

A Comparative Analysis of the Long Short-Term Memory (LSTM) and Prophet Models in Gold Price Forecasting Using Time Series Data

Faqih Rifaldy ^{1*}, Reza Ade Putra ²

^{1*,2} Computer Science Study Program, Faculty of Science and Technology, Universitas Islam Negeri Sumatera Utara, Deli Serdang Regency, North Sumatra Province, Indonesia

Email: faqihrifaldy03@gmail.com ^{1*}, rezaadeputra@uinsu.ac.id ²

Article history:

Received September 18, 2026

Revised October 8, 2026

Accepted October 12, 2026

Abstract

Gold price forecasting is crucial for investors in mitigating financial risk, particularly due to high market volatility. This study aims to evaluate and compare the performance of two distinct time series modeling approaches, representing additive statistical models. The research utilizes 4,733 daily closing price records of PT Aneka Tambang Tbk (ANTAM) golds spanning from Januari 4, 2010, to November 20, 2025 partitioned into an 80:20 chronological ratio for training (3,786 days) and testing (947 days) sets. Unlike previous studies that included lag features, this paper evaluates both models in a pure univariate setup using only dates and prices to reflect their true intrinsic forecasting capabilities. Models were configured through systematic hyperparameter tuning: LSTM utilized a 90-day window size, 128 units, a single layer with 0.2 Dropout, and Adam optimizer with a learning rate of 0.001, evaluated via a one-step walk-forward scheme. Prophet was configured with a linear growth, multiplicative seasonality, and a changepoint prior scale of 0.8. Evaluation metrics demonstrate that LSTM significantly outperforms Prophet across all metrics, achieving a Mean Absolute Error (MAE) of Rp15,038, Root Mean Square Error (RMSE) of Rp23,637, and Mean Absolute Percentage Error (MAPE) of 0.89%, compared to Prophet's MAE of Rp130,982, RMSE of Rp174,014, and MAPE of 8.62%. The findings indicate that LSTM's dynamic memory cell mechanism is superior in adapting to non-linear market trends, offering valuable technological insights for automated commodity decision-support systems.

Keywords:

Gold Price Forecasting; Long Short-Term Memory; Prophet Model; Time Series Analysis; PT Antam.

1. INTRODUCTION

Antam's gold price is one of the investment instruments that is highly sought after by the Indonesian public. Antam's gold price tends to fluctuate over time, influenced by various economic factors, both global and domestic. These unpredictable price fluctuations make it difficult for the public and investors to determine the right time to buy or sell gold (Siagian, 2025). Gold price fluctuations refer to changes in the price of gold over time, whether in the form of increases (appreciation) or decreases (depreciation) (Sunani et al., 2025). Since prices are recorded sequentially on a daily basis, gold price data falls under the category of time-series data. The dynamic nature of this data makes forecasting based on historical data crucial in helping investors and industry players mitigate risks and make more rational and measured investment decisions (Sagala & Dewi, 2023).

As information technology has advanced, various methods in the fields of machine learning and deep learning have been widely used to analyze time series data. One widely used model is the Long Short-Term Memory (LSTM). LSTM is an extension of the Recurrent Neural Network (RNN) designed to capture long-term dependency patterns in sequential data. LSTM is a type of Recurrent Neural Network (RNN) capable of

capturing long-term data patterns and overcoming RNN limitations such as gradient vanishing (Putra & Andryana, 2025). In addition to deep learning approaches, another method widely used in time series forecasting is Prophet, a forecasting model developed by the technology company Meta Platforms. The Prophet algorithm is designed with key components to handle the unique characteristics of time series data, such as trends, seasonal effects, and holidays (Simamora et al., 2024).

Although both methods are used to make time series predictions, each takes a different approach to modeling patterns. LSTM uses a neural network-based approach capable of learning complex nonlinear patterns. LSTM is a type of neural network architecture that can handle long-range dependencies; it is widely used to model complex relationships and patterns over time in datasets (Michel, 2024). Meanwhile, Prophet uses a model-based approach that identifies cause-and-effect relationships based on trend and seasonal components (Muzakki et al., 2021).

A comparison between LSTM and Prophet was conducted due to their differing characteristics, which may result in variations in performance when predicting gold price data. The LSTM model is capable of capturing patterns that do not need to be sequential and employs a gate mechanism that adapts to the sequence of historical data. Meanwhile, the Prophet model operates based on an additive decomposition approach that explicitly separates the data into trend, seasonal, and holiday components (Dalimunthe et al., 2025).

