Jurnal Krisnadana (Jurnal Komputer, Sistem Kendali & Jaringan)
Vol 5, Issue 3, May 2026, pp. 550-569
https://doi.org/10.58982/avs92s73

E-ISSN: 2808-3563
P-ISSN: 2808-3520

Comparison of Deep Learning Methods for Product Sales Forecasting at The Catur Sasih Cooperative Four Seasons Hotel

I Putu Agus Eka Darma Udayana [1*], Putu Yoka Angga Prawira [2], Ni Made Jeni Aprilia Dewi [3]

[1*,2,3] Informatics, Institut Bisnis Dan Teknologi Indonesia,Denpasar, Bali, Indonesia
[1*]agus.ekadarma@gmail.com ; [2] yokaangga@rocketmail.com ; [3] jeniapriliaa12@gmail.com

ARTICLE INFO

Article history:
Received 3 June 2026
Revised 13 July 2026
Accepted 15 July 2026

***Keywords**:*
Comparison;
Sales Forecasting;
LSTM;
Autoformer;

ABSTRACT

This study aims to assist Koperasi Catur Sasih Hotel Four Seasons in determining the most appropriate sales forecasting method based on the characteristics of the cooperative's sales data, as fluctuating demand creates challenges in determining optimal inventory levels. The methods compared in this study are Long Short-Term Memory (LSTM) and Autoformer, using monthly sales data from January 2020 to December 2024 for three products, namely Yakult, Bavarois Roti Pizza/Sisir, and Marlboro Lights 20. The results indicate that the Autoformer method provides more accurate sales predictions than the LSTM method in forecasting product sales at Koperasi Catur Sasih, as evidenced by the lowest error values according to the RMSE, MAE, and MAPE metrics. For the Yakult product, the Autoformer method achieved an RMSE of 82.4660, an MAE of 72.2140, and a MAPE of 9.65%. For the Bavarois Roti Pizza/Sisir product, the Autoformer method produced an RMSE of 18.2666, an MAE of 15.4194, and a MAPE of 11.37%. Furthermore, for the Marlboro Lights 20 product, the Autoformer method resulted in an RMSE of 44.3490, an MAE of 37.0785, and a MAPE of 12.36%. In addition, noise reduction testing by removing the first three months of 2020 as an anomalous data period resulted in a decrease in error values for both methods. Nevertheless, Autoformer consistently outperformed LSTM in sales forecasting, both before and after the noise removal process.

1. Introduction

The A cooperative is a business organization that is collectively owned and operated by its members for mutual benefit. It aims to fulfill the economic needs of its members by upholding the principles of cooperation and a family-based approach [1]. One common form of cooperative is the employee cooperative, which typically provides various essential goods for its members. With these characteristics, employee cooperatives play an important role in improving the economic welfare of their members.

Koperasi Catur Sasih is an employee cooperative located at the Four Seasons Resort Jimbaran Hotel, Kuta District, Badung Regency. Currently, the cooperative has 566 active members and manages approximately 1,000 items or products, including basic necessities, snacks and processed foods, beverages, personal care products, household supplies, clothing, and cigarettes. In its daily operations, the cooperative employs four staff members who handle hundreds of transactions each day. In addition to registered members, family members of cooperative members are also allowed to shop at the cooperative, meaning that transaction volumes are not solely generated by official members.

Based on interviews with the cooperative supervisor, Koperasi Catur Sasih often experiences spikes in demand during certain periods, such as ahead of the Galungan holiday and on the 14th to 15th of each month due to the payroll deduction system. However, these demand patterns are not consistent over time, resulting in demand fluctuations that pose challenges in determining optimal stock levels. In several cases, this uncertainty causes the cooperative to experience stock shortages, which can hinder the optimal fulfillment of members' needs. This condition indicates that actual consumer demand does not always align with demand estimates. Therefore, data-driven approaches based on historical sales records, such as sales forecasting, are required to more accurately capture demand patterns and fluctuations.

For time series prediction, various methods can be applied, including deep learning–based approaches such as Long Short-Term Memory (LSTM) and Autoformer. LSTM is an extension of the Recurrent Neural Network (RNN) designed to address long-term dependency issues in sequential data, such as monthly sales records. In a study conducted by [2], LSTM was applied to sales forecasting for inventory optimization in a distribution company and achieved high accuracy, indicated by a MAPE value of 3.60%. On the other hand, Autoformer is a method specifically developed for long-term time series forecasting by utilizing series decomposition and an auto-correlation mechanism. Research by [3] demonstrated that Autoformer performed well in predicting retail sales demand, achieving an accuracy level of 1.097 for MASE and 0.272 for WQL. Given that both methods have shown strong performance in sales forecasting, this creates uncertainty in determining which method is more suitable for the characteristics of cooperative sales data. Therefore, this study conducts a comparative analysis using LSTM and Autoformer to identify which method produces more accurate sales predictions for products at Koperasi Catur Sasih based on the lowest error values.

To evaluate the accuracy of each method, this study employs three model evaluation metrics: Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE). RMSE measures the magnitude of prediction errors in the same units as the actual data, helping to assess how far the predicted values deviate from the true values. MAE calculates the average absolute difference between actual and predicted values without considering the direction of the error, providing an overall measure of the model's average error. Meanwhile, MAPE indicates the prediction accuracy level, where lower MAPE values represent higher forecasting accuracy [4] .

2. Method

This section explains the research stages, preprocessing data, forecasting modeling using the Long Short-Term Memory (LSTM) and Autoformer methods, as well as evaluation methods including RMSE, MAE, and MAPE.

Research Stages

This research was conducted through several stages: data collection, data preprocessing, forecasting modeling using Long Short-Term Memory (LSTM) and Autoformer methods, and model performance evaluation. The research stages were designed to produce a forecasting model capable of accurately representing sales patterns based on historical data. The general research flow includes sales data collection, data preprocessing, model development, model evaluation, and model performance comparison.

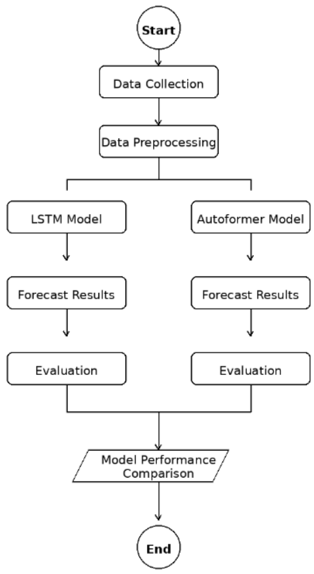

Figure 1. Research Stage Architecture.

Preprocessing Data

Data preprocessing is performed to ensure the data is ready for use in time series modeling. This stage includes checking and handling empty values and removing duplicate data to avoid data redundancy. Next, the sales data is transformed using a pivot table technique to form a time series structure per product based on chronological order. The transformed data is then normalized using the Min-Max Scaling method so that all values are in the range of 0 to 1, allowing the model to learn more stably and not be biased towards larger values [5]. After the normalization process, the dataset is divided into training data and test data with proportions of 80% and 20%. This process aims to evaluate the model's generalization ability to new data. This data splitting process aims to evaluate the model's generalization ability to new data [6].

