ISSN 1816-6075 (Print), 1818-0523 (Online)
Journal of System and Management Sciences
Vol. 13 (2023) No. 5, pp. 470-482
DOI:10.33168/JSMS.2023.0530

Forecasting System Analysis using Gated Recurrent Unit Neural Network

Ni Luh Wiwik Sri Rahayu Ginantra, I Made Dwi Putra Asana, Wayan Gede Suka Parwita, Made Lanang Diegoray Wiadnyana

Institut Bisnis dan Teknologi Indonesia, Bali, Indonesia

wiwik@ko2pi.org (Corresponding author)

Abstract. Sales competition is tightening, and people's wants and wishes are becoming more complex. Retail organizations must plan strategically when buying items. Retail X Company started in 2004. Due to lack of planning, the warehouse's stock of items runs out quickly. Purchasing division in this company usually relies on instinct when purchasing items for stock. It always becomes a problem by using this method because that make inaccuracy in stock planning. For better planning, forecasting the number of items sales is needed so that planning for purchasing inventory can be more accurate to reduce the risk of unsold items due to excess purchases. Forecasting system is a tool or methodology used to predict future items sales based on historical data. It is faster and more efficient for processing sales forecasting. In this research the forecasting system used the Gated Recurrent Unit (GRU) model to anticipate product sales for more accurate inventory planning. The model was trained and validated using 2017-2021 time series data. The best model has batch size 64 and 64 hidden neurons, with MSE train scores of 138.8019 and validation scores of 136.3658. Using January 2022 and February 2022 sales data as actual data, MAPE evaluations were 46% and 60%. With this finding, Retail Company X can use GRU Neural Network to get reasonable forecasting for sales forecasting.

Keywords: Gated Recurrent Unit, Forecasting, Time Series

1. Introduction

Competition in product sales between supermarkets is getting tighter. People's demands and wants are becoming more complex and difficult to predict. Therefore, companies engaged in the retail industry must have strategic planning in procuring goods, including a forecasting sales system(Burinskienė & Leonavičienė, 2022; Muna et al., 2023). Inventory of goods is essential to notice because it ensures the availability of goods desired by consumers to increase customer satisfaction. Therefore, the role of goods sales forecasting is needed as the basis for company decisions to plan the company's future strategy to estimate the inventory or stock of goods that must be prepared in the future(Amara et al., 2022).

Retail X Company is a company engaged in the retail business and has been operating since 2004. This company has approximately 15,000 items or products with an average daily transaction of more than 1000. All transactions are recorded in a sales system integrated with the company's warehouse system. The problem faced by this company is that the available items stock in the warehouse suddenly runs out due to the absence of planning for items to be purchased at a time. This problem occurs because the purchasing division in this company usually only relies on instinct when buying items or decides to purchase inventory when the number of items in the warehouse system is running low. It becomes a problem because of the inaccuracy of stock planning. It can cause losses because purchased items will be in vain if unsold products exist(Lisaria Putri et al., 2023). In addition, it can disappoint consumers because they cannot buy the desired products due to being out of stock(Rasiobar & Alfiannor, 2023). Therefore, it requires a forecasting system for the number of items sales so that planning for purchasing inventory can be more accurate to reduce the risk of unsold items due to excess purchases. Besides, to increase consumer satisfaction so that they are interested in buying their needs at this company.

Gated Recurrent Unit (GRU) is an RNN variation. Kumar (Arun Kumar et al., 2022) compared GRU, LSTM, ARIMA, and SARIMA algorithms to forecast COVID-19 situations with time series models. Deep learning-based models (LSTM and GRU) outperformed statistical models (ARIMA and SARIMA) with RMSE values 40 times less than ARIMA. Retail Company X's sales data is sequence data since it includes the date and time of each sale. Time series data is a set of sequential measurements (Adhikari & Agrawal, 2013). Wu (Wu et al., 2019) used the GRU algorithm to estimate electricity prices based on power loads. This study found that GRU computing is faster and more accurate than LSTM. Ardianti (Arfianti et al., 2021) also predicted current sunspot numbers. The GRU approach had a higher MAPE score than the LSTM method, 7.171%.

2. Related Works

2.1. Neural Network (NN)

Neural Network (NN) is an algorithm in machine learning that imitates the workings of human nerves, which are the main part of the brain. NN consists of an output and input layer. Each layer contains one or more neuron units with an activation function that determines the output of that unit. The NN capability can be improved by adding a hidden layer. NN is trained using training data. The NN performance will be better if using more training data. However, the NN capability is also limited by the number of layers. The more layers, the higher the NN capacity. The number of layers will also impact the number of training or iterations required. Thus, a deep learning technique was designed to overcome this problem (Batubara & Awangga, 2020).

