LITERATURE SURVEY

Prediction of future crude oil price is considered a significant challenge due to the extremely complex and dynamic nature of the market and stakeholder's perception. We did a survey over the possible sources that we could access. In our exploration, we did find the authors.

Predictive analytics is a piece of advanced research used to gage obscure future opportunities by statistical models and other scientific approaches aimed at making practical forecasts (as opposed to the theory-only forecasts), as well as tools for testing the accuracy of these prediction in reality. Predictive analytics has not only been applied in the financial sector but also has been used in user profiling, health, and many others. Several algorithms and methods related to predictive analytics include adaptive neuro-fuzzy system (ANFIS), spiking neural network, Recurrent Neural Network, Long-Short Term Memory (RNN-LSTM) and others. These methods are useful to produce an insight towards data such as historical crude oil prices to be used by experts[1].

Similar to many other products, crude oil prices depend on the supply and demand curves. In other words, if the demand is more than the supply, it means the oil market is currently in shortage condition, hence the price of oil increases due to its limited supply and availability. Conversely, if the supply is more than the demand, the oil supply and availability are in surplus; hence the price becomes cheaper as it is widely available, and it has only a few buyers. Equilibrium price exists when there are no surpluses or deficiencies in the market. Other factors, such as the supply-demand curve and competitive factors, have a short-term effect[2].

The neural network contains a set of neurons (or perceptron) which acts as processing units, interlinked, and may reside within an extensive network. The network can be designed to handle deep learning. The working mechanism of a neuron combines inputs from the data with a set of coefficients that either amplify or reduce that input. Upon reaching a threshold value, an output is produced. The number of hidden layers may vary based on the complexity of computation. Due to the popularity of neural networks, many researchers have introduced different approaches for predictive analytics task including spiking neural network data modelling , ANFIS, deep learning, EEMD-LSTM, and many other[3].

RNN-LSTM NETWORK: Despite being an effective method for predictive tasks, the traditional neural networks method is not capable of holding memories. The Recurrent Neural Network (RNN), on the other hand, is a chunk of a neural network made into a loop— giving its capability to retain information from its previous state.The architecture of a standard RNN is related to sequences and lists. In the unrolled state, an input x0 fed into a chunk of neural network A, which gives out an output , preserving information to the next chunk of neural network. A special kind of RNN that works better than the standard RNN utilizes LSTM[4].

Hochreiter & Schmidhuber have introduced the concept of Long-Short Term Memory (LSTM), which has proven its accuracy across various domains. LSTM is a type of Recurrent Neural Network (RNN) that can learn longterm dependencies and is useful for a sequence to sequence prediction—such as prediction of upcoming crude oil prices using time-series data. RNN has a powerful potential to increase input information by refeeding the output as the input. Therefore, the output is created based on the previous iteration output. After the output is created, it is duplicated and sent back to the intermittent system. In order to create a decision, RNN relies on the new information fed into the network and the output produced from the previously fed information[5].

ANN has gained much attention for its computational intelligence approach and its capability to make prediction. It is popular for capable on modelling the nonlinearity, which results to a class of general function approximators. The development of this ANN-Q model is based on a process development suggested by and presented. There are three steps of development for this model; (i) objective determination, (ii) data pre-processing and (iii) ANN modelling[6].

(i)   Objective Determination: The objectives are determined to focus on developing a suitable and an accurate prediction tool for the crude oil market and predict its price for every barrel of crude oil in US Dollar (USD). USD is used as a standard price in NYMEX where WTI is traded. More over at this phase, we determined and analysed the selection of inputs and output to be used in the prediction so to find the fittest and right observations for the prediction. Nonetheless, it is important to ensure ANN capable of learning the connections between inputs effectively so to successfully achieve the final objective of the model. The networks' performance will be demoted if the selection of input and output of variables are not carefully selected. Simultaneously, the selection of input and output is a process of development and errors in mapping the correct input and output will make the prediction tool be less reliable.

(ii)   Data Pre-processing All selected data in this phase are given extensive deliberation to ensure the cleanliness of data from noises and errors before further transformation into data representation stage. Furthermore, [16] suggests to collect data from reliable and premiere source, determine the series of time in specific range of date and have full cleansing on all data. These are to ensure no missing or inappropriate value of data present so to achieve fitter sets of data for the prediction. In this research, data used for training are time-series data and normalised data. Both of these types of data will be tested against each other to determine the best training model. The normalised data were represented with One-Step Returns function.

(iii)   ANN Modelling In this study, we used BPNN for training the input variables. Detail on BPNN can be referred. The networks can capture relatively complex environment as it contains multiple layers of interacting nonlinear neurons. Nevertheless, the networks are composed by historical observations with assigned weight values for input and the future values as the output. the numbers of input and output variables selected are in parallel with the number of input and output layer in ANN. Selecting the right and appropriate numbers for hidden layers is important and its capability of reasoning will fail if extensive numbers of hidden layers are used in the model.

REFERENCES:

[1]. A. H. Bukhari, M. A. Z. Raja, M. Sulaiman, S. Islam, M. Shoaib, and P. Kumam, "Fractional neuro-sequential ARFIMA-LSTM for financial market forecasting," IEEE Access, vol. 8, pp. 71326–71338, 2020.

[2]. R. Zazo, P. Sankar Nidadavolu, N. Chen, J. Gonzalez-Rodriguez, and N. Dehak, "Age Estimation in Short Speech Utterances Based on LSTM Recurrent Neural Networks," IEEE Access, vol. 6, pp. 22524– 22530, 2018.

[3]. M. Othman, S. A. Mohamed, M. H. A. Abdullah, M. M. Yusof, and R. Mohamed, "A Framework to Cluster Temporal Data Using Personalised Modelling Approach," in Advances in Intelligent Systems and Computing, vol. 700, 2018, pp. 181–190.

[4]. M. H. A. Abdullah, M. Othman, and S. Kasim, "A Spiking Neural Networks Model with Fuzzy-Weighted k-Nearest Neighbour Classifier for Real-World Flood Risk Assessment," Adv. Intell. Syst. Comput., vol. 978, pp. 1–9, 2020.

[5]. M.F. Nasrudin, "Pembangunan Model dan Aplikasi Ramalan Pasaran Saham BSKL Menggunakan Rangkaian Neural Perambat Balik," Msc. of Computer Science, Faculty of Technology and Science Information, Universiti Kebangsaan Malaysia, Bangi, Selangor, 2001.

[6]. M. Negnevitsky, Book Artificial Intelligence: A Guide to Intelligent System, Pearson Education Ltd., 2005, p. 415