International Journal of Engineering & Technology, 7 (4.19) (2018) 114-120

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper

Improving Oil Price forecasting using Regression-Relief-bases Indicators Selection Model

Eman S. Al-Shamery[1*], Hussein A. Al-Gashamy[2]

[1]IT College. Babylon University, Iraq
[2]MSC. Student - IT College. Babylon University, Iraq
**Corresponding Author E-mail: [1]emanalshamery@itnet.uobabylon.edu.iq*

Abstract

Crude oil holds a vital and growing role in the local and global economy. The main goal of this study is to explore the effect of technical indicators in enhancing the capability of Sequential Minimal Optimization (SMO) to forecast the precise oil price. In addition, Relief algorithm reduces the dimensional space and eliminates irrelevant factors. 10-fold Cross validation testing method is used to train two main series of crude oil price, Brentand West Texas Intermediate (WTI). Further, Mean Absolute Error and Root Mean Squared Error are considered as evaluation criteria. A comparison has been implemented on input features with and without TIs. The results show that using Technical Indicators satisfy better results of prediction. The Accuracy rate is raised with MAE ratio 14:1 for Brent and 15:1 for WTI, while RMSE improved with ratio17:1 for Brent and 20:1 for WTI. Finally, The experimental result proved that the optimized model was superior when compared with Linear regression, MLP regression, and Gaussian in terms of prediction errors.

Keyword: *SMO, Crude oil, Feature Selection, technical indicator, Ranker*

1. Introduction

The crude oil price has a great effect on both local and global areas. It meets about two-thirds of the global energy demand[1]. stock levels, weather, shipping between countries, Gross Domestic Product (GDP) growth, political aspects, these factors lead to strongly oil price fluctuation, which has the characteristics of high irregularity, nonlinearity and dynamic variation[2]. The rapid swing in the price of oil, up and down have led to an enormous effects on global geopolitics and economies, so that oil price spikes have stunt economic growth and a sudden price plunge might wreak on cash-strapped oil companies. In the field of power research crude oil consider one of the most important topics with various economic, statistical and intelligent approaches to forecasting oil prices. The nature of prices between past and present make the relationship between them nonlinear [3].the linear approach such as Auto Regressive Integrated Moving Average ,autoregressive , and the Auto Regressive Moving Average models are clearly premature when compared to nonlinear approaches like the Generalized Auto Regressive Conditional Heteroskedasticity and Revolutionary technique such as ANN, Genetic Algorithm, and SVM, In this paper, a typical competitive learning algorithm is empirically investigated to verify the capability and feasibility of SVM in crude oil price forecasting. Support Vector Machin were developed for pattern recognition and regression estimation problems[4]. Recently, SVM has been applied in various areas of time series forecasting and successfully used for financial forecasting[5-7]. SVM called support vector regression (SVR) in term of non-linearity, It has a great use in prediction the volatility of different markets' energy resources such as oil, gas, and electricity.

[8]adopted various kernel functions of Support Vector Machine with a Particle Swarm algorithm used to optimize the parameters(coefficient, epsilon error, alpha, and bias) to build thecrude oilprice prediction model.
[9] proposed multiple kernel regression to forecast crude oil for WTI and Brent markets with MACD indicator for feature extraction which considered superior to benchmark methods in (RMSE) and average percentage profit (APP).in this study the information gained from another market not useful to those from multiple timeframes for prediction. On the other hand, Economist has been used technical indicator (TI) to analyze crude oil price and stocks by capturing the behavior, however. [10]provide that, Regardless of the complexity of the formula required to compute TI, Technical Indicatorshas been provided some gaudiness to find out the future trending of price activity. Many traders and investors use technical indicators to predict the direction of the future prices. It was very hopeful to forecast oil prices.
The goal of this paper is to explore the predictability of crude oil price with Sequential Minimal Optimization enhanced by TIs. The remainder of this paper is organized as follows; Section (2) discusses some of the literature reviews. Section (3) introduces the basic theory of SVR and technical indicator. The experiment results are shown in Section 4. The conclusions are drawn in Section 5.

2. Literature Review

1- In [11] work introduces a comparative study between the Gene Expression Programming and neural network models as an evolutionary technique to forecast oil prices in US$ over the period from 2-Jan-1986 to 12-June-2012.in addition to that,Auto Regressive Integrated Moving Averagemodel is used as evolutionary model. Experimental results showed that the Gene

Expression Programming model outperforms the Neural Network and ARIMA models for all mentioned error measures.

2- In [12] paper Proposed a hybrid method that based on Discrete Wavelet Transforms technique and SVM model named as (WSVM) for forecastingcrude oil price. the wavelet theory decomposes the original time series of the crude oil price at 3 decomposition levels (2,4and8 months), then these time series were taken as input to the SVM. The Experiment was implemented over the period from 1986 through 2006 and find out that The performance of the proposed model is considered superior in comparison to the regular SVM model.

