Modern Management based on Big Data V
A.J. Tallón-Ballesteros (Ed.)

doi:10.3233/FAIA240263

Machine Learning-Based Sales Prediction Using Bayesian Optimized XGBoost Algorithms

Xue-xia Liang [a], Ying Lin [b], Chao Deng [a,1], Yu-hua Mo [a], Bin Lu [a], Jiong Yang [c], Liang-ben Xu [a], Pai-yu Li [a], Xi-peng Liu [a], Kai-di Chen [a] and Meng Liu [a]
[a] *China Tobacco Guangxi Industrial Co., Ltd., Nanning Guangxi 530001, China*
[b] *China National Tobacco Corporation Guangdong Company, Guangzhou 510610, China*
[c] *China Tobacco Fujian Industrial Co., LTD., Xiamen 361012, China*
ORCiD ID: Xue-xia Liang https://orcid.org/0009-0002-9184-6062
Chao Deng https://orcid.org/0000-0001-9484-8923

Abstract. Market forecasting is crucial for the success of the commodity sales, and with the development of machine learning, a lot of companies are increasingly relying on data-driven models for market analysis and forecasting. In order to explore the application of machine learning for the commodity sales prediction, this paper constructs a product market prediction framework based on extreme gradient enhancement. Taking data-driven artificial intelligence machine learning as a method, it improves the accuracy and speed of predicting product sales in the market industry using mathematical modeling and manual analysis generally. Traditional manual analysis requires a long time and high cost to collect data, build models, and solve models among other processes. This article uses a Bayesian optimized limit gradient boosting intelligent prediction model to train and predict the ordering data of Zhenlong (Zhenpin) products in Guilin City in 2022 and predicts the sales of time series data and feature data respectively. The results showed that making full use of machine learning models for prediction has higher accuracy and faster speed comparing to traditional prediction methods, greatly optimizing the prediction process. This method provides a new research approach for predicting sales in the future commodity industry.

Keywords. Market forecast, XGBoost, Sales forecast, Big data, Evaluation

1. Introduction

The consumer commodity industry is currently facing unprecedented challenges and opportunities due to the rapid development of the global economy and the diversification of consumer demands. In order to formulate effective marketing strategies, it is vital to accurately grasp market demand and predict consumer behavior in a highly competitive market environment. Therefore, N. Caglayan and J. Huber [1-2] believe that demand forecasting (DF) has become increasingly important in the industry. For the sake of their economic benefits, companies can make decisions based on demand, price, competition,

[1] Corresponding Author: Chao Deng, China Tobacco Guangxi Industrial Co., Ltd., Nanning Guangxi 530001, China; Email: dengch021@hotmail.com

and consumer behavior, and establish rational production targets such as output levels, which are crucial for achieving optimal procurement and distribution strategies so that they can avoid overstocking [3-4]. Over the past few decades, demand forecasting methods and techniques have developed rapidly. From simple qualitative forecasting methods to time series analysis methods, causal analysis methods, and machine learning prediction methods [5], the accuracy and practicality of demand forecasting have continuously improved. Commonly used machine learning methods include linear regression, XGBoost, decision tree models, K-nearest neighbor models, MLP regression models, and support vector machine models.

In recent years, a lot of scholars have delved into in-depth exploration and research on the aforementioned common machine learning methods. Ji et al built upon the XGBoost model, developed a C-A-XGBoost forecasting model that considers both the sales features of commodities and the trend of data series, which indicates that C-A-XGBoost can enhance the accuracy of sales forecasting for e-commerce companies [6]. Jiang et al. proposed an adaptive univariate support vector machine (AUSVM) model to address the uncertainty and intermittency in predicting heavy vehicle spare parts demand. Through empirical validation, AUSVM demonstrated superior computational efficiency compared to basic SVM and neural networks. Furthermore, it achieved significant enhancements in forecasting accuracy and inventory management effectiveness when dealing with non-smooth demand series [7]. Käck and Freitag conducted a comprehensive empirical study to assess the performance of demand forecast using the local k-nearest neighbor model, which is grounded in dynamical systems theory, which indicates that this model yields higher prediction accuracy within a shorter time frame compared to the established benchmark methods [8]. Massaoudi et al. proposed a novel ensemble-based hybrid model that combines Light Gradient Boosting Machine (LGBM), extreme Gradient Boosting machine (XGB), and Multi-Layer Perceptron (MLP) through stacked generalization. Through verification on data sets at different locations, it is proved that its performance is superior to the existing benchmark techniques and hybrid models, which provides an effective method to solve the problem of short-term load forecasting [9]. Mohanty et al. introduced an efficient crop price forecasting framework, leveraging machine learning methodologies. Their methods incorporated time series method, statistical approaches, and machine learning techniques to anticipate crop prices. The findings of their study indicated that decision tree regressor emerged as the optimal forecasting model, providing farmers with the capability to proactively assess their potential profitability and financial risks [10]. In Özmen's research, he investigated generating natural gas demand forecasts of residential users for the distribution system operators and utilized simple multiple-linear regression (LR), multiple adaptive regression splines (MARS), and least absolute shrinkage and selection operator (LASSO) models to facilitate a comparative analysis. The research findings indicate that while the efficacy of these methods diminishes as the forecasting period extends, MARS maintains optimal performance [11].

The extant literature underscores the multifaceted applications of demand forecasting across diverse domains. Each field has its own optimal model tailored to specific scenarios. When forecasting commodity demand, it is crucial to take into account the unique characteristics of each commodity in order to select the most appropriate forecasting method [12]. In the intricate and unpredictable consumer goods market, demand forecasting still faces numerous challenges. It is crucial to focus on the impact of distinctive variables and take into account the influence of other factors on demand forecasting. Meanwhile, in the exploration of time series and machine learning

models, optimizing the design of model parameters should be prioritized. Buckell et al. incorporated the displayed preference data into the model to calibrate the market forecast [13]. Furthermore, numerous researchers have employed Bayesian optimization techniques to optimize hyperparameters [14-16].

If product sales are influenced by seasonal, trend, or cyclical factors, linear regression models cannot directly capture these time effects. Linear regression models assume that independent variables are independent of each other. If there are interaction effects between independent variables, the model will not accurately reflect this relationship. Besides requiring a sufficient sample size to ensure predictive accuracy. In the case of small samples, the model's predictive power may decrease. When building the model, it is necessary to manually select independent variables that significantly affect product sales. Incorrect variable selection can lead to inaccurate model predictions. In order to overcome these limitations, researchers may employ more complex models, such as time series analysis, nonlinear regression, or machine learning algorithms, to improve the accuracy and robustness of predictions.

The researchers have verified the accuracy of algorithm models based on artificial intelligence. They established a prediction model for the SC-CO2 fracturing effect in coal rock masses using the SVM algorithm [17], and optimized the hyperparameters of SVM with the dragonfly algorithm and differential evolution algorithm. The study points out that the predictive accuracy of neural networks depends on their architectural design, and overly deep networks can increase training time while suffering from poor interpretability. SVM performs well with a small sample size, but its predictive accuracy decreases as the amount of data increases. In contrast, XGBoost [18] is an extensible extreme gradient boosting model framework that offers higher computational efficiency, is suitable for incremental learning, maintains high accuracy, and has better interpretability. In the XGBoost algorithm, grid search or population-based optimization algorithms are commonly used to select the best combination of hyperparameters, such as genetic algorithms or particle swarm optimization. However, grid search requires enumerating all possible parameter combinations, which can lead to large computational costs and potential curse of dimensionality issues [19]. Population-based optimization algorithms require a large number of initial sample points and have lower optimization efficiency. In comparison, Bayesian optimization algorithms, based on Bayesian theorem and Gaussian process regression [20], can use prior knowledge and historical information to guide the search direction, quickly converging to the optimal solution with higher efficiency and accuracy [21]. Dong Na and colleagues [22] applied the Artificial Bee Colony algorithm to refine the parameter estimation for Support Vector Machines (SVMs), thereby establishing an ABC-SVM based model for construction cost estimation. This model has shown a high level of predictive precision and broad usability. Nevertheless, SVMs are greatly affected by their internal parameters, leading to slow computation when dealing with extensive training datasets, thus limiting its practicality for large-scale data processing [23].

