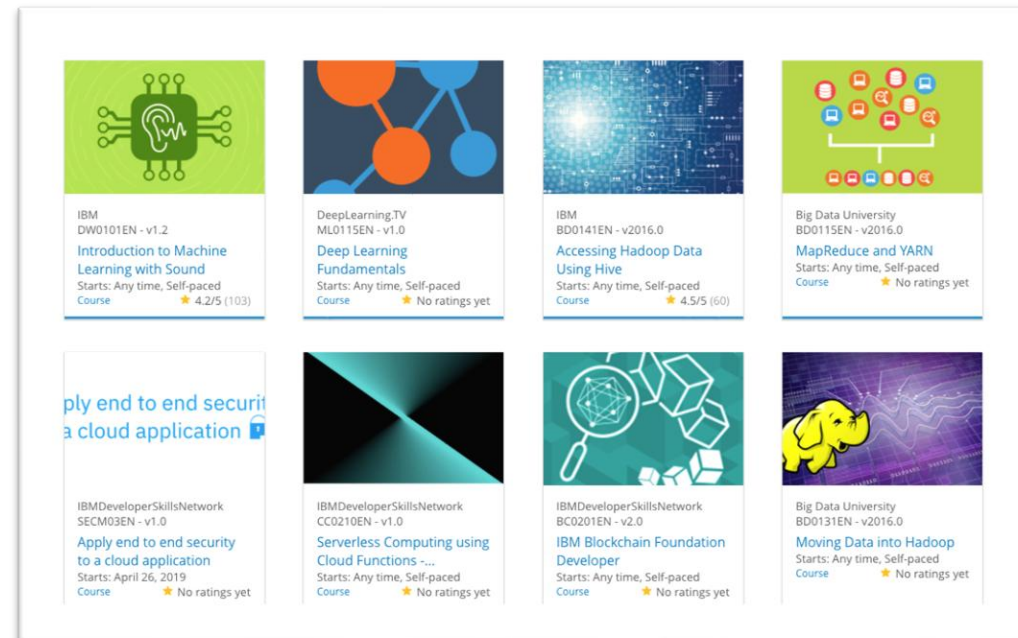


Build a Personalized Online Course Recommender System with Machine Learning

Rodrigo Senna Muhlbauer
23/12/2024



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

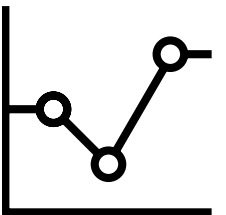
- Background and context

A course recommendation system is designed to help users discover better courses that align with their individual interests. The system aims to recommend the most suitable courses by considering users' personal preferences, their friends' interests, and the courses they are currently enrolled in

- Problem states and hypotheses

However, building such a system comes with challenges, including the variety of possible approaches and the differing assumptions inherent in each methodology.

