Article

Spot Charter Rate Forecast for Liquefied Natural Gas Carriers

Dimitrios V. Lyridis [1,*]

[1]   Laboratory for Maritime Transport, National Technical University of Athens, Athens, Greece; dsvlr@mail.ntua.gr

*   Correspondence: dsvlr@mail.ntua.gr

Abstract: Recent maritime legislations demand the transformation of the sector to greener and more energy efficient transportation. Liquified Natural Gas (LNG) seems a promising alternative fuel solution that could replace the conventional fuel sources. Various studies have been focused on the prediction of LNG price, however, no previous work has been made on the forecast of spot charter rate of LNG carrier ships. An important knowledge for the maritime industries and companies when it comes to decision-making. Therefore, this study is focused on the development of a machine learning pipeline to address the aforementioned problem by: (i) forming a dataset with variables relevant to LNG; (ii) identifying the variables that impact on the freight price of LNG carrier; (iii) developing and evaluating regression models for short and mid-term forecast. The results showed that the General Regression Neural Network presented a stable overall performance for 2, 4 and 6 months forecast.

Keywords: machine learning; forecast; regression models; Liquified Natural Gas; maritime transportation

1. Introduction

Maritime fuel combustion is estimated to contribute 3% of the annual global greenhouse gas emissions [1]. The International Maritime Organization (IMO) regulations on the marine fuel impose the need for greener transportation. These regulations include the limitation to Sulphur Emission Control Areas (SECAs) and Nitrogen Oxide Emission Control Areas (NECAs) while they encourage alternative fuel sources that will contribute to the increase of greenhouse gas emissions and capital investments [2]. A promising alternative solution for fuel is the Liquefied Natural Gas (LNG) [3,4]. In 2020, the International Gas Union (IGU) reported a significant increase in the number of terminals for LNG liquefaction and regasification. Moreover, the European Commission, in the context of Clean Power for Transport Directive, has supported the deployment of alternative fuels, recharging and refueling infrastructures. Furthermore, the evidence shows the gradually development of short term and sport LNG market and consequently the corresponding shipping market (Fig. 1) [5]. Following the unconventional gas revolution, the forecasting of natural gas prices and moreover the freight prices of LNG bunkering ships have become important due to the low association of these prices with those of crude oil [6].

Figure 1.Share of spot and short term versus the total LNG trade (MTPA / %) [7].

Machine learning and artificial intelligence analytics have been commonly employed for forecasting prices of fuels [8,9] in the energy or marine sector. Recent studies have

highlighted the necessity of adopting alternative fuels for a more sustainable marine transportation, proposing the liquefied natural gas (LNG) as a greener ship fuel [10,11]. To this end, the scientific community has turned its attention to the study of LNG as ship fuel in the marine sector, focusing on economic feasibility, safety analysis and decision-making models regarding the use of LNG [11–14].

In the literature various studies have been conducted regarding the forecasting of price of LNG. Specifically, in [15] hybrid models based on the combination of wavelets, time series, and Artificial Neural Networks (ANNs) have been proposed in order to predict the price of natural gas. An Autoregressive Neural Network (ARNN) model was presented in [16] for predicting natural gas spot prices. A least squares regression boosting (LSBoost) algorithm was applied for a data-driven daily, weekly and monthly forecast of natural gas spot price on Henry Hub time series [17]. In the context of the implementation of thesis, an analysis of potential influence factors regarding the spot price movements of natural gas was conducted based on a Nonlinear Autoregressive Neural Network with eXogenous inputs (NARX)　[18]. However, the study was limited to the market area NetConnect Germany (NCG). Time series and various Artificial Neural Network (ANN) models were also adopted in another thesis [19] for predicting the price of natural gas in the United States market. Another study for a specific market was conducted for the Regional Basis in Turkey [20], where an artificial bee colony-based artificial neural networks (ANN-ABC) was developed to forecast the day-ahead demand of natural gas. From company perspective, a stochastic programming approach was adopted for an optimal planning of LNG purchase for oil and gas companies [21]. Hence, the model aimed to predict the demand and the prices of LNG in a planning horizon.

In the literature various studies have been conducted regarding the forecasting of price of LNG. In [22], a combination of two non-parametric methods, such as a rescaled range analysis and a multifractal detrended fluctuation analysis, was presented for statistical analysis, with respect to the correlation, fluctuation and scaling, of the freight process in the Liquid Petroleum Gas shipping market. Another statistical analysis was performed in [23]. The aim of this study was to identify the multiple financial and shipping related measures that have statistically significant contribution on the prediction of the spot voyage time charter price of P1A Panamax shipping route. In [24], linear and non-linear methods were evaluated for short-term forecasts in the dirty tanker shipping market. Serial time series and neural networks were involved in this study for a more accurate prediction of freights that will support the decision making of maritime companies. An extension of [24] is presented on freight rates derivatives to improve the prediction accuracy of the models [25]. ANNs were also employed for modeling the Baltic dry bulk shipping market trained by macroeconomic factors and shipping market parameters [26]. Specifically, an ANN was trained by using real data for a twenty-year period for a wide range of macroeconomic factors (nineteen) and maritime indexes (four). A similar study focusing on the dry bulk shipping market on the BPI T/C and BCI C7 routes was conducted in [27], where vector autoregression model and vector error connection model were applied for identifying the dynamics and interactions between spot and forward freight agreement prices. In [28], adynamic probit model was developed to forecast the future weekly, quarterly and biyearly changes of spot freight rates for Panamax dry bulk ships [28]. To forecast the Value-at-Risk (*VaR*) of dry bulk shipping markets, nine different risk models were developed and evaluated in [29]. Regarding the crude oil market, the periodic variation law of the tanker market was studied based on quantitative methods. The paper investigates the cycle duration and amplitude of different scales of Aframax tanker's freight to predict the long-term variation trend of freight rate on that basis.

Even if various studies have been conducted for the forecast of LNG price and the charter rate in various shipping markets, to the best of our knowledge, no previous work has been implemented regarding the forecast of the charter rate for the LNG shipping market. This is especially important as any increase in LNG demand as a marine fuel will be interconnected with increased demand of LNG transportation service via LNG carriers,

from liquefaction terminals to local storage and bunkering hubs. To this end, this study contributes to the development of a machine learning approach for:

identifying the features that contribute to the accurate prediction;

employing various neural networks for 2-months, 4-months and 6-months forecasting the charter rate of LNG carriers;

evaluation of the predictive models and comparisons with respect to the forecasting of LNG 145K CDM Spot Rate on 01/03/2017, 01/05/2017 and 01/07/2017 based on time series data from 01/01/2010 until 01/01/2017.

