International Journal of Computer Information Systems and Industrial Management Applications.
ISSN 2150-7988, Volume 7 (2015) pp. 001-013

Comparison of Hybrid Intelligent Approaches for Prediction of Crude Oil Price

Lubna A.Gabralla[1] and Ajith Abraham[2, 3]

[1] Faculty of Computer Science & Information Technology,
Sudan University of Science and Technology,
Khartoum, Sudan
lubnagabralla@gmail.com

[2]Machine Intelligence Research Labs, WA, USA
[3]IT4Innovations, VSB - Technical University of
Ostrava, Czech Republic
ajith.abraham@ieee.org

Abstract: **Crude oil price prediction is a challenging task due to its complex nonlinear and chaotic behavior. There is a great need for oil price volatility measuring and modeling of oil price chaotic behavior. During the last couple of decades, both academicians and practitioners have devoted proactive knowledge to address this issue. Combined predictors are one of the most promising forms in Machine learning (ML). It can be found in different styles in the literature such as Meta learning, Ensemble based prediction, Hybrid methods and more. The aim of this paper to conduct comprehensive comparisons among the combined prediction model in order to improve the performance**

***Keywords*:** Crude oil prediction, Meta prediction models, Hybrid models, Ensemble prediction model, ANFIS, PSO**.**

I. Introduction

Crude oil formation occurs by a mix of hydrocarbons that exists in liquid phase in natural underground reservoirs with certain minerals such as sulfur under extreme pressure and remains liquid at atmospheric pressure after passing through surface separating facilities [1, 2]. The price of a particular crude oil is usually set at a discount or at a premium to a marker or reference price (benchmarks) for buyers and sellers of crude oil [3]. There are several international benchmarks of pricing system and the most popular are West Texas Intermediate (WTI). Recently world had suffered from political instability, wars and conflicts, especially in the Middle East oil-rich areas, such as the Arab Spring movements in Tunisia, Libya, Egypt, Syria and Yemen. With the acceleration of technological development, these factors and others had influences on the oil market and volatile behavior of trading. Therefore crude oil prices are characterized by high volatility and some drastic shocks [4], and the dominant feature of the behavior of the oil prices is becoming is very chaotic. The crude oil prediction problem is one case of data mining using regression approach. [5, 6] showed that in the last years this particular area of research development. There are various successful applications of data mining techniques in real-life situations. In the banking and financial services domain, identifying customers who are most likely interested in a new credit product is one instance of a data mining application. Another example is telecommunications fraud prediction in mobile telephony and services [7]. According to Gorunescu [7], machine learning(ML) represents an extremely important scientific discipline in the development of data mining, using techniques that allow the computer to learn with training. The role of the learning task is to search more efficiently for a solution of problems. Learning systems include different components such as a set of examples, a set of possible learning results and a learning algorithm [8]. ML algorithms take a different scenario, according to the scope of the level of adaptation. Combined predictors are one of the most promising forms in ML. It can be found in different styles in the literature such as Meta learning [9], Ensemble based prediction [10], Hybrid methods [11] and more. Meta-learning succeeded on the appropriate selection of a suitable predictive model (or combination of models) depending on the domain of application by providing mapping automatic for a suitable model to a particular task. The hybridization of the artificial intelligence techniques can provide solutions to complex, nonlinear, and volatile crude oil price prediction. Neuro-fuzzy approach refers to combinations of artificial neural network learning and fuzzy logic. Neuro-fuzzy incorporates the human-like reasoning style of fuzzy systems through the use of fuzzy sets and a linguistic model consisting of a set of *if-then* fuzzy rules. The main strength of neuro-fuzzy systems is that they are universal approximates with the ability to solicit interpretable *if-then* rules and accuracy [12]. Moreover ensemble methods are one of the latest techniques that promise results more effective. The ensemble method depends on the behavior that a collection of predictor such as machine learning algorithms (neural network, support vector machine, decision trees and so on) can do better than the individual approaches. Verikas, et al. [10] explained a distinction between a hybrid system and an ensemble of predictors. A system is considered as hybrid if several soft computing approaches are exploited for data analysis, but only one direct predictor is applied to make a final decision and to obtain a final decision. In an ensemble, outputs of multiple predictors are combined in various ways. The aim of this study is to conduct the comparison between combined prediction models in order to explain the advantages and disadvantages of each method to help the researchers to design better models to predict oil prices and problems similar to it. The rest of this paper is organized as follows: literature review is presented in Section 2 while Section3 describes the research methodology in detail. The data used and their divisions are found in Section 4, experimental results are reported in Section 5, Comparison analysis of combined prediction models presented in Section 6 and finally section 7contains the concluding remarks.

II. Literature Review

Hybrid methods have emerged in the recent years, the basic idea of which was to complement the disadvantages of the individual models and generate synergy effect on the results to predict oil prices. Wang, et al. [13] developed a hybrid AI system of neural networks rule based expert system and web–based text mining using historical data of monthly spot prices of crude oil collected from January 1970 to December 2002. The results illustrated that the performance of hybrid forecast were more accurate than single neural network forecasting. Mathematical and statistical methods are not well suitable for expression of human experiences such as perception, logic and uncertain concepts. A fuzzy inference system employing fuzzy rules can then provide a framework to model human. On the other hand, Artificial Neural Network (ANN) learning mechanism is hard to extract structured knowledge from either the weights or the configuration of the ANN. To overcome these drawbacks and to take advantages of these two approaches an integrated neuro-fuzzy system was built called ANFIS. Panella, et al. [14] collected data from Europe (Brent crude oil) and the US (West Texas Intermediate crude oil) from 2001 to 2010 to forecast crude oil, natural gas, electricity, and coal prices using three different models radial basis function neural networks, adaptive neuro-fuzzy inference system networks and least-square approximation. The experimental results showed the superiority of adaptive neuro-fuzzy inference system. Ghaffari and Zare [15] applied an adaptive network-based fuzzy inference system for forecasting WTI crude oil spot price. Using daily data from 5 January 2004, to 30 April 2007, 68.18% prediction accuracy was achieved. In order to enhance the effectiveness of the artificial intelligence techniques, an ensemble machine learning approach was built for the prediction of crude oil price Yu, et al. [16] constructed an empirical mode decomposition (EMD) based on neural network ensemble learning. They used daily West Texas Intermediate (WTI) data from 1/1/1986 to 30/9/2006 as training and Brent from 20/5/1987 to 30/9/2006 as test data. Results proved that EMD based neural network ensemble could be used for oil price prediction. Yu, et al. [17] obtained different prediction results based on different training sets and variety of models including (ARIMA) model, support vector machine regression (SVMR) model, back-propagation neural network (BPNN) model and, radial basis function network (RBFN) model, and then the results were combined using a fuzzy ensemble model. Their results indicated that the proposed model was superior to the single models for oil price prediction.

III. Research Methodology

A. Meta prediction models

Meta-learning involves several algorithms. We used a popular set of this technique namely Bagging, Random subspace, Ensemble selection, Voting and Stacking. We grouped Meta depend on their mechanism into two parts: Bagging, Random subspace which separate data into subparts each part train by same predictor. Another part includes Ensemble selection, Voting, and Stacking, which providing same input to a number of predictors and combine their output using a given decision logic. As mentioned previously, Meta learning helps to create optimal predictive models and reuse previous experience from analysis of other problems, such that the modified learner is better than the original learner at learning from additional experience.

1) Bagging

Breiman [18] introduced bagging methods, which is basically a multiple predictor combined to get the final result through bootstrap replicates. As illustrated in Figure 1, we first generated yield subset $SBDS_i$ by replacing the original data set DS (bootstrapping) many times, then compute a sequence predictor Pi by classifier C_i and then we used the same machine learning scheme for each sub-dataset SBDSi. Finally, aggregate the results by voting (averaging) to access the last from P_{final}. This method is characterized that it could enhance performance[19]and reducevarianceto improve generalization [20] [21].

2) The Random Subspace Method (RSM)

RSM is the combining technique proposed by [22]. We selected an r-dimensional random subspace from the original p-dimensional feature space X , where r < p.
$X^b = \left(X_1^b, X_2^b, \ldots\ldots, X_r^b\right)$ .Then one constructs predictor in the random subspaces X^b and combines them by simple majority voting in the final decision rule. We repeated this process for b = 1, 2, 3, 4, 5 according to number of direct predictors.

