Product Demand Analysis Using the XGBoost Algorithm at PT Atmadjaya Sembada Anugerah

[1]Okky Julian Atmajaya Tarmoko, [2]Rinabi Tanamal
Information System for Business, School of Information and Technology
Ciputra University Surabaya, Citraland CBD Boulevard, Indonesia, 60219
[1]Okkyjulian1202@gmail.com, [2]r.tanamal@ciputra.ac.id

Abstract— PT Atmadjaya Sembada Anugerah is a frozen food manufacturing company that faces challenges in stock management due to unpredictable daily fluctuations in product demand. Inaccurate demand forecasting can lead to inefficiencies in distribution and storage operations. This study aims to apply the Extreme Gradient Boosting (XGBoost) algorithm to forecast product demand using historical daily sales data. The process involves exploratory data analysis, data cleaning, feature engineering for time and statistical variables, and time-based data splitting. The model is trained using features selected through Recursive Feature Elimination and optimized using hyperparameter tuning with Optuna. Evaluation is conducted through TimeSeriesSplit cross-validation and assessed using three standard performance metrics. The results indicate that the model effectively captures seasonal patterns and general demand trends, although it remains limited in responding to sudden demand spikes. These findings support the use of XGBoost as a foundational approach for demand forecasting systems in stock planning, with potential for further improvement through the integration of external data and expanded feature sets.
.

Keywords— *XGBoost, Machine Learning, Demand Forecasting, Time Series, Sales Prediction, Historical Data, Stock Planning.*

I. INTRODUCTION

In the modern era, companies are required to respond quickly and accurately to market changes. One crucial approach that supports strategic decision-making is demand forecasting, which refers to the process of predicting future product demand based on historical data. Demand forecasting plays a significant role in aligning production processes and inventory management efficiently [1].

As the volume of data and complexity of demand patterns increase, conventional forecasting approaches are being replaced by machine learning algorithms. These technologies enable more accurate and adaptive modeling of demand patterns. One widely used algorithm in regression and time series forecasting is Extreme Gradient Boosting (XGBoost), which is known for its speed, efficiency, and predictive accuracy.

Previous studies have shown that XGBoost delivers superior performance compared to traditional methods. For example, [2] demonstrated the effectiveness of XGBoost in predicting retail sales with high accuracy based on historical data. Furthermore, research by [3] proved that XGBoost can outperform Long Short-Term Memory (LSTM) models in forecasting microgrid energy demand. Similarly, [4] found that XGBoost is also effective in forecasting pharmaceutical needs, especially when external features such as weather are included.

Moreover, ensemble learning approaches such as Random Forest have also shown competitive predictive performance. A study by [5] showed that the Random Forest model could predict house prices with up to 88% accuracy. These findings further reinforce the viability of the XGBoost algorithm as a promising approach for improving forecasting accuracy.

In this context, PT Atmadjaya Sembada Anugerah, a frozen food manufacturing company based in Sidoarjo, faces real challenges in stock management due to unpredictable daily demand fluctuations. Inaccurate forecasts can result in stockouts or overstocking, which in turn disrupt distribution and increase operational costs. Therefore, the company needs a data-driven solution to predict demand more accurately.

This study aims to implement the XGBoost algorithm for demand forecasting based on the daily sales data of PT Atmadjaya Sembada Anugerah. The predictive model is expected to assist the company in planning stock levels more efficiently and responding more effectively to dynamic market demand.

Previous studies have demonstrated the effectiveness of the XGBoost algorithm in improving demand prediction accuracy. One example is the study by Sukarsa, which applied XGBoost to forecast gourami fish supply in a restaurant. By combining features such as lag, rolling window, and mean encoding, the model achieved an accuracy of up to 97.54%, with an MAE of 0.63 and a MAPE of 2.64% [6]. Another study by [7] compared the performance of XGBoost with LSTM and ARIMA models in the context of retail sales forecasting. The results showed that XGBoost outperformed the others in capturing demand patterns based on historical data, making it more suitable for supporting data-driven operational decisions. Research by [2], using sales data from 45 Walmart stores, further supports this claim. The XGBoost model in their study captured complex patterns more accurately than linear regression, making it highly relevant for demand forecasting in the retail and logistics sectors.

Machine learning is an approach within artificial intelligence that enables systems to learn patterns from data and make predictions without being explicitly programmed. The learning process involves two main stages: training and testing, which aim to improve the model’s decision-making accuracy based on data (Roihan et al., 2020). One of the standout algorithms in this approach is Extreme Gradient Boosting (XGBoost), a tree-based ensemble method designed for efficient and accurate prediction. XGBoost excels at handling large-scale datasets, supports parallel processing and out-of-core computing, and offers flexibility in parameter tuning, making it well-suited for time-series-based demand forecasting [2].

