Jurnal ELTIKOM:
Jurnal Teknik Elektro, Teknologi Informasi dan Komputer

Vol. 7, No. 2, December 2023, page. 125-132
ISSN 2598-3245 (Print), ISSN 2598-3288 (Online)
DOI: http://doi.org/10.31961/eltikom.v7i2.770
Available online at http://eltikom.poliban.ac.id

MODEL ANALYSIS OF GATED RECURRENT UNIT FOR MULTIVARIATE RICE PRICE FORECASTING

Muhammad Ikhsan Ananda
Departement of Computer Science, IPB University, Bogor, Indonesia
e-mail: ikhsanananda@apps.ipb.ac.id.com

Received: 1 June 2023 – Revised: 8 July 2023 – Accepted: 10 July 2023

ABSTRACT

Food security, especially in the agricultural sector in the form of food price stability of rice as a national food ingredient is a strategic issue for Indonesia. Rice price forecasting is needed to mitigate rising rice food prices. Rice price fluctuations can be caused by internal factors such as bad weather or external factors such as the low selling price of rice, resulting in losses for farmers. This study aims to carry out multivariate rice price forecasting in DKI Jakarta by involving rice prices, weather, economic, and health factors using the Gated Recurrent Unit (GRU) algorithm where the accuracy test is based on the MAPE value between forecasting results and actual data. As a result of the GRU algorithm for multivariate rice price forecasting, the MAPE for training and testing is 0.964% and 2.628%, indicating that all models in the measurement category are very well represented.

Keywords*: gated recurrent unit, multivariate forecasting, rice prices.*

I. INTRODUCTION

FOOD security can be interpreted as a condition that describes an individual's ability to access healthy and nutritious food physically, socially and economically [1]. Indonesia ranks 63rd in terms of food security compared to 113 other countries [2]. One sector that plays an important role in food security is agriculture, namely the rice sub-sector. Rice is a staple food commodity for the Indonesian people so that fluctuations in food prices have an impact on the stability of Indonesia's staple food nationally [3]. These price fluctuations are caused by several internal factors including the production of basic commodities that have failed due to weather, pest or disease disturbances, and obstructions in food distribution channels [4]. In addition to these factors, external factors in the economic and health sectors also affect food prices. The economic factors that cause fluctuations in food prices are the inflation rate, rising fuel prices, and the Rupiah exchange rate against the US Dollar [5]. The COVID-19 pandemic has reduced people's purchasing power along with the sluggish economy and increased spending on the health sector [6]. Fluctuations in inflation rates are in line with fluctuations in food and fuel prices. For instance, if rice prices increase by 1%, it will push up inflation by 0.8%, while if fuel prices increase by 1%, it will raise inflation rate by 0.07%, accompanied by an increase in other food prices due to rising production costs [7].

Multivariate forecasting of food prices needs to be done to prevent a food crisis due to an imbalance in demand and supply as well as weakening purchasing power in line with rising food prices. Forecasting food prices for the next few months has been carried out by applying certain algorithms to support the readiness of related parties to changes in food prices [8]. The ARIMA algorithm produces an RMSE value of 313.379 for forecasting food prices for 30 days [9]. Furthermore, forecasting for 12 days with the Double Exponential Smoothing Holt and Brown algorithm yielded MSE values of 21,328.60 and 188,086.86, respectively [10]. Then, forecasting for 8 months with the Holt-Winters Exponential Smoothing algorithm produces a MAPE value of 1.2% [11]. Next, the Weighted Moving Average algorithm for forecasting for 1 month produces a MAPE value of 1.90% [12]. After that, the Naïve Bayes algorithm can predict whether the price of a food commodity is likely to rise or fall compared to the next period [13]. The implementation of the forecasting algorithm in previous studies was only univariate forecasting or only considering one research variable and a comparison of the error values of several algorithms for forecasting food prices instead of testing the accuracy of forecasting. One of the studies using the multivariate forecasting algorithm is local soybean price forecasting where multivariate is only done in the form of 2 inputs by looking at prices from the previous two days, but only uses one variable which is local soybean price [14].

