Advanced International Journal for Research (AIJFR)
E-ISSN: 3048-7641 ● Website: www.aijfr.com ● Email: editor@aijfr.com

Fashion Inventory Management and Demand Forecasting Using Gradient Boosted Trees

Mr. N.M Shahane[1], Asma Fatima[2], Shreya Karode[3], Nikita Jadhav[4], Shrushti Jadhav[5]

[1]Associate Professor, Department of Computer Engineering, K.K Wagh Institute of Engineering Education & Research, Nashik, Maharashtra
[2,3,4,5]Student, Department of Computer Science & Design, K.K Wagh Institute of Engineering Education & Research, Nashik, Maharashtra

Abstract:
Efficient inventory management and accurate demand forecasting are persistent challenges in the fashion retail industry. Traditional forecasting methods often fail to account for the complex interplay between product attributes, and customer preferences, leading to issues such as overstocking, stockouts, and missed sales opportunities. These inefficiencies increase operational costs and negatively impact customer satisfaction. This project presents a machine learning–based framework to predict monthly product demand by analyzing historical sales data and product attributes. The proposed system applies Gradient Boosted Trees (LightGBM) to predict demand based on product attributes including type, color, and category. By forecasting sales volumes at the product and attribute level, the approach provides retailers with actionable insights to recognize high-demand products, optimize inventory levels, and improve decision-making. The results demonstrate enhanced predictive accuracy compared to conventional methods, enabling more efficient and customer-centric inventory management.

Keywords: Inventory Management, Demand Forecasting, Machine Learning, Light- GBM, Feature Engineering, Retail Analytics, Product Attribute Analysis, Sales Prediction

1. Introduction:

In the fashion industry, predicting what customers will buy is always a challenge. Trends shift quickly, seasons influence buying behaviour, and consumer preferences keep evolving. Traditional forecasting methods usually depend on simple statistics or human judgement, which often fail to capture the deeper patterns in sales data—such as the impact of colour, style, or price on product demand.
Because of these limitations, retailers often face issues like overstocking or stock-outs, leading to financial losses and unhappy customers. This makes it difficult for them to make confident and accurate inventory decisions.

To overcome these challenges, our project introduces a machine learning–based demand prediction system using LightGBM. By analysing historical sales data along with key product attributes, the system delivers more accurate demand forecasts. This helps retailers maintain the right inventory levels, reduce waste, and improve overall operational efficiency.

2. Literature Review:

In recent years, the fashion retail industry has increasingly turned to machine learning and big data analytics to improve inventory management and demand forecasting. Researchers have explored advanced techniques such as neural networks, ensemble models, and hybrid deep learning systems to enhance sales prediction accuracy, reduce overstocking and stock-outs, and streamline supply chain operations. Together, these developments form a strong foundation for building intelligent, data-driven inventory systems that align product availability more closely with customer demand.

Author: Mahya Seyedan & Fereshteh Mafakheri [1]

Publication: Journal of Big Data, 2020, Vol. 7(1) — "Predictive Big Data Analytics for Supply Chain Demand Forecasting: Methods, Applications, and Research Opportunities."

Summary: The authors examined a wide range of predictive analytics methods used in supply chain forecasting. They emphasized how big data tools can significantly improve accuracy and responsiveness. Their research showed that combining real-time data with predictive algorithms helps businesses manage demand volatility and make better decisions in fast-changing markets.

Author: İlker Güven & Fuat Şimşir [2]

Publication: Computers & Industrial Engineering, 2020, Vol. 147 — "Demand Forecasting with Color Parameter in Retail Apparel Industry Using Artificial Neural Networks and Support Vector Machines."

Summary: This study introduced a forecasting model that considered visual features—specifically color—when predicting apparel sales. Using ANN and SVM models, the authors achieved higher accuracy than traditional statistical methods. Their findings highlight the significant role of aesthetic attributes in shaping fashion demand.

Author: Chandadevi Giri & Yan Chen [3]

Publication: Forecasting, 2022, Vol. 4(2) — "Deep Learning for Demand Forecasting in the Fashion and Apparel Retail Industry."

Summary: The researchers developed deep learning models using RNNs and CNNs to forecast demand based on time-series sales data. Their approach captured both temporal trends and product-level features, resulting in strong and adaptive performance for forecasting in a rapidly changing retail environment.

Author: M.S. Sousa, A.L.D. Loureiro, & V.L. Migueis [4]

Publication: Expert Systems with Applications, 2024 — "Predicting Demand for New Products in Fashion Retailing Using Censored Data."

