DOI: https://doi.org/10.59294/HIUJS.VOL.5.2023.552

Applying deep learning to forecast the demand of a Vietnamese FMCG company

Le Duc Dao* and Le Nguyen Khoi

Ho Chi Minh City University of Technology, Vietnam National University, Vietnam

ABSTRACT

In the realm of Fast-Moving Consumer Goods (FMCG) companies, the precision of demand forecasting is essential. The FMCG sector operates in a highly uncertain environment marked by rapid market shifts and changing consumer preferences. To address these challenges, the application of deep learning techniques, particularly Long Short-Term Memory (LSTM) networks, has emerged as a vital solution for enhancing forecast accuracy. This research paper focuses on the critical role of demand forecasting in FMCG, emphasizing the need for LSTM-based deep learning models to deal with demand uncertainty and improve predictive outcomes. Through this exploration, we aim to illuminate the link between demand forecasting and advanced deep learning, enabling FMCG companies to thrive in a highly dynamic business landscape.

Keywords: *demand forecast, ARIMA, deep learning, long-short term memory, FMCG*

1. INTRODUCTION

Within the domain of Fast-Moving Consumer Goods (FMCG), the importance of precise demand prediction remains of paramount significance [1]. The nature of the FMCG industry is represented by swift market fluctuations and ever-shifting consumer preferences. As product life cycles grow ever shorter and consumers become familiar with greater product variety, FMCG companies face increasing pressure to accurately anticipate future demand in order to optimize production schedules, inventory levels, supply chain coordination, promotional campaigns, workforce allocation, and other key operations that can make profit for them. However, the complex factors influencing product demand in the FMCG space often proves difficult to model using traditional statistical techniques. Demand drivers may include broad economic conditions, consumer confidence, competitive landscape, channel dynamics, weather patterns, commodity prices, cultural trends, and a myriad of other variables that can be difficult to quantify. While ARIMA (Autoregressive Integrated Moving Average) and other traditional forecasting techniques have been valuable tools for prediction in various fields, they often struggle to cope with the complexities of today's rapidly changing and highly dynamic world [2]. Such methods rely heavily on historical sales patterns continuing into the future. When conditions or consumer preferences shift suddenly, traditional models fail to account for new realities. Consequently, the adoption of advanced deep learning methodologies, particularly the Long Short-Term Memory (LSTM) networks, has gained prominence as an essential method for improving the precision of forecasts [3]. LSTMs and related recurrent neural network architectures possess provide advantages in processing time series data, identifying subtle patterns across long time lags, and adapting predictions based on newly available information. Inspired by the workings of human memory, LSTM models can learn context and discard outdated assumptions in light of updates, much as a supply chain manager would after noticing an impactful new trend. By combining the basic statistical foundation of methods like ARIMA with the pattern recognition capabilities of deep learning, FMCG forecasting stands to become significantly more accurate and responsive to fluctuations in consumer demand. Stems from the fact that ARIMA's ability to model linear historical patterns and LSTM's ability for uncovering nonlinear relationships, a combined forecast of ARIMA and LSTM were proposed to guide the direction of this research, considering the nature of products. Further investigations into optimal model architectures, hyperparameter tuning, and

Corresponding author: Le Duc Dao
Email: lddao@hcmut.edu.vn

ensemble techniques offer rich potential to enhance predictive power even in turbulent markets. As the FMCG landscape grows more complex each year, harnessing both statistical and machine learning will only increase in necessity to keep up with the pace of change.

