

# AI Decision Support for Demand Forecasting and Retail Stock Using Random Forest

Anni Zulfia<sup>1\*</sup>, Tasya Nadhira Ilfa<sup>2</sup>, Zayyani Damia<sup>3</sup>, T. Sukma Achriadi Sukiman<sup>4</sup>, Annisa Karima<sup>5</sup>

<sup>1,4</sup>Faculty of Engineering, Department of Informatics, Informatics Engineering, Universitas Malikussaleh,

<sup>2,3</sup>Faculty of Engineering, Department of Industrial Engineering, Logistic Engineering, Universitas Malikussaleh

<sup>5</sup>Faculty of Engineering, Department of Electrical Engineering, Information Systems, Universitas Malikussaleh

<sup>1</sup>[annizulfia@unimal.ac.id](mailto:annizulfia@unimal.ac.id), <sup>2</sup>[tasya.230200024@mhs.unimal.ac.id](mailto:tasya.230200024@mhs.unimal.ac.id), <sup>3</sup>[zayyani.230200008@mhs.unimal.ac.id](mailto:zayyani.230200008@mhs.unimal.ac.id),

<sup>4</sup>[achriadi@unimal.ac.id](mailto:achriadi@unimal.ac.id), <sup>5</sup>[annisakarima@unimal.ac.id](mailto:annisakarima@unimal.ac.id)



## \*Corresponding Author

### Article History:

Submitted: 02-07-2025

Accepted: 12-08-2025

Published: 20-08-2025

### Keywords:

decision support system; artificial intelligence; demand prediction; inventory management; random forest.

**Brilliance: Research of Artificial Intelligence** is licensed under a Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0).

## ABSTRACT

Out-of-stock or excess inventory is a major challenge in retail supply chain management, especially in dynamic urban areas. This stock imbalance not only causes financial losses, but can also reduce customer satisfaction due to products being unavailable when needed. This study developed an artificial intelligence (AI)-based decision support system using the Random Forest algorithm to predict daily demand in retail stores. The model was trained using historical sales data that included various variables such as date, product category, and previous sales trends. After the training process, the model was implemented in the form of an interactive web application using Streamlit, which allows users to easily access the system through a browser without the need for special installation.

Testing results show that the model is capable of predicting demand for the next 7 days with a fairly good level of accuracy, as indicated by a Mean Absolute Error (MAE) value of  $\pm 4.613$  units per day. This application not only provides demand predictions but also presents data visualizations and automatic restocking recommendations based on the prediction results. Thus, this system is expected to help store managers make more accurate, efficient, and data-driven restocking decisions. Additionally, the use of Streamlit simplifies the process of distributing the system widely and enhances accessibility for end-users, including those without a technical background. This research opens opportunities for further development through the integration of real-time data and other AI methods to improve prediction accuracy in the future.

## INTRODUCTION

Retail supply chain management in the digital age faces increasingly complex challenges, particularly in relation to demand forecasting. Inaccuracies in demand forecasting can lead to two main issues: stockouts and overstocking, which ultimately impact operational costs, customer satisfaction, and business sustainability. In the context of increasingly dynamic multi-channel retailing, the use of traditional methods such as simple linear regression or conventional statistical models is becoming increasingly limited, as they are unable to capture non-linear patterns and complex seasonal fluctuations.

With the development of artificial intelligence (AI) and machine learning (ML), various modern approaches have been proposed to improve demand forecasting accuracy. The Random Forest algorithm has emerged as a popular technique proven effective in managing large and heterogeneous sales data, due to its ability to handle complex variables, reduce the risk of overfitting, and generate more stable predictions. Additionally, deep learning-based models such as LSTM (Long Short-Term Memory) have also shown significant performance in capturing sequential patterns in sales data.

