ATLANTIS PRESS

Advances in Intelligent Systems Research, volume 166

7th Scientific Conference on Information Technologies for Intelligent Decision Making Support (ITIDS 2019)

DSS for Oil Price Prediction Using Machine Learning

Guzel Khuziakhmetova
International Center For Computational Logic
TU Dresden
Dresden, Germany
guzelkhuz@gmail.com

Vitaly Martynov
Department of Economic Informatics
Ufa State Aviation Technical University
Ufa, Russian Federation
vvmartynov@bk.ru

Kai Heinrich
Chair of Business Informatics
TU Dresden
Dresden, Germany
kai.heinrich@tu-dresden.de

***Abstract*—The oil price affects the economic situation of many countries in the world, therefore, there is always an increased interest. A number of efforts have been made by researchers towards developing efficient methods for forecasting oil prices. In this paper, three types of models for forecasting oil prices were created: Linear Regression, Support Vector Machine (SVM) and Convolutional Neural Network (CNN). The root mean square error and standard error were chosen to estimate the constructed models by quantitative characteristics. For visual analysis the graphs depicting the actual and forecast values were plotted. According to the interpretation of the results to the evaluation criteria of the models, when using the price of Brent oil as input data, the SVM has the best predictive ability. This makes it a good tool for forecasting dynamically changing data of large volumes. Also a model of the decision support system (DSS) architecture, a forecasting subsystem and a forecasting module are designed to show how the results of the study can be used in the work of commodity market traders.**

Keywords—oil prices, time series, prediction, neural network, support vector machine, machine learning, energy resources, deep learning

I. INTRODUCTION

Crude oil is one of the main resources in the energy sector, and the efficiency of this industry partly depends on the price of this resource. Therefore, it is important to buy it at minimum price, which requires tools for market price forecasting. As well it can be useful in Strategic Trade Theory that can be improved by predictions [1].

The complexity of predicting is as follows. The main feature of the oil market is the paramount role of political factors (wars, revolutions, terrorist acts, the situation in the Middle East, the strategy of the OPEC oil-producing countries, etc.) [2], which can only be predicted with great difficulty. The so called “black gold,” is one of the world's most precious commodities: its price affects the economic ecosystem at every level, from family budgets to corporate earnings to the nation's gross domestic product (GDP). Therefore, current oil exchange participants, industrial enterprises and governments of countries are extremely interested in prediction of oil prices. In addition to the need for developing models for measuring the intensity of these oscillations and predict future prices in the short term and long term [2].

In this paper, we examine Brent crude oil prices using daily data for the period 2008 - 2018 and Urals crude oil prices using daily data for the period 2016–2017 from the open sources. Three types of models for forecasting oil prices were created: Linear Regression, SVM and CNN. Our purpose is to build a model to predict daily oil prices, so we could select the most suitable one.

II. PROBLEM DESCRIPTION

A. Factors affecting oil price

The problem we are going to look at is the choice between algorithms of machine learning that is the most accurate and efficient for crude oil price prediction, so it could be used in economic information systems. Nowadays dynamic of oil prices is one of the most discussed topics. After the world's largest collapse in oil prices in the second part of 2014 [3], the whole world felt how dependent it is on this black substance extracted from the bowels of the earth. When the dollar sharply "jumped" upwards [4], the majority of the oil companies had to lower their prices even for the most elite category of fuel, thereby incurring serious losses. Until the last year, the situation on the oil market was considered catastrophic, but already in 2018, experts promise an improvement that will undoubtedly affect the economies of many countries [5]. Oil prices in 2018 will increase significantly, - experts of most analytical and economic agencies assure.

In the current year analysts and investors are more optimistic about the fact that this year the supply of oil will decrease even more, which will support the growth of oil prices [5]. However, the growth of shale production in the US is likely to significantly reduce the price increase.

The main factors contributing to the increase in prices for oil products include the reduction in fuel production by the OPEC cartel [2], as well as future interruptions in supplies from the leading exporting countries, such as Libya, Venezuela, Nigeria, etc. OPEC and a number of non-member countries in late 2016 agreed to cut oil production from the level of October of the same year, totalling 1.8 million barrels per day, of which 300,000 are in Russia. The agreement was first concluded for the first half of 2017, then extended until the end of March 2018, and then until the end of 2018. However, the participants of the agreement do not exclude the revision of its parameters in June.