3- In [13]: this work uses Different Decision treealgorithms like the Random forest, Random tree, Decision stump REP Tree, and M5P all are investigated for forecasting crude oil price. Historical data collected over the period from 1986 till 2009, They are made up of eight input attributes with the corresponding WTI weekly oil price. Results showed that the Random Tree DT have the fastest computational times of 0.01s whilst M5P performed the best in terms of error metrics.

3. Methodologies

3.1 Sequential Minimal Optimization (SMO):

Due to large size in the objective function for the optimization problem in Support Vector Regression(SVR):

$$\text{Max } W(\alpha) = \sum_{i=1}^{n} \alpha_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{i=1}^{n} \alpha_i \alpha_j y_i y_j K(X_i, X_j) \quad (1)$$

Subject to:

$\sum_{i=1}^{n} \alpha_i y_i = 0$

$0 \le \alpha_i \le C$; i=1,2,…….e

the Quadratic problemthat arises from SVRs cannot be solved efficiently via standard numerical Quadratic Problem techniques especially for large size problem that has a matrix with the number of elements equal to the square of the number of training instance.

Different algorithms are proposed for solving the problem of dual function. Platt proposed a Sequential Minimal Optimization algorithm for classification problems, that iteratively selecting the working set of size two and optimize the objective function according to them by solving sub problem analytically[14]. the process repeated iteratively till all training instance satisfy KKT conditions. Smola and Schölkopf extend the SMO algorithm for solving the regression problems[15].

Four different cases must be distinguish (α_i, α_j), $(\alpha_i, \acute{\alpha}_j)$, $(\acute{\alpha}_i, \alpha_j)$, $(\acute{\alpha}_i, \acute{\alpha}_j)$ for regression problems, thus one obtain from the summation of the constrains:

$$(\alpha_i - \acute{\alpha}_i) + (\alpha_j - \acute{\alpha}_j) = \gamma \quad (2)$$

Where $\alpha_j \in[0,Cj]$, $\alpha_i \in[l,h]$ where l and h are defined as in Table 5:

Table 1: Boundary of the feasible region

	α_j	$\acute{\alpha}_j$
α_i	l = Max(o, γ – C_j) h=Min(C_i , γ)	l = Max(o, γ) h=Min(C_i , $\acute{C}_j$ + γ)
$\acute{\alpha}_i$	l = Max(o, -γ) h=Min($\acute{C}_i$, -γ + C_j)	l=Max(0, -γ -$\acute{C}_j$) h=Min($\acute{C}_i$, -γ)

In this algorithm two threshold parameters value b_{up} and $b_{low}$and checking optimality with the condition :

$$B_{low} \le B_{up} + 2_{\varrho} \quad (3)$$

SMO chooses two elements α_i and α_j to jointly optimize, then find the optimal values for those two parameters given that all the others are fixed, and updates theα vector accordingly. A heuristic search is employed to select the two points through two separate heuristics: the first Lagrange multiplier provide the outer loop of the algorithm that iterates over the entire training one to determine whether each instance violates the Karush-Kuhn-Tucker conditions. If an instance violates these conditions, then it is qualified for optimization. After the first pass through the training set the outer loop iterates over all instance with non-bound Lagrange multipliers ($0 < \alpha_i < C$). This loop repeated over all the non-bound instances until all of them obey the Karush-Kuhn-Tucker within ε. Then the outer loop goes back and iterates over the entire training set. The outer loop keeps switching between one passes over the entire training set and multiple passes over the non-bound subset until the entire training set obeys the KKT conditions within ε. while the optimization of the two multipliers is performed analytically. The rate of convergence of the algorithm is strongly affected by the order in which the data points are chosen for updating. Heuristic measures such as the degree of violation of the KKT conditions can be used to ensure very effective convergence rates in practice.

3.2. Multiple Layer Perceptron (MLP):

MLP is a class of artificial neural network with one or more hidden layer, the nodes of each layer are neurons with a nonlinear activation function (often used sigmoid function). As shown inFig.1, MLP consists of layer of input, one or more hidden layer, and output layer of nodes interconnected in a feed forward direction[16].

MLP is a deep learning method that utilize back propagation(BP) as technique for training, BP consist of forward and backward phases, the first phase fed the inputs to the nodes of first hidden layer to perform the activities according to the activation function from input to output layer while the second phase exploits the error between desired and actual value from the output layer to propagate the input layer in order to modify the learning weights.

3.3. Technical Indicator

Mathematical formula using the concept of statistical based learning, applied to interpret stock market trend and investing decision by exploiting historical data to predict the movement of price over different time series ,usually shown in a graphical way above or below a price chart .different performance for different indicator ,some of them improve the accuracy or time and another both of them when using to enhance the performance of artificial intelligence calculation .

