DOI: https://doi.org/10.59294/HIUJS.VOL.4.2023.380

Forecasting market demand using ARIMA and Holt - Winter method: A case study on canned fruit production company

Le Duc Dao* and Pham Linh Chi
Ho Chi Minh City University of Technology (HCMUT), VNU-HCM, Ho Chi Minh City, Vietnam

ABSTRACT
Consumer demand is an important factor in any business, especially in the food retail industry whose products are perishable and have a short life cycle. The daily demand for a food product is affected by external factors, such as seasonality, price reduction and holidays. To satisfy the stochastic demand, product characteristics vary with customer are required to be timely updated based on market dynamicity. According to previous research, to choose suitable forecasting model is the main concern of enterprises on demand management issue. Proper demand forecasting provides organization with valuable information regarding their prospective in their current market, allowing to make appropriated production portfolio. By applying ARIMA and Holt-Winter, this paper aims to forecast the canned fruit demand at a specific company to help them eliminate waste of lean related to production and distribution. Results are evaluated according to forecasting errors (MAD, MSE, MAPE). By comparing the aforementioned methods, it can be concluded that ARIMA outperforms Holt-Winter related to prediction accuracy.

Keywords: *Demand forecast, ARIMA, Holt-Winter model, food industry, canned fruit*

1. INTRODUCTION

Demand forecasting is an imperative force to drive the business operation and management efficiency. The volatility of demand directly affects the predictability [1]. Stable demand eases the task with the implementation of basic forecast methods while randomly varying one challenges the business from reaching the ideal accuracy. To be specific, the demand variance in the fresh food market results from short period factors (such as holiday, weather, promotion and sales attempts, and so on) [2]. Hence, the daily forecast has been seen as the most suitable periodic method to minimize the risk of product redundancy or shortage [3]. Considering the characteristics of the data in the study, ARIMA and Holt-Winter model were selected as potential candidates for the research.

2. CASE STUDY

In the article, research subject is, on point, a canned fruit production company. Customer demand of this company is conducted from actual data of company during 6 months. The fruit for producing final products is provided by a supplier who always meets the quantity and delivery time as negotiated. The quantity of fruit ordered is decided based on manager's experience. Consequently, the company has encountered some problems such as not providing enough order quantity. Proceeding to examine the quantity of delayed orders during the timeframe from 10/2022 to 03/2023, as indicated in Table 1, it is shown that the rate of overdue shipments within a week fluctuates between 8% and 23%. However, the company's set limit for acceptable delayed orders is only below 5%. The system has recorded some typical reasons contributing to the occurrence of delayed orders, such as customer appointment delays, out-of-stock items, customers being unreachable on the delivery day, and more. Table 2 presents the respective percentages of these reasons among the total number of delayed orders. Notably, the issue of inventory shortage for customer deliveries accounts for the largest percentage (63.41%) among the overall delayed orders.

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

Table 1. Number of late orders and total orders

Month	Week	Late orders	Total orders	Rate %
Oct-2022	1	292	1369	21.4%
	2	137	1321	10.4%
	3	227	1362	16.7%
	4	283	1553	18.2%
Nov-2022	1	291	1249	23.3%
	2	203	1349	15.1%
	3	189	1286	14.7%
	4	382	2006	19.0%
Dec-2022	1	160	781	20.5%
	2	253	1339	18.9%
	3	211	1371	15.4%
	4	338	2587	13.1%
Jan-2023	1	135	1621	8.3%
	2	203	1399	14.5%
	3	200	1446	13.8%
	4	235	1727	13.6%
Feb-2023	1	175	1092	16.1%
	2	244	1671	14.6%
	3	266	1484	17.9%
	4	257	2133	12.0%
Mar-2023	1	157	1167	13.5%
	2	130	1615	8.1%
	3	150	1668	9.0%
	4	329	2862	11.5%

Table 2. Reasons for delayed orders

Reasons	Orders	% delayed orders
Appointment delays	1142	21%
Out of stock	3455	63%
Customer unavailability	615	11%
Others	236	4%
Total	**5448**	

Therefore, the decision-making process regarding ordering from suppliers is crucial. The quantity of orders placed will directly impact the order fulfillment rate, the level of service, and the company's profitability. Since, this research focuses on analyzing the old demand data with the aim of developing a forecast for future demand and also providing the most appropriated order quantity.

