Crude Oil Price Forecasting Using Long Short-Term Memory and Support Vector Regression

Rifdah Amelia[1], Ahmad Zuhdi[2], Abdul Rochman[3]
[1, 2, 3] Informatics Engineering Department, Trisakti University, Jakarta, Indonesia
[1]rifdah064001900019@std.trisakti.ac.id, [2]zuhdi@trisakti.ac.id, [3]abdul.rochman@trisakti.ac.id

Accepted 02 November 2022
Approved 06 December 2022

***Abstract*— Crude oil or petroleum is a non-renewable energy source derived from organic materials whose formation process is lengthy. Crude oil is a commodity whose price often fluctuates. When there is a fluctuation, a nation's economy will be affected. The crude oil price datasets are categorized as non-linear. This research used two models to compare the performance of those two models to find the best model to predict Brent crude oil prices. The models used in this research are Long Short-Term Memory (LSTM) and Support Vector Regression (SVR). Those two methods are widely used for similar cases, such as forecasting the stock price. The dataset used in this study is the price of Brent crude oil from May 1987 to May 2022. The result of this study indicates that the deep learning algorithm, LSTM, performs better in forecasting the price of Brent crude oil, with a root mean squared error value of 1.543.**

***Index Terms*— Crude Oil; Deep Learning; Forecasting; Long Short-Term Memory (LSTM); Support Vector Regression (SVR).**

I. INTRODUCTION

Crude oil is a non-renewable energy source. Crude oil is categorized as a non-renewable energy source because the formation process takes quite a long time [1]. Crude oil reached through the fractionation process will provide various final products, such as fuel, liquefied gas, diesel, and others [1].

Indonesia is a country rich in natural resources. As a country rich in natural resources, Indonesia is among the countries capable of producing crude oil. In the early 2022 period, Indonesia's crude oil consumption level fluctuated substantially. However, it tends to increase when the level of crude oil consumption in Indonesia reaches 1,710.52 thousand barrels per day [2].

From December 2021 to January 2022, there has been a decline in the level of crude oil production. Indonesia can only produce 651.68 thousand barrels per day, while in January 2022, Indonesia only produced 616.06 thousand barrels of crude oil per day [3].

The data presented in the previous paragraph shows that there has been a plunge in the level of crude oil production in Indonesia. On the contrary, an increase in crude oil consumption impacts the government because the government needs to import petroleum from other countries.

Petroleum is a mining product that has sensitive properties [4]. So, when fluctuations emerge, a country's economy (both micro and macro) can be affected. Consequently, we must pay close attention to crude oil prices to predict future prices.

The changes in oil prices can be recorded from time to time and aggregated to form a time series data. The time series data for crude oil prices are non-stationary because their average value, variance, and covariance change over time [5].

The comparison of the performance of the LSTM and SVR models in the case of stock prices of several companies, such as the NYSE, NSE, BSE, NASDAQ, and others, we have learned that the LSTM model provides the best forecasting results, with a MAPE value of 0.86. In comparison, the SVR model gives a MAPE error value of 1.44 [5].

In a similar case, the other research tries to forecast the closing price of the iShares MSCI United Kingdom fund, which is a non-linear and non-stationary dataset. The results of this study reveal that the LSTM model is the best-performing model. The MAE value given from this research is 0.210350. Then the SVR model followed, which has an MAE value of 0.24002, the random forest method, and the ANN method [6].

The Long Short-Term Memory (LSTM) model is an expansion model of the deep learning model, namely the Recurrent Neural Network (RNN) [7]. Hochrieter and Schmidhuber introduced this model in 1997. The RNN model has a long-term dependency, so the LSTM model builds to finish the problem that happens in the RNN model [7].

The Support Vector Regression (SVR) model is a model of the Support Vector Machine used to resolve regression cases [8]. The SVR model is used to solve regression cases by applying machine learning theory to maximize the level of prediction accuracy while automatically avoiding over-fitting to data [8].

This research aims to compare the level of accuracy generated by the deep learning model, videlicet LSTM, and the machine learning model, SVR, in forecasting the price of crude oil, which is a non-linear time series data.

