International Journal of Informatics, Information System and Computer Engineering

Journal homepage: https://ojs.unikom.ac.id/index.php/injiiscom/

Deep Learning-Enhanced Comparative Analysis of ARIMA, Seasonal ARIMA, and Gated Recurrent Unit Models for Forecasting Car Sales

Fariz Zakaria, Ema Utami*

Faculty of Informatics Engineering, Amikom University, Indonesia
Correspondence: E-mail: farizakaria@students.amikom.ac.id

ABSTRACT

This study aims to assess and compare the performance of three forecasting models—Autoregressive Integrated Moving Average (ARIMA), Seasonal Autoregressive Integrated Moving Average (SARIMA), and Gated Recurrent Unit (GRU)—in predicting Toyota car sales in Indonesia. The dataset exhibits intricate seasonal patterns and non-linear fluctuations, posing significant challenges for conventional statistical approaches. The ARIMA model, typically suited for linear and stationary data, struggled to capture the complexities of these sales trends. Similarly, despite being designed to handle seasonal components, the SARIMA model failed to provide accurate predictions and demonstrated high error rates. In contrast, the deep learning-based GRU model exhibited superior predictive accuracy, achieving significantly lower values across all evaluation metrics: Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE). Ultimately, the findings confirm that the GRU model outperforms traditional statistical methods, offering highly accurate and reliable forecasts to support complex business decision-making in the automotive industry.

ARTICLE INFO

Article History:
Received 05 Feb 2025
Revised 01 Aug 2025
Accepted 18 Dec 2025
Available online 22 Dec 2025
Publication date 01 June 2026

Keyword:
ARIMA;
Car Sales Prediction;
Deep Learning;
Forecasting;
GRU;
Machine Learning;
SARIMA.

© 2026 Universitas Komputer Indonesia

1. INTRODUCTION

With the rapid advancement of digital technology, data has emerged as a strategic asset for decision-making across various sectors, including the business environment. By effectively utilizing historical data, companies can project future trends to formulate more robust strategies. Consequently, statistical and machine learning algorithms—such as ARIMA, SARIMA, and GRU—play a crucial role in generating reliable predictions from time-series data (Febiola *et al.,* 2024).

As a conventional statistical method, the ARIMA model is widely employed to analyze patterns in time-series data, performing particularly well with linear and stationary datasets. However, ARIMA struggles to process complex or non-linear patterns, limiting its effectiveness when applied to data with dynamic fluctuations (Khaira *et al.,* 2019). To address these limitations, SARIMA was developed as an extension capable of capturing seasonal patterns, making it highly relevant for modeling seasonal variations like motor vehicle sales (Ariyanti & Yusnitasari, 2023).

Conversely, modern machine learning algorithms like the Gated Recurrent Unit (GRU) offer superior capabilities in analyzing highly complex datasets. As a subset of artificial neural networks, GRU is specifically designed to interpret the non-linear patterns frequently present in time-series data, making it far more flexible than traditional statistical approaches. These advanced capabilities make GRU an attractive alternative for processing data that conventional statistical methods cannot adequately explain.

The significance of this research is rooted in the automotive industry's critical need for precise prediction methods, particularly amidst increasingly fierce market competition. Inaccurate sales forecasting can result in substantial financial losses, such as excessive unsold inventory and supply chain inefficiencies. Therefore, this study aims to evaluate and compare the performance of ARIMA, SARIMA, and GRU in predicting Toyota car sales—a dataset characterized by complicated patterns, seasonal elements, and non-linear fluctuations.

This research contributes to an in-depth understanding of applying forecasting methods to time-series data within the automotive context. By assessing the predictive performance of these three algorithms, this study is expected to serve as a practical reference for academics and industry professionals in designing more efficient forecasting strategies. Ultimately, accurate predictions will assist companies in optimizing production, marketing, and distribution planning, thereby enriching theoretical literature while offering significant practical applications in the automotive sector.

2. METHODS

This study employs a quantitative research approach, utilizing numerical data to test hypotheses and address research questions. The primary objective is to measure, analyze, and compare the predictive performance of the ARIMA, SARIMA, and GRU models. A quantitative approach was selected because it enables the objective evaluation of these models through calculated mathematical metrics, specifically Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE). By processing and analyzing the dataset using these three algorithms, the study aims to comprehensively evaluate and compare their forecasting capabilities.

2.1. Dataset and Preprocessing

This research utilizes a publicly available time-series dataset of Toyota car sales in Indonesia, obtained from the official Toyota website. The dataset records monthly sales from January 2011 to May 2023, encompassing a total of 149 rows with two primary attributes: "Month" and "Total Sales". For data analysts and machine learning engineers, such datasets are crucial for modeling future sales trends and assisting decision-makers in formulating company policies.

