IMPLEMENTATION OF GRU AND ADAM OPTIMIZATION METHOD FOR STOCK PRICE PREDICTION

Muhammad Mada[1], Andi Farmadi[2], Irwan Budiman[3], Mohammad Reza Faisal[4], Muhammad Itqan Mazdadi[5]
[12345]Ilmu Komputer FMIPA ULM
A. Yani St. KM 36 Banjarbaru, South Kalimantan
Email: j1f115039@mhs.ulm.ac.id[1] andifarmadi@ulm.ac.id[2], irwan.budiman@ulm.ac.id[3], reza.faisal@ulm.ac.id[4], mazdadi@ulm.ac.id[5]

Abstract

In terms of their potential, stocks are one of the most profitable investment options today. If done well and right, stocks can be a very profitable investment. However, volatile stock prices make it necessary to predict stock prices to make a profit. Gated Recurrent Unit (GRU) is a method for predicting time series data such as stock prices. The Optimization method is needed to get accurate prediction results. The weight renewal optimization method such as Adam is implemented to obtain the best weight in the Gated Recurrent Unit (GRU) and to find out the best loss function value generated by the Adam optimization method. The GRU-Adam implementation is carried out on two stock data, namely ICBP and YULE. The results of this research are that the ICBP data yields the respective loss function values, namely train loss 0.0016 and validation loss 0.0007. Whereas the YULE data resulted in a train loss value of 0.0051 and a validation loss of 0.0031. The MAPE generated in the ICBP stock data is 0.97%. While the YULE data is 3.00%.

Keywords: *GRU, Adam, Prediction, Stock Price*

1. INTRODUCTION

In predicting time-series data, a good method is needed so that the prediction results do not deviate far from the true value. There are many methods for predicting time series data. One method that can be used to make predictions is the Recurrent Neural Network (RNN) method. RNN is suitable for processing sequential data (continuous data) and has the advantage of being able to store information from the past by looping through its architecture. Some of the variants of RNN that have developed to date are the Gated Recurrent Unit (GRU) [3].

Based on the publication journal Cho, et al. (2014) GRU has advantages because of the gating concept, so it can avoid missing problems from the gradient that might occur in RNN. GRU can be used to forecast time-series data, for example in a foreign currency exchange rate forecast research written by [3], more than 60%. A precision value of more than 60% is sufficient to solve predictive cases such as foreign exchange, stocks , or gold whose value can fluctuate depending on the market situation [3].

Adaptive moment estimation (Adam) optimization method is an optimization method that develops by taking advantage of the Adaptive Gradient (AdaGrad) and Root Mean Square Propagation (RMSprop) methods. The LSTM model with Adam's optimization obtained the best results, namely with the composition of training data 70% and test data 30%, parameter 1 time series pattern, number of 25 hidden neurons, and max epoch is 100 with an average accuracy of 95.36% training data and data testing 93.5% [2].

In this journal, Gated Recurrent Unit (GRU) and Adam optimization method is proposed to predict stock price of Indonesian stock market. We use TensorFlow to help us design the model. At first, we will collecting data and normalize them because each feature has a big difference, otherwise it will affect the result. Then, set one parameter is closing price as input.

2. RESEARCH METHODOLOGY

The flow of this research can be seen in Figure 1 below:

