Optimization of logistics operator planning and dock allocation in a warehouse

DOUZI Ilham[1] and MESSOUS Romaissa[1]
Supervised by : MOULAI Ratiba [1] and REZGUI Wail[2]

Abstract

This article aims to enhance the productivity of Numilog, a logistics and transportation service provider, through two main objectives. The first step involves using traditional forecasting methods to estimate goods flow. These forecasts form the basis for applying artificial intelligence techniques to predict the optimal number of logistics operators needed daily. The second step focuses on developing an algorithm dedicated to optimizing dock allocation within the warehouse. The goal is to select the nearest dock for each order preparation and stock placement operation. The results obtained show a significant improvement in the accuracy of flow forecasts, as well as in the prediction of the number of necessary operators. Additionally, the dock allocation algorithm has reduced the distance traveled, thereby improving logistical efficiency.

Keywords : Artificial Intelligence, Forecasts, Dock Allocation, Warehouse, Productivity, Python, R.

1 Introduction

Logistics represents a central pillar of modern industry, essential for the smooth flow of commercial exchanges and customer satisfaction. At the heart of this chain, logistics warehouses play a crucial role by ensuring the efficient receipt, storage, and distribution of goods. However, to remain competitive in a dynamic and demanding economic environment, it is essential to adopt advanced methods aimed at optimizing and improving these distribution centers [**1**].

Warehouse optimization goes far beyond simply maximizing storage space or reducing operational costs. It is a strategic process aimed at aligning logistics capabilities with market demands while ensuring optimal operational efficiency[**2**].

Accurately forecasting the flow of goods and the necessary personnel is essential for maintaining smooth and efficient operations. In this context, the use of appropriate forecasting methods is of paramount importance. Traditional approaches analyze past trends to provide robust estimates of future needs in the flow of goods. Similarly, advanced artificial intelligence methods exploit historical and real-time data, facilitating informed decision-making in logistics operations planning[**3**].

An efficient allocation of docks also significantly improves stock management and reduces operational costs by minimizing the distances traveled for stocking and order preparation.

This article explores these practical challenges through the example of the Numilog warehouse, highlighting the strategies and technologies used to optimize its logistics operations and enhance its productivity.

2 Related works

Many studies utilize time series analysis techniques such as ARIMA models, exponential smoothing, or TBATS to forecast inventory levels and demand patterns. These techniques help warehouses anticipate future stock requirements, optimize storage space, and streamline supply chain operations.

[1] *Department of Logistics and Transport Engineering,National Higher School of Advanced Technologies (ENSTA) Dergana, Algeria.*
[2] *Department of Industrial Engineering and Maintenance,National Higher School of Advanced Technologies (ENSTA) Dergana, Algeria.*

G. Rivera, R. Florencia-Juárez, J. P. Sánchez-Solís, V. García, and C. D. Luna [**4**] implemented time-series models to forecast the demand for parts in an assembly plant warehouse. They applied Holt-Winters Seasonal and SARIMA models, using AIC and BIC metrics to identify the best models. Their results showed that the SARIMA model outperformed the Holt-Winters Seasonal model, providing more accurate demand forecasts for improving the relocation process in the warehouse.

Jamal Fattah, Latifa Ezzine, Zineb Aman Haj El Moussami, and Abdeslam Lachhab[**5**] applied the ARIMA model to forecast demand in a food company. They used historical demand data to develop several ARIMA models using the Box–Jenkins time series procedure.The chosen ARIMA (1, 0, 1) model was validated with additional historical data, proving its effectiveness in forecasting future demand. These results provide reliable guidelines for decision-making by managers in the food manufacturing sector.

Claudimar Pereira Da Veiga, Cássia Rita Pereira Da Veiga, Anderson Catapan, Ubiratã Tortato, and Wesley Vieira Da Silva [**6**]compared the Holt-Winters (HW) and ARIMA models for demand forecasting in food retail. They focused on perishable dairy products, where accurate forecasting is crucial due to the volatile demand patterns. The study found that the HW model outperformed the ARIMA model. The HW model provided better adjustment and effectively captured the linear behavior of the time series, making it a more suitable choice for predicting demand in this context. Many studies also utilize machine learning techniques, such as Linear regression, random forest and Xgboost models, to optimize various aspects of business operations.

