ISSN: 2710-0170

Research on Low-Carbon Economic Reform of Construction Industry Based on Stacking-IGA Modeling

Xulin Chen*, Yiming Wang

School of Southwest Petroleum University, Chengdu 610500, China

*Corresponding Author

Abstract

Low-carbon economy is an economic development model centered on reducing greenhouse gas emissions, which has become an inevitable trend in China. Low-carbon economic reform of the construction industry is an important part of China's realization of low-carbon economy, and the key lies in integrating and coordinating the relationship between low-carbon development and economic development of the construction industry. In order to achieve this goal, this paper aims to minimize the carbon emission and maximize the gross output value of the construction industry, and constructs the Stacking-IGA model, which includes data-driven, Stacking integrated model and improved genetic algorithm. First, based on the data-driven and Stacking integration model, the planning problem of low-carbon economic development target of the construction industry is defined. Secondly, the improved genetic algorithm is used to solve the problem. Finally, the Pareto frontier surface is used to provide a reference basis for the low-carbon economic reform of the construction industry. The research results show that the model circumvents the dependence of the traditional objective planning model on the assumption of the objective function and the assumption of the background conditions, and is suitable for uncertainty objective planning modeling and solving. The model shows good results in sample expansion, sample equalization, feature redundancy, model accuracy and stability, objective solving efficiency and solving quality. It is difficult to realize the reduction of carbon emission and the increase of total output value in China's construction industry at the same time, and to realize the reform of low-carbon economy in the construction industry, we should start from the production of building materials and promote the new construction technology at the same time.

Keywords

low carbon economy; data-driven; goal planning; Stacking integration model; improved genetic algorithm

1. Introduction

Ecological and environmental change is one of the serious challenges facing all countries in the world today. To cope with this challenge, the concept of low carbon economy has emerged. The Kyoto Protocol adopted in Japan in 1997 has led to the formation of the concept of low carbon economy[1] , and the concept of low carbon economy was firstly put forward in the White Paper on Energy issued by the United Kingdom in 2003[2] . A low carbon economy is a development approach that simultaneously considers low carbon development and sustainable economic growth[3] , an economic model based on low energy consumption, low emissions and low pollution[4] , the essence of which is to increase the productivity per unit of carbon and reduce carbon emissions per unit of GDP[5] .

ISSN: 2710-0170

As a responsible big country, China is actively trying the low-carbon economic strategy[6,7] , which is not only in line with the intrinsic needs of China's sustainable development, but also highly compatible with the global trend of green and low-carbon transition set by the Paris Agreement. In terms of low carbon development: China's 14th Five-Year Plan clearly states: "By 2025, energy consumption per unit of GDP will be 13.5% lower than in 2020, total energy consumption will be reasonably controlled, and the total emissions of chemical oxygen demand, ammonia nitrogen, nitrogen oxides, and volatile organic compounds will be reduced by more than 8%, 8%, 10%, and 10%, respectively, compared with the 2020 level. "; in terms of economic development: the construction industry as the cornerstone of the national economy, in 2020, the national construction industry completed the added value of 7,299.6 billion yuan, the industry contribution rate of 9.9%, increasing the industry pull rate of 6.9%, the role of the pillar industry is obvious. In this context, realizing the low-carbon economic development of the construction industry has been China's path towards a low-carbon economy.

In terms of low-carbon development in the construction industry: Du et al.[8] pointed out that carbon emissions from the construction industry are mainly determined by indirect emissions and increase with the increase of industrial GDP. Higher economic growth rate will lead to more carbon emissions, but the carbon intensity will decrease with the increase of economic growth rate; Xu et al.[9] pointed out that the most powerful driver to promote the low-carbon development of the construction industry is the low-carbon laws and regulations, followed by the industrial structure and organizational characteristics related to the development of the industry; Yao et al.[10] pointed out that in the process of the low-carbon transition of the construction industry, the efficacy of the drivers will change over time, and the effectiveness of government intervention will decrease with the low-carbon transition process of the construction industry. The effectiveness of government intervention decreases with the low-carbon transition process of the construction industry, on the contrary, technology and market factors play a dominant role in the later stage of the transition; Fan Jianshuang et al.[11] pointed out that the level of development of the construction industry and the size of the population working in the construction industry are the main contributing factors to the increase of the carbon emissions of the construction industry; Zhao Hongyan et al.[12] pointed out that the increase of the resident population, the rate of urbanization, the output of steel, the average transportation mileage, and the increase of the labor productivity of the construction enterprises will increase the carbon emissions of the construction industry. While the rise of per capita GDP and value added of tertiary industry is conducive to the reduction of carbon emissions from construction, Yuan Runsong et al.[13] pointed out that the progress of energy-saving technology and output technology is an important factor to inhibit the growth of carbon emissions from the construction industry, and at the same time the growth of gross output value of the construction industry, and the reduction of the efficiency of the scale of the construction industry are the main factors of the increase of carbon emissions from the construction industry. Obviously, the carbon emission of the construction industry is affected by many factors, and the analysis of the carbon emission of the construction industry needs to comprehensively consider the technological progress, policy measures, market technology, production of building materials, population and other factors.

[15]In terms of the economic development of the construction industry: Du et al.[14] pointed out that carbon regulation has a great impact on the economic development of the construction industry, and that there is a significant relationship between carbon emissions and the economic growth of China's construction industry, and that the level of economic development of the majority of provinces in China is positively correlated with carbon emissions. Economic growth still promotes carbon emissions; Wang Zhendong et al.[16] point out that environmental regulation affects the cost of regulation as well as the benefits of regulation for construction firms, which in turn affects the development of the firms; Wang Junping et al.[17]

ISSN: 2710-0170

point out that environmental regulation inhibits the economic growth of the construction industry in the short term; An Bowen et al.[18] et al. point out that there are large regional variations in the productivity of China's construction industry, and that the distribution of the output value of the industry is not equal among the various regions of China.

In summary, low carbon economy has become an inevitable trend in China, and the construction industry has a considerable impact on both China's low carbon transition and economic development. There is a certain contradiction between the economic growth of China's construction industry and the carbon emission of the construction industry, economic growth will bring about the rise of carbon emission, at the same time, carbon regulation will cause the shrinkage of economic growth, however, the existing scholars tend to analyze how to realize the low-carbon development or economic development of the construction industry from a single level, and few analyze the two in a coordinated manner, pointing out how to realize the low-carbon economic development of China's construction industry. As a result, in order to coordinate the relationship between low-carbon development and economic development of the construction industry, and to realize the low-carbon economic development of the construction industry, this paper establishes a mathematical model that can comprehensively consider various elements affecting the low-carbon economic development of the construction industry and take into account the dual objectives of low-carbon development and economic development, to provide a clear evaluation criterion for the low-carbon economic development of the construction industry, so as to realize the low-carbon development and economic development of the construction industry. Coordination and optimization.

2. A coordinated optimization model for low-carbon economic development in the construction industry

Objective planning is a commonly used model for solving coordinated optimization problems, and defining the low-carbon economic development model of the construction industry in a traditional way has obvious shortcomings, which requires human-specified weights and parameters, and has a high dependence on the hypothetical background and experimental results, and is able to accommodate fewer features[19-20] ,However, for the low-carbon economic development problem of the construction industry, which involves numerous influencing factors, which may result in the traditional method of defining the objective function cannot accurately reflect the nature of the problem, which in turn leads to the problem of subjectivity, uncertainty and one-sidedness of the experimental results. As a result, this paper utilizes the data-driven and Stacking integration model to construct an objective planning model, and defines a more accurate and objective objective function by analyzing and modeling a large amount of actual data[21~24] . Firstly, in the form of data-driven can avoid the artificial assumption of background conditions, and can analyze more features, and secondly, in the form of Stacking integrated model can avoid the artificial specification of weights and parameters, and reduce the degree of dependence of the experimental results on the objective function model. Among them, Stacking integrated model is an integrated learning method, commonly used in the prediction problem, regression problem, classification problem, etc. feature data fitting[25-27] , which can combine several different base models, make full use of their respective advantages and characteristics, and more accurately define the objective function.

