https://journal.uinsgd.ac.id/index.php/kubik/index
Vol. 10, No. 1, 2025, pp. 34-47
e-ISSN: 2686-0341 p-ISSN: 2338-0896

Implementation of Bidirectional Long Short-Term Memory (BiLSTM) to Predict World Crude Oil Prices

Firda Yunita Sari[1], Nurissaidah Ulinnuha[2, *)]

[1,2]*Department of Mathematics, Faculty of Science and Technology, UIN Sunan Ampel, Surabaya, Indonesia*

[*)]email: nuris.ulinnuha@uinsa.ac.id

Abstract

The primary energy source worldwide is crude oil, which almost all countries use as an energy source. Crude oil is key in driving the global economy, especially in the industrial and transportation sectors. Due to the growth in technology, artificial intelligence techniques such as Bidirectional Long-Term Memory (BiLSTM) that use past and future information to be more accurate can be used to forecast crude oil prices. The objective of this study is to use BiLSTM with 57 test cases to forecast global crude oil prices for one year based on the change of parameters such as activation function, batch size, number of neurons, and sharing data. The BiLSTM model achieved the lowest MAPE value of 0.09% and the optimal performance when the data splitting was 90:10, the number of neurons was 100, the batch size was 4, and the activation function was ReLU. These imply that BiLSTM can effectively forecast crude oil prices with precision. It is hoped that this study will be a reference for developing more accurate crude oil price predictive models to be utilized by industry and research for the coming years.

Keywords: BiLSTM, MAPE, Parameters, Prediction, World Crude Oil Prices
MSC2020: 62P05, 62P20

Abstrak

Minyak mentah adalah sumber energi utama di dunia dan digunakan oleh hampir semua negara. Perekonomian global secara signifikan didorong oleh minyak mentah, terutama di sektor transportasi dan industri. Seiring perkembangan teknologi, prediksi harga minyak mentah dapat dilakukan dengan metode kecerdasan buatan, salah satunya adalah Bidirectional Long Short-Term Memory (BiLSTM), yang menggabungkan informasi masa lalu dan masa depan untuk meningkatkan akurasi. Penelitian ini bertujuan untuk memprediksi harga minyak mentah dunia selama satu tahun menggunakan BiLSTM dengan 57 skenario pengujian berdasarkan variasi parameter seperti pembagian data, jumlah neuron, batch size, dan fungsi aktivasi. Hasil terbaik diperoleh pada model BiLSTM dengan pembagian data 90:10, jumlah neuron 100, ukuran batch 4, dan fungsi aktivasi ReLU, yang menghasilkan nilai MAPE terkecil sebesar 0.09%. Temuan ini menunjukkan bahwa BiLSTM mampu memberikan prediksi harga minyak mentah dengan tingkat akurasi tinggi. Penelitian ini diharapkan dapat menjadi referensi dalam pengembangan model prediksi harga minyak mentah yang lebih akurat serta bermanfaat bagi industri dan penelitian di masa depan.

Kata kunci: BiLSTM, MAPE, Parameter, Prediksi, Harga Minyak Mentah Dunia

*) Corresponding Author
Received: 06-11-2024, Accepted: 11-05-2025, Published: 30-05-2025

MSC2020: 62P05, 62P20

Citation: F. Y. Sari, and N. Ulinnuha, “Implementation of Bidirectional Long Short-Term Memory (BiLSTM) to Predict World Crude Oil Prices”, *KUBIK J. Publ. Ilm. Mat.*, Vol. 10, No. 1, pp. 34-47, 2025.

Introduction

Crude oil is the primary energy source in the world and is utilized by practically every nation [1]. Crude oil is significant in propelling the world economy, particularly in the transportation and industrial sectors. Crude oil is a naturally occurring yellow-black liquid beneath the Earth's surface. It is the primary fuel for almost one-third of the world's energy consumption and the world's primary energy source. Refined crude oil is also used to make petroleum products. Because of their indisputable contribution to global warming, the use of fossil fuels like crude oil is declining in popularity [2]. Many nations, both oil producers and consumers, are concerned about oil prices, which frequently fluctuate significantly and have a significant effect on the economy [3]. Having the ability to make correct crude oil price predictions is therefore crucial in helping nations with policy decisions, economic planning, and controlling the effects of oil price volatility on economic stability [4].

With the development of technology, more advanced techniques based on artificial intelligence can be used to estimate crude oil prices [5]. Using data from January 2018 to August 2023, some studies use the Support Vector Regression (SVR) technique to forecast crude oil prices. Crude oil provided the best results, with a MAPE value of 49.73%, which was considered sufficient [6]. SVR has limitations in collecting complex patterns of time series data influenced by numerous external variables, although it might yield excellent results. Therefore, processing increasingly complicated temporal dependencies and modeling non-linear relationships require an increasingly adaptive approach [7].

Using artificial neural networks, which can rapidly and precisely process and analyze historical oil price data, is an effective technique. Bidirectional Long Short-Term Memory (BiLSTM) is one form of artificial neural network technique [8]. Long Short-Term Memory (LSTM) technique was developed into the BiLSTM technique, which employs both forward and backward LSTM simultaneously to achieve maximum accuracy in processing sequence data by integrating past and future information [9]. Input to output and output to input are used for the model training. BiLSTM splits each unit into two units, where both have the same input but are connected to the same output. Hence, BiLSTM gives the benefit of learning immense time series data [10].