Leading and lagging are TIs classes, Leading (Oscillators) indicator Precede the price movement with predictive ability, e.g. Commodity Channel Index and relative strength index which generate signals prior to new trend happen.[10]While lagging follow the price action with less predictive abilities, which informs after new trend takes place. E.g. moving average convergence divergence(MACD),.[Fernandez et al.,2008]. In this paper, lagging technical indicators are used and these indicators are (AROON, Linear Regression Slope, Linear Regression Intercept, Variance, Moving Average Convergence Divergence(MACD), Time series Forecasting. And leading technical indicators are(Relative Strength Index,Commodity Channel Index, Absolute Price Oscillator, plus(minus)Directional Movement,Plus(Minus)Directional Movement Index .

2. Exponential Moving Average (EMA):[17]

EMA can be measured by giving more importance to recent prices relative to older prices. The weighting applied to the most recent price depends on the specified period of the moving average. EMA can be calculated as follow:

$$EMA_{current} = ((C_{current} - EMA_{previous}) \times \text{Multiplier}) + EMA_{previous} \quad (4)$$

Multiplier = $\left(\frac{2}{1+n}\right)$
where N is the predefined number of periods.

2. Moving Average Convergence Divergence[18]
There are three different equations can be calculated; the first one The MACD line which represents the difference between 26 days and 12 days (EMAs). The signal line indicates 9days-EMA of the line while Histogram represents the difference between them.

MACD = EMA (12) – EMA (26) (5)

Signal=EMA(MACD,9) (6)

Histogram=MACD - Signal (7)

3. Relative Strength Index(RSI) [18]
The Relative strength is measured by dividing the change of an average upward price by the change of an average downward price, and signals are generated by comparing predetermined entry thresholds to the Relative Strength Index , which have values between [0,100].

RSI=100 - (100/(1 + RS)) (8)

Where
RS=(EMA(Upper,n))/(EMA(Down,n))
Upper =$P_c(t)$ - $P_c(t-1)$
Down = $P_c(t-1)$ - $P_c(t)$
Upper=trading period in upward change
Down=trading period in Downward change

4. Commodity Channel Index (CCI)[19]
Commodity Channel Index measures the differences between price change and the average price change for selected period . The calculation of CCI is presented below.

$$\text{CCI} = \frac{\text{Typical Price} - \text{SimpleMovingAverage}}{0.015 - \text{Mean Deviation}} \quad (9)$$

Tp=(low + high +close) / 3

5. The plus and the minus directional movement indicator (+DI,⁻DI): (Labs, 2016)
Directional movement measures the direction of price moves upward or downward. Calculating by smoothing moving average of (DM) over the number of periods selected divided by average true range.as in the formula below:

$$\text{Plus DM}_t = (\text{Plus DM})_{t-1} - \frac{(\text{Plus DM})_{t-1}}{n} + (\text{Plus DM})_t$$

$$\text{Minus DM}_t = (\text{Minus DM})_{t-1} - \frac{(\text{Minus DM})_{t-1}}{n} + (\text{Minus DM})_t$$

$$\text{Plus DI}_t = \frac{(\text{Plus DM})t}{\text{ATR}} * 100 \quad (10)$$

$$\text{Minus DI}_t = \frac{\text{Minus DM}}{\text{ATR}} * 100 \quad (11)$$

6. The plus and the minus directional indicator(+DM,⁻DM): (Labs, 2016)
Which shows the difference between closing price with high and low price.as in the formula below:

$\Delta H = H_t - H_{t-1}$
$\Delta L = L_{t-1} - L_t$
If($\Delta H > \Delta L$, $\Delta H > 0$) then plus DM=ΔH, Minus DM=0
If($\Delta H < 0$ and $\Delta L > 0$) or $\Delta H = \Delta L$ then plus DM=0, Minus DM=0
If($\Delta H < \Delta L$, $\Delta L > 0$) then plus DM=0, Minus DM=ΔL

7. Aroon oscillator[20]
It is calculated by subtracting Aroon down from Aroon up, Readings above the central zero point indicate that an uptrend. A downtrend is indicated by readings below zero. with a range between -100 and 100.calculated by:

Aroon Osc =Aroon Up – Aroon Down (14)

8. Aroon:[20]
It determine if the market are trend and how strong the trend, based on High and Low price over specific period ,calculated by:

$$\text{Aroon UP} = 100 \times \left(\frac{n - P_H(n)}{n}\right) \quad (15)$$

$$\text{Aroon Down} = 100 \times \left(\frac{n - P_L(n)}{n}\right) \quad (16)$$

Where $P_H(n)$,$P_L(n)$ represent the heist price and lowest price ,respectively over(n) period

9. Time Series Forecasting(TSF) [21]
Time Series Forecast indicator shows the statistical trend of a market's price over
a predefined time period(n),using the "least squares fit" methodto calculate a linear regression trend line, which fits data in the bar value .

$$\text{Slope} = \frac{\sum_{t=i-n}^{i}(X_{0t} - M_{x0n}) \cdot (X_{1t} - M_{x1n})}{\sum_{t=i-n}^{i}(X_{0t} - M_{x0n})2} \quad (17)$$

$$\text{Intercept} = M_{x1n} - \text{slop}_i * M_{x0n} \quad (18)$$

$$\text{TSF} = \text{Slope} * pc_t + \text{Intercept}_t \quad (19)$$

10. midpoint[Steven,1994[21]]
Mid-Point(MID) is indicator in technical analysis , It's calculated by findingthe lowest low and the highest high through the time period being analyzed.