Long Short-Term Memory (LSTM)

Long Short-Term Memory (LSTM) is a Recurrent Neural Network (RNN) architecture developed to address the problems of vanishing gradients and exploding gradients in sequential data modeling with long-term dependencies. LSTM incorporates a memory cell structure regulated by three main gates, namely the forget gate, input gate, and output gate. These gates control the flow of information that is retained, updated, and released from the cell state [7].

Mathematically, the LSTM mechanism can be explained as follows. The forget gate determines the information from the previous cell state that needs to be retained:

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

The input gate regulates the amount of new information added to the cell state:

$$i_t = \sigma(W_i[h_{t-1}, x_t] + b_i) \quad (2)$$

$$\tilde{C}_t = tanh(W_C[h_{t-1}, x_t] + b_C) \quad (3)$$

The cell state at time t is then updated as:

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t \quad (4)$$

The output gate controls the information exposed as the hidden state output:

$$o_t = \sigma(W_o[h_{\boldsymbol{t-1}}, x_t] + b_o) \quad (5)$$

$$h_t = o_t * tanh(C_t) \quad (6)$$

Through this three-gate mechanism, LSTM is able to preserve long-term information while discarding irrelevant patterns, making it effective for modeling sequential data such as time series, and forecasting based on historical data [8].

Autoformer

Autoformer is a Transformer-based architecture specifically designed for long-term time series forecasting. Unlike conventional Transformers, which use a point-wise self-attention mechanism, Autoformer introduces two key components, series decomposition and an auto-correlation mechanism, to capture trend and seasonal patterns more effectively and with improved computational efficiency [9].

In the series decomposition stage, the input time series X is decomposed into two components using a moving average–based decomposition block, namely the trend component (X_t) and the seasonal component (X_s). The trend component is obtained through a moving average operation, which can be formulated as:

$$X_t = \mathrm{AvgPool}(\mathrm{Padding}(X)) \quad (7)$$

The seasonal component is then computed by removing the trend component from the input series:

$$X_s = X - X_t \quad (8)$$

Thus, the decomposition process can be expressed as:

$$X_s, X_t = SeriesDecomp(X) \quad (9)$$

The seasonal component is subsequently modeled using a series-wise auto-correlation mechanism to identify recurring patterns and periodic dependencies within the time series. To improve computational efficiency, the auto-correlation computation is performed using

Fast Fourier Transform (FFT), reducing the computational complexity to $O(L \log L)$, which can be expressed as:

$$R_{XX}(\tau) = F^{-1}(F(X).F(X)*) \tag{10}$$

The resulting auto-correlation scores are used to identify dominant time lags, and the value representations are shifted accordingly to generate seasonal representations relevant for forecasting. In the decoder stage, Autoformer integrates both the seasonal and trend components to produce the final prediction. The trend component is directly propagated and incrementally updated across decoder layers, while the seasonal component is processed through the auto-correlation mechanism. By integrating time series decomposition with an auto-correlation-based modeling approach, Autoformer is able to effectively capture long-term dependencies and seasonal patterns, making it well suited for sales forecasting tasks involving complex historical data.

RMSE

Root Mean Square Error (RMSE) is an evaluation metric used to measure the extent to which a model's predictions deviate from the actual values. RMSE is calculated by summing the squared differences between the actual and predicted values, dividing by the number of data points, and finally taking the square root of the result. The RMSE formulation is given as follows[10] :

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}} \tag{11}$$

MAE

Mean Absolute Error (MAE) measures the average magnitude of prediction errors by calculating the mean of the absolute differences between actual and predicted values. A higher MAE value indicates a larger average prediction error. The MAE is defined as follows [11] :

$$MAE = \frac{1}{n}\sum_{i=1}^{n}[y_i - \hat{y}_i] \tag{12}$$

MAPE

Mean Absolute Percentage Error (MAPE) is an evaluation metric used to measure the accuracy of a predictive model. MAPE calculates the average absolute error over a given period, which is then expressed as a percentage of the actual value. The MAPE formulation is given as follows [10] :

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \times 100\% \tag{13}$$

3. Result and Discussions

This study uses sales time series data from the Catur Sasih Cooperative with an observation period of five years, namely from January 2020 to December 2024. For the purpose of model testing, three products were selected as test samples, namely Yakult, Bavarois Roti Pizza/Sisir, and Marlboro Lights 20. The selection of these three products aims to represent the variation in sales pattern characteristics, ranging from products with high sales volumes to products with sharper fluctuations. Thus, these three products are used to compare the performance of the LSTM and Autoformer models in handling various sales data patterns in the cooperative environment.

To prepare for the modeling process, data preprocessing was performed. The preprocessing phase included data cleaning, data transformation, data normalization, and data division into training and test data. The data cleaning process was performed by checking for missing values and duplicate data. The results showed that all sales data used was complete and free of missing values, so no data removal was necessary at this stage. The amount of data used remained consistent throughout the cleaning process, namely 180 monthly sales data sets from the three research products.

Next, the sales data were transformed into a time series by arranging the data in monthly order for each product. This transformation aimed to create a data structure suitable for time series modeling, allowing for temporal analysis of sales patterns for each product. After the transformation process, the data was normalized using the Min-Max Scaling method to equalize the scale of sales values between products within the range of 0 to 1. This normalization process aimed to increase the stability of the model training process and prevent the dominance of larger-scale values.

In the final stage, the normalized data was divided into training and test data sets, with proportions of 80% and 20%, respectively. The training data were used to build the forecasting model, while the test data were used to evaluate the model's performance on previously unseen data. This data split resulted in 48 training and 12 test data sets for each product, allowing for consistent evaluation of model performance across all test samples.

Long Short-Term Memory (LSTM)

After data preprocessing, a Long Short-Term Memory (LSTM) model was implemented using the TensorFlow Keras framework. The training data was formatted into a sequence format based on a specific time step length, where historical data was used as input to predict sales values for the next period. This approach enabled the model to learn temporal dependencies in sales time series data.

The LSTM model architecture used was a stacked LSTM consisting of three LSTM layers, with the first two layers using 64 neurons each and the last layer using 32 neurons. A dropout rate of 0.2 was applied to each LSTM layer to reduce the risk of overfitting. The output from the LSTM layer was then processed through a Dense layer with 16 neurons and a ReLU activation function before being mapped to a single neuron in the output layer to generate the predicted sales value. The model was trained using the Adam optimizer with a Mean Squared Error (MSE) loss function.

The training process was carried out with several hyperparameter configurations, including learning rate 0, 001, time steps of 2 and 4, batch sizes of 8 and 16, and epochs of 50, 75, and 100. During the training process, the training data was used to update the model weights, while the test data was used as validation data to evaluate the model's performance. Training was performed without data randomization to maintain the temporal order of the time series data.

