ICIC Express Letters ICIC International ©2019 ISSN 1881-803X
Volume **13**, Number **5**, May **2019** pp. **427–434**

FORECASTING HIGH-DIMENSIONAL MULTIVARIATE REGRESSION OF BALTIC DRY INDEX (BDI) USING DEEP NEURAL NETWORKS (DNN)

Imam Mustafa Kamal[1], Hyerim Bae[2,*], Sunghyun Sim[2], Hyemee Kim[2]
Dohee Kim[2], Yulim Choi[2] and Heesung Yun[3]

[1]Department of Big Data
[2]Department of Industrial Engineering
Pusan National University
2, Busandaehak-ro 63 beon-gil, Geumjeong-gu, Busan 46241, Korea
{ imamkamal; ssh; khm0219; doheedohee; flamethrower }@pusan.ac.kr
*Corresponding author: hrbae@pusan.ac.kr

[3]Korea Maritime Institute
26, Haeyang-ro 301 beon-gil, Yeongdo-gu, Busan 49111, Korea
heesung@kmi.re.kr

Received October 2018; accepted January 2019

Abstract. *Forecasting of the Baltic Dry Index (BDI) is very important, as it provides an indicator of worldwide trade and manufacturing activity. This index fluctuates due to the influences of global factors involving many countries around the world. Deep Neural Networks (DNN) is a powerful method for dealing with complex problems. Therefore, in this paper, we propose its employment with 135 input layers and five hidden layers for BDI forecasting. Experiments with various DNN architectures and optimized hyper-parameters were conducted, yielding an average RMSE of 0.1764109. This result outperforms the previous trials using Long Short-Term Memory (LSTM) (RMSE: 0.2312339). Indeed, it indicates that DNN is suitable for BDI forecasting.*

Keywords: Forecasting, Baltic Dry Index, Regression, Deep neural networks

1. **Introduction.** The Baltic Dry Index (BDI) is a shipping and trade index created by the London-based Baltic Exchange. In a nutshell, it provides shipment costs for dry bulk cargoes consisting of commodities such as grain, coal, iron, ore, and copper. The BDI is a composite of three sub-indices, namely Capesize, Panamax, and Supramax. Those indices have different bulk carrier capacities, 150,000, 65,000-80,000, and 50,000-60,000 dwt, respectively. The BDI has been used as a world trade economic indicator on a global scale. Many stakeholders make serious efforts to forecast it, precisely, so as to be able to make smart investment decisions. However, the fluctuation of BDI is known to be similar to the fluctuation of stock exchange prices. Therefore, it is a challenging task to perform prediction against BDI value.

Among the many fluctuation-effecting factors are vessel supply, commodity demand, seasonal pressures, bunker prices, choke points, and market sentiment, all of which involve many countries around the world. Even accounting carefully for all of those factors is not sufficient for precise BDI forecasting. Many studies have undertaken to uncover other factors influencing the BDI. Lin and Sim tried to find associations between trade (the BDI) and income improvements for certain Least Developed Countries (LDCs) [1]. Ruan et al. found cross-correlations between the BDI and crude oil prices [2]. Bildirici et al. determined that the BDI can be used as a crisis indicator of U.S. GDP growth [3]. Bildirici et al. also revealed, in a subsequent study, a positive correlation among the BDI,

DOI: 10.24507/icicel.13.05.427

the price of gold, and economic growth in the U.S. [4]. In China, which is the biggest iron consumer in the world, the price fluctuation of iron ore is positively correlated with the price of crude oil and the BDI [5]. All of the relevant previous studies serve to demonstrate that BDI forecasting, which entails consideration of a number of global factors, not only is quite challenging, but also might require consideration of many other factors as well.

Recently, deep learning has been prominent in outperforming other machine-learning methods in image recognition, speech recognition, and object detection [6]. In ImageNet classification for example, the deep-learning architecture tends go ever deeper to obtain the best results. Given such outcomes, it can be inferred that more complex architecture yields better results. However, many previous researches still utilize machine learning to forecast fluctuating indices such as stock markets and the BDI [7-10]. Therefore, in this research we proposed to employ a deep learning to predict BDI.