This study aims to forecast rice prices which also involve the weather, economic, and health factors as independent variables for multivariate rice price forecasting using the Gated Recurrent Unit (GRU) algorithm besides the variable of rice prices itself. The GRU algorithm can remember a collection of information that has been stored for a long time and delete information that is no longer relevant to a simpler computation than LSTM but has the same accuracy and is still effective enough to avoid the vanishing gradient problem. The results of the multivariate rice price forecasting will be tested for accuracy with actual data using the MAPE value to determine the quality of the model's accuracy. Therefore, the following research questions (RQ) are proposed in this study: (RQ1) In terms of economic and health factors, what variables affect the price of rice? and (RQ2) How does the MAPE value category resulting from rice price forecasting compared to actual data as a measure of model accuracy?

This study contributes to data science, especially in testing the implementation of forecasting algorithms with time series data, in the form of model performance from the resulting algorithm in the form of accuracy using MAPE values and model fit. In addition, it also provides insight into market managers like Perumda Pasar Jaya and the government like Perum BULOG who have responsibility for food price stability regarding the use of technology that can be developed to help mitigate the causes of fluctuations in food prices.

II. Research Method

A. Research Data

The primary research data were rice food prices with the categories of Low-Quality Rice 1, Low-Quality Rice 2, Medium-Quality Rice 1, Medium-Quality Rice 2, Super-Quality Rice 1, and Super-Quality Rice 2. The data came from the results of the daily recording of the National Strategic Food Price Information Center (PIHPSN) at Jatinegara Main Market, Kramat Jati, and Pasar Minggu Market from 1 January 2017 to 28 February 2023.

The research supporting data used as predictors were data of weather, inflation, fuel prices, Rupiah exchange rate, and number of positive cases of COVID-19. The weather data came from the daily recording of the Halim Perdana Kusuma Meteorological Station. Inflation data came from monthly publications on the official website of Bank Indonesia to see the monthly inflation rate in Indonesia. Fuel price data came from news publications on Pertamina's official website when there was an increase

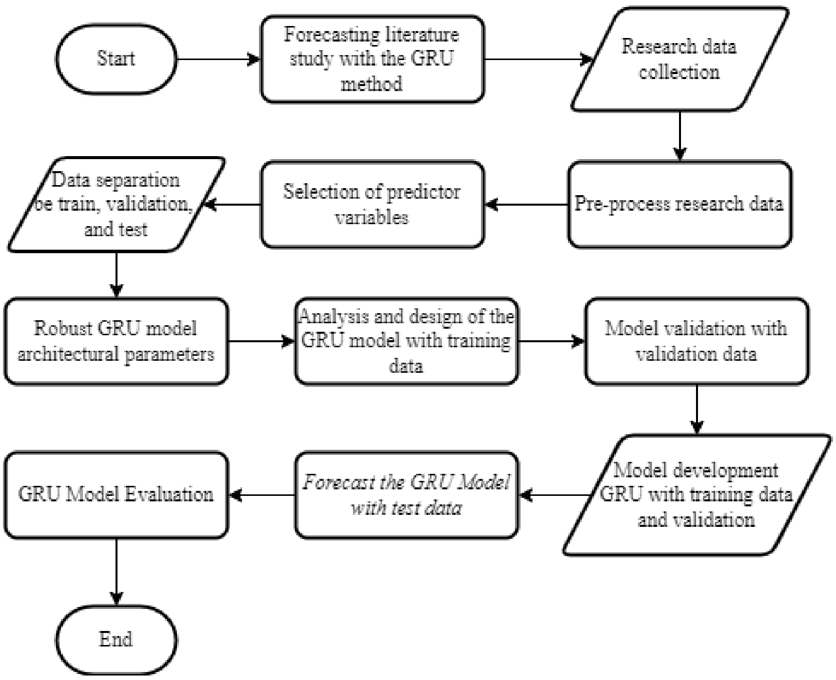

Figure. 1. Stages of forecasting the price of rice

or decrease in the price of fuel oil (BBM) at RON 90 and RON 92. Rupiah exchange rate data came from the annual publication of the Rupiah exchange rate against the US Dollar on the official website of Statistics Indonesia. Data on the number of positive cases came from the daily publication of the DKI Jakarta Information and Communication Service on the Corona Jakarta website.