2.2. Artificial Neural Network

An Artificial Neural Network (ANN) draws inspiration from the biological structures of the human brain. This information processing technique is characterized by its ability to customize the structure of an information system to fit the requirements of a specific application. The primary concept behind ANN is to emulate the way the human brain is able to take input data, recognize patterns, tolerate mistakes and process different types of signals in parallel (Putro & Awangga, 2020).

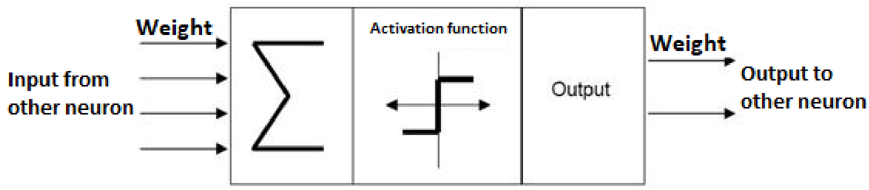

Fig.1: ANN Structure (Putro & Awangga, 2020)

2.3. Recurrent Neural Network

Recurrent Neural Network (RNN) is a particular type of Artificial Neural Network (ANN) that is used for sophisticated data analysis and has the capacity to learn sequences of data and perform sequential data processing. The connections between the neurons in RNN can form a cycle, allowing it to remember patterns from previous input, as well as scale images with large dimensions, although this process takes longer than usual for an unspecialized network (Nugraha et al., 2020).

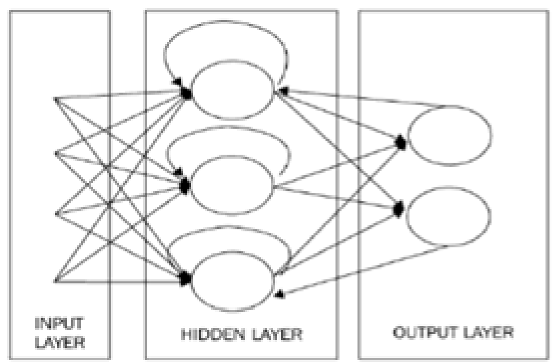

Fig.2: RNN Structure (Nugraha et al., 2020)

In operation performed by an RNN, the parameters are divided differently than the convolutional network. Each output member is a function of the previous output member. Each output member is created using the same update rules applied to the previous output. This repeated formulation results in the distribution of parameters through a very deep computational graph (Nugraha et al., 2020) .

2.4. Gated Recurrent Unit

GRU is an RNN improvement designed to better capture dependencies over different timescales, address the missing gradient problem, and simplify the complex LSTM structure. It has a single gate unit to decide when to forget and update the state, making it more persistent and efficient for forecasting scenarios requiring quick consumption (Chung et al., 2014; Dawani, 2020a; Y. jiang Li et al., 2021).

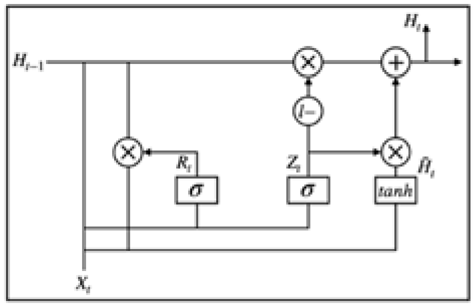

Fig.3: GRU Structure (Dawani, 2020b)

These components are as follows:

a. Z_t is *update gate*
b. R_t is *reset gate*
c. $\widetilde{H}_t$ is *new memory*
d. H_t is *hidden state*

To generate the current hidden state, the GRU uses these operations:

a. $Z_t = \sigma(W_z X_t + U_z h_{t-1})$
b. $R_t = \sigma(W_r X_t + U_r H_{t-1})$
c. $\widetilde{H}_t = \tanh(R_t * U h_{t-1} + W X_t)$
d. $H_t = (1 - Z_t) * \widetilde{H}_t + Z_t * H_{t-1}$

The following are the sequences of the previous equation to get a better idea of what the GRU does with its two inputs. It can be seen as follows:

a. GRU takes the present input (X_t) and the earlier hidden state (H_{t-1}) and uses the information it has about the prior words or memory to produce $\widetilde{H}_t$ the updated memory state.
b. Gate R_t determines the importance of the prior hidden state when figuring out the current hidden state, deciding if it is applicable to gain new memory and support short-term relationships.
c. The gate (Z_t) dictates how much of the previous hidden states are used to inform the current state, preserving long-term dependencies. When $Z_t \approx 1$, most of the previous memory is retained in the new state, but when $Z_t \approx 0$, new information is given more priority.
d. The hidden state of the current time step (H_t) is established based on the evaluation of the update gate and the combination of the new memory and the previous hidden state.