The rest of this paper is organized as follows. Section 2 presents the related work. Sections 3 and 4 discuss the proposed method and experimental results, respectively. Finally, Section 5 draws conclusions and looks forward to future work.

2. **Related Work.** In previous research, Kim et al. [11] forecasted the BDI using Long Short-Term Memory (LSTM). Their LSTM architecture is with 25 input variables, various hidden layers scenarios, and four outputs. Afterward, the four outputs are calculated by using a weighting factor to calculate the BDI. This research yielded the best RMSE (0.0673027). Lu et al. analyzed a series of BCI (Baltic Capesize Index) day earnings ratios and Baltic Panamax Index (BPI) using a new GARCH model, and also scrutinized the freight-rate characteristics of world dry bulk shipping market fluctuation using a statistical method [12]. Wei applied fractal theory to an analysis of the features of international dry bulk shipping's long memory [13]. He argued that long memory provides a more accurate forecast than short memory can. However, statistical analysis, though a good method, is inadequate for analysis of the complex problems that the BDI presents.

Chou applied fuzzy time-series model to a forecast of the BDI's next month, achieving the impressive RMSPE (Root Mean Squared Prediction Error) of 4.28% [14]. One of the drawbacks of fuzzy methods remains their reliance on human assessment for determination of thresholds; indeed, this would have incurred a problem in our present case, as we were dealing with 135 features. Bao et al. proposed Support Vector Machine (SVM) combined with Correlation-based Feature Selection (CFS) to forecast the BDI [7]. They compared their result with that of a neural network model, and found that their proposed method had shown better performance with respect to both the trend and forecast precision. Meesad and Rasel forecast stock market prices using different types of windowing functions and Support Vector Regression (SVR) [15]. SVR is a powerful machine-learning method, though choosing the correct kernel to solve non-linear cases for such a fluctuating index as the BDI is daunting.

3. **Proposed Method.** Time-series data with numerical values was gathered from the Korea Maritime Institute. The values represent the shipping costs of various raw materials for different routes around the world. The costs considered were all in USD as derived from the Baltic Capesize Index, the Baltic Panamax Index, the Baltic Supramax Index, Bunker prices, the Bulk-carrier-time charter rate, Capesize Earning, Panamax Earning, and Supramax Earning. The data is weekly starting from 2005-07-03 and ending at 2018-02-11. The overall data consists of 365 features. We employ those high dimensional features for predicting BDI.

Data preprocessing was run before utilizing Deep Neural Networks (DNN), as shown in Figure 1. Feature selection was conducted to select the best features among 365 total features. To calculate the correlation between each of the features with the BDI, we used Pearson correlation (the Pearson correlation coefficient being a measure of the strength

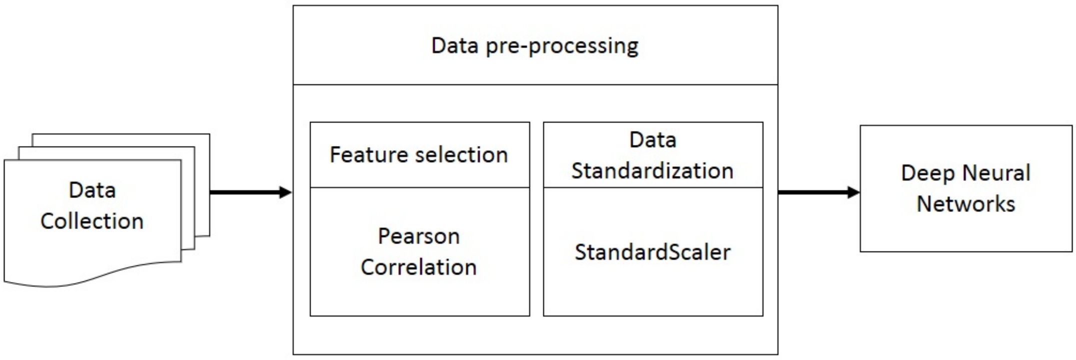


Figure 1. Designed system of the proposed method

of the linear relationship between dependent and independent variables), as shown in Equation (1), where r_{xy} is the correlation value in the range of 0-1, X is the independent variable, Y is the dependent variable, and $\overline{X}$ and $\overline{Y}$ are the means of the X and Y variables, respectively. We set the threshold for selection of variables having a correlation value higher than ± 0.7. The experimental results indicated that 135 of the 365 features were highly correlated with the BDI.

