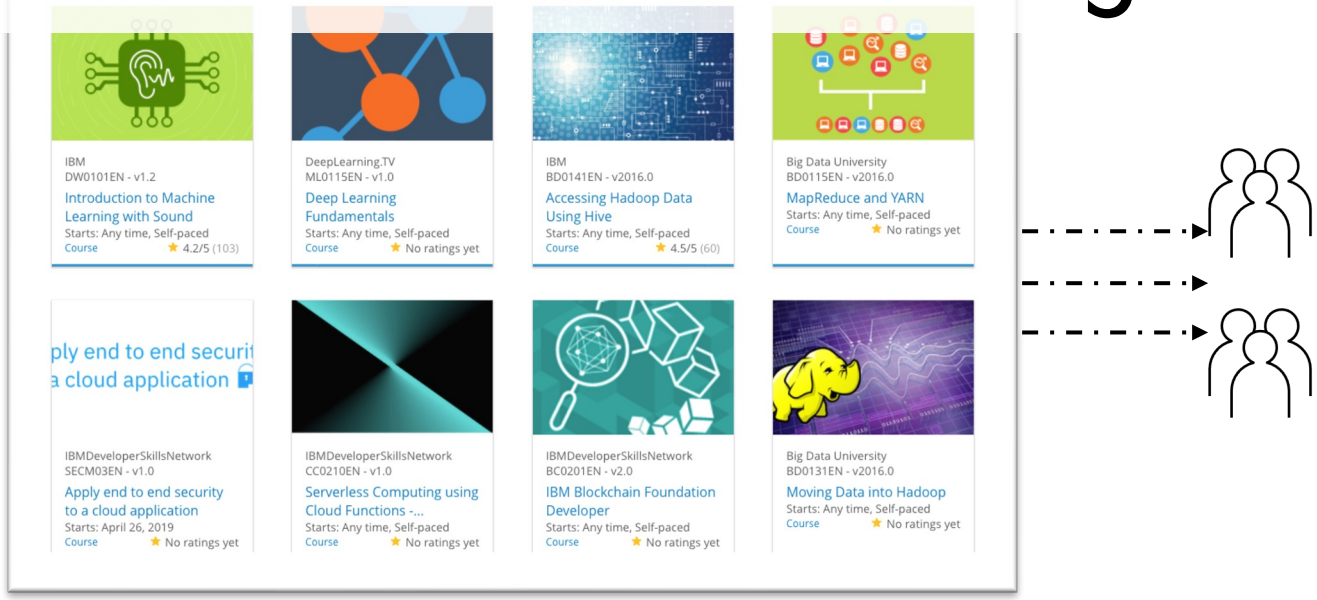
Build a Personalized Online Course Recommender System with Machine Learning

Mohammed Qasim Karjatwala

31st October 2022



Outline

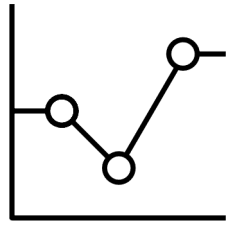
- Introduction and Background
- Exploratory Data Analysis
- Content-based Recommender System using Unsupervised Learning
- Collaborative-filtering based Recommender System using Supervised learning
- Conclusion
- Appendix