3) Ensemble selection algorithm [23]:

The first step we created a "model library". This library should be a large and diverse set of direct prediction models and the presence of a number of Prediction Models (PM) are denoted by PM1, PM2, PM3, PMn with different parameters. The second step is to combine the outputs of these models from our library with the Ensemble Selection algorithm. To prevent over-fitting, we have an ensemble with only one model in it. Then, we added models one at a time to our ensemble, to figure out which model to add, each time: individually average the predictions of each model from the library currently being considered with the current ensemble. Then pick the model that provided the most performance improvement Figure 2 explains these steps.

4) Voting

The simplest kind of ensemble is the way of aggregating a collection of prediction values with each base level giving different voting power for its prediction. The final prediction obtains the highest number of votes. Voting includes the weighted average (of each base classifier holds) when using regression problem and majority voting when doing classification, the weighted-majority output is:

$$\text{Argmax}\left(\sum_{j=1}^{k} p_i(x), w_i\right) \quad (1)$$

Pi(x) is the results of the prediction of i[th] prediction model and Pi(x, w) is indicator function defined as:

$$p_i(x,w) = \begin{cases} 1 & x = w \\ 0 & otherwise \end{cases} \quad (2)$$

Majority voting has some benefits that it does not require any additional complex computation and any previous knowledge. However, this approach leads to the result that it is difficult to analyze and interpret. The second strategy is un-weighted, which gives some predictor higher weight if they achieve more accuracy than others (the winner is the one with the most number of votes) [24, 25].

5) Stacking:

Another popular approach to combine predictors is called stacking or stacked generalization. It is a function that depends on the Meta learner (level-1 model), which concerns with the combination of the predictions of the numerous predictors generated by using different learning processes on a direct dataset, and base classifiers (level-0 models) to obtain the final prediction [26], according to which features and algorithms are used in the Meta. According to Figure 3, we used four direct predictors in level-0 model as inputs to Random subspace in level-1 and we generated training data for level-1 by using cross-validation model.

B. Hybrid prediction model

Hybrid intelligence techniques are a combination of multiple methods to build an efficient solution to deal with a particular problem and in recent years it is considered as a powerful tool to improve the accuracy [9]. ANFIS is one distinct example of the hybrid system. It is a good model to explore and propose a decision making system by extracts information (input) and compute it in the system automatically, thus producing a decision (output) based on information from the extracted crude oil market's rules.

1) Adaptive Neuro fuzzy inference system (ANFIS)

Mathematical and statistical methods are not well suitable for expression of human experiences such as perception, logic and uncertain concepts. A fuzzy inference system [27] employing fuzzy if-then rules can provide a framework to model human knowledge. Takagi, Sugeno and Kang (TSK) [28] proposed a fuzzy inference method in which the conclusion of a fuzzy rule is constituted by a weighted linear combination of the crisp inputs rather than a fuzzy set. There is no systematic way to transform experiences of knowledge of human experts to the knowledge base of a fuzzy inference system (FIS). On the other hand, Artificial Neural Network (ANN) learning mechanism hard to extract structured knowledge from either the weights or the configuration of the ANN. To overcome these drawbacks and to take advantages of these two approaches integrated system was built by combining the concepts of (FIS) and (ANN) called Adaptive Neuro Fuzzy Inference System (ANFIS). ANFIS implements a Takagi Sugeno Kang (TSK) fuzzy inference system. For a first order TSK model, a common rule set with two fuzzy if-then rules is represented as follows:

Rule 1: If x is A1 and y is B1, then f1 = p1x + q1y + r1
Rule 2: If x is A2 and y is B2, then f2 = p2x + q2y + r2

Where x and y are linguistic variables and A1, A2, B1, B2 are corresponding fuzzy sets and p1, q1, r1 and p2, q2, r2 are linear parameters. ANFIS makes use of a mixture of back propagation to learn the premise parameters and least mean square estimation to determine the consequent parameters. Figure 4 illustrates the ANFIS structure.

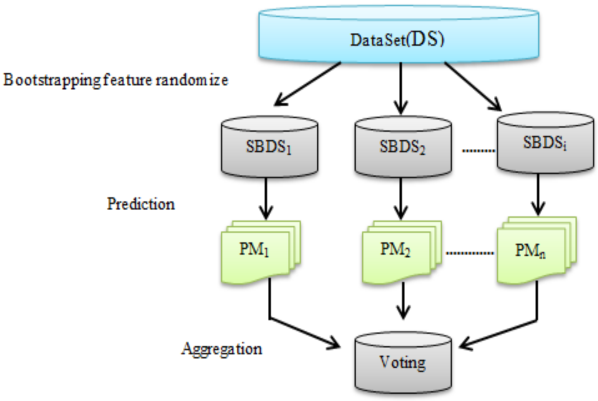

Figure 1. Bagging ensemble methods

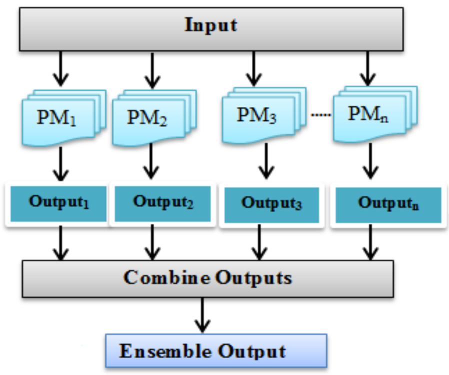

Figure 2. Ensemble method framework

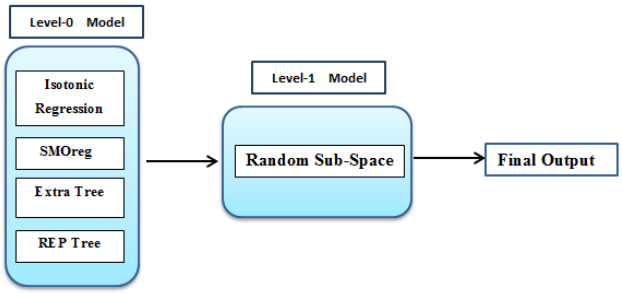

Figure 3. Stacking structure for prediction crude oil prices

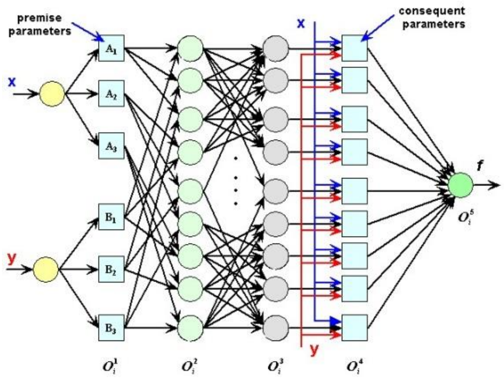

Figure 4. The architecture of the ANFIS [29]

C. Ensemble prediction model

In an ensemble, outputs of multiple predictors are combined in various ways. Ensemble methods are one of the latest techniques that promises results more effective in different applications such as pattern recognition [30], machine learning, data mining [31] and medical applications [32]. Ensemble methods have two phases: the first phase is the production of the different models [33] such as Bagging and Boosting. Sometime this phase is also recognized as Meta learning for instance Vilalta, et al. [34] considered that Stacked generalization and Boosting a form of meta-learning while Maclin and Opitz [35] indicated that Bagging and Boosting are two popular methods for creating accurate ensembles in addition to Džeroski and Ženko [26] used Stacking to ensemble of classifiers. Similarly Blachnik [36] presented Voting, Staking, Bagging and Boosting as examples of ensemble learning. Menahem, et al. [37] defined Meta based on Ensemble "Meta-learning is the process of learning from basic classifiers (ensemble members); the inputs of the meta-learner are the outputs of the ensemble-member classifiers". The second phase of an Ensemble method is the combination of the models [33]. The basic ensemble method and generalized ensemble method are the most popular techniques used in this phase.

1) Basic ensemble method (BEM)

We employed the basic ensemble method (BEM) as defined by:

$$F_{BEM} = \frac{1}{n} \sum_{i=1}^{n} F_i(X) \tag{3}$$

Where Fi (x) is the output produced by the different models. This approach by itself can lead to improved performance, but does not take into account the fact that some networks may be more accurate than others. It has the advantage of being easy to understand and implement and is often found not to increase the expected error [38]. The ensemble method depends on the behavior that a collection of predictor such as machine learning algorithms (neural network, support vector machine, decision trees and so on) can do better than the individual approaches. Predictors are combined through some weighted average or weighted combination.

