21% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

Match Groups

- **94** Not Cited or Quoted 19%
  Matches with neither in-text citation nor quotation marks
- **7** Missing Quotations 1%
  Matches that are still very similar to source material
- **4** Missing Citation 1%
  Matches that have quotation marks, but no in-text citation
- **0** Cited and Quoted 0%
  Matches with in-text citation present, but no quotation marks

Top Sources

- 14% Internet sources
- 15% Publications
- 9% Submitted works (Student Papers)

Match Groups

- **94** Not Cited or Quoted 19%
  Matches with neither in-text citation nor quotation marks
- **7** Missing Quotations 1%
  Matches that are still very similar to source material
- **4** Missing Citation 1%
  Matches that have quotation marks, but no in-text citation
- **0** Cited and Quoted 0%
  Matches with in-text citation present, but no quotation marks

Top Sources

14% Internet sources
15% Publications
9% Submitted works (Student Papers)

Top Sources

The sources with the highest number of matches within the submission. Overlapping sources will not be displayed.

1 Publication
Dhikayla Azizah Putri, Didit Adytia. "Time Series Wave Forecasting with Transfor... **2%**

2 Publication
Eka Briza Erdyana, Ris Yuwono Yudo Nugroho. "Inflation and policy response: A c... **1%**

3 Student papers
Universitas Khairun **1%**

4 Internet
usir.salford.ac.uk **1%**

5 Internet
stat.ethz.ch **1%**

6 Internet
www2.mdpi.com **1%**

7 Internet
journal.universitasbumigora.ac.id **1%**

8 Student papers
Universitas Negeri Semarang - iTh **1%**

9 Student papers
University of York **1%**

10 Student papers
University of Exeter **1%**

11 Publication

Yung-Chen Li, Hsiao-Yun Huang, Nan-Ping Yang, Yi-Hung Kung. "Stock Market For... 1%

12 Publication

"Natural Language Processing and Chinese Computing", Springer Science and Bu... 0%

13 Internet

ijssr.ridwaninstitute.co.id 0%

14 Internet

www.diva-portal.org 0%

15 Internet

ejournal.unkhair.ac.id 0%

16 Internet

www.utupub.fi 0%

17 Publication

Moses Olabhele Esangbedo, Blessing Olamide Taiwo, Hawraa H. Abbas, Shahab H... 0%

18 Internet

journal.uinjkt.ac.id 0%

19 Internet

publikationen.ub.uni-frankfurt.de 0%

20 Internet

jutif.if.unsoed.ac.id 0%

21 Publication

Yiming Dong, Li Liang, Tongze Hou, Xin Ma, Youqing Wang. "A process monitorin... 0%

22 Internet

medikom.iocspublisher.org 0%

23 Internet

www.ijain.org 0%

24 Student papers

De Montfort University 0%

25	Student papers	
University of East London		0%
26	Internet	
journal.nurulfikri.ac.id		0%
27	Internet	
ebin.pub		0%
28	Internet	
journal.tarbiyahiainib.ac.id		0%
29	Internet	
jurnal.fikom.umi.ac.id		0%
30	Internet	
mdpi-res.com		0%
31	Publication	
Hassan Anjum, Farooq Malik. "Forecasting Risk in the US Dollar Exchange Rate un...		0%
32	Publication	
Xin Ma, Zhanzhan Liu, Mingxing Zheng, Youqing Wang. "Application and explorat...		0%
33	Internet	
www.stiemuttaqien.ac.id		0%
34	Publication	
"Database and Expert Systems Applications", Springer Science and Business Medi...		0%
35	Internet	
journal.stmikjayakarta.ac.id		0%
36	Student papers	
Gisma University of Applied Sciences GmbH		0%
37	Internet	
link.springer.com		0%
38	Student papers	
Corvinus University of Budapest		0%

39 Internet

www.geeksforgeeks.org 0%

40 Internet

www.atlantis-press.com 0%

41 Student papers

Al Akhawayn University in Ifrane 0%

42 Publication

I Made Wirawan, Aji Prasetya Wibawa, Triyanna Widiyanintyas. "Photovoltaic Ene... 0%

43 Internet

www.scribd.com 0%

44 Publication

Mukhlis Fadhli, Indra Jaya, Sri Nurdianti, Karlisa Priandana, Irman Hermadi. "Inte... 0%

45 Internet

www.frontiersin.org 0%

46 Publication

Eugeniu Strelet, You Peng, Ivan Castillo, Ricardo Rendall, Zhenyu Wang, Mark Jos... 0%

47 Publication

Karimi, Y.. "Application of support vector machine technology for weed and nitro... 0%

48 Publication

Xiaofeng Zhou, Andri Pranolo, Yingchi Mao. "AB-LSTM: Attention Bidirectional Lo... 0%

49 Publication

Furizal Furizal, Aldi Bastiatul Fawait, Hari Maghfiroh, Alfian Ma'arif, Asno Azzawa... 0%

50 Publication

Gaddadevara Matt Siddesh, Ganesh Chandra Deka, Krishnarajanagar GopalaIyen... 0%

51 Publication

Zeshan Ali Ali. "Research Chinese-Urdu Machine Translation Based on Deep Learn... 0%

52 Internet

gdic.unja.ac.id 0%

JIKO (Jurnal Informatika dan Komputer) Accredited KEMENDIKBUD RISTEK, No.105/E/KPT/2022
Vol. x, No. x, April 2024, pp. x-x p-ISSN: 2614-8897
DOI: 10.33387/jiko e-ISSN: 2656-1948

TRANSFORMER WITH LAGGED FEATURES FOR HANDLING LONG-TERM DATA DEPENDENCY IN TIME SERIES FORECASTING

Eko Verianto[1], Annisa Fikria Shimbun[2]

[1]Program Studi Sistem Informasi, Fakultas Sains dan Komunikasi Visual, Universitas Cendekia Mitra Indonesia
[2]Program Studi Informatika, Fakultas Sains dan Komunikasi Visual, Universitas Cendekia Mitra Indonesia
[*]Email: [1]eko@unicimi.ac.id, [2]niesashimbun@unicimi.ac.id

(Received: dd mmm yyyy, Revised: dd mmm yyyy, Accepted: dd mmm yyyy)

Abstract (10pt)

Data with long-term dependencies play an important role in time series forecasting, but learning data with long-term dependencies in time series data is a constraint for most algorithms. Some of these algorithms can forecast time series data, but not all algorithms can model data that has long-term dependencies in a better way. The algorithm that is usually used to perform the task of forecasting data with long-term dependencies is Long Short-Term Memory (LSTM). This algorithm uses a recurrent structure that is built and designed to overcome the problems of vanishing gradients and exploding gradients. However, LSTM can still experience vanishing gradient problems and as a result will have difficulty identifying long-term dependencies in data sets that are too long. The purpose of this research is to develop an effective solution in forecasting time series data with long-term dependencies. The approach proposed in this research is a transformer with lagged features and uses a time series cross-validation validation technique. The results of this research indicate that the transformer model is not optimal in capturing long-term temporal patterns in the data, so that transformers with lagged features have not been able to provide their best performance in forecasting long-term time series data.