2. CASE STUDY

The researched product is pre-packaged, and historical market demand data has been collected from January 2022 to May 2023. The current demand forecast is generated annually, using a one-month time bucket. Consequently, the company has encountered issues related to an excess of finished goods, resulting in overcapacity in the warehouse. These problems have adversely affected supply chain efficiency and financial flow. Days Inventory Outstanding (DIO) is among the key performance indicators used to evaluate the operational efficiency of the company. DIO stands for Days Inventory Outstanding and measures the average number of days that a company's inventory is held before it is sold or used up. This metric provides valuable insights into the efficiency of a company's inventory turnover and helps evaluate the effectiveness of the supply chain and inventory management process. In fact, the company has encountered a high DIO, around 50 days, with the target of reducing it to about 20 days. DIO is comprised of many factors, one of which is having accurate demand forecasts to ensure on-hand inventory is kept at appropriate levels. Therefore, a comprehensive analysis of demand forecasting is necessary to develop a new forecasting model with the purpose of improving the forecast accuracy for the company. To conduct this analysis, the first step will be data preparation and cleaning to ensure the demand data is accurate and consistent over the given time period. Statistical analysis such as trend, seasonality and residual decomposition will then be performed to understand the demand patterns. Potential forecasting methods to explore further include time series models like ARIMA models or advanced forecasting technique, including LSTM or the combined model. The parameters and fit of each model will be evaluated to select the one that optimizes error metrics like MAPE, MSE, MAD. Once an appropriate model is selected, it will be tested by forecasting by using the historical demand data. By improving demand planning, the company can better align production, inventory and distribution plans. This will increase supply chain agility, reduce waste, enable cost savings and ultimately provide better customer service. The overall goal is an integrated and intelligent demand forecasting approach customized for the business based on statistical best practices.

2.1. Data processing

Data has been collected from January 2022 to May 2023, in a weekly basis (74 observations). The data then being pre-processed to eliminate error and N/A values. *(Detail in Table 1)*.

Table 1. Demand from January 2022 to May 2023

Period	Demand	Period	Demand	Period	Demand	Period	Demand
1	78102	20	198599	39	140682	58	114350
2	112797	21	135898	40	78210	59	132432
3	132570	22	155856	41	102881	60	119826
4	65469	23	115008	42	104850	61	121932
5	39270	24	212886	43	101356	62	129282
6	120738	25	128238	44	98298	63	148685
7	126173	26	200184	45	103759	64	149196
8	169288	27	117263	46	112511	65	127824
9	180010	28	225381	47	115666	66	189222
10	131364	29	89120	48	108126	67	165828
11	107148	30	154791	49	96533	68	177150
12	177275	31	111870	50	120708	69	198114
13	163092	32	80339	51	129684	70	205284
14	147462	33	176513	52	150497	71	189090
15	154049	34	138088	53	62202	72	138092

16	156886	35	99042	54	114276	73	154218
17	174881	36	117530	55	53964	74	218382
18	98908	37	138040	56	130188		
19	142319	38	93319	57	148680		

- **Data analysis**

From the descriptive analysis, the dataset exhibits the following characteristics: The data ranges from 27,240 to 242,922, with an interquartile range of 109,062 to 155,590 and the presence of two outliers, detected using the 1.5 interquartile range rule [4]. The 1.5 interquartile range (IQR) rule is a statistical method to detect outliers in a dataset. It works by identifying any data points that fall more than 1.5 times the range from the first quartile to the third quartile. Points which fall outside those limits usually indicate unusual fluctuations in demand. Therefore, it is necessary to replace these values using the 1.5 interquartile range rule [4]. The final time series, after treating the outliers, is shown in Figure 1.