In its implementation, the GRU method can be used in various fields and has successfully achieved good results. Such as prediction of the remaining life of Lithium-ion battery by Ardeshiri (Ardeshiri et al., 2022), prediction of dissolved oxygen present in ponds to apply artificial aeration with low energy and cost by Li (W. Li et al., 2021), and forecasting of runoff for flood mitigation carried out by Gao (Gao et al., 2020).

3. Research Method

This study employed the Cross Industry Standard Process Model for Data Mining (CRISP-DM) as a blueprint for developing a forecasting system. The CRISP-DM methodology is depicted in Fig 4.

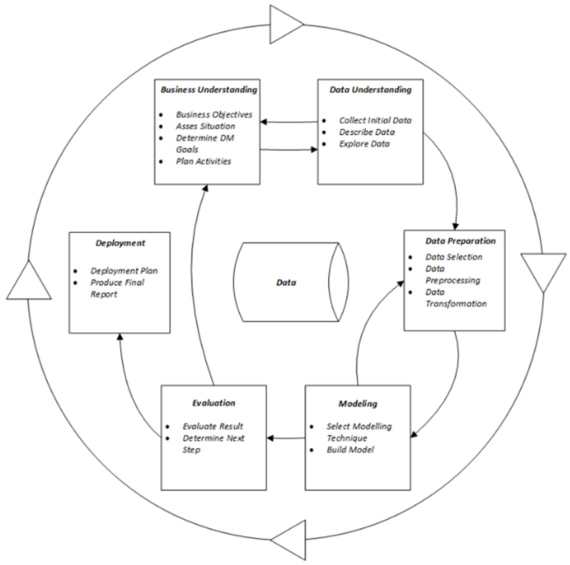

Fig.4: The Flow of CRISP-DM Method

a. Business *Understanding*
The introduction of business problems in this research proposal referred to the sale of goods that occur based on a certain period in Retail Company X. This stage required an understanding of the importance of using a data warehouse. Thus, it could be used for forecasting sales of goods to improve the quality of planning stock purchases.

b. Data Understanding
This research used data from the data warehouse of Retail Company X, which was data on sales of goods in the period 2017 to early 2022. The dataset had more than ten million rows and 25 columns of data. Each row of data represented the number of transactions on each item. The columns contained in the dataset include nonota, barcode, day (*hari*), month (*bulan*), year (*tahun*), item name (*nama barang*), unit (*satuan*), qty, and so on.

c. Data Preparation
This stage was carried out to overcome the problems in the dataset before the data was processed further to the modeling stage to get optimal modeling performance.

d. Data Selection
At this stage, columns from the obtained dataset will be selected so that only the attributes left will be used in the modeling process. The columns used are date (*tgl*), itemname (*namabarang*), unit (*satuan*), qty, barcode, customer code (*kodecustomer*), type code (*kodejenis*), type (*jenis*), category_code (*kodekategori*), and category (*kategori*). The attribute specified as the label or forecast target was the qty attribute.

e. Data Preprocessing
This stage aimed to analyze the data quality by changing, deleting, and checking less relevant data for research needs. Missing value, noise, and redundant data contained in the dataset will be removed from the dataset to avoid interfering with the modeling process.

f. Data Transformation
At this stage, the data that has been selected and cleaned will be converted into data that is in accordance with the modeling process. Before being used in the training process, the time series data will be rearranged into supervised learning by changing the input and output into a sequence. For example,

product sales data from January to December 2021 will be used as input to predict product sales in January 2022. Further, the dataset will be divided into training and validation data. Sales data in 2017 and 2021 will be used as training and validation data, while sales data in January and February 2022 will be used as testing data to evaluate the forecasting stage. The training data amounted to 90% of the total dataset. After the model was trained to learn data patterns using the data train, then the model was validated using data validation which was 10% of the total amount of data available in the dataset to determine the model performance.

g. Modelling
At this stage, the process was to create predictive models to produce information that could facilitate interested parties in making decisions. The hidden neurons parameter used in modelling were 8, 16, 32, and 64 with batch size were 8, 16, 32, and 64.

h. Select Modeling Technique
The GRU technique was employed for the modeling process. This strategy was well-suited for this task because it was specifically created to process temporal data with identifiable patterns over consecutive time intervals, such as time-based data (Noh et al., 2020).

