

Build a Personalized Online Course Recommender System with Machine Learning

Tim Jhung
July 11, 2025

The image shows a grid of 10 course cards, likely from a platform like Coursera or IBM Skills Network. Each card contains a thumbnail, the course title, the provider, start date, course type, rating, and number of reviews. To the right of the grid, there are three stylized human icons connected by dashed arrows, suggesting a personalized learning path or recommendation system.

Course Title	Provider	Starts	Type	Rating	Reviews
Introduction to Machine Learning with Sound	IBM DW0101EN - v1.2	Any time, Self-paced	Course	4.2/5	(103)
Deep Learning Fundamentals	DeepLearning.TV ML0115EN - v1.0	Any time, Self-paced	Course	No ratings yet	
Accessing Hadoop Data Using Hive	IBM BD0141EN - v2016.0	Any time, Self-paced	Course	4.5/5	(60)
MapReduce and YARN	Big Data University BD0115EN - v2016.0	Any time, Self-paced	Course	No ratings yet	
Apply end to end security to a cloud application	IBMDeveloperSkillsNetwork SEC03EN - v1.0	April 26, 2019	Course	No ratings yet	
Serverless Computing using Cloud Functions ...	IBMDeveloperSkillsNetwork CC0210EN - v1.0	Any time, Self-paced	Course	No ratings yet	
IBM Blockchain Foundation Developer	IBMDeveloperSkillsNetwork BC0201EN - v2.0	Any time, Self-paced	Course	No ratings yet	
Moving Data into Hadoop	Big Data University BD0131EN - v2016.0	Any time, Self-paced	Course	No ratings yet	

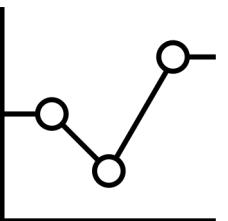
Outline

- Introduction and Background
- Exploratory Data Analysis
- Content-based Recommender System using Unsupervised Learning
- Collaborative-filtering based Recommender System using Supervised learning
- Conclusion
- Appendix