2) The generalized ensemble method

Find weights for each output that minimizes the Root Mean Square Error (RMSE) or Mean Absolute Error (MAE) of the ensemble. The general ensemble model (GEM) is defined by:

$$F_{GEM} = \sum_{i=1}^{n} \alpha F_i(X) \tag{4}$$

$$\sum_{i=1}^{m} \alpha = 1 \tag{5}$$

Where α Fi (x) are chosen to minimize the MAE between the outputs and the desired values. Finding the optimal values of α is not an easy task. We used a Particle swarm optimization (PSO) method to determine the optimal weights. Particle swarm optimization [39] is a technique for simulating the social and cooperative behavior of different types such as birds, fish, bees and human beings. The PSO composed of a population (swarm) of possible solutions called particles. These particles move through the search domain with a specified velocity in search of optimal solutions. Each particle maintains a memory, which helps it in keeping the track of its previous best position.

IV. Data set and Experimental Environment

A. Dataset Description

The dataset for experiments are obtained cooperative by Faculty of Management and Economic Sciences, Sousse University, Tunisia. It consists of 3337 records as instances and 14 variables as attributes to predict the West Taxes Intermediate (WTI) as output. The data set was taken from different sources such as [40, 41]. Attributes are listed as below:

Date (DT): The daily data from 4 January 1999 to 10 October 2012. Dates are converted to numeric form when the input file is read.

- **West Texas Intermediate (WTI):** It is the most famous benchmark [16], and plays an important role as a reference point to determine the price, and it constitutes a crucial factor in the configuration of prices of all other commodities [42].

- **Federal Fund rate (FFR):** One of the most influential interest rates in the U.S. economy, because it effects on monetary and financial conditions, which in turn have an impact on fundamental aspects of the broad economy including employment, growth and inflation [43].

-
- **Volatility Implied Equity Index (VIX):** Measures the contribution of the instability of the market.
- **The regional Standard and Poor's equity index (SPX):** Represent the market performance.
- **New York Harbor conventional gasoline spot prices (GPNY):** As example to assesses oil products.
- **US Gulf Coast conventional gasoline spot prices (GPUS):** As example to assesses oil products.
- **New York Harbor No. 2 heating oil spot price (HP):** As indication of seasonality in the energy market.
- **Future contracts (FC1- FC2- FC3- FC4):** For WTI to maturity traded on NYMEX
- **Exchange rate (ER):** The price of oil and exchange rates of other currencies against the U.S. Dollar price.
- **Gold prices (GP):** Gold is that less volatile than crude oil and could reflect the real trend in the commodity market rather than the noise and gold used as the results of investors hedge against inflation caused by the oil price shock [44].

B. Data preprocessing

Before constructing a model we selected several aspects of initial preparation of data. Feature selection, normalization and data partition are used for preparation the inputs. It is worth mentioning that these steps are often used when designing any model in this research. We implemented first feature selection methods which is defined as a process of selecting a subset of features, d, out of the larger set of D features, which maximize the classification or prediction performance of a given procedure over all possible subset data. The second method is normalization, which shifts the instance values in specific and obviously means to represent information contained within the data and the data set [45]. Finally divided the dataset to groups according to deferent percentages of training and testing.

1) Feature selection methods

We formulated 7 different sub datasets, which were derived from the original dataset after implementing the several attribute selection algorithms. For instance SBDS1 and SBDS2 are as a result of Correlation based Feature Selection (CFS) algorithm by evaluating the value of a group of attributes by concerning the individual predictive ability of each feature as well with the possibility of redundancy among the features with several search methods such as best-first, which keeps a list of all attribute subsets evaluated so far, sorted in order of the performance measure. We used Forward selection, where we start with no attributes and add them one at a time and Backward, where we start with all the attributes and delete each one at a time, stops when the addition/deletion of any residual attributes results in a decrease in evaluation. In a case of one, begin with all the attributes or with none of them and this called bidirectional search method [46]. In SBDS3 and SBDS4, we utilized Genetic algorithm, which is based on search processes on the principle of natural selection [46].We performed forward selections with a limited number of k attributes, based on the ranking using training data to decide, which attribute is added in each iteration of forward selection, and the test data is only used to evaluate the “best”, P best subsets of a particular size. To determine the “optimal” subset size, we average the P scores on the test data for each subset size, and choose the size with the highest average. Then, a final forward selection is performed on the complete data set to find a subset of that optimal size and SBDS6 is created. We used ranker as search method, which Ranked the list of attributes based on individual evaluation of each attribute [47]. SBDS9and SBDS10 used wrapper algorithm, which evaluate attribute sets by using SMOreg algorithm. It is called wrapper because the learning algorithm is wrapped into a selection task [46]. We implemented the best-first search method in two directions: forward and backward respectively. Table 1 illustrates the categories and attributes for each algorithm.

2) Normalization

Most models work well with normalized data sets the data were normalized using Eq. (6) by scaling the instance to the range between -1 and 1 to improve prediction accuracy and CPU processing time [48] .

$$n_o = \frac{k_i - x_{min}}{p_{max} - x_{min}} \tag{6}$$

Where n_0 = normalized dataset k_i = raw dataset, x_{min} = minimum value of the dataset and p_{max} = maximum value of the dataset.

3) Data Partition

There are various alternatives to recognize the training and testing split process such as cross-validation, bootstrap and holdout [46]. According to holdout method, we divided dataset randomly into two parts, one half of training and the other half for testing. It is common to hold out one-third of the data for testing and use the remaining two-thirds for training [46]. However, several researchers achieved good results with other divisions, for example Lai, et al. [49] created their model using 60% for training and 40% for testing while Yu, et al. [50] utilized 80% for training and 20% for testing. We investigated the effect of training and testing data by randomly splitting them as follows.

90% - 10% (A)
80% - 20% (B)
70% - 30% (C)
60% - 40% (D)

We used several percentages to increase the opportunities for achieving better results. In the literature, there are also some studies conducted by using such divisions for training and test data [51] .

Table 1. Attribute selection methods and their features

Sub Dataset	Attributes evaluator	Search method	Attributes
$SBDS_1$	Correlation based Feature Selection subset evaluator	Best-first-Forward	WTI,SPX,FG1
$SBDS_2$	Correlation based Feature Selection subset evaluator	Best-first- Backward	DT,VIX,WTI,SPX,GPNY GPUS,HP,ER,FC1,FC2,FC3,FC4
$SBDS_3$	Correlation based Feature Selection subset evaluator	Genetic	VIX,WTI,GPNY, ER, FC1
$SBDS_4$	Correlation based Feature Selection subset evaluator	Genetic	WTI,GPNY,FC1
$SBDS_6$	Correlation based Feature Selection subset evaluator	Subset Size Forward Selection	VIX,WTI,GPNY, FC1
$SBDS_9$	Wrapper subset evaluator (SMOreg)	Best-first- Forward	WTI,GPUS
$SBDS_{10}$	Wrapper subset evaluator (SMOreg)	Best-first- Backward	WTI,FC1

V. Experimental Results

On the one hand, we used seven sub datasets (SBSD1, SBDS2, SBDS3, SBDS4, SBDS6, SBDS9 and SBDS10) which is derived from the original dataset by using several attribute selection algorithms mentioned in Table 1 and on the other hand we used four groups (A-B-C-D), which contain different training and testing percentages as displayed previously. We used five direct prediction models namely Isotonic regression, SMOreg, IBK ExtraTree, REPTree, the experiments for each algorithm were already published in [52] and [53]. It is worth mentioning that we repeated the training and testing experiments ten times with different random sample for each sub dataset to guarantee that the full dataset represented in the training and testing sets in the correct way and the error rates on the different iterations are averaged to yield an overall error rate. To judge the prediction performances and evaluate the accuracy of prediction, there are two basic criteria: the Mean Absolute Error (MAE) and Root Mean Square error (RMSE).The smaller the value of the evaluation indexes, the higher the performance of the algorithm.

A. Meta Prediction Experiments

We grouped Meta prediction models depending on two parts: Bagging and Random subspace, which separate data into subparts and each part is trained by the same predictor. Another part including Ensemble selection, Voting, and Stacking, which provides the same input to a number of predictors and combine their output using a given decision logic. First we implemented Bagging and Random-subspace with five direct prediction models. In this experiment, we search to improve prediction model results by using the Bagging model. We used 7 sub-datasets with four categories of training and testing and finally calculated the error by using MAE and RMSE for all prediction models as illustrated in Tables 2 and 3 respectively. According to Figure 5, we compared five prediction models with Bagging. As evident, bagging is significantly more accurate than the prediction models except IBK, which is also less accurate comparing with another direct prediction models, therefore we removed it from next set of experiments.

