Forecasting of Dry Freight Index Data by Using Machine Learning Algorithms

Kemal Akyol
Department of Computer Engineering, Kastamonu University, Kastamonu, 37100, Turkey
E-mail: kakyol@kastamonu.edu.tr

Abstract—Discovery of meaningful information from the data and design of an expert system are carried out within the frame of machine learning. Supervised learning is used commonly in practical machine learning. It includes basically two stages: a) the training data are sent to as input to the classifier algorithms, b) the performance of pre-learned algorithm is tested on the testing data. And so, knowledge discovery is carried out through the data. In this study, the analysis of Lloyd data is performed by utilizing Gradient Boosted Trees and Multi-Layer Perceptron learning algorithms. Lloyd data consist of the Baltic Dry Index, Capesize Index, Panamax Index and Supramax Index values, updated daily. Accurate prediction of these data is very important in order to eliminate the risks of commercial organization. Eight datasets from Lloyd data are obtained within the frame of two scenarios: a) the last three index values in the freight index datasets; b) the last three index values in both crude oil price and freight index datasets. The results show that the models designed with Gradient Boosted Trees and Multi-Layer Perceptron algorithms are successful for Lloyd data prediction and so proved their applicability.

Index Terms—Crude oil price, freight index data, machine learning, Gradient Boosted Trees, Multi-Layer Perceptron.

I. Introduction

Maritime transportation plays an important role in the global trade system. Mari-time statistics open a window to global economic actions because overseas trading is too much. Lloyd data presents the shipping prices of bulk cargoes that is generated from daily reported fixtures or estimations of the Baltic Exchange. It is generally hard to forecast these indexes because they are volatile, complex, and cyclic [1]. The prediction of the trend of dry bulk market becomes difficult since the affecting factors of price of dry bulk market are complexities [2].

Lloyd's data includes of the world's largest dedicated marine and energy data. This database consists of the Baltic Dry Index (BDI), Baltic Capesize Index (BCI), Baltic Panamax Index (BPI), and Baltic Supramax Index (BSI) values, updated daily. These indexes project the volatility of dry bulk shipping market and demonstrate the status of the global economy and trend of international trade. They are called as the "barometer" of dry bulk shipping market. Many investors and experts have tried to forecast the future trend of dry bulk shipping market. These indexes are useful for them in order to set their investment portfolio and management risk strategy [3]. BDI, a powerful tool for shipping industry, has been the average of BCI, BPI and BSI since 2006. It reflects bulk shipping market sentiment as a price reference for transaction platform of bulk shipping companies and investors [4].

The aim of this study is to investigate the performance of machine learning algorithms and find the most successful model for analysis of freight index data. In this con-text, the analysis of Lloyd data is performed by utilizing Gradient Boosted Trees (GBT) and Multi-Layer Perceptron (MLP) learning algorithms which are introduced in [5,6] respectively. These algorithms are suitable for nonlinear time-series prediction. By sending the training data as input to these algorithms, supervised machine learning is performed. The performance of pre-learned classifier algorithm is tested on the testing data. And so, knowledge discovery is carried out through the data. The BSI out of all data are collected from March 2009 to January 2018. BSI data is collected from August 2012 to January 2018. Eight datasets from Lloyd data are obtained within the frame of two scenarios: a) the last three index values in the freight index datasets (four datasets); b) the last three index values in both crude oil price and freight index datasets (four datasets).

This paper is organized as follows. Section 2 reviews the studies on the shipping freight market. Section 3 introduces the data collection. Section 4 presents performance evaluation criteria. The designed models are introduced in Section 5. The performances of the models are compared and analysed in Section 6. Finally, the conclusion is presented in Section 7.

II. Related Work

There are many studies on analysis of freight index data. Şahin et al. [1] introduced an artificial neural network approach for BDI forecasting. According to their studies, the ANN is a considerable method for modelling and forecasting of BDI. Han et al. [2] presented the improved support vector machine model which is combined model of wavelet transform and support vector

machine in order to forecast dry bulk freight index. Wong [4] introduced that Autoregressive integrated moving average fits better than Fuzzy heuristic model for the prediction of BDI in their studies. Ming-Tao Chou [7] applied fuzzy time series model to forecast the BDI. According to their studies, the fuzzy time series model is suitable for the BDI's prediction and better than other models. Giannarakis et al. [8] investigated the effect of economic leading indicator of Baltic dry index on stock returns of socially responsible stock index. According to their studies, BDI affect positively Dow Jones sustainability index world. Geman and Smith [9] aimed to describe and explain the key features of shipping markets; and analyse the behaviour of freight rates. Tsioumas et al. [10] aimed to enhance the forecasting accuracy of the BDI by means of developing a multivariate Vector Autoregressive model with exogenous variables. The model incorporates the Chinese steel production, the dry bulk fleet development and a new composite indicator, the Dry Bulk Economic Climate Index. Ruan et al. [11] investigated the cross-correlation properties of BDI and crude oil prices, using cross-correlation statistics test and multifractal detrended cross-correlation analysis. They verified that the multifractality of the cross-correlations of BDI and crude oil prices are both attributable to the persistence of fluctuations of time series and fat-tailed distributions. Baltyn [12] aimed to identify the level of diffusion index changes between United States Gross Domestic Product and the BDI for different periods of time which are 90 days. Zeng et al. [13] proposed Empirical Mode Decomposition for forecasting of BDI. For this, they decomposed the BDI into three distinct components representing short-term changes, long-run trends, and external shocks respectively. Bao et al. [14] presented a new BDI forecasting model based on Support Vector Machine combined with Correlation-based Feature Selection and introduced the macroeconomic fundamental indicators of BDI. Apergis and Payne [15] disclosed the importance of the BDI in predicting the future course of the real economy, yielding a link between financial asset markets and the macro economy. Yang et al. [16] developed a model which is combination of the wavelet transforms and support vector machine in

