KARYA ILHAM PUBLISHING
CONNECTING MINDS, BRIDGING IDEAS, INSPIRING READERS

Journal of Advanced Research in Computing and Applications

Journal homepage:
https://karyailham.com.my/index.php/arca/index
ISSN: 2462-1927

Study of E-commerce Sale Prediction Based on Machine Learning Methods

Rong Liu[1], Annie Anak Joseph[1,*], Wanzhen Wang[2]

[1] Department of Electrical and Electronic Engineering, Faculty of Engineering, Universiti Malaysia Sarawak (UNIMAS), Kota Samarahan, Sarawak, Malaysia
[2] Institute of Intelligent Manufacturing, Qilu Institute of Technology, Jinan, PR. China

ARTICLE INFO

Article history:
Received 11 October 2025
Received in revised form 24 November 2025
Accepted 30 November 2025
Available online 8 December 2025

Keywords:
Sales forecasting; time series data; model performance comparison; practical; metrics

ABSTRACT

Precise overall sales forecasting is essential in the sales domain for controlling slow-moving commodities and cutting inventory expenses. However, seasonality, trends, and multi-product scenarios provide challenges for established methods of sales forecasting. For time series data and complicated patterns, models such as gated cycle unit (GRU), recurrent neural network (RNN), and short- and long-term memory network (LSTM) were chosen to increase processing power. To find the best models for sales forecasting, the performance of these models was compared using metrics (MAE, RMSE, and R^2). It is found that GRU model is the best model in this field. In order to assure the research's suitability from a scientific and practical standpoint, these additional components have been added to increase the study's scope, address the issue of previous research using these models sparingly or not at all, and look for more efficient ways to forecast sales.

1. Introduction

With the vigorous development of e-commerce, competition in the commercial field has become increasingly fierce, and the market environment has become more and more complex [1]. In this rapidly changing context, merchants need to predict the total sales volume of multiple products quickly and accurately, as well as the time trend of sales, to adapt to changing consumer needs. This is essential for raising operational effectiveness and enhancing inventory control. Through accurate short-term forecasts, merchants can adjust supply chains, inventory, and marketing strategies in a timely manner to make them more adaptable to market dynamics and meet customer needs to the greatest extent. Traditional time series forecasting methods are limited by the assumptions of linear models and the limitations of feature engineering and often difficult to capture complex nonlinear relationships and long-term dependencies. In the complex market environment of e-commerce, these traditional methods often perform poorly. In the last few years, technology for deep learning, Recurrent neural network models in particular, such the long short-term memory network (LSTM) [2]

* *Annie Anak Joseph.*
E-mail address: jannie@unimas.my

https://doi.org/10.37934/arca.41.1.3648

and gated recurrent unit (GRU) [3], have shown promise. These models offer a strong tool for time series forecasting because they can model lengthy sequences and more effectively handle nonlinear interactions. The aim of this research is to use time series forecasting technology, specifically the deep learning-based recurrent neural network (RNN) model, to perform short-term time series forecasting of retail shop sales [4]. This study introduces time series prediction models based on recurrent neural networks (RNN), with special focus to long short-term memory networks (LSTM) and gated recurrent units (GRU). This research aims to make full use of these advanced deep learning models to achieve accurate short-term time series forecasting of retail store sales. Such predictions can provide merchants with more insightful sales decision support and help them remain invincible in the fierce market competition.

2. Related Works

This section will focus on several important machine learning models, such as GRU (Gated Recurrent Unit), RNN (Recurrent Neural Network), LSTM (Long Short-Term Memory Network), and ARIMA (Auto Regressive Integrated Moving Average) [5]. One of them is the statistical model ARIMA. The traditional recurrent neural network models RNN, GRU, and LSTM are adept in processing sequence data and possess the internal loop architecture needed to identify sequence dependencies. In this chapter, previous case studies will be summarized to analyze the advantages and disadvantages of each model, and then identify the research gaps in store sales forecasting under different time series application scenarios.

2.1 Autoregressive Integrated Moving Average (ARIMA)

Model of autoregressive smoothing, also known as ARIMA model. Proposed by American statisticians Jen Kins and Box. In previous studies by Toğa *et al.*, [6] strong support for health forecasting is provided by the use of ARIMA and Artificial Neural Network (ANN) to predict the number of infection cases, fatalities, and recovered cases in Turkey. To increase the predictability of time series data, the enhanced seasonal autoregressive integrated moving average (ESARIMA) is presented and paired with DWT technology [7]. The effectiveness of time series models is predicted to increase as a result of this invention. As a classic time sequence analysis method, the ARIMA model has made important breakthroughs after years of development, especially in model selection, seasonal modeling and long-term dependence modeling.

2.2 Recurrent Neural Network (RNN)

RNN, or recurrent neural network was proposed by Elman and plays a key role in deep learning, especially in time series data processing and sales time series analysis. For the application of energy storage systems, a technique for anticipating future loads considering the energy storage effect was studied, using LSTM-RNN and a two-charge and two-discharge operation strategy to reduce the peak load [8]. The N-BEATS-RNN model of Sbrana *et al.*, [9] reduces training time through an efficient weight sharing search mechanism, bringing satisfactory results to time series prediction. The benefit of RNN is that it can process time-dependent data, which is especially suitable for tasks such as sales time series. It automatically captures patterns and trends in data, improving forecast accuracy. However, RNN models also have some limitations. It is still challenging to deal with long-term dependence problems, and there may be vanishing or exploding gradient problems. In addition, deep RNN models usually require massive quantities of data and computing resources and are not suitable for all prediction situations. In practical applications, model structure and hyperparameters need to be carefully selected to obtain optimal performance.