The research data is organized in a tabular format with 18 columns and 2251 records. Data recorded in the research consist of the recording date column, average temperature, average humidity, rainfall, sunshine duration, average wind speed, inflation rate, fuel prices in the RON 90 and RON 92 categories, the Rupiah to US Dollar exchange rate, COVID-19 active cases and daily positive cases, and rice prices from categories Low-Quality Rice 1 to Super-Quality Rice 2.

B. Research Stages

The development of the forecasting model employing the Gated Recurrent Unit method consisted of five sequential stages: literature review, data pre-processing, initialization of the architectural parameters of the GRU model, analysis and design of the GRU model, and evaluation of the GRU model between forecasting results and actual food prices for rice in DKI Jakarta. The stages of the research method can be seen in Figure 1.

C. Data Preprocessing

Multivariate forecasting required data preprocessing to manage missing values and minimize errors in order to improve the accuracy of the resulting data. The handling of missing values in time series data for rice prices used the Next Observation Carried Backward (NOCB) method. By using the Next Observation Carried Backward method, missing values were calculated by replacing them with values in the nearest index of missing values from the missing values index. It was appropriate to use this method to deal with missing values relating to rice prices in the initial data recording, where there were no data at the start of recording in January 2017. Meanwhile, the handling of missing values in other data used spline interpolation. The number of missing values for each variable in the data is shown in Table 1.

Furthermore, efforts to minimize errors used the normalization method with min-max scaling techniques on the dataset by changing the values of the independent and dependent variables at intervals from 0 to 1. In addition, the preprocessed data were separated into training, validation, and test data in proportions of 60%, 20%, and 20%, respectively.

D. Initialize GRU Model Architecture Parameters

The initialization of the parameters needed to develop the GRU algorithm model was the number of hidden layers, input layers, output layers, epochs, and batches. This stage was needed to see the relationship that occurred between changes in each parameter value to validation loss with validation data. This stage was illustrated by testing each layer with each epoch and batch value. This was also true for epochs, where each epoch was tested separately with a batch value. Therefore, this step was carried

TABLE 1
THE NUMBERS OF MISSING VALUES

Variable	Number of Missing Values
Date of Records	0
Average Temperature	16
Average Humidity	17
Rainfall Rate	472
Sunshine Duration	37
Average Wind Speed	13
Inflation Rate	0
Fuel RON 90	0
Fuel RON 92	0
Exchange rate	727
Active Covid-19	0
Daily Covid-19	0
Low-Quality Rice 1	395
Low-Quality Rice 2	395
Medium-Quality Rice 1	395
Medium-Quality Rice 2	395
Super-Quality Rice 1	395
Super-Quality Rice 2	395

out repeatedly to get the best parameters where the parameters selected were the parameters with the smallest validation loss to produce a multivariate forecasting model for rice prices.

E. GRU Model Analysis and Design

The analysis and design of the GRU model was carried out after the data preprocessing and initialization of the GRU architecture parameters were completed. The development of the GRU model used training data to train the GRU model with sequential functions. In addition, the model was built using the best parameters found during the GRU parameter initialization stage. Then, the resulting GRU model was tested with the RMSE loss function and the Adam optimizer. Next, the resulting model was trained with the fit function to fit the model with validation data.