ISSN: 2710-0170

2.1. Data-driven

2.1.1. Data collection

Referring to the research of historical scholars[8-18,28,29] , based on the literature analysis method to collect the characteristic variables affecting the economic development and low carbon development of the construction industry, considering the difficulty of obtaining the data of the characteristic variables and the comprehensiveness of the characteristic variables, we finally collect the data according to the following characteristics.

Table 1 : Analysis of Characteristic Variables

name (of a thing)	serial number	unit (of measure)	name (of a thing)	serial number	unit (of measure)
Gross construction output	Y_1	myriads	Gross value of construction	X_{11}	myriads
Carbon emissions from the construction sector	Y_2	hundred million tons	Level of development of assembled buildings	X_{12}	/
Quality of the construction labor force	X_1	ten thousand dollars	Population size by province	X_{13}	hundred million
Technical assembly rate in the construction industry	X_2	10MW/hour	Per capita disposable income	X_{14}	ten thousand dollars
Number of Assembly Building Patents	X_3	thousands	Provincial GDP	X_{15}	trillion
Average assembly rate	X_4	%	Effectiveness of Assembly Building Policy Advancement	X_{16}	ingredient
Scale of assembly building construction	X_5	hundred million square meters	cement production	X_{17}	hundred million tons
Location entropy index	X_6	/	Steel production	X_{18}	hundred million tons
Percentage of state-owned enterprises	X_7	%	timber production	X_{19}	Million cubic meters
People working in the construction industry	X_8	million people	timing	X_{20}	/
Number of assembly building production bases	X_9	classifier for individual things or people, general, catch-all classifier	as suffix city name, means prefecture or county (area administered by a prefecture level city or county level city)	X_{21}	/
Assembly building production base capacity	X_{10}	cubic meter (unit of volume)			

The data include data on relevant characteristic variables of 30 provinces in China from 2015 to 2019, which are obtained from the National Bureau of Statistics, Prefabricated Building Network, Assembled Building Network, China Carbon Accounting Database, China Statistical Yearbook, China Construction Industry Yearbook, China Building Materials Industry Yearbook and related literature, totaling 150 sets of data and 21 sets of characteristic variables, which is a relatively small scale of data, but contains a characteristic The data size is small, but contains a large number of variables, and the ratio of data sample size to degrees of freedom is 7.1 (150:21), which requires further preprocessing.

ISSN: 2710-0170

2.1.2. Handling of outliers

Data outlier detection is performed using IQR (Interquartile Range) outlier identification method[30] . The principle is to calculate the first quartile, third quartile and interquartile range to get a range to determine whether the data points are outliers or not. Finally, the mean replacement method is used to replace the detected outliers. The implementation process of outlier processing is as follows:

1) Arrange the data set in ascending order and determine the number of data points n .

2) Calculate the first quartile (Q_1): $Q_1 = (n+1)/4$.

3) Calculate the third quartile (Q_3): $Q_3 = 3*(n+1)/4$.

4) Calculate i.e. Quartile Distance (IQR): $IQR = Q_3 - Q_1$.

5) Define the lower and upper bounds: $下界 = Q_1 - 1.5*IQR$, $上界 = Q_3 + 1.5*IQR$.

6) Based on the upper and lower bounds, determine whether the data point is an outlier: if the data point is smaller than the lower bound or larger than the upper bound, it is considered an outlier.

7) Calculate the average value of the data set: Sum all the data in the data set except the outliers and divide by the number of non-outliers to get the average value of the data set.

8) Use of averages in place of outliers

2.1.3. SMOTE-based sample data expansion

In view of the imbalance of the level of development of assembled buildings in Chinese provinces and the problem of small number of provincial samples, which leads to the imbalance of the sample data and the shortage of the number of samples, the SMOTE intelligent algorithm is used to expand and equalize the samples.The SMOTE algorithm is an improved method of the random oversampling technique, which overcomes the shortcomings of the over-fitting of the random oversampling technique, and the basic idea of it is that the samples of a few categories will be analyzed and the samples from the minority class samples are used to synthesize new samples artificially and added to the data set[31] . The specific SMOTE algorithm process is as follows:

1) For each sample in the minority class X_i , calculate its distance to all samples in the minority class sample set in terms of Euclidean distance to obtain its k-nearest neighbor.

2) Set a sampling ratio based on the sample imbalance ratio to determine the sampling multiplicity N. For each minority class sample X_i , randomly select a number of samples from its k-nearest neighbors, assuming that the selected nearest neighbors are X_n .

3) For each randomly selected nearest neighbor X_n , respectively, construct a new sample with the original sample according to the following formula.

$$X_{\text{new}} = X_i + r(X_i - X_n) \tag{1}$$

Where X_{new} , X_i , X_n denote the feature (coordinate) values of new samples, few class samples and their near neighbor samples respectively, and r denotes a random number from 0 to 1.

Table 2 : Analysis of Main Characteristic Data after Abnormal Value Handling and Smote Sample Expansion

variable name	sample size	maximum values	minimum value	average value	(statistics) standard deviation	upper quartile	variance (statistics)
Y_1	225	9.084	0.142	3.06	1.909	3.255	3.646

ISSN: 2710-0170

Y_2	225	6.326	0.049	2.249	1.467	1.915	2.152
X_1	225	0.827	0.412	0.635	0.075	0.624	0.006
X_2	225	2.714	0.372	1.18	0.413	1.163	0.171
X_3	225	3.954	0.005	1.267	1.027	1.086	1.055
X_4	225	0.76	0.12	0.415	0.173	0.449	0.03
X_5	225	3.897	0.724	2.336	0.669	2.371	0.447
X_6	225	1.773	0.279	1.075	0.322	1.076	0.104
X_7	225	0.299	0.01	0.151	0.065	0.159	0.004
X_8	225	4.567	0.074	1.431	0.999	1.243	0.997
X_9	225	50	1	13.7	11.745	12.209	137.938
X_{10}	225	1659.5	14.5	520.784	394.374	477	155530.926
X_{11}	225	5.226	0.055	1.696	1.095	1.826	1.2
X_{12}	225	5	1	3	1.417	3	2.009
X_{13}	225	1.011	0.058	0.455	0.256	0.428	0.066
X_{14}	225	4.205	1.347	2.431	0.578	2.376	0.334
X_{15}	225	7.054	0.201	2.534	1.53	2.478	2.34
X_{16}	225	4.2	0.3	1.333	0.962	1.044	0.925
X_{17}	225	1.806	0.032	0.795	0.509	0.811	0.259
X_{18}	225	0.725	0	0.249	0.156	0.257	0.024
X_{19}	225	6.48	0	1.769	1.548	1.56	2.398
X_{20}	225	2019	2015	2016.907	1.348	2017	1.817
X_{21}	225	30	1	14.283	9.347	13	87.359