In scenarios with small data size, numerous parameters, and lack of prior knowledge, the Bayesian linear regression prediction model exhibits higher predictive accuracy and applicability compared to the BP neural network [24].

This study utilized the ordering data of "Zhenlong (Zhenpin)" products in Guilin City for the year 2022 as the validation dataset for subsequent experiments. To avoid accidental errors caused by manual data extraction and enhance the model's training and performance, a systematic random sampling method was employed for data selection. We established three model evaluation parameters and introduced six machine learning

model algorithms. Based on the sampled dataset, we conducted agent fitting training on the numerical simulation process and the fitting accuracy and speed is as evaluation index. After comparison, the XGBoost model was chosen as the subsequent model for predicting the consumer market. To further improve prediction accuracy, we applied the Bayesian optimization algorithm (BOA) to adaptively optimize the hyperparameters of the crack simulation agent model based on the extreme gradient boosting model. Using the posterior probability model, a hyperparameter combination with the greatest improvement potential was selected through a sampling strategy. Through multiple iterations, a set of hyperparameters close to the optimal was found. For fracturing datasets that exceeded the fitting range, we constructed a highly expressive crack simulation agent model that achieves a balance between fitting ability and generalization ability, thereby enhancing the generalization prediction accuracy and efficiency of the machine learning model simulation. The innovation points of this article are that by using the system random sampling method to select the validation dataset, unexpected errors that may be caused by manual data extraction are avoided, and the training and performance of the model are improved. Meanwhile, adopting Bayesian optimization algorithm to adaptively optimize hyperparameters further improves the predictive accuracy of the model. Through the research on the above innovative points, this article has made significant progress in data selection and optimization, model selection and comparison, and model construction, providing strong support for predicting and making decisions in the consumer market.

The advantage of combining BOA and XGBoost lies in the fact that BOA, as an efficient hyperparameter optimization algorithm, can automatically adjust the hyperparameters of the XGBoost model to achieve optimal model performance. In comparison to Support Vector Machine (SVM), linear regression, and Bayesian optimization of linear models, Bayesian optimization of machine learning models has the following innovative points: The Bayesian optimization algorithm can better handle nonlinear problems and complex models, while SVM and linear regression are mainly suitable for linear problems. BOA considers the uncertainty of the model during the optimization process, allowing for a more comprehensive exploration of the hyperparameter space to find better parameter combinations. The Bayesian optimization algorithm does not require tedious grid search or random search; instead, it guides the search direction through a probabilistic model, thereby improving optimization efficiency. BOA can automatically adjust the model complexity to avoid overfitting and underfitting issues, thus enhancing the model's generalization ability. In summary, the advantage of combining BOA with XGBoost is the ability to automatically adjust the model's nonlinear hyperparameters, improve model performance and generalization ability, and provide an efficient method for predicting and decision-making in complex problems. We have applied this method for the first time in the tobacco industry to achieve sales forecasting.

2. Method

With the rapid development of artificial intelligence (AI), machine learning (ML) can provide a promising and effective way to deal with challenges in commodity prediction. The benchmark results of this study are those of a specific machine learning model from the XGBoost, which is widely regarded as a cutting-edge machine learning model. It is possible that the model’s performance can be improved and it may have underperformed

in the previous study due to suboptimal hyperparameter tuning procedures [25-27]. This paper extracts sampling data features for different data sources in different situations, and selects a portion of data samples from the overall dataset for the process of model training research. Based on the construction and training of machine learning agent models, the quality and quantity of dataset samples have a significant impact on the training and performance of the model. Therefore, in order to avoid accidental errors caused by human data extraction, a systematic random sampling method was used to extract and screen the ordering data of Zhenlong (Zhenpin) products in Guilin City in 2022, and the corresponding dataset was generated through data preprocessing.

This article identified three model evaluation parameters and introduced six machine learning model algorithms. Based on sample datasets, model was conducted as the numerical simulation process was trained through surrogate fitting. The fitting accuracy and speed were compared, and the limit gradient enhancement model was selected as the subsequent model for predicting the product market. Continuously modifying hyperparameters and iterating through model optimization, the prediction accuracy was improved. The fitting accuracy and speed were compared, and the limit gradient enhancement model was selected as the subsequent model for predicting the tobacco market. By modifying hyperparameters and iterating through model optimization, the prediction accuracy was improved continuously.

2.1. Data Source

Tobacco companies can collect data from multiple sources including sales records and consumer surveys, which need to be integrated and cleaned up in order to build a dataset that can be used for model training. The model will be modeled for each type of product in each city, which is to predict next week's order volume based on historical retail data as shown in table 1. It is not open to all readers. Private datasets may contain specific domain information that is crucial for model training and performance tuning. Tobacco data may include specific attributes related to tobacco production, sales, etc., which may not be available in other public datasets. Using real, domain-specific datasets relevant to actual application scenarios can better simulate real-world conditions, thus enabling models to have better performance and generalization when deployed in real-world applications. Public datasets may not be as relevant to our specific problem, and the data quality may not be as good as your proprietary tobacco data. Using our own dataset can better reflect real-world scenarios, improve model accuracy, and enhance generalization capabilities. Utilizing exclusive internal data can provide you with a competitive advantage, as you can leverage this data to train more targeted and effective models, thereby achieving better performance in the market.

Table 1. Example of data

Number	Week	Month	Retailer	Product	Sale (million box)
1	202237	202209	33020111949	Mudan(soft)	0.06
2	202230	202207	330204106168	Zhonghua (Jinzhong) Branch)	0.02
3	202231	202207	330203116684	Furongwang(hard)	0.02

The predicted order quantity can be achieved through normalization. The feature inputs of the model are mainly divided into three parts: derived features obtained from

the historical order volume of the retailer, the characteristics of the retailer and other features. The derived features obtained from the historical order volume of retail households are mainly based on the corresponding calculations of the order volume of retail households in the past 4 weeks, 13 weeks, and 26 weeks. We stipulate that the order quantities for 4 weeks, 13 weeks, and 26 weeks will be used as monthly, quarterly, and semi-annual features for subsequent machine learning respectively. Different statistical parameters were selected for calculation based on monthly, quarterly, and semi -annual time dimensions so that the different characteristics of order quantity can be obtained. For calculating the basic statistical parameters of order quantity, data such as maximum, minimum, mean, standard deviation, variance, skewness, mode, and total order quantity from derived features is selected [26]. We extracted the derived features of weeks with order quantity of 0 and weeks with order quantity of non to study the impact of week as the time dimension on the change of order quantity [28]. For the sake of studying the impact of holidays on order quantity, the order quantity within all holidays in the derived features was extracted. In order to study the impact of cigarettes at different price points on order quantity, we used 100, 500, and 1000 as price differentiation points to calculate the order quantity of specifications for cigarettes with a price point greater than 1000, the order quantity of specifications for cigarettes with a price point between 500 and 1000, the order quantity of specifications for cigarettes with a price point between 100 and 500, and the order quantity of specifications for cigarettes with a price point below 100, in order to study the trend of change of order quantity, the change values of order quantity for products of different specifications were extracted from this week to the first two weeks. The 53 weeks' sales curves of ten retailers were randomly extracted and plotted as shown in Figure 1, and it can be found that there seems to be a certain correlation in their changing trends. By assembling the ordering data of Zhenlong (Zhenpin) in 2022, a data body was formed, consisting of 420255 samples and 22 time series. 80% of the data was used as the training set, and the remaining 20% was used as the validation set for validation. Finally, 333385 training data and 83347 validation data sets were formed.