3.1. Research Stages

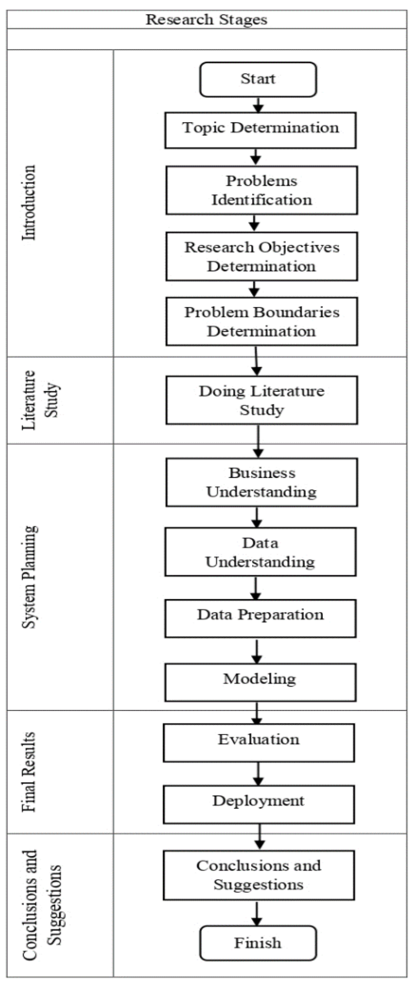

Fig.5: Research Stages

3.2. System Overview

The system overview (Fig 6) describes the system flow and business processes in general in designing the sales forecasting model using the Gated Recurrent Unit method. There were two important processes in this system: the training process – data validation, and the forecasting process. The training and validation processes were carried out to produce a GRU model that could accurately predict the number of sales of goods. If the best model has been found, then the model will be used in the forecasting process.

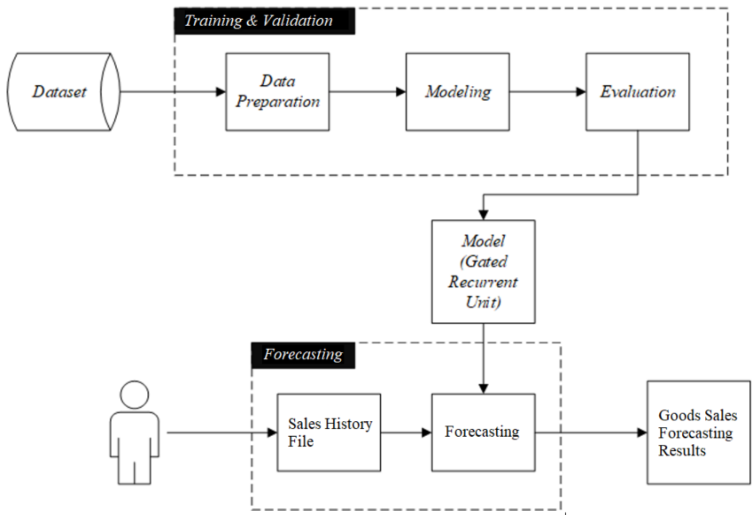

Fig.6: System Overview

4. Results and Discussion

Focus of research is to find the best GRU model with different number of hidden neurons and batch sizes. The comparison will be elaborated for each combination result of number of hidden neurons and batch, so that in the end the research got the best combination of neurons and batches. The parameter values for hidden neurons tested were 8, 16, 32, and 64. Similar to hidden neurons, the batch size values tested were 8, 16, 32, and 64. Based on all the training and validation experiments in Table 1, the lowest loss is indicated by a green label, while a yellow label indicates the highest loss. The lowest loss was achieved using batch size 64 and hidden neuron 64 with a train MSE value of 138.8019 and MSE validation of 136.3658. While the highest loss was achieved using batch size 8 and hidden neuron 8 with a train MSE value of 143.0221 and a validation MSE of 141.4689. Each of these values explained that the training and validation process, which had the lowest loss value, performs the data pattern recognition process better than the training and validation process, which had a large loss value.

Table 1. Parameter Value Testing

Early Stopping	Batch Size	Hidden Neuron	Train MSE	Validation MSE
32	8	8	143.0221	141.4689
38		16	142.2886	137.9702
25		32	141.8339	141.5539
34		64	141.3339	139.3387
48	16	8	142.5603	137.9322
51		16	141.6289	137.5091
38		32	141.2016	138.7038
19		64	141.0043	137.4425
79		8	141.8690	137.5619

47	32	16	141.3768	139.4229
31		32	140.8004	136.5194
39		64	139.3680	137.1299
108	64	8	141.1476	139.6360
61		16	140.7559	136.7161
46		32	140.0804	137.7550
46		64	138.8019	136.3658