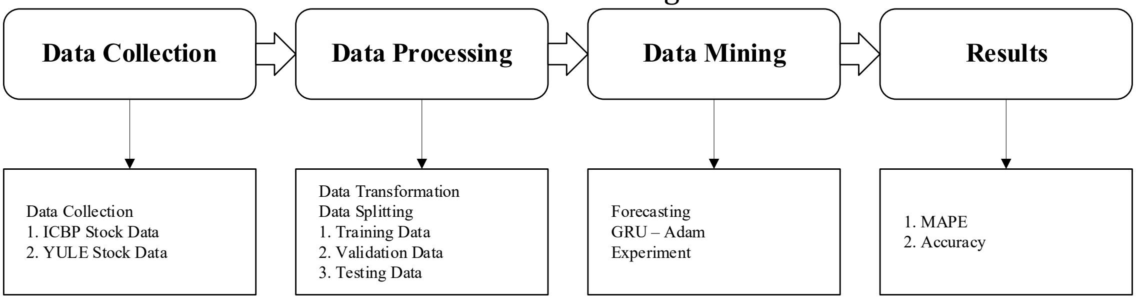

Figure 1. Research flow of stock price predictions

2.1 Data Collection

Data collection is a process for obtaining stock price data that will be used in predictions. This research will use stock closing price data to predict stock prices.

2.2 Data Processing

Data processing is divided into two stages, the first stage is a data transformation and the second stage is data splitting. Data transformation is changing the form of data into many features or time-series data with a certain lag size. Data splitting is the process of splitting data into training data, validation data, and testing data based on predetermined ratios.

2.3 Data Mining

The data mining stage implements predictions using the Gated Recurrent Unit (GRU) method then uses the Adam weight optimization method to get the best weight. Experiments were carried out to get the best parameter values such as lag value, hidden unit and learning rate.

2.3.1 GRU-*Adam* Implementation

The time-series data input of the stock price is one of the stock data which will then be normalized using a formula (1):

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

x' : result of normalization
x : current data value
Min(x) : smallest data value
Max(x) : the largest data value

The process of implementing the Gated Recurrent Unit method in this research is as follows:

a. Initialize weights using normal random initialization of weights. The weights contained in the GRU unit are the gate update weights *(U_z dan W_z)*, reset gate weights *(U_r dan W_r)*, dan activation candidate weights($U_{\tilde{h}}$ dan $W_{\tilde{h}}$). There is also the output weight of the model, namely output weight (V_{out}).

b. The GRU unit also has a bias value, namely bias on the update gate (b_z), bias on the reset gate (b_r) and bias on the activation candidate ($b_{\tilde{h}}$). There is also an output bias from the model which is called output bias (b_{out}).

c. Calculate the update gate value on the time-step with the following formula:
$z_t = \sigma(U_z x_t + W_Z h_{t-1} + b_Z)$ (1)

d. Calculate the gate research value on the timestep with the following equation:
$r_t = \sigma(U_r x_t + W_r h_{t-1} + b_r)$ (2)

e. Calculate activation candidate ($\tilde{h}t$) the following formula:
$\tilde{h} = \tanh(U_{\tilde{h}} x_t + W_{\hat{h}}(h_{t-1} * r_t) + b_{\tilde{h}}$ (3)

f. For each timestep, the activation value (ht) is used as the output of the GRU unit:
$h_t = (1 - z_t) * \tilde{h}_t + z_t * h_{t-1}$ (4)

g. The activation value is then used for the update gate, reset gate, and candidate activation on the next GRU timestep unit.

h. In the last timestep, the output of the GRU unit is multiplied by the weight of the output to produce a predictive value from the model. The predictive value calculation is calculated with the following formula:
$\hat{y} = V_{out} h_2 + b_{out}$ (5)

i. The predicted output value ($\hat{y}$) of the model is compared with the current data (y) with the mean squared error function. This value is called the loss value. The loss value calculated in the example case is as follows:
$loss = J = (y - \hat{y})^2$ (6)

j. Followed by the Adam optimization method. Calculate the estimated gradient
$\nabla_\theta \sum_{i=1}^{m} J(\theta_t \,;\, x^{(i)}, y^{(i)})$ (7)

k. Update the bias first-moment estimate : $m_t = \beta_1 m_{t-1} + (1 - \beta_1) g_t$ (8)

l. Correcting the first-moment of bias $\hat{m}_t = \frac{m_t}{1 - \beta_1^t}$ (9)

m. Correcting the second-moment of bias $\hat{v}_t = \frac{v_t}{1 - \beta_2^t}$ (10)

n. Calculate updates $\theta_{t+1} = \theta_t - \frac{\eta}{\sqrt{\hat{v}_t + \epsilon}} \hat{m}_t$ (11)

o. Apply weight updates $\theta_{(t+1)} = \theta_t - \Delta\theta_{(t+1)}$ (12)

p. Apply the updated weights using the GRU method

q. Done.

