Crude oil price prediction using artificial neural network-backpropagation (ANN-BP) and particle swarm optimization (PSO) methods

Fitri Amalia Langgundi¹, Aji Purwinarko²
¹,²Department of Computer Science, Universitas Negeri Semarang, Indonesia

ABSTRACT
Crude oil price fluctuations significantly affect commodity market price fluctuations, so a sudden drop in oil prices will cause a slowdown in the economy and other commodities. This is very important for Indonesia, one of the world's oil-producing countries, to gain multiple benefits from oil exports when world oil prices increase and increase economic growth. Therefore, a system is needed to predict world crude oil prices. In this case, the Particle Swarm Optimization (PSO) algorithm is applied as the optimization of the weight parameters in the Artificial Neural Network-Backpropagation (ANN-BP) method. We compared the ANN-BP–PSO and ANN-BP methods to obtain the method with the best causation value based on the MAPE and MSE results. PSO aims to find the best weight value by iterating the process of finding and increasing position, speed, Pbest, and Gbest until the iteration is complete. The results showed that the ANN-BP-PSO process was classified as very good and had a lower predictive error rate than the ANN-BP method based on the MAPE and MSE values, which is 5.02007% and 7.15827% compared to 6.28323% and 13.86345.

Keywords: Prediction, Backpropagation, Neural network, Particle swarm optimization, Crude oil

INTRODUCTION
One of the petroleum products found in nature is crude oil, which comprises unprocessed hydrocarbon reserves and other organic substances [1]–[3]. Crude oil is one of the most strategic commodities traded on the world market because it plays an essential role in society, including economy, politics, and technology [4]–[7]. Crude oil has been traded internationally by oil-producing countries, companies, and individuals to oil-importing countries. Crude oil prices are influenced by stock levels, economic growth, political aspects, political instability, Organization of the Petroleum Exporting Countries (OPEC) decisions, and different psychology from traders [8]–[10].

The fluctuations in crude oil prices affect commodity market price fluctuations, so any sudden increase or decrease in crude oil prices causes a slowdown in the economy and other commodities [11]. Therefore, the predictive ability of crude oil prices helps manage most industrial sectors worldwide [8], [12]
because this is very important for Indonesia, one of the world's oil-producing countries, to gain multiple benefits from oil exports when world oil prices increase and to increase economic growth.

Predicting or estimating the future prices of crude oil is complex and challenging because numerous factors impact its behavior and fluctuations within the worldwide market for this commodity [13]–[15]. These factors could include supply and demand dynamics, geopolitical events, economic indicators, environmental regulations, and natural disasters [15]. Therefore, accurately forecasting the price of crude oil requires a comprehensive understanding of the various factors at play in the global oil market.

Predictions can be made using various methods and tools, including computer technology. Prediction involves estimating something that will happen by studying the patterns of existing data [16]. One commonly used methods for making predictions is the Artificial Neural Network (ANN) method. ANN is a computational method that attempts to replicate the workings of a neural network [17]. Backpropagation is one of the methods in ANN [18], a supervised learning algorithm often used to make predictions from time series data.

However, Ramyar and Kianfar [19] showed that the ANN model with multi-layer perceptron architecture predicts crude oil prices more accurately than the Vector Autoregressive model. ANN Backpropagation (ANN-BP) is generally used in networks with multi-layer perceptions [20], [21]. ANN using the Backpropagation algorithm will work in a supervised manner so that this method can be used for prediction and data classification [17], [22], [23]. Determining the parameters used in ANN-BP forecasting greatly influences the performance of the ANN-BP process. So, we need an algorithm to optimize the determination of the ANN-BP input parameters so that ANN-BP can work optimally [19].