Numerous research studies have confirmed the advantage of BiLSTM compared to other prediction techniques. BiLSTM was the best in terms of prediction when compared to LSTM, Autoregressive Integrated Moving Average (ARIMA), Support Vector Regression (SVR), and BiLSTM models in financial prediction. The BiLSTM model outdid other approaches with the lowest Root Mean Square Error (RMSE) of 0.6056 and the smallest Mean Absolute Percentage Error (MAPE) of 0.4482. Aside from that, BiLSTM's optimal goodness-of-fit value of 0.9967 demonstrated its capability to identify intricate patterns in financial time series data effectively. Because economic data commonly contains intricate temporal correlations, BiLSTM's edge is due to its ability to consider bidirectional dependencies. BiLSTM gives a more general and robust strategy to time series forecasting compared to SVR, which can falter at long-range connections, and ARIMA, which hints at linearity [10].

Based on GNSS-adjusted tide gauge (TG) vertical land movement (VLM) data, sea level rise (SLR) was examined. A deep learning stacked BiLSTM model was developed based on input oceanographic features pre-processed by SVMD and BRO. With >0.99 correlation values and minimum error, the results show that the stacked BiLSTM possesses maximum accuracy compared to AdaBoost, SVR, and MLR [11]. Besides, BiLSTM was superior to LSTM and RNN in studies on the prediction of sea level rise with extended prediction coverage from 48 to 168 hours and accuracy from 0.9665 to 0.9787 [12] cases. The current research compares BiLSTM with Support Vector Regression (SVR), Autoregressive Integrated Moving Average (ARIMA), LSTM and Gated Recurrent Unit (GRU). The BiLSTM gives the lowest error value with MAE 0.0070 and RMSE 0.0077. Past studies indicate that BiLSTM works well in forecasting time series data patterns.

Based on the background and supported by previous research, this study aims to predict world crude oil prices for one year using historical data with the BiLSTM method to produce more accurate predictions related to daily movements in crude oil prices. In forming the best model, a combination of parameters is used: data division, batch size, number of neurons, and type of activation function. It is expected that the results of this study can provide a more accurate and reliable prediction model in analyzing crude oil price trends, so that it can be used as a reference in decision-making for the government, industry players, and investors in the energy sector.

Methods

The data in this study are daily data on world crude oil prices for 1 year from 2022 to 2023 obtained from investing.com [13]. The method used is one architecture of RNN, namely Bidirectional Long Short-Term Memory (BiLSTM) which is a development of the LSTM algorithm commonly used for deep learning problems [14]. BiLSTM processes sequential data with the use of both future and past information. Since the use of forward and backward LSTM, BiLSTM is capable of generating more accurate predictions such as crude oil price forecasts [10]. Figure 1 shows the flowchart used in this research.

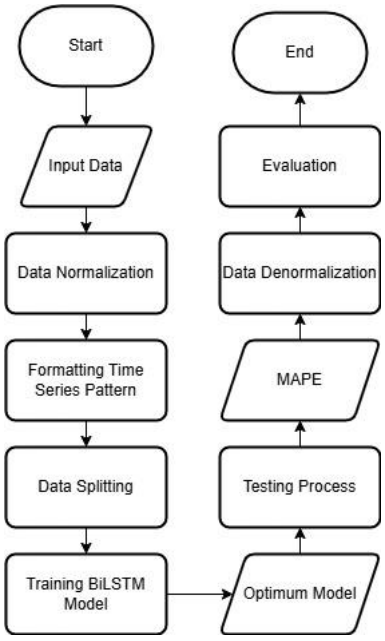

Figure 1. Research Flowchart

The research stages based on the flowchart in Figure 1 are as follows:

1. Collecting raw data that will be used in the training and testing process.

2. Performing data normalization to ensure uniform feature scaling and improve model training stability.
3. Formatting the data into time series patterns to suit the needs of the BiLSTM model.
4. Divide the data into training and testing data with a ratio of 70:30, 80:20, or 90:10.
5. Training the BiLSTM model with the training data. The model uses a BiLSTM architecture consisting of forward and backward layers at this stage. The training process is performed by applying the parameters listed in Table 1, including the activation function, number of neurons, and various batch sizes. During training, the model optimizes the weights and biases using ReLU or Sigmoid activation functions to calculate forget gate, input gate, and output gate to produce an optimal data representation.
6. Using test data to test the trained model, calculate predictions, and evaluate model performance based on metrics such as MAE and RMSE.
7. Denormalize the prediction results to get the original value of the model results.
8. Evaluate the model by comparing the prediction results with actual data to assess the accuracy and effectiveness of BiLSTM.

Data Normalization

Normalization is a method of performing linear transformation of the original data with the output in the form of a of each variable in the same range, namely 0 to 1, so that when calculating the similarity value, each variable tested gives the same level of importance or has the same influence [15].

$$x_{normalized} = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (1)$$

Information:

x = the data of each column

x_{min} = the minimum value of data in each column

x_{max} = the maximum value of data in each column

Time series pattern formation

This stage aims to organize data serially based on time series [16]. By looking at the correlation between previous values (input) and expected future values (output), it is possible to predict future trends based on historical data [17].