F. GRU Model Evaluation

After the analysis and design stages of the GRU model were done, the evaluation stage of the GRU model was carried out, which was denoted by the minimum validation loss value that had been obtained from many efforts to apply the GRU model parameters. The method used was to calculate the Mean Absolute Percentage Error (MAPE) to determine the accuracy between the testing data and the predicting findings. The testing data was gathered during the data preprocessing stage, when the data had not been used for training or model optimization in the previous phases. After obtaining the MAPE value, the value was analyzed with the condition that if it is less than 10%, the model accuracy is very good, if it is between 10% and 20%, the model accuracy is good, if it is between 20% and 50%, the model accuracy is sufficient, and if it is greater than 50%, the model accuracy is poor [15].

III. RESULT AND DISCUSSION

A. Selection of Economic and Health Factor Variables

Variables belonging to economic factors are inflation rates, fuel prices for both Fuel RON 90 and Fuel RON 92, as well as the middle exchange rate of the Rupiah against the US Dollar. Meanwhile, the variable belonging to the health factor is the number of positive daily and active COVID-19 cases in DKI Jakarta. The selection of variables on economic and health factors was carried out on the variables of fuel prices and the number of DKI Jakarta COVID-19 cases. The method used is the Pearson correlation coefficient where the selected variable is the variable with the highest average correlation value for rice prices as shown in Table 2.

Table 2 shows that the Fuel RON 92 variable and the number of active COVID-19 cases have a higher average Pearson correlation coefficient on the price of rice than the Fuel RON 90 variable and the number of daily COVID-19 cases, indicating that these variables have a greater impact on rice price fluctuations.

B. Robust GRU Model Architecture Parameters

Robust model architecture parameters are an attempt to obtain the best architectural parameters of the

TABLE 2
THE PEARSON CORRELATION AVERAGE VALUE

Variable	Average Correlation
Fuel RON 90	0.5017
Fuel RON 92	0.5273
Number of Daily COVID-19 Cases	0.3285
Number of Active COVID-19 Cases	0.3422

TABLE 3
THE EXAMPLE OF SELECT BEST ARCHITECTURE FOR LOW-QUALITY RICE 1

Layer	Epoch	Batch	Validation Loss
150	110	8	0.00831
150	110	16	0.00913
150	110	32	0.00065
150	110	64	0.00207
150	110	128	0.02040
150	110	256	0.00660
150	110	512	0.02555
150	110	1024	0.01591

model by running experiments that are repeated numerous times, then picking the best model by looking at the minimum validation loss value. In addition, this stage also aims to maximize forecast results with the hope that the resulting MAPE value will be smaller to improve the quality of the model.

The flow of robust GRU model architecture parameters begin with initializing layer, epochs, and batches values, initializing early stopping callback parameters and check point models, and ending with selecting the best model parameters based on minimum validation loss. Repeatedly examined parameters include layers, epochs, and batches with values ranging from 10 to 250 for layers and epochs and 8 to 512 for batches. This stage is illustrated by the fact that each layer is tested with the appropriate epoch and batch data. This also applies to epochs where each epoch is checked individually with batch values.

The best GRU model architecture hyperparameter optimization results are based on the smallest validation loss value between training data and validation data. Table 3 illustrates the selection of the optimum GRU model architectural parameters for each rice commodity.

As shown in Table 3, the validation loss for Low-Quality 1 rice occurs at layers 150 and epochs 110. When the batch value is 32, there is the smallest validation loss, which means that 150, 110, and 32 are the best layers, epochs, and batch parameters for Low-Quality 1.

The results of the optimization of the optimal architectural parameters of the GRU model are based on the minimum validation loss value between the training data and the validation data. Table 4 shows the optimum parameters obtained by the GRU algorithm for forecasting rice prices: the number of layers, epochs, batches, and validation loss.

In addition, Table 4 also shows that the best parameters of the GRU model for each rice categories are not entirely the same. This is due to the different movements of food prices so that the architectural parameters of the model adjust the movement patterns of each food commodity.

C. GRU Model Analysis of Training Data

The GRU model developed with the best values of layers, epochs, and batches in rice prices was then tested on training data. The analysis aims to see the percentage of model output errors against the training data. The GRU model is used to make predictions using training data, and the resulting MAPE values are then examined. The MAPE values for the training data are shown in Table 5.

