Problem Statement: Online Shoppers Purchase Intentions

Objective

The aim of this project is to analyze user browsing behavior and session-level data to predict whether a user will complete a purchase during an online shopping session. This will be achieved by combining Exploratory Data Analysis (EDA) with a powerful machine learning model, XGBoost, to uncover insights and build an accurate predictive model.

Dataset Overview

We will be using the Online Shoppers Purchasing Intention Dataset from the UCI Machine Learning Repository.

Dataset Link: https://archive.ics.uci.edu/ml/datasets/Online+Shoppers+Purchasing+Intention+Dataset
File Name: online_shoppers_intention.csv

Features Description

This dataset consists of 12,330 sessions and includes 18 behavioral features and 1 target variable (Revenue) indicating whether a purchase was made. Key features include:

- Administrative, Administrative_Duration: Count and time spent on admin pages
- Informational, Informational_Duration: Count and time on informational pages
- ProductRelated, ProductRelated_Duration: Interaction with product pages
- BounceRates, ExitRates: Session bounce/exit behavior
- PageValues: Assigned value of pages visited
- SpecialDay: Closeness to special dates (like holidays)
- OperatingSystems, Browser, Region, TrafficType: Technical user/session info
- VisitorType: New, Returning, or Other
- Weekend: Whether the session occurred on a weekend
- Revenue (Target): True if purchase occurred, otherwise False

Tasks & Workflow

Problem Statement: Online Shoppers Purchase Intentions

1. Data Cleaning & Preprocessing

- Handle missing values
- Encode categorical variables (e.g. VisitorType, Weekend)
- Feature scaling (if needed)

2. Exploratory Data Analysis (EDA)

- Analyze user behavior patterns
- Compare sessions that led to purchases vs those that didn't
- Visualize key features using histograms, boxplots, heatmaps, and bar plots

3. Modeling with XGBoost

- Split data into training and testing sets
- Train a binary classification model using XGBoost
- Tune hyperparameters using GridSearchCV or RandomizedSearchCV
- Extract and visualize feature importance

4. Model Evaluation

- Accuracy
- Precision & Recall
- F1-Score
- ROC-AUC Score
- Confusion Matrix

5. Final Deliverables

- Trained and optimized XGBoost model
- Insightful EDA visualizations and findings
- List of most influential features on purchase behavior
- Report or dashboard summarizing predictions and insights

Learning Outcomes

Problem Statement: Online Shoppers Purchase Intentions

- Understand how user browsing behavior impacts online purchase decisions
- Apply EDA techniques on a real-world e-commerce dataset
- Learn to train and tune XGBoost models for classification tasks
- Interpret model performance using multiple metrics
- Gain insight into feature importance in predictive modeling