

E-commerce Return Rate Prediction Dashboard

Abstract

The focus of the project is the analysis and prediction of the return behavior in the e-commerce platform. With data preprocessed using Python and visualized using Power BI, some of the major factors of influence identified are Product category, Payment methods, Shipping type, and Customer demographics. The Logistic Regression model was derived to predict the probability of returning a product for proactive decision-making and operational efficiency. The final Power BI dashboard shows actionable insights into return trends, customer behavior, and category-wise performance that help identify ways for stakeholders to reduce return rates and increase business profitability.

Introduction

Product returns are inevitable in the e-commerce industry but, by all standards, very costly. High levels of return may affect inventory management, logistics expenses, and customer satisfaction. Understanding why customers return products, which categories, or demographics contribute most will make it possible for companies to improve policies and operations. This project brings together data analysis and predictive modeling to identify crucial factors driving returns and predict future return probabilities, while visualizing these insights in an interactive Power BI dashboard that will help decision-makers enhance their strategies, identify risky products, and optimize supply chain performance.

Tools Used

- Python (Pandas, NumPy, Scikit-learn): Data cleaning, encoding, feature engineering, and predictive modelling.
- Matplotlib & Seaborn: Data exploration and visualization.
- Power BI: Creating dashboards and visualizing business insights.
- Excel / CSV: Dataset storage and import/export between tools.
- Jupyter Notebook: Data preprocessing and model execution

Steps Involved in Building the Project

1. Data Collection: The dataset contains transactional information such as Order ID, Product ID, Order Date, Return Date, Product Category, Product Price, Order Quantity, User Age, Gender, Location, Payment Method, Shipping Method, Discount Applied, and Return Status.

2. Data Preprocessing in Python: - Filled missing Return_Reason with "No Return". - Replaced null Return_Date values with NaT. - Created derived feature Days_to_Return_Calc. - Encoded categorical columns and created a binary 'Returned' column. - Removed duplicates and invalid entries.

3. Predictive Model Development: - Used Logistic Regression for prediction. - Applied SMOTE to balance data. - Evaluated with Accuracy, Precision, Recall, F1-Score, and AUC. - Generated Return Probability for each transaction.

4. Data Export and Integration: - Saved results as data_with_predictions.csv. - Imported into Power BI for visualization and dashboard creation.

5. Power BI Dashboard Design: - KPIs: Total Orders, Total Returns, Return Rate, Avg. Discount, Avg. Price, Return Probability. - Visuals: Bar (Category), Pie (Payment), Line (Trends), Column (Location), Scatter (Price vs Probability). - Filters for Category, Gender, Location, and Payment Method.

Conclusion

The E-Commerce Return Rate Prediction Dashboard shows the integration of data preprocessing, machine learning, and business intelligence. Logistic Regression helps understand which transactions are more likely to result in returns. Power BI visualizations make these predictions actionable. This project demonstrates how data-driven insights can help e-commerce companies reduce return rates and enhance customer satisfaction. Future improvements may include exploring advanced algorithms such as Random Forest or XGBoost and building real-time dashboards.