Artificial intelligence for supporting forecasting in maritime sector

F. Briatore*, R. Revetria*

*Mechanical, Industrial and Transport Engineer Department (D.I.M.E.), University of Genoa, Genoa, Ge 16126. Italy (federico.briatore@gmail.com, roberto.revetria@unige.it)

Abstract: The importance of the time series of data has always been of great relevance. A main use of them is the prediction of the future values of the quantities of interest. On this purpose, a lot of models have been created so far, as AR, MA, ARMA, ARMAX, ARIMA and so on. In the last years, the interest on Artificial Intelligence and Neural Network has grown a lot and a lot of studies were conducted to enable their use in different fields. This paper has the aim to show the possibility to use a system based on Artificial Intelligence to analyze the time series of index and future on the chartering of ships in order to predict the future values of them. The Neural Network is trained with the data of the last 3 years and the results obtained have be compared with those coming from ARIMA model and Carbon Copy model. The first aim of this paper is thus showing if the Neural Network performs better than the other 2 models and what day (first, third or fifth) is the best for the prevision made. The second purpose of this paper is establishing if the knowledge of the trend of the quantity value influences the results. The Neural Network has been trained both with a bullish trend and a bearish trend, then the results have been compared to prove if setting the right trend improve the quality of the prediction.

Keywords: Artificial Intelligence, Neural Network, Forecasting, Time Series, Maritime Sector

I. INTRODUCTION

Forecasting has a great value as it enables to estimate the future value of a phenomenon. The sectors where it can be applied are very wide and different from each other, like economics, business intelligence and industrial applications [6][9][10]. The base of forecasting are the time series, a sequence of measurements collected along time and chronologically ordered [1] that, through Big Data, are in rapid growth [6]. Time series are used in linear problems with regression methods and also in more advanced models like Autoregressive Models (AR), Moving Average Models (MA), Autoregressive Moving Average Models (ARMA) and Autoregressive Integrated Moving Average Models (ARIMA) [2]. As the computer computational capacity has increased in the last decades, Machine Learning can be widely implemented, like Support Vector Regression and Multi-Layer Perceptron (MLP) [6], with Artificial neural networks (ANN). Machine Learning with MLP has proved to provide better results than ARIMA, which underfits test data in long-term [6]. Artificial Intelligence (AI), using ANN, is able to give a great amount of knowledge to support managerial decision making [7], especially as a manager has a lot of tasks in order to achieve better performances of the company. AI can enhance the results of the leadership by providing better information to evaluate decisions and to assess the current used methods of decision making [7]. This kind of decisions involves all type of enterprise: human resources, finance, productivity and efficiency are only some examples. That can be applied not only to production businesses, but also in services. This paper studies the forecasting in the maritime sector for what concerns the value at 1, 3 and 5 days of ship freights. Financial investments have a time series nature and this is the reason why ANN applications have a great interest in this sector and, in this paper, a Neural Network is used for forecasting. One of the fields where AI is used in finance is in fact the forecasting of stock prices and stock market index values in order to support investments decisions [7].