Introduction

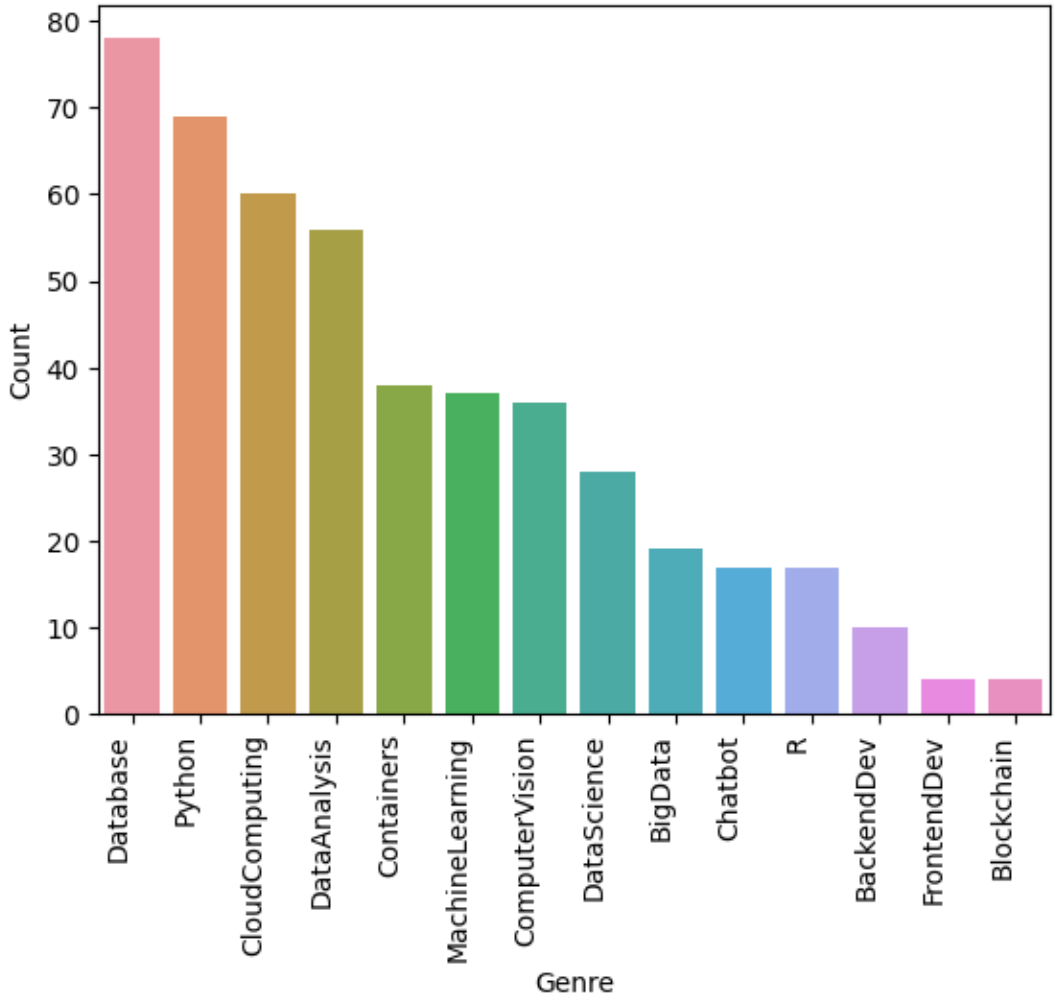
- AI Training Room is a a Massive Open Online Courses (MOOCs) startup, reaching millions of learners. It is working on a recommender system project. The main goal of the project is to help learners quickly find new interested courses. With learners interacting with more courses via the recommender systems, the company's revenue may also be increased.

