

# Machine Learning Approach for Predicting Crude Oil Price Using Fuzzy Rule Based Time Series Method and Sentimental Analysis

<sup>1</sup>Dona Sara Jacob; <sup>2</sup>Sam G Benjamin; <sup>3</sup>Dr. Radhakrishnan B

<sup>1</sup> CSE Department, BMCE  
APJ Abdul Kalam Technological University, Sasthamcotta, Kerala, India

<sup>2</sup> CSE Department, BMCE  
APJ Abdul Kalam Technological University, Sasthamcotta, Kerala, India

<sup>3</sup> CSE Department, BMCE  
APJ Abdul Kalam Technological University, Sasthamcotta, Kerala, India

**Abstract** - The movements in oil prices are very complex and, therefore, seem to be unpredictable. The continuous usage of statistical and econometric techniques for crude oil price prediction might demonstrate demotions to the prediction performance. Crude oil price prediction depends on heavily on uncertainty in the crude oil price fluctuation. The proposed approach uses a fuzzy rule based system embedded in fuzzy time series application to accurately extract a feature weight that can predict crude oil price prediction accurately. Another major parameter used for crude oil price prediction is news feeds. Our proposed approach extracts features weights from news feeds using a sentimental analysis based on latent dirichlet allocation topic model that can distinguish various online news topics. Both these feature weights along with the quantitative key factors are feed in to recurrent neural network.

**Keywords** - Crude oil price prediction, fuzzy rule based system, sentimental Analysis, Latent Dirichlet Allocation topic Model

## 1. Introduction

Crude oil is a key commodity for global economy. It is a vital component for the economic development and growth for industrialized and developing countries. Moreover, political events, extreme weather, speculation in financial market, amongst others are major characteristics of crude oil market which increase the level of price volatility in the oil markets. Therefore, to reduce the negative impact of the price fluctuations, it is very important to predicate the price direction. Due to strong chain effects owned by this crude oil market, any changes in the factors involved will have exclusive impact to the price. Therefore, every increment and decrement that occurs to the crude oil price will then also give impact to the price of petroleum and later correspond to the global economy. A good prediction tool is crucial to be developed for this matter. Moreover, the variable used for this forecast is the average price of internationally traded oil, West Texas Intermediate (WTI).

The drastic increment of global crude oil market in July 2008 had indeed affects the economy around the globe. Started with USD\$69.00 per barrel in April, 2006 and rising up to 50% increment of USD\$134.00 per barrel in July, 2008, this phenomenon had then gave an exclusive economic impact to the oil importing and exporting

countries. In addition, crude oil products are one of the world's major commodities with high volatility level. The volatility of this crude oil price is depending on demand and supply of the commodity, level of inventories, economic indicators and finally, the population of the world.

Artificial Neural Network (ANN) is the science of Artificial Intelligence (AI) that being applied in various fields effectively, for example, pattern recognition, prediction, control, optimization and clustering. An ANN is based on a collection of connected units or nodes called artificial neurons, which loosely model the neurons in a biological brain. Each connection, like the synapses in a biological brain, can transmit a signal from one artificial neuron to another. An artificial neuron that receives a signal can process it and then signal additional artificial neurons connected to it. In common ANN implementations, the signal at a connection between artificial neurons are a real number, and the output of each artificial neuron is computed by some non-linear function of the sum of its inputs.

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 approximates.

Recurrent Neural Networks (RNN) is a powerful and robust type of neural networks, because they are the only ones with an internal memory. Because of their internal memory, RNN's are able to remember important things about the input they received, which enables them to predicting what's coming next. The Long Short-Term Memory, or LSTM, network is perhaps the most successful RNN. Because, it overcomes the problems of training a recurrent network and in turn has been used on a wide range of applications.

Although there are several methods used to predict crude oil price. In many cases, a prediction of future price has many limitations in terms of accuracy. Fuzzy rule based time series method is useful in handling the uncertainty in the data. This method enhances the efficiency of time-series forecasting, which would in turn increase the accuracy of the predictions.