MID =(highest value + lowest value)/2 (20)

12. variance:[22]
measure of variability that represents how far members of a group are spread out, that describes the variability of observations from its arithmetic mean

$$\sigma^2 = \frac{\sum_{i=1}^{n}(xi - M)^2}{n} \quad (21)$$

13. standard deviation:[23]
Standard deviation is a statistical measurement measures the volatility ,generally.that quantifies the amount of dispersion of the observations in a dataset around an average, calculated by:

$$SD = \sigma = \sqrt{\frac{\sum_{i=1}^{n}(xi - M)^2}{n}} \quad (22)$$

14. The Absolute Price Oscillator (APO)[24]
displays the difference between two exponential moving averages of a security's price and is expressed as an absolute value.

APO=Shorter Period EMA – Longer Period EMA (23)

Table 2: Shows the period of TI's

NO.	TI	Time period(n)	NO.	TI	Time period(n)
1	Low	5(30)	11	High	5(30)
2	Aroon	7(14d)	12	LR-intercept	22(14d)
3	Aroon Osc	7(14d)	13	LR-Linearge	22(14d)
4	Midpoint	5(14d)	14	RSI	14

5	APO	14	15	CCI	7(14d)
6	Plus-DM	14	16	Var.	7(5d)
7	Minus-DM	14	17	MACD	35
8	Plus-DI	14	18	TSF	14
9	Minus -DI	14	19	EMA	30
10	LR-slope	22(14d)			

*d represents the default value

3.3. Regression-Relief Feature Selection(R reliefF):

The key idea of the original Relief algorithm [25] is to estimate the attributes quality according to how well their values distinguish between the instances that are close to each other, the original Relief algorithm was limited to classification problems with two classes. Then its extension to ReliefF to deal with multiclass problemsand it is more robust to deal with noisy data. But cannot deal with regression problems. ReliefF has been improved to RReliefF (Regression ReliefF) has been adapted for regression problems (continuous class). The algorithm estimate W[T] of the quality of attribute T is an approximation of the following difference of probabilities:

W[T] = P(diff. value of T| nearest inst. from diff. class) − P(diff. value of T| nearest inst. from same class) (24)

the estimate of probability that the attribute differentiate between the instances with different class values are representing the positive updates of the weights, the probability that the attribute separates the instances with the same class value are formingthe negative updates.

In classification problem identifying the nearest hit and nearest miss represent the core of Reliff jobs but In regression problems the predicted value $\tau(\cdot)$ is continuous, so that the previous issue of nearest hit and misses doesn't work. instead of that, a kind of probability that targets of two instances are different is introduced. Which can be modeled with the relative distance between the predicted (class) values of two instances. Rrelief F select random instance z_i and its k nearest instances Ri. The weights for different prediction value $\tau(\cdot)$, different attribute , and different prediction & different attribute are collected in*Eq.(25-27)*, respectively. Eq. (24) is reformulated, so that it can be directly evaluated using the probability that predicted values of two instances are different.

$$P_{diffT} = P(\text{different value of } T| \text{ nearest instances}) \quad (25)$$

$$P_{diffC} = P(\text{different prediction} \mid \text{nearest instances}) \quad (26)$$

$$P_{diffC|diffT} = P(\text{diff. prediction} \mid \text{diff. value of } T \text{and nearest instances}) \quad (27)$$

By using Bayes' rule, obtain from Eq.(24)

$$W[T] := \left(\frac{N_{DC\&DT}[T]}{N_{DC}}\right) - \left(\frac{(N_{DT}[T] - N_{DC\&DT}[T])}{(m - N_{DC})}\right) \quad (28)$$

The term $d(i, j)$ in Eq.(29)represent the distance between two instances *zi*and *Rj*. which takes into account that the closer instances should have greater influence on the weight of the attribute. exponentially decrease the influence of the instance *Rj*with the distance from the given instance *zi* as shown below.

$$d(i\,,j) = \frac{e^{-\left(\frac{rank(zi,Rj)}{\sigma}\right)^2}}{\sum_{l=1}^{k} e^{-\left(\frac{rank(zi,Rj)}{\sigma}\right)^2}} \quad (29)$$

4. Proposed System

Fig. 1 illustration the architecture of the proposed system for crude Oil.