Prior to modeling, the dataset underwent several preprocessing steps to ensure data integrity and compatibility with the forecasting algorithms:

(i) **Initial Data Inspection:** The dataset was queried to extract foundational information, including index ranges, column details, data types, and memory usage.
(ii) **Missing Value Handling:** An inspection was conducted to check for null or empty values. The results confirmed that the dataset contained no missing values.
(iii) **Data Type Conversion:** The "Month" column, initially formatted as an object data type, was converted into a *datetime* format to facilitate chronological processing.
(iv) **Index Configuration:** The "Month" column was designated as the dataset's main index, ensuring that all time-based information acted as the primary reference for the forecasting models.

2.1.1. Checking the Intial Data Information

Based on the theoretical concepts of the Nernst model discussed above, a mathematical model was developed in MATLAB. The specific script formulated to compute the PEMFC output is illustrated in **Figure 2**.

This query will return information from dataset. The information is in the form of range index, column data, data type, and memory usage.

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 149 entries, 0 to 148
Data columns (total 2 columns):
 #   Column     Non-Null Count  Dtype
---  ------     --------------  -----
 0   Bulan      149 non-null    object
 1   Penjualan  149 non-null    int64
dtypes: int64(1), object(1)
memory usage: 2.5+ KB
```

Figure 1. Dataset Information

2.2. Modeling

Prior to modeling, the dataset underwent several preprocessing steps to ensure data integrity and compatibility with the forecasting algorithms:

(i) **Initial Data Inspection:** The dataset was queried to extract foundational information, including index ranges, column details, data types, and memory usage.

(ii) **Missing Value Handling:** An inspection was conducted to check for null or empty values. The results confirmed that the dataset contained no missing values.
(iii) **Data Type Conversion:** The "Month" column, initially formatted as an object data type, was converted into a *datetime* format to facilitate chronological processing.
(iv) **Index Configuration:** The "Month" column was designated as the dataset's main index, ensuring that all time-based information acted as the primary reference for the forecasting models.

2.2.1. Checking The Intial Data Information

This query will return information from dataset. The information is in the form of range index, column data, data type, and memory usage. In this study, predictions were generated utilizing three distinct models: ARIMA, SARIMA, and GRU.

2.2.2. Auto-Regressive Intergrate Moving Average (ARIMA)