Keywords: *Time Series Forecasting, Transformer, Self-Attention, Time Series Cross-Validation, Lagged Features*

**Corresponding Author:* Author1*

1. INTRODUCTION

The development of technology in the last few decades has triggered a digital revolution that affects every aspect of human life. Digital development is basically dependent on internet technology [1]. Today's internet technology has changed the economy in various parts of the world. This means that the internet can make products and services reach a wider market share. This expansion of market share also has an impact on the increasing volume of data known as big data. One of the causes of the increasing volume of data is the phenomenon of data being changed continuously, even changes can occur in a short time interval.

Currently, data is the main commodity for business people and stakeholders in decision making. The existence of data is often referred to as oil in the 21st century [2]. Data has a very beneficial impact on business development and progress. One of the benefits of data that can be felt is the ability to make more precise and effective decisions, this includes how data is analyzed and modeled for certain purposes such as forecasting.

Forecasting is defined as estimating future information using past data [3]. Forecasting is an important part of running a business because it is able to predict future trends and events accurately and is useful in many contexts, including business management [4].

Forecasting activities cannot be separated from past data, this data is the main raw material in the modeling process. The data currently available tends to be more complex and has different characteristics, one example is data with long-term trends. Data like this usually has long-term dependencies that play an important role in time series forecasting [5].

Learning data with long-term dependencies in time series data is a constraint for most algorithms [5]. Some algorithms can forecast time series data, but not all algorithms can model data with long-term dependencies in a better way. This will certainly affect the quality of the model which tends to be unable to capture long-term patterns in time series data, so that decision making will be difficult to make considering the poor quality of the model in predicting trends and changes in long-term data.

The algorithm that is usually used to forecast data with long-term dependencies is Long Short-Term Memory (LSTM). This algorithm uses a recurrent structure that is built and designed to overcome the problems of vanishing gradients and exploding gradients [3]. However, LSTM can still experience the problem of vanishing gradients and as a result will have difficulty identifying long-term dependencies in data sets that are too long [6].

Previous research proposed a combination of fractional differentiation parameter (and/or hurst parameter) estimation methods with recurrent conditional networks to learn and predict long-term dependencies on information. This research evaluated four different architectures, including simple RNN, LSTM, BiLLSTM and GRU. The results of this research showed that accurate prediction results can be made one step ahead of the long-term memory parameters, especially the BiLLSTM network which obtained the best results when using the proposed methodology, but there are still obstacles such as the recurrent neural network that fails to capture points that are very far away [5].

Other studies describe the inefficiency of LSTM in handling long-term dependencies due to the problem of vanishing gradients in LSTM networks. This research provides an empirical analysis using a case research of NASA turbofan engine degradation. This research shows that the longer the input sequence, the more difficult it is for the LSTM model to remember all relevant information [6].

Referring to the problems of previous studies, the main focus of this research is how to develop effective solutions and build a forecasting model for time series data with long-term dependencies. Based on this, an approach is proposed that can overcome data with long-term dependencies. The approach proposed in this research is a transformer with lagged features and time series cross-validation. Transformers with self-attention mechanisms can analyze internal characteristics of data well and focus on important information globally and locally [7]. Based on this, the self-attention mechanism is suitable for use in forecasting problems for time series data with long-term trends. This research also proposes lagged features that are used to capture temporal dependencies in data and also proposes time series cross-validation that is used to maintain the temporal order of data.

2. RESEARCH METHOD

This research method uses a quantitative approach with an experimental design. Experimental research provides an opportunity for researchers to directly influence research variables and is the only type of research that can test hypotheses about causal relationships [8]. The method in this research is structured in the form of structured, systematic, planned and clear stages. The method in this research is presented in the diagram in Figure 1 below:

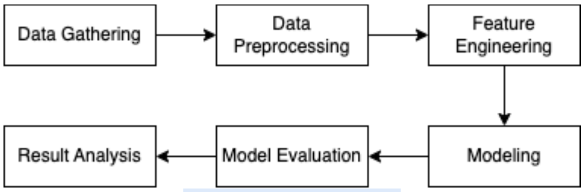

Figure 1. Research stages

2.1 Data Gathering

The first stage in this research is data gathering. Data gathering or data collection is a technique that can be used by researchers to collect data [9]. This research uses internet search techniques to collect data. The data used in this research is a financial dataset consisting of historical data on BBCA stock closing prices and historical data on the closing price of the US Dollar exchange rate against the Indonesian Rupiah. Each of these datasets has a different trend. Figure 2 shows the long-term trend of BBCA stock prices over a period of 10 years which tends to increase.

Figure 2. BBCA stock closing price chart

In contrast to the long-term trend of BBCA stock prices, the trend of the closing price of the US Dollar exchange rate against the Indonesian Rupiah tends to be more volatile, although there are indications that it will increase. Figure 3 shows the trend of the closing price of the US Dollar exchange rate against the Indonesian Rupiah over 10 years.

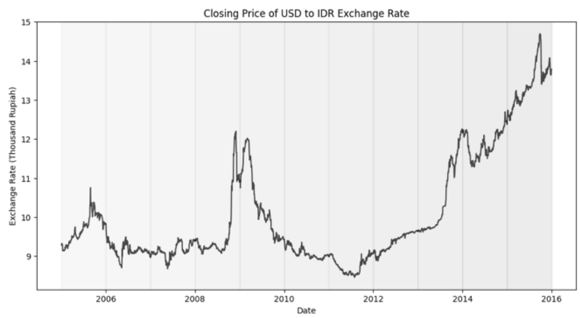

Figure 3. USD/IDR exchange rate closing price chart

2.2 Data Preprocessing

The second stage in this research is data preprocessing. Data preprocessing is the initial data processing process used to change raw data obtained from various sources into cleaner information that can be used for further analysis [10]. Data preprocessing in this research consists of data cleaning and data adjustment to needs. At this stage, data is sorted according to date order, management of data duplication and also filling in data gaps using the linear interpolation method. Linear interpolation is an interpolation that assumes values form a straight line [11]. The linear interpolation equation for time series data is as in equation 1.

$$y = y_0 + \frac{y_1 - y_0}{t_1 - t_0} \times (t - t_0) \qquad (1)$$

The value to be searched is expressed by y, y_0 is the known value at time t_0, y_1 is the known value at time t_1. t is the time of the interpolated value y, t_0 is the time of the interpolated value y_0 and t_1 is the time of the interpolated value y_1.

