Capstone Project : 10

Problem statement:

People often spend a lot of time browsing through online shopping websites, but the coversion rate into purchases is low. Determine the likelihood of purchase based on the given features in the datsaset. The dataset consists of feature vectors belonging to 12,330 online sessions. The purpose of this project is to identify user behaviour patterns to effectively understand features that influence the sales.

Data Description:

The dataset cointains the following features:

|  |  |  |
| --- | --- | --- |
| **Features** | **Description** | **DataType** |
| Administrative | Number of pages visited by the user for user account management related activities | Discrete values from 0 to 27 |
| Administrative\_Duration | Time spent on Admin pages by the user | Continuous value (time in seconds) |
| Informational | Number of pages visited by the user about the website | Discrete values from 0 to 24 |
| Informational\_Duration | Time spent on Informational pages by the user | Continuous value (time in seconds) |
| ProductRelated | Number of product related pages visited by the user | Discrete values from 0 to 705 |
| ProductRelated\_Duration | Time spent on Product related pages by the user | Continuous value (time in seconds) |
| BounceRates | Average bounce rate of the pages visited by the user | Continuous value |
| ExitRates | Average exit rate of the pages visited by the user | Continuous value |
| PageValues | Average page value of the pages visited by the user | Continuous value |
| SpecialDay | Closeness of the visiting day to a special event like Mother’s Day or festivals like Christmas | Discrete values (0, 0.2, 0.4, 0.6, 0.8, 1.0) |
| Month | Month of the visit from Jan to Dec | Categorical |
| OperatingSystems | OperatingSystems of the visitor | Discrete values from 0 to 7 |
| Browser | Browser of the visitor | Discrete values from 0 to 12 |
| Region | Geographic region from which the session has been started by the visitor | Discrete values from 0 to 8 |
| TrafficType | Traffic source through which user has entered the website | Discrete values from 0 to 19 |
| VisitorType | Visitor type as New visitor, Returning user or Others | Categorical |
| Weekend | If the user visited on a weekend or not | Boolean |
| Revenue | If the user visit resulted with a transaction | Boolean |

Instructions:

1. Perform the required data pre-processing to treat for missing values and outliers.
2. Perform exploratory data analysis to visualise the spread of each of the X variables and the relationship between the various X variables and the Y variable
3. Divide the given data into train and test sets
4. Predict how likely it is for a customer to make a purchase by building classification models
5. Interpret how each of the X variables influence the conversion propensity
6. Evaluate the model performance measures and choose the most optimum model
7. Enlist your key findings based on the most optimum model and the respective feature importance