2.3 Long Short-Term Memory (LSTM)

The issue of the vanishing gradient is avoided when processing long sequences using the RNN variation known as LSTM. It is widely used in sales forecasting and can capture long-term dependencies. In order to accurately estimate the probabilistic power demand and assess the model's level of uncertainty, the modified deterministic LSTM method put out by Zhu *et al.*, [10] employs a two-stage model. Xue *et al.*, [11] combined the LSTM and XGBoost models and proposed the XGBoost-LSTM photovoltaic power prediction model, which has high prediction accuracy and is suitable for photovoltaic grid-connected and off-grid applications. Based on the CNN-LSTM model, Wensheng *et al.*, [12] proposed a regional comprehensive energy load prediction method distinguished by user energy tags, which takes into account different user behaviors and is suitable for regional energy planning. The Bi-LSTM model, which has better prediction accuracy than the conventional recurrent neural network, was put up by Luo *et al.*, [13] and introduced the bidirectional LSTM network. Bilgili *et al.*, [14] used deep learning technology based on LSTM neural network to predict Turkey's total energy consumption (GEC). The findings demonstrate that the LSTM model has higher accuracy than the SARIMA model. The benefit of LSTM is that it can manage long-term dependencies, employ memory units to store information, enable parallel computing, and is appropriate for time series prediction of a variety of jobs, according to prior research and analysis. Although sensitive to hyper parameters, LSTM also has a high computational complexity and parameter volume and a sophisticated decision-making procedure.

2.4 Gated Recurrent Unit (GRU)

GRU, or gated recurrent unit is a variant of RNN commonly used in time series data modeling and sales forecasting, which has fewer parameters and faster training speed. A series of studies have shown that GRU has achieved significant success in multiple fields [15]. The integration method of EEMD-ABGRU model is adopted to increasing the precision of power load forecasting, which is especially suitable for short-term load data with randomness and periodicity. Cheng *et al.*, [16] introduced a hybrid neural network (FNN and GRU) to precisely forecast the electric vehicle charging load, which helps the adjustment of power grid sites during peak hours. In addition, Gu *et al.*, [17] combined digital twin technology and the GRU-TCN model to achieve high-precision prediction of digital grid load, while Guo *et al.*, [18] used K-means and genetic algorithm optimization. The GRU network successfully predicts short-term fluctuations in electric vehicle charging load. In addition to power load forecasting, GRU also excels in other areas. [19] introduced an energy prediction model based on hybrid sequential learning, combining CNN and GRU to build a unified framework, which improves the accuracy and generalization ability of energy consumption prediction.

In comparison to LSTM, gated recurrent unit (GRU) has a simpler structure, fewer gated units, and great computational efficiency, according to prior research and analysis. The training speed of GRU will be faster than that of LSTM because there are less parameters, allowing it to quickly converge a model. However, unlike long short-term memory networks (LSTM), it is not as good at preserving long-term information and capturing long-term dependencies, which limits its ability to represent complex sequence patterns.

To sum up, previous studies have conducted in-depth research on sales forecast models for instance ARIMA, RNN, GRU and LSTM. However, how to further improve the accuracy and robustness of predictions is still an underexplored area. This study will use the GRU model to predict sales data and compare it with the ARIMA model used in the Kaggle data set for sales prediction.

3. Proposed Method

In the study, historical data including sales date (or time) and sales quantity were first collected. Ensure that the date field format is correct and parse the data that does not conform to the time series format. Secondly, the data is sorted according to the sales date or time, using the sales date as the index of the time series, and the corresponding sales volume as the value of the time series, thus obtaining time-based sales time series data. On this basis, GRU is used as a decoder to extract key historical time information by assigning the weight of the GRU hidden state. Through this method, the correlation of sales historical time series can be mined to predict sales in the next 42 days. In the comparative analysis, this method is compared with LSTM and RNN to evaluate the performance differences of different models on sales forecast tasks, paying special attention to their ability to obtain key information from historical time series and the accuracy and loss function of sales forecasts. Finally, a comparative analysis of the root mean square error is performed with the ARIMA time series prediction model used in the Kaggle competition. Predict a future buying preferences based in the historical sale data. Therefore, the research goal of this study is to predict future buying preferences based on historical sales data.

3.1 Data Set

Figure 1 reflects the relationship between variables in the data set, so this study uses heat maps to explore the pairwise correlation between multiple variables. As a kind of heat map, heat map can use color differences to present data effects. The bright colors in the picture represent situations where time occurrence frequency is higher or the distribution density of things is higher, while dark colors represent the opposite. Figure 1 shows this clearly, Sales correlate most strongly, so the experiment uses the Sales column to predict the next 42 days. Perform retrograde training on the store's historical sales data set named 'train.csv'.

