IJEGEO

International Journal of Environment and Geoinformatics 10(3):077-085 (2023)

Research Article

A Model on Charter Rate Prediction in Container Shipping

Tolga Tuzcuoğlu[1,], Hüseyin Gençer[2,*]

[1] Turkish German University, Faculty of Economics and Administrative Sciences, Department of Business, Istanbul, TÜRKİYE
[2] Piri Reis University, Faculty of Economics and Administrative Sciences, Department of Maritime Business Management, Istanbul, TÜRKİYE

* Corresponding author: H. Gençer
E-mail: hgencer@pirireis.edu.tr

Received 17.08.2022
Accepted 06.09.2023

How to cite: Tuzcuoğlu and Gençer (2023). A Model on Charter Rate Prediction in Container Shipping, *International Journal of Environment and Geoinformatics (IJEGEO)*, 10(3): 000-000. doi. 10.30897/ijegeo.1345053

Abstract

The maritime industry has witnessed numerous challenges in recent years after the global pandemic, primarily characterized by sharp fluctuations in the daily charter rates. Considering an unpredictable business environment, this study aims to suggest a financial forecasting model on charter rates, creating added value for the stakeholders of the maritime trade business. The empirical analysis utilized the data from the Clarksons Research Portal, which encompassed 7,409 charter chartering transactions of container ships from 01.01.2018 to 10.03.2023. After examining seven different linear and ensemble regressions, it was revealed that the XGBoost regressor resulted in the least RMSE value of 0.1833 with an R2 of 0.9015. The selected predictors were the TEU, container fixture type, charter time, charter time multiplied by TEU, ship age, laycan year, and laycan month, respectively. In addition to coping with the limitations of linear regression, the model revealed that the laycan years, charter time, and charter time multiplied with TEU were the essential variables in charter rate prediction. As a result, the model developed in the study can be used as an important decision support tool for stakeholders in the container shipping industry.

Keywords: Forecasting, Charter market, Container shipping, Ensemble regression, Ship financing strategy

Introduction

Maritime industry has witnessed many challenges such as financial crisis, pandemic and war risks. As in many areas of the maritime industry, the container shipping is also affected by similar risks and uncertainties. Container shipping has profound implications for the maritime industry, as more than 50% of world trade in value is handled by container shipping (UNCTAD, 2021). It is one of the most important catalysts of globalization thanks to the safe, reliable, fast handling and intermodal transportation opportunities it provides (Ma, 2021).

One of the main challenges in container shipping is the sharp fluctuations in daily charter rates. Fluctuations in charter rates have a direct impact on the price of shipping products, which impacts companies' expenses for import and export. On the other hand, charter rates directly affect the profitability of ship owners and operators. Figure 1 depicts that the average daily charter rates have been highly volatile since the end of 2020, leading to sudden and unforeseen fluctuations. Container shipping is a highly competitive industry with low profit margins. Accurate forecasting allow companies to adjust their pricing strategies and are important in estimating their profits. Moreover, accurate forecasting helps companies manage risk against losses and industry fluctuations (Petersen, 2016). For that purpose, this study proposes a forecasting model on charter rates, creating added value for the stakeholders of the container shipping industry.

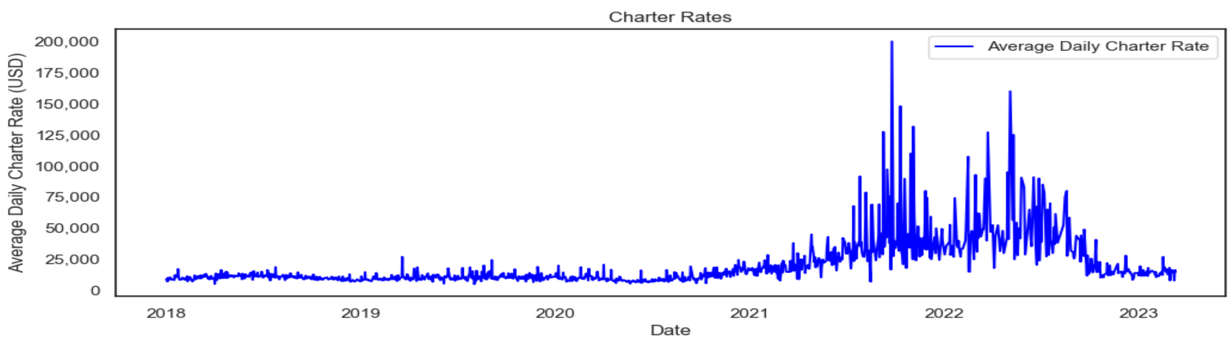


Fig. 1. Average Daily Charter Rates between 2018 and 2023

The charter market in maritime transport is divided into two as voyage charter and time charter. Voyage charter, which is common especially in dry bulk cargo and tanker transportation, is the charter of a ship for the transportation of a certain cargo between two ports. A time charter is the charter of ships for a certain period of time (from a few months to several years) and is very common in container shipping. A bareboat charter is a variant of a time charter where the charterer undertakes the crew and maintenance of the ship. Bareboat charters are frequently utilized when the owner is a financial investor uninvolved in shipping operations (Hübner, 2016).

