5th South American Conference on Industrial Engineering and Operations Management
Bogota, Colombia, May 7-9, 2024
Publisher: IEOM Society International, USA DOI: 10.46254/SA05.20240077
Published: May 9, 2024

Improvement Proposal Based on Machine Learning, Big Data and DDMRP to Improve Forecasting Compliance in a Consumer Goods Company

Diana Micaela MARISCAL CRUZADO and Bruno Piero GONZÁLEZ JÁUREGUI
Carrera de Ingeniería Industrial
Universidad de Lima
Lima, Peru
20191214@aloe.ulima.edu.pe, 20162088@aloe.ulima.edu.pe

José Antonio TAQUÍA GUTIÉRREZ
Instituto de Investigación Científica
Universidad de Lima
Lima, Peru
jtaquia@ulima.edu.pe

Abstract

Nowadays, is crucial to accurately forecast products, especially for a company that imports its goods. Having an accurate forecasting enables the company to optimize resource management, increasing productivity and preventing overselling or underselling of products. Additionally, establishing a demand-based material planning model is essential to ensure that our suppliers meet their service level commitments. In this research project, Machine Learning and Big Data are employed to enhance de forecasting methods of consumer goods company. The data collected from the company's sales over the last four years for "the hair category" has been trained and the Arima method will be employed to predict the first 8 months of the year 2023. Furthermore, the Demand Driven Material Requirement Plan (DDMRP) is implemented to improve suppliers' service level. The impact of the proposed model will be evaluated using indicators such as Forecast Bias (FB), Forecast Accuracy (FA), Mean Absolute Percentage Error (MAPE) and Service Level Agreement (SLA).

Keywords
Demand Forecast, DDMRP, Forecast Accuracy, Forecast Bias, Service Level Agreement

1. Introduction
In most cases, forecast research assumes that forecasts are an end in themselves, without considering the subsequent stages of calculation needed to transform forecasts into replenishment decisions and decisions that support all the processes encompassing the supply chain. (Goltsos et al, 2022) The planner of the company was consulted, and this issue was not foreign to the company. His explanation will be used for the case study. The real name of the company will be omitted due to confidentiality issue and the present research paper will study the most dominant category of the company, the one that provided the highest revenue in the year 2022. In a consumer goods company (CGC) having an excessive quantity of products due to underselling can result in product losses, while inventory shortages from overselling can lead to a low fill rate for their customers. Either of them, due to excess or lack of inventories, represents an opportunity cost for the company. As we continue to gather data with the assistance of the planner, four general causes for the imprecision of demand forecasting have emerged. The first is inadequate data management, followed by improper handling of inbound logistics, external factors, and lastly, the influence of competition. Furthermore, it

is worth emphasizing that the company heavily relies on its suppliers, referred to as 'Source Units (SU),' which are plants within the same company but located in other countries in Latin America. Once again, the importance of demand forecasting is highlighted, as importing products introduces lead time considerations that can impact product availability. Therefore, precise demand forecasting and a commitment to the service level of the SU are crucial.

1.1 Objectives

The objective of this work is to enhance the accuracy of monthly demand forecasting by implementing big data and machine learning. With this improvement, we aim to construct a Demand-Driven Material Requirements Planning (DDMRP) model to enhance the Service Level Agreement (SLA) of the Source Units (SU).

2. Literature Review

In the present research project had the need to find agile and precise methods to develop a demand forecasting in the hair category of a CGC. Ifraz et al. (2023) asserts that improvement management would amount to demand forecasting carried out using methods based on regression (multivariate linear regression, multivariate nonlinear regression, Gaussian process regression, additive regression, discretization regression, support vector regression), rule-based methods, tree-based methods (random forest, M5P, Random tree, REPTree), and artificial neural networks. However, Thivakaran and Ramesh (2022), in their research article, indicate that the proposed supervised artificial neural network algorithm produces reliable results compared to other learning methods. Nevertheless, Pavlyshenko (2019) conductes a diagnosis showing that combining different forecasts methods produced by different algorithms can improve forecast accuracy. In the study, various conditions were examined.