II. THEORY

A. *Long Short-Term Memory*

There is a problem that the Recurrent Neural Network (RNN) model needs to face, that is, a long-term dependency problem. So, to deal with the issues that ensued in the precursory model (RNN), the Long Short-Term Memory model was developed in 1997 [7]. In Figure 1, we can see the architecture of the LSTM model. The LSTM model has three types of gates: forget gates, input gates, and output gates.

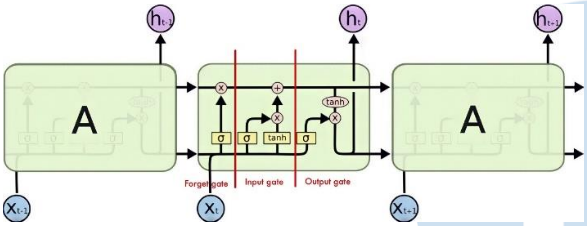

Fig. 1. The architecture of a cell in the LSTM unit

The charge of the forget gate component is to determine which data or information should be scraped or stored from the cell state [7]. From equation 1, we got the formulation of the forget gate.

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

Explanation :

- f_t = forget gate
- σ = sigmoid activation function
- W_f = forget gate weight
- h_{t-1} = previous layer's output
- x_t = new input vector at t time
- b_f = forget gate bias

We used the input gate to determine which new data would be stored in the cell state and determine the gate to write memory [7], [9]. To calculate the input gate, we can use equation (2), and to calculate the new candidate value to be added to the state, we use equation (3).

$$i_t = \sigma\left(W_f \times [h_{t-1}, x_t] + b_f\right) \quad (2)$$

$$\tilde{c}_t = tanh\left(\sigma\left(W_c \times [h_{t-1}, x_t] + b_c\right)\right) \quad (3)$$

Explanation :

- i_t = input gate
- $\tilde{c}_t$ = new candidate values to be added to the state
- σ = sigmoid activation function
- tanh = tanh activation function
- W_f = forget gate weight
- W_c = cell state weight
- h_{t-1} = previous layer's output
- x_t = new input vector at t time
- b_f = forget gate bias
- b_c = cell state bias

Aside from various gates, an LSTM unit has a cell state. The task of the cell state is to remove information about the subject from the past and add the information formulated in equation (4) [7]. The operations performed on the cell state can be seen in Figure 2.

$$c_t = f_t \times c_{t-1} + \tilde{c}_t \times i_t \quad (4)$$

Explanation :

- c_t = cell state
- f_t = forget gate
- c_{t-1} = cell state value at time t - 1
- $\tilde{c}_t$ = new candidate value
- i_t = input gate

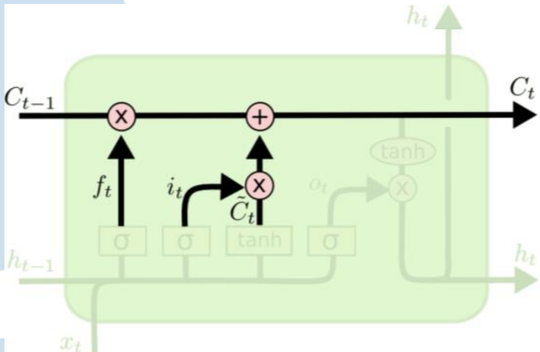

Fig. 2. LSTM cell state illustration [7]

The output gate is a gate that will decide what information should be issued and read from memory. We can use equation (5) to calculate the output gate, and we can use equation (6) to calculate the hidden state value [7].

$$o_t = \sigma\left(W_f \times [h_{t-1}, x_t] + b_f\right) \quad (5)$$

$$h_t = o_t \times tanh\,(c_t) \quad (6)$$

Explanation :

- h_t = hidden state
- o_t = output gate
- σ = sigmoid activation function
- W_f = forget gate weight
- h_{t-1} = previous layer's output
- x_t = new input vector at t time
- b_f = forget gate vias
- c_t = cell state

B. *Support Vector Regression*

Support Vector Regression (SVR) is one of the machine learning models, scilicet Support Vector Machine (SVM), which is applied to solve regression problems. SVM is a model used to examine linear predictors in a high-dimensional feature space where the feature space's high dimensions can increase the sample's complexity and challenge the computational complexity [10].