Data Splitting

The process starts with dividing time series data into training and testing data to examine the model's generalization ability. The data is split at some ratio, i.e., 70:30, 80:20, or 90:10, where most of the data is used to train the BiLSTM model. In contrast, the rest of the data evaluates the model's prediction ability. This division ensures that the model is trained based on past data and validated using data that has never been seen before.

Prediction modeling using the BiLSTM method

Bidirectional Long Short-Term Memory (BiLSTM) is a two-layer extension of the LSTM model that functions simultaneously in opposite directions. The BiLSTM architecture is shown in Figure 2.

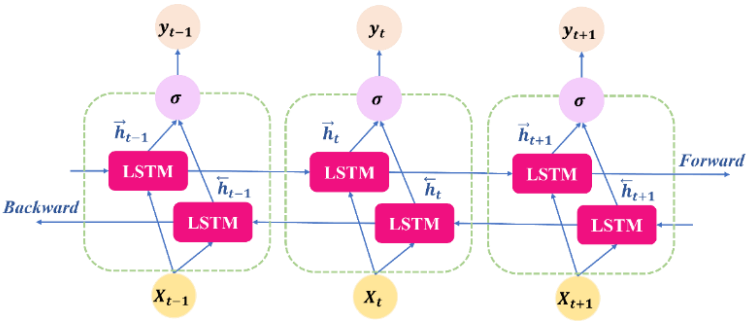

Figure 2. Architecture of BiLSTM

The BiLSTM model formation step begins with the LSTM process shown in the following equation [18].

$$f_t = \sigma\left(W_f . [h_{t-1}, x_t] + b_f\right) \tag{2}$$

f_t is the forget gate, where σ is the default recurrent activation function of sigmoid, thich has a range of [0,1] with the formula $\sigma(x) = \frac{1}{1+e^{-x}}$. x_t is the input value in order t, W_f is the forget gate weight and b_f is the forget gate bias. h_t is is the output values in t. The model output at time t is obtained from the previous model output, t-1. Equation (2) does not directly use the previous cell state C_{t-1} as input, but instead computes a gating factor f_t, which will be applied C_{t-1} in Equation (5) to determine how much of the previous memory should be carried forward.

$$i_t = \sigma(W_i . [h_{t-1}, x_t] + b_i) \tag{3}$$

i_t the input gate regulates how much new information should be written to the cell state. Similar to the forget gate, it uses h_{t-1} and x_t as inputs, and applies a sigmoid function. Equation (3) produces the input modulation factor for the cell state update in Equation (5).

$$\overline{C_t} = \mathrm{r}(W_c . [h_{t-1}, x_t] + b_c) \tag{4}$$

$\overline{C_t}$ represents the candidate cell state**,** a new content vector proposed for addition to the cell state. The function $r(x)$ is a non-linear activation function, commonly ReLU $r(x) = \max(0, x)$ in the range [0,∞] and Sigmoid with the formula $r(x) = \frac{1}{1+e^{-x}}$ in the range [0,1] [19]. Equation (4) computes potential new information to be integrated into the cell state.

$$C_t = f_t . C_{t-1} + i_t . \overline{C_t} \tag{5}$$

C_t is the updated cell state, calculated by combining the scaled previous state f_t . $C_{t-1}$with the scaled new candidate state $i_t . \overline{C_t}$. Equation (5) is the core of the LSTM memory mechanism, determining what to forget and what to add to the cell memory.

$$o_t = \sigma(W_o . [h_{t-1}, x_t] + b_o) \tag{6}$$

The bias value b_c in the cell state, and o_t is the output gate result with the bias value b_o. The final output of the hidden forward layer or LSTM process is at $\overrightarrow{h_t}$. o_t controls how much the cell state should influence the output hidden state. Like other gates, it is based on h_{t-1}, x_t and passes through a sigmoid activation. Equation (6) determines the visibility of the internal memory (cell state) to the next layer or time step.

$$\overrightarrow{h_t} = o_t . \mathrm{r}(C_t) \tag{7}$$

$\overrightarrow{h_t}$ the forward hidden state output at time t, obtained by applying an activation function r (e.g., ReLU) to the current cell state C_t and scaling it by the output gate o_t. Equation (7) produces the forward-direction output of the LSTM layer at time t.

$$\overleftarrow{h_t} = o_t . \sigma(C_t) \tag{8}$$

In equation (8), $\overleftarrow{h_t}$ is the final output value of the hidden backward layer. $\overleftarrow{h_t}$ is the backward hidden state output at time t, where the cell state C_t is passed through a sigmoid activation and modulated by the output gate o_t. Equation (8) is used for the reverse-time computation in BiLSTM, capturing future context.

$$y_t = U_y\overrightarrow{h_t} + W_y\overleftarrow{h_t} + b_y \quad (9)$$

The prediction result of the BiLSTM output gate is described at y_t, the weight values U_y and W_y are at the output gate $\overrightarrow{h_t}$ and $\overleftarrow{h_t}$. y_t is the final BiLSTM output at time t, computed by combining the forward hidden state $\overrightarrow{h_t}$ and backward hidden state $\overleftarrow{h_t}$ through weight matrices U_y, W_y and bias b_y. Equation (9) integrates information from both directions of the sequence to produce the prediction or representation for time step t.

