Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 170 (2020) 642–647

Procedia Computer Science

www.elsevier.com/locate/procedia

The 3rd International Conference on Emerging Data and Industry 4.0 (EDI40)
April 6-9, 2020, Warsaw, Poland

Crude Oil Price Prediction using Artificial Neural Network

Nalini Gupta, Shobhit Nigam*

School of Liberal Studies, Pandit Deendayal Petroleum University, Gandhinagar, India

Abstract

Crude oil is amongst the most important resources in today's world, it is the chief fuel and its cost has a direct effect on the global habitat, our economy and oil exploration, exploitation and other activities. Prediction of oil prices has become the need of the hour, it is a boon to many large and small industries, individuals, the government. The evaporative nature of crude oil, its price prediction becomes extremely difficult and it is hard to be precise with the same. Several different factors that affect crude oil prices. We propose a contemporary and innovative method of predicting crude oil prices using the artificial neural network (ANN). The main advantage of this approach of ANN is that it continuously captures the unstable pattern of the crude oil prices which have been incorporated by finding out the optimal lag and number of the delay effect that controls the prices of crude oil. Variation of lag in a period of time has been done for the most optimum and close results, we then have validated our results by evaluating the root mean square error and the results obtained using the proposed model have significantly outperformed.

Keywords: Artificial Neural Network, Crude Oil Price, Prediction Model, Optimal Lag

1. Introduction

Crude oil is a yellow-black naturally occurring liquid found in geological formations beneath the Earth's surface, it can be separated into various kinds of consumer fuels through the process of fractional distillation. Crude oil is amongst the most important energy resources on earth right now. So far, it remains the world's leading fuel, with nearly one-third of global energy consumption. Petroleum products are also made of refined crude oil. Encouraging usage of fossil fuels is getting highly unpopular as they're irrefutably responsible for global warming, and other severe impacts on ecosystems [1]. A conscious effort to phase-out fossil fuels is being made throughout the world to act upon the climate crisis. Petroleum is of utmost importance to industries, civilization, it accounts for a

* Corresponding author. Tel: +91 9428488296
Email address: shobhit.nigam@sot.pdpu.ac.in

1877-0509
10.1016/j.procs.2020.03.136

significant percentage of the world's energy consumption which makes it an underlying factor in world politics and international relations[2]. Current estimates suggest that the world usage of Petroleum ranges up to 95 million barrels per day. Crude oil price prediction has a scope larger than we can think of, the forecasting used is relevant for big and small industries along with the government benefitting from the predicted prices, but due to the evaporative nature of oil, it becomes very challenging to achieve accuracy. In the current scenario where technology is taking over our lives and efforts are being made to minimize human labour the Artificial Neural Network Technique has become one of the most effective methods used for prediction of any data. In this paper, we propose a technique which can be used to forecast the oil prices using the Artificial Neural Network (Sigmoid Function with the Learning Algorithm). Prices of the crude oil and the economy have a strong correlation along with the economic and political factors affecting the crude oil price.
This work proposes a model to determine the lag in a given data set. Lag calculation in any prediction model is extremely important as it makes you one step ahead, in any prediction model the maximum impact is due to the previous year's values, hence lag values carry a lot of information of the future series of price and time. ANN has some major advantages that make it extremely suitable for a prediction model, ANNs can model non-linear and complex relationships between input and output[3].ANN can also generalize, after learning from the inputs it can infer relationships without seeing the data or the input. ANN also gains knowledge from hidden relationships in the data without imposing any fixed relationships in the data, hence making it a reliable method for making predictions[4].

There are innumerable ways and approaches which are being used and have been used for predicting the prices of crude oil, one of the common methods is the one based on intuitions wherein the experiences, knowledge and opinions of experts and professionals who have worked for a long time in this industry is used to predict the prices for the coming future[5]. A lot of economists and analysts also work on forecasting the crude oil cost, they use data transformation and regression which include autoregressive moving average (ARMA) models and vector autoregressive (VAR) models with a different value for the input each time and then plot the graph with their predicted prices taking in the major economic factors into account[6]. Another and one of the most common techniques which are used is by accepting the current statistics and prices and assuming it to be the same for future without any change at all in the prices, however, such a prediction never works now as the nature of oil is extremely volatile.

