Indonesian Journal of Artificial Intelligence and Data Mining (IJAIDM)
Vol 6, No.1, March 2023, pp. 71 – 83
p-ISSN: 2614-3372 | e-ISSN: 2614-6150

Forecasting The Value of Indonesian Oil-Non-Oil and Gas Imported Using The Gated Recurrent Unit (GRU)

[1*]Dian Kurniasari, [2]Sulistian Oskavina, [3]Wamiliana, [4]Warsono
[1,2,3,4]Jurusan Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam,
Universitas Lampung, Indonesia
Email: [1]dian.kurniasari@fmipa.unila.ac.id, [2]sulistianoskavina21@email.com,
[3]wamiliana.1963@fmipa.unila.ac.id, [4]warsono.1963@fmipa.unila.ac.id

Article history:
Received Dec 23th, 2022
Revised Feb 14th, 2023
Accepted Jun 13th, 2023

Keyword:
Gas Imported
Gated Recurrent Unit
Import
MAPE
Oil-Non-Oil

ABSTRACT

In Indonesia, various factors play a role in economic development. Oil-non-oil and gas imports are one of the main factors. However, the value of oil-non-oil and gas imports in Indonesia fluctuates monthly. Therefore, an appropriate method is required to monitor changes in the value of oil-non-oil and gas imports in Indonesia so that the government can make the right choices. This study uses the GRU method to estimate the amount of oil-non-oil and gas imports in Indonesia. The best model for forecasting over the next two years has an optimum structure of 32 GRU units, 16 batch sizes, and 100 epochs, with a dropout of 0.2 and uses 80% training data and 20% test data. The MAPE value obtained is 0.999955%, with an accuracy of 99.000044%. Forecast results suggest an improvement from June 2022 to July 2024.

Corresponding Author:
Dian Kurniasari
Departement of Mathematics, Faculty of Mathematics and Natural Sciences
University of Lampung
Jl. Sumantri Brojonegoro No 1, Bandar Lampung, Indonesia.
Email: dian.kurniasari @fmipa.unila.ac.id

DOI: http://dx.doi.org/10.24014/ijaidm.v6i1.20651

1. INTRODUCTION

The act of importing is crucial in bolstering economic growth in Indonesia. It has a significant impact on the lives of the populace. Import involves the purchase of goods from foreign countries for distribution within Indonesia. However, before these goods can be circulated locally, they must be reported to the Directorate General of Customs and Excise under the Ministry of Finance [1]. Two types of imports impact Indonesia's economic growth: oil-non-oil and gas.

Nonetheless, the value of both types of imports fluctuates monthly. Hence, an appropriate method must be employed to track changes in the value of imports so that the government can implement suitable policies for future economic development. The method used to predict future scenarios is referred to as forecasting. Predictive models generally employ classical statistical models to ensure accuracy [2]. Time series forecasting is a popular technique used in weather forecasting, financial indices, energy consumption, medical monitoring, retail sales, anomaly detection, and traffic prediction. However, this method is only ideal for short-term forecasting due to several factors that affect fluctuations and cannot be entirely mitigated in time series forecasting. Additionally, time series models are susceptible to overfitting, and inaccurate prediction results may arise if outliers are not correctly handled [3].

Due to the progress made in Artificial Intelligence (AI) and machine learning (ML), Deep Learning (DL) has become the most favoured method for tackling a range of machine learning problems, including time series forecasting [4]. That is due to its exceptional and precise performance across various domains [5]. Various architectures can be employed for implementing deep learning, including Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN). RNN is the most popular deep learning architecture for predicting sequential data and time series. In contrast, CNN is primarily utilized for image-processing assignments like object recognition and semantic segmentation.

Recurrent Neural Network is a deep learning architecture that processes input repeatedly to enable the learning of information sequences [6]. Although RNN is helpful in various deep learning tasks, its short-term memory nature limits its use, making it challenging to recall previous information due to the vanishing gradient phenomenon. Therefore, we need an approach to mitigate the effect of the vanishing gradient during the training process. One of the most common ways is to change the network structure. Hochreiter-Schmidhuber has demonstrated this in his Long Short-Term Memory (LSTM) and by Cho et al. with the introduction of Gated Recurrent Units (GRU) [7].

Long Short-Term Memory and Gated Recurrent Units have a similar architecture. However, LSTM is more complex and has many parameters compared to GRU, which has more straightforward computations because it only has two gates, namely the reset gate and the update gate [8]. Despite being more straightforward, GRU can still achieve comparable accuracy and efficiently handle the issue of disappearing gradients [9]. Therefore, in this study, we will use the GRU method with time series data on the value of Indonesia's oil-non-oil and gas imports (million US$) to produce accurate predictions with more straightforward computations.

The GRU Algorithm has been applied in various research fields before and has proven effective in forecasting tasks, such as health, transportation, finance, and other fields. Lie et al. [10] used the GRU method to predict the water level of the Luoc River, Vietnam, affected by tides and achieved an accuracy rate of 94-96%, even with a small sample of data. Mateus [11] compared the performance of the LSTM and GRU models to predict the condition of the pulp paper press machine. The results show that the GRU model is generally more effective in operating with fewer data and gives better results with a broader range of parameters than the LSTM model.