Chartering of container ships has become even more important due to the start of container liner shipping in the 1970s and its widespread use, especially after the 1990s (Otani and Matsuda, 2023) Time charter markets ultimately reflect the balance of supply and demand for shipping on a global scale. On the other hand, while the trends in freight rates are similar to charter rates, there are minor differences. If too many new ships are built, liner carriers may not charter all of these ships, so if demand increases slightly, freight rates may remain stable, even if chartering rates fall (Lemper). Container ship charter rates remained high from the late 1980s to the mid-1990s. Subsequently, there was an oversupply due to the boom in new shipbuilding and the effects of the 1997 Asian financial crisis on trade and shipping. Because of this, charter rates fell to very low levels in 1999. Afterwards, the rates reached quite high levels in the mid-2000s (Otani and Matsuda, 2023).

The charter rates, which fell sharply because of the financial crisis in 2008, and remained at similar levels until 2020. Due to the halt of production at the beginning of the Covid-19 pandemic and then its resumption, charter rates reached an all-time high. After the pandemic, charter rates decreased sharply as a result of the decrease in demand in world trade.

Container shipping is mainly applicable for freight markets with parcel shipment sizes. Whereas the charter market pertains to the overall transport capacity of the entire ship, the freight market focuses solely on parcels smaller than a complete vessel (Ma, 2021). Most of the forecasting studies in container shipping are on the freight market.

Luo et al. (2009) investigated the fluctuation in freight rates due to supply and demand in container shipping. In the study, the relationship between ship new order and time charter was also examined. Fan and Yin (2015) examined the dynamic interactions between container ship newbuilding and second-hand prices, and time charter rate. The authors also discussed the impact of different sized container ships such as feeder, handymax, sub-panamax and panamax. Jeon et al. (2019) analyzed the CCFI (China Containerized Freight Index) cycles using the system dynamics approach. Their study can be used as an important guideline for decision makers in ship investment time. Chen et al. (2021) proposed a combined approach consisting of empirical mode decomposition and grey wave methods to predict CCFI. Koyuncu et al. (2021) proposed SARIMA and an Exponential Smoothing State Space models to forecast container throughput index of RWI/ISL considering the impact of COVID-19. Schramm and Munim (2021) introduced an autoregressive integrated moving average (ARIMA) model to forecast China Containerized Freight Index (CCFI) and later on the Shanghai Containerized Freight Index (SCFI). Munim (2022) used SARIMA (Seasonal Autoregressive Integrated Moving Average), SNNAR (Seasonal Neural Network Autoregression) and the state-space TBATS models to forecast CCFI. Hirata and Matsuda (2022) developed a long short-term memory (LSTM) model and a seasonal autoregressive integrated moving average (SARIMA) model to forecast the Shanghai Containerized Freight Index (SCFI). Saeed et al. (2023) analyzed the data between 2010 and 2020 using machine learning and natural language processing, and determined six important factors in container shipping such as congestion, peak demand, policy, price up, overcapacity, and coronavirus. The authors used the prophet method to forecast the container freight rates on the six main container routes, considering these six factors.

As can be seen, no research in the literature has considered a model by taking into accountdry bulk the variables in this study. The second part of the study describes the data and methodology used in the study. In the third part, empirical findings are presented. The fourth part concludes the study with the discussions.

Table 1. Descriptive Statistics

	Count	Mean	Std	Min	25%	50%	75%	Max
Built year	7,409	2007	4.9	1989	2005	2007	2010	2024
TEU	7,409	2,671.9	1,945.3	340.0	1,221.0	1,970.0	3,450.0	14,952.0
Charter Rate	7,409	14,851.8	15,785.3	820.0	8,000.0	10,250.0	15,000.0	235,000.0
Charter Time	7,409	324.4	356.0	5.0	150.0	210.0	360.0	5,400.0
Transaction Year	7,409	2020	1.3	2018	2018	2020	2020	2023
Transaction Month	7,409	6.1	3.5	1.0	3.0	6.0	9.0	12.0
Laycan Year	7,409	2020	1.3	2018	2018	2020	2021	2024
Laycan Month	7,409	6.4	3.4	1.0	3.0	6.0	10.0	12.0
Ship Age	7,409	12.5	5.0	-3.0	10.0	12.0	15.0	32.0

Data and Methods

The data used in this study was acquired from Clarksons Research Portal, a prominent worldwide data supplier in the shipping business. The dataset extracted from the research portal encompassed 8,231 container ship charter rate transactions of diverse charterers between 01.01.2018 and 10.03.2023. To be more precise, the observations included ten features, which were the transaction date, name of the container ship, built year, twenty-foot equivalent unit (TEU), charterer name, starting laycan date, charter time, daily charter rate in United States Dollar (USD), ship owner and the type of the container fixture, respectively. During the exploratory data analysis phase, the dataset was refined to 7,409 observations due to missing charter rate values. The refined dataset consisted of 227 unique charterers, 461 diverse owners, and 2,468 different container ships. As for container fixture types, there were six classes; Feeders 100-999 TEU, Feeders 1-1,999 TEU, Feeders 2-2,999 TEU, Narrow Beams 3,000+ TEU, Wide Beams 3-5,999 TEU and Containerships 6,000+ TEU.