This paper is organized as follows. Section 2 gives a description of the dataset that was used in our paper. In Section 3, the proposed methodology along with the necessary feature selection and validation mechanisms, are presented. Results are given in Section 4. Conclusions and future work are finally drawn in Section 5.

2. Materials and Methods

2.1 LNG Data

The data relevant to LNG were collected from multiple data sources (Table 1). The data collection was focused on the chronological period from 01/01/2010 until 01/01/2017, since the short-term market was significantly increased from 2010 onwards

Table 1. Data sources and the time series used in the methodology

Data source	Data description of time series
Clarkson PLC – Shipping Intelligence Network	LNG 145K CBM Spot Rate ($/day): the desired prediction variable. It represents the price of the daily fare for an LNG tanker with a capacity of 145,000 CBM and a steam turbine vessel.
	LNG 160K CBM Spot Rate ($/day): the price of the daily fare for an LNG tanker with a capacity of 160,000 CBM, Tri-Fuel Diesel Electric (TFDE).
	LNG 160K CBM 1 Year Timecharter Rate ($/day): it presents the price of the daily fare for one-year contracts for a ship with the same characteristics as above.
	World Seaborne LNG Trade (million tonnes): it reveals the demand for LNG regarding the quantity that is traded internationally.
	World Seaborne LNG Trade (billion tonne-miles): it represents the trade of LNG, multiplied by the distance that the commodity has travelled.
	Import LNG Japan Price ($/mmbtu): the import price of LNG in Japan.
GIIGNL - International Group of LNG Importers	Total LNG Fleet: it reveals the number of vessels that transport LNG.
	Total Shipping Capacity (m^3 - CBM): it is related to the offer and shows the total capacity of all LNG vessels.
	Operational Capacity (m^3 - CBM): it presents the total operating capacity for trading LNG. Its combination with the operating capacity shows the percentage of ships that are inactive at a specific time in the market.
	New Orders Placed: the new ship orders indicate the attitude of shipowners toward the future of the LNG market.
	Orderbook: the orderbook shows reflects the capacity and the ability of shipyards to accept new orders in near future.
	Ships Delivered That Year: it presents the number of ships that the shipyards deliver in that year.
	Liquefaction Plants/Liquefaction (Million Tonnes Per Annum - MTPA): it presents the amount of gas that is liquefied. While new liquefaction plants are being built, it shows that the market is on the rise.

	Liquefaction Plants/Storage (m³ - CBM): it directly affects the short-term purchase of LNG. The storage capacity was one of the main factors that led to the rise of the short-term market, allowing sellers to keep the quantities they produce and dispose of them whenever they consider it necessary.
	Regasification Plants/Storage (m³ - CBM): it shows the evolvement of the ability to store LNG in regasification stations.
	Regasification Plants/Sent Out (Billion Cubic Meters - bcm/year): it refers to the annual quantities of LNG that is gasified.
	Spot LNG Imports (million tonnes): it is linked with the quantities of LNG imported under the direct delivery regime.
U.S. Energy Information Administration (EIA)	Price of Liquefied U.S. Natural Gas Exports ($/thousand cubic feet): the price of LNG exported by the USA.
	Henry Hub Natural Gas Spot Price ($/million btu): Henry Hub is a gas pipeline located in Louisiana, USA. It is the pricing reference point for gas contracts traded on the New York Mercantile Exchange (NYMEX). Settlement prices are used as benchmarks for the entire North American gas market as well as for parts of the global LNG market. It is an important indicator as the price of natural gas is based on real supply and demand as a stand-alone commodity.
	WTI Oil Price ($/barrel): West Texas Intermediate (WTI) crude oil is the basis for New York oil futures contracts. This indicator is important as it is a reference point for buyers and sellers of oil.
	Brent Oil Price ($/barrel): Brent is a blend of crude oil exported from the North Sea. It is the reference point for most of the crude oil in the Atlantic basin and it is used to price two thirds of the crude oil traded internationally.
BP Statistical Review of World Energy	Worldwide Natural Gas Production (Billion Cubic Meters - bcm): it shows the global production of natural gas.
	Worldwide Natural Gas Consumption (Billion Cubic Meters - bcm): it shows the global consumption of natural gas.

2.2 Methodology

To predict a price of a specific product/index in the market, a common approach is to identify the correct data to use, adopt a pre-process methodology to transform them and identify certain patterns for knowledge extraction. These approaches are commonly implemented via machine learning techniques[30,31]. The proposed machine learning pipeline consists of the following steps: (i) variables selection process; and (ii) development and evaluation of prediction models. The selection of features is realized based on the Pearson Product-Moment Correlation Coefficient (PCC) whereas various prediction models, such as Regression models and Neural Networks are applied and compared with respect to the forecast of the price of the daily charter rate for an LNG tanker with a capacity of 145,000 CBM (LNG 145K CBM Spot Rate).

2.2.1 Variables selection

In this stage the correlation between the prediction variable, LNG 145K CBM Spot Rate, and the other independent variables, which are described in Table 1, is calculated based on Pearson Product-Moment Correlation Coefficient, $\rho_{X,Y}$, [32,33] for a time horizon of 2, 4 and 6 months. PCC is used to identify the variables with high linear correlation with respect to the selected decision variable in order to be used for the development of the prediction models.

Let X and Y be two zero-mean real-valued random variables. The PCC is defined as:

$$\rho_{X,Y} = \frac{cov(X,Y)}{\sigma_X \sigma_Y} \qquad (1)$$

Where $cov(X, Y)$ is the covariance of the two variables X and Y, and σ_X, σ_Y are the standard deviation of X and Y, respectively. The covariance $cov(X, Y)$ is given by:

$$cov(X, Y) = E[(X - \mu_X)(Y - \mu_Y)] \tag{2}$$

Where μ_X, μ_Y are the mean of X and Y, respectively. The values of the PCC ranges: $-1 < \rho_{X,Y} < 1$.

