Wholesale Inventory Management Optimization: Methodological Approach with XGBoost, SVR, and Random Forest Algorithms

Carli Apriansyah Hutagalung1*, Gisela Anastacia Rosalind2, Dewi Masito Setyo Tuhu3, Ayu Agustianingsih4

1Computer Science Program, Faculty of Computer Science, Universitas Media Nusantara Citra, Indonesia.
2,3,4Mathematics Education Program, Faculty of Teacher Training and Education, Universitas Media Nusantara Citra, Indonesia.
1carli.apriansyah@mncu.ac.id, 2giselaanastacia08@gmail.com, 3dewimst5@gmail.com, 4ayuagustianingsih9@gmail.com

ABSTRACT

This research aims to optimize wholesale inventory management at PT Primafaoood International Pasir Putih 2 by implementing leading algorithms, namely XGBoost, Support Vector Regression (SVR), and Random Forest. In the wholesale industry, effective inventory management plays a crucial role in maintaining smooth production processes and enhancing company profitability. Despite the acknowledged benefits of inventory management, there are aspects that remain not fully disclosed, particularly concerning demand uncertainty and market fluctuations. This study addresses these gaps by exploring the potential of these three algorithms. Experimental methods with a quantitative approach were employed to shape and prepare the dataset. The analysis and predictions' results using XGBoost, SVR, and Random Forest were evaluated using metrics such as Mean Squared Error (MSE), F1-Score, and Accuracy. The evaluation indicates that XGBoost and SVR exhibit optimal performance with low MSE values of 7714.446 and 119.315, high F1-Scores (0.92), and good accuracy levels (0.86 and 0.85), respectively. While Random Forest shows a higher MSE, it still delivers solid performance with an F1-Score of 0.89 and an accuracy rate of 0.81. These findings suggest that all three algorithms can be considered to enhance inventory management performance at PT Primafaoood International Pasir Putih 2, with significant potential benefits for overall industry development. This research provides valuable insights for decision-making at the business and industrial levels, highlighting the effectiveness of each algorithm in the context of predicting stock level.

INTRODUCTION

In the realm of industry, inventory management is a critical aspect that requires serious attention. Optimal inventory not only helps maintain the smoothness of the production process but also has a direct impact on the profit and operational efficiency of a wholesale company, particularly at PT Primafaoood International Pasir Putih 2. In this context, this research aims to optimize wholesale inventory management by applying three leading algorithms: XGBoost, Support Vector Regression (SVR), and Random Forest. Wholesale inventory involves a large number of items that need to be carefully managed. Understanding this subject involves comprehension of the complexity and dynamics of inventory in the wholesale industry. In the literature, it is known that effective inventory management can improve operational efficiency, reduce storage costs, and prevent detrimental stockouts (Jakkraphobyothin, Srifa, & Chinda, 2018; S et al., 2023).

Despite the growing knowledge of the benefits of good inventory management, there are still aspects that have not been fully revealed, especially in dealing with the challenges of handling demand uncertainty, market fluctuations, and seasonal variations in the wholesale industry. Conventional methods commonly used have proven to be less effective in addressing these challenges. Therefore, this research is directed at filling this gap by exploring the potential of the XGBoost, SVR, and Random Forest algorithms. Literature findings indicate that machine learning techniques, such as the application of the XGBoost algorithm in operational management, show an increasing trend in the use of machine learning in the supply chain, providing effectiveness in predicting backorders. This provides a strong foundation that this approach can improve inventory management efficiency (Bertolini, Mezzogori, Neroni, & Zammori, 2021; Lolli et al., 2019; Ni, Xiao, & Lim, 2020; Ntakolia, Kokkotis, Karlsson, & Moustakidis, 2021; Svoboda & Minner, 2022).
This research aims to fill the knowledge gap related to wholesale inventory management by applying advanced algorithms such as XGBoost, SVR, and Random Forest. The expected solution should be more accurate and adaptive to the constantly changing market dynamics. The goal of this research is to contribute to the improvement of efficiency and effectiveness in wholesale inventory management, which, in turn, will support operational smoothness and increase company profits. Thus, this research is significant in presenting a better solution for wholesale inventory management, with the potential to enhance competitiveness and resilience in a dynamic market.