2.3 Feature Engineering

The third stage of this research is feature engineering. Feature engineering is the process of extracting features from raw data and converting them into a format that is in accordance with the machine learning model [12]. At this stage, temporal feature extraction is carried out using lagged features. Lagged features are features that contain data from the previous time step [13]. Figure 4 is an example of the application of lagged features for univariate time series data on BBCA closing stock price data.

Date	*Closing Price*
05/01/15	2640
06/01/15	2620
07/01/15	2625
08/01/15	2595
09/01/15	2585
12/01/15	2560

→

t-1	t-2	t-3
NaN	NaN	NaN
2640	NaN	NaN
2620	2640	NaN
2625	2620	2640
2595	2625	2620
2585	2595	2625

Figure 4. Illustration of Lagged Features in BBCA Stock Data

The illustration of lagged features in Figure 4 uses a lag size of three, this shows the stock prices on the previous day (t-1), two days before (t-2) and three days before (t-3). These lagged features are used to help the model learn temporal patterns in the data, so that it can improve prediction accuracy.

2.4 Modeling

The fourth stage in this research is modeling. This stage focuses on the development of a transformer model designed to overcome long-term data dependence on time series data forecasting. At this stage, the model is designed using a transformer architecture with lagged features. The use of lagged features in the transformer architecture aims to strengthen the model's ability to capture temporal patterns from the data. Figure 5 shows the transformer architecture with lagged features.

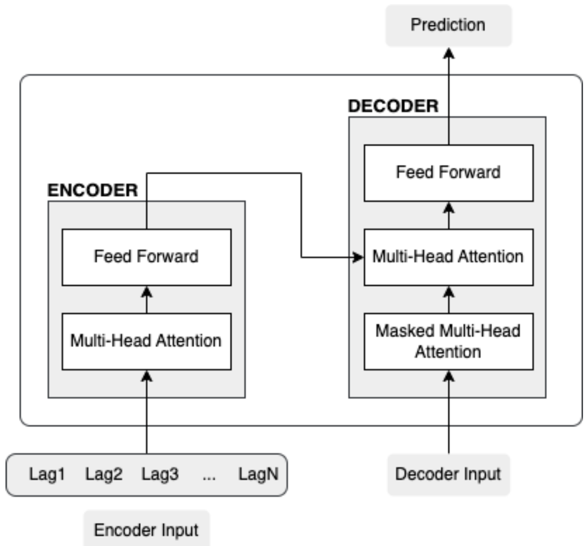

Figure 5. Arsitektur Transformer dengan *Lagged Features*

Figure 5 shows that the encoder section of the transformer architecture receives input in the form of past values of time series data expressed in lags. Furthermore, the encoder uses a self-attention mechanism to capture the temporal relationship between lags in the input sequence. Then it will be continued with a feed-forward process to strengthen the feature representation. Masked multi-head attention is used to prevent future information leakage, where only previous values are used. The output of the encoder will then be processed in the multi-head attention on the decoder (cross-attention). In this section, all information from the encoder output and the decoder will be processed to be able to capture the relationship of each data. After going through the cross-attention stage, the last step is the feed-forward process to learn the non-linear relationship between data.

2.5 Model Evaluation

The fifth stage of this research is model evaluation. This research applies time series cross-validation to evaluate the performance of the forecasting model. Time series cross-validation focuses on data division that maintains its temporal order. The characteristic of time series cross-validation is that the validation sample consists of sequential observations [14]. It works by ensuring that the training data only covers information available at a certain point in time, while the testing data is taken from the future after the training data. Figure 6 shows how time series cross-validation works.

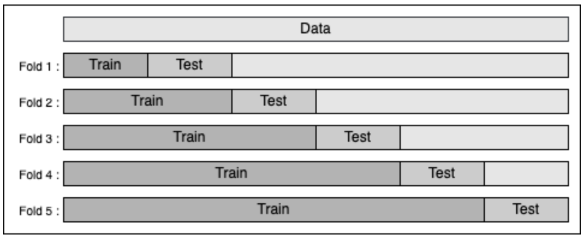

Figure 6. Illustration of how time series cross-validation works

Figure 6 shows how time series cross-validation works. Each fold involves training data and testing data. The subset of training data is gradually expanded by incorporating previous data. By implementing time series cross-validation like this, each fold preserves the temporal order of the data.

At this stage, an evaluation will be carried out on each model. Model evaluation is carried out by measuring how well a model performs in making predictions. Model performance evaluation is carried out using the Mean Absolute Percentage Error (MAPE). MAPE is used to measure the average prediction error proportionally to the actual value [15]. The mean absolute percentage error equation is as shown in equation 2.

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{\hat{y}_i}\right| \quad (2)$$

Equation 2 consists of the absolute value sigma obtained through the difference between the actual value y_i and the predicted value $\hat{y}$ then divided by the predicted value $\hat{y}$. This value will then be multiplied by one hundred percent.

2.5 Result Analysis

The final stage of this research is the analysis of the results. This research focuses on the analysis of model performance in predicting long-term time series data. Observations focus on evaluation metrics to measure the extent to which the model is able to predict values at each future time step. Comparison between the prediction results of the transformer model and the comparison model provides more depth regarding the advantages and limitations of the model in handling long-term time series data. Visualization of prediction graphs and error rate graphs can also provide an understanding of the error patterns that occur in the model.

This analysis stage aims to provide a clearer picture of the ability and stability of the transformer model in predicting data movement. The results of this analysis are also used to identify potential improvements and further developments in the transformer model to handle long-term time series data.

3. RESULT AND DISCUSSION

3. 1 Dataset Processing

The dataset in this research consists of historical BBCA stock price data from January 2005 to December 2015 with daily frequency. The file format of the data used in this research is comma-separated values (csv). Figure 7 shows a sample of BBCA stock price data used in this research.

1	Date	Open	High	Low	Close	Adj Close	Volume
2	3-Jan-05	295	295	292.5	295	198.73	126,180,000
3	4-Jan-05	295	302.5	295	300	202.1	291,730,000
4	5-Jan-05	300	315	300	307.5	207.15	564,730,000
5	6-Jan-05	310	310	302.5	307.5	207.15	161,150,000
6	7-Jan-05	305	305	292.5	295	198.73	592,910,000

Figure 7. BBCA stock data sample

In addition to the closing price of BBCA shares, this research also uses historical data on the closing price of the US Dollar exchange rate against the Indonesian Rupiah from January 2005 to December 2015 with a daily frequency. Figure 8 shows a sample of US Dollar exchange rate data against the Indonesian Rupiah.

