Study on Freight Volume Prediction of Routes Based on Random Forest Model

Tianyu Du*, Qingwen Yin

School of Computer, Jiangsu University of Science and Technology, Zhenjiang, China, 212100

* Corresponding Author Email: 13305147201@163.com

Abstract. With the rise of e-commerce and the ever-changing consumer demand, logistic network sorting centers face increasingly complex and significant challenges in cargo volume forecasting. To predict the cargo volume for various routes, this paper selects the Random Forest model for forecasting. Feature data were extracted, and the dataset was divided into a training set and a test set. The model was evaluated using the test set. Ultimately, a high coefficient of determination and low mean squared error were achieved, demonstrating the accuracy and reliability of the Random Forest model's predictions. The transportation scenario was extended to changes in transportation routes, focusing on the impact of adding and canceling transportation routes on cargo volume forecasting. By establishing model features, conducting feature synthesis, and calculating impacts, the adjusted cargo volume for each sorting center was obtained, and an analysis of changes in transportation routes was conducted. The influence of added and canceled routes was integrated into the cargo volume forecast, resulting in the final cargo volume prediction. By distributing the adjusted volume evenly on a daily basis, the accuracy and reliability of the forecast results were ensured. Finally, a visual representation of the forecast results was provided, intuitively showcasing the model's effectiveness.

Keywords: Random Forest Model, Volume Forecasting, Route Alteration.

1. Introduction

In today's rapidly developing society, the logistics industry is indispensable, supporting global trade and the modern economic system. With the rise of e-commerce and changing consumer demands, logistics network sorting centers face complex challenges in volume forecasting and personnel scheduling. Accurate volume forecasting is crucial for optimizing inventory management, reducing costs, and enhancing customer satisfaction. A March 2020 paper developed a combined ARIMAX-SVR model for improving port cargo throughput forecasting at the Port of Tianjin [1]. A 2021 paper explored the time-varying characteristics of cargo transportation volume and used the Bagging+BP integrated learning method to improve multimodal freight prediction accuracy [2-5]. This article utilized the random forest algorithm to establish a predictive model, considering temporal features, historical volumes, seasonal variations, holidays, and promotional activities [6-9]. This model, optimized by adjusting parameters like the number of trees and maximum features, helps allocate resources more effectively, reduce wait times, lower costs, and improve logistics efficiency. In the logistics network, changes in transport routes impact sorting center volumes. This article reassess volume data before and after route changes to adjust or retrain the predictive model, ensuring it reflects the current transport network state [10]. The data for this article comes from http://www.mathorcup.org/ detail/2428.

2. Modeling and Solving

2.1. Modeling of daily cargo volume forecasts

2.1.1. Plotting selected time series

The data in this paper has been preprocessed. In selecting the forecasting model, this paper first plots its time series and observes whether there is seasonal cyclical change, if so, it is more appropriate to use seasonal time series analysis model. The plotted time series are shown in Figure 1.