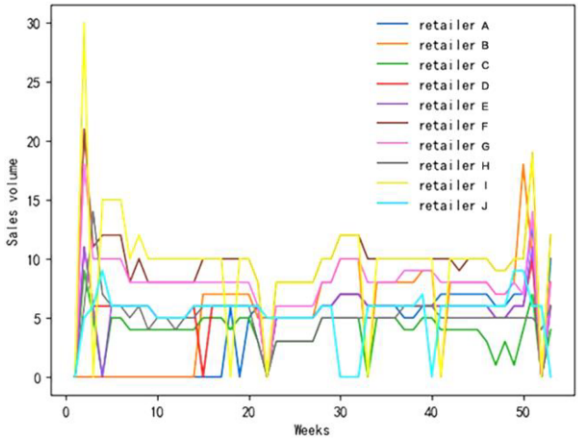


Figure 1. Schematic diagram of partial retail data

Therefore, the average and median sales curves of 24546 retail households, and plot the relative errors between all curves and these two sets of curves to obtain Figure 2, with a control confidence interval of 0.9. From the Figure, it can be found that the error

between the overall sales curve and the average and median curves is generally controlled within 3, which has strong temporal characteristics, Therefore, it can be trained and predicted through time series machine learning models [27, 29].

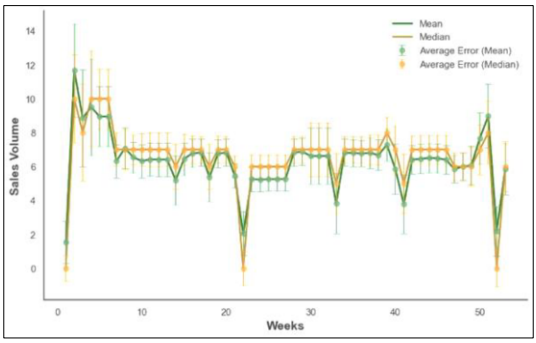


Figure 2. Comparison of Retail Household Data with Median and Average Values

Features include transportation facilities service, Accommodation, Sports and leisure, Public facilities, company, Medical care, dwelling, address, Indoor facilities, Motorcycle maintenance, Automobile sales and so on. After that, feature engineering involves transforming, combining, and processing raw data, such as handling missing values, standardizing, normalizing, and one-hot encoding, to generate more useful features. Then, appropriate machine learning models like random forests are selected. The processed feature data is input into the model for training, and the trained model is used for sales forecasting. After model training, feature importance analysis helps understand the impact of each feature on sales prediction. Finally, by calculating performance metrics like mean squared error, R-squared value, etc., the model's predictive accuracy is evaluated to ensure effective sales forecasting.

2.2. Evaluation coefficient

This article selected different evaluation coefficients to assess the accuracy of machine learning models in predicting the tobacco industry market. Evaluation coefficient is an important indicator in data science and machine learning, often used to evaluate the performance of models and the prediction accuracy of models. This article selects the judgment coefficient as the evaluation coefficient [30].

The coefficient of determination is a statistical measure used to evaluate the results of regression analysis, defined as the proportion of the dependent variable that can be explained by the independent variable to the total variance. In other words, the determination coefficient reflects the degree to which the regression equation can explain the variability of the dependent variable. In simple linear regression, the determination coefficient is equal to the square of the sample correlation coefficient. In multiple linear regression, the determination coefficient is equal to the square of the multiple correlation coefficient. The value range of the judgment coefficient is between 0 and 1. The closer the value is to 1, the greater the proportion of variance explained by the regression equation, and the better the regression effect; The closer the value is to 0, the smaller the proportion of variance explained by the regression equation, and the worse the regression effect.

Assuming a dataset includes$y_1, y_2, \ldots, y_n$, the corresponding predicted values of the

model are $f_1, f_2, \ldots, f_n$, defining residuals $e_i = y_i - f_i$. In statistical analysis, we use the concept of sums of squares to evaluate the fit of a model and determine its effectiveness. We calculate three types of sums of squares: total sum of squares SS_{tot}, regression sum of squares SS_{reg}, and residual sum of squares SS_{res}.The total sum of squares $SS_{tot}$is calculated as the sum of the squared differences between each observed value y_i and the mean of the observed values $\bar{y}$.

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i \quad (1)$$

$$SS_{tot} = \sum_i (y_i - \bar{y})^2 \quad (2)$$

$$SS_{reg} = \sum_i (f_i - \bar{y})^2 \quad (3)$$

$$SS_{res} = \sum_i (y_i - f_i)^2 = \sum_i {e_i}^2 \quad (4)$$

$$R^2 \equiv 1 - \frac{SS_{res}}{SS_{tot}} \quad (5)$$

The closer the statistic is to 1, which indicates a high degree of fit, indicating that the model is a good representation of the underlying data. Statistical software typically calculates these values automatically, allowing researchers to assess the effectiveness of their models and make informed decisions about their choice of predictive models. The higher the goodness of fit of the model.

2.3. Model building

Six machine learning model algorithms including Linear Regression, Extreme Gradient Boosting, Decision Tree Model, K-Nearest Neighbor Model, MLP Regression Model and Support Vector Machine Model were introduced for modeling based on sampled datasets, and pre fitting training was conducted for the numerical simulation process.

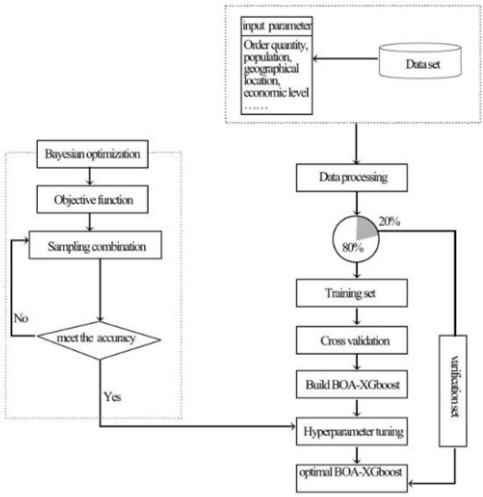
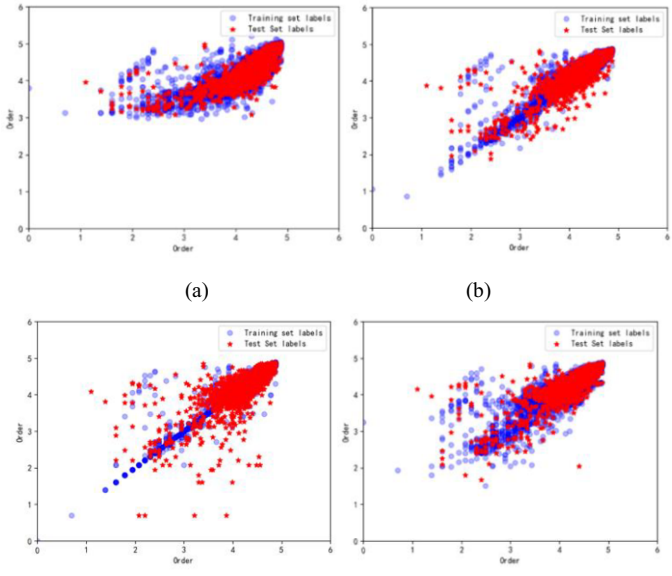


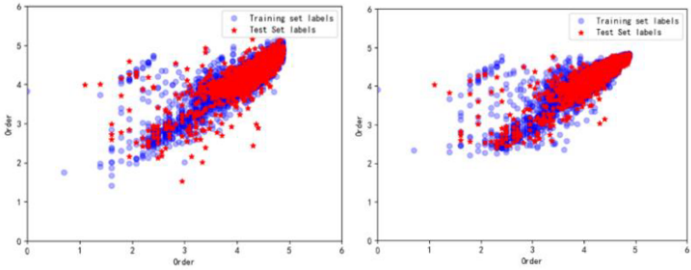
Figure 3. The training process of the model

The process of solving the prediction model is shown in Figure 3. The first step is to split the dataset into training and testing sets, 20% is the validation set, and 80% is the training set. usually using cross-validation to evaluate the model performance. Next, we utilize Bayesian optimization to optimize parameters for a given objective function while

ensuring a specified accuracy threshold. we train the model on the training set and evaluate its performance on the validation set. The objective function, often representing a metric like accuracy or another performance measure, guides the optimization process. Bayesian optimization iteratively suggests new parameter configurations based on previous evaluations, aiming to find the optimal set of parameters that maximizes the objective function within the defined accuracy constraints. This iterative process continues until a satisfactory model with the desired level of accuracy is obtained the XGBoost model hyperparameters are tuned using methods such as grid search or random search to find the best parameter combination. Finally, the XGBoost model is trained using the training set, continuously optimizing the loss function, resulting in a powerful ensemble model.