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

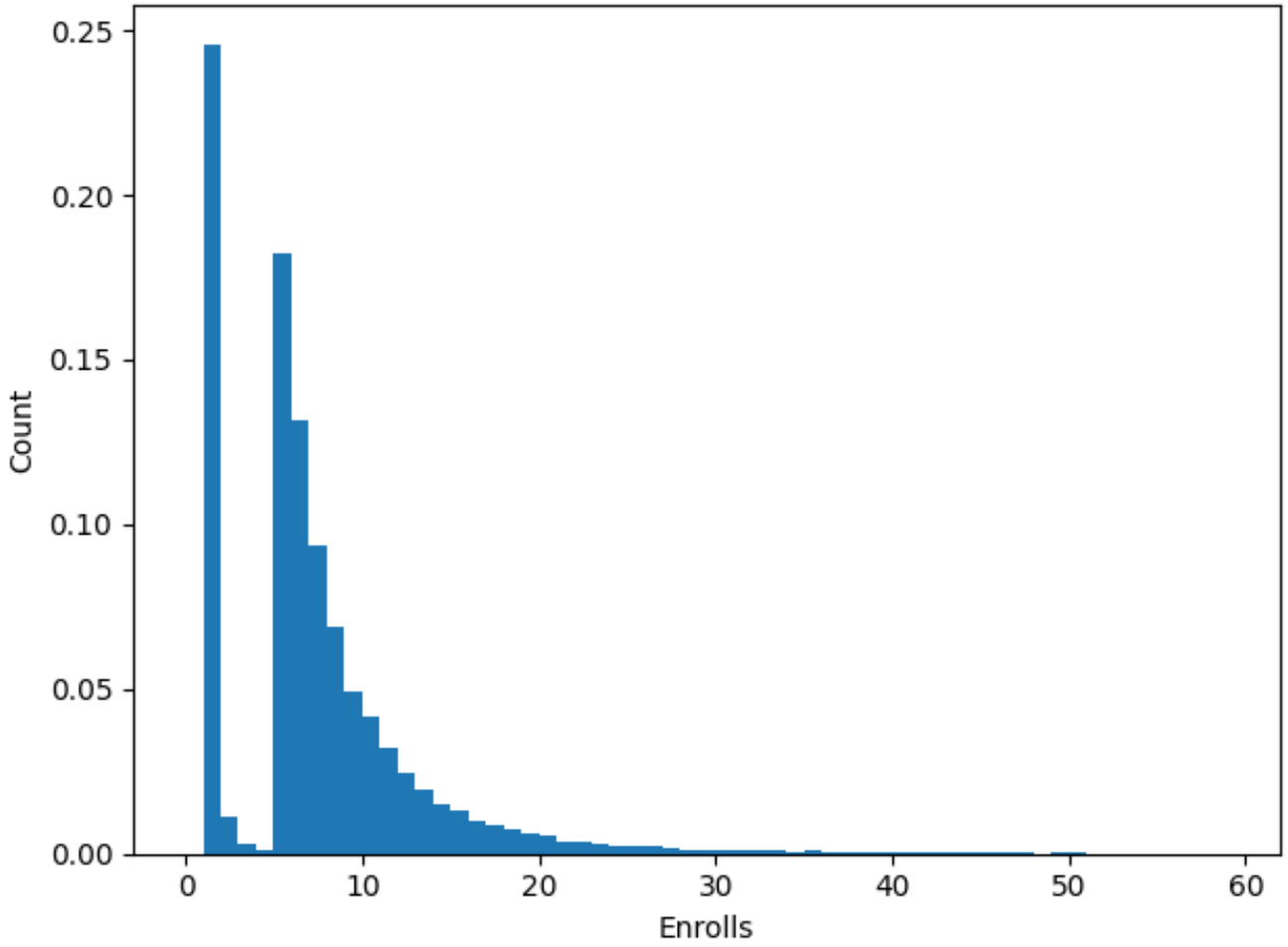
Exploratory Data Analysis



Course counts per genre



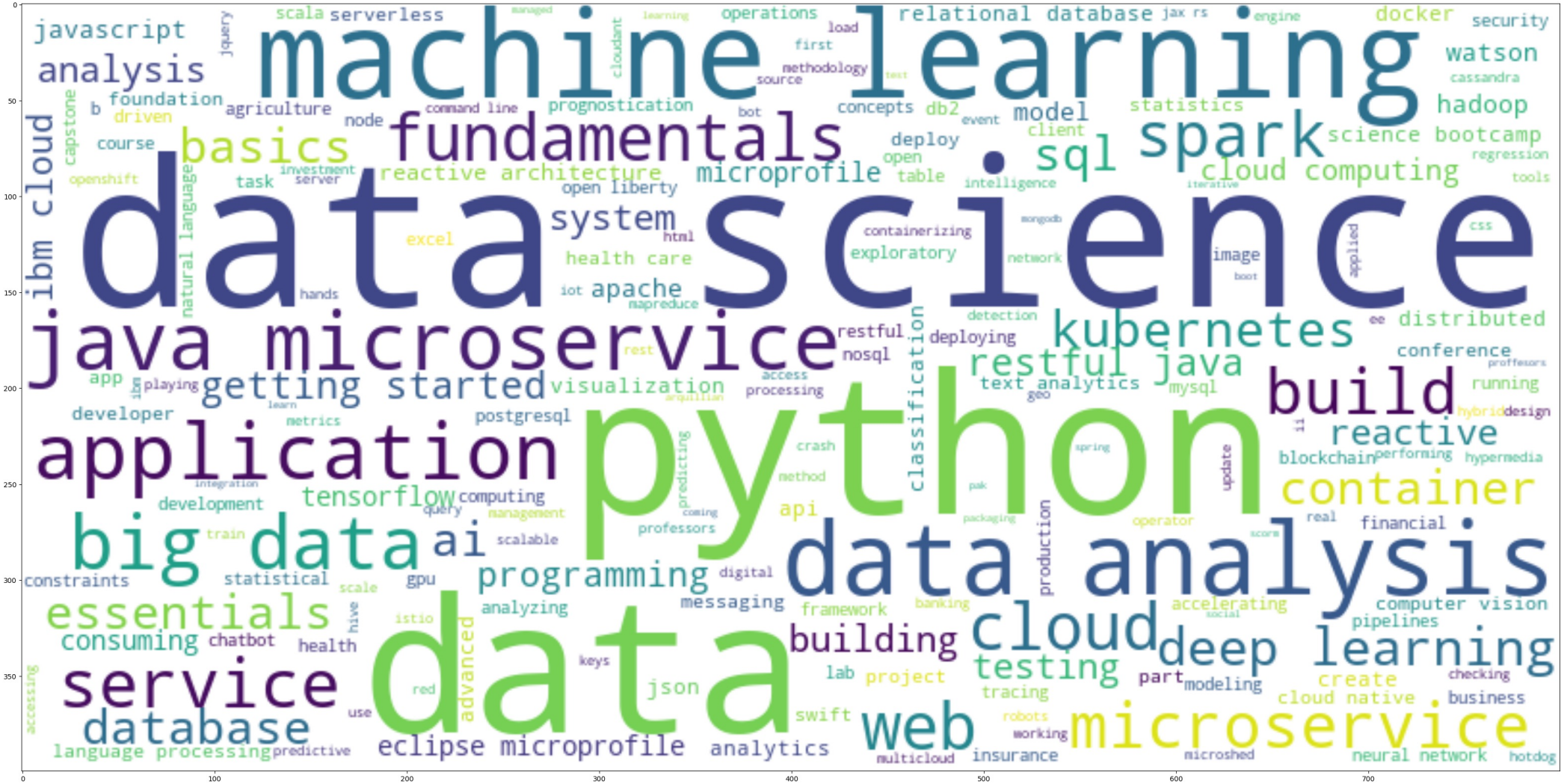
Course enrollment distribution



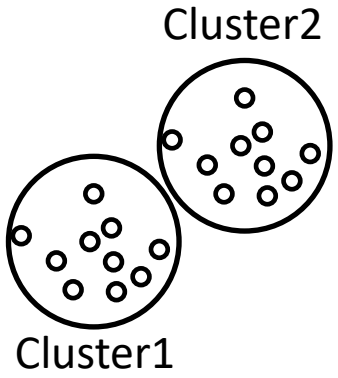
20 most popular courses