This model was raised by Boser, Guyon, and Vapnik in 1992 [8]. SVR can be used to solve the case of the prediction of a value or price in the future [10]. The objective of the SVR model is to find a value that is close to the best within a certain margin (ε – tube). This model adjusts the error to a certain threshold. Moreover, this model also aims to insert as many data points as possible into the margins without breaking or crossing the margins. An illustration of how the SVR algorithm works and its components can be seen in Figure 3.

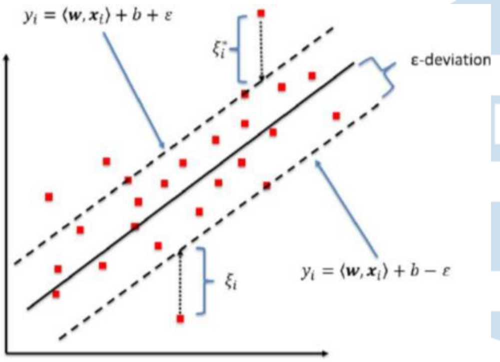

Fig. 3. Support Vector Regression component [11]

The basic concept of the SVR model is to map the x data into a feature space with high dimensions through non-linear mapping and execute linear regression in this space [12]. So, this model has the basic formula written in equation (7).

$$f(x) = \langle \omega, x \rangle + b \qquad (7)$$

Explanation :

- $f(x)$ = the argument of the function
- ω = weight
- x = x value
- b = bias

Equation (7) shows that the training data were taken as $\{(x_1, y_1), \ldots, (x_n, y_n)\} \subset \aleph \times \Re$, where $\aleph$ indicating the space of the input pattern. In equation (7), there is an operator $\langle .,. \rangle$. This operator indicates that the product dot operation ensues.

To minimize equation (7), we require the Euclidean norm using $\|\omega\|^2$ so that equation (7) turns into equation (8).

$$\text{Min} \quad \frac{1}{2}\|\omega\|^2 \qquad (8)$$

$$\text{Subject to} \quad \begin{cases} y_i - \langle \omega, x_i \rangle - b \le \varepsilon \\ \langle \omega, x_i \rangle + b - y_i \le \varepsilon \end{cases}$$

Explanation :

- y_i = dependent variable
- x_i = independent variable
- ω = weight
- b = bias
- ε = deviation or intensive loss function

We reconstruct equation (8) by introducing the slack variable (ξ_i, ξ_i^*) to solve the unfeasible optimization issue [12]. So, equation (9) is the result of the rebuild equation.

$$\text{Min} \quad \frac{1}{2}\|\omega\|^2 + C\sum_{i=1}^{l}(\xi_i + \xi_i^*) \qquad (9)$$

$$\text{Subject to} \quad \begin{cases} y_i - \langle \omega, x_i \rangle - b \le \varepsilon + \xi_i \\ \langle \omega, x_i \rangle + b - y_i \le \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \ \ge 0 \end{cases}$$

Explanation :

- C = Constant
- ξ_i = Slack variable

Linear and non-linear cases are the cases that the SVR model can handle. There are different formulas, whereas in the non-linear case, it is not explicitly given [12]. We can see the formula used for the non-linear case in equation (10) and Figure 4 to visualize the non-linear SVR case study.

$$f(x) = \sum_{i=1}^{N}(\alpha_i^* - \alpha_i)k(x_i, x) + b \qquad (10)$$

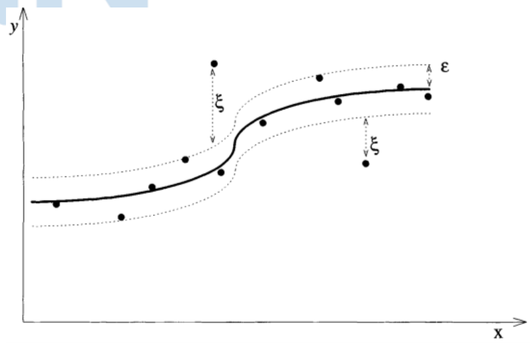

Fig. 4. SVR non-linear illustration [13]

From equation 8, we can see a kernel function $k(x_i, x)$, while α_i^* and α_i is a Lagrange multiplier and b is bias. The purpose of a kernel or kernel function is to overcome non-linear problems by involving linear classification [14].