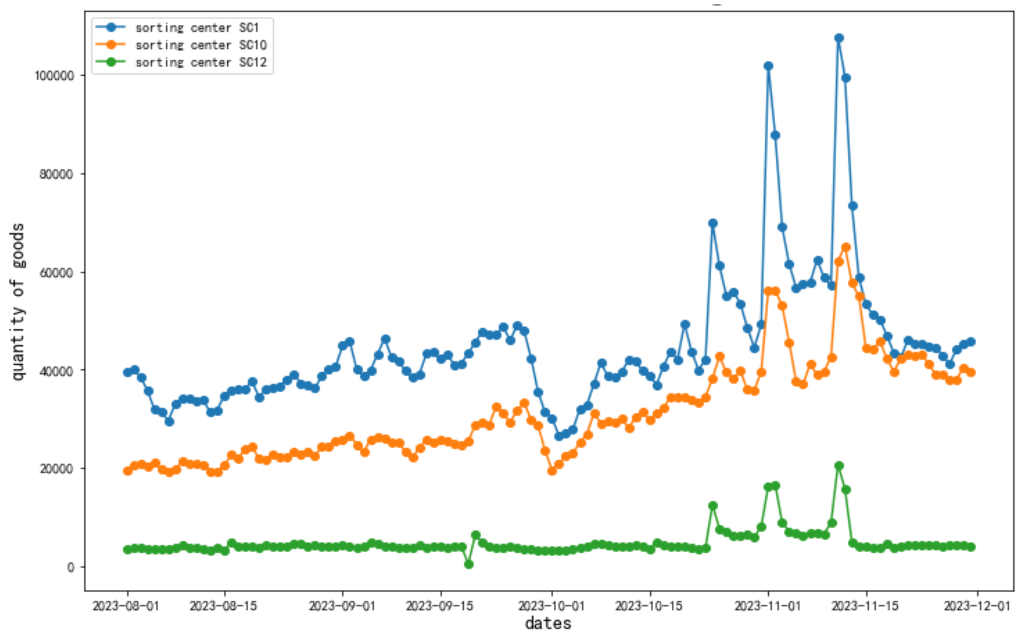

Figure 1. Time Series Chart - Selected Sorting Centers

It can be seen that all three of the selected sorting centers experienced more significant increases around November, making this question unsuitable for prediction using a time series analysis model.

Since the change of cargo volume does not satisfy the seasonal change, we decided to use the random forest model for prediction. First, we need to extract the feature data. We first divide the dataset into training set and test set, and we choose the time near data as the test set, which accounts for 20%. After social investigation, we finally choose seven variables such as year, month, day, and day of the week for training. The process of building a random forest model is shown in Figure 2.

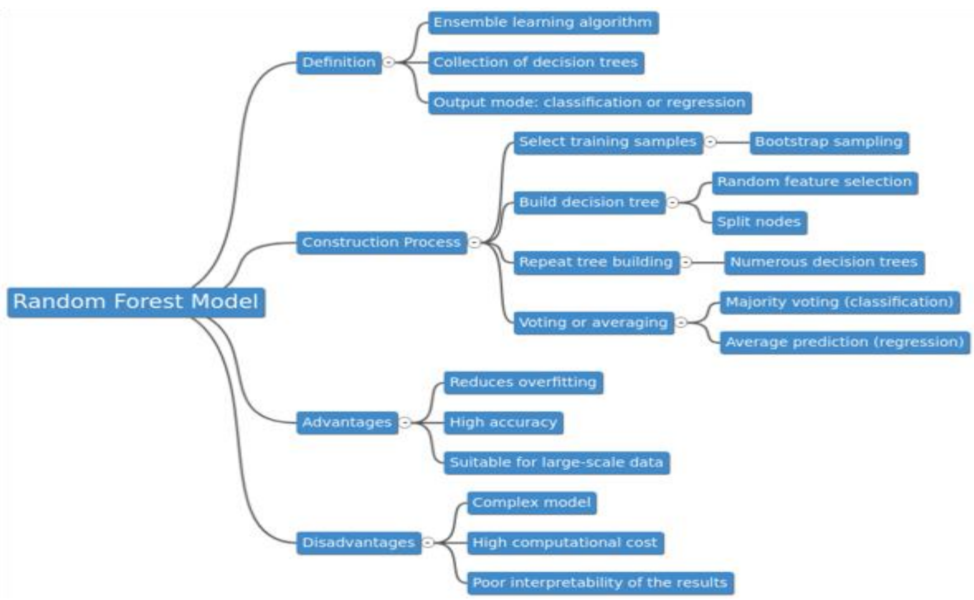

Figure 2. The process of building a random forest model.

2.2. Model solution

After the training, this paper evaluates the model through the test set, and we get the value of mean square error (MSE) is 17, 266, 658.03, and the value of coefficient of determination () reaches 0.963, as the closer the coefficient of determination is to 1, the more it can show the high prediction accuracy, which shows that the prediction accuracy of the random forest model is good enough to be used 5for prediction. The table of exported predictions is shown in Table.1.

Table 1. Predicted data using the random forest model.

	Sorting Center	Date	Year	Month	Day	Day of the Week	Previous day's cargo volume	Rolling average of 7 days	Rolling average 30 days	Forecasted volume
1	SC1	2023-12-1	2023	12	1	4	45707	44043.28571	58057.03333	52016.93
2	SC1	2023-12-2	2023	12	2	5	45707	44043.28571	58057.03333	46847.63
3	SC1	2023-12-3	2023	12	3	6	45707	44043.28571	58057.03333	44581.26
4	SC1	2023-12-4	2023	12	4	0	45707	44043.28571	58057.03333	45419.75
5	SC1	2023-12-5	2023	12	5	1	45707	44043.28571	58057.03333	46581.62
6	SC1	2023-12-6	2023	12	6	2	45707	44043.28571	58057.03333	46454.08
7	SC1	2023-12-7	2023	12	7	3	45707	44043.28571	58057.03333	46351.73
8	SC1	2023-12-8	2023	12	8	4	45707	44043.28571	58057.03333	46027.4
9	SC1	2023-12-9	2023	12	9	5	45707	44043.28571	58057.03333	46123.79
10	SC1	2023-12-10	2023	12	10	6	45707	44043.28571	58057.03333	44526.47
11	SC1	2023-12-1	2023	12	1	4	45707	44043.28571	58057.03333	52016.93

We plot the residuals of the predicted and actual values as Figure 3 and it is clear from the figure that the prediction is good.