	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

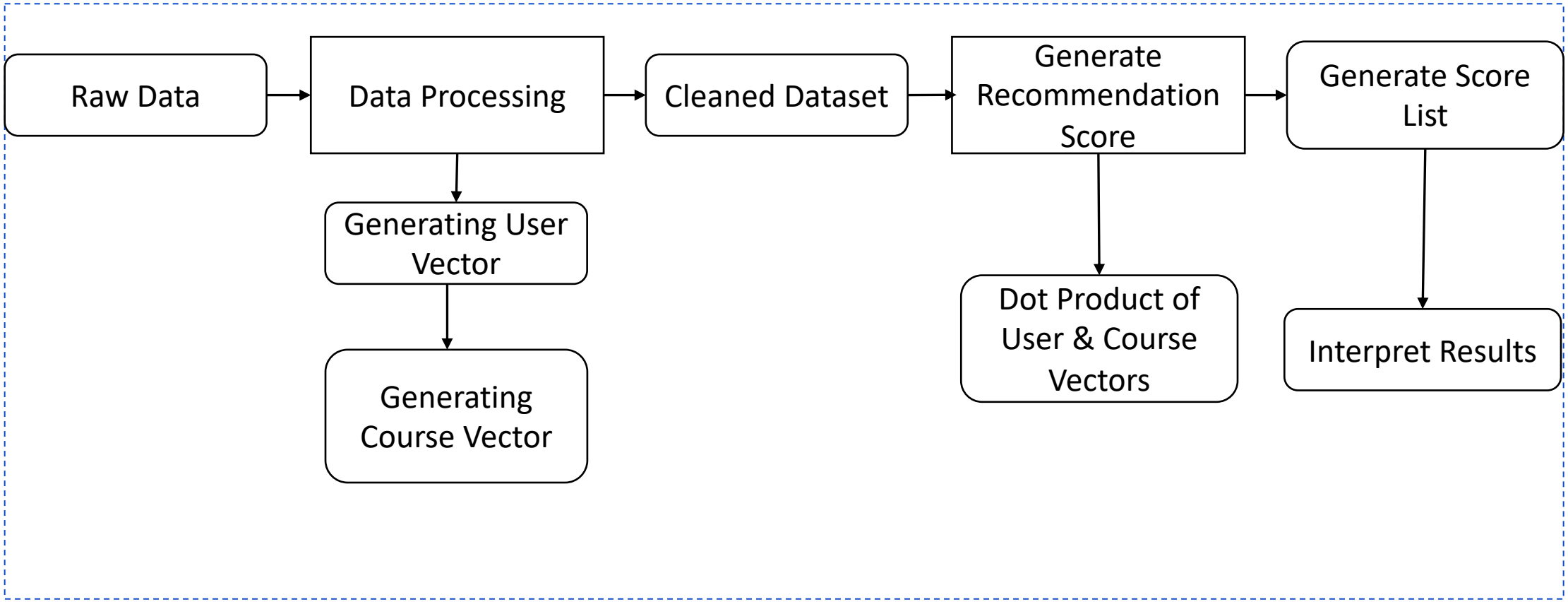
Word cloud of course titles



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

Courses with a recommendation score over a *score_threshold of* 10.0 were only recommended for each test user