**METHOD**

This research employs a quantitative approach to optimize inventory management at PT Primafood International Pasir Puth 2. Experimental methods are applied using three leading algorithms, namely XGBoost, Support Vector Regression (SVR), and Random Forest. The research workflow is illustrated in Fig. 1.

![Figure 1. Research Flowchart](image)

**Data Set Preparation and Formation**

The obtained data consists of an Excel soft copy from PT Primafood International Pasir Puth 2, containing daily data spanning from the beginning of July to the end of September 2023. Sales information is available in Table 1 of the Data Set.

<table>
<thead>
<tr>
<th>No</th>
<th>ID Item</th>
<th>Product Name</th>
<th>Stock</th>
<th>1 Jul 2023</th>
<th>...</th>
<th>30 Sept 2023</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>brg dtg</td>
<td>sales stock</td>
<td>brg dtg sales stock</td>
</tr>
<tr>
<td>1</td>
<td>10000018</td>
<td>ZODA 250 ML/BTL</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>TEH KOTAK JASMINE 300 ML</td>
<td>23</td>
<td>0</td>
<td>1</td>
<td>23</td>
</tr>
<tr>
<td></td>
<td></td>
<td>FIESTA CHIC SLICE FRIED RICE 320 GR/PAC</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>1201</td>
<td>21310102</td>
<td>FIESTA CHIC SLICE FRIED RICE 320 GR/PAC</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

The dataset in Table 1 is a compilation of daily data obtained by the SO AUDIT PFM Pasir Puth 2 team. It consists of 1201 different items, where each row presents relevant information. The 'ID Item' column provides a unique identification number for each item, while the 'Product Name' column lists the names of the items recorded in the dataset. The 'Stock' column reflects the total stock quantity at the beginning of the period on July 1, 2023. The data is proportionally split into 80% training and 20% testing.
Additionally, the 'Incoming Goods' column records the quantity of items shipped by the distributor to the store on that day. Sales information is reflected in the 'Sales' column, while the 'Stock' column indicates the availability of items after deducting the sales quantity. The columns 'Incoming Goods,' 'Sales,' and 'Stock' represent grouped data for each day within the timeframe from July 1, 2023, to September 30, 2023. The next step involves data cleaning to optimize the analysis.

The data set formation identifies key variables such as incoming goods, sales, and remaining stock. The focus of the prediction is the daily available stock quantity. This is intended to assist managers in determining the quantity of items that need to be sent by the distributor to the store for each item per week in the future.

Application of Machine Learning Algorithms

This stage aims to standardize variables within the equations of the XGBoost, Support Vector Regression (SVR), and Random Forest algorithms to achieve results aligned with the objectives of this research. The following are explanations of the equations of the algorithms used.

a. XGBoost Algorithm

XGBoost (Extreme Gradient Boosting) is an ensemble algorithm that has proven to be effective in various business contexts. The strength of XGBoost lies in its ability to model the relationship between attributes and the target variable in an adaptive and robust manner (Gupta, Yadav, Jha, & Pathak, 2022; Jain & Prasad, 2020). This formula can be seen in Equation 1.

\[ y_i = \frac{1}{K} \sum_{k=1}^{K} f_k(X_i) \]  

(1)

Explanation;

- \( i \) : Index of the data row.
- \( k \) : Total number of trees in the XGBoost model.
- \( y_i \) : Predicted target variable.
- \( f_k(X_i) \) : Model mapping the attributes of incoming goods, sales, and remaining stock.