Summary: This paper tackled the challenge of forecasting demand for newly launched products, which often lack historical sales data. Using censored regression models, the authors were able to predict early product demand more accurately, helping retailers reduce forecast bias and make more informed inventory decisions for new items.

Author: T. Zhou [5]

Publication: arXiv preprint arXiv:2306.11245, 2023 — "Improved Sales Forecasting Using Trend and Seasonality Decomposition with LightGBM."

Summary: The author proposed a hybrid model that applies trend and seasonality decomposition along with LightGBM. This combination helped capture nonlinear and seasonal patterns in sales data, outperforming traditional forecasting techniques. The study highlights LightGBM's effectiveness in managing the complexities of retail demand forecasting.

3. Materials and Methods:

In recent years, machine learning has become an essential part of modern demand forecasting systems. The proposed fashion demand forecasting and inventory optimization system follows a structured, data-driven workflow that includes data collection, preprocessing, model training, and decision-support logic. The overall process is shown in the block diagram in Figure 1.

The workflow begins by collecting data from the H&M fashion dataset, which contains transaction records, product information, and customer details. These datasets are combined into a unified format and passed through the preprocessing module. This module handles tasks such as filling in missing values, encoding categorical variables, and generating useful time-based features.

Once the data is cleaned, it is fed into the Light Gradient Boosting Machine (LightGBM) model. Using product attributes and temporal features, the model performs regression-based demand forecasting. The predicted demand values are saved in the backend database and simultaneously passed to the inventory management module. This component evaluates current stock levels, identifies risks such as potential overstocking or stock-outs, and suggests appropriate replenishment actions.

Finally, a Flask-based dashboard retrieves the stored predictions and presents them to managers and analysts in a visual and easy-to-understand format. It displays insights such as demand trends, feature importance, and inventory recommendations. Together, these components form an intelligent, end-to-end pipeline that enhances decision-making in fashion retail and enables more accurate, data-driven forecasting.