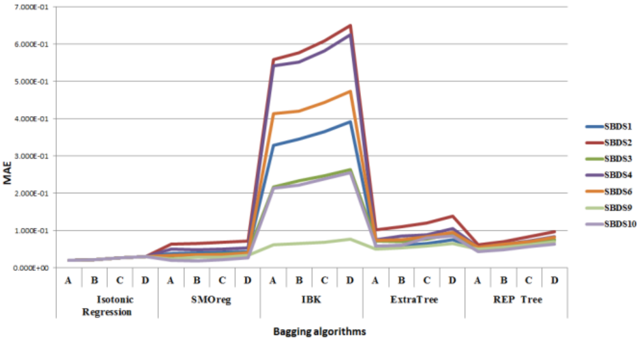

Figure 5. Bagging algorithm with 5 base prediction models

Similar to the Bagging experiments, 7 sub-data sets with 4 categories was exposed to Random subspace method. Number of single predictors was squeezed to four algorithms when Random subspace is used as a result of the exclusion of IBK algorithm due to their poor performance. MAE and RMSE results are illustrated in Tables 3 and 4 respectively. The main goal for this experiment is to investigate the best results for modeling oil prices. We combined the four prediction models as mentioned above using random subspace and then compared their results with bagging methods. Figure 6 shows that Random subspace works better than bagging with Extra Tree and REP Tree models, while bagging better than Random subspace with Isotonic Regression and SMOreg.

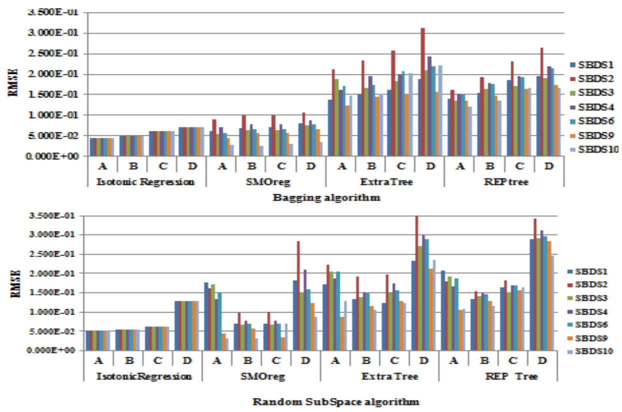

Figure 6. Comparison between Bagging and Random Subspace using 7 sub datasets and 4 categories of training and testing.

International Journal of Computer Information Systems and Industrial Management Applications.
ISSN 2150-7988, Volume 7 (2015) pp. 001-013
© MIR Labs, www.mirlabs.net/ijcisim/index.html

Table 2. Bagging algorithms with 5 base prediction models using MAE

Prediction Model	Data	$SBDS_1$	$SBDS_2$	$SBDS_3$	$SBDS_4$	$SBDS_6$	$SBDS_9$	$SBDS_{10}$
Bagging Isotonic Regression	**A**	**2.060E-02**	**2.060E-02**	**2.060E-02**	**2.060E-02**	**2.060E-02**	**2.060E-02**	**2.060E-02**
	B	2.230E-02	2.230E-02	2.230E-02	2.230E-02	2.230E-02	2.230E-02	2.230E-02
	C	2.600E-02	2.600E-02	2.600E-02	2.600E-02	2.600E-02	2.600E-02	2.600E-02
	D	3.020E-02	3.020E-02	3.020E-02	3.020E-02	3.020E-02	3.020E-02	3.020E-02
Bagging SMOreg	**A**	3.840E-02	6.400E-02	3.200E-02	4.930E-02	3.330E-02	2.390E-02	1.970E-02
	B	3.970E-02	6.570E-02	3.450E-02	4.880E-02	3.550E-02	2.820E-02	**1.840E-02**
	C	4.180E-02	6.830E-02	3.600E-02	5.050E-02	3.590E-02	2.830E-02	2.250E-02
	D	4.380E-02	7.190E-02	3.950E-02	5.410E-02	3.930E-02	3.330E-02	2.620E-02
Bagging IBK	**A**	3.290E-01	5.592E-01	2.169E-01	5.416E-01	4.131E-01	**6.160E-02**	2.141E-01
	B	3.451E-01	5.764E-01	2.327E-01	5.526E-01	4.208E-01	6.450E-02	2.212E-01
	C	3.650E-01	6.080E-01	2.468E-01	5.817E-01	4.433E-01	6.870E-02	2.391E-01
	D	3.927E-01	6.501E-01	2.638E-01	6.249E-01	4.738E-01	7.660E-02	2.560E-01
Bagging Extra Tree	**A**	5.630E-02	1.023E-01	7.120E-02	7.500E-02	7.360E-02	**4.950E-02**	5.910E-02
	B	6.010E-02	1.105E-01	6.930E-02	8.440E-02	7.450E-02	5.360E-02	6.050E-02
	C	6.520E-02	1.209E-01	7.710E-02	8.900E-02	8.390E-02	5.810E-02	7.980E-02
	D	7.490E-02	1.381E-01	8.970E-02	1.049E-01	9.510E-02	6.510E-02	8.450E-02
Bagging REP Tree	**A**	5.230E-02	6.160E-02	5.110E-02	5.570E-02	5.520E-02	4.980E-02	**4.330E-02**
	B	5.800E-02	6.980E-02	5.840E-02	6.380E-02	6.290E-02	5.540E-02	4.850E-02
	C	6.690E-02	8.310E-02	6.480E-02	7.140E-02	7.040E-02	6.160E-02	5.690E-02
	D	7.590E-02	9.640E-02	7.520E-02	8.300E-02	8.160E-02	7.040E-02	6.290E-02

Table 3. Bagging algorithms with 5 base prediction models using RMSE

Prediction Model	Data	$SBDS_1$	$SBDS_2$	$SBDS_3$	$SBDS_4$	$SBDS_6$	$SBDS_9$	$SBDS_{10}$
Bagging Isotonic Regression	**A**	**4.560E-02**	**4.560E-02**	**4.560E-02**	**4.560E-02**	**4.560E-02**	**4.560E-02**	**4.560E-02**
	B	4.910E-02	4.910E-02	4.910E-02	4.910E-02	4.910E-02	4.910E-02	4.910E-02
	C	6.160E-02	6.160E-02	6.160E-02	6.160E-02	6.160E-02	6.160E-02	6.160E-02
	D	7.160E-02	7.160E-02	7.160E-02	7.160E-02	7.160E-02	7.160E-02	7.160E-02
Bagging SMOreg	**A**	6.060E-02	9.050E-02	5.380E-02	7.040E-02	5.640E-02	4.530E-02	2.700E-02
	B	6.910E-02	9.910E-02	6.380E-02	7.690E-02	6.670E-02	5.570E-02	**2.610E-02**
	C	7.070E-02	9.980E-02	6.480E-02	7.780E-02	6.680E-02	5.660E-02	3.070E-02
	D	7.970E-02	1.078E-01	7.500E-02	8.740E-02	7.770E-02	6.700E-02	3.510E-02
Bagging IBK	**A**	4.920E-01	7.841E-01	3.486E-01	7.776E-01	6.475E-01	**1.310E-01**	3.673E-01
	B	5.310E-01	8.412E-01	4.121E-01	8.091E-01	6.679E-01	1.447E-01	3.757E-01
	C	5.491E-01	8.831E-01	4.278E-01	8.476E-01	6.976E-01	1.485E-01	4.030E-01
	D	5.936E-01	9.560E-01	4.635E-01	9.324E-01	7.707E-01	1.701E-01	4.309E-01
Bagging Extra Tree	**A**	1.383E-01	2.129E-01	1.888E-01	1.622E-01	1.724E-01	**1.231E-01**	1.472E-01
	B	1.514E-01	2.342E-01	1.661E-01	1.964E-01	1.735E-01	1.444E-01	1.510E-01
	C	1.610E-01	2.581E-01	1.825E-01	2.003E-01	2.061E-01	1.521E-01	2.035E-01
	D	1.882E-01	3.129E-01	2.099E-01	2.441E-01	2.195E-01	1.572E-01	2.225E-01
Bagging REP Tree	**A**	1.401E-01	1.621E-01	1.353E-01	1.527E-01	1.509E-01	1.343E-01	**1.204E-01**
	B	1.556E-01	1.926E-01	1.641E-01	1.775E-01	1.773E-01	1.468E-01	1.354E-01
	C	1.848E-01	2.301E-01	1.724E-01	1.949E-01	1.937E-01	1.648E-01	1.661E-01
	D	1.949E-01	2.648E-01	1.903E-01	2.189E-01	2.152E-01	1.746E-01	1.662E-01