These two models represent two opposing philosophies in time series forecasting: LSTM learns from the data autonomously without any structural assumptions, while Prophet explicitly models the data structure based on interpretable components (Yustiana et al., 2026). For example, PT Antam's gold price data exhibits two characteristics that simultaneously highlight the strengths of each model: Prophet captures long-term trends with significant shifts in rate such as during the 2020 COVID-19 pandemic while LSTM accounts for nonlinear daily fluctuations driven by exchange rates and market sentiment. Since it is not known for certain which approach provides more accurate predictions for Antam's specific data characteristics, a direct comparison of the two models on the same dataset is the focus of this study.

Several previous studies have utilized machine learning and statistical methods to forecast commodity prices, including gold. One of the most recent studies comparing the LSTM and Prophet models for gold price forecasting was conducted by (Primawati, 2025), who used global gold futures price data from 2020 to 2025 (1,387 daily data points). Based on the analysis, it was found that they used only two-time steps from the data; the study incorporated the prices from 1 and 2 days prior ($y-1$, $y-2$) as features into both models. The results showed that Prophet outperformed LSTM with a MAPE of 0.21% and an R^2 of 0.99971, while LSTM actually had a lower MAPE (0.0%) but a slightly lower R^2 (0.9567); thus, the conclusion that "Prophet is better" depends on which metric is prioritized. The researchers themselves acknowledge that this very high R^2 value may be due to overfitting, given that the dataset used is relatively short and covers a period of relatively stable prices.

This study notes that the $y-1$ and $y-2$ features have different methodological impacts on the two models. In LSTM, these features are essentially equivalent to a window as short as 2 days (in accordance with the $\text{TimeStep}=2$ parameter in their architecture); they do not represent additional information beyond the natural LSTM mechanism, but rather the window itself, which is explicitly written as a column. However, in Prophet which natively lacks any windowing mechanism or history of previous values the addition of $y-1$ and $y-2$ as regressors introduces new information from outside its original structure. Consequently, Prophet's exceptionally high accuracy risks reflecting the predictive power of "yesterday's price" rather than Prophet's inherent ability to extrapolate trends and seasonality.

Given this gap, research that specifically examines and conducts a comparative analysis between the LSTM and Prophet models using Antam gold price data in Indonesia remains limited; therefore, further research is needed to evaluate both models using purely univariate variables specifically, date (ds) and gold price (y) without any lag-based regressor features. Thus, the performance of LSTM and Prophet obtained purely reflects the intrinsic capabilities of each architecture, without the aid of additional features that risk disproportionately favoring one model, particularly Prophet, as described above. This study also uses domestic historical gold price data from PT Antam covering a much longer time span January 2010 through November 2025 totaling 4,893 days of raw data, unlike the 5-year global data set used by Primawati et al (2025). The evaluation was conducted using three equally weighted metrics MAE, RMSE, and MAPE to avoid conclusions dependent on a single metric, while also serving as additional validation of the overfitting findings identified by (Primawati, 2025).

Based on these issues, this study aims to conduct a comparative analysis between the Long Short-Term Memory (LSTM) and Prophet models in forecasting the price of gold at PT Aneka Tambang (Antam) using time-series data. The results of this study are expected to provide insight into which forecasting model performs best in modeling gold price movements, thereby serving as a reference for developing commodity price prediction systems and aiding future investment decision-making.

2. RESEARCH METHOD

2.1. Research Approach and Design

This research was designed as a quantitative experimental study using a comparative approach, aimed at measuring and comparing the accuracy of two time-series forecasting models—namely, the Long Short-Term Memory (LSTM) algorithm and the Prophet model—in forecasting PT Antam's gold prices. A comparative approach was chosen because this study does not merely develop a single prediction model but runs two models in parallel under identical data conditions, so that any differences in performance can be directly attributed to the characteristics of each algorithm rather than to external factors such as differences in training data or testing periods.

2.2. Population, Sample and Data Collection

The population and sample in this study consist of secondary data from historical time-series records of gold bar prices from PT Aneka Tambang Tbk (ANTAM). The data were downloaded from a public repository on the GitHub platform published by the HuggingFace community (account: theonegareth). The raw data set consists of 4,893 rows of intraday records covering a 15-year period, from January 4, 2010, to November 20, 2025. The data inclusion criteria applied covered all daily transaction data on PT Antam's gold prices recorded in Indonesian Rupiah per gram (Rp/gram), along with valid, zero, or negative timestamps. Since this research is based on experimental computation using publicly available secondary data and does not involve human subjects, no subject demographic criteria or informed consent procedures are required.