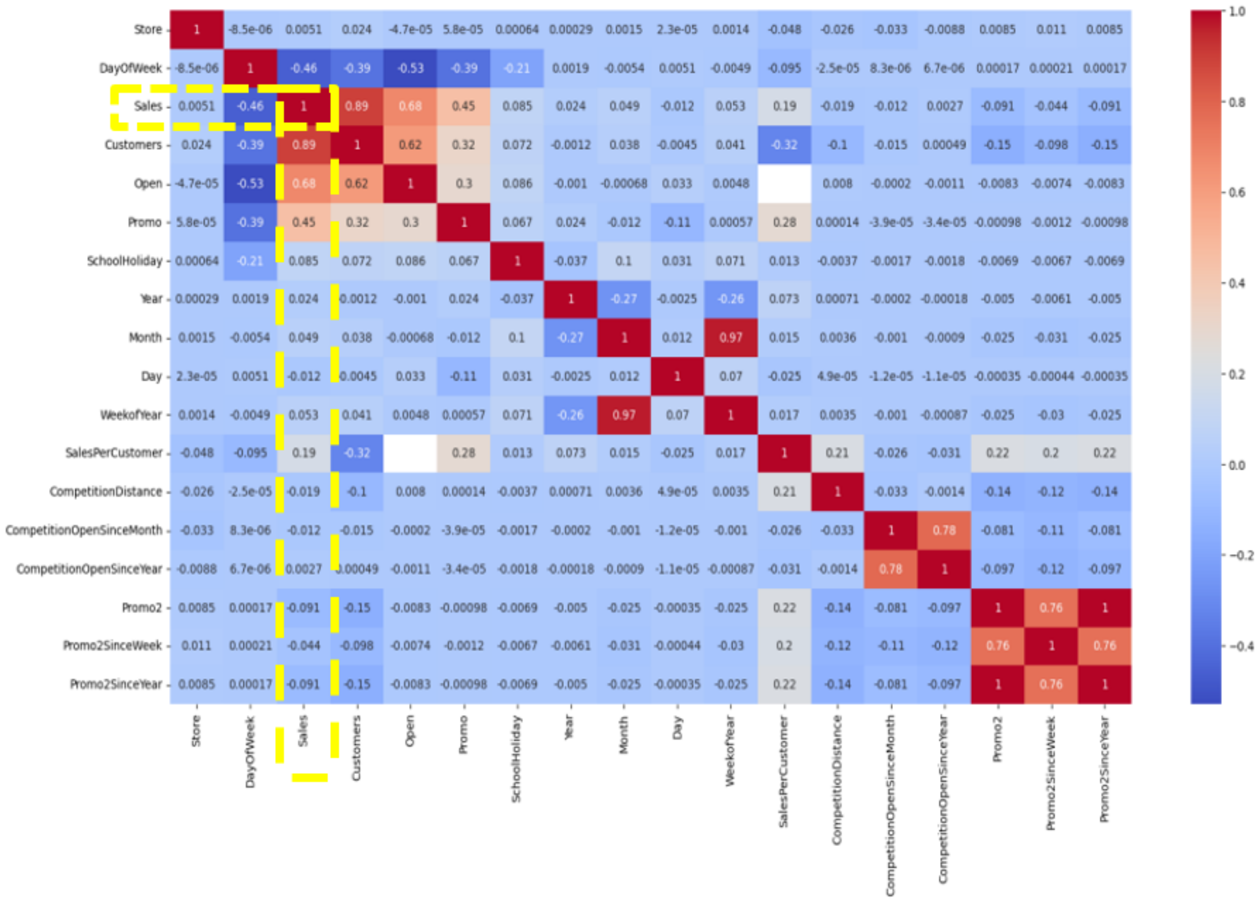

Fig. 1. Heat map of the data set [20]

3.2 Partition of the Data Set

The training set is divided into 80% and the test set into 20% when the entire set of data is treated as 1. 50% of the test set is the validation set, and 50% is the test set.

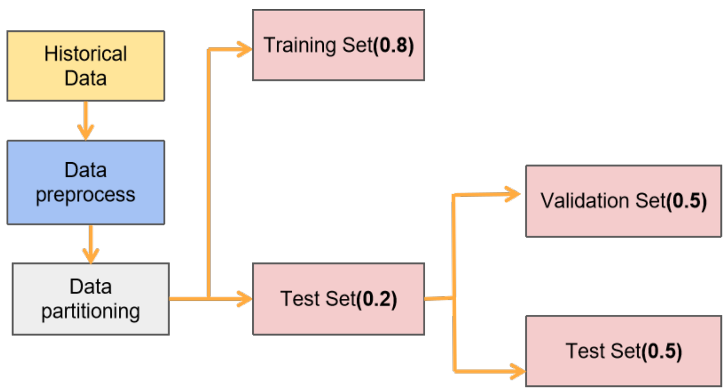

Fig. 2. Data set division

3.3 Preprocessing

First, use Pandas to import sales data and select the first 15,000 sales data for processing. Normalize sales data to the range of 0 to 1 so that the model can learn better. Secondly, create a sliding window data set: define the function create sliding window dataset to convert the normalized sales data into a sliding window data set to adapt to the training sample format of the model. Thirdly, the module uses Kera’s to build a sequence model, including GRU layer, flat layer, and fully connected layer, and configures the model's input shape, number of neurons and activation function. When compiling the model, specify Adam as the optimizer, mean square error (MSE) as the loss function, and mean square error, mean absolute error, and score as the evaluation indicators. The validation set was used to assess model performance after the model had been trained using the training set for 120 epochs with a batch size of 35. The trained model is then applied to forecast the test set, and the mean square error (MSE), mean absolute error (MAE), root mean square error (RMSE), and score between the forecasted outcomes and the actual values are computed.