On the other hand, Duhem et al. (2023) explained that the complexity of production planning is greatly emphasized, compounded by the low tolerance level of the customer; hence the importance of using the DDMRP tool. Lahrichi and Damand (2023) further adds that DDMRP places buffer stocks that are used when the flow falls below a certain level. Additionally, Martin and Lauras (2023) highlight the importance of correctly utilizing buffer stocks to maintain a high level of service and adapt to the constant market changes.

The company studied in this paper entered Peru in 1960 through another established company; however, they controlled 51% of the shares. The company is dedicated to the manufacturing, marketing, and distribution of food and consumer goods for personal care such as margarines, oils, hydrating beverages, shampoos, and toiletry items. (EMIS, n.d.). This research paper will focus on the hair care category, which is the most important category of the company with an annual turnover of $ 30,952,521.50 by 2022. Nowadays, in terms of company shares, the company is the second most important in Peru.

As the other categories within the company, the hair category used the same method of forecating. We found that the forecast was off by 12.88% during 2022. Additionally, there was an over-forecast of 10.3% when the company's accepted margin is 4%. Finally, the service level of the SU was on average 72%, with a gap of 18% being the company's established goal of 90% (Company´s Planner, 2023). Additionally, the abundance of information sources contributes to the complexity. Moreover, relying on human intervention rather than automated processes for data processing, cleaning, and management increases the likelihood of errors.

3. Methods

Throughout the study, various methods have been employed to achieve the results. To determine the most crucial category, an ABC analysis was used. Subsequently, a sampling process was conducted to identify the most important products lines for further focus.

As for the experimentation, a mixed-methods approach was employed, combining both quantitative and qualitative methods to leverage the strengths of each while minimizing their weaknesses. In the quantitative aspect of the study, understanding the application of Machine Learning (ML) in improving the accuracy of demand forecasting and DDMRP is essential. Furthermore, a univariate model has been chosen, where only sales will be considered as the variable to estimate. Other potential factors that may affect monthly demand will not be taken into account in this model.

4. Data Collection

4.1 Plan

The problem of inaccuracy in demand forecasting was quantitatively identified. For the year 2022 in the hair category, a Mean Absolute Percentage Error (MAPE) of 12.88% was found. Also, the forecast bias was -10.3%, while the company's policy accepts a deviation of ± 2%. Additionally, there was a forecast accuracy of 88.52%, below the acceptable threshold of 90% and a SLA of 72% of the SU.

On the other hand, following the interviews with those in charge of the areas involved in demand forecasting, the general causes of deviation in demand forecasting were identified. In this way, both general and specific causes were found and are represented in the Pareto Diagram in Figure 1.

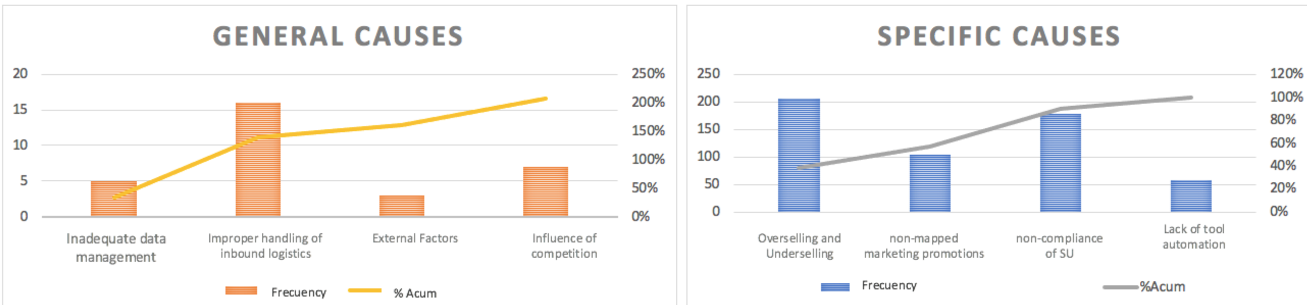

Figure 1. Pareto Diagram

4.2 Development

In this phase, the aim was to implement a simulation using Python for demand forecasting and create a model using the Arena program for DDMRP. Afterward, we evaluated the KPIs' results to confirm the performance of our model.