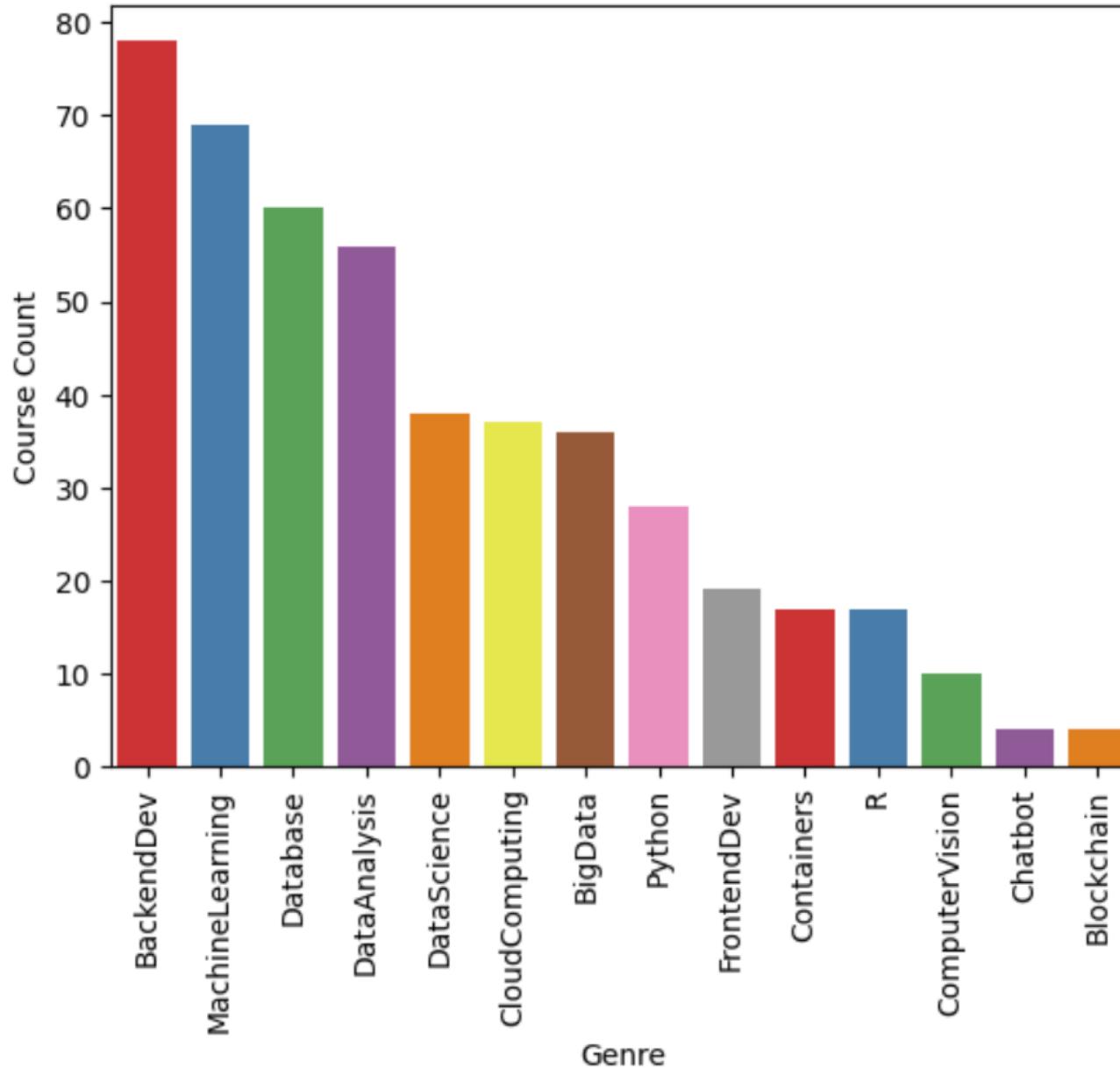
Introduction

- A Massive Open Online Courses (MOOCs) startup, AI Training Room, provides learners across the world with courses to learn leading technologies such as Machine Learning, AI, Data Science, Cloud, App development, etc.
- AI Training Room's machine learning engineering team is working very hard on a recommender system project. The main goal of this project is to improve learners' learning experience by helping them quickly find new interested courses and better paving their learning paths.
- This project is currently at the Proof of Concept (PoC) phase and the 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.