Moreover, Sonali [12] suggests utilizing Auto-Regressive Integrated Moving Average (ARIMA), LSTM, and Hybrid GRU-Neural Networks to resolve the issue of IoT traffic. It is essential to address the problem of IoT traffic to augment channel capacity and diminish network latency. The study's outcomes reveal that the GRU-Neural Network provides the most precise predictions based on performance evaluation metrics such as Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and Mean Squared Error (MSE).

Ming-Che [13] integrates the GRU-Attention Mechanism model to predict trends in financial commodities on the stock market. The study focuses on Taiwan Semiconductor Manufacturing Co., Ltd., the most extensive semiconductor processing plant globally. They also compared their approach with other research methods such as K-Nearest Neighbors (KNN), Support Vector Machine (SVM), and Random Forest. The findings demonstrate that the proposed GRU-Attention Mechanism model has an average prediction accuracy of over 70%. Compared to other models, the proposed technique significantly improves prediction accuracy.

Additionally, GRU has started to be employed in digital domains such as sentiment analysis. Nowadays, many individuals spend their time on social networking platforms like Twitter. Every day, numerous tweets with a diverse structures are posted. In this scenario, semantic analysis alone cannot classify and comprehend sentences. Consequently, Aditya [14] proposes a modified GRU method for classifying and analyzing sentiment on tweets. At 35 epochs, their trials achieved the highest precision of 97.7% by categorizing sentences as either positive or negative. Additionally, they juxtaposed the accuracy of the suggested model with that of LSTM and Bidirectional-LSTM models. The results of the experiments demonstrate that the altered GRU model outperforms both LSTM and Bidirectional-LSTM models.

2. RESEARCH METHOD

2.1 Literature Review

2.1.1 Data Mining

Discovering fascinating patterns and insights from vast data sets is the essence of data mining. The data could be derived from databases, data reservoirs, online platforms, various repositories, or dynamically generated data. The process of data mining is an essential component of knowledge discovery from data (KDD), which employs intelligent methodologies to detect pertinent data patterns [15].

In the data mining process, the first step is to sort large data sets, identify patterns and create relationships to analyze the data and solve problems. Some of the tasks in data mining are as follows [16]:

1. Association: seeking a connection between one occurrence and another.
2. Pathway analysis: searching for trends where a particular occurrence results in another.
3. Classification: grouping a label into specific groups according to their similarities.
4. Clustering: identifying information groups that were previously unknown.
5. Forecasting: finding patterns in data that can be used to predict the future.

2.1.2 Deep Learning

Deep Learning (DL) is a knowledge-based approach that employs intricate neural networks and computation [17]. Three terms are often used interchangeably to describe intelligent systems or software: Artificial Intelligent (AI), Machine Learning (ML), and DL. As shown in Figure 2, DL is part of ML and AI. In general, AI refers to the ability of a machine or system to imitate human behaviour and intelligence [18], whereas ML is a technique of learning from data [19] that enables the automation of analytical model creation.

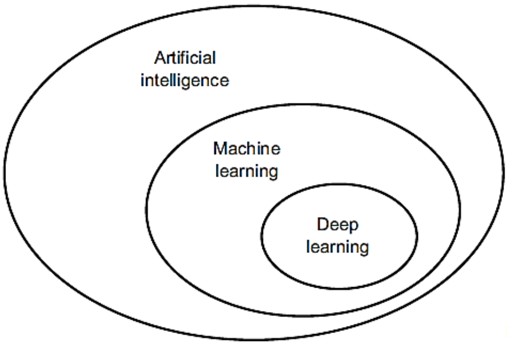

Figure 1. Illustration of the position of DL, compared to ML and AI

2.1.3 Gated Recurrent Unit

As a deep learning model, RNN applies a loop structure to identify the temporal features of the input sequence. The GRU and LSTM models are an enhanced RNN architecture. Compared to LSTM, the GRU architecture has fewer training parameters, resulting in higher training efficiency with the same level of accuracy as LSTM [20]. According to [21], GRU can achieve optimal performance even in intricate tasks when trained extensively. The main components of the GRU consist of an update gate and a reset gate, as shown in the illustration in Figure 2.

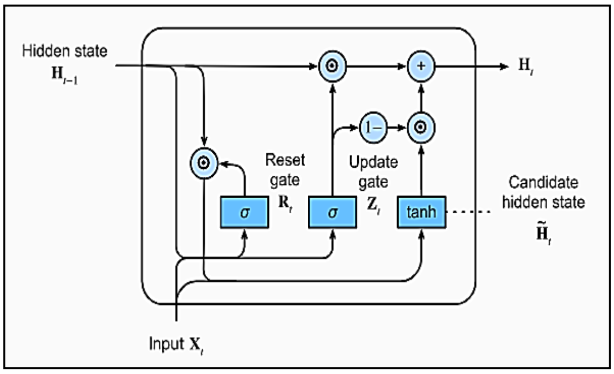

Figure 2. GRU Architecture

1. Z_t indicates the update gate used to determine the amount of stored past information. The greater the update gate value, the more information is retained [22]. The mathematical equation for the update gate is as follows:

$$Z_t = \sigma(W_z * [h_{t-1}, x_t] + b_z) \quad (1)$$

2. The second gate is the reset gate, which plays a role in determining the amount of past information that cannot be passed on to the next state. The reset gate function is similar to a combination of the input gate and forgets gate in LSTM [23]. The reset gate calculation is shown in Equation 2:

$$R_t = \sigma\left(W_r * [h_{t-1,} x_t] + b_r\right) \quad (2)$$

3. Next, the hidden state is determined using the tanh activation function with the following equation:

$$\widetilde{H_t} = \tanh(W * X_t + (R_t * h_{t-1})W + b_h) \tag{3}$$

Ultimately, the computation of the result is executed using the following formula:

$$H_t = \left(1 - Z_t\right) * \widetilde{H_t} + Z_t * h_{t-1}\Big) \tag{4}$$

2.1.4 Hyperparameter Tuning

Hyperparameters refer to the parameters that are predetermined prior to commencing the learning process. Conversely, hyperparameter tuning determines the most suitable value when training a dataset for a learning algorithm's hyperparameter [24]. Enhanced learning outcomes are achieved if the hyperparameter is optimized [25]. However, hyperparameter tuning is a time-consuming process. Consequently, hyperparameter optimization involves early stopping to terminate the model training process when certain conditions are met. The parameters utilized in the hyperparameter tuning process include epoch, batch size, and dropout.

1. Dropout
   Utilizing the dropout method during model training may enhance the architecture of the network. The implementation of dropouts can restrict the functionality of specific neurons, thereby decreasing interdependence among them, minimizing the interconnections between neurons, and avoiding reliance on particular characteristics. That can effectively prevent overfitting and enhance optimization at a structural level [26].

2. Batch Size
   As per [27], the magnitude of the batch is a crucial factor that requires calibration prior to commencing training, signifying the count of training examples employed in every gradient estimation iteration.

3. Epoch
   One epoch refers to the training data undergoing both forward and backward processes within the neural network. Insufficient epochs may lead to an imprecise model as the neural network lacks adequate knowledge to tackle the problem effectively. Conversely, excessive epochs can result in overfitting, leading to complications. Thus, the optimal number of epochs should be determined to achieve the best outcomes [28].

2.1.5 Model Evaluation

Evaluation of the predictive model is critical to determine how well the model fits the data. Several metrics can be used to evaluate predictive models, according to [29], including:

1. Mean Absolute Percentage Error
   The Mean Absolute Percentage Error (MAPE) calculates the percentage variance between the predicted and actual values.

2. Root Mean Squared Error
   Root Mean Squared Error (RMSE) is a measure that calculates the average difference between actual and predicted values and emphasizes more significant errors over more minor errors in evaluating model performance. The best value for MAPE and RMSE is zero.

   MAPE and RMSE mathematical equations are as follows:

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{Y_i - X_i}{Y_i}\right| \tag{5}$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(X_i - Y_i)^2$$
$$RMSE = \sqrt{MSE} \tag{6}$$

where:
X_i = predicted value in period i
Y_i = actual value in period i
n = the number of observations.

3. Accuracy
   Accuracy can be determined by comparing the number of accurate forecasts to the total number of forecasts, which are then expressed as a percentage.

$$a = \frac{t}{n} \times 100\% \tag{7}$$

with:

a is the accuracy in per cent,
t is the number of correct prediction attempts, and
n is the number of trials.

2.2 Methodology

This study uses the GRU method. Gated Recurrent Unit is a type of RNN based on optimized LSTM [30]. Although inspired by the LSTM model, GRU is considered more superficial to train and implement.

2.2.1 Research Stages

The stages that will be carried out in the research process are as follows:

Step 1: Input Data

The initial stage is to input data on the value of oil-non-oil and gas imports in Indonesia (in US$ million) into Google Colab. The data used in this research is monthly data on the value of Indonesia's oil-non-oil and gas imports (in US$ million) obtained from the Agency's website. Statistics Center (BPS): https://www.bps.go.id/indicator/8/1754/1/nilai-impor-migas-nonmigas.html, from January 2000 to June 2022 in millions of US $.

Step 2: Data Preprocessing

Second, an important step that must be taken is to preprocess the data by checking for missing values and then transforming the data with min-max normalization on the data.

Step 3: Splitting Data

Concerning data splitting, data samples are often divided into two data groups, namely training data to train models and testing data to validate models [31]. Overall, the training data size significantly impacts the predictive ability of the model built [32]. Many researchers suggest dividing the dataset with a 70/30 or 80/20 scheme (data training/testing). However, through their research, Pham et al. [33] found that increasing the training data size can improve training performance and make the model more stable. To achieve the most excellent potential model performance, we split the data distribution using a pattern of 60% for training data and 40% for testing data, 70% for training data and 30% for testing data, 80% for training data and 20% for testing data.

Step 4: Hyperparameter Tuning

Before the training process, do hyper-tuning on the parameters that have been set. Parameters to be hyper-tuned include GRU units, batch size, dropout and epoch.

Step 5: Build Model

Build the GRU model based on the best parameters from the hyper-tuning process.

Step 6: Prediction

If the best model has been obtained in the training process, the next step is to make predictions and compare the predicted results to actual data. However, the first denormalization is done on the actual data because the prediction results obtained from the built model are still in the form of a range of intervals during data normalization [34].

Step 7: Model Evaluation

Evaluate the model using MAPE, RMSE, and accuracy values.

