Churn Prediction Analysis Report

Submitted by - GAJAL GARG

Introduction

This report outlines the steps, methodology, and findings from the churn prediction analysis conducted on the provided e-commerce dataset. The goal was to identify user churn and develop actionable insights to improve user retention.

Approach

1. Exploratory Data Analysis (EDA)

- **Dataset Overview**: The dataset contained 372,490 rows and 9 columns. Missing values were observed in category_code, brand, and user_session, which were addressed through imputation.
- **Distributions**:
  - Event type distribution was visualized using count plots.
  - Price and category code distributions were analyzed with histograms, ensuring clean and interpretable visualizations.
- **User Behavior**:
  - Aggregated user events to understand interactions such as view, cart, and purchase behaviors.

2. Churn Definition

- **Definition**: A user was considered churned if no purchase activity occurred within 30 days of their last interaction.
- **Implementation**:
  - Identified the last purchase date for each user.
  - Computed the difference in days between the dataset's latest event time and each user’s last purchase.
  - Users exceeding the 30-day threshold were labeled as churned.

3. Feature Engineering

- Created user-level features to capture meaningful churn signals:
  - **Recency**: Days since the last event.
  - **Frequency**: Total number of events per user.
  - **Monetary**: Total monetary value of transactions.
  - **Session Count**: Unique sessions per user.

- Encoded categorical data such as category_code for modeling.

4. Model Building

- **Model Used**: Random Forest Classifier.
- **Train-Test Split**: Data split into 80% training and 20% testing sets.
- **Performance Metrics**:
  - **Classification Report**:

	precision	recall	f1-score	support
0	1.00	1.00	1.00	148882
1	0.98	0.99	0.99	28144

  - accuracy 1.00 177026

macro avg 0.99 1.00 0.99 177026 weighted avg 1.00 1.00 1.00 177026

* AUC Score: 0.9988

5. Feature Importance

- Identified the most important features influencing churn:
  - **Recency**: 0.42
  - **Monetary**: 0.25
  - **Frequency**: 0.18
  - **Session Count**: 0.15

Recommendations

1. **Personalized Discounts**:
   - Offer discounts to users with high cart-to-purchase ratios to incentivize purchases.
2. **Targeted Engagement**:
   - Reach out to inactive users via marketing emails or app notifications within 30 days of inactivity.
3. **Top Product Focus**:
   - Analyze and improve the top categories or products associated with churn.
4. **Loyalty Programs**:

- Reward frequent purchasers and high-spending users with loyalty points or exclusive offers.

5. **Session Analysis**:
    - Improve session experiences for users with low session counts to increase engagement.
6. **Brand Collaboration**:
    - Collaborate with popular brands to ensure availability and exclusive deals.

Conclusion

The Random Forest model effectively predicted user churn with a high level of accuracy and AUC score. The insights derived from feature importance and user behaviour analysis provide actionable strategies to reduce churn and enhance user retention.

Future Work

1. **Incorporate Time-Series Analysis**:
    - Explore sequential patterns in user behaviour.
2. **Experiment with Advanced Models**:
    - Use Gradient Boosting or Neural Networks to potentially improve predictive performance.
3. **Dynamic Churn Definition**:
    - Adjust the churn threshold dynamically based on user segment behaviour.

Final Metrics:

- **Accuracy**: 1.00
- **AUC Score**: 0.9988