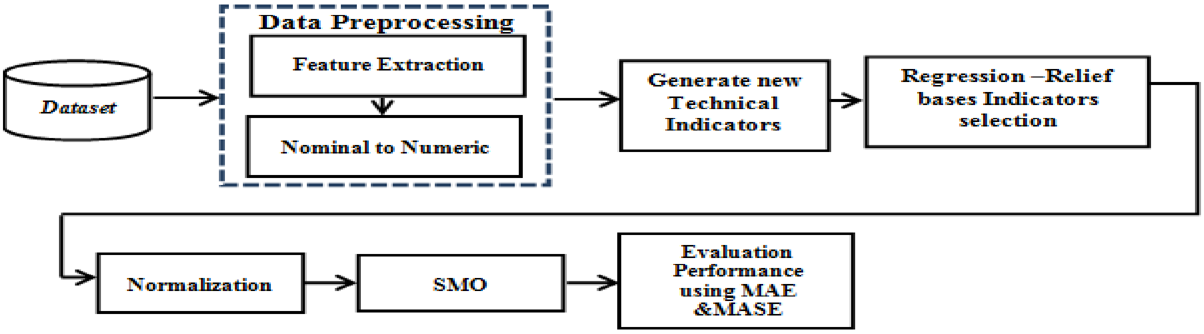

Fig. 1: The proposed system

4.1 Dataset

The dataset for our study is provided from U.S Energy Information Administration historical data. two global market West Texas Intermediate and Brent crude oil price had been taken for our experimental tests based on weekly periods, the period of WTI from 25-Dec-1987 to 29-Dec-2006for Brent and from 1986to29-Dec-2006 was selected for WTI as shown in figure (3a) and (3b) respectively.

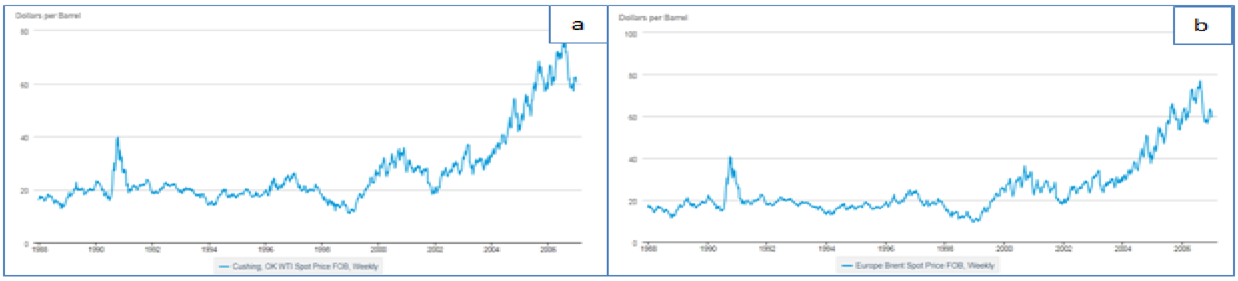

Fig. 2: the period time for a) WTI b) Brent

4.2. Pre-processing Stage

This stage consists of two steps include features extraction and transforming nominal to numeric:

Feature extraction:

In this step, Traditional weekly features (date, price) are taken as input to proposed model. Table(4-1) show the type of these features.

Feature's name	Type
Date	Nominal
Close	Numeric

Nominal to Numeric:

In this step, a transformation have been applied to convert categorical attribute into a numerical form. This step convert our (date) attribute into test then to numeric by merging the (day-month-year) and removing separator digit(/) .

Input: for each training instance a vector of attribute values T_i , i = 1, . . . , t, labeled with the target value τ j .

Output: the vector W of estimations of the qualities of attributes

begin

-set all weights W[A] and all N_{DC}, $N_{DT}[T]$, $N_{DT,DC}[T]$:=0;

-Set selected Sample =Z; and threshold value = 0 ;

Step.1. For each instance value (z_i) in Z

Randomly select instance z_i;

Step.2. Select k instances Ri nearest to z_i;

For each value in k do

A- compute difference between two instance by:

$$\text{diff}(T,z1,z2) = \frac{|\text{value}(T,z1) - \text{value}(T,z2)|}{\max(T) - \min(T)}$$

B- Compute the distance between the two instances zi and Rj by:

$$d(i\ ,j) = \frac{e^{-\left(\frac{\text{rank}(zi,Rj)}{\sigma}\right)^2}}{\sum_{l=1}^{k} e^{-\left(\frac{\text{rank}(zi,Rj)}{\sigma}\right)^2}}$$

C- update $N_{DC}$by:

N_{DC}:= N_{DC}+ diff(τ (・), z_i, R j) ・ d(i, j);

for each attribute in T

update $N_{DT}[T]$ and $N_{DC\&DT}[T]$by:

$N_{DT}[T]$:= $N_{DT}[T]$ + diff(T, z_i , R j) ・ d(i, j);

$N_{DC\&DT}[T]$:= $N_{DC\&DA}[T]$ + diff(τ (・), z_i , Rj) ・ diff(T, z_i , R j) ・ d(i, j);