After training the LSTM model with various hyperparameter configurations, the next step is to evaluate the model's performance using RMSE, MAE, and MAPE evaluation metrics. This evaluation aims to assess the accuracy of the prediction results and determine the best hyperparameter configuration for each product studied. A comparison of the error values from each configuration is used as the basis for selecting the best model, which is then used in the analysis and visualization of the prediction results.

Based on the evaluation results for the Yakult product, the use of time step 4 generally resulted in lower error values than time step 2. This indicates that Yakult sales patterns can be better predicted by considering longer historical data. The best configuration was obtained at time step 4, batch size 8, and epoch 50, with an RMSE of 109.1014, MAE of 95.2959, and MAPE of 13.83%, which is the lowest error value compared to other configurations.

For the Bavarois Roti Pizza/Sisir product, the evaluation results showed that the use of time step 2 provided better performance than time step 4, especially at batch size 8. This indicates that the sales patterns of this product are more effectively studied using a shorter historical data range, given its relatively high sales fluctuations. The best configuration was obtained at time step 2, batch size 8, and epoch 100, with an RMSE of 27.9321, MAE of 23.9474, and MAPE of 16.88%.

Meanwhile, the evaluation results for the Marlboro Lights 20 product showed that the RMSE, MAE, and MAPE values were relatively stable across all tested configurations. Changes in time step, batch size, and number of epochs did not result in significant differences in error. This indicates that the Marlboro Lights 20 sales pattern tends to be consistent and exhibits low fluctuations, making it easy for the LSTM model to learn. The best configuration was obtained at time step 2, batch size 8, and epoch 100, with an RMSE of 50.9264, MAE of 42.8373, and MAPE of 14.61%.

Overall, the evaluation results indicate that the performance of the LSTM model is influenced by the characteristics of each product's sales pattern and the choice of hyperparameters used. Therefore, the best configuration was selected specifically for each product based on the lowest error value. The best configuration obtained was then used in the visualization stage to compare the model's prediction results with actual sales data.

After obtaining the best configuration, the prediction results were analyzed through a visual comparison between actual sales data and the LSTM model's predictions for each product. Based on the visualization, the LSTM model was able to follow the general sales direction and pattern for all three products, both in the training and test data

For Yakult, the predictions tended to be more stable than the actual data, which exhibited sharp fluctuations, resulting in some extreme spikes and drops not being fully captured by the model. For Bavarois Roti Pizza/Sisir, the LSTM model was able to follow the sales trend quite well, but in some periods with high sales spikes, there was still a difference between the predicted values and the actual data. Meanwhile, for Marlboro Lights 20, the predictions showed a relatively consistent pattern and were close to the actual data, in line with its relatively stable sales characteristics.

Overall, the visualization results indicate that the LSTM model is effective in representing historical sales patterns and following the direction of data movements, but tends to produce smoother predictions, especially for data with high fluctuations. The differences in sales pattern characteristics for each product affect the model's ability to capture extreme changes. Therefore, another model approach is needed that is more adaptive to seasonal patterns and long-term fluctuations to improve prediction accuracy.

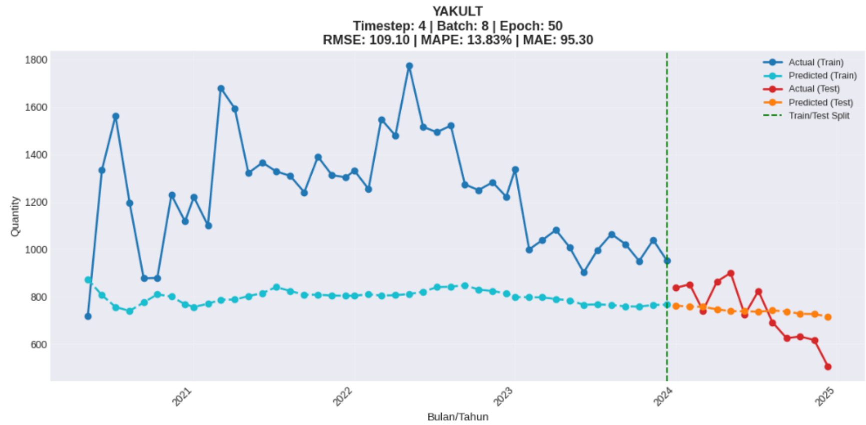

Figure 2. LSTM Yakult Prediction Chart

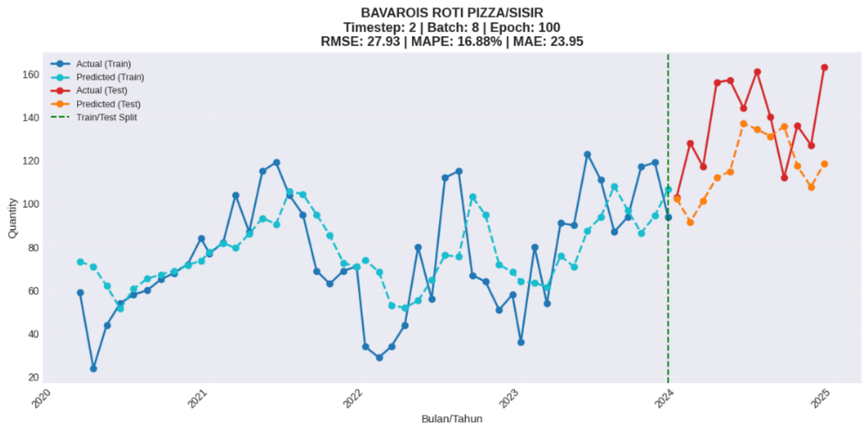

Figure 3. LSTM Bavarois Roti Pizza/Sisir Prediction Chart

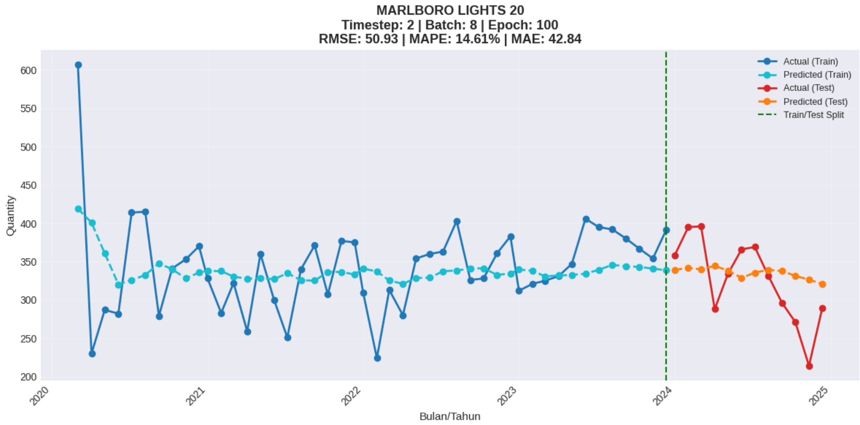

Figure 4. LSTM Marlboro Lights 20 Prediction Chart

After evaluating the performance of the LSTM model on the overall data, this study continued with a re-evaluation using data that had undergone a noise removal process. The noise removal process involved removing the first three months of 2020, which were considered an anomalous period due to extraordinary external conditions and did not represent normal sales patterns. This step aimed to assess the effect of noise removal on the performance of the LSTM model in predicting sales.

