Problem Statement: The objective of this project is to predict online shoppers' intention to make a purchase based on various features related to their browsing behavior. The dataset comprises 12,330 sessions, each belonging to a different user, over 1 year. The features include both numerical and categorical attributes, providing insights into users' interactions with the website.

The dataset consists of feature vectors belonging to 12,330 sessions. The dataset was formed so that each session would belong to a different user in a 1-year period to avoid any tendency to a specific campaign, special day, user profile, or period.

You are supposed to do EDA, various feature transformations, and apply the SVM model with hyperparameter tuning. Since it is imbalanced data, the evaluation metric will be f1 score. The benchmark f1 score is 0.55

The dataset is available in the Downloads

Features

1. Administrative: The number of pages of this type (administrative) visited by the user in that session.
2. Administrative Duration: The total amount of time (in seconds) spent by the user on administrative pages during the session.
3. Informational: The number of informational pages visited by the user in that session.
4. Informational Duration: The total time spent by the user on informational pages.
5. Product Related: The number of product-related pages visited by the user.
6. Product-Related Duration: The total time spent by the user on product-related pages.
7. Bounce Rates: The average bounce rate of the pages visited by the user. The bounce rate is the percentage of visitors who navigate away from the site after viewing only one page.
8. Exit Rates: The average exit rate of the pages visited by the user. The exit rate is a metric that shows the percentage of exits from a page.
9. PageValues: The average value of the pages visited by the user. This metric is often used as an indicator of how valuable a page is in terms of generating revenue.
10. SpecialDay: This indicates the closeness of the site visiting time to a specific special day (e.g., Mother’s Day, Valentine's Day) in which the sessions are more likely to be finalized with a transaction.
11. Month: The month of the year in which the session occurred. Operating Systems: The operating system used by the user.
12. Browser: The browser used by the user.
13. Region: The region from which the user is accessing the website.

14. Traffic Type: The type of traffic (e.g., direct, paid search, organic search, referral).
15. Visitor Type: A categorization of users (e.g., Returning Visitor, New Visitor).
16. Weekend: A Boolean indicating whether the session occurred on a weekend.
17. Revenue: A binary variable indicating whether the session ended in a transaction (purchase).