Akinbola S. M. and Buoye P. A.[**7**] explored predictive analytics, focusing on random forest models, to optimize resource allocation in data warehousing and data mining.The study aimed to use random forest models to predict resource utilization trends based on historical data, enabling proactive optimization strategies. Historical data on resource usage metrics, including timestamps and data warehouse storage, were collected and analyzed. The random forest model successfully identified trends in past data, facilitating predictions of future storage requirements.

M. Kavitha, R. Srinivasan, R. Kavitha, M. Suganthy [**8**]conducted a study focused on sales demand forecasting for retail marketing using the XGBoost algorithm. The primary objective was to predict the demand for goods and services available at different times of the year, aiming to satisfy consumer needs and maximize retailer profits. The paper utilized time series analysis alongside the XGBoost algorithm to forecast sales data, capturing trends and fluctuations in sales over various periods. The model achieved a high level of accuracy, demonstrating its effectiveness in making well-trained and precise predictions.

3 Data Exploration and Methodology

In this article, we have chosen the Numilog warehouse as a case study, which plays a central role as a leader in Algeria in the field of logistics. The company is committed to maintaining its competitiveness by precisely meeting its customers' expectations in terms of cost, quality, and deadlines. However, it seeks to improve its productivity, particularly due to the lack of forecasts regarding the number of logistics operators, which directly affects operational efficiency. Additionally, the arbitrary allocation of loading docks leads to significant time losses and operational inefficiencies, requiring additional distances. These inefficiencies have a direct impact on the company's productivity.

To address these challenges, we have developed decision support tools that address three sub-problems :

- Forecasting of Goods Flow : Using traditional forecasting methods to estimate daily volumes of goods flow.
- Forecasting the Number of Daily Operators : Applying machine learning algorithms to historical data to determine the necessary number of logistics operators based on goods flow forecasts.
- Dock Allocation : Implementing an algorithm to optimize dock assignment during loading operations.

3.1 Goods flow forecasting problem

3.1.1 Data preparation

In this section, we present a detailed analysis of the data for forecasting the flow of goods at the Numilog logistics warehouse, based on data collected from January 2023 to May 2024. This data includes the total sum of incoming and outgoing goods flows.

The study begins with a thorough visualization of the data in the form of daily time series with a 7-day period. This initial analysis highlights a general downward trend in goods flow over time, as well as pronounced seasonality characterized by regular peaks. The visualization used helps detect outliers, requiring intervention to maintain data integrity.

To identify and manage these anomalies, an approach combining statistical tests and visualization techniques such as box plots was employed. Outliers were replaced with missing values (NA :Not Available) and imputed using linear interpolation, ensuring data consistency throughout the analysis.

Statistical validation was conducted using the Fisher test to confirm significant trends and seasonality. Subsequently, the Buys-Ballot test was used to determine the multiplicative nature of seasonal variations.

This study provides a solid methodological foundation for modeling and forecasting goods flows at the Numilog warehouse.

3.1.2 Methodology and results

In this section, we will conduct a comparative evaluation of three specific forecasting models to determine which one provides the most accurate predictions. To achieve this, we divided our dataset into two parts : training data and test data (which correspond to the last 28 days of the series).

Box-Jenkins Model

To effectively apply the Box-Jenkins model, we first adjusted the time series by removing its seasonality. Next, we checked for stationarity using the Dickey-Fuller test. The results showed a significant trend with a p-value of 0.016, indicating the absence of a unit root in the series (p-value = 3.65e-05). Therefore, the initial time series is non-stationary (type ts), requiring the removal of trend as well. A second Dickey-Fuller test was conducted to validate the stationarity of the transformed data.

Given that our model was of the "ts" type, implying a choice between SARMA or ARMA, we presumed the use of a SARMA model. Subsequently, we conducted a search for optimal parameters (p, d, q) and (P, D, Q) to minimize the Akaike Information Criterion (AIC). The results revealed that the optimal SARMA model had parameters (5, 0, 7)(2, 0, 0), indicating the presence of a significant seasonal component that precludes the use of an ARMA model.