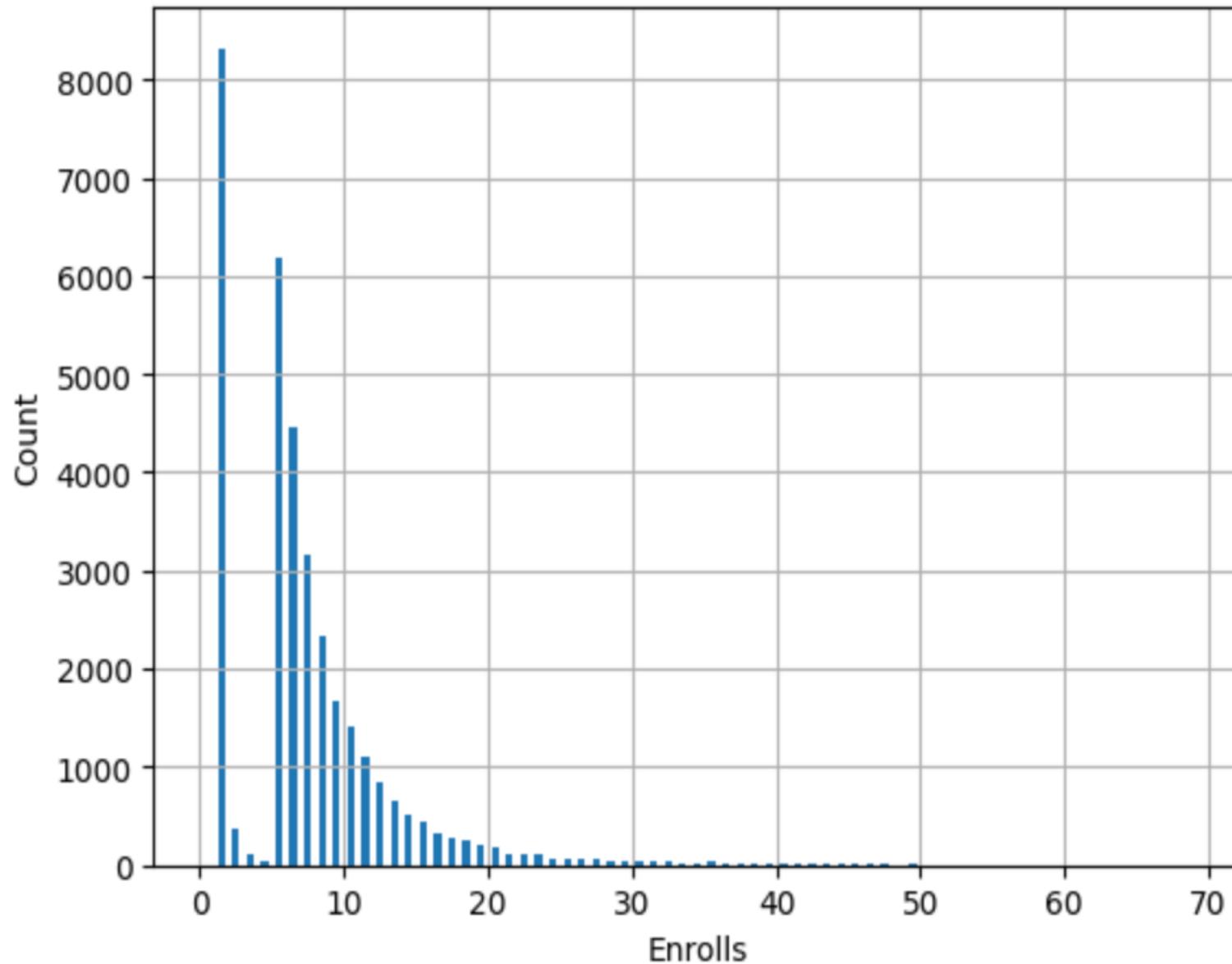
Exploratory Data Analysis



Course counts per genre



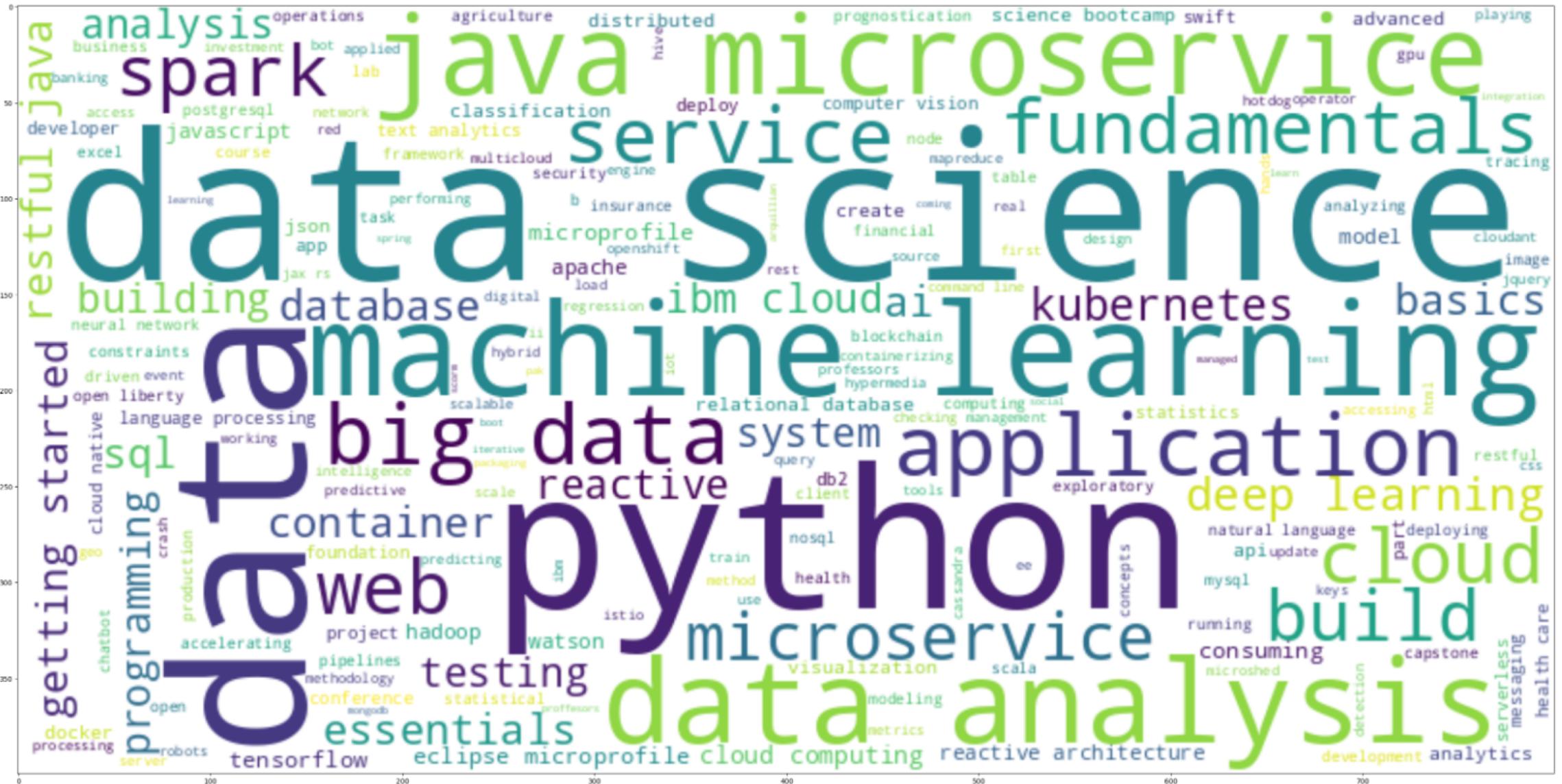
Course enrollment distribution



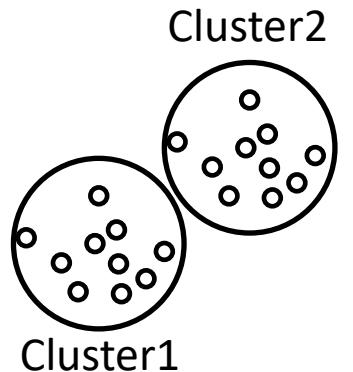
20 most popular courses

	TITLE	Enrolls
0	python for data science	14936
1	introduction to data science	14477
2	big data 101	13291
3	hadoop 101	10599
4	data analysis with python	8303
5	data science methodology	7719
6	machine learning with python	7644
7	spark fundamentals i	7551
8	data science hands on with open source tools	7199
9	blockchain essentials	6719