$$r_{xy} = \frac{\sum_{i=1}^{n} \left(X_i - \overline{X}\right)\left(Y_i - \overline{Y}\right)}{\sqrt{\sum_{i=1}^{n} \left(X_i - \overline{X}\right)^2} \sqrt{\sum_{i=1}^{n} \left(Y_i - \overline{Y}\right)^2}} \tag{1}$$

$$z = \frac{X - \mu}{\sigma} \tag{2}$$

The original data contained noise and had a different scale, such as unit values in dollar cost, DWT (Deadweight Tonnage), and indices. Therefore, we scaled the data by harnessing a standard scaler using Equation (2), where X is the data, and μ and σ are the mean and standard deviations, respectively. In other words, standardization transforms data such that its distribution will have a mean value 0 and standard deviation of 1. Standardization of a dataset is commonly requiree for many machine learning estimators: they might behave badly if the individual feature does not more or less look like standard normally distributed data.

Based on feature selection using Pearson correlation, 135 important variables highly correlated with the BDI were derived. Therefore, the input layer in our DNN had 135 neurons. We utilized five hidden layers with different numbers of neurons, as indicated in Table 1 and Figure 2. The output is a single neuron representing the BDI value. Due to the complexity of the architecture, we used bias and drop-out functions in each layer against overfitting. AdagradOptimizer was used to optimize the hyper-parameters in the training phase. And, various activation functions were examined for each layer as shown

Table 1. RMSEs of testing data for various DNN architectures

Architecture	RMSE		
	Tanh	ReLu	Sigmoid
135-640-512-256-128-64-1	3.0011148	2.4517422	0.2782829
135-512-512-256-256-128-1	7.5426189	10.9884484	**0.0869994**
135-512-512-512-128-128-1	**2.7025432**	**1.6023951**	0.1236289
135-128-256-256-512-512-1	3.2097196	18.4451349	0.2878388
135-64-128-256-512-640-1	5.3379959	34.7570137	0.1465725