Not only economical, but also political and technical factors may influence on international oil market [2]. A significant period of growth in oil quotations was the second half of 2017. The first sharp jump, immediately at $ 2, of oil price in late September in less than a day, rising to $ 59 per barrel of Brent against, is the background of the conflict

between Iraq and Iraqi Kurdistan. In November, on the eve of the extension of the OPEC + deal, quotations for the first time since July 2015 rose to $ 60, and in late December, oil overcame a mark of $ 66 after the explosion of the oil pipeline in Libya.

In turn, the decline in oil quotations will occur if the weak demand for products continues, the return to the previous volumes of production and the impossibility on the part of OPEC to perform a significant reduction in oil production.

B. Neural networks for time-series prediction

Artificial Intelligence models are extremely popular among researchers at the moment [2]. They allow to use different approaches for data mining and analytics, also provide a wide range of tools beginning from simple statistical and econometric models and ending with tools for recognition ability on complex patterns and decision-making.

The task of forecasting time series has been and remains relevant, especially recently, when powerful means of collecting and processing information became available. Forecasting time series is an important scientific and technical problem, as it allows to predict the behaviour of various factors in ecological, economic, social and other systems. In recent decades, many models and methods, procedures and forecasting techniques have appeared. According to experts, there are already over one hundred forecasting methods, and therefore, the task is to select methods that would provide adequate predictions for the studied processes or systems. Hard statistical assumptions about the properties of time series often limit the possibilities of classical prediction methods. Neural network methods of information processing began to be used several decades ago [2]. Over time, interest in neural network technologies has weakened, then revived again. This inconsistency is directly related to the practical results of ongoing researches. The ability of a neural network to differentiate information comes from its capability to generalize and isolate the hidden dependencies between input and output data. The great advantage of neural networks is that they are capable of learning and generalizing the accumulated knowledge.

The use of neural networks for the financial information analysis is a promising alternative (or complement) for traditional research methods. Due to their adaptability, the same neural networks can be used for analysing several instruments and markets, while the regularities found by the player for a particular tool using technical analysis methods may work worse or not work at all for other tools. Some characteristics of the research object imposes certain features on the use of neural networks for data analysis. This feature is the choice of the neural network error function, which is different from the traditional root-mean-square error. It should be noted that one of the important components of the data analysis with the help of neural networks is data pre-processing, aimed at reducing the dimension of network inputs, increasing the joint entropy of input variables and normalizing input and output data.