10	data visualization with python	6709
11	deep learning 101	6323
12	build your own chatbot	5512
13	r for data science	5237
14	statistics 101	5015
15	introduction to cloud	4983
16	docker essentials a developer introduction	4480
17	sql and relational databases 101	3697
18	mapreduce and yarn	3670
19	data privacy fundamentals	3624

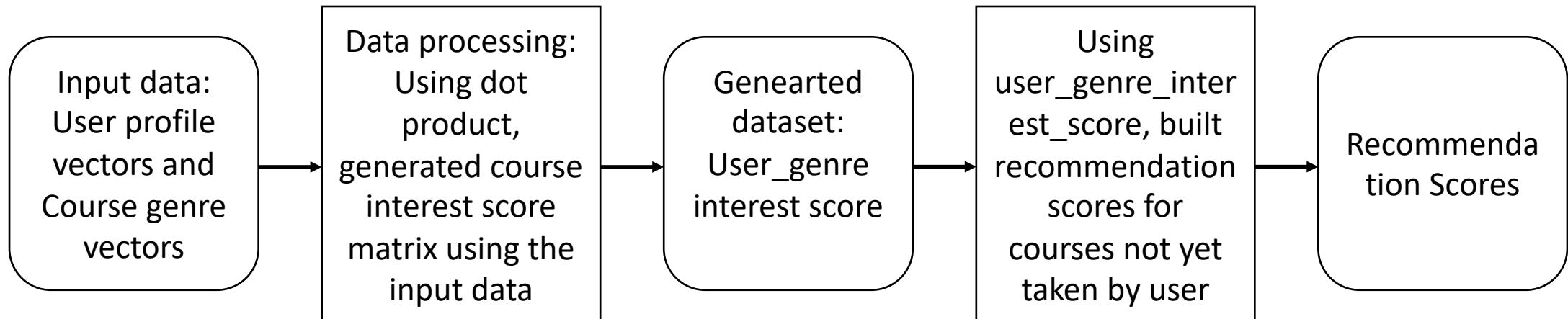
Word cloud of course titles



Content-based Recommender System using Unsupervised Learning



Flowchart of content-based recommender system using user profile and course genres