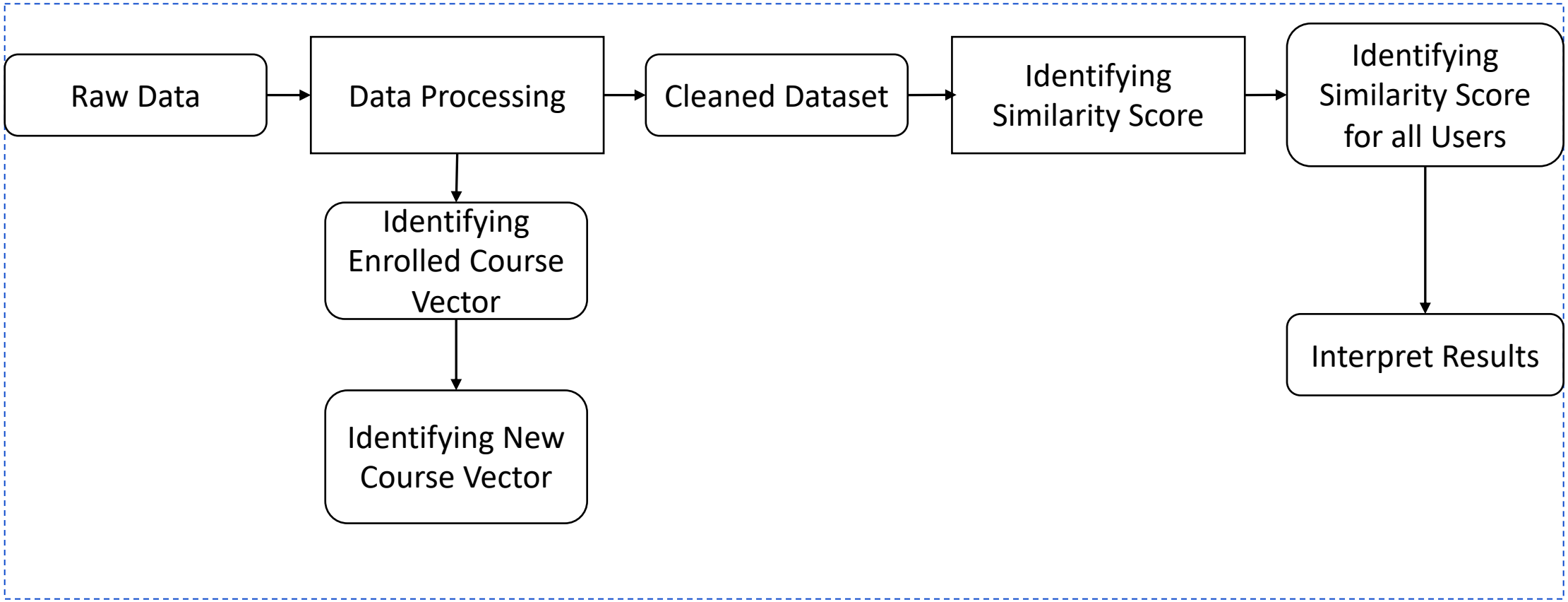
On average, how many new/unseen courses have been recommended per user?

66.923 ~ 67 courses

What are the top 10 most frequently recommended courses?

RP0103EN	962
RP0103	927
RP0105EN	914
BD0133EN	900
CC0271EN	890
GPXX06RFEN	886
excourse74	867
BD0135EN	859
excourse73	842
BD0145EN	835

Flowchart of content-based recommender system using course similarity



Evaluation results of course similarity based recommender system

Courses with a similarity score over a *sim_threshold of* 0.6 were only recommended for each test user

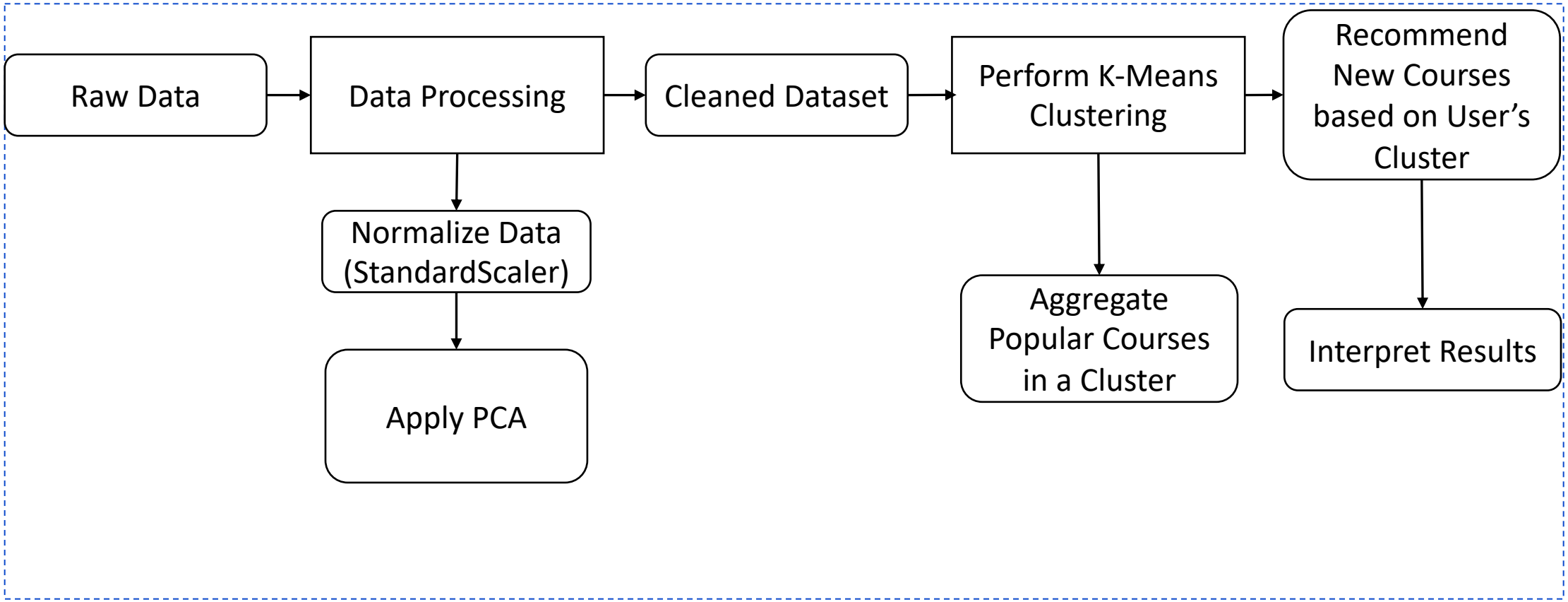
On average, how many new/unseen courses have been recommended per user?