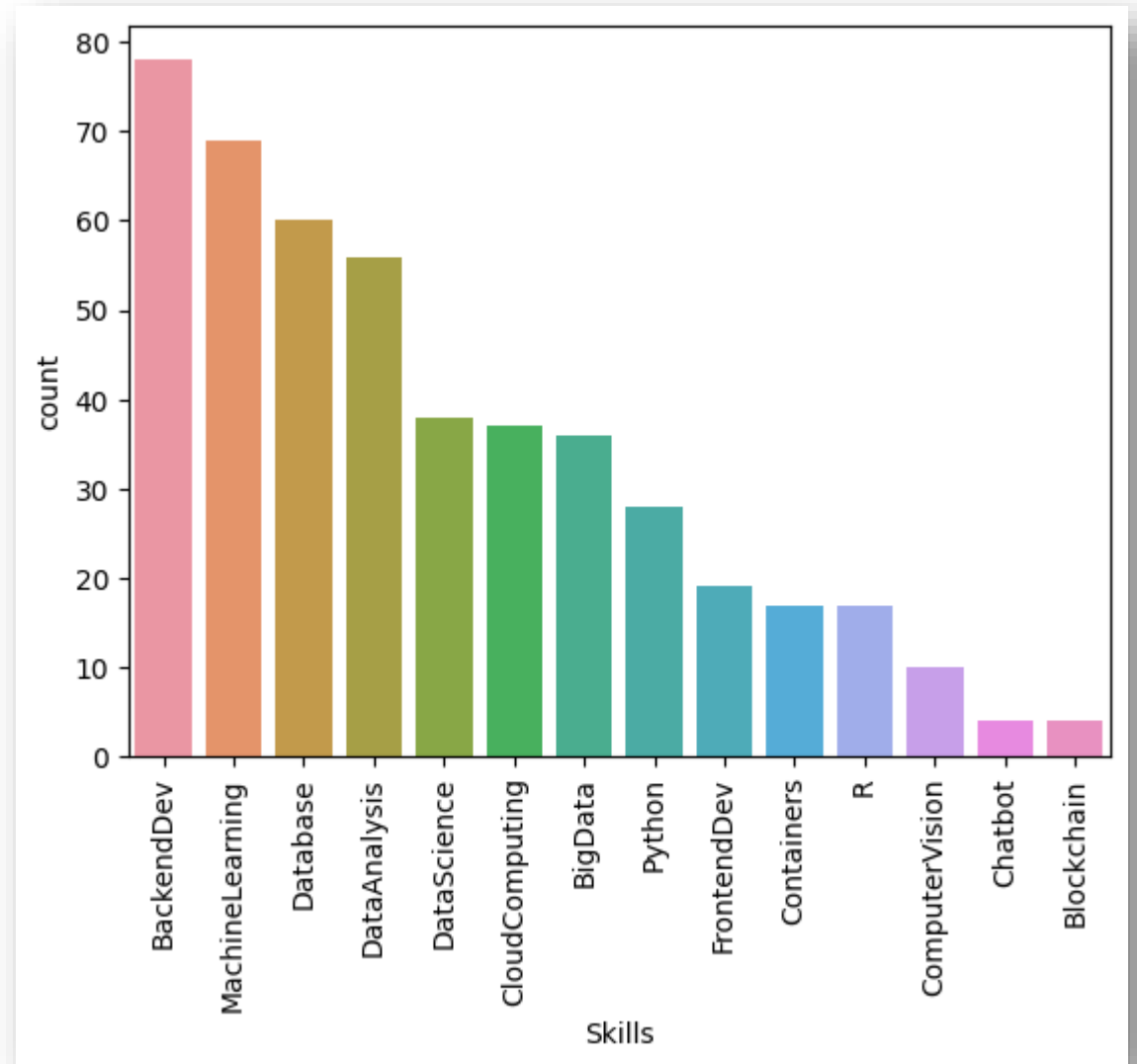
Exploratory Data Analysis



Course counts per genre

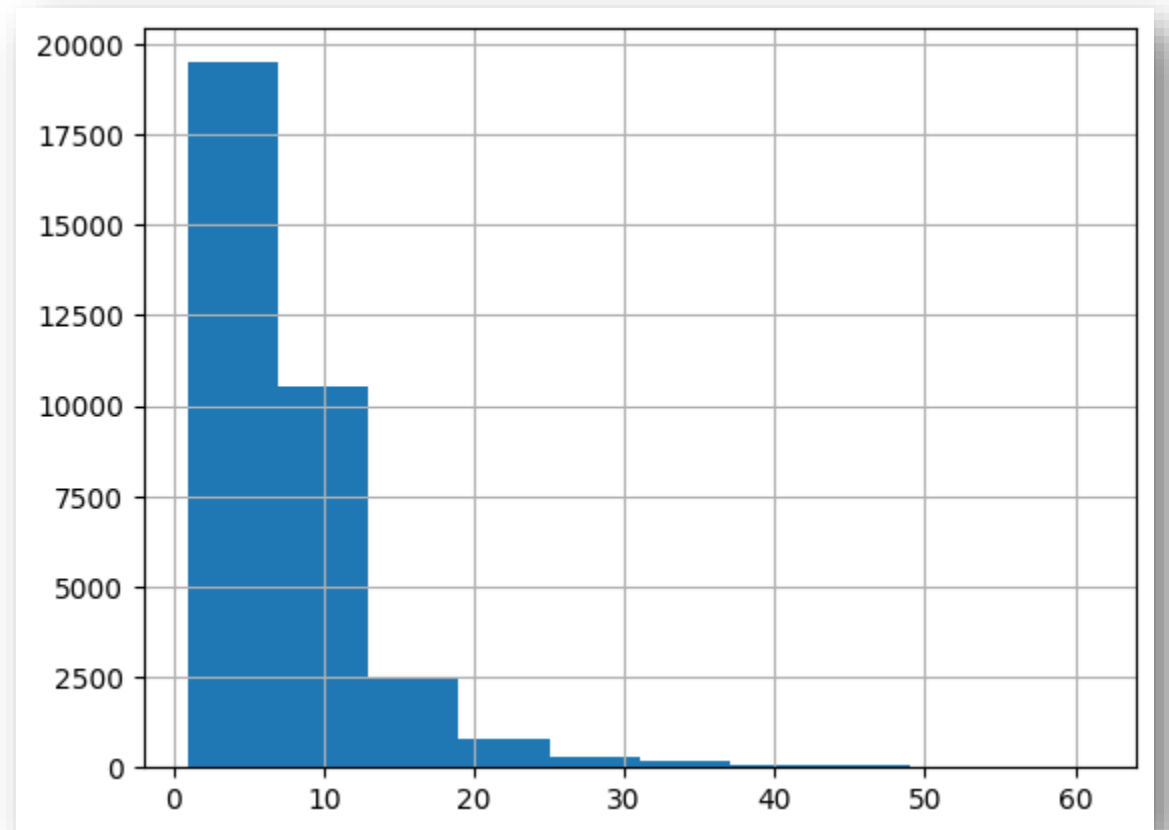
From the barchart we can see that:

- The most popular course genres are those related to Backend development, Machine Learning, Database and Data Analysis.
- On the other side, the less popular course genres are those related to Computer Vision, Chatbot and Blockchain.



Course enrollment distribution

From the histogram we clearly see that the biggest part of users rated only a few courses as it is clearly skewed to the left.



