JOURNAL OF EMERGING ECONOMIES AND POLICY 2023 8(2) 416-423

JOEEP

e-ISSN: 2651-5318
Journal Homepage: http://dergipark.org.tr/joeep

Araştırma Makalesi ● Research Article

Forecasting the China Container Freight Index with Ensemble Models

Ensemble Modelleriyle Çin Konteyner Navlun Endeksi'nin Tahmin Edilmesi

Tolga Tuzcuoğlu [a], Hüseyin Gençer [b, *]

[a] Dr., Turkish German University, Faculty of Economics and Administrative Sciences, Department of Business, 34820, Istanbul/Turkiye
ORCID: 0000-0002-5269-9701
[b] Dr., Piri Reis University, Faculty of Economics and Administrative Sciences, Department of Maritime Business Management, 34940, Istanbul/Turkiye
ORCID: 0000-0002-4945-4420

M A K A L E B İ L G İ S İ

Makale Geçmişi:
Başvuru tarihi: 7 Eylül 2023
Düzeltme tarihi: 24 Ekim 2023
Kabul tarihi: 25 Ekim 2023

Anahtar Kelimeler:
Navlun Endeksi Tahmini
Toplu Modeller
Finansal Tahminleme

ÖZ

Konteyner taşımacılığında navlun oranlarının nasıl değişeceğini sektördeki paydaşların öngörebilmesi oldukça önemlidir. Bu çalışma, literatürde ilk defa olmak üzere, konteyner taşımacılığında navlun oranlarının değişimini gösteren en önemli göstergelerden biri olan CCFI (Çin Konteyner Navlun Endeksi) 'nin tahmini için toplu zaman serisi modelleri sunmaktadır. Çalışmanın sonuçları, modellerin CCFI'nin tahmininde oldukça iyi sonuçlar verdiğini ve önemli bir karar destek sistemi olarak kullanılabileceğini göstermektedir.

A R T I C L E I N F O

Article history:
Received: Sep 7, 2023
Received in revised form: Oct 24, 2023
Accepted: Oct 25, 2023

Keywords:
Freight Index Prediction
Ensemble Models
Financial Forecasting

A B S T R A C T

Stakeholders in the sector need to be able to predict how freight rates will change in container transportation. For the first time in the literature, this study presents aggregate time series models for the prediction of CCFI (China Container Freight Index), one of the most critical indicators showing the change of freight rates in container shipping. The study results show that the models provide promising results in forecasting CCFI and can be used as an essential decision support system.

1. Introduction

Containerization has been one of the most important catalysts of globalization. It significantly impacts world trade as more than 50% of the cargo in terms of value is transported by container shipping (UNCTAD, 2021). The maritime industry has faced various challenges in the last 15 years, such as the global financial crisis, the COVID-19 pandemic, Evergiven's closure of the Suez Canal, and the Russia-Ukraine conflict. Figure 1 shows the global container shipping in terms of million tonnes and TEU between 1996 and 2022. As can be seen, there has been an increase in container shipping almost every year, except for the global financial crisis in 2009 and the beginning of the COVID-19

* *Sorumlu yazar/Corresponding author*.
e-posta: *hgencer@pirireis.edu.tr*

Atıf/Cite as: Tuzcuoğlu, T. & Gençer, H. (2023). Forecasting the China Container Freight Index with Ensemble Models. *Journal of Emerging Economies and Policy*, 8(2), 416-423.

pandemic in 2020.