Evaluation results of user profile-based recommender system

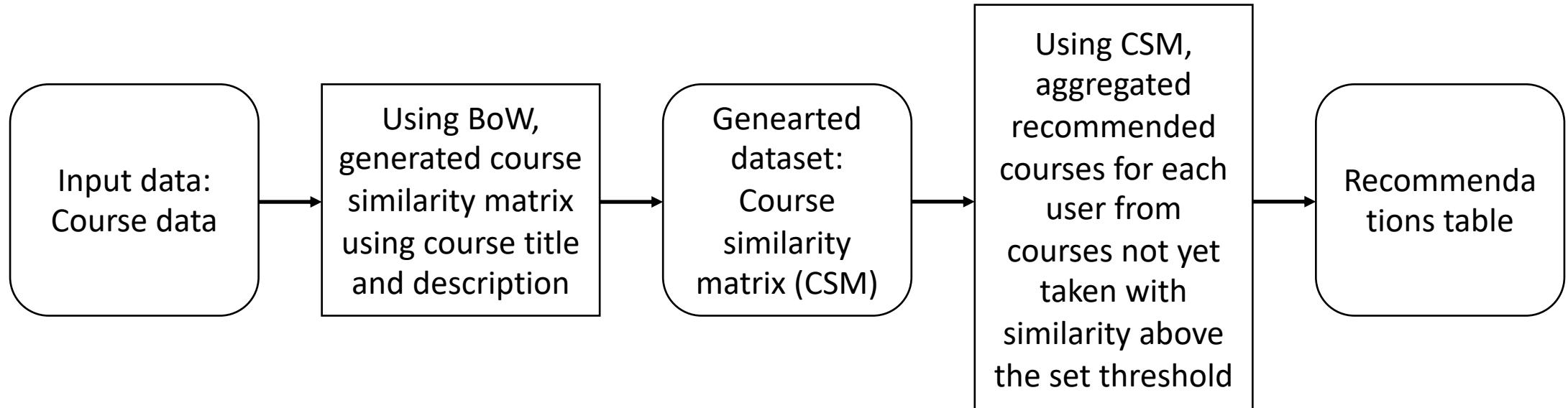
Using a recommendation score threshold of 10, the following results have been observed.

On average, 9 new/unseen courses have been recommended per user (in the test user dataset)

Top 10 commonly recommended courses

	course id	recommendation count
1	ML0122EN	27824
59	TA0106EN	17390
155	ML0101EN	16451
234	excourse21	15656
235	excourse22	15656
38	GPXX0IBEN	15644
219	excourse06	15062
217	excourse04	15062
35	RP0103	14921
186	GPXX0TY1EN	14689

Flowchart of content-based recommender system using course similarity



Evaluation results of course similarity based recommender system

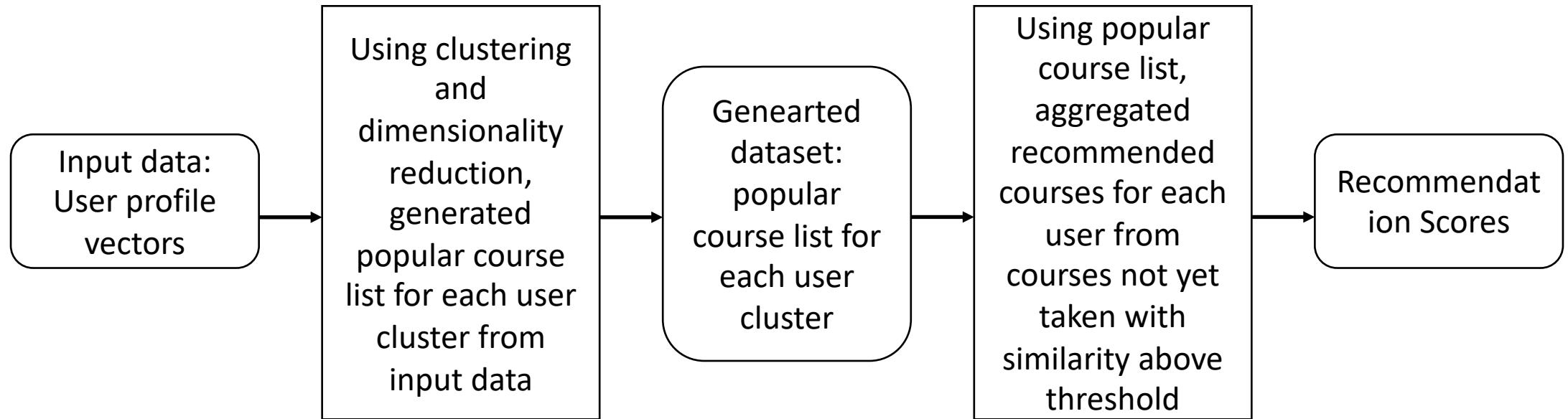
Using a similarity score threshold of 0.6, the following results have been observed.