Judgment coefficient R^2, mean absolute error MAE, and mean square error MSE are as evaluation parameters. The prediction accuracy of the models is ranked from high to low, namely the limit gradient improvement model, support vector machine model, K-nearest neighbor model, MLP regression model, decision tree model, and Linear Regression model. In order to select the model with the highest prediction accuracy for the next step of learning and prediction, this article conducted pre validation on the six different models selected. Pre validation is a comparison of multiple models before selecting the model. By comparing the prediction accuracy and generalization ability for the same dataset, the optimal model is selected and the risk of overfitting is avoided. This article will select the model with the highest accuracy and the best generalization ability for further learning and prediction. In order to ensure the reduction of errors between data learning and prediction, this paper used the same hardware equipment to process the dataset accordingly, and used the train of the Scikit learn library in Python_ Test_ When using the split function, the same random number is selected to ensure that different models use the same training and testing sets for learning and prediction. The comparison of prediction accuracy of different models can be obtained through calculation as shown in Figure 4.





(e) (f)

Figure 4. Different model prediction accuracies (a: Linear Regression regressions model; b: Extreme Gradient Boosting model; c: Decision Tree model; d: K-Nearest neighbor model; e: MLP regression model; f: Support Vector Machine model)

Figure 4 respectively show the predictive performance of models established by Linear-Regression, Extreme Gradient Boosting, Decision Tree Model, K-Nearest Neighbor Model, MLP Regression Model, and Support Vector Machine Model on data. The results in Figure 4 show that after parameter tuning and optimization to the best result, the predictive accuracy of the models is ranked from high to low as follows: Extreme Gradient Boosting Model, Support Vector Machine Model, K-Nearest Neighbor Model. The MLP regression model, decision tree model, and Linear Regression model have the highest prediction accuracy compared to other models, with a maximum gradient improvement coefficient of 0.741. The choice of an R^2 (coefficient of determination) threshold is often based on a balance between model fit and practical significance. In many scientific and statistical analyses, a commonly accepted threshold for R^2 is around 0.7, which indicates a strong linear relationship between variables. However, this can vary depending on the field: Social Sciences: A threshold of 0.6 can be considered acceptable, as per some studies [31], where a higher value might be reserved for more rigorous research. Economics: In some economic models, a higher R^2 (e.g., >0.8) is expected, as the goal is to explain a significant portion of the variance [32]. Psychology: While there is no strict standard, R^2 above 0.5 is often considered acceptable, but researchers might prefer higher values for more reliable generalizations [33]. Engineering and Physics: In predictive models, an R^2 of 0.9 or higher is often sought, as these fields demand high accuracy [34]. Ultimately, the choice should be informed by the research question, the complexity of the data, and the ability to generalize the findings. It's crucial to also consider the trade-off between model fit and model simplicity (e.g., avoiding overfitting).Therefore, the R^2 value of prediction accuracy should be greater than 0.7 to meet the simulation accuracy requirements. Taking into account various results, this article ultimately chooses the limit gradient enhancement model as the subsequent training and prediction model. We chose the extreme gradient enhancement model for subsequent training and prediction due to its highest prediction accuracy after parameter adjustment and optimization, as well as its strong gradient enhancement effect. The model accurately predicts trends in the tobacco market and provides reliable results for decision-makers. Overall, it meets accuracy requirements, exhibits high interpretability and stability, making it the most suitable choice for our needs.

XGBoost (extreme Gradient Boosting) extreme gradient boosting algorithm is based on gradient boosting algorithm to achieve parallel computing and sparse data processing, and achieve a balance between model prediction performance and computational speed.

Compared with gradient boosting ensemble algorithms, it can significantly improve computational speed.

The objective function of the XGBoost model is the traditional loss function and model complexity:

$$\text{Obj} = \sum_{i=1}^{m} l(y_i, \hat{y}_i) + \sum_{k=1}^{K} \Omega(f_k) \quad (6)$$

Where i represents the i-th sample in the dataset, m represents the number of samples imported into the k-th tree dataset, and K represents the total number of trees established.

Bayesian Optimization is a global optimization algorithm mainly used to optimize black box functions, that is to say, to find their minimum or maximum values without understanding the specific form of the function. Bayesian optimization utilizes the concept of Bayesian inference in the optimization process, and approximates the objective function by establishing probability models (such as Gaussian process regression) on experimental data, thereby achieving efficient optimization of the objective function. Bayesian optimization has a wide range of applications in machine learning, such as

$$p(z|D) = \frac{p(D|z)p(z)}{p(D)} \quad (7)$$

Where, z represents the unknown objective function, D represents the set of known observations D = {(x1, y1), (x2, y2), ..., (xn, yn)}, yn=z (xn)+ ε n. Among them ε It is an observation error. P (D | z) is the likelihood distribution of y, p (z) is the prior probability distribution of z, representing the assumption of the unknown function state, and p (z | D) represents the posterior probability distribution of z, describing the confidence level of the unknown objective function after correcting the prior probability based on the observation dataset.

Bayesian optimization is an efficient hyperparameter optimization method that can find the optimal combination of hyperparameters within a limited number of times. By using Bayesian optimization techniques, the limit gradient boosting model can be optimized for hyperparameters. During the model training and fitting stage, the hyperparameter results obtained through Bayesian optimization are shown in table 2.

Table 2. Hyperparameter results of XGBoost model based on Bayesian optimization during model training phase

Hyper-parameters	Range	Optimal parameter results
Number of decision trees	1-300	78
The maximum depth of the tree	1-25	3
The minimum number of split samples	2-10	2
Bootstrap	True/False	True
k-fold cross validation	/	40
The number of Bayesian iterations	/	5

For time series data, LSTM can process continuous time series data and capture temporal dependencies. It is necessary to convert time series data into discrete sequences, XGBoost can be used for modeling and prediction. XGBoost is a gradient boosting tree based algorithm commonly used for processing structured data. It can process data that has been transformed into discrete features, select features, and train models. It is suitable for extracting patterns from data and making predictions. LSTM is used to process raw time series data and capture temporal dependencies. XGBoost can be used to process discretized time series data for feature selection and model training. These two methods can be combined to comprehensively model and predict time series data. In some cases, the original time series data may contain a large amount of details and noise, and

discretization can convert continuous numerical data into discrete categories or intervals, reducing the complexity of the data. Discretization can transform time series data into discrete features, better adapting to some machine learning algorithms while preserving key feature information. The purpose of introducing BOA optimization into XGBoost is to further improve its performance. Traditional XGBoost uses grid search or random search to select the optimal combination of hyperparameters, but this method requires traversing the entire parameter space and requires a large amount of computation. BOA optimization can intelligently select parameter combinations, reduce unnecessary computational costs, and find better parameter combinations within a limited number of evaluations.

By using BOA optimization, XGBoost can quickly find the optimal combination of hyperparameters, improving the performance and generalization ability of the model. This optimization method can help XGBoost better adapt to different datasets and tasks, and improve the effectiveness of the model. The data samples calculated each time are random. In order to better reflect the characteristics of the entire dataset and reduce dependence on specific samples. The correlation coefficient used in the study is the average. However, it should be noted that each partition of samples is random, representative, and there is no overlap between samples to avoid introducing bias or duplicate calculations.