2.3. Research Steps and Data Preprocessing

All stages of the experiment and data processing were conducted in an academic environment using Python on the Google Colaboratory (Google Colab) platform. The systematic steps for data preprocessing were as follows:

- Data Cleaning & Resampling.** A check for missing values revealed that there were no empty values in the raw dataset. Since the data is intraday (there are 139 days with more than one price recording per day), daily resampling was performed by taking the last daily closing price for each day. This process resulted in 4,733 unique daily data rows.
- Anomaly Detection.** Anomaly checking was performed based on an indicator of extreme daily price spikes (>15% in a single day) to identify data recording errors. The evaluation results confirmed that no extreme spikes were found, so all 4,733 daily data points were retained.
- Split into Training and Test Datasets (80:20).** The data was split chronologically (not randomly) in an 80:20 ratio to preserve temporal dependency. The training data consists of 3,786 days (January 4, 2010 – March 13, 2023), while the test data consists of 947 days (March 14, 2023 – November 20, 2025).
- Data Normalization (LSTM-Specific).** Specifically for LSTM modeling, the training data is normalized to the range [0,1] using MinMaxScaler to ensure the stability of the neural network training. The parameters are calculated exclusively from the training data to avoid data leakage.

2.4. Model Architecture and Analysis Technique

2.4.1. Model Long Short-Term Memory (LSTM)

The normalized training data is converted into a sequence structure using a sliding window scheme. Mathematical computations in the LSTM cell units are performed using the following gate equations:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (2.1)$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (2.2)$$

$$\tilde{C}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \quad (2.3)$$

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \quad (2.4)$$

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o); h_t = o_t * \tanh(C_t) \quad (2.5)$$

Gradual tuning determined the optimal LSTM configuration: window size = 90 days, 128 LSTM units, 1 hidden layer with a dropout rate of 0.2, the Adam optimizer (learning rate = 0.001), batch size = 32, and training for a maximum of 100 epochs using EarlyStopping (patience of 20 epochs). Evaluation was performed using the one-step walk-forward scheme.

2.4.2. Model Prophet

The Prophet model operates directly on the original Rupiah scale by formulating forecasts through additive/multiplicative decomposition:

$$y(t) = g(t) + s(t) + h(t) + \epsilon_t \quad (2.6)$$

The linear trend growth function $g(t)$ and the Fourier-based seasonal component $s(t)$ are calculated using equations (2.7) and (2.8):

$$g(t) = (k + a(t)^T \delta)t + (m + a(t)^T \gamma) \quad (2.7)$$

$$s(t) = \sum_{n=1}^N \left(a_n \cos\left(\frac{2\pi i n t}{P}\right) + b_n \sin\left(\frac{2\pi i n t}{P}\right) \right) \quad (2.8)$$

Gradual tuning establishes the optimal Prophet configuration: growth = linear, changepoint_prior_scale = 0.8, seasonality_mode = multiplicative, changepoint_range = 0.1, and weekly_seasonality = True. The parameters are estimated using the Maximum A Posteriori (MAP) method with L-BFGS optimization. A summary of the experimental architecture of the two models is presented in Table 1.

Table 1. Summary of the Model's Architecture and Experimental Configuration

Configuration Components	LSTM Model	Prophet Model
Input Data Scale	Univariate Normalized [0,1]	Univariate Original Scale (Rupiah)
Structuring/Timestep	Sliding Window (90) days	ds and y column formats
Architecture/Growth	1 LSTM Layer (128 Units) + 0.2 Dropout	Linear Growth
Seasonality Component	Implicit via Memory Cell	Fourier Series ($N=10$, $P=365.25$)
Trend Flexibility	Adam Optimizer ($lr=0.001$)	change_prior_scale = 0.8
Solver/Estimation	Backpropagation Through Time	MAP Estimation (L-BFGS)

2.4.3. Performance Evaluation Metrics

The performance of the two models was evaluated using 947 days of test data and three standard quantitative metrics:

$$MAE = \frac{1}{n} \sum_{i=1}^n |A_i - P_i| \quad (2.9)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (A_i - P_i)^2} \quad (2.10)$$

$$MAPE = \left(\frac{1}{n} \right) \sum_{i=1}^n \left| \frac{A_i - F_i}{A_i} \right| \times 100\% \quad (2.11)$$

2.4.4. Time Series Cross-Validation

To improve methodological rigor, a time series cross-validation procedure was implemented in addition to the validation procedure was implemented in addition to the standard 80:20 holdout evaluation. Unlike conventional k-fold cross-validation which shuffles data randomly and is therefore inappropriate for temporal data, time series cross-validation preserves chronological order by using a walk-forward expanding-window scheme. This method is implemented using TimeSeriesSplit from the scikit-learn library with $k=5$ folds.