Table 4. MAE for Random Subspace

Prediction Model	Data	$SBDS_1$	$SBDS_2$	$SBDS_3$	$SBDS_4$	$SBDS_6$	$SBDS_9$	$SBDS_{10}$
Random subspace Isotonic Regression	A	**2.220E-02**	**2.220E-02**	**2.220E-02**	**2.220E-02**	**2.220E-02**	**2.220E-02**	**2.220E-02**
	B	2.420E-02	2.420E-02	2.420E-02	2.420E-02	2.420E-02	2.420E-02	2.420E-02
	C	2.780E-02	2.780E-02	2.780E-02	2.780E-02	2.780E-02	2.780E-02	2.780E-02
	D	3.250E-02	3.250E-02	3.250E-02	3.250E-02	3.250E-02	3.250E-02	3.250E-02
Random subspace SMOreg	A	4.350E-02	6.510E-02	3.040E-02	4.850E-02	3.520E-02	**2.210E-02**	2.320E-02
	B	4.000E-02	6.770E-02	3.660E-02	4.930E-02	3.680E-02	2.630E-02	2.440E-02
	C	3.960E-02	6.880E-02	3.820E-02	4.970E-02	3.760E-02	2.860E-02	2.650E-02
	D	4.540E-02	7.120E-02	3.780E-02	5.270E-02	4.010E-02	3.100E-02	2.230E-02
Random subspace Extra Tree	A	4.120E-02	7.550E-02	5.120E-02	5.670E-02	5.360E-02	**3.870E-02**	4.010E-02
	B	4.790E-02	8.280E-02	5.420E-02	6.090E-02	5.830E-02	4.120E-02	4.320E-02
	C	4.910E-02	8.970E-02	5.890E-02	6.810E-02	6.310E-02	4.700E-02	5.060E-02
	D	5.830E-02	1.028E-01	6.830E-02	7.540E-02	7.270E-02	5.340E-02	5.930E-02
Random subspace REP Tree	A	4.710E-02	5.370E-02	4.610E-02	5.130E-02	4.880E-02	4.610E-02	**4.290E-02**
	B	5.330E-02	6.090E-02	5.490E-02	5.710E-02	5.720E-02	5.340E-02	4.630E-02
	C	6.480E-02	7.080E-02	6.270E-02	6.590E-02	6.540E-02	6.030E-02	5.580E-02
	D	7.230E-02	8.600E-02	7.350E-02	7.860E-02	7.450E-02	7.120E-02	6.190E-02

Table 5. RMSE for Random Subspace

Prediction Model	Data	$SBDS_1$	$SBDS_2$	$SBDS_3$	$SBDS_4$	$SBDS_6$	$SBDS_9$	$SBDS_{10}$
Random subspace Isotonic Regression	A	**5.270E-02**	**5.270E-02**	**5.270E-02**	**5.270E-02**	**5.270E-02**	**5.270E-02**	**5.270E-02**
	B	5.340E-02	5.340E-02	5.340E-02	5.340E-02	5.340E-02	5.340E-02	5.340E-02
	C	6.160E-02	6.160E-02	6.160E-02	6.160E-02	6.160E-02	6.160E-02	6.160E-02
	D	1.294E-01	1.294E-01	1.294E-01	1.294E-01	1.294E-01	1.294E-01	1.294E-01
Random subspace SMOreg	A	1.760E-01	1.618E-01	1.714E-01	1.332E-01	1.512E-01	**4.490E-02**	3.100E-02
	B	6.990E-02	9.870E-02	6.650E-02	7.720E-02	6.890E-02	5.630E-02	3.180E-02
	C	7.020E-02	9.910E-02	6.770E-02	7.730E-02	6.910E-02	3.530E-02	7.020E-02
	D	1.809E-01	2.834E-01	1.505E-01	2.102E-01	1.597E-01	1.233E-01	8.870E-02
Random subspace Extra Tree	A	1.711E-01	2.217E-01	2.053E-01	1.865E-01	2.041E-01	**8.770E-02**	1.290E-01
	B	1.345E-01	1.916E-01	1.382E-01	1.506E-01	1.482E-01	1.149E-01	1.064E-01
	C	1.228E-01	1.976E-01	1.483E-01	1.740E-01	1.572E-01	1.292E-01	1.228E-01
	D	2.322E-01	4.093E-01	2.719E-01	3.003E-01	2.897E-01	2.127E-01	2.363E-01
Random subspace REP Tree	A	2.084E-01	1.787E-01	1.933E-01	1.666E-01	1.874E-01	1.051E-01	**1.078E-01**
	B	1.327E-01	1.546E-01	1.416E-01	1.498E-01	1.457E-01	1.286E-01	1.154E-01
	C	1.643E-01	1.826E-01	1.509E-01	1.700E-01	1.690E-01	1.555E-01	1.643E-01
	D	2.879E-01	3.428E-01	2.927E-01	3.130E-01	2.967E-01	2.836E-01	2.469E-01

Another part of Meta prediction experiments contains Ensemble selection, Voting, and Stacking. These algorithms combine several prediction models using different ways as explained previously. The main idea behind this Section is to promote diversity among prediction models and then combine them by different methods to obtain better predictive performance. In this experiment, we used four different prediction models: SMOreg, Isotonic Regression, Extra-Tree and REPtree. Then we implemented three different combining techniques including stacking, voting, and ensemble selection. We also illustrate the results based on the different percentages of training and testing. MAE and RMSE were used as a performance measure for all the prediction models. Tables 4 and 5 show results for each technique. According to [54] NNs accomplished distinguished results and hence we tried to use NNs with Meta prediction models in order to improve results. We concluded that Bagging does not work with NNs and the error was increased with Random subspace when compared with direct NNs results, as shown in Table 6. It is important to notice that most Meta prediction models consumed a long time up to 4 hours or more. Therefore, we used hybrid model as another type of combined predictor models and their results are illustrated in next Section.

International Journal of Computer Information Systems and Industrial Management Applications.
ISSN 2150-7988, Volume 7 (2015) pp. 001-013

Table 6. NNs results with Bagging and Random subspace

Combination Model	Data	$SBDS_1$	$SBDS_2$	$SBDS_3$	$SBDS_4$	$SBDS_6$	$SBDS_9$	$SBDS_{10}$
Bagging Multilayer Perceptron	**A**	5.210E+01	2.228E+01	5.218E+01	5.765E+01	5.166E+01	5.039E+01	5.013E+01
	B	5.195E+01	2.556E+01	5.234E+01	4.942E+01	5.197E+01	5.054E+01	5.027E+01
	C	5.224E+01	2.510E+01	5.234E+01	5.154E+01	5.193E+01	5.055E+01	5.027E+01
	D	5.204E+01	2.342E+01	5.239E+01	4.849E+01	5.227E+01	5.068E+01	5.035E+01
Random-subspace Multilayer Perceptron	**A**	2.270E-02	4.450E-02	9.600E-03	6.600E-03	2.640E-02	2.066E-01	1.926E-01
	B	2.570E-02	5.070E-02	1.020E-02	8.700E-03	1.970E-02	2.670E-01	2.211E-01
	C	2.130E-02	4.900E-02	9.300E-03	7.900E-03	4.150E-02	2.301E-01	2.182E-01
	D	3.260E-02	5.050E-02	8.200E-03	1.220E-02	2.330E-02	2.351E-01	2.588E-01

B. Hybrid prediction models

A neuro-fuzzy model is developed to predict the crude oil price. The neural network learning method is used for building a fuzzy model and used 7 sub datasets and 100 learning epochs. To design ANFIS model we followed these steps: Specify number and type of membership functions. Different types of membership functions (MF) include Trapezoidal, Guassian, Gbell and Triangular shapes were tested for the inputs and output. A two (Trapezoidal) MF for each input variable type ANFIS resulted in high accurate modeling and minimal training time. The membership function, rule base and ANFIS structure are displayed in Figures 7, 8 and 9 respectively for SBDS7-B (three inputs). Tables 7 and 8 show that the best results for each sub-datasets based on Datasets A and B (bold font) and overall data SBDS10. Dataset (B) achieved best results with MAE= 4.70906E-07 and RMSE=7.382E-07. Training by ANFIS completed in about 3 seconds and Table 9 illustrates ANFIS time for training. According to Figure 8, a sample rule would appear as follows:

If (WTI is 1.74) and (GPNY is 0.208) and (FC1 is 2.08) then (WTI is 1.81)

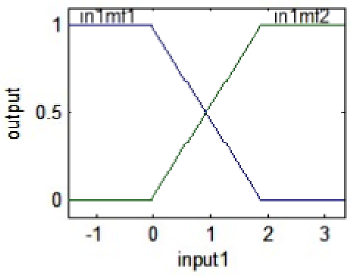

Figure 7. Trapezoidal-shaped membership function for the first Input

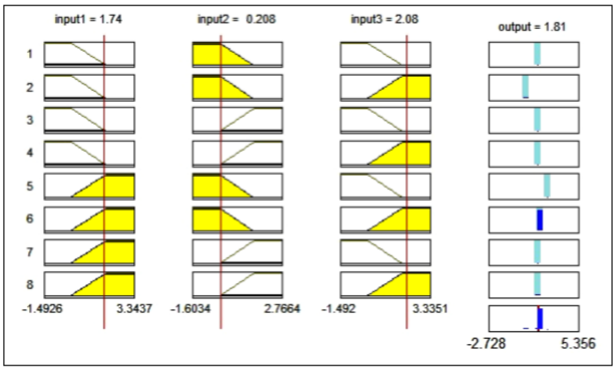

Figure 8. Developed TSK FIS using 3 inputs

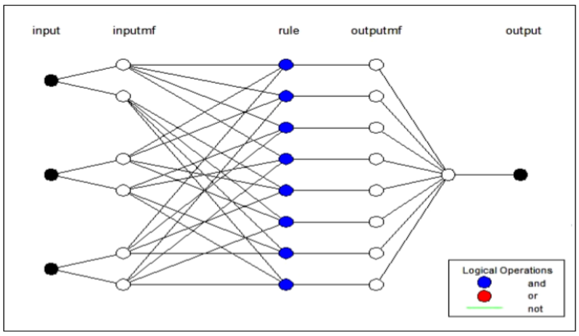

Figure 9. Developed ANFIS structure with 3 inputs

The learned eight *if-then* rules appear as follows:

***1.** If (WTI is in1mf1) and (GPNY is in2mf1) and (FC1 is in3mf1) then (WTI is out1mf1)*
***2.** If (WTI is in1mf1) and (GPNY is in2mf1) and (FC1is in3mf2) then (WTI is out1mf2)*
***3.** If (WTI is in1mf1) and (GPNY is in2mf2) and (FC1 is in3mf1) then (WTI is out1mf3)*
***4.** If (WTI is in1mf1)and (GPNY is in2mf2) and (FC1 is in3mf2) then (WTTI is out1mf4)*
***5.** If (WTI is in1mf2) and (GPNY is in2mf1) and (FC1 is in3mf1) then (WTI is out1mf5)*
***6.** If (WTI is in1mf2) and (GPNY is in2mf1) and (FC1 is in3mf2) then (WTI is out1mf6)*
***7.** If (WTI is in1mf2) and (GPNY is in2mf2) and (FC1 is in3mf1) then (WTI is out1mf7)*
***8.** If (WTI is in1mf2) and (GPNY is in2mf2) and (FC1 is in3mf2) then (WTI is out1mf8)*

Table 7. ANFIS results (MAE) for 7 sub-data sets

DATA	**Mean Absolute Error (MAE)**						
	SBDS$_1$	**SBDS$_2$**	**SBDS$_3$**	**SBDS$_4$**	**SBDS$_6$**	**SBDS$_9$**	**SBDS$_{10}$**
A	**1.23522E-05**	**8.94070E-06**	3.75200E-02	3.44406E-04	**9.96863E-04**	**2.39222E-06**	7.95274E-06
B	1.56636E-05	1.32849E-05	**2.89157E-02**	3.69172E-04	1.50347E-03	6.78925E-06	**4.70906E-07**
C	5.81167E-05	2.87358E-05	2.12495E-01	8.19114E-04	7.59600E-03	9.50917E-06	1.77291E-06
D	4.84699E-05	2.90189E-01	1.12921E-01	**1.90370E-04**	1.23274E-02	1.90200E-05	2.47023E-02

Table 8. ANFIS results (RMSE) for 7 sub-datasets

DATA	**Root Mean Square Error (RMSE)**						
	SBDS$_1$	**SBDS$_2$**	**SBDS$_3$**	**SBDS$_4$**	**SBDS$_6$**	**SBDS$_9$**	**SBDS$_{10}$**
A	**2.799E-05**	**1.1752E-05**	8.680E-02	5.860E-01	**2.193E-03**	**4.69999E-06**	7.626E-07
B	3.0375E-05	2.8262E-05	**7.802E-02**	1.646E-03	6.549E-03	1.497E-05	**7.382E-07**
C	2.8134E-04	1.415E-04	4.394E-01	2.506E-02	3.378E-02	2.116E-05	3.955E-06
D	8.0161E-05	2.540E-03	5.860E-01	**6.648E-04**	1.210E-01	2.390E-05	4.942E-02

Table 9. ANFIS training time

DATA	**Time (hh: mm: ss)**						
	SBDS$_1$	**SBDS$_2$**	**SBDS$_3$**	**SBDS$_4$**	**SBDS$_6$**	**SBDS$_9$**	**SBDS$_{10}$**
A	00:00:10	00:01:55	00:02:17	00:00:09	00:00:28	00:00:05	00:00:05
B	00:00:08	00:01:46	00:02:02	00:00:08	00:00:25	00:00:04	00:00:04
C	00:00:07	00:01:40	00:01:42	00:00:07	00:00:22	00:00:04	00:00:04
D	00:00:06	00:01:10	00:01:33	00:00:06	00:00:20	00:00:03	00:00:03

C. Ensemble prediction models

In this research, we employed the basic ensemble method and generalized ensemble method, which are the most popular techniques used in this phase:

1) The basic ensemble method

We used basic ensemble (average) method with NNs, according to [54] Best results for RBF and RCN were achieved with dataset (B) and for FFN when using (A).To combine three models, we used dataset (B) for all NNs and then combined them by using the average method for creating ensembles and the results are depicted in Table 10.

Table 10. MAE for basic ensemble results for NNs

Neural Networks	**MAE**	**RMSE**
RCN	3.9480E-05	2.300E-03
RBF	**2.2065E-05**	**1.291E-03**
FFN	6.0465E-05	2.569E-03
Ensemble Average	3.1780E-05	2.172E-03

Table 11. Ensemble using Average method for Data (A)

Data (A)	**MAE**	**RMSE**
SBDS$_1$	1.2352E-05	2.799E-05
SBDS$_2$	8.9407E-06	1.175E-05
SBDS$_6$	9.9686E-04	2.193E-03
SBDS$_9$	**2.3922E-06**	**4.699E-06**
Ensemble Average	2.5011E-04	1.097E-03

Table 12. Ensemble using Average method for Data (B)

Data (B)	**MAE**	**RMSE**
SBDS3	2.89157E-02	7.802E-02
SBDS10	**4.70906E-07**	**7.382E-07**
Ensemble Average	1.44578E-02	3.901E-02

It is important to notice that Average method improved the individual results for RCN and FFN but RBF still had the best overall results. In order to improve ANFIS results we applied again Ensemble Average with ANFIS using datasets (A) and (B). The MAE and RMSE results are displayed in Tables 11 and 12 for group (A) and for group (B) respectively.

Likewise ANFIS Ensemble average results are better than some sub-dataset such as SBDS6 in DATA (A) and SBDS3 in DATA (B). However, there was no clear superiority on the other sub datasets such as SBDS1, SBDS2 and SBDS9 in data (A) and SBDS10 in data (B).

2) The generalized ensemble method (GEM)

Another important concept regarding the performance of a predictive model is the GEM method. Based on equations 4 and 5 we need to find the optimal values of weight α to minimize the MAE or RMSE between the outputs and the desired values. We used a PSO algorithm to determine the optimal weights. As evident from Tables 13 and 14 the GEM model is better in predicting ANFIS results than the average method for Data (A) and Data (B) respectively.