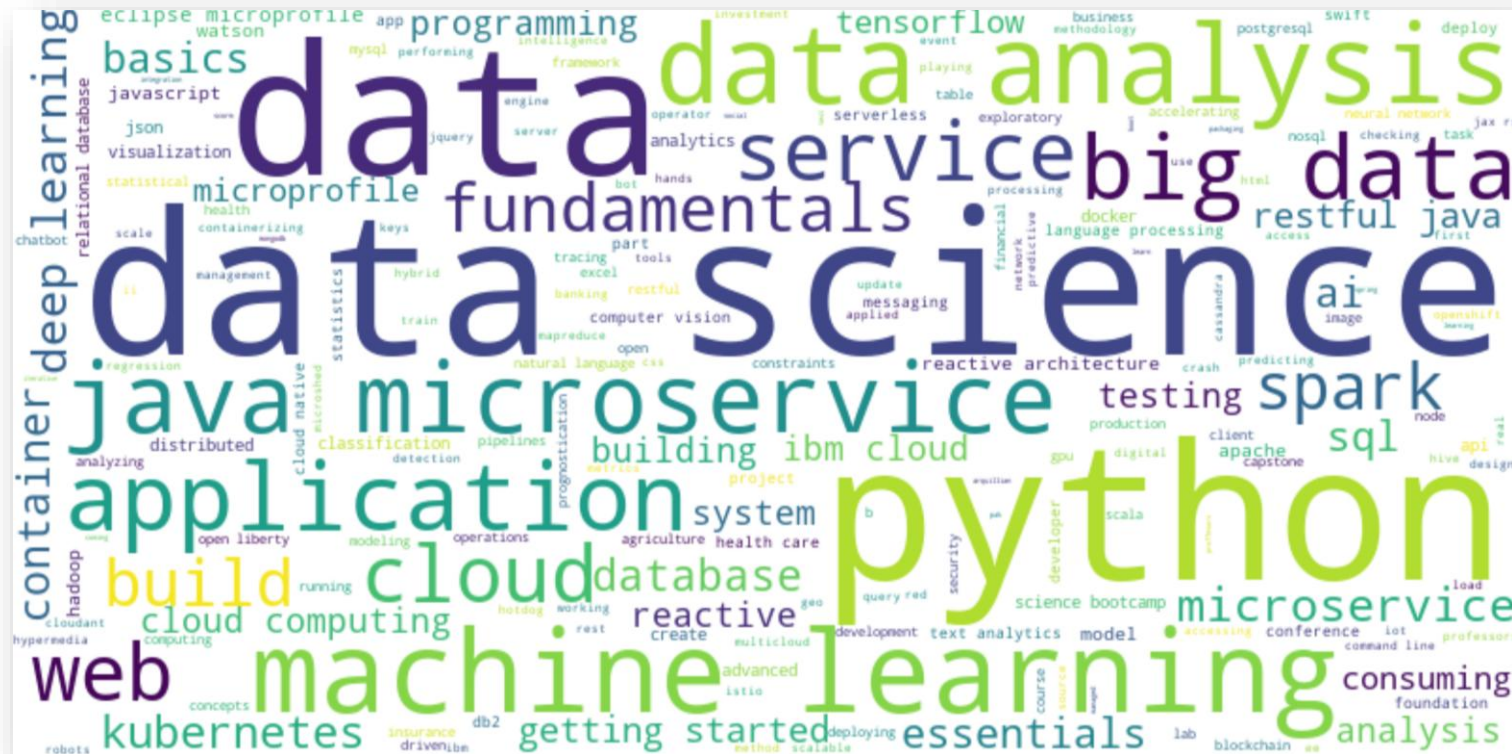
20 most popular courses

Here are the 20 courses with the highest enrollment rates among users.

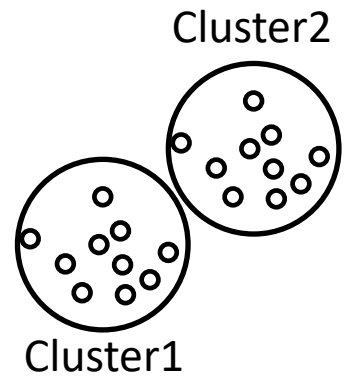
	COURSE_ID	count	TITLE
0	DS0301EN	3624	data privacy fundamentals
1	BD0115EN	3670	mapreduce and yarn
2	DB0101EN	3697	sql and relational databases 101
3	CO0101EN	4480	docker essentials a developer introduction
4	CC0101EN	4983	introduction to cloud
5	ST0101EN	5015	statistics 101
6	RP0101EN	5237	r for data science
7	CB0103EN	5512	build your own chatbot
8	ML0115EN	6323	deep learning 101
9	DV0101EN	6709	data visualization with python
10	BC0101EN	6719	blockchain essentials
11	DS0105EN	7199	data science hands on with open source tools
12	BD0211EN	7551	spark fundamentals i
13	ML0101ENv3	7644	machine learning with python
14	DS0103EN	7719	data science methodology
15	DA0101EN	8303	data analysis with python
16	BD0111EN	10599	hadoop 101
17	BD0101EN	13291	big data 101
18	DS0101EN	14477	introduction to data science
19	PY0101EN	14936	python for data science