The most common type of kernel used for the SVM case is the Radial Basis Function (RBF) kernel. RBF evolves as the most common type of kernel used because the RBF kernel is dependent on the Gaussian distribution [15]. The RBF kernel intends to acquire an estimate of similarity or determine how close two points are, i.e., points x_1 and x_2 [15]. Equation (11) formulates the RBF kernel.

$$K(x_1, x_2) = \exp\left(-\gamma \|x_1 - x_2\|^2\right) \tag{11}$$

III. Research Method

This research consists of several stages that will be carried out, including data collection, pre-processing, parameter selection, training, testing, and predicting the future for 180 days. In this research, Nasdaq provides data for Brent crude oil so that we can collect the data from the website [16]. This dataset consists of 8887 rows of data and two columns.

The next step is pre-processing the data by dividing the dataset and normalizing the dataset.

A. Pre-Processing

The dataset utilized is Brent crude oil price data for May 1987 to May 2022. Overall, the dataset consists of 8887 rows and two columns, scilicet the Date column, and the Value column.

In this research, we saved 180 rows of data to predict the future after finishing the validating process. The testing data that will be forecasted is oil price data for the period of September 2021 to May 2022. This research will use the remaining 8707 rows of data for training and testing. This research will use 80 percent of the data for training, and we will use the rest of the dataset for validation.

TABLE I. Dataset Splitting

	Training Data	Validating Data
Amount and Percentage	6965 80%	1742 20%
Period	1987-05-20 to 2014-10-31	2014-11-03 to 2021-09-06

After splitting the data, we will normalize the data. The method we use to normalize the data is min-max scaling. With the normalization of min-max scaling, the actual price data will change in intervals between 0 and 1.

B. Parameter Selection

After going through the pre-processing process, the next step is specifying the parameters used in the experiments on the two models. The parameters used in each model will be explained in sections 1) to 2).

1) *Long Short-Term Memory Model*

In this research, we will combine several parameters, i.e., the number of LSTM units, batch sizes, and epochs. Table 2 contains various parameters to be combined and tested.

TABLE II. Estimation of LSTM Model Parameters

Parameter	Amount
Batch size	[32, 64, 128]
Layer	3 (LSTM, dropout, output)
Unit LSTM	[10, 30, 50]
Epoch	[5, 10, 15, 20, 25, 30]

2) *Support Vector Regression Model*

For the SVR model experiment, we will combine the parameters that we can see in table 3. The experiment uses a combination of these parameters to obtain the best results that the model can give in forecasting the price of Brent crude oil. The kernel used in this experiment is the Radial Basis Function (RBF) kernel.

TABLE III. Estimation of SVR Model Parameters

Parameter	Amount
C	[0.001, 0.01, 0.1, 1, 10, 100]
γ	[0.001, 0.01, 0.1, 1, 10, 100]

C. Training Model

This research used 80 percent of the dataset in the training process from May 20, 1987, to October 31, 2014. For the experiment, we used the various combinations of parameters listed in Table 2 and Table 3.

1) *Long Short-Term Memory Model*

The LSTM model uses three layers: an input layer, a dropout layer, and an output layer. In the input layer, we determine the number of LSTM units, then in the next layer, there is a dropout layer with a value of 0.05, and the output layer is dense, consisting of 1 unit neuron.

In the LSTM model, the activation function is tanh, and the optimizer is Adam optimizer. The epochs and batch size used in each experiment were also specified. To train the model, we need to execute the program by calling the fit function, which contains the parameters used.

To visualize the model's performance in each experiment, we create the MSE loss function graph. Moreover, using a loss function graph, we can use the RMSE value to determine how well the model performs. Also, we assemble a plot to make it easier to compare the actual and forecasted prices by the model.

2) *Support Vector Regression Model*

For this model, we trained the model using the various combination of C and γ. To perform the training of this model, we need to call the fit function after specifying the kernel type, C value, and γ value.

We will use the RMSE loss function to evaluate how accurate the model is. We will also visualize the plot by displaying the actual and forecasted prices.

D. Validating Model

At the validating step, we need to call the predict function and use the validation set that contains 20 percent of the total data using a model that has passed the training process. The period used for data validating is November 3, 2014, to September 6, 2021. After passing the forecasting process, we will calculate the RMSE value to determine the model's performance.

E. Testing Model

After passing the validation stage, the parameter that can provide the lowest error value from both models will be used at the testing stage. We use the testing data, which contains 180 days for the period from September 2021 to May 2022.