The highlights of the extracted dataset are further described in Table 1. Accordingly, the built year of the container ships ranged from 1989 to 2024, which was, on average, 2007. Similarly, the average ship age at the time of the conclusion of the chartering agreement was 12.5 years, which was between -3 and 32 years. Negative ship age implied that the transaction covered a future laycan date upon constructing a newly built vessel. Overall, out of 7,409 transactions, 5,840 of them (78.8%) included a future laycan date at the time of the conclusion of the contract; 1,009 instances (13.6%) covered a retrospective laycan date, which can be thought of as a renewal or extension of an existing agreement; and 560 of the observations (7.6%) had the same transaction and laycan date. In addition, the TEU of the vessels ranged from 340.0 to 14,952.0 TEU with a mean value of 2,671.9, all in line with the defined thresholds of the fixture types. As for the charter time, the acquired data, in most instances, illustrated a period with lower and upper values instead of a single discrete value. In such cases, the upper boundary was taken as a single charter time, expressed in days. Hence, the calculated charter days were between 5 and 5,400, with a median of 210 days. The transaction and laycan dates in the observations were further expressed in years and months, giving rise to additional four features. The transaction years were between 2018 and 2023, as expected, with the date dimension of the acquired dataset. On the contrary, the laycan year additionally covered a future time frame in 2024, indicating the conclusion of prospective contracts.

The charter rate ranged from 820 to 235,000 USD daily, in which the mean and median values were 14,851.8 and 10,250.0 USD, respectively. A higher mean value than the median implied a skewed distribution, plotted in Figure 2.

As it may be inferred from the first plot in Figure 2, the distribution of the charters rates, which is the target variable, had a right-skewed distribution. A log transformation was carried out to bring this to a normal distribution, depicted in the second plot above.

Following the transformation of the target variable, the exploratory data analysis continued investigating the association between the variables. In this sense, the correlation between the numerical variables, the TEU, charter time, ship age, and the target charter rate, was examined with Pearson correlation. The correlation matrix presented in Figure 3 reported that the TEU and charter time were positively and weakly correlated with the charter rate at 0.27 and 0.31, respectively. In addition, the analysis revealed that the ship age was not associated with the charter rate. As for the association between the predictors, the TEU and the charter time were again positively and weakly correlated with 0.32, in which the ship age showed no association with the rest of the numerical predictors.

The correlation between the categorical variables in the dataset, the laycan year, laycan month, container fixture type, transaction year, and transaction month, was analyzed with the Cramer's V approach. As expected, the laycan month and the transaction month, along with the laycan year and the transaction year, resulted in a high positive association with each other with 0.61 and 0.92, respectively. The rest of the categorical variables did not show any significant association. The results of the correlation matrix between the categorical variables are summarized in Figure 4. As for the next step, the focus was on the effect of container fixture types on charter rates. As expected, it was determined that as the container ship's capacity

increased, the range of the daily charter rate also rose. The median value of the charter rate was 6,600 USD for feeders 100-999 TEU, 9,000 USD for feeders 1-1,999 TEU, 10,250 USD for feeders 2-2,999 TEU, 11,350 USD for narrow beams 3,000+ TEU, 16,000 USD for wide beams 3-5,999 TEU, and 22,000 USD for containerships 6,000+ TEU.
To conduct an eyeball test, the boxplot of charter rates by container fixture types was plotted, depicted in Figure 5. In addition, to further analyze whether container fixture types significantly affect daily charter rates, an ANOVA test was executed. The reported p-value was almost zero, concluding that at least one mean value differed significantly from the rest. Tukey's test was additionally conducted to carry out multiple comparisons of container fixture type means. Tukey's HSD results indicated that except for the wide beams 3-5,999 TEU and containerships 6,000+ TEU, all paired mean values significantly differed from each other at FWER=0.05. Using the same method of eyeball tests through boxplots, ANOVA, and Tukey's test for multiple comparisons of means, charter rates were examined against laycan years and laycan months. The boxplot of charter rates by laycan years, illustrated in Figure 6, provided an impression of differing mean charter rate values in the laycan years of 2021 to 2024. The ANOVA test, with a p-value of 0.06, concluded no significant differences among the means. However, Turkey's test at FWER=0.05 level rejected no significant differences except for the pairs 2018-2019, 2018-2020, 2019-2020, and 2022-2024. The ANOVA test for laycan months yielded a p-value of 0.3 and reported no significant differences among the means. The following Tukey's test stated significant mean differences mainly for May and August, with most of the rest. The boxplot of charter rates by laycan months is shown in Figure 7.