II. NEURAL NETWORKS

Neural networks are inspired by the neural structures of intelligent organisms and its main property is the ability to learn. This kind of networks are formed by neuros, information processing units which are interconnected and able to recognize patterns in the data provided. A neuron receives an input from all the neighbouring neuros, elaborates it and the output is the input for the following neurons which are connected to it. The connections among neurons are weighted and the learning algorithm has the task of adjusting the weights [1]. In an ANN architecture, there are basically 3 kinds of layers: input, hidden and output. The first one provides the signal to hidden layers and doesn't process information. The hidden layers analyse the features or the input layer, that are then processed by the output layer [1]. The main skill of ANN is understanding and reproducing complex behaviour using the real data without any limiting hypothesis in the model. An indicator of the potential use of ANN in industrial field is the result got in telecommunication company, where the energy consumption estimation error was lower than 3% [9]. The neural network must be at first trained and then tested to evaluate the prediction power. In this paper, the time series has been cut in 2 files, the first days for the train and the last for the training. A special ANN architecture is the Deep Belief Network (DBN), where multiple layers of ANN are used following the concept named Restricted Boltzmann Machine (RBM). A BM is a stochastic network of neurons with 2 layers: visible, made of the collected data, and hidden, whose purpose is to learn features from the previous layer in order to represent a probabilistic distribution of the data [4]. The BM produces several outputs with its probability from an input, this creates a distribution [9]. For what concerns the data treated, they are binary ones [3]. RBM had initially problems with time series, one of this is the difficult to find dependencies between input parameters [6]. There are a lot of variants of the RBM, like the Conditional Restricted Boltzmann Machine (CRBM) [5], the Temporal Restricted Boltzmann machine (TRBM) [4], the Recurrent Temporal Restricted Boltzmann Machine (RTRBM) [4] and the Dynamic BM (DyBM). To overcome the limit of the binary data, a Gaussian distribution has been applied to the visible layer, creating a Gaussian-Bernoulli RBM (GBRBM) and a G-DyBM with components able to capture long term dependencies [8]. The G-DyBM has been then extended by adding a Recurrent Neural Network and the result is a RNN-G-DyBM [8]. The RNN-G-DyBM has been used to forecast the price of gasoline and diesel taking the time series of the weekly retail and then the results have been compared with AR model. RNN-G-DyBM outperforms AR more than 30% on the prediction of the first week. Moreover, the RNN-G-DyBM is better than G-DyBM by more than 21%. When multiple RBM are staked, a Deep Belief Network (DBN) is created. In this disposal, a hidden layer of a RBM is connected to the visible layer of the next one[1]. An approach used for Time Series Prediction is the hybrid one, which utilise a Feedforward Neural Network (FNN) that takes as input the hidden layer of the last RBM. In a paper [1] FNN and a GBRBM+FNN are trained and compared to forecast the future values of 3 datasets: Australia Energy Production, Dollar to Libra Conversion and North Atlantic Oscillation. For what concerns the first 2 datasets, the best model is the GBRBM+FNN [1]. Another approach used is combining RBM with MLP. In this combination, the RBM is used to find the initial weights for the MLP network [6]. A pre-training with RBM can improve the final performance, but only if the time series features aren't lagged [6].

III. Methodology

The aim of this study is to apply a ANN to the maritime sector to forecast the future value of BCI_5TC, Baltic Exchange Capesize Index, using the time-series of the last 3 years. Of this data, the last 350 have been used as test. The software used for this work is Attrasoft Predictor^{TM} (NN). The predictions considered are at 1, 3 and 5 days and have been compared to those coming from an ARIMA model, a Carbon Copy model (CC) (this one only for the first day), an Encorder-Decorder (EC) model and an Augmented Forecast (AF) model. The choice of using the forecast at 1, 3 and 5 days is due to the fact that only these values were available for each model. Calling t the last known value, t+1 is the first future day value predicted. The CC assigns at the t+1 day the value of the t-1 day. The first day predicted was 24/11/2021 and the last was 08/04/2022. The comparison was on the percentual error, the squared percentual error and the absolute percentual error committed in the 3 forecasting. The first one is the prediction value minus the real value divide for the real value, the second one is the previous value squared and the third one is the absolute value of the first. The data have been then proved to be distributed as a Gaussian. Once checked this behaviour, the values of the 3 kind of error have been averaged. These data are reported in TABLE I. 1 day prediction outstands the others. It also appears the difference between the percentual error in the 3 trends. It is important to underline that the neutral trend model failed the forecast for most of the days. So its analysis was postponed.

TABLE I
Average percentual error divided per trend

Trend	1 day	3 day	5 day
Bullish %	*-0,53%*	*5.89%*	*11.48%*
Bearish %	*6,19%*	*23.62%*	*50.36%*
Neutral %	*1,04%*	*12,93%*	*16,78%*
Bullish $\%^2$	*1.60%*	*11.01%*	*30.18%*
Bearish $\%^2$	*1.10%*	*14.49%*	*50.66%*
Neutral $\%^2$	*0,48%*	*7,99%*	*22,97%*
Bullish |%|	*8.74%*	*23.77%*	*37.05%*
Bearish |%|	*8.34%*	*30.96%*	*56.73%*
Neutral |%|	*5,22%*	*21,23%*	*39,92%*

To overcome the problem of the sign, there are the other 2 kind of error. One of the aims of this paper is in fact testing if the knowledge on the trend has effect on the final result. To do that, the 3 kind of errors made by ANN trained with a bullish, bearish and no trend have been compared with statistical methods: ANOVA and direct mean comparison to check the results. It was also used an estimate of the trend for the days to see if it was useful or not. To each day, it has been associated the trend of the last day of the test data set. That means that the wrong trend can be associated. This issue influences the utility of the information about the trend. This test was made for all the forecasts. This second analysis has been performed before the comparison of the models, to be able to state if there is a difference among the models. Once got the result, the main analysis has been performed using the same methods. In particular, the first ANOVA is with 2 ways in order to detect if the day forecasted has an influence. The models have been proved to have a gaussian distribution. Due to the Russia-Ukraine war, the volatility of the market increased, that could lead to a higher error in the forecast.