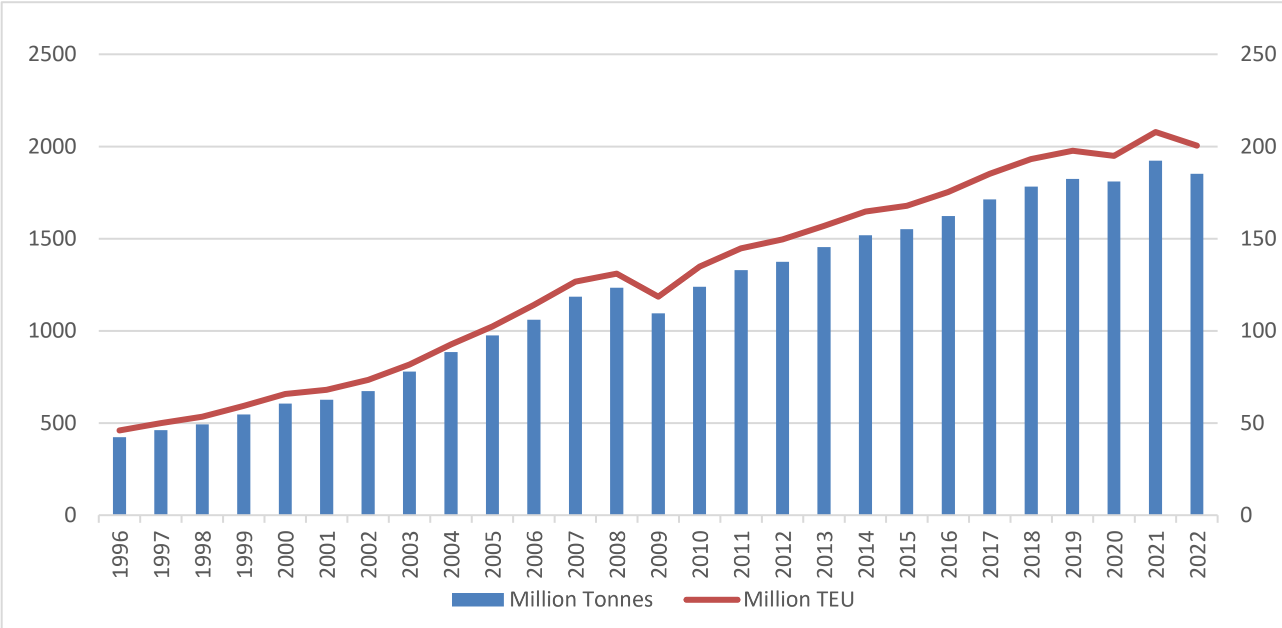

Figure 1. Global Container Shipping 1996-2022 **Source**: Clarksons Research (2023)

Container shipping is one of the best examples of liner shipping in the maritime industry. The demand for liner services is a result of the derived demand stemming from the demand for imported commodities. It is worth noting that transportation costs generally constitute a very minor component of market prices. Consequently, trade flows exhibit less sensitivity to fluctuations in transportation costs than market price fluctuations. All else being equal, an increase in the fraction of transportation costs in the sale price will significantly impact trade flows. Shippers opt for liner services primarily due to the relatively modest size of their consignments compared to the larger vessels employed in maritime trade. Ensuring a stable supply of consumer goods and industrial inputs across all markets is of utmost importance for manufacturers. If shipping services were to function irregularly, the expenses associated with warehousing commodities at both ends of the trade would be substantial, particularly during elevated interest rates. The ability to provide precise delivery schedules is of utmost importance for capital goods providers throughout the bidding process for a project (ICS, 2015).

The predictability of transportation costs to potential export markets is crucial since a significant variability in these costs can amplify the risk associated with market development. This is particularly important for manufactured products since they are sold at predetermined or negotiated prices several months in advance. Conversely, this is less applicable to bulk cargoes, as their prices exhibit more significant fluctuations than freight rates (Stopford, 2008).

Container shipping is primarily suitable for freight markets with smaller parcel transportation sizes. While the charter market encompasses the aggregate carrying capability of the entire vessel, the freight market specifically deals with smaller pieces than a complete vessel. Liner carriers consider many factors, such as voyage cost, BAF (Bunker Adjustment Factor), CAF (Currency Adjustment Factor), port handling charges, and container imbalance costs, while determining the freight price (Notteboom, 2012). Although such costs are considered when proposing the freight price to the shippers, the final price is determined according to the market. In other words, the demand for container shipping and the supply of vessels determines the freight price. Nevertheless, freight prices are also immediately affected by extraordinary situations such as natural disasters, pandemics, and financial crises. Many organizations have initiated the production of their freight rate level indices in response to the need for more transparency regarding the fluctuations of freight costs. Establishing an index within the shipping sector, particularly concerning freight rates and market circumstances, is essential due to many factors. Shipping indexes are of utmost importance in the maritime industry as they are vital for enhancing transparency, establishing benchmarks, managing risks, and facilitating informed decision-making (ICS, 2015). This study proposes ensemble methods for forecasting CCFI (China Containerized Freight Index), the most used index in container shipping.

CCFI is commonly utilized as a reliable indicator of the container shipping market's conditions, making it a widely adopted tool. It considers the freight rates on 12 important routes of 23 domestic and foreign shipping companies with high international prestige and large market shares. Over the past decade, the CCFI has effectively fulfilled its role in capturing market trends, resulting in significant global impact and notable economic and social consequences. Due to its rigorous scientific methodology and authoritative stance, the CCFI is widely recognized as the second most prominent freight index globally, behind only the Baltic Dry Bulk Freight Index. In addition to the composite index, CCFI, individual indexes are published for 12 vital trade lanes (SSE, 2023).