2.2.2 Data regression

A typical form of a prediction/forecast task based on time series can be summarized as: Given a set of n samples $\{y(t_1), y(t_2), \ldots, y(t_n)\}$ in a time sequence $t_1, t_2, \ldots, t_n$, the goal is to predict the sample $y(t_m)$ at some future time t_m, with $m > n$. Forecasting or prediction has significant impact on the decision making in science, business or engineering.

Therefore, Machine learning (ML) has been widely applied to regression estimations in various domains. In contrast with the traditional statistical approaches, ML techniques extract prior knowledge by restricting the space of assumed dependencies without making any distributional assumptions [34]. Typical regression approaches include Moving average and ARIMA models [35]. Moving average is a classical method. A moving average of order n can be written as:

$$\hat{Y}_t = \frac{1}{n} \sum_{j=-k}^{k} y_{t+j} \tag{3}$$

Where $n = 2k + 1$. It expresses the estimation of the trend-cycle at time t, obtained by averaging values of the time series within k periods of t.

One of the most powerful ML algorithms is the Artificial Neural Networks (ANNs) [36]. Even if they are usually used for classification tasks, ANNs are also easily adaptable to regression models [37,38]. ANN consists of a collection of processing elements, such as neurons or nodes, which are fully or partially interconnected. Its architecture resembles to a directed graph where each node n performs a process described by a transfer or else activation function f_n:

$$y_n = f_n\left(\sum_{i=1}^{m} \omega_{ni} x_i + b_n\right) \tag{4}$$

where y_n is the output of the node, n, x_i is the ith input to the node, ω_{ni} is the connection weight between the n and i nodes, and b_n is the threshold or else bias of the node. The activation function is usually non-linear, such as the sigmoid, Heaviside or Gaussian functions. Through this process, a set of inputs is transformed to a set of desired outputs. To get the desired output, the weights are adjusted through the learning via examples. This process is called 'training' and the examples 'training set'. The learning rule is a weight-updating rule which determines how connection weights are changed. Popular learning rules include the Hebbian rule, the delta rule, the antiHebbian rule, and the competitive learning rule [39,40].

To address regression problems with time series data, the Recurrent Neural Networks (RNN) are gaining place. RNNs have been used for various application with time series data [41–43]. RNNs are neural networks for sequential data, such as time series. Their operation is based on the use of the input data combined with previous outputs for making a prediction. RNN models that present a high level performance constitute the long short-term memory (LSTM) and the gated recurrent unit (GRU) [44,45].

Elman networks and Jordan networks are popular simple RNNs (SRN) that are used in this study. Elman [46], as well as, Jordan [47] networks consist from 3 layers. Bellow, the mathematical formulations for the hidden and layer vector and output vector are given for both networks.

Elman network

$$h_t = \sigma_h(W_h x_t + U_h h_{t-1} + b_h) \tag{5}$$
$$y_t = \sigma_y(W_y h_t + b_y) \tag{6}$$

Jordan network

$$h_t = \sigma_h(W_h x_t + U_h y_{t-1} + b_h) \tag{7}$$
$$y_t = \sigma_y(W_y h_t + b_y) \tag{8}$$

Where x_t is the input vector, h_t is the hidden layer vector, y_t is the output vector, σ_h and σ_y are the activation functions and W_h, W_y, U_h, U_y, b_h and b_y are the parameter matrices and vectors, respectively.

Multilayer Perceptron (MLP), is the most common neural network which generates a nonlinear model for prediction based on supervised training procedures. MLP is a layered feedforward neural network where the information is transferred from the input layer unidirectionally to the output layer via the hidden layers. The weighting step consists of the following steps:

(i) the multiplication of features with the corresponding weight, $\{x_i w_i\}_{i=0}^n$, where $\{x_i\}_{i=0}^n$ is the input;

(ii) their sum:

$$z = \sum_{i=0}^n x_i w_i \tag{9}$$

where $\{w_i\}_{i=0}^n$ are the weights of the neuron; and

(iii) the transfer step where the output y is produced by the application of an activation function f to the sum, $y = f(z)$. Commonly used activation functions include linear or logistic (sigmoid) or the unit step (Heaviside) functions [48].

Time-Lag recurrent networks (TLRN) are MLPs with short term memory structures and local recurrent connections. The input layer uses the inputs delayed by multiple time points before presented to the network [49]. The memory structures are characterized by the Laguerre memory and delay operator:

$$L_i(z, u) = \sqrt{1 - (1-u)^2} \frac{\left(z^{-1} - (1-u)\right)^{i-1}}{(1 - (1-u)z^{-1})^i} \tag{10}$$

Where L_i is the ith Laguerre function in the z-domain, u is a free parameter that represents the memory resolution and z^{-1} is the delay operator [50].

Generalized Regression Neural Network (GRNN) consists of an improvement of Radial Basis Neural Network (Fig. 2). The advantages of the GRNN include strong nonlinear mapping ability and learning speed. GRNN can have very good prediction effect with small or unstable data. Although GRNN does not have radial basis accuracy, it can also perform well in classification and fitting, especially in cases where the data accuracy is poor [51–53]. The prediction value $Y(x)$ of input x is given from:

$$Y(x) = \frac{\sum_{i=1}^N w_i K(x, x_i)}{\sum_{i=1}^N K(x, x_i)} \tag{11}$$

where w_i is the activation function for the pattern layer neuron at i and $K(x, x_i)$ is the Radial basis function kernel, such as the Gaussian kernel:

$$K(x, x_i) = e^{-d_k/2\sigma^2} \tag{12}$$

where $d_k = (x - x_i)^T (x - x_i)$ is the squared Euclidean distance between the training samples x_i and the input x.