Initially, the BiLSTM model is trained with the input training data. To ensure the effectiveness of the model that has been built in the training stage, model testing is carried out using test data [20]. The parameters tested in the development of the BiLSTM model are shown in Table 1.

Table 1. Parameters of Testing

Architecture	Values
Data Splitting	70:30, 80:20, 90:10
Number of Neurons	25, 50, 100
Batch Size	4, 16, 32
Activation Function	ReLu, Sigmoid

Table 1 presents the parameters used in the testing phase of the BiLSTM model development. The data splitting ratios (70:30, 80:20, and 90:10) indicate the proportion of training and testing data, which is essential for evaluating the model's generalization capability. The number of neurons (25, 50, and 100) represents the complexity of the model in learning patterns from the data. The batch size (4, 16, and 32) determines the number of samples processed before updating the model weights, influencing the training stability and convergence speed. Two activation functions—ReLU and Sigmoid—are also utilized to study their impact on model performance and optimization. These activation functions are applied to the cell state candidate ($\overline{C_t}$), which lies within the input gate pathway of the LSTM architecture, and directly influences how new information is integrated into the memory cell.

Evaluation of prediction results with MAPE

Model evaluation aims to determine a forecasting method's error rate or to determine its level of accuracy [21]. Model testing has many types, while this research utilizes a model testing, i.e., Mean Absolute Percentage Error (MAPE). MAPE is an error measurement that calculates the percentage deviation between actual data and forecast data, with the smallest MAPE value assumed to be the best method [22].

The formula for calculating MAPE is as in equation (10).

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \times 100\% \quad (10)$$

Information:

y_i = Actual data

i = sequence-i data

$\hat{y}_i$ = Prediction data

n = Number of data

Interpretation of MAPE values as in Table 2:

Table 2. Accuracy of MAPE Results

MAPE Values	Prediction Accuracy
MAPE $\leq$ 10%	Very Good
10% < MAPE $\leq$ 20%	Good
20% < MAPE $\leq$ 50%	Fair
MAPE > 50%	Poor

Data Denormalization

After the BiLSTM model performs the prediction process, the output results need to be returned to their original scale through a denormalization process so that they can be compared directly with the actual data [15]. The denormalization formula is:

$$x = x_{normalized}.(x_{max} - x_{min}) + x_{min}$$

Where x is the value on the original scale, $x_{normalized}$ is the normalized prediction result, x_{min} and x_{max} are the minimum and maximum values of the original data. With this denormalization, the prediction results become more meaningful and can be evaluated accurately against the actual value.

Results and Discussion

Data Collecting

This research uses world crude oil price data, with as many as 268 data points, for 1 year from September 20, 2022, to September 20, 2023. Data on Saturdays and Sundays is empty because they are holidays, so they are ignored. The data sample is shown in Table 3. Day 0 is the day the data starts to be taken and day 268 is the day the last data is taken.

Table 3. Preprocessing Data

Day	Price ($)
268	90.40
364	90.48
363	90.58
⋮	⋮
2	83.49
1	82.94
0	84.45

Data Normalization

Table 4: Data Normalization

Day	Price ($)
365	0.91457
364	0.91766
363	0.92153
⋮	⋮
2	0.64746
1	0.62620
0	0.68457

Table 4 shows the normalization results. The minmax normalization formula aims to reduce the data range by converting the price values into a range between 0 and 1. This aims to help the BiLSTM model converge faster during the training process and avoid numerical problems that may arise when the data is extensive.

Formation of Time Series Pattern

Table 5. Formation of Prediction Data Series

Input	Output
0.9145	0.9176
0.9176	0.9215
0.9215	0.9288
⋮	⋮
0.4638	0.6474
0.6474	0.6262
0.6262	0.6845

Table 5 is the result of time series pattern formation. The pattern is formed in the order {X_t, X_{t+1}} where X is used as input data while X_{t+1} is the prediction output or target.

BiLSTM model building

Implementing the BiLSTM model in this study refers to the mathematical stages described through Equations (2) to (9). To ensure this model is theoretically described and implemented, the following shows the manual calculation process at the first three time steps as a real illustration of applying the equations in forming model predictions.

In the process of forming the BiLSTM model, the first step begins with the calculation of Forget Gate (f_t) using Equation 2. Forget gate determines how much information from the previous state cell (C_{t-1}) will be stored in the calculation process. The higher the f_t value, the more information is stored. Randomly initialize the weight value $\boldsymbol{W_f}$=[0.5,0.3], bias b_f=0.2 at the forget gate. Initial value of h_{t-1} is usually set to zero (0) because there is no information from the previous step. In the next iteration, the value of h_{t-1} is obtained from the calculation of the previous hidden state.