C. Literature overview

There are studies where different algorithms for a continuous time series value of crude oil prices prediction are compared. W. Xie et al. [7] used in their study monthly spot prices of West Texas Intermediate (WTI) crude oil from January 1970 to December 2003 with a total of n = 408 observations. Support Vector Machine (SVM) model is estimated with other two models: AutoRegressive Integrated Moving Average (ARIMA) and Backpropogation Neural Networks (BPNN). According to the evaluation SVR has better performance than the other two algorithms [6]. Nevertheless, the authors do admit that BPNN was also outperformed in some evaluated periods of time and has the ability of detecting the nonlinear dynamics of crude oil price. There are still some scarceness and restrictions that can be improved for a better performance tool, in spite both statistical and AI models performed well in their individual approach. Later, hybrid models are introduced to treat these inconveniences. Study [8] introduced TEI@I methodology to hybrid four models for the prediction. Based on the idea of combining Text Mining, Econometrics, Intelligence (intelligent algorithm) components and integrating (@) of the mentioned, this study integrates Web-based Text Mining (WTM), Auto-Regressive Integrated Moving Average (ARIMA), Artificial Neural Network (ANN), and Rules based Expert System (RES) to predict the price. The dynamic movements of crude oil price market are also due to related irregular unexpectedly occurred events. Therefore, ARIMA and ANN are used to handle the linear and nonlinear components respectively in crude oil price while WTM and RES as the news or irregular events retriever. Later, these four models are integrated with Nonlinear Integrated Forecasting approach based on BPNN training where it makes the sum of squared errors minimal. This approach performed quite well in forecasting the crude oil price, by using monthly WTI price together with the online news as the training data. Another paper about hybrid models related to crude oil price prediction is [9], a rough-set refined text mining approach where text mining and rough-set are combined to produce useful knowledge that can be used to configure and predict the tendency of crude oil market. The benefit of this method is that it can consider both the quantitative and the qualitative factors. Variables which were used as input data are all possible events that affect the crude oil market. These events are extracted via internet and internal file system using the rough-set refined text mining approach. Other than that, world oil demand and supply, crude oil production and crude oil stock level are selected as the input variables and monthly WTI price as the output variable used. Nonetheless, this approach has shown a promising tool for forecasting the movements of crude oil market where it outperformed the other models in the evaluation process. Finally, research [10] integrates Empirical Mode Decomposition (EMD) with Feed-forward Neural Network (FNN) and Adaptive Linear Neural Network (ALNN); EMD-FNN-ALNN to formulate a combined output for the original crude oil price series. This paper used daily WTI and Brent oil price from January, 1986 to September, 2003, without taking into account public holidays. According to the experiment evaluation, we can conclude that this approach offers an alternative prediction tool to crude oil price forecasting. It also proved that the decomposition and ensemble techniques used in EM (Decomposition)-FNN (Prediction)-ALNN (Ensemble) had improved the limitations carried by other previous single models. Details of this approach can be referred in [11].

Authors of the article already have an experience in development and implementation of information

technologies in an oil field. The article [12] considers one of the most dynamically developing technologies in the world oil refining technology of delayed coking. Mathematical models of coking heavy oil residues are proposed, which allow to control the output of products, by changing the ratio of raw materials at the input, taking into account the dynamics of coking of each component. Another article [13] presents an information technology to describe the properties of oil wells in an oil production in the ontological form. Obtained application of the results can solve the problem of efficiently searching for scientific information in this direction in Internet sources.

As can be seen it is obvious that researches on finding the best method of oil price prediction are of current interest. Classical and modern as well as combined methods are using in investigations.

III. Experiment and Results

In this part, we first explain why we have chosen RapidMiner software, then describe data, define evaluation metrics for prediction purposes. Finally, results of evaluation are presented.

A. Software choice

Hierarchy analysis method was applied to find an optimal Data Science tool for our research purposes. Criteria for evaluation are cost efficiency, ease of use, customer support, functionality, performance, integration. Among alternatives are RapidMiner, KNIME, BigML and DataRobot. Tool have to be open-source for educational purposes, have a simple and intuitive interface, provide users a clear and quality documentation and forum, be able to cope with processing of big amount of data in a shortest time, have enough appropriate functions and features for forecasting and data analysis. Results presented on the table I shows that RapidMiner meets the needs. It implements the principle of visual programming, i.e. the analyst does not need to write the program code himself, as he does not need to carry out complex mathematical calculations. Everything is as follows: the user loads the data into the working field, and then simply drag and drop the operators into the GUI, forming the data processing process.

TABLE I. Results of the Hierarchy Analysis Method

RM	0,3382
KNIME	0,2918
BIGML	0,1882
DATAROBOT	0,1816

B. Data preparation

For the first set we use daily data from 25.01.2008 to 25.01.2018. Oil prices data for Brent are obtained from www.investing.com. Our sample starts from 2008 because period of ten years provides the necessary size (2583 observations) of a set for forecasting (Fig. 1). A line plot of the series is then created.