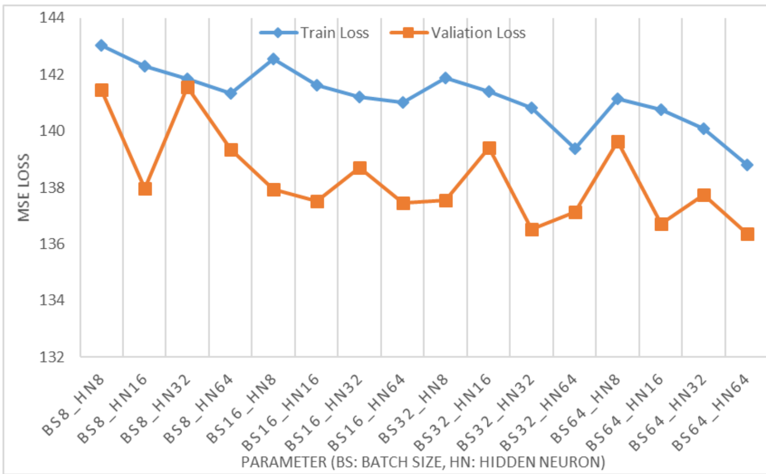

Fig.7: Loss Model Comparison Graph

Fig 7 shows that the number of hidden neuron values used can affect the loss value, which will decrease. Likewise, the number of batch size values used can affect the loss value, which will decrease. Based on all the batch size and hidden neuron experiments above, the author will use the model generated in the batch size 64 and hidden neuron 64 experiments during the training and validation process to predict the number of product sales.

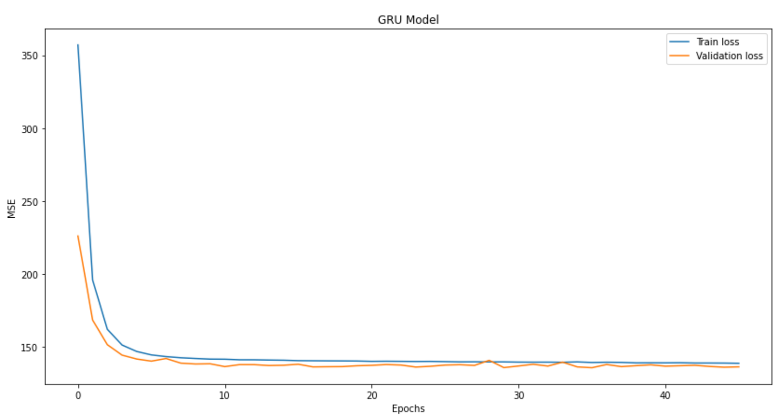

Fig.8: Best Train Loss - Validation Loss Model Graph

Fig 8 shows the visualization or plot of the training and validation processes of the batch size and hidden neuron values with the lowest loss. The plot shows that the trained model has a good learning process. The process is marked by a graph of validation loss which decreases until the last epoch, and the training loss also decreases as the epoch increases.

Fig 9 shows the eliminating data process with a value above 400 in the correct qty column. Based on the graph above, the model obtained from the process is indicated to underfit, so the model cannot be used to carry out the forecasting process. In addition to conducting a parameter testing process using data that has been eliminated, the author also carried out a training and validation process using sales data for the ten best-selling categories from 2017 to 2021. Each category was tested using hidden neuron parameter values of 8, 16, 32, and 64 and batch size values of 8, 16, 32, and 64.

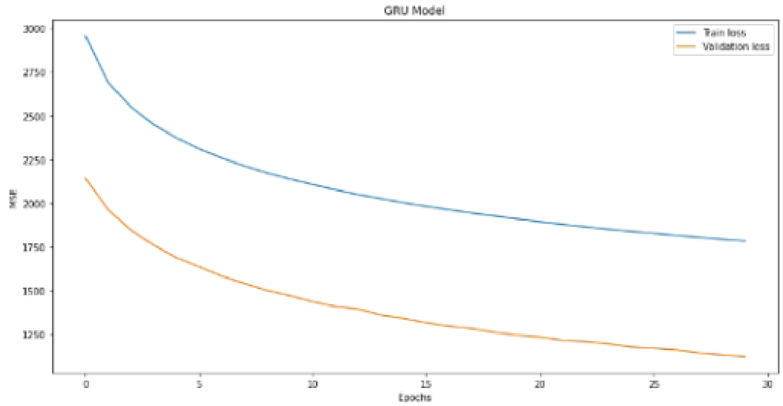

Fig.9: The Graph of Training - Validation Before Qty Elimination

Table 2. Best Testing Results Per Category

Category Name	Early Stopping	Batch Size	Hidden Neuron	Train MSE	Validation MSE
Food	75	64	64	138.8019	136.3658
Healt & Beauty Care	66	8	64	57.4221	74.3991
Milk (*Susu*)	181	64	64	453.0387	557.6100
Snack (*Makanan Ringan*)	279	32	64	334.9719	389.4208
Biscuit (*Biskuit*)	198	32	64	420.3798	417.5112
Clothes Care (*Perawatan Pakaian*)	172	32	64	284.4926	250.7087
Noodle (*Mie*)	144	16	64	1807.4934	1972.9906
Body Care (*Perawatan Tubuh*)	272	64	32	145.3492	144.3678
Beverage	43	16	64	221.8452	286.6234
Cleaning	117	16	32	143.8858	134.0355

Based on all the best testing results for each category in Table 2, the lowest loss was achieved in the Health & Beauty Care category using batch size 8, hidden neurons 64 with a train MSE value of 57.4221, and validation MSE of 74.3991. While the highest loss was achieved by the Noodle (Mie) category using a batch size 16 and hidden neurons 64 with a train MSE value of 1807.4934 and a validation MSE of 1972.9906.