A self-organizing feature map (SOFM) is a type of ANN that uses unsupervised learning in the training process to produce a map with reduced dimensionality compared to the input. Map is a low-dimensional, discretized representation of the input space of the training samples. SOFM apply competitive learning by using a neighbourhood function to preserve the topological properties of the input space. Common ANNs use error-correction learning, such as back propagation with gradient descent [54].
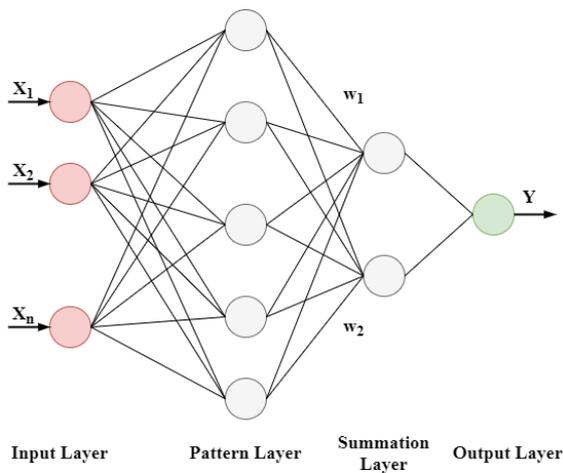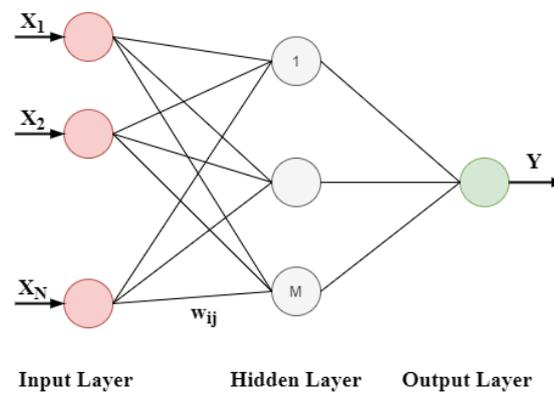
Figure 2. General structure of GRNN   Figure 3. General structure of GFNN

The generalized feedforward neural network (GFNN) architecture follows the one of the feedforward neural networks (Fig. 3). The Feed forward neural networks consist of at least 3 layers, such as the input layer, the hidden layer and the output layer. GFNN uses a generalized shunting neuron (GSN) model as the basic computing unit. The ability of the GSNs to form complex and nonlinear decision boundaries make GFNN suitable for regression and classification [53,55]. In GFNN the activity of the neurons can be described by the non-linear expression:

$$\frac{dx_i}{dt} = I_i - a_i x_i - f\left(\sum_j w_{ij} x_j\right) x_i + b_i \tag{13}$$

where x_i is the activity of the ith neuron, I_i is the input to the ith neuron, a_i is a positive constant that represents the passive decay rate of the neuron, w_{ij} is the weight from the jth input to the ith neuron, b_i is the bias and f is the activation function.

3. Results

3.2 Evaluation Methodology

The proposed methodology was applied on the case study for the forecast of the LNG carrier charter rates and it was implemented using the time series data presented in Section 2. Time series data from 01/01/2010 until 01/01/2017 were used. The aim is to predict the desired prediction variable, LNG 145K CBM Spot Rate, for the following dates: (i) 01/03/2017; (ii) 01/05/2017; and 01/07/2017. In Table 2 these values are presented.

Table 2. Real values for the time series LNG 145K CBM Spot Rate for the prediction dates.

LNG 145K CBM Spot Rate	Value ($/day)
01/03/2017	31,681
01/05/2017	34,768
01/07/2017	37,854

Initially the variables that will contribute to the development of the prediction models are identified based on the PCC analysis. Table 3 shows the values of the PCC and their interpretation with respect to the correlation of the variables in this study.

Table 3. PCC values and correlation

Value	Correlation
1	total positive linear correlation

−1	total negative linear correlation
> |0.2|	no linear correlation
|0.2 – 0.3|	low-medium linear correlation
|0.3 – 0.5|	medium linear correlation
|0.5 – 0.6|	medium-high linear correlation
< |0.6|	high linear correlation

The variables that are round zero are considered as variables with no or low linear correlation with the decision variable and, thus, they will be excluded from our analysis.

Following the variable selection, the training of the prediction models was performed. To evaluate the performance of the models the mean squared error (MSE) was used:

$$MSE = \frac{\sum_{j=0}^{P} \sum_{i=0}^{N}(d_{ij} - y_{ij})^2}{N \cdot P} \tag{9}$$

Where P is the number of output process elements, N is the number of iterations, y_{ij} is the output of ith iteration in the process element j and d_{ij} is the desired output for the ith iteration in process element j.

In Table 4 the ANN models with their parameter settings, which are used in this study, are presented.

Table 4. Summarized table of the parameters and the data used in the study.

Neural Networks	
Name	**Number of Parameters**
Multilayer Perceptron (MLP)	4 (Hidden Layers)
Generalized Feed Forward (GFFN)	4 (Hidden Layers)
Modular Neural Network (MNN)	4 (Types)
Jordan/Elman Network	4 (Types)
General Regression Neural Network (GRNN)	4 (Hidden Layers)
Self-Organizing Feature Map Network (SOFM)	4 (Hidden Layers)
Time-Lag Recurrent Network (TLRN)	4 (Hidden Layers)
Variables for Prediction	
Name	**Prediction Time**
LNG 145K CBM Spot Rate	2 Months
LNG 145K CBM Spot Rate	4 months
LNG 145K CBM Spot Rate	6 months
Data allocation	
Name	**Percentage**
Training data	75 %
Testing data	10 %
Cross Validation data	15 %
Number of Epochs	
1000	

3.2 Results

From the PCC analysis we get the following results, shown in Table 5. From the results, we observe that some variables show low to moderate correlation or even not at all. That is, they are unrelated to the predictor variable. Therefore, these variables were excluded from our study. Specifically, the variables World Seaborne LNG Trade, Price of Liquefied U.S. Natural Gas Exports, Henry Hub Natural Gas Spot Price and New Orders Placed that will be discarded due to no or low linear correlation with the prediction variable, LNG 145K CBM Spot Rate, as it is shown in the Table 6.

Table 5. Correlation coefficients between the prediction variable and the other variables for 2, 4 and 6 months.