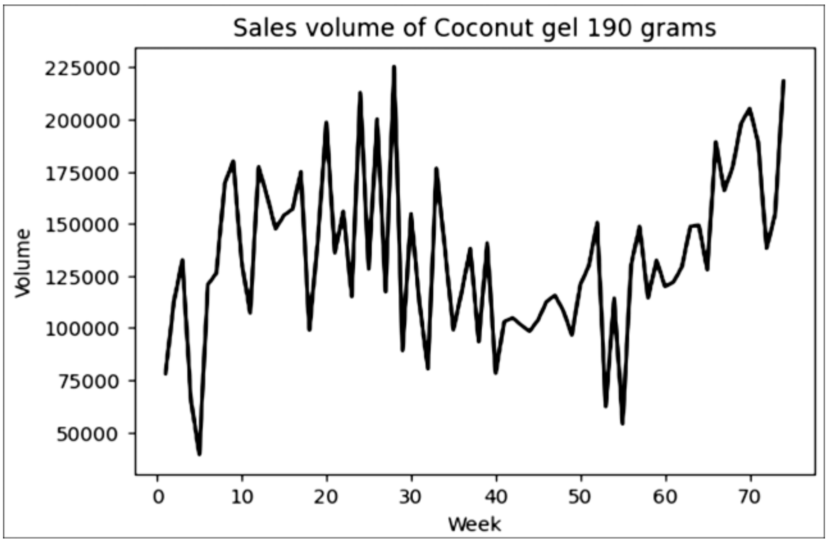

Figure 1. Time series plot after replacing outliers

The time series has been decomposed in Figure 2. Time series decomposition has enabled us to separate the time series into three main components: trend, seasonality, and residual. From the trend component, it appears that there is a slight upward trend. Seasonality occurs with a period of two, as many retailers tend to import the company's products on a bi-monthly basis. Additionally, numerous irregular fluctuations in demand result in the variation of residual data points.

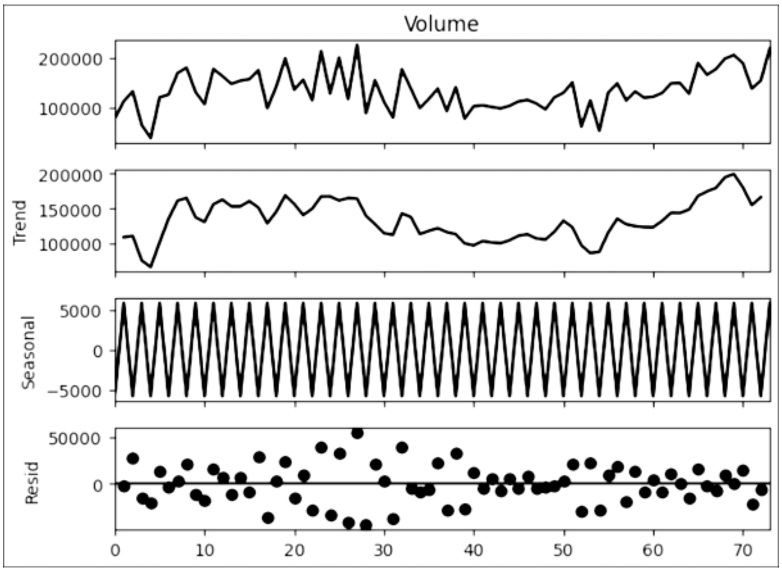

Figure 2. Time series decomposition

2.2. Model selection and evaluation

There are many methods that can be used to work well with the time series that has slight trend and seasonality with a strong irregular patter. ARIMA and LSTM models have been widely applied for time series forecasting tasks across domains. For instance, Williams et al [5] have developed seasonal ARIMA models to forecast traffic flow. The models outperformed historical average benchmarks. Ediger et al [6] have applied ARIMA forecast primary energy demand in Turkey by fuel type. The models were able to accurately forecast primary energy demand for each fuel type one to five years ahead, with lower errors than alternative extrapolation methods. On the other hand, LSTM also being applied in many research. Abbasimehr et al [7] have proposed an optimized LSTM model for product demand forecasting and compare performance against statistical methods. The optimized LSTM model significantly outperforms the statistical methods across all forecast horizons, while the ARIMA and SARIMA performance degrades significantly for longer horizon forecasts. In finance, Jiang et al [8] have developed a LSTM model to predict the stock market. The result states that the LSTM model outperformed the ARIMA model in forecasting stock prices in term of RMSE and MAPE metrics. However, some researches have provided the superiority of a combined LSTM-ARIMA model. G. Peter Zhang [9] has built a hybrid model combining ARIMA and neural networks for time series forecasting. The hybrid ARIMA-NN model significantly outperforms both individual models across all forecast horizons on the two datasets. Similarly, Dave et al [10] have developed a hybrid ARIMA-LSTM model to forecast Indonesia's monthly export values and compare performance to individual models. The ARIMA-LSTM hybrid model provides the most accurate forecasts with lowest MAPE and RMSE scores across all horizons. It improves on individual models by 3-10%. By referring these researches, ARIMA, LSTM and a hybrid ARIMA-LSTM model is selected for this paper.