In evaluating forecasting results, the Mean Absolute Percentage Error (MAPE) and Mean Squared Error (MSE) methods were used to calculate errors based on actual data with forecasting data. After the calculation, the MAPE value from the January forecast was 46%, while in February was 60%. Meanwhile, the MSE value from forecasting results in January was 84,7150, while in February was 85,0732. When the MAPE value achieved is greater than 50, it can be categorized as Inaccurate Forecasting. If the MAPE value is between 20 and 50, it is categorized as Reasonable Forecasting, and if the MAPE value is between 10 and 20, it is categorized as Good Forecasting. In addition, if the MAPE value is less than 10, it is categorized as Highly Accurate Forecasting (Lewis, 1982).

Based on the calculation results of MAPE and MSE with the GRU model that has been built, it could be categorized as Reasonable Forecasting and forecast sales in the following month. With this finding, Retail Company X can use GRU Neural Network to get reasonable forecasting with hidden neurons 64 and batch sizes 64. Fig 10 to 17 is a graph of the MAPE and MSE score for forecasting annual sales, starting from 2018 to 2021.

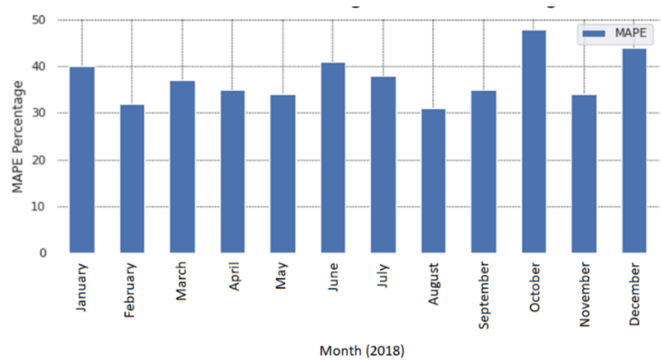

Fig.10: MAPE Graph of 2018

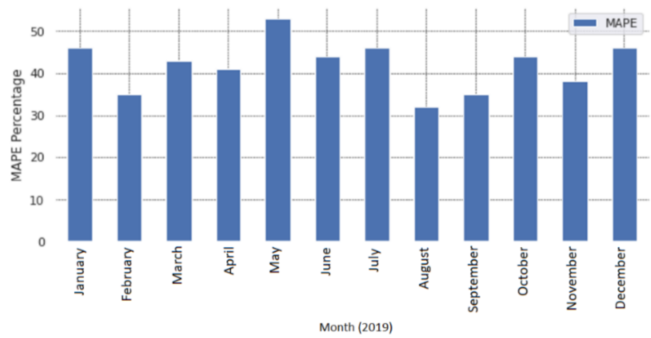

Fig.11: MAPE Graph of 2019

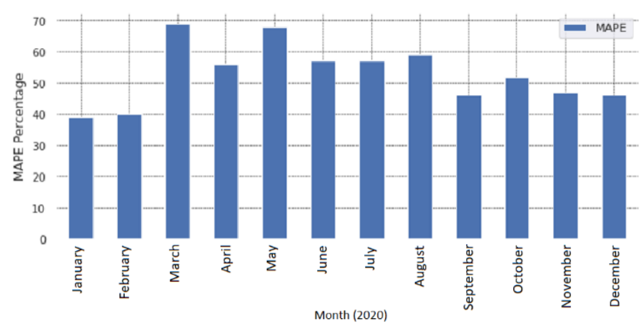

Fig.12: MAPE Graph of 2020

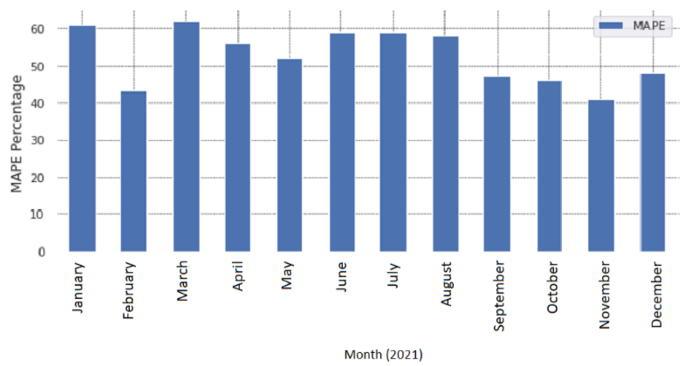

Fig.13: MAPE Graph of 2021

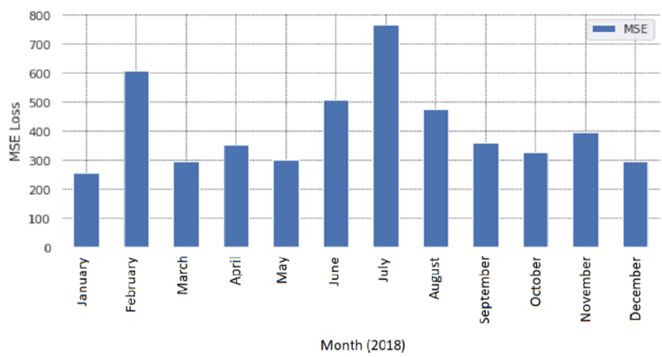