2.1.4. Feature redundancy

When highly correlated features are present, the redundant features do not add information to the data, but it has an impact on the confidence of the model, and the effects produced by the redundant features are superimposed on the model, which makes the model performance degrade. In order to ensure that the features are as independent of each other as possible, the features need to be filtered[32] . The smaller the correlation value, the greater the difference between the feature and the rest of the features, so the feature independence weight is calculated by the formula (2), where: λ_1 indicates the size of the independence weight of the first i feature; b_i is the absolute value of the correlation coefficient of the first i feature and the features other than itself (according to the Spearman correlation coefficient to calculate the correlation between the features), which indicates the size of the correlation between the feature and the rest of the features. Main feature independence table:

$$\lambda_i = \frac{\sum_{i=1}^{n} b_i}{n}, \quad n = 21 \tag{2}$$

ISSN: 2710-0170

Table 3: Independence Table of Main Features

serial number	Independence weighting	serial number	Independence weighting	serial number	Independence weighting
X_1	0.3108	X_8	0.5251	X_{15}	0.5747
X_2	0.1996	X_9	0.4786	X_{16}	0.4031
X_3	0.5275	X_{10}	0.4546	X_{17}	0.4124
X_4	0.4244	X_{11}	0.5397	X_{18}	0.4304
X_5	0.1867	X_{12}	0.5317	X_{19}	0.3712
X_6	0.2239	X_{13}	0.4910	X_{20}	0.1331
X_7	0.2940	X_{14}	0.4259	X_{21}	0.4973

According to the formula to screen out the features with poor independence, considering that the number of feature variables after screening will have a certain impact on the Stacking integrated model, too few feature variables will lead to the decline in the accuracy of the Stacking integrated learning model, thus this paper takes 0.5 as the threshold to eliminate redundant features (X_3 , X_8 , X_{11} , X_{12} , X_{15}), and makes a heat map of feature variables:

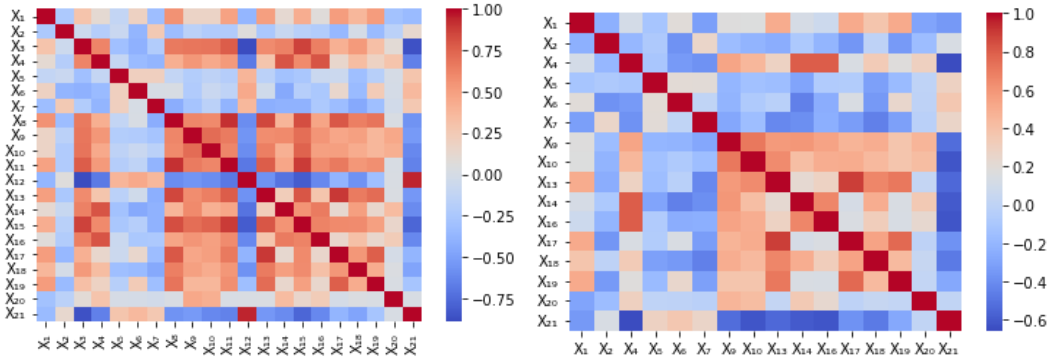

Figure 1 (a): Thermal diagram before removal

Figure 1 (b): Thermal diagram feature redundancy after feature redundancy removal

Figure1: Thermal Diagram of Characteristic Variables

Analyzing Fig. (1), it can be seen that there are a large number of highly correlated feature variables and highlighted blocks in the heat map before feature redundancy, which indicates that there is a large amount of repetitive information among these feature variables. However, after feature redundancy, the number of highlighted blocks decreases significantly and the overall color of the heat map becomes lighter, which indicates that the repetitive information among the feature variables is reduced significantly. By using equation (2) and setting 0.5 as the threshold for feature redundancy, a better feature redundancy effect can be achieved.

2.1.5. Data set segmentation

After the Smote algorithm processing as well as feature de-redundancy processing, the data sample size to degrees of freedom ratio is 14, which is an improvement of about 100%, and the data morphology is substantially improved. The purpose of this paper is to study how to realize the low-carbon economic development of the construction industry, with the goal of minimizing the amount of carbon emissions from the construction industry as well as maximizing the total output value of the construction industry, to coordinate the economic development of the construction industry and the coordination of low-carbon development,

ISSN: 2710-0170

and for this reason, the dataset is divided into dataset A and dataset B, which provides the data base for the Stacking integration model.

Dataset A: $\{Y_1 : X_i\}_{i=1\sim2,4\sim7,9\sim10,13\sim14,16\sim21}$

Dataset B: $\{Y_2 : X_i\}_{i=1\sim2,4\sim7,9\sim10,13\sim14,16\sim21}$

2.2. Stacking Integration Model

Stacking is an integrated learning method in machine learning, which is used to combine multiple different base models to build a more powerful predictive model. It is a meta-learning method that can combine the predictive power of multiple models to improve the overall generalization ability, which not only avoids the problem of poor performance of traditional machine learning models on different datasets and problems, but also makes full use of the advantages of each base model to reduce the bias and variance of the model. It not only avoids the problem of poor performance of traditional machine learning models on different datasets and problems, but also makes full use of the advantages of each base model to reduce the bias and variance of the model and improve the prediction performance and generalization ability.

The basic idea of Stacking is to take the predictions of multiple base models as input features and train them using real labels in order to construct a secondary learner. First, the original dataset is divided into a training set and a test set. Then, in the training phase, the training set is divided into several folds, and each fold is used to train a different base model. Each base model is trained using the training data and generates predictions for the test set. These prediction results will be used as input features along with real labels to train the secondary learners. Finally, in the testing phase, each base model generates prediction results for the test set, which will be fed into the secondary learners for final prediction.

2.2.1. Model evaluation criteria

(1) R^2 called the coefficient of determination

R2 is a measure of the proportion of the characteristics that can be explained by the independent variable, and thus the performance of the regression model. r2 compares the predicted values to the case where only the mean is used, and the closer the result is to 1 the more accurate the model is. Let the mean true value $\bar{y}$ be:

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i \tag{3}$$

The total sum of squares SS_{tot} is:

$$SS_{tot} = \sum_{i=1}^{n}(y_i - \bar{y})^2 \tag{4}$$

The residual sum of squares SS_{res} is:

$$SS_{\text{res}} = \sum_{i=1}^{n}(y_i - \hat{y}_l)^2 \tag{5}$$

Thus, the R-square (R^2) is:

$$R^2 = 1 - \frac{SS_{\text{res}}}{SS_{\text{tot}}} \tag{6}$$

(2) Root mean square error

MAE can reflect the actual situation of the prediction value error, the smaller the value, the higher the accuracy of the model. Its calculation formula is:

ISSN: 2710-0170

$$MAE = \frac{\sum_{i=1}^{n} |y_i - \hat{y}_l|}{n} \tag{7}$$

where y_i is the true value and $\hat{y}_l$ is the predicted value, which avoids positive and negative canceling each other out due to absolute valorization.

(3) Root mean square error

MSE is used to measure the degree of difference between predicted and observed values in regression analysis, the larger the MSE, the larger the difference between predicted and observed values, indicating that the regression model is poorly fitted; the smaller the MSE, the smaller the difference between predicted and observed values, and the better the model is fitted. The formula is:

$$\mathrm{MSE} = \sqrt{\frac{\sum_{i=1}^{n} (P_i - O_i)^2}{n}} \tag{8}$$

where P_i is the predicted value of the ith observation in the dataset, O_i is the measured value of the ith observation in the dataset, and n is the number of samples.