The average MAPE value for the training data is 0.964% indicating an average error in the model output in the form of a larger or smaller model output of 0.964% for the training data. In addition, Table 5 also shows that the MAPE value for each food commodity for rice is all below 10%, meaning that the GRU model that has been produced is in the very good category for multivariate rice price forecasting.

D. GRU Model Analysis of Test Data

Analysis of the GRU model using test data aims to see the accuracy of forecasting rice prices. The method used is to look at the MAPE value between the model output and the test data. The MAPE value of the test data is shown in Table 6.

TABLE 4
THE GRU PARAMETER OPTIMIZATION RESULTS

Rice Categories	Layer	Epoch	Batch	Validation Loss
Low-Quality Rice 1	150	110	32	0.00065
Low-Quality Rice 2	10	258	32	0.00182
Medium-Quality Rice 1	300	90	16	0.00073
Medium-Quality Rice 2	90	259	64	0.00137
Super-Quality Rice 1	110	130	8	0.00531
Super-Quality Rice 2	10	130	32	0.00147

TABLE 5
THE MAPE VALUE OF TRAINING DATA

Rice Categories	MAPE (%)
Low-Quality Rice 1	2.821
Low-Quality Rice 2	0.169
Medium-Quality Rice 1	0.439
Medium-Quality Rice 2	1.598
Super-Quality Rice 1	0.536
Super-Quality Rice 2	0.964

TABLE 6
THE MAPE VALUE OF TESTING DATA

Rice Categories	MAPE (%)
Low-Quality Rice 1	9.361
Low-Quality Rice 2	0.216
Medium-Quality Rice 1	0.610
Medium-Quality Rice 2	3.508
Super-Quality Rice 1	1.731
Super-Quality Rice 2	0.341