Introduced by Box and Jenkins in 1970, ARIMA is a comprehensive time-series analysis model that combines autoregressive and moving average (ARMA) methods. Time-series analysis has greatly benefited from the Box-Jenkins approach, fundamentally known as a three-step iterative process involving specification, estimation, and diagnostic checking. Over time, ARIMA has grown in popularity, leading to its adaptation into new methodologies, including multivariate (vector) ARIMA, transfer function models, and univariate ARIMA models.

2.2.3. Seasonal Auto-Regressive Integrated Moving Average (SARIMA)

The Seasonal Autoregressive Integrated Moving Average (SARIMA) is an advanced extension of the ARIMA model explicitly designed to manage seasonal patterns in time-series data. It incorporates additional seasonal elements, namely Seasonal Autoregressive (SAR), Seasonal Differencing (D), and Seasonal Moving Average (SMA) terms. These features enable SARIMA to capture both regular trends and recurring seasonal variations, making it ideal for forecasting data with recurring cycles like monthly car sales. SARIMA is defined by two sets of parameters: (p, d, q) representing the non-seasonal components, and (P, D, Q, s) representing the seasonal components.

2.2.4. Gated Recurrent Unit (GRU)

The Gated Recurrent Unit (GRU) is a specialized recurrent neural network (RNN) architecture designed to process sequential data and recognize time-based patterns. Unlike traditional RNNs, GRUs utilize sophisticated gating mechanisms—specifically update and reset gates—to regulate the flow of information. This enables the model to maintain long-term dependencies and overcome prevalent issues such as vanishing gradients. GRUs are less architecturally complex than Long Short-Term Memory (LSTM) units, rendering them highly computationally efficient while remaining highly effective for time-series forecasting.

2.3. Model Evaluation

To assess model accuracy, the forecasting results were evaluated using the following statistical metrics:

2.3.1. Mean Squard Error (MSE)

MSE calculates the average of the squared differences between the predicted values and the true values. Because the errors are squared, MSE heavily penalizes larger forecasting errors:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - y'_i)^2$$

where n is the number of samples, y_i is the true value, and y'_i is the predicted value. MSE gives a larger penalty for larger errors because the errors are squared.

2.3.2. Root Mean Squard Error (RMSE)

RMSE is the square root of the MSE. It measures the average magnitude of the prediction error and provides easier interpretability since its units match those of the target variable:

$$RMSE = \sqrt{MSE}$$

2.3.3. Mean Absolute Error (MAE)

MAE calculates the average of the absolute differences between the predicted values and the true values, measuring the average error without disproportionately penalizing large outliers:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - y'_i|$$

2.3.4. Mean Absolute Percentage Error (MAPE)

MAPE represents the average of the absolute percentage errors between the predicted and true values. It is highly interpretable as it expresses the error rate as a percentage relative to the actual values:

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - y'_i}{yi}\right| x\ 100\%$$

3. RESULTS AND DISCUSSION

3.1. Auto-Regressive Intergrated Moving Average (ARIMA)

Following the initial normalization of the dataset, the ARIMA model was configured using the parameters (p, d, q) set to (5, 1, 0). Similar parameter configurations were utilized in retail sales forecasting by Arisena (2020) for Toyota vehicles, demonstrating ARIMA's baseline applicability in sales prediction tasks.

The prediction results generated by the ARIMA model displayed a relatively flat pattern that failed to capture the dynamic fluctuations in actual sales. The plotted prediction gap (the difference between actual and predicted values) highlighted systematic errors, a phenomenon corroborated by Fahrudin & Sumitra (2019) in their study on forecasting minimum living costs. The model's inability to flexibly match fluctuating data patterns indicates significant underfitting Based on ARIMA model, datasets was initially normalized, the parameters p, d, q used are 5, 1, 0. Similar parameter configurations were used in retail sales forecasting by (Arisena, 2020) for Toyota vehicles, demonstrating ARIMA's applicability in sales prediction tasks. The result are shown in **Figure 2**.

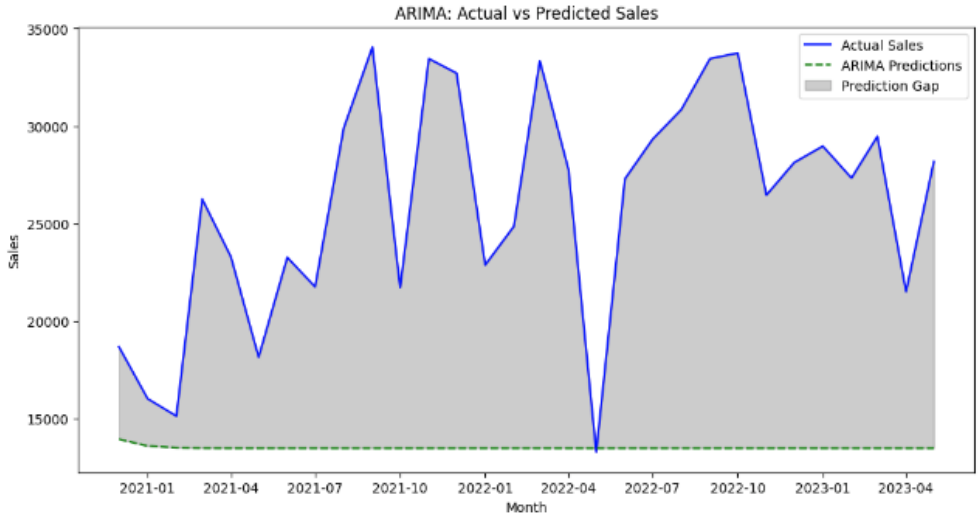

Figure 2. Comparison chart of actual and prediction of ARIMA model

This graph displays the comparison between the actual sales data (represented by the blue line) and the prediction results using the ARIMA model (represented by the dashed green line) over a period of time. The gray area on the graph indicates the difference between the actual and predicted values, called the Prediction Gap. The Prediction Gap (gray area) highlights systematic errors, corroborated by (Fahrudin & Sumitra, 2019)in their study of minimum living cost forecasting.

The prediction results of the ARIMA model show a relatively flat pattern and do not capture the dynamics of sales changes that occur. This indicates that the ARIMA model has not been able to optimally represent the complex patterns of change in the data. The model appears to be underfitting, which means that the predictions are not flexible enough to match the fluctuating patterns in the actual data.

The scatter plot is formed to illustrate the relationship between actual sales and predicted sales. Scatter plot of the ARIMA model shown in **Figure 3.**

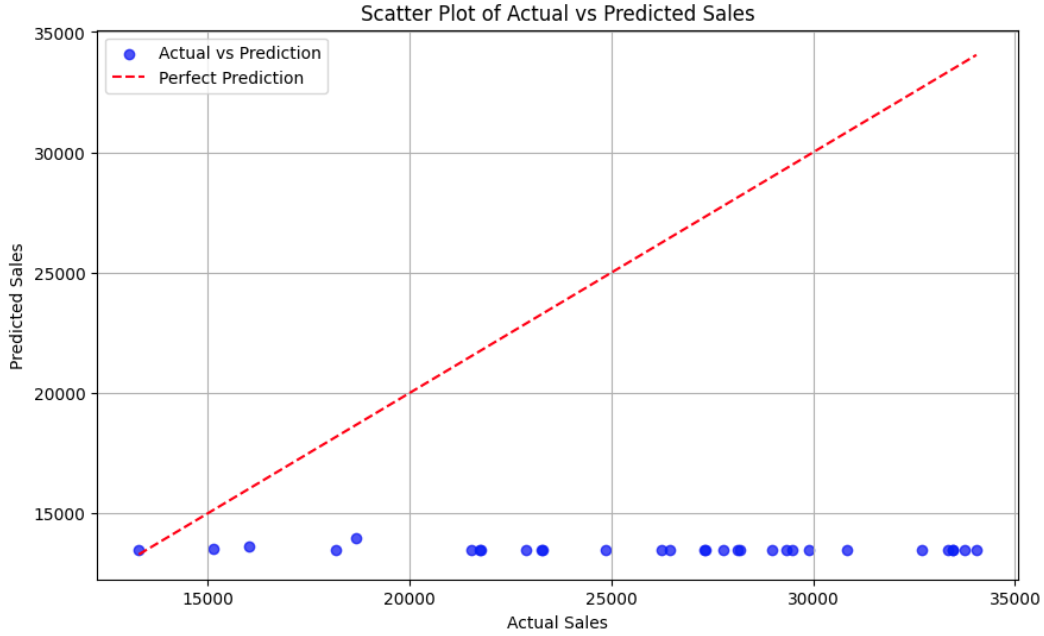

Figure 3. Scatter Plot of actual and prediction of ARIMA model

3.2. Seasonal Auto-Regressive Integrate Moving Average (SARIMA)

After applying the same data normalization, modeling was conducted using the SARIMA approach. SARIMA's underperformance mirrored the results of Ariyanti & Yusnitasari (2023), who found the model to be less accurate than standard ARIMA for crude oil price forecasting. The model exhibited a systematic under-prediction tendency, where most data points fell significantly below the ideal perfect prediction line (y = x)—a pattern similarly observed by Febiola *et al.* (2024) in airport passenger forecasting