In the literature, there is no study using ensemble methods for machine learning to forecast freight rates or indexes in container shipping. This study makes the forecasting of CCFI with ensemble methods and compares these methods. The second part of the study reveals some critical studies conducted in the literature on freight forecasting. The third section of the study describes the data and methodology. The fourth section presents the empirical results. The last section concludes the study with discussions.

2. Literature Review

Since estimating future freight rates has been an intriguing issue in the shipping business for a significant amount of time, academics have put a variety of econometric models to the test throughout the past several decades. While there has been a greater focus on studying freight forecasting in the bulk sector, primarily due to the abundance of data and the maturity of the market, there has also been a notable growth in research on freight estimating in container shipping in recent years (Munim and Schramm, 2021).

Luo et al. (2009) examined the fluctuations in container freight rates resulting from the interplay between the demand for container shipping services and the capacity of container vessel fleets. The model developed in their study demonstrated a significant ability to mimic the historical fluctuations in the container shipping industry accurately.

Xin (2010) analyzed the process of compiling the CCFI. The author examined several aspects related to the selection of calculating formula, identification of freight type, choice of shipping line samples, and regulations governing the revision of the index. Nielsen et al. (2014) studied the correlation between aggregated market prices, namely the Shanghai Containerized Freight Index (SCFI), and individual liner rates. The model under consideration emphasizes the aspects of performance and robustness, specifically the adequacy of observations and the forecasting timeframe. Fan and Yin (2015) analyzed the dynamic connections between the costs of new and second-hand container ships and the time charter rates.

In their study, Munim and Schramm (2017) utilized the ARIMA model and a hybrid approach combining ARIMA and autoregressive conditional heteroscedasticity (ARCH) models, referred to as ARIMARCH. They employed these models to analyze the SCFI and the CCFI at monthly and weekly intervals. Yifei et al. (2018) proposed a daily container freight index based on the data taken from the E-platforms by analyzing the freight prices. Jeon et al. (2019) employed the system dynamics technique to examine the cycles of the CCFI. Their study can be a valuable reference for decision-makers involved in ship investment timing. Chen et al. (2021) introduced a novel methodology integrating empirical mode decomposition and grey wave techniques to forecast the CCFI.

Munim and Schramm (2021) presented a comparative analysis of artificial neural networks and traditional models in the context of forecasting container freight rates on essential trade routes. The particular details of the findings are not explicitly mentioned in their study. However, the primary objective of their research is to emphasize the possible benefits associated with using sophisticated neural network models as opposed to conventional forecasting techniques within the marine industry.

Koyuncu et al. (2021) studied the impact of COVID-19 on maritime trade, focusing on the RWI/ISL Container Throughput Index. They used the Seasonal Autoregressive Integrated Moving Average (SARIMA) model and Exponential Smoothing State Space Model (ETS) methodologies to forecast the index, finding a sustained downward trend. Deng and Yang (2021) examined the co-integration relationship between China's coastal bulk freight index and the freight rates of selected routes inside the country. The study constructed a vector auto-regressive (VAR) model using weekly data from January 2010 to November 2019. The empirical study reveals a co-integration connection between the two variables, suggesting a bidirectional causal link. Both variables are responsive to changes in each other, and error correction models are available for analysis. Schramm and Munim (2021) presented a new methodology for forecasting freight rates in container shipping by including qualitative data from external sources. The research demonstrates that incorporating qualitative factors can enhance the precision of freight rate forecasting.

Hirata and Matsuda (2022) investigated using deep learning algorithms, particularly the long short-term memory (LSTM) technique, to forecast container freight rates. The authors compared the Long Short-Term Memory (LSTM) model and the Seasonal Autoregressive Integrated Moving Average (SARIMA) technique in the context of forecasting the SCFI. The results indicate that Long Short-Term Memory (LSTM) models perform better than Seasonal Autoregressive Integrated Moving Average (SARIMA) models across most datasets. Specifically, LSTM models have significantly decreased forecasting errors by as much as 85% for specific routes.