Word cloud of course titles

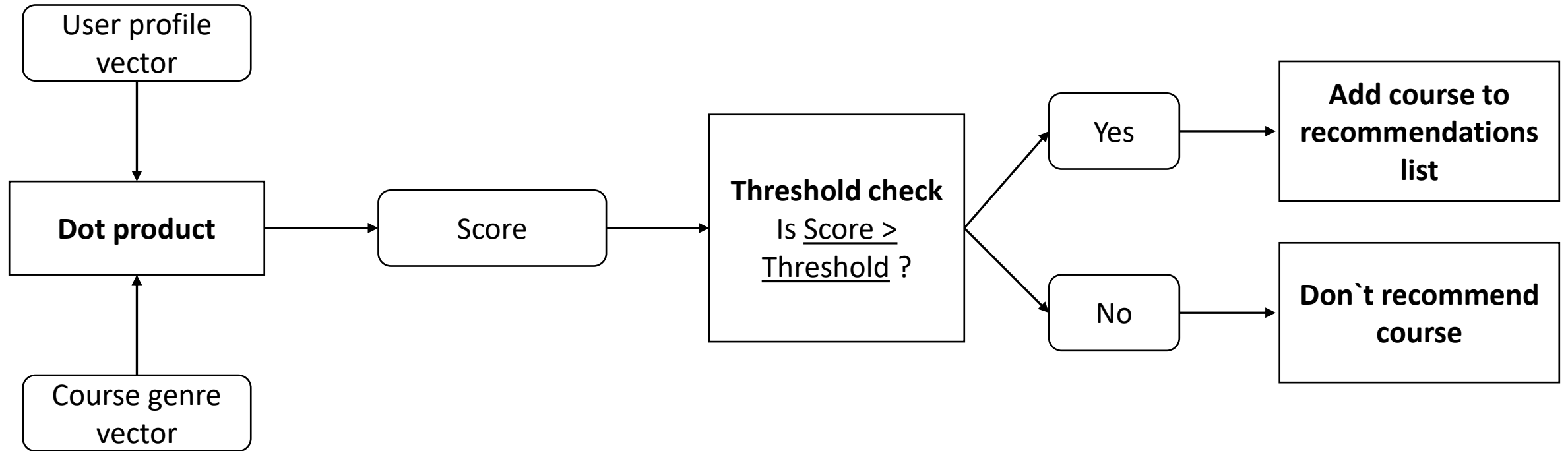
This word cloud visually represents the most frequently occurring words or topics related to courses or content. Larger words, such as data, science, python, machine learning, and java, indicate higher frequency or importance.



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

The threshold can be fine-tuned to adjust the size of generated recommendations
`score_threshold = 10.0`

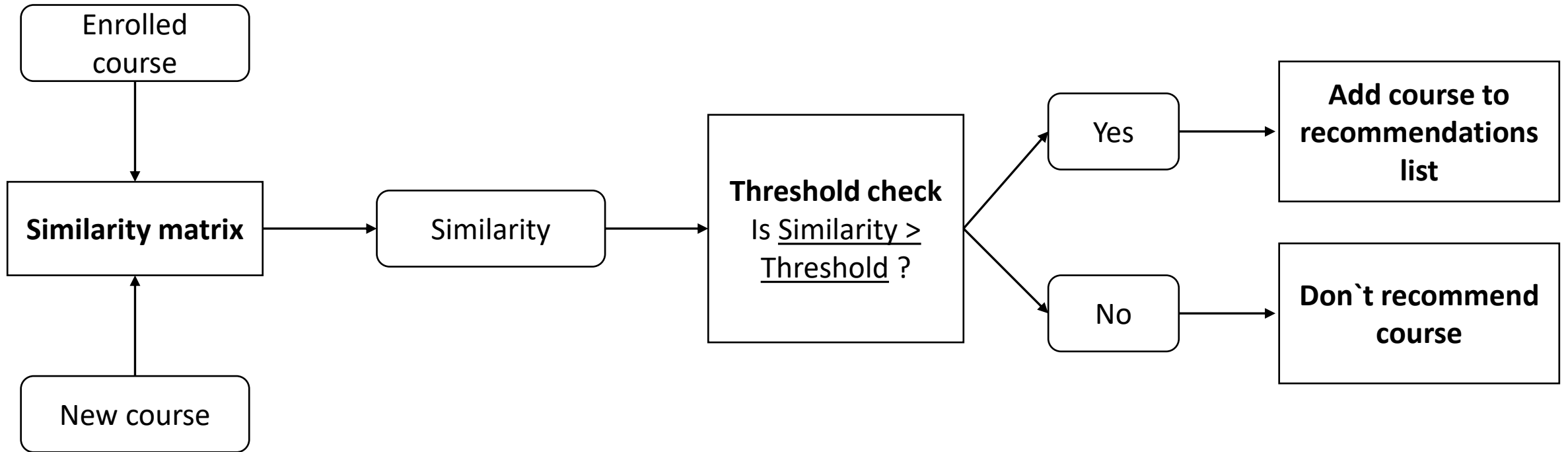
On average, how many new/unseen courses have been recommended per user (in the test user dataset)

What are the most frequently recommended courses? Return the top-10 commonly recommended courses across all users

19

COURSE_ID	
TA0106EN	608
GPXX0IBEN	548
excouse22	547
excouse21	547
ML0122EN	544
excouse06	533
excouse04	533
GPXX0TY1EN	533
excouse31	524
excouse73	516