Table 6. Forget Gate Calculation

t	$\boldsymbol{X_t}$	$\boldsymbol{h_{t-1}}$	$\boldsymbol{W_f}.[\boldsymbol{h_{t-1}}, \boldsymbol{x_t}]$	$\mathbf{f_t} = \boldsymbol{\sigma}(\mathbf{W_f}.[\mathbf{h_{t-1}}, \mathbf{x_t}] + \mathbf{b_f})$
1	0.9145	0	(0.5 x 0) + (0.3 x 0.9145) = 0.2743	σ (0.2743 + 0.2) = 0.6163
2	0.9176	0.12	(0.5 x 0.12) + (0.3 x 0.9176) = 0.3313	σ (0.3313 + 0.2) = 0.6295
3	0.9215	0.195	(0.5 x 0.195) + (0.3 x 0.9125) = 0.3954	σ (0.5954 + 0.2) = 0.6443

Table 6 shows that the value of f_t increases over time, indicating that the model retains more information from the previous cell state as the iteration increases. Forget gate gives a value between 0 and 1, indicating how much information from the last state cell should be retained. Forget gate tends to have a value of 0.6, meaning that about 60% of the information from the previous iteration is retained. The following calculation is performed from the input gate (i_t), candidate cell state ($\overline{C_t}$), cell state (C_t), output gate (o_t), hidden state forward ($\overrightarrow{h_t}$), hidden state backward ($\overleftarrow{h_t}$), until final output (y_t). Table 7 presents the results of the manual calculation for each component. The weight and bias values used are randomly initialized for illustration purposes,

and sigmoid and ReLU activation functions are used as per the functions in the equation. The weight and bias values used in the manual calculation are summarized in Table 8.

In general, at the first time step (t = 1), the input gate (i_t = 0.5825) indicates that 58.25% of the new information from the candidate cell state will be integrated into the cell state, and the output gate (o_t = 0.5925) indicates that about 59.25% of the cell state will affect the hidden state. A similar trend is seen at time steps t = 2 and t = 3, where the percentage of retained and integrated values is 58-64%, indicating that the model maintains a balance between retaining previous memory and incorporating new information at each time step.

Table 7. The Results of the Manual Calculation

Time (t)	f_t	i_t	$\overline{C_t}$	C_t	o_t	$\overrightarrow{h_t}$	$\overleftarrow{h_t}$	y_t
1	0,6163	0,5825	0,4658	0,2714	0,5925	0,1608	0,3362	0,4479
2	0,6295	0,5909	0,5390	0,4894	0,6071	0,2971	0,3763	0,5714
3	0,6443	0,6019	0,5856	0,6519	0,6157	0,4014	0,4048	0,6643
...	...	...	...	...	...	...	...	...

Table 8. The Weight and Bias Values Used in the Manual Calculation

Component	Weight	Bias	Activation Function
Forget gate (f_t)	[0.5; 0.3]	0.2	Sigmoid
Input gate (i_t)	[0.4; 0.2]	0.15	Sigmoid
Candidate cell state ($\overline{C_t}$)	[0.6; 0.4]	0.1	ReLU
Output gate (o_t)	[0.5; 0.3]	0.1	Sigmoid
Output layer (y_t)	$U_y =$ W_y= 0.7, b_y = 0.1	0.1	ReLU dan Sigmoid

After this manual calculation, the following process becomes more complex as it has to perform iterations for each data point in the time series. Therefore, the next stage is done automatically with the help of Python software using the TensorFlow and Keras libraries. In the training stage, the BiLSTM model is trained using training data that has been normalized and formatted into time series patterns according to the BiLSTM input structure. During this process, backpropagation through time is performed, where the weights and biases in each layer are gradually updated using optimization algorithms, that is Adam, to minimize the loss function value.

The next step after the model has been trained is a testing phase using test data previously separated from the training data. The prediction process is carried out on the test data, and the prediction results obtained are then denormalized so that the values can be returned to the original scale. This phase is necessary to compare the prediction findings directly with the real data. Lastly, the accuracy of the generated predictions is assessed by utilizing the Mean Absolute Percentage Error (MAPE) measure to evaluate the model's performance.

Model Evaluation

Table 9: The Results of Model Testing

Data splitting	Number of neurons	Batch Size	Activation Function	MAPE (%)
70:30	25	4	ReLu	0.56
			Sigmoid	0.49
		16	ReLu	1
			Sigmoid	3.77

Data splitting	Number of neurons	Batch Size	Activation Function	MAPE (%)
		32	ReLu	2.79
			Sigmoid	3.9
	50	4	ReLu	0.34
			Sigmoid	0.93
		16	ReLu	0.56
			Sigmoid	0.95
		32	ReLu	2.29
			Sigmoid	3.45
	100	4	ReLu	0.29
			Sigmoid	0.31
		16	ReLu	0.36
			Sigmoid	0.68
		32	ReLu	1.89
			Sigmoid	4.43
80:20	25	4	ReLu	0.37
			Sigmoid	0.41
		16	ReLu	1.12
			Sigmoid	4.81
		32	ReLu	3.93
			Sigmoid	4.79
	50	4	ReLu	0.3
			Sigmoid	0.45
		16	ReLu	0.42
			Sigmoid	4.74
		32	ReLu	2.35
			Sigmoid	3.67
	100	4	ReLu	0.18[c]
			Sigmoid	1.11
		16	ReLu	0.31
			Sigmoid	2.78
		32	ReLu	1.07
			Sigmoid	2.09
90:10	25	4	ReLu	0.48
			Sigmoid	0.3
		16	ReLu	0.52
			Sigmoid	1.76
		32	ReLu	1.59
			Sigmoid	6.08
	50	4	ReLu	0.18[b]
			Sigmoid	0.69
		16	ReLu	0.37
			Sigmoid	1.88