2.1. Data processing

First of all, Data has been collected over 6 months period (182 data points), from Oct-2022 to Mar-2023 (Table 3). The data set contains information about the product's demand by day. After that, data characteristic has been manually pre-processed before applying Minitab and Python to run predictive model.

Table 3. Customer demand from Oct-2022 to March-2023

t	Y_t	t	Y_t	t	Y_t	t	Y_t	t	Y_t	t	Y_t
01-Oct	149	01-Nov	215	01-Dec	192	01-Jan	202	01-Feb	216	01-Mar	245
02-Oct	136	02-Nov	206	02-Dec	205	02-Jan	210	02-Feb	214	02-Mar	230
03-Oct	142	03-Nov	199	03-Dec	195	03-Jan	210	03-Feb	208	03-Mar	225
04-Oct	145	04-Nov	199	04-Dec	189	04-Jan	204	04-Feb	235	04-Mar	230
05-Oct	150	05-Nov	243	05-Dec	179	05-Jan	196	05-Feb	219	05-Mar	237
06-Oct	148	06-Nov	187	06-Dec	176	06-Jan	184	06-Feb	234	06-Mar	239
07-Oct	184	07-Nov	194	07-Dec	184	07-Jan	210	07-Feb	232	07-Mar	215
08-Oct	155	08-Nov	200	08-Dec	184	08-Jan	205	08-Feb	246	08-Mar	242
09-Oct	160	09-Nov	196	09-Dec	214	09-Jan	188	09-Feb	241	09-Mar	232
10-Oct	167	10-Nov	192	10-Dec	200	10-Jan	188	10-Feb	234	10-Mar	227
11-Oct	180	11-Nov	177	11-Dec	202	11-Jan	193	11-Feb	253	11-Mar	232
12-Oct	189	12-Nov	195	12-Dec	198	12-Jan	189	12-Feb	231	12-Mar	228
13-Oct	195	13-Nov	195	13-Dec	204	13-Jan	199	13-Feb	229	13-Mar	235
14-Oct	226	14-Nov	182	14-Dec	189	14-Jan	234	14-Feb	222	14-Mar	232
15-Oct	175	15-Nov	173	15-Dec	174	15-Jan	208	15-Feb	206	15-Mar	245
16-Oct	189	16-Nov	185	16-Dec	214	16-Jan	210	16-Feb	196	16-Mar	237
17-Oct	185	17-Nov	174	17-Dec	199	17-Jan	204	17-Feb	178	17-Mar	233
18-Oct	196	18-Nov	182	18-Dec	193	18-Jan	183	18-Feb	215	18-Mar	241
19-Oct	181	19-Nov	201	19-Dec	187	19-Jan	198	19-Feb	238	19-Mar	245
20-Oct	193	20-Nov	189	20-Dec	176	20-Jan	204	20-Feb	231	20-Mar	241
21-Oct	243	21-Nov	194	21-Dec	185	21-Jan	243	21-Feb	242	21-Mar	223
22-Oct	180	22-Nov	203	22-Dec	188	22-Jan	204	22-Feb	235	22-Mar	256
23-Oct	184	23-Nov	202	23-Dec	242	23-Jan	198	23-Feb	240	23-Mar	237
24-Oct	171	24-Nov	208	24-Dec	202	24-Jan	202	24-Feb	225	24-Mar	231
25-Oct	167	25-Nov	210	25-Dec	209	25-Jan	210	25-Feb	250	25-Mar	238
26-Oct	191	26-Nov	237	26-Dec	190	26-Jan	225	26-Feb	245	26-Mar	242
27-Oct	194	27-Nov	195	27-Dec	192	27-Jan	224	27-Feb	230	27-Mar	241

28-Oct	223	28-Nov	193	28-Dec	200	28-Jan	243	28-Feb	235	28-Mar	237
29-Oct	184	29-Nov	174	29-Dec	188	29-Jan	213			29-Mar	239
30-Oct	208	30-Nov	190	30-Dec	210	30-Jan	202			30-Mar	242
31-Oct	215			31-Dec	218	31-Jan	208			31-Mar	235