The Flowchart of stock price predictions using GRU-Adam method is shown in figure 2.

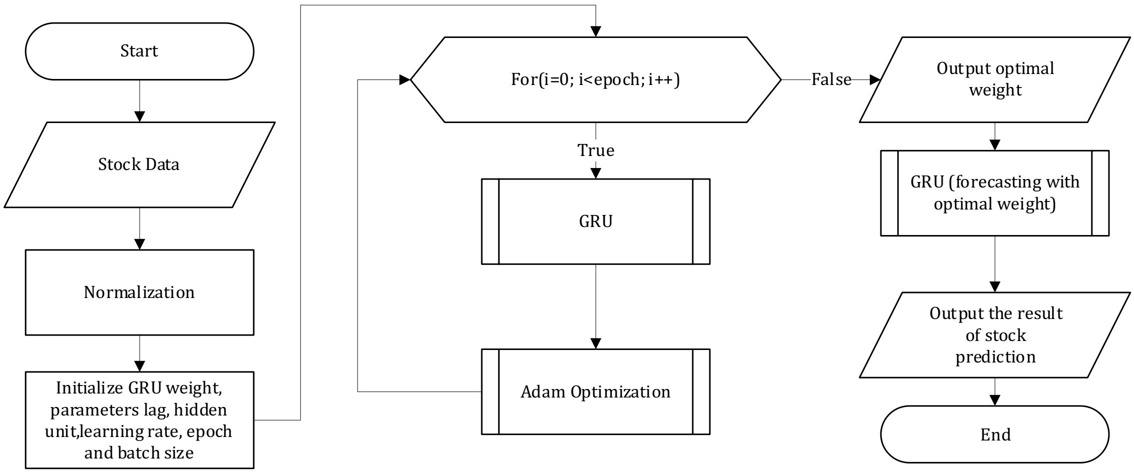

Figure 2. Flowchart GRU-Adam

2.4 Evaluation

The evaluation will be carried out using MAPE calculations to measure the error rate and calculate the accuracy value. Accuracy is obtained from the calculation of 100% minus the MAPE value. Then compare the MAPE results and the accuracy of each method used. Then it can be determined which method is better. If the MAPE is less than 10%, it can be said that the results are accurate.

3. RESULTS AND DISCUSSION

3.1 Results

The data used in this research are PT Indofood CBP Sukses Makmur Tbk (ICBP) stock data from December 2017 to November 2019 and PT. Yulie Sekuritas Indonesia Tbk (YULE) from October 2018 to November 2020 obtained from yahoo finance. Each stock data totals 519 data. In the stock data, the data used are only the closing price (Close) and the date (Date) only.

Table 1. Amount of Data Obtained

No	Stock Data	Range Data	Total Data
1.	ICBP	01 December 2017 - 29 November 2019	519
2.	YULE	26 October 2018 - 17 November 2020	519

Data transformation is carried out in data processing, where the data from the close price column in the sequence is converted into data in the form of sliding windows. Data predicted y, predicted based on the previous n data as much as the specified lag. The process of forming sliding windows is carried out on training data, validation data and testing data. The lag being tested is 1 to 5. Then divide the data into training data, validation data and testing data with a ratio of 70%: 10%: 20%. Here's an example of sharing data with lag sliding windows 2.