The evaluation results showed that noise removal improved model performance across all tested products. Based on the evaluation results for Yakult, using a longer time step tended to produce lower error values than using a shorter time step. This indicates that Yakult sales patterns can be better predicted when considering longer historical data after noise removal. The best configuration was obtained at a time step of 4, a batch size of 16, and epochs of 75, with an RMSE of 107.2046, an MAE of 92.4461, and a MAPE of 13.24%. This MAPE value indicates an increase in accuracy compared to before noise removal.

For the Bavarois Roti Pizza/Sisir product, evaluation results showed that using time step 4 provided better performance than time step 2, especially at a batch size of 16. This indicates that noise removal helps the model capture medium-term historical patterns in fluctuating sales data. The best configuration was obtained at time step 4, batch size 16, and epoch 50, with an RMSE of 24.3316, MAE of 19.0352, and MAPE of 13.21%, the lowest error value for this product.

Meanwhile, for the Marlboro Lights 20 product, the LSTM model's performance was relatively stable across the various parameter combinations tested. The difference in error values between time step 2 and time step 4 was not significant. The best configuration was obtained at time step 2, batch size 16, and epoch 75, with an RMSE of 48.2141, MAE of 40.5960, and MAPE of 13.42%. These results indicate that noise removal helps the model learn historical sales patterns more effectively, although the effect varies by product.

A visualization comparing actual sales data and predicted results shows that the LSTM model is able to follow the general sales direction and pattern for all three products after noise removal. Predictions tend to be more stable than actual data, especially for products with high sales fluctuations, so some extreme spikes and drops are not fully captured by the model. However, the MAPE value, which is around 13%, indicates that the prediction error rate is still within acceptable limits. Overall, noise removal produces more consistent predictions and improves the LSTM model's ability to represent historical sales patterns, although the model still tends to produce smooth predictions during periods with sharp sales changes.

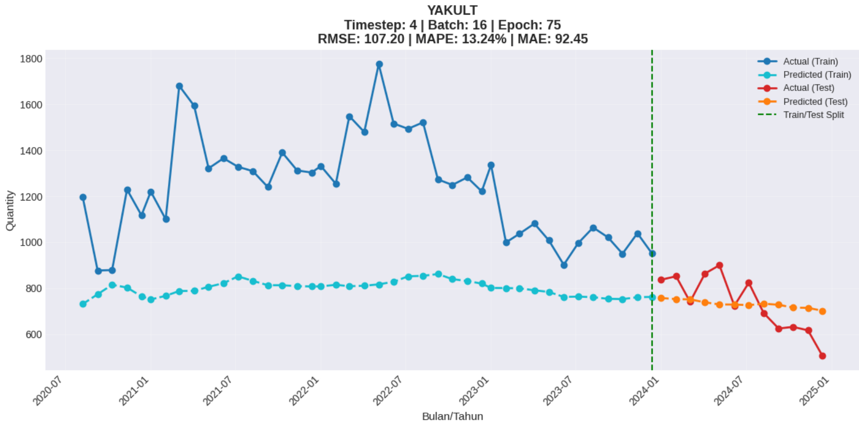

Figure 5. LSTM Yakult Noise Removal Chart

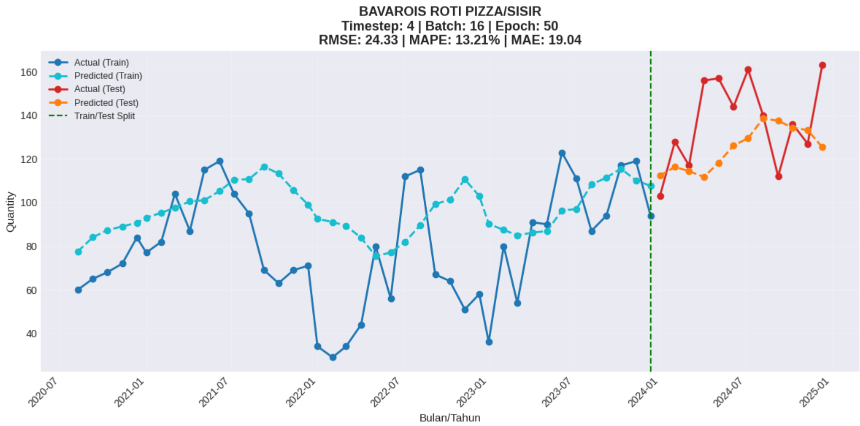

Figure 6. LSTM Bavarois Roti Pizza/Sisir Noise Removal Chart

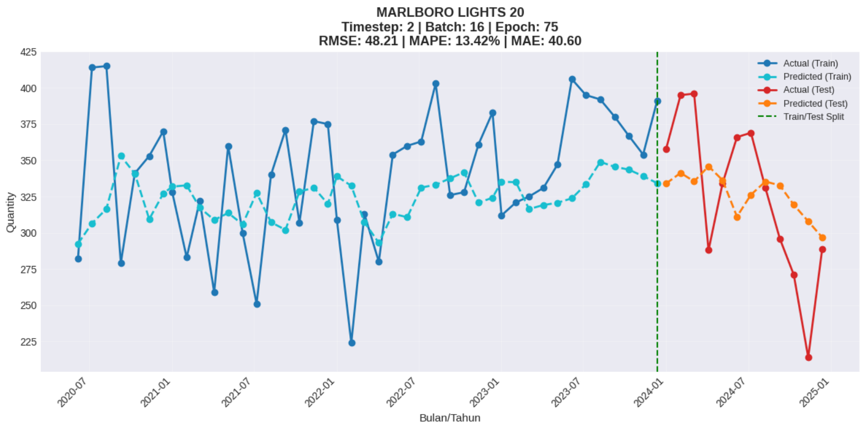

Figure 7. LSTM Marlboro Lights 20 Noise Removal Chart

Autoformer

After the preprocessing stage is complete, the data is used in the Autoformer model implementation stage. In this study, Autoformer was implemented using the PyTorch framework, which supports the development of neural network architectures, time series data processing, and efficient model training and evaluation. The model optimization process was performed using the Adam optimizer, while the loss function used was the Mean Squared Error (MSE) to measure the difference between predicted and actual values.

The Autoformer model architecture was designed using an encoder-decoder approach, consisting of two encoder layers and one decoder layer. Each layer is equipped with an Auto-Correlation mechanism to capture periodic patterns and long-term dependencies in the time series data. The model dimensionality was set at 64, with a feed-forward network size of 256, and a GELU (Gaussian Error Linear Unit) activation function. To reduce the risk of overfitting, a dropout rate of 0.1 was applied. At the end of the model, one neuron in the output layer was used to generate predicted sales values.