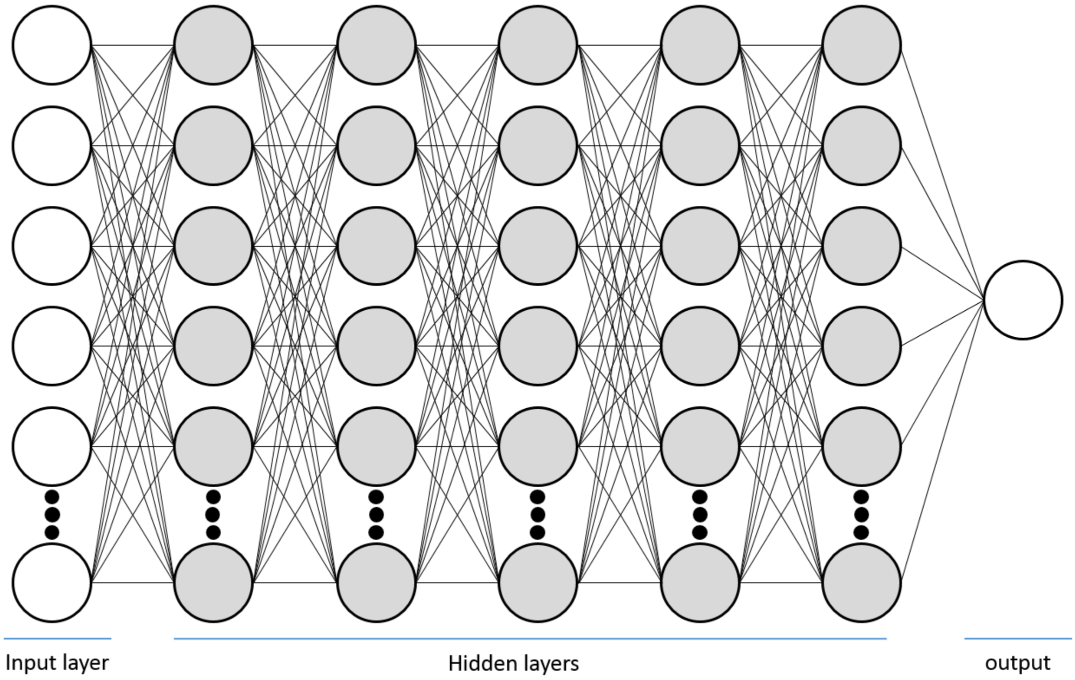


FIGURE 2. Deep Neural Networks (DNN) architecture

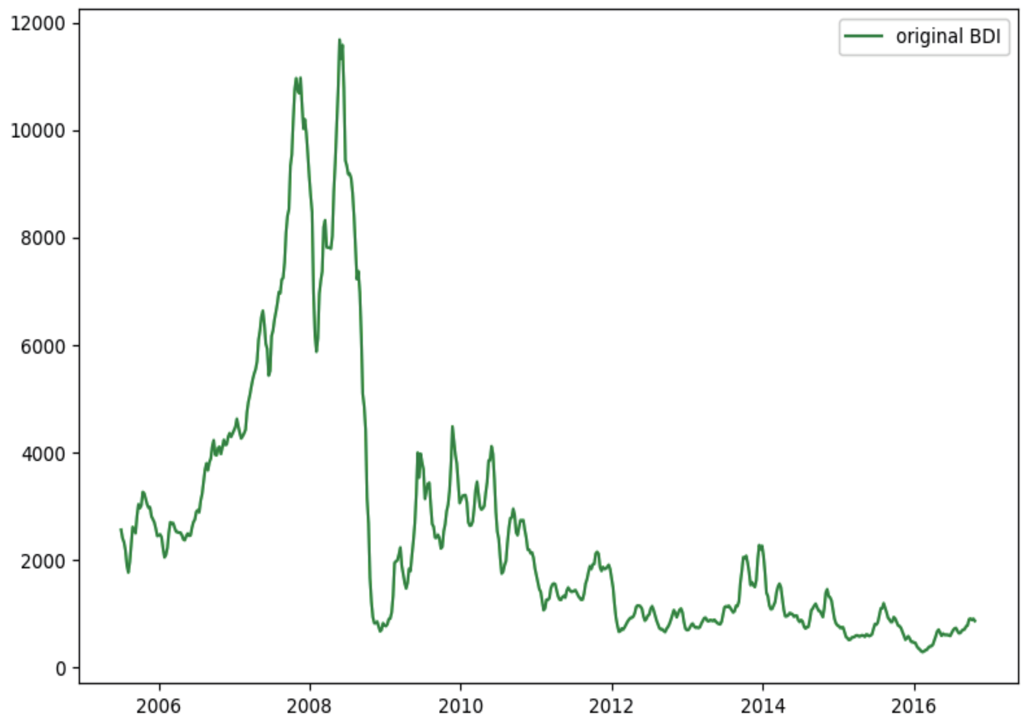
TABLE 2. RMSEs for training and testing batch scenarios, using 135-512-512-256-256-128-1 architecture and Sigmoid activation function for each hidden layer (best DNN architectures based on Table 1)

Data Partition		RMSE	
Training (%)	Testing (%)	Training	Testing
60	40	0.0080379	0.2220079
70	30	0.0098774	0.1910654
80	20	0.0071754	0.0869994
Average		**0.0084391**	**0.1764109**