2.2.1. ARIMA model

The ARIMA model relies on three fundamental parameters-p, d, and q-each representing a crucial aspect of the forecasting process. The variable "p" corresponds to the count of autoregressive terms (AR), indicating the reliance on past observations for predicting future values while "d" signifies the number of nonseasonal differences incorporated into the model, capturing the extent of data transformation needed to achieve stationarity. Lastly, "q" denotes the quantity of lagged forecast errors (MA), reflecting the influence of past errors on the current prediction. By analyzing ACF and PACF plots, optimal parameters are chosen based on the information criteria (AIC), so the most suitable model is the ARIMA (4,0,4) [11].

The ACF of residuals (Figure 3) shows that there is no lag value that fall outside the significant limits. Furthermore, the p-value (Figure 4) for lag 12, 24, 36, and 48 all greater than 0.05. Therefore, there is not enough evidence to reject the null hypothesis of no autocorrelation in the residuals, which can conclude that errors are random.

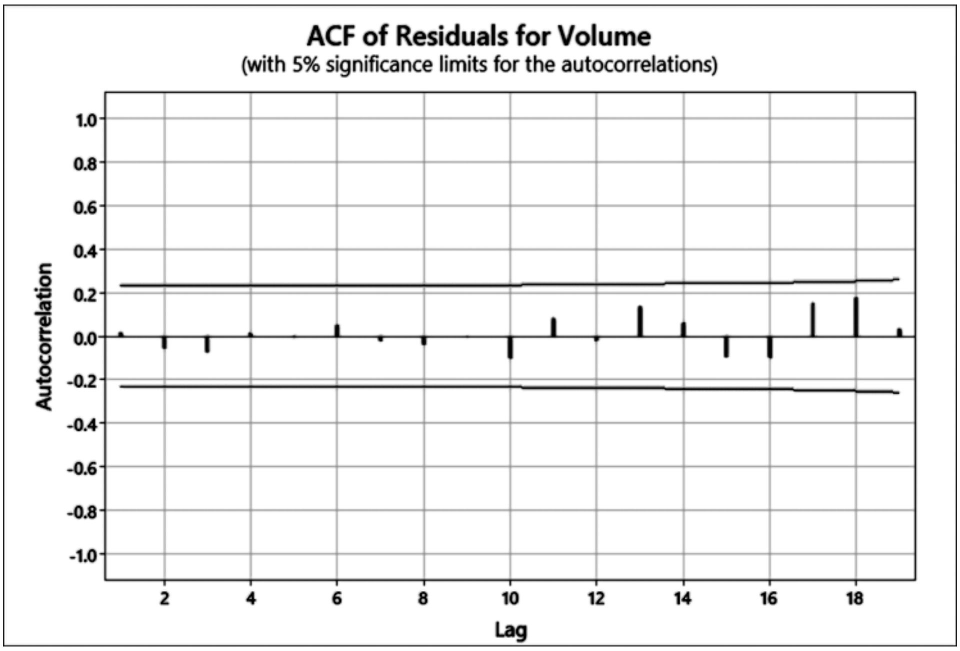

Figure 3. The ACF plot of ARIMA's residuals

Modified Box-Pierce (Ljung-Box) Chi-Square Statistic

Lag	12	24	36	48
Chi-Square	2.66	14.45	20.62	30.56
DF	4	16	28	40
P-Value	0.616	0.565	0.841	0.859