IV. RESULTS

In this research, we compare two models: scilicet LSTM, a deep learning model, and SVR, a machine learning model. From the Brent crude oil price data acquired from the Nasdaq website, we learned that the lowest price of crude oil was on December 10, 1998, at 9.1 USD per barrel. Meanwhile, the highest price was on July 3, 2008, at 143.950 USD per barrel. Figure 5 visualizes the crude oil price from May 20, 1987, to May 23, 2022.

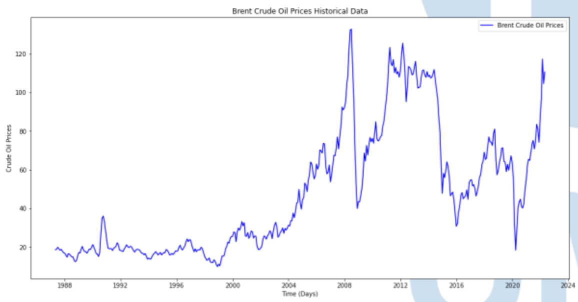

Fig. 5. Brent crude oil price chart for May 20, 1987, to May 23, 2022

A. Long Short-Term Memory Model

Using the combination of the parameters in Table 2, we conducted this research. The combination of these parameters gave 54 trials.

TABLE IV. LSTM TRAINING AND VALIDATION BEST RESULT

No	Unit LSTM	Epoch	Batch Size	RMSE	
				Train	Validation
1	10	20	64	1.968	1.543
2	30	20	64	2.046	1.588
3	50	20	64	1.948	1.742

Based on the results of the trials, we found that the numbers of LSTM units, the number of epochs, and batch sizes that give the best results are 20 and 64.

Table 4 shows that the LSTM model parameters that can give the most optimal results are the models with 10 LSTM units, 20 epochs, and a batch size of 64.

The root mean squared error value in the training process is 1.968. This RMSE value means that the model performs well at the training stage because the prediction values are close to the actual price.

To evaluate the model, aside from using the RMSE value, we can visualize the predicted value and the actual value from this model, which we can see in Figure 6. The blue line shows the actual price, while the red shows the forecast result. From Figure 6, it can be seen that the two lines have a small distance.

Fig. 6. LSTM training chart comparison of actual prices and predicted prices

In the validation process, the RMSE value was 1.543. Compared with the RMSE value at the training stage and the validation stage, we've got the RMSE value at the validation stage to have a smaller value; this means that the model's performance is getting better and has a good generalization ability. Figure 7 visualizes the results of the model validation, which show that the actual prices and forecasted prices have a small difference.

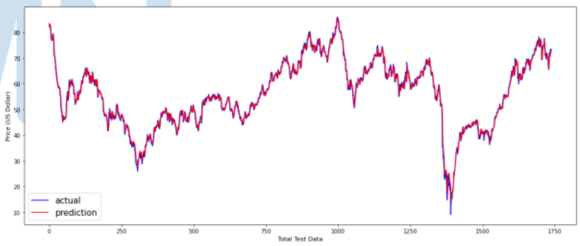

Fig. 7. LSTM validating chart comparison of actual prices and predicted results

From the validation outcomes above, the parameters that give the best results are used to forecast crude oil prices for the next 180 days. Figure 8 shows that the price for the next 180 days, visualized using a cyan-colored line, tends to fluctuate.

Figure 8 shows that the forecast results for the next 180 days have a relatively small difference from the actual price. The RMSE error value obtained is 3.628. Furthermore, we can see that the forecast price line almost covers the dark blue actual price line for the next 180 days.