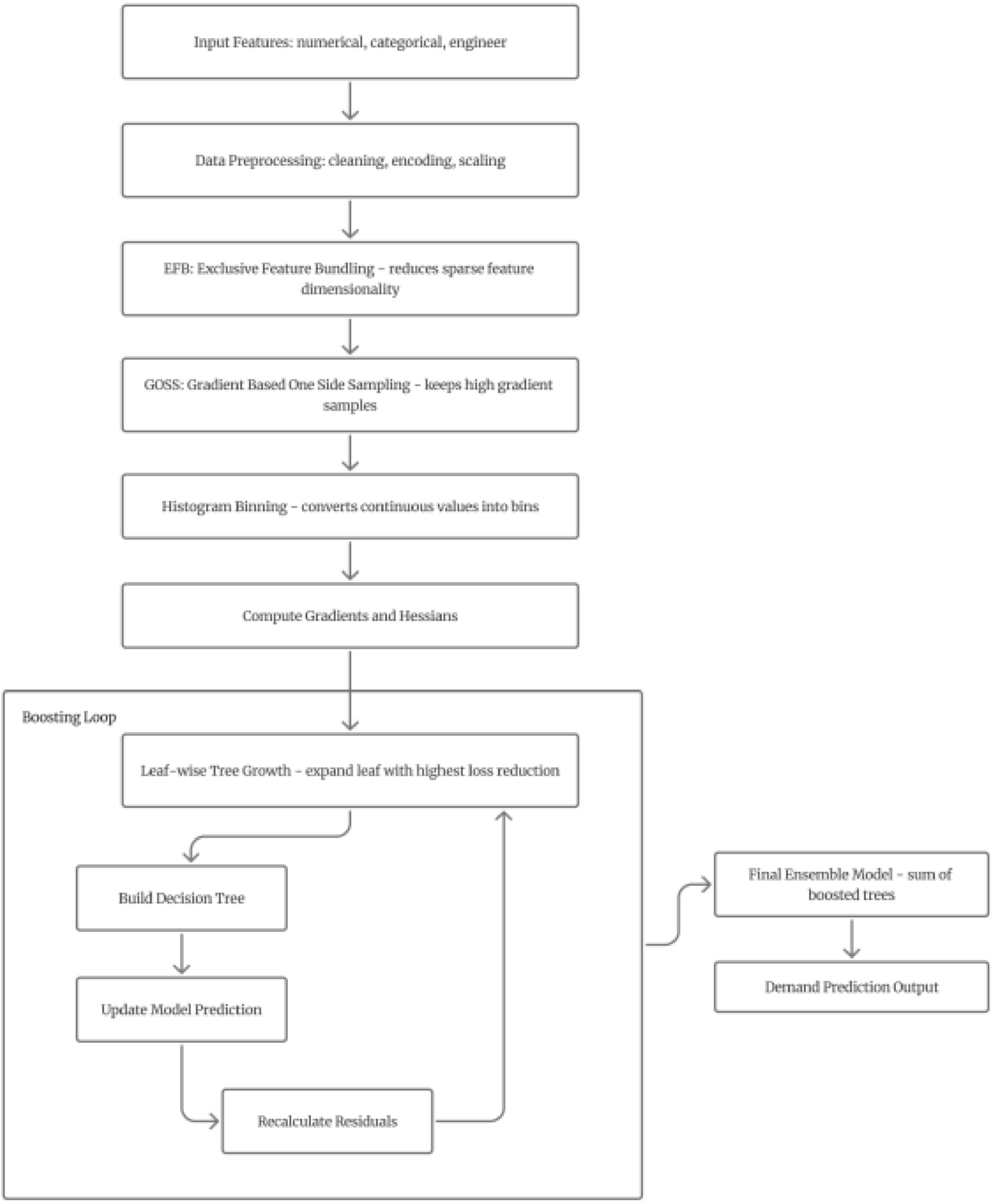

Figure 1: System Architecture

Method of Operation:

The proposed system operates through an automated predictive modelling pipeline supported by a well-structured data flow. It begins by importing the raw datasets—customers, articles, and transactions—and merging them using unique identifiers. Once the data enters the processing stage, the pre-processing module addresses missing values, standardizes fields, encodes categorical attributes, and aggregates monthly sales data for each product.

The engineered dataset is then passed to the LightGBM model, which learns complex relationships between various features such as product type, color, price, season, and historical sales. After training, the model predicts the monthly demand for each article. These predictions are sent to the inventory logic module, where they are compared with the current stock levels stored in the database.

If the expected demand is higher than the available stock, the system flags the product for replenishment. If the predicted demand is lower, it warns of potential overstock. All forecast results and inventory decisions are stored in a central PostgreSQL database.

A real-time dashboard retrieves this data through API calls and presents it to users through interactive charts, tables, and filters. This creates an end-to-end loop—from data ingestion to forecasting to inventory decision-making—reducing the need for manual intervention and enabling more efficient inventory management based on predicted demand.

Traditional retail inventory decisions often rely on intuition or simple rule-based methods, which can lead to overstocking, stock-outs, and missed sales opportunities. These approaches fail to account for demand fluctuations influenced by product attributes, seasonal changes, and shifting consumer behaviour. In contrast, the proposed machine-learning-driven system provides a more accurate, efficient, and data-driven approach to demand forecasting and inventory optimization in the fashion retail industry.

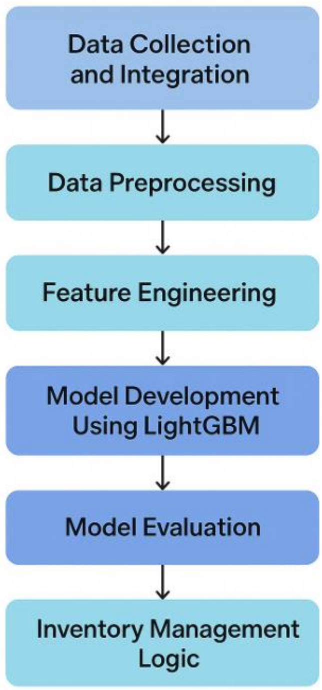

Figure 2: Methodology Flow for Demand Forecasting and Inventory Optimization

4. Result:

The project makes use of the **H&M Personalized Fashion Recommendation Dataset** from Kaggle, a large-scale dataset that connects customers, products, and their purchase histories. It is divided into three major files:

• **customers.csv**: Contains about 1.37 million customer records, including demographic information and membership details.
• **articles.csv**: Includes around 105,000 unique fashion items with attributes like product type, color, department, and price.
• **transactions train.csv**: Consists of more than 317 million purchase entries, each providing details such as the transaction date, product ID, customer ID, and price.
After cleaning, filtering, and merging the raw files, a refined working dataset of roughly **50,000 records with 35 attributes** was prepared for model development. The preprocessing phase involved handling missing values, encoding categorical variables, and aggregating monthly sales for each article. The key feature groups used were:
• **Temporal**: year, month, week, weekend.
• **Categorical**: product type, color, garment category.
• **Numerical:** price, age, membership duration.
• **Target Variable**: monthly sales (units per article per month)