order to forecast the BPI. Yuwei et al. [17] proposed a model which includes empirical evaluations in order to analyse the leverage effect among the Handysize, Supramax, Panamax and Capesize index in the dry bulk shipping market. As can be seen, there are very studies which were performed, using BDI data, while there are few studies which were carried out, using other indexes. Even though BDI is expressed as the average of these four indexes, the contribution of this study to the literature is conducted by focusing on each of them separately in this study.

III. DATA COLLECTION

Lloyd data which were gathered from Quandl offers the BDI, BCI, BPI and BSI values updated daily. Quandl's data contain various objects including time-series and tables. Researchers can access the premium data through the APIs and various tools. Also, the crude oil price (COP) data are collected from the Federal Reserve Economic Data (West Texas Intermediate (WTI) - Cushing, Oklahoma). The information about the daily index data is given in Table 1. For example, the time-series of BSI consist of data between August 2012 and January 2018. The sample size, weekly average value, is 280 for this dataset.

Table 1. The data range for daily index data.

Datasets	Time Interval	
	Start Date	End Date
BCI	2009-03-16	2018-01-28
BPI	2009-03-16	2018-01-28
BDI	2009-03-16	2018-01-28
BSI	2012-08-21	2018-01-28
COP	2012-08-21	2018-01-28

First, the weekly average values are obtained from the daily index data. This information is presented in Table 2 in detail. For example, there are 456 records for BDI. The 319 prior consecutive records are assigned as the training set, and the last 137 records are assigned as the testing set.

Table 2. Weekly average data for datasets.

Datasets	Start Date – End Date	Total size	Train size	Test size
BPI	Week 13,2009 - Week 4,2018	456 weeks	319 weeks	137 weeks
BPI	Week 13,2009 - Week 4,2018	456 weeks	319 weeks	137 weeks
BDI	Week 13,2009 - Week 4,2018	456 weeks	319 weeks	137 weeks
BSI	Week 35,2012 - Week 4,2018	280 weeks	196 weeks	84 weeks
COP	Week 35,2012 - Week 4,2018	280 weeks	196 weeks	84 weeks

IV. PERFORMANCE EVALUATION

To assess the performance of the learning algorithm, it is determined how well the method produces estimates that match the actual results. Mean Absolute Error (MAE), Root Mean Squared Error (RMSE) and R-squared (R^2) evaluation metrics which are expressed

respectively by the following equations are used in order to assess the prediction accuracy of the models.

$$MAE = \left[n^{-1} \sum_{i=1}^{n} |e_i| \right] \qquad (1)$$

$$RMSE=\left[n^{-1}\sum_{i=1}^{n}\left|e_i\right|^2 \right]^{1/2} \qquad (2)$$

where e_i is defined as individual model prediction error usually ($e_i=P_i–O_i$). n is the number of data samples, P_i and O_i are the predicted and observed values respectively in these equations [18].

$$R^2=1-\frac{\sum_{i=1}^{n}(y-y)^2}{\sum_{i=1}^{n}(y-\bar{y})^2} \qquad (3)$$

where y is the observed response variable, $\bar{y}$ its mean and ŷ is the corresponding predicted values. R^2, commonly used in statistics and also known as the coefficient of determination, is a measure of how well a model fits a dataset. It measures the degree of variation in the target variable; this is explained by the variation in the input features. This coefficient generally takes a value between 0 and 1, where 1 equates to a perfect fit of the model [19].

V. DESIGNED MODELS

This study basically consists of three stages:

1. *Preparation of data for each freight index.*

   a. The datasets which include weekly average Lloyd values as time-series are obtained.
   b. The whole data normalized by using min-max normalization technique in order to avoid the training and prediction error.

2. *Preparation of datasets within the frame of the past observations.*

   a. Past observations in t-2, t-1 and t times.
   b. Past observations included COP data in t-2, t-1 and t times.

3. *The Designed Models for Lloyd Data Prediction*

A. Model 1 (1st approach)

The last two freight index values in t-2 and t-1 times are used as an input parameter. The output parameter is the datum in t time for each freight index.

BDI dataset

Input parameters are BDI(t-2) and BDI(t-1), and output parameter is the BDI(t). The sample size as weekly average data is 456. The portion of the 70% of dataset, prior consecutive data of 319 weeks in between Week 13, 2009 to Week 21, 2015 is reserved for training. The remained portion of dataset, data of 137 weeks in between Week 22, 2015 to Week 4, 2018 is reserved for prediction.

