A Comparison of Seasonal ARIMA and LSTM in Forecasting Demand Time Series

Kevin Fernando
Mathematics Department
School of Computer Science
Bina Nusantara University
Jakarta, Indonesia 11480
kevin.fernando004@binus.ac.id

Kevin Tanoto
Mathematics Department
School of Computer Science
Bina Nusantara University
Jakarta, Indonesia 11480
Kevin.tanoto001@binus.ac.id

Alexander Agung Santoso Gunawan
Mathematics Department
School of Computer Science
Bina Nusantara University
Jakarta, Indonesia 11480
**Corresponding author:*
aagung@binus.edu

***Abstract*—Time series forecasting has been used in many subjects. While planning, the more accurately we predict the future, the better we can position ourselves for it. For some time ago, to forecast the time series we used Moving Average, Exponential Smoothing, and foremost among them is Autoregressive Integrated Moving Average (ARIMA). As technology advances, we continue to employ the conventional methods on a regular basis, but machine learning has led many firms to desire higher profits with lower expenses. For demand forecasting, a variety of statistical models and machine learning have been widely used. This paper will investigate whether the traditional statistic model ARIMA performs better in forecasting demand time series with the advanced model of machine learning such as Long Short-Term Memory (LSTM). Some papers suggest that machine learning is not always better than traditional model in forecasting time series. The dataset is taken from a publicly used data resource such as Kaggle. From that dataset, the third section then will simulate each model. The final section will show the simulation results.**

Keywords*— Autoregressive Integrated Moving Average (ARIMA), Machine Learning, Forecasting, Long Short-Term Memory (LSTM), Time Series.

I. Introduction

In many scientific fields, including economics, business, marketing, and others, forecasting is a crucial activity. For instance, accurate sales predictions will undoubtedly make production planning easier. The aim of forecasting time series is to make predictions and guide strategic decision-making which involves examining time series data using statistic modeling. The prediction is not always accurate, and the probability of the prediction value could change a lot depending on the variables, for example if the variable often goes up and down in the data with the influence from outside. The more complete data available the more accurate the model forecast. To analyze the time series, it requires the creation of the model in order to examine the data and determine the problem.[1]. Analysis can describe the reason things happen. The next step after discovering the information from forecasting is to decide what to do with the information itself.

For a single timeseries data, the most popular technique is univariate "Auto-Regressive Moving Average (ARIMA). ARIMA stands for Auto-Regressive (AR) and Moving Average (MA). The ARIMA model is a forecasting tool that fully disregards independent variables. To generate precise short-term forecast, ARIMA model uses the dependent variable's historical and current values. It is appropriate to use ARIMA if the time series is statistically connected to one another [10].

Through machine learning and deep learning algorithms, the approach to prediction problems where relationships between variables are expressed in deep and complex structure. LSTM is a popular deep learning algorithm that has been used many in time series predictions [10].

This study aims to investigate which forecasting methods produce a better result for predicting time series based on fewer forecast error. This is done because not many papers cover the prediction performance between traditional model and machine learning model. The prediction model uses statistical models such as ARIMA and machine learning models such as LSTM.

This paper consists of 5 sections. Section 1 will cover the background and research objectives. Section 2 will highlight previous studies. Section 3 will explain the model the author used and the description of the dataset. Section 4 will be the result of the experiment and finally the conclusion of this paper will be on section 5.

II. Literature Review

In the paper [11], Spyros carried out the time series analysis by making a comparison between econometric and ML methods. They make the comparison by predicting 1,045 monthly time series for 18 horizons, using both ML and statistical methods. Based on the standards for "Mean Absolute Percentage Error" and "Mean Absolute Relative Error", they concluded that statistical methods are better predictors than the ML methods for all horizons.

In the study [10], is the closest article related to this paper. The study is easy to understand, and the study also attached the algorithm of both ARIMA and LSTM. The validation they used is Root Square Means Error. The study demonstrates that LSTM outperforms ARIMA by far in forecasting historical finance time series.

In the study [17], the author also compares ARIMA model and LSTM model for forecasting temperature. The author uses MAE as the evaluation. The study also demonstrates that LSTM performs better than ARIMA when forecasting temperature.

Based on the study, there is not much research showing the performance difference between statistical model and machine learning model. This paper will use ARIMA and LSTM model and evaluate the model using MAPE.