Table 2. ICBP training data

Index	Date	Day-2	Day-1	Actual
0	12/5/2017	8450	8750	8650
1	12/6/2017	8750	8650	8850
2	12/7/2017	8650	8850	8675
...	...	...	...	...
360	4/23/2019	9100	9025	9200

Table 3. ICBP validate data

Index	Date	Day-2	Day-1	Actual
361	4/24/2019	9025	9200	9200
362	4/25/2019	9200	9200	9075
363	4/26/2019	9200	9075	9175
...	...	...	...	...
411	7/3/2019	10125	10125	10125

Table 4. ICBP testing data

Index	Date	Day-2	Day-1	Actual
412	7/4/2019	10125	10125	10100
413	7/5/2019	10125	10100	10100
414	7/8/2019	10100	10100	10025
...	...	...	...	...
516	11/29/2019	11375	11325	11325

3.1.1 GRU-Adam Implementation

Data normalization with formula (1) and the results are as in the following table:

Table 5. ICBP training data after normalization

Index	Date	Day-2	Day-1	Actual
0	12/5/2017	0.1257	0.1943	0.1714
1	12/6/2017	0.1943	0.1714	0.2171
2	12/7/2017	0.1714	0.2171	0.1771
...	...	...	...	...
360	4/23/2019	0.2743	0.2571	0.2971

Table 6. ICBP validate data after normalization

Index	Date	Day-2	Day-1	Actual
361	4/24/2019	0.2571	0.2971	0.2971
362	4/25/2019	0.2971	0.2971	0.2686
363	4/26/2019	0.2971	0.2686	0.2914
...	...	...	...	...
411	7/3/2019	0.5086	0.5086	0.5086

Table 7. ICBP testing data after normalization

Index	Date	Day-2	Day-1	Actual
412	7/4/2019	0.5086	0.5086	0.5029
413	7/5/2019	0.5086	0.5029	0.5029
414	7/8/2019	0.5029	0.5029	0.4857
...	...	...	...	...
516	11/29/2019	0.7943	0.7829	0.7829

Experiments were carried out to find the optimal hyper parameter based on validation data. Hyper parameters tested are as in the following table.

Table 8. Parameters test

Parameters test	Parameters
Learning rate	0,1; 0,01; 0,001
lag sliding window	Lag 1 – lag 5
hidden layer	15, 20, 25, 30

Apart from the above parameters, the other parameter values are fixed. This aims to avoid too many parameters being tested. Parameters that are not changed in value are listed in the following table.

Table 9. Fixed parameters

Fixed parameters	Parameters
Maximum Epoch	300
Batch Size	128

The experimental mechanism was carried out in accordance with what was done by [1] by summarizing several steps in determining lag and hidden units and adding experiments to determine learning rates. In order, experiments were carried out on the parameters tested as follows:

a. Conduct experiments to determine the appropriate learning rate using fixed architecture first.
b. Followed by the best sliding windows lag experiment, namely between lag 1 to lag 5. The lag was determined 4 times each experiment with the number of hidden units 15, 20, 25, & 30.
c. After obtaining the best lag parameters, an experiment was conducted to determine the best number of hidden units. Experiments were carried out on hidden units 15, 20, 25, & 30.
d. Apply each of the data with Adam optimization.

The results of the experiments that have been carried out are in the form of the best value learning rate, lag sliding windows, and hidden unit which will be used as parameters in the model that has been created. To determine this value, first the training data and validation data are used. Then it is applied to the testing data so that the comparison of the predicted data with the real data is obtained and the MAPE value is obtained.

Following are the results of the learning rate parameter test using ICBP stock data and Adam's optimization.

Table 10. Best learning rate on the GRU-Adam method ICBP stock data

Learning rate	Best Validation loss	Best Training loss
0.1	0.0031	0.0037
0.01	0.0008	0.0020
0.001	**0.0006**	**0.0018**

The best learning rate on the GRU-Adam method with ICBP stock data is 0.001. Continue with the experiment with the best lag sliding windows and the best hidden unit. Following are the test results for the best sliding windows and hidden unit lag parameters.