BPI dataset

Input parameters are BPI(t-2) and BPI(t-1), and output parameter is the BPI(t). The sample size as weekly average data is 456. The portion of the 70% of dataset, prior consecutive data of 319 weeks in between Week 13, 2009 to Week 21, 2015 is reserved for training. The remained portion of dataset, data of 137 weeks in between Week 22, 2015 to Week 4, 2018 is reserved for prediction.

BCI dataset

Input parameters are BCI(t-2) and BCI(t-1), and output parameter is the BCI(t). The sample size as weekly average data is 456. The portion of the 70% of dataset, prior consecutive data of 319 weeks in between Week 13, 2009 to Week 21, 2015 is reserved for training. The remained portion of dataset, data of 137 weeks in between Week 22, 2015 to Week 4, 2018 is reserved for prediction.

BSI dataset

Input parameters are BSI(t-2) and BSI(t-1), and output parameter is the BSI(t). The sample size as weekly average data is 280. The portion of the 70% of dataset, prior consecutive data of 196 weeks in between Week 35, 2012 to Week 21, 2016 is reserved for training. The remained portion of dataset, data of 84 weeks in between Week 22, 2016 to Week 4, 2018 is reserved for prediction.

B. Model 2 (2nd approach)

COP data is included to the input parameters dataset. That is, the last two freight index data and COP data in t-2 and t-1 times are selected for input parameter. The output parameter is the datum in t time for each freight index.

BDI_COP dataset

Input parameters are COP(t-2), BDI(t-2), COP(t-1) and BDI(t-1), and output parameter is the BDI(t).

BPI_COP dataset

Input parameters are COP(t-2), BPI(t-2), COP(t-1) and BPI(t-1), and output parameter is the BPI(t).

BCI_COP dataset

Input parameters are COP(t-2), BCI(t-2), COP(t-1) and BCI(t-1), and output parameter is the BCI(t).

BSI_COP dataset

Input parameters are COP(t-2), BSI(t-2), COP(t-1) and BSI(t-1), and output parameter is the BSI(t).

The clauses established for training and prediction information in Model 1 are not included to this section in order to avoid repeated statements.

VI. RESULTS

Training sets are sent to the GBT and MLP one by one. After that, test sets are sent to the pre-learning models. So,

the learning and prediction successes of these algorithms are evaluated. Our proposed models predicted quite close values to the actual values on the datasets. These results indicate that the actual and the predicted values are consistent with each other. Prediction results of the models and actual values of the datasets based on Model 1 approach are plotted in between Fig. 1 and Fig. 4. Besides, pred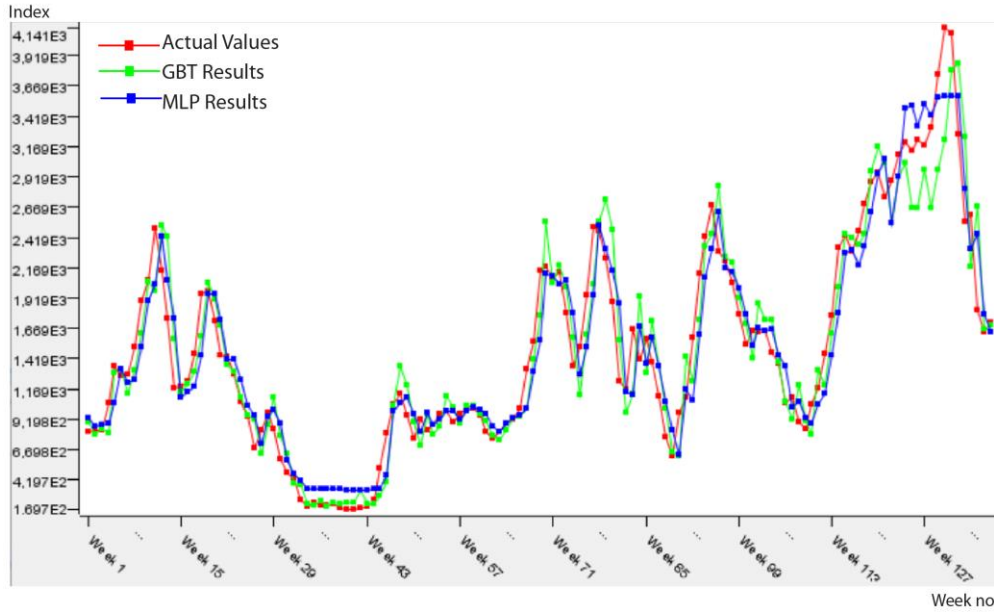iction results of the models and actual values of the datasets based on Model 2 approach are plotted in between Fig. 5 and Fig. 8.

Fig.1. Predicted results on the BCI testing data.