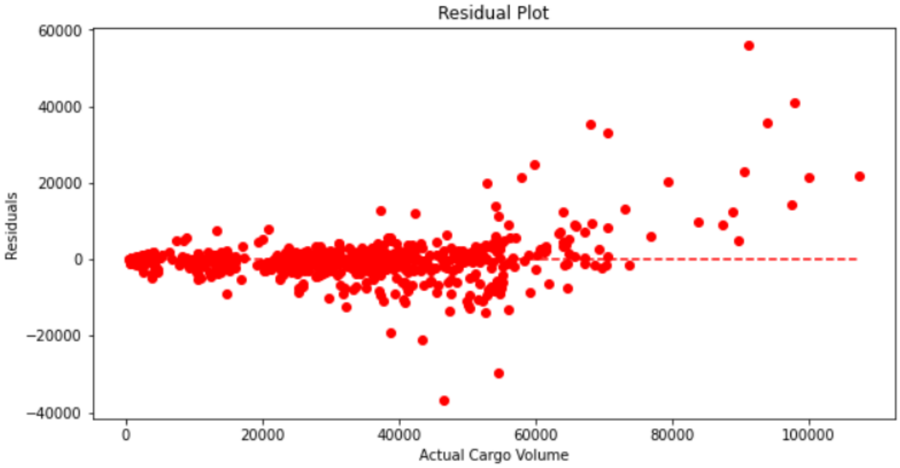

Figure 3. The process of building a random forest model

2.3. Mathematical model to address the impact of route changes on cargo flow forecasts

In this section, we focus on how changes in new transportation routes will affect the daily and hourly cargo flow over the next 30 days. To address this issue, we need to follow several steps:
First, import the data into the analytical tool to gain a deep understanding of the average cargo flow on transportation routes over the past 90 days, as well as any anticipated changes in transportation routes over the next 30 days. Then, perform an impact assessment. Delve into how these route changes will affect the cargo flow at each sorting center, and then proceed with feature adjustment. Adjust the input features in the forecasting model according to the changes in transportation routes. Finally, conduct prediction correction. Use the adjusted features and the previously trained model to predict future cargo flow.

To handle the impact of route changes on cargo flow predictions, this paper adopts a simplified mathematical model approach to estimate the changes in daily and hourly cargo flow over the next 30 days. Below are the detailed theoretical foundations and mathematical modeling steps of the process.

2.3.1. Model Feature Construction

Anticipated volume of shipments on new transportation routes P_{new} Estimated formula

$$P_{new} = \sum_i h_i \tag{1}$$

Where h_i is the historical average volume of shipments on the same routes.

Reduced volume for canceled transportation routes P_{drop} Estimated formula

$$P_{drop} = \sum_i h_i \tag{2}$$

Where h_j is the average volume of cargo on historically canceled routes.

2.3.2. Feature synthesis and impact calculations

For each sorting center, all relevant new and canceled line loads are summarized to calculate the total load adjustment:

$$\Delta\mathrm{P}: \Delta\mathrm{P} = P_{new} + P_{drop} \tag{3}$$

Final adjusted shipments per sorting center:

$$\mathrm{L_{last}}: \mathrm{L_{last}} = \mathrm{L_{pre}} + \Delta P \tag{4}$$

Where $\mathrm{L_{pre}}$ is the original predicted cargo volume based on the model's failure to account for route changes.

2.3.3. Practical Solution

In this subsection, an analysis is presented of how changes in transit railroads over the next 30 days may affect the volume of shipments at the sorting centers, which includes consideration of new routes added and routes that may be eliminated. In this paper, these changes are combined to characterize the future cargo volume forecasts.

We look for additional transportation routes as shown in Table.2.

Table 2. Predicted data using the random forest model.

Outgoing Sorting Center	Terminal Sorting Center
SC31	SC9
SC31	SC53

Eliminated transportation routes as shown in Table.3.

Table 3. Eliminated Transportation Routes.