Step.3. For each attribute t in T compute the weight by:

$$W[T] := \left(\frac{N_{DC\&DT}[T]}{N_{DC}}\right) - \left(\frac{(N_{DT}[T] - N_{DC\&DT}[T])}{(m - N_{DC})}\right)$$

Fig. 3: Algorithm of Regression -Relief Feature selection

4.3 Normalization:

Normalization in the feature space is not applied on the input vectors directly but can be seen as a kernel interpretation of the preprocessing. The main reason of this process is to avoid the variation of large values.

where:

$$\tilde{\boldsymbol{\varphi}}\ (\mathbf{x}) = \frac{\boldsymbol{\varphi}\ (\mathbf{x})}{|\boldsymbol{\varphi}(\mathbf{x})|} = \frac{\boldsymbol{\varphi}\ (\mathbf{x})}{\sqrt{\boldsymbol{K}(\boldsymbol{x},\boldsymbol{x})}} \quad (30)$$

stands for the "normalized" mapping and thus the expression above satisfies the conditions of Mercer's theorem.

4.4 Indicator Selection

In this stage Ranker Method have been used as indicators selection that's Rankattributes by their individual evaluations. Use in conjunction with different attribute evaluators like (Relief F, Gain Ratio, Entropy),that Specify a threshold by which attributes may be discarded from the ranking.in this paper Regression ReliefFeatures Selection algorithm has been selected as evaluator for the merit of the attributes.

The experiment result of Relief algorithm and the data of WTI and Brent from EIA is imported to calculate feature weights. the data weight value of attributes for Brent (Var., High, Midpoint, CCI, TSF, LOW, Lr-Lineage , Lr-Intercept , Plus-DM,APO, RSI,Lr-Slope, Aroon, Minus-DM, MACD, Minus-DI, Plus-DI)bigger than the weight value of observation, evaporation and crop whose value are below the threshold value. Therefore three indicators namely (EMA, Date, Aroon Osc) are rejected from the features subset, table(3)shows the Merits of attributes for Brent. While the merits of TI's for WTI reveal that three indicators (Date, Aroon, Aroon Osc) are rejected, as shown in Table(4).

Table 3: shows the Merit of TI's for Brent

NO.	TI	Merit	NO.	TI	Merit
1	Var.	0.031	11	LR-Slop	0.015
2	High	0.028	12	Minus-DM	0.013
3	Midpoint	0.028	13	CCI	0.012
4	TSF	0.027	14	Plus-DI	0.01
5	Low	0.027	15	Minus-DI	0.005
6	LR-Linearge	0.027	16	Aroon	0.003
7	LR-intercept	0.025	17	MACD	0.002
8	Plus-DM	0.02	18	**EMA**	**0**
9	APO	0.017	19	**Date**	**0**
10	RSI	0.015	20	**Aroon Osc**	**-0.001**

Table 4: Shows the Merits of TI's for WTI

NO.	**TI**	**Merit**	**NO.**	**TI**	**Merit**
1	0.034	Midpoint	**12**	0.012	Lr-slope
2	0.034	Min	**13**	0.011	RSI
3	0.034	Max	**14**	0.008	CCI
4	0.032	Lr-Linerage	**15**	0.008	TSF
5	0.03	Lr-Intercept	**16**	0.007	Plus-DI
6	0.03	Var.	**17**	0.003	Minus-Di
7	0.018	Plus-Dm	**18**	**0**	**Date**
8	0.017	EMA	**19**	**-0.004**	**Aroon**
9	0.014	APO	**20**	**-0.004**	**Aroon Osc**
10	0.013	MACD	**21**		
11	0.012	Minus-DM			

4.6. Evaluating Performance:

In this study Root Mean squared error (RMSE) and Mean Absolute error (MAE)are used here as the indicators of model performance to compare the results achieved by the proposed model with those obtained using other models.

$$\text{MAE} = \frac{1}{N}\sum_{i=1}^{N}|(X_t - \bar{X})^2 - \bar{p}| \quad (31)$$

$$\text{RMSE} = [\frac{1}{N}\sum_{i=1}^{N}(X_t - \bar{X})2\ - \bar{p}]^{1/2} \quad (32)$$

Where:

X_t: The return of the horizon before the current time t

$\bar{X}$: The average return

p_t: Is the forecast value of the conditional variance over n steps ahead horizon of the current time t.

5. Experimental Results

In order to evaluate the efficiency of the SMOreg, this paper have been implemented a crude oil forecasting system using Java programming language, Two metrics (RMSE and MAE) are used to evaluate the regression predictor model. The model has been applied with datasets contain only standard features(SF) and standard features with indicators(SFI), respectively as they are summarized in table (5).

