Penerbit Akademia Baru

Journal of Advanced Research Design

Journal homepage:
https://akademiabaru.com/submit/index.php/ard
ISSN: 2289-7984

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

Rong Liu[1,2], Annie Anak Joseph[1,*], Wanzhen Wang[2], Feng Cao[2], Na Li[2], Taihao Zhang[3]

[1] Department of Electrical and Electronic Engineering, Faculty of Engineering, Universiti Malaysia Sarawak (UNIMAS), Kota Samarahan, Sarawak, Malaysia
[2] Department of Electronic Information, Qilu Institute of Technology, Jinan, PR. China
[3] Faculty of Computer Science and Technology, Qilu University of Technology, Jinan, PR. China

ARTICLE INFO

Article history:
Received 27 March 2025
Received in revised form 18 August 2025
Accepted 29 September 2025
Available online 10 October 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,

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

https://doi.org/10.37934/araset.145.1.136148

Recurrent neural network models in particular, such the long short-term memory network (LSTM) [2] 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