The study of Saaed et al. (2023) elucidated that the accurate projection of freight rates might assist cargo owners and shipping lines in making prompt judgments about their asset management strategies. Their research highlights the importance of forecasting in the maritime industry's operational decision-making context. Tu et. Al (2023) investigated the CCFI as an indicator of the Chinese shipping industry. Their research gathered data about six variables that impact the shipping industry. Subsequently, an analytical framework was developed utilizing the DNN (Deep Neural Network), CatBoost regression, and robust regression models. The deep neural DNN model exhibited superior prediction ability concerning the CCFI. The results of the robust regression analysis demonstrate that the variable "Global: Aluminum (minimum purity of 99.5%, LME spot price): UK landed price" exhibits the highest level of significance with its influence on the CCFI. Fei and Zhou (2023) examined the utilization of technical indicators to offer valuable perspectives to investors engaged in stock market analysis. More specifically, the study concentrates on the Shanghai Composite Index, spanning November 1994 to March 2022. The research used 20 technical indicators to predict the Shanghai Composite Index's excess return rate, providing high accuracy during the economic cycle contraction phase. This helps investors achieve enhanced returns without transaction costs.

There is a lack of research utilizing ensemble methods in machine learning for forecasting freight rates or indices in container shipping. This study will be the first endeavor in this particular field.

3. Data and Methods

The study utilized a dataset from the Shipping Intelligence Network Timeseries component of the Clarksons Research Database, a renowned global provider of shipping-related data. The dataset included weekly CCFI values from March 14, 2003, to August 18, 2023, with 1,038 observations. Figure 2 illustrates the temporal evolution of the index values during the specified timeframe. The analysis of Figure 1 reveals significant fluctuations in the CCFI between 2020 and 2022. These variations can be attributed to the profound influence exerted by the COVID-19 pandemic on the container shipping industry.

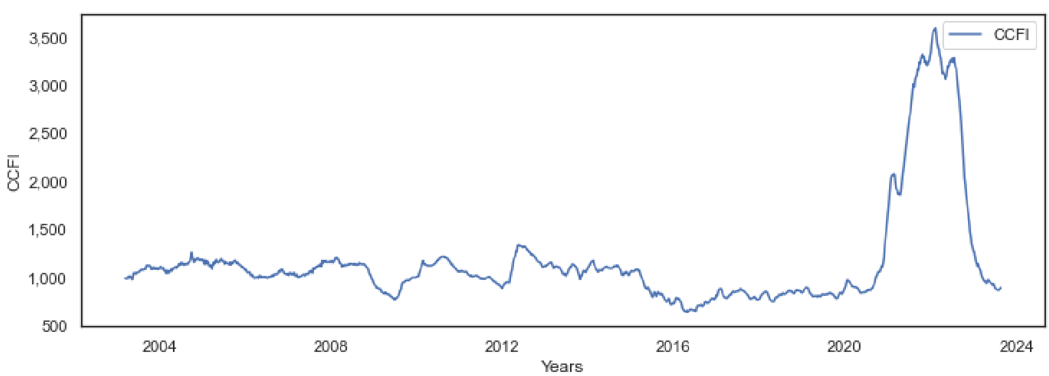

Figure 2. The Development of CCFI from 2003 to 2023 **Source**: Clarksons Research (2023)

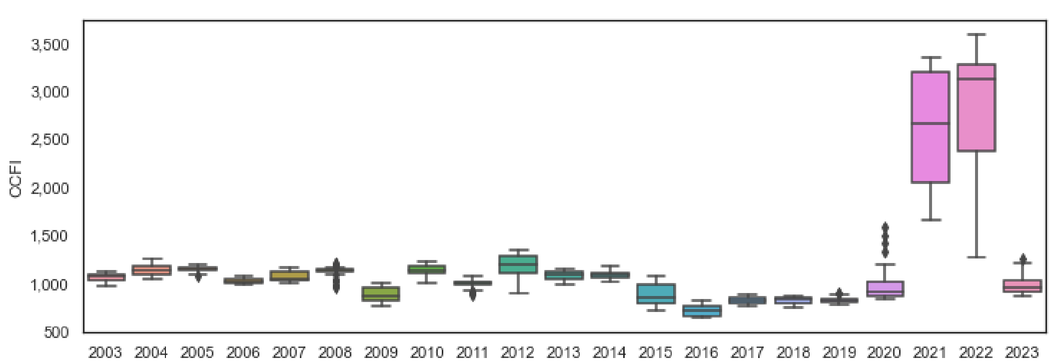

Figure 3. Boxplots of CCFI Years

The boxplots in Figure 3 further analyze the yearly fluctuations in the index values. The graphic representation demonstrates notable volatility in the weekly CCFI values for the period spanning from 2021 to 2023. To provide a comprehensive analysis, it is essential to present the statistical measures of the CCFI during the whole duration under consideration. Specifically, the mean, median, and standard deviation of the CCFI were 1,161.93, 1,044.56, and 567.05, respectively. In contrast, the CCFI exhibited mean values of 970.93, 2,597.49, and 2,792.14 in 2020, 2021, and 2022, respectively. However, this figure declined to 979.50 during the initial eight months of 2023. The median values observed over the last four years were 901.34, 2,653.32, 3,123.10, and 949.53. Correspondingly, the standard deviations for these successive years were 170.43, 581.86, 722.64, and 104.39.