Table 11. Best lag and hidden unit on the GRU-Adam method ICBP stock data

Lag/Hidden unit	15	20	25	**30**
1	1.3578	1.4621	1.4428	**0.9676**
2	1.7296	1.3307	1.2494	1.2269
3	1.482	1.6406	1.4468	1.6371
4	1.7765	1.375	1.3325	1.6554
5	2.0255	2.0988	1.295	1.1809

The best lag for sliding windows is lag = 1 and the best hidden unit is 30 with a MAPE of 0.9676%.

Testing on the GRU method with Adam's optimization using the best hyper parameters obtained, namely learning rate = 0.001, lag = 1, hidden units = 30 and fixed parameters such as batch size = 128 and epoch = 300. The following is the training loss and validation loss for each epoch obtained using the GRU-Adam method to get the best loss value, namely training loss = 0.0016 and validation loss = 0.0007.

Table 12. Training loss and validation loss on GRU-Adam

Epoch	Training loss	Validation loss
1	0.0864	0.1519
2	0.0762	0.1354
3	0.0675	0.1194
...	...	...
300	0.0018	0.0008
Best (Minimal)	**0.0016**	**0.0007**

The following are the results of the stock price prediction with the best parameters :

Table 13. The results of the GRU-Adam prediction of ICBP stock data

Date	Actual	Prediction Results
7/8/2019	10025	10083.97
7/9/2019	10100	10009.31
7/10/2019	10125	10083.97
...	...	...
11/29/2019	11325	11304.43

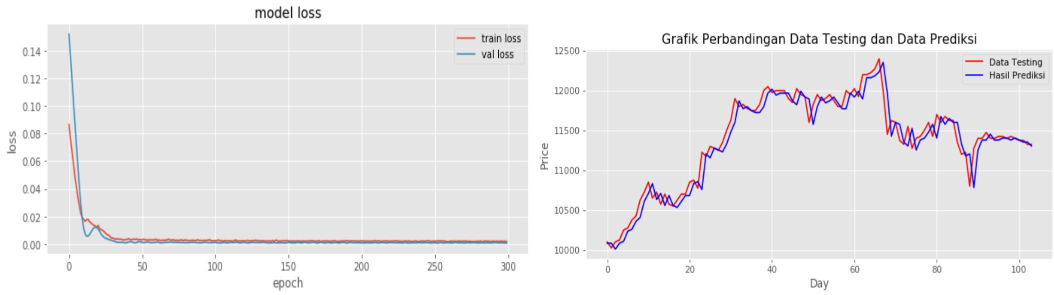

Figure 3. Loss model chart and the prediction result of GRU-Adam ICBP data

The following are the results of parameter testing using the YULE stock data and the GRU-Adam method.

Table 14. Best learning rate on the GRU-Adam method YULE stock data

Learning rate	Best validation loss	Best training loss
0.1	**0.0032**	**0.0051**
0.01	0.0032	0.0052
0.001	0.0032	0.0052

The best learning rate for the GRU-Adam method with YULE stock data is 0.1. Continue with the experiment with the best lag sliding windows and the best hidden unit. Following are the test results for the best sliding windows and hidden unit lag parameters.

Table 15. Best lag and hidden Units in the YULE data GRU-Adam Method

Lag/Hidden unit	15	20	25	**30**
1	4.5396	3.3094	5.0015	**3.0011**
2	3.8905	3.7135	3.7346	4.0537
3	3.9634	4.4402	4.4750	4.6988
4	5.4237	4.0294	9.0665	4.3532
5	6.5267	9.9422	7.0164	7.3641

The best lag for sliding windows is lag = 1 and the best hidden unit is 30 with a MAPE of 3.0011%.