In this procedure, the entire dataset of 4,733 days is divided into five sequential folds. In each fold i , the model is trained on all data from day 1 through the end of fold i (expanding training window), and evaluated on the immediately following held-out period.

Table 2. Fold Boundaries

Fold	Training Set (days)	Validation Set (days)
1	1 – 2,840	2,841 – 3,029
2	1 – 3,029	3,030 – 3,218
3	1 – 3,218	3,219 – 3,407
4	1 – 3,407	3,408 – 3,596
5	1 – 3,596	3,597 – 3,786

Note: Fold boundaries are defined proportionally; each validation set covers approximately 189 days.

For the LSTM model, each fold applies the same preprocessing pipeline. MinMax Scaller is fit exclusively on the training portion of that fold (preventing data leakage), and predictions are generated using the one-step walk-forward scheme. For the Prophet model, the training DataFrame for each folds is reconstructed chronologically and the model is refitted from scratch. Performance is measured using MAE, RMSE, and MAPE on each validation fold, the mean and standard deviation across folds is reported to assess stability.

3. RESULTS AND DISCUSSION

3.1. Results

3.1.1. Preprocessing Results and Descriptive Statistical Analysis

The raw dataset, consisting of 4,893 rows of intraday price records, was first processed through data cleaning and daily resampling (based on daily closing prices). A total of 139 days with more than one price observation were consolidated by taking the last closing price of the day, resulting in 4,733 unique daily data points (January 4, 2010 – November 20, 2025). An anomaly test using a daily surge threshold of >15% confirmed that there were no data recording errors in the dataset. A summary of the descriptive statistics for the cleaned dataset is presented in Table 3.

Table 3. Descriptive Statistics on PT Antam's Daily Gold Prices (2010–2025)

Statistic	Value (Rp/gram)
Number of Observations (Count)	4733
Mean	Rp843.426
Standard Deviation (Std. Dev)	Rp382.458
Minimum Price (Min)	Rp394.000
1 st Quartile (25%)	Rp572.000
Median (50%)	Rp670.000
3 rd Quartile (75%)	Rp994.000
Maximum Price (Max)	Rp2.487.000

Table 3 shows a very wide range of fluctuations in the price of gold, ranging from a minimum of Rp394,000 to a maximum of Rp2,487,000 per gram, with a standard deviation of Rp382,458 (equivalent to 45% of the mean). This indicates the presence of a significant long-term exponential upward trend, particularly during the 2023–2025 period, which encompasses the test data range.

The data was then divided chronologically in an 80:20 ratio. A total of 3,786 days (January 4, 2010 – March 13, 2023) were allocated as the training set, and 947 days (March 14, 2023 – November 20, 2025) as the test set. In the LSTM model, the training data was normalized using MinMaxScaler to the range [0,1], with $X_{min} = \text{Rp}394,000$ and $X_{max} = \text{Rp}1,065,000$. Due to the sharp upward trend from 2023 to 2025, the normalized test data range falls between 0.961 and 3.119 (exceeding the value of 1.0).

3.1.2. Model Development Results and Parameter Tuning

To obtain the optimal forecasting performance from both algorithms, a systematic hyperparameter tuning process was conducted under identical data conditions.

3.1.2.1. Model Tuning Parameters for the Prophet Model

The hyperparameter tuning for the Prophet model was performed in four sequential stages to evaluate the impact of trend flexibility, seasonality mode, changepoint range, and weekly seasonality.

Table 3. Configuration Parameters

Configuration Parameter	Result
Growth	Linear
yearly_seasonality	True
weekly_seasonality	True
daily_seasonality	False
changepoint_prior_scale	0,8
seasonality_mode	multiplicative
changepoint_range	1,0

Table 4 summarize the final optimall hyperparameter values selected for the Prophet model after completing all experimental tuning stages. The model utilizes a linear growth function combined with a multiplicative seasonality mode, which accounts for seasonal fluctuations that scale proportionally with the rising gols price. A changepoint_prior_scale of 0.8 was selected to provide high flexibility for the trend line to adjust to historical price shifts, while changepoint_range was set to 1.0 to allow trend changes to be detected

across the entire training dataset. Annual and weekly seasonalities were set to active, whereas daily seasonality was disabled due to the daily frequency of the closing price data.