IV. RESULTS

A. Trend

The first result obtained concerns the trend. Looking at Table I, the difference in the averages suggested a significant importance of the knowledge of the trend.

Percentual error

The first error tested is the percentual one. The ANOVA is reported in Table II. SS are the Sum Squares, dof are the degrees of freedom, MS are the Mean Squares, FO is the Fisher value calculated and F(alfa) is the Fisher distribution value corresponding to the alfa chosen. The ANOVA shows that there is difference between the ANN trained with plus trend and the one with minus trend with a security level of 97,5%.

TABLE II
% ERROR ANOVA 1ST DAY TREND

Source	SS	dof	MS	FO	F 0,025
trend	*0,09*	*1*	*0,09*	*6,97*	*5,29*
error	*0,91*	*74*	*0,01*		
Tot	*1,00*	*75*			

TABLE III
% ERROR MEAN COMPARISON 1ST DAY TREND PLUS(P)/MINUS(M)

H0	Ha	Z	alfa Z
$\mu(p) = \mu(m)$	$\mu(p) \neq \mu(m)$	*2,68*	*0,007*
$\mu(p) = \mu(m)$	$\mu(m) > \mu(p)$	*2,68*	*0,004*

The mean comparison shows that the 2 means differs with a significance level of 99,3% and that the minus trend has a higher error with a confidence level of 99,6%. The same analysis was performed for the 3rd and 5th day.

TABLE IV
% ERROR ANOVA 3RD AND 5TH DAY TREND

Day	FO	F 0,05	F 0,025	F 0,01
3rd	*6,91*	*4*	*5,29*	*7*
5th	*12.28*	*4*	*5,29*	*7*

TABLE V
% ERROR MEAN COMPARISON 3RD/5TH DAY TREND PLUS(P)/MINUS(M)

Day	H0	Ha	Z	alfa Z
3rd	$\mu(p) = \mu(m)$	$\mu(p) \neq \mu(m)$	*2,66*	*0,008*
	$\mu(p) = \mu(m)$	$\mu(m) > \mu(p)$	*2,66*	*0,004*
5th	$\mu(p) = \mu(m)$	$\mu(p) \neq \mu(m)$	*3,54*	*0,0004*
	$\mu(p) = \mu(m)$	$\mu(m) > \mu(p)$	*3,54*	*0,0002*

The results are reported in Table IV, V. As it can be seen from the ANOVA, also for the 3rd and 5th days there are differences between the 2 trends and the plus performs better than the minus with a confidence level superior than 95%.

Squared percentual error

The results of the squared error are reported in the following tables. These data shows that there is no significant difference between the trends. The only values that statistically differ are at 5 days with a single tailed test, but only with a confidence level of 91%.

TABLE VI
$\%^2$ ERROR ANOVA TREND

Day	FO	F 0,1
1st	*0,53*	*2,79*
3rd	*0,62*	*2,79*
5th	*1,84*	*2,79*

TABLE VII
$\%^2$ ERROR MEAN COMPARISON 3RD/5TH DAY TREND PLUS(P)/MINUS(M)

Day	H0	Ha	Z	alfa Z
1st	$\mu(p) = \mu(m)$	$\mu(p) \neq \mu(m)$	*0,74*	*0,36*
	$\mu(p) = \mu(m)$	$\mu(p) > \mu(m)$	*0,74*	*0,18*
3rd	$\mu(p) = \mu(m)$	$\mu(p) \neq \mu(m)$	*0,79*	*0,43*
	$\mu(p) = \mu(m)$	$\mu(m) > \mu(p)$	*0,79*	*0,21*
5th	$\mu(p) = \mu(m)$	$\mu(p) \neq \mu(m)$	*1,37*	*0,17*
	$\mu(p) = \mu(m)$	$\mu(m) > \mu(p)$	*1,37*	*0,09*

Absolute percentual error

TABLE VIII
|%| ERROR ANOVA TREND

Day	FO	F 0,10	F 0,05	F 0,01
1st	*0,0003*	*2,79*	*4*	*7*
3rd	*2,16*	*2,79*	*4*	*7*
5th	*4,87*	*2,79*	*4*	*7*

TABLE IX
|%| ERROR MEAN COMPARISON 3RD/5TH DAY TREND PLUS(P)/MINUS(M)

Day	H0	Ha	Z	alfa Z
1st	$\mu(p) = \mu(m)$	$\mu(p) \neq \mu(m)$	*0,02*	*0,98*
	$\mu(p) = \mu(m)$	$\mu(m) > \mu(p)$	*0,02*	*0,49*
3rd	$\mu(p) = \mu(m)$	$\mu(p) \neq \mu(m)$	*1,49*	*0,14*
	$\mu(p) = \mu(m)$	$\mu(m) > \mu(p)$	*1,49*	*0,07*
5th	$\mu(p) = \mu(m)$	$\mu(p) \neq \mu(m)$	*2,23*	*0,026*
	$\mu(p) = \mu(m)$	$\mu(m) > \mu(p)$	*2,26*	*0,013*

The absolute error ANOVA is reported in Table VIII and in Tables IX the mean comparison. For the 1st day there isn't any statistical differences, while on the 3rd only with the 1 tailed test. At the contrary, on the forecast of the 5th day, the ANN trained with a plus type outstands the ANN with a minus type with a security level of more than 95%.