3.4 GRU Model

Similar to LSTM, GRU is a conventional RNN variation that is capable of capturing the semantic association between lengthy sequences and mitigating the phenomenon of gradient explosion or disappearance. Its computation and structure are also less complicated than those of LSTM. Its core structure can be divided into two parts: update gate and reset gate.

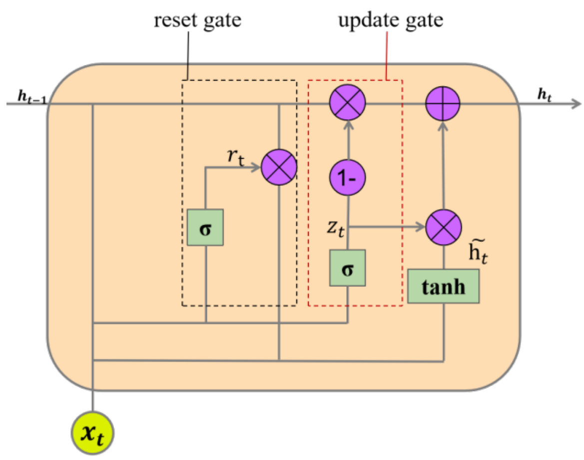

Fig. 3. GRU model

r_t in the figure is the reset gate. By controlling the opening or closing of the gate, the purpose of past information is achieved. The existence of the forgetting gate allows GRU to investigate short-term dependencies between data, thereby further strengthening the correlation between investigation data. The update gate z_t determines which information of the previous moment should be retained or forgotten at the current moment. The update gate exists to help GRU obtain the long-term dependency between time series data. The calculation process of each GRU is as follows:

$$r_t=\sigma(W_r x_t + U_r h_{t-1}+b_r) \quad (1)$$

$$z_t=\sigma(W_z x_t + U_z h_{t-1}+b_z) \quad (2)$$

$$h_t=\tanh(r_t \odot U_h h_{t-1} + W_h x_t + b_h) \quad (3)$$

$$h_t=(1-z_t)\odot h_t + z_t \odot h_t \quad (4)$$

Equations (1) through (4) are connected and cannot be utilized separately. The reset gate r_t gets smaller, the more information you have to forget in the last moment. The closer the update gate z_t

is to 1, the more past information is retained.ht represents the candidate hidden layer state, which displays the input information at t time and the selective retention of the result ht-1 at time t-1.The hidden layer's output at time t is represented by ht. σ is a Sigmoid function, which means that the output value of r_t and z_t at time t is controlled between [0,1]. 1 means that all input information has passed through the "gate", and 0 means that the information has been forgotten. tanh represents the function of activation, which is used to compress the source value at time t and the candidate state value for the memory unit h_t output by r_t at time t-1 to between [-1,1], and then control the state of the memory unit through z_t.⊙is the product of the matrix, and its corresponding position is subjected to dot multiplication operation. After h_t passes through the tanh function, it performs a dot multiplication calculation with the update gate zt , and finally obtains the state vector h_t output at time t.W_r、W_z、U_r、U_z、U_h、W_h are matrixes of parameters trained within the network, and b_r、b_h、b_z are biases.

Compared with RNN and LSTM, GRU solves the problem of RNN's difficulty in long-term memory information and gradient disappearance through structural improvements, reduces training parameters, and accelerates the convergence speed of the model.

3.5 Overall Flow Chart of the Model

When processing time series data, it is often necessary to consider information at multiple time steps. For the prediction of total store sales, the input data x can represent various influencing factors at different time points, such as seasonal changes,promotional activities, etc. These factors undergo a series of processing at each time step, including dropout and GRU processing, to obtain multiple feature representations, such as xt1, xt2 until xtn. Similarly, for the output data y, it can represent the actual sales data at different time points. After the same processing steps, the corresponding feature representations yt1, yt2 until ytn are obtained. This series of feature representations captures the changes and correlations in the data at each time step, such as the relationship between sales volume and seasonal changes and promotional activities. This study combines these feature representations, using stacking, connection, or other methods, to obtain a comprehensive feature representation, which can be used to build a sales forecast model. This comprehensive feature representation is very useful in subsequent analysis, modeling, or prediction tasks, because it can comprehensively consider the information of the data at different time steps to better understand and predict the behavior of the store's total sales. This combination process helps to making the most of the key knowledge in the time series data to more effectively deal with various factors and problems in the specific application of store sales forecasting. Figure 4 represents a sales forecast graph using the GRU model.