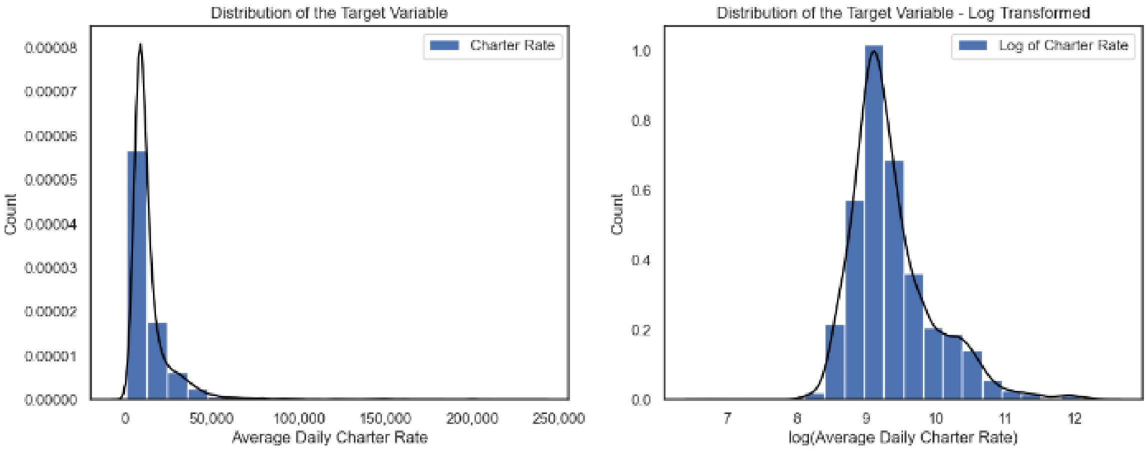


Fig. 2. The Distribution of Charter Rates: Nominal and Log Transformed

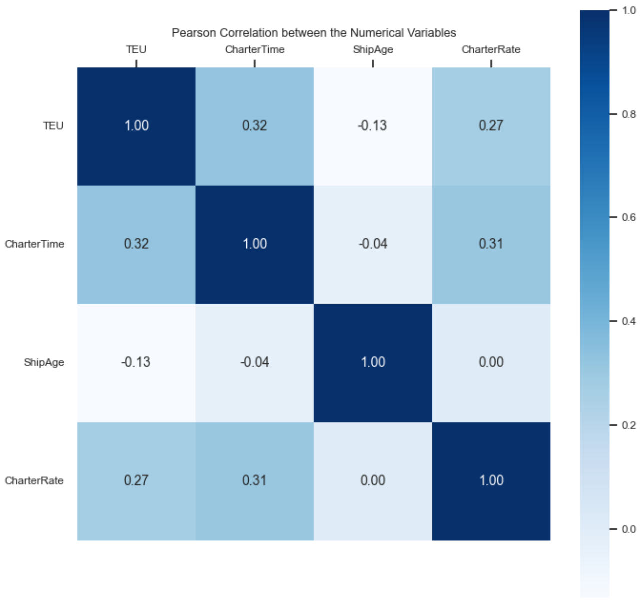


Fig. 3. Correlation Matrix of Numerical Variables

Fig. 4. Correlation Matrix of Categorical Variables

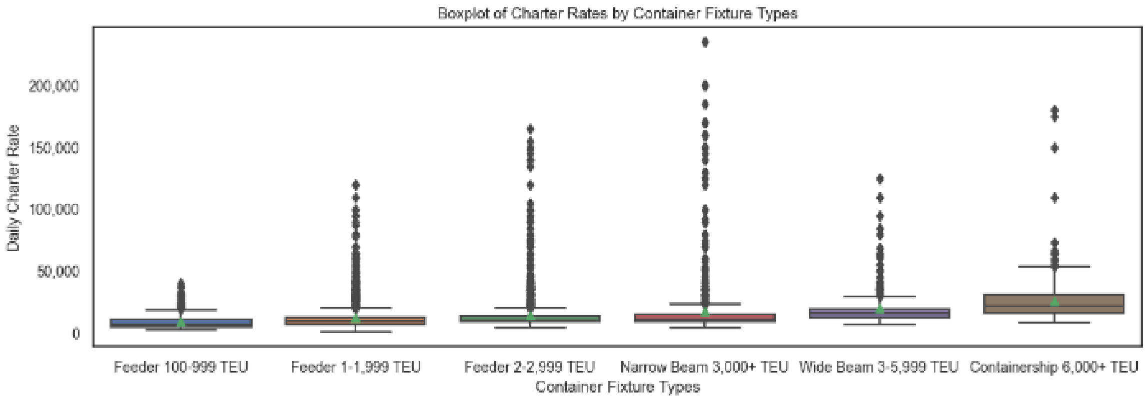


Fig. 5. Boxplot of Charter Rates by Container Fixture Types

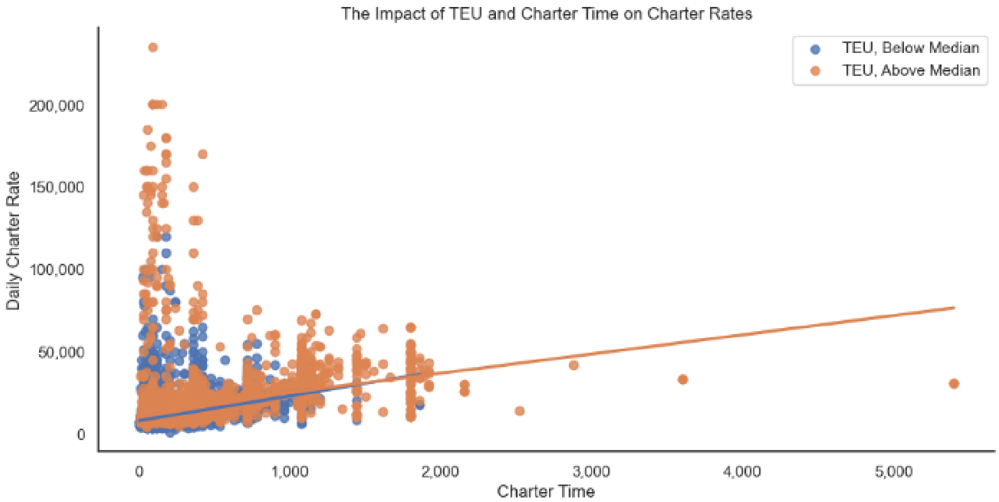


Fig.6. Boxplot of Charter Rates by Laycan Years

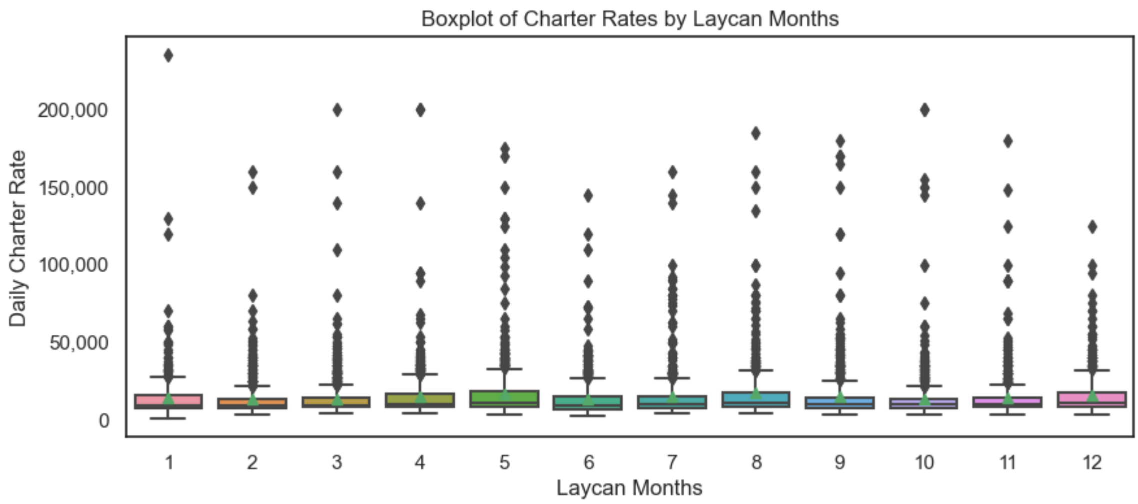


Fig. 7. Boxplot of Charter Rates by Laycan Months

The last step of the exploratory data analysis investigated whether selected two predictors mutually affect the target variable. One predictor was divided into two equal instances by taking the median value as the cut-off point. Subsequently, the target variable was plotted against the other predictor showing the initially paired predictor. To illustrate, the interaction between the TEU and the charter time on charter rates was analyzed by dividing the TEU instances into two subsets by taking the median observation as the cut-off point. Then, charter rates were plotted against the charter time in the interacting pair of below-median and above-median TEU observations, depicted in Figure 8. The crossing lines in the figure implied that the charter time interacted with the TEU to impact the charter rates. The same analysis was conducted for ship age and TEU and ship age and chartering days. However, no crossing lines were detected.

In summary, the exploratory data analysis concluded with the following findings:

- The use of the log-transformed form of the target variable charter rate.