The predicted results approximately same to actual values in Fig. 1 show that performances of the prediction algorithms are close to each other. In addition, Table 3 presents the performance measures of two prediction algorithms within the frame of 1st approach (Model 1). According to this table, the most suitable model for BCI data is MLP because of the minimum error (RMSE: 0.06, MAE: 0.047) and the perfect fit of R^2 value (0.926).
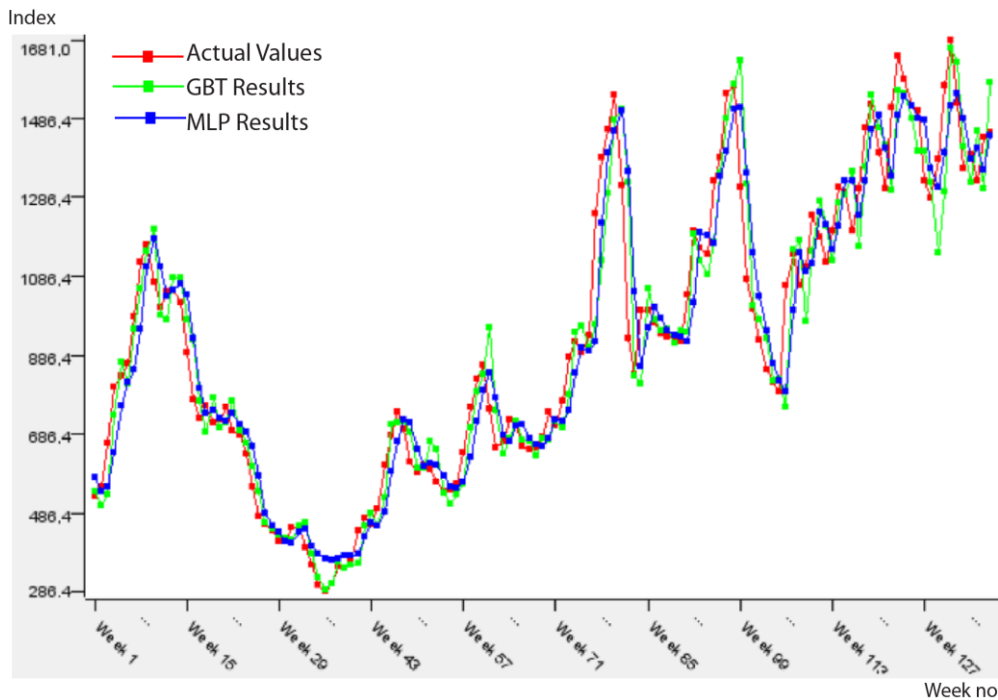

Fig.2. Predicted results on the BPI testing data.

Table 3. Performance results of the algorithms on the BCI testing data.

| Datasets | Performance Measure | Prediction Algorithms | |
		GBT	MLP
BCI	R^2	0.894	**0.926**
	MAE	0.053	**0.047**
	RMSE	0.072	**0.06**

Note: Best performances are bolded.

The predicted results approximately same to actual values in in Fig. 2 show that performances of the prediction algorithms are close to each other. In addition, Table 4 presents the performance measures of two

prediction algorithms within the frame of 1st approach (Model 1). According to this table, the most suitable model for BPI data is GBT because of the minimum error (RMSE: 0.072, MAE: 0.049) and the perfect fit of R^2 value (0.923).

Table 4. Performance results of the algorithms on the BPI testing data.

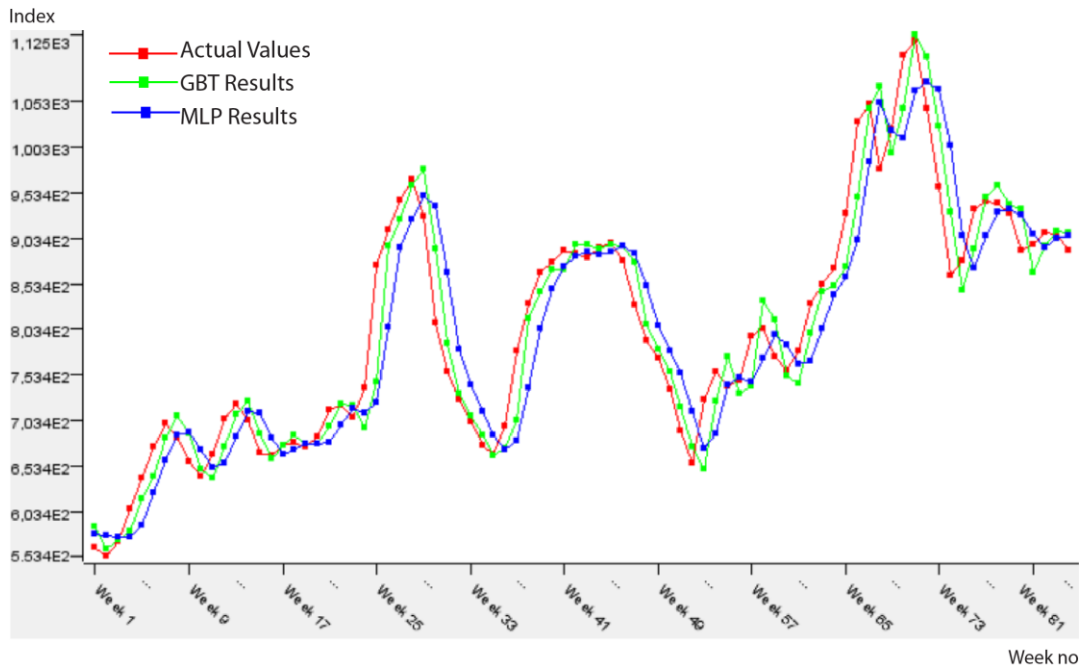| Datasets | Performance Measure | Prediction Algorithms | |
		GBT	MLP
BPI	R^2	**0.923**	0.891
	MAE	**0.049**	0.064
	RMSE	**0.072**	0.086

Note: Best performances are bolded.