2.2.2. Model comparison

(1) Base Model Comparison: The base model of Stacking plays an important role in integrated learning, which can learn different feature representations, generate meta-features, and reduce the risk of overfitting to provide more comprehensive as well as more accurate prediction ability for Stacking integration model, thus improving the generalization ability and prediction performance of Stacking integration model. In order to improve the performance of Stacking integration model, this paper uses a comparison from: Random Forest (RF) model, Extra Trees model (ET), Gradient Boosting Decision Tree (GBDT), CatBoost model, LightGBM model, AdaBoost model, and XGBoost model for base model comparison, and the specific comparison results are as follows:

Table 4 : Comparison Table of Dataset A-based Models

Data set A	RF	ET	Adaboost	XGB	Catboost	GBDT	LightGBM
R2	0.928	0.962	0.909	0.928	0.941	0.925	0.939
MAE	0.296	0.203	0.437	0.298	0.254	0.317	0.319
MSE	0.273	0.144	0.343	0.270	0.221	0.282	0.230

Table 5: Comparison and Selection Table of Dataset B-based Models

Data set B	RF	ET	Adaboost	XGB	Catboost	GBDT	LightGBM
R2	0.903	0.914	0.833	0.825	0.837	0.873	0.787
MAE	0.307	0.276	0.511	0.361	0.395	0.364	0.426
MSE	0.203	0.179	0.348	0.365	0.340	0.265	0.444

Firstly, from the R2 metrics, it can be seen that: on dataset A, the ET, Catboost and LightGBM models have relatively high R2 values of 0.962, 0.941 and 0.939, respectively. whereas on dataset B, the ET, RF and GBDT models have higher R2 values of 0.914, 0.903 and 0.873, respectively. these models on the corresponding dataset can fit the target variables better and can explain a larger proportion of the variance. Second, from the MAE metrics, it can be seen that the ET, Catboost and RF models have small MAE values of 0.203, 0.254 and 0.296 on dataset A. On dataset B, the ET, RF and GBDT models have relatively small MAE values of 0.276, 0.307 and 0.364, respectively. the predictive accuracies of these models on corresponding datasets

ISSN: 2710-0170

are higher prediction accuracy. Finally, from the MSE metrics, it can be seen that the ET, Catboost, and LightGBM models have relatively small MSE values of 0.179, 0.221, and 0.230, respectively, on dataset A. On dataset B, the ET, RF, and GBDT models have relatively small MSE values of 0.179, 0.203, and 0.265, respectively. the predictive accuracy of these models is higher on the corresponding dataset have better prediction performance. As a result, ET, Catboost, and LightGBM are used as base models for dataset A, and ET, RF, and GBDT are used as base models for dataset B.

(2) Metamodel Comparison: Stacking's metamodel is the model used to integrate the base model, and its main role is to combine the prediction results of the base model to generate the final integrated prediction results.Stacking's metamodel plays an important role in the integrated learning, which improves the performance and prediction capability of the integrated model by combining the prediction results of the base model, learning the weighting strategy, dealing with the nonlinear relationship, and avoiding overfitting. the performance and prediction ability of the integrated model. In order to improve the performance of the Stacking integrated model, this paper adopts a comparative selection method from: Linear Regression, Support Vector Regression (SVR), k-Nearest Neighbors (KNN), Random Forest, Gradient Enhancement, and Gradient Enhancement. Forest, Gradient Boosting Decision Tree (GBDT), and XGBoost model for meta-model comparison, the results are as follows:

Table 6: Dataset A Metamodel Comparison Table

Data set A	linear regression (math.)	SVR	KNN	RF	GBDT	XGB
R2	0.9655	0.9618	0.9610	0.9644	0.9697	0.9577
MAE	0.1959	0.2091	0.2441	0.2184	0.2358	0.2642
MSE	0.1303	0.1441	0.1471	0.1343	0.1142	0.1597

Table 7: Dataset B Metamodel Comparison Table

Data set B	linear regression (math.)	SVR	KNN	RF	GBDT	XGB
R2	0.9135	0.9123	0.9010	0.8697	0.8290	0.8329
MAE	0.2821	0.2931	0.3345	0.3641	0.4044	0.4001
MSE	0.1807	0.1831	0.2068	0.2722	0.3572	0.3491

First, from the R2 metric, it can be seen that: on dataset A, the GBDT model has the highest R2 value of 0.9697. On dataset B, the linear regression model has the highest R2 value of 0.9135. Second, from the MAE metric, it can be seen that: on dataset A, the linear regression model has the smallest MAE value of 0.1959. And on dataset B, the linear regression model has the smallest MAE value of 0.2821. Finally, from the MSE metrics, it can be seen that the XGB model has the smallest MSE value of 0.1142 on dataset A. The linear regression model has the smallest MSE value of 0.1807 on dataset B. The predictive performance of these models is better on the corresponding datasets. Thus dataset A has GBDT as the metamodel and dataset B has linear regression as the metamodel.

2.2.3. Stacking Integration Models

Based on the model comparison, the Stacking integration model is established.

(1) Data set division. In this paper, 225 sets of datasets expanded by SMOTE algorithm are randomly sampled, and the test set and training set are divided according to 2:8, the training set is used for model training, and the test set is used for training model validation.

(2) Base model training process. A 5-fold cross-validation is used to avoid model overfitting, i.e., during base model training, the training data are divided into 5 equal parts, 4 of them are used to train the base model, and the trained model is used to predict the remaining 1 part of the data, and a total of 5 predictions of the base model are obtained.

ISSN: 2710-0170

(3) Metamodel training process. The combination of prediction results and actual values obtained from the base model is used as a new training set for the learning and training of the meta-model, and the prediction results of the base learner are given weights (η_n), so that the prediction results can be more accurate, and thus, the training process is finished.

(4) Model test output process.

2.2.4. Level of importance of features

Analyzing feature importance can provide insight into how models make predictions based on different features, leading to better explanations of model behavior and prediction results. Additionally, feature importance analysis can provide insights about key factors and mechanisms in the problem domain that can help provide a deeper understanding of the data and the nature of the problem.

This paper analyzes the feature importance based on the base model weighting method. The principle is that in Stacking integration model, each base model contributes to the final prediction result, and these base models will have different weights. By calculating the weight of each base model, the importance of features can be evaluated. If a feature gets a higher weight in multiple base models, it means that this feature has a higher impact on the final result. The specific calculation formula is as follows:

$$\xi_i = \sum_{n=1}^{3} \eta_n * \gamma_{(n,i)} \tag{9}$$

Where ξ_i denotes the feature importance value of the ith feature, $\gamma_{(n,i)}$ denotes the feature importance value of the ith feature in the nth base model, and η_n is the integration weight of the nth base model.

2.3. Coordinated Optimization Model Establishment for Low-Carbon Economic Development in the Construction Industry

Adopting Stacking stacking model, using dataset A and dataset B to define the construction industry economic development objective function and the construction industry low carbon development objective function respectively, and co-ordinate this into the construction industry low carbon economic development objective planning function, which is used to study how to realize the low carbon economic development of the construction industry.

$$\begin{cases} Max \quad Y_1 = F(X_i) \\ Min \quad Y_2 = G(X_i) \\ st \quad X_i \in R \quad i=1\sim 2,4\sim 7,9\sim 10,13\sim 14,16\sim 21 \end{cases} \tag{10}$$

Where Y_1 is the economic development goal of the construction industry, Y_2 is the low-carbon development goal of the construction industry, and X_i is the residual feature variable after feature de-redundancy.

Comparing with Literature 19, Literature 20, the traditional multi-objective planning model requires artificial assumptions on background conditions, feature parameters, and objective functions, and is more dependent on experimental assumptions, whereas the data-driven Stacking-based integrated model circumvents the artificial specification of weights and parameters, is suitable for objective planning modeling in uncertain environments, and is capable of accommodating more features, and is less dependent on assumptions of background and experimental results Low.