Flowchart of content-based recommender system using course similarity



Evaluation results of course similarity based recommender system

threshold = 0.6

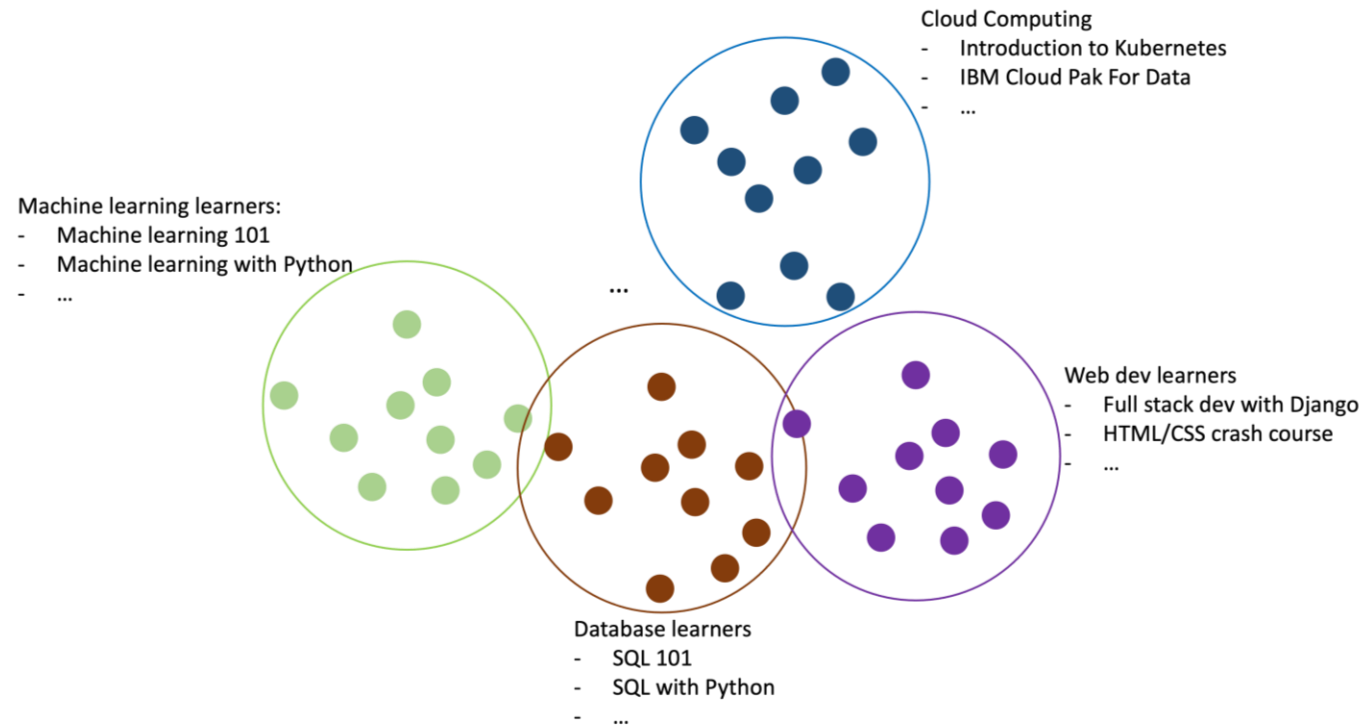
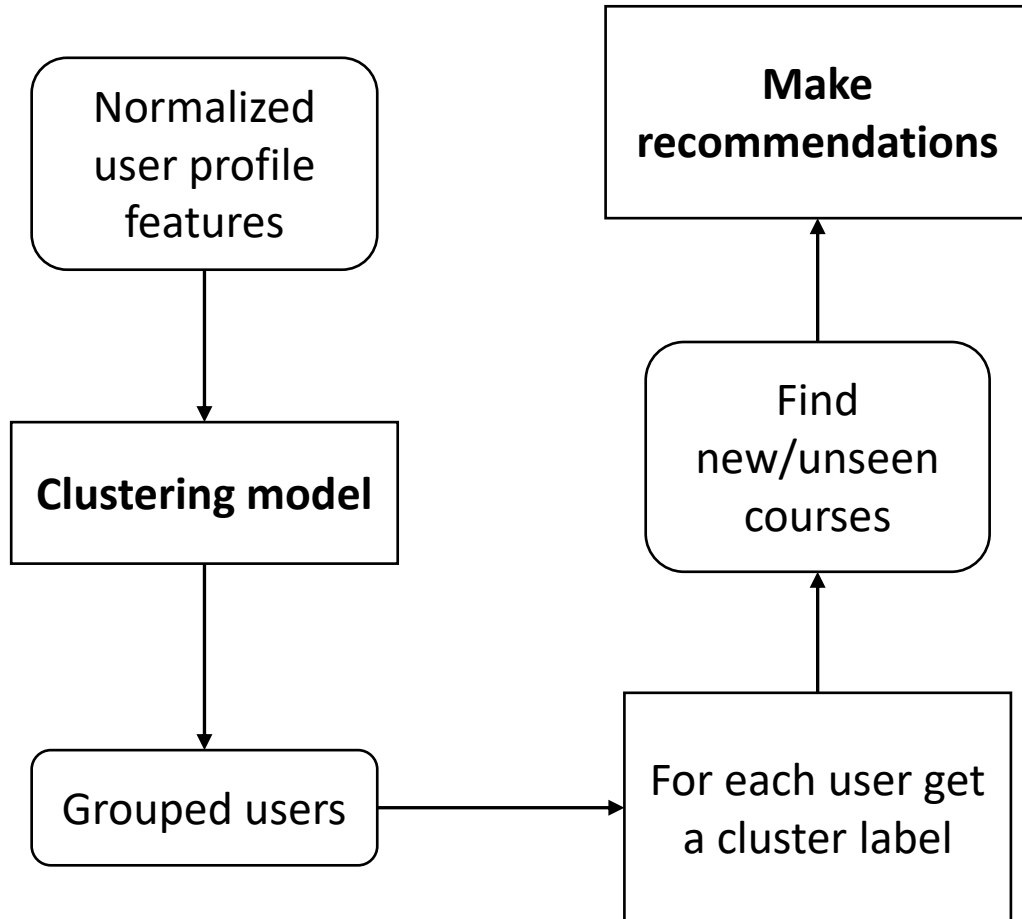
On average, how many new/unseen courses have been recommended per user (in the test user dataset)

10

What are the most frequently recommended courses? Return the top-10 commonly recommended courses

excourse22	579
excourse62	579
DS0110EN	562
excourse65	555
excourse63	555
excourse72	551
excourse68	550
excourse67	539
excourse74	539
BD0145EN	506