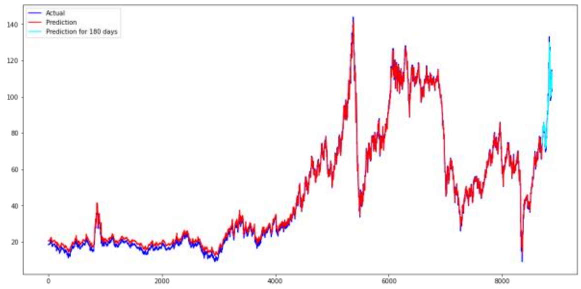

Fig. 8. Chart of LSTM model forecasting results for the next 180 days

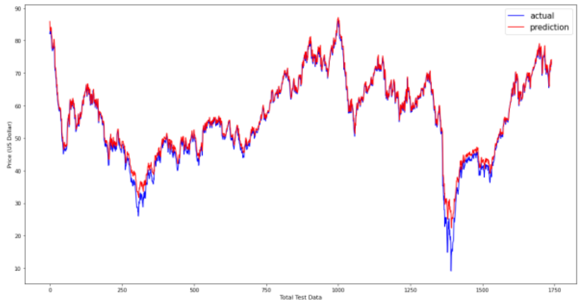

Fig. 10. SVR testing chart comparison of actual prices and predicted results

B. *Support Vector Regression Model*

We trained the Support Vector Regression model using the parameters in Table 3. From the combination of various parameters, we did 36 training processes, so we have 36 SVR models.

A model with a validation error value smaller than its training error value has a good generalization capability. So, depending on Table 5, we determine that the second experiment is the best model for forecasting the crude oil price.

After determining the best model, we use those parameters to predict the 180 days ahead. We can see the visualization of the future prediction in Figure 11, where the cyan-colored line represents the forecast value for the next 180 days.

TABLE V. SVR TRAINING AND VALIDATION BEST RESULT

No	C	Gamma	RMSE	
			Train	Validation
1	1	10	5.660	4.816
2	10	1	6.596	2.104
3	100	0.01	4.355	2.766

Table 5 displays that the model with the best performance is the model with a value of C 10 and Gamma 1. The training process indicates that the 3rd experiment's smallest root means squared error, where this model gives a 4.355 RMSE value. Furthermore, from Figure 9, we notice that the forecasting prices are close to the actual prices.

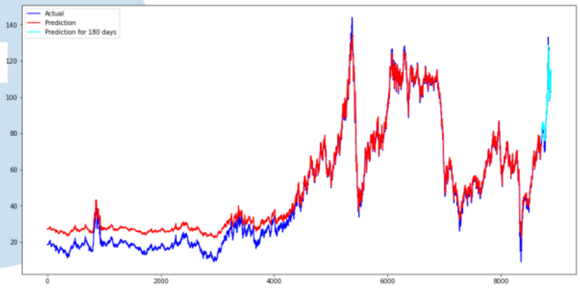

Fig. 11. Chart of SVR model forecasting results for the next 180 days

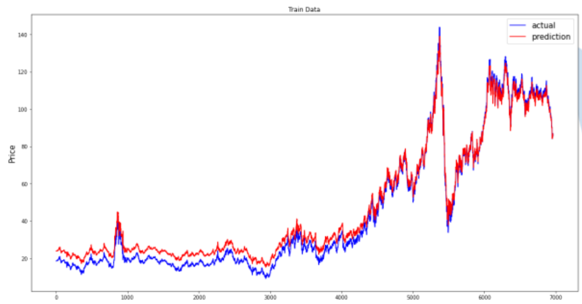

Fig. 9. SVR training chart comparison of actual prices and predicted prices

The RMSE value generated in the forecasting process for the next 180 days is 2.709. The visualization in Figure 11 shows that the price of Brent crude oil is volatile, and the actual price line and the forecasted price tend to have only a slight difference.

We got the second experiment from the training phase, which uses C 10 and Gamma 1 as the best model. So we used those parameters in the validation phase. Then, after we evaluated the model, we got the minimum error value RMSE is 2.104. Figure 10 visualizes the result of the SVR testing model.

V. CONCLUSIONS

Brent crude oil price data recorded from time to time is time series data, and in this case, this dataset can be categorized as a non-linear dataset. In this research, we compare two models from deep learning and machine learning algorithms: Long Short-Term Memory and Support Vector Regression. We use the Keras library for building the Long Short-Term memory model and the Sklearn library for building the Support Vector Regression model.

The loss function Root Mean Squared Error (RMSE) is used to evaluate the model's performance. The time series data used is Brent crude oil price data collected from May 20, 1987, to February 28, 2022.

In this research, the LSTM model's evaluation results for the training process have an RMSE value of

1,968. While in the testing process, the RMSE value is 1.543. The parameters used are models with 10 LSTM units, 20 epochs, and 64 batch sizes.