Name of Variable	LNG 145K CBM Spot Rate *2 Months*	LNG 145K CBM Spot Rate *4 Months*	LNG 145K CBM Spot Rate *6 Months*
LNG 145K CBM Spot Rate	0.987	0.964	0.933
LNG 160K CBM Spot Rate	0.964	0.929	0.889
World Seaborne LNG Trade (million tonnes)	-0.676	-0.690	-0.699
World Seaborne LNG Trade (billion tonne-miles)	*-0.115*	*-0.172*	*-0.222*
LNG 160K CBM 1 Year Timecharter Rate	0.933	0.897	0.855
Price of Liquefied U.S. Natural Gas Exports	*0.217*	*0.206*	*0.193*
Henry Hub Natural Gas Spot Price	*0.221*	*0.248*	*0.272*
Import LNG Japan Price	0.789	0.742	0.700
WTI Oil price	0.666	0.626	0.593
Brent Oil Price	0.729	0.683	0.639
Total LNG Fleet	-0.915	-0.929	-0.935
Total Shipping Capacity	-0.901	-0.915	-0.921
Operational Capacity	-0.909	-0.932	-0.945
New Orders Placed	*0.294*	*0.265*	*0.243*
Orderbook	-0.726	-0.776	-0.816
Ships Delivered That Year	-0.934	-0.925	-0.906
Liquefaction Plants/Liquefaction	-0.815	-0.827	-0.834
Liquefaction Plants/Storage	-0.885	-0.906	-0.917
Regasification Plants/Storage	-0.838	-0.871	-0.895
Regasification Plants/Sent Out	-0.839	-0.874	-0.901
Spot LNG Imports	-0.700	-0.746	-0.783
Worldwide Natural Gas Production	-0.821	-0.858	-0.886
Worldwide Natural Gas Consumption	-0.804	-0.837	-0.863

Table 6. Variables with no or low correlation with the prediction variable.

Name of Variable	LNG 145K CBM Spot Rate *2 Months*	LNG 145K CBM Spot Rate *4 Months*	LNG 145K CBM Spot Rate *6 Months*
World Seaborne LNG Trade (billion tonne-miles)	no correlation	no correlation	low-medium correlation
Price of Liquefied U.S. Natural Gas Exports	low-medium correlation	low-medium correlation	no correlation
Henry Hub Natural Gas Spot Price	low-medium correlation	low-medium correlation	low-medium correlation
New Orders Placed	low-medium correlation	low-medium correlation	low-medium correlation

The results show that all the rate (spot and time-charter) variables has very high positive correlation with the desired output. However, high negative correlation was observed between the desired decision variable and the variables that show the total shipping capacity (Total LNG fleet, Total Shipping Capacity, Operational Capacity etc). To this end, these variables were excluded from the training of the predictive models.

The rest variables were used to train the prediction models for the 2 months, 4 months and 6 months forecast of the LNG 145K m^3 spot rate. The Tables 7-9 show the results for the trained models.

Table 7. Results of forecast for LNG 145K CBM Spot Rate 2 months.

Neural Model	Type	LNG 145K CBM Spot Rate 2 months	
		Mean squared error (MSE)	
		Training	*Cross Validation*
Multilayer Perceptron	1 Hidden Layer	2.04177E-06	0.000126952
Multilayer Perceptron	2 Hidden Layers	1.02071E-06	0.000271574
Multilayer Perceptron	3 Hidden Layers	8.30945E-08	0.000249368
Multilayer Perceptron	4 hidden Layers	4.04827E-06	0.00019093
Generalized Feed Forward	1 Hidden Layer	8.13606E-05	0.0006953
Generalized Feed Forward	2 Hidden Layers	8.51169E-07	0.000492954
Generalized Feed Forward	3 Hidden Layers	1.22216E-25	0.000114536
Generalized Feed Forward	4 Hidden Layers	4.80591E-16	**9.0676E-05**
Modular Neural Network	Type 1	1.57247E-10	0.000137449
Modular Neural Network	Type 2	1.34044E-08	0.000268479
Modular Neural Network	Type 3	7.46558E-09	0.000295709
Modular Neural Network	Type 4	1.12069E-09	0.000300228
Jordan/Elman Network	Type 1	7.34076E-06	0.000284278
Jordan/Elman Network	Type 2	0.000400029	0.00022034
Jordan/Elman Network	Type 3	6.49497E-05	0.000267897
Jordan/Elman Network	Type 4	4.18497E-06	0.000191801
Generalized Regression Neural Network	1 Hidden Layer	8.96735E-07	0.000241314
Generalized Regression Neural Network	2 Hidden Layers	1.21131E-09	0.000150232
Generalized Regression Neural Network	3 Hidden Layers	3.44779E-09	0.000230269
Generalized Regression Neural Network	4 Hidden Layers	1.76621E-07	9.18787E-05
Self-Organized Feature Map Network	1 Hidden Layer	5.71591E-13	0.001645705
Self-Organized Feature Map Network	2 Hidden Layers	4.08174E-29	0.003566318
Self-Organized Feature Map Network	3 Hidden Layers	4.60911E-11	0.000205648
Self-Organized Feature Map Network	4 Hidden Layers	1.16882E-09	0.131393527
Time-Lag Recurrent Network	1 Hidden Layers	0.002074106	0.00163206
Time-Lag Recurrent Network	2 Hidden Layers	0.017725692	0.00023123
Time-Lag Recurrent Network	3 Hidden Layers	0.010047504	0.000537872
Time-Lag Recurrent Network	4 Hidden Layers	0.014389702	0.000238308

Table 8. Results of forecast for LNG 145K CBM Spot Rate 4 months.

Neural Model	Type	LNG 145K CBM Spot Rate 4 months	
		Mean squared error (MSE)	
		Training	*Cross Validation*
Multilayer Perceptron	1 Hidden Layer	3.741E-06	0.000372187
Multilayer Perceptron	2 Hidden Layers	1.42506E-07	0.000427877
Multilayer Perceptron	3 Hidden Layers	4.28696E-08	**8.09547E-05**
Multilayer Perceptron	4 Hidden Layers	1.99132E-08	0.000233475
Generalized Feed Forward	1 Hidden Layer	6.39153E-06	0.000101767
Generalized Feed Forward	2 Hidden Layers	2.82904E-11	0.00024182
Generalized Feed Forward	3 Hidden Layers	2.39175E-25	0.000120044
Generalized Feed Forward	4 Hidden Layers	4.44123E-26	9.26663E-05