Fig.3. Predicted results on the BSI testing data.

The predicted results approximately same to actual values in Fig. 3 show that performances of the prediction algorithms are close to each other. In addition, Table 5 presents the performance measures of two prediction algorithms within the frame of 1st approach (Model 1). According to this table, the most suitable model for BSI data is GBT because of the minimum error (RMSE: 0.063, MAE: 0.047) and the perfect fit of R^2 value (0.921).

Table 5. Performance results of the algorithms on the BSI testing data.

| Datasets | Performance Measure | Prediction Algorithms | |
		GBT	MLP
BSI	R^2	**0.921**	0.86
	MAE	**0.047**	0.066
	RMSE	**0.063**	0.084

Note: Best performances are bolded.

The predicted results approximately same to actual values in Fig. 4 show that performances of the prediction algorithms are close to each other. In addition, Table 6 presents the performance measures of two prediction algorithms within the frame of 1st approach (Model 1). According to this table, the most suitable model for BDI data is GBT because of the minimum error (RMSE: 0.051, MAE: 0.041) and the perfect fit of R^2 value (0.952).

Table 6. Performance results of the algorithms for Model 1 approach.

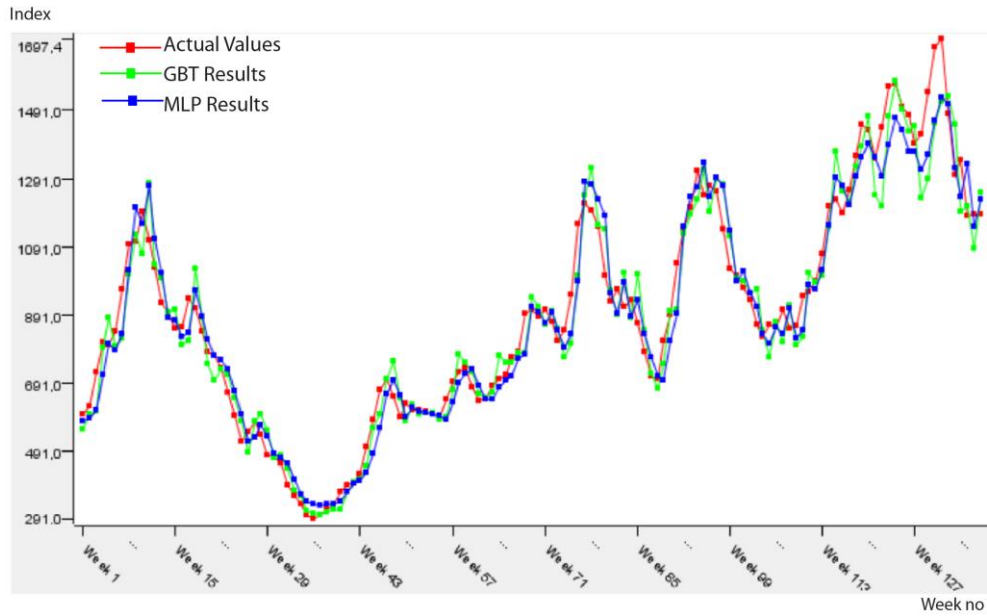| Datasets | Performance Measure | Prediction Algorithms | |
		GBT	MLP
BDI	R^2	0.942	**0.952**
	MAE	0.042	**0.041**
	RMSE	0.056	**0.051**

Note: Best performances are bolded.
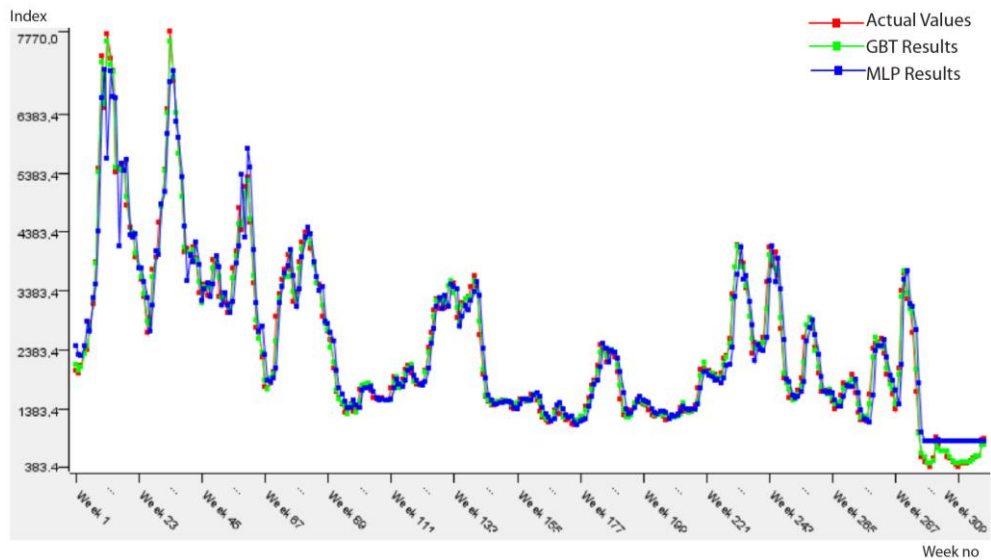
Fig.4. Predicted results on the BDI testing data.