Neutral trend

TABLE X
% ERROR MEAN COMPARISON TREND PLUS(P)/NEUTRAL(N)

Day	H0	Ha	Z	alfa Z
1st %	$\mu(p) = \mu(n)$	$\mu(p) \neq \mu(n)$	*0,68*	*0,50*
	$\mu(p) = \mu(n)$	$\mu(n) > \mu(p)$	*0,68*	*0,25*
3rd %	$\mu(p) = \mu(n)$	$\mu(p) \neq \mu(n)$	*0,92*	*0,36*
	$\mu(p) = \mu(n)$	$\mu(n) > \mu(p)$	*0,92*	*0,18*
5th %	$\mu(p) = \mu(n)$	$\mu(p) \neq \mu(n)$	*0,34*	*0,73*
	$\mu(p) = \mu(n)$	$\mu(n) > \mu(p)$	*0,34*	*0,37*

TABLE XI
%^2 ERROR COMPARISON TREND PLUS(P)/ NEUTRAL(N)

Day	H0	Ha	Z	alfa Z
1st %	$\mu(p) = \mu(n)$	$\mu(p) \neq \mu(n)$	*1,90*	*0,06*
	$\mu(p) = \mu(n)$	$\mu(p) > \mu(n)$	*1,90*	*0,03*
3rd %	$\mu(p) = \mu(n)$	$\mu(p) \neq \mu(n)$	*0,60*	*0,55*
	$\mu(p) = \mu(n)$	$\mu(p) > \mu(n)$	*0,60*	*0,27*
5th %	$\mu(p) = \mu(n)$	$\mu(p) \neq \mu(n)$	*0,51*	*0,61*
	$\mu(p) = \mu(n)$	$\mu(p) > \mu(n)$	*0,51*	*0,31*

TABLE XII
|%| ERROR COMPARISON TREND PLUS(P)/ NEUTRAL(N)

Day	H0	Ha	Z	alfa Z
1st %	$\mu(p) = \mu(n)$	$\mu(p) \neq \mu(n)$	*1,74*	*0,08*
	$\mu(p) = \mu(n)$	$\mu(p) > \mu(n)$	*1,74*	*0,04*
3rd %	$\mu(p) = \mu(n)$	$\mu(p) \neq \mu(n)$	*0,39*	*0,70*
	$\mu(p) = \mu(n)$	$\mu(p) > \mu(n)$	*0,39*	*0,35*
5th %	$\mu(p) = \mu(n)$	$\mu(p) \neq \mu(n)$	*0,30*	*0,76*
	$\mu(p) = \mu(n)$	$\mu(n) > \mu(p)$	*0,30*	*0,38*

These data show how the only statistical difference is on the forecast of the first day, where the Neutral trend is better than the plus one on the basis of the squared and the absolute percentual error.

Taking into consideration the problem of the high rate of failure in the predictions, the comparison was made on the means of the plus trend with the neutral one.

Conclusion of the 1st analysis

Standing the previous results, the plus and minus trends are considered basically equal. Moreover, as the neutral trend often failed and the minus one appeared to be sometimes worse than the plus one, the models will be compared using the plus trend to train the ANN.

B. Trend estimate

An estimate of the trend has been used to detect if this kind of information was useful to reduce the error.

TABLE XIII
|%| ERROR ANOVA TREND

Day	FO	F 0,10
1st	*0,00003*	*2,79*
3rd	*0,01380*	*2,79*
5th	*0,05550*	*2,79*

TABLE XIV
|%| ERROR COMPARISON TREND RIGHT(R)/WRONG(W)

Day	H0	Ha	Z	alfa Z
1st %	$\mu(r) = \mu(w)$	$\mu(r) \neq \mu(w)$	*0,005*	*1*
	$\mu(r) = \mu(w)$	$\mu(w) > \mu(r)$	*0,005*	*0,5*
3rd %	$\mu(r) = \mu(w)$	$\mu(r) \neq \mu(w)$	*0,12*	*0,90*
	$\mu(r) = \mu(w)$	$\mu(w) > \mu(r)$	*0,12*	*0,45*
5th %	$\mu(r) = \mu(w)$	$\mu(r) \neq \mu(w)$	*0,24*	*0,81*
	$\mu(r) = \mu(w)$	$\mu(r) > \mu(w)$	*0,24*	*0,41*

The tests on these a data have been carried out on the absolute percentual error. As FO is lower than F(alfa=0,1), it means there is no difference in using the information on the estimate of the trend. This result is confirmed with the following direct comparisons of the means with a 2 tailed test and a 1 tailed test. This result, joined with the one of the previous subsection, has been used for the following step, that is to say the comparison among the different models.

