PROJECT 1 - Customer Journey Analysis Using Clustering and Dimensionality Reduction

INTRODUCTION

In today's data-driven business environment, understanding customer behavior across different touchpoints is crucial for improving customer experience and increasing retention. Customer Journey Analysis involves tracking and analyzing the steps customers take when interacting with a business, from initial awareness to post-purchase behavior. This process generates vast and complex datasets, often with high dimensionality, making analysis challenging.

Key Features & Technologies

- Collects data across web, mobile, email, in-store, and support channels.
- Builds a comprehensive view of
- Captures the sequence and timing of customer actions (e.g., page views, clicks, purchases).
- Helps reconstruct and analyze sessions or journeys.
- Groups customers based on actual interaction patterns rather than static demographics.
- Enables tailored marketing and personalization strategies.

Key Technologies:

1. **Data Collection & Integration:**

- **Tools:** Google Analytics, Adobe Analytics, Mixpanel, Segment, Snowplow.
- **Technologies:** ETL pipelines (Apache NiFi, Airflow), data lakes (Amazon S3, Azure Data Lake).

2. **Data Processing & Transformation:**

- **Languages:** Python, R, SQL
- **Libraries:** Pandas, NumPy, PySpark

- **Functions:** Sessionization, feature engineering, event stream processing.

3. **Clustering Algorithms (Unsupervised Learning)**

- **K-Means** – Simple and fast for large datasets
- **DBSCAN** – Good for detecting noise and clusters of varying shape
- **Hierarchical Clustering** – Useful for building customer journey trees
- **Gaussian Mixture Models (GMM)** – For soft clustering with probability distributions

Application

- Customer Segmentation: Groups customers based on behavior patterns for targeted marketing.
- Churn Prediction: Identifies customers likely to leave based on journey deviations.
- Funnel Optimization: Detects where users drop off to improve conversion paths.
- Product Recommendation: Enhances recommendations by clustering similar user journeys.
- Campaign Analysis: Evaluates marketing effectiveness across different customer clusters.

Strategic Impact

Customer Journey Analysis empowered by clustering and dimensionality reduction offers significant strategic benefits to organizations aiming to become more customer-centric and data-driven. By uncovering hidden patterns in complex, multi-touchpoint customer data, businesses can make more informed decisions and tailor their strategies to meet evolving customer needs.

Advantages

- Reveals deep behavioral patterns and preferences across touchpoints.
- Groups customers based on journey data, enabling precise targeting and personalization.
- Dimensionality reduction simplifies complex data for clearer, more intuitive insights.
- Empowers teams with data-driven insights to guide strategy and resource allocation.

Disadvantages

- Customer journey data is often vast and unstructured, making preprocessing time-consuming and resource-intensive.
- Dimensionality reduction techniques like t-SNE or UMAP can produce results that are hard to interpret or explain to non-technical stakeholders.
- Clustering results can vary significantly based on the choice of algorithm and parameters (e.g., number of clusters in K-Means).

Conclusion

Customer Journey Analysis, when enhanced by clustering and dimensionality reduction techniques, provides powerful insights into how customers interact with a brand across multiple touchpoints. These methods allow organizations to uncover hidden patterns, segment users based on real behavior, and visualize complex data in a clear and actionable way.