Table 4. Model Tuning Parameters for the Prophet Model

Stages	Parameters Tested	Best	MAPE
Baseline	changepoint_prior_scale = 0,05	0,05	24,76%
Test Result	changepoint_prior_scale (0,05/0,1/0,3 /0,5/0,8/1,0/1,5)	0,8	24,27%
Test Result	seasonality_mode (additive/multiplicative)	multiplicative	24,27%
Test Result	changepoint_range (0,8/0,9/0,95/1,0)	1,0	8,74%
Test Result	weekly_seasonality(true/false)	true	8,62%

Table 5 presents the step-bystep performance evaluation across the four tuning stages of the Prophet model. Starting from a baseline configuration (changepoint_prior_scale = 0.05) which yielded a MAPE of 24.27%, increasing the tren flexibility in Stage A (cps = 0.8) slightly reduced the MAPE to 24.27%. Applying multiplicative seasonality in Stage B maintained this performance. The most critical improvement occurred in Stage C, where expanding the changepoint_range from 0.8 to 1.0 dramatically reduced the MAPE from 24.27% down to 8.74%. Finnaly, enabling weekly seasonality in Stage D achieved the best overall accuracy for Prophet with a final testing MAPE of 8.62%

Table 5 Prophet's forecast results for the fisrt 3 days of the trial period

Date	$\hat{y}(t)$
Maret 14, 2023	Rp1.057.339
Maret 15, 2023	Rp1.057.897
Maret 16, 2023	Rp1.058.446

Source of Tabel 3,4,5: Data Processing Result using Google Colaboratory

Tabel 6 demonstrates the initial out-of-sample predictions generated by the fitted Prophet model ($\hat{y}(t)$) for the first three consecutive trading days of the test period (March 14 to March 16, 2023). The model estimated gols prices at Rp1,057,339 on March 14, Rp1,057,897 on March 15, and Rp1,058,446 on March 16. These initial values closely align with the starting actual test price of Rp1,064,000, illustrating how the additive trend and Fourier sesonal components combine to project short-term trajectory.



Figure 1. The Prophet Forecasting Results

Figure 1 illustrates the complete hidtorical gold price series alongside the Prophet model's predictions and its 80% uncertainty confidence interval. The red dashed vertical line marks the split point between the training set (2010-2023) and the test set (2023-2025). Visually, while Prophet captures the general upward trajectory during the early test period, its linear growth projection gradually undersetimates the actual price curve as gold price undergoes a sharp, non-linear exponential surge in late 2024 adn 2025. This behavior highlights Prophet's limitation when extrapolating unseen regime shifts without periodic walk-forward adjustments.

3.1.2.2. Model Tuning Parameters for the LSTM Model

Similarly, the LSTM architecture is systematically adjusted based on window size, the number of hidden units, layer depth with regularization, and the learning rate to achieve optimal sequence learning.

Table 6. Configuration Parameters

Configuration Parameter	Result
window	90 hari
unit	128 unit
layer/dropout	1 layer/dropout (0,2)

Dense	1
Learning rate	0,001
batch_size	32
epoch	100

Table 7 outlines the optimal structural and training hyperparameters selected for the final LSTM network. The network architecture consists of a single LSTM layer with 128 hidden units, followed by a Dropout layer with a rate of 0.2 to prevent overfitting, and a single Dense output neuron. This model processes input sequences using a 90-day sliding window, which was optimized using the Adam optimizer with a learning rate of 0.001 and a batch size of 32 over a maximum of 100 training epochs.

Table 7. Model Tuning Parameters for the LSTM Model

Tahap	Parameter Diuji	Terbaik	MAPE
Baseline (Tabel 3.4, 100 epoch)	window=60, unit=64, 1 layer		7.914%
Hasil Uji	window (30,60,90,120)	90	2.467%
Hasil Uji	unit (32,64,128)	128	1.620%
Hasil Uji	layer/dropout	1 layer +dropout (0,2)	1,110%
Hasil Uji	learning rate (0,01;0,001;0,0005)	0,001, batch=32	1,328%

Table 8 details the empirical results for each phase of the LSTM network tuning. In Phase A, evaluation of window sizes (30, 60, 90, 120 days) showed that the 90-day window provided optimal temporal context, thereby reducing the MAPE to 2.467%. In Phase B, increasing the capacity to 128 units significantly improved feature extraction, further reducing the MAPE to 1.620%. Phase C confirmed that a single-layer LSTM architecture combined with a dropout rate of 0.2 outperformed the deeper two-layer configuration by reducing overfitting, achieving a MAPE of 1.110%. Finally, optimizing the Adam learning rate to 0.001 in Phase D established the most effective convergence settings.