Before the training process, the training data was formatted into a sequence format based on a predetermined time step length. A set of historical data is used as input to predict sales values for the next period. This approach allows the Autoformer model to learn temporal patterns and relationships between consecutive periods. The model training process is carried out by testing several hyperparameter combinations: a learning rate of 0.001, time steps of 2 and 4, batch sizes of 8 and 16, and training epochs of 50, 75, and 100. At each epoch, the model generates a predicted value, which is then compared with the actual value to calculate the loss value. This loss value is used to update the model weights through backpropagation. Next, the model is evaluated using validation data to obtain a validation loss value, an indicator of the model's generalization ability to previously unseen data. All tested configurations are then compared to determine the Autoformer configuration with the best performance. The Autoformer model evaluation results indicate that prediction performance is influenced by the combination of time step, batch size, and number of epochs used, as well as the characteristics of the sales patterns for each product. The evaluation used RMSE, MAE, and MAPE metrics to determine the best configuration for each product. For Yakult, using time step 4 generally resulted in lower error values than using time step 2. This indicates that Yakult sales patterns can be better predicted by considering longer historical data. The best configuration was obtained at time step 4, batch size 8, and epoch 100, with an RMSE of 82.4660, MAE of 72.2140, and MAPE of 9.65%. MAPE values below 10% indicate that the Autoformer model has excellent forecasting capabilities for Yakult.

For Bavarois Roti Pizza/Sisir, the evaluation results showed significant performance variation between configurations. The best configuration was obtained at time step 2, batch size 16, and epoch 50, with an RMSE of 18.2666, MAE of 15.4194, and MAPE of 11.37%. These results indicate that products with high sales fluctuations can be predicted more optimally using short-term historical data. Increasing the number of epochs in some configurations actually increased the error value, indicating potential overfitting.

Meanwhile, for the Marlboro Lights 20 product, the Autoformer model's performance was relatively stable across the various parameter combinations tested. The differences in RMSE, MAE, and MAPE values between configurations were not significant, reflecting consistent sales patterns. The best configuration was obtained at time step 4, batch size 8, and epoch 50, with an RMSE of 44.3490, MAE of 37.0785, and MAPE of 12.36%. The consistency of the MAPE values in the 12%–15% range indicates that the Autoformer model is capable of predicting sales for this product quite well.

Overall, the evaluation results showed that the Autoformer model performed well on all three products with varying parameter configurations, depending on the characteristics of each product's sales pattern. The choice of time step, batch size, and number of epochs significantly affected the prediction accuracy. The best configuration was then used in the visualization analysis phase to compare the prediction results with actual sales data.

After obtaining the best model configuration for each product based on RMSE, MAE, and MAPE values, the next step is to analyze the prediction results generated by the Autoformer model using this best configuration on the test data. This analysis is conducted through a visual comparison between actual sales data and predicted results to evaluate the Autoformer model's ability to capture sales patterns, trends, and dynamics for each product.

Based on the graphical visualization of the comparison between actual data and predicted results for the three products, it can be seen that the Autoformer model is generally able to follow the direction of movement and basic sales patterns in both the training and test data. The resulting prediction line tends to move in the same direction as the actual data, indicating that the model is able to learn the characteristics of the historical sales patterns of each product. However, in certain periods, there are still differences between the predicted and actual values, especially when there are sharp sales fluctuations.

For Yakult, the Autoformer model's prediction results are able to follow the general sales trend, both during periods of increasing and decreasing sales. However, at certain points that show significant spikes or decreases in sales, the model's predictions tend to be smoother and do not fully reflect the extreme changes that occur in the actual data. This indicates that the Autoformer model focuses more on learning global patterns than short-term fluctuations. Although there are still differences in some periods, the relatively low MAPE value indicates that the overall prediction error rate is still within acceptable limits.

For the Bavarois Roti Pizza/Sisir product, sales data exhibits relatively high fluctuations in both the training and test data. The visualization shows that the Autoformer model is able to follow the direction of sales movements and recognize general patterns from historical data. However, during periods with sharp sales fluctuations, the prediction results are not fully able to accurately represent the magnitude of these changes. The resulting predictions tend to be more stable than the actual data, resulting in a discrepancy between the predicted line and the actual data in some periods. Nevertheless, the evaluation results show that the prediction error rate is still within acceptable limits, so the Autoformer model is still able to provide a general overview of the sales movement pattern for this product.

Meanwhile, for the Marlboro Lights 20 product, which has a relatively consistent sales pattern, the visualization results show that the Autoformer model's predictions align better with the actual data than products with high fluctuations. The prediction line followed historical sales patterns quite well, both on the training and test data. Although differences between predicted and actual values were still visible during some extreme periods, the model was overall able to maintain the direction and trend of the sales pattern. This indicates that the Autoformer model performs more optimally on data with more stable fluctuations.

Overall, the graphic visualization results indicate that the Autoformer model has a fairly good ability to capture general patterns and sales trends for the three products tested. The model tends to produce smoother predictions, thus representing medium- to long-term trends well. However, on data with sharp sales fluctuations, the model still has limitations in capturing extreme changes in certain periods. Differences in the characteristics of sales patterns for each product affect the accuracy of the resulting predictions, with products with more stable sales patterns tending to produce predictions closer to actual data than products with high fluctuations.