*-**t:** Time period*

-Y_t: Demand

• Evaluation of data characteristics

Results of the descriptive analysis of the dataset includes: The interquartile range (187; 227), Min = 64, Max = 259 and seven outliers (as known as the points out of interquartile range). These abnormal values occurred during holiday and flash sales and led to the significant volatility in demand. Thus, it is necessary to remove these outliers by replacing them with the median value of 199 [4 - 5]. Then, the clean time series and time series plot are shown in Figure 1 below:

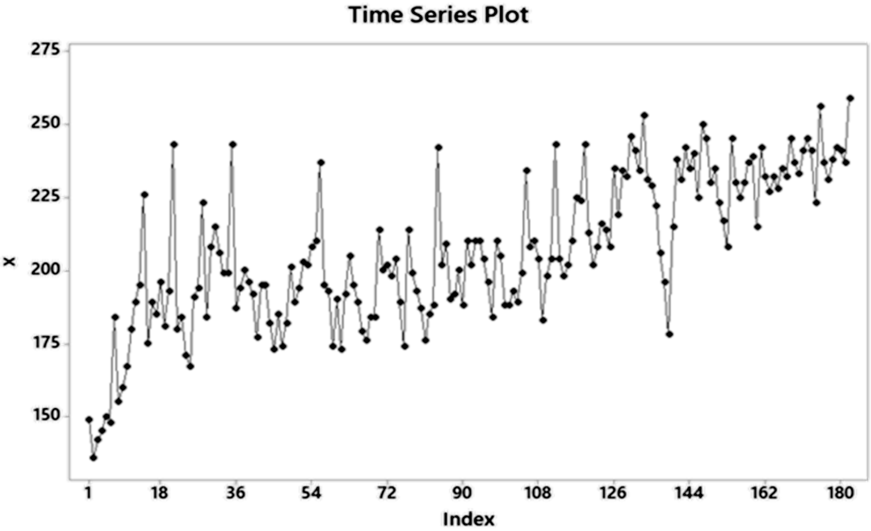

Figure 1. Time series plot of data after removing outliers

Based on the autocorrelation shown in figure 2, the first lags have high correlation and are significantly non-zero, and as the lag increases, ACF value decreases to 0. Therefore, it is possible to confirm trend data. At the same time, the Autocorrelation coefficient occurs at the number of season lengths and their multiples. Therefore, the data is seasonal with a season length of 7.