On average, 9 new/unseen courses have been recommended per user (in the test user dataset)

Top-10 commonly recommended courses

course id	recommendation count
131 DS0110EN	15003
235 excourse22	14937
275 excourse62	14937
276 excourse63	14641
278 excourse65	14641
281 excourse68	13551
285 excourse72	13512
280 excourse67	13291
287 excourse74	13291
18 BD0145EN	12497

Flowchart of clustering-based recommender system



Evaluation results of clustering-based recommender system

Your hyper-parameter settings, such as a score or similarity threshold

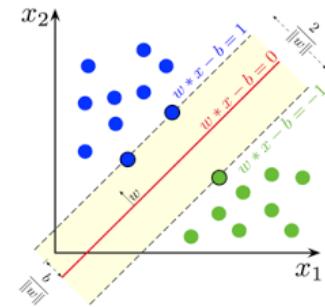
Note if you have tried multiple hyper-parameters, you may show your results in a grouped bar chart

On average, 18 new/unseen courses have been recommended per user (in the test user dataset)

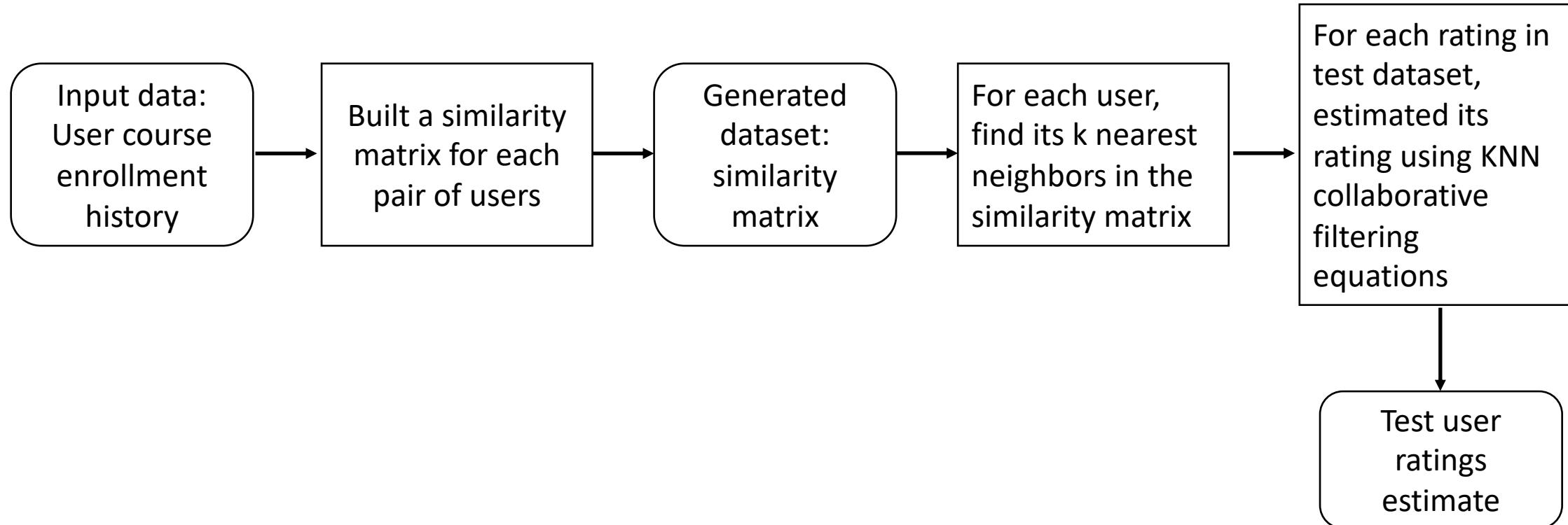
Top-10 commonly recommended courses

	course id	recommendation count
106	ST0101EN	28101
142	ML0115EN	27072
53	WA0101EN	26825
178	DS0301EN	26284
81	BC0101EN	26129
134	DS0105EN	25989
135	DS0103EN	25655
191	DB0101EN	25467
203	RP0101EN	24545
181	BD0111EN	22860