Comparative analysis with literature 21, literature 22, literature 23, and literature 24 shows that when scholars adopt data-driven integration with goal optimization, they often ignore the correlation between feature parameters, which may lead to errors in the experimental results,

ISSN: 2710-0170

and the assumed objective function lacks adaptability, ignoring the changes in the relationship between the target feature variables and the objective function caused by the dynamic changes and uncertainties of the environment, whereas this paper proposes a data-driven integration model based on the data-driven Stacking integration model, not only deals with the correlation problem among feature variables, but also makes the goal planning model more universal due to the assumption of the objective function by using a machine learning model.

3. Coordinated Optimization Model Solution for Low Carbon Economic Development in Construction Industry

The bi-objective planning problem constructed by the data-driven and Staking integration model contains multiple feature variables without any conditional constraints. Genetic algorithm is one of the commonly used optimization algorithms in solving such problems[33-35] . Genetic algorithm is an optimization algorithm that simulates the evolutionary process, which searches for the optimal solution by mimicking the mechanism of natural selection and genetic inheritance, and has been widely used in the fields of combinatorial optimization, machine learning parameter optimization, scheduling problems, and engineering design. In solving multi-objective planning problems, genetic algorithms have the advantages of maintaining diversity, non-inferior solution set generation and expandability. For this unconstrained bi-objective planning problem, the genetic algorithm is improved by simulated binary crossover, random restart mutation, elite strategy and tournament selection method to ensure the convergence speed and retrieval effect of the algorithm.

In this paper, we set the initial population as 100, the maximum number of iterations as 30, the crossover frequency as 0.9, and the mutation frequency as 0.1, and improve the genetic algorithm by using the improved genetic algorithm with simulated binary crossover, random restart mutation, elite strategy, and tournament selection method to improve the performance of the model.

3.1. Elite Strategy

The elite strategy of genetic algorithm refers to the method of retaining the optimal individuals of the previous generation in the evolutionary process of each generation. By selecting and retaining the optimal individuals, it enables the optimal individuals to remain in the population and pass on their excellent genetic information to the next generation, which is realized by directly copying the optimal individuals to the next generation of the population without crossover and mutation operations when the selection operation is carried out.

By retaining the optimal individuals, the elite strategy can avoid the risk of losing the excellent solution due to the randomness of the selection operation, thus effectively preventing the algorithm from converging to the suboptimal solution prematurely, which can improve the search ability and convergence speed of the genetic algorithm, and at the same time, the elite strategy can also maintain the diversity of the population.

3.2. Tournament selection method

Tournament selection method in genetic algorithms is a method used to select individuals that mimics the competition of a tournament. In the tournament selection method, a number of individuals are randomly selected from the population as competitors, and then the best-performing individual from these competitors is selected as the parent and used to generate the next generation of the population. This is accomplished by randomly selecting a fixed-size population of competitors, choosing the best-adapted individual from the population of competitors as the parent, and so on until a sufficient number of parents are selected.

The significance of the tournament selection method is to increase the variety and randomness of selection operations. The tournament selection method avoids overly focusing on a few

ISSN: 2710-0170

individuals with high fitness while not neglecting individuals with low fitness. Each selection is carried out in a population of competitors, which can better explore the diversity in the population and improve the search efficiency of the algorithm. At the same time, the tournament selection method also has a certain effect of pressure selection. By setting the size of the competitor group, the probability that the more adapted individuals are selected in the selection operation can be controlled. When the competitor group is small, the individual with higher adaptability is more likely to win in the tournament and then become the parent. This accelerates the propagation and retention of superior genes and helps improve the convergence and search quality of the algorithm.

4. Model Testing

4.1. Analysis of the effectiveness of the SMOTE algorithm

Comparison of the data set before and after processing of SMOTE algorithm is shown in Table 8, the total number of sample set is expanded from 150 to 225, with an expansion rate of 50%, the number of samples after processing is improved significantly, the sample ratio of each category of assembly building development level is changed from 4:5:7:9:5 to 1:1:1:1:1, the number of samples in the sample set and the degree of balance after processing is improved, and it provides data basis for the establishment of the prediction model. Provide data basis.

Table 8: Comparison of Data before and after Smote Algorithm Processing

data set	Original data set	SMOTE algorithm processed dataset
Total sample set	150	225
balance	4:5:7:9:5	1:1:1:1:1:1

For dataset A, before applying the Smote algorithm, the model has an R2 score of 0.9461, an MAE of 0.2906, and an MSE of 0.1686. whereas after applying the Smote algorithm, the model's R2 score improves to 0.9697, and the MAE decreases to 0.2358, and the MSE decreases to 0.1142. For dataset B, similarly, the model's R score is 0.7802, and the MSE is 0.4112, and the MSE is 0.4819 before applying the Smote algorithm, the R2 score of the model is 0.7802, MAE is 0.4112, and MSE is 0.4819. whereas, after applying Smote algorithm, the performance of the model improves significantly and the R2 score increases to 0.9135, the MAE reduces to 0.2821, and the MSE reduces to 0.1807. this indicates that by using Smote algorithm to balance the sample category imbalance problem, the prediction accuracy of the model on dataset A, and dataset B is significantly improved.

Table 9: Performance Analysis of Stacking Model Driven by Data before and after Smote

data set	Pre-Smote			Post Smote		
	R2	MAE	MSE	R2	MAE	MSE
Data set A	0.9461	0.2906	0.1686	0.9697	0.2358	0.1142
Data set B	0.7802	0.4112	0.4819	0.9135	0.2821	0.1807

4.2. Feature De-Redundancy Effectiveness Analysis and Stacking Integration Model Performance Analysis

Table 10: Analysis of feature redundancy removal effectiveness

	R2	RF	ET	Adaboost	XGB	Catboost	GBDT	LightGBM	Stacking
Data set A	pre-redundancy	0.9111	0.9548	0.9189	0.9075	0.9685	0.9076	0.9303	0.9344
	after redundancy	0.9277	0.9619	0.9092	0.9285	0.9415	0.9253	0.9392	0.9697

ISSN: 2710-0170

Data set B	pre-redundancy	0.6006	0.7597	0.4714	0.8297	0.6739	0.7678	0.6477	0.8144
	after redundancy	0.9028	0.9143	0.8334	0.8255	0.8374	0.8733	0.7874	0.9135

(1) Analysis of the effectiveness of feature redundancy

For dataset A, the R2 values after feature redundancy are improved in most models. In particular, in RF, ET, XGB, Catboost, LightGBM and Stacking models, the R2 value after redundancy is significantly higher than that before redundancy, with Stacking model having the highest enhancement rate of 3.78%. However, in the Adaboost model, the R2 value after redundancy is slightly lower than before redundancy. This may be due to the fact that in these models, certain redundant features actually contribute to the predictive power of the model, and removing them leads to a decrease in the performance of the model, but the overall performance of the model remains at a high level.

For dataset B, the effect of feature de-redundancy on the models is more pronounced, and the R2 values after de-redundancy are generally higher than those before de-redundancy in all the models, with the RF model having the highest enhancement rate of 52.21%.

In summary, in most cases, feature de-redundancy has a positive effect on the model, and removing redundant features improves the performance of the model.

(2) Performance Analysis of Stacking Integration Model after Feature De-redundancy

In dataset A, the Stacking integration model has a significant performance improvement relative to the other single models. The accuracy of each single model is above 0.9, while the Stacking integrated model has a higher accuracy of 0.9697, which is about 0.0282 compared to the best single model (XGB), which is an improvement of about 2.82%, although this improvement is relatively small, it still indicates the performance advantage of Stacking on this dataset.