C. Model comparison (older data)

The first comparison among models has been performed on the forecast of the 1st day between the Carbon Copy model (CC) and the ANN trained both with a plus trend. Firstly, the percentual errors made by CC have been tested to be gaussian. Then, the percentual error, the squared percentual error and the absolute percentual error have been compared with a direct mean comparison. Here the results:

TABLE XV
% ERROR COMPARISON 1ST DAY ANN PLUS(P) AND CC

H0	Ha	Z	alfa Z
$\mu(p) = \mu(CC)$	$\mu(p) \neq \mu(CC)$	*2,50*	*0,012*
$\mu(p) = \mu(CC)$	$\mu(p) > \mu(CC)$	*2,50*	*0,006*

TABLE XVI
%^2 ERROR COMPARISON 1ST DAY ANN PLUS(P)/MINUS(M) AND CC

H0	Ha	Z	alfa Z
$\mu(p) = \mu(CC)$	$\mu(p) \neq \mu(CC)$	*5,21*	*0,00004*
$\mu(p) = \mu(CC)$	$\mu(p) > \mu(CC)$	*5,21*	*0,00002*

TABLE XVII
|%| ERROR COMPARISON 1ST DAY ANN PLUS(P)/MINUS(M) AND CC

H0	Ha	Z	alfa Z
$\mu(p) = \mu(CC)$	$\mu(p) \neq \mu(CC)$	*0,30*	*0,76*
$\mu(p) = \mu(CC)$	$\mu(p) > \mu(CC)$	*0,30*	*0,38*

These results show how the CC copy model outstands the ANN for what concerns the percentual error and the squared percentual error. Related to the absolute percentual error, the models are statistically the same. As there is a statistical difference on 2 tests on 3, the test on newer data was carried out and the results are reported in the next subsection.

D. Model comparison (Newer data)

The last tests have been carried out on the last collected data on the basis of squared and absolute percentual error. They concern the days from 07/04/2022 to 26/04/2022. This choice was made as only for these ones the values forecasted by each model were available.

1st day

TABLE XVIII
%^2 ERROR METRICS OF 1ST DAY

	ARIMA	CC	E-D	AF	NN
μ %	*0,600*	*0,6200*	*4,33*	*19,23*	*0,94*
Var	*0,0001*	*0,0001*	*0,001*	*0,099*	*0,0001*

TABLE XIX
|%| ERROR METRICS OF 1ST DAY

	ARIMA	CC	E-D	AF	NN
μ %	*5,64*	*6,31*	*20,21*	*31,14*	*8,40*
Var	*0,003*	*0,002*	*0,003*	*0,104*	*0,003*

From data in Table XVIII and Table XIX, it appears clear that for the means AF and E-D are the worst. The 2 ways ANOVA analysis shows a difference among the models and not among the days, Table XX and XXI.

TABLE XX
%^2 ERROR ANOVA WITH 2 WAYS 1ST DAY MODEL

Source	FO	F 0,1	F 0,05	F 0,025	F 0,01
Model	*3,92*	*2,09*	*2,61*	*3,13*	*3,83*
Day	*1,03*	*1,76*			

TABLE XXI
|%| ERROR ANOVA WITH 2 WAYS 1ST DAY MODEL

Source	FO	F 0,1	F 0,05	F 0,025	F 0,01
Model	*7,19*	*2,09*	*2,61*	*3,13*	*3,83*
Day	*1,58*	*1,76*			

The day has no influence, while it can be stated with a security level of 99% that there is at least 1 model which performs differently then the others. To prove it, an ANOVA with 1 way was performed and the sum square of the days has been added to the sum square of the error. In Table XXII and XXIII it is possible to see that the security level of 99% for the models is confirmed.