Fig.5. Predicted results on the BCI_COP testing data.

The predicted results approximately same to actual values in Fig. 5 show that performances of the prediction algorithms are close to each other. In addition, Table 7 presents the performance measures of two prediction algorithms within the frame of 1st approach (Model 2). According to this table, the most suitable model for BCI_COP data is MLP because of the minimum error (RMSE: 0.065, MAE: 0.052) and the perfect fit of R^2 value (0.913).

Table 7. Performance results of the algorithms for Model 2 approach.

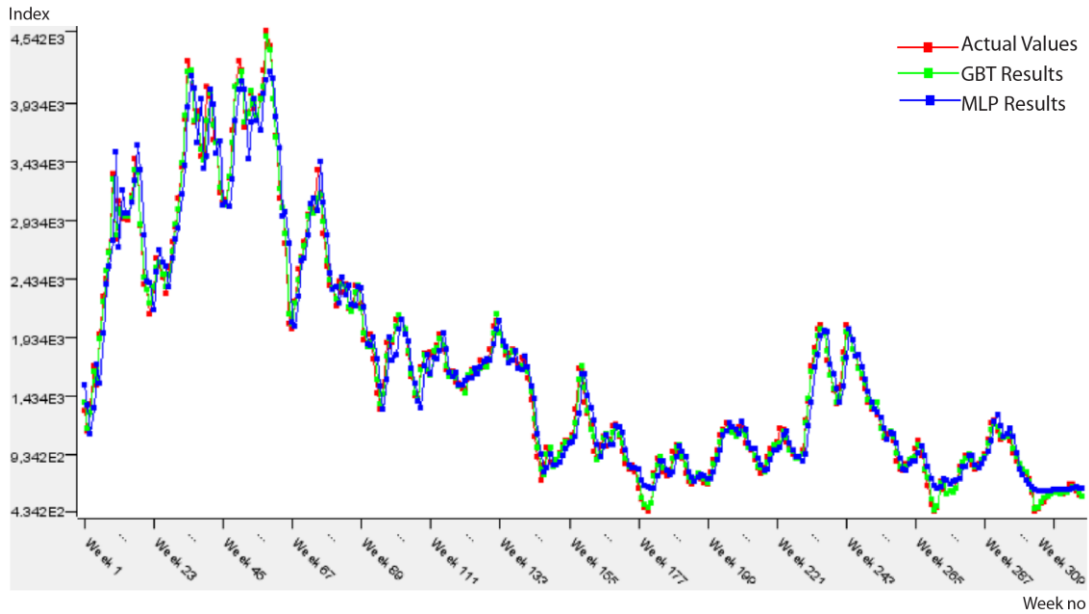| Datasets	Performance Measure	GBT	MLP
	R^2	0.906	**0.913**
BCI_COP	MAE	0.052	**0.052**
	RMSE	0.067	**0.065**

Note: Best performances are bolded.

Fig.6. Predicted results on the BPI_COP testing data.

The predicted results approximately same to actual values in Fig. 6 show that performances of the prediction algorithms are close to each other. In addition, Table 8 presents the performance measures of two prediction algorithms within the frame of 1st approach (Model 2). According to this table, the most suitable model for BPI_COP data is MLP because of the minimum error (RMSE: 0.074, MAE: 0.057) and the perfect fit of R^2 value (0.918).

Table 8. Performance results of the algorithms for Model 2 approach.

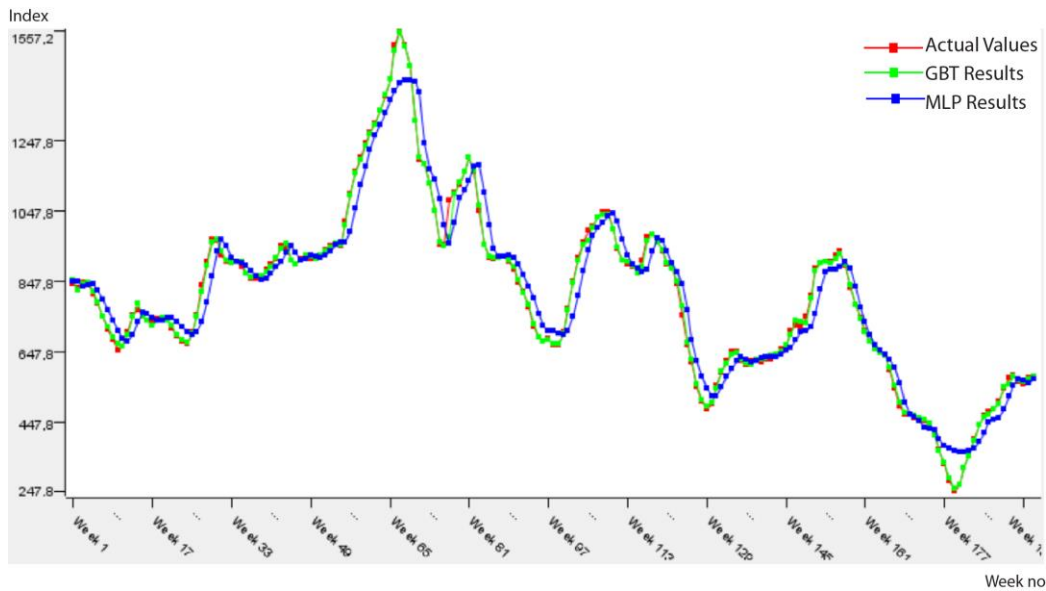| Datasets	Performance Measure	GBT	MLP
	R^2	0.911	**0.918**
BPI_COP	MAE	0.058	**0.057**
	RMSE	0.077	**0.074**

Note: Best performances are bolded.