Step 8: Forecasting

Furthermore, forecasting the value of Indonesia's oil-non-oil and gas imports (millions of US$) for the next two years if the accuracy value obtained shows that the model built is good enough.

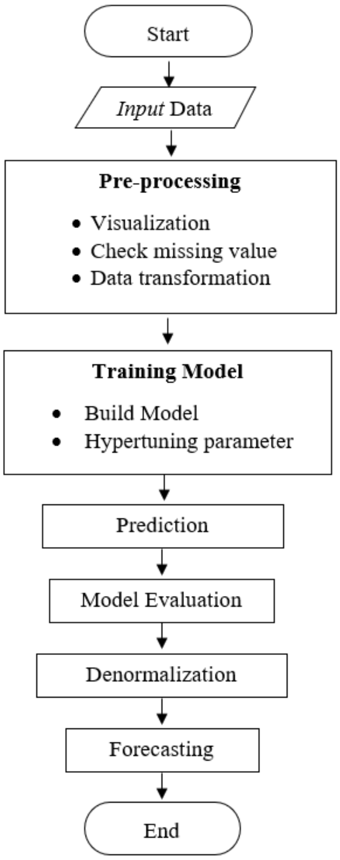

Figure 3. GRU Method Flowchart

3 RESULT AND ANALYSIS

First: input data on the value of Indonesia's oil-non-oil and gas imports (in US$ million), which includes monthly data from January 2000 to June 2022, totalling 270.

Second: the data that has been input is preprocessed. Data preprocessing can be done in several stages, including:

1. Visualization of time series data plots to see data concentration and distribution patterns. Based on Figure 2, the data on the value of Indonesia's oil-non-oil and gas imports (millions of US $) which will be analyzed, shows data fluctuations. That means that the data increases or decreases every month

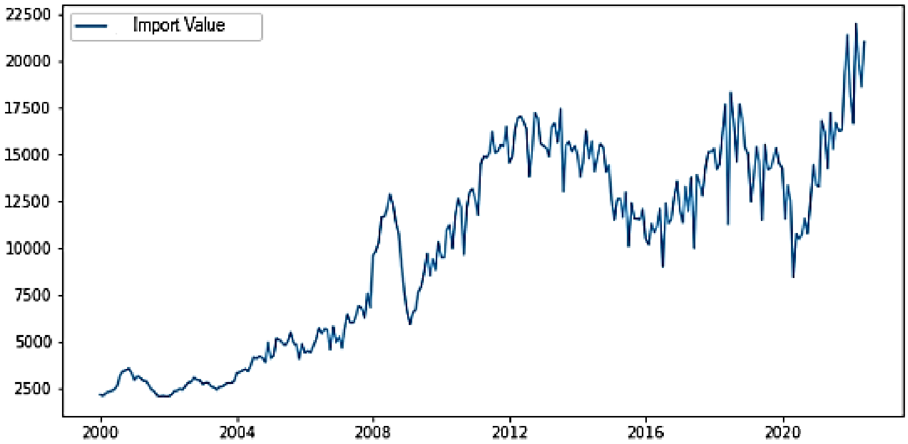

Figure 4. Time Series Plot of Oil-Non-Oil and Gas Import Value

2. Check missing value
   Converting data is carried out to adjust the data within a consistent value range. In this study, a min-max scaler was utilized to transform the data. By applying the min-max scaler, the original data is modified to attain a value range of 0-1. The ensuing outcome displays the normalized data on Indonesia's oil-non-oil and gas imports (in millions of US $).

Table 1. Data Normalization Results

Import Value (Million US$)	Normalization Results
2169.5	0.0045
2120.4	0.0020
2265.1	0.0093
2338.9	0.0130
2383.9	0.0153
…	…
16638.5	0.8113
21962.4	0.7322
19757.4	0.8891
18606.3	0.8313
21003.4	0.9517

Third: Different ratios will be used to splitting data into training and testing sets. The data training and testing ratios are 60% for training and 40% for testing, 70% for training and 30% for testing, 80% for training and 20% for testing.

Table 2. Splitting Data into Training and Testing

Splitting Data	Amount Data
60% training and 40% testing	162 data training and 108 data testing
70% training and 30% testing	189 data training and 81 data testing
80% training and 20% testing	216 data training and 54 data testing

Fourth: Build the GRU model. In this study, the GRU model consists of an input layer, a hidden layer, an output layer, and a sigmoid activation function. In addition, this model also has several other parameters, such as GRU unit, batch size, epoch and dropout. The following are the parameters used.

Table 3. Parameters used

Parameter Type	Amount
GRU Unit	16 dan 32
Batch size	16 dan 32
Epoch	50 dan 100
Dropout	0.2 dan 0.3

Fifth: determining the best parameters through hyper tunning and early stopping to stop the learning model when the conditions have been met. After the hyper-tuning process, the best parameters are obtained as follows.

Table 4. Optimal Parameters for 60% Training Data and 40% Testing Data

Parameter Type	Amount
GRU Unit	16
Batch size	16
Epoch	50
Dropout	0.2
Loss Function	0.529058
Running time (s)	385.134 s

Based on Table 4, it can be concluded that the best parameters for 60% training data and 40% testing data are 16 GRU Units, 16 batch sizes, 50 epochs, and 0.2 dropouts after going through the hyper tuning process. The optimal parameter still has a reasonably significant loss function value of 0.529058.