	Outgoing Sorting Center	Terminal Sorting Center	Quantity of Goods
11	SC36	SC8	97
16	SC19	SC15	336
20	SC4	SC15	15
21	SC51	SC15	12
36	SC36	SC47	113
42	SC55	SC7	128

According to the addition and cancellation of transportation routes, adjust the volume characteristics of the relevant sorting centers. For example, if a route is canceled, we may need to reduce the expected volume for the sorting centers affected by that route; if a route is newly added, we might need to increase the expected volume.

The feature adjustment method adopted in this paper is: integrating historical data and using the historical average volume data to estimate the possible volume of the newly added routes and the possible reduction of volume for the canceled routes.

Code execution steps: First, create a volume change table for each sorting center, then initialize the volume adjustment for all possible sorting centers to 0, then calculate the volume increase for each sorting center due to newly added routes, then calculate the volume decrease for each sorting center due to canceled routes, and finally calculate the final adjusted volume impact. The results obtained after solving are as shown in Table.4.

Table 4. Impact of final adjustments to cargo volumes.

Sorting Center	Additional cargo	Reduction of cargo volume	Volume adjustment
0	SC1	0	-254
1	SC10	0	0
2	SC12	0	0
3	SC14	0	0
4	SC15	0	0

Integrate these adjustments into the shipment forecast and use these adjustments to update the daily and hourly shipment forecast for the next 30 days.

When applying the adjustments, we should ensure that the total daily adjustment is spread out over each hour, rather than subtracting the entire daily adjustment for each hour. This means that the "reduction in volume" should be distributed evenly over the number of hours.

Using the features that have been tuned, combined with the previously trained model to predict the next 30 days of shipments. For the daily prediction, after the prediction we perform a visualization as shown in Figure 4.

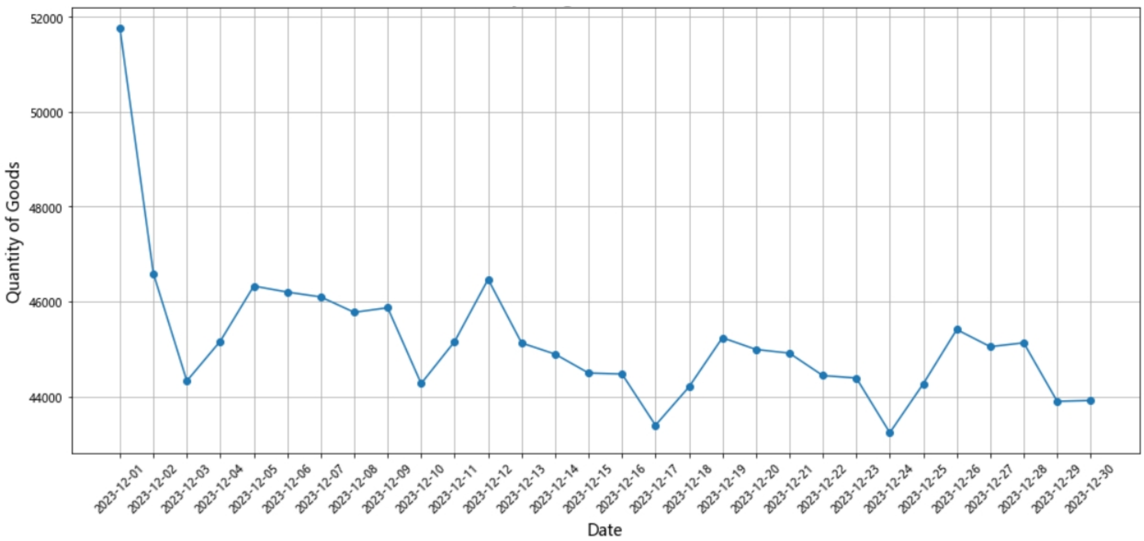

Figure 4. Daily Cargo Forecast-SC1.

For hourly forecasts, we can get similar results, visualized as as shown in Figure 5.