3. Result and analysis

We have constructed a reliable commodity market prediction model through data processing and feature extraction. This model is based on machine learning agent models, which have been trained and optimized to accurately predict trends and changes in the market, providing valuable information and reference for decision-makers. Features include mean, max, min, std and median. All artificial features from 4, 13, and 26 weeks were selected to assemble, predicting the ordering data for 27 weeks based on the assembled dataset that consists of a total of 420255 samples and 22 time series. Optimizing the hyperparameters of the limit gradient enhancement model and verify the optimization results using k-fold. The optimal hyperparameters are obtained in table 3. The prediction performance of models established by Linear Regression, Extreme Gradient Boosting, Decision Tree Model, K-Nearest Neighbor Model, MLP Regression Model, and Support Vector Machine Model on data was demonstrated. As shown in the Figure 4, after parameter tuning and optimization to the best result, the prediction accuracy of the four models was ranked from high to low as follows: Extreme Gradient Boosting Model, Support Vector Machine Model, K-Nearest Neighbor Model, MLP Regression Model The decision tree model and Linear Regression model, with the highest accuracy, have a judgment coefficient of 0.741 for the improvement of the limit gradient, which has extremely high prediction accuracy compared to other models. Therefore, the R^2 value of prediction accuracy should be greater than 0.7 to meet the simulation accuracy requirements. Taking into account various results, this article ultimately chooses the limit gradient enhancement model as the subsequent training and prediction model.

Table 3. Hyperparameter optimization results of XGBoost model

Hyper-parameters	Range	Optimal parameter results
Number of decision trees	1-300	100
The maximum depth of the tree	1-25	12
The minimum number of split samples	2-10	3
Bootstrap	True/False	True
k-fold cross validation	/	5
random_state	/	42

When modeling data with different regression models, we usually assess the predictive power of the models based on performance metrics such as root mean square error and correlation coefficients. A higher correlation coefficient which is used to evaluate the correlation between features and predicted values implies a stronger linear relationship between the model's predicted values and the actual observed values, indicating that the XGBoost model can more accurately capture the relationships within the data when making predictions. XGBoost models often provide information about the importance of features, i.e., which features have the most significant impact on the prediction outcomes. This can help us better understand the data and the model, facilitating feature selection or model optimization.

Without considering the scale of the dataset or the distribution of errors can lead to misleading conclusions about the model's performance. Additionally, R^2does not distinguish between systematic errors and random errors, nor does it penalize the complexity of the model. In cases where overfitting is a concern, a high R^2 value can be deceptive, indicating a good fit to the training data but not necessarily good generalization to unseen data. R^2scores can be significantly impacted by outliers. A few outlier predictions that are significantly different from the actual values can distort the R^2 score, making it sensitive to outliers or indicating unreliability in datasets with heavy-tailed distributions. Due to these limitations, using other metrics that consider different aspects of model performance in addition to R^2 can provide a more comprehensive perspective. Metrics such as Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE) can be valuable complements to R^2 in evaluating model performance. When discussing the size of errors, it is important to consider metrics like MSE, RMSE, MAE, and MAPE, as they provide insights into the magnitude of errors in the predictions. These metrics help in understanding how far off the model's predictions are from the actual values, giving a clearer picture of the model's accuracy. In terms of the model's robustness to outliers, it is crucial to assess metrics like MAE and MAPE, which are less sensitive to outliers compared to R^2 Models that are robust to outliers will have lower MAE and MAPE values, indicating that the model's predictions are not heavily influenced by extreme values in the dataset. Regarding prediction accuracy with respect to the dataset size, it is essential to consider metrics like MSE and RMSE, which provide a measure of the overall error in the predictions. As the dataset size increases, these metrics can help in evaluating how well the model generalizes to larger datasets and whether the model's performance remains consistent as the data volume grows. In conclusion, while R^2 can be a useful metric for assessing model performance, it is important to complement it with other metrics that provide a more nuanced understanding of the model's accuracy, robustness to outliers, and scalability to different dataset sizes. By considering a range of metrics, one can gain a more comprehensive insight into the strengths and limitations of the predictive model.

From Figure 5, it can be obtained that the R^2score is 0.86, Mse is 2.3, Rmse is 1.5, and Mae is 0.78 by calculation, which meets the requirement for model prediction accuracy exceeding 0.7 in the previous text. In addition, in the comparison scatter plot of prediction accuracy, the prediction accuracy is relatively high within the range of data size less than 20, which basically coincides with the standard regression line y=x, proving that the prediction accuracy is high in this interval, In the range where the model size is greater than 30, the degree of deviation from the y=x regression line increases. This article believes that the significant increase in buyer purchase volume at this time is mainly due to holidays or other accidental factors, which cannot clearly collect corresponding feature parameters, leading to an increase in prediction errors.

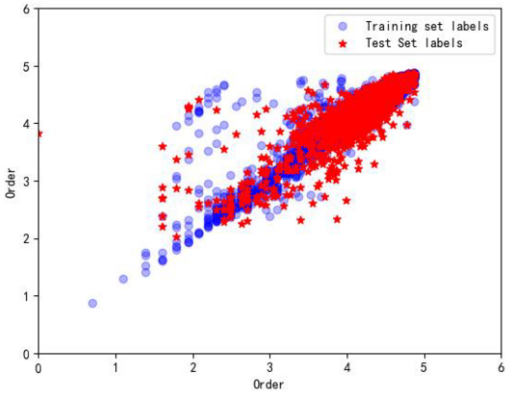


Figure 5. Comparison of prediction accuracy of extreme gradient boosting models

As depicted in the graph, the loss function evolves over the course of training, with the number of epochs. The epochs represent the number of times the training dataset is iterated over during the training process. As the model learns from the data and its parameters are updated, the loss function gradually decreases, indicating better performance. During the initial epochs, the loss function tends to be high, as the model struggles to learn the underlying pattern in the data. As the training progresses and the model becomes more sophisticated, the loss function decreases, indicating that the model is making more accurate predictions. However, it is essential to note that the loss function does not reach zero, as there is always some degree of discrepancy between the predicted and actual outputs. The rate at which the loss function decreases depends on various factors, such as the complexity of the problem, the quality of the training data, and the choice of the loss function and optimization algorithm. In some cases, the loss function may plateau or even increase slightly before decreasing again, which is known as overfitting or saturation. To address this issue, researchers often employ techniques such as early stopping, batch normalization, and dropout to prevent the model from overfitting. In conclusion, the graphical representation of the loss function over epochs provides valuable insights into the training process and the model's performance. By analyzing the trends and patterns in the loss function, researchers can determine the appropriate number of epochs and adjustments to the model's architecture, loss function, and optimization strategy to achieve the best possible performance as shown in Figure 6.

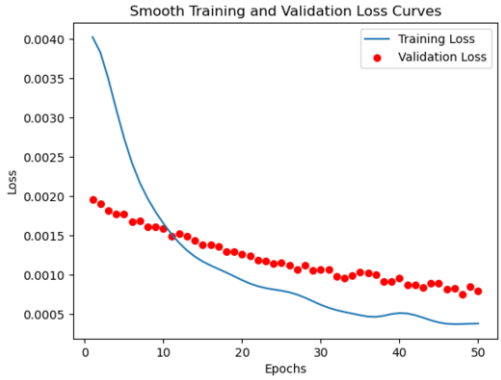


Figure 6. The loss function of the model

The judgment coefficient for the highest accuracy limit gradient improvement is 0.741, Compared to other models, it has extremely high prediction accuracy. The R^2 value of the project prediction accuracy in this article should be greater than 0.7 to meet the simulation accuracy requirements. Taking into account various results, this article ultimately chooses the limit gradient enhancement model as the subsequent training and prediction model.