Table 5. Optimal Parameters for 70% Training Data and 30% Testing Data

Parameter Type	Amount
GRU Unit	32
Batch size	16

Parameter Type	Amount
Epoch	50
Dropout	0.3
Loss Function	0.534714
Running time (s)	413.961 s

Meanwhile, Table 5 shows that the best parameters for 70% of the training and 30% of the testing data are 32 GRU Units, 16 batch sizes, 50 epochs, and 0.3 dropouts after the hyper-tuning process. However, as shown in Table 4, the optimal parameter still has a reasonably significant loss function value of 0.534714.

Table 6. Optimal Parameters for 80% Training Data and 20% Testing Data

Parameter Type	Amount
GRU Unit	32
Batch size	16
Epoch	100
Dropout	0.2
Loss Function	0.373256
Running Time(s)	487.318 s

Unlike the two previous tables, Table 6 shows that the best parameters for 80% of the training and 20% of the testing data are 32 GRU Units, 16 batch sizes, 100 epochs, and 0.2 dropouts after going through the hyper-tuning process. This optimal parameter is suitable because it has a low loss function value of 0.373256. This loss function value is significant in optimizing the model during training.

Sixth: after obtaining the best model from the hyper-tuning process, the model built will be trained and evaluated using loss values and validation losses. The loss function used when building the model is MSE. The following is the loss curve and validation loss from the GRU model.

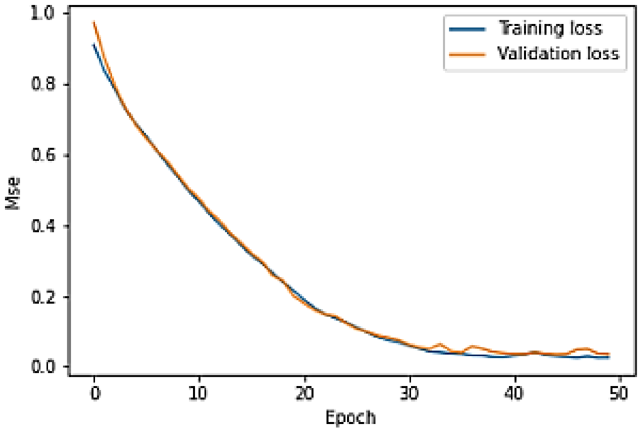

Figure 5. Loss Graph for Data Training 60%

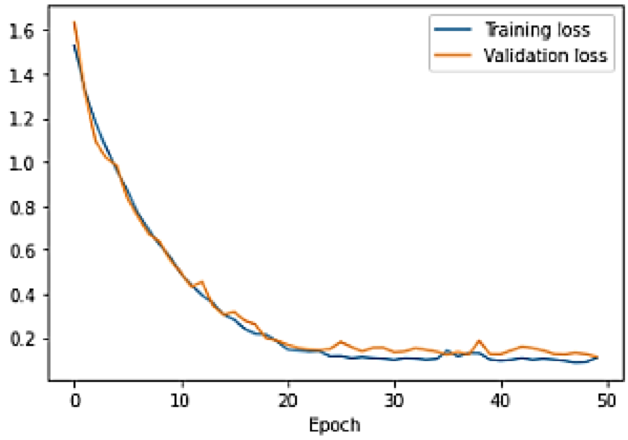

Figure 6.   Loss Graph for Data Training 70%

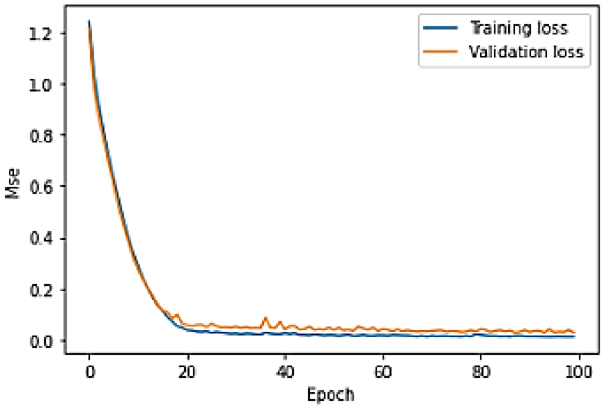

Figure 7. Loss Graph for Data Training 80%

Based on Figures 5, 6, and 7, it can be concluded that the resulting model is not overfitting because it can be seen that the loss value and validation loss converge at a certain point. The table compares the data distribution's composition and the parameters employed in the model.

Table 6. Comparison of Composition of Data Distribution and Parameters Used

Figure	Training	GRU Unit	Batch Size	Epoch	Dropout	Runtime
5	60%	16	16	50	0.2	385.143 *s*
6	70%	32	16	50	0.3	413.961 *s*
7	80%	32	16	100	0.3	487.318 *s*

Seventh: After getting the best model, the next step is to make predictions to evaluate the model built. Before making predictions, data needs to be normalized and compared with actual data. A comparison of these values is then visualized in the plot. If the predicted value is close to the actual value, then the model can predict the value of Indonesia's oil-non-oil and gas imports for the next 24 months.