Modular Neural Network	Type 1	8.50244E-09	0.000371133
Modular Neural Network	Type 2	0.000458156	0.001759781
Modular Neural Network	Type 3	1.78094E-18	0.000283729
Modular Neural Network	Type 4	9.76073E-10	0.000127877
Jordan/Elman Network	Type 1	1.1566E-05	0.000118301
Jordan/Elman Network	Type 2	2.96119E-05	0.000151788
Jordan/Elman Network	Type 3	0.000364442	0.000533405
Jordan/Elman Network	Type 4	1.95111E-05	0.000620016
Generalized Regression Neural Network	1 Hidden Layer	1.36243E-06	0.000134973
Generalized Regression Neural Network	2 Hidden Layers	7.19216E-10	0.00033396
Generalized Regression Neural Network	3 Hidden Layers	9.0162E-08	0.000203275
Generalized Regression Neural Network	4 Hidden Layers	2.48142E-08	0.000530602
Self-Organized Feature Map Network	1 Hidden Layer	9.06784E-27	0.167926899
Self-Organized Feature Map Network	2 Hidden Layers	2.23852E-06	0.001855077
Self-Organized Feature Map Network	3 Hidden Layers	8.6554E-27	0.059753386
Self-Organized Feature Map Network	4 Hidden Layers	1.16815E-27	0.088058387
Time-Lag Recurrent Network	1 Hidden Layer	0.021580527	0.000480302
Time-Lag Recurrent Network	2 Hidden Layers	0.00679889	0.00044482
Time-Lag Recurrent Network	3 Hidden Layers	0.017249335	0.001708667
Time-Lag Recurrent Network	4 Hidden Layers	0.019813585	0.00067057

Table 9. Results of forecast for LNG 145K CBM Spot Rate 6 months.

		LNG 145K CBM Spot Rate 6 months	
		Mean squared error (MSE)	
Neural Model	**Type**	*Training*	*Cross Validation*
Multilayer Perceptron	1 Hidden Layer	2.3057E-06	0.000274236
Multilayer Perceptron	2 Hidden Layers	3.57161E-07	4.2437E-05
Multilayer Perceptron	3 Hidden Layers	1.28371E-06	0.000494272
Multilayer Perceptron	4 Hidden Layers	2.80881E-06	0.00072802
Generalized Feed Forward	1 Hidden Layer	4.12988E-06	0.000340456
Generalized Feed Forward	2 Hidden Layers	2.9611E-26	0.000159281
Generalized Feed Forward	3 Hidden Layers	1.47921E-29	0.004853003
Generalized Feed Forward	4 Hidden Layers	3.96904E-27	8.58066E-05
Modular Neural Network	Type 1	5.34594E-08	0.00027619
Modular Neural Network	Type 2	2.00206E-16	0.002561659
Modular Neural Network	Type 3	8.28519E-10	0.000444795
Modular Neural Network	Type 4	8.34674E-07	0.000759397
Jordan/Elman Network	Type 1	1.02852E-05	0.000557087
Jordan/Elman Network	Type 2	7.33229E-05	0.000547481
Jordan/Elman Network	Type 3	0.000305968	0.000983395
Jordan/Elman Network	Type 4	5.30217E-06	0.000605895
Generalized Regression Neural Network	1 Hidden Layer	3.00376E-07	7.78351E-05
Generalized Regression Neural Network	2 Hidden Layers	2.59021E-08	0.001267711
Generalized Regression Neural Network	3 Hidden Layers	1.23496E-19	**3.51658E-05**
Generalized Regression Neural Network	4 Hidden Layers	0.056960491	0.008906992
Self-Organized Feature Map Network	1 Hidden Layer	3.77435E-15	0.003702385
Self-Organized Feature Map Network	2 Hidden Layers	3.99304E-27	0.000916829

Self-Organized Feature Map Network	3 Hidden Layers	1.50628E-23	0.077536511
Self-Organized Feature Map Network	4 Hidden Layers	2.97364E-15	0.008736496
Time-Lag Recurrent Network	1 Hidden Layer	0.037065748	0.000772208
Time-Lag Recurrent Network	2 Hidden Layers	0.012002993	0.000210496
Time-Lag Recurrent Network	3 Hidden Layers	0.014963015	0.000778788
Time-Lag Recurrent Network	4 Hidden Layers	0.028511914	0.000755658

4. Discussion

The following neural network models have been applied to forecast the LNG 145K CBM Spot Rate for 2 moths (01/03/2017), 4 months (01/05/2017) and 6 months (01/07/2017):

- Multilayer perceptron (MLP)
- Generalized feedforward (GFFN)
- Modular (programming)
- Jordan/Elman
- General Regression Neural Network (GRNN)
- Self-organizing map (SOM)
- Time-lag recurrent network (TLRN)

The results from Table 7 show that the GFFN architecture, with 4 hidden layers, results to a better forecast (MSE 9.0676E-05) regarding a very short forecast (2 months), however the GRNN with 4 hidden layers presented a competitive performance (MSE 9.18787E-05). Regarding, the short-term prediction of 4 moths shown in Table 8, the MLP with 3 hidden layers achieved the best performance (MSE 8.09547E-05). However, the GRNN with 4 hidden layers (MSE 9.26663E-05). In case of mid-term forecast (6 months), as it is shown in Table 9, GRNN networks reached the best performance with 3 hidden layers (MSE 3.51658E-05). However, MLP (MSE 4.2437E-05) and GFFN (MSE 8.58066E-05) networks presented competitive performance.

Overall, we can say that the GRNN architecture presented a more stable performance with respect to the forecast period in all cases. Thus, a more accurate prediction model for the freight price of LNG carriers can be built by using GRNN networks.

5. Conclusions

In this study, a machine learning pipeline was presented to forecast the freight price of the LNG carriers. The proposed methodology covers the gap in the literature regarding LNG carriers to facilitate the decision-making of maritime industries. Specifically, the study focused on the prediction of the price of the daily fare for an LNG tanker with a capacity of 145,000 CBM and a steam turbine vessel. The methodology incorporates: (i) the collection of data relevant to LNG to form a dataset; (ii) the identification of the variables of the dataset that significantly contribute to the accurate prediction of the selected decision variable; (iii) the development of prediction machine learning models; and (iv) the evaluation of the models' performance with respect to the mean square error.

Overall, data from 01/01/2010 until 01/01/2017 were collected for 23 variables. The PCC showed that there was no linear correlation among the decision variable and the variables relevant to trades, contracts trades and new orders. Hence, from the 23 variables only 4 were excluded from the analysis. The formed dataset was used to train various machine learning models, such as Multilayer MLP, GFFN, Jordan/Elman, GRNN, SOM and TLRN. The results showed that GFFN with 4 hidden layers had the best performance for 2 moths forecast, MLP with 3 hidden layers had the best performance for 4 months forecast and GRNN with 4 hidden layers had the best performance for 6 months forecast. However, GRNN presented a stable and comparative performance in all cases. Therefore, a GRNN architecture can be considered as a suitable machine learning approach to develop a forecast model for the freight price of LNG carriers.