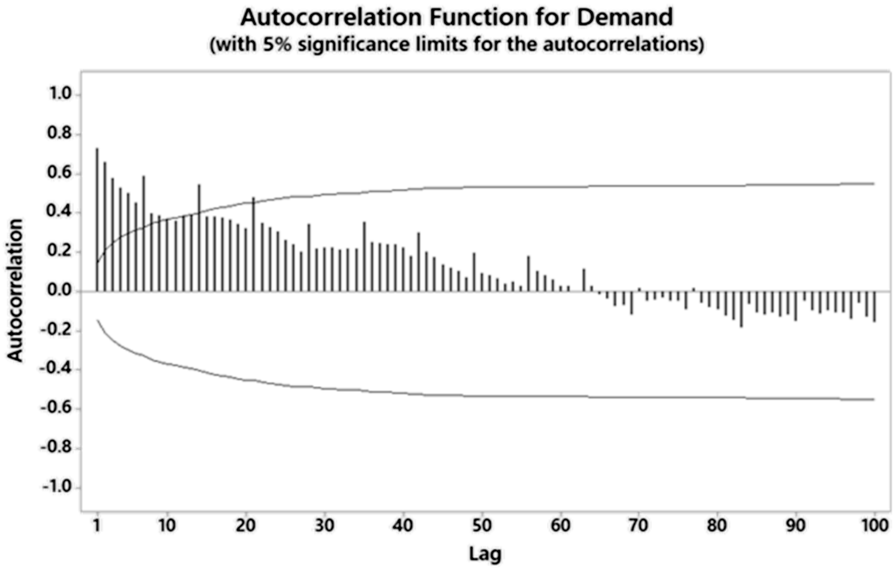

Figure 2. ACF of the data after removing outliers

2.2. Discussing

After defining the trend of the data, several methods have been proved the efficiency to run the season model such as Holt Winte's and Arima [5, 6, 8]. For example**,** Jamah Fattah et al [5]. have developed ARIMA method to predict demand for a food company by the Box-Jenkins time series approach. The positive outcome has proved the efficiency of the method as it can be applied to build a production plan [6]. In 2022, Surindar Gopalrao Wawale et al. [6] researched the usefulness of the agriculture products price prediction model compared to rice price in India using the ARIMA method. After that, the best model is selected based on the forecast error [7]. Ünal et al. [8] conducted demand forecast in the food processing industry by comparing Holt-Winter model, trend analysis and decomposition. In the study, Holt-Winter and decomposition models obtained better results regarding the forecast error. By referring these previous research, Holt Winter's and Arima are chosen as the predictive method in our model.

2.2.1. Holt Winter's Method

Step 1: Determine the alpha, beta, and gamma

Alpha, beta and gamma are smoothing parameters for the model. Alpha helps to remove the effect of random factors in time series, beta represents the change in the underlying level that we expect to occur between current and next month, gamma shows the tendency of time-series data to exhibit behavior that repeats itself every periods [9]. They are sometimes determined by using Excel solver with min MSE error [9]. In this study, the result of parameters is shown in Table 4.

Table 4. Optimal result of three smoothing numbers

Alpha	Beta	Gamma	MSE
0.70598003	0.0416291	0.82221199	152.512167

Step 2: Forecast on Minitab

Result of forecasting with additive and multiplier models on Minitab is displayed as below:

- **Multiplicative method:** The results of the model represent smoothing constants and accuracy measure (MAPE: 4.926, MAD: 9.937, MSD: 146.788), along with ACF plot as Figure 3 below:

Autocorrelations

Lag	ACF	T	LBQ
1	38.133	0.51	0.27
2	-56.874	-0.77	0.87
3	-207.723	-2.79	8.94
4	-122.373	-1.58	11.76
5	-22.283	-0.28	11.85
6	-48.924	-0.62	12.31
7	-60.788	-0.77	13.02
8	76.697	0.97	14.15
9	4.861	0.06	14.15
10	-42.141	-0.53	14.50

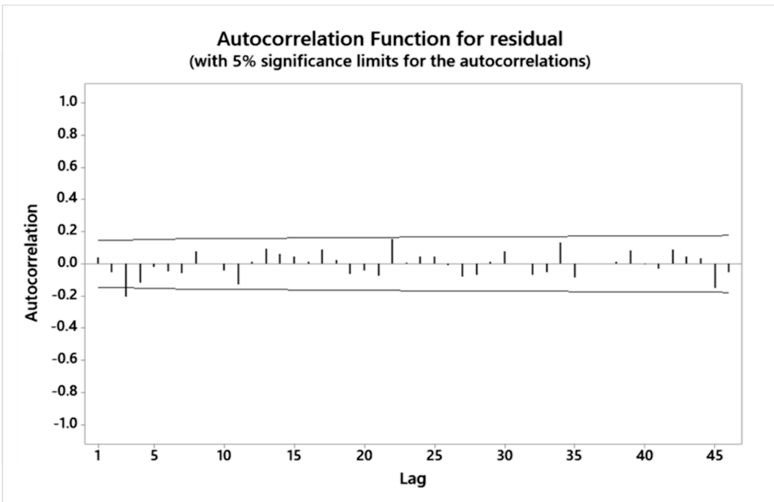

Figure 3. ACF plot of Winter (x) method

The ACF shows no significant autocorrelation coefficients other than zero and the LBQ value of the first 4 lags is larger than the Chi-square table value (11.76 > 9.488). Therefore, it can be concluded that the error is not random.