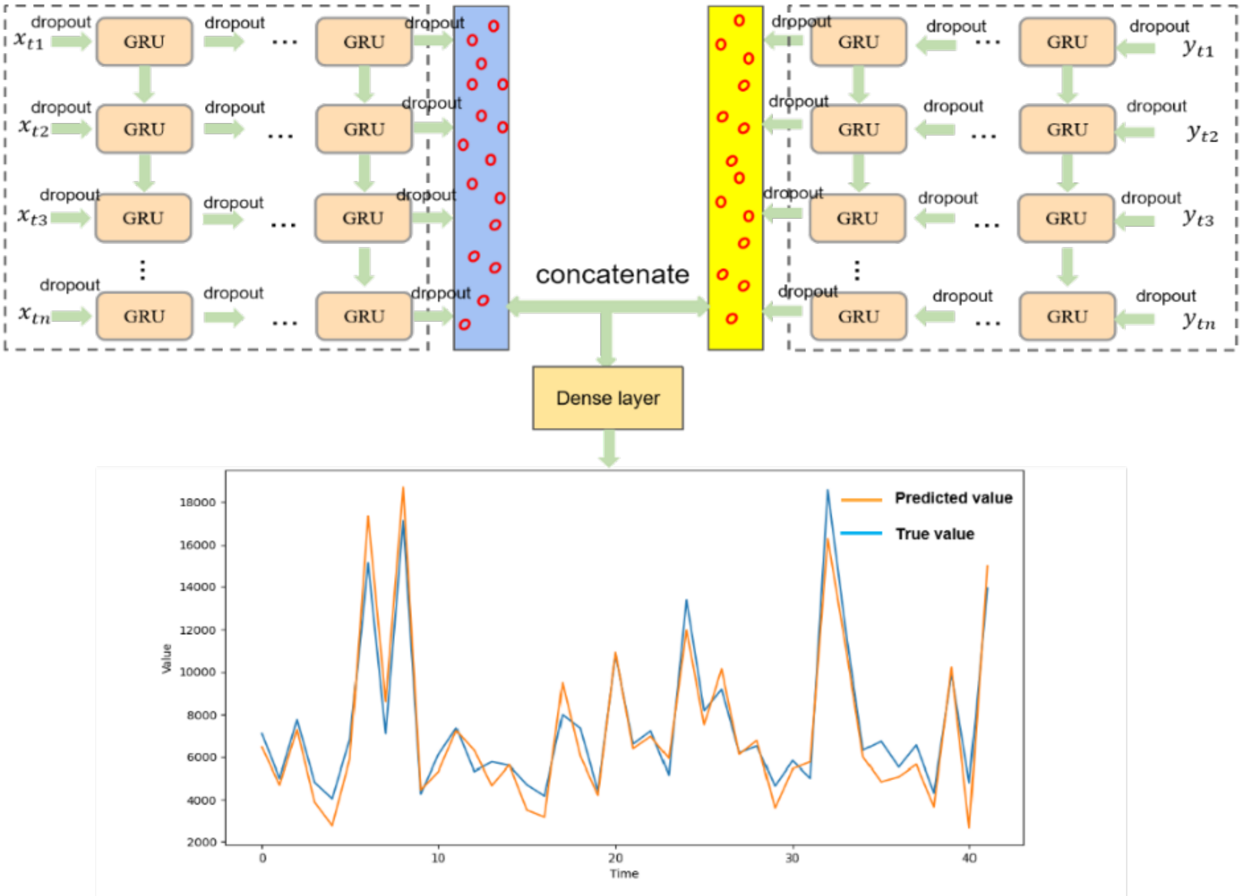

Fig. 4. GRU model training

4. Results and Discussion

The earlier chapters, the research methods are described in detail, including data preprocessing, model construction, training, and prediction. In this part, the main findings of the research and their relevance to the research objectives will be analyzed and discussed. Explore the performance of the model, especially in the prediction of sales data, and compare the different recurrent neural network (RNN) models used (such as GRU, LSTM, etc.). At the same time, the prediction error, trend capturing ability and feasibility of practical application of the model will be discussed in depth. Finally, the research results are summarized, emphasizing their importance and potential applications in the field of sales forecasting.

4.1 RNN Prediction

Below are the specific experimental results, comparing the predicted data with the raw data, where Figure 5 shows the results of the sales prediction using the RNN (recurrent Neural network) models:

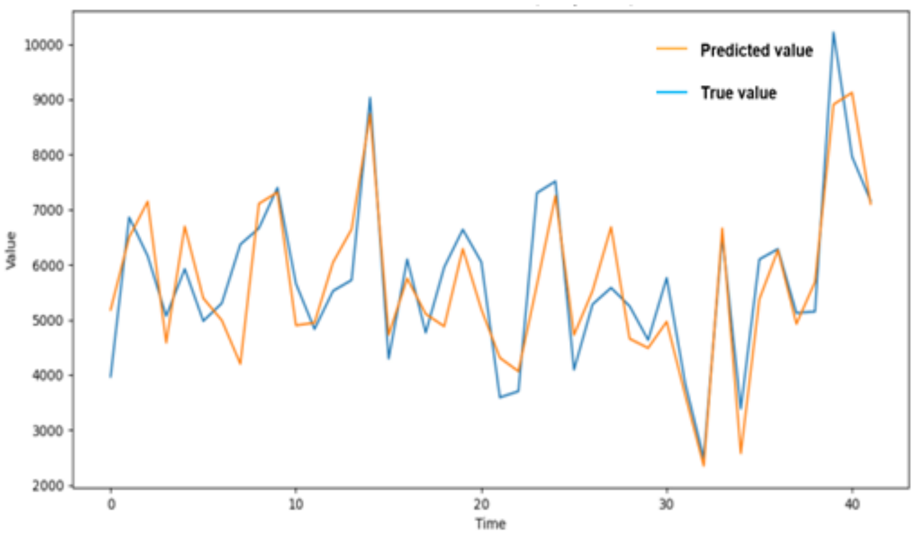

Fig. 5. RNN prediction results

The training process of RNN found that when running from 120 epochs to 30 epochs, the RNN model has converged. The model has adjusted most of its parameters, and the adjustment space has become limited, making it difficult to further reduce the loss function and evaluation indicators. The RNN model is prone to gradient disappearance or gradient explosion problems, resulting in the inability to effectively propagate gradients and making it impossible to continue optimizing the model. At 20 epochs, the model is mainly learning the general features of the data. As the training progresses, the model begins to overfit the training data, resulting in reduced generalization performance. Therefore, measures for model adjustment and gradient clipping have been proposed to address the training problem, but for the data set used in this study, the training performance is still not good.