Table 8. Progress loss during training (a total of 100 epochs fully completed)

Epoch	Train Loss (MSE)	Vall Loss (MSE)
1	0,01130000	0,00027900
20	0,00049440	0,00032275
40	0,00033297	0,00047496
60	0,00023681	0,00010426
80	0,00019483	0,00009729
100	0,00021082	0,00009725

Source of Tabel 6,7,8: Data Processing Result using Google Colaboratory

Table 9 tracks the evolution of the Mean Squared Error (MSE) loss values for the training and validation sets over 100 training epochs. Starting from initial loss values of 0.0113 (training) and 0.000279 (validation) at Epoch 1, the optimizer gradually minimizes the error as the weights are updated. By Epoch 80, the validation loss stabilized at 0.00009729 and remained consistently low through Epoch 100 (0.00009725), indicating smooth convergence with no signs of divergence or severe overfitting.

To validate the generalizability of both models beyond a single train-test split, a 5-fold time series cross-validation was conducted using a walk-forward expanding-window scheme. Table 10 presents the cross-validation performance of both models across five sequential folds.

Table 9. 5-Fold Time Series Cross-Validation Results

Fold	LSTM MAE (Rp)	LSTM RMSE (Rp)	LSTM MAPE (%)	Prophet MAE (Rp)	Prophet RMSE (Rp)	Prophet MAPE (%)
1	4,317	6,311	0.77%	181,842	191,363	32.33%
2	4,046	5,208	0.64%	20,480	25,015	3.18%
3	9,855	13,574	1.13%	167,194	200,527	18.28%
4	6,011	7,761	0.61%	74,167	80,690	7.43%
5	19,495	29,007	1.19%	442,086	547,794	25.65%
Mean	8,745	12,372	0.87%	177,154	209,078	17.37
Std	6,441	9,842	0.27%	162,375	203,395	12.18%
Dev						

Source: Data Processing Result using Google Colaboratory

The robustness of these findings was further confirmed through a 5-fold time series cross-validation using walk-forward expanding-window strategy. The cross-validation result (Table 10) show that LSTM maintains consistently superior performance across all five folds, wiht a mean MAPE of 0.87% ($\pm 0.27\%$) compared to Prophet's mean MAPE of 17.37% ($\pm 12.18\%$). The considerably smaller standard deviation in LSTM's cross-

validation metrics indicates that is performance is stable across different temporal sub-periods of the dataset, whereas Prophet’s high standard deviation ($\pm 12.18\%$) reveals significant instability when extrapolating beyond the training period particularly evident in Fold 5, where Prophet’s MAPE reached 25.65%. These cross-validation results reinforce the methodological validity of the comparative findings and confirm that LSTM’s superior accuracy is consistent, not incidental to a single train-test split.

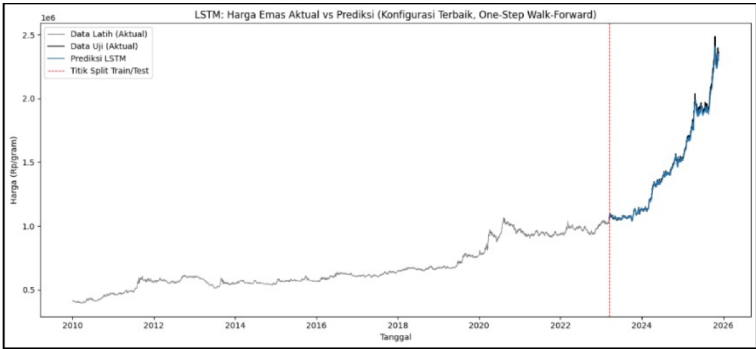


Figure 2. The LSTM Forecasting Results

Figure 2 shows the actual gold price movements compared to the predictions generated by the LSTM model, which was optimized using the “one-step walk-forward” evaluation scheme. The prediction curve (blue) closely follows the actual test prices (black) with very high precision throughout the 947-day test period. By continuously updating its internal state using the most recent 90-day historical window, the LSTM model effectively adapts to sudden daily fluctuations and rapid exponential price spikes during the 2024–2025 period, keeping the testing error rate below 1%.