2. Methodology for Proposed Model

Artificial Neural Network also the connectionist systems are computing systems that are based on and are theoretically alike, but not exactly identical to, biological neural networks of a human body. An ANN performs its task by taking in examples and requires no programming with task-specific rules [7]. The purpose of a neural network is to construct or design an output pattern when given an input pattern. An artificial neural network (ANN) has an architecture which is parallelly-distributed with large number of nodes (neurons) and connections[9].

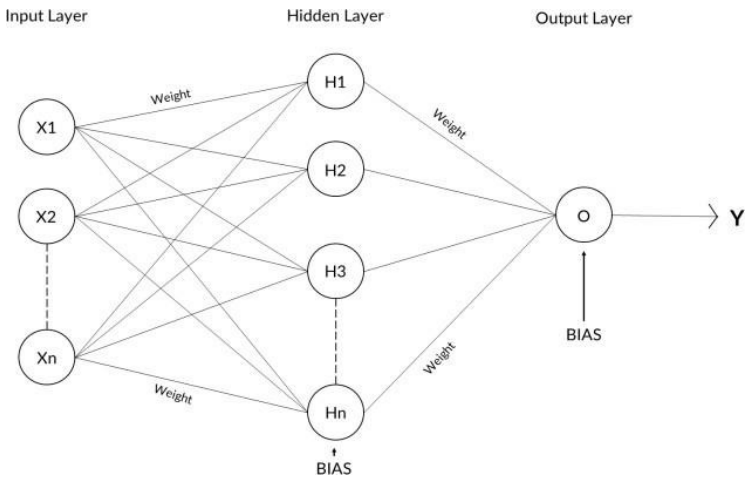

Fig. 1. Structure of the Artificial Neural Network

We use the Back-propagation learning algorithm and the error signal is cultivated through the network in the backward direction by changing and managing weights of the network to maximize the performance of the network[8]. The procedure is done until the network is able to provide desired responses.

The ANN model development steps for our prediction are as follows:

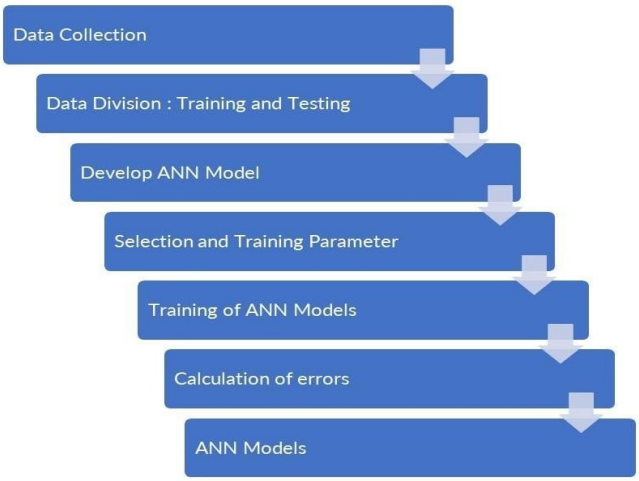

Fig. 2. Model development Diagram

In the suggested model, there is only one dependent variable, the closing price of crude oil which has been considered, since it's a time series, we have followed the model for general time series forecasting in conducting the experiments, which have been represented in the form as follows[10]:

$$Y_t = f(X')$$

where X' is vector of lagged variables $\{x_{t-1}, x_{t-2}, \dots, x_{t-p}\}$. The input variables depicted in figure 1 are the lagged variables. The estimating problem is to approximate the function f [10]. This can be done by iteratively adjustingthe weights in the modelling process. The diagram for the proposed model is depicted in figure 2.

As an illustration, we have taken four phases in the formation of this suggested predictive model, which are as follows:

(a) Collection of data: We have taken the crude oil price data from *investing.com*. A period of 5 years and 11 months is the period of the collected data. The closing price of crude oil was comprised and approximately a total of 1500 records were extracted.