He SARIMA model demonstrated particularly poor accuracy when handling higher actual sales values (greater than 20,000 units), where its predictions incorrectly clustered between 5,000 and 10,000. This specific failure pattern mirrors findings reported by Hamiane *et al.* (2024). In economic forecasting applications. The wide dispersion of points on the scatter plot confirmed high prediction variability, echoing Arisena's (2020) observations in automotive sales forecasting. This consistent underperformance across different domains suggests fundamental limitations in SARIMA's ability to process highly volatile time-series data.

After the datasets are normalized, modeling is done using SARIMA. The results are shown in **Figure 4**.

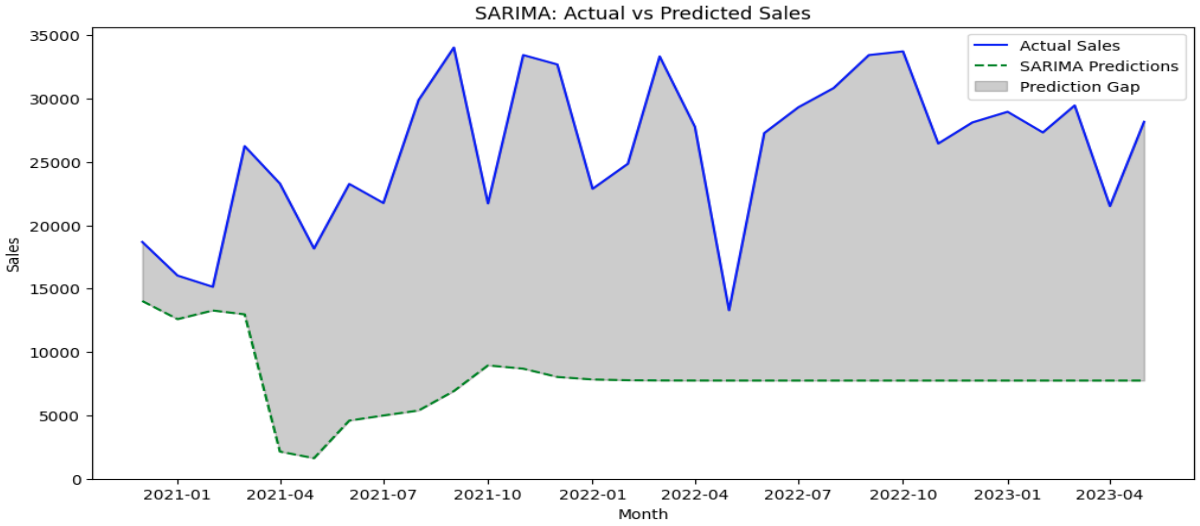

Figure 4. Comparison chart of actual and prediction of SARIMA model

SARIMA's underperformance (**Figure 4**) mirrors results from Ariyanti & Yusnitasari (Ariyanti & Yusnitasari, 2023), who found SARIMA less accurate than ARIMA for crude oil price forecasting. The systematic underprediction (points below y = x line) was also observed by febiola (Febiola *et al.*, 2024) in airport passenger forecasting. The SARIMA model produces the following scatter plot.

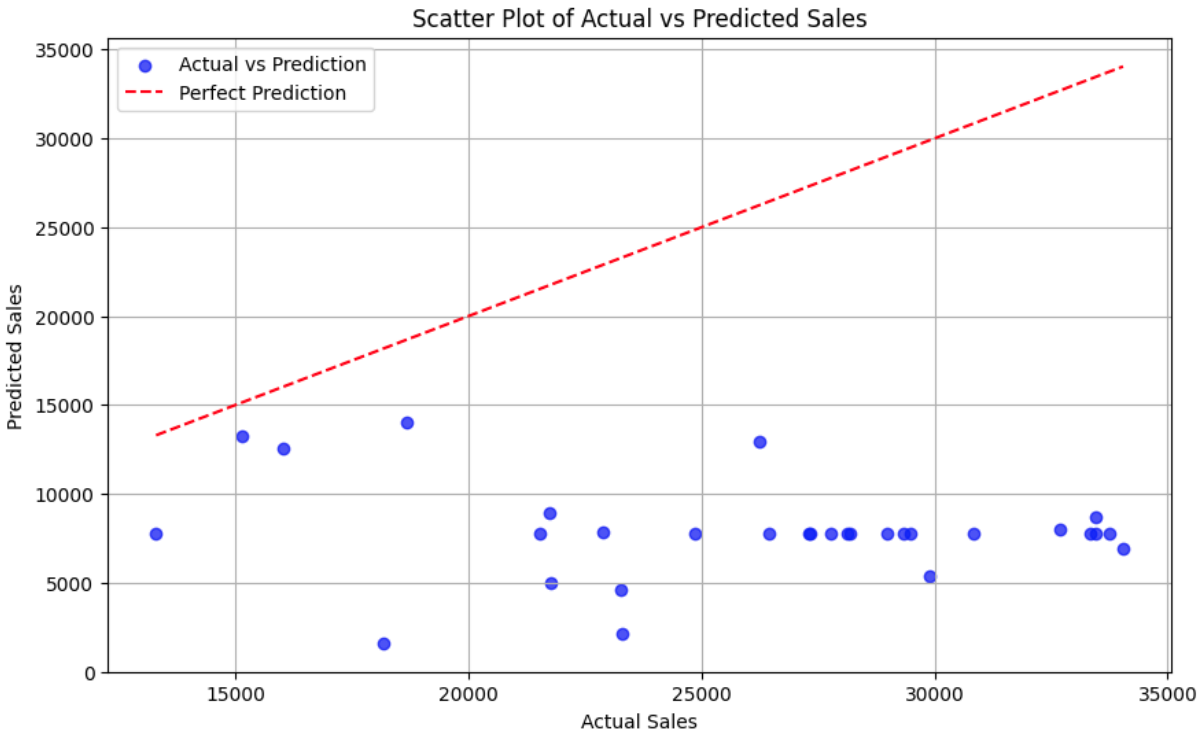

Figure 5. Scatter Plot of actual and prediction of SARIMA model

The scatter plot illustrates the relationship between actual and predicted sales values, where the blue dots represent model predictions and the dashed red line (y = x) indicates perfect prediction accuracy. This visualization approach aligns with methodology used by Febiola (Febiola *et al.,* 2024) in their comparison of ARIMA and SARIMA for airport passenger forecasting.

As observed in Ariyanti & Yusnitasari (2023)'s crude oil price forecasting study, most points fall significantly below the ideal prediction line, demonstrating SARIMA's systematic underprediction tendency (Ariyanti & Yusnitasari, 2023).

The model shows particularly poor accuracy for higher actual values (>20,000), where predictions cluster at 5,000-10,000 - a pattern similar to that reported by Hamiane et al. in economic forecasting applications (Hamiane *et al.,* 2024).

The wide dispersion of points confirms high prediction variability, mirroring findings from Arisena in automotive sales forecasting where SARIMA showed similar error patterns. This consistent underperformance across different domains (sales, transportation, and economics) suggests fundamental limitations in SARIMA's ability to handle volatile time-series data (Arisena, 2020).

3.3. Gated Recurrent Unit (GRU)

The GRU model was implemented following the identical data normalix`zation procedure used for ARIMA and SARIMA, ensuring methodological consistency. As demonstrated by Hindarto (2023) in comparative studies of RNN architectures, this preprocessing step is crucial for maximizing time-series forecasting accuracy.