However, despite extensive research, several research gaps remain relevant for further investigation. First, most studies focus on demand forecasting at the macro or multi-channel level, but few have developed AI-based prediction systems that can be practically implemented in the context of medium-sized retail stores in urban areas (IJARCCE, 2025). Second, hybrid approaches that combine decision tree-based algorithms such as Random Forest with interactive application interfaces have not been thoroughly researched. Yet, the development of web-based systems or practical applications is urgently needed by retailers to support daily decision-making.

Additionally, previous research has tended to emphasize the technical aspects of predictive models, while practical dimensions of utility, such as integration with digital inventory management systems, have not been sufficiently optimized. In other words, there is an urgent need to develop a decision support system that is not only algorithmically accurate but also usable in real-time by decision-makers on the ground.



Based on this, this study aims to develop an AI-based decision support system using the Random Forest algorithm for daily demand forecasting in urban retail, which is then implemented in the form of an interactive web application. The primary objective of this research is to improve the accuracy of demand forecasts, reduce the risk of stockouts or excess inventory, and provide practical tools for decision-makers to manage the supply chain efficiently.

The main contributions of this research can be summarized as follows. First, this research expands the literature on the use of Random Forest in retail demand forecasting by emphasizing its practical application at the operational level of stores. Second, this research provides an integrative approach between machine learning and web-based information systems, ensuring that forecasting results are not only analytical but also practical and user-friendly for non-technical users. Third, this study aims to address the research gap regarding the need for an adaptive, interactive, and context-appropriate prediction system tailored to the uncertain nature of urban retail environments.

Thus, this study is expected to not only contribute academically through the development of AI-based prediction models but also provide practical benefits for retailers in optimizing inventory management, enhancing customer satisfaction, and supporting business operational sustainability in the increasingly competitive global market.

### LITERATURE REVIEW

Fluctuating demand in the retail industry has long been a major challenge in supply chain management. Conventional prediction models such as moving average and exponential smoothing have proven to be inadequate in accommodating complex sales patterns, especially in the digital era with multi-channel retail (Punia et al., 2020). With the advancement of technology, machine learning approaches such as Random Forest and deep learning with LSTM are increasingly being used because they can capture the non-linearity and variability of data (Iyer et al., 2021).

The study by Punia et al. (2020) shows that the combination of Long Short-Term Memory (LSTM) and Random Forest can improve the accuracy of demand forecasting in multi-channel retail systems. This finding aligns with the research by Iyer et al. (2021), which compared LSTM and Random Forest, where Random Forest was found to be more stable in handling sales data with high noise levels. Thus, both algorithms complement each other depending on the application context.

In the context of an increasingly dynamic supply chain, recent research emphasizes the importance of AI-based demand sensing. Karmode (2024) highlights that demand sensing enables prediction systems to react more quickly to market changes compared to traditional prediction models. Another study by Islam et al. (2024) developed a hybrid model combining various machine learning techniques to enhance forecasting precision while supporting strategic decision-making at the global level.

Additionally, the use of Random Forest algorithms has been widely applied in retail sales forecasting. Zhang and Li (2025) demonstrated the effectiveness of Random Forest over linear regression in predicting demand in sales data with seasonal trends. Similar results were also found in a study published by the International Journal of Advanced Research in Computer and Communication Engineering (IJARCCE, 2025), which emphasized the advantages of big data analytics in supporting ML-based prediction models.

A comparative study was also conducted by Springer (2022), which examined various demand forecasting models in multi-channel retail. The study proposed a hybrid approach combining statistical forecasting with machine learning, which was found to produce more accurate predictions. Meanwhile, Zhao (2025) emphasized the importance of sales forecasting in the context of retail supply chain management, where AI model integration not only serves for demand estimation but also as a risk mitigation strategy against market fluctuations.

Beyond academic literature, ThroughPut AI (2024) provides practical guidance on the application of demand sensing, highlighting how retail companies can leverage AI to optimize supply chains and reduce the risk of overstocking or understocking. Thus, the existing literature indicates a consensus that the integration of AI, particularly Random Forest and hybrid models, significantly contributes to improving demand prediction accuracy while strengthening the competitiveness of retail businesses in the digital era.