The equation above illustrates how the final prediction \( y_i \) is calculated as the sum of predictions provided by each tree \( f_k(X_i) \). In other words, XGBoost constructs a series of decision trees to model the relationship between attributes and the target variable in an adaptive and robust manner.

b. Support Vector Regression (SVR) Algorithm

The selection of Support Vector Regression (SVR) is based on the need to handle non-linear and complex characteristics in wholesale inventory data. SVR has the ability to adapt to patterns that cannot be accommodated by linear models (Cao, Xu, Goodman, Bao, & Zhu, 2020; Mukherjee & Ramachandran, 2018; Sun, Ding, Zhang, & Jia, 2021). Therefore, it is expected that SVR prediction results can provide higher accuracy, making it a strategic choice in the context of optimizing wholesale inventory management (Panahi, Sadhasivam, Pourghasemi, Rezaie, & Lee, 2020; Tao, Yan, Gao, Sun, & Li, 2019; Zheng, Wang, Li, & Chen, 2021). The SVR formula is as follows in Equation 2.

\[ Y = f(X) \sum_{i=1}^{N}(\alpha_i - \alpha_i^*)K(X_i, X) + b \]  

(2)

Explanation;

- \( Y \) : Predicted target variable.
- \( N \) : Number of samples in the training data.
- \( \alpha_i \) and \( \alpha_i^* \) : Dual variables of the optimal solution.
- \( K(X_i, X) \) : Kernel function measuring the similarity between two feature vectors \( X_i \) and \( X \).
- \( b \) : Bias.

This model utilizes a kernel function to map data into a higher dimension, enabling the handling of non-linear patterns.

c. Random Forest Algorithm

In this stage, the research implements the Random Forest algorithm to predict the quantity of items that should be sent by the distributor to the store per item per day. Random Forest is chosen because it can accommodate the diversity and fluctuations in wholesale inventory data (J.-M. Zhu, Geng, Li, Li, & He, 2022). By leveraging multiple decision trees, Random Forest can capture nuances and patterns that may be challenging to identify by a single model. This significantly improves the accuracy of predicting the quantity of items to be sent, supporting more effective
inventory management and responsiveness to market changes (Lohrmann & Luukka, 2019; Roy, Chopra, Lee, Spampinato, & ivatlood, 2020; Sandag, 2020; Sari, Firdausi, & Azhar, 2020). The formula for this algorithm is as follows in Equation 3.

\[
Y = \frac{1}{N} \sum_{i=1}^{N} T_i(X)
\]

Explanation:
- \(Y\): Predicted target variable.
- \(N\): Number of decision trees in the Random Forest model.
- \(T_i(X)\): Prediction given by the \(i\)-th decision tree, depending on attributes \(X\) such as incoming goods, sales, and remaining stock.

This formula reflects that the final prediction \(Y\) is calculated as the average of predictions given by each decision tree in the Random Forest. This ensemble approach helps address variability and uncertainty in wholesale inventory data (Saadah & Salsabila, 2021).

Model Evaluation
Evaluating the model is the final stage of this research, using evaluation metrics to measure the performance of the algorithms applied. Firstly, Mean Squared Error (MSE) is employed to gauge the average squared error between predictions and actual values, reinforcing theoretical findings (Tang, Yan, Tang, & Chien, 2022). Subsequently, F1-Score is useful for assessing the model's performance in identifying the quantity of items to be sent, a metric particularly suitable for the algorithm used in this research (Aditama & Munir, 2022; Mouhajir, Nechba, & Sedjari, 2023; Sairam & K, 2022). Accuracy is examined using the built-in library, specifically the accuracy_score from the matrix, aiming to provide an overall overview of how well the model classifies the items. The results of this evaluation stage will be analyzed to determine the model that best aligns with the research objectives. Comparisons between the XGBoost, Random Forest, and SVR methods will provide insights into the relative performance of each algorithm.

RESULT AND DISCUSSION

Dataset Preparation and Formation
The data cleaning process in this research serves as a crucial initial step to ensure the quality of the dataset used. This stage is commonly referred to as the preprocessing stage, with the goal of identifying missing data in each variable. Next, the variables with the highest number of missing values are determined, and imputation is performed using both conventional and modern methods. As a result, the processed data retains 608 items. Finally, the initial dataset is converted to Comma Separated Values (CSV) format to facilitate the algorithm analysis process. The data that has undergone the cleaning stage can be seen in Table 2 Training Data.