In the context of machine learning and statistical modeling, the loss function is a crucial component that measures the discrepancy between the predicted output and the actual output. A well-designed loss function helps optimize the model's parameters to minimize this discrepancy, thus improving the model's performance. here are various types of loss functions, each with its unique characteristics and applications. Some common loss functions include the Mean Squared Error (MSE), Mean Absolute Error (MAE), and Cross-Entropy Loss. The MSE and MAE are widely used in regression and classification problems, respectively, as they penalize the squared or absolute differences between the predicted and actual values. In summary, the loss function plays a vital role in the optimization and performance improvement of machine learning models. By selecting an appropriate loss function and ensuring its balance ability, researchers can develop more robust and accurate models that generalize well to new data.

Using the order quantity of Zhenlong (Zhenpin) in Guilin City in 2022 as the time series dataset as an example in Figure 7, 494064 sets of samples were grouped based on retail ID, and assembled into a 494064 x 27 time series dataset. The next week's order quantity was predicted using 26 weeks as the time window. GRU and LSTM models were used for corresponding predictions, and the results are as shown in Figure 7. 80% of the data was used as the training set and 20% as the prediction set, resulting in 395251 training data and 98813 validation data.

	time_1	time_2	time_3	time_4	time_5	time_6	time_7	time_8	time_9	time_10	...	time_18	time_19	time_20	time_21	time_22	time_23	time_24	tin
0	18	10	10	10	8	8	8	8	8	8	...	8	6	6	6	6	6	6	
1	10	10	10	8	8	8	8	8	8	8	...	6	6	6	6	6	6	8	
2	10	10	8	8	8	8	8	8	8	8	...	6	6	6	6	6	8	8	
3	10	8	8	8	8	8	8	8	8	8	...	6	6	6	6	8	8	10	
4	8	8	8	8	8	8	8	8	8	8	...	6	6	6	8	8	10	10	
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
494780	7	7	7	6	6	6	6	6	7	8	...	7	7	8	8	8	8	8	
494781	7	7	6	6	6	6	6	7	8	8	...	7	8	8	8	8	8	7	
494782	7	6	6	6	6	6	7	8	8	8	...	8	8	8	8	8	7	10	
494783	7	6	6	6	6	6	6	6	7	7	...	7	7	7	7	6	7	10	
494784	6	6	6	6	6	6	6	7	7	7	...	7	7	7	6	7	10	10	

494064 rows × 27 columns

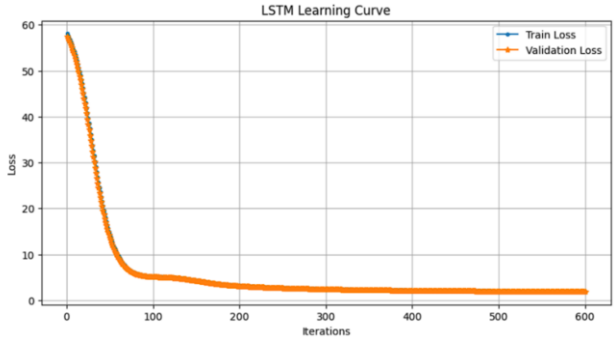
Figure 7. Schematic diagram of time series analysis data body

The parameters of LSTM and GRU neural network models were optimized and k-fold multi fold validation was used to obtain parameter combinations. LSTM and GRU were trained using the same parameters, and the optimization results are as follows in table 4.

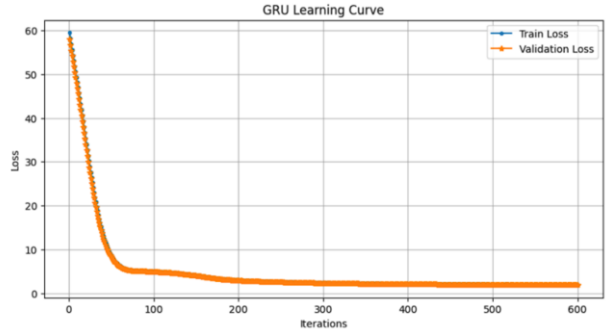
Table 4. Optimization Results of LSTM Model Parameters

Parameter	Optimal parameter results
input_size	26
hidden_size	64
num_layers	2
output_size	1
Epoch	600
k-fold cross validation	5

By comparing the learning curves of two different models, it can be found that the error of both models decreases to below 10% after about 60 iterations. After reaching the set 600 iterations, the error of the model has decreased to below 3%, proving that the model has high training accuracy at this time. By using the trained models separately to predict the original data results, the comparison of prediction accuracy can be obtained in Figure 8 and table 5.



(a)



(b)

Figure 8. Comparison chart of model prediction accuracy. (a)LSTM, (b)GRU hematic diagram of time series analysis data body

Table 5. Evaluation coefficients for LSTM/GRU model prediction results

Parameter	LSTM Model	GRU Model
MSE	1.92	1.91
RMSE	1.39	1.38
MAE	0.74	0.78
R^2	0.82	0.79

By comparing the evaluation coefficients between different models, it can be found that the prediction accuracy of LSTM and GRU models in this project is similar. By observing the accuracy comparison scatter plot, it can be found that except for some noisy data, LSTM and GRU neural network models have small errors in the prediction results of the entire range of data, R^2 scores is above 0.8, meeting the accuracy requirements of this article. Moreover, the scatter distribution plot effectively converges to both sides of the regression line y=x, proving that in the process of time series prediction, both LSTM and GRU models can effectively predict the results in Figure 9 and 10.

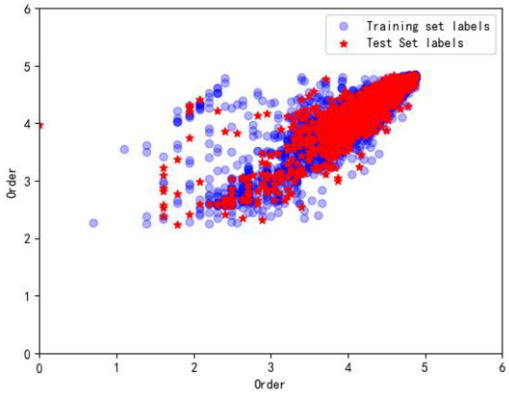


Figure 9. Comparison of prediction accuracy of LSTM model

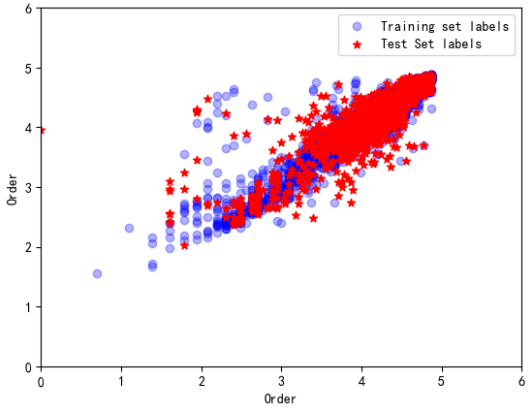


Figure 10. Comparison of prediction accuracy of GRU model

Use the trained model to validate the random raw data and predict the data for the next week every 20 weeks. The comparison of the prediction results is shown in Figure 11. From Figure 11, it can be seen that although there are some significant errors at some high or low sales nodes, the overall prediction trend is close to the same, which can generally meet the accuracy requirements of the prediction and meet the basic market demand.