Table 13. Ensemble of PSO-ANFIS for Data (A)

Data (A)	MAE	RMSE
$SBDS_1$	1.2352E-05	2.7993E-05
$SBDS_2$	8.9407E-06	1.1752E-05
$SBDS_6$	9.9686E-04	2.1932E-03
$SBDS_9$	2.3922E-06	4.69999E-06
Ensemble –PSO	**2.39215E-06**	**4.69295E-06**

Table 14. Ensemble of PSO-ANFIS for Data (B)

Data (B)	MAE	RMSE
$SBDS_3$	2.89157E-02	7.8021E-02
$SBDS_{10}$	4.70906E-07	7.3868E-07
Ensemble –PSO	**4.62053E-07**	**7.2736E-07**

VI. Comparison analysis of combined prediction models

We compared the results of the previous experiments in Section5 to determine the best Meta learning for the prediction of crude oil prices. The results are summarized in Table 15. Ensemble selection achieved the best results with MAE 1.420E-02 and RMSE 2.42E-02 when compared to Bagging using SMOreg. By comparing all the results obtained from Meta learning and ANFIS, the optimal result with the minimum MAE and RMSE values were derived from ANFIS with data (B). This set produced MAE = 4.70906E-07 and RMSE = 7.382E-07 for its RMSE with a competitive training time of 04 sec. Based on the performance comparison in Table 16, the Ensemble-PSO-ANFIS leads the other Ensemble approaches in terms lowest MAE and RMSE values. Another aspect is comparison of the (Ensemble–PSO-ANFIS) prediction model results with other machine learning approaches in order to measure the performance of the GEM model using ANFIS-PSO method, a comparison is made with other machine learning methods and is shown in Table 17.

VII. Conclusions

This paper presented the experiments using combined models and the contributions are summarized, as follows: Meta-learning empirical results were derived from two parts: Bagging and Random subspace . Further, Ensemble selection, Voting, and Stacking which combines the outputs from several predictors were also presented. We used ANFIS, which achieved better results than Meta-learning models and NNs approaches in terms of accuracy and training time. In order to improve the results, we operated all the four training and testing datasets and used for the basic ensemble and generalized ensemble methods. The best training result was obtained from the data that were trained using 80% training and 20% for testing and obtained a mean absolute error (MAE) value of 4.62053E-07, and root mean squared error (RMSE) value of 7.2736E-07 using Ensemble PSO -ANFIS for SBDS3 and SBDS10. From the implementation of combined prediction models, it is evident that VIX, WTI, GPNY, ER, and FC1 are the most important factors to determine the crude oil price and ANFIS is a good interpretable model to explore and explain crude oil market's *if-then* rules. Finally comparison with different results from the literature as presented in Table 6.18 further illustrates the effectiveness and superiority of the Ensemble method using ANFIS PSO for the prediction of WTI crude oil price.

Table 15. Comparison among Meta learning models

Meta learning model	Data	MAE	RMSE	Sub-dataset	Time
Bagging SMOreg	B	1.840E-02	2.610E-02	SBDS10	00:00:10
Random Subspace SMOreg	A	2.210E-02	4.490E-02	SBDS9	00:01:06
Stacking	A	2.710E-02	5.510E-02	SBDS1	00:05:41
Voting	A	5.400E-02	7.670E-02	SBDS9	00:01:02
Ensemble selection	**A**	**1.420E-02**	**2.420E-02**	SBDS3	00:06:17

Table 16. Performance comparison between Ensemble prediction models

Ensemble prediction models	Data	MAE	RMSE
Ensemble Average	(A)	2.5011E-04	1.0970E-03
	(B)	1.4458E-02	3.9010E-02
Ensemble –ANFIS-PSO	(A)	2.3922E-06	4.6930E-06
	(B)	**4.6205E-07**	**7.2736E-07**

Table 17. Comparison of models used in the literature to predict WTI crude oil price using the ANFIS-PSO Ensemble

Reference	Algorithm	Task	Performance
[13]	Hybrid AI method	Prediction of WTI crude oil price	RMSE =2.040E+00
[55]	Support Vector Machine	Prediction of WTI crude oil price	RMSE =2.192E+00
[56]	EMD-SVM-ELM	Prediction of WTI crude oil price	RMSE =4.770E-01 MAE =3.630E-01
[57]	Orthogonal wavelet support vector machine	Prediction of WTI crude oil price	MAE= 3.560E-04
[58]	GA–NN	Prediction of WTI crude oil price	RMSE =1.15E-03
[59]	Co-active neuro fuzzy inference system	Prediction of WTI crude oil price	RMSE= 2.0864E-04 MAE =8.5321E-04
[60]	Ensemble –ANFIS-PSO	Prediction of WTI crude oil price	RMSE 7.2736E-07 MAE =4.6205E-07

References

[1] Petroleum. (2013). *Petroleum Formation*. Available: http://www.petroleum.co.uk/formation [Accessed 27/12/2013]

[2] U. S. E. I. A. EIA. (2013). *Defintions, Sources and Explanatory Notes -Crude Oil Production*. Available: http://www.eia.gov/dnav/pet/TblDefs/pet_crd_crpdn_tbldef2.asp [Accessed 23/11/2014]

[3] B. Fattouh, *An anatomy of the crude oil pricing system*: Oxford Institute for Energy Studies Oxford, England, 2011.

[4] A. Charles and O. Darné, "Volatility persistence in crude oil markets," *Energy Policy*, vol. 65, pp. 729-742, 2014.

[5] H. Chiroma, S. Abdulkareem, A. ABUBAKAR, and J. U. MOHAMMED, "Computational Intelligence Techniques with Application to Crude Oil Price Projection: A literature Survey from 2001-2012," *Neural Network World*, vol. 23, pp. 523-551, 2013.

[6] L. A. Gabralla and A. 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 pp. 729-740, 2013.

[7] F. Gorunescu, *Data Mining: Concepts, models and techniques* vol. 12: Springer, 2011.

[8] M. M. Gaber, *Scientific data mining and knowledge discovery: Principles and foundations*: Springer, 2009.

[9] H. S. Own and A. Abraham, "A Novel-weighted Rough Set-based Meta Learning for Ozone Day Prediction," *Acta Polytechnica Hungarica*, vol. 11, 2014.

[10] A. Verikas, Z. Kalsyte, M. Bacauskiene, and A. Gelzinis, "Hybrid and ensemble-based soft computing techniques in bankruptcy prediction: a survey," *Soft Computing*, vol. 14, pp. 995-1010, 2010.

[11] F. A. Anifowose, "Advances in Hybrid Computational Intelligence Application in Oil and Gas Reservoir Characterization," in *SPE Saudi Arabia section Young Professionals Technical Symposium*, 2012.

[12] A. Abraham, "Adaptation of fuzzy inference system using neural learning," in *Fuzzy systems engineering*, ed: Springer, 2005, pp. 53-83.

[13] S. Wang, L. Yu, and K. K. Lai, "A novel hybrid AI system framework for crude oil price forecasting," in *Data Mining and Knowledge Management*, ed: Springer, 2005, pp. 233-242.

[14] M. Panella, F. Barcellona, and R. L. D'Ecclesia, "Forecasting energy commodity prices using neural networks," *Advances in Decision Sciences*, vol. 2012, 2012.

[15] A. Ghaffari and S. Zare, "A novel algorithm for prediction of crude oil price variation based on soft computing," *Energy Economics*, vol. 31, pp. 531-536, 2009.

[16] L. Yu, S. Wang, and K. K. Lai, "Forecasting crude oil price with an EMD-based neural network ensemble learning paradigm," *Energy Economics*, vol. 30, pp. 2623-2635, 2008.

[17] L. Yu, S. Wang, and K. K. Lai, "A generalized Intelligent-agent-based fuzzy group forecasting model for oil price prediction," in *Systems, Man and Cybernetics, 2008. SMC 2008. IEEE International Conference on*, 2008, pp. 489-493.

[18] L. Breiman, "Bagging predictors," *Machine learning*, vol. 24, pp. 123-140, 1996.

[19] E. Rashedi and A. Mirzaei, "A hierarchical clusterer ensemble method based on boosting theory," *Knowledge-Based Systems*, vol. 45, pp. 83-93, 2013.

[20] P. Yang, Y. Hwa Yang, B. B Zhou, and A. Y Zomaya, "A review of ensemble methods in bioinformatics," *Current Bioinformatics*, vol. 5, pp. 296-308, 2010.

[21] E. Bauer and R. Kohavi, "An empirical comparison of voting classification algorithms: Bagging, boosting, and variants," *Machine learning*, vol. 36, pp. 105-139, 1999.