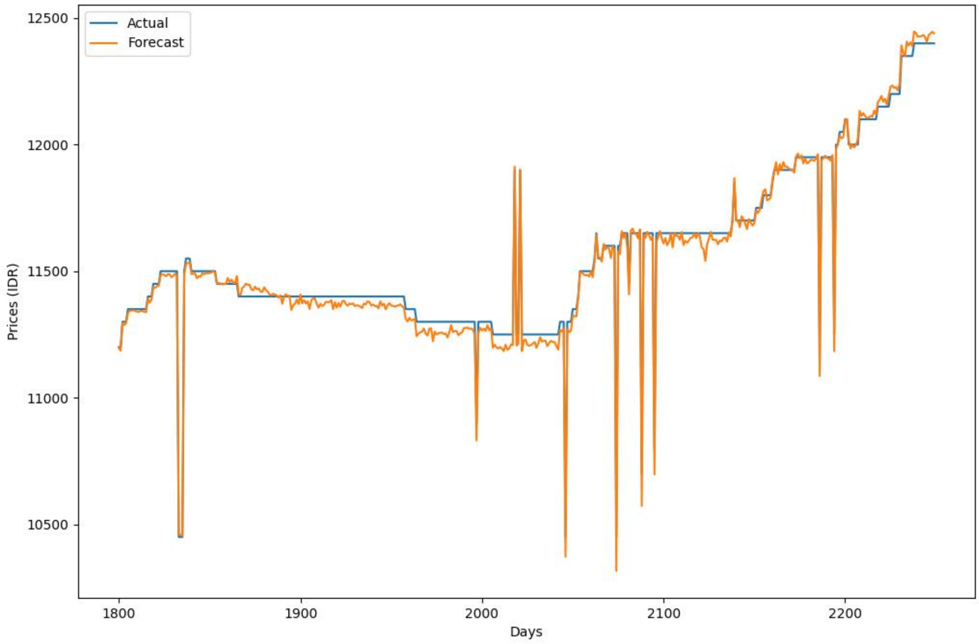

Figure. 2. Forecasting result with Low-Quality Rice 2 test data

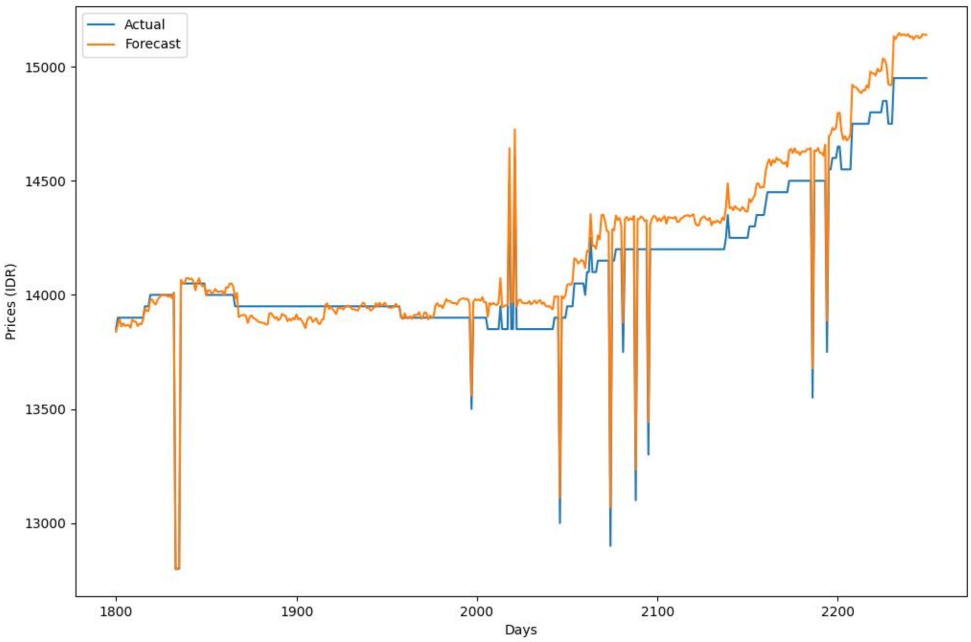

Figure. 3. Forecasting result with Medium-Quality Rice 1 test data

The average MAPE value of the test data is 2.628% indicating an average error in the model output in the form of a larger or smaller model output of 2.628% of the test data. In addition, Table 6 also shows that the MAPE value for each food commodity for rice is all below 10%. This suggests that the GRU model developed falls into the very good category for multivariate rice price forecasting. Figures 2 to 4 depict the visualization of the results of the GRU model analysis on test data for rice, where the blue

TABLE 7
THE DEVIATION BETWEEN MAPE TRAINING AND TEST DATA

Rice Categories	MAPE (%)		
	Train	Test	Deviation
Low-Quality Rice 1	2.821	9.361	6.540
Low-Quality Rice 2	0.169	0.216	0.047
Medium-Quality Rice 1	0.439	0.610	0.171
Medium-Quality Rice 2	1.598	3.508	1.909
Super-Quality Rice 1	0.536	1.731	1.196
Super-Quality Rice 2	0.964	0.341	0.118