- The primary consideration of the laycan months and laycan years instead of the transaction months and transaction years due to their highly correlated nature, the boxplot analysis of laycan variables as well as Tukey's test results. In addition, since the laycan year was available one year further compared to the transaction year in the dataset, a prospective model with the inclusion of laycan variables would extend the availability of the prediction horizon for more than one year. In other words, if the transaction month and year were included in the model, the prediction horizon of such a model would be limited to the most recent available transaction date, which was 10.03.2023 in the dataset.
- The addition of container fixture types as a categorical variable in the model since they impacted charter rates.
- The inclusion of the multiplication of TEU with charter time due to their combined effect on charter rates.
- It was decided to consider the ship age in the model selection phase along with the abovementioned predictors.

Accordingly, the variables which were considered in the regression model development phase to predict the charter rate were the TEU, charter time, charter time multiplied with the TEU, ship age, container fixture type, laycan year, and the laycan month, in which the latter three were categorical variables. After defining the variables, the dataset was divided into two sub-datasets to train (80%) and test (20%) diverse regression models. The model evaluation phase included a five-fold cross-validation to examine model performance in which the numerical variables were standardized, and the categorical ones were one-hot encoded. Seven different regression models were trained in the empirical analysis, and their performance was compared with the RMSE metric. These models were the linear regression, Ridge regressor, Lasso regressor, stochastic gradient descent (SGD) regressor, random forest regressor, extreme gradient-boosting (XGBoost) regressor, and light gradient-boosting machine (LightGBM) regressor, respectively.