Regarding the methodology employed, the prediction of CCFI values was carried out using five ensemble time series models. The ensemble regressors examined in this study were the random forest, light gradient-boosting machine (LightGBM), extreme gradient-boosting (XGBoost), adaptive boosting (AdaBoost), and categorical boosting (CatBoost) models. The models are implemented by calling libraries from the Python programming language. The performance metric utilized in this study was the root mean square error (RMSE). The dataset was initially partitioned into separate train and test datasets to compare the algorithms' RMSE values. Out of 1,038 observations, a subset of 24 observations was designated as the test dataset. The models underwent training using an initial dataset consisting of 1,014 observations. Subsequently, their predictions for the next 24 weeks were evaluated by comparing them to the actual values, employing the RMSE metric.

Ensemble models integrate the judgments made by numerous weak learners to construct a more robust learner that exhibits enhanced predictive capability and stability. This phenomenon is sometimes referred to as the wisdom of the crowd. While the random forest model utilizes bootstrap aggregating (bagging), XGBoost, LightGBM, Adaboost, and Catboost apply a boosting methodology. In the bagging method, the process of resampling the training set is conducted in a manner that is not influenced by the performance of the prior classifiers. In contrast, boosting algorithms leverage the knowledge acquired from previously trained models to iteratively adjust the weights assigned to data points (Kunapuli, 2023).

4. Empirical Results

The empirical analysis began with determining the baseline RMSE value derived from the naïve forecast. The naïve method can be considered the anchor point of the study since it takes the last actual observation as the predicted value over the forecast horizon. Any RMSE value lower than that of the naïve method is regarded as improved performance. Hence, the rest of the models are evaluated to the extent that they outperform the RMSE of this baseline model.

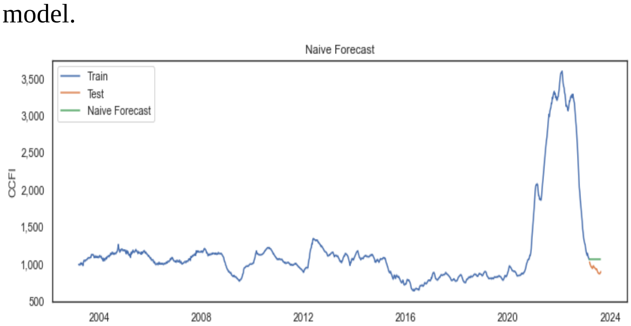

Figure 4. The Naïve Forecast

The empirical study calculated the baseline RMSE value obtained from the naïve forecast. The naïve technique may be regarded as the baseline approach in the study, as it utilizes the most recent observed value as the anticipated value for the whole forecast period. Improved performance is having an RMSE value lower than the naïve technique's. Therefore, the remaining models are assessed based on their ability to surpass the RMSE of the baseline model. Consequently, the RMSE at the baseline level was computed to be 137.34. The resulting forecast is visually represented in Figure 4, presented as follows.

After establishing the baseline value, further predictors were generated based on the date component of the series. To clarify, the existing index dates were expanded to include additional columns that indicate various temporal aspects such as the day of the week, quarter, month, year, day of the year, day of the month, and week of the year. Consequently, a further seven columns were incorporated into the dataset.

The random forest method was the first ensemble model to be examined in the following stage. The model was instantiated using the default hyperparameters, with the number of trees (estimators) set at 500. The RMSE obtained from the evaluation of the test dataset was 192.02, indicating a significantly worse performance compared to the baseline. The projected outcomes are depicted in Figure 5.