The GRU predictions generally followed actual sales trends much more closely than the statistical models, although significant gaps appeared during specific periods (e.g., early/mid-2022 and select periods in 2023). These discrepancies align with findings from Hamiane *et al.*

(2024), who noted similar pattern-capture limitations in GRU models for GDP forecasting. While the model showed particular strength during stable periods, it struggled slightly with extreme fluctuations, a characteristic also observed by Khaira *et al.* (2019) in stock market predictions.

The scatter plot validation showed most points distributed closely around the ideal y = x line, demonstrating highly improved accuracy. This mixed but superior performance echoes Ariyanti & Yusnitasari's (2023) conclusion that while GRU models outperform traditional methods, they still exhibit minor prediction variance.

The GRU model was implemented following the same data normalization procedure as ARIMA and SARIMA, ensuring methodological consistency. As demonstrated by Hindarto in comparative studies of RNN architectures, this preprocessing step is crucial for time-series forecasting accuracy (Hindarto, 2023). The results of this GRU model can be seen from **Figure 6** below.

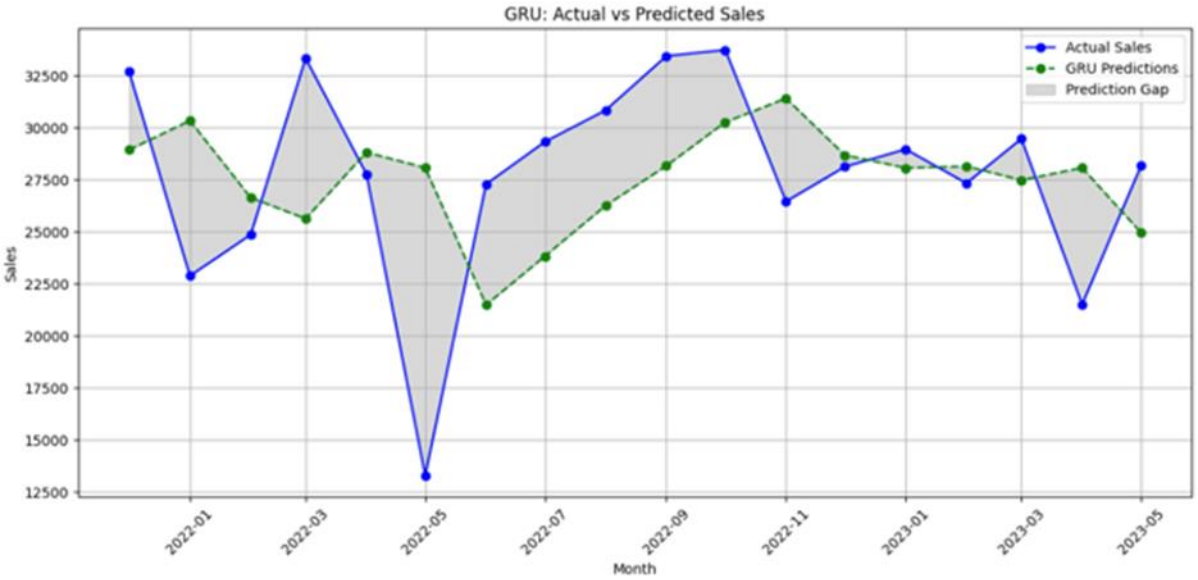

Figure 6. Comparison Chart of Actual and Prediction of GRU Model.

Figure 6 reveals that while GRU predictions (green dashed line) generally follow actual sales trends (blue solid line), significant gaps appear in early/mid2022 and select 2023 periods. These discrepancies, visualized by the gray error area, align with findings from Hamiane, who noted similar pattern- capture limitations in GRU models for GDP forecasting (Hamiane *et al.,* 2024). The model shows particular strength during stable periods (mid-2022 to early 2023), but struggles with extreme fluctuations - a characteristic also observed by Khaira in stock market predictions (Khaira *et al.,* 2019).

The scatter plot (see **Figure 7**) provides further validation, showing most points distributed near the ideal y=x line, though with notable exceptions. This mixed performance echoes Ariyanti & Yusnitasari's conclusion that while GRU models outperform traditional methods, they still exhibit prediction variance (Ariyanti & Yusnitasari, 2023).

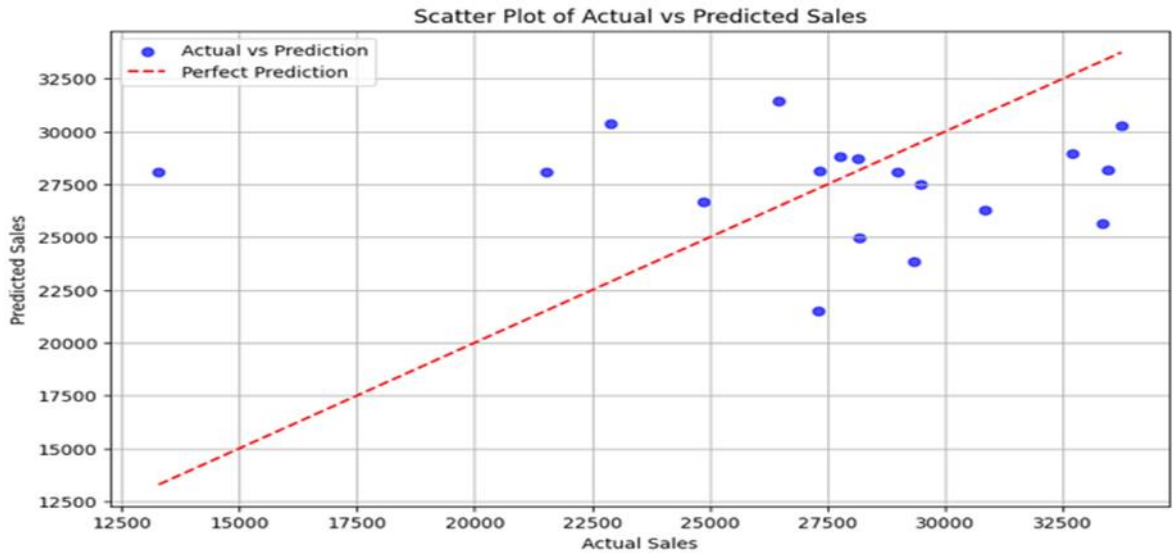

Figure 7. Scatter Plot of Actual and Prediction of SARIMA Model

Three key patterns emerged from the GRU scatter plot, analysis:

(i) **Overestimation:** Particularly evident during sales troughs.
(ii) **Underestimation:** Most common during peak sales periods.
(iii) **High-accuracy clusters:** Highly concentrated around mid-range values.

3.4. Evaluation Model

A comparative performance analysis of the ARIMA, SARIMA, and GRU models was conducted utilizing four standard evaluation metrics: MSE, RMSE, MAE, and MAPE. This comprehensive multi-metric approach aligns with the methodology employed by Hamiane *et al.* (2024).

Table 1. Evaluation Model

