Terbit online pada laman web jurnal: **https://ejurnalunsam.id/index.php/jicom/**

Jurnal Informatika dan Teknologi Komputer

Vol. 06 No. 02 (2021) 149 - 152 E-ISSN :2774-7115 P-ISSN: 2775-2089

Analysis of the Regularization Techniques in the GRU Method for Forecasting Toyota Car Sales

Fariz Zakaria[1], Ema Utami[2]

[1] *Master of Informatics EngineeringUniversitas Amikom Yogyakarta*

[2] *Master of Informatics EngineeringUniversitas Amikom Yogyakarta*

[1] farizakaria@students.amikom.ac.id* Ema@amikom.ac.id

Abstract

In the automotive industry, sales prediction is an important element in supporting strategic decision making, including production planning and supply chain management. This research focuses on applying the Gated Recurrent Unit (GRU) method to predict Toyota car sales in Indonesia based on monthly data from January 2011 to May 2023. GRU, as part of an artificial neural network, offers advantages in handling complex non-linear patterns in time series data. However, GRU also has challenges such as the risk of overfitting and the need for complex parameter tuning. Therefore, in this study, various regularization techniques such as dropout, batch normalization, L1 and L2 penalties, and early stopping are applied to improve model generalization. The evaluation results showed that the application of regularization techniques significantly improved the model performance, with a decrease in MAE by 20.9%, MSE by 23.1%, RMSE by 12.3%, and MAPE by 18.5% compared to the initial model. These findings suggest that the combination of GRU with regularization techniques can be an effective approach for sales prediction in the context of time series data.

Keywords: sales prediction, GRU, Optimization, Regularization,

1. Introduction

In the automotive industry, making sales predictions is a very important aspect to support strategic decision-making, such as production scheduling and supply chain management. Toyota, as one of the largest automotive manufacturers in the world, has sales data that shows seasonal patterns and fluctuations influenced by economic factors and market dynamics. As more historical data becomes available, data-driven analytics approaches are increasingly needed to improve prediction accuracy and reduce risks in business processes.

Predicting time series data presents its own challenges, especially related to stationarity issues, overfitting risks, and model fit to non-linear data patterns. Time series generally have seasonal elements and long-term trends that need to be carefully accounted for in the modeling and predictive analysis process.

Some statistical algorithms that are commonly used in time series data prediction include ARIMA, SARIMA, and LSTM. In addition, there are also other algorithms such as the Gated Recurrent Unit (GRU), which belongs to the modern machine learning category and is known to perform well in processing data with high complexity. GRU is part of an artificial neural network architecture designed to identify non-linear patterns in time series data, making it more adaptive than conventional statistical methods.

While GRU has the advantage of recognizing non-linear patterns and handling short-term fluctuations effectively, it also has its own challenges, such as the potential for overfitting and the need for a more complicated parameter tuning process. Compared to LSTM, GRU tends to be more efficient in some aspects, but less optimal in capturing long-term patterns with precision.

To overcome the weaknesses of the GRU method, regularization techniques are applied. The application of regularization techniques to the GRU (Gated Recurrent Unit) method aims to prevent overfitting and improve model generalization, especially in sequential data processing tasks such as NLP (Natural Language Processing) or time-series prediction.

2. Research Metdhology

2.1. Dataset

The research started with a dataset of Toyota car sales in Indonesia. The data is public data obtained from Kaggle. The dataset contains data on Toyota car sales in Indonesia per month from January 2011 to May 2023.

Diterima Redaksi : 14-09-2025 | Selesai Revisi : 25-09-2025 | Diterbitkan Online : 01-10-2025

The dataset contains attributes such as month and total sales with a total of 149 rows of data.

For data analysts and machine learning engineers, this dataset can be used to model and predict sales in the future. Various machine learning algorithms, such as SARIMA, ARIMA, GRU, etc., can be used to predict car sales that can help decision makers to make and determine company policies.