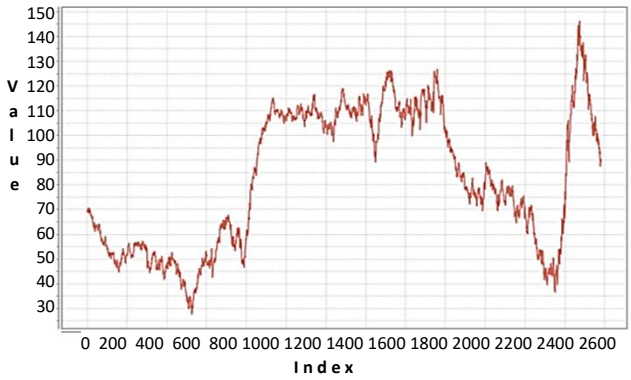

Fig. 1. The daily oil price of Brent

od of the second data set starts from the 29.11.2016 and itinues until 09.06.2017 (127 observation). This is due to fact that the first trading of the export futures for Rus Urals oil was officially opened at the St. Petersburg Inte ional Commodity Exchange (SPIMEX) in 29. 016, and lasted until 09.06.2017. Data was taken from the -site www.spimex.com.

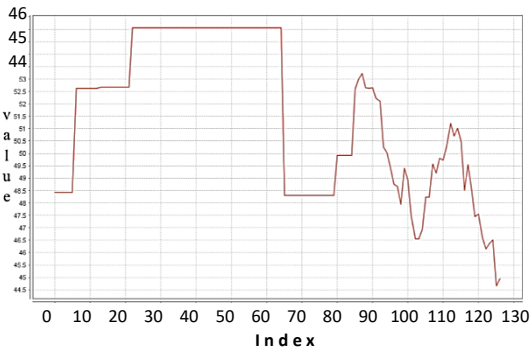

Fig. 2. The daily oil price of Urals

This choice of dataset is because of their dissimilarity. The first dataset is a lot bigger than the second one, 2583 compared to 127 observations (Fig. 2). Brent oil prices are more dynamic – price constantly changes along all the period of ten years. On the other hand, Urals oil prices have a period of several weeks when data keep the same from day to day. It would be interesting to discover what forecasting results are at the end.

C. Building of models

The process of data analysis for time series prediction consists of data collection, data selection, windowing, data division, determination of model and its architecture, subset training, subset validation, prediction and evaluation. No normalization is used in this investigation for simplicity [3].

After selecting necessary attribute, windowing needs to be applied. Windowing allows to take any time series data and transform it into a "cross-sectional" format. The traditional machine learning approach is to split an available historic dataset into two smaller sets to train a model and to further validate its performance against data that a machine hasn't seen before. The oil prices dataset was splited into two parts: a training and a test set. How to find the optimal splitting proportion can also be a good research problem. Many similar researches advise to try a series of runs with different amounts of training data. But all of them agree, that

the more observations you have the less this proportion is important. In this research 90% of data will be taken for the training dataset and the remaining 10% of data will be used for the test set. In a dataset a training set is implemented to build up a model, while a test set is to validate the model.

Linear Regression and Support Vector Machine already have ready to work operators in RM, but Convolutional Neural Network model requires to define its architecture. It consists of levels of different types which are convolutional (Fig. 3), pooling and core (dropout, flatten, dense).

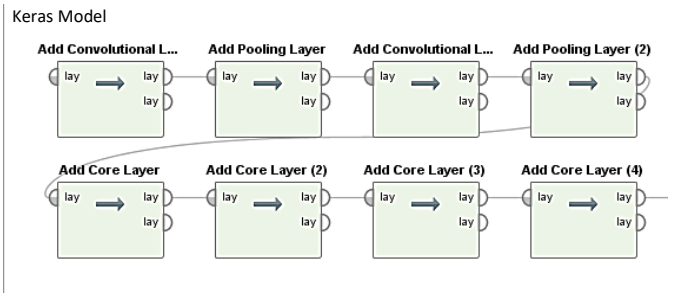

Fig. 3. Convolutional Neural Network model

D. *Evaluation metrics*

To estimate the accuracy of any measurements means to determine on the basis of the results obtained comparable numerical (quantitative) characteristics expressing the qualitative side of the measurements themselves and the conditions for their conduct. Performance evaluation metrics used in this research are Root Mean Squared Error (RMSE) and Squared Error (SE). The concept of the mean square error was introduced by Gauss, and is now accepted as the main characteristic of the prediction accuracy. The mean square error is the mean square of the sum of the error squares of the individual measurements.