Future work includes further investigation of various types of machine learning models and extension of the work focusing on other types of LNG tankers, such as LNG tanker with a capacity of 160,000 or 174,000 CBM. Finally, future work can include updated datasets, where further development of the spot LNG shipping market will be reflected.

Institutional Review Board Statement: Not applicable

Informed Consent Statement: Not applicable
Data Availability Statement: Not applicable

Conflicts of Interest: The authors declare no conflict of interest

References

1. Awoyomi, A.; Patchigolla, K.; Anthony, E.J. Process and Economic Evaluation of an Onboard Capture System for LNG-Fueled CO2 Carriers. *Industrial and Engineering Chemistry Research* **2020**, doi:10.1021/acs.iecr.9b04659.

2. Chu Van, T.; Ramirez, J.; Rainey, T.; Ristovski, Z.; Brown, R.J. Global Impacts of Recent IMO Regulations on Marine Fuel Oil Refining Processes and Ship Emissions. *Transportation Research Part D: Transport and Environment* **2019**, doi:10.1016/j.trd.2019.04.001.

3. Seithe, G.J.; Bonou, A.; Giannopoulos, D.; Georgopoulou, C.A.; Founti, M. Maritime Transport in a Life Cycle Perspective: How Fuels, Vessel Types, and Operational Profiles Influence Energy Demand and Greenhouse Gas Emissions. *Energies* **2020**, doi:10.3390/en13112739.

4. Pavlenko, N.; Comer, B.; Zhou, Y.; Clark, N.; Rutherford, D. The Climate Implications of Using LNG as a Marine Fuel. *ICCT Working Paper 2020-02* **2020**.

5. International Gas Union *WORLD LNG REPORT*; 2020;

6. International Energy Agency *Global Gas Security Review 2019*; 2019;

7. GIIGNL - International Group of Liquefied Natural Gas Importers *GIIGNL Annual Report - The LNG Industry*; 2020;

8. Atsalakis, G.S. Using Computational Intelligence to Forecast Carbon Prices. *Applied Soft Computing Journal* **2016**, doi:10.1016/j.asoc.2016.02.029.

9. Dario, U. Forecasting Energy Market: An Artificial Neural Network Approach, Universita Ca'Foscari Venezia, 2015.

10. Wu, Y.H.; Hua, J.; Chen, H.L. Economic Feasibility of an Alternative Fuel for Sustainable Short Sea Shipping: Case of Cross-Taiwan Strait Transport. In Proceedings of the Proceedings of the World Congress on New Technologies; 2018.

11. Le Fevre, C.N. A Review of Demand Prospects for LNG as a Marine Fuel. **2018**.

12. Merien-Paul, R.H.; Enshaei, H.; Jayasinghe, S.G. Guessing to Prediction-a Conceptual Framework to Predict LNG Bunker Demand Profile in Australia. In Proceedings of the IAMU AGA 17-Working together: the key way to enhance the quality of maritime education, training and research; 2016; pp. 244–252.

13. Aronietis, R.; Sys, C.; Van Hassel, E.; Vanelslander, T. Forecasting Port-Level Demand for LNG as a Ship Fuel: The Case of the Port of Antwerp. *Journal of Shipping and Trade* **2016**, *1*, 2.

14. Arnet, N.M.L. LNG Bunkering Operations: Establish Probabilistic Safety Distances for LNG Bunkering Operations., Institutt for energi-og prosessteknikk, 2014.

15. Jin, J.; Kim, J. Forecasting Natural Gas Prices Using Wavelets, Time Series, and Artificial Neural Networks. *PLoS ONE* **2015**, doi:10.1371/journal.pone.0142064.

16. Al-Fattah, S.M.; Startzman, R.A. Predicting Natural Gas Production Using Artificial Neural Network.; 2001.

17. Su, M.; Zhang, Z.; Zhu, Y.; Zha, D. Data-Driven Natural Gas Spot Price Forecasting with Least Squares Regression Boosting Algorithm. *Energies* **2019**, doi:10.3390/en12061094.

18.   Busse, S.; Helmholz, P.; Weinmann, M. Forecasting Day Ahead Spot Price Movements of Natural Gas - An Analysis of Potential Influence Factors on Basis of a NARX Neural Network. In Proceedings of the Multikonferenz Wirtschaftsinformatik 2012 - Tagungsband der MKWI 2012; 2012.

19.   Hosseinipoor, S. Forecasting Natural Gas Prices in the United States Using Artificial Neural Networks, UNIVERSITY OF OK-LAHOMA, 2016.

20.   Akpinar, M.; Adak, M.F.; Yumusak, N. Day-Ahead Natural Gas Demand Forecasting Using Optimized ABC-Based Neural Network with Sliding Window Technique: The Case Study of Regional Basis in Turkey. *Energies* **2017**, doi:10.3390/en10060781.

21.   Moraes, L.A.M.; Faria, L.F.T. A Stochastic Programming Approach to Liquified Natural Gas Planning. *Pesquisa Operacional* **2016**, doi:10.1590/0101-7438.2016.036.01.0151.

22.   Engelen, S.; Norouzzadeh, P.; Dullaert, W.; Rahmani, B. Multifractal Features of Spot Rates in the Liquid Petroleum Gas Shipping Market. *Energy Economics* **2011**, doi:10.1016/j.eneco.2010.05.009.

23.   Giannakopoulou, P.; Chountas, P. Forecasting the Spot Price of P1A Shipping Route. In Proceedings of the 2019 Big Data, Knowledge and Control Systems Engineering, BdKCSE 2019; 2019.

24.   von Spreckelsen, C.; von Mettenheim, H.J.; Breitner, M.H. Short-Term Trading Performance of Spot Freight Rates and Derivatives in the Tanker Shipping Market: Do Neural Networks Provide Suitable Results? In Proceedings of the Communications in Computer and Information Science; 2012.

25.   von Spreckelsen, C.; von Mettenheim, H.-J.; Breitner, M.H. Spot and Freight Rate Futures in the Tanker Shipping Market: Short-Term Forecasting with Linear and Non-Linear Methods. In; 2014.