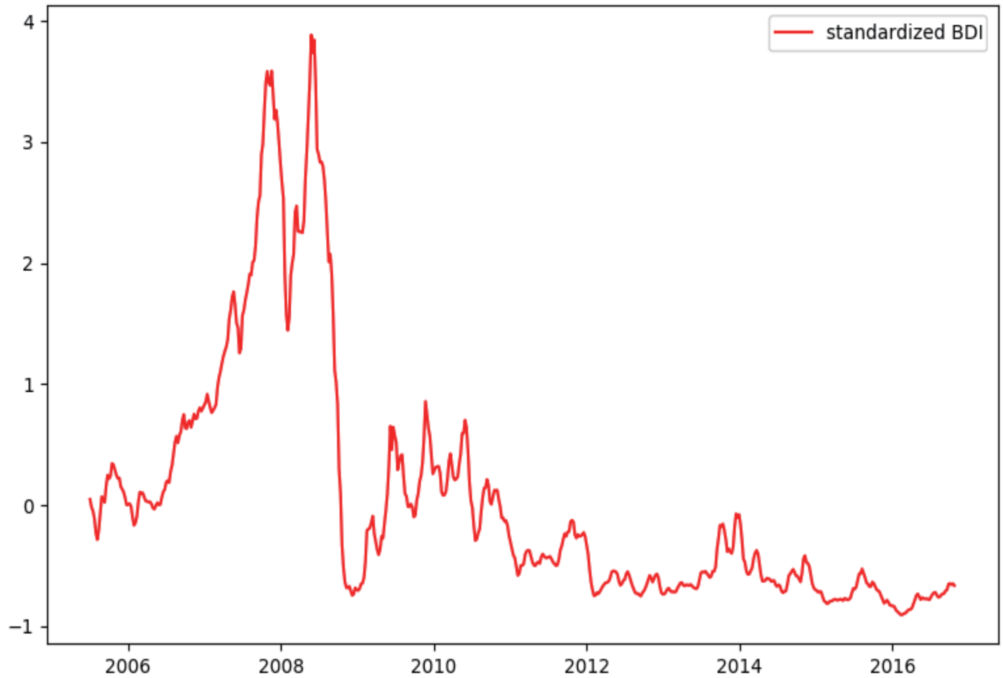
in Table 1, except for the last layer, for which we used the linear function. These various procedures were conducted so as to obtain the best configuration of DNN architecture. Finally, to assess how good our model was based on Table 1, some scenarios for portions number of training and testing data are examined as shown in Table 2.

4. **Experimental Results.** In this research, the data was partitioned according to two scenarios, as indicated in Table 2. The experiment was conducted using a PC with an Intel Core i7 4790K CPU, 32 Gb RAM, a GPU NVIDIA GeForce GTX 1080Ti, Python and TensorFlow as the deep learning framework. We restricted the number of epochs to 10,000 for training, which took roughly 3.5 minutes.

In order to enhance the training process to make the network faster to converge, we transformed the data to zero centers by using a standard scaler. As depicted in Figure 3, the BDI range, originally 0-12000 was -1-4 after implementing the standard scaler. In this research, the experiment was conducted using various architectures and different activation functions, as shown in Table 1, for example, 135 neurons in the input layer, 640 in the first hidden layer, 512 in the second hidden layer, and so on. Our DNN model was designed for both the pyramid form (the number of neurons decreases as the number of hidden layers increases) and the reversed pyramid form (the number of neurons increases as the number of hidden layers increases). The third architecture model obtained the best results: using the Tanh or ReLu activation function and the Tanh and ReLu



(a)



(b)

FIGURE 3. Baltic Dry Index (BDI) for 2005-2017: (a) original BDI, (b) standardized zero-centered BDI

activation function, achieving RMSEs of 2.7025432 and 1.6023951, respectively. The best RMSE (0.0869994) was obtained using the second architecture model with Sigmoid as the activation function: (135-512-512-256-256-128-1).

To assess how well our model forecast the BDI, we ran different scenarios for the time-series data. By using rolling time window (see Table 2), in the first experiment, we split the data 60/40% for training/testing; in the second experiment, 70/30%, and in the final experiment, 80/20%. All of those scenarios obtained RMSEs. As shown in Table 2, the

