Predicting User Churn for an E-commerce Platform

1. Feature Engineering

Key features were engineered to capture user behavior and churn signals:

- **Recency, Frequency, Monetary (RFM):**
  - **Recency:** Time since the last activity.
  - **Frequency:** Number of events by the user.
  - **Monetary:** Total expenditure by the user.
- **Session Insights:** Derived session-level data to assess user activity patterns.
- The processed features were saved to user_features.csv.

2. Predictive Modeling

- **Model:** A Gradient Boosting Classifier was trained to predict churn (binary classification: churned or not).
- **Data Splitting:** The dataset was split into training and testing sets.
- **Evaluation Metrics:** The model achieved:
  - **Accuracy:** 100%
  - **Precision, Recall, and F1-Score:** 1.00 for both classes.
  - **ROC-AUC Score:** 1.0

3. Model Output

The trained model was saved as churn_model.pkl for deployment. Predictions are highly accurate based on validation and testing results.

Key Results

Model Evaluation

- **Confusion Matrix:**
  - True Negatives (non-churn): 6,553
  - True Positives (churn): 74,904
  - No false positives or false negatives, indicating perfect predictions.
- **Classification Report:**
  - Both precision and recall achieved a perfect score (1.00), reflecting the model's ability to predict churn with absolute accuracy.
- **ROC-AUC:** A score of 1.0 indicates perfect discrimination between churn and non-churn users.

Recommendations

1. **Proactive Retention Campaigns:**
    - Target churn-prone users based on the model's predictions.
2. **Enhanced User Experience:**
    - Address bottlenecks in user journeys, especially for users frequently viewing but not purchasing.
3. **Data-Driven Marketing:**
    - Use insights from product categories and user sessions for tailored offers and promotions.

Files

1. **Code Files:**
    - data_analysis.py: Dataset exploration and summary.
    - feature_engineering.py: Feature creation and processing.
    - modeling.py: Model training and evaluation.
2. **Output Files:**
    - user_features.csv: Feature data for all users.
    - churn_model.pkl: Trained churn prediction model.
3. **Performance Metrics:**
    - Accuracy: 100%
    - Precision, Recall, F1-Score: 1.00
    - ROC-AUC: 1.0