TABLE XXII
%^2 ANOVA WITH 1 WAY 1ST DAY MODEL

Source	FO	F 0,1	F 0,05	F 0,025	F 0,01
Model	*3,89*	*2,09*	*2,61*	*3,13*	*3,83*

TABLE XXIII
|%| ANOVA WITH 1 WAY 1ST DAY MODEL

Source	FO	F 0,1	F 0,05	F 0,025	F 0,01
Model	*6,45*	*2,09*	*2,61*	*3,13*	*3,83*

At this point, the 3 best models, ARIMA, CC and NN have been compared in order to detect statistical differences among the means. However, as the dataset are made of only 12 days, the sample is small. That means that a t-Student is needed and that, before comparing the means, a test on the variances must be performed. As all the variances are statistically the same, it was possible to compare the means. The results are reported in Table XXIV and Table XXV. These results show how these 3 models doesn't differ for the mean.

TABLE XXIV
%^2 MEAN COMPARISON 1ST DAY MODEL

H0	Ha	t	t 0.1
μ(NN) = μ(ARIMA)	*μ(NN) ≠ μ(ARIMA)*	*0,789*	*1,717*
μ(NN) = μ(ARIMA)	*μ(NN) > μ(ARIMA)*	*0,789*	*1,321*
μ(NN) = μ(CC)	*μ(NN) ≠ μ(CC)*	*0,814*	*1,717*
μ(NN) = μ(CC)	*μ(NN) > μ(CC)*	*0,814*	*1,321*
μ(CC) = μ(ARIMA)	*μ(CC) ≠ μ(ARIMA)*	*0,041*	*1,717*
μ(CC) = μ(ARIMA)	*μ(CC) > μ(ARIMA)*	*0,041*	*1,321*

TABLE XXV
|%| MEAN COMPARISON 1ST DAY MODEL

H0	Ha	t	t 0.1
μ(NN) = μ(ARIMA)	*μ(NN) ≠ μ(ARIMA)*	*1,276*	*1,717*
μ(NN) = μ(ARIMA)	*μ(NN) > μ(ARIMA)*	*1,276*	*1,321*
μ(NN) = μ(CC)	*μ(NN) ≠ μ(CC)*	*1,035*	*1,717*
μ(NN) = μ(CC)	*μ(NN) > μ(CC)*	*1,035*	*1,321*
μ(CC) = μ(ARIMA)	*μ(CC) ≠ μ(ARIMA)*	*0,312*	*1,717*
μ(CC) = μ(ARIMA)	*μ(CC) > μ(ARIMA)*	*0,312*	*1,321*

3rd day

The same procedure was performed for the forecasts at 3 days. The metrics are reported in Table XXVI and Table XXVII and they show that the best 3 models are, in this case, ARIMA, AF and NN.

TABLE XXVI
$\%^2$ ERROR METRICS OF 3RD DAY

	ARIMA	E-D	AF	NN
μ %	*2,28*	*8,02*	*5,98*	*4,99*
Var	*0,002*	*0,002*	*0,011*	*0,005*

TABLE XXVII
|%| ERROR METRICS OF 3RD DAY

	ARIMA	E-D	AF	NN
μ %	*9,32*	*26,38*	*16,45*	*17,75*
Var	*0,015*	*0,012*	*0,036*	*0,02*

Analyzing the squared percentual error, the first ANOVA, Table XXVIII, shows how there is no statistical difference among the models concerning the squared but there is an important one among the days, with a security level of 95%.

TABLE XXVIII
%^{2} ERROR ANOVA WITH 2 WAYS 3RD DAY MODEL

Source	FO	F 0,1	F 0,05	F 0,025	F 0,01
Model	*1,82*	*2,28*	*2,92*	*3,59*	*4,51*
Day	*2,22*	*1.82*	*2,16*	*2,51*	*2,98*

Again, it's not the model which has an effect on the error made, but it's the day forecasted. This result implies that the monovalent couldn't be made.

TABLE XXIX
%^{2} MEAN COMPARISON 3RD DAY MODEL

H0	Ha	t	t 0.1
μ(NN) = μ(ARIMA)	*μ(NN) ≠ μ(ARIMA)*	*1,177*	*1,717*
μ(NN) = μ(ARIMA)	*μ(NN) > μ(ARIMA)*	*1,177*	*1,321*
μ(NN) = μ(AF)	*μ(NN) ≠ μ(AF)*	*0,274*	*1,717*
μ(NN) = μ(AF)	*μ(NN) > μ(AF)*	*0,274*	*1,321*

Table XXIX confirms the results of the ANOVA.

TABLE XXX
|%| ERROR ANOVA WITH 2 WAYS 3RD DAY MODEL

Source	FO	F 0,1	F 0,05	F 0,025	F 0,01
Model	*4,37*	*2,28*	*2,92*	*3,59*	*4,51*
Day	*3,16*	*1.82*	*2,16*	*2,51*	*2,98*

Related to the absolute percentual error, the model is significant with a security level of 97,5%, while the day with a security level of 99%. This implies a monovalent can't be performed.