In dataset B, the Stacking integration model has a notable performance improvement over the other single models.The accuracy of the Stacking integration model is 0.9135, which is an improvement of about 0.0117 compared to the best single model (Adaboost), which is an improvement of about 1.17%.

Taken together, in both datasets, the Stacking integration model achieved some degree of performance improvement relative to a single model after feature de-redundancy. It shows that the Stacking integration model is able to utilize the strengths of multiple models to combine predictions, thus improving the overall accuracy.

4.3. Cross-validation

In the Stacking model, the prediction results of the Stacking model may fluctuate significantly across datasets due to the differences between base models and the randomness of dataset partitioning. Therefore, stability analysis is needed to assess the consistency and reliability of the model. Stability analysis is an assessment of the consistency of the performance of an integrated learning model on different datasets to determine whether its prediction results are stable and credible.

Table 11: Model Stability Analysis Table

50% off cross-checking	Stacking Integration Model			cross-validation		
	R2	MAE	MSE	R2	MAE	MSE
Data set A	0.9697	0.2358	0.1142	0.9003	0.4102	0.3576
Data set B	0.9135	0.2821	0.1807	0.810108	0.4005	0.4078

For dataset A, the R2 value of the original model is 0.9697, and the R2 value of the Stacking model in cross-validation is 0.9004. It can be seen that the Stacking model is slightly lower than

ISSN: 2710-0170

the R2 value of the original model in cross-validation, which indicates that the Stacking model has a stronger stability over different data subsets relative to the original model.

For dataset B, the R2 value of the original model is 0.9135, and the R2 value of the Stacking model in the cross-validation is 0.8101. The R2 value in the cross-validation is relatively low, but the overall level still remains high, and the stacking integration model has a certain stability.

4.4. Performance analysis of improved genetic algorithm

The horizontal coordinate in Fig. 2 indicates the number of iterations, and the vertical coordinate indicates the optimal fitness value. After the improved genetic algorithm iterates to 11 times, the optimal fitness value stabilizes and the genetic algorithm converges. This is because in each generation of selection operation, the excellent individuals have a greater probability of being selected and thus retained to participate in the reproduction of the next generation, and at the same time, the crossover and mutation operation introduces new gene combinations and increases diversity, which helps to jump out of the local optimal solution and move towards the global optimal solution.

The horizontal coordinates in Figure 3 represent individuals and the vertical coordinates represent crowding degrees. The crowding degree gap between genetic individuals does not exceed 2.5, the crowding degree gap between individuals is not large, and genetic individuals with higher crowding degrees account for the majority. Individuals with larger crowding distances indicate that the closer they are to the boundary, the higher their diversity, which shows that the improved genetic algorithm has a good performance in maintaining the diversity of individuals.

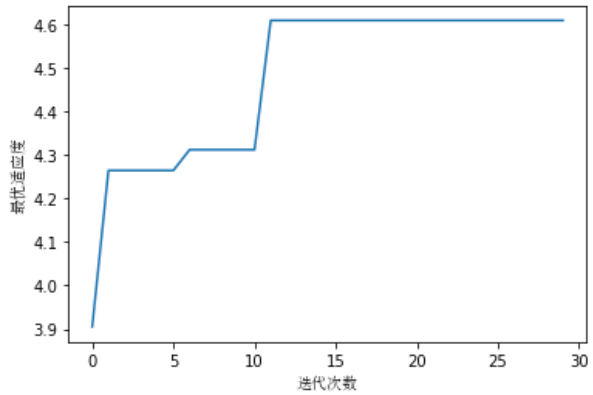

Figure 2: Convergence Trend Chart

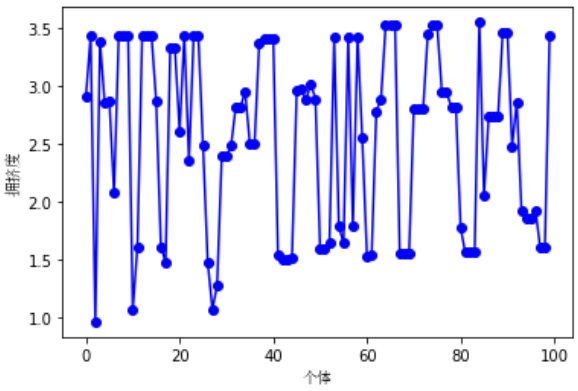

Figure 3: Congestion Ranking Chart

The improved genetic algorithm using the elite strategy and tournament selection method performs well in maintaining individual diversity, accelerating the convergence of the genetic algorithm, and jumping out of local optima.

ISSN: 2710-0170

5. Analysis of results

5.1. Feature importance analysis

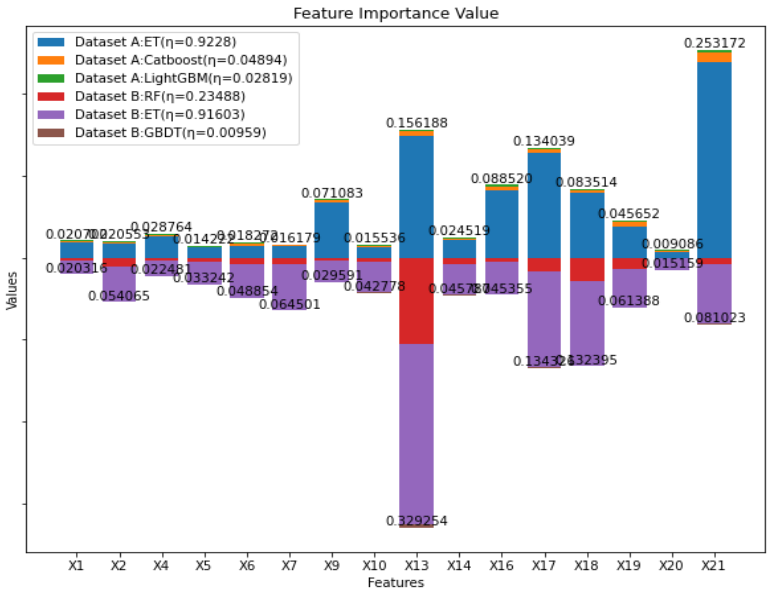

Figure 4: Analysis of Feature Importance

The horizontal coordinates in Figure 4 indicate the different feature classifications and the vertical coordinates indicate the importance values of the different features, reflecting the weights of the different base models and the importance of the different features.

Similar to the findings of Literature 8 to Literature 13, the population size of each province (X_{13}), cement production (X_{17}), steel production (X_{18}), the proportion of state-owned enterprises (X_7), and timber production (X_9) are the key factors influencing the carbon emissions of the construction industry. Similar to the findings of Literature 18, Literature 28, the population size of each province, cement production, steel production, the number of assembly building production bases (X_9), and the region (X_{21}) are the key factors affecting the total output value of the construction industry. It is undeniable that provincial population size, region, cement production, steel production and other factors have a considerable impact on both the gross value of construction industry and the amount of carbon emissions from the construction industry.

Regional factors, the number of assembly building production bases (X_9), and the effectiveness of assembly building policy promotion (X_{16}) have a significantly greater impact on the gross output value of the construction industry than on the carbon emissions of the construction industry. Regional factors have a greater impact on the total output value of the construction industry, probably because economically developed regions may have more construction projects and investments, which drive the growth of total output value, and developed regions tend to adopt more advanced construction technologies, which leads to their ability to maintain a lower level of total carbon emissions despite the increase in the number of projects. The number of assembly building production bases and the effectiveness of assembly building policy promotion may contribute to the development of assembly buildings, thus affecting the total output value of the construction industry, however, they are mainly related to the production methods of construction and the strength of policy implementation, and do not directly involve carbon emission-related factors such as energy consumption, material

ISSN: 2710-0170

utilization and construction methods in the process of construction, thus resulting in a smaller impact on the carbon emissions of the construction industry.