1	Date	Close	Open	High	Low	Vol.	PercentageChange%
2	3-Jan-05	9.283,0	9.325,0	9.335,0	9.275,0		0,01%
3	4-Jan-05	9.300,0	9.300,0	9.310,0	9.265,0		0,18%
4	5-Jan-05	9.310,0	9.310,0	9.325,0	9.285,0		0,11%
5	6-Jan-05	9.327,5	9.292,5	9.327,5	9.264,5		0,19%
6	7-Jan-05	9.310,0	9.326,0	9.326,0	9.285,0		-0,19%

Figure 8. Sample data of the exchange rate of the United States Dollar against the Indonesian Rupiah

The two datasets used will then be processed to select the values used for modeling. The values used for modeling are the closing prices on each dataset. Figure 9 shows a sample vector of closing prices on the BBCA stock dataset and the US Dollar exchange rate against the Indonesian Rupiah.

```
Closing Stock Price [BBCA]:
0    295.0
1    300.0
2    307.5
3    307.5
4    295.0
Name: Close, dtype: float64
Currency Exchange Rate [USDIDR]:
0    9.2830
1    9.3000
2    9.3100
3    9.3275
4    9.3100
```

Figure 9. Sample vector data of BBCA stock closing and US Dollar exchange rate against Indonesian Rupiah

After performing feature selection on both datasets, the next step is to perform data preprocessing. One of the activities in data preprocessing is data imputation. The technique used to perform data imputation is linear interpolation. Linear interpolation will fill in the gaps in values by estimating the values between two known data points, the linear interpolation calculation process in equation 1.

The next step is the data preprocessing stage, which is the feature engineering stage. At this stage, the data will be adjusted for modeling needs, such as determining the lag and target shown in Figure 10.

```
Closing Stock Price [BBCA]
Lags [n=5]:
[[295.   307.5 307.5 300.   295. ]
 [282.5 295.   307.5 307.5 300. ]
 [282.5 282.5 295.   307.5 307.5]
 [277.5 282.5 282.5 295.   307.5]
 [280.   277.5 282.5 282.5 295. ]]
Target:
[282.5 282.5 277.5 280.   282.5]
Currency Exchange Rate [USDIDR]
Lags [n=5]:
[[9.31   9.3275 9.31   9.3    9.283 ]
 [9.162  9.31   9.3275 9.31   9.3   ]
 [9.1415 9.162  9.31   9.3275 9.31  ]
 [9.195  9.1415 9.162  9.31   9.3275]
 [9.2875 9.195  9.1415 9.162  9.31  ]]
Target:
[9.162  9.1415 9.195  9.2875 9.3025]
```

Figure 10. Sample lag and target data for BBCA stock closing and the exchange rate of the US Dollar against the Indonesian Rupiah

Figure 10 shows historical data on BBCA stock closing prices and historical data on the closing prices of the US Dollar exchange rate against the Indonesian Rupiah in lag and target form.

3. 2 Modeling and Evaluation

The model is developed with data training. The data used in the training process is training data with hyperparameter variations to find the best model. The first experiment conducted in this research was an experiment on determining the number of lags and folds. This experiment uses variations in the number of folds found in time series cross-validation and variations in lags found in lagged features.

Table 1 shows the results of the lag experiment on the BBCA dataset and the US Dollar exchange rate against the Indonesian Rupiah. This experiment aims to observe how much influence the variation in the amount of lag has on the error rate and also the execution time.

Table 1. Lag Experiments on Transformer Models

Dataset	Number of Folds	Number of Lags	MAPE	Execution Time (*s*)
BBCA Stock Price	5	5	0,0381	3312
		10	0,0400	3427
		15	0,0417	3815
		20	0,0574	3847
Currency Exchange Rates (USD/IDR)	5	5	0,0335	3506
		10	0,0536	3677
		15	0,0500	3836
		20	0,0546	3822

The lag experiment in Table 1 was evaluated using MAPE. The MAPE results show that increasing lag does not always affect the error rate. The results also show that the dominant use of lag=5 produces a minimum error rate. This shows that the use of lag with a small size has quite good relevance in model formation. This experiment also shows that the number of lags can affect the execution time of the model, where the greater the number of lags, the potential to increase the execution time.

This research also conducted experiments on variations in the number of folds on both datasets. The purpose of this experiment was to observe how much influence the variation in the number of folds had on the error rate and execution time produced. Table 2 shows the results of the experiment on the number of folds measured using MAPE.

Table 2. Fold Experiment on Transformer Model

Dataset	Number of Lag	Number of Folds	MAPE	Execution Time (*s*)
BBCA Stock Price	5	5	0,0381	3312
		10	0,0471	6781
		15	0,0518	10521
		20	0,0402	13581
Currency Exchange Rates (USD/IDR)	5	5	0,0335	3506
		10	0,0321	7019
		15	0,0285	10598
		20	0,0364	13202

Table 2 shows that increasing the number of folds does not always give good results on the error rate, this shows that the number of folds does not really affect the generalization ability of the model, while the execution time always increases with the increasing number of folds. Observations on the influence of variations in the number of lags and folds on the error rate and execution time, then variations of both were made as shown in Table 3.

Table 3. Fold and Lag Experiments on Transformer Model

Dataset	Number of Folds	Number of Lags	Avg. MAPE Across all Folds	Execution Time (*s*)
BBCA Stock Price	5	5	0,0381	3312
		10	0,0400	3427
		15	0,0417	3815
		20	0,0574	3847
	10	5	0,0411	6599
		10	0,0537	6943
		15	0,0484	7512
		20	0,0544	7897
	15	5	0,0518	10521
		10	0,0506	10636
		15	0,0416	10960
		20	0,0525	10985
	20	5	0,0402	13581
		10	0,0520	14483
		15	0,0559	15418
		20	0,0536	14679
Currency Exchange Rates (USD/IDR)	5	5	0,0335	3506
		10	0,0536	3677
		15	0,0500	3836
		20	0,0546	3822
	10	5	0,0338	6798
		10	0,0540	7398
		15	0,0556	7898
		20	0,0512	7898
	15	5	0,0285	10598
		10	0,0610	14583
		15	0,0568	15463
		20	0,0506	11380
	20	5	0,0364	13202
		10	0,0587	14716
		15	0,0530	17317
		20	0,0600	14959

Table 3 shows the fold and lag experiments on the BBCA stock dataset and the US Dollar exchange rate dataset against the Indonesian Rupiah. The BBCA stock dataset shows that the effect of fold and lag on the error rate varies and increasing the fold and lag values does not always affect the error rate, while a smaller number of folds and lags can provide better results.