TABLE XXXI
|%| ERROR MEAN COMPARISON 3RD DAY MODEL

H0	Ha	t	t 0.1
μ(NN) = μ(ARIMA)	*μ(NN) ≠ μ(ARIMA)*	*1,551*	*1,717*
μ(NN) = μ(ARIMA)	*μ(NN) > μ(ARIMA)*	*1,551*	*1,321*
μ(NN) = μ(AF)	*μ(NN) ≠ μ(AF)*	*0,190*	*1,717*
μ(NN) = μ(AF)	*μ(NN) > μ(AF)*	*0,190*	*1,321*
μ(AF) = μ(ARIMA)	*μ(AF) ≠ μ(ARIMA)*	*1,094*	*1,717*
μ(AF) = μ(ARIMA)	*μ(AF) > μ(ARIMA)*	*1,094*	*1,321*

The tests on the variances showed that they were statistically the same for ARIMA/ANN and AF/ANN, so these 2 comparisons have been made. The result is: the 2 tailed tests always state that there is no difference among the models and the 1 tailed test only find a difference between ARIMA and the ANN with a security level of 90 %.

5th day

The last tests are made for the forecast at 5 days. In this case Augmented Forecast failed in predict the values of the index from day 19/04/2022 to day 22/04/2022. The result is a smaller dataset. The choice was to use only the data of the days for which every model had a prediction.

TABLE XXXII
%^{2} ERROR METRICS OF 5TH DAY

	ARIMA	E-D	AF	NN
μ %	*5,63*	*10,01*	*5,90*	*8,23*
Var	*0,005*	*0,007*	*0,0026*	*0,008*

TABLE XXXIII
|%| ERROR METRICS OF 5TH DAY

	ARIMA	E-D	AF	NN
μ %	*19,08*	*27,30*	*21,64*	*24,80*
Var	*0,022*	*0,028*	*0,014*	*0,023*

Looking the metrics of the models related to the squared percentual and absolute error in Table XXXII and Table XXXIII, it appears that the best 3 model are again ARIMA, AF and ANN. The bivalent ANOVA in Table XXXIV shows there is no difference among the models. All the effect on the squared percentual error comes from the day forecasted. This result is the same of the direct comparison among the means in Table XXXV.

TABLE XXXIV
%^{2} ERROR ANOVA WITH 2 WAYS 5TH DAY MODEL

Source	FO	F 0,1	F 0,05	F 0,025	F 0,01
Model	*0,34*	*2,36*	*3.07*	*3,82*	*4,87*
Day	*5,44*	*2,02*	*2,49*	*2,97*	*3,64*

TABLE XXXV
%^{2} ERROR MEAN COMPARISON 5TH DAY MODEL

H0	Ha	t	t 0.1
μ(NN) = μ(ARIMA)	*μ(NN) ≠ μ(ARIMA)*	*0,646*	*1,761*
μ(NN) = μ(ARIMA)	*μ(NN) > μ(ARIMA)*	*0,646*	*1,345*
μ(NN) = μ(AF)	*μ(NN) ≠ μ(AF)*	*0,788*	*1,761*
μ(NN) = μ(AF)	*μ(NN) > μ(AF)*	*0,788*	*1,345*
μ(AF) = μ(ARIMA)	*μ(AF) ≠ μ(ARIMA)*	*0,106*	*1,761*
μ(AF) = μ(ARIMA)	*μ(AF) > μ(ARIMA)*	*0,106*	*1,345*

For what concerns the absolute percentual error, the ANOVA in Table XXXVI shows again that all the effect is on the day. This is shown also in Table XXXVII with the mean comparisons.

TABLE XXXVI
|%| ERROR ANOVA WITH 2 WAYS 5TH DAY MODEL

Source	FO	F 0,1	F 0,05	F 0,025	F 0,01
Model	*0,17*	*2,36*	*3.07*	*3,82*	*4,87*
Day	*3,56*	*2,02*	*2,49*	*2,97*	*3,64*

TABLE XXXVII
|%| ERROR MEAN COMPARISON 5TH DAY MODEL

H0	Ha	t	t 0.1
μ(NN) = μ(ARIMA)	μ(NN) ≠ μ(ARIMA)	0,768	1,761
μ(NN) = μ(ARIMA)	μ(NN) > μ(ARIMA)	0,768	1,345
μ(NN) = μ(AF)	μ(NN) ≠ μ(AF)	0,572	1,761
μ(NN) = μ(AF)	μ(NN) > μ(AF)	0,572	1,345
μ(AF) = μ(ARIMA)	μ(AF) ≠ μ(ARIMA)	0,472	1,761
μ(AF) = μ(ARIMA)	μ(AF) > μ(ARIMA)	0,472	1,345