Population size of each province, technical assembly rate of the construction industry (X_2), location entropy index (X_6), proportion of state-owned enterprises (X_7), and production capacity of assembly building production bases (X_{10}) have a significantly greater impact on carbon emissions from the construction industry than on the total output value of the construction industry. The population size of each province represents the difference in population size, which is closely related to the demand for construction, and an increase in construction demand leads to an increase in construction activities, which in turn brings more carbon emissions. This shows the variation in the impact of regional and demographic factors on the total value of construction output and carbon emissions. Carbon emissions are related to the production process of the construction industry, which is mainly related to energy consumption, material use and construction methods. Factors such as the technical assembly rate of the construction industry and the production capacity of assembly building production bases directly affect the energy consumption and material utilization efficiency in the construction process, and thus have a more significant impact on carbon emissions. The location entropy index reflects the degree of aggregation of construction enterprises, and the degree of aggregation will affect the frequency of construction activities and thus lead to changes in carbon emissions. The higher share of state-owned enterprises in the construction sector may mean that the government has greater control and can adopt more stringent environmental policies and emission reduction measures, thus having a more significant impact on carbon emissions.

5.2. Pareto frontier analysis

The horizontal coordinate in Fig. 5 indicates the gross output value of the construction industry, and the vertical coordinate indicates the carbon emission of the construction industry. The convergence effect of Pareto Front in the figure is good, the improved genetic algorithm runs in a more satisfactory way, and the quality of the objective solution is not bad.

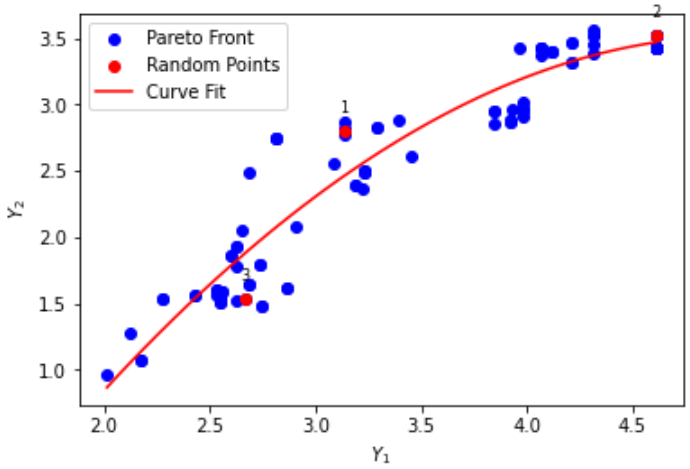

Figure 5 : Pareto Frontal Surface

Obviously it is difficult to realize the reduction of carbon emission and the increase of total output value in China's construction industry at the same time, with the growth of total output value of the construction industry, the weight of carbon emission of the construction industry increases, and coordinating the economic development of the construction industry with the low carbon development is a problem that China must face. The Curve Fit in the figure is made by fitting the Pareto Front with higher-order curves, which shows the best relationship curve

ISSN: 2710-0170

between carbon emission and total output value of the construction industry that can be achieved by eliminating the characteristic variables X_3 , X_8 , X_{11} , X_{12} , X_{15} , and the remaining characteristic variables, which provides a theoretical basis for the low-carbon economic development of China's construction industry, and the decision makers can make trade-offs according to the actual situation.

Combined with the analysis in Fig. 6, it can be seen that when the decision maker's decision does not satisfy the optimal Pareto solution, it is more reasonable to start adjusting from the factors that have a greater impact on the target planning, such as the cement production, the steel production, the effectiveness of the policy of advancing the assembled building, the number of production bases of assembled building, and the timber production (X_{19}).

The following table shows the corresponding eigenvalues for some of the Pareto front points:

Table 12 : Pareto Frontier Display Table

Number	Point 1	Point 2	Point 3	Number	Point 1	Point 2	Point 3
X_1	0.02388	0.42093	0.84920	X_{13}	0.77570	0.92247	0.44339
X_2	0.62552	0.20049	0.47623	X_{14}	0.22826	0.40871	0.06871
X_4	0.66304	0.54057	0.55580	X_{16}	0.70634	0.79119	0.61254
X_5	1.09312	0.90336	0.55811	X_{17}	0.50277	0.85191	0.33666
X_6	0.64298	0.13158	0.66503	X_{18}	0.92996	0.24041	0.70250
X_7	0.72710	0.21742	0.05519	X_{19}	0.39872	0.54907	0.22109
X_9	0.36083	0.95220	0.35074	X_{20}	0.07181	0.06071	0.70001
X_{10}	0.56620	0.28325	0.47405	X_{21}	0.56281	0.21307	0.77126

6. Conclusions and Implications

In order to solve the contradiction between low-carbon development and economic development of China's construction industry and realize the low-carbon economic development of the construction industry, this paper combines the data-driven, Stacking integrated model and Improved Genetic Algorithm (IGA) with each other, and puts forward the data-driven Stacking-IGA model, which researches the strategy of the construction industry's low-carbon economic development, and provides China's low-carbon economic transformation of the construction industry with a theoretical basis, and the research conclusions are as follows:

(1) The data-driven Stacking-IGA model is applicable to the modeling and solving of uncertain objective planning, and the combination of data-driven and Stacking integrated model makes the model reflect the actual problem more realistically, and the use of improved genetic algorithms to solve the unconstrained planning problem has the advantages of high solving efficiency and high solving quality.

(2) In order to address the problem that the data sample size and the ratio of degrees of freedom of the sample data are low, this paper adopts the Smote algorithm combined with the feature de-redundancy to preprocess the sample data, after the preprocessing, the data sample size and the ratio of degrees of freedom of the sample data are improved from 7.1 to 14, with the improvement rate of close to 100%, and the total number of the sample set is expanded from 150 to 225, with the expansion rate of 50%, and the sample proportion of each category of assembly building development level is improved from 4:5:7:9:5 to 1:1:1:1:1:5. 50%, the sample proportion of each category of assembly building development level is improved from 4:5:7:9:5 to 1:1:1:1:1:1. The method shows good results in sample expansion, sample

ISSN: 2710-0170

equalization and feature de-redundancy, and it also improves the performance of the Stacking integration model, and enhances the ability of the low-carbon development target planning of the construction industry based on the Stacking integration model to reflect the actual problems. It improves the ability of the low-carbon development target planning for the construction industry constructed based on Stacking integration model to reflect the actual problems.

(3) The low-carbon development target planning of the construction industry constructed by the Stacking integrated model reflects the actual problems better.The Stacking integrated model has achieved a certain degree of performance improvement compared with the single model, indicating that the Stacking integrated model can utilize the advantages of multiple models to combine the prediction, thus improving the overall accuracy, and its reflection of the data characteristics is more comprehensive and accurate. The integrated model has a certain degree of stability. The integrated model has a certain degree of stability, which indicates that the prediction results of the model have less fluctuation in different data sets, and the experimental results are less accidental.

(4) The improved genetic algorithm constructed by the elite strategy and tournament selection method shows remarkable performance in both goal solving efficiency and solution quality when solving the target planning problem of low-carbon economic development in the construction industry. It is difficult to realize the reduction of carbon emission and the increase of total output value in China's construction industry at the same time. With the increase of total output value of the construction industry, the weight of carbon emission of the construction industry also increases, so the coordination of the economic development of the construction industry and the low-carbon development is a problem that China must face. The Pareto frontier solved by the improved genetic algorithm provides a theoretical basis for the solution of this problem.