The US Dollar exchange rate dataset against the Indonesian Rupiah shows varying results. Increasing folds and lags also does not always affect the error rate, but increasing the number of folds and lags will increase the execution time. These results also show that the number of folds and lags can produce a better error rate. The experimental results in table 3 also provide information that larger variations in folds and lags will result in longer execution times.

The next experiment is to compare the performance of the transformer model and the LSTM model per fold. This test uses both datasets with all fold variations in the previous test, while for the lag variation only uses the lag value that produces the minimum error rate in Table 3. Table 4 shows comparison of the performance of the transformer model and the LSTM model on the BBCA stock dataset.

Table 4. Comparison of Transformer Model and LSTM Model Performance on BBCA Stock Dataset with fold=5 and lag-5

Fold	MAPE per Fold (Transformer Model)	MAPE per Fold (LSTM Model)
Fold 1	0,0390	0,0190
Fold 2	0,0329	0,0226
Fold 3	0,0207	0,0148
Fold 4	0,0554	0,0148
Fold 5	0,0423	0,0111

Table 4 shows comparison of the performance of the transformer model and the LSTM model with the configuration of fold=5 and lag=5. Figure 11 shows comparison of the MAPE per fold of the transformer model and the LSTM model in a graph.

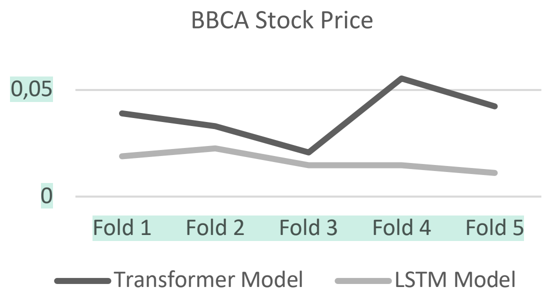

Figure 11. MAPE comparison chart on BBCA shares with fold=5 and lag=5

Based on observations in the graph Figure 11 shows that all folds in the LSTM model have a smaller error rate compared to the transformer model. The transformer model shows fluctuating pattern, while the LSTM model shows tendency to decrease the error rate. Further testing was conducted using the US Dollar exchange rate dataset against the Indonesian Rupiah. The results of the testing are shown in Table 5.

Table 5. Comparison of Transformer Model and LSTM Model Performance on USD/IDR Currency Exchange Rate Dataset with fold=5 and lag=5

Fold	MAPE per Fold (Transformer Model)	MAPE per Fold (LSTM Model)
Fold 1	0,0273	0,0043
Fold 2	0,0431	0,0065
Fold 3	0,0498	0,0025
Fold 4	0,0236	0,0035
Fold 5	0,0237	0,0073

Table 5 shows the performance comparison of the transformer model and the LSTM model with the configuration of fold=5 and lag=5. Figure 12 shows the comparison of MAPE per fold in the transformer model and the LSTM model in a graph.

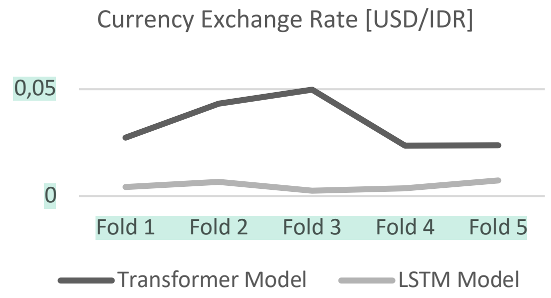

Figure 12. MAPE comparison chart on USD/IDR currency exchange rate dataset with fold=5 and lag=5

Based on the observation of the graph shown in Figure 12, the error rate in both models shows tendency of fluctuating patterns, but the LSTM model has a difference in error rate that is not so significant when compared to the transformer model. Further testing is carried out by increasing the number of folds as shown in Table 6.

Table 6. Comparison of Transformer Model and LSTM Model Performance on BBCA Stock Dataset with fold=10 and lag=5

Fold	MAPE per Fold (Transformer Model)	MAPE per Fold (LSTM Model)
Fold 1	0,1023	0,0214
Fold 2	0,0225	0,0183
Fold 3	0,0697	0,0299
Fold 4	0,0593	0,0204
Fold 5	0,0284	0,0210
Fold 6	0,0299	0,0182
Fold 7	0,0137	0,0114
Fold 8	0,0261	0,0163
Fold 9	0,0366	0,0144
Fold 10	0,0229	0,0126

Table 6 shows the performance comparison of the transformer model and the LSTM model with the configuration of fold=10 and lag=5. Figure 13 shows the comparison graph of MAPE per fold in the transformer model and the LSTM model.

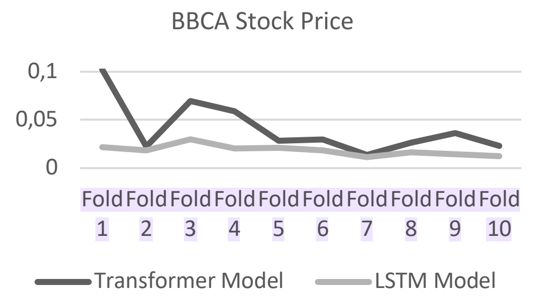

Figure 13. Grafik perbandingan MAPE pada saham BBCA dengan fold 10 dan lag=5

Based on the observations made on the graph shown in Figure 13, the transformer model and also the LSTM model show a tendency to decrease the error rate. Further testing was carried out on the US Dollar exchange rate dataset against the Indonesian Rupiah as shown in Table 7.

Table 7. Comparison of Transformer Model and LSTM Model Performance on USD/IDR Currency Exchange Rate Dataset with fold=10 and lag=5

Fold	MAPE per Fold (Transformer Model)	MAPE per Fold (LSTM Model)
Fold 1	0,0232	0,0115
Fold 2	0,0111	0,0038
Fold 3	0,0468	0,0054
Fold 4	0,0266	0,0064
Fold 5	0,0529	0,0026
Fold 6	0,0684	0,0029
Fold 7	0,0267	0,0032
Fold 8	0,0570	0,0032
Fold 9	0,0178	0,0036
Fold 10	0,0075	0,0081

Table 7 shows performance comparison of the transformer model and the LSTM model with the configuration of fold=10 and lag=5. The graph in Figure 14 is a comparison of MAPE per fold in the transformer model and the LSTM model.