Flowchart of clustering-based recommender system



Evaluation results of clustering-based recommender system

Number of clusters = 20

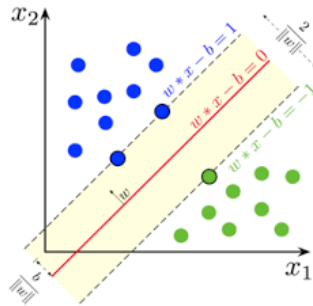
On average, how many new/unseen courses have been recommended per user (in the test user dataset)

What are the most frequently recommended courses? Return the top-10 commonly recommended courses

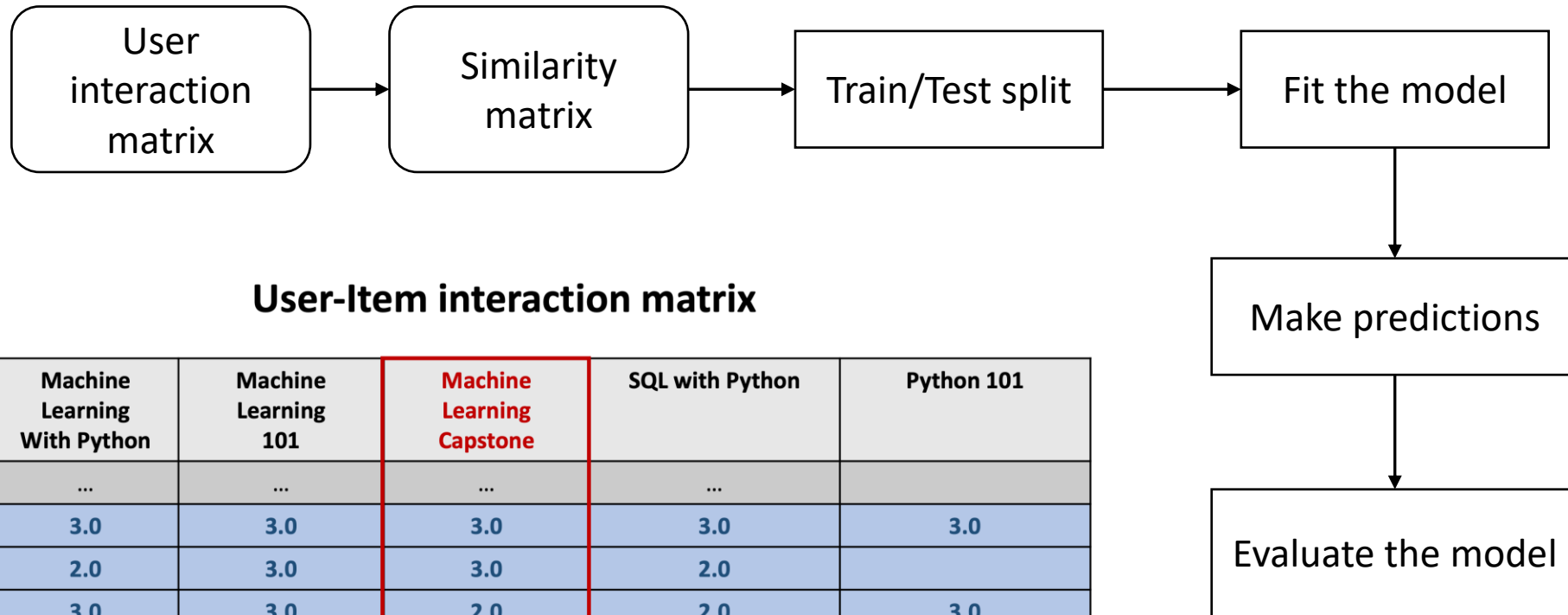
6

DS0103EN	579
DA0101EN	532
BD0111EN	456
DS0101EN	444
BD0101EN	428
PY0101EN	386
DS0105EN	319
ML0101ENv3	299
BC0101EN	296
ML0115EN	286