4.5. Modified SMOreg:

Input: training instance a vector of attribute values X_i , i = 1, . . . , m, labeled with the target value y, C(say 1.0), kernel(RBF Kernel), error cache, epsilon(tolerance=0.001).
Output: sum of weight vector w,α array
Initialize: b, w and all α's to 0
Step1: Repeat until KKT satisfied (to within epsilon):
A-Find an instance(i1) that violates KKT (with unbound Lagrange multiplier ($0<\alpha_i<C$), choose randomly among those)
B-Choose a second instance(i2). that maximize step size (in practice, faster to just maximize $|E_1 - E_2|$).
- If that fails to result in change, randomly choose unbound example.
- If that fails, randomly choose example. If that fails, re-choose i1.
Step2: Compute b_{up} and b_{low}.
Until $b_{up} \geq b_{low}+2\varepsilon$
Step3: find optimal values for these two multipliers
Step4: Update the threshold value b_{low} and b_{up} and store theα_1, α_2
Step5: Repeat Step 1 through Step 4 until the algorithm converges

Fig. 4: The algorithm of modified SMO

Table 5: MAE &RMSE for both markets

Dataset	*Features*	*Time(ms)*	*MAE*	*RMSE*	*No. Instance*
Brent	SF	1.05	8.7496	14.7726	993
Brent	SFI	2.42	0.9088	1.1434	993
WTI	SF	1.18	8.6532	14.7254	1025
WTI	SFI	2.02	0.6666	0.8449	1025

Additionally, a comparative study with two standard regression predictors Gaussian process based RBF kernel and Linear regression(LR). The Gaussian, LR, and SMO testing with cross-validation (10-fold) and compared to checking the performance. The results of RMSE and MAE to these regression predictors have applied on tow datasets with all features and technical indicators are summarized in tables (6).

Table 6: MAE &RMSE for both markets

Dataset	Algorithm	Time(ms)	MAE	RMSE	No. Instance
Brent	SMOreg	1.75	0.7813	1.021	993
WTI	SMOreg	1.39	0.5054	0.6978	1025
WTI	MLP regressor	274.83	1.4991	1.962	1025
Brent	MLP regressor	0.15	0.3447	0.5016	993
WTI	Gaussian	3.24	1.1589	1.5161	1025
Brent	Gaussian	3.65	1.8296	2.735	993
WTI	Linear regression	4.83	9.6837	13.5475	1025
Brent	Linear regression	11.74	9.7375	13.6152	993

Table (7) show that the ranker search method with relief algorithm can be successfully used for indicators selection for both markets with better results. Fig.(5) show visually the bar charts of RMSE and MAE.

Table 7: shows the effect of IS on WTI &Brent

Dataset	Algorithm	Time	MAE	RMSE	No. Instance
Brent	With-FS	1.18	0.6798	0.8687	993
WTI	With-FS	1.18	0.5654	0.721	1025

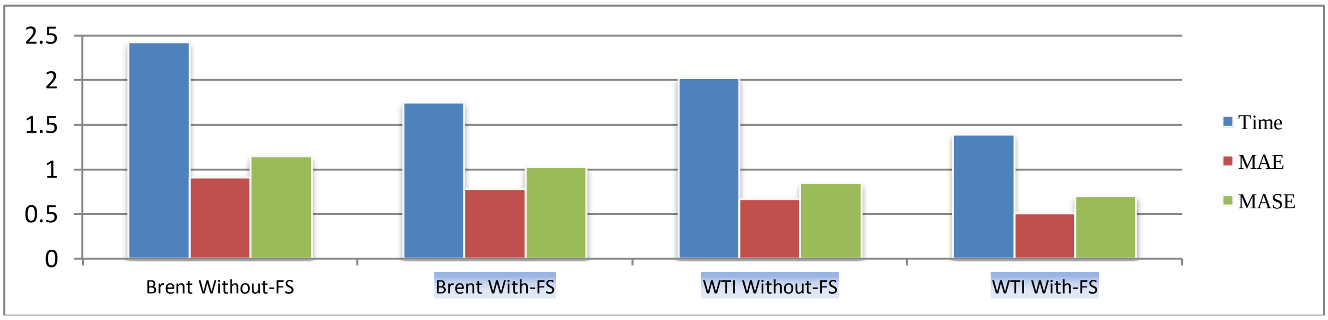

Fig. 5: Shows the effect of IS on both market

6. Conclusion

In this paper, the sequential minimal optimization algorithm (SMO) has been employed to be an effective method for training support vector regression (SVR) on regression tasks defined on Crude oil data sets. This paper exploits technical indicators which have been used by economists and traders to predicatethe price of the most two global market(WTI and Brent)by extracting hidden patterns.TI's are demonstrated impact on the efficiency of the prediction model, then ranker search method with relief algorithmis proposed to select the best collection of indicators for each market. According to our results, indicators improve the performance of SVR with MAE (0.7) and (0.5) for Brent and WTI,respectively.

References

[1] J. Alvarez-Ramirez, A. Soriano, M. Cisneros, and R. Suarez, "Symmetry/anti-symmetry phase transitions in crude oil markets," Physica A: Statistical Mechanics and its Applications, vol. 322, pp. 583-596, 2003/05/01/ 2003.