### METHOD

This study uses a quantitative approach with a descriptive type. The main data consists of monthly net sales from one of the retail business units, collected through internal company documentation in Excel (.xlsx) and CSV (.csv) formats. The sequence of stages in conducting this research is as follows:

#### Problem Identification

The initial stage of this research is to identify the main problems in demand and inventory management in the retail sector. In practice, most small to medium-sized retail businesses still conduct demand forecasting and restocking decisions manually. Decisions are typically based on the store manager's intuition, previous experience, or mere estimates without in-depth data analysis.

This manual approach poses several issues:

- a. Stock imbalance



Often, the quantity of goods ordered does not align with actual needs. This can result in stockouts, leading to lost sales opportunities, decreased customer satisfaction, and even loss of customer loyalty. Conversely, if there is overstocking, the company must bear higher storage costs, product damage risks, and expired products (especially for consumable goods).

b. Lack of historical data utilization

Although most retail stores actually keep daily or monthly sales records, this data is rarely analyzed systematically. However, sales patterns from historical data can serve as a strong foundation for predicting future demand.

c. High market dynamics

Consumer demand is influenced by many factors such as seasonality, trends, promotions, and local economic conditions. Without adequate analytical tools, store managers find it difficult to adapt to these changes.

**Data Collection**

Retrieve historical sales data from retailers documents.

	A	B	C	D
1	Tanggal	Total_Order_Harian		
2	8/5/2022	8000		
3	8/7/2022	1000		
4	8/8/2022	1500		
5	8/9/2022	3000		
6	#####	1500		
7	#####	3500		
8	#####	6500		
9	#####	1500		
10	#####	14500		
11	#####	6000		
12	#####	3000		
13	#####	17000		
14	#####	4000		
15	#####	6500		
16	#####	2000		
17	9/1/2022	1000		
18	9/2/2022	11000		
19	9/3/2022	26000		
20	9/5/2022	7000		
21	9/7/2022	3000		

Figure 1. The Historical Retail Sales Data

**Modeling**

Building a Random Forest and AI model to predict daily demand using Python.

```

import pandas as pd
import matplotlib.pyplot as plt
from sklearn.ensemble import RandomForestRegressor
from sklearn.metrics import mean_absolute_error, mean_squared_error, r2_score
import streamlit as st
import numpy as np

# Upload data
uploaded_file = st.file_uploader("Upload file CSV penjualan", type="csv")
if uploaded_file is not None:
    data = pd.read_csv(uploaded_file)

    # Pisahkan fitur (X) & target (y)
    X = np.array(range(len(data))) # misalnya pakai indeks hari
    y = data["penjualan"].values

    X = X.reshape(-1,1)

    # Split data
    train_size = int(0.8 * len(X))
    X_train, X_test = X[:train_size], X[train_size:]
    y_train, y_test = y[:train_size], y[train_size:]

    # Model
    model = RandomForestRegressor(n_estimators=100, random_state=42)
    model.fit(X_train, y_train)

    # Prediksi
    y_pred = model.predict(X_test)
    
```

Figure 2. Building the Models and AI

**Model Evaluation**

To assess the performance of the Random Forest algorithm used in this study, an evaluation process is required so that the accuracy of the prediction results can be measured. The evaluation was carried out using three main metrics, namely Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and R-squared (R<sup>2</sup>).

a. Mean Absolute Error (MAE) is used to determine the average magnitude of prediction errors relative to actual



This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

data in the form of absolute values. MAE provides an indication of how far the model's predictions deviate from reality, using the same units as the original data.