Applying similar approach to the additive model, the results are shown as below:

- **Additive model:** The results of the model represent smoothing constants and accuracy measure (MAPE: 4.911, MAD: 9.881, MSD: 149.515), along with ACF plot as Figure 4 below:

Autocorrelations

Lag	ACF	T	LBQ
1	16.146	0.22	0.05
2	-52.280	-0.71	0.56
3	-185.951	-2.50	07.03
4	-108.803	-1.42	9.25
5	-22.491	-0.29	9.35
6	-67.727	-0.87	10.22
7	-65.950	-0.84	11.05
8	86.179	1.10	12.48
9	4.557	0.06	12.49
10	-30.761	-0.39	12.67

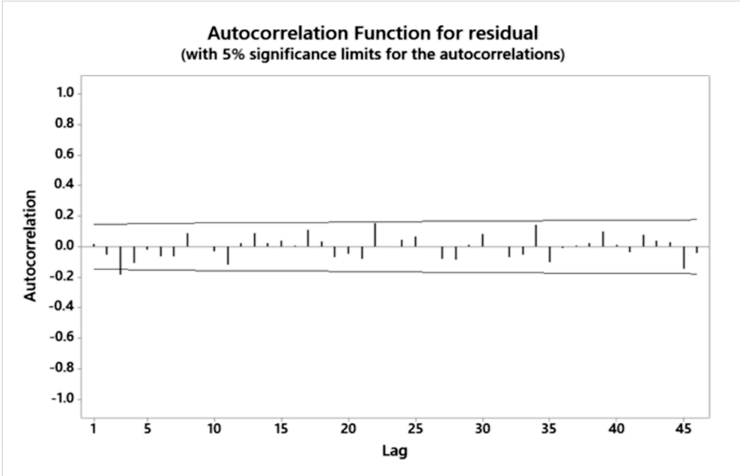

Figure 4. ACF plot of Winter (+) method

The ACF shows no significant autocorrelation coefficients other than zero and the LBQ value of the first 10 lags is larger than the Chi-square table lookup value (12.67 < 18.3). Therefore, it can be concluded that the error is random.

2.2.2. The ARIMA method

Step 1: Determine the parameters p, d, q, P, D, Q in the model using Python.

Three core parameters of the ARIMA model are p, d and q, which symbolize the number of autoregressive terms (AR), the number of nonseasonal differences, and the number of lagged forecast errors (MA) in the prediction equation, respectively. On the other hand, the SARIMA(p,d,q) (P,D,Q)[s] model considers the seasonality according to three aforementioned indexes [10]. By identifying the optimal parameters and selecting a model based on the information criteria (AIC), the model for the study is determined as ARIMA [(1,0,0), (0,1,1)] [7].

Step 2: Forecast in Minitab 18 and record the results.

The ACF histogram (Figure 5) shows that there is no significant non-zero autocorrelation coefficient and the LBQ value of the first 10 lags is larger than the Chi-square table lookup value (3.47 < 18.3). Therefore, it can be concluded that the random of errors.

Autocorrelations

Lag	ACF	T	LBQ
1	-0.062663	-0.83	0.70
2	0.057980	0.76	1.30
3	-0.076437	-1.00	2.35
4	-0.019738	-0.26	2.42
5	0.056718	0.74	3.01
6	-0.014422	-0.19	3.05
7	0.002043	0.03	3.05
8	-0.034513	-0.45	3.27
9	-0.030597	-0.40	3.44
10	-0.010840	-0.14	3.47

