

**Skills
Network**

Capstone Overview

In this capstone course, you will apply various machine learning knowledge and skills that you have learned as part of the previous courses to solve some real-world industrial challenges.

Project Scenario

Assume you are a new machine learning engineer in a Massive Open Online Courses (MOOCs) startup called AI Training Room. In AI Training Room, learners across the world can learn leading technologies such as Machine Learning, AI, Data Science, Cloud, App development, etc. Your company grows rapidly and reaches millions of learners in a very short period.

The learning topics of AI Training Room can be summarized in the following word cloud:



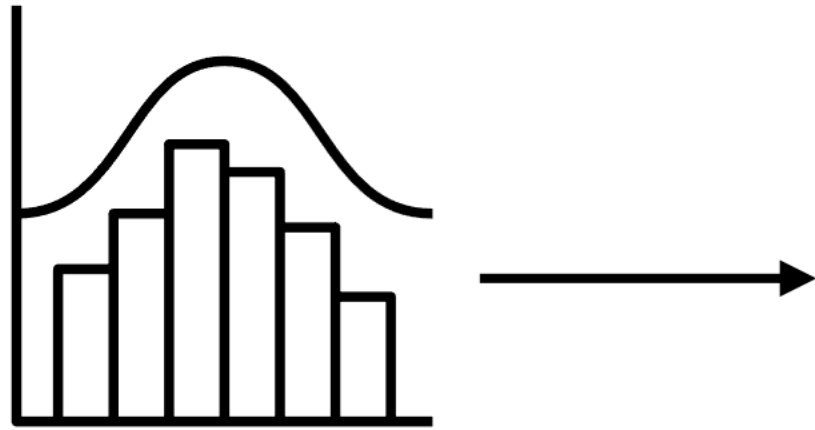
Starting this year, your machine learning engineer team is working very hard on a recommender system project. The main goal of this project is to improve learners' learning experience

via helping them quickly find new interested courses and better paving their learning paths. Meanwhile, with more learners interacting with more courses via your recommender systems, your company's revenue may also be increased.

This project is currently at the Proof of Concept (PoC) phase so your main focus at this moment is to explore and compare various machine learning models and find one with the best performance in off-line evaluations.

Your Tasks

Your tasks in this project are summarized in the following workflow, and you will be guided through them in hands-on labs.



Exploratory Data Analysis



Raw Data



Feat

More specifically, you will undertake the tasks of:

- Collecting and understanding data
- Performing exploratory data analysis on online course enrollments datasets
- Extracting Bag of Words (BoW) features from course textual content
- Calculating course similarity using BoW features
- Building content-based recommender systems using various unsupervised learning algorithms, such as:
 - Distance/Similarity measurements, K-means, Principal Component Analysis (PCA), etc.
- Building collaborative-filtering recommender systems using various supervised learning algorithms
 - K Nearest Neighbors, Non-negative Matrix Factorization (NMF), Neural Networks, Linear Regression, Logistic Regression, RandomForest, etc.
- Creating an insightful and informative slideshow and presenting it to your peers

If you have extra bandwidth, you can also deploy and demonstrate your models via a web app built with `streamlit`. `Streamlit` is an open-source app framework for Machine Learning and Data Science to quickly demonstrate their works.

Your course recommender app where you select different recommendation models and generate recommendations, may look like the following screenshot:



Personalized Learning Recommender

1. Select recommendation models

Select model:

Course Similarity

2. Tune Hyper-parameters:

Top courses

10

0100

Course Similarity Threshold %

50

0100

3. Training:

Train Model

4. Prediction

Recommend New Courses

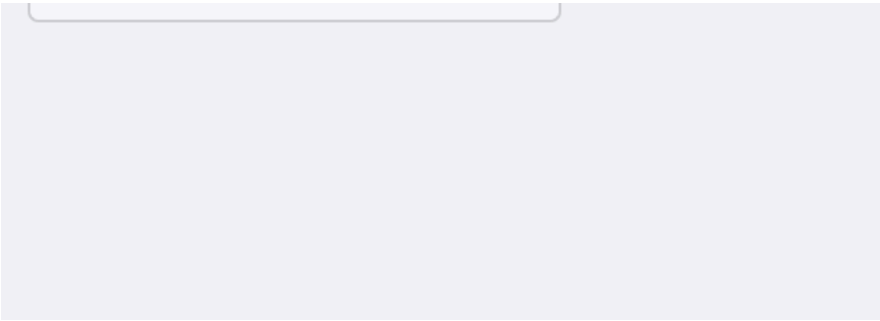
Select course

COURSE_ID	
<input type="checkbox"/>	DW0101EN
<input type="checkbox"/>	ML0111EN
<input type="checkbox"/>	GPXX0ZMZEN
<input type="checkbox"/>	ML0101EN
<input type="checkbox"/>	ML0109EN
<input type="checkbox"/>	ML0101ENv3
<input type="checkbox"/>	ML0151EN
<input type="checkbox"/>	excourse21
<input type="checkbox"/>	excourse40
<input checked="" type="checkbox"/>	excourse46
<input checked="" type="checkbox"/>	excourse47
<input type="checkbox"/>	excourse48
<input type="checkbox"/>	excourse49

Your courses

	COURSE_ID
0	excourse46
1	excourse47

Recommendation



	SCORE	TITI
0	0.6893	Mac Witl

This project is a great opportunity to showcase your machine learning skills, and demonstrate your proficiency to potential employers.

Grading Schema

- Graded Quizzes: **30 pts**
- Final presentation, peer-review: **70 pts**

Development Environments

In this project, you have at least three development environments you may choose from:

Skills Network Labs

Skills Network Labs is a virtual lab environment reserved for the exclusive use by the learners on IBM Developer Skills Network portals and its partners.

Use your Python and Jupyter Environment

If you experience any issues with the above two cloud environments, you may install Python and JupyterNotebook / JupyterLab on your own environments like a desktop or laptop computer. All the notebooks and data used in the capstone can be downloaded and executed locally.

Next Steps

Now you should have a basic understanding of this capstone project.

In the next step of your project, you will start with collecting and exploring the datasets.

Author(s)

[Yan Luo](#)

Other Contributor(s)

Changelog

Date	Version	Changed by	Change Description
2023-05-11	1.3	Eric Hao & Vladislav Boyko	Updated Page Frames
2023-05-10	1.2	Eric Hao & Vladislav Boyko	Updated Page Frames
2023-05-10	1.1	Eric Hao & Vladislav Boyko	Updated Page Frames
2022-03-18	1.0		Initial version created