To evaluate the performance of the identified SARMA model, we generated forecasts for the next 28 days using the forecast function in R, reintegrating the previously removed trend and seasonality.

Date	Actual	Forecast
14/04/2024	1385	1876.239
15/04/2024	1032	1918.668
16/04/2024	1877	1819.141
17/04/2024	1653	1721.004
18/04/2024	2016	1656.073
19/04/2024	2082	1645.249
20/04/2024	1777	1449.56
21/04/2024	1337	1761.34
22/04/2024	1654	1768.567
23/04/2024	2091	1653.863

Table 1 – SARMA Model Results

Exponential Smoothing Model

Following our analysis, it became evident that our time series exhibits both trend and seasonality. To effectively model these characteristics, we opted for the Holt-Winters model, renowned for its ability to capture trends and seasonal variations in time series data. The process of tuning the parameters of the Holt-Winters model involves finding optimal values for these parameters, which we accomplished using the Holt-Winters function in R. Subsequently, to generate our forecasts, we utilized the forecast function available in R.

Date	Actual	Forecast
14/04/2024	1385	1404.241
15/04/2024	1032	1769.181
16/04/2024	1877	1760.352
17/04/2024	1653	1741.794
18/04/2024	2016	1615.572
19/04/2024	2082	1616.784
20/04/2024	1777	13/9.398
21/04/2024	1337	1406.472
22/04/2024	1654	1771.993
23/04/2024	2091	1763.148

Table 2 – Holt-Winters Model Results

TBATS Model

We also chose to employ the TBATS model, recognized for its capability to capture complex trends and seasonal variations in time series data. Tuning the parameters of the TBATS model involves determining optimal values for its various parameters, a task we completed using the Tbats function in R. For generating our forecasts, we utilized the forecast function available in R.

Date	Actual	Forecast
14/04/2024	1385	1465.345
15/04/2024	1032	1581.515
16/04/2024	1877	1508.159
17/04/2024	1653	1485.802
18/04/2024	2016	1640.993
19/04/2024	2082	1670.865
20/04/2024	1777	1523.712
21/04/2024	1337	1571.055
22/04/2024	1654	1672.581
23/04/2024	2091	1581.343

Table 3 – TBATS Model Results

3.1.3 Comparison of model performances

The evaluation of model fits was conducted using the FTSA library, revealing the following table :

Model	MAE	RMSE	MAPE (%)
Box-Jenkins	285	345.41	20.54
Exponential Smoothing	275.69	342.02	20.16
TBATS	252.67	308.20	17.95

Table 4 – Table of Metrics

The results of this evaluation strongly support the recommendation to use the TBATS method for warehouse flow forecasting applications. This approach will be particularly useful for estimating the required number of operators using machine learning methods.

3.2 Forecasting daily operator numbers

3.2.1 In-Depth Data Exploration

The data collected by the company from January 2023 to May 2024 allowed us to compile a detailed history of goods flows, including dates and quantities of goods processed each day (combined receipts and shipments). We also gathered historical data related to productivity, a key indicator for assessing warehouse operational efficiency, along with the daily count of logistics operators .

We systematically prepared our data by starting with the removal of duplicates and replacing missing values with the mean of available data within the same column. This ensures data continuity while minimizing impact on analyses. Concurrently, we converted dates into a suitable format to facilitate chronological analysis.

To stabilize data variance, we applied a logarithmic transformation followed by Min-Max Scaling normalization. This approach standardizes the scale of values between 0 and 1, enhancing comparability across the data.

Outliers were handled by first replacing them with NaN (Not a Number) and then imputing them with the mean of each column, ensuring data consistency and reliability. This rigorous preprocessing process was essential to optimally prepare the data for integration into machine learning models.

3.2.2 Implementation of machine learning methods

Among the various algorithms evaluated in the literature for forecasting applications, we have chosen to present those that provide the best estimates for our data : Random Forest (RF), eXtreme Gradient Boosting (XGBoost), and Linear Regression (LR).