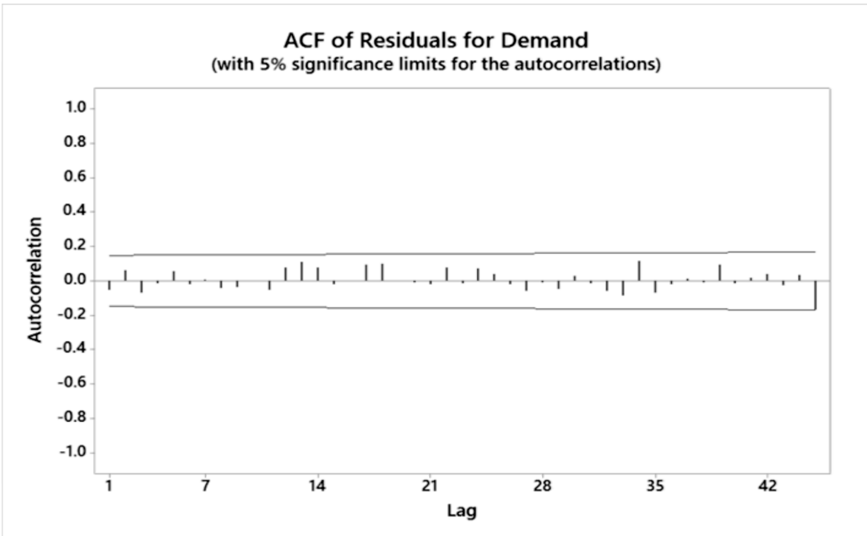

Figure 5. ACF plot of ARIMA method

2.3. Error comparison

The forecast errors of the methods are summarized in Table 5.

The forecasting methods give nearly the same error. However, ARIMA model has smaller values (MAD, MSE and MAPE), compared to the other methods. Therefore, ARIMA model method is chosen to forecast for the next period.

Table 5. Prediction error of models

	Winters (MH plus)	Winters (MH multiply)	ARIMA
MAD	9.937	9.886	8.502
MSE	146.788	145.565	115.657
MAPE	4.926	4.911	4.17

3. FORECAST

After defining the most appropriate model, Forecasting is implemented by Minitab program. Table 5 presents the result of the customer demand in the next 1 month with the ARIMA model (1,0,0) (0,1,1) [7]. We can clearly see that the chosen model is appropriated for forecasting. Model operation will forecast unobserved quantity by generate the demand trend according to historical data. The quality of trend is evaluated by risk function (MSE, MAD. MAPE). The projections generated through modeling recommend this food company for decision-making about production. Once we obtain a demand forecast, it will be much simpler and clearer to decide on the best manufacturing strategy, eliminating significant cost overruns.

Table 6. Result of the customer demand in the next 1 month with the ARIMA model (1,0,0) (0,1,1) [7]

Period	Forecast	Period	Forecast	Period	Forecast
01-Apr	230	14-Apr	259	27-Apr	240
02-Apr	230	15-Apr	245	28-Apr	265
03-Apr	235	16-Apr	242	29-Apr	250
04-Apr	237	17-Apr	245	30-Apr	247
05-Apr	238	18-Apr	245		
06-Apr	230	19-Apr	245		
07-Apr	256	20-Apr	237		
08-Apr	242	21-Apr	262		
09-Apr	239	22-Apr	248		
10-Apr	242	23-Apr	244		
11-Apr	242	24-Apr	247		
12-Apr	242	25-Apr	247		
13-Apr	234	26-Apr	248		

4. CONCLUSION

Demand forecasting is an important aspect of managing supply chain. Because it is the input for every other function of the company to operate effectively and efficiently. This study aims to compare two forecasting methods ARIMA and Holt-Winters based on error analysis and comparison of error values (MAPE, MSE, and MAD). It can be concluded that ARIMA model, which is applied in the time series analyzed in the daily study, has been effective in the demand forecasting process of the company. The results obtained proves that the application of this model has helped business to improve the delay rate within the allowable range (0%-5%); these results will support company the decisions making for producing product. Future research can be extended in some field, we can generate new models by combining qualitative and quantitative methods in order to provide accurate forecasts and raise forecast reliability. Furthermore, it is necessary to monitor and analyze changes and impacts by external factors such as trend, weather, etc.

REFERENCES

[1] Gilliland, M., & Sglavo, U., *Worst practices in business forecasting*, 2010.

[2] Armstrong, J. Scott. "Selecting forecasting methods" *Principles of forecasting: A handbook for researchers and practitioners*, 365-386, 2001.

[3] Dekker, M., Van Donselaar, K., & Ouwehand, P. "How to use aggregation and combined forecasting to improve seasonal demand forecasts," *International Journal of Production Economics*, 90(2), 151-167, 2004.

[4] Liu, H., Shah, S., & Jiang, W., "On-line outlier

detection and data cleaning," *Computers & chemical engineering*, *28*(9), 1635-1647, 2004.

[5] FATTAH, Jamal, et al, "Forecasting of demand using ARIMA model," *International Journal of Engineering Business Management*, 2018, 10: 1847979018808673.