Forecasting is the process of estimating future values based on historical trends. In a business context, forecasting plays a critical role in supporting decisions related to inventory management and production planning [9]. Several evaluation metrics are commonly used to assess prediction models, including Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Weighted Absolute Percentage Error (WAPE). MAE measures the average absolute difference between actual and predicted values, while RMSE penalizes larger errors more heavily due to squaring the residuals, lower RMSE values indicate more accurate predictions [10]. Meanwhile, WAPE calculates the total absolute error relative to the total actual value in percentage form. This metric is scale-independent, unaffected by zero values, and delivers stable results in fluctuating demand scenarios, making it suitable for demand forecasting in retail and digital sectors [11].

To enhance model performance, feature selection techniques such as Recursive Feature Elimination (RFE) are often used. RFE works by iteratively removing less relevant features until the best subset is found, which yields the most optimal prediction results [12]. Additionally, hyperparameter tuning plays a crucial role in building accurate predictive models. Optuna is an automated hyperparameter optimization framework that applies Bayesian optimization via the Tree-structured Parzen Estimator (TPE). This framework enables efficient and flexible parameter search, supporting various strategies such as grid search and random search [13].

II. METHOD

This study employs a quantitative approach by implementing the XGBoost algorithm to predict product demand based on daily historical sales data from PT Atmadjaya

Sembada Anugerah. The data were collected from the company's internal recording system, covering the period from January 2021 to December 2024. The dataset was analyzed through several main stages, including data exploration, data cleaning, feature engineering, modeling, and evaluation.

The initial stage involved Exploratory Data Analysis (EDA) to understand the structure, distribution, and potential anomalies in the data. This was followed by a data cleaning process, such as removing duplicate entries, rows with negative values, and extreme outliers. Additionally, transaction data were aggregated so that each date corresponds to a single daily demand entry.

The next step was feature engineering, where derived features were created to enrich the context of the data. Time-based features included: day of year, quarter, year, day of the week, and weekend indicators. Rolling statistics such as MA (Moving Average), STD (Standard Deviation), and Delta were used to capture trends and daily fluctuations. Sinusoidal transformations (sin_doy and cos_doy) were added to represent annual seasonality continuously. Moreover, national holidays and major events (e.g., Eid al-Fitr and Christmas) were encoded with a ±5-day window around the main date.

Once the features were prepared, the data were split using a time-based approach, with records before October 1, 2024 used for training and data from October 1 to December 31, 2024 for testing. This approach preserved the chronological order of the time series and avoided data leakage.

An initial model was built using XGBoost with default parameters. Feature selection was then performed using the Recursive Feature Elimination (RFE) method, followed by hyperparameter tuning with Optuna. The optimized parameters included n_estimators, learning_rate, max_depth, subsample, colsample_bytree, and gamma. The final model utilized the best parameter combination and the selected features.

Model validation was conducted using TimeSeriesSplit with five folds. Performance evaluation employed three main metrics: Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Weighted Absolute Percentage Error (WAPE).

A diagram outlining the overall research workflow is presented below to provide a general overview of the methodology used in this study:

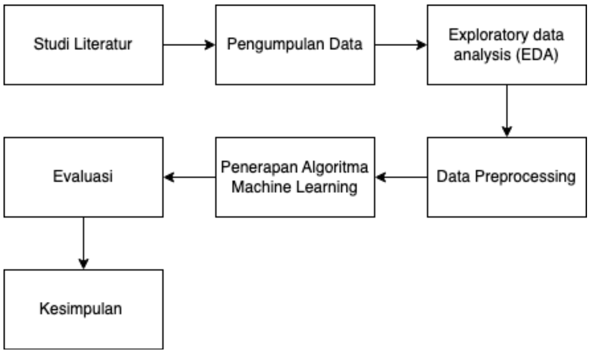

Figure 1 Workflow of the Demand Forecasting Methodology.

The implementation of this research was carried out using the Python programming language within the Jupyter Notebook environment. Several libraries were utilized, including Pandas for data manipulation, NumPy for numerical operations, Matplotlib and Seaborn for data visualization, and XGBoost and Optuna for modeling and hyperparameter tuning. All processes were executed on a MacBook Pro equipped with the Apple M3 Pro chip and macOS operating system.

The implementation began with the loading of the daily sales dataset obtained from the internal system of PT Atmadjaya Sembada Anugerah. Following an initial exploration, the data were cleaned by removing duplicates, missing values, and anomalies. Subsequently, feature engineering was conducted, which included the creation of time-based features, rolling statistics (MA, STD), daily value differences (Delta), and the encoding of national holidays and major events.