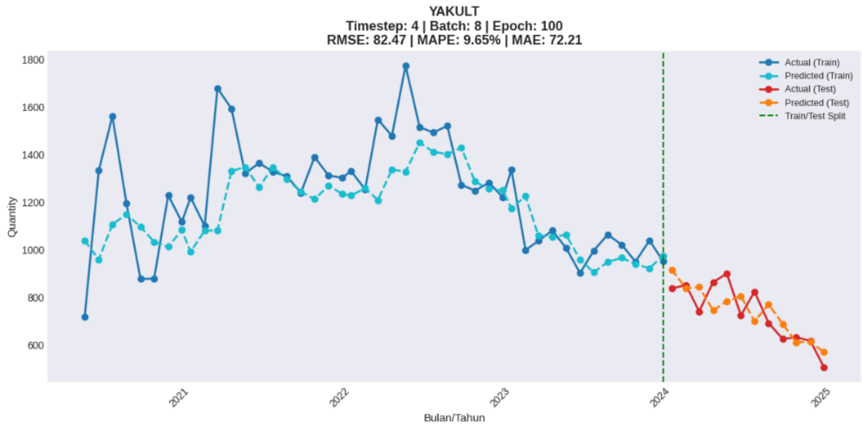

Figure 8. Autoformer Yakult Prediction Chart

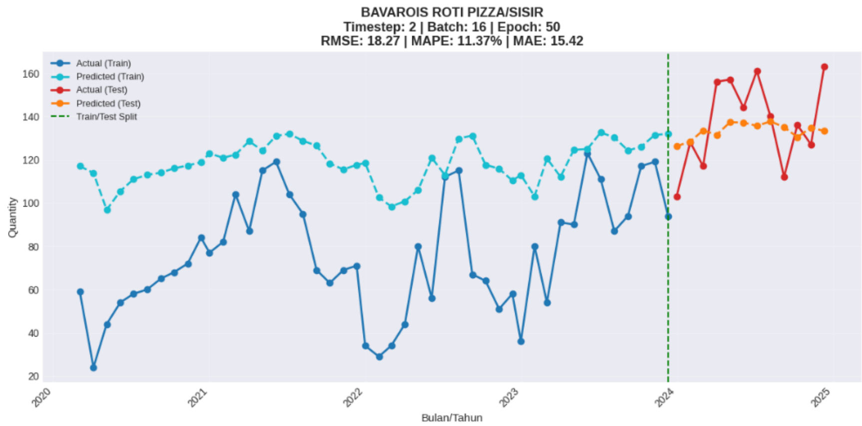

Figure 9. Autoformer Bavarois Roti Pizza/Sisir Prediction Chart

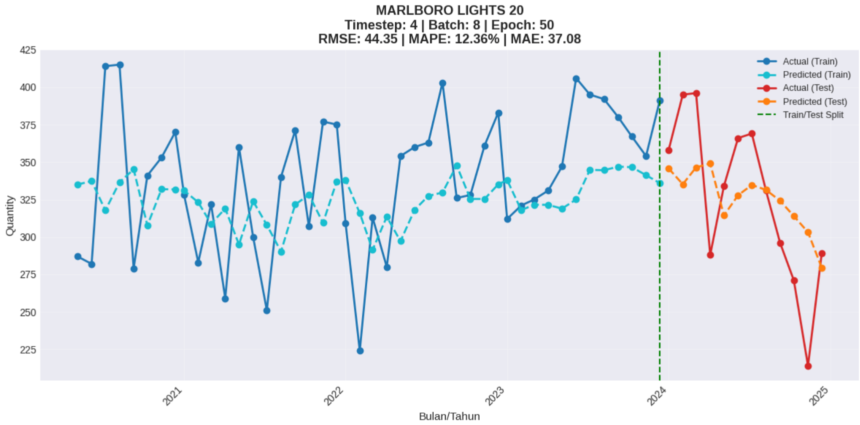

Figure 10. Autoformer Marlboro Lights 20 Prediction Chart

After evaluating the performance of the Autoformer model on the overall data, the next step was to evaluate the data after denoising. This denoising was done by removing the first three months of 2020, which were considered an anomalous period due to extraordinary external conditions and did not represent normal sales patterns. This step aimed to determine changes in the performance of the Autoformer model and the impact of denoising on prediction results.The Autoformer model performance was evaluated using the Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE) metrics for each product with various parameter combinations.

The evaluation results for the Yakult product showed that model performance was significantly influenced by the parameter combination used. In general, using time step 4 resulted in lower error values than using time step 2, indicating that using longer historical data can improve prediction accuracy. The best configuration was obtained at time step 4, batch size 8, and epoch 100, with an RMSE of 82.3681, MAE of 71.5334, and MAPE of 9.57%. A MAPE value below 10% indicates that the Autoformer model has excellent forecasting ability for Yakult products after noise removal.

The evaluation results for the Bavarois Roti Pizza/Sisir product show variations in performance for each parameter combination. In general, using time step 4 yields better results, especially at batch size 8, which yields lower error values than other configurations. The best configuration was obtained at time step 4, batch size 8, and epoch 75, with an RMSE of 18.4781, MAE of 15.5646, and MAPE of 11.19%. These results indicate that the Autoformer model is capable of predicting Bavarois product sales patterns quite well after noise removal, although increasing the number of epochs in some configurations actually leads to increased error values.

Evaluation results on the Marlboro Lights 20 product showed that the Autoformer model's performance was relatively stable across various parameter combinations. In general, the combination of a time step of 4 and a batch size of 8 provided more optimal results than other configurations. The best configuration was obtained with a time step of 4, a batch size of 8, and 100 epochs, with an RMSE of 43.9789, an MAE of 38.4103, and a MAPE of 12.59%. The MAPE value, which was in the 10–15% range, indicated that the

Autoformer model had quite good forecasting performance for the Marlboro Lights 20 product after noise removal.

Overall, the evaluation results using denoised data showed that the Autoformer model was capable of producing better predictive performance with appropriate parameter selection. Differences in RMSE, MAE, and MAPE values for each product confirmed that parameters such as time step, batch size, and number of epochs significantly influence prediction accuracy. The best configuration obtained for each product is then further analyzed through visualization of the comparison between actual sales data and predicted results.