Fig.7. Predicted results on the BSI_COP testing data.

The predicted results approximately same to actual values in Fig. 7 show that performances of the prediction algorithms are close to each other. In addition, Table 9 presents the performance measures of two prediction algorithms within the frame of 1st approach (Model 2). According to this table, the most suitable model for BSI_COP data is GBT because of the minimum error (RMSE: 0.079, MAE: 0.059) and the perfect fit of R^2 value (0.878).

Table 9. Performance results of the algorithms for Model 2 approach.

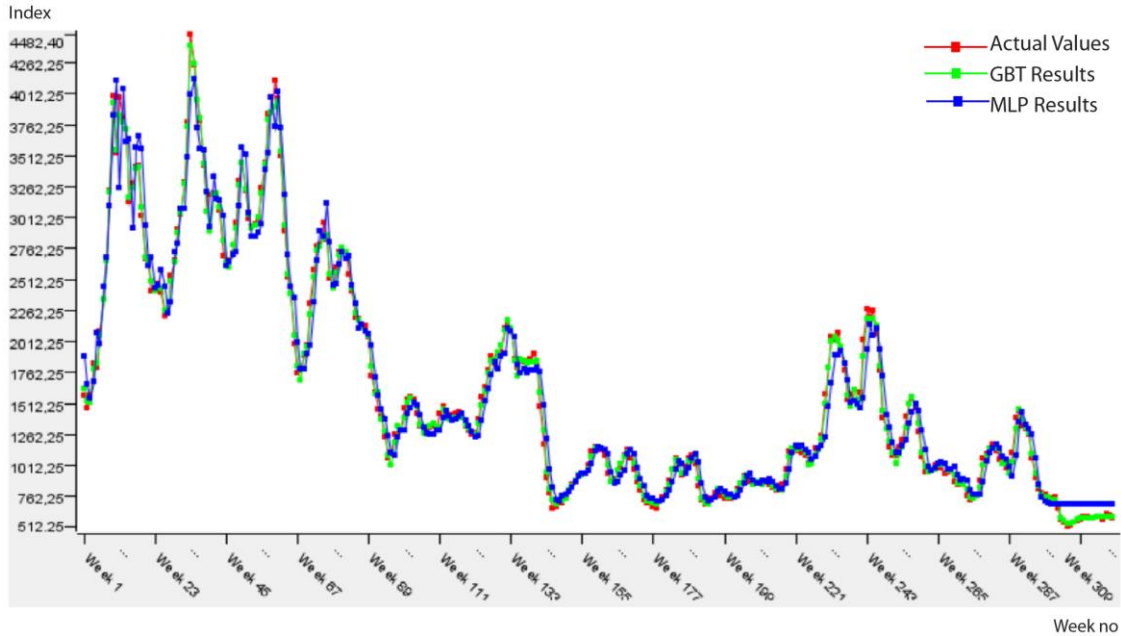| Datasets	Performance Measure	GBT	MLP
BSI_COP	R^2	**0.878**	0.747
	MAE	**0.059**	0.085
	RMSE	**0.079**	0.114
Note: Best performances are bolded.			

Fig.8. Predicted results on the BDI_COP testing data.

The predicted results approximately same to actual values in Fig. 8 show that performances of the prediction algorithms are close to each other. In addition, Table 10 presents the performance measures of two prediction algorithms within the frame of 1st approach (Model 2). According to this table, the most suitable model for BDI_COP data is MLP because of the minimum error (RMSE: 0.054, MAE: 0.042) and the perfect fit of R^2 value (0.947).

Table 10. Performance results of the algorithms for Model 2 approach.

Datasets	Performance Measure	GBT	MLP
BDI_COP	R^2	0.933	**0.947**
	MAE	0.044	**0.042**
	RMSE	0.06	**0.054**
Note: Best performances are bolded.			

VII. CONCLUSION

Maritime transportation highly affects the global economic system due to overseas shipping too much. Lloyd data presents the shipping prices of bulk cargoes that are generated from daily reported fixtures or estimations of the Baltic Exchange. Forecast of these index generally is hard because they are volatile, complex, and cyclic. In shipping sector ship owners, ship brokers, operators, traders, etc. depend on their past experience for their future decisions as a rule. Previous studies usually investigate the forecast of BDI data by considering diverse statistical regression methods. In this study, the prediction performances of the models designed with GBT and MLP algorithms are presented for four types of freight index data, Lloyd data. This study investigates the effectiveness of the prediction algorithms for the prediction of all freight index datasets and its applicability. In addition, GBT is the most consistent algorithm for BPI and BSI data. Moreover, MLP is the most convenient algorithm for BCI, BDI and BPI data while GBT is the most suitable algorithm for BSI data within the frame of 2nd approach. This study could be a powerful and beneficial reference in order to avoid risk in the shipping industry. In the future, the deep neural networks based models will be designed and the performances of these models will be compared.

