang Xianzhi, Zeng Siming, Zhou Xueqing, Chen Tian-ying, GUO Shao-fei, ZHANG Wei-ming. Photovoltaic power prediction based on XGBoost Combined Model [J]. Acta Energiae Solaris Sinica, 2022, 43(04):236-242. (In Chinese) DOI:10.19912/ J.0254-0096.tynxb.2020-0890.

[15] Ke Guolin, Meng Qi, Thomas Finley, et al. LightGBM: a highly efficient gradient boosting decision tree[C] //In Proceedings of the31st International Conference on Neural Information Processing Systems (NIPS '17), Red Hook: Curran Associates Inc,2017:3149-3157.

[16] Li Xinzhi, Zhang Jianqin, Hu Hao et al. Prediction model of Heavy Metal Contamination Site Risk level based on CatBoost [J]. Green science and technology, 2022, 24 (24) : 140-145 + 151. DOI: 10.16663 / j.carol carroll nki LSKJ. 2022.24.011.