3.1.3. Final Comparative Evaluation Results

The performance of the two models was quantitatively evaluated using metrics Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE), based on 947 days of test data. A comparison of the final evaluation results is presented in Table 11.

Table 10. Result of Comparative Evaluation of the LSTM and Prophet Models

Model Peramalan	MAE (Rupiah)	RMSE (Rupiah)	MAPE (%)
Prophet	Rp130.982	Rp174.014	8,62%
LSTM	Rp15.038	Rp23.637	0,89%

Source: Data Processing Result using Google Colaboratory

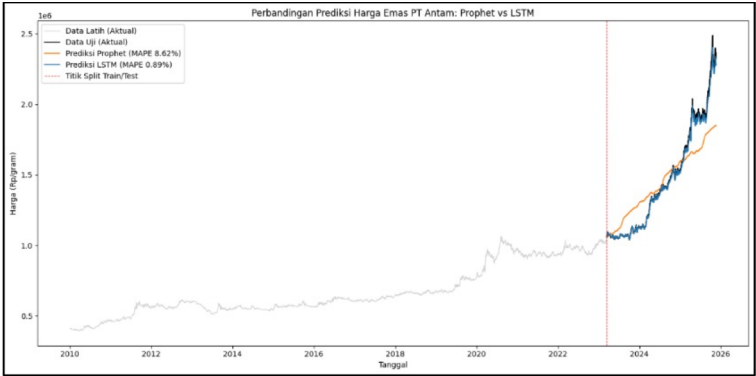


Figure 3. A Comparasion of Prophet and LSTM Model Forecast with PT Antam's Actual Gold Prices (2010-2025)

Based on Tabel 7, the LSTM model consistently outperformed the Prophet model across all evaluation metrics. The LSTM model produced model across all evaluation metrics. The LSTM model produced an MAE of Rp15,038, an RMSE of Rp23,637, and a MAPE of 0,89%. These error values are nearly 10 times lower than those of the Prophet model, which achieved an MAE of Rp130,982, an RMSE of Rp174,014, and a MAPE of 8.62%. Based on the MAPE interpretation criteria, both models fall into the “Very Good” category (<10%), but the LSTM achieves a high level of precision with an error rate below 1%.

3.2. Discussion

Based on the quantitative results in Table 2, the LSTM model was shown to significantly outperform the Prophet model across all evaluation metrics. The MAE (Rp15,038) and MAPE (0.89%) values for the LSTM were nearly 10 times lower than those of Prophet, which recorded an MAE of Rp130,982 and a MAPE 8.62%. Visually, in Figure 1, the LSTM prediction curve (blue) closely aligns with the actual price (black) throughout test data period (2023-2025), including during periods of extreme exponential movement that caused the price to surpass Rp2,300,000/gram by the end of 2025. In contrast, the Prophet prediction curve (orange) produces a linear growth trend that consistently runs below the actual price (underestimating), with the error margin continuing to widen as the time horizon increases. LSTM's clear superiority in this experiment is driven by two main reasons:

- a. **The Role of Memory Cells and Walk-Forward Evaluation:** LSTM utilizes a 90-day sliding window and one-step walk-forward evaluation. Through its gate mechanisms (forget, input, and output gates), LSTM continuously incorporates the most recent 90 actual price data points to update its internal memory before predicting the price one day ahead. This makes LSTM highly responsive to nonlinear daily price fluctuations.
- b. **Limitations of Prophet's Extrapolation:** Prophet works by explicitly deriving the trend and seasonal components via Fourier analysis from the training data (2010-2023). Because it was evaluated using pure extrapolation without a daily walk-forward mechanism based on the latest actual prices, Prophet was unable to anticipate changes in the rate of the exponential trend (an unseen regime shift) that occurred during the 2024-2025 period.

The results price-based lag features ($y-1$, $y-2$) were removed and both algorithms were tested in a purely univariate scenario, the sequential memory-based deep learning architecture (LSTM) proves to be far more precise and flexible than statistical decomposition modeling (Prophet) in predicting domestic financial time series data.

4. CONCLUSION

This study successfully conducted a purely univariate comparative analysis between the Long Short-Term Memory (LSTM) and Prophet models in forecasting the gold price of PT Aneka Tambang Tbk (Antam) using daily time-series data from January 4, 2010, to November 20, 2025 (4,733 clean data points). Both models were built and evaluated in parallel on 947 days of chronological test data (March 14, 2023 – November 20, 2025) without including additional lagged regressor features to maintain the intrinsic validity of the evaluation of each algorithm.