4.2.1 Techniques

The information gathered for the forecasting salutation was introduced to python as structured data containing historical sales of the last four year for the chosen product line. The simulation will use the modular code method which uses packages already created. We will be using the libraries of panda, numpy, seaborn and matplotlib. pyplot and "cobbled together like building blocks to create a larger application." (Sturtz,2018) To perform a statistical analysis of the data, the following methods were used:

Augmented Dickey-Fuller Test (ADF): The test provides a statis test and a p-value. Under the null hypothesis, the test statistic does not follow a standard distribution, and percentiles tables are used (Fuller, 1976). If the analysis uses standard significance levels such as 0.01, 0.05, or 0.1, the test statistic is compared to the corresponding critical value. If the test statistic is less than or equal to the critical value, the null hypothesis is rejected. (Minitab, 2021)

- Null Hypothesis: The data is not stationary.
- Alternative Hypothesis: The data is stationary.

Ljung-Box: This test checks for autocorrelation in a time series. Ideally, the null hypothesis should be rejected. This means that the p-value of the test is greater than 0.05, indicating that the model residuals are independent. (Stalogos, 2021)

- H0: The residuals are distributed independently.
- HA: The residuals are not distributed independently; they exhibit serial correlation.

The performance evaluation of results for demand forecasting will be carried out using Forecast Bias (FB), Forecast Accuracy (FA) and Mean Absolute Percentage Error (MAPE). According to the Company's Planner (2023), these are the indicators the company employs

- Forecast Accuracy: How accurate the forecast is. (Institute of Business Forecasting and Planning, 2021)

$$FA = 1 - \frac{ABS(Venta - Forecast)}{Venta}$$

- Forecast Bias: Measures the tendency to over or under forecast. If the value is negative, a company tends to over forecast, and if the value is positive, it tends to under forecast. (Institute of Business Forecasting and Planning, 2021)

$$FB = \frac{Ventas}{Forecast} - 1$$

- Mean Absolute Percentage Error:

$$MAPE = \frac{\sum \%\, de\ error\ abs}{periodo}$$

To perform the simulation of DDMRP, the company's planner was consulted regarding the material importation process in order to create a demand-driven planning model. Crucial data considered for this model includes the coverage policy period and lead time.

The company identifies the SLA with the name International Dispatch Rate (IDR), measuring the level of service from the Source Units. It is the multiplication of Conformance to Production (CTP), which indicates the percentage of the order that the Source Units can supply, and Conformance to Dispatch (CTD), which indicates the percentage of fulfillment by the Source Unit based on the committed dispatch percentage. (Personal interview with the Planner, 2023).

$$IDR = CTP \times CTD$$

5. Results and Discussion

5.1 Numerical Results

The statistical analysis of the data was carried, for this case, the ADF test tells us that the p-value = 0.169229, which means that it is very likely that the data are non-stationary and that the null hypothesis is accepted. According to the Ljung-Box test, the value of 1.60 along with the probability value (Prob(Q)) of 0.21 suggest that there is no significant evidence of autocorrelation in the model residuals. These results are shown in Figure 2

```
Resultados de la Pruena de Dickey Fuller:
Test Statistic                 -2.308537
p-value                         0.169229
#Lags Used                      2.000000
Number of Observations Used    45.000000
Critical Value (1%)            -3.584829
Critical Value (5%)            -2.928299
Critical Value (10%)           -2.602344
dtype: float64
```

```
==============================================================================
Dep. Variable:                        y   No. Observations:                  48
Model:              SARIMAX(0, 1, 0, 12)  Log Likelihood                -114.626
Date:                   Thu, 28 Sep 2023  AIC                            233.253
Time:                           19:08:35  BIC                            236.420
Sample:                       01-31-2019  HQIC                           234.358
                            - 12-31-2022
Covariance Type:                     opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
intercept     -2.5832      0.976     -2.646      0.008      -4.497      -0.670
sigma2        34.1327     10.641      3.208      0.001      13.276      54.990
===================================================================================
Ljung-Box (L1) (Q):                   1.60   Jarque-Bera (JB):                 1.12
Prob(Q):                              0.21   Prob(JB):                         0.57
Heteroskedasticity (H):               1.00   Skew:                             0.08
Prob(H) (two-sided):                  1.00   Kurtosis:                         2.15
===================================================================================
```

Figure 2. Satistic results

After conducting the simulation using Python, the following demand forecast was found for the first 8 months of 2023 shown in Figure 2 and Table 1.