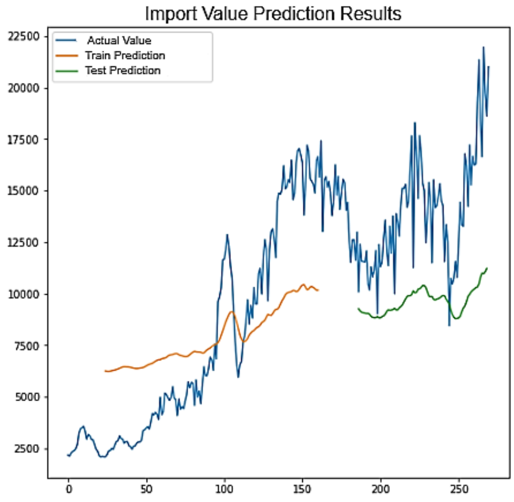

Figure 8. 60% Training Data Prediction Plot

The blue line in Figures 8, 9 and 10 represent the factual data, the orange line shows the predicted value of the training data, and the green line displays the predicted value of the testing data. The predictive visualization in Figure 6 is a ratio of 60% training and 40% testing data. Nevertheless, the graph of the predicted outcomes is dissimilar from the actual data distribution pattern, demonstrating a substantial variance between the predicted value and the actual data. Thus, it can be concluded that the constructed model is inadequate.

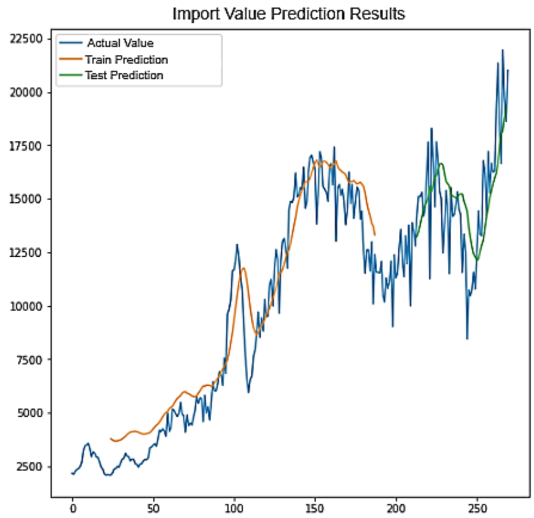

Figure 9. 70% Training Data Prediction Plot

Nonetheless, based on Figure 9, when employing a 70% training data approach and 30% testing data, the resulting graph adheres closely to the genuine data distribution, implying that the predicted value approximates the actual value. Hence, it can be deduced that the utilized model is considerably robust.

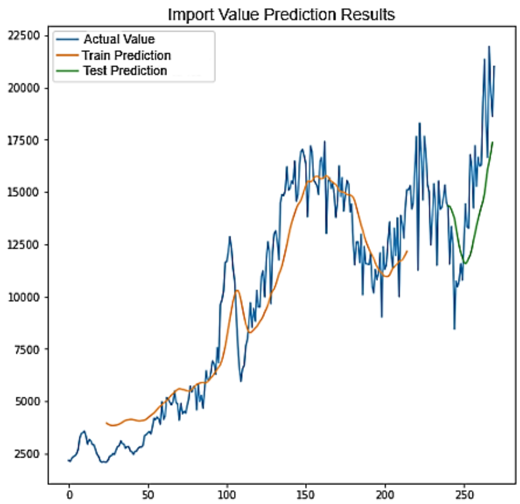

Figure 10. 80% Training Data Prediction Plot

In Figure 8, employing 80% of the data for training and 20% for testing yields a graph that closely adheres to the pattern of the actual data distribution. That suggests that the predicted value is nearly identical to the actual value. Hence, it can be deduced that the model utilized is appropriate.

Eighth: After making predictions, the next step is to evaluate the model using RMSE and MAPE values to determine the GRU model's accuracy in predicting the value of Indonesia's oil-non-oil and gas imports (millions of US$).

Table 7. Accuracy Comparison

Data Comparison	Model Evaluate		
	RMSE	MAPE	Accuracy
60% training and 40 testing	0.2462	0.999972%	99.000027%
70% training and 30 testing	0.1085	0.999959%	99.000040%
80% training and 20 testing	0.1509	0.999955%	99.000044%

According to Table 7, it can be inferred that utilizing an 80% data training ratio and 20% data testing yields the highest accuracy of 99.000044%. Therefore, this model can be trusted with exceptional forecasting proficiency for prediction purposes.

The Last: after evaluating the model, the next step is forecasting for the next two years, from June 2022 to July 2024. The following is a visualization of the results of forecasting the value of Indonesia's oil-non-oil and gas imports (in a million US$) using the 80% data training scheme and data testing 20%.

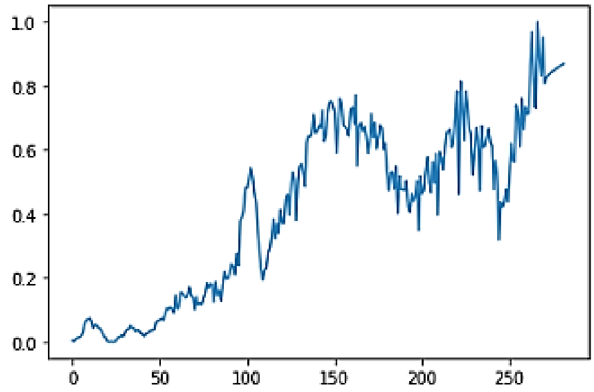

Figure 11. The plot of GRU Forecasting Results

According to Figure 11, it is evident that there has been an upsurge in the predicted amount of Indonesia's oil-non-oil and gas imports (measured in millions of US$) for the upcoming two years.