Performance Parameters:
The performance of the LightGBM model was assessed using standard regression metrics:

• **Root Mean Squared Error (RMSE):** Shows how much prediction errors deviate from the actual values.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

• **Mean Absolute Error (MAE):** Represents the average difference between predicted demand and actual demand.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|$$

• **Coefficient of Determination (R2):** Measures how well the model captures and explains the variability in the data.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

Where,

- y_i= actual demand value
- $\hat{y}_i$= predicted demand value
- n= number of samples

Table 1: Model Performance Metrics

Metric	Value
Root Mean Squared Error (RMSE)	1.4065
Mean Absolute Error (MAE)	0.2471
Coefficient of Determination (R2)	0.9661

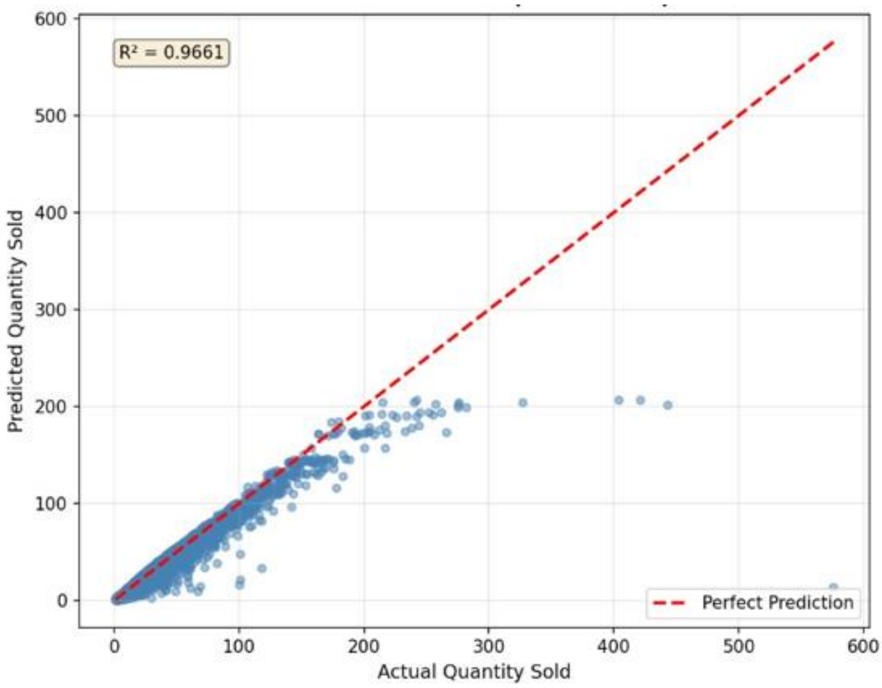

Figure 3: Actual v/s Predicted (Scatter Plot)