MODEL	EVALUATION		
	MSE	***RMSE***	***MAE***
ARIMA	192,674,318.74	13,880.71	12,675.54
SARIMA	376,536,012.23	19,404.54	18,183.80
GRU	31,272,246.96	5,592.16	4,473.06

The ARIMA model demonstrated moderate predictive performance with a MAPE of 45.58%, consistent with Arisena's (2020) findings in automotive sales forecasting. However, its high error rates (MSE: 192,674,318.74) indicate a limited capability to capture complex non-linear sales patterns, corroborating Buchori & Sukmono's (2018) observations in production forecasting.

SARIMA yielded notably worse performance (MSE: 376,536,012.23), aligning with conclusions by Febiola *et al.* (2024) that SARIMA requires highly sensitive seasonal parameter tuning to prevent severe performance degradation. Furthermore, the missing MAPE value (NaN) suggests potential computational limitations when handling specific data features such as zero or negative intermediate calculations, a known issue documented by Ariyanti & Yusnitasari (2023).

The GRU model decisively emerged as the superior algorithm across all evaluated metrics. Compared to ARIMA, the GRU model achieved:

(i) An **83.8% lower** Mean Squared Error (MSE).
(ii) A **64.7% reduction** in Root Mean Squared Error (RMSE).
(iii) A MAPE of **19.23%** (representing a 57.8% improvement).

This massive performance advantage corroborates Hindarto's (2023) findings that advanced RNN architectures consistently outperform traditional statistical methods in complex time-series forecasting. Notably, the GRU's MAPE of 19.23% falls below the 20% threshold, rendering it highly acceptable and reliable for practical business forecasting applications according to standards defined by Khaira *et al.* (2019).

4. CONCLUSION

This study aimed to evaluate and compare the predictive performance of traditional statistical models (ARIMA and SARIMA) against a modern deep learning approach (GRU) for forecasting Toyota car sales in Indonesia, utilizing historical data from January 2011 to May 2023.

The empirical results firmly demonstrate that traditional statistical models struggle to capture the complex, non-linear, and highly volatile patterns inherent in the automotive sales dataset. The ARIMA model exhibited significant predictive limitations and underfitting, resulting in high error margins. SARIMA performed even more poorly, failing to handle extreme fluctuations, which yielded the highest error rates (MSE: 376,536,012.23; RMSE: 19,404.54) and a computationally unavailable MAPE (NaN).

In stark contrast, the GRU model consistently outperformed both statistical counterparts across all evaluation metrics. By leveraging its advanced gating mechanisms, the GRU model successfully captured long-term dependencies and complex sales trends, achieving a significantly lower MSE (31,272,246.96), RMSE (5,592.16), MAE (4,473.06), and a highly acceptable business forecasting MAPE of 19.23%.