Table 1. Demand Forecast

Month	Forecast (tons)
Enero	14.37

Febrero	20.77
Marzo	22.38
Abril	12.80
Mayo	10.11
Junio	16.31
Julio	13.77
Agosto	12.77

Using FB,FA and MAPE the following results were obtained with the average results for the studied product line in each month up to August, as shown in Table 2.

Table 2. FB, FA and MAPE results

KPI	Result
FB	1.90%
FA	80.3%
MAPE	20.43%

Taking into account the previously found demand, we proceeded to determine the SLA of the Source Units (SU) with the assistance of the simulation conducted in Arena. We used five different models to determine which one provided the best result for the SLA. Model number four yielded the best result, as seen in the table below Table 3.

Table 3. DDMRP results

Model	Coverage Period Policy (months)	Lead Time (días)	SLA
1	2	UNIF (33,40)	80.30%
2	4	UNIF (33,40)	67.50%
3	3	UNIF (19,20)	76.60%
4	2	UNIF (19,20)	82.60%
5	4	UNIF (19,20)	68.60%

5.2 Graphical Results

The statistical results of the data will be presented graphically below Figure 3.

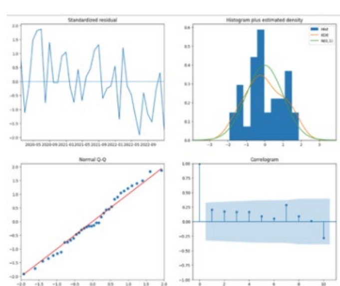

Figure 3. Graphic satistic results.

To determine the best model, Input Analyzer, which is part of the Arena program where the simulation was conducted, was utilized. Figure 4 shows the SLA result obtained from the best model.

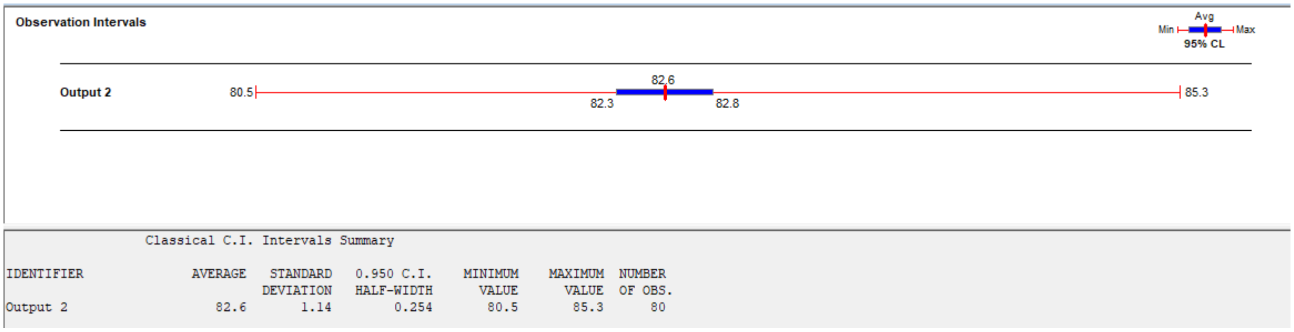

Figure 4. Input Analyzer results

5.3 Proposed Improvements

Drawing on the improvements observed in other papers and the companies planner input, we set objectives for our investigation to achieve, as indicated in Table 4. It is important to mention that these figures were measured at the aggregate level of the entire hair category. However, our research specifically targets one product line. Therefore, we will use the improvement percentage as the objective to achieve.

Table 4. Objectives

Tools	KPI´s	AS IS	TO BE	% of improvement	Authors
Arima	FB	-10.30%	(+-)2%	81.00%	Companie´s planner, 2023
	MAPE	12.88%	7%	46%	Companie´s planner, 2023
	FA	88.52%	94.20%	4.00%	Ifraz Et al, 2023
DDMRP	SLA	72%	83%	15.00%	Duhem L. Et al, 2023

5.4 Validation

Taking into account the results of the studied product line and the expected improvement percentage, the values of the current model used by the company and the proposed model in this paper are presented. Achieving the objectives and showing improvements in all KPIs, as shown in Table 5.