III. Methodology

This section will review the model technique that will be used in this paper. Techniques used in our analysis include:

- Auto-Regressive Moving Average (ARIMA)
- Long short-term memory (LSTM)

A. ARIMA

Autoregressive Integrated Moving Average Model or ARIMA is an old and widely used for univariate time series forecasting. ARIMA notes as *(p, d, q)* with 3 parameters autoregression (AR), differencing (I), and moving average (MA). The AR model can be described in a simple form as:

$$x_t = c + \sum_{i=1}^{p} \phi_i x_{t-i} + \varepsilon_t \quad (1)$$

Which:

x_t : Stationary variable
c : Constant
ϕ_i : Autocorrelation coefficients
ε_t : The residuals

The MA model can be described as:

$$x_t = \mu + \sum_{i=0}^{q} \theta_i \varepsilon_{t-i} \quad (2)$$

Which:

μ : The expectation of x_t
θ_i : The weights applied
ε_t : The residuals.

Then combine models (1) and (2) by adding them to form ARIMA model:

$$x_t = c + \sum_{i=1}^{p} \phi_i x_{t-i} + \varepsilon_t + \sum_{i=0}^{q} \theta_i \varepsilon_{t-i} \quad (3)$$

Where $\phi_i \neq 0, \theta_i \neq 0$, and $\sigma_\varepsilon^2 > 0$. ARIMA can handle data with a trend, but it does not work with seasonal data. Which is why the author will use SARIMA or Seasonal ARIMA. SARIMA model considers both seasonal and non-seasonal elements in multiplication model. The notation for SARIMA is $(p, d, q)\,(P, D, Q)m$ where m is number observation per year, (p, d, q) is the non-seasonal parts of the model and $(P, D, Q)m$ is the seasonal parts of the model. The seasonal parts of the model are like the non-seasonal parts except that include backshifts and seasonal period.

Finding the value of *(p, d, q)* and *(P, D, Q)* is crucial in estimating SARIMA [10]. To find the best parameters for the model. The author uses grid search to find the lowest AIC. An estimate of the overall value of statistical models for a specific data set is the Akaike information criterion (AIC). AIC evaluates a model's ability to fit the data while also taking into consideration the model's overall complexity. By using AIC the model uses fewer features yet has a similar fit. That's why the lower the AIC the better.

B. LSTM

An artificial neural network is a system of interconnected neurons that have been trained with specific activation functions using a particular optimization method. Artificial neural networks are composed of individual units called neurons, which are organized in layers. The neurons in each subsequent layer receive the output from the neurons in the layer above them. This structure ensures that the flow of brain activations is unidirectional, layer by layer. Typically, A neural network consists of at least two layers, one for input and one for output. However, To enhance the network's computational abilities, extra layers referred to as hidden layers can be included between the input and output layers. These hidden layers can serve as a "universal approximator" if they have enough units.

An advanced type of Deep learning and AI applications utilize a type of artificial neural network known as Long Short-Term Memory (LSTM). Unlike traditional feedforward neural networks, LSTM networks have feedback connections. This capability allows them to handle not only individual data points but also sequences of data as a whole. Applications of LSTM such as machine translation, handwriting recognition, robot control, video games, healthcare, and speech recognition.

The structure of an LSTM unit typically includes a cell, an input gate, an output gate, and a forget gate. These gates regulate the flow of information into and out of the cell, where data is stored for an extended period of time.

LSTM networks are ideal for working with time series data as they can handle lag times of varying duration between important events. LSTMs were designed to overcome the gradient disappearance issue that could happen when training traditional RNNs. They are also less sensitive to gap lengths than RNNs, hidden Markov models, and other sequence learning techniques, which makes them well-suited for a variety of use cases.

LSTMs are a specific type of RNN that have additional capabilities for remembering sequence data. Sequences of data are collected and saved in the LSTM's collection of cells, also called system modules. Every cell has a pathway that connects one module to another, enabling data to flow from the past and be gathered for the present. The purpose of gates in every cell allows for the manipulation of data within the cell, such as removing, filtering or adding data for the following cells. The gates, which are based on a sigmoidal neural network layer, enable the cells to decide whether to let input pass through or discard it.

Every single sigmoid layer generates values between 0 and 1, which indicates the proportion of data that should be permitted to pass through in each cell. Specifically, a value of 0 means "block all data" and a value of 1 means "allow all data". [10].

C. Dataset

The data is taken from Kaggle [1]. The dataset consists of historical product demand for a manufacturing corporation. Within dozens of product categories, the corporation offers thousands of products. There are 4 categories of warehouse which is Whse_A, Whse_S, Whse_C, and Whse_J, with Whse_J has largest number of samples. There are also 33 product categories and 2160 unique product codes. The considered datasets are gathered from 2011-01-08 until 2017-01-01. The following is table 1, which shows an example of the dataset.