26.   Lyridis, D. V.; Manos, N.D.; Zacharioudakis, P.G. Modeling the Dry Bulk Shipping Market Using Macroeconomic Factors in Addition to Shipping Market Parameters via Artificial Neural Networks. In Proceedings of the International Journal of Transport Economics; 2014.

27.   Yin, J.; Luo, M.; Fan, L. Dynamics and Interactions between Spot and Forward Freights in the Dry Bulk Shipping Market. *Maritime Policy and Management* **2017**, doi:10.1080/03088839.2016.1253884.

28.   Yip, T.L. Predicting the Shipping Market by Spreads of Timecharter Rates. *Zeszyty Naukowe Akademii Morskiej w Szczecinie* **2018**.

29.   Du, Q. Forecasting and Backtesting of <I>VaR</I> in International Dry Bulk Shipping Market under Skewed Distributions. *American Journal of Industrial and Business Management* **2019**, doi:10.4236/ajibm.2019.95079.

30.   Cao, J.; Li, Z.; Li, J. Financial Time Series Forecasting Model Based on CEEMDAN and LSTM. *Physica A: Statistical Mechanics and its Applications* **2019**, doi:10.1016/j.physa.2018.11.061.

31.   Sagheer, A.; Kotb, M. Time Series Forecasting of Petroleum Production Using Deep LSTM Recurrent Networks. *Neurocomputing* **2019**, doi:10.1016/j.neucom.2018.09.082.

32.   Schober, P.; Schwarte, L.A. Correlation Coefficients: Appropriate Use and Interpretation. *Anesthesia and Analgesia* **2018**, doi:10.1213/ANE.0000000000002864.

33.   CHEN, P.Y.; POPOVICH, P.M. CORRELATION: PARAMETRIC AND NONPARAMETRIC MEASURES. In *Correlation*; 2011.

34.   Herbrich, R.; Graepel, T.; Obermayer, K. Regression Models for Ordinal Data⊚: A Machine Learning Approach. *Technical Report, TU Berlin. TR-99/03* **1999**.

35.   Madsen, H. *Time Series Analysis*; 2007; ISBN 9781420059687.

36.   Wang, K.; Li, K.; Zhou, L.; Hu, Y.; Cheng, Z.; Liu, J.; Chen, C. Multiple Convolutional Neural Networks for Multivariate Time Series Prediction. *Neurocomputing* **2019**, doi:10.1016/j.neucom.2019.05.023.

37.   Lathuiliere, S.; Mesejo, P.; Alameda-Pineda, X.; Horaud, R. A Comprehensive Analysis of Deep Regression. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **2020**, doi:10.1109/TPAMI.2019.2910523.

38.   Cheng, J.; Wang, Z.; Pollastri, G. A Neural Network Approach to Ordinal Regression. In Proceedings of the Proceedings of the International Joint Conference on Neural Networks; 2008.

39.   Hassoun, M.H. Fundamentals of Artificial Neural Networks. *Proceedings of the IEEE* **2005**, doi:10.1109/jproc.1996.503146.

40.   Buscema, P.M.; Massini, G.; Breda, M.; Lodwick, W.A.; Newman, F.; Asadi-Zeydabadi, M. Artificial Neural Networks. In *Studies in Systems, Decision and Control*; 2018.

41.   Laptev, N.; Yosinski, J.; Li, E.L.; Smyl, S. Time-Series Extreme Event Forecasting with Neural Networks at Uber. In Proceedings of the ICML 2017 Time Series Workshop; 2017.

42.   Hsu, D. Time Series Forecasting Based on Augmented Long Short-Term Memory. *arXiv preprint arXiv: 1707.00666 [cs.NE]* **2017**.

43.   Che, Z.; Purushotham, S.; Cho, K.; Sontag, D.; Liu, Y. Recurrent Neural Networks for Multivariate Time Series with Missing Values. *Scientific Reports* **2018**, doi:10.1038/s41598-018-24271-9.

44.   Caterini, A.L.; Chang, D.E. Recurrent Neural Networks. In *SpringerBriefs in Computer Science*; 2018.

45.   Hewamalage, H.; Bergmeir, C.; Bandara, K. Recurrent Neural Networks for Time Series Forecasting: Current Status and Future Directions. *International Journal of Forecasting* **2020**, doi:10.1016/j.ijforecast.2020.06.008.

46.   Elman, J.L. Finding Structure in Time. *Cognitive Science* **1990**, doi:10.1016/0364-0213(90)90002-E.

47.   Jordan, M.I. Chapter 25 Serial Order: A Parallel Distributed Processing Approach. In *Advances in Psychology*; 1997.

48.   Taud, H.; Mas, J.F. Multilayer Perceptron (MLP). In; 2018.

49.   Zhou, Y.; Xia, J.; Shen, H.; Zhou, J.; Wang, Z. Extended Dissipative Learning of Time-Delay Recurrent Neural Networks. *Journal of the Franklin Institute* **2019**, doi:10.1016/j.jfranklin.2019.08.003.

50.   Dayhoff, J.; Omidvar, O. Locally Recurrent Networks: The Gamma Operator, Properties, and Extensions. In *Neural Networks and Pattern Recognition*; 1998.

51.   Cigizoglu, H.K. Generalized Regression Neural Network in Monthly Flow Forecasting. *Civil Engineering and Environmental Systems* **2005**, doi:10.1080/10286600500126256.

52.   Specht, D.F. A General Regression Neural Network. *IEEE Transactions on Neural Networks* **1991**, doi:10.1109/72.97934.

53.   Firat, M.; Gungor, M. Generalized Regression Neural Networks and Feed Forward Neural Networks for Prediction of Scour Depth around Bridge Piers. *Advances in Engineering Software* **2009**, doi:10.1016/j.advengsoft.2008.12.001.

54.   Nikoo, M.; Sadowski, Ł.; Khademi, F.; Nikoo, M. Determination of Damage in Reinforced Concrete Frames with Shear Walls Using Self-Organizing Feature Map. *Applied Computational Intelligence and Soft Computing* **2017**, doi:10.1155/2017/3508189.

55.   Arulampalam, G.; Bouzerdoum, A. A Generalized Feedforward Neural Network Architecture for Classification and Regression. In Proceedings of the Neural Networks; 2003.