Fig.14: MSE Graph of 2018

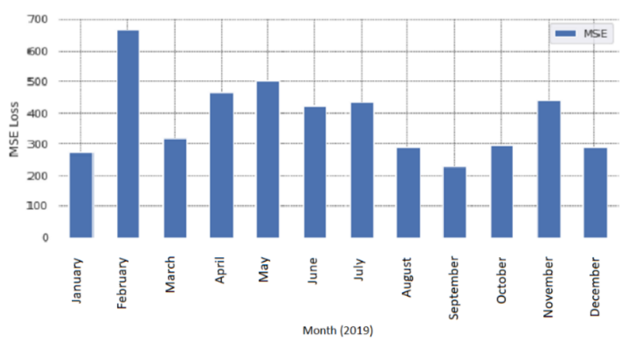

Fig.15: MSE Graph of 2019

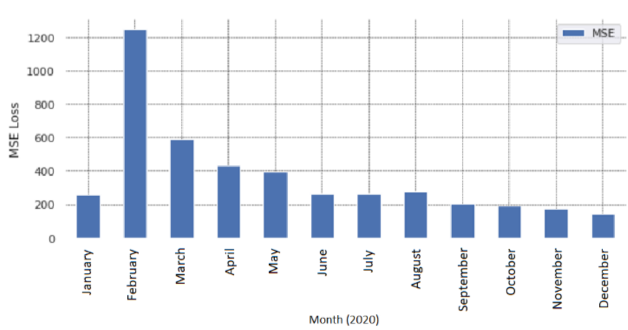

Fig.16: MSE Graph of 2020

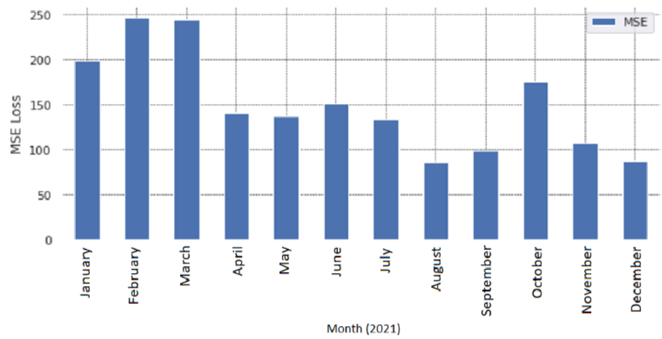

Fig.17: MSE Graph of 2021

5. Conclusion

During the training and validation process, parameter values were very influential in producing the best model performance. The lowest loss was achieved using batch size 64 and hidden neuron 64 with a train MSE value of 138.8019 and validation MSE of 136.3658. While the highest loss was achieved using batch size 8, hidden neuron 8 with a train MSE value of 143.0221 and a validation MSE of 141.4689.

The forecasting model built using the Gated Recurrent Unit (GRU) method could be categorized as Reasonable Forecasting and capable of predicting sales in the following month. The MAPE evaluation in January 2022 was 46%, while in February 2022 was 60%. Meanwhile, the MSE value in January 2022 was 84,7150, while in February 2022 was 85,0732. With this finding, Retail Company X can use GRU Neural Network to get reasonable forecasting with hidden neurons 64 and batch sizes 64.

Based on the results, there are suggestions that the authors give to help further research, wich is adding some attributes or variables that might affect the number of product sales, such as holiday seasons.

References

Adhikari, R., & Agrawal, R. K. (2013). An Introductory Study on Time Series Modeling and Forecasting Ratnadip Adhikari R. K. Agrawal. *ArXiv Preprint ArXiv:1302.6613*, *1302.6613*, 1–68.