Testing on the GRU method with Adam's optimization using the best hyper parameters obtained, namely learning rate = 0.1, lag = 1, hidden unit = 30 and fixed parameters such as batch size = 128 and epoch = 300. The following is the training loss and validation loss for each epoch obtained using the GRU-Adam method to get the best loss values, namely training loss = 0.0051 and validation loss = 0.0031.

Table 16. Training loss and validation loss on GRU-Adam

Epoch	Training loss	Validation loss
1	0.3575	0.0191
2	0.0388	0.0908
3	0.0677	0.0324
...	...	...
300	0.0056	0.0036
Best (Minimal)	**0.0051**	**0.0031**

The following are the results of the stock price prediction with the best parameters :

Table 17. The results of the GRU-Adam prediction of YULE stock data

Date	Actual	Prediction Results
6/22/2020	156	162.40
6/23/2020	152	155.90
6/24/2020	155	152.42
...	...	...
11/17/2020	328	282.19

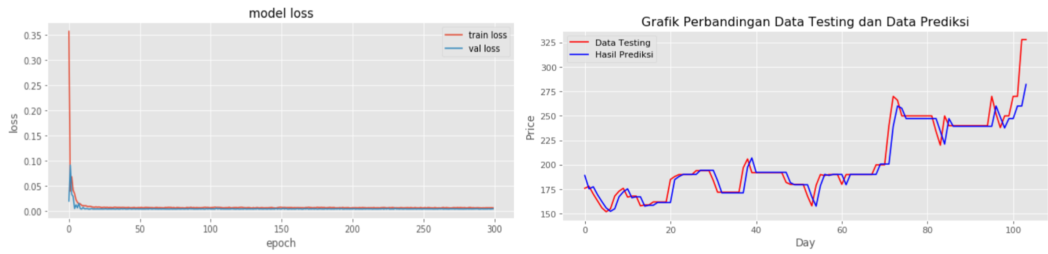

Figure 4. Loss model chart and the prediction result of GRU-Adam YULE data

3.1.2 Calculation of Accuracy

Accuracy can be calculated by the equation 100% minus MAPE and the following results:

Table 18. Accuracy Results

No	Data	MAPE (%)	Accuracy (%)
1	ICBP	0.97	99.03
2	YULE	3.00	97.00

3.2 Discussion

In this research used 2 stock price data those are ICBP data and YULE data with a total of each data is 519 data. The implemented method is GRU-Adam to predict the stock price data. GRU method is used to predict time series data. Adam is used to optimize the weights for high accuracy. The results of the GRU-Adam show accuracy rate more than 90%, therefore this method can be said to have high accuracy.

4. CONCLUSION

The loss function value in the Gated Recurrent Unit (GRU) method with the Adam optimization method is the ICBP data, the train loss values are 0.0016 each and the validation loss is 0.0007. While in the YULE data the train loss value is 0.0051 and the validation loss is 0.00031.

The MAPE obtained from the ICBP data from the GRU-Adam method is 0.97%. The MAPE obtained in the YULE data from the GRU-Adam method is 3.00%.

REFERENCES

[1] Aji, Henridha Wisnu, et al. 2019. *Prediksi Harga Bitcoin Dengan Menggunakan Metode Gated Recurrent Unit (GRU)*. Universitas Lambung Mangkurat.
[2] Aldi, Muhammad Wildan Putra, Jondri dan Annisa Aditsania. 2018. *Analisis dan Implementasi Long Short Term Memory Neural Network untuk Prediksi Harga Bitcoin*. ISSN : 2355-9365 Vol.5, No.2. Universitas Telkom.
[3] Kholish, Tyas N. 2017. *Gated Recurrent Unit – Recurrent Neural Network Untuk Peramalan Nilai Tukar Mata Uang Rupiah Terhadap Dolar Amerika*. Universitas Gadjah Mada.
[4] Mada, Muhammad. 2021. "*Perbandingan Metode Optimasi AdaGrad, RMSprop dan Adam pada Metode Gate Recurrent Unit (GRU) untuk Prediksi Harga Saham*"