Figure 4. The modified Box-Pierce Chi-Square statistic result

2.2.2. LSTM model

The LSTM model operates with distinctive parameters that shape its architecture and influence its forecasting capabilities. Essential elements such as the number of memory cells, layers, and other architectural features play a pivotal role in capturing intricate temporal dependencies within the sequential data [12]. The LSTM model used in this paper is constructed with a sequential architecture, featuring input layers with a shape of (4, 1). The core part of the model lies in the LSTM layer with 256 units and a recurrent dropout of 0.2, allowing it to capture temporal dependencies and patterns within the input data. The subsequent dense layers, each with 64 units and ReLU activation, add non-linearity to the model, enhancing its capacity to learn complex relationships. The model is designed to predict a single output. During training, the mean squared error (MSE) is employed as the loss function, with the Adam optimizer utilizing a learning rate of 0.005. The model's performance is evaluated using mean absolute error as a metric. Training occurs over 200 epochs, with a batch size of 32. This architecture, through its LSTM structure and subsequent dense layers, is tailored to effectively capture and learn intricate patterns within sequential data, making it a potent tool for forecasting and prediction tasks.

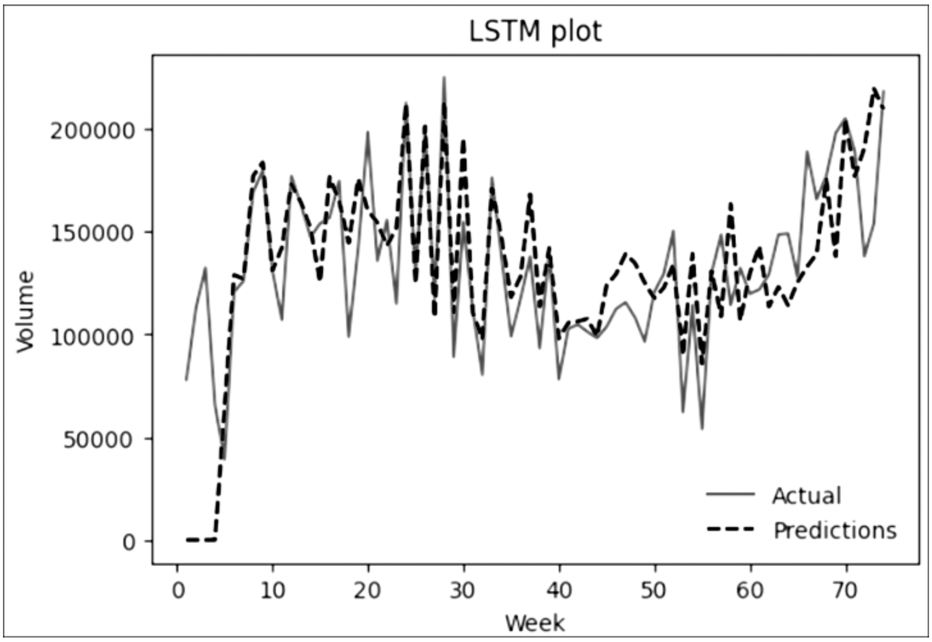

Figure 5. The LSTM model

2.2.3. The hybrid ARIMA-LSTM model

In general, both ARIMA and LSTM models have demonstrated success within their respective linear or nonlinear domains, but these methods can't be applied to all scenarios. ARIMA's approximation capabilities may fail to address complex nonlinear challenges, while LSTM, although suitable for handling both linear and nonlinear time series data, are hindered by prolonged training times and a lack of clear parameter selection guidelines [10]. Recognizing the limitations of each model, a hybrid approach is employed, leveraging the individual strengths of ARIMA and neural networks. This hybrid model aims to enhance prediction accuracy by allowing the models to complement each other, overcoming their individual weaknesses. This strategy recognizes the composite nature of time series, considering a linear autocorrelation alongside a nonlinear component [9]. With trend component, ARIMA is selected to capture that linearity. The model for seasonality and residuals is formed using LSTM to leverage its strength in non-linearity data. The final model is then combined from those three decomposed components.