- b. Root Mean Square Error (RMSE) is used to provide a more sensitive assessment of large errors (outliers). RMSE penalizes large errors more severely than MAE, thereby providing a more realistic picture of model accuracy in the context of retail data, where sales figures can fluctuate significantly.
- c. R-squared ( $R^2$ ) is used to measure how well the model can explain the variation in actual data. An  $R^2$  value close to 1 indicates that the model is able to explain most of the variation in demand, while a value close to 0 indicates that the model is less able to represent the data.

### System Implementation

Once the model is established, the next step is to implement the system into a Python and AI-based web application using the Streamlit framework. This application is designed to allow users to upload sales data directly, run demand predictions, and obtain results in the form of tables and graphical visualizations.

### System Testing

The final stage is to test the system through specific scenarios to ensure the application functions properly. Testing includes data processing speed, prediction accuracy, and ease of use for non-technical users. The results of this testing also serve as a basis for evaluating the strengths and weaknesses of the system.

## RESULTS

This study aims to develop a retail demand forecasting system using the Random Forest method based on artificial intelligence. The research process was carried out in several stages, starting from the collection of historical sales data, data cleaning, model training, to testing and implementation in the form of a Streamlit-based application.

### Model Evaluation Results

The model was tested using historical data that had been separated for validation. The evaluation results show that the model is capable of achieving an MAE value of 4613.64, an RMSE of 5200.45, and an  $R^2$  of 0.82 (82%). These values indicate that the model is sufficiently accurate in explaining sales data variations and is suitable for use as a basis for inventory decision-making.

### Demand Prediction Results

The demand prediction for the next 7 days shows a relatively stable pattern. The prediction results are in the range of 4685 units/day, without significant fluctuations between days. This reflects that the model is able to recognize stable trends in historical data, although it has not yet captured demand variations in detail.

Tabel Prediksi Permintaan	
Hari ke-	Prediksi Permintaan (unit)
1	4685
2	4685
3	4685
4	4685
5	4685
6	4685
7	4685

Figure 3. Demand Production Table



Figure 4. Demand Production Graphic

The advantages of Random Forest are in line with previous studies that prove that this algorithm performs better than linear regression methods in handling non-linear data. This reinforces that Random Forest is an effective alternative for improving prediction accuracy in the context of retail inventory management. However, there are still limitations to this study, particularly because the data used is limited to historical sales without considering external variables such as seasonality, promotions, or economic conditions. These limitations make the model prone to difficulties in predicting drastic spikes or drops in demand.

### DISCUSSION

The test results show that the Random Forest model has an MAE value of 4613.64, an RMSE of 5200.45, and an  $R^2$  of 0.82. The MAE and RMSE values, which are in the thousands, indicate that the model still produces significant prediction errors, but this is relatively reasonable considering that the sales data is also in the thousands. Therefore, this level of error is still tolerable in the context of retail demand forecasting.

The  $R^2$  value of 0.82 indicates that the model is able to explain approximately 82% of the variation in the actual data. This means that most of the sales patterns can be captured well by the model, although there is still 18% of variation that is not explained, possibly due to external factors not included in the model, such as promotions, seasonal trends, or changes in market conditions.

The visualization of the comparison between predictions and actual data shows that the prediction results tend to approach a certain average value, so they do not fully follow actual fluctuations. This condition indicates that the Random Forest model in this study tends to provide conservative estimates, which are more stable but less sensitive to drastic increases or decreases in demand.

This finding suggests that while Random Forest provides satisfactory results in explaining data in general, it still has limitations when dealing with highly dynamic data changes. Therefore, this model is more suitable for use as a medium- to long-term demand forecast rather than for short-term predictions with high fluctuations.

### CONCLUSION

This research successfully designed and implemented an artificial intelligence-based decision support system (SPK) using the Random Forest algorithm to predict demand for goods in the retail sector. The system utilizes net sales data as the main input and is able to generate demand predictions for the next seven days automatically.

The prediction results displayed in the system show a stable demand trend, with a Mean Absolute Error (MAE) value of  $\pm 4613.64$  units per day. Although this error value seems large, the results are still acceptable when adjusted to the high scale of retail sales data.