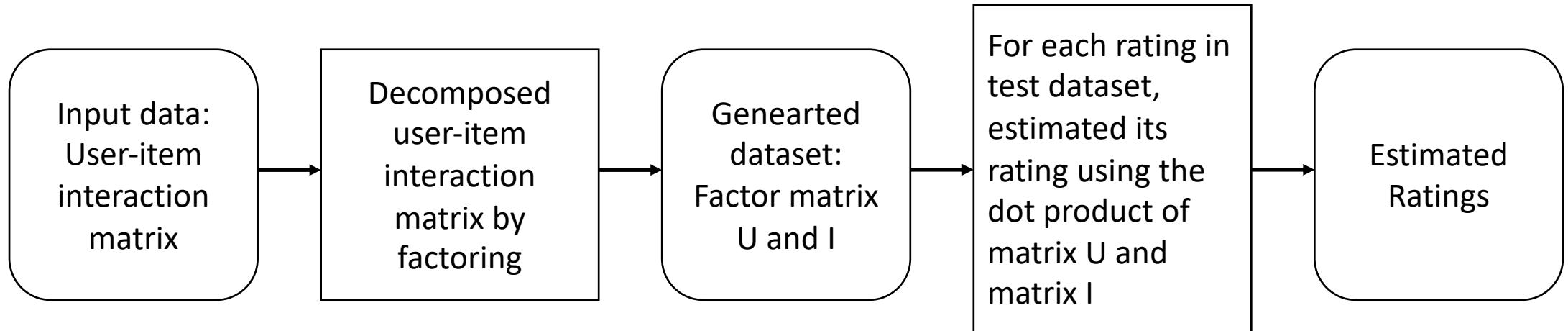
Collaborative-filtering Recommender System using Supervised Learning



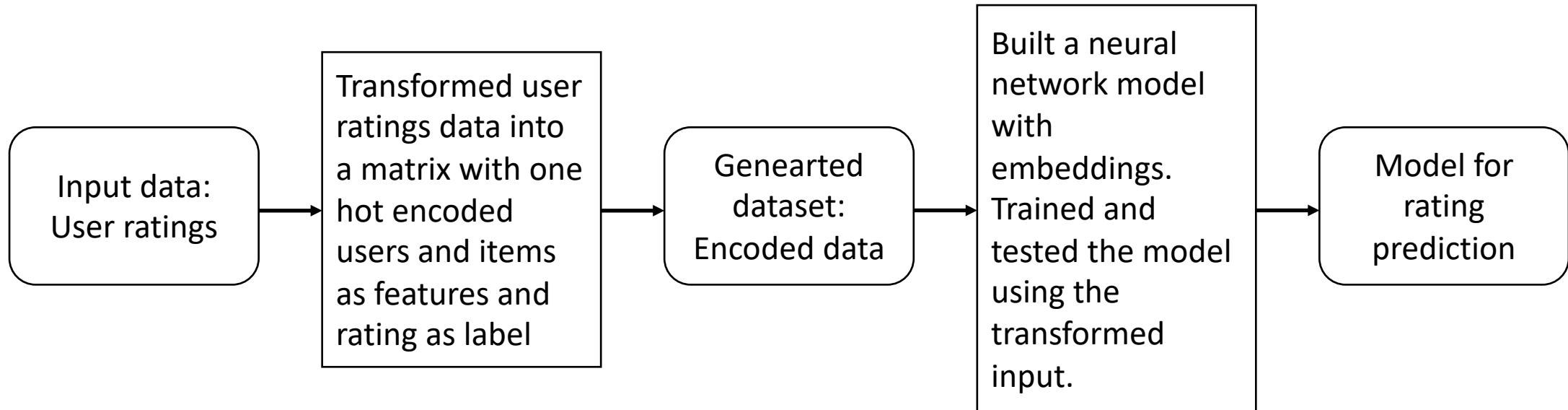
Flowchart of KNN based recommender system