Data splitting	Number of neurons	Batch Size	Activation Function	MAPE (%)
	100	32	ReLu	1.87
			Sigmoid	5.94
		4	ReLu*	**0.09**[d]
			Sigmoid	0.19[a]
		16	ReLu	0.32
			Sigmoid	1.51
		32	ReLu	1.11
			Sigmoid	2.17

The results of testing the model with different data splitting, neuron count, batch size, and hidden layer activation function are shown in Table 9. The 70:30, 80:20, and 90:10 data divides are utilized. The findings indicate that the activation function, batch size, and neuron count all significantly impact the model's performance. The model with 25 neurons using Sigmoid activation generated a lower error rate than ReLu activation at a 70:30 data split. Nonetheless, a setup like 100 neurons with ReLu activation at a 90:10 data split can deliver outstanding performance with a MAPE of 0.09 and is the best model result in these 57 tests. Therefore, optimal parameter selection and model sensitivity to changes in data split are crucial to achieving accurate and reliable model predictions.

The Sigmoid function generally produces higher MAPE than the ReLU activation function, particularly when bigger neurons are used in parameter combinations. This implies that ReLU rather than Sigmoid would work better in this situation. Furthermore, models with a small batch size (e.g., 4) and more neurons (e.g., 100 neurons) appear to produce improved prediction accuracy. This suggests that the model's performance can be enhanced and its prediction accuracy increased by employing lower batch sizes and additional neurons.

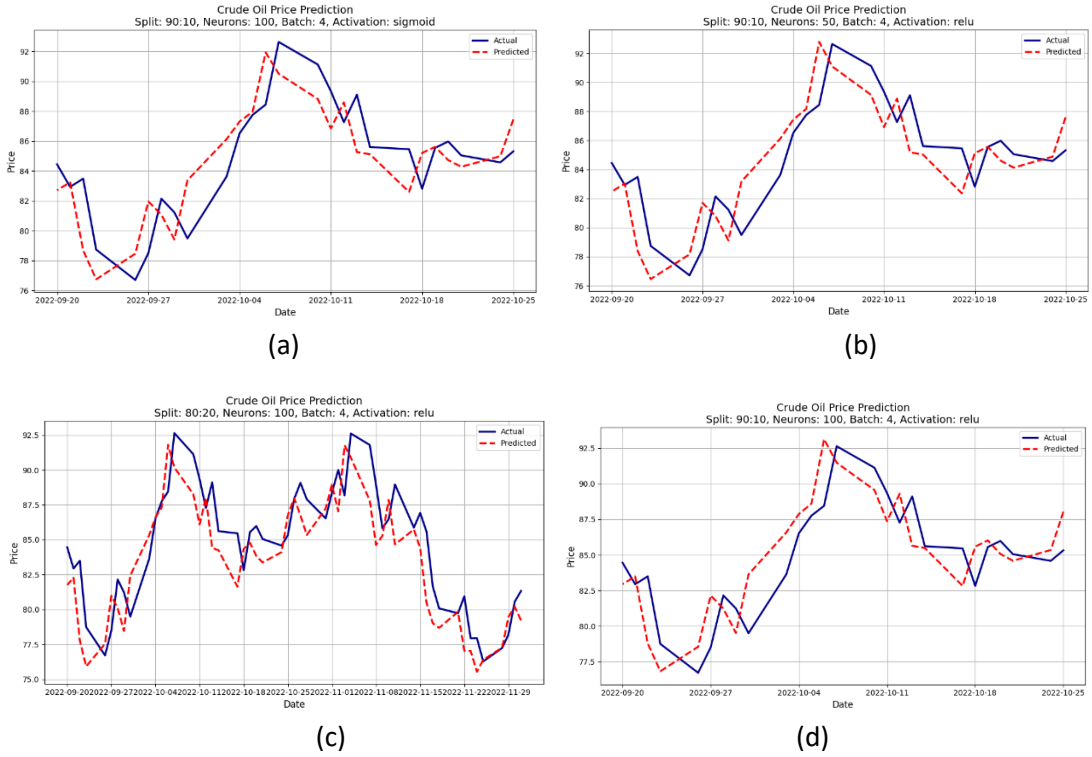

Figure 2. Plot of testing results with the four best models (a) Split: 90:10, Neurons: 100, Batch size: 4, Activation: sigmoid (highest MAPE), (b) Split: 90:10, Neurons: 50, Batch

size: 4, Activation: relu, (c) Split: 80:20, Neurons: 100, Batch size: 4, Activation: relu, (d) Split: 90:10, Neurons: 100, Batch size: 4, Activation: relu (lowest MAPE)

The four models with the lowest MAPE value are shown in Figure 2, where the order from (a) to (d) is from the largest to smallest MAPE. In general, all the models can track fluctuating trends of actual prices well. The differences between the four models' forecasts are not graphically sensational, yet it is observable that the model in sub-figure (d)-with 90:10 split configuration data, 100 neurons, batch size 4, and ReLU activation function-forecasts the greatest similarity with the actual price trend.