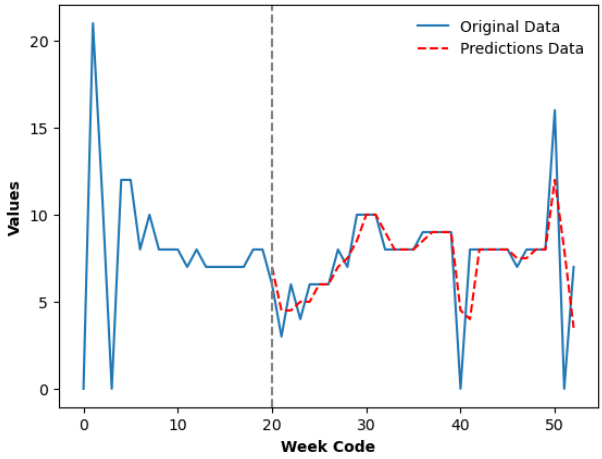


Figure 11. Comparison of prediction accuracy of GRU model

During the training process, the model may go through different stages and learn different features and patterns. Therefore, the performance of the model trained each time on evaluation metrics may vary. By training 600 times, a series of models can be obtained and their values on evaluation metrics can be calculated. The changes in these values can help us understand the stability, generalization ability, and performance of the model under different data distributions. When using the average value for model evaluation, it is necessary to comprehensively consider the properties of evaluation indicators, numerical changes, and the performance differences of the model under different training times. The numerical variation is very small, and our final conclusion is based on the average.

After the completion of model training, I performed a Spearman correlation analysis on the data to assess the relationship between the predicted price and various features. The analysis revealed the strength of correlation between the indicators and the predicted price, with a threshold of $|\rho s| > 0.7$ indicating a strong correlation. The results were ranked based on the degree of correlation, with positive values indicating a positive correlation and negative values indicating a negative correlation in Figure 12.

The graphs derived from this analysis provide valuable insights into the model's training and the importance of each feature in predicting the target variable. By identifying the indicators with a strong correlation, we can gain a better understanding of the factors that contribute to the model's performance.

For instance, shopping services and company Enterprise has a high positive correlation with the predicted price, it means that the model considers this factor as influential in determining the target variable. On the other hand, if a feature has a high negative correlation, it implies that the model believes that this factor has a detrimental impact on the target variable.

Moreover, these graphs can help us identify redundant or irrelevant features that may have a negative impact on the model's performance. By removing or adjusting these features, we can improve the model's efficiency and accuracy.

In conclusion, the Spearman correlation analysis graphs serve as a useful tool for interpreting the model's training results and understanding the importance of various features in predicting the target variable. This information can be utilized to refine the model's architecture and ensure that it is making the most accurate predictions possible in Figure 12.

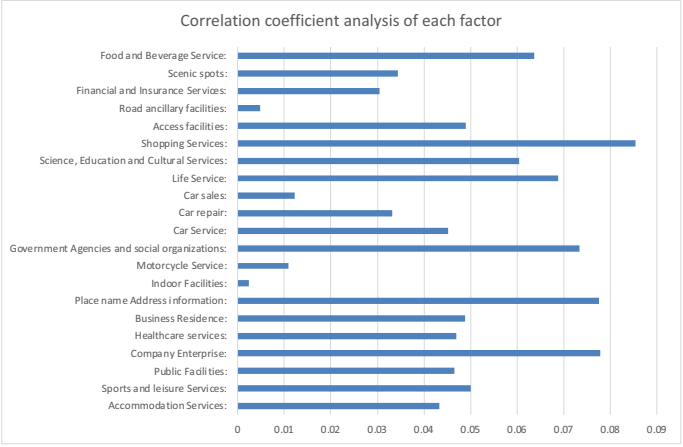


Figure 12. Correlation coefficient analysis of each factor

4. Conclusion

In data processing, we conducted random sampling and obtained various data features related to tobacco market conditions. We then preprocessed the data, which included cleaning, handling missing values, and standardizing features. We utilized a machine learning proxy model for feature extraction based on Transformer architecture. After evaluating different model algorithms, we selected XGBoost for predicting future product markets. Continuous modification and optimization of hyperparameters improved the model's prediction accuracy. Extreme gradient boosting models for sales is used for forecasting and analysis. By calculation, the R^2 score is 0.86, with a mean square error of 2.3, and a root mean square error of 1.5, and an average absolute error of 0.78, which meets the requirement of the model's prediction accuracy exceeding 0.7 in the previous text. In addition, in the scatter plot of prediction accuracy comparison, within the range of data size less than 20, the prediction accuracy is relatively high, basically in line with the standard regression line y=x, proving that the prediction accuracy is high within this interval. Within the range of model size greater than 30, the degree of deviation from the y=x regression line will increase due to the presence of other accidental factors.

LSTM and GRU models are used to predict order sales under different time series. By comparing the evaluation coefficients between different models, it was found that in this project, the prediction accuracy of LSTM and GRU models was similar. By observing the scatter plot of accuracy comparison, it can be seen that except for some noisy data, LSTM and GRU neural network models have small errors in the prediction results across the entire data range, with R^2 scores above 0.8, which can meet the accuracy requirements of this article well. Moreover, they can effectively converge to both sides of the regression line y=x in the scatter plot, proving that they can effectively predict sales in time series. In addition, The Spearman correlation analysis helps to identify the crucial factors influencing the predicted price, enabling developers to refine the model and enhance its predictive accuracy.

References

[1] N. Caglayan, S. I. Satoglu, and E. N. Kapukaya, (2020) "Sales forecasting by artificial neural networks for the apparel retail chain stores- An application," Journal of Intelligent & Fuzzy Systems, vol. 39, no. 5, pp. 6517-6528. doi: 10.3233/jifs-189115.

[2] J. Huber and H. Stuckenschmidt, (2020) "Daily retail demand forecasting using machine learning with emphasis on calendric special days," International Journal of Forecasting, vol. 36, no. 4, pp. 1420-1438. doi: 10.1016/j.ijforecast.2020.02.005.

[3] Y. Huang and Y. H. Zhang, (2022) "Novel market decision approach using data management technique with short supply," Computers & Electrical Engineering, vol. 100, no. 4, pp. 856-872. doi: 10.1016/j.compeleceng.2022.107881.

[4] X. D. Wang, Y. C. Su, W. T. Xu, S. Zhang, and Y. J. Zhang, (2023) "A market decision-making model for load aggregators with flexible load," Frontiers in Energy Research, vol. 10, no. 7, pp.1030076. doi: 10.3389/fenrg.2022.1030076.

[5] M. A. Khan, S. Saqib, T. Alyas, A. U. Rehman, Y. Saeed, A. Zeb, M. Zareei, and E. M. Mohamed, (2020) "Effective Demand Forecasting Model Using Business Intelligence Empowered with Machine Learning," Ieee Access, vol. 8, pp. 116013-116023. doi: 10.1109/access.2020.3003790.

[6] S. W. Ji, X. J. Wang, W. P. Zhao, and D. Guo, (2019) "An Application of a Three-Stage XGBoost-Based Model to Sales Forecasting of a Cross-Border E-Commerce Enterprise," Mathematical Problems in Engineering, vol. 2019, no. 8, pp.1361-1379. doi: 10.1155/2019/8503252.

[7] P. Jiang, Y. B. Huang, and X. Liu, (2021) "Intermittent demand forecasting for spare parts in the heavy-duty vehicle industry: a support vector machine model," International Journal of Production Research, vol. 59, no. 24, pp. 7423-7440. doi: 10.1080/00207543.2020.1842936.

[8] M. Käck and M. Freitag, (2021) "Forecasting of customer demands for production planning by local k-nearest neighbor models," International Journal of Production Economics, vol. 231, no. 4, pp.375-383. doi: 10.1016/j.ijpe.2020.107837.

[9] M. Massaoudi, S. S. Refaat, I. Chihi, M. Trabelsi, F. S. Oueslati, and H. Abu-Rub, (2021)"A novel stacked generalization ensemble-based hybrid LGBM-XGB-MLP model for Short-Term Load Forecasting," Energy, vol. 214, no. 9, pp.13-23. doi: 10.1016/j.energy.2020.118874.

[10] M. K. Mohanty, P. K. G. Thakurta, and S. Kar, (2023) "Agricultural commodity price prediction model: a machine learning framework," Neural Computing & Applications, vol. 35, no. 20, pp. 15109-15128. doi: 10.1007/s00521-023-08528-7.