Table 1. Sample from the dataset

Product Demand	Product Code
	Warehouse
	Product Category
	Date
	Order Demand

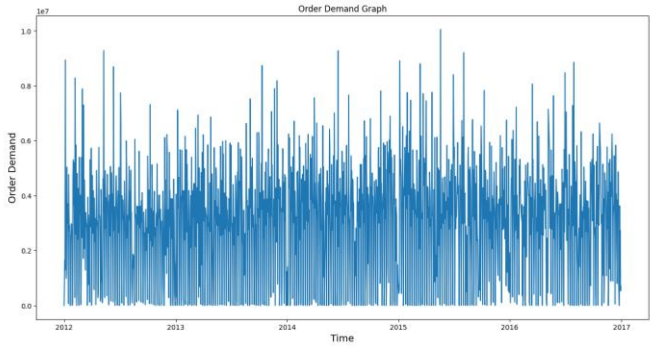

Figure 1. Product Demand over the period

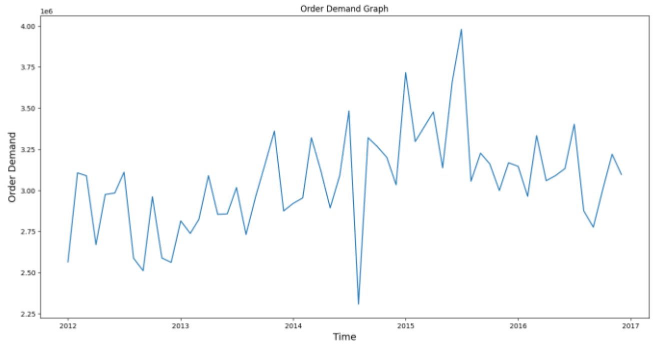

Figure 2. Product Demand Monthly

The objective of each model is to forecast demand for customer orders. The main challenge from the data is that the dataset itself is highly skewed and lacks normal distribution. While checking the data it is also shown that the are some missing values with null to dataset ratio in date is about 1%, which is why the missing values are taken out. According to Fig. 2 the sales are low in the beginning of each year, but the demand peaked every year in the last quarter with the year 2014-2016 were the highest and then reducing. In the initial data processing, the year 2011 is taken out because there are some missing values between the dates and there are few data before the year 2012.

In the initial data processing for each model, the author split the dataset into monthly demand and then split the dataset into two for LSTM model in which 80% data set was used as training and the remaining 20% was used as testing.

D. Evaluation Metric

For this paper, the author will use Mean Absolute Percentage Error (MAPE).

$$MAPE = \frac{1}{n}\sum_{i=1}^{n} \left| \frac{\left(y_i - y_{pred}\right)}{y_i} \times 100 \right| \quad (4)$$

Which:

y_i : The true value

y_{pred} : The predicted value

n : The data

The value of MAPE is in percentage, the prediction of each model will be evaluated with this metric. If the value of MAPE is small, it means the model can predict better.

IV. Result

The results of demand forecasting for the last 12 months are shown in Fig. 3 and Fig. 4. The orange line represents the prediction value, and the blue line represents the true value.

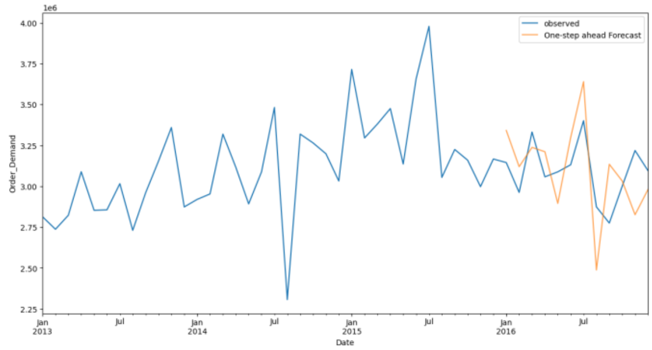

Figure 3. Forecasting results by SARIMA model

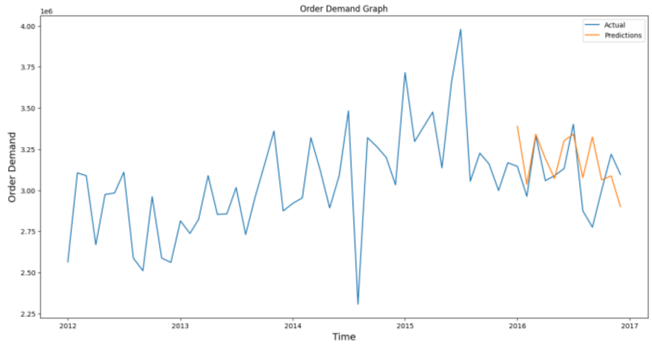

Figure 4. Forecasting results by LSTM model

As shown in Table 2, the author uses equation (4) as an evaluation to validate the accuracy of both SARIMA and LSTM. The SARIMA model shows a better performance by having a smaller value of MAPE which is 7.9706. The error obtained by SARIMA is 15% lower than the LSTM model. It can be concluded in this case the SARIMA model performs better than LSTM model.