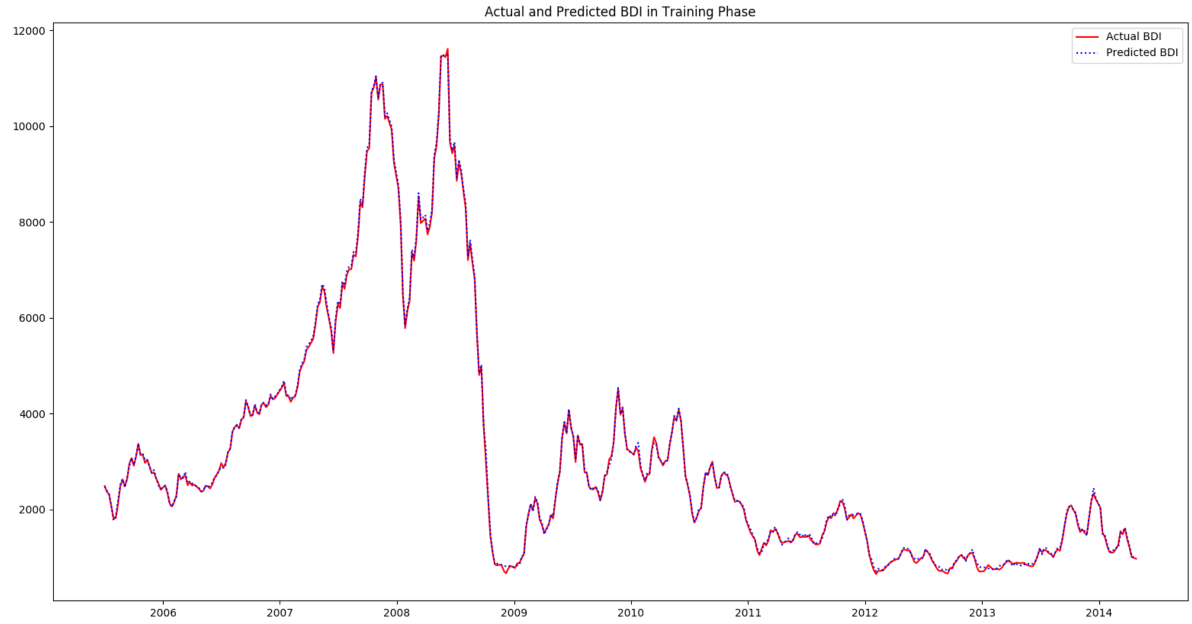


FIGURE 4. Forecast BDI using DNN (dot), and original BDI data (line) in training phase. The model is fit well with the training data.

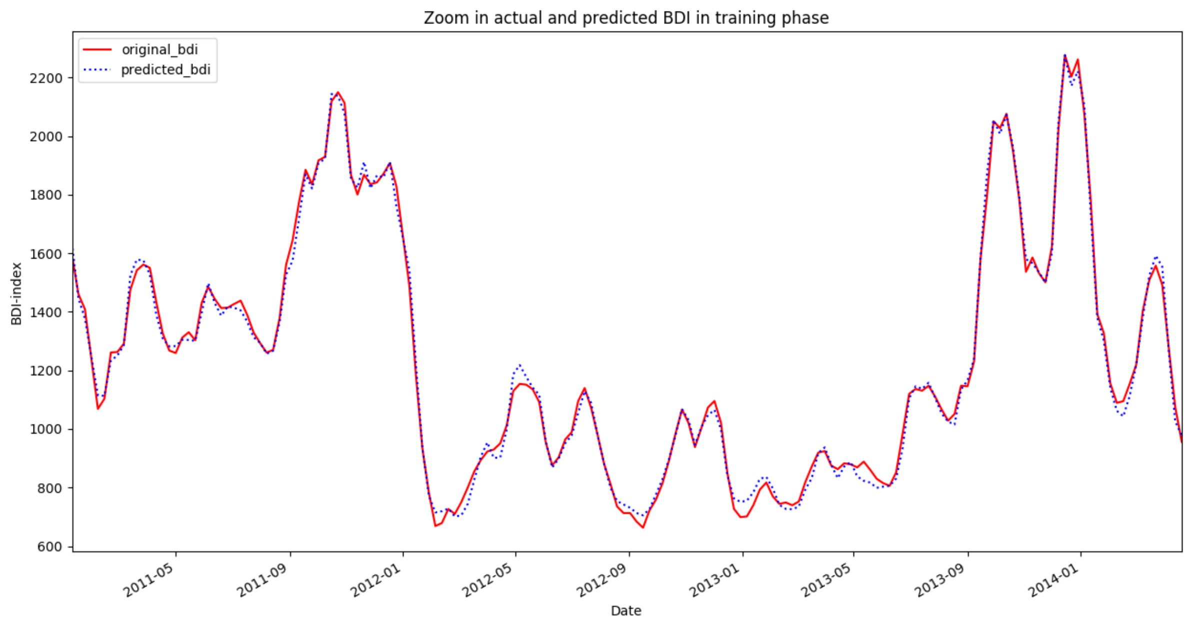


FIGURE 5. Zoom-in of Figure 4, detailed view of the difference between forecast and actual model

average RMSE in training was 20 times lower than the RMSE in testing. This indicated that our model suffered from overfitting: that is, the model was fit too well for training data. The detailed model and actual BDI are depicted in Figure 4, and zoomed-in version is shown in Figure 5.

