FIREBLAZE AI SCHOOL

ENLIGHTEN YOUR CAREER

Mini Project - Machine Learning

Domain : Logistic

About Dataset

- A retail company “ABC Private Limited” wants to understand the customer purchase behavior (specifically, purchase amount) against various products of different categories.
- They have shared purchase summaries of various customers for selected high volume products from last month.
- The data set also contains customer demographics (age, gender, marital status, citytype, stayincurrentcity), product details (productid and product category) and Total purchase amount from last month.
- Now, they want to build a model to predict the purchase amount of customers against various products which will help them to create personalized offers for customers against different products.

Features Information :

Variable	Definition
User_ID :	User ID
Product_ID :	Product ID
Gender :	Sex of User
Age :	Age in bins

- **Occupation :** Occupation(Masked)
- **City_Category :** Category of the City (A,B,C)
- **StayInCurrentCity Years :** Number of years stay in current city
- **Marital_Status :** Marital Status
- **ProductCategory1 :** Product Category (Masked)
- **ProductCategory2:** Product may belongs to other category also (Masked)
- **ProductCategory3 :** Product may belongs to other category also (Masked)
- **Purchase** : Purchase Amount (Target Variable)

Deliverables :

Data Preprocessing : This step performs all pre-processing steps such as data manipulation, data filling,converting categorical into numeric, and all processes.

Exploratory Data Analysis :

- The data preprocessing stage extracts useful information statistically.

- Such as check outliers, skewness, compare the features by graph and many more. Do all the required steps as well.

The EDA process involves performing

1. **Univariate Analysis :** In this part, first check target features and start univariate analysis.
2. **Bivariate analysis :** This analysis involves studying two variables and their relationship, recalling some of the hypotheses that we generated earlier.
3. **Removing Missing values if any / Outlier treatment :** After exploring all the variables in our data, we can now impute the missing values and treat the outliers because missing data and outliers can have an adverse effect on the model performance and accuracy.

   Machine Learning : Probability of purchase. and also check if the model is to be underfitting or overfitting if it has then solves this by using cross-validation technique, or perform hyperparameter tuning to improve model performance.

   Deadline to submit the project : One Week.

ALL THE BEST