4 CONCLUSION

Based on data from https://www.bps.go.id/indicator/8/1754/1/nilai-impor-migas-nonmigas.html, this study uses the GRU algorithm to predict the growth of Indonesia's oil-non-oil and gas imports value from January 2000 to June 2022 in millions of US$. In practice, we use a variety of data-sharing schemes for training and testing, including 60%: 40%, 70%: 30%, and 80%: 20%. In addition, hyperparameter tuning is also carried out on GRU unit, dropout, batch size, and epoch parameters to obtain the most fitting parameters. The results show that 32 GRU units, a dropout of 0.2, 16 batch sizes, and 100 epochs with 80% training data scheme and 20% testing data are the most suitable criteria for estimating the value of Indonesia's oil-non-oil and gas imports (million us $) using GRU. The model has a good predictive ability with a 99.000044% accuracy and a MAPE of 0.9999955% to forecast the next two years. According to forecasting, Indonesia's oil-non-oil and gas imports (million US$) will increase from June 2022 to July 2024.

REFERENCES

[1] Hamdani and Haikal, Seluk Beluk Perdagangan Ekspor Impor, Edisi ke-2, Jakarta: Bushindo, 2018.
[2] S. M. Shakeel, N. S. Kumar, P. P. Madalli, R. Srinivasaiah and D. R. Swamy, "COVID-19 Prediction Models: a Systematic Literature," *Osong Public Health and Research Perspective,* vol. 12, no. 4, pp. 215-229, 2021.
[3] A. S. Bharatpur, "A Literature Review on Time Series Forecasting Methods," United Kingdom, 2022.
[4] P. Lara-Benitez, M. Carranca-Garcia and J. C. Riquelme, "An Experimental Review on Deep Learning Architectures for Time Series Forecasting," *International Journal of Neural Systems,* vol. 31, no. 3, 2021.
[5] J. F. Torres, D. Hadjout, A. Sebaa, F. Martinez-Alvarez and A. Troncoso, "Deep Learning for Time Series Forecasting: A Survey," *Big Data,* vol. 9, no. 1, pp. 3-21, 2021.
[6] T. Katte, "Recurrent Neural Network and its Various Architecture Types," *International Journal of Research and Scientific Innovation (IJRSI),* vol. 5, no. 3, pp. 124-129, 2018.
[7] K. Zarzycki and M. Lawrynczuk, "LSTMand GRUNeural Networks as Models of Dynamical Processes Used in Predictive Control: AComparison of Models Developed for Two Chemical Reactors," *Sensors,* vol. 21, no. 16, 2021.
[8] M. V. Wuthrich and M. Merz, Statistical Foundations of Actuarial Learning and its Applications, Switzerland: Springer, 2023.
[9] J. Chung, C. Gulcehre, K. Cho and Y. Bengio, "Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling," *arXiv: Neural and Evolutionary Computing,* 2014.
[10] X. H. Lie, H. H. Viet and G. Lee, "Application of Gated Recurrent Unit (GRU) Network for Forecasting River Water Levels Affected by Tides," in *APAC 2019*, Singapore, Springer Singapore, 2020, pp. 673-680.
[11] A. Mathew, A. Arul and S. Sivakumari, "Deep Learning Techniques: An Overview," in *Advanced*