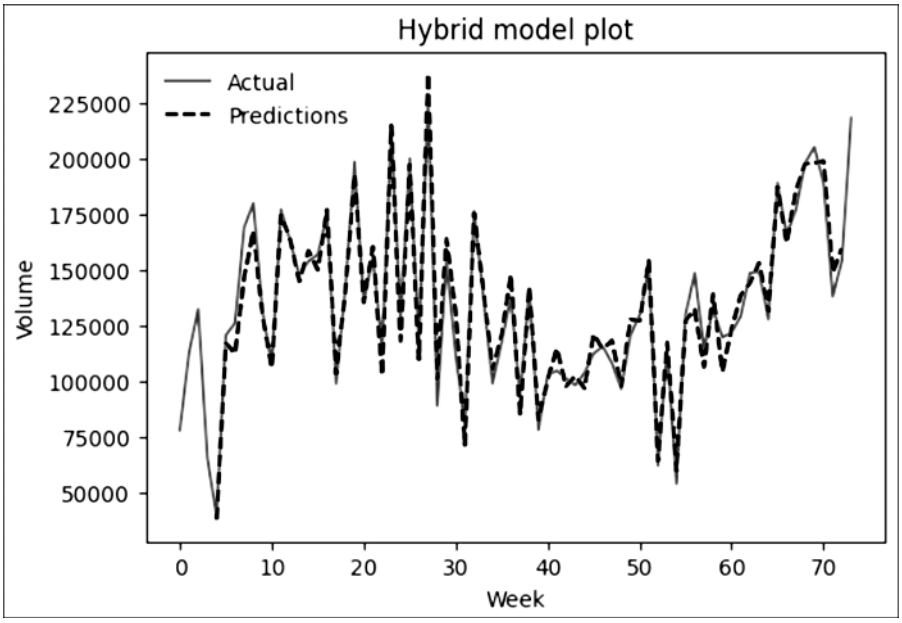

Figure 6. The hybrid model

2.3. Error comparison

For comparing different forecasting models, the Mean Absolute Percent Error (MAPE) metric is used, since it is scale-independent and appropriate for comparing forecast methods on a single series [13]. The forecast errors are summarized in Table 2:

Table 2. Model error

	ARIMA	LSTM	Hybrid model
MAPE	24	15	4
MSE	1076819036	607819036	58030857
MAD	26808	17851	3548

The forecast errors measured by Mean Absolute Percentage Error (MAPE), Mean Squared Error (MSE), and Mean Absolute Deviation (MAD) indicate that the hybrid model performs the best out of the three models - ARIMA, LSTM, and the hybrid model. Specifically, the hybrid model has the lowest MAPE of 4%, compared to 24% for ARIMA and 15% for LSTM. This means that on average, the hybrid model deviates by only 4% from the actual values, while ARIMA and LSTM deviate by 24% and 15% respectively. Similarly, the MSE of 58030857 for the hybrid model is significantly smaller than 1076819036 for ARIMA and 607819036 for LSTM. MSE represents the squared differences between predicted and actual values, so a lower value indicates the hybrid model makes more accurate predictions with smaller errors. Finally, the MAD of 3548 for the hybrid model compares favorably to 26808 for ARIMA and 17851 for LSTM. With the smallest errors across all three metrics - MAPE, MSE, and MAD, the hybrid model produces the most accurate forecasts out of the three models. The superior accuracy demonstrates the hybrid model's ability to leverage the strengths of both ARIMA and LSTM methodologies, while minimizing their individual weaknesses. Given its significantly lower errors, the hybrid model is the optimal choice to generate forecasts for the next time period. Relying on the hybrid model over ARIMA or LSTM alone will provide forecasts with the lowest deviations from actuals.