Several researchers have used the Particle Swarm Optimization (PSO) algorithm to optimize a method of finding the best parameters for a forecast. Zhu et al. [24] employed the PSO algorithm to determine the optimal weight and bias parameters in an Extreme Learning Machine (ELM) model for predicting daily evapotranspiration. Their results showed that using the PSO algorithm led to a smaller RMSE value than not using it, with a difference of 17.1% to 17.9% for the radiation-based model and 20.6% to 21.7% for the temperature-based model. Rusmalawati et al. [25], the PSO algorithm has helped optimize Support Vector Regression (SVR) to form a forecasting model through sequential learning SVR. The evaluation results are obtained and classified as having a high level of accuracy because it has a MAPE value of 0.8195% with a fitness of 0.5496. PSO is a heuristic optimization technology to find the optimal solution based on an objective function based on the behavior of a flock of birds in carrying out their duties [26]–[28].

Furthermore, this study proposes the ANN-BP prediction model with the PSO algorithm to optimize weight parameters in predicting world crude oil prices. The model made in this study is expected to provide low MSE and MAPE results to obtain accurate predictions.

2. METHOD

The method used in this study is ANN-BP as a prediction method with optimization of weight parameters using the PSO algorithm to predict world crude oil prices. This study has six main stages of increasing accuracy through the ANN-BP process and the PSO algorithm. The six stages are the data collection stage, data normalization, parameter testing, model determination, prediction, and conclusion drawing. The following flowchart for determining the model using ANN-BP – PSO is shown in Figure 1.

2.1. Dataset Collection

The dataset used in this study is public, namely the Crude Oil WTI standard (CL=F) dataset obtained from the finance.yahoo.com website. A total of 1058 data were used in the study from January 3, 2017 to March 31, 2021 which was accessed on April 5, 2021 with West Texas Intermediate (WTI) standard size in U.S Dollars. The price used in this study is the close price because it is the price that can be a reference for predicting the open price on the next day. The data is divided into 70% training and 30% test data. The distribution of this data is based on research conducted by [9], which managed to get an accuracy rate of 99.25%. Table 1 shows the daily close price of world crude oil.
Table 1 World crude oil close price

<table>
<thead>
<tr>
<th>Date</th>
<th>Close</th>
</tr>
</thead>
<tbody>
<tr>
<td>1/3/2017</td>
<td>52.33</td>
</tr>
<tr>
<td>1/4/2017</td>
<td>53.26</td>
</tr>
<tr>
<td>1/5/2017</td>
<td>53.76</td>
</tr>
<tr>
<td>1/6/2017</td>
<td>53.99</td>
</tr>
<tr>
<td>1/9/2017</td>
<td>51.96</td>
</tr>
<tr>
<td>1/10/2017</td>
<td>50.82</td>
</tr>
<tr>
<td>1/11/2017</td>
<td>52.25</td>
</tr>
<tr>
<td>1/12/2017</td>
<td>53.01</td>
</tr>
<tr>
<td>1/13/2017</td>
<td>52.37</td>
</tr>
<tr>
<td>1/17/2017</td>
<td>52.48</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>3/31/2021</td>
<td>59.16</td>
</tr>
</tbody>
</table>

Figure 1. ANN-BP and PSO Model Determination Flowchart
2.2. Data Normalization
For the method to be used to recognize data as input, it is necessary to normalize the data using a scale in the interval [0.1] in Equation 1.

\[ X'_i = \frac{x_i - \text{min}}{\text{max} - \text{min}} \]  \hspace{1cm} (1)

where,

- \( X'_i \): normalization data
- \( x_i \): data to be normalized
- \( \text{min} \): smallest data
- \( \text{max} \): biggest data

Table 2 shows the world crude oil price dataset before and after normalization in the interval [0.1].

<table>
<thead>
<tr>
<th>Date</th>
<th>Original data</th>
<th>Normalized Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>1/3/2017</td>
<td>52.33</td>
<td>0.637349</td>
</tr>
<tr>
<td>1/4/2017</td>
<td>53.26</td>
<td>0.651355</td>
</tr>
<tr>
<td>1/5/2017</td>
<td>53.76</td>
<td>0.658886</td>
</tr>
<tr>
<td>1/6/2017</td>
<td>53.99</td>
<td>0.662349</td>
</tr>
<tr>
<td>1/9/2017</td>
<td>51.96</td>
<td>0.631777</td>
</tr>
</tbody>
</table>

2.3. Parameter Testing
At this stage, the parameters for ANN-BP and PSO parameters are tested. ANN-BP parameters tested include testing the number of input neurons and hidden neurons, the number of iterations (epochs), and the learning rate. At the same time, the PSO parameters tested include epochs and values of r1 and r2. Parameter testing was carried out using the ANN-BP training process with 70% dataset. After testing the parameter values, the best parameter values are selected through the lowest MSE and MAPE results in Table 3. The ANN-BP algorithm is shown in Figure 2 [18].

