

Smart Learning Path Recommender for E-Learning

Objective:

This project aims to design and develop an interactive platform that enables learners to discover optimal learning paths on e-learning platforms. By analyzing course metadata (price, duration, subject, level, reviews, subscribers), the system generates **personalized recommendations** and **structured progression paths (Beginner → Intermediate → Expert)**. The goal is to improve learner satisfaction, reduce dropouts, and encourage efficient skill development.

Dataset Details:

Feature	Description
Dataset Source	Udemy Online Education Courses dataset (Kaggle)
Key Attributes	Course ID, course title, URL, is_paid (boolean), price, number of subscribers, number of reviews, number of lectures, course level (Beginner, Intermediate, Expert, All Levels), content duration (in hours), publication timestamp, subject category
Granularity	Course-level metadata
Time Coverage	Variable, includes published timestamps covering various years (parsed as datetime)
Data Types	Mix of numeric (price, subscribers, reviews, lectures, duration), categorical (level, subject), datetime (published_timestamp)
Price Range	From free courses (price 0) to paid courses (prices vary up to 200 USD)
Subjects Covered	Business Finance, Graphic Design, Musical Instruments, Web Development
Levels	Courses categorized into levels such as Beginner, Intermediate, Expert, and All Levels. Levels normalized for analysis.
Content Duration	Course length available in decimal hours

Dashboard Highlights:

1. Overview

- KPIs: Total filtered courses, most popular subject, average price, average duration.
- Visuals:
 - Course distribution by subject & level (bar chart)
 - Courses published over time (line chart)
 - Feature correlation heatmap

2. Personalized Recommendations

- **Recommended Courses:** Content-based filtering (TF-IDF + cosine similarity) for similar courses.
- **Suggested Learning Path:** Ordered Beginner → Intermediate → Expert courses filtered by subject and budget.
- Card-style course previews (title, subject, level, price).

3. Learner Dashboard

- **Quick Picks (Top 3 courses)** personalized by skill level and weekly study capacity.
- **Skill Gap Radar:** Visual comparison between learner's current and target level.
- **Completion Time Estimator:** Duration of top 3 courses vs learner's weekly availability.

4. Course Insights

- Price vs. Course Score scatterplots
- Top 10 Best Value Courses (Score per \$)
- Popularity vs Engagement (Subscribers vs Reviews)

5. Clustering & Trends

- K-Means clusters (based on price, duration, subscribers, reviews, score).
- Cluster profiling (average price, engagement, duration).
- Cluster-wise course distribution and narratives.

6. Pricing & Popularity

- Price distribution histogram
- Average price by subject
- Price vs Subscribers (demand sensitivity)
- Duration vs Reviews (engagement measure)

7. Downloads

- Export full dataset, filtered dataset, and personalized learning path in CSV.

Technical Stack & Tools:

- Language: Python (Pandas, NumPy)
- Visualization: Plotly Express, Matplotlib, Seaborn
- Machine Learning: Scikit-learn (KMeans clustering, TF-IDF vectorization, cosine similarity)
- UI/Interactive: Streamlit framework for an accessible web app interface
- Caching for performance optimization on data loading and model computations

Steps Followed:

1. **Data Extraction:** Kaggle Udemy dataset (course titles, metadata, subscribers, reviews, prices, etc.).
2. **Data Transformation:** Cleaning nulls, normalizing levels (Beginner, Intermediate, Expert), deriving scores.
3. **Exploratory Data Analysis (EDA):** Distributions, correlations, trends.
4. **Recommendation Modeling:**
 - TF-IDF + Cosine similarity for content-based filtering.
 - K-Means clustering for course grouping.
 - Rule-based sequencing for structured learning paths.
5. **Visualization & UI:** Interactive Streamlit dashboard with Plotly, Seaborn, Matplotlib.
6. **Interactivity:** Filters, sliders, learner profile inputs, path download options.

Outcome:

The final system delivers a **personalized, interactive recommendation engine** for e-learning. Learners can:

- Discover top recommended courses based on similarity & budget.
- Visualize insights into pricing, engagement, and value.
- Download their customized learning path for tracking progress.

This platform prototype demonstrates how **data-driven personalization** can improve learner satisfaction, guide structured upskilling, and serve as a foundation for real-world e-learning integration.