(b) Data normalization: The process of Data Normalisation takes place before the start of the training process. The normalisation range of the closing price is [0.001, 0.005] using the following equation

$$P' = \frac{p - Min}{Max - Min}\ (l - m) + m$$

Where we have taken as P' the normalized value; p is the value that is to be normalized; *Min* is the minimum value of the series that needs to be normalized; *Max* is the maximum value of the series that is to be normalized; *m* is the minimum value of the range *l* is the maximum value of the range[11].

(c) Activation function: The activation function or popularly known as the transfer function finds the relationship among the inputs node and the output nodes of a network. The sigmoid function given below has been used to carry out this work

$$f(x) = \frac{1}{1+e^{-x}}$$

(d) Training algorithm: The training for a neural network is extremely complex, it is an unconstrained nonlinear (change in output is not proportional to the change in input) problem of minimization where the arc weights of the neutral network are modified iteratively to minimize the total mean or overall squared error between the desired and the actual output values for the output nodes overall input patterns. The usual BP algorithm which is based on approach of steepest descent gradient has been taken to train the model and minimize the errors, and error function E is defined in

$$E = \frac{1}{2N}\sum_{p=1}^{n}(y_p - y_p^d)^2$$

where y_p is taken as the output of the network and y_p^d is the output of desired output for the p^{th} input pattern. Used here is the steepest descent gradient approach and the partial derivate is solved using the chain rule, the renovated rules for the weights and biases of this model are obtained using the following equations:

$$w_i^{new} = w_i^{old} + \Delta w_i$$

$$b_i^{new} = b_i^{old} + \Delta b_i$$

Where $\Delta w_i = -\eta \frac{dE}{dw_i}$

$$= -\eta \frac{1}{n} \sum_{p=1}^{n} \left((y_p - y_p^d)\, y_p\, (1 - y_p) \frac{u}{w_i x_i + b_i} x_i\right)$$

$$\Delta b_i = -\eta \frac{dE}{db_i}$$

$$= -\eta \frac{1}{n} \sum_{p=1}^{n} \left((y_p - y_p^d)\, y_p\, (1 - y_p) \frac{u}{w_i x_i + b_i}\right)$$

where η is taken as the parameter for learning that controls the convergent speed.

3. Results and Discussions

The crude oil prices are taken out by implementing the performance of the proposed model. The proposed model is here used to predict the closing price of crude oil. The performance check of the suggested model has been evaluated using the Root Mean Square Error (RMSE) criterion which is given as:

$$RMSE = \sqrt{\frac{1}{n}\,(y_i - \widehat{y_i})^2}$$

where y_i and $\widehat{y_i}$ are taken as the original price and predicted price, n is taken as the total number of observations[12].

Table 1: Delay with its corresponding RMSE Values

LAG	RMSE VALUE
2	10.23
3	7.68
4	8.50
5	9.71

The data set has been split into 2 parts of 70% and 30%, respectively for instructing, training and testing. The training data is used to find the optimal lagged value for our proposed model. The results are taken out by varying the lag value, the optimal lag is obtained with the least RMSE values. Table 1 shows the RMSE values corresponding to the lag values 2,3,4 and 5. Since the RMSE value is least for lag value 3, it has been used for our prediction.