[6] WAWALE, Surindar Gopalrao, et al, "Minimizing the error gap in smart framing by forecasting production and demand using ARIMA model," *Journal of Food Quality*, 2022.

[7] Tirkes, G., Guray, C., & Celebi, N. E., "Demand forecasting: a comparison between the Holt Winters, trend analysis and decomposition models," *Tehnicki Vjesnik-Technical Gazette*, 24, 2017

[8] Ünal, Y. Z., Al-Turki, U. M., Zaim, S., Demirel, Ö. F., & Görener, R., "Developing spreadsheet models of Holt-Winter methods and solving with Microsoft Excel solver and differential evaluation technique: An application to tourism sector," In *2015 International Conference on Industrial Engineering and Operations Management (IEOM),* IEEE, 2015.

[9] ArunKumar, K. E., Kalaga, D. V., Kumar, C. M. S., Chilkoor, G., Kawaji, M., & Brenza, T. M., "Forecasting the dynamics of cumulative COVID-19 cases (confirmed, recovered and deaths) for top-16 countries using statistical machine learning models: Auto-Regressive Integrated Moving Average (ARIMA) and Seasonal Auto-Regressive Integrated Moving Average (SARIMA)," *Applied soft computing*, *103*, 107161, 2021.

[10] Wang, M., Zhang, Q., Tai, C., Li, J., Yang, Z., Shen, K., & Guo, C., "Design of PM2. 5 monitoring and forecasting system for opencast coal mine road based on internet of things and ARIMA Mode," *Plos one*, *17*(5), e0267440, 2022.

[11] CONTRERAS, Javier, et al, "ARIMA models to predict next-day electricity prices," *IEEE transactions on power systems,* 18.3: 1014-1020, 2003.

[12] MATSUMOTO, Mitsutaka; IKEDA, Akira, "Examination of demand forecasting by time series analysis for auto parts remanufacturing," *Journal of Remanufacturing*, 5: 1-20, 2015.

[13] TIRKES, Guzin; GURAY, Cenk; CELEBI, Nes'e, "Demand forecasting: a comparison between the Holt-Winters, trend analysis and decomposition models," *Tehnicki Vjesnik-Technical Gazette*, 24, 2017.

[14] KUMAR, T. L. M., et al, "Holt-winters exponential smoothing and sesonal ARIMA time-series technique for forecasting of onion price in Bangalore market," *Mysore Journal of Agricultural Sciences*, 45.3: 602-607, 2011.

Dự báo nhu cầu khách hàng cho một công ty sản xuất trái cây đóng hộp sử dụng phương pháp ARIMA và Holt-Winter

Lê Đức Đạo và Phạm Linh Chi

TÓM TẮT

Nhu cầu khách hàng là một trong những yếu tố cực kỳ quan trọng trong một doanh nghiệp, đặc biệt là trong ngành thực phẩm. Các sản phẩm trong ngành này chủ yếu là sản phẩm có hạn sử dụng ngắn, dễ hư hỏng và sẽ bị loại bỏ ngay nếu không đạt chất lượng. Như vậy, việc cần nắm bắt được nhu cầu khách hàng đúng số lượng, đúng thời điểm là rất cần thiết. Việc dự báo nhu cầu sẽ hữu ích để giảm thiểu lãng phí liên quan đến các vấn đề như đáp ứng đơn hàng, sản xuất sản phẩm. Dựa trên các nghiên cứu trước đây, tác giải đã thử nghiệm áp dụng mô hình ARIMA và Holt-Winter để dự báo dòng sản phẩm trái cây đóng hộp cho một công ty thực phẩm. Kết quả được đánh giá dựa trên các sai số dự báo như MAD, MSE, MAPE. Sau khi thực hiện dự báo, kết quả cho thấy mô hình ARIMA là phù hợp nhất với nhu cầu khách hàng ở thời điểm hiện tại.

Từ khóa: *Dự báo nhu cầu khách hàng, ARIMA, Holt-Winter, công nghiệp thực phẩm, trái cây đóng hộp*

Received: 16/05/2023
Revised: 09/06/2023
Accepted for publication: 10/06/2023