4.2 LSTM and GRU Predictions

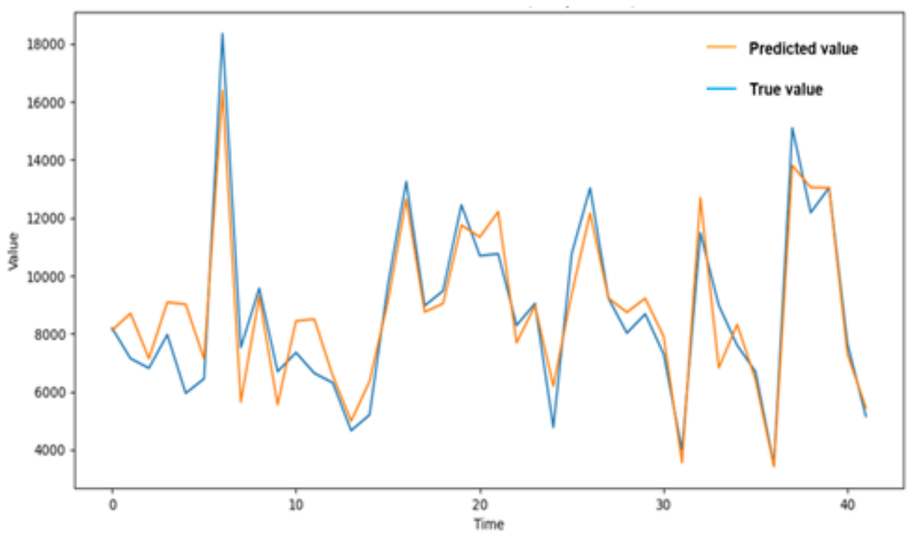

Fig. 6. LSTM prediction results

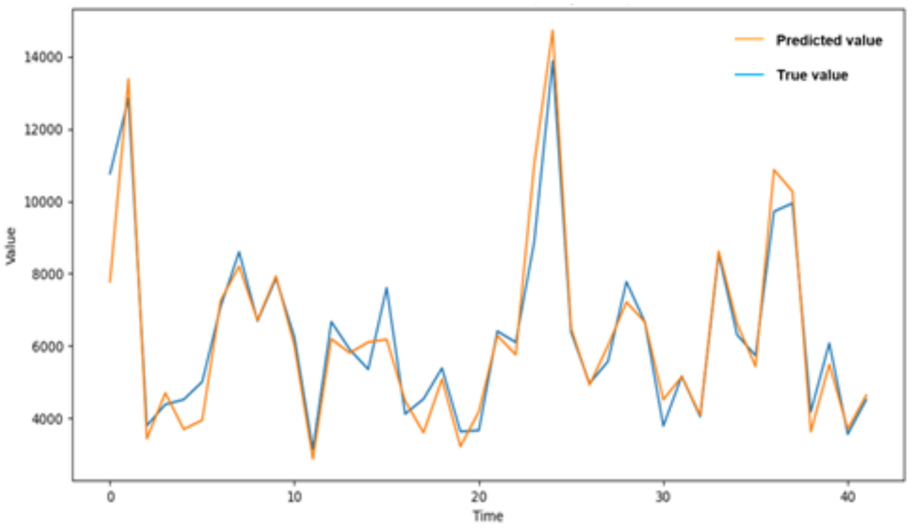

Fig. 7. GRU prediction results

During the experimental training process, the training performance of LSTM (long short-term memory network) and GRU (gated recurrent unit) is not similar, but GRU performs better. According to the characteristics of the data set used in this experiment, it may be more suitable for the structure of the GRU model. GRU may be more suitable for some sequence data than LSTM because GRU has a simpler structure and is more suitable for processing simpler sequence patterns. LSTM has more parameters, making it easier to overfit on the current data set. Because GRU has fewer parameters, it has better generalization ability and avoids the over-fitting problem encountered by LSTM.LSTM is more likely to encounter the problem of gradient disappearance or gradient explosion, and GRU encounters fewer gradient problems during the training process due to fewer gated units. In the experiment, tried to adjust the parameters of LSTM and GRU, such as the number of layers, the number of hidden units, etc., to make them better adapt to the characteristics of the data set. But the GRU model still performs better than LSTM.

