Predicting Purchase Intent in E-commerce

Ayush Chaudhary (120429125)

May 6, 2025

1 Introduction

In the e-commerce industry, understanding and predicting user purchase intent is crucial for optimizing user experience and increasing conversion rates. I aim to develop a machine learning model that predicts the likelihood of a user purchasing a product they've viewed. This prediction can help e-commerce platforms implement targeted strategies to convert views into purchases.

2 Dataset Overview

I will utilize the Retail Rocket E-commerce Dataset [1], which contains 2,756,101 behavioral events (views, add-to-cart, transactions), 417,053 unique items with 20,275,902 property records, 1,407,580 unique visitors, and 1,669 category relationships. This rich dataset provides comprehensive information about user interactions and product characteristics.

3 Proposed Approach

3.1 Target Variable

I will create a binary classification model where the target variable indicates whether a user will purchase a product they've viewed (1) or not (0). This will be derived from the transaction events in the dataset, providing clear ground truth for model training and evaluation.

3.2 Feature Engineering

I will engineer features from three main categories. User behavior features will include the number of product views, time spent viewing, previous purchase history, add-to-cart events, and time between views. Product characteristics will encompass category information, price changes over time, availability status, and product properties. Temporal features will capture time of day/week of view, time since last purchase, and session duration.

[1] `https://www.kaggle.com/datasets/retailrocket/ecommerce-dataset`

3.3 Algorithms

I will evaluate multiple machine learning algorithms, starting with Logistic Regression as a baseline model. I will then implement and compare Random Forest, XGBoost, and Neural Networks to identify the most effective approach for this prediction task.

4 Hypothesis

I hypothesize that purchase intent can be predicted using user behavior patterns, product characteristics, and temporal features. Users who view a product multiple times are more likely to purchase, while products with stable prices tend to have higher conversion rates. Certain product categories consistently show higher purchase rates, and time-based patterns significantly influence purchase likelihood.

5 Evaluation and Impact

I will evaluate the model using AUC-ROC, Precision-Recall, F1 Score, and Confusion Matrix. The project's practical implications include improved conversion rates, better user experience, optimized marketing strategies, and enhanced inventory management. These improvements can lead to significant business value for e-commerce platforms.