As shown in Table 10, the comparison of MAPE values in other studies is used to evaluate the advantages and disadvantages of the model proposed in this study and to see the extent of the improvement in accuracy achieved compared to previous methods. The studies used for comparison all focus on predicting world crude oil prices, so they are relevant in assessing the performance of the BiLSTM model developed in this study.

Table 10: Comparison with Other Research

Year	Method	MAPE (%)
2018-2023	SVR [6]	49.73
1987-2017	Combination ARIMA [24]	0.2597
2001-2021	ARIMA-Artificial Neural Network (ANN) [25]	13.95
2009-2022	Naive Bayes [26]	0.9463
Our Research	BiLSTM	0.09

Based on Table 7, the BiLSTM method in this study shows the best performance with a MAPE value of 0.09, much smaller than other methods such as SVR (0.4973), ARIMA (0.2597), and Naïve Bayes (0.9463). This shows that BiLSTM is superior in capturing complex historical crude oil price data patterns. Further, the ARIMA-ANN combined methodology yielded a MAPE value of 13.95 due to its inadequacy in coping with more complex non-linear trends. Subsequent studies of oil price forecasting may also explore hybrid frameworks combining BiLSTM with statistical methods such as ARIMA or other deep networks to improve forecasting performance, as implied by prior investigations. Also, suppose the external variables, like OPEC production, world oil inventories, and macroeconomic variables like inflation and interest rates, are included in the model. In that case, it can better capture the direction of crude oil prices.

Conclusion

Using 57 tests, the present study forecasts crude oil prices from September 2022 to September 2023. The best MAPE of 0.09% is obtained using BiLSTM with a 90:10 data split, ReLU activation, 10 neurons, and batch size 4. These findings confirm that BiLSTM is a precise predictor of oil prices and can be a trusted source of knowledge for investors and the energy sector. However, this study has limitations because it has not considered external variables such as OPEC production levels, global oil reserves, and macroeconomic factors, which could impact oil prices. Therefore, future research can combine external variables or use a hybrid approach with other methods to improve prediction accuracy. The contribution of this research to society is to provide insight into the utilization of artificial intelligence in crude oil price analysis, which can be a reference in economic and business decision-making.

Acknowledgments

The author would like to thank state-owned enterprises (Indonesian: Badan Usaha Milik Negara/BUMN) for providing the opportunity for a research scholarship so that this research can be completed properly.

References

[1] V. G. Yadav, G. D. Yadav, and S. C. Patankar, "The Production of Fuels and Chemicals in The New World: Critical Analysis of The Choice Between Crude Oil And Biomass Vis-À-Vis Sustainability and The Environment," *Clean Technol Environ Policy*, vol. 22, no. 9, pp. 1757–1774, 2020, doi: 10.1007/s10098-020-01945-5.

[2] C. W. Su, S. W. Huang, M. Qin, and M. Umar, "Does Crude Oil Price Stimulate Economic Policy Uncertainty in BRICS?," *Pacific Basin Finance Journal*, vol. 66, no. September 2020, p. 101519, 2021, doi: 10.1016/j.pacfin.2021.101519.

[3] Z. Jia, S. Wen, and B. Lin, "The Effects and Reacts of COVID-19 Pandemic and International Oil Price on Energy, Economy, and Environment in China," *Appl Energy*, vol. 302, no. August, p. 117612, 2021, doi: 10.1016/j.apenergy.2021.117612.

[4] S. Karasu, A. Altan, S. Bekiros, and W. Ahmad, "A New Forecasting Model with Wrapper-Based Feature Selection Approach Using Multi-Objective Optimization Technique for Chaotic Crude Oil Time Series," *Energy*, vol. 212, p. 118750, 2020, doi: 10.1016/j.energy.2020.118750.

[5] A. Pourdaryaei *et al.*, "Recent Development in Electricity Price Forecasting Based on Computational Intelligence Techniques in Deregulated Power Market," *Energies (Basel)*, vol. 14, no. 19, pp. 1–28, 2021, doi: 10.3390/en14196104.

[6] D. Suryani and M. Fadhilla, "Indonesian Crude Oil Price (ICP) Prediction Using Support Vector Regression Algorithm Des," *Jurnal Resti*, vol. 5, no. 158, pp. 127–134, 2024, doi: https://doi.org/10.29207/resti.v8i1.5551.

[7] A. Mahmoud and A. Mohammed, "Leveraging Hybrid Deep Learning Models for Enhanced Multivariate Time Series Forecasting," *Neural Process Lett*, vol. 56, no. 5, pp. 1–25, 2024, doi: 10.1007/s11063-024-11656-3.

[8] J. Wu, Z. Wang, Y. Hu, S. Tao, and J. Dong, "Runoff Forecasting using Convolutional Neural Networks and optimized Bi-directional Long Short-term Memory," *Water Resources Management*, vol. 37, no. 2, pp. 937–953, 2023, doi: 10.1007/s11269-022-03414-8.