(5) Different characteristics have different impacts on the issue of low-carbon economic development in the construction industry. Factors such as population size, region, cement production, steel production, etc. of each province have a greater impact on both the total output value of the construction industry and the amount of carbon emissions from the construction industry. When the decision makers do not satisfy the optimal Pareto solution, it is more reasonable to start from the cement production, steel production, timber production, the effectiveness of the assembly building policy, the number of assembly building production bases and other factors that have a greater impact on the target planning, and the realization of the low-carbon economic development of the construction industry should be started from the production of building materials, and at the same time, to promote the new construction technology.

References:

[1] Bao Jianqiang, Miao Yang and Chen Feng, "Low-carbon economy: a new change in the way of human economic development", China Industrial Economy, No. 4, 2008.

[2] UK Government. Energy White Paper, Our Energy Future: Creating a Low Carbon Economy [R]. 2003.

[3] Lin Boqiang, 2011: China's Low Carbon Transition, Science Press.

[4] Zhang Kunmin, Pan Jiahua and Cui Dapeng, 2008: Theory of Low-Carbon Economy, China Environmental Science Press.

[5] He Jiankun, 2009: "Development of low carbon economy, the key lies in low carbon technology innovation", Green Leaf, No. 1.

[6] YU Zhuangxiong, CHEN Jie, DONG Jiemiao. The road to a low-carbon economy: a perspective on industrial planning[J]. Economic Research,2020,55(05):116-132.

ISSN: 2710-0170

[7] Linhui Wang, Hui Wang, and Zhiqing Dong, "Policy conditions for the compatibility of economic growth and environmental quality - A test of policy bias effect under the perspective of the direction of environmental technology progress", Management World, No. 3, 2020.

[8] Du Q, Shao L, Zhou J, et al. Dynamics and scenarios of carbon emissions in China's construction industry[J]. Sustainable Cities and Society, 2019, 48: 101556.

[9] Xu P, Wang Y, Yao H, et al. An exploratory analysis of low-carbon transitions in China's construction industry based on multi-level perspective[J]. Sustainable Cities and Society, 2023, 92: 104460.

[10] Yao H, Xu P, Wang Y, et al. Exploring the low-carbon transition pathway of China's construction industry under carbon-neutral target: A socio-technical system transition theory perspective[J]. technical system transition theory perspective[J]. Journal of Environmental Management, 2023, 327: 116879.

[11] FAN Jianshuang,ZHOU Lin. Spatial and temporal characteristics of carbon emissions from China's construction industry and provincial contributions[J]. Resource Science,2019,41(05):897-907.

[12] ZHAO Hongyan, HUO Zhenggang, ZHA Xiaoting et al. Analysis of factors affecting building carbon emissions and scenario prediction[J]. Journal of Yangzhou University (Natural Science Edition),2022,25(06):65-70.

[13] YUAN Runsong,FENG Chao. Decomposition of carbon emission drivers in China's construction industry based on common frontier production theory: 2000-2020[J]. China Population-Resources and Environment,2023,33(01):161-170.

[14] [14] Du Q, Zhou J, Pan T, et al. Relationship of carbon emissions and economic growth in China's construction industry[J]. Journal of cleaner production, 2019, 220: 99-109

[15] [15]Long L,Yinting L. The Spatial Relationship between CO2 Emissions and Economic Growth in the Construction Industry: Based on the Tapio Decoupling Model and STIRPAT Model[J]. Sustainability,2022,15(1).

[16] WANG Zhendong,JIN Tianlin. Mechanisms and paths of environmental regulation affecting the development of the construction industry[J]. People's Forum-Academic Frontier,2020, No.202 (18):140-143.

[17] WANG Junping,ZHAO Chenjing,NIU Bo. Research on the impact of environmental regulation on the economic growth of construction industry[J]. Construction Economy,2021,42(09):19-22.

[18] AN Bowen, HOU Zhenmei, WANG Qingjie. Analysis of spatial divergence and convergence mechanism of productivity in China's construction industry[J]. Statistics and Decision Making,2023, 39(08):114-119.

[19] P. Zhang, Y. Xiong, J. Jian. Research on false data injection attack for smart grid based on multi-objective bi-layer planning[J]. Operations Research and Management,2023,32(01):22-26.

[20] Wang S.I., Wu J., Luo H. et al. Optimization and application of biomass power generation subsidy policy based on multi-objective nonlinear programming[J]. Arid Zone Resources and Environment, 2020,34(12):65-71.

[21] Liang Yuan, Xu Enyong, Meng Yanmei. A data-driven optimization method based on car body uncertainty[J]. Journal of Chongqing University of Technology:Natural Science,2021,35(5):85-92.

[22] Yan Jun, Zhang Hengrui, Li Wenbo et al. Data-driven multi-objective optimization of flexible riser skeleton layer[J]. Journal of Harbin Engineering University,2023,44(05):689-697.

[23] WANG Ling, EMPLOYEE Hua, NA Wenbo et al. Optimization of wheelset turning repair strategy and remaining life prediction based on wear data-driven model[J]. Systems Engineering Theory and Practice,2011,31(06):1143-1152.

[24] H. Zhang, K. Sun, Y. R. Shi, et al. A review of power system scheduling methods considering flexible resources and modulo-driven methods[J/OL]. High Voltage Technology:1-13

[25] LIU Xiao,ZHOU Rongxi,LI Yuru. China's credit bond default prediction based on the integration of Stacking algorithm[J]. Operations Research and Management,2023,32(03):163-170.

ISSN: 2710-0170

[26] Ran Yaxin,Han Hongqi,Zhang Yunliang et al. A large-scale text hierarchical classification method based on Stacking integrated learning[J]. Intelligence Theory and Practice,2020,43(10):171-176+182.

[27] WANG X ,FENG Z ,CAI J et al. All-van der Waals stacking ferroelectric field-effect transistor based on In2Se3 for high-density memory[J/OL]. Science China(Information Sciences):1-8.

[28] He Yuejie. Research on the prediction and development of gross output value of construction industry based on machine learning [D]. Changjiang University,2023.

[29] JI Yingbo, DUAN Zhaohui, ZHAO Likun et al. Analysis of comprehensive differences in regional development level of assembled buildings in China - Based on PP-DEA modeling

[30] Swami R ,Dave M ,Ranga V .IQR-based approach for DDoS detection and mitigation in SDN[J].Defence Technology,2023,25(07):76-87.

[31] ZHOU Z, CHEN H, LI G, et al. Data-driven fault diagnosis for residential variable refrigerant flow system on imbalanced data environments [J]. International Journal of Refrigeration, 2021, 125: 34-43.

[32] Tan WK et al. Rockburst intensity classification prediction based on four integrated learning[J]. Journal of Rock Mechanics and Engineering,2022,41(S2):3250-3259.

[33] Tian Y ,Feng Y ,Zhang X , et al. A Fast Clustering Based Evolutionary Algorithm for Super-Large-Scale Sparse Multi-Objective Optimization[J].IEEE /CAA Journal of Automatica Sinica,2023,10(04):1048-1063.

[34] Wang Z ,Li Y ,Shuai K , et al. Multi-objective Trajectory Planning Method based on the Improved Elitist Non-dominated Sorting Genetic Algorithm[J]. Chinese Journal of Mechanical Engineering,2022,35(01):81-95.

[35] RenFang H ,XianWu L ,Bin J , et al.Multi-objective optimization of a mixed-flow pump impeller using modified NSGA-II algorithm[J].Science China(Technological Sciences),2015,58(12):2122-2130.