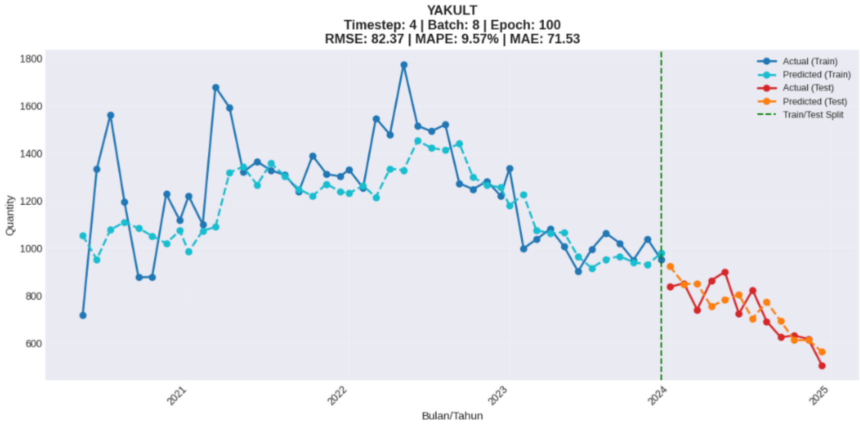

Figure 11. Autoformer Yakult Noise Removal Chart

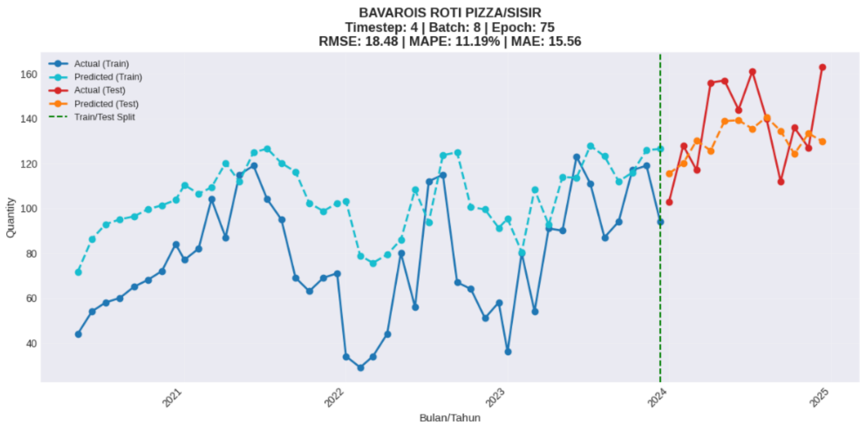

Figure 12. Autoformer Bavarois Roti Pizza/Sisir Noise Removal Chart

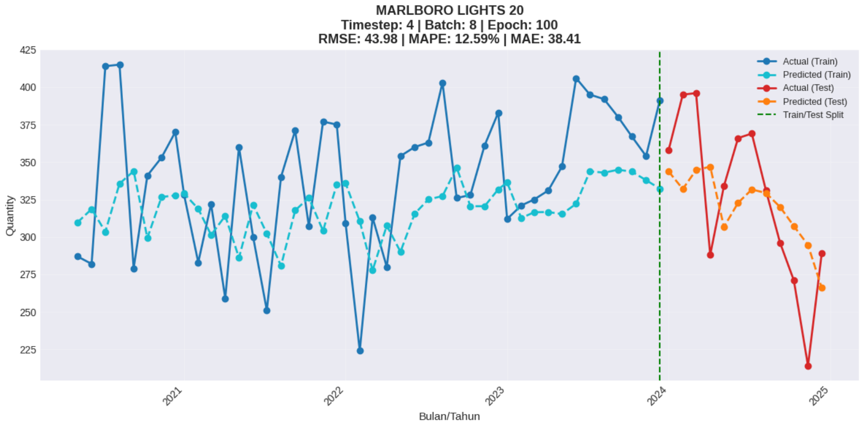

Figure 13. Autoformer Marlboro Lights 20 Noise Removal Chart

Performance Comparison between LSTM and Autoformer for Sales Forecasting

To determine the more accurate forecasting method, a comparison was conducted between the LSTM and Autoformer models for each product. This comparison aims to evaluate the capability of both models in predicting sales patterns based on the resulting prediction error values. Model performance evaluation was carried out using three metrics, namely Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE). A summary of the performance comparison between the LSTM and Autoformer models for each product is presented in Table 1 and Table 2.

Table 1. Model Accuracy Comparison

Product	Method	RMSE	MAE	MAPE
Yakult	LSTM	109.1014	95.2959	13.83%
	Autoformer	82.4660	72.2140	9.65%
Bavarois	LSTM	27.9321	23.9474	16.88%
	Autoformer	18.2666	15.4194	11.37%
Marlboro	LSTM	50.9264	42.8373	14.61%
	Autoformer	44.3490	37.0785	12.36%

Based on Table 4.1, the model performance comparison was conducted using RMSE, MAE, and MAPE on sales data before noise removal. The results show that for all three tested products, the Autoformer model consistently produced lower error values than the LSTM model.

For the Yakult product, the Autoformer model achieved an RMSE of 82.4660, MAE of 72.2140, and MAPE of 9.65%, which are lower than those obtained by the LSTM model, which recorded an RMSE of 109.1014, MAE of 95.2959, and MAPE of 13.83%. These results indicate that, compared to LSTM, Autoformer is more responsive to changes in Yakult sales and is therefore able to better represent sales fluctuations.

For the Bavarois Roti Pizza/Sisir product, Autoformer also demonstrated superior performance, with an RMSE of 18.2666, MAE of 15.4194, and MAPE of 11.37%, while the LSTM model produced an RMSE of 27.9321, MAE of 23.9474, and MAPE of 16.88%. This difference indicates that Autoformer is better able to follow the fluctuating sales pattern of

Bavarois Roti Pizza/Sisir. Meanwhile, for the Marlboro Lights 20 product, Autoformer again resulted in lower error values, with an RMSE of 44.3490, MAE of 37.0785, and MAPE of 12.36%, compared to the LSTM model, which achieved an RMSE of 50.9264, MAE of 42.8373, and MAPE of 14.61%. This suggests that Autoformer provides better predictive performance than LSTM for products with relatively consistent sales patterns.

To provide an overall view of model performance, the average error values of the best-performing models for the three products were calculated. The results show that the Autoformer model consistently outperformed LSTM, with average values of RMSE 48.3605, MAE 41.5706, and MAPE 11.13%. These values are lower than those produced by the LSTM model, which yielded average RMSE, MAE, and MAPE values of 62.6533, 54.0269, and 15.11%, respectively.

Table 2. Model Accuracy Comparison after Noise Removal

Product	Method	RMSE	MAE	MAPE
Yakult	LSTM	107.2046	92.4461	13.24%
	Autoformer	82.3681	71.5334	9.57%
Bavarois	LSTM	24.3316	19.0352	13.21%
	Autoformer	18.4781	15.5646	11.19%
Marlboro	LSTM	48.2141	40.5960	13.42%
	Autoformer	43.9789	38.4103	12.59%

After noise removal, which involved excluding the first three months of 2020 considered as anomalous periods that did not represent normal sales patterns, model performance evaluation was conducted again using RMSE, MAE, and MAPE, as shown in Table 2. The results indicate that for all tested products, the Autoformer model continued to produce lower error values than the LSTM model, consistent with the results obtained before noise removal.

For the Yakult product, Autoformer achieved an RMSE of 82.3681, MAE of 71.5334, and MAPE of 9.57%, which are lower than those obtained by the LSTM model, which recorded an RMSE of 107.2046, MAE of 92.4461, and MAPE of 13.24%. The reduction in error values for both models indicates that removing anomalous data periods had a positive impact on prediction accuracy, although the performance gap between Autoformer and LSTM remained consistent. For the Bavarois Roti Pizza/Sisir product, Autoformer again demonstrated better performance, achieving an RMSE of 18.4781, MAE of 15.5646, and MAPE of 11.19%, while the LSTM model produced higher error values of RMSE 24.3316, MAE 19.0352, and MAPE 13.21%. These results indicate that although noise removal reduced overall prediction errors, Autoformer remained superior in capturing the fluctuating sales pattern of the Bavarois product.

Similarly, for the Marlboro Lights 20 product, Autoformer recorded an RMSE of 43.9789, MAE of 38.4103, and MAPE of 12.59%, which are lower than those of the LSTM model, which achieved an RMSE of 48.2141, MAE of 40.5960, and MAPE of 13.42%. This shows that for products with relatively stable sales patterns, Autoformer is also able to maintain better predictive performance than LSTM after noise removal.

Based on the average error values shown in Table 2, Autoformer continued to outperform LSTM after noise removal, with average RMSE, MAE, and MAPE values of 48.2750, 41.8361, and 11.12%, respectively. These values are lower than those produced by the LSTM model, which yielded average RMSE of 59.9168, MAE of 50.6924, and MAPE of 13.29%. These findings indicate that although noise removal improves overall prediction accuracy, it does not change the relative superiority of Autoformer over LSTM.