<table>
<thead>
<tr>
<th>Line Number</th>
<th>brg_dtg1,sales1,stock1,,,...,...,brg_dtg92,sales92,stock92</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.0,0.0,28.0,0.0,0,0,28.0,0.0,0,0,28.0,0.0,0.0,0,0,0.0,6.0</td>
</tr>
<tr>
<td>...</td>
<td>..... ..... ..... ..... ..... ..... ..... ..... ..... ..... .....</td>
</tr>
<tr>
<td>608</td>
<td>0.0,0.0,5.0,0.0,0,0,5.0,0.0,0,0,5.0,0.0,0.0,5.0,0.0,0.0,0,0,1.0</td>
</tr>
</tbody>
</table>

The next step involves dividing the time variable into two parts, namely, the variable for training data and the variable for testing data, with an 80:20 ratio.

Algorithm Modeling
a. XGBoost Algorithm
In the modeling stage, this research optimizes key parameters of the XGBoost algorithm to enhance model performance (Zhang, Gao, Lou, Yan, & Hu, 2019). Some parameters utilized in this study include the number of decision trees in the model set to 200, the learning rate set to 0.05, the maximum depth for each tree set to 5, the fraction of samples used to train each tree set to 0.8, the fraction of features used to train each tree set to 0.8, and the random state set to 42. This model successfully accommodates the prediction of daily data, and the prediction results can be seen in the Table 3.
Table 3. Results of XGBoost

<table>
<thead>
<tr>
<th>Line Number</th>
<th>brg_dtg93</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2.532690</td>
</tr>
<tr>
<td>2</td>
<td>1.008038</td>
</tr>
<tr>
<td>3</td>
<td>12.698670</td>
</tr>
<tr>
<td>...</td>
<td>..</td>
</tr>
<tr>
<td>608</td>
<td>8.605600</td>
</tr>
</tbody>
</table>

The graph of actual and predicted data can be seen in Figure 2.

![Confusion Matrix (XGBoost)](image)

Figure 2. Confusion Matrix (XGBoost)

In the Confusion Matrix graph (XGBoost), the model's prediction results show True Negative with a count of 5, True Positive with a count of 95, False Negative with a count of 9, and False Positive with a count of 9. True Negative and True Positive values indicate the model's success in predicting the corresponding categories, while False Negative and False Positive indicate errors in the model's data classification. The results of this processing yield an accuracy rate of 0.819, an MSE of 1142.393, and an F1-Score with a value of 0.896.

b. Support Vector Regression (SVR) Algorithm

Much like XGBoost, the Support Vector Regression (SVR) algorithm requires parameter adjustments in this research to enhance model performance. The study sets the parameters, adjusting the C value to 1.0, epsilon to 0.1, and utilizing a kernel function (Yaqin et al., 2022; Yigit, Gunel, & Gunel, 2018; Zhan, Zhang, & Liu, 2021; H. Zhu & Hu, 2019). The results can be seen in Table 4.

Table 4. Results of SVR

<table>
<thead>
<tr>
<th>Line Number</th>
<th>brg_dtg93</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>7,759795</td>
</tr>
<tr>
<td>2</td>
<td>5,347782</td>
</tr>
<tr>
<td>3</td>
<td>5,270168</td>
</tr>
<tr>
<td>...</td>
<td>..</td>
</tr>
<tr>
<td>608</td>
<td>5,227086</td>
</tr>
</tbody>
</table>

The graph of actual and predicted data can be seen in Figure 3.
The confusion matrix graph evaluates the performance of the Support Vector Regression (SVR) model on two classes in the test data. From the graph, it can be observed that the model correctly predicted 104 observations as class 1 (True Positive), while there were no observations wrongly predicted as class 0 (True Negative). However, there were 18 observations incorrectly predicted as class 1 (False Positive), and no observations were wrongly predicted as class 0 (False Negative). This information provides an overview of the model's success in recognizing class 1 but also highlights errors in predicting class 0. The obtained MSE is 6.5086, F1-Score with a value of 0.92, and an accuracy rate of 0.852.

c. Random Forest Algorithm

In the modeling stage using the Random Forest algorithm, adjustments were made to several key parameters to enhance model performance (Demidova & Ivkina, 2019; Kurniawati, Novita Nurmala Putri, & Kurnia Ningsih, 2020). The number of decision trees used is 100, and the assigned random state value is 42. This model successfully provides predictions for daily data, as seen in Table 5.

