PROJECT SYNOPSIS

Predicting the Customer's Behavior in online shopping

Submitted towards the partial fulfillment of the criteria for award of Post Graduate In Data Analytics by Imarticus

Submitted By: S.Muveeshkumar

Course and Batch: PGA 28

Scope & Objective:

The aim is to predict the purchasing intention of new customers in online shopping based on existing customer data.

Business Problem Statement:

To be successful in a highly competitive eCommerce environment, it is essential to understand customers' online purchase intentions, their inclination to buy from a website rather than just browse. Concentrating solely on customers who come to browse would be futile. By identifying purchasing behavior customers, online stores can increase their productivity.

Data Sources

Online Shoppers Purchasing Intention Dataset - UCI Machine Learning Repository

Analytics Tools

Python (Jupyter Notebook)

Analytics Approach:

K-Means Clustering (Unsupervised Machine Learning): It is used to cluster the data into groups based on user behavior.

Process:

1. Feature Scaling
2. Missing Values
3. Feature Encoding
4. K-Means Clustering (Without Ground Truth)
5. Silhouette Score

RandomForestClassifier (Classification-Ensemble Model): Used to predict which group the new customers belong to, with the target variable as the cluster.

Process:

1. Model Fit
2. Model Prediction
3. Accuracy Score

Finally, the Percentage of Revenue from each cluster will be calculated.

KPI (Key Performance Indicator): Calculate the percentage of revenue contributed by each cluster to understand the impact of different customer segments on overall revenue. This provides insights into the effectiveness of targeting specific clusters for marketing and sales strategies.

Feature Description:

1. Administrative - The number of pages visited by the user in the administrative section, such as Customer Support, Product Tracking, Profile Settings, and Refund Requests.

2. Administrative Duration - The time spent by the user in administrative pages.

3. Informational - The number of pages visited by the user in Product Blogs, Customer Reviews, and Product Videos.

4. Informational Duration - The time spent by the user in informational pages.

5. Product Related - The list of particular product pages. For example, if I am searching for a watch, the page will only contain watches.

6. Product Duration - The time spent on the product page by the user.

7. Bounce rate - Calculated by the ratio of the total number of one-page views to the total number of visits. It represents when a visitor enters a page without triggering any action and then exits to another page or exits the website.

8. Exit Rate - The rate at which users exit the website from a particular page.

9. Page value - This represents the likelihood of a user completing a transaction or goal on a specific page. It is calculated based on the unique page that leads to a transaction.

10. Special day - A value ranging from 0 to 1, indicating whether it is a special day or not. For example, if it is zero before and after a particular date like Mother's Day, it signifies a non-special day.

11. Month - Sales are highest in May and November on a monthly basis.

12. Operating system - The type of operating system used by the customer.

13. Browser - The kind of browser used by the customer.

14. Region - The type of region the customer belongs to.

15. Traffic type - The type of traffic the website receives.

16. Visitor type - The different types of visitors, such as new visitors or returning visitors.

17. Weekend - A Boolean value indicating whether the customer visits on weekends or not.

18. Revenue - A Boolean value indicating whether there is revenue from the customer or not.

WORKFLOW ARCHITECTURE