The dataset was split using a time-based approach, with data prior to October 2024 used for training, and data from October to December 2024 used for testing. An initial model was built using the XGBoost algorithm with default parameters. Feature selection was then performed using Recursive Feature Elimination (RFE) to identify the most relevant features for the prediction target.

Hyperparameter tuning was carried out using the Optuna library, with the search space including n_estimators, learning_rate, max_depth, subsample, colsample_bytree, and gamma. The best parameters obtained from the tuning process were then used to train the final model.

The following table presents the optimal hyperparameter values used in the final XGBoost model:

Table 1 Optimal Hyperparameters Obtained from Optuna Tuning

n_estimators	916
learning_rate	0.02362064267141174
max_depth	8
subsample	0.8583881123060892
colsample_bytree	0.9584689141642643
gamma	1.0512305644151259

III. RESULTS AND DISCUSSION

A. RESULT

The XGBoost model was evaluated using historical daily demand data from October to December 2024. The evaluation employed three primary performance metrics:

Table 2 Evaluation Results of the XGBoost Model.

MAE	72.76
RMSE	102.84
WAPE	15.19%

To visualize the model's performance, the predicted values were plotted alongside actual daily demand:

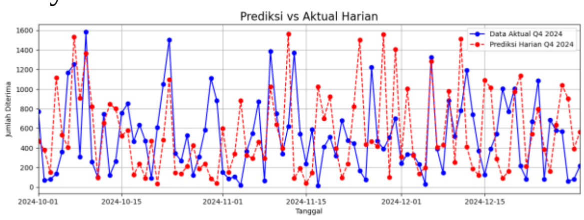

Figure 2 Workflow of the Demand Forecasting Methodology.

In addition to daily evaluation, the forecasts were aggregated weekly to observe overall trend stability across Q4 2024:

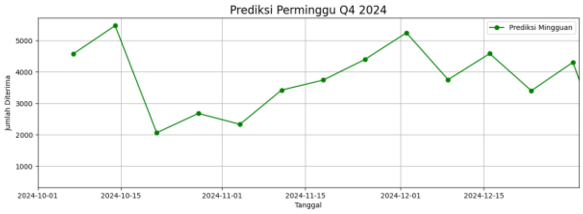

Figure 3 Workflow of the Demand Forecasting Methodology.

A combined visualization was also developed to display prediction accuracy across time:

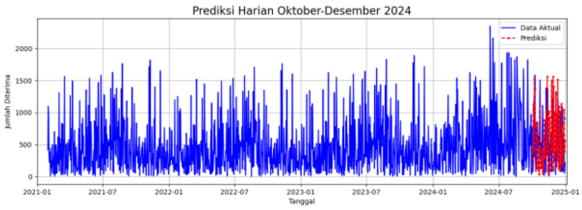

Figure 4 Workflow of the Demand Forecasting Methodology.

B. DISCUSSION

The evaluation results indicate that the XGBoost model achieved a WAPE of 15.19%, which is acceptable in the context of daily demand forecasting where fluctuations are often extreme. The MAE of 72.76 and RMSE of 102.84 suggest that the model's prediction errors are moderate and consistent with real-world applications.

From the visualizations, it is evident that the model successfully captured the general seasonal and temporal patterns, such as weekly trends and overall fluctuations. However, it exhibited noticeable deviations during sudden spikes, particularly following extended holidays.

These deviations imply that the model may lack contextual awareness of sudden demand shifts. This limitation suggests a potential area for improvement, such as incorporating external features e.g., promotional campaigns, weather data, or competitor behavior into future iterations of the model.

Overall, the results confirm that XGBoost is an effective algorithm for demand forecasting in the frozen food industry and can be integrated into operational decision-making processes like inventory planning and production scheduling. However, further refinement is necessary for it to handle unpredictable, high-impact events with greater sensitivity.

IV. CONCLUSION

This study successfully implemented the Extreme Gradient Boosting (XGBoost) algorithm, optimized with Optuna and Recursive Feature Elimination (RFE), to address the product demand forecasting challenges at PT Atmadjaya Sembada Anugerah. The final model demonstrated robust performance during the evaluation period (Q4 2024), achieving a Weighted Absolute Percentage Error (WAPE) of 15.19%, a Mean Absolute Error (MAE) of 72.76, and a Root Mean Squared Error (RMSE) of 102.84. These metrics indicate that the model provides a highly reliable and structurally sound baseline for daily demand prediction within the frozen food industry.

The empirical findings confirm that the inclusion of time-based features, rolling statistics, and continuous seasonal variables enables the XGBoost model to effectively learn complex temporal relationships and weekly demand patterns. Consequently, this system can serve as a data-driven foundation to assist management in optimizing stock distribution, mitigating overstocking or stockout risks, and improving operational efficiency.