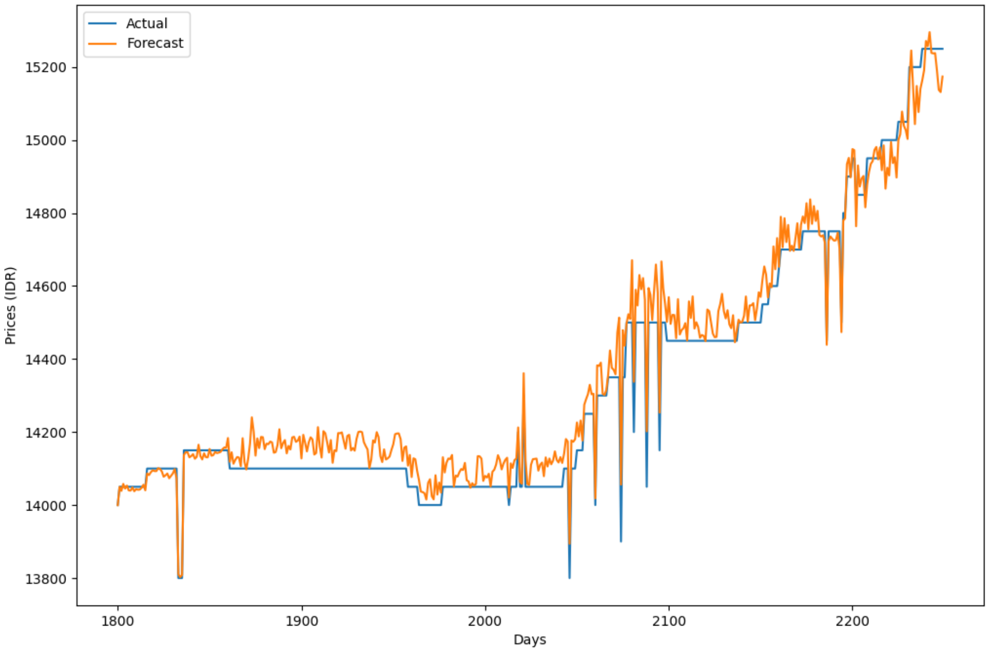

Figure. 4. Forecasting result with Super-Quality Rice 2 test data

line represents the actual data and the orange line represents the forecast results. In addition, Figures 2 to 4 aim to illustrate the accuracy of the rice price forecast, where the closer the gap between the actual data (blue line) and the forecasted data (red line), the higher the accuracy of the actual data in the forecasting results.

E. Analysis of GRU Model Accuracy Against Training and Test Data

Analysis of the accuracy of the GRU model on the training and test data aims to see whether the resulting model is underfitting, overfitting, or best fitting. This analysis was carried out by looking at the deviation between the MAPE values in the training data and the MAPE in the test data. Comparison of MAPE values to training and test data where the value in the deviation column is the subtraction of the MAPE against the training data with the test data is shown in Table 7.

Table 7 demonstrates that the difference between the MAPE values of the training data and the test data is not excessive. In addition, the MAPE value for both training and test data is below 10% which indicates that both are very good at understanding and performing multivariate rice price forecasting.

F. Discussion

For the answer to RQ1, the Pearson correlation coefficient average of the Fuel RON 92 variable is greater by 0.0256 than that of the Fuel RON 90 variable, and the number of active COVID-19 cases is greater by 0.0137 than the number of daily COVID-19 cases on the price of rice. This suggests that the Fuel RON 92 and number of active COVID-19 cases variables will have a greater impact on the price fluctuations of rice.

For the answer to RQ2, the average MAPE value of the test data is 2.628%, where the highest value is 9.361% in Low-Quality Rice 1 and the lowest value is 0.047 in Low-Quality Rice 2, indicating an average error in the form of a larger or smaller model output of 2.628% of the test data. In addition, all of the MAPE values of the test data are below 10%. This indicates that the developed GRU model for multivariate rice price forecasting is in the very good category.

IV. Conclusion

The resulting parameters for each rise categories are not the same, because they adjust the pattern of price movements for each rice categories. As a result of the GRU algorithm, the MAPE averages for training and test data are 0.964% and 2.628%, respectively. As a result of the MAPE value for all models being less than 10%, it can be concluded that the model developed is the most accurate and best-fitting model for multivariate rice price forecasting.

References

[1] F. Ulirrahmi and A. Yazid, "Wakaf Berbasis Akad Muamalah untuk Meningkatkan Ketahanan Pangan di Indonesia," *Al-Mustashfa J. Penelit. Huk. Ekon. Syariah*, vol. 7, no. 2, pp. 230–243, 2022.

[2] A. R. Herawati, T. Yuniningsih, and I. H. Dwimawanti, "Assesing the Impact of Digital Technologies on Governance Policies for Food Security: A Case Study of Indonesia," *KnE Soc. Sci.*, pp. 166–184, 2023.