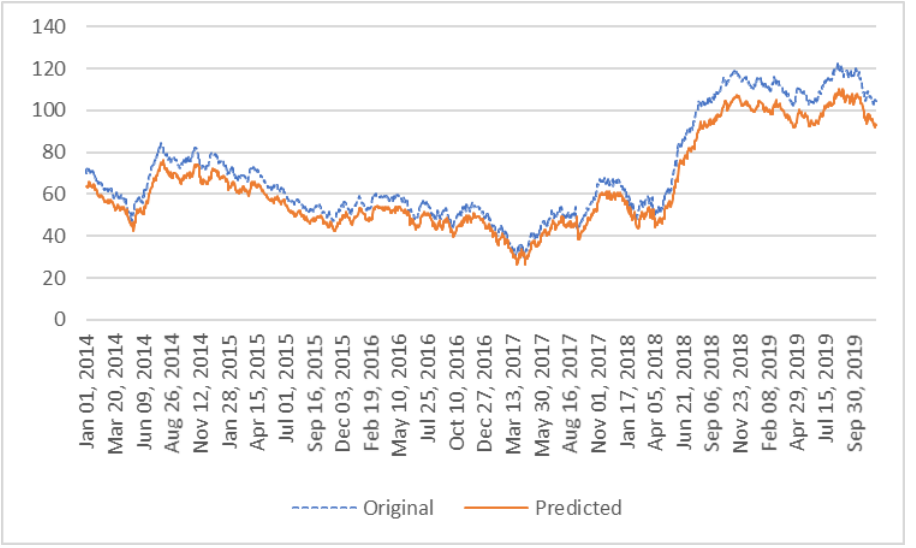

Fig. 3. Original and Predicted Closing prices with time

4. Conclusion and Future Scope

In this paper, an artificial neural network model is presented with the task of determining the most favourable lag in the crude oil price data. It is evident, the result is shown in the figure, the prediction is accurate till there is a massive and sudden change in the actual data, where it becomes challenging to predict the exact new price with the change, however, the proposed model has efficiently taken into consideration these patterns. Else ways, this also proves the theory that financial markets are unpredictable and change anytime because of known and unknown factors[13]. This work indicates that the ANN model is an effective tool for crude oil price prediction and can be efficiently used for short term price forecasting by determining the optimal lags. The proposed model is powerful and highly suggested because investors can use it not only to initiate trades but also as an effective tool to judge various strategies relating

to investments. This work is carried out on the closing price of crude oil; however, there are various other factors which also affect the crude oil prices like change in the prices and quantities (demand and supply), change in the economy and current affairs as shown by the media. The main advantage of this research is in capturing the changing pattern of these prices. In the coming future, fundamental indicators and market trends have been planned to be incorporated into a model which will help the proposed model perform more efficiently.

5. Acknowledgement

This work has been supported by School of Liberal Studies, Pandit Deendayal Petroleum University, Gandhinagar, I would also like to thank all the faculties of Mathematics Department for their valuable suggestions which significantly contributed in increasing my knowledge in this field.

References

[1] Kaufmann, R. K., & Ullman, B. (2009). Oil prices, speculation, and fundamentals: Interpreting causal relations among spot and futures prices. Energy Economics, 31(4), 550–558.
[2] Shobhit Nigam. "Chapter 84 Single Multiplicative Neuron Model in Reinforcement Learning" , Springer Science and Business Media LLC, 2019
[3] "Harmony Search and Nature Inspired Optimization Algorithms" , Springer Science and Business Media LLC, 2019
[4] Shuang Gao, Yalin Lei. "A new approach for crude oil price prediction based on stream learning" , Geoscience Frontiers, 2017
[5] Ramakanta Mohanty. "Software Reliability Prediction Using Group Method of Data Handling" , Lecture Notes in Computer Science, 2009
[6] Kulkarni, S., Haidar, I., 2009. Forecasting model for crude oil price using artificial neural networks and commodity future prices. International Journal of Computer Science and Information Security 2 (1).
[7] Haykin, S. (1999). Neural Networks: A Comprehensive Foundation, 2nd edition, Prentice Hall, 842 pages
[8] Onur Dursun. "3 Methodology" , Walter de Gruyter GmbH, 2014
[9] Lakshmanan, Indhurani, and Subburaj Ramasamy. "An Artificial Neural-Network Approach to Software Reliability Growth Modeling" , Procedia Computer Science, 2015.
[10] Haykin, S. (2009). Neural Networks and Learning Machines, 3rd edition, Pearson, 938 pages
[11] N. Raj Kiran, V. Ravi. "Software reliability prediction by soft computing techniques" , Journal of Systems and Software, 2008
[12] Lean Yu. "An EMD-Based Neural Network Ensemble Learning Model for World Crude Oil Spot Price Forecasting" , Studies in Fuzziness and Soft Computing, 2008
[13] Xin Wang, Ji Wu, Chao Liu, Senzhang Wang, Wensheng Niu. "A Hybrid Model Based on Singular Spectrum Analysis and Support Vector Machines Regression for Failure Time Series Prediction" , Quality and Reliability Engineering International, 2016