3. FORECAST

When an appropriate model for the time series has been defined, the next step is to generate predictions for the upcoming periods. For the LSTM model, the rolling forecast technique, wherein new data is incorporated into the model as it becomes available to make updated predictions, is suitable for time series forecasting [14]. The rolling forecast method takes advantage of the LSTM's ability to learn from new data by continually updating the model with the latest observations. As each new actual data point becomes available, it can be fed into the trained LSTM model to make a one-step-ahead forecast. This one-step-ahead forecast is then compared to the actual value to calculate the error. This allows the model to adapt in real-time to changes in the time series. A major advantage of the rolling forecast is that model performance can be monitored on an ongoing basis. If the one-step-ahead errors start increasing, this signals that the model may need adjustment to account for changes in the data dynamics scheme. When using a rolling forecast approach, it is recommended to generate only short-term predictions. Since the new demand data is comprised into the current model and tuned its parameter, its recommended to generate only short-term forecasts to ensure accuracy [15]. While the rolling method allows leveraging new data to keep predictions up-to-date, the accuracy still tends to degrade further into the future. Rather than creating multi-month or annual forecasts, the prediction horizon should be limited to a few weeks or months ahead. For example, generating weekly forecasts for the upcoming month or quarterly forecasts for the next year. This shorter time frame aligns better with the model retraining frequency. By focusing predictions only on the near-term, the rolling LSTM model can make the most accurate and relevant forecasts for strategic planning needs before the next retraining update. Going too far ahead risks decreasing relevance due to reduced precision. Table 3 shows the predicted value for the next 2 months. The forecasted values by the hybrid model can serve as references for the company's strategic planning.

Table 3. The forecasted value for demand in the next 2 months

Period	Forecast	Period	Forecast
75	212714	79	221429
76	223681	80	210314
77	230102	81	207514
78	233174	82	224560

4. CONCLUSION

In the real world, demand forecasting holds an important role in operating a supply chain. It serves as a crucial input for the smooth and efficient functioning of every aspect within a company. Accurate predictions empower businesses to strategically allocate resources, optimize production, and streamline distribution efforts, ensuring a seamless response to customer needs. This paper undertakes an analysis of the company's historical demand, aiming to propose an appropriate model for forecasting future sales volumes. By analyzing three different models—ARIMA, LSTM, and the hybrid ARIMA-LSTM model—it's worth noticing that the hybrid model outperforms the other approaches. The results indicate a notable reduction in forecast error to 4%, offering the company valuable insights to adjust their production strategy in alignment with actual demand. To further enhance the robustness of the demand forecasting, external factors influencing customer purchasing behaviors could be incorporated as additional model inputs. Examples of relevant external factors include economic indicators like employment levels and interest rates, competitive landscape, demographics, seasons and holidays, and disruptive events like natural disasters. By integrating a broader set of demand drivers, the sensitivity of the model can be improved to provide even more accurate forecasts. On the other hand, ongoing model validation should become an integral part of the research's future work. Continual tracking of forecast accuracy paired will enable leveraging the latest sales data and trends. In conclusion, this study demonstrates the tangible benefits of statistical forecasting, yielding a 4% forecast error reduction for the company's demand planning and supply chain processes. The suggested hybrid machine learning methodology and extensions provide a strategic roadmap for interfacing advanced analytics with operational supply chain decisions toward increased revenue, reduced costs, and streamlined customer.

REFERENCES

[1] Gruen, T. W., Corsten, D. S., & Bharadwaj, S., "Retail out-of-stocks: A worldwide examination of extent, causes and consumer responses," Washington, DC: Grocery Manufacturers of America, 2002.

[2] Khashei, M., & Bijari, M., "An artificial neural network (p, d, q) model for timeseries forecasting", *Expert Systems with applications*, 37(1), 479-489, 2010.

[3] Hewamalage, H., Bergmeir, C., & Bandara, K., "Recurrent neural networks for time series forecasting: Current status and future directions", *International Journal of Forecasting*, 37(1), 388-427, 2021.

[4] Seo, S., A review and comparison of methods for detecting outliers in univariate data sets (Doctoral dissertation, University of Pittsburgh), 2006.