67.755 ~ 68 courses

What are the top 10 most frequently recommended courses?

SC0101EN	1000
RP0101EN	999
ML0120EN	998
ML0122ENv1	998
ML0120ENv2	998
ML0120ENv3	998
TMP0105EN	997
TA0105	996
BD0151EN	996
TA0105EN	996

Flowchart of clustering-based recommender system



Evaluation results of clustering-based recommender system

Courses with enrollments over a *enrollment_threshold* of 10 were only recommended for each test user

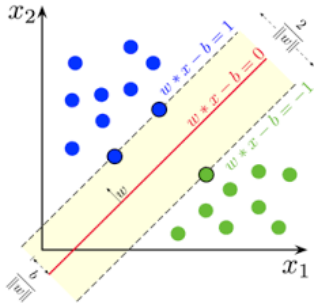
On average, how many new/unseen courses have been recommended per user?

43.725 ~ 44 courses

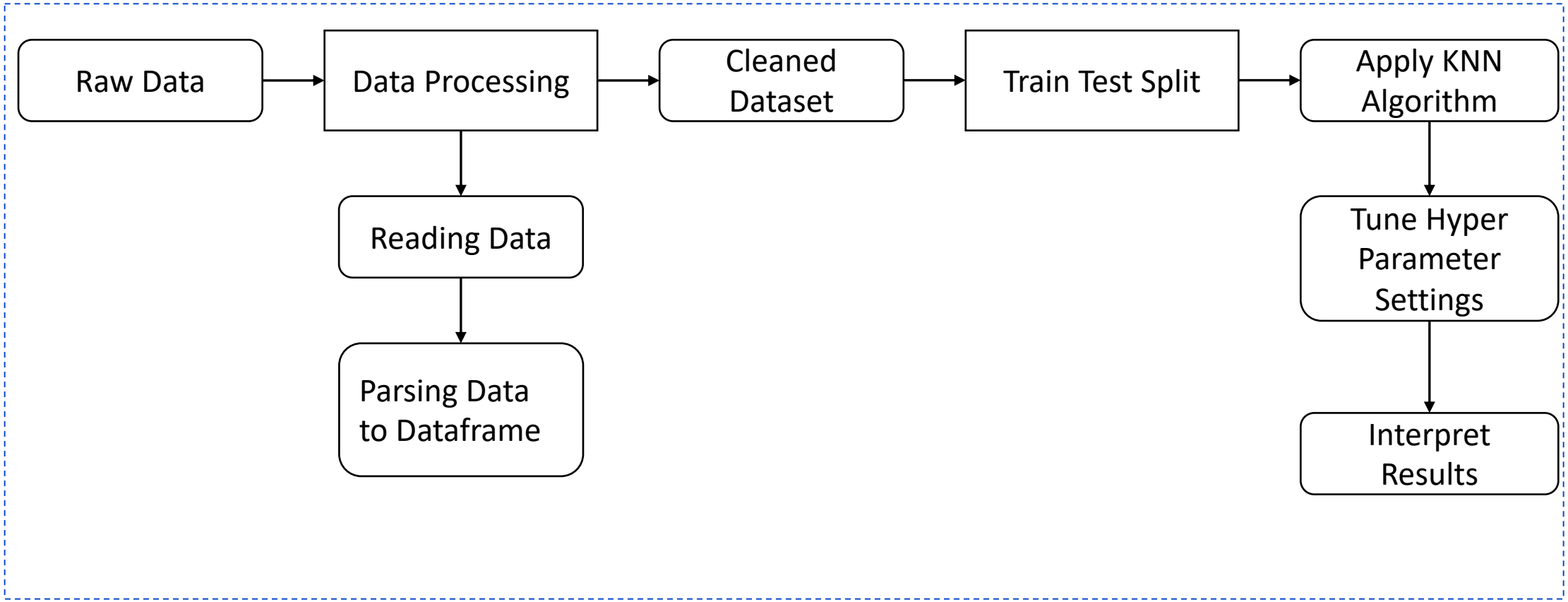
What are the top 10 most frequently recommended courses?

DS0101EN	1758
BD0101EN	1526
PY0101EN	1447
CC0101EN	1350
DA0101EN	1338
CC0201EN	1256
CC0103EN	1213
DS0105EN	1210
DS0103EN	1188
ML0101ENv3	1171