[3] S. Sen, D. Sugiarto, and A. Rochman, "Komparasi metode multilayer perceptron (MLP) dan long short term memory (LSTM) dalam peramalan harga beras," *Ultim. J. Tek. Inform.*, vol. 12, no. 1, pp. 35–41, 2020.

[4] D. Z. Rizaldy, "Pengaruh Harga komoditas pangan terhadap inflasi di Kota Malang Tahun 2011-2016," *J. Ekon. Pembang.*, vol. 15, no. 2, pp. 171–183, 2017.

[5] P. S. Silaban, D. N. Harefa, J. I. M. Napitupulu, and J. P. B. Sembiring, "The Impact of BI Interest Rate and Amount of Money Supply on Inflation in Indonesia During 2017-2019," *Media Ekon. dan Manaj.*, vol. 36, no. 1, 2021.

[6] N. D. Pramanik, "Dampak bantuan paket sembako dan bantuan langsung tunai terhadap kelangsungan hidup masyarakat padalarang pada masa pandemi covid 19," *J. Ekon. Sos. Hum.*, vol. 1, no. 12, pp. 113–120, 2020.

[7] H. S. Sundoro, "Pengaruh Harga Komoditas Pangan Dan Bensin Terhadap Tingkat Inflasi Selama Pemerintahan Jokowi," *E-Jurnal Ekon. Dan Bisnis Univ. Udayana*, vol. 10, no. 02, pp. 73–82, 2021.

[8] N. P. Dewi and I. Listiowarni, "Implementasi Holt-Winters Exponential Smoothing untuk Peramalan Harga Bahan Pangan di Kabupaten Pamekasan," *Digit. Zo. J. Teknol. Inf. Dan Komun.*, vol. 11, no. 2, pp. 219–231, 2020.

[9] I. Mardianto, M. I. Gunawan, D. Sugiarto, and A. Rochman, "Perbandingan Peramalan Harga Beras Menggunakan Metode ARIMA pada Amazon Forecast dan Sagemaker," *J. Resti*, vol. 4, no. 3, 2020.

[10] A. Gunaryati, F. Fauziah, and S. Andryana, "Perbandingan Metode-metode Peramalan Statistika untuk Data Indeks Harga Pangan," *STRING (Satuan Tulisan Ris. dan Inov. Teknol.*, vol. 2, no. 3, pp. 241–248, 2018.

[11] N. P. Dewi and I. Listiowarni, "Peramalan Harga Bahan Proyek Menggunakan Metode Least Square (Studi Kasus: CV Rizky Mulya)," *J. Tek. Inform.*, vol. 2, no. 1, pp. 28–33, 2019.

[12] R. Ramadania, "Peramalan Harga Beras Bulanan di Tingkat Penggilingan dengan Metode Weighted Moving Average," *Bimaster Bul. Ilm. Mat. Stat. dan Ter.*, vol. 7, no. 4, 2018.

[13] B. E. W. Asrul and S. Zuhriyah, "Sistem Informasi Peramalan Harga Pangan Dengan Menggunakan Metode Naïve Bayes Di Kota Makassar," *E-JURNAL JUSITI J. Sist. Inf. dan Teknol. Inf.*, vol. 7, no. 2, pp. 163–171, 2018.

[14] F. Fatkhuroji, S. Santosa, and R. A. Pramunendar, "Prediksi Harga Kedelai Lokal Dan Kedelai Impor Dengan Metode Support Vector Machine Berbasis Forward Selection," *J. Cyberku*, vol. 15, no. 1, pp. 61–76, 2019.

[15] Y.-F. Chang, C.-J. Lin, J. M. Chyan, I. M. Chen, and J.-E. Chang, "Multiple regression models for the lower heating value of municipal solid waste in Taiwan," *J. Environ. Manage.*, vol. 85, no. 4, pp. 891–899, 2007.