[5] Williams, B. M., & Hoel, L. A., "Modeling and forecasting vehicular traffic flow as a seasonal ARIMA process: Theoretical basis and empirical results", *Journal of transportation engineering,* 129(6), 664-672, 2003.

[6] Ediger, V. Ş., & Akar, S., "ARIMA forecasting of primary energy demand by fuel in Turkey", *Energy policy*, 35(3), 1701-1708, 2007.

[7] Abbasimehr, H., Shabani, M., & Yousefi, M., "An optimized model using LSTM network for demand forecasting", *Computers & industrial engineering*, 143, 106435, 2020.

[8] Jiang, Q., Tang, C., Chen, C., Wang, X., & Huang, Q., "Stock price forecast based on LSTM neural network. In Proceedings of the Twelfth International Conference on Management Science and Engineering Management (pp. 393-408)", *Springer International Publishing*, 2019.

[9] Zhang, G. P., "Time series forecasting using a hybrid ARIMA and neural network model", *Neurocomputing*, 50, 159-175, 2003.

[10] Dave, E., Leonardo, A., Jeanice, M., & Hanafiah, N., "Forecasting Indonesia exports using a hybrid model ARIMA-LSTM", *Procedia Computer Science*, 179, 480-487, 2021.

[11] Jain, G., & Mallick, B., "A study of time series models ARIMA and ETS", *Available at SSRN*, 2898968, 2017.

[12] Greff, K., Srivastava, R. K., Koutník, J., Steunebrink, B. R., & Schmidhuber, J., "LSTM: A search space odyssey", *IEEE transactions on neural networks and learning systems*, 28(10), 2222-2232, 2016.

[13] Hyndman, R. J., & Koehler, A. B., "Another look at measures of forecast accuracy", *International journal of forecasting*, 22(4), 679-688, 2006.

[14] Bergmeir, C., Hyndman, R. J., & Koo, B., "A note on the validity of cross-validation for evaluating autoregressive time series prediction", *Computational Statistics & Data Analysis*, 120, 70-83, 2018.

[15] Taieb, S. B., Taylor, J. W., & Hyndman, R. J., "Coherent probabilistic forecasts for hierarchical time series", *In International Conference on Machine Learning,* pp. 3348-3357, PMLR, 2017.

Ứng dụng khai phá dữ liệu trong việc dự báo nhu cầu của một công ty hàng tiêu dùng nhanh Việt Nam

Lê Đức Đạo và Lê Nguyên Khôi

TÓM TẮT

Trong lĩnh vực hàng tiêu dùng nhanh (FMCG), việc có được một dự báo với độ chính xác cao là điều cần thiết. Đặc điểm của lĩnh vực FMCG là việc môi trường có nhiều sự biến động bất thường như sự thay đổi nhanh chóng của thị trường hay sự thay đổi sở thích của người tiêu dùng. Để giải quyết những thách thức này, việc áp dụng các kỹ thuật học sâu, đặc biệt là mạng bộ nhớ ngắn hạn dài (LSTM), đã được sử dụng như một giải pháp quan trọng để nâng cao độ chính xác của dự báo. Bài nghiên cứu này sẽ tập trung vào vai trò quan trọng của dự báo nhu cầu trong ngành hàng FMCG, nhấn mạnh sự cần thiết của các mô hình học sâu dựa trên LSTM để giải quyết sự không chắc chắn về nhu cầu và cải thiện kết quả dự đoán. Thông qua bài nghiên cứu này, chúng tôi mong muốn làm rõ thêm về mối liên hệ giữa dự báo nhu cầu và công nghệ học sâu nâng cao, từ đó cho phép các công ty FMCG có thể phát triển trong bối cảnh thị trường kinh doanh có nhiều sự biến động.

Từ khóa: *dự báo nhu cầu khách hàng, ARIMA, học sâu, mạng bộ nhớ ngắn hạn dài, FMCG*

Received: 24/11/2023
Revised: 11/12/2023
Accepted for publication: 12/12/2023