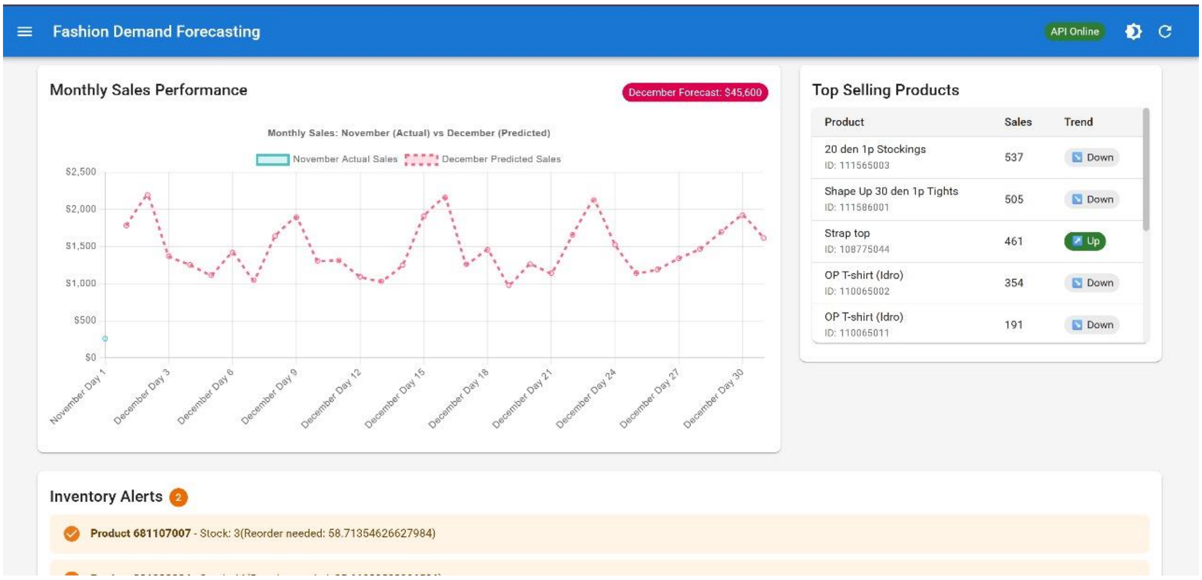

Figure 4: Demand Forecasting Dashboard

Advanced International Journal for Research (AIJFR)
E-ISSN: 3048-7641 ● Website: www.aijfr.com ● Email: editor@aijfr.com