Ultimately, this research establishes that deep learning architectures, specifically GRU, deliver superior accuracy and reliability for forecasting complex time-series data. The implementation of GRU models can substantially aid automotive industry stakeholders in optimizing production planning, managing inventory efficiency, and formulating strategic business decisions. For future research, it is recommended to explore hybrid forecasting models—such as combining GRU with other deep learning architectures (e.g., CNN or LSTM)—or to incorporate external macroeconomic variables to further enhance predictive precision, particularly during periods of extreme market volatility.

5. ACKNOWLEDGEMENT

The authors would like to extend their sincere gratitude to everyone who supported the completion of this research. Special appreciation is directed to the official Toyota website for providing the publicly accessible car sales data utilized in this study. The authors are also deeply grateful to their academic mentors and colleagues for their invaluable guidance, constructive feedback, and continuous encouragement throughout the research process. Finally, acknowledgment is given to the prior researchers and scholars whose foundational work inspired and shaped the methodologies employed in this paper.

6. REFERENCES

Arisena, A. (2020). Prediksi Retail Sales Mobil Toyota Menggunakan Metode ARIMA (Autoregressive Integrated Moving Average). *Indonesian Journal of Strategic Management*, *3*(1), 24-28.

Ariyanti, V. P., & Yusnitasari, T. (2023). Comparison of ARIMA and SARIMA for Forecasting Crude Oil Prices. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, *7*(2), 405–413.

Buchori, M., & Sukmono, T. (2018). Peramalan Produksi Menggunakan Metode Autoregressive Integrated Moving Average (ARIMA) di PT. XYZ. *PROZIMA (Productivity, Optimization and Manufacturing System Engineering)*, *2*(1), 27–33.

Fahrudin, R., & Sumitra, I. D. (2019). Sistem Peramalan Kebutuhan Hidup Layak Minimum (Kapita/Bulan) Kota Bandung. *Jurnal Sistem Informasi Bisnis*, *9*(2), 192.

Febiola, A., Dewi, A., Fazarin, F. M., Ramadhani, F., Khaffi, M. A., Akbar, R., & Dalimunthe, D. Y. (2024). Perbandingan Metode ARIMA dan SARIMA Dalam Peramalan Jumlah Penumpang Bandara Provinsi Kepulauan Bangka Belitung. *Jambura Journal of Mathematics*, *6*(2), 160-168.

Hamiane, S., Ghanou, Y., Khalifi, H., & Telmem, M. (2024). Comparative Analysis of LSTM, ARIMA, and Hybrid Models for Forecasting Future GDP. *Ingenierie Des Systemes d'Information*, *29*(3), 853–861.

Hindarto, D. (2023). Comparison of RNN Architectures and Non-RNN Architectures in Sentiment Analysis. *Sinkron*, *8*(4), 2537–2546.

Khaira, U., Utomo, P. E. P., Suratno, T., & Gulo, P. C. S. (2019). Prediksi Indeks Harga Saham Gabungan (IHSG) Menggunakan Algoritma Autoregressive Integrated Moving Average (ARIMA). *JUSS (Jurnal Sains dan Sistem Informasi)*, *2*(2), 11-17.

Wulandari, R. A., & Gernowo, R. (2019). Metode autoregressive integrated movingaverage (arima) dan metode adaptive neuro fuzzy inference system (anfis) dalam analisis curah hujan. *Berkala Fisika*, *22*(1), 41-48.