Amara, G., Abdulrahman, S., Alkhateeb, S., & Alabbas, S. (2022). Scarcity Marketing: The Impact of Limited-Edition Packaging on Purchase Intentions in the Syrian Market. *Journal of Service, Innovation and Sustainable Development*, *3*(2), 21–38. https://doi.org/10.33168/SISD.2022.0202

Ardeshiri, R. R., Razavi-Far, R., Li, T., Wang, X., Ma, C., & Liu, M. (2022). Gated recurrent unit least-squares generative adversarial network for battery cycle life prediction. *Measurement*, *196*, 111046. https://doi.org/https://doi.org/10.1016/j.measurement.2022.111046

Arun Kumar, K. E., Kalaga, D. V, Mohan Sai Kumar, Ch., Kawaji, M., & Brenza, T. M. (2022). Comparative analysis of Gated Recurrent Units (GRU), long Short-Term memory (LSTM) cells, autoregressive Integrated moving average (ARIMA), seasonal autoregressive Integrated moving average (SARIMA) for forecasting COVID-19 trends. *Alexandria Engineering Journal*, *61*(10), 7585–7603. https://doi.org/https://doi.org/10.1016/j.aej.2022.01.011

Batubara, N. A., & Awangga, R. M. (2020). *TUTORIAL OBJECT DETECTION PLATE NUMBER WITH CONVOLUTION NEURAL NETWORK (CNN)*. Kreatif.

Burinskienė, A., & Leonavičienė, E. (2022). Risk Management in International Business Development Projects. *Journal of Service, Innovation and Sustainable Development*, *3*(2), 51–64. https://doi.org/10.33168/SISD.2022.0204

Chung, J., Gulcehre, C., Cho, K., & Bengio, Y. (2014). *Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling*. 1–9.

Dawani, J. (2020a). *Hands-On Mathematics for Deep Learning: Build a solid mathematical foundation for training efficient deep neural networks*. Packt Publishing.

Dawani, J. (2020b). *Hands-On Mathematics for Deep Learning: Build a solid mathematical foundation for training efficient deep neural networks*. Packt Publishing.

Gao, S., Huang, Y., Zhang, S., Han, J., Wang, G., Zhang, M., & Lin, Q. (2020). Short-term runoff prediction with GRU and LSTM networks without requiring time step optimization during sample generation. *Journal of Hydrology*, *589*, 125188. https://doi.org/https://doi.org/10.1016/j.jhydrol.2020.125188

Li, W., Wu, H., Zhu, N., Jiang, Y., Tan, J., & Guo, Y. (2021). Prediction of dissolved oxygen in a fishery pond based on gated recurrent unit (GRU). *Information Processing in Agriculture*, *8*(1), 185–193. https://doi.org/https://doi.org/10.1016/j.inpa.2020.02.002

Li, Y. jiang, Yang, Y., Zhu, K., & Zhang, J. (2021). Clothing Sale Forecasting by a Composite GRU-Prophet Model With an Attention Mechanism. *IEEE Transactions on Industrial Informatics*, *3203*(c), 1–9. https://doi.org/10.1109/TII.2021.3057922

Lisaria Putri, R., Chasan Amrulloh, M., Sutrisno, J., Saptana, S., & Dani, N. L. (2023). Government Policy Strategy towards Empowering MSMEs in the Covid-19 Pandemic at the Bulak Fish Market Center, Surabaya City. *Athena: Journal of Social, Culture and Society*, *1*(3), 93–101. https://doi.org/10.58905/ATHENA.V1I3.43

Muna, A., Ausat, A., Al Bana, T., & Gadzali, S. S. (2023). Basic Capital of Creative Economy: The Role of Intellectual, Social, Cultural, and Institutional Capital. *Apollo: Journal of Tourism and Business*, *1*(2), 42–54. https://doi.org/10.58905/APOLLO.V1I2.21

Noh, J., Park, H., Kim, J. S., & Hwang, S. (2020). *Gated Recurrent Unit with Genetic Algorithm for Product Demand Forecasting in Supply Chain Management*. 1–15. https://doi.org/10.3390/math8040565

Nugraha, F. A., Harani, N. H., Habibi, R., & Awangga, R. M. (2020). *Analisis Sentimen Terhadap Pembatasan Sosial Menggunakan Deep Learning*. Kreatif.

Putro, E. C., & Awangga, R. M. (2020). *TUTORIAL GENDER CLASSIFICATION USING THE YOU LOOK ONLY ONCE (YOLO)*. Kreatif.

Rasiobar, M. R., & Alfiannor, A. (2023). The Effect of Lifestyle, Self-Concept and Perceived Value on Purchasing Decisions Ice Cream Baskin Robbins in Giant Express Banjarmasin. *Apollo: Journal of Tourism and Business*, *1*(2), 55–63. https://doi.org/10.58905/APOLLO.V1I2.46

Wu, W., Liao, W., Miao, J., & Du, G. (2019). Using gated recurrent unit network to forecast short-term load considering impact of electricity price. *Energy Procedia*, *158*, 3369–3374. https://doi.org/10.1016/j.egypro.2019.01.950