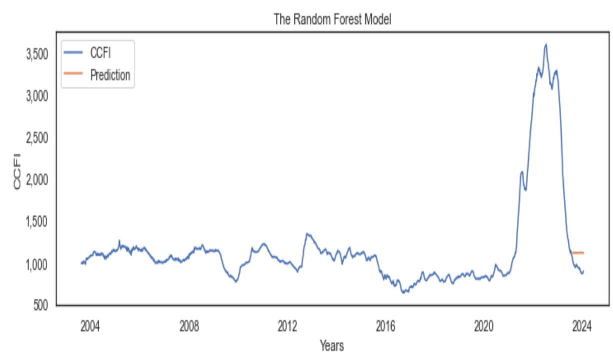

Figure 5. Forecasting Results of the Random Forest Regressor

The LightGBM model was evaluated using 500 estimators and default settings in the subsequent phase. The model yielded an RMSE value of 2,296.88, as seen in the graphic representation below.

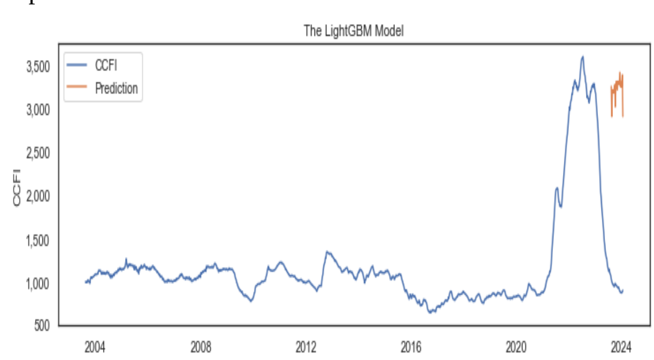

Figure 6. Forecasting Results of the LightGBM Model

The RMSE of the XGBoost model demonstrated superior performance compared to the baseline, with a value of 53.43 as opposed to 137.34. Similarly, the model utilized the default parameters, consisting of 500 trees. Figure 7 illustrates the outcomes of the forecasting analysis.

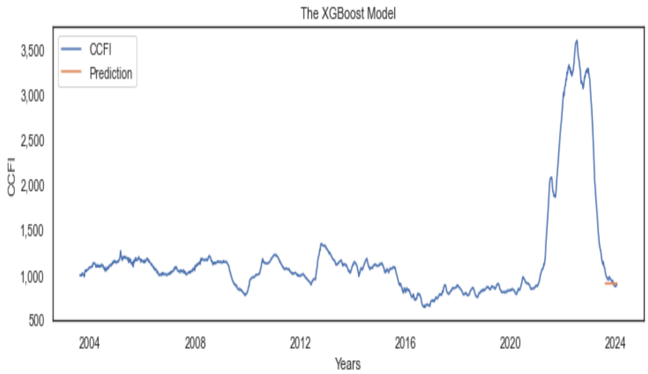

Figure 7. Forecasting Results of the XGBoost Model

The AdaBoost and CatBoost models, utilizing default settings and consisting of 500 trees, did not perform better than the baseline. The reported RMSE values for AdaBoost and CatBoost were 170.99 and 721.31, respectively. The visual examination of the model findings may be conducted by referring to Figures 8 and 9.

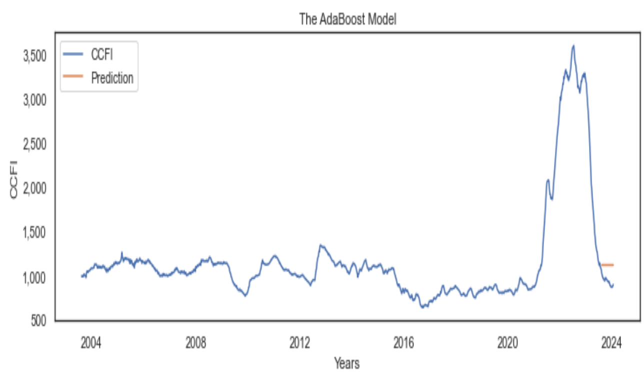

Figure 8. Forecasting Results of the AdaBoost Model

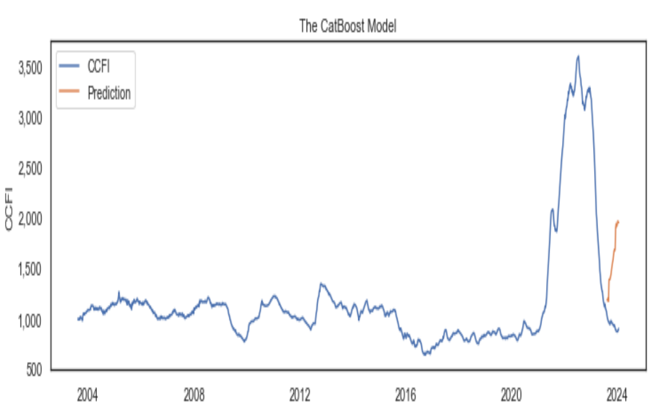

Figure 9. Forecasting Results of the CatBoost Model

Table 1 displays the RMSE values for the examined ensemble time series models. In summary, the XGBoost model exhibited better results than the baseline model, as evidenced by its lower RMSE of 53.43 compared to the baseline model's RMSE of 137.34. Furthermore, the analysis of feature importance, as depicted in Figure 10, indicated that the year and the day of the year exerted the most significant influence on the forecast of the CCFI.