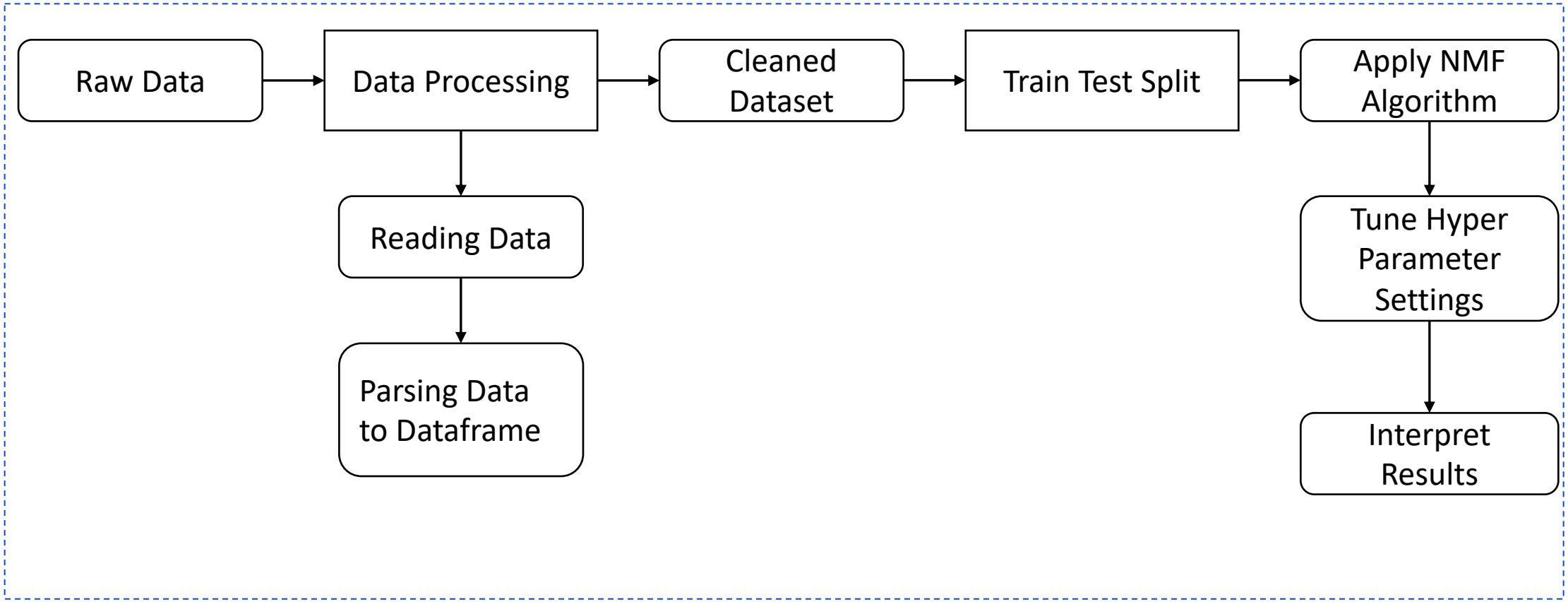
Collaborative-filtering Recommender System using Supervised Learning



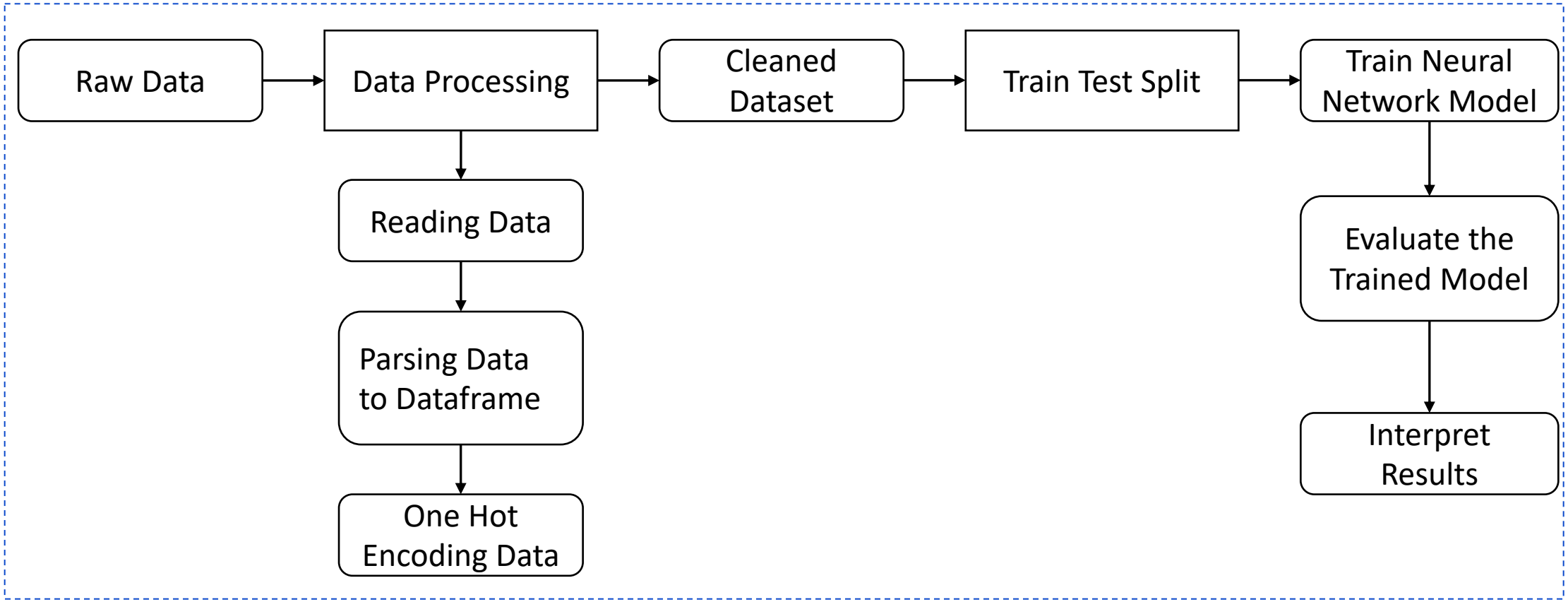
Flowchart of KNN based recommender system