Table 2. Comparison of MAPE for each model

Model	MAPE
SARIMA	7.9706
LSTM	9.4080

V. Conclusion

This study is focused on monthly forecasting demand time series with traditional model and machine learning model. Through analysis of the demand time series, we can read patterns to tell a story which helps a lot in planning business decisions. The purpose of this study is to compare the machine learning model performance by looking at the MAPE performance. The results show that ARIMA performance are better than LSTM in forecasting monthly demand time series. This proves that the traditional model could rival the machine learning model in forecasting time series.

VI. References

[1] D. Ni, Z. Xiao and M. K. Lim, "Int. J. Mach. Learn. & Cyber," *A systematic review of the research trends of machine learning in supply chain management,* p. 1463–1482, 2020.

[2] F. Chen, Z. Drezner, J. K. Ryan and D. Simchi-Levi, "Management Science," *Quantifying the Bullwhip Effect in a Simple Supply Chain: The Impact of Forecasting, Lead Times, and Information,* pp. 436-443, 2000.

[3] J. Feizabadi, "International Journal of Logistics Research and Applications," *Machine learning demand forecasting and supply chain performance,* vol. 25, no. 2, pp. 119-142, 2022.

[4] N. Vandeput, Data Science for Supply Chain Forecasting, Berlin/Boston: Deutsche Nationalbibliothek, 2021.

[5] R. Sharma, S. S. Kamble, A. Gunasekaran, V. Kumar and A. Kumar, "Computers & Operations Research," *A systematic literature review on machine learning applications for sustainable agriculture supply chain performance,* vol. 119, p. 104926, 2020.

[6] H. Stadtler, Supply Chain Management — An Overview, Berlin: Springer, Berlin, Heidelberg, 2008.

[7] S. Makridakis, E. Spiliotis and V. Assimakopoulos, "The Accuracy of Machine Learning (ML) Forecasting Methods versus Statistical Ones: Extending the Results of the M3-Competition," 2017.

[8] A. Sagheer and M. Kotb, "Time series forecasting of petroleum production using deep LSTM recurrent networks," *Neurocomputing,* vol. 323, pp. 203-213, 2019.

[9] C. Chatfield, Time-Series Forecasting (1st ed.), New York: Chapman and Hall/CRC, 2000.

[10] S. Siami-Namini, N. Tavakoli and A. S. Namin, "A Comparison of ARIMA and LSTM in Forecasting Time Series," *Conference on Machine Learning and Applications (ICMLA),* 2018.

[11] M. Spyros, S. Evangelos, Assimakopoulos and Vassilis, "The Accuracy of Machine Learning (ML) Forecasting Methods versus Statistical Ones: Extending the Results of the M3-Competition," 2017.

[12] Y. Yu, X. S, C. Hu and J. Zhang, "A Review of Recurrent Neural Networks: LSTM Cells and Network Architectures," *Neural Comput,* vol. 31, no. (7), p. 1235–1270, 2019.

[13] E. S. G. Jr., "Exponential smoothing: The state of the art—Part II," *International Journal of Forecasting,* vol. 22, p. 637 – 666, 2006.

[14] A. Syntetos, Z. B. J. Boylan and S. Kolassa, "Supply Chain Forecasting: Theory, Practice, their Gap and the Future," *European Journal of Operational Research,* vol. 252, no. 1, pp. 1-26, 2016.

[15] A. Aamer, L. P. E. Yani and I. M. A. Priyatna, "Data Analytics in the Supply Chain Management: Review of Machine Learning Applications in Demand Forecasting," *OSCM Publications,* vol. 14, no. 1, pp. 1-13, 2021.

[16] F. Zhao. [Online]. Available: https://www.kaggle.com/datasets/felixzhao/productdemandforecasting.

[17] E. D. Saa and L. Ranathunga, "Comparison between ARIMA and Deep Learning Models for Temperature Forecasting," 2020.