DATA ANALYSIS AND DATA SCIENCE USING PYTHON

TASK - 3

Task 3: Clustering Analysis – Customer Segmentation

Objective

Perform customer segmentation using clustering techniques to group customers based on their purchasing behavior, allowing businesses to target each segment effectively.

Project Steps

Step 1: Dataset Selection

- **Dataset Name**: `customer_data.csv`
- **Columns**:
  - **Customer ID**: Unique identifier for each customer.
  - **Age**: Age of the customer.
  - **Annual Income**: Income in $ (or any currency).
  - **Spending Score**: A score assigned to customers based on their spending patterns and behavior.

Step 2: Tasks to Perform

1. Load the Dataset

- Use libraries such as **Pandas** to load the dataset into a DataFrame.
- Inspect the dataset by checking:
  - Shape, missing values, duplicates, and data types.
  - Summary statistics to understand ranges of values.

2. Data Preprocessing

- **Standardize the data**:
  - Use a scaler (e.g., `StandardScaler` or `MinMaxScaler` from sklearn) to ensure all features are on the same scale.
  - This step is crucial since clustering algorithms are sensitive to feature magnitudes.

3. Clustering

- **Determine the optimal number of clusters**:
    - Use the **Elbow Method**:
        - Plot the Within-Cluster Sum of Squares (WCSS) against the number of clusters.
        - The "elbow point" indicates the optimal number of clusters.
    - Alternatively, use the **Silhouette Score** for evaluation.
- **Apply K-Means Clustering**:
    - Use the optimal number of clusters identified from the Elbow Method.
    - Assign a cluster label to each customer.

4. Visualization

- Create visualizations to represent the clusters:
    - **2D Scatter Plot**:
        - Use **PCA (Principal Component Analysis)** or **t-SNE** to reduce dimensions to two for visualization.
        - Plot clusters in different colors.
    - **Pair Plots**: Visualize relationships between features within clusters.
    - **Centroid Visuals**: Show the centroid of each cluster for better interpretation.

Deliverables

1. **Clustered Dataset**:
    - Add a new column to the dataset with the assigned **cluster labels** for each customer.
2. **Visualizations**:
    - A 2D scatter plot representing customer clusters.
    - Elbow Method plot showing the optimal number of clusters.
    - Any additional visuals (e.g., pair plots, heatmaps).
3. **Recommendations**:
    - Insights and actions based on clusters:
        - Which groups to target for promotions or premium products.
        - Identify high-spending customers and propose loyalty programs.
        - Tailored marketing strategies for different age or income segments.

Main Flow Services and Technologies Pvt. Ltd.
Contact Us. +91 9389641586, +91 97736 99074
Email-Add. contact.mainflow@gmail.com
www.mainflow.in

Expected Outcomes

- Segment customers into meaningful groups based on purchasing behavior.
- Identify key patterns or trends for actionable insights.
- Provide a visual understanding of customer groups for strategic planning.
- Facilitate better customer targeting, marketing strategies, and resource allocation.

Deadline Compliance

- **Restriction**: **Submit the project within 7 days** from the start date.
- **Reason**: Meeting deadlines is crucial in the real-world software development environment. This restriction helps students practice **time management** and **task prioritization**. In professional settings, tight deadlines are often the norm, and learning to meet them without compromising quality is an essential skill.
- **Learning Outcome**: Students will learn to manage their time effectively, complete projects under pressure, and **deliver results on time**, which are all important skills in the workplace.

Main Flow Services and Technologies Pvt. Ltd.
Contact Us. +91 9389641586, +91 97736 99074
Email-Add. contact.mainflow@gmail.com
www.mainflow.in