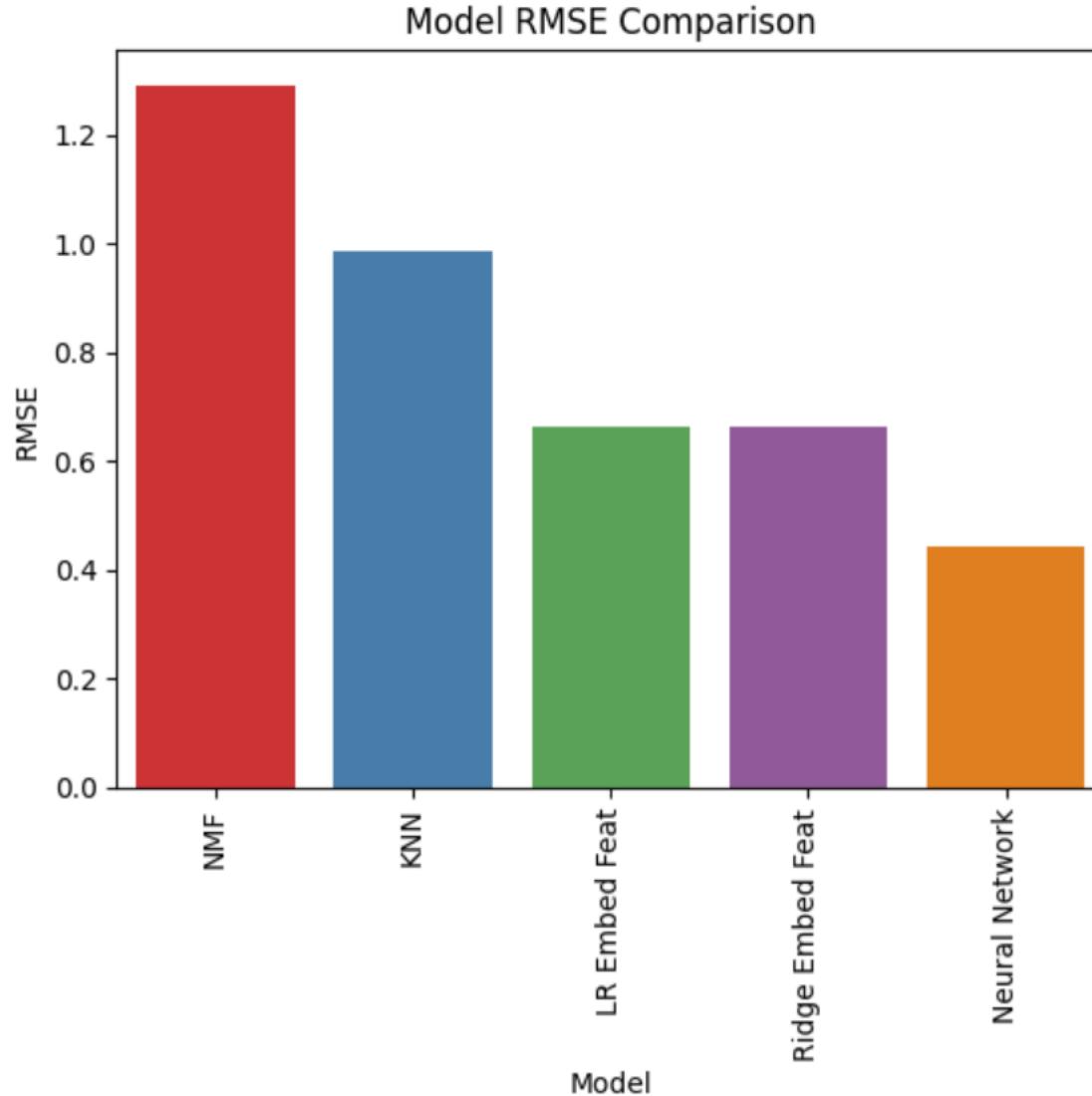
Flowchart of NMF based recommender system



Flowchart of Neural Network Embedding based recommender system



Compare the performance of collaborative-filtering models



Conclusions

- Among the different models tested, a neural network (NN) showed the best performance for course rating prediction as measured by root mean squared error.
- As a neural network model does not make assumptions about features, using NN model could also obviate the need for manual feature selection in model building which could be an advantage.
- If additional resource and time allow, further exploration of regression based or classification based models using embedding features from NN model might improve performance as those models build on the embedding layer weights from NN model.
- Based on the PoC, it is recommended that AI Training Room ML engineering team pursue utilizing NN model in its course recommendation system project.

Appendix

- Exploratory Data Analysis [notebook in github](#)
- Content-based recommender system using user profile and course genres [notebook in github](#)
- content-based recommender system using course similarity [notebook in github](#)
- clustering-based recommender system [notebook in github](#)
- KNN based recommender system [notebook in github](#)
- NMF based recommender system [notebook in github](#)
- Neural Network Embedding based recommender system [notebook in github](#)