Table 5. Results

KPI	Actual	Model	% of improvement	The objective was achieve
FB	-22%	1.90%	100%	Yes
FA	61%	80.30%	32%	Yes
MAPE	38.96%	20.43%	48%	Yes
SLA	72.00%	82.60%	15%	Yes

6.Conclusion

During the investigation, it was possible to identify the main causes of forecast deviation for the hair category in the company, with inadequate inbound logistics management being the primary cause. Using the model, better results w ere obtained in all studied KPIs, including forecast bias (FB), forecast accuracy (FA), and mean absolute percentage error (MAPE). However, according to the literature studied, with additional data, it could be possible to combine different techniques to achieve an even better result than the one already obtained. Therefore, it is recommended to the company to begin storing marketing or promotion data to understand consumer behavior much better. Finally, model number four yields the best result for increasing the SU's SLA, thus recommending a lead time of UNIF (19.20) days and a Coverage Period Policy of two months. The utilization of machine learning, big data, and DDMRP techniques can provide exceptional results to the company by creating a more precise inbound logistics, better demand forecasting, and improving the SLA, thus addressing the demand deviation issue.

References

Youssef Lahrichi; David Damand. A first MILP model for the parameterization of Demand-Driven MRP. *Computers & Industrial Engineering.* 2023.*Doi: 10.1016/j.cie.2022.108769.*

Metin İfraz, Adnan Aktepe, Süleyman Ersöz, Tahsin Çetinyokuş. Demand forecasting of spare parts with regression and machine learning methods: Application in a bus fleet.Journal of Engineering *Research,Volume 11, Issue 2.* 2023. *https://doi.org/10.1016/j.jer.2023.100.*

Thivakaran, T. K., & Ramesh, M. Exploratory data analysis and sales forecasting of bigmart dataset using supervised and ANN algorithms. *Measurement: Sensors, 23,* 2022. *doi:10.1016/j.measen.2022.100388 (2022)*

Pavlyshenko, B.M. Machine-Learning Models for Sales Time Series Forecasting. *Data 2019, 4, 15.* 2019. *https://doi.org/10.3390/data4010015.*

Guillaume Martin; Matthieu Lauras. Dynamical multi-parameter sizing of DDMRP buffers in finite capacity flow-shops. *Computers & Industrial Engineering.* 2023. *Doi: 10.1016/j.cie.2022.108858.*

Louis Duhem; Maha Benali. Parametrization of a demand-driven operating model using reinforcement learning. *Computers in Industry.* 2023. *Doi: 10.1016/j.compind.2023.103874.*

EMIS,Lima.Available:https://www.emis.com/php/companyprofile/PE/Unilever_Andina_Peru_SA_es_2023331, 2023.

Minitab. Example of Augmented Dickey-Fuller Test. 2023. *https://support.minitab.com/en-us/minitab/21/help-and-how-to/statistical-modeling/time-series/how-to/augmented-dickey-fuller-test/before-you-start/example/*

Sturtz. Introduction to Python 3. *Realpython.com; Real Python. https://realpython.com/python-introduction/*

Institute of Business Forecasting & Planning., What Is Forecast Accuracy? 2018. *https://ibf.org/knowledge/glossary/forecast-accuracy-2-128, June 22, 2021.*

Goltsos, T. E., Syntetos, A. A., Glock, C. H., & Ioannou, G. Inventory – forecasting: Mind the gap. *European Journal of Operational Research, 299(2), 397-419.* 2022. *doi:10.1016/j.ejor.2021.07.040.*

Company´s Planner. *Anonymous Company.,* 2023.

Biographies

Diana Micaela Mariscal Cruzado: Bachelor on industrial engineer from the Faculty of Engineering at Universidad de Lima, Lima, Peru.

Bruno Piero Gonzalez Jáuregui: Bachelor on industrial engineer from the Faculty of Engineering at Universidad de Lima, Lima, Peru.

José Antonio Taquía is a Doctoral Researcher from Universidad Nacional Mayor de San Marcos and holds a Master of Science degree in Industrial Engineering from University of Lima. He is a member of the School of Engineering, teaching courses on quantitative methods, predictive analytics, and research methodology. He is also a member of the Scientific Research Institute at the Universidad de Lima. His main research interests are on statistical learning, predictive analytics, and industry 4.0.