XGBoost and Random Forest are popular algorithms belonging to the ensemble methods family. These algorithms combine multiple decision trees to predict a target variable while minimizing overall error. XGBoost sequentially builds multiple relatively weak models, each attempting to correct errors made by its predecessor. In contrast, Random Forest utilizes multiple decision trees in parallel, with predictions computed by averaging predictions from all trees.

Linear Regression models the relationship between a dependent variable and one or more independent variables in a linear form, aiming to minimize the difference between predicted and observed values of the dependent variable.

3.2.3 Model parameter optimization

We utilized cross-validated Grid Search CV to select the best hyperparameters and evaluate model performance. This method automatically divides the training dataset into multiple subsets for rigorous model evaluation.

3.2.4 Evaluation and results

To assess the models, we employed two primary criteria :

- Error Evaluation : We calculated the training error on predictions (ypred) and compared it with the test data error (ytest). This allowed us to determine which model learned best to generate the most accurate and reliable forecasts.

- Success Rate in Achieving Productivity Goal : In addition to error evaluation, we used model-generated forecasts to calculate expected productivity, set at a minimum threshold of 20 units. The success rate measured how often each model achieved or exceeded this goal.

For this study, we excluded data from the last 28 days to evaluate forecast performance and used these days' forecasts to calculate productivity. Features such as date, goods flow, and productivity were integrated as input variables into each model. For instance, the "Flow" feature represents goods flow forecasts obtained from the TBATS model described earlier, while the "Productivity" feature reflects our minimum target to achieve.Thus, we included the company's logistics operators forecasts in our evaluation. This approach allows us to compare all the models, including the company's, to determine the one most suited to our needs.

The summary table below presents the calculated errors and the success rate of achieving the productivity goal for each model :

Model	MAE	RMSE	MAPE (%)	Accuracy (%)
Random Forest	2.03	2.69	13.12	68
XGBoost	1.84	2.47	12.05	82
Linear Regression	1.75	2.32	11.22	64
Company	3.07	3.60	28.08	29

Table 5 – Summary Table of Model Performances

By analyzing the data in the table, we observe that linear regression is characterized by better accuracy in terms of errors. Meanwhile, XGBoost stands out for its notable efficiency, achieving a rate of 80%, which is considered satisfactory and appropriate. Additionally, it is clear that the forecasts generated by XGBoost have shown a significant improvement in productivity compared to the forecasts generated by the company.

Based on these results, we recommend the use of XGBoost as the forecasting model to determine the number of necessary logistics operators, due to its superior ability to achieve the productivity targets set by the company and the analysis of forecast errors indicating acceptable performance for our use case.

3.3 Loading dock allocation problem

3.3.1 Data related to dock allocation

In our data management strategy for loading dock assignment,, we adopted a rigorous approach by constructing a detailed matrix that integrates specific information obtained from each location, including rows, aisles, levels, and positions. Subsequently, we assigned the nearest dock to each location.

3.3.2 Implementation of the solution

In this part of our process, we addressed the selection of the nearest dock for each order. The objective is to determine the optimal dock to minimize the required time by reducing the distance traveled for order preparation.

Firstly, we assigned the nearest dock to each location in the order. This approach was chosen to work directly with the docks rather than the locations themselves. Next, we calculated the distance between these docks and the available docks in the warehouse using the Manhattan distance.

Subsequently, we searched for which available dock in the warehouse minimizes the sum of distances, while adhering to the constraint of dock availability. We developed an algorithm summarizing this approach :

Algorithm 1 Dock Selection

Input :
File containing locations and corresponding docks
File containing the order
List `availableDocks` containing available docks for selection
Output :
The appropriate dock
Begin :

```
Create a dictionary from the file containing locations and corresponding docks, where the key is the location details and the value is the corresponding dock.
For each line in the order file :
    Use the dictionary to find the corresponding dock.
Create a list docks from the new dock column of the order data.
Define a matrix of dimensions (size of docks, size of availableDocks).
For each dock in docks :
    For each available dock in availableDocks :
        Calculate the distance between the dock and the available dock.
Calculate the sum of each column of the matrix to get total distances per availableDocks.
Create a list of tuples (dock, sum of distances).
Sort the list by the sum of distances in ascending order.
For each tuple (dock, distance) in the list :
    Display the dock.
    Ask the user if the dock is available.
    If the dock is available :
        Display the dock as optimal.
        Exit the loop.
If no dock is available :
    Display a message indicating no dock is available and to wait.
```