Before the dataset is used, several things are done, namely:

a. Checking the initial data information

```
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
```

Fig1. Dataset Information

b. Checking for empty value

```
df.isnull().sum().sum()
```

```
0
```

Fig 2. Missing value information

c. Change the required data type
The column "Month" in the data is known to be an object data type, so that it can be used in the next process, the data type is changed to datetime data type.

```
# Mengubah kolom 'Bulan' ke tipe datetime
df['Bulan'] = pd.to_datetime(df['Bulan'])

# Mengatur kolom 'Bulan' sebagai indeks
df.set_index('Bulan', inplace=True)
```

Fig 3. Changing the data type of the month column

d. Setting as index

```
# Mengatur kolom 'Bulan' sebagai indeks
df.set_index('Bulan', inplace=True)
```

Fig 4. Setting the month column to index

2.2. Modeling

A. Gated Recurrent Unit (GRU)

The Gated Recurrent Unit (GRU) is a specialized type of recurrent neural network (RNN) designed to process sequential data and understand time-based patterns. Unlike traditional RNNs, GRUs use gating mechanisms—update and reset gates—to regulate the flow of information, which helps the model maintain long-term dependencies and overcome issues like vanishing gradients. GRUs are less complex compared to Long Short-Term Memory (LSTM) units, making them more efficient while still being effective for tasks like time series forecasting and sequence analysis.

B. Regularization Technique

The application of regularization techniques to the GRU (Gated Recurrent Unit) method aims to prevent overfitting and improve model generalization, especially in sequential data processing tasks such as NLP (Natural Language Processing) or time-series prediction.

The optimization techniques used are

a) L1 and L2 regularization:

Added to kernel and recurrent weights to prevent overfitting and combination of L1 (0.01) and L2 (0.01) helps simplify the model

b) Dropout:

Dropout layers (0.2 and 0.1) to prevent co-adaptation of neurons. Recurrent dropout (0.2 and 0.1) specific to RNN/GRU

c) Batch Normalization:

Added after each GRU layer to stabilize learning

d) Deeper Architecture:

Using 2 GRU layers (64 and 32 units) with an additional Dense layer. More variable layer sizes for stepwise feature extraction

e) Smaller Learning Rate:

Adam optimizer with learning rate 0.001 for more stable convergence

f) Early Stopping:

Stop training when validation loss does not improve for 15 epochs and Restore best weights

g) Validation Split:

20% of training data is used for validation during training

After optimization, the model should show: Lower MAE, MSE, RMSE, and MAPE values, More stable training and validation loss curves, Better generalization to test data, and Fewer signs of overfitting

2.3. Model Evaluation

A. MSE

MSE is the average of the squares of the difference between the predicted value and the true value.

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

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.

B. RMSE

RMSE is the square root of MSE. This provides an easier interpretation as the units of RMSE are the same as the units of the target variable:

$$RMSE = \sqrt{MSE}$$

RMSE measures how large the prediction error is on average.

C. MAE

MAE is the average of the absolute value of the difference between the predicted value and the true value.

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

MAE provides a measure of the average error without penalizing larger errors like MSE.

D. MAPE

MAPE is the average of the absolute value of the percentage error between the predicted value and the true value.

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

MAPE gives the average percentage error, making it easier to understand the error rate relative to the true value.

3. Result and Discussion

3.1. Before Optimization

The results of this GRU model before optimization can be seen from figure 5 below

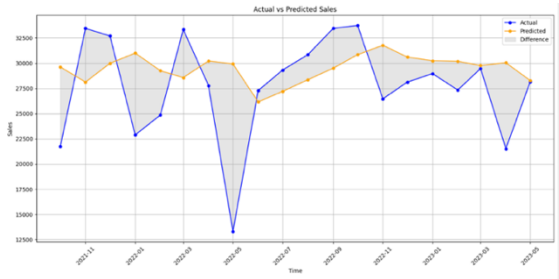


Fig 5. Result Of GRU before Optimization

The actual sales data is colored blue while the GRU before Optimize is orange. If seen, there is still a slight difference between the actual data and the prediction data which is marked in gray.

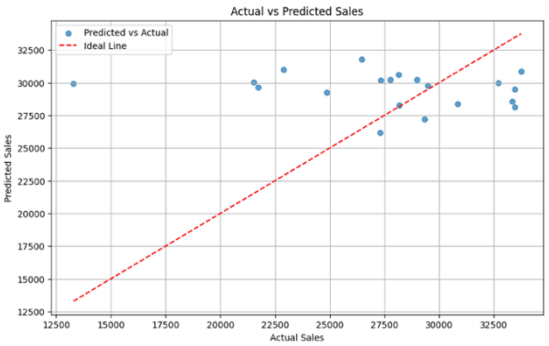


Fig 6. Scatter Plot GRU Before Optimization

3.2. After Optimization

Cara mudah membuat layout adalah dengan menggunakan panduan ini secara langsung.