[2] P. André and W. G.C., "How volatile are crude oil prices?," OPEC Review, vol. 18, no. 4, pp. 431-444, 1994.

[3] B. Lebaron and J. Alexandre Scheinkman, Nonlinear Dynamics and Stock Returns. 1989, pp. 311-37.

[4] V. L. Nadh and G. S. Prasad, "Support vector machine in the anticipation of currency markets," Int. J. Eng. Technol, vol. 7, no. 2-7, p. 66, 2018.

[5] L. Cao and F. E. H. Tay, "Financial Forecasting Using Support Vector Machines," Neural Computing & Applications, journal article vol. 10, no. 2, pp. 184-192, May 01 2001.

[6] C. Shiyi, K. H. r. Wolfgang, and J. Kiho, "Forecasting volatility with support vector machine-based GARCH model," Journal of Forecasting, vol. 29, no. 4, pp. 406-433, 2010.

[7] K.-j. Kim, "Financial time series forecasting using support vector machines," Neurocomputing, vol. 55, no. 1, pp. 307-319, 2003/09/01/ 2003.

[8] Z. Xiao-lin and W. Hai-wei, "Crude Oil Prices Predictive Model Based on Support Vector Machine and Particle Swarm Optimization," Berlin, Heidelberg, 2012, pp. 645-650: Springer Berlin Heidelberg.

[9] S. Deng and A. Sakurai, Crude Oil Spot Price Forecasting Based on Multiple Crude Oil Markets and Timeframes. 2014, pp. 2761-2779.

[10] P. Fernandez-Blanco, D. J. Bodas-Sagi, F. J. Soltero, and J. I. Hidalgo, "Technical market indicators optimization using evolutionary algorithms," presented at the Proceedings of the 10th annual conference companion on Genetic and evolutionary computation, Atlanta, GA, USA, 2008.

[11] M. M. Mostafa and A. A. El-Masry, "Oil price forecasting using gene expression programming and artificial neural networks," Economic Modelling, vol. 54, pp. 40-53, 2016/04/01/ 2016.

[12] R. Samsudin and A. Shabri, "Crude oil price forecasting with an improved model based on wavelet transform and support vector machines," EJ. Artif. Intell. Comput. Sci, vol. 1, pp. 9-15, 2013.

[13] N. I. Nwulu, "A decision trees approach to oil price prediction," in 2017 International Artificial Intelligence and Data Processing Symposium (IDAP), 2017, pp. 1-5.

[14] J. C. Platt, "Fast training of support vector machines using sequential minimal optimization," in Advances in kernel methods, S. Bernhard, lkopf, J. C. B. Christopher, and J. S. Alexander, Eds.: MIT Press, 1999, pp. 185-208.

[15] A. J. Smola, B. Sch, #246, and lkopf, "A tutorial on support vector regression," Statistics and Computing, vol. 14, no. 3, pp. 199-222, 2004.

[16] S. Pandharipande, A. Akheramka, A. Singh, and A. Shah, "Artificial Neural Network Modeling of Properties of Crude Fractions with its TBP and Source of Origin and Time," International Journal of Computer Applications, vol. 52, no. 15, 2012.

[17] D. S. Grebenkov and J. Serror, "Following a trend with an exponential moving average: Analytical results for a Gaussian model," Physica A: Statistical Mechanics and its Applications, vol. 394, pp. 288-303, 2014/01/15/ 2014.

[18] S. M. Nor and G. Wickremasinghe, "The profitability of MACD and RSI trading rules in the Australian stock market," Investment Management and Financial Innovations, vol. 11, no. 4, p. 194, 2014.

[19] M. Maitah, P. Procházka, and M. Cermak, "Commodity Channel Index: Evaluation of Trading Rule of Agricultural Commodities," International Journal of Economics and Financial Issues, vol. 6, no. 1, 2016.

[20] P. Selvam, "A study on Market Trend Prediction using “Aroon Oscillator” with special reference to the Indian private sector banks," Journal of Advance Management Research, 2017.

[21] A. Steven B, Technical Analysis from A to Z. Chigaco: Probus Publishing, 1995.

[22] Y. Zhang, H. Wu, and L. Cheng, "Some new deformation formulas about variance and covariance," in Modelling, Identification & Control (ICMIC), 2012 Proceedings of International Conference on, 2012, pp. 987-992: IEEE.

[23] M. Walker Helen, Studies in the history of statistical method. The Williams And Wilkins Company; Baltimore, 1929.

[24] Z. Bitvai and T. Cohn, "Day trading profit maximization with multi-task learning and technical analysis," Machine Learning, vol. 101, no. 1-3, pp. 187-209, 2015.

[25] K. Kira and L. A. Rendell, "A Practical Approach to Feature Selection," presented at the Proceedings of the Ninth International Workshop on Machine Learning, 1992.