Table 1. Summary of the Times Series Model Results

Method	RMSE
Naive Method	137.34
Random	188.57
LightGBM	2,296.88
XGBoost	53.43
AdaBoost	192.72
CatBoost	721.31

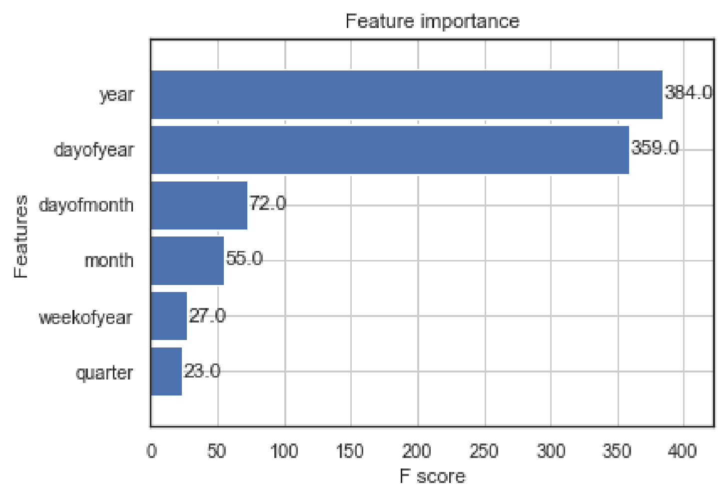

Figure 10. Feature Importances of the XGBoost Model

To make the most accurate forecasting, the entire data set was trained with the XGBoost method, which has the lowest RMSE, and the 24-week results are given in Figure 11. Considering the trend of the forecast results in Figure 11, it can be seen that freight rate increases will occur due to the increase in demand before Christmas and New Year.

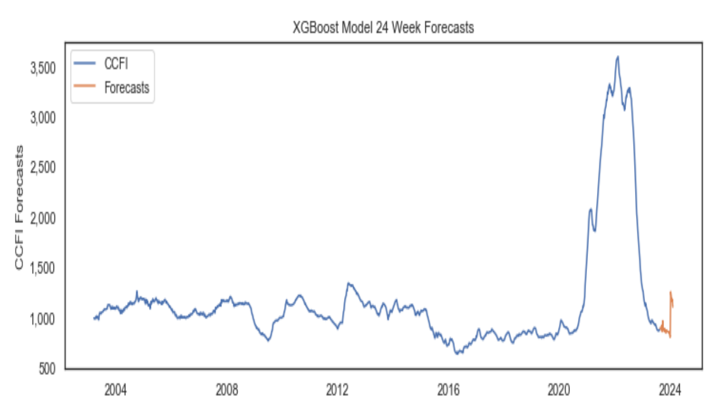

Figure 11. 24-Week Future Forecasts of the XGBoost Model

5. Conclusion

Container shipping has been one of the most important keystones of globalization. In container shipping, which is a mirror of world trade, it is crucial to recognize that freight rates are subject to the influence of many factors. The interaction between these elements can lead to complex and perhaps unpredictable price fluctuations. Liner carriers, shippers, and freight forwarders diligently observe these elements to make well-informed judgments and develop effective supply chain and shipping operations plans. In this regard, it is vital to estimate freight index trends correctly. Shipping indexes serve as a standardized and dependable mechanism for monitoring and assessing freight prices, market dynamics, and other pertinent information within the maritime sector. Stakeholders are provided with the means to enhance their decision-making capabilities, effectively mitigate risks, and contribute to the overall efficacy and transparency of the shipping market.

Various models have been used in the literature to forecast freight rates and indexes. In these studies, traditional time series methods were generally used. This study employed ensemble methods never used before in forecasting container freight rates or indexes. It aimed to forecast composite CCFI. According to the results, XGBoost emerged as the model with the least error. Based on the projections generated by this model, freight rates are anticipated to exhibit an upward trend by the end of 2023. Consequently, the models can be used as an essential decision support tool for both carriers, especially shippers and shipowners who rent their ships in case of fluctuations in the container transportation sector. All these parties in container transportation can estimate more accurate freight rates by looking at this index.