In the Support Vector Regression model, the best RMSE value is 4.355 at the training stage. In comparison, the best RMSE value that can be generate at the validation stage is 2.104, where the parameters used in the SVR model that can give the best results are C 10 and γ 1.

The results of this study indicate that the Long Short-Term Memory (LSTM) model can predict the price of Brent crude oil, which is classified as data with non-linear characteristics better than the Support Vector Regression model. Moreover, from the experimental results, it is also known that the LSTM and SVR models have relatively small error value differences.

By doing this research, we hope that this research can be used as a reference for conducting research with other case studies or trying to combine the models that have been tested with other models.

REFERENCES

[1] G. A. Olah, Molnár Árpád, and P. G. K. Surya, "Hydrocarbon chemistry" 3rd ed. United States of America. 2018, ch. 1, pp 7-19.

[2] Knoema. Indonesia petroleum consumption, 2020-2022. Available: https://knoema.com/atlas/Indonesia/topics/Energy/Oil/Petroleum-consumption#:~:text=In%20January%202022%2C%20petroleum%20consumption,per%20day%20in%20January%202022.

[3] Knoema. Indonesia production of crude oil, 2020-2022. Available: https://knoema.com/atlas/Indonesia/topics/Energy/Oil/Production-of-crude-oil#:~:text=Indonesia%20production%20of%20crude%20oil%20was%20at%20level%20of%20616.06,is%20production%20of%20crude%20oil%3F

[4] Nizar, M. A., "Dampak fluktuasi harga minyak dunia terhadap perekonomian Indonesia". Buletin Ilmiah Litbang Perdagangan, 6(2), pp 189-210, 2012.

[5] G. Bathla, "Stock price prediction using LSTM and SVR", *Sixth International Conference on Parallel, Distributed and Grid Computing (PDGC)*, 2020, https://doi.org/10.1109/pdgc50313.2020.9315800.

[6] S. Carollo, *Understanding oil prices*. Cornwall, UK: John Wiley & Sons, 2012, pp. 89-101, https://doi.org/10.1002/9781118467251.

[7] C. Olah, "Understanding LSTM Networks", 27 August 2015, [Online]. Available: http://colah.github.io/posts/2015-08-Understanding-LSTMs.

[8] Jakkula, V.R., "Tutorial on support vector machine (SVM)", [Online]. Available: https://course.ccs.neu.edu/cs5100f11/resources/jakkula.pdf.

[9] J. Firoujazee and P. Khaliliyan, "LSTM architecture for oil stocks prices prediction", *arXiv.org*, 2022. [Online]. Tersedia: https://arxiv.org/abs/2201.00350.

[10] M. Nikou, G. Mansourfar dan J. Bagherzadeh, "Stock price prediction using DEEP learning algorithm and its comparison with machine learning algorithms", *Intelligent Systems in Accounting, Finance and Management*, vol. 26, no. 4, pp. 164-174, 2019. Available: https://doi.org/10.1002/isaf.1459.

[11] T. Kleynhans, M. Montanaro, A. Gerace, and C. Kanan, "Predicting top-of-atmosphere thermal radiance using MERRA-2 atmospheric data with Deep Learning," *Remote Sensing*, vol. 9, no. 11, p. 1133, 2017.

[12] Basak, D., Pal, S., & Patranabis, D.C., "Support vector regression", [Online]. Available: https://static.aminer.org/pdf/PDF/000/337/560/uncertainty_support_vector_method_for_ordinal_regression.pdf.

[13] N. Cristianini and J. Shawe-Taylor, "An introduction to support vector machines and other kernel-based learning methods". Cambridge University Press, Cambridge, 2000.

[14] A. Gupta. 1 June 2021. "Kernel tricks in support vector machine". [Online]. Available: https://medium.com/geekculture/kernel-methods-in-support-vector-machines-bb9409342c49.

[15] S. Sreenivasa. 12 October 2020. "Radial basis function (RBF) kernel: the go-to kernel". [Online]. https://towardsdatascience.com/radial-basis-function-rbf-kernel-the-go-to-kernel-acf0d22c798a.

[16] Nasdaq. 25 May 2022. "Crude Oil Prices: Brent - Europe". [Online]. Available: https://data.nasdaq.com/data/FRED/DCOILBRENTEU-crude-oil-prices-brent-europe.