5. Evaluation Index

The study uses MAE, RMSE, and R^2 to evaluate the model. To compare the effectiveness of various time series forecasting algorithms, research evaluate models using three common criteria, mean absolute error (MAE), namely root square error (RMSE), and R^2, which are widely used in regression tasks. The following definitions describe the equations for both measurements.

$$\text{RMSE}=\sqrt{\frac{\sum_{i=1}^{n}(y_t^i-\hat{y}_t^i)^2}{n}} \tag{5}$$

The root means square error (RMSE), a measurement of the discrepancy between expected values and actual observed values, is frequently used to assess the predictive power of prediction models. y_t^i and $\hat{y}_t^i$ stand for the corresponding actual and predicted sales among them, whereas n denotes the total number of anticipated days.

$$\text{MAE}=\frac{\sum_{i=1}^{n}|y_t^i-\hat{y}_t^i|}{n} \tag{6}$$

Mean Absolute Error, or MAE, is a frequently used metric to assess a model's prediction error. It is employed to calculate the mean absolute error between the observed value and the expected value.y_t represents the actual sales volume at time t, while $\hat{y}_t$ represents the predicted sales volume at time t.

$$R_t^2=1-\frac{\sum_{i=1}^{n}(y_t^i-\hat{y}_t^i)^2}{\sum_{i=1}^{n}(y_t^i-\overline{y_t})^2} \tag{7}$$

To determine how well the model fits the observed data at a particular time point t, the R-squared at that point is represented by the letter R_t^2 . The mean at moment t is depicted by $\overline{y_t}$.

Table 1
RNN, LSTM, GRU evaluation model performance

Model	MAE	RMSE	R^2
RNN	1608	2439	0.3092
LSTM	596	868	0.8773
GRU	511	691	0.9124
Model	MAE	RMSE	R^2

6. Model Selection

The GRU model yields the lowest MAE and RMSE on this problem, hence the study suggests using it based on the MAE and RMSE results. LSTM comes next, and RNN follows. However, other factors should also be considered in the final model selection, such as model complexity, training time, etc.

In view of the issues that should be considered in the final model, the experiment also conducted a comparative analysis of the loss function Loss. In this experiment, the GRU model exhibits the lowest loss function value (0.0009), relative to RNN (0.0056) and LSTM (0.0021), which indicates that the GRU model does the best job of fitting the training data. A lower loss function means that the model can be closer to the true value of the training data, and a smaller loss usually means that the model's predictions are more accurate. The LSTM model's loss function value, 0.0021, falls between RNN and GRU. This shows that even if LSTM performs better than RNN, there is still room for improvement when compared to GRU. In comparison to LSTM and GRU, the RNN model has the greatest loss function value (0.0056), possibly indicating that it has the least capacity for fitting. The predictions of the ARIMA model are also based on the linear combination of observations and model parameters at previous time points for the ARIMA time series analysis and prediction approach utilized in the (Rejaul Islam Royel) Kaggle data set. According to the study, GRU outperforms ARIMA in terms of sales forecast because its RMSE 691 is lower than ARIMA's 867.

7. Conclusion

This study compared the performance of three different recurrent neural network architectures - RNN, LSTM, and GRU - when predicting e-commerce sales. Through research and analysis it can be made clear that the GRU works well, showing smaller errors and higher accuracy. This demonstrates the important advantages of estimating e-commerce revenue. Through this research, the e-commerce industry will be able to more accurately predict sales trends, effectively manage inventory, and improve operational efficiency. Future research will focus on a larger range of activities and data sets, as well as exploring more complex network topologies and optimization methods to further improve the efficiency and adaptability of the model.

Acknowledgement
Appreciative of Professor Annie Anak Joseph's advice, which has improved my research. We thank Wang Wanzhen for her essential assistance with the experimental design. A special thank you for your steadfast support, family and friends. This research was not funded by any grant.

References

[1] Verstraete, Gylian, El-Houssaine Aghezzaf, and Bram Desmet. "A data-driven framework for predicting weather impact on high-volume low-margin retail products." *Journal of Retailing and Consumer Services* 48 (2019): 169-177. https://doi.org/10.1016/j.jretconser.2019.02.019

[2] Succetti, Federico, Francesco Di Luzio, Andrea Ceschini, Antonello Rosato, Rodolfo Araneo, and Massimo Panella. "Multivariate prediction of energy time series by autoencoded LSTM networks." In *2021 IEEE International Conference on Environment and Electrical Engineering and 2021 IEEE Industrial and Commercial Power Systems Europe (EEEIC/I&CPS Europe)*, pp. 1-5. IEEE, 2021. https://doi.org/10.1109/EEEIC/ICPSEurope51590.2021.9584744

[3] Zhai, Naiju, Peifu Yao, and Xiaofeng Zhou. "Multivariate time series forecast in industrial process based on XGBoost and GRU." In *2020 IEEE 9th Joint International Information Technology and Artificial Intelligence Conference (ITAIC)*, vol. 9, pp. 1397-1400. IEEE, 2020. https://doi.org/10.1109/ITAIC49862.2020.9338878

[4] Bui, Van, Tung Lam Pham, Joongheon Kim, and Yeong Min Jang. "RNN-based deep learning for one-hour ahead load forecasting." In *2020 International conference on artificial intelligence in information and communication (ICAIIC)*, pp. 587-589. IEEE, 2020. https://doi.org/10.1109/ICAIIC48513.2020.9065071

[5] Heinemann, Gerrit. "Intelligent Retail." In *Intelligent Retail: The Future of Stationary Retail*, pp. 109-229. Wiesbaden: Springer Fachmedien Wiesbaden, 2022. https://doi.org/10.1007/978-3-658-38316-9_3