Collaborative-filtering Recommender System using Supervised Learning

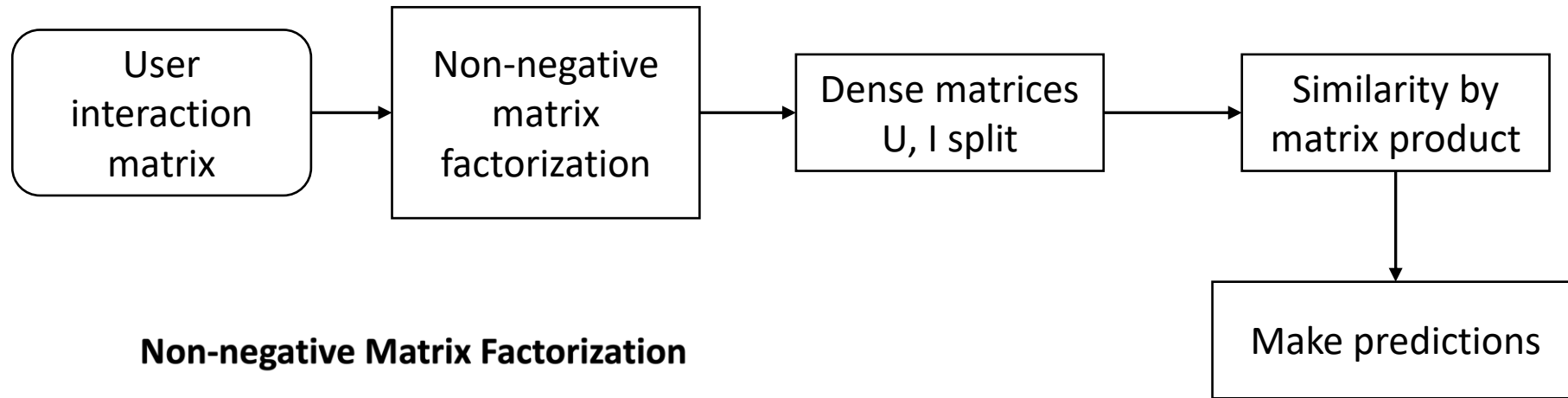


Flowchart of KNN based recommender system



Predict the rating of user *user6* to item *Machine Learning Capstone*

Flowchart of NMF based recommender system



Non-negative Matrix Factorization

User-item interaction matrix: **A** 10000 x 100

	item1	...	item100
user1	
user2	3.0	3.0	3.0
user3	2.0	2.0	-
user4	3.0	2.0	3.0
user5	2.0	-	-
user6	3.0	-	3.0
...	

