Report: Predicting User Churn for an E-commerce Platform

Title:

Predicting User Churn for an E-commerce Platform.

Objective:

The objective of this project is to predict which users are most likely to churn (i.e., stop returning or purchasing) on an e-commerce platform and to provide actionable insights that can improve user retention. By analyzing user behavior data, we aim to create a model that identifies at-risk users and recommend strategies to reduce churn effectively.

Dataset Overview:

- **File Name: events.csv**
- **Columns:**
    - **event_time: Datetime of user activity.**
    - **event_type: Activity type (view, cart, purchase).**
    - **product_id: Product identifier.**
    - **category_id: Product category identifier.**
    - **category_code: Human-readable category code.**
    - **brand: Product brand.**
    - **price: Price of the product.**
    - **user_id: User identifier.**
    - **user_session: Session identifier.**

Approach:

1. **Churn Definition:**
   Churn is defined as users who do not perform any activity (view, cart, or purchase) within a 30-day window from their last recorded activity. This threshold aligns with typical e-commerce behavior patterns and helps capture inactive users accurately.
2. **Feature Engineering:**
    - **RFM Metrics:**
        - **Recency:** Days since the user's last activity.
        - **Frequency:** Number of interactions (e.g., views, carts, purchases).

        - **Monetary:** Total monetary value of purchases.
    - **Behavioral Patterns:**
        - View-to-Cart Ratio.
        - Cart-to-Purchase Ratio.
    - **Session Metrics: Number of sessions per user.**
3. **Modeling:**
    - **Built and evaluated two models: Logistic Regression and Random Forest.**
    - **Metrics evaluated:**
        - **Precision, Recall, F1 Score, Accuracy.**
        - **Log Loss, Brier Loss, and ROC-AUC.**
4. **Interpretability:**
    - **Feature importance analysis using SHAP values and bar plots.**

Key Findings:

Metric	Logistic Regression	Random Forest
Precision	**0.9979**	**1.0**
Recall	**0.9999**	**1.0**
F1 Score	**0.9989**	**1.0**
Accuracy	**0.9983**	**1.0**
ROC-AUC	**0.9999**	**1.0**
Log Loss	**0.0146**	**0.00007**
Brier Loss	**0.0035**	**0.000001**

Best Model:
The Random Forest model achieved perfect classification results with an AUC of 1.0 and minimal probabilistic error metrics (Log Loss = 0.00007). Logistic Regression performed well but was slightly less accurate.

Interpretability & Insights:

1. **Key Features Influencing Churn:**
    - **Recency: Users with higher days since their last activity are more likely to churn.**
    - **View-to-Cart Ratio: High ratios indicate users who browse but do not engage further, a strong churn signal.**
    - **Cart-to-Purchase Ratio: A low ratio suggests users abandon items in their cart without purchasing.**
2. **Behavior Patterns:**
    - **Users with low monetary engagement and infrequent sessions are at greater risk of churn.**

Business Recommendations:

1. **Targeted Campaigns:**
    - Offer personalized discounts to users with high view-to-cart ratios to encourage purchases.
    - Re-engage users inactive for 20+ days through email campaigns or app notifications.
2. **Improve Conversion Rates:**
    - Optimize product pages and cart experience to reduce friction during checkout.
3. **Product Recommendations:**
    - Leverage user preferences (e.g., most viewed categories) for personalized suggestions.
4. **Retention Strategies:**
    - Introduce loyalty programs to reward frequent purchases.
    - Provide special offers for users close to churning based on predictive scores.

Conclusion:

This project successfully identifies at-risk users and provides actionable insights to reduce churn. By implementing these recommendations, the e-commerce platform can improve customer retention, enhance user experience, and increase revenue.