E. *Results*

The model result with the smallest RMSE and SE values will be the best tool for prediction. The analysis supposes that SVM has the best forecasting power to forecast crude oil price and this implies a good prediction tool, while the Convolutional Neural Network model shows the worst performance (table II).

TABLE II. RESULTS

Input factors	**Criteria**	**Models**		
		Linear Regression	***Support Vector Machine***	***Convolutional Neural Network***
Brent	RMSE	7.240 +/- 0.408	2.130 +/- 0.107	27.592 +/- 0.000 761.322 +/- 778.987
	SE	52.590 +/- 5.845	4.549 +/- 0.446	
Urals	RMSE	0.997 +/- 0.678	0.805 +/- 0.640	23.534 +/- 0.000 553.850 +/- 723.385
	SE	1.454 +/- 1.571	1.058 +/- 1.551	

The easiest way to show the accuracy of the forecasting is to represent data as chart. As can be seen on the Fig. 4-6 it is obvious that SVM outperforms other two algorithms. Red line is input data, blue - predicted data.

Fig. 4. Linear Regression chart

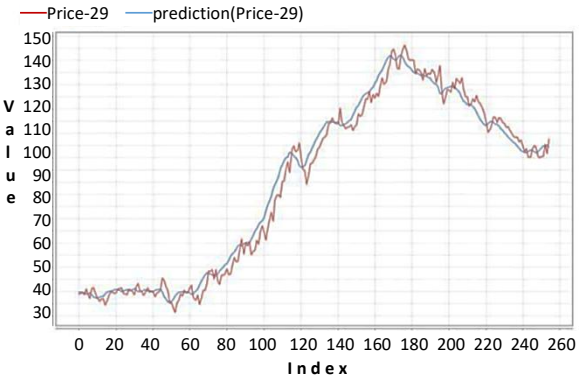

Fig. 5. SVM chart

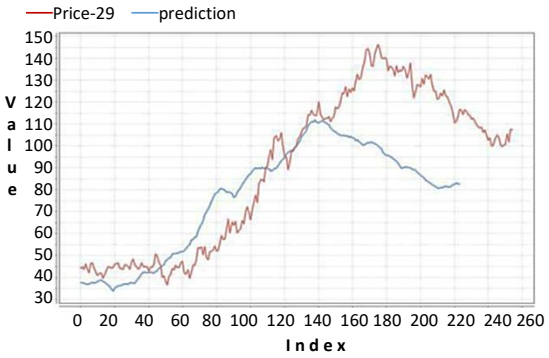

Fig. 6. Convolutional Neural Network model chart

It is common in the field of forecasting, when the used model can very well study the structure of the data set, but at the same time build a bad forecast. In this case, the model is likely to be overfit. Overfitting occurs when the model learns not only patterns in the dataset, but also the noise in data. Linear Regression and Convolutional Neural Network model tend to suffer from over-fitting problem. On the other hand, SVM is more resistant to the over-fitting problem in this case and can model nonlinear relations in an efficient and stable way.

F. *Architecture of DSS*

Considered DSS consists of a data store, a forecasting and analytical system and a hardware and technical system (Fig. 7).

The DSS data warehouse is intended for accumulation and storage of historical data on indicators such as opening price, closing price, calculation price, average price, number of transactions. The subsystem of information exchange,

which is included in the data warehouse, is designed to provide information exchange; import data from established formats and various databases into the data Warehouse, export data to established formats. For example .csv,.xml,.xlsx.

The subsystem of forecast-analytical calculations is intended for monitoring, analysis of the current situation, scenario and target forecasting of trends depending on control actions. In it the subsystem of development of administrative decisions on the set criteria supports development of administrative decisions on the basis of a complex of the set criteria.

The hardware-technological system is intended for interaction of the user with DSS, in particular for display of data in tabular, graphic and text forms.