In this paper we proposed a fuzzy rule based method embedded in fuzzy time series application and extracts features weight from news feeds using sentimental analysis based on Latent dirichlet allocation model that can distinguish various news headlines. Both these weights are given to a recurrent neural network. The proposed method outperforms with an accuracy of 98.6%.

## 2. Research Background

Saeed Moshiri et al [2], model and forecast daily crude oil futures prices from 1983 to 2003, listed in NYMEX, applying ARIMA and GARCH models. Then test for chaos using embedding dimension, BDS(L), Lyapunov exponent, and neural networks tests. Finally, set up a nonlinear and develop a feed forward multilayer ANN model to estimate and forecast crude oil futures prices, and compare the results with those obtained by standard ARMA and GARCH models. The ANN model outperforms the linear and nonlinear models.

Mayuree Sompui et al [5], The prediction model based on artificial neural network (ANN) to forecast and compared with least square method (LSM). The results show that on the short-term, the best prediction model for ANN of four, three, two and one hidden layers, respectively. The ANN of one - four hidden layers is found to be able to forecast better than the LSM. The most widely used activation functions for ANN in the hidden layer are the sigmoid functions and the hyperbolic tangent.

All data sets retrieved from Energy Information Administration. The network architecture is 4 layers feed forward with 9 neurons in the hidden layer. The network was trained for 1000 iterations or until one of the stopping criteria is met. The learning rate is 0.01 and training algorithm is Levenberg Marquardt. The results of the prediction model for ANN show that the outputs of ANN are found to be able to forecast better than the LSM.

Nur Fazliana Rahim et al [3], presented the area of data driven Fuzzy Rule Based System in the time series forecasting of CPO prices. This research showed that this new approach has several advantages that can be applied to strengthen the previous method. The development of the Weighted Subset hood Based Algorithm (WSBA) – a preliminary data-driven FRBS – from the fuzzy subset hood values, provided ease of use because default fuzzy rules could be churned out without the requirement of threshold values. The outcomes of the method have shown an increase in the accuracies of the CPO price forecasts.

Xuerong Li et al [1], proposed a new text based crude oil price forecasting method using deep learning techniques, sentimental analysis and topic extractions. This study further proposes a feature grouping method based on the Latent Dirichlet Allocation (LDA) topic model for distinguish effects from various online news topics. Optimized input variable combination is constructed using lag order selection and feature selection methods. S. N. Abdullah, X. Zeng [4], machine learning and computational intelligence approach through HC and ANN-Q model is applied to predict the monthly WTI crude oil price for every barrel in USD. The result obtained from simulation study validates the effectiveness of data selection process by HC model. HC model successfully extracts a comprehensive list of key factors that cause the crude oil price market to volatile. The effectiveness and accurateness of data selection also helps to extensively deliberate the input variables combination for ANN-Q model. By using monthly WTI price together with the online news as the training data, this approach performed very well in predicting the crude oil price.

In our system we extract features weights from news feeds using a sentimental analysis based on latent dirichlet allocation topic model that can distinguish various online news topics. Both these feature weights along with the quantitative key factors are feed in to recurrent neural network.

