ENGR – E 516: Engineering Cloud Computing

EduSync – A Distributed Analysis and Recommendation Engine for Udemy Courses

1. Project Overview

1.1 Need for Project:

The global e-learning market has experienced exponential growth, with millions of learners accessing online courses. Udemy, as one of the leading platforms, hosts an extensive library of courses catering to diverse interests and skill levels. However, with this vast selection, learners often struggle to discover courses that best align with their goals and preferences.

1.2 Objective:

To address the challenges and meet the growing demand for personalized and efficient course recommendations, we propose the development of a comprehensive recommendation system. This system will leverage Udemy's API, employ advanced data processing techniques, and integrate with a user-friendly front-end interface.

2. Project Introduction

In today's rapidly evolving educational landscape, online learning platforms have become indispensable tools for acquiring new skills and knowledge. Udemy, as one of the world's leading online learning platforms, offers a vast and diverse catalog of courses spanning various disciplines. It caters to a global audience of millions of learners, ranging from individuals seeking to upskill for career advancement to lifelong learners pursuing their passions.

Our project will aim to tackle challenges like Information Overload, Personalization, Quality Assurance, and Scalability. To address these challenges and enhance the learning experience on Udemy, we propose the development of a recommendation system. This system will leverage Udemy's extensive dataset, including course details, user reviews, and curriculum content, to provide learners with personalized course suggestions.

By analyzing user behavior, preferences, and historical interactions, the recommendation engine will curate a list of courses that align with each learner's individual goals. Learners will receive tailored recommendations that not only save time but also improve the quality of their learning journeys.

By leveraging cutting-edge technologies, including distributed data processing, serverless computing, and efficient caching, we seek to provide Udemy users with a recommendation system that is not only accurate and personalized but also scalable and responsive to the platform's ever-growing user base.

The development of this recommendation engine is a response to the evolving needs of modern learners and Udemy's commitment to delivering a superior online learning experience. This project is poised to make a significant impact on how learners discover and engage with educational content on the platform, ultimately empowering them to achieve their learning objectives more effectively and efficiently.

3. Related Work and Gap Analysis:

In enhancing the online learning experience and developing a recommendation system for Udemy, several related work and research areas are personalized recommendation systems, which have been widely studied and implemented in various domains, including e-commerce, music, and video streaming platforms. Also, scalability and performance in online platforms; as online platforms like Udemy continue to grow, scalability and responsiveness become critical concerns. Some of the studies conducted in this field are as mentioned:

- W. Yunmin, C. Chaoying, A. YouLi, X. Min and P. K. Pareek, "Chinese Personalized Course Content Push Algorithm in Online Vocational Education Based on Big Data," 2023 International Conference on Distributed Computing and Electrical Circuits and Electronics (ICDCECE), Ballar, India, 2023, pp. 1-7, doi: 10.1109/ICDCECE57866.2023.10150656.
- Y. Saleh, N. El-Khalili, N. Otoum, M. A. -S. Hasan, S. Abu-Aishah and I. Matar, "Analysis of Teachers' Satisfaction With Online Learning During the Covid-19 Pandemic," 2022 29th International Conference on Systems, Signals and Image Processing (IWSSIP), Sofia, Bulgaria, 2022, pp. 1-4, doi: 10.1109/IWSSIP55020.2022.9854477.
- X. Pan, X. Li, and M. Lu, "A MultiView Courses Recommendation System Based on Deep Learning," 2020 International Conference on Big Data and Informatization

Education (ICBDIE), Zhangjiajie, China, 2020, pp. 502-506, doi: 10.1109/ICBDIE50010.2020.00124.

- C. Panse and S. Rastogi, "Impact of Online Teaching on the Placement in Professional Courses in India," 2022 2nd International Conference on Information Technology and Education (ICIT&E), Malang, Indonesia, 2022, pp. 102-106, doi: 10.1109/ICITE54466.2022.9759545.

Referring to the critical and significant aspects of the studies listed above, we have decided to implement advanced data processing techniques and integrate with a user-friendly front-end interface using Udemy's API.

4. Proposed Tasks:

4.1 Data Ingestion and Storage:
Tools/Technologies: Python (for scripting), Amazon S3 (for storage)
- Utilize Python scripts to interact with Udemy's API
- Fetch data from relevant endpoints (e.g., course details, reviews, curriculum)
- Store raw JSON data in Amazon S3 for scalable and efficient storage

4.2 Data Processing, Pre-processing and Analysis:
Technology: Python
- Clean and preprocess data, addressing missing values, duplicates, and erroneous records
- To better understand the dataset and the Udemy course ecosystem, we will conduct exploratory data analysis. This will involve generating various plots and visualizations to uncover insights, such as identifying the most in-demand courses, popular topics, and user preferences
- Extract key features required for recommendation algorithms
- Transform data into a suitable format for recommendation model training

4.3 Recommendation Engine Development:
Tools/Technologies: Python, scikit-learn, or TensorFlow (based on algorithm choice).
- Choose and implement recommendation algorithms (e.g., collaborative filtering, content-based filtering)
- Train and validate the recommendation model using a subset of the data
- Optimize the model for accuracy and relevance.

4.4 Serverless Backend and API Gateway:
Tools/Technologies: AWS Lambda, Amazon API Gateway
- Set up AWS Lambda functions to host recommendation engine logic

- Interface Lambda functions with Amazon API Gateway to provide endpoints for the front-end application

4.5 Caching for Enhanced Performance:
Tools/Technologies: Not decided yet
- Cache frequent database queries and recommendation results
- Periodically update the cache to ensure up-to-date recommendations

4.6 Front-End Development:
Tools/Technologies: Not decided yet
- Sketch the initial wireframes detailing the user flow and interface elements.
- Design high-fidelity mockups using tools like Figma or Adobe XD.
- Develop the interface components using React.js or another chosen framework.
- Integrate the front-end application with the serverless backend using API calls.

5. **Team members and workload allocation:**

Nikhil Prashant Kulkarni:	4.3	4.5
Venkhatesh Arunachalam:	4.1	4.5
Raj Patel:	4.2	4.6

6. **Planned timeline:**

End of September: Project Kick-off and Planning
Mid-October: Data Ingestion and Storage Setup - Develop Python scripts for data ingestion from Udemy's API, implement data storage in Amazon S3, organizing data into logical buckets, begin initial data collection
End of October: Data Processing and Pre-processing, EDA, Recommendation Engine Development
Mid-November: Serverless Backend, API Gateway, and Front-End Development
15 Nov – 25 Nov: QA + Testing
End of November: Complete final stages, Project Closure