

E-commerce Return Rate Reduction Analysis

Introduction:

The E-commerce industry faces frequent product returns, which impact revenue, logistics, and customer satisfaction. This project aims to identify the root causes of returns, analyze patterns across categories, sellers, and channels, and build a predictive system to flag high- return-risk products. The end goal is to help stakeholders take preventive actions and optimize inventory decisions.

Abstract:

This project implements a complete return risk reduction pipeline using Python for data processing and machine learning, and Power BI for dashboarding. The dataset was cleaned and explored to identify return trends by category and seller. A Random Forest model was trained to predict the likelihood of returns. The model's output was used to generate a list of high-risk products, visualized in an interactive Power BI dashboard. This helps non-technical users understand which products are at higher risk of return and take action accordingly.

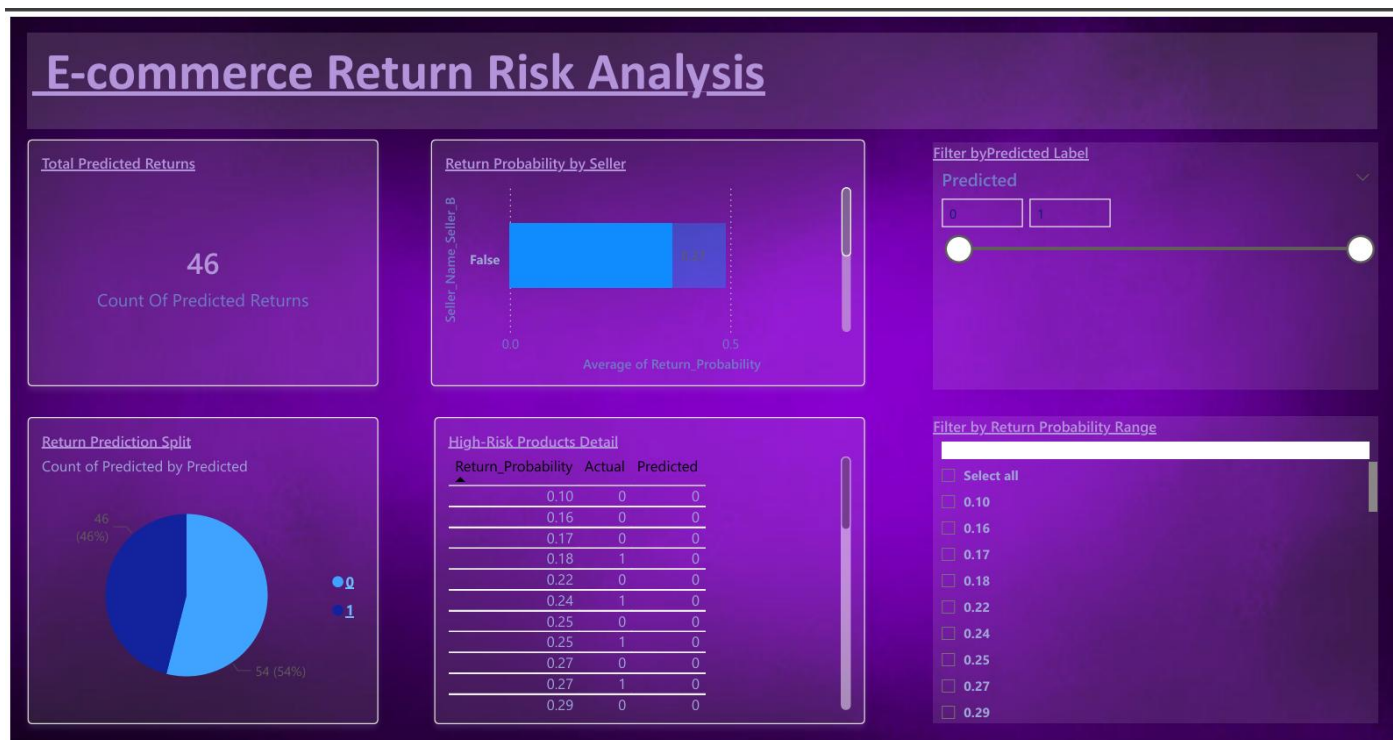
Tools Used:

- **Python (Google Colab)** – Data cleaning, feature engineering, modeling
- **Scikit-learn** – Machine Learning (Random Forest Classifier)
- **Matplotlib/Seaborn** – Visualizing category/seller return trends
- **Power BI** – Building interactive visual dashboard
- **Pandas** – CSV generation and processing

Steps Involved in Building the Project:

1. **Load and Clean Data**
The dataset was checked for nulls, duplicates removed, and column names standardized.
2. **Analyze Return Trends**
Return % was calculated and visualized by `Category` and `Seller_Name` to spot high- return segments.
3. **Predict Return Probability**
A Random Forest model was trained using features like `Price`, `Rating`, and `Name_Length`. The model predicted high-return-risk products with ~54% accuracy.
4. **Export Prediction Results**
The output included actual vs predicted labels and return probability. The full test results were exported as `high_risk_products.csv`.
5. **Create Power BI Dashboard**
The exported CSV was visualized in Power BI. Components included:
 - Pie chart of return prediction split (0 vs 1)
 - Bar chart of return probability by seller
 - Card showing total high-risk predicted products
 - Table with prediction details
 - Slicer to dynamically filter views by Predicted, Actual and Probability

Power BI Dashboard:



Conclusion:

This project provides a full pipeline from return trend analysis to predictive modeling and visualization. The dashboard enables stakeholders to filter and identify products most at risk of being returned. These insights can help reduce losses, improve supply chain performance, and enhance customer satisfaction. Future improvements may include integrating real-time feedback or return reasons to further optimize predictions.