Machine Learning Technologies and Applications, Singapore, Springer, 2020, pp. 599-608.
[12] S. A. Patil, L. A. Raj and B. K. Singh, "Prediction of IoT Traffic Using the Gated Recurrent Unit Neural Network- (GRU-NN-) Based Predictive Model," *Security and Communication Networks,* vol. 2021, 2021.
[13] M.-C. Lee, "Research on the Feasibility of Applying GRU and Attention Mechanism Combined with Technical Indicators in Stock Trading Strategies," *Applied Sciences*, vol. 12, no. 3, 2022.
[14] A. Agarwal, P. Dey and S. Kumar, "Sentiment Analysis using Modified GRU," in *IC3-2022: Proceedings of the 2022 Fourteenth International Conference on Contemporary Computing*, New York, 2022.
[15] J. Han, J. Pei and H. Tong, Data Mining Concepts and Techniques, 4th Edition, Elsevier, 2022.
[16] S. K. C. Rudraraju, N. Desai, M. Krishna and B. S. B. P. Rani, "Data Mining in Cloud Computing: A Review," *Journal of Advanced Research in Dynamical and Control Systems,* vol. 9, 2019.
[17] I. H. Sarker, "Deep Learning: A Comprehensive Overview on Techniques, Taxonomy, Applications and Research Directions," *SN Computer Science,* vol. 2, no. 6, pp. 1-20, 2021.
[18] I. H. Sarker, M. H. Furhad and R. Nowrozy, "AI-Driven Cybersecurity: An Overview, Security Intelligence Modeling and Research Directions," *SN Computer Science,* vol. 2, no. 3, pp. 1-18, 2021.
[19] I. H. Sarker, "Machine Learning: Algorithms, Real-World Applications and Research Directions," *SN Computer Science*, vol. 2, no. 3, pp. 1-21, 2021.
[20] X. Su, Y. Shan, C. Li, Y. Mi, Y. Fu and Z. Dong, "Spatial-temporal Attention and GRU Based Interpretable Condition Monitoring of Offshore Wind Turbine Gearboxes," *IET Renewable Power Generation,* vol. 16, no. 2, pp. 402-415, 2021.
[21] D. V. Trivedi and S. Patel, "An Analysis of GRU-LSTM Hybrid Deep Learning Models for Stock Price Prediction," *International Journal of Scientific Research in Science, Engineering and Technology,* vol. 9, no. 3, pp. 47-52, 2022.
[22] N. Li, L. Hu, Z.-L. Deng, T. Su and J.-W. Liu, "Research on GRU Neural Network Satellite Traffic Prediction Based on Transfer Learning," *Wireless Personal Communications,* vol. 118, no. 1, pp. 815-827, 2021.
[23] C. Zeng, C. Ma, K. Wang and Z. Cui, "Parking Occupancy Prediction Method Based on Multi Factors and Stacked GRU-LSTM," *IEEE Access,* vol. 10, pp. 47361-47370, 2022.
[24] L. Amusa, D. North and T. Zewotir, "Optimal Hyperparameter Tuning of Random Forests for Estimating Causal Treatment Effects," *Songklanakarin Journal of Science and Technology,* vol. 43, no. 4, pp. 1004-1009, 2021.
[25] E. K. Hashi and M. S. Uz Zaman, "Developing a Hyperparameter Tuning Based Machine Learning Approach of Heart Disease Prediction," *Journal of Applied Science & Process Engineering,* vol. 7, no. 2, pp. 631-647, 2020.
[26] X. Xie, M. Xie, A. J. Mosyahedi and M. H. N. Skandari, "A Hybrid Improved Neural Networks Algorithm Based on L2 and Dropout Regularization," *Mathematical Problems in Engineering,* vol. 2022, pp. 1-19, 2022.
[27] I. Kandel and M. Castelli, "The Effect of Batch Size on the Generalizability of the Convolutional Neural Networks on a Histopathology Dataset," *ScienceDirect,* vol. 6, no. 4, pp. 312-315, 2020.
[28] I. Cardoza, J. P. Garcia-Vazquez, A. Diaz-Ramirez and V. Quintero-Rosas, "Convolutional Neural Networks Hyperparameter Tunning for Classifying Firearms on Images," *Applied Artificial Intelligence,* vol. 36, no. 1, 2022.
[29] D. Kim, T. Kim, J. Jeon and Y. Son, "Soil-Surface-Image-Feature-Based Rapid Prediction of Soil Water Content and Bulk Density Using a Deep Neural Network," *Applied Sciences,* vol. 13, no. 7, pp. 1-14, 2023.
[30] A. S. Santra and J.-L. Lin, "Integrating Long Short-Term Memory and Genetic Algorithm for Short-Term Load Forecasting," *Energies,* vol. 12, no. 11, pp. 1-11, 2019.
[31] Q. H. Nguyen, H.-B. Ly, S. L. Ho, N. Al-Ansari, H. V. Le, V. Q. Tran, I. Prakash and B. T. Pham, "Influence of Data Splitting on Performance of Machine Learning Models in Prediction of Shear Strength of Soil," *Mathematical Problems in Engineering,* vol. 2021, 2021.
[32] C. Qi, A. Fourie, Q. Chen and Q. Zhang, "A Strength Prediction Model using Artificial Intelligence for Recycling Waste Tailings as Cemented Paste Backfill," *Journal of Cleaner Production,* vol. 183, pp. 566-578, 2018.
[33] B. T. Pham, C. Qi, L. S. Ho, T. Nguyen-Thoi, N. Al-Ansari, M. D. Nguyen, H. D. Nguyen, H.-B. Ly, H. V. Le and I. Prakash, "A Novel Hybrid Soft Computing Model Using Random Forest and Particle Swarm Optimization for Estimation of Undrained Shear Strength of Soil," *Sustainability,* vol. 12, no. 6, 2020.
[34] U. I. Arfianti, D. C. R. Novitasari, N. Widodo, M. Hafiyusholeh and W. D. Utami, "Sunspot Number

Prediction Using Gated Recurrent Unit (GRU) Algorithm," *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, vol. 15, no. 2, pp. 141-152, 2021.

BIBLIOGRAPHY OF AUTHORS

Dian Kurniasari is a Lecturer in the Mathematics Department of FMIPA at the University of Lampung. She obtained his bachelor's degree from Gadjah Mada University and his master's degree from Curtin University.

Wamiliana is a Lecturer in the Mathematics Department of FMIPA at the University of Lampung. She completed his bachelor's degree at the University of Indonesia, master's at Curtin University, and doctoral degree at the University of Montana.

Sulistian Oskavina completed his undergraduate education at the Mathematics Department of FMIPA, University of Lampung and graduated in 2022.

Warsono is a Lecturer in the Mathematics Department of FMIPA at the University of Lampung. He completed his bachelor's and master's degrees at IPB and his doctoral degree at the University of Alabama.