E-commerce Shopper's Behaviour Understanding

Problem Statement:-Let's say, we have collected the users' session data for a year of a E-commerce company, where each row belongs to a different user. Whether the user has made a purchase or not during that year is indicated by an indicator. We will predict that column using other attributes of the users.

Problem classification:- Logistic regression

Libraries to be used:
-Scikit-learn
-XGBoost
-imblearn
-NumPy
-Pandas
-Seaborn
-Matplotlib
-catBoost

Dataset:-
There are 21 attributes and their descriptions are given below.
Files:
train.csv - the training set
test.csv - the test set
sample_submission.csv - a sample submission file in the correct format
Columns:
The first six columns represent the different pages in the e-commerce website visited by a user from other sites.
•   HomePage: Number of times visited this page
•   HomePage_Duration: Total number of duration spent on this page.
•   LandingPage: Number of times visited this page
•   LandingPage_Duration: Total number of duration spent on this page.
•   ProductDesriptionPage: Number of times visited this page
•   ProductDescriptionPage_Duration: Total number of duration spent on this page.
•   GoogleMetric-Bounce Rate: Whenever a user comes to any one web-page of the website and he/she does not go to any other page and exits from the website from the same page, then this activity done by the user is called Bounce. And the percentage of the total number of times the user visiting our website and bounce it, is called Bounce Rate

•   GoogleMetric-Exit Rate: The bounce rate is calculated based on the user exiting a website after visiting one page. But some users exit from the second, third, fourth, or any other page of our website, then those visitors' data help determine the exit rate. The percentage of the total number of times the user to our website who do not exit from the first page (Landing Page) but exit after exploring other website pages is called the Exit Rate.
•   GoogleMetric- Page Value is the average value for a page that a user visited before landing on the goal page or completing an Ecommerce transaction.

• SeasonalPurchase: It is a weight indicator to track the seasonal purchase. If a user makes a purchase during any seasonal time (Holi,Christmas), we will assign based on internal heuristic.
• Month_ SeasonalPurchase: Month of the special day considered for seasonal purchase.
The other attributes like, OS, Search Engine, Zone, Type of Traffic, Customer Type, Gender, Cookies Setting, Education, Marital Status and Weekend Purchase are self-explanatory variables

Methodology:-
We aim to build a machine learning model that will predict whether an online customer of a retail shop made a purchase during that year or not.The major steps included the following:

Import the required libraries and load the data into your Python environment.

Preprocess the data by removing any unnecessary columns or missing values, and converting any categorical variables into dummy variables.

Split the data into training and testing sets, using a 70-30 or 80-20 split.Instantiate a logistic regression model using the scikit-learn library in Python.Fit the model to the training data using the fit method.

Use the predict method to predict the probability of a user making a purchase on the testing data.

Evaluate the model's performance using metrics such as accuracy, precision, recall, and F1-score.

Models used:-

Logistic Regression
Gradient Boosting
AdaBoost
CatBoost
Stacking Classifier

Results:-

Based on the analysis, the best performing model for predicting customer purchases was found to be the Stacking Classifier, which combined the

Gradient Boosting and AdaBoost classifiers. The Stacking Classifier achieved an accuracy score of 0.66 on the test data.

This model can be used by the online retailer to predict which customers are likely to make a purchase, allowing them to target their marketing efforts more effectively.