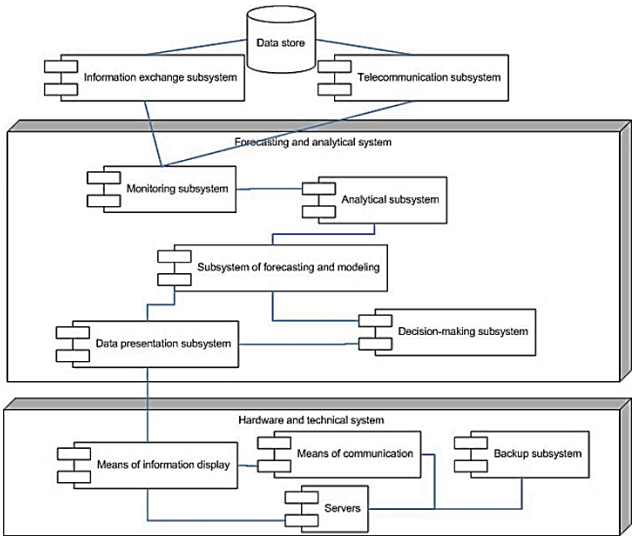

Fig. 7. DSS architecture

The architecture of the forecasting subsystem (Fig. 8) includes:

- emulator that simulates the state of the environment using various algorithms for changing system parameters in the operational database;
- machine learning-based forecasting module;
- multi-agent module of RL-training, consisting of a group of independent agents, each of which is trained on the basis of one of the developed TD-methods (TD(0), TD(λ), SARSA, Q-training), as well as used for the accumulation of knowledge about the environment and capable of adaptation, modification and accumulation of knowledge;
- Decision analysis module, can be used to analyze the data coming from the forecasting module, RL-training and make decisions on further actions.

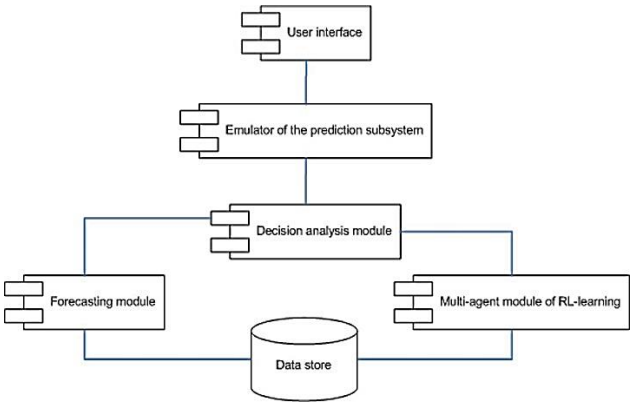

Fig. 8. Architecture of the forecasting subsystem

The forecasting module consists of a server and a user parts (Fig. 9).

The server part includes submodules responsible for import and validation of data, generation and preparation of reports, server and service procedures, and libraries (methods of forecasting, risk assessment).

The user part is designed to display the results of forecasting and evaluation, charting and reporting.

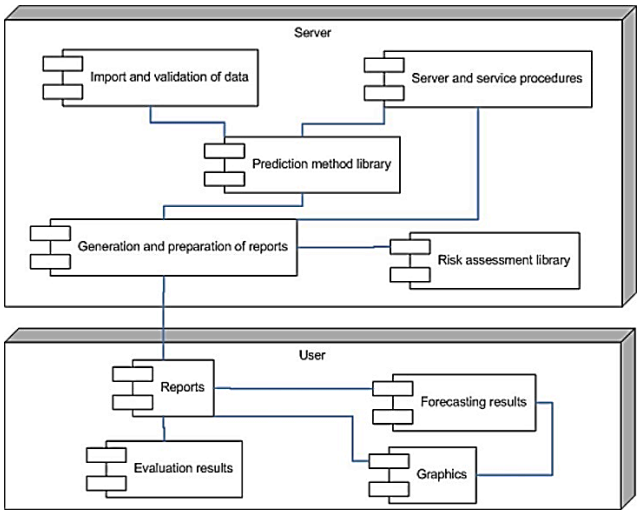

Fig. 9. Forecasting module

As a result of this section, a model of the DSS architecture, a forecasting subsystem and a forecasting module were designed to show the practical value of the study.

IV. Discussions and Future Research

Crude oil plays a significant role in the world economy, and especially in the energy sector. Therefore, accurate predictions of oil prices are vital. Of late, Neural Networks are being widely used as an effective and efficient tool for forecasting purposes. We guess, that there is an increased trend amongst the researchers to apply them for the forecasting efficiency and accuracy.