<table>
<thead>
<tr>
<th>Line Number</th>
<th>brg_dtg93</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>6,41</td>
</tr>
<tr>
<td>2</td>
<td>2,29</td>
</tr>
<tr>
<td>3</td>
<td>10,682</td>
</tr>
<tr>
<td>...</td>
<td>..</td>
</tr>
<tr>
<td>608</td>
<td>4,858</td>
</tr>
</tbody>
</table>

The graph of actual and predicted data can be seen in Figure 4.
In the Confusion Matrix graph (Random Forest), the model's prediction results show True Negative with a count of 1, True Positive with a count of 104, False Negative with a count of 0, and False Positive with a count of 17. True Negative and True Positive values indicate the model's success in predicting the corresponding categories, while False Negative and False Positive indicate errors in the model's data classification. The results of this processing yield an accuracy rate of 0.861, MSE of 7714.446, and an F1-Score with a value of 0.924. The performance evaluation of the model in the confusion matrix provides a comprehensive overview of the model's success and weaknesses in recognizing existing classes.

Therefore, the results from these three algorithms can serve as a foundation for selecting the best model according to the analysis and prediction needs desired in the context of this research.

**Evaluation**

The prediction results from each algorithm have been successfully obtained, and the models have been created for use in analyzing future stock availability. This research records the Mean Squared Error (MSE), Accuracy, and F1-Score for each algorithm used, and the results can be found in Table 6.

<table>
<thead>
<tr>
<th>Evaluation Algorithm Perform</th>
<th>XGBoost</th>
<th>SVR</th>
<th>Random Forest</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSE</td>
<td>7714.446</td>
<td>119.315</td>
<td>1142.393</td>
</tr>
<tr>
<td>F1-Score</td>
<td>0.92</td>
<td>0.92</td>
<td>0.89</td>
</tr>
<tr>
<td>Akurasi</td>
<td>0.86</td>
<td>0.85</td>
<td>0.81</td>
</tr>
</tbody>
</table>

After obtaining the evaluation results, it is evident that each algorithm provides different performance in predicting stock availability. XGBoost and SVR show relatively lower MSE and sufficiently high F1-Score and Accuracy values, while Random Forest yields a higher MSE but still maintains good F1-Score and Accuracy values. This evaluation provides further insight into the effectiveness of each algorithm in the context of predicting stock availability.

**CONCLUSION**

Based on the research results, the three machine learning algorithms—XGBoost, Support Vector Regression (SVR), and Random Forest—contribute differently to predicting stock availability at PT PRIMA FOOD. XGBoost and SVR stand out with optimal performance, indicated by low MSE, as well as high F1-Score and Accuracy values, signifying good prediction accuracy. Despite Random Forest having a higher MSE, it still demonstrates solid performance with adequate F1-Score and Accuracy values. This evaluation provides a comprehensive overview of the effectiveness of each algorithm in the context of stock prediction, offering significant benefits for decision-making at the business and industry levels. XGBoost, with its superior performance, emerges as the primary choice as a highly effective tool for improving inventory management performance at PT Primafood International Pasir Putih 2. Consequently, this research contributes not only to the advancement of knowledge in applying machine learning to inventory management but also opens up significant potential benefits for the development of the industry and society as a whole.

**REFERENCES**


Telecom Networks with time series data. In 2020 8th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO) (pp. 893–897). IEEE. doi:10.1109/ICRITO48877.2020.9197864