[9] Z. Hameed and B. Garcia-Zapirain, "Sentiment Classification Using a Single-Layered BiLSTM Model," *IEEE Access*, vol. 8, pp. 73992–74001, 2020, doi: 10.1109/ACCESS.2020.2988550.

[10] M. Yang and J. Wang, "Adaptability of Financial Time Series Prediction Based on BiLSTM," *Procedia Comput Sci*, vol. 199, pp. 18–25, 2021, doi: 10.1016/j.procs.2022.01.003.

[11] N. Raj, "Prediction of Sea Level with Vertical Land Movement Correction Using Deep Learning," *Mathematics*, vol. 10, no. 23, 2022, doi: 10.3390/math10234533.

[12] F. Shahid, A. Zameer, and M. Muneeb, "Predictions for COVID-19 with Deep Learning Models of LSTM, GRU and Bi-LSTM," *Chaos Solitons Fractals*, vol. 140, p. 110212, 2020, doi: 10.1016/j.chaos.2020.110212.

[13] Investing, "Crude Oil Historical Data," investing.com. Accessed: Oct. 02, 2023. [Online]. Available: https://www.investing.com/commodities/crude-oil-historical-data

[14] S. Sivamohan, S. S. Sridhar, and K. S., "An Effective Recurrent Neural Network (RNN) based Intrusion Detection via Bi-directional Long Short-Term Memory," *IEEE Access*, 2021.

[15] Q. T. Bui, Q. H. Nguyen, X. L. Nguyen, V. D. Pham, H. D. Nguyen, and V. M. Pham, "Verification of Novel Integrations of Swarm Intelligence Algorithms Into Deep Learning Neural Network for Flood Susceptibility Mapping," *J Hydrol (Amst)*, vol. 581, p. 12479, 2020, doi: 10.1016/j.jhydrol.2019.124379.

[16] A. H. Elsheikh *et al.*, "Deep Learning-Based Forecasting Model for COVID-19 Outbreak in Saudi Arabia," *Process Safety and Environmental Protection*, vol. 149, pp. 223–233, 2021, doi: 10.1016/j.psep.2020.10.048.

[17] C. Magazzino, M. Mele, and N. Schneider, "A Machine Learning Approach on The Relationship Among Solar and Wind Energy Production, Coal Consumption, GDP, and CO2 Emissions," *Renew Energy*, vol. 167, pp. 99–115, 2021, doi: 10.1016/j.renene.2020.11.050.

[18] P. Singla, M. Duhan, and S. Saroha, "An Ensemble Method to Forecast 24-H Ahead Solar Irradiance Using Wavelet Decomposition and BiLSTM Deep Learning Network," *Earth Sci Inform*, vol. 15, no. 1, pp. 291–306, 2022, doi: 10.1007/s12145-021-00723-1.

[19] D. Kent and F. Salem, "Performance of Three Slim Variants of the Long Short-Term Memory (LSTM) Layer," *Midwest Symposium on Circuits and Systems*, vol. August, no. 7, pp. 307–310, 2019, doi: 10.1109/MWSCAS.2019.8885035.

[20] M. Rhanoui, M. Mikram, S. Yousfi, and S. Barzali, "A CNN-BiLSTM Model for Document-Level Sentiment Analysis," *Mach Learn Knowl Extr*, vol. 1, no. 3, pp. 832–847, 2019, doi: 10.3390/make1030048.

[21] M. Piekutowska *et al.*, "The Application of Multiple Linear Regression and Artificial Neural Network Models for Yield Prediction of Very Early Potato Cultivars before Harvest Magdalena," *Agronomy*, vol. 11, no. 5, pp. 1–17, 2021, doi: 10.3390/agronomy11050885.

[22] R. Ahmed, V. Sreeram, Y. Mishra, and M. D. Arif, "A Review and Evaluation of The State-of-The-Art in PV Solar Power Forecasting: Techniques And Optimization," *Renewable and Sustainable Energy Reviews*, vol. 124, no. June 2019, p. 109792, 2020, doi: 10.1016/j.rser.2020.109792.

[23] Al-Khowarizmi, R. Syah, M. K. M. Nasution, and M. Elveny, "Sensitivity of MAPE using Detection Rate for Big Data Forecasting Crude Palm Oil on K-Nearest Neighbor," *International Journal of Electrical and Computer Engineering*, vol. 11, no. 3, pp. 2696–2703, 2021, doi: 10.11591/ijece.v11i3.pp2696-2703.

[24] R. Bollapragada, A. Mankude, and V. Udayabhanu, "Forecasting the Price of Crude Oil," *Decision*, vol. 48, no. 2, pp. 207–231, 2021, doi: 10.1007/s40622-021-00279-5.

[25] H. Alrweili and H. Fawzy, "Forecasting Crude Oil Prices Using an ARIMA-ANN Hybrid Model," *J Stat Appl Probab*, vol. 11, no. 3, pp. 845–855, 2022, doi: 10.18576/jsap/110308.

[26] T. P. Ogundunmade, A. A. Adepoju, and A. Allam, "Predicting Crude Oil Price in Nigeria with Machine Learning Models," *Modern Economy and Management*, vol. 1, no. 4, pp. 1–8, 2022, doi: 10.53964/mem.2022004.