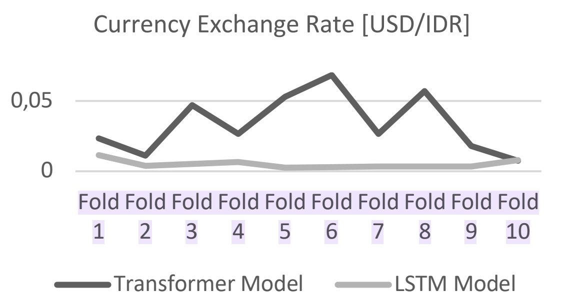

Figure 14. MAPE comparison chart on USD/IDR currency exchange rate dataset with fold=10 and lag=5

Based on the graph shown in Figure 14, the transformer model and also the LSTM model show a fluctuating pattern in the error rate, although both show a fluctuating pattern, the LSTM model does not show a significant change when compared to the transformer model. In further testing, the number of folds will be increased as shown in Table 8.

Table 8. Comparison of Transformer Model and LSTM Model Performance on BBCA Stock Dataset with fold=15 and lag=15

Fold	MAPE per Fold (Transformer Model)	MAPE per Fold (LSTM Model)
Fold 1	0,0535	0,0263
Fold 2	0,0882	0,0260
Fold 3	0,0297	0,0274
Fold 4	0,0496	0,0336
Fold 5	0,0715	0,0525
Fold 6	0,0459	0,0281
Fold 7	0,0432	0,0188
Fold 8	0,0321	0,0165
Fold 9	0,0376	0,0171
Fold 10	0,0429	0,0293
Fold 11	0,0189	0,0258
Fold 12	0,0379	0,0168
Fold 13	0,0300	0,0158
Fold 14	0,0180	0,0127
Fold 15	0,0255	0,0140

Table 8 shows comparison of the performance of the transformer model and the LSTM model with the configuration of fold=15 and lag=15. The performance comparison graph of the transformer model and the LSTM model is shown in Figure 15.

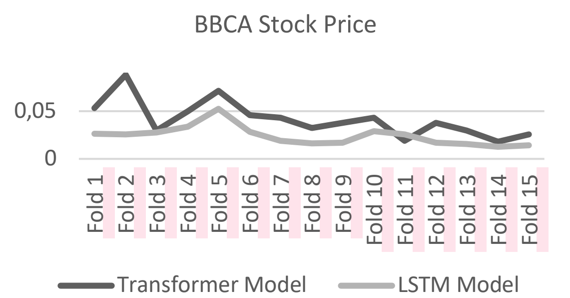

Figure 15. MAPE comparison chart on BBCA stock dataset with fold=15 and lag=15

The graph in Figure 15 shows that the error rate in both models tends to decrease. Furthermore, testing was carried out on the US Dollar exchange rate dataset against the Indonesian Rupiah as shown in Table 9.

Table 9. Comparison of Transformer Model and LSTM Model Performance on USD/IDR Currency Exchange Rate Dataset with fold=15 and lag=5

Fold	MAPE per Fold (Transformer Model)	MAPE per Fold (LSTM Model)
Fold 1	0,0165	0,0062
Fold 2	0,0069	0,0086
Fold 3	0,0073	0,0042
Fold 4	0,0070	0,0040
Fold 5	0,0542	0,0080
Fold 6	0,0185	0,0076
Fold 7	0,0462	0,0033
Fold 8	0,0349	0,0014
Fold 9	0,0532	0,0040
Fold 10	0,0396	0,0050
Fold 11	0,0569	0,0011
Fold 12	0,0208	0,0037
Fold 13	0,0157	0,0040
Fold 14	0,0328	0,0045
Fold 15	0,0176	0,0052

Table 9 is a comparison of the performance of the transformer model and the LSTM model with the configuration of fold=15 and lag=15. The performance comparison graph of the transformer model and the LSTM model is as shown in Figure 16.

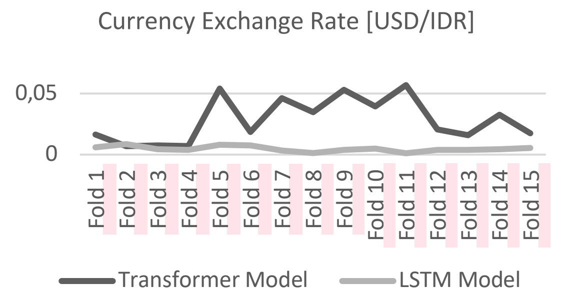

Figure 16. MAPE comparison chart on USD/IDR currency exchange rate dataset with fold=15 and lag=5

Based on the graph in Figure 16, both models show a fluctuating pattern but the change in error rate is not so significant in the LSTM model, this is inversely proportional to the transformer model. The final test in this research will increase the number of folds again as shown in Table 10.

Table 10. Comparison of Transformer Model and LSTM Model Performance on BBCA Stock Dataset with fold=20 and lag=5

Fold	MAPE per Fold (Transformer Model)	MAPE per Fold (LSTM Model)
Fold 1	0,0198	0,0154
Fold 2	0,1276	0,0184
Fold 3	0,0299	0,0131
Fold 4	0,0185	0,0154
Fold 5	0,0345	0,0268
Fold 6	0,0402	0,0266
Fold 7	0,0536	0,0346
Fold 8	0,0388	0,0200
Fold 9	0,0503	0,0178
Fold 10	0,0419	0,0171
Fold 11	0,0272	0,0217
Fold 12	0,0303	0,0190
Fold 13	0,0151	0,0105
Fold 14	0,0253	0,0110
Fold 15	0,0597	0,0171
Fold 16	0,0276	0,0192
Fold 17	0,0233	0,0121
Fold 18	0,0398	0,0111
Fold 19	0,0584	0,0128
Fold 20	0,0429	0,0204

Table 10 shows comparison of the performance of the transformer model and also the LSTM model with the configuration of fold=20 and lag=5. The graphical presentation of the table is shown in Figure 17.

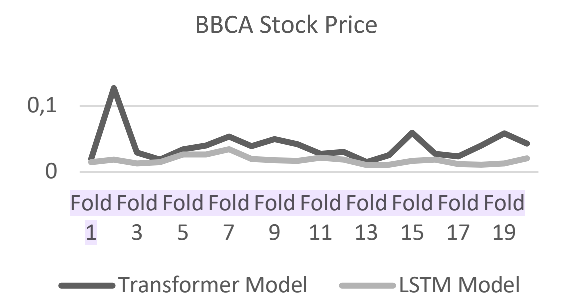

Figure 17. MAPE comparison chart on BBCA stock dataset with fold=20 and lag=5

Figure 17 shows that the error rate in both models has a fluctuating pattern, only the LSTM model has a difference in error rate that is not so significant when compared to the transformer model. Further testing uses the US Dollar exchange rate dataset against the Indonesian Rupiah as shown in Table 11.