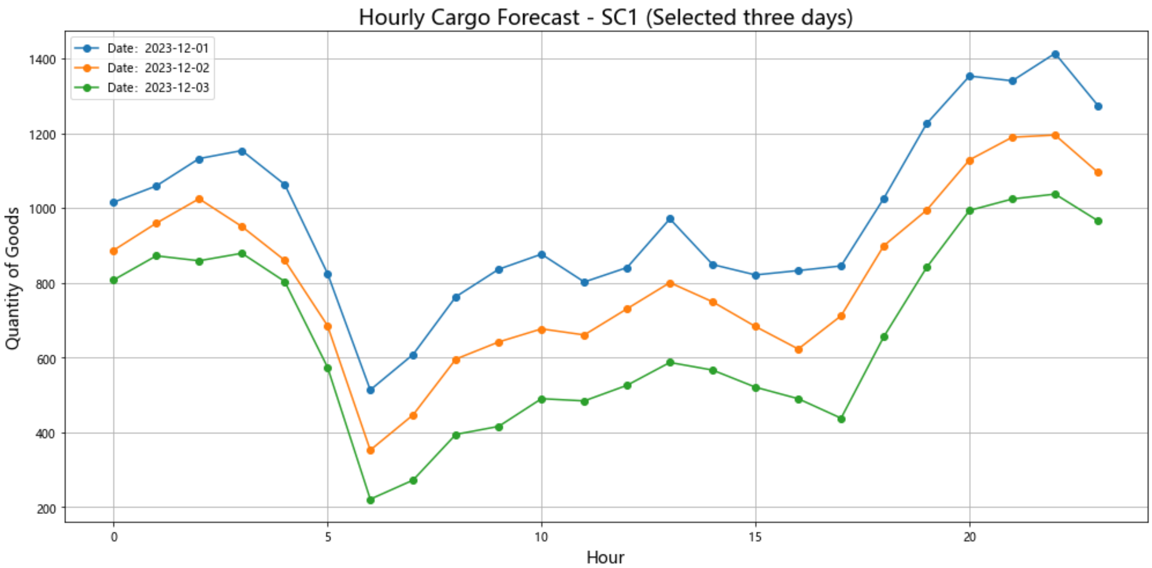

Figure 5. Hourly Cargo Forecast - SC1 (Selected three days).

3. Conclusions

In conclusion, this study effectively demonstrates the significant role of mathematical models in addressing practical challenges within the logistics industry, particularly under the rapid growth of e-commerce and dynamic market conditions. By leveraging advanced mathematical tools such as the random forest algorithm for cargo volume forecasting and other optimization techniques for personnel scheduling, the research highlights substantial improvements in operational efficiency and cost reduction for logistics sorting centers. The integration of these methodologies not only enhances the processing capacity and responsiveness of sorting centers but also provides robust decision support for management. This underscores the practical application value of mathematical modeling in solving intricate business problems, ultimately contributing to more resilient and efficient logistics operations. Moreover, the study emphasizes the importance of accurate cargo volume forecasting in anticipating demand fluctuations and making informed staffing decisions, which can significantly impact the overall performance of the logistics centers. The findings suggest that continuous refinement and adaptation of these models are essential for keeping pace with the evolving demands of the logistics sector, ensuring sustained efficiency and competitiveness. By implementing these advanced mathematical approaches, logistics companies can better manage resources, minimize operational costs, and improve service levels, ultimately driving higher customer satisfaction. The successful application of these models demonstrates their potential to transform logistics operations, making them more agile and capable of meeting the challenges posed by the fast-paced growth of e-commerce. Therefore, the ongoing development and application of mathematical models remain crucial for the continued advancement and optimization of logistics operations in today's competitive market

References

[1] Wang S, Li HQ, Zhang WL. Research on logistics cargo volume prediction model based on big data [J]. China Management Science, 2020, 28: 102-110.

[2] Liu Q, Zhao TY. An empirical analysis of the application of machine learning methods to logistics volume forecasting [J]. Systems Engineering Theory and Practice, 2020, 30: 88-94.

[3] Chen LQ, Ma TT, Zhao L. Research on optimization of logistics network and cargo volume prediction [J]. Logistics Technology, 2020, 39: 45-50.

[4] Huang X, Yang Y. A comparative study of cargo volume forecasting models integrating time series and machine learning [J]. Computer Engineering and Applications, 2020, 56: 123-128.

[5] Li N, Zhang H. Logistics cargo volume forecasting based on SARIMA model and its application [J]. Statistics and Decision Making, 2021, 36: 159-162.

[6] Zhao XY, Wang SY. Construction of logistics cargo volume prediction model based on deep learning [J]. Computer Systems Applications, 2021, 30: 214-219.

[7] Zheng W, Zhou GF. Research on cargo volume prediction model and application in logistics centers [J]. Systems Engineering, 2021, 39: 153-158.

[8] Xu HH, Li XJ. Research on gray system theory of logistics cargo volume forecasting [J]. China Logistics and Purchasing, 2022, 23: 76-79.

[9] Zhang R, Wu TT. Research on optimization model of cargo volume prediction and staff arrangement in logistics enterprises [J]. Logistics Technology, 2022, 41: 112-116.

[10] Zhou J, Liu YN. Improved neural network-based logistics cargo volume prediction method [J]. Computer Engineering and Applications, 2022, 58: 134-139.