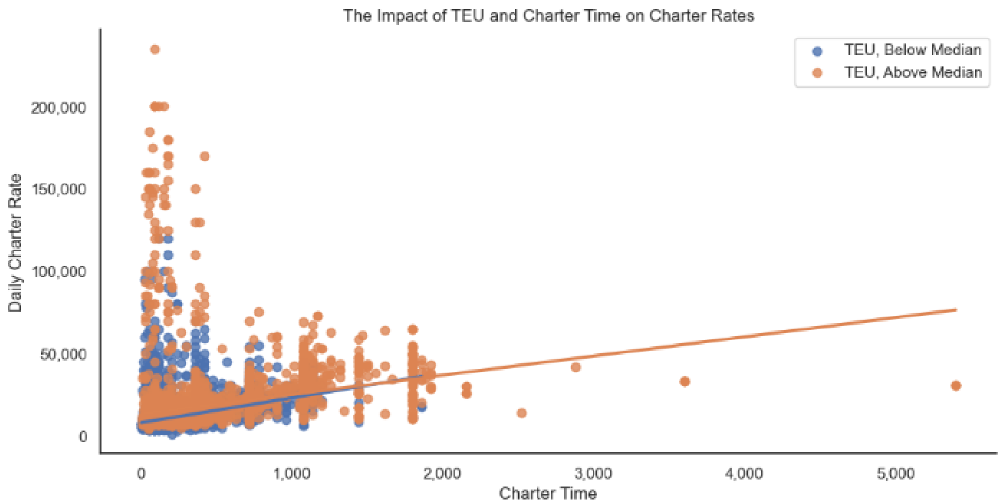


Fig. 8. The Impact of TEU on Charter Rates and Charter Time

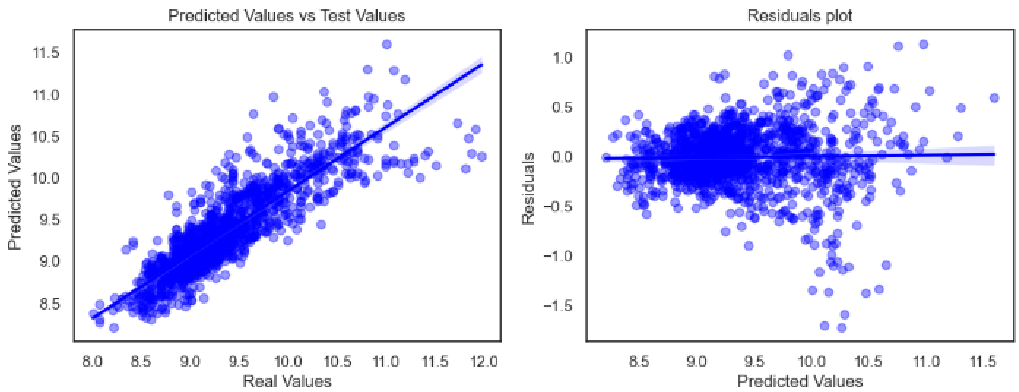


Fig. 9. The Plots of Test Values and Residuals vs. Predicted Values

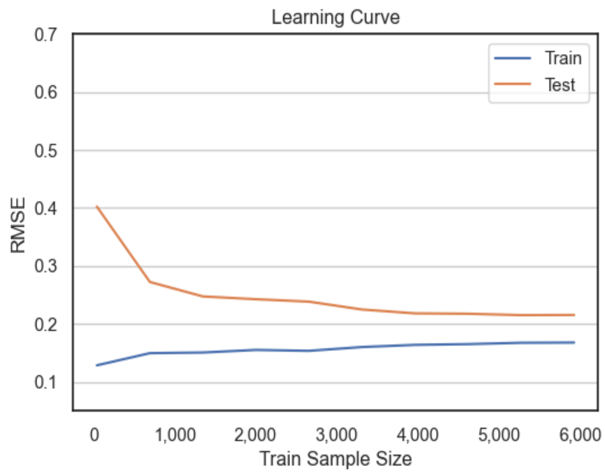
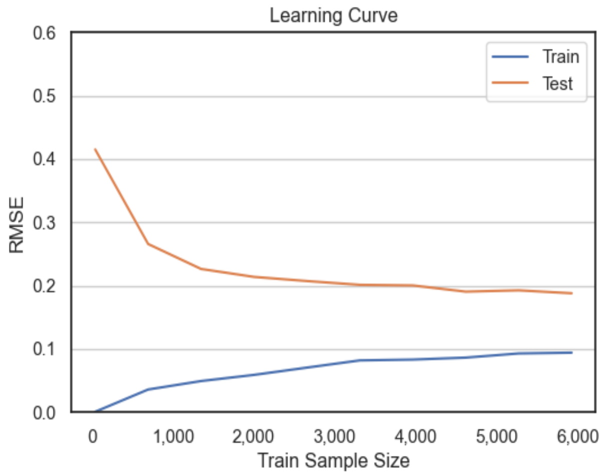


Fig. 10. The Learning Curve of the XGBoost Regressor



Emprical Results

The empirical analysis started with implementing linear regression, considered the baseline model. The conventional assumptions of linear regression are fivefold; independence (the prediction errors are independent of each other), homoscedasticity (constant variance of the errors), collinearity (predictors are not perfectly associated with each other), normality (the errors have a normal distribution with zero mean value), and linearity (the target variable and the predictors do not have a curved relationship) (Hoffman, 2022). In practice, if the primary objective of linear regression is to predict the target variable, then most of the assumptions stated above will not be relevant; nevertheless, such a case will lead to deriving misleading statistical interpretation (Matloff, 2017).

The linear regression model is presented below in the Equation .

$$\mathrm{Log(Target)} \sim \mathrm{TEU} + \mathrm{C(ContainerFixtureType)} + \mathrm{CharterTime} + \mathrm{CharterTime} * \mathrm{TEU} + \mathrm{C(LaycanYear)} + \mathrm{C(LaycanMonth)} + \mathrm{ShipAge}$$

Where

Log(Target): Daily Charter Rate, log-transformed

C(): Categorical variable

The zero mean of the residuals and homoscedasticity were first analyzed to examine whether the model fulfilled the assumptions of linear regression. Whereas the mean value of the residuals in the test dataset was 0.0073, almost close to zero, the p-value of the Breusch-Pagan test was zero, leading to the rejection of the hypothesis that the variance of the residuals was constant. The cone-shape of the residuals plot presented on the right side of Figure 9 also confirmed the test finding. Secondly, the normality of the residuals was examined by conducting Jarque-Berra and Kolmogorov-Smirnov tests.