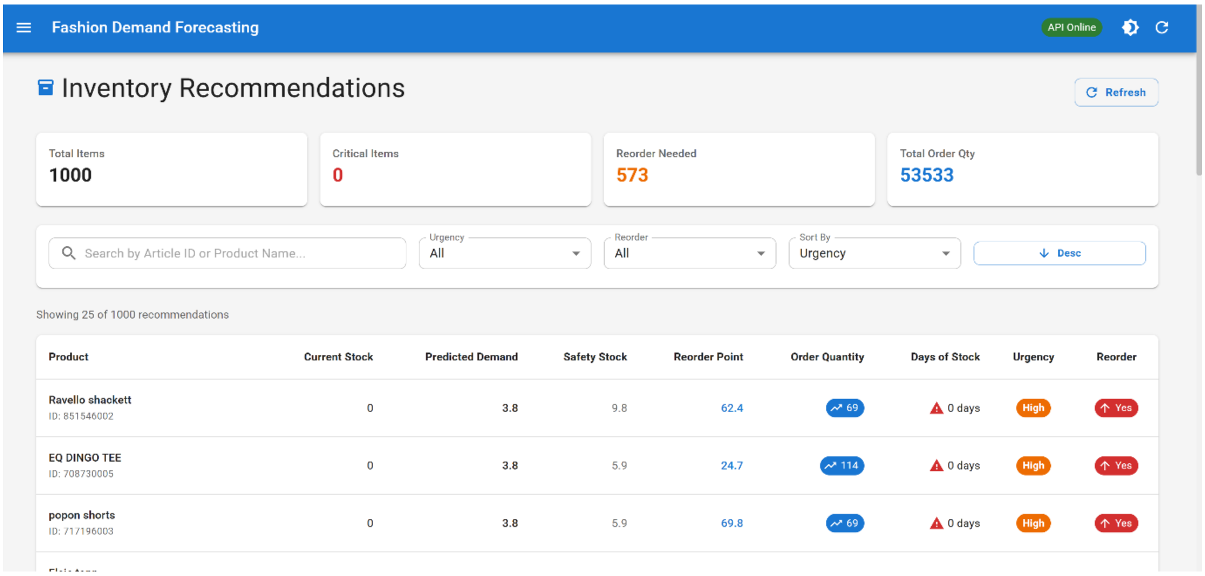

Figure 5: Inventory Management

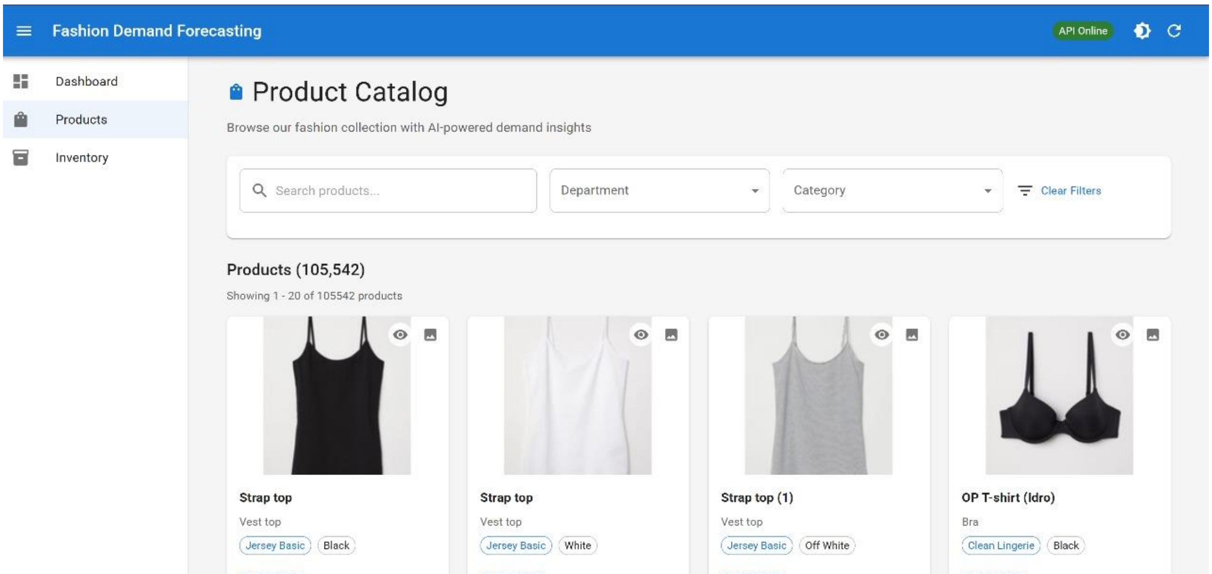

Figure 6: Product Catalog

5. Discussion:

The proposed LightGBM-based system provides an effective way to forecast monthly fashion demand by combining product attributes with customer-related information. The model delivers strong and consistent performance, closely reflecting real sales patterns—especially for products affected by seasonal trends. By capturing a wide range of features such as color, price, and category, the system achieves higher accuracy and better interpretability than traditional manual forecasting methods.

With support from a PostgreSQL backend and an interactive dashboard, retailers can monitor real-time demand, spot potential stock risks like overstocking or stock-outs, and understand which product attributes influence demand the most. Overall, this system offers a reliable, data-driven solution that helps retailers make smarter purchasing decisions, reduce inventory costs, and avoid revenue losses resulting from inaccurate demand forecasting.

6. Conclusion:

The proposed fashion demand forecasting and inventory optimization system has been successfully developed using machine learning and structured retail data. By combining product attributes, customer information, and historical sales trends, the system provides a reliable way to predict monthly product demand in the fashion retail industry. Powered by the Light Gradient Boosting Machine (LightGBM) algorithm, the model handles complex, high-dimensional data efficiently while delivering strong and consistent forecasting performance.

The integrated dashboard further enhances the system by allowing retailers to monitor product performance, view real-time predictions, and make informed stocking decisions. With interactive charts, error metrics, and feature-importance insights, the platform promotes transparency and supports data-driven decision-making.

Overall, this solution offers an accurate, scalable, and cost-effective framework that helps retailers optimize inventory levels, reduce losses from poor stock management, and improve overall operational efficiency in fashion retail

References:

1. M. Seyedan and F. Mafakheri, "Predictive Big Data Analytics for Supply Chain Demand Forecasting: Methods, Applications, and Research Opportunities," Journal of Big Data, vol. 7, no. 1, 2020. https://doi.org/10.1186/s40537-020-00329-2.
2. I. Güven and F. Şimşir, "Demand Forecasting with Color Parameter in Retail Apparel Industry Using Artificial Neural Networks and Support Vector Machines," Computers & Industrial Engineering, vol. 147, 2020. https://doi.org/10.1016/j.cie.2020.106669.
3. C. Giri and Y. Chen, "Deep Learning for Demand Forecasting in the Fashion and Apparel Retail Industry," Forecasting, vol. 4, no. 2, pp. 565–581, 2022. https://doi.org/10.3390/forecast4020031.
4. M. Abreu et al., "Predicting Demand for New Products in Fashion Retailing Using Censored Data," Expert Systems with Applications, 2024, p. 125313. https://doi.org/10.1016/j.eswa.2024.125313.
5. T. Zhou et al., "Improved Sales Forecasting Using Trend and Seasonality Decomposition with LightGBM," arXiv preprint arXiv:2306.11245, 2023.
6. H&M Group, "H&M Personalized Fashion Recommendations Dataset," Kaggle, 2023. Available at: https://www.kaggle.com/competitions/h-and-m-personalized-fashion-recommendations.