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value/Amount</th>
</tr>
</thead>
<tbody>
<tr>
<td>PSO</td>
<td></td>
</tr>
<tr>
<td>Number of particles</td>
<td>15</td>
</tr>
<tr>
<td>Popsize</td>
<td>5</td>
</tr>
<tr>
<td>c1</td>
<td>1</td>
</tr>
<tr>
<td>c2</td>
<td>1.5</td>
</tr>
<tr>
<td>Inertia weight</td>
<td>0.5</td>
</tr>
<tr>
<td>Epoch</td>
<td>16</td>
</tr>
<tr>
<td>ANN-BP</td>
<td></td>
</tr>
<tr>
<td>Input layer</td>
<td>5</td>
</tr>
<tr>
<td>Hidden layer</td>
<td>3</td>
</tr>
<tr>
<td>Output layer</td>
<td>1</td>
</tr>
<tr>
<td>Epoch</td>
<td>60</td>
</tr>
<tr>
<td>Learning rate</td>
<td>0.2</td>
</tr>
</tbody>
</table>

2.4. Model Determination
Determination of the model using the ANN-BP - PSO method through a training process. The dataset used is 70% of the total data. The optimization carried out by PSO in ANN-BP aims to produce the lowest error rate. PSO optimizes the ANN-BP parameters, namely weight updates so that it is expected to increase prediction accuracy. This process continues until the ANN-BP and PSO epochs have reached their limit. The execution process using this combined method takes quite a long time to adjust the number of epochs used.
2.5. Prediction

The prediction process is carried out using datasets as much as 30% of the total data used. The prediction process follows the ANN-BP testing stages based on the model results from the ANN-BP and PSO training processes.

The method's success in this study is determined using indicators of predictive accuracy. These indicators are Mean Square Error (MSE) and Mean Absolute Percentage Error (MAPE). MSE is a method used to evaluate forecasting models through each error or residual squared, then summed and added to the number of observations [29]–[31]. The MSE formula can be seen in Equation 2.

\[
MSE = \frac{1}{n} \sum_{t=1}^{n} (Y_t - \hat{Y}_t)^2
\]

where,
\(n\) : number of data points
\(Y_t\) : observed value
\(\hat{Y}_t\) : Predicted value

Due to its ability to be applied to various contexts, easily understood, and dependable, MAPE is regarded as the most widely used method for measuring accuracy [32], [33]. MAPE indicates how big the error is in forecasting compared to the actual value [34], [35]. The MAPE formula can be seen in Equation 3.

\[
MAPE = \frac{1}{n} \sum_{t=1}^{n} \left| \frac{X_t - \hat{X}_t}{X_t} \right| \times 100\%
\]

where,
\(X_t\) : time series value in the t-period
\(\hat{X}_t\) : forecast value in the t-period
\(n\) : total number of observations

3. RESULTS AND DISCUSSIONS

The results of determining the model using the ANN-BP and ANN-BP - PSO methods, with 70% of the training data getting MSE and MAPE values, are given in Table 5. The PSO algorithm has succeeded in...
optimizing the weight parameter (w) in the ANN-BP training process. ANN-BP and PSO training runs in each iteration. The best position is obtained, followed by updating the weight, speed, position, Pbest, and fitness to determine Gbest until the iteration is complete. The selection of parameter values determined through parameter testing has improved accuracy. The parameters obtained from the testing process were the architecture of the ANN-BP model and the PSO parameter values. The PSO parameter values comprised 15 particles, 5 popsize, an epoch value of 16, a c1 value of 1, a c2 value of 1.5, and an inertia weight value of 0.5. Meanwhile, the ANN-BP model architecture comprised 5 input layers, 3 hidden layers, 1 output layer, an epoch value of 60, and a learning rate value of 0.2.