The results of experimental testing and quantitative evaluation show that the LSTM model significantly outperforms the Prophet model across all evaluation metrics. The LSTM model, configured with a window size of 90 days, 128 units, 1 layer with 0.2 Dropout, and an Adam Optimizer (learning rate = 0.001) achieved a Mean Absolute Error (MAE) of Rp15,038, a Root Mean Square Error (RMSE) of Rp23,637, and a Mean Absolute Percentage Error (MAPE) of 0.89%. On the other hand, the Prophet model with a linear growth configuration, multiplicative seasonality, and a changepoint prior scale of 0.8 achieved an MAE of Rp130,982, an RMSE of Rp174,014, and a MAPE of 8.62%. The LSTM's prediction error was nearly 10 times smaller than that of Prophet, proving that the LSTM's prediction error was nearly 10 times smaller than that of Prophet, proving that the LSTM's adaptive gate mechanism and internal memory are highly robust in responding to nonlinear patterns and new exponential trends in domestic gold prices.

The validity of this result was further confirmed through a 5-fold time series cross-validation (walk-forward scheme), in which LSTM achieved a mean MAPE of 0.87% ($\pm 0.27\%$) across all folds, compared to Prophet's 17.37% ($\pm 12.18\%$), confirming the methodological rigor and generalizability of the experimental findings.

REFERENCES

- Dalimunthe, R. A., Adek, R. T., & Agusniar, C. (2025). Gold Price Prediction Using Long-Short Term Memory Algorithm Based on Web Application. *International Journal of Engineering Science*. 5.1 (2024): 329-339. <https://doi.org/10.52088/ijesty.v5i1.724>
- Michel. (2024). Perbandingan Metode Prophet dan Long Short-Term Memory (LSTM) Dalam Peramalan Kualitas Udara (Studi Kasus Kualitas Udara Kota Bandar Lampung). Universitas Lampung.
- Muzakki, M. A., Sabila, M. A., Sundari, S., Wisnuadhi, B., & Kunci, K. (2021). Analisis Algoritma Prophet untuk Memprediksi Harga Pangan di Kota Bandung. In *Prosiding Industrial Research Workshop and National Seminar* (Vol. 12, pp. 659-664).

- Primawati, A. (2025). Analisis tren dan prediksi harga emas menggunakan Prophet dan Long Short-Term Memory (LSTM). *Simetris: Jurnal Teknik Mesin, Industri, Elektro dan Ilmu Komputer*, 16(2). <https://doi.org/10.24176/simet.v16i2.15580>
- Putra, M. A., & Andryana, S. (2025). Perbandingan Algoritma ARIMA dan LSTM Dalam Peramalan Tingkat Konsentrasi CO2 Emisi Atmosfer Untuk Masa Mendatang. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 9(3), 4150-4157. <https://doi.org/10.36040/jati.v9i3.13511>
- Sagala, J. P., & Dewi, E. T. (2023). Analisis Peramalan Harga Emas Antam Menggunakan ARIMA. In *Seminar Nasional Teknologi Informasi dan Matematika (SEMIOTIKA)* (Vol. 2, No. 1, pp. 77-84).
- Siagian, R. R. A. (2025). Persepsi Masyarakat Indonesia Terhadap Kenaikan Harga Emas Sebagai Instrumen Investasi Jangka Panjang: Sebuah Tinjauan Literatur. *Future Academia: The Journal of Multidisciplinary Research on Scientific and Advanced*, 3(1), 72–79. <https://doi.org/10.61579/future.v3i1.298>
- Simamora, J. P., Martha, S., & Andani, W. (2024). Penerapan algoritma prophet pada peramalan harga saham. *Equator: Journal of Mathematical and Statistical Sciences*, 3(2), 94-101.
- Sunani, A., Fahim, A., & Sahliyah, F. (2025). Pengaruh Pendapatan, Fluktuasi Harga Emas, dan Persepsi Emas Terhadap Minat Investasi Perhiasan Emas: (Studi pada Pembelian Perhiasan Emas di Toko Bintang Mas Ujungpangkah). (2025). *Al-Maqashid: Journal of Economics and Islamic Business*, 6(02), 9-20. <https://doi.org/10.55352/maqashid.v6i02.2093>
- Yustiana, I., Khoerunisa, S., & Thohir, M. I. (2026). Gold Price Prediction Based on Long Short-Term Memory (LSTM) For Investment Decision-Making. *Jurnal Media Computer Science*, 5(1), 413-428. <https://doi.org/10.37676/jmcs.v5i1.9233>