V. CONCLUSION

This study has the aim to evaluate the performance of a Neural Network based on a Boltzmann Machine in the forecasting of the future values of an index in the maritime sector. Firstly, the importance of the knowledge of the trend has been tested. Related to the percentual error, there is a statistical difference for all the 3 days. However, on the basis of the squared percentual error there is only a slight difference on the 5th day, where the ANN trained with plus trend outstands the one with minus trend. Looking at the absolute percentual error there is a difference on the 3rd day and on the 5th day. These results mean that there isn’t a statistically difference on all the analysis. However, as when the difference appears always states that ANN with plus trend outstands ANN with minus trend, we can conclude that the information on the trend has a value in reducing the error made in the forecasting. Once obtained this first result, the following step was comparing the different models on the base of the squared and absolute percentual error made. For older data, ANN trained with plus trend has been compared with CC on the 1st day forecasted. In this case the comparison was also on the percentual error. The analysis showed how CC outstands ANN related to the percentual error and the squared percentual error, while on the absolute percentual error there is no statistical difference. So it can be stated that CC can be better than an ANN. Concerning the newer data, the models analysed are: ARIMA, Carbon-Copy, Encoder-Decoder, Augmented Forecasted and the Neural Network of Predictor. What could be found out by comparing the best 3 models averages for prediction at 1 and 5 days was that, statistically, the Neural Network doesn’t differ from the other 2 best models. The comparison of biggest interest was with ARIMA, which is, for all the 3 predictions, the model with a lower average. However, although apparently ARIMA is the best, the statistical tests showed that there is not a significant difference between ARIMA, the best, and the Neural Network, the worst of the best 3, in the forecast at 1 and 5 days. At the contrary, at 3 days, ARIMA statistically performs better than the Neural Network with a confidence level of 90%. This is the only difference found comparing the best 3 models for each forecasted day. About this result, there must be paid attention to the fact that the market has a bigger volatility due to the Russia-Ukraine war. This made the forecast harder and in fact all the models got worse performances in the last days reaching peaks over 20% of error and, in the worst cases, even up to 40% in the forecast at 5 days. This situation was easily detected by the bivalent ANOVA in the forecast at 3 and 5 days, where in some the day has more effect than the model or even the model has no effect at all, like the for the 5th day. This makes clearly understand how huge the effect of an external situation, like the Russia-Ukraine war, is on the market and so it makes the model a lot less performing than expected. In order to have a better comparison among the models, especially as the final aim is to state if the Neural Networks has the ability to outclass the others, the analysis should be performed once again when the global situation comes again in a state of quiet.

REFERENCES

[1] Rafael Hraskoa, Andr´e G. C. Pachecoa, Renato A. Krohling (2015). Time Series Prediction using Restricted Boltzmann Machines and Backpropagation. Procedia Computer Science 55 (2015) 990 – 999.

[2] Box G, Jenkins G M, Reinsel G C. (2015). Time series analysis: Forecasting and control. Prentice Hall, New Jersey.

[3] T. Osogami (2017). Boltzmann machines and energy-based models. Technical Report RT0979, IBM Research - Tokyo

[4] Takayuki Osogami. (2017). Boltzmann machines for time-series. IBM Research - Tokyo.

[5] G. W. Taylor, G. E. Hinton, and S. T. Roweis. Modeling human motion using binary latent variables. In P. B. Sch• olkopf, J. C. Platt, and T. Ho man, editors, Advances in Neural Information Processing Systems 19, pages 1345{1352. MIT Press, 2007.

[6] D. Vengertsev (2014). Deep Learning Architecture for Univariate Time Series Forecasting. Department of Computer Science Stanford University.

[7] S. Walczak (2016). Artificial Neural Networks and other AI Applications for Business Management Decision Support. International Journal of Sociotechnology and Knowledge Development.

[8] S. Dasgupta and T. Osogami (2017). Nonlinear Dynamic Boltzmann Machines for Time-series Prediction.

[9] P. Giribone, R. Revetria, M. Antonetti, R. Tabolacci (2000). Use of Artificial Neural Networks as Support for Energy Saving Procedures in Telecommunications.

[10] L. Cassettari, I. Bendato, M. Mosca, R.Mosca. A new stochastic multi source approach to improve the accuracy of the sales forecasts. foresight, Vol. 19 Iss 1 pp. 48 - 64. http://dx.doi.org/10.1108/FS-07-2016-0036