In future studies, forecasting can be made for individual indexes on the major trade lanes where CCFI is used. Moreover, other performance criteria for forecasting can be taken into account. Machine learning applications in the maritime industry are a remarkably untouched field. Therefore, they can be used effectively in freight index forecasting and many areas of the maritime industry.

References

Chen, Y., Liu, B., Wang, T. (2021). Analysing and forecasting China containerized freight index with a hybrid decomposition–ensemble method based on EMD, grey wave and ARMA, Grey Systems: Theory and Application, 11(3), 358-371.

Clarksons Research (2023). Shipping Intelligence Network, (Erişim:14.07.2023), Retrieved from https://sin.clarksons.net/

Deng, Y., and Yang, J. (2021). Research on the Co-integration Relationship between China Coastal Bulk Freight Index and the Freight Rate of Sample Route, 3rd International Academic Exchange Conference on Science and Technology Innovation (IAECST), Guangzhou, China, pp. 1373-1377.

Fan, L. and Yin, J. (2015). Analysis of structural changes in container shipping, Maritime Economics & Logistics, 18(2), 174–191.

Fei, Y., and Zhou Y. (2023). Intelligent Prediction Model of Shanghai Composite Index Based on Technical Indicators and Big Data Analysis, Highlights in Business, Economics and Management, 17.

Ghareeb, A. (2023). Time Series Forecasting of Stock Price for Maritime Shipping Company in COVID-19 Period Using Multi-Step Long Short-Term Memory (LSTM) Networks, Scieondo, 17, 1728 – 1747.

Hirata, E., Matsuda, T. (2022). Forecasting Shanghai Container Freight Index: A Deep-Learning-Based Model Experiment", Journal of Marine Science and Engineering, 10 (593).

ICS, Institute of Chartered Shipbrokers (2015). Liner Trades, London: Institute of Chartered Shipbrokers.

Jeon, J., Duru, O., Yeo, G. (2019). Modeling cyclic container freight index using system dynamics, Maritime Policy & Management, 47(3), 287–303.

Koyuncu, K., Tavacıoğlu, L., Gökmen, N., Arıcan, U.Ç. (2021). Forecasting COVID-19 impact on RWI/ISL container throughput index by using SARIMA models, Maritime Policy and Management, 48(8), 1–13.

Kunapuli, G. (2023). Ensemble Methods for Machine Learning. Manning Publications.

Luo, M., Fan, L., Liu, L. (2009). An econometric analysis for container shipping market, Maritime Policy and Management, 36(6), 507–523.

Munim, Z.H., and H.J. Schramm. (2017). Forecasting container shipping freight rates for the Far East Northern Europe trade lane. Maritime Economics & Logistics 19 (1), 106–125.

Munim, Z.H. and Schram, H.J. (2021). Forecasting container freight rates for major trade routes: a comparison of artificial neural networks and conventional models, Maritime Economics & Logistics, 23, 310–327.

Nielsen, P., L. Jiang, N.G.M. Rytter, and G. Chen. (2014). An investigation of forecast horizon and observation ft's influence on an econometric rate forecast model in the liner shipping industry, Maritime Policy & Management 41 (7), 667–682.

Notteboom, T. (2012). Container Shipping. in The Blackwell Companion to Maritime Economics, Wiley-Blackwell, Oxford, UK, 230-262.

Saeed, N., Nguyen, S., Cullinane, K., Gekara, V., Chhetri, P. (2023). Forecasting container freight rates using the Prophet forecasting method, Transport Policy, 133, 86-107.

Schramm, H. J. and Munim, Z.H. (2021). Container freight rate forecasting with improved accuracy by integrating soft facts from practitioners, Research in Transportation Business & Management, 41.

SSE, Shanghai Shipping Exchange. (2023). (Erişim: 15.08.2023), Retrieved from: https://en.sse.net.cn/indices/introduction_ccfi_new.jsp

Stopford, M. (2008). Maritime Economics, 3rd ed. London: Routledge.

Tu, X., Yang, Y., Lin, Y., Ma, S. (2023). Analysis of influencing factors and prediction of China's Containerized Freight Index. Front. Mar. Sci. 10:1245542.

UNCTAD. (2021). Review of Maritime Transport. Geneva.

Yifei, Z., Z. Dali, and T. Yanagita (2018). Container liner freight index based on data from e-booking platforms. Maritime Policy & Management 45 (6): 739–755.