Despite its strong performance in capturing general trends, the model exhibits a notable limitation in predicting sudden, volatile demand spikes, particularly those occurring immediately after extended holiday periods. This suggests that relying solely on historical transaction trends is insufficient for extreme event adaptation. For future research, it is highly recommended to expand the feature space by integrating external market factors, such as real-time promotional data, weather variations, and competitor dynamics, or to explore hybrid deep learning frameworks to further increase sensitivity toward sudden market fluctuations.

REFERENCES

[1] D. Karthika and K. Karthikeyan, "a Recent Review Article on Demand Forecasting," *Xi'an University of Architecture & Technology*, vol. 12, no. 3, pp. 5769–5777, 2020.

[2] Y. F. Akande, J. Idowu, A. Misra, S. Misra, O. N. Akande, and R. Ahuja, "Application of XGBoost Algorithm for Sales Forecasting Using Walmart Dataset," *Lecture Notes in Electrical Engineering*, vol. 881, no. August, pp. 147–159, 2022, doi: 10.1007/978-981-19-1111-8_13.

[3] M. Mohamed, F. E. Mahmood, M. A. Abd, M. Rezkallah, A. Hamadi, and A. Chandra, "Load Demand Forecasting Using eXtreme Gradient Boosting (XGboost)," *2023 IEEE Industry Applications Society Annual Meeting, IAS 2023*, no. December, 2023, doi: 10.1109/IAS54024.2023.10406613.

[4] Muhammad Dzul Asmi Alhamdi, Herman, and Wistiani Astuti, "Peramalan Kebutuhan Obat Menggunakan XGBoost Studi Kasus pada Rumah Sakit XYZ," *Indonesian Journal of Computer Science*, vol. 12, no. 5, pp. 2757–2764, 2023, doi: 10.33022/ijcs.v12i5.3344.

[5] R. Tanamal, N. Minoque, T. Wiradinata, Y. Soekamto, and T. Ratih, "House Price Prediction Model Using Random Forest in Surabaya City," *TEM Journal*, vol. 12, no. 1, pp. 126–132, 2023, doi: 10.18421/TEM121-17.

[6] I. M. Sukarsa, N. N. Pandika Pinata, N. Kadek Dwi Rusjayanthi, and N. W. Wisswani, "Estimation of Gourami Supplies Using Gradient Boosting Decision Tree Method of XGBoost," *TEM Journal*, vol. 10, no. 1, pp. 144–151, 2021, doi: 10.18421/TEM101-17.

[7] D. Swami, A. D. Shah, and S. K. B. Ray, "Predicting Future Sales of Retail Products using Machine Learning," pp. 1–6, 2020, [Online]. Available: http://arxiv.org/abs/2008.07779

[8] A. Roihan, P. A. Sunarya, and A. S. Rafika, "Pemanfaatan Machine Learning dalam Berbagai Bidang: Review paper," *IJCIT (Indonesian Journal on Computer and Information Technology)*, vol. 5, no. 1, pp. 75–82, 2020, doi: 10.31294/ijcit.v5i1.7951.

[9] S. S. Mahadik, I. Technology, and N. Mumbai, "A New AGE Forecasting Model PROPHET," no. June, 2023, doi: 10.55041/IJSREM15522.

[10] L. S. Ihzaniah, A. Setiawan, and R. W. N. Wijaya, "Perbandingan Kinerja Metode Regresi K-Nearest Neighbor dan Metode Regresi Linear Berganda pada Data Boston Housing," *Jambura Journal of Probability and Statistics*, vol. 4, no. 1, pp. 17–29, 2023, doi: 10.34312/jjps.v4i1.18948.

[11] J. Rombouts, M. Ternes, and I. Wilms, "Cross-temporal forecast reconciliation at digital platforms with machine learning," *Int J Forecast*, vol. 41, no. 1, pp. 321–344, 2024, doi: 10.1016/j.ijforecast.2024.05.008.

[12] A. M. Priyatno and T. Widiyaningtyas, "a Systematic Literature Review: Recursive Feature Elimination Algorithms," *JITK (Jurnal Ilmu Pengetahuan dan Teknologi Komputer)*, vol. 9, no. 2, pp. 196–207, 2024, doi: 10.33480/jitk.v9i2.5015.

[13] V. Sinap, "Improving Machine Failure Prediction with Grey Wolf, Whale Optimization, and Optuna Techniques," *Gazi University Journal of Science Part A Engineering and Innovation*, vol. 12, pp. 154–174, Mar. 2025, doi: 10.54287/gujsa.1544942.