Table 11. Comparison of Transformer Model and LSTM Model Performance on USD/IDR Currency Exchange Rate Dataset with fold=20 and lag=5

Fold	MAPE per Fold (Transformer Model)	MAPE per Fold (LSTM Model)
Fold 1	0,0203	0,0083
Fold 2	0,0200	0,0074
Fold 3	0,0482	0,0051
Fold 4	0,0332	0,0111
Fold 5	0,0176	0,0045
Fold 6	0,0149	0,0023
Fold 7	0,0681	0,0101
Fold 8	0,0144	0,0081
Fold 9	0,0275	0,0054
Fold 10	0,0274	0,0025
Fold 11	0,0443	0,0016
Fold 12	0,0728	0,0064
Fold 13	0,0651	0,0067
Fold 14	0,0436	0,0035
Fold 15	0,0484	0,0029
Fold 16	0,0171	0,0049
Fold 17	0,0404	0,0107
Fold 18	0,0246	0,0064
Fold 19	0,0588	0,0055
Fold 20	0,0208	0,0042

Table 11 shows comparison of performance on the transformer model and also the LSTM model. The configuration of both models also uses fold=20 and lag=5. The graph of the table is shown in Figure 18.

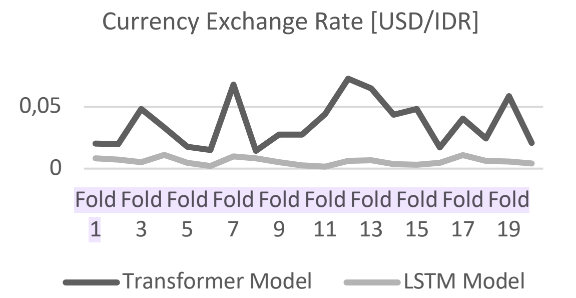

Figure 18. MAPE comparison chart on USD/IDR currency exchange rate dataset with fold=20 and lag=5

Based on the experiments that have been conducted on the best models of all fold variations. The transformer model does not show good performance when compared to the LSTM model in all experiments. The LSTM model has a more consistent performance with a relatively stable error rate in almost all folds, while the transformer model shows fluctuating pattern with significant changes, this indicates that the transformer model is not good enough in capturing temporal patterns in long-term data.

This is because the basic architecture of the transformer relies more on self-attention. Self-attention is an attention mechanism found in the basic architecture of the transformer. The self-attention mechanism tends to be less effective in handling data with long temporal dependencies and without any components that can be explicitly used to handle sequential processing such as those in LSTM. LSTM has a mechanism that allows the model to remember important information and ignore less relevant information.

The following will show the results of the best model predictions per fold on the BBCA stock dataset and the US Dollar exchange rate against the Indonesian Rupiah with a configuration of fold = 5 and lag = 5. This visualization is a time series cross-validation which is a model validation technique used in this research. In addition, a graph of the error rate per fold of each model will also be displayed. Figure 19 shows a visualization of the test results on the transformer model on the BBCA dataset.

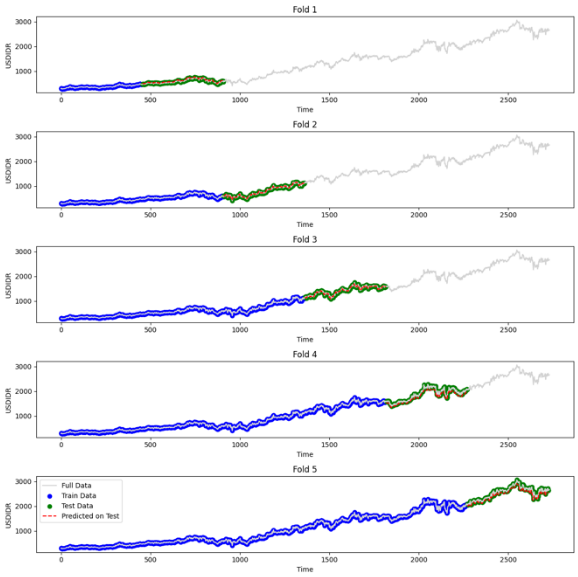

Figure 19. Visualization of BBCA stock dataset prediction with fold=5 and lag=5 configuration

Figure 19 is a visualization of the prediction results with the configuration fold=5 and lag=5, while the visualization of MAPE per fold is shown in Figure 20.

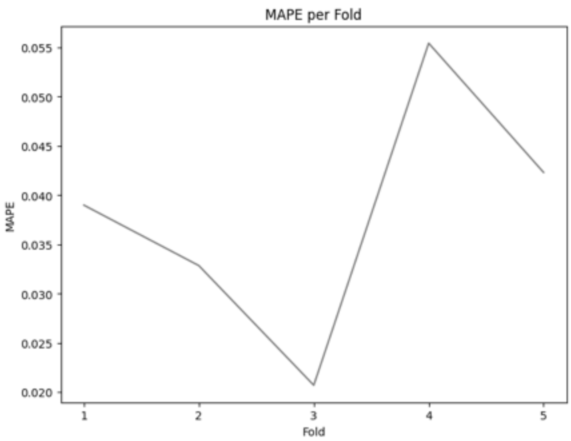

Figure 20. MAPE on BBCA stock dataset with fold=5 and lag=5 configuration

Next is a visualization of the test results on the US Dollar exchange rate dataset against the Indonesian Rupiah as shown in Figure 21.

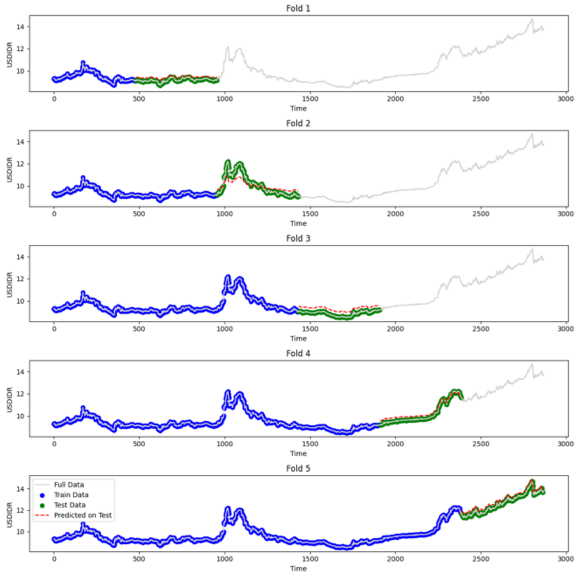

Figure 21. Visualization of USD/IDR currency exchange rate dataset prediction with fold=5 and lag=5 configuration

Figure 21 shows the visualization of the prediction results with the configuration fold=5 and lag=5, while the visualization of MAPE per fold is shown in Figure 22.