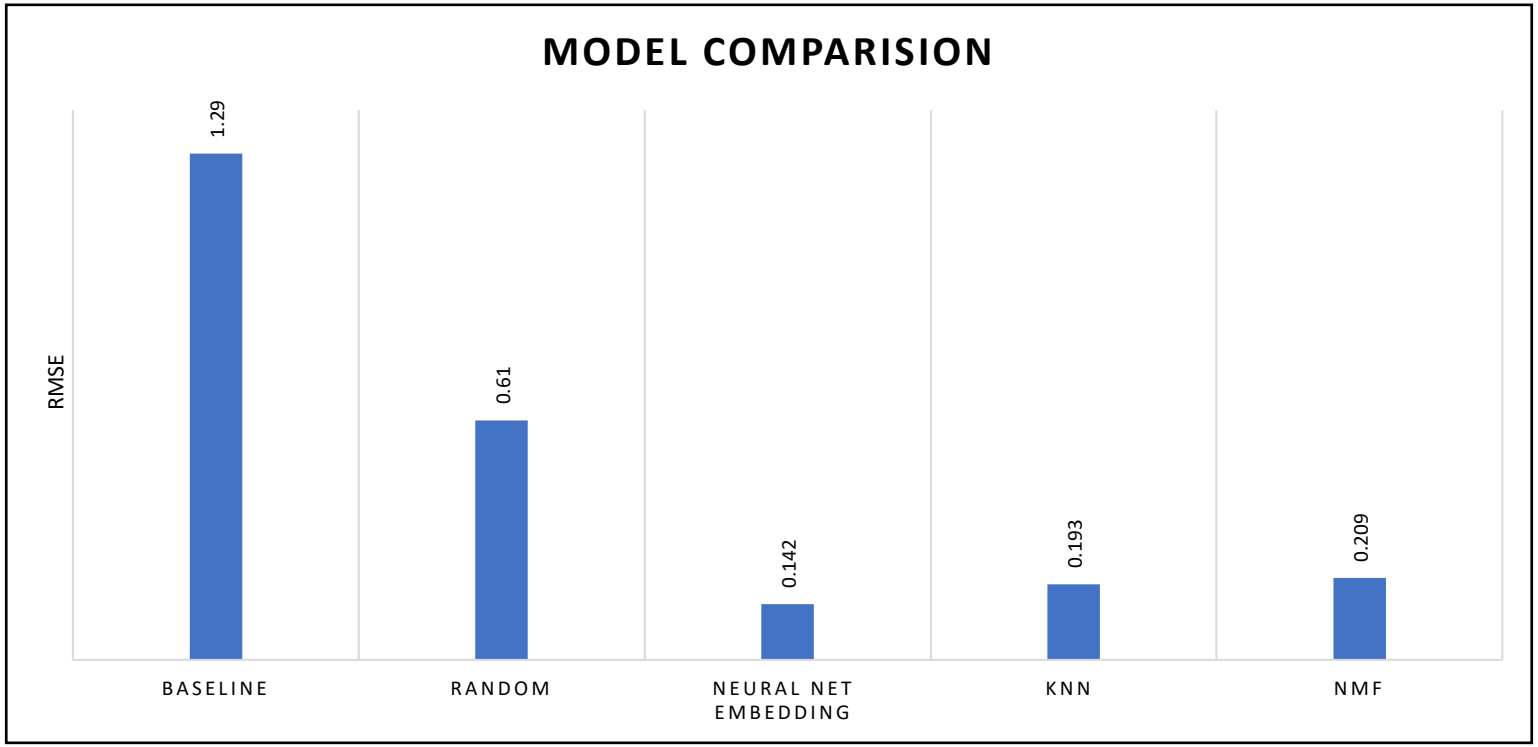
Flowchart of NMF based recommender system



Flowchart of Neural Network Embedding based recommender system



Compare the performance of collaborative-filtering models



Conclusions

- Content, User and Course similarity based recommender systems provide with plausible list of courses to suggest to the user on average recommend more than 44 courses per user.
- KNN, NMF and other models used provide an accurate course rating prediction.
- The recommended courses combined with an accurate rating prediction can be used to recommend a couple courses to the user instead of an enormous list.

Appendix

- https://github.com/moqa19/IBM-Machine-Learning/tree/main/ML%206%20Upload