📄 E-commerce Product Return Analysis and Prediction

1. Abstract

Product returns pose a significant challenge for e-commerce businesses, leading to financial losses and reduced customer satisfaction. This project aims to analyze return patterns using a synthetic dataset and build a logistic regression model to predict return probabilities. Through data visualization, Power BI dashboards, and predictive modeling, high-risk products are identified for proactive management.

2. Introduction

In the e-commerce sector, understanding customer return behavior is crucial for optimizing operations and enhancing user experience. This project utilizes a synthetic dataset of e-commerce transactions to:

- Analyze return trends by product category, payment, and shipping method.
- Build a machine learning model to predict the likelihood of a return.
- Visualize insights interactively using Power BI dashboards.
- Identify high-risk products based on prediction probabilities.

3. Tools Used

- **Python** – Data manipulation and model building.
- **Pandas & NumPy** – Data preprocessing and numerical operations.
- **Matplotlib & Seaborn** – Data visualization.
- **Scikit-learn** – Machine learning modeling and evaluation.
- **Power BI** – Interactive data visualization and dashboard creation.

4. Steps Involved in Building the Project

a. Data Preparation

- Loaded data from a CSV file.
- Converted date columns to datetime format.
- Created a binary target variable Returned.
- Filled missing values in Return_Reason and Days_to_Return.

- Dropped irrelevant columns like IDs and raw date fields.

b. Exploratory Data Analysis (EDA)

- Visualized return rates by **Product Category**, **Payment Method**, and **Shipping Method**.

- Used bar plots, box plots, and pie charts for insights.

c. Feature Engineering

- Performed one-hot encoding on categorical features.

- Separated features (X) and target (y).

d. Model Building

- Used **Logistic Regression** for binary classification.

- Split data into training and test sets (80/20).

- Trained the model and predicted return probabilities.

e. Evaluation

- Evaluated using **classification report**, **ROC AUC score**, and **confusion matrix**.

- ROC AUC Score indicates the model's capability to distinguish returned items.

f. Insights & Export

- Flagged products with a predicted return probability above 70%.

- Exported high-risk product data to high_risk_products.csv.

g. Power BI Dashboard

- Built an interactive Power BI dashboard for stakeholders.

- Visualized key metrics: return rate by category, payment method trends, and return probabilities.

- Enabled dynamic filtering to explore patterns across variables and time periods.

5. Conclusion

This project successfully demonstrated the use of logistic regression and Power BI to understand and predict e-commerce return behavior. With clear insights and predictive capabilities, the approach aids in reducing return rates and improving strategic decisions in inventory and customer engagement.