The results of GRU model after optimization can be seen from figure 7 below

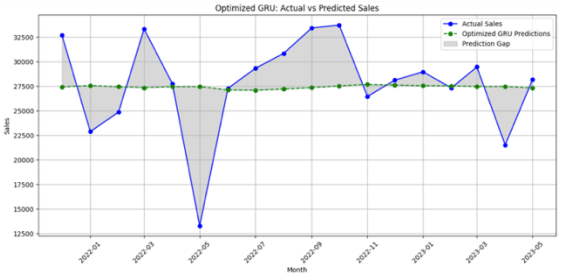


Fig 7. Result Optimized GRU

The actual sales data is colored blue while the Optimized GRU prediction is green. Scatter Plot in figure 8

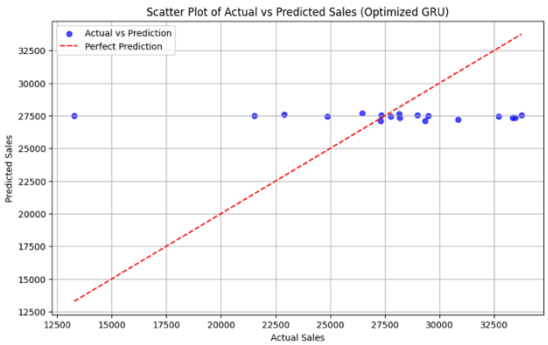


Fig 8. Scatter plot of optimized GRU

3.3. Comparisson Result

In comparing the models, the author uses MSE, RMSE, MAE, and MAPE to evaluate them. To describe the evaluation of the three models can be seen from table 1

Tabel 1.Model Evaluation

Evaluation	*GRU*	*Optimized GRU*	*Improvement*
MSE	31191952.90	23981185.64	23.1%
RMSE	5584.98	4897.06	12.3%

Evaluation	*GRU*	*Optimized GRU*	*Improvement*
MAE	4460.72	3528.87	20.9%
MAPE	19.02%	15.50%	18.5%

4. Conclusion

This research proves that the use of GRU models combined with regularization techniques can significantly improve the accuracy of Toyota car sales predictions. Various optimization methods such as dropout, batch normalization, L1/L2 penalty, as well as the application of early stopping proved effective in reducing the risk of overfitting and improving the generalization ability of the model to test data. Compared to the initial model, the optimized model showed better evaluation performance on all metrics, including MAE, MSE, RMSE, and MAPE. Thus, this approach can be relied upon as a tool in historical data-based business planning, especially in the automotive sector which has complex market dynamics and seasonal patterns. This research also opens up opportunities for further development by integrating the GRU architecture into hybrid or ensemble models for more robust and accurate prediction results.

References

[1] Arisena, A. (2020). PREDIKSI RETAIL SALES MOBIL TOYOTA MENGGUNAKAN METODE ARIMA (AUTOREGRESSIVE INTEGRATED MOVING AVERAGE). *Indonesian Journal of Strategic Management*, *3*(1). https://doi.org/10.25134/ijsm.v3i1.2817

[2] Ba, J. L., Kiros, J. R., & Hinton, G. E. (2016). Layer normalization. arXiv preprint, arXiv:1607.06450. https://doi.org/10.48550/arXiv.1607.06450

[3] Gal, Y., & Ghahramani, Z. (2016). A theoretically grounded application of dropout in recurrent neural networks. arXiv preprint, arXiv:1512.05287. https://doi.org/10.48550/arXiv.1512.05287

[4] 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. https://doi.org/10.18280/isi.290306

[5] Hindarto, D. (2023). Comparison of RNN Architectures and Non-RNN Architectures in Sentiment Analysis. *Sinkron*, *8*(4), 2537–2546. https://doi.org/10.33395/sinkron.v8i4.13048

[6] Krueger, D., Maharaj, T., Kramár, J., Pezeshki, M., Ballas, N., Ke, N. R., Goyal, A., Bengio, Y., Courville, A., & Pal, C. (2016). ZONEOUT: Regularizing RNNs by randomly preserving hidden activations. arXiv preprint, arXiv:1606.01305. https://doi.org/10.48550/arXiv.1606.01305

[7] Mateus, B. C., Mendes, M., Farinha, J. T., Assis, R., & Cardoso, A. M. (2021). Comparing LSTM and GRU models to predict the condition of a pulp paper press. *Energies*, *14*(21). https://doi.org/10.3390/en14216958

[8] Nusaiba Yulisa, P., Al Haris, M., & Rismawati Arum, P. (2023). Peramalan Nilai Ekspor Migas di Indonesia dengan Model Long Short Term Memory (LSTM) dan Gated Recurrent Unit (GRU). In *J Statistika* (Vol. 16, Issue 1)