The MSE and MAPE values generated from the training process in the search for the best parameter model using the ANN-BP and PSO methods are 1.97637 and 1.85356%, respectively. While the MSE and MAPE values using only the ANN-BP method in the training process are 2.25938 and 3.03976%, with PSO fitness results of 0.9818. These results indicate that PSO has optimized ANN-BP to get a minor error value, so the prediction model is tested in the ANN-BP training process. MSE and MAPE results from the prediction process are shown in Table 6.

The results obtained in the prediction process are the MSE and MAPE values. The MSE and MAPE values generated by the prediction process using the ANN-BP and PSO methods are 7.15827 and 5.02007%, respectively. Meanwhile, the results of MSE and MAPE, which only used the ANN-BP method, were 13.86345 and 6.28323%. The smaller the MSE value obtained, the better the forecasting performance [36].

Although the PSO algorithm can improve the accuracy and minimize the error value in the ANN-BP method, the training process is quite time-consuming [37]. This is because each epoch in ANN-BP performs weight update calculations in each PSO epoch. Therefore, as the value of the ANN-BP and PSO epochs increases, the weight update process will also take longer.

The study using the ANN-BP-PSO model obtained better forecasting results with a high level of accuracy in predicting crude oil prices based on daily time series compared to studies that used the ARIMA method [38]. Edited Nearest Neighbor (ENN) [39], Local Mean Decomposition (LMD)-ARIMA [40], and Naive [39], [40].

4. CONCLUSION

The application of the PSO algorithm in optimizing the weight parameter (w) of ANN-BP makes the prediction quality of crude oil prices increase, as evidenced by the results of MSE and MAPE ANN-BP – PSO is better than using only ANN-BP. Based on the results of the MAPE and MSE values, the testing process using the PSO algorithm in the ANN-BP method, which is 7.15827 and 5.02007%, indicates that the ANN-BP – PSO method is classified as very good and has a smaller error rate compared to using only ANN-BP method only. The prediction error value obtained decreased by 1.26316% compared to using only the ANN-BP model, which had MSE and MAPE values of 13.86345 and 6.28323% on the WTI standard Crude Oil object (CL=F).

REFERENCES


Table 4. Results of MSE and MAPE training process

<table>
<thead>
<tr>
<th>No</th>
<th>Method</th>
<th>MSE</th>
<th>MAPE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>ANN-BP</td>
<td>2.25938</td>
<td>3.03976%</td>
</tr>
<tr>
<td>2</td>
<td>ANN-BP-PSO</td>
<td>1.96737</td>
<td>1.85356%</td>
</tr>
</tbody>
</table>

Table 5. Prediction results

<table>
<thead>
<tr>
<th>No</th>
<th>Method</th>
<th>MSE</th>
<th>MAPE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>ANN-BP</td>
<td>13.86345</td>
<td>6.28323%</td>
</tr>
<tr>
<td>2</td>
<td>ANN-BP-PSO</td>
<td>7.15827</td>
<td>5.02007%</td>
</tr>
</tbody>
</table>
T. no. 14, pp. 16701
E. Windsor and W. Cao, “Improving exchange rate forecasting via a new deep multimodal fusion model,”
10.1109/TSG.2020.2971339.
Assessment in Local Energy Markets,”
S. Schreck, I. Prieur de La Comble, S. Thiem, and S. Niessen, “A Methodological Framework to support Load Forecast Error
US,”
M. Kumar, S. Gupta, K. Kumar, and M. Sachdeva, “SPREADING OF COVID
Ser.
S. Amar, A. Sudiarso, and M. K. Herliansyah, “The Accuracy Meas
Smart Grid,”
Y. H
in Robotic Path Planning,” in
Swarm Optimization,”
V. Rusmalawati, M. T. Furqon, and Indriati, “Peramalan Harga Saham Menggunakan Metode Support Vector Regression (SVR)