[22] T. K. Ho, "The random subspace method for constructing decision forests," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 20, pp. 832-844, 1998.

[23] R. Caruana, A. Niculescu-Mizil, G. Crew, and A. Ksikes, "Ensemble selection from libraries of models," in *Proceedings of the twenty-first international conference on Machine learning*, 2004, p. 18.

[24] C. W. Wang, "New ensemble machine learning method for classification and prediction on gene expression data," in *Engineering in Medicine and Biology Society, 2006. EMBS'06. 28th Annual International Conference of the IEEE*, 2006, pp. 3478-3481.

[25] M. Panda and M. R. Patra, "Ensemble voting system for anomaly based network intrusion detection," *International journal of recent trends in engineering*, vol. 2, pp. 8-13, 2009.

[26] S. Džeroski and B. Ženko, "Is combining classifiers with stacking better than selecting the best one?," *Machine learning*, vol. 54, pp. 255-273, 2004.

[27] L. A. Zadeh, "Roles of soft computing and fuzzy logic in the conception, design and deployment of information/intelligent systems," in *Computational intelligence: soft computing and fuzzy-neuro integration with applications*, ed: Springer, 1998, pp. 1-9.

[28] T. Takagi and M. Sugeno, "Fuzzy identification of systems and its applications to modeling and control," *Systems, Man and Cybernetics, IEEE Transactions on*, pp. 116-132, 1985.

[29] J.-S. Jang, "ANFIS: adaptive-network-based fuzzy inference system," *Systems, Man and Cybernetics, IEEE Transactions on*, vol. 23, pp. 665-685, 1993.

[30] H.-B. Shen and K.-C. Chou, "Ensemble classifier for protein fold pattern recognition," *Bioinformatics*, vol. 22, pp. 1717-1722, 2006.

[31] Z. Zhao and Y. Zhang, "Design of ensemble neural network using entropy theory," *Advances in Engineering Software*, vol. 42, pp. 838-845, 2011.
[32] D. West, P. Mangiameli, R. Rampal, and V. West, "Ensemble strategies for a medical diagnostic decision support system: A breast cancer diagnosis application," *European Journal of Operational Research*, vol. 162, pp. 532-551, 2005.
[33] I. Partalas, G. Tsoumakas, I. Katakis, and I. Vlahavas, "Ensemble pruning using reinforcement learning," in *Advances in Artificial Intelligence*, ed: Springer, 2006, pp. 301-310.
[34] R. Vilalta, C. G. Giraud-Carrier, P. Brazdil, and C. Soares, "Using Meta-Learning to Support Data Mining," *IJCSA*, vol. 1, pp. 31-45, 2004.
[35] R. Maclin and D. Opitz, "Popular ensemble methods: An empirical study," *arXiv preprint arXiv:1106.0257*, 2011.
[36] M. Blachnik, "Ensembles of instance selection methods based on feature subset," *Procedia Computer Science*, vol. 35, pp. 388-396, 2014.
[37] E. Menahem, L. Rokach, and Y. Elovici, "Combining one-class classifiers via meta learning," in *Proceedings of the 22nd ACM international conference on Conference on information & knowledge management*, 2013, pp. 2435-2440.
[38] I. Maqsood, M. R. Khan, and A. Abraham, "An ensemble of neural networks for weather forecasting," *Neural Computing & Applications*, vol. 13, pp. 112-122, 2004.
[39] R. C. Eberhart and J. Kennedy, "A new optimizer using particle swarm theory," in *Proceedings of the sixth international symposium on micro machine and human science*, 1995, pp. 39-43.
[40] M. M. Inc. (2012). *Growing gold and silver producer in the Americas*. Available: http://www.mcewenmining.com/ [Accessed 15/5/2013]
[41] EIA. (2012). *PETROLEUM & OTHER LIQUIDS*. Available: http://www.eia.gov/petroleum/data.cfm [Accessed 23/4/2013]
[42] A. Alexandridis and E. Livanis, "Forecasting crude oil prices using wavelet neural networks," *Published in the proc. of 5th FSDET, Athens, Greece* vol. 8, 2008.
[43] investopedia. (2012). *Federal Funds Rate*. Available: http://www.investopedia.com/terms/f/federalfundsrate.asp [Accessed 7/8/2014]
[44] I. El-Sharif, D. Brown, B. Burton, B. Nixon, and A. Russell, "Evidence on the nature and extent of the relationship between oil prices and equity values in the UK," *Energy Economics*, vol. 27, pp. 819-830, 2005.
[45] S. Zhang, C. Zhang, and Q. Yang, "Data preparation for data mining," *Applied Artificial Intelligence*, vol. 17, pp. 375-381, 2003.
[46] I. H. Witten and E. Frank, *Data Mining: Practical machine learning tools and techniques*: Morgan Kaufmann, 2005.
[47] I. Guyon and A. Elisseeff, "An introduction to variable and feature selection," *The Journal of Machine Learning Research*, vol. 3, pp. 1157-1182, 2003.
[48] O. Kaynar, I. Yilmaz, and F. Demirkoparan, "Forecasting of natural gas consumption with neural network and neuro fuzzy system," *Energy Education Science and Technology Part A–Energy Science and Research*, vol. 26, pp. 221-238, 2011.
[49] K. K. Lai, K. He, and J. Yen, "Modeling VaR in crude oil market: a multi scale nonlinear ensemble approach incorporating wavelet analysis and ANN," in *Computational Science–ICCS 2007*, ed: Springer, 2007, pp. 554-561.
[50] L. Yu, Y. Zhao, and L. Tang, "A compressed sensing based AI learning paradigm for crude oil price forecasting," *Energy Economics*, vol. 46, pp. 236-245, 2014.
[51] X. Zeng, "Machine Learning Approach for Crude Oil Price Prediction," Doctor of Philosophy Doctoral level ETD - final, Faculty of Engineering and Physical Sciences, School of Computer Science, The University of Manchester Manchester, UK, 2014.
[52] L. A. Gabralla, R. Jammazi, and A. Abraham, "Oil price prediction using ensemble machine learning," in *Computing, Electrical and Electronics Engineering (ICCEEE), 2013 International Conference on*, 2013, pp. 674-679.
[53] L. A. Gabralla and A. Abraham, "Prediction of Oil Prices Using Bagging and Random Subspace," in *Proceedings of the Fifth International Conference on Innovations in Bio-Inspired Computing and Applications IBICA 2014*, 2014, pp. 343-354.
[54] L. A. Gabralla, H. Mahersia, and A. Abraham, "Ensemble Neurocomputing Based Oil Price Prediction," in *Afro-European Conference for Industrial Advancement*, 2015, pp. 293-302.
[55] W. Xie, L. Yu, S. Xu, and S. Wang, "A New Method for Crude Oil Price Forecasting Based on Support Vector Machines," in *Computational Science–ICCS 2006*, ed: Springer, 2006, pp. 444-451.
[56] R. A. Ahmed and A. B. Shabri, "A Hybrid of EMD-SVM Based on Extreme Learning Machine for Crude Oil Price Forecasting," *Australian Journal of Basic and Applied Sciences*, vol. 8(15), pp. 341-351, 2014.
[57] H. Chiroma, S. Abdul-Kareem, A. Abubakar, A. M. Zeki, and M. J. Usman, "Orthogonal Wavelet Support Vector Machine for Predicting Crude Oil Prices," in *Proceedings of the First International Conference on Advanced Data and Information Engineering (DaEng-2013)*, 2014, pp. 193-201.
[58] H. Chiroma, S. Abdulkareem, and T. Herawan, "Evolutionary Neural Network model for West Texas Intermediate crude oil price prediction," *Applied Energy*, vol. 142, pp. 266-273, 2015.
[59] H. Chiroma, S. Abdulkareem, A. Abubakar, A. Zeki, A. Gital, and M. Usman, "Co—Active neuro-fuzzy inference systems model for predicting crude oil price based on OECD inventories," in *Research and Innovation in Information Systems (ICRIIS), 2013 International Conference on*, 2013, pp. 232-235.
[60] L. A. Gabralla, T. M. Wahby, V. K. Ojha, and A. Abraham, "Ensemble of Adaptive Neuro-Fuzzy Inference System Using Particle Swarm Optimization for Prediction of Crude Oil Prices," presented at the International Conference on Hybrid Intelligent Systems (HIS), Kuwait, 2015 b.