Smart Learning Path Recommender for E-Learning

1. Executive Summary :
This capstone project focuses on designing a personalized learning path recommendation system for e-learning platforms using data from the Udemy Online Education Courses dataset. The system will analyze user preferences, course content, and engagement metrics to suggest tailored learning paths, enhancing user satisfaction and course completion rates.

2. Problem Statement :

Background:
E-learning platforms host thousands of courses across various subjects, making it overwhelming for learners to identify an optimal learning path tailored to their needs. Without proper guidance, users may enroll in irrelevant or disjointed courses, reducing their overall learning effectiveness and satisfaction.

Objective:
To build a recommendation system that suggests personalized learning paths by analyzing course metadata, user preferences, popularity metrics, and topic progression patterns.

Scope:
The project will utilize course data from the Udemy dataset available on Kaggle, focusing on building a recommendation engine based on course similarity, user intent (e.g., beginner to advanced), and pricing/popularity trends.

3. Data Sources :
Primary Data:
N/A

Secondary Data:
The dataset titled "Udemy Online Education Courses" from Kaggle
Link: https://www.kaggle.com/datasets/yusufdelikkaya/udemy-online-education-courses

The dataset contains course attributes such as:
- Course title and subject
- Number of subscribers and reviews
- Level (Beginner, Intermediate, Expert)
- Price and content duration
- Publication date

4. Methodology :

Data Collection & Preparation:
- Load and inspect the dataset using pandas
- Handle missing/null values and normalize categorical data
- Create derived features such as course score (combination of reviews, rating, duration, etc.)

Exploratory Data Analysis (EDA):
- Visualize course distributions by subject, level, and price
- Identify correlations between course popularity and attributes

Modeling & Recommendation Techniques:
- **Clustering**: Group similar courses using K-Means based on features
- **Content-Based Filtering:** Recommend similar courses using cosine similarity (TF-IDF on course titles/descriptions)
- **Path Sequencing:** Construct likely learning paths (Beginner → Intermediate → Advanced) based on metadata and enrollment patterns

Tools & Technologies:
- Python, Jupyter Notebook
- Libraries: pandas, numpy, matplotlib, seaborn, scikit-learn, nltk, sklearn.metrics.pairwise

5. Expected Outcomes :

- A functioning personalized course recommendation system tailored to learner profiles.
- Identification of typical learning paths within a subject.
- Data dashboard visualizing trends in course popularity, pricing, and learner behavior.
- Insights into which course features drive engagement and retention.

6. Risks and Challenges :

- Absence of actual user behavior data (clickstream, ratings, etc.) may limit personalization depth.
- Text data in course titles/descriptions may require extensive preprocessing and NLP techniques.
- Imbalanced course levels or subjects could skew learning path recommendations.

7. Conclusion :

This project aims to improve the online learning experience by recommending structured and personalized course paths. By leveraging metadata from Udemy courses, the system will empower learners to make informed decisions and follow coherent progression strategies. It can serve as a prototype for integration into real-world e-learning platforms for improving user satisfaction and completion rates.