Overall, the evaluation results demonstrate that the Autoformer model provides better forecasting performance than the LSTM model across all tested products. This is evidenced by consistently lower RMSE, MAE, and MAPE values achieved by Autoformer. The superiority of Autoformer in this study lies in its ability to capture dynamic and fluctuating sales patterns more effectively. Based on the visualization and prediction results before and after noise removal, the LSTM model tends to produce overly smooth forecasts and is less responsive to sharp fluctuations, as observed in the Yakult sales data. In contrast, Autoformer, through its Auto-Correlation mechanism, is able to capture long-term trends and sales fluctuations more effectively. This capability enables Autoformer to model inter-period sales pattern changes more accurately, resulting in consistently lower error values across all products. Therefore, based on the results of this study, the Autoformer model is proven to be more effective and accurate than the LSTM model in forecasting product sales at Koperasi Catur Sasih.

4. Conclusion

Based on the results of the comparative analysis between the Long Short-Term Memory (LSTM) and Autoformer models for sales forecasting at Koperasi Catur Sasih using monthly sales data of three products—Yakult, Bavarois Roti Pizza/Sisir, and Marlboro Lights 20—it can be concluded that the Autoformer model consistently demonstrates superior performance compared to the LSTM model. Model performance was evaluated using the Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE) metrics. The findings indicate that Autoformer produces lower prediction error values across all tested products. This is reflected in the average error values obtained from the best-performing configurations of each model, where Autoformer achieved an average RMSE of 48.3605, MAE of 41.5706, and MAPE of 11.13%, which are lower than those recorded by the LSTM model with an average RMSE of 62.6533, MAE of 54.0269, and MAPE of 15.11%. These results demonstrate that Autoformer exhibits superior forecasting capability compared to LSTM within the context of this study.

For the Yakult product, the Autoformer model consistently produced lower error values, with an RMSE of 82.4660, MAE of 72.2140, and MAPE of 9.65%, compared to the LSTM model, which yielded an RMSE of 109.1014, MAE of 95.2959, and MAPE of 13.83%. These results indicate that Autoformer is more responsive to variations in Yakult sales and is therefore better able to represent sales fluctuations. This improved performance suggests that Autoformer is more effective in capturing dynamic changes in demand patterns than LSTM.

Similar results were also observed for the Bavarois Roti Pizza/Sisir and Marlboro Lights 20 products, where the Autoformer model consistently demonstrated higher predictive accuracy than the LSTM model. For the Bavarois Roti Pizza/Sisir product, Autoformer achieved an RMSE of 18.2666, MAE of 15.4194, and MAPE of 11.37%, whereas the LSTM model produced higher error values with an RMSE of 27.9321, MAE of 23.9474, and MAPE of 16.88%. For the Marlboro Lights 20 product, Autoformer also yielded lower error values, with an RMSE of 44.3490, MAE of 37.0785, and MAPE of 12.36%, compared to the LSTM model, which recorded an RMSE of 50.9264, MAE of 42.8373, and MAPE of 14.61%. These results indicate that Autoformer is capable of delivering more accurate forecasts for products with both relatively stable sales patterns and those exhibiting inter-period fluctuations.

Furthermore, after applying a noise removal process by excluding the first three months of 2020, which were identified as anomalous periods, a reduction in error values was observed for both forecasting models across all products. For the Yakult product, the exclusion of anomalous data resulted in lower RMSE, MAE, and MAPE values for both LSTM and Autoformer. A similar trend was observed for the Bavarois Roti Pizza/Sisir and Marlboro Lights 20 products, indicating that the removal of anomalous periods contributed positively to improving overall forecasting accuracy. Nevertheless, despite the general reduction in error values, the Autoformer model consistently maintained superior performance compared to the LSTM model, both before and after the noise removal process.

Overall, the results of this study confirm that the Autoformer model is more accurate for forecasting product sales at Koperasi Catur Sasih than the LSTM model, as evidenced by consistently lower error values across all evaluation metrics and experimental scenarios.

References

[1] Z. D. Widodo, J. P. Purwaningrum, I. Purbasari, and G. P. Rini, "Manajemen Koperasi dan UMKM," 2022.

[2] D. Susilo, A. Chusyairi, and M. I. Saputra, "Prediksi Penjualan Untuk Optimasi Stock Produk Menggunakan Algoritma Long Short Term Memory," *Sistematis*, vol. 1, no. 2, pp. 119–134, 2025.

[3] J. M. Oliveira and P. Ramos, "Evaluating the Effectiveness of Time Series Transformers for Demand Forecasting in Retail," *Mathematics*, vol. 12, no. 17, Sep. 2024, doi: 10.3390/math12172728.

[4] R. Dwiki, P. Hamas, N. Y. Setiawan, and D. E. Ratnawati, "PREDIKSI PENJUALAN MAKANAN RESTORAN MENGGUNAKAN METODE ARIMA: STUDI KASUS WAROENG MARISUKAKOI," 2025. [Online]. Available: http://j-ptiik.ub.ac.id

[5] A. Alsarhan, F. Hussein, S. Moh, and F. S. El-Salhi, "The Effect of Preprocessing Techniques, Applied to Numeric Features, on Classification Algorithms' Performance," 2021, doi: 10.3390/data.

[6] J. Khatib Sulaiman, A. Khoirala Ermy Pily, F. Aprilia, L. Efrizoni, and I. Artikel Abstrak, "Komparasi Algoritma K-Nearest Neighbors dan Naïve Bayes dalam Klasifikasi Penyakit Diabetes Gestasional STMIK Amik Riau," *Indonesian Journal of Computer Science Attribution*, vol. 13, no. 1, pp. 2024–1195, 2024.

[7] A. P. Meriani and A. Rahmatulloh, “Perbandingan Gated Recurrent Unit (Gru) Dan Algoritma Long Short Term Memory (Lstm) Linear Refression Dalam Prediksi Harga Emas Menggunakan Model Time Series,” *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 12, no. 1, 2024.

[8] D. S. Ameera, A. T. Damaliana, and M. Idhom, “PENERAPAN MODEL HIBRIDA ARIMA-LSTM PADA PREDIKSI INFLASI DI INDONESIA,” *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 9, no. 3, pp. 3743–3749, 2025.

[9] H. Wu, J. Xu, J. Wang, and M. Long, “Autoformer: Decomposition Transformers with Auto-Correlation for Long-Term Series Forecasting,” 2021. [Online]. Available: https://github.com/thuml/Autoformer.

[10] T. Lattifia, P. W. Buana, and N. K. D. Rusjayanthi, “Model Prediksi Cuaca Menggunakan Metode LSTM,” *JITTER J. Ilm. Teknol. dan Komput*, vol. 3, no. 1, pp. 994–1000, 2022.

[11] D. Ramatika, L. Marlina, and A. P. U. Siahaan, “MODEL PREDIKTIF DAN ANALISIS TREN PENYEWAAN KONTAINER DENGAN MACHINE LEARNING DALAM MENGOPTIMALKAN OPERASIONAL PERUSAHAAN,” *Jurnal Informatika Teknologi dan Sains (Jinteks)*, vol. 7, no. 2, pp. 937–946, 2025.