To prevent overfitting, we implemented dropout [16] in our DNN model, whereby the number of neurons in a particular layer is reduced based on predefined probability. This mechanism tended not to have a significant impact in our case, due to the BDI data containing an outlier, specifically a peak index at around 2008, which was assumed to correspond to the world economic crisis at that time. This outlier caused the BDI data variance to be high. Therefore, the forecast result in the testing phase was not as precise

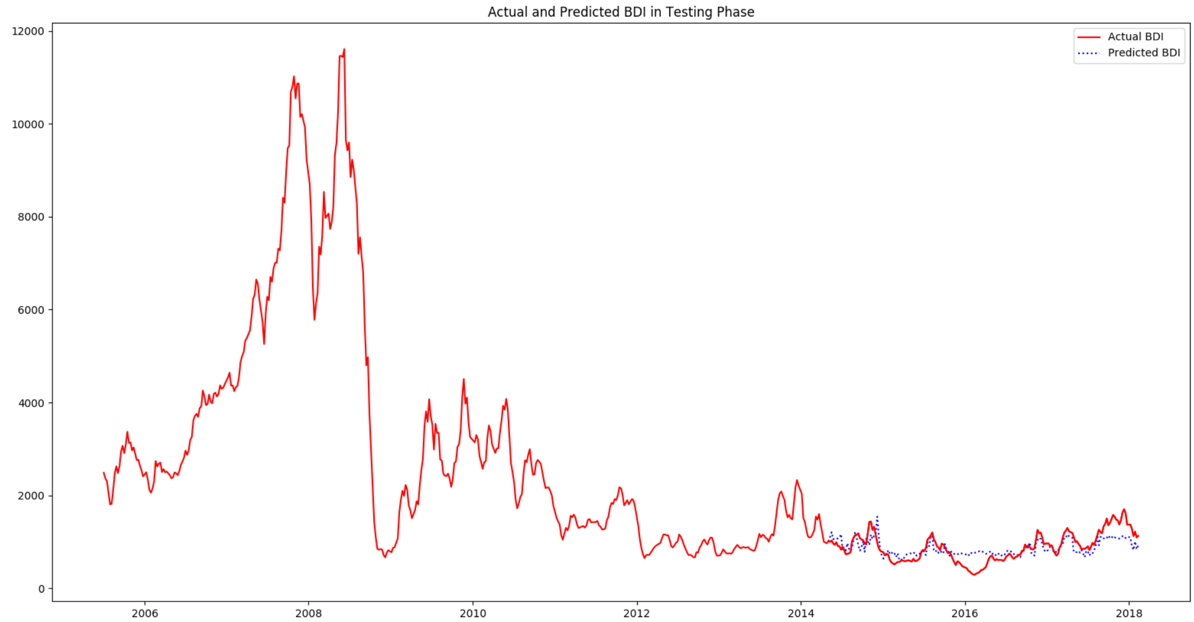


Figure 6. Actual BDI model (line) and testing data to forecast BDI started from 2015 to 2018 (dot)

as in the training phase, as shown in Figure 6. Compare with the previous research by using LSTM [11] as described in Section 2. The testing RMSEs were 0.2312339 and 0.1764109 for the LSTM and our DNN, respectively. Accordingly, this research conducted a better approach to perform prediction against BDI.

5. **Conclusions and Future Work.** The BDI presents a complex regression problem. This fact notwithstanding, there is an opportunity to determine patterns using a machinelearning approach and harnessing historical data. In this research, we utilized a proposed DNN of complex architecture having 135 input variables, the experimental results for which were highly correlated with the BDI. Those results indicated that the best architecture was 135-512-512-256-256-128-1 with the Sigmoid activation function in each of the hidden layers. Using this architecture, average RMSEs of 0.0084391 and 0.1764109 were obtained for training and testing, respectively. These results reflected the fact that our model suffered from overfitting, it having performed better in the training phase than in the testing phase. Nonetheless, the testing RMSE of 0.1764109 was shown, in comparison with relevant previous research utilizing LSTM (RMSE: 0.2312339), to be quite robust for BDI forecasting. In future work, we will compare this result with those of other machine-learning approaches such as ensemble methods and SVM. Further, as the BDI is a composite of the Capesize, Panamax, and Supramax indices, we will try to forecast those indices first by means of a newly designed hierarchical model, on which basis we will then forecast the BDI.

