Customer Journey Analysis Using Clustering and Dimensionality Reduction: Enhancing User Experience

Phase 1: Problem Definition and Data Understanding

College Name: Sapthagiri College Of Engineering

Group Members:

- **Name: Karthik L**
  CAN ID Number: CAN_33260330
- **Name: Dhanush J R**
  CAN ID Number: CAN_33274239
- **Name: Rakesh R**
  CAN ID Number: CAN_33260330

1.1 Project Overview

The objective of this project is to analyze customer journeys by leveraging clustering and dimensionality reduction techniques. A detailed understanding of customer journeys helps businesses optimize user experience and increase engagement. Traditional clustering techniques often falter with complex and high-dimensional data. To address this, dimensionality reduction methods like Principal Component Analysis (PCA) and t-SNE are employed to extract key latent features, enhancing the clustering process. By segmenting customer journeys into distinct patterns, businesses can identify critical touchpoints, pain points, and opportunities for improvement, leading to more effective strategies for user retention and satisfaction.

1.2 Objective of the Project

 Objectives:

- To segment customer journeys into meaningful clusters by analyzing their behaviors, interactions, and preferences.
- This involves unsupervised learning methods to discover hidden structures in customer data without prior labels. The goal is to gain actionable insights that can guide strategic decisions to enhance user experience.

 Target Users:

- Marketers seeking to personalize campaigns.
- UX designers improving product flows.
- Data scientists applying clustering techniques to real-world scenarios.

 Potential Applications:

- **Journey Segmentation:** Identifying common patterns and deviations in customer interactions to target improvements effectively.
- **Customer Retention:** Spotting at-risk customers based on their journey data to implement preemptive retention strategies.

- **Product Optimization:** Informing feature development by understanding how users navigate products or services.

1.3 Dataset Overview and Data Requirements

The analysis requires a dataset rich in behavioral, interactional, and demographic features that map the customer journey across touchpoints.

- **Features**:
  - Behavioral Data: Frequency and timing of interactions, abandonment rates, and conversion metrics.
  - Engagement Data: Page views, clicks, session duration, and navigation paths.
  - Demographics: Customer age, location, and other profile attributes.
  - Contextual Data: Device type, access time, and geographical location.
- **Labels:** As this is an unsupervised task, labels are not required. Instead, the goal is to extract meaningful segments from unlabeled data.
- **Dataset Format:**
  - Tabular structure (e.g., CSV or SQL).
  - Rows representing unique customer journeys and columns denoting attributes.

1.4 Data Sources

Data for this project can be sourced from:

- **Public Datasets**:
  - Kaggle datasets with user journey data.
  - UCI Machine Learning Repository offering behavioral datasets.
- **Web Analytics Tools**: Tools like Google Analytics providing data on customer navigation patterns.
- **Proprietary Systems**: Internal CRM and analytics platforms storing customer engagement data.

1.5 Initial Data Exploration

- **Data Quality Checks**: Addressing missing values and ensuring data completeness.
- **Feature Distribution**: Understanding the spread of key features and detecting outliers.
- **Correlation Analysis**: Identifying relationships between variables for dimensionality reduction.
- **Visual Exploration**: Employing heatmaps, scatter plots, and path visualizations to identify trends.

1.6 Preprocessing Objectives

- **Dimensionality Reduction:** Techniques like PCA or t-SNE to reduce noise while preserving meaningful structures.
- **Feature Scaling:** Standardizing data to ensure uniformity across features.
- **Data Transformation:** Encoding categorical variables and normalizing numerical values.

1.7 Conclusion of Phase 1

This phase establishes a roadmap for customer journey analysis by defining objectives, understanding data needs, and outlining preprocessing strategies. Subsequent phases will focus on implementing clustering algorithms and evaluating their effectiveness in uncovering actionable insights from customer journey data.