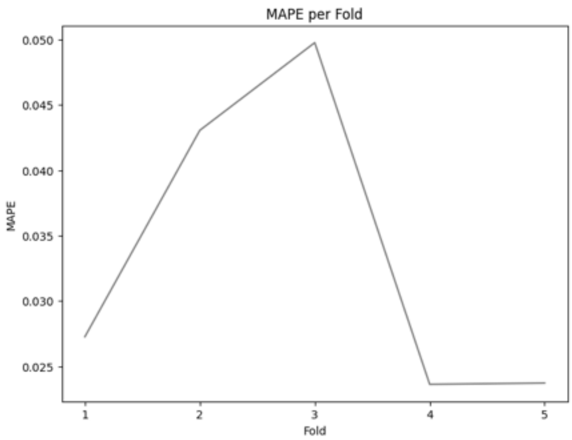

Figure 22. MAPE on BBCA stock dataset with fold=5 and lag=5 configuration

4. CONCLUSION

The results of this research indicate that the transformer model with lagged features and time series cross-validation has not been able to provide its best performance in predicting long-term time series data. In all experiments and tests, the transformer model has a relatively higher error rate compared to the LSTM model. The transformer model also shows fluctuating pattern in each fold, there is no tendency to decrease the error rate from the initial fold to the final fold, this means that the transformer model is not optimal in capturing long-term temporal patterns in the data. This is because the self-attention mechanism in the transformer is not specifically designed to handle data with temporal dependence and also the size of the data used in this research is still limited. The development of effective solutions requires a mechanism that is specifically developed to deal with data with long-term temporal dependence on the data, in addition it is necessary to engineer and simplify the transformer architecture that is specifically designed for forecasting long-term time series data by reducing computational complexity and increasing the size of longer sequence data for the testing process. So it can be concluded that the transformer model with lagged features and time series cross-validation cannot be used as an effective solution in handling data forecasting with long-term dependence.

Acknowledgment [if any]

We would like to express our deepest appreciation to the Ministry of Education, Culture, Research and Technology of the Republic of Indonesia for providing financial support for this research through the Penelitian Dosen Pemula Program. We would like to express our gratitude to the research team for their dedication and collaboration. We also extend our thanks to the entire community of Universitas Cendekia Mitra Indonesia for their support, enabling the successful completion of this research.

5. REFERENCE

[1] A. K. Sharma dan D. P. S. Shanmugaboopathi, "Digital Revolution and Its Nature and Extent in the Contemporary World," *The Technoarete Transactions on Advances in Social Sciences and Humanities,* vol. 2, no. 4, pp. 27-32, 2022.

[2] C. Stach, "Data Is the New Oil–Sort of: A View on Why This Comparison Is Misleading and Its Implications for Modern Data Administration," *Future Internet,* vol. 15, no. 2, pp. 2-49, 2023.

[3] A. W. Saputra, A. P. Wibawa, U. Pujianto, A. B. P. Utama dan A. Nafalski, "LSTM-based Multivariate Time-Series Analysis: A Case of Journal Visitors Forecasting," *ILKOM Jurnal Ilmiah,* vol. 14, pp. 57-62, 2022.

[4] D. P. Khairunnisa, N. Halwatunnissa dan D. H. Nurdiansyah, "Analysis Forecasting Of Operational Expense Of PT. Bank Rakyat Indonesia (Persero) Tbk," *Jurnal Ekonomi dan Bisnis,,* vol. 9, pp. 480-487, 2022.

[5] C. Ubal, G. Di-Giorgi, J. E. Contreras-Reyes dan R. Salas, "Predicting the Long-Term Dependencies in Time Series Using Recurrent Artificial Neural Networks," *Machine Learning and Knowledge Extraction,* vol. 5, no. 4, pp. 1340-1358, 2023.

[6] S. M. Al-Selwi, M. F. Hassan, S. J. Abdulkadir dan A. Muneer, "LSTM Inefficiency in Long-Term Dependencies Regression Problems," *Journal of Advanced Research in Applied Sciences and Engineering Technology,* vol. 30, no. 3, pp. 16-31, 2023.

[7] X. Ma, Z. Liu, M. Zheng dan Y. Wang, "Application and exploration of self-attention mechanism in dynamic process monitoring," *IFAC-PapersOnLine,* vol. 55, no. 6, pp. 139-144, 2022.

[8] R. Akbar, Weriana, R. A. Siroj dan M. W. Afgani, "Experimental Reseacrch Dalam Metodologi Pendidikan," *Jurnal Ilmiah Wahana Pendidikan,* vol. 9, no. 2, pp. 465-474, 2023.

[9] Sukmawati, Sudarmin dan Salmia, "Development of Quality Instruments and Data Collection Techniques," *Jurnal Pendidikan dan Pengajaran Guru Sekolah Dasar,* vol. 6, no. 1, pp. 119-124, 2023.

[10] R. R. Salam, M. F. Jamil, Y. Ibrahim, Rahmaddeni, Soni dan Herianto, "Analisis Sentimen Terhadap Bantuan Langsung Tunai (BLT) Bahan Bakar Minyak (BBM) Menggunakan Support Vector Machine," *MALCOM: Indonesian Journal of Machine Learning and Computer Science,* vol. 3, no. 1, pp. 27-35, 2023.

[11] M. R. Ismail, A. Zain, Jamaludin, F. Dewantoro dan D. Pratiwi, "Perhitungan Data Curah Hujan yang Hilang dengan Menggunakan Metode Interpolasi Linier," *Jurnal SENDI,* vol. 4, no. 2, pp. 60- 66, 2023.

[12] M. R. Suherlan, Asriyanik dan A. Pambudi, "UMMIBOT sebagai Media Layanan Informasi Penerimaan Mahasiswa Baru Universitas Muhammadiyah Sukabumi," *Jurnal Informatika Terpadu,* vol. 9, no. 2, pp. 82-91, 2023.

[13] E. Sokolova, O. Ivarsson, A. Lillieström, N. K. Speicher, H. Rydberg dan M. Bondelind, "Data-driven models for predicting microbial water quality in the drinking water source using E. coli monitoring and hydrometeorological data," *Science of the Total Environment,* vol. 802, 2022.

[14] N. H. Setiawan dan Zulkarnain, "Forecasting Palm Oil Production using Long Short-term Memory (LSTM) with Time Series Cross Validation (TSCV)," *International Journal of Social Service and Research (IJSSR),* vol. 4, no. 5, pp. 1237-1251, 2024.

[15] W. j. Gu, Y. h. Zhong, S. z. Li, C. s. Wei, L. t. Dong, Z. y. Wang dan C. Yan, "Predicting Stock Prices with FinBERT-LSTM: Integrating News Sentiment Analysis," dalam *Proceedings of the 2024 8th International Conference on Cloud and Big Data Computing*, Oxford, United Kingdom, 2024.