### 3. Design & Architecture

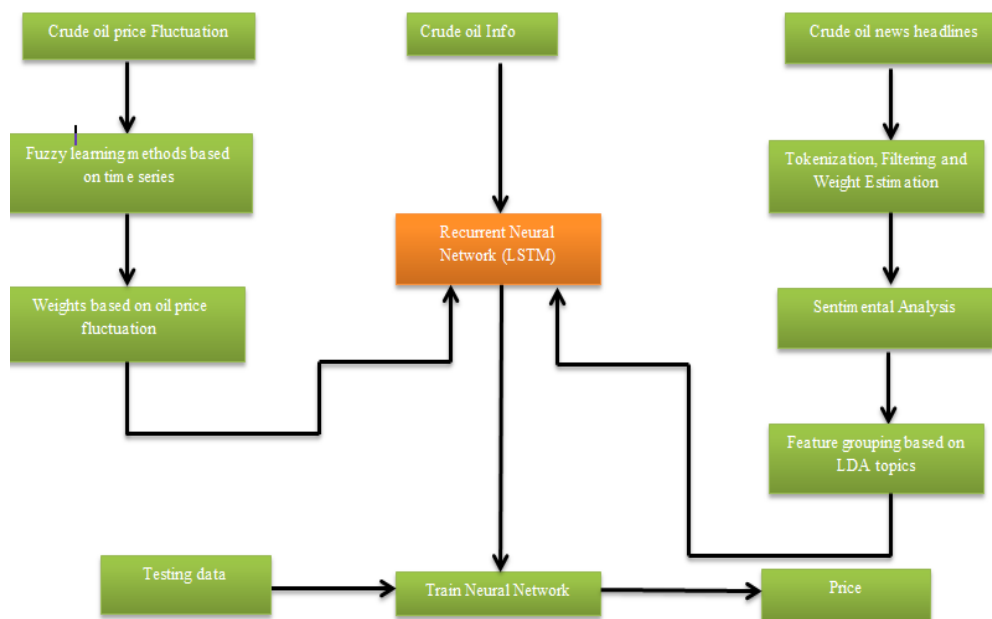


Fig.1.Proposed System

This system jointly combined the features of a news feeds processing system along with the features that of a price and production fluctuation detection unit. The system combines the strength of a fluctuation detection unit to an optimizer, and news feeds processing unit that can optimize the system to achieve better output. In news feeds processing unit, the news headlines are tokenized and filtered then each headlines are designated with a weight. The weighted words undergo a sentimental analysis and features are extracted and grouped using LDA (Latent Dirichlet Allocation). In Parallel crude oil price and production are analysed based on a fuzzy learning method and both features are feed into a neural network to build the system.

### 4. Experimental Result

Table.1. Experimental Result

Model	Accuracy
LSTM	98.46%
ARMA	76%
GARCH	70%

There are 28 quantitative key factors with an external support of standardized data using Linear Dirichlet Allocation method are used as input to the neural

network. The result shows that the proposed model is better as compared with the previous models and also it outperforms with an accuracy of 98.46%. The experimental results of the forecasted crude oil price obtained are discussed. Table 1 illustrates the comparison of the outcome obtained by using previous methods and those by the present method. NYFE data is used for training. For training, here we use recurrent neural network. In recurrent neural network, Long Term Short Memory method is used for training the network. The activation function used is SOFTMAX.

### 5. Conclusion

This study proposes a new prediction model based on fuzzy rule based time series and sentimental analysis method. Here, the crude oil price fluctuation, online news headlines and other crude oil news headlines are considered as the input. In news feeds processing unit, the news headlines are tokenized and filtered then each headlines are designated with a weight. The weighted words undergo a sentimental analysis and features are extracted and grouped using LDA (Latent Dirichlet Allocation). In Parallel crude oil price and production are analysed based on a fuzzy learning method and both features are feed into a recurrent neural network. In recurrent neural network Long short Term Memory (LSTM) method is used for training. The result shows that the proposed model is better as compared with the

previous models. Also, this model gives an consistent accurate result.

## References

- [1] Xuerong Li, Wei Shang, Shouyang Wang. Text based crude oil price forecasting: A deep learning approach
- [2] Moshiri, Saeed and Foroutan, Faezeh, Forecasting Nonlinear Crude Oil Future Prices.
- [3] Nur Fazliana Rahim, Mahmud Othman, Rajalingam Sockalingam, and Evizal Abdul Kadir. Forecasting Crude Palm Oil Prices using Fuzzy Rule-Based Time Series Method.
- [4] S. N. Abdullah, X. Zeng. Machine Learning Approach for Crude Oil Price Prediction with Artificial Neural Networks-Quantitative (ANN-Q) Model
- [5] Mayuree Sompui and Wullapa Wongsinlatam. Prediction Model for Crude Oil Price Using Artificial Neural Networks



**Sam G Benjamin** is working as Assistant Professor in Computer Science and Engineering. His research interests focuses on image processing, data mining, and image mining.



**Dr. Radhakrishnan B** is working as the Head of CSE department. He has more than 14 years experience in teaching and has published papers on data mining and image processing

## Authors Profile



**Dona Sara Jacob** received her B.Tech (Computer Science and Engineering) degree from University of Kerala, Trivandrum in 2017. She is currently pursuing his Masters in Computer Science & Engineering from KTU