In this research on the chosen algorithms (Linear Regression, Support Vector Machine, Convolutional Neural Network) three models that can be applied for short-term forecasting of oil prices were constructed. Although SVM has quite good results, we are sure that the choice of the

forecasting algorithm for price is individual and depends on the amount and structure of the input data. The idea was put forward to create an information system that could choose an algorithm for given values of the input parameters or combine several algorithms to further predict prices, so it could be used in economic information systems by stock exchange participants, industrial enterprises and governments of countries.

More research should be dedicated to developing more effective and efficient methods in feature selection and prediction of oil price to assess more successful outcomes in the future. Among the most promising is the development of an information system that could choose an algorithm itself depending on the given values of the input parameters and also the development of hybrid models that could combine the advantages of several algorithms, and thus mutually exclude each other's shortcomings.

REFERENCES

[1] Anselmo, Peter & Hovsepian, Karen & Ulibarri, Carlos & Kozloski, Michael, "Automated Options Trading Using Machine Learning", 2018.

[2] Lubna A Gabralla, Ajith Abraham, "Computational Modeling of Crude Oil Price Forecasting: A Review of Two Decades of Research," International Journal of Computer Information Systems and Industrial Management Applications, vol. 5 (2013), pp. 729-740.

[3] Dietrich Domanski, Jonathan Kearns, Marco Lombardi, Hyun Song Shin, "Oil and debt," BIS Quarterly Review, March 2015.

[4] Stefan Avdjiev, Valentina Bruno, Catherine Koch, Hyun Song Shin, "The dollar exchange rate as a global risk factor: evidence from investment1," The IMF 18th Jacques Polak Annual Research Conference, 2017.

[5] Bassam Fattouh, Andreas Economou, "Oil Price Paths in 2018: The Interplay bertween OPEC, US Shale and Supply Interruptions," The Oxford Institute for Energy Studies, publications, 13.02.2018.

[6] S.N. Abdullah, "Machine learning approach for crude oil price prediction," A thesis submitted to The University of Manchester for the degree of Doctor of Philosophy in the Faculty of Engineering and Physical Sciences, 2013.

[7] Xie, W., Yu, L., Xu, S., & Wang, S. "A new method for crude oil price forecasting based on support vector machines." In Computational Science–ICCS ,Springer Berlin Heidelberg,pp. 444-451, 2006.

[8] S. Wang, L. Yu, & K.K. Lai, "Crude Oil Price Forecasting With TEI@I Methodology," Journal of Systems Science and Complexity, vol. 18, no. 2, pp. 145-166, 2005.

[9] L. Yu, S. Wang & K.K. Lai, "A Rough-Set-Refined Text Mining Approach for Crude Oil Market Tendency Forecasting," International Journal of Knowledge and Systems Sciences, vol. 2, no. 1, 2005.

[10] L. Yu, S. Wang & K.K. Lai, "Forecasting crude oil price with an EMD-based neural network ensemble learning paradigm," Energy Economics, vol. 30, no. 5, pp. 2623-2635, 2008.

[11] S. N. Abdullah and X. Zeng, "Machine learning approach for crude oil price prediction with Artificial Neural Networks-Quantitative (ANN-Q) model," Proceedings of the International Joint Conference on Neural Networks, pp. 1-8, 2010.

[12] V. P. Zaporin, S. V. Sukhov, M. Yu. Dolomatov, N. A. Zhuravleva, A. R. Galiakbirov, V. V. Martynov, and A. V. Kutueva, "Mathematical model for predicting yield of heavy oil residue carbonization products," Chemistry and Technology of Fuels and Oils, vol. 53, No. 2, May, 2017.

[13] V.V. Martynov, E.I. Filasova, Yu.V. Sharonova., L.P. Fandrova, "Development of the ontological analysis technology for the description of oil production processes," Bulletin of the USATU: Collection of scientific papers, vol. 17, No. 5 (58), pp. 188-194, 2013. (in Russian).