Acknowledgment. This work was partly supported by the MSIT (Ministry of Science and ICT), Korea, under the Grand Information Technology Research Center support program (IITP-2018-2016-0-00318) supervised by the IITP (Institute for Information & Communications Technology Promotion) and National Research Foundation of Korea (NRF) grant funded by the Korean Government (MEST) (No. NRF-2015R1D1A1A09061331).

REFERENCES

[1] F. Lin and N. C. S. Sim, Trade, income and the Baltic Dry Index, *European Economic Review*, vol.59, pp.1-18, 2013.

[2] Q. Ruan, Y. Wang, X. Lu and J. Qin, Cross-correlations between Baltic Dry Index and crude oil prices, *Physica A: Statistical Mechanics and Its Applications*, vol.453, pp.278-289, 2016.

[3] M. E. Bildirici, F. Kayıkçı and I. Ş. Onat, Baltic Dry Index as a major economic policy indicator: The relationship with economic growth, *Procedia – Social and Behavioral Sciences*, vol.210, pp.416-424, 2015.

[4] M. E. Bildirici, F. Kayıkçı and I. Ş. Onat, BDI, gold price and economic growth, *Procedia Economics and Finance*, vol.38, pp.280-286, 2016.

[5] H. Li, B. Wang, E. Ren and C. Wu, Empirical analysis of the influencing factors on iron ore prices, *The 2nd International Conference on Artificial Intelligence, Management Science and Electronic Commerce (AIMSEC)*, Deng Leng, pp.3004-3008, 2011.

[6] A. Krizhevsky, I. Sutskever and G. E. Hinton, ImageNet classification with deep convolutional neural networks, *Proc. of the 25th International Conference on Neural Information Processing Systems*, pp.1097-1105, 2012.

[7] J. Bao, L. Pan and Y. Xie, A new BDI forecasting model based on support vector machine, *IEEE Information Technology, Networking, Electronic and Automation Control Conference*, Chongqing, pp.65-69, 2016.

[8] G. S. Atsalakis and K. P. Valavanis, Surveying stock market forecasting techniques – Part II: Soft computing methods, *Expert Systems with Applications*, vol.36, no.3, pp.5932-5941, 2009.

[9] S. Soni, Applications of ANNs in stock market prediction: A survey, *International Journal of Computer Science & Engineering Technology*, vol.2, no.3, pp.71-83, 2011.

[10] D. Olson and C. Mossman, Neural network forecasts of Canadian stock returns using accounting ratios, *International Journal of Forecasting*, vol.19, no.3, pp.453-465, 2003.

[11] D. Kim, H. Kim, S. Sim, H. Bae, Y. Choi and H. Yun, Predicting BDI (Baltic Dry Index) using LSTM, *KIIE 2018 Korea Institute of Industrial Engineers*, 2018.

[12] J. Lu, F. Wei and Q. Chen, Study on the fluctuant characteristics of freight index in world dry bulk shipping market subdivided with ARCH model, *IEEE/SMC International Conference on System of Systems Engineering*, Los Angeles, CA, p.6, 2006.

[13] F. Wei, Analysis on long memory of the volatilities of international dry bulk freight index using fractal theory, *International Conference on Wireless Communications, Networking and Mobile Computing*, Shanghai, pp.6333-6336, 2007.

[14] M. T. Chou, A fuzzy time series model to forecast the BDI, *The 4th International Conference on Networked Computing and Advanced Information Management*, Gyeongju, pp.50-53, 2008.

[15] P. Meesad and R. I. Rasel, Predicting stock market price using support vector regression, *International Conference on Informatics, Electronics and Vision (ICIEV)*, Dhaka, pp.1-6, 2013.

[16] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever and R. Salakhutdinov, Dropout: A simple way to prevent neural networks from overfitting, *Journal of Machine Learning Research*, pp.1929-1958, 2014.