The p-values of both tests were almost zero, which led to the rejection of the null hypothesis that the residuals are normally distributed. Thirdly, the Durbin-Watson test was conducted to report whether there was no autocorrelation in the residuals. The test statistic was 1.9532, which could be accepted as relatively usual since falling in the threshold from 1.5 to 2.5. The last step was to check whether there was no perfect multicollinearity between the predictors by calculating the VIF values. For almost all predictors, the VIF values were infinite, revealing solid collinearity between the predictors, hence leading to misinterpretation of coefficients. The linear regression model did not satisfy the residuals' homoscedasticity and normality and the independent variables' non-multicollinearity. Despite this, the calculated RMSE value of the test data was 0.2898, in which the R2 was 0.7538. The regression model was re-trained using regularization via Ridge and Lasso in the next step. Regularization penalizes one of the highly collinear predictors by pushing them to zero to handle the multicollinearity problem; hence predictors are subject to shrinking and getting too large values (Matloff, 2017). The penalty factor (alpha) was 2.9763 in the Ridge regression, in which the model picked 29 independent variables without eliminating any of them. The reported test RMSE value was 0.2893 with an R2 of 0.7546. The Lasso regression revealed equivalent results; the alpha was this time close to zero, in which 27 variables were picked up, and the other two were eliminated. The Lasso test RMSE and R2 values were 0.2892 and 0.7548, respectively. As for the last step of the linear regression approach, a Stochastic Gradient Boosting (SGD) regression model was constructed. With the selected hyperparameters of alpha: 0.001, epsilon: 0.001, eta: 0.1, loss: squared error, and the penalty factor: no, the SGD regressor revealed an RMSE of 0.2970 and R2 of 0.7415 in the test dataset. In conclusion, the application of regularization and a stochastic method showed no improvement in the RMSE value.

Given that the assumptions of linear regression are not entirely fulfilled, the second phase of the empirical analysis focused on non-parametric regression techniques, including tree-based regressors. Ensemble regression models offer better prediction accuracy, robustly handle outliers, and unveil complex relationships in the data; nevertheless, they are inclined to overfit (Kunapuli, 2023). The analysis started with random forest regression. The hyperparameters determined as a result of the cross-validation were bootstrap: true, max depth: 10, max features: auto, min samples leaf: 1, min samples split: 2, and the number of estimators: 50. The model disclosed an RMSE value of 0.2154 with an R2 of 0.8640 in the test dataset, quite outperforming the linear regression models. As for the top five most important features, the model reported the CharterTime_TEU, TEU, LaycanYear 2022, LaycanYear 2021, and CharterTime, respectively. On the other hand, the RMSE and R2 values in the training dataset were 0.1678 and 0.9212, which implied overfitting. Overfitting was further checked with the learning curve depicted in Figure 10. As it may be inferred from this chart, the test and train RMSE lines became parallel as the sample size increased.

As for the second tree-based model, the Extreme Gradient Boosting (XGBoost) regression was implemented. The model considered the hyperparameters as a result of the cross-validation as follows; colsample by level: 0.8, colsample by tree: 0.8, learning rate: 0.1, max depth: 9, number of estimators: 75, and subsample: 0.8. The test RMSE value of the model was 0.1833, and the R2 was 0.9015, which reported better results compared to the random forest regression. The top-five feature rankings stated the laycan years 2022, 2021, and 2023 as well as CharterTime and CharterTime_TEU. Besides, the RMSE and R2 in the training dataset were 0.1151 and 0.9629, which signaled to overfit. The learning curve of the model presented in Figure 11 plotted train and test RMSE lines that were not wholly parallel but approached each other as the sample size increased. It led to the impression that had we more observations, the test RMSE line would come closer to the train RMSE line.

