****

**Assessment Report**

on

**“Customer Behavior Prediction”**

submitted as partial fulfillment for the award of

**BACHELOR OF TECHNOLOGY**

**DEGREE**

SESSION 2024-25

in

**CSE(AI)**

By

Name : Aman Singh Yadav

Roll Number : 202401100300031

Section: A

**Under the supervision of**

“Bikki Kumar”

**KIET Group of Institutions, Ghaziabad**

**May, 2025**

**1. Introduction**

As online retail continues to expand, understanding and predicting customer behavior is essential. This project addresses the problem of classifying customers into 'bargain hunters' or 'premium buyers' using their purchasing patterns. By analyzing features like total spending, average purchase value, and visit frequency, the goal is to build a predictive model that helps businesses target customers more effectively.

**2. Problem Statement**

To classify customers into two categories—'bargain hunters' or 'premium buyers'—based on purchase behavior. This helps in tailoring marketing strategies and enhancing customer experience.

**3. Objectives**

* Preprocess the dataset for training a machine learning model.
* Train a Logistic Regression model to classify customer behavior.
* Evaluate model performance using accuracy, precision, and recall.
* Visualize the confusion matrix using a heat map for interpretability.

**4. Methodology**

* **Data Collection**: The user uploads a CSV file containing the dataset.
* **Data Preprocessing**:  
  + Handling missing values using mean and mode imputation.
  + One-hot encoding of categorical variables.
  + Feature scaling using StandardScaler.
* **Model Building**:  
  + Splitting the dataset into training and testing sets.
  + Training a Logistic Regression classifier for customer classification.
* **Model Evaluation**:  
  + Evaluating accuracy, precision, recall, and F1-score.
  + Generating a confusion matrix and visualizing it with a heatmap.

**5. Data Preprocessing**

The dataset is cleaned and prepared as follows:

* Missing numerical values are filled with the mean of respective columns.
* Categorical values are encoded using one-hot encoding.
* Data is scaled using StandardScaler to normalize feature values.
* The dataset is split into 80% training and 20% testing.

**6. Model Implementation**

Logistic Regression is used due to its simplicity and effectiveness for binary classification. The model is trained on a dataset containing customer spending features and predicts whether a customer is a bargain hunter or a premium buyer.

**7. Evaluation Metrics**

The following metrics are used to evaluate the model:

* **Accuracy**: Measures overall correctness.
* **Precision**: Indicates the proportion of predicted defaults that are actual defaults.
* **Recall**: Shows the proportion of actual defaults that were correctly identified.
* **F1 Score**: Harmonic mean of precision and recall.
* **Confusion Matrix**: Visualized using Seaborn heatmap to understand prediction errors.

**8. Results and Analysis**

* The model provided reasonable performance on the test set.
* Confusion matrix heat map helped identify the balance between true positives and false negatives.
* Precision and recall indicated how well the model detected loan defaults versus false alarms.

**9. Conclusion**

The logistic regression model was able to classify customers based on their purchase behavior with modest accuracy. This project highlights how machine learning can be applied in retail analytics to better understand customer types. Future improvements can include trying other models and addressing class imbalance for better prediction.

**10. References**

* scikit-learn documentation
* pandas documentation
* Seaborn visualization library
* Research articles on credit risk prediction