End

3.3.3 Results

After implementing our resolution method on a real case at Numilog warehouse, we obtained the following results :

	Chosen Loading Dock	Distance (distance unit)
Result obtained by our resolution	21	45
Result obtained by the company	19	73

These results demonstrate that our method identified Dock 21 with a distance of 45 units. In contrast, the company, by arbitrarily choosing Dock 19, obtained a distance of 73 units. This significant difference highlights the effectiveness of our approach. By minimizing the distance, our method significantly reduces travel times and consequently improves productivity.

4 Conclusion

In conclusion, continuous improvement in distribution logistics is essential to maintain competitiveness in today's market. This article has highlighted the importance of this approach by focusing on optimizing the productivity of a logistics warehouse through three major challenges : forecasting goods flows, predicting daily logistics operator needs, and optimizing loading dock allocation.

To address the challenge of forecasting goods flows, various approaches were evaluated, among which the TBATS model proved most suitable after a thorough analysis of relevant error measures.

Regarding the forecasting of the number of daily logistics operators, the TBATS model results were integrated with machine learning techniques such as Linear Regression, XGBoost, and Random Forest. This comparison showed that XGBoost offered superior productivity, justifying its adoption in our solution approach.

Finally, a specific algorithm was developed to enhance loading dock allocation, demonstrating its advantages through a comparative case study with existing company practices.

The results confirmed a significant improvement in warehouse productivity, enhancing its ability to effectively and proactively address contemporary logistics challenges. This integrated approach underscores the crucial importance of continuous innovation in logistics to ensure not only competitiveness but also long-term operational sustainability.

For future work, it would be beneficial to further explore proactive management of material resources using predictive models to predict failures and improve preventive maintenance. Additionally, integrating our solution into other Numilog warehouses could standardize practices and enhance productivity and responsiveness to fluctuating market demands.

Bibliography

[1] KUCHARAVY, D., DAMAND, D., BARTH, M., DERROUICHE, R., MORNAY, S. ENTREPOT DU FUTUR : CARTOGRAPHIE DE CONTRADICTIONS.

[2] Gwynne, R. (2014). Warehouse management : a complete guide to improving efficiency and minimizing costs in the modern warehouse. Kogan Page Limited.

[3] Odimarha, A. C., Ayodeji, S. A., Abaku, E. A. (2024). Machine learning's influence on supply chain and logistics optimization in the oil and gas sector : a comprehensive analysis. Computer Science IT Research Journal, 5(3), 725-740.

[4] Rivera, G., Juárez, R. F., Sánchez-Solís, J. P., García, V., Luna, C. D. (2020). Forecasting the Demand of Parts in an Assembly Plant Warehouse using Time-Series Models. Polibits, 62, 59-67.

[5] Fattah, J., Ezzine, L., Aman, Z., El Moussami, H., Lachhab, A. (2018). Forecasting of demand using ARIMA model. International Journal of Engineering Business Management, 10, 1847979018808673.

[6] Da Veiga, C. P., Da Veiga, C. R. P., Catapan, A., Tortato, U., Da Silva, W. V. (2014). Demand forecasting in food retail : A comparison between the Holt-Winters and ARIMA models. WSEAS transactions on business and economics, 11(1), 608-614.

[7] Akinbola, S. M., Buoye, P. A. (2024). Predictive Analytics for Resource Optimization in Data Warehousing and Data Mining Using Random Forest. International Journal of Women in Technical Education and Employment, 5(1), 33-39.

[8] Kavitha, M., Srinivasan, R., Kavitha, R., Suganthy, M. (2023). Sales Demand Forecasting for Retail Marketing Using XGBoost Algorithm. Intelligent and Soft Computing Systems for Green Energy, 127-140.