[11] A. Özmen, (2023) "Sparse regression modeling for short- and long-term natural gas demand prediction," Annals of Operations Research, vol. 322, no. 2, pp. 921-946. doi: 10.1007/s10479-021-04089-x.

[12] J. Zhao, H. Zhu, and S. Zheng, (2018) "What is the value of an online retailer sharing demand forecast information?," Soft Computing, vol. 22, no. 16, pp. 5419-5428. doi: 10.1007/s00500-018-3091-3.

[13] J. Buckell and S. Hess, (2019) "Stubbing out hypothetical bias: improving tobacco market predictions by combining stated and revealed preference data," Journal of Health Economics, vol. 65, no. 20, pp. 93-102. doi: 10.1016/j.jhealeco.2019.03.011.

[14] A. Biswas, R. Vasudevan, M. Ziatdinov, and S. V. Kalinin, (2023) "Optimizing training trajectories in variational autoencoders via latent Bayesian optimization approach *," Machine Learning-Science and Technology, vol. 4, no. 1, pp.1147-1153. doi: 10.1088/2632-2153/acb316.

[15] J. Z. Chen, M. M. Aral, S. H. Kim, C. Park, and Y. Xie, (2023) "Constrained Bayesian optimization and spatio-temporal surveillance for sensor network design in the presence of measurement errors," Engineering Optimization, vol. 55, no. 3, pp. 510-525. doi: 10.1080/0305215x.2021.2014475.

[16] L. Zou, K. J. Zhuang, A. Zhou, and J. Hu, (2023) "Bayesian optimization and channel-fusion-based convolutional autoencoder network for fault diagnosis of rotating machinery," Engineering Structures, vol. 280, no. 9, pp.1105-1113. doi: 10.1016/j.engstruct.2023.115708.

[17] H. Yan, J. Zhang, N. Zhou, P. Shi, (2023) "SC-CO2 Fracturing Effect Prediction of Coal and Rock Mass Based on Da-De-Svm Intelligent Model," Chinese Journal of Geotechnical Engineering, vol. 45, no. 2, pp. 362-368. doi: 10.11779/CJGE20211287

[18] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," presented at the Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '16). New York, USA, 2016. doi: 10.1145/2939672.2939785.

[19] C. Ding, X. Wang, T. Chen, and T. Yu, (2023) "Application of Bayesian Optimized XGBoost in Seismic Interpretation of Small-scale Faults," Journal of China Coal Society, vol. 48, no. 6, pp. 2530-2539. doi: 10.13225/j.cnki.jccs.2021.1932.

[20] W. Hanittinan, "Resilient Modulus Prediction Using Neural Network Algorithm," Columbus: The Ohio State University, 2007. http://rave.ohiolink.edu/etdc/view?acc_num=osu1190140082.

[21] D. R. Jones, M. Schonlau, and W. J. Welch, (1998) "Efficient Global Optimization of Expensive Black-box Functions," Journal of Global Optimization, vol. 13, no. 4, pp. 455-492. doi: 0.1023/A:1008306431147.

[22] N. Dong, S. Lu, F. Xiong, (2021) " Cost prediction of construction projects based on ABC-SVM under the big data context, "Technology Economics, vol. 40, no. 8, pp. 25-32. https://kns.cnki.net/kcms2/article/abstract?v=C_qHzaiysLWUTinV_n7KJ4qa9kKhk5aLhZdBSTXz8ZBYFKTB7TYwwoSkpUNTd2gAiTXxgabgaj72mn46oMkQFn7Vg4_uO8nAAlArdNn8A2iyvvizUOsc660SzBXJrOTF1NCRisulSzLiPlKUbpTASA==&uniplatform=NZKPT&language=CHS.

[23] Y. Chen, S. Zheng, W. Liu, (2021) "A study on the cost prediction of industrial construction projects based on machine learning," Journal of Wuhan University of Technology (Information and Management Engineering), vol. 43, no. 4, pp. 314-321. https://kns.cnki.net/kcms2/article/abstract?v=C_qHzaiysLUECmG_LylHK5uxyOMtWje4i8QK44sPUMeIvtebxv5sIbqJPTTf5hOuHfUA67F5qkuY5P29E_j5qxVVbmU1tpLTvIQCRY_IxcxFsr30RcMGrO2Z-FfqNPmr4ImL5VWOEgZgXoI_WoaWAw==&uniplatform=NZKPT&language=CHS.

[24] J. Yuan, P. Guo, T. Zeng, (2024) " A cost prediction model based on Bayesian linear regression," Journal of Wuhan University of Technology (Information and Management Engineering), vol. 46, no. 1, pp. 90-94. https://kns.cnki.net/kcms2/article/abstract?v=C_qHzaiysLXmt6KG7dzyVonzHsORK-lBln6N6xaZUX4yckhiBB_T3nAyVZM3BQLONy3ud9gIMAvBjEX-g2Zdvi8beg4oEFWzacCy2qkpzGPCq5hEN8CQ6yi9Q-nwBSNWa4WvC2fu4g9c_Jmig9XA_g==&uniplatform=NZKPT&language=CHS

[25] F. Shafie-Khorassani, M. E. Piper, D. E. Jorenby, T. B. Baker, N. L. Benowitz, T. Hayes-Birchler, R. Meza, and A. F. Brouwer, (2022) "Associations of Demographics, Dependence, and Biomarkers With Transitions in Tobacco Product Use in a Cohort of Cigarette Users and Dual Users of Cigarettes and E-cigarettes," Nicotine & Tobacco Research, vol. 25, no. 3, pp. 462-469, doi: 10.1093/ntr/ntac207.

[26] P. Sharma, R. Kalaskar, (2021) "Evaluation of Various Factors Which Motivate Children in Ashram Schools of Central India Region to Consume Tobacco Products," International journal of clinical pediatric dentistry, vol. 3, no. 1, pp. 554-557. doi: 10.5005/jp-journals-10005-1977.

[27] T.W. Mapuwei, J. Ndava, M. Kachaka, and B. Kusotera, (2022) "An Application of Time Series ARIMA Forecasting Model for Predicting Tobacco Production in Zimbabwe". American Journal of Modeling and Optimization, vol. 9, no. 1, pp. 15-22. doi: 10.12691/ajmo-9-1-3.

[28] N. Q. K. Le, (2019) "Fertility-GRU: Identifying Fertility-Related Proteins by Incorporating Deep-Gated Recurrent Units and Original Position-Specific Scoring Matrix Profiles," Journal of Proteome Research, vol. 18, no. 9, pp. 15-22. doi: 10.1021/acs.jproteome.9b00411.

[29] E. M. Dewi, T. L. R. Mengko, H. Zakaria and K. Astami, (2019) "Increased Arterial Stiffness in Catheterization Patient by Photoplethysmography Analysis," International Conference on Electrical Engineering and Informatics (ICEEI), Bandung, Indonesia, vol. 9, no.6, pp. 619-623. doi: 10.1109/ICEEI47359.2019.8988783.

[30] M. Potrykus, (2023) "Price bubbles in commodity market – A single time series and panel data analysis," The Quarterly Review of Economics and Finance, vol. 87, pp.110-117. doi: 10.1016/j.qref.2022.12.002.

[31] A. Field, (2005) "Discovering statistics using SPSS," London: Sage. https://psycnet.apa.org/record/2005-05622-000.

[32] J. H. Stock, M. W. Watson, (2015) "Introduction to econometrics," https://scholar.harvard.edu/stock/pages/introduction-econometrics.

[33] B. G. Tabachnick, L. S. Fidell, (2021) "Using Multivariate Statistics (7th ed.)," Boston, MA: Pearson. https://www.pearson.com/en-us/subject-catalog/p/using-multivariate-statistics/P200000003097/9780137526543.

[34] R. J. Hyndman, G. Athanasopoulos, (2018) " Forecasting: principles and practice (2nd ed)," OTexts. https://otexts.com/fpp2/.