[6] Toğa, Gülhan, Berrin Atalay, and M. Duran Toksari. "COVID-19 prevalence forecasting using autoregressive integrated moving average (ARIMA) and artificial neural networks (ANN): case of Turkey." *Journal of infection and public health* 14, no. 7 (2021): 811-816. https://doi.org/10.1016/j.jiph.2021.04.015

[7] Amshi, Ahmad Hauwa, and Rajesh Prasad. "Time series analysis and forecasting of cholera disease using discrete wavelet transform and seasonal autoregressive integrated moving average model." *Scientific African* 20 (2023): e01652. https://doi.org/10.1016/j.sciaf.2023.e01652

[8] Li, Yuanzheng, Shangyang He, Yang Li, Leijiao Ge, Suhua Lou, and Zhigang Zeng. "Probabilistic charging power forecast of EVCS: Reinforcement learning assisted deep learning approach." *IEEE Transactions on Intelligent Vehicles* 8, no. 1 (2022): 344-357. https://doi.org/10.1109/TIV.2022.3168577

[9] Sbrana, Attilio, André Luis Debiaso Rossi, and Murilo Coelho Naldi. "N-BEATS-RNN: deep learning for time series forecasting." In *2020 19th IEEE International Conference on Machine Learning and Applications (ICMLA)*, pp. 765-768. IEEE, 2020. https://doi.org/10.1109/ICMLA51294.2020.00125

[10] Zhu, S., J. Yao, and S. Z. Djokic. "Comparison of two modified deterministic LSTM models with a probabilistic LSTM model for a day-ahead forecasting of electricity demands." In *2023 IEEE Belgrade PowerTech*, pp. 1-6. IEEE, 2023. https://doi.org/10.1109/PowerTech55446.2023.10202847

[11] Xue, Jiaxiang, Xuan Hu, Haifeng Chen, and Gang Zhou. "Research on lstm-xgboost integrated model of photovoltaic power forecasting system." In *2022 14th international conference on intelligent human-machine systems and cybernetics (IHMSC)*, pp. 22-25. IEEE, 2022. https://doi.org/10.1109/IHMSC55436.2022.00014

[12] Wensheng, Li, Wu Kuihua, Feng Liang, Li Hao, Wang Yanshuo, and Cui Can. "A region-level integrated energy load forecasting method based on CNN-LSTM model with user energy label differentiation." In *2020 5th International Conference on Power and Renewable Energy (ICPRE)*, pp. 154-159. IEEE, 2020. https://doi.org/10.1109/ICPRE51194.2020.9233226

[13] Luo, Qing, Yi Chen, Chengyao Gong, Yuanyu Lu, Yi Cai, Yan Ying, and Guanhong Liu. "Research on short-term air conditioning cooling load forecasting based on bidirectional LSTM." In *2022 4th International Conference on Intelligent Control, Measurement and Signal Processing (ICMSP)*, pp. 507-511. IEEE, 2022. https://doi.org/10.1109/ICMSP55950.2022.9858975

[14] Bilgili, Mehmet, and Engin Pinar. "Gross electricity consumption forecasting using LSTM and SARIMA approaches: A case study of Türkiye." *Energy* 284 (2023): 128575. https://doi.org/10.1016/j.energy.2023.128575

[15] Bin, Wang, Wang Yang, Cheng Yan, Yu Min, and Wang Zhen. "A short-term power load forecasting method based on eemd-abgru." In 2020 Chinese Control And Decision Conference (CCDC), pp. 5574-5579. IEEE, 2020. https://doi.org/10.1109/CCDC49329.2020.9164880

[16] Cheng, Yali, Haiqing Chang, Kai Tang, Jianhang Zou, Jianming Zhuo, and Yijun Cai. "Multistep electricity load forecasting method based on the hybrid GRU neural network." In *2022 International Applied Computational*

Electromagnetics Society Symposium (ACES-China), pp. 1-3. IEEE, 2022. https://doi.org/10.1109/ACES-China56081.2022.10065306

[17] Gu, Yu, Fandi Wang, Mukun Li, Lu Zhang, and Wenlong Gong. "A digital load forecasting method based on digital twin and improved GRU." In *2022 Asian Conference on Frontiers of Power and Energy (ACFPE)*, pp. 462-466. IEEE, 2022. https://doi.org/10.1109/ACFPE56003.2022.9952254

[18] Guo, Lei, Peiran Shi, Yong Zhang, Zhengfeng Cao, Zhuping Liu, and Bin Feng. "Short-term EV charging load forecasting based on GA-GRU model." In *2021 3rd Asia Energy and Electrical Engineering Symposium (AEEES)*, pp. 679-683. IEEE, 2021. https://doi.org/10.1109/AEEES51875.2021.9403141

[19] Sajjad, Muhammad, Zulfiqar Ahmad Khan, Amin Ullah, Tanveer Hussain, Waseem Ullah, Mi Young Lee, and Sung Wook Baik. "A novel CNN-GRU-based hybrid approach for short-term residential load forecasting." *Ieee Access* 8 (2020): 143759-143768. https://doi.org/10.1109/ACCESS.2020.3009537

[20] Liang, Zhang, Zheng Chengyuan, Zhao Zhengang, and Zhang Dacheng. "Short-term load forecasting based on kalman filter and nonlinear autoregressive neural network." In *2021 33rd Chinese Control and Decision Conference (CCDC)*, pp. 3747-3751. IEEE, 2021.. https://doi.org/10.1109/CCDC52312.2021.9602793