REFERENCES

[1]   B. Şahin, S. Gürgen, B. Ünver, and İ. Altın, "Forecasting

the baltic dry index by using an artificial neural network approach," *Turk J Elec Eng & Comp Sci,* vol. 26, 2018, pp. 1673 – 1684.

[2] Q. Han, B. Yan, G. Ning, and B. Yu, "Forecasting dry bulk freight index with improved SVM," *Math Probl in Eng,* vol. 2014, Article ID 460684, 2014. doi: dx.doi.org/10.1155/2014/460684.

[3] F. Guan, Z. Peng, K. Wang, X. Song, and J. Gao, "Multi-step hybrid prediction model of baltic supermax index based on support vector machine," *Neural Netw World,* vol. 26, 2016, pp. 219-232. doi: 0.14311/NNW.2016.26.012.

[4] H.L. Wong, "BDI forecasting based on fuzzy set theory, grey system and ARIMA," In: Ali M., Pan JS., Chen SM., Horng MF. (eds) Modern Advances in Applied Intelligence, *Lect Notes Comput Sc,* vol. 8482, 2014, Springer, Cham.

[5] J.H. Friedman, "Greedy Function Approximation: A Gradient Boosting Machine," *The Annals of Statistics,* vol. 29, 2001, pp. 1189-1232.

[6] F. Rosenblatt, "Principles of neurodynamics: Perceptions and the theory of brain mechanism," *Spartan Books,* Washington, DC, 1961.

[7] C. Ming-Tao, "A fuzzy time series model to forecast the BDI," *Fourth International Conference on Networked Computing and Advanced Information Management,* vol. 2, 2008, pp. 50-53.

[8] G. Giannarakis, C. Lemonakis, A. Sormas, and C. Georganakis, "The effect of baltic dry index, gold, oil and USA," *Trade Balance on Dow Jones Sustainability Index World,* vol. 7, 2017; pp. 155-160.

[9] H. Geman and W.O. Smith, "Shipping markets and freight rates: an analysis of the baltic dry index," *The Journal of Alternative Investments,* vol. 5, 2012, pp. 98-109.

[10] V. Tsioumas, S. Papadimitriou, Y. Smirlis, and S. Z. Zahran, "A novel approach to forecasting the bulk freight market," *The Asian Journal of Shipping and Logistics,* vol. 33, 2017, pp. 33-41. doi: doi.org/10.1016/j.ajsl.2017.03.005.

[11] Q. Ruan, Y. Wang, X. Lu, and J. Qin, "Cross-correlations between baltic dry index and crude oil prices," *Physica A,* vol. 453, 2016, pp. 278-289. doi: doi.org/10.1016/j.physa.2016.02.018.

[12] P. Baltyn, "Baltic dry index as economic leading indicator in the United States," *Managing Innovation and Diversity in Knowledge Society Through Turbulent Time Proceedings of the Make Learn and TIIM Joint International Conference,* 25-27 May 2016; Timisoara, Romania.

[13] Q. Zeng, C. Qu, A.K. Ng, and X. Zhao, "A new approach for baltic dry index forecasting based on empirical mode decomposition and neural networks," *Marit Econ Logist,* vol. 18, 2016, pp. 192-210.

[14] J. Bao, L. Pan, and Y. Xie, "A new BDI forecasting model based on support vector machine," *Information Technology, Networking, Electronic and Automation Control Conference,* 20-22 May 2016, Chongqing, China: IEEE.

[15] N. Apergis and J.E. Payne, "New evidence on the information and predictive content of the baltic dry index," *International Journal of Financial Studies,* vol. 1, 2013, pp. 62-80. doi: 10.3390/ijfs1030062.

[16] Z. Yang, L. Jin, and M. Wang, "Forecasting baltic panamax index with support vector machine," *Journal of Transportation Systems Engineering and Information Technology,* vol. 11, 2011, pp. 50-57. doi: doi.org/10.1016/S1570-6672(10)60122-5.

[17] X. Yuwei, Y. Shaoqiang, F. Yonghui, and Y. Hualong, "Leverage effect analysis of baltic dry index based on EGARCH model," *Journal of Chemical and Pharmaceutical Research,* vol. 6, 2014, pp. 289-294.

[18] C.J. Willmott, and K. Matsuura, "Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance," *Clim Res,* vol. 30, 2005, pp. 79–82. doi: 10.3354/cr030079.

[19] D.L.J. Alexander, A. Tropsha, and D.A. Winkler, "Beware of R^2: simple, unambiguous assessment of the prediction accuracy of QSAR and QSPR models," *J Chem Inf Model,* vol. 55, 2015, pp. 1316–1322. doi: 10.1021/acs.jcim.5b00206.

Authors' Profiles

Kemal Akyol: He received his B.Sc. in Computer Science Department from Gazi University, Ankara/Turkey in 2002. He received his M.Sc. degree from Natural and Applied Sciences, Karabuk University, Karabuk/Turkey and Ph.D. degree from the same department. His research interests include data mining, decision support systems and expert systems.