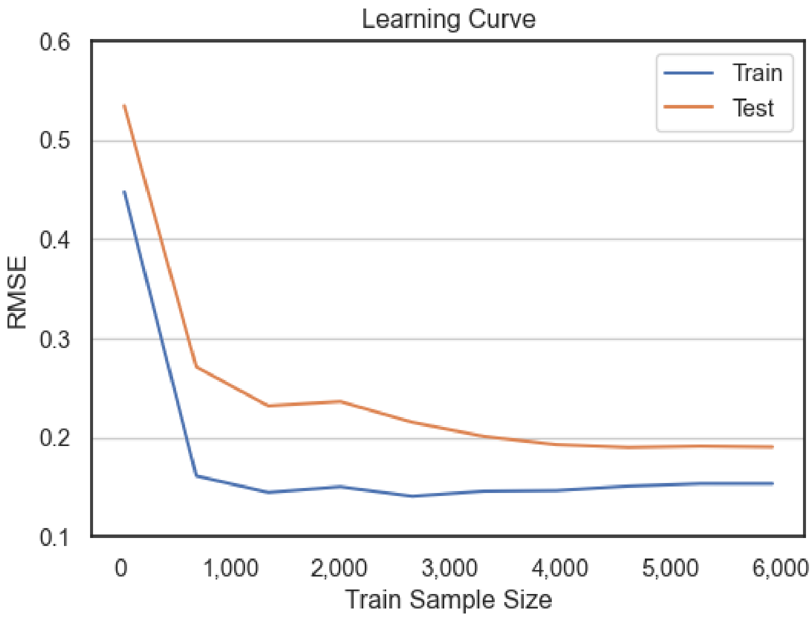


Fig. 11. The Learning Curve of the LightGBM Regressor

The third and last tree-based model trained was the Light Gradient-Boosting Machine (LightGBM) regression. The tuned hyperparameters in the model were colsample by tree: 0.5, learning rate: 0.2, max depth: 10, metric: l1, min child samples: 10, min child weight: 0.01, number of estimators: 300, number of leaves: 9, alpha: 0, lambda: 0.1, and task: train. The LightGBM disclosed a test RMSE of 0.1972, in which the test R2 was 0.8860, performing better than the random forest but worse than the XGBoost regressor. The RMSE and R2 values in the training dataset were 0.1458 and 0.9405, which implied overfitting again. Overfitting was verified with the parallel RMSE lines in Figure 12. The top-five feature rankings included the TEU, CharterTime_TEU, CharterTime, ShipAge, and the laycan year 2021.

The result of the empirical analysis is summarized in Table 2. The best-performing model of the trained regression models was the XGBoost regressor with the least RMSE value.

Table 2. Summary of the Regression Model Results

Regressor	RMSE	R-squared
Linear Regression	0.2898	0.7538
Ridge	0.2893	0.7546
Lasso	0.2892	0.7548
SGD	0.2970	0.7415
Random Forest	0.2154	0.8640
XGBoost	0.1833	0.9015
LightGBM	0.1972	0.8860

Discussion and Conclusion

Container shipping is one of the most important components of international trade. Charter rates play a significant role when making decisions and taking action in container shipping. However, it is quite a challenge to have a clear opinion on how the rates will develop in the future, especially in today's turbulent economic environment. This study strived to propose a model for predicting charter rates to create added value for the decision-makers and stakeholders in the maritime business. The predictors considered in the analysis were the TEU, charter time, charter time multiplied by the TEU, ship age, container fixture type, laycan year, and the laycan month. The analysis first focused on the traditional linear regression model and strived for its enhancement by applying regularization and stochastic methods. In this sense, the RMSE values of the linear regression, ridge, lasso, and SGD regressor were 0.2898, 0.2893, 0.2892, and 0.2970, respectively. However, limitations arise due to the fulfillment of the assumptions of linear regression. To cope with this challenge, tree-based models were further analyzed by constructing models with random forest, XGBoost, and LightGBM algorithms. In addition to offering flexibility for the assumptions of linear regression, the tree-based models reported improved RMSE values of 0.2154 (random forest), 0.1833 (XGBoost), and 0.1972 (LightGBM). Hence, out of the trained models, the least RMSE was derived from the XGBoost regressor, where the R2 was 0.9015. Besides, the overfitting problem was comparatively less evident in the XGBoost regression. It was also disclosed that the TEU, charter time, ship age, and laycan year were the industry's most significant determinants of charter rates. In sum, the study highlighted that the XGBoost regression could be adapted as a method for charter rate prediction with its outstanding prediction accuracy and flexibility.

As for the study's limitations, it would be worth mentioning two main aspects. Firstly, the limited number of observations could not be verified how the overfitting trend would look as the training data volume increased. Secondly, the model employed the laycan year and the laycan month as predictors. Accordingly, the prediction charter rate is limited to the extent that these two variables are available for the requested time frame. Taking these limitations as a basis for future research, the volume of observations could be increased to analyze possible overfitting situations in tree-based regressors. In addition, further research can focus on LSTM (long short-term memory) and neural network-based models such as CNN (convolutional neural network).

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.

Hoffman, J. P. (2022). *Linear regression models: Applications* in R. CRC Press.

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

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

Hoffman, J. P. (2022). *Linear regression models: Applications* in R. CRC Press.

Hübner, J.H. (2016). Shipping Markets and Their Economic Drivers, In: Kavussanos, M. G., Visvikis, I. D. (Eds), *The International Handbook of Shipping Finance* : 6-7, Palgrave Macmillan.

Jeon, J., Duru, O., Yeo, G. (2019). Modelling cyclic container freight index using system dynamics, *Maritime Policy and 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*.

Lemper, B. (2008). Volkswirtschaftliche Aspekte der Schifffahrt, In: Winter, H., Hennig, C., Gerhard, M. (Eds), *Grundlagen der Schiffsfinanzierung* (113-162), Frankfurt School Verlag.

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

Ma, S. (2021). *Economics of Maritime Business*, Routledge.

Matloff, N. (2017). Statistical Regression and Classification: From *Linear Models to Machine Learning.* CRC Press.

Munim, Z.H. (2022). State-space TBATS model for container freight rate forecasting with improved accuracy, *Maritime Transport Research*, 3.

Otani, S. and Matsuda, T. (2023). Unified container shipping industry data from 1966: Freight rate, shipping quantity, newbuilding, secondhand, and scrap price, *Transportation Research Part E*, 176.

Petersen, H.B. (2016). Asset Risk Assessment, Analysis and Forecasting in Asset Backed Finance, In: Kavussanos, M. G., Visvikis, I. D. (Eds), *The International Handbook of Shipping Finance*, 53-59, Palgrave Macmillan.

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

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