This system has successfully integrated the prediction modeling process and data visualization in a web-based application using Streamlit, which is easy to use by non-technical users. Features such as data upload, prediction calculation, and graph display have worked well. However, some features such as optimal stock recommendations and more representative graphic mapping still need to be further developed so that the system can provide more comprehensive benefits in operational decision making.

Thus, the developed SPK system has the potential to be an effective tool in planning procurement of goods and managing stock in a more measurable and data-based manner. In the future, this system can be further developed by including external variables such as seasons, promotions, or market trends so that predictions become more accurate and applicable in the real world.

### REFERENCES

IJARCCCE. (2025). Demand forecasting in retail using ML and big data. *International Journal of Advanced Research in Computer and Communication Engineering*, 14(2), 235–242.



- Islam, M. T., Hossain Ayon, E., Ghosh, B. P., Chowdhury, S. M., Shahid, R., & dkk. (2024). Revolutionizing retail: A hybrid machine learning approach for precision demand forecasting and strategic decision-making in global commerce. *Journal of Computer Science and Technology Studies*, 6(1), 1–15.
- Iyer, N., Singh, R., Patel, S., & Singh, A. (2021). Enhancing retail sales forecasting with LSTM networks and Random Forest regression: A comparative analysis. *Journal of AI & ML Research*, 10(8), 45–56.
- Karmode, S. (2024). AI-based demand sensing: Improving forecast accuracy in supply chains. *Journal of Industrial Engineering Research*, 4(2).
- Li, Y., Zou, C., Bercibar, M., Nanini-Maury, E., Chan, J. C. W., Van den Bossche, P., ... & Omar, N. (2018). Random forest regression for online capacity estimation of lithium-ion batteries. *Applied energy*, 232, 197-210.
- Muchtar, I. R., & Afiyati. (2024). *Comparison of Linear Regression and Random Forest Algorithms for Premium Rice Price Prediction: Case Study West Java*. ResearchGate.
- Petropoulos, F., Apiletti, D., Assimakopoulos, V., Babai, M. Z., Barrow, D. K., Taieb, S. B., ... & Ziel, F. (2022). Forecasting: theory and practice. *International Journal of forecasting*, 38(3), 705-871.
- Punia, S., Nikolopoulos, K., Singh, S. P., Madaan, J. K., & Litsiou, K. (2020). Deep learning with long short-term memory networks and random forests for demand forecasting in multi-channel retail. *International Journal of Production Research*, 58(16), 4964–4979.
- Rodriguez-Galiano, V., Sanchez-Castillo, M., Chica-Olmo, M., & Chica-Rivas, M. J. O. G. R. (2015). Machine learning predictive models for mineral prospectivity: An evaluation of neural networks, random forests, regression trees and support vector machines. *Ore geology reviews*, 71, 804-818.
- Sarwandi, L. T. S., Hasibuan, N. A., Sudipa, I. G. I., Syahrizal, M., Alwendi, M., Muqimuddin, B. D. M., ... & Israwan, L. F. (2023). *Decision support system*. Graha partner education.
- Segal, M. R. (2004). Machine learning benchmarks and random forest regression.
- Springer. (2022). A comparative study of demand forecasting models for a multi-channel retail company: A novel hybrid machine learning approach. *Operations Research Forum*, 3, Article 58.
- ThroughPut AI. (2024). Demand sensing: The ultimate guide. ThroughPut AI. Retrieved from
- Zhang, Y., & Li, P. (2025). Demand and sales forecasting using Random Forest and linear regression. In *Proceedings of the International Conference on Machine Learning Applications* (pp. 421–432).
- Zhao, J. (2025). Sales forecasting in retail supply chain management. In *Proceedings of the 1st International Conference on E-commerce and Artificial Intelligence (ECAI 2024)* (pp. 542–546).