User matrix: **U** 10000 x 16

	feature1	...	feature16
user1
user2
user3
user4
...
...
user6

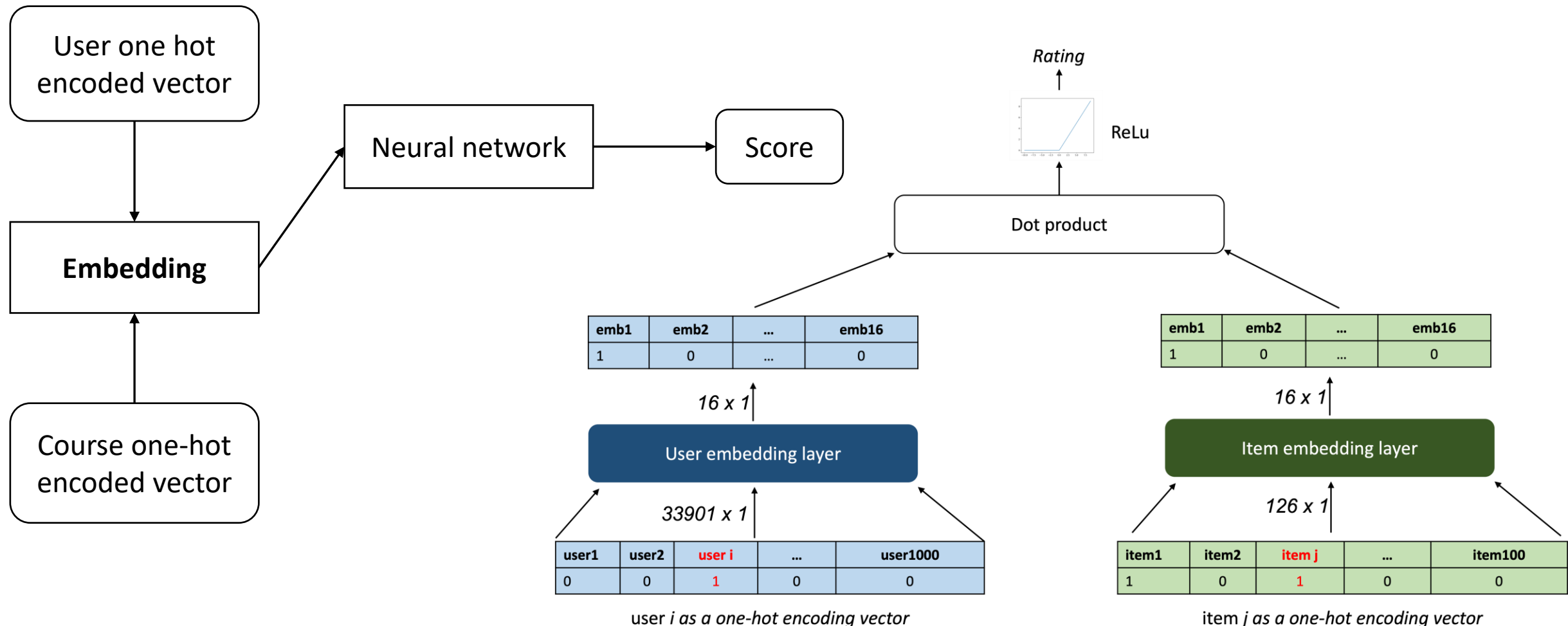
≈

Item matrix: **I** 16 x 100

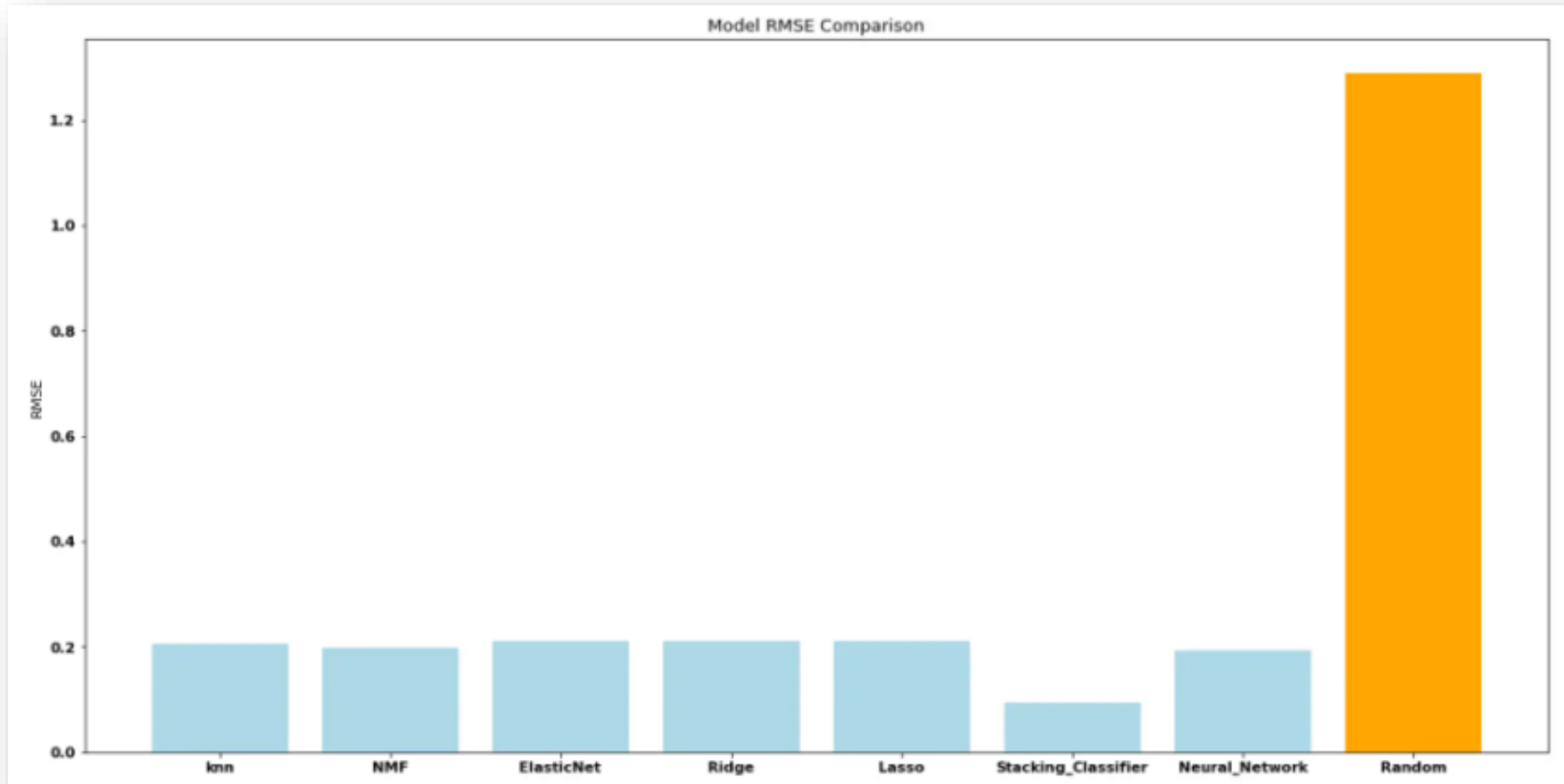
	item1	...	item100
feature1
feature2
...
feature16

X

Flowchart of Neural Network Embedding based recommender system



Compare the performance of collaborative-filtering models



Optional: Build a course recommender system app with Streamlit

Personalized Learning Recommender

1. Select recommendation models

Select model:

Course Similarity

2. Tune Hyper-parameters:

Top courses

1 100

Course Similarity Threshold %

0 50 100

3. Training:

Train Model

4. Prediction

Recommend New Courses

Select courses that you have completed:

COURSE_ID	TITLE	DESCRIPTION
<input checked="" type="checkbox"/> ML0201EN	Robots Are Coming Build Iot Apps With Watson Swift And Node Red	have fun with
<input type="checkbox"/> ML0122EN	Accelerating Deep Learning With Gpu	training com
<input checked="" type="checkbox"/> GPXX0ZG0EN	Consuming Restful Services Using The Reactive Jax Rs Client	learn how to
<input type="checkbox"/> RP0105EN	Analyzing Big Data In R Using Apache Spark	apache spar
<input type="checkbox"/> GPXX0Z2PEN	Containerizing Packaging And Running A Spring Boot Application	learn how to
<input type="checkbox"/> CNSC02EN	Cloud Native Security Conference Data Security	introduction
<input checked="" type="checkbox"/> DX0106EN	Data Science Bootcamp With R For University Professors	a multi day i
<input type="checkbox"/> GPXX0FTCEN	Learn How To Use Docker Containers For Iterative Development	learn how to
<input type="checkbox"/> RAVSCTEST1	Scorm Test 1	scron test co
<input type="checkbox"/> GPXX06RFEN	Create Your First MongoDB Database	in this guide
<input type="checkbox"/> GPXX0SDXEN	Testing Microservices With The Arquillian Managed Container	learn how to
<input type="checkbox"/> CC0271EN	Cloud Pak For Integration Essentials	in this short

Your courses:

	COURSE_ID	TITLE
0	ML0201EN	Robots Are Coming Build Iot Apps With Watson Swift And Node Red
1	GPXX0ZG0EN	Consuming Restful Services Using The Reactive Jax Rs Client
2	DX0106EN	Data Science Bootcamp With R For University Professors

Recommendations generated!

	USER	COURSE_ID	TITLE
0	2103073	TMP0106	Data Science Bootcamp

Personalized Learning Recommender

1. Select recommendation models

Select model:

KNN

2. Tune Hyper-parameters:

Top courses

1 100

Number of Neighbors

1 20 50

3. Training:

Train Model

4. Prediction

Recommend New Courses

GPXX0HZ2EN Deploying Microservices To Kubernetes

GPXX04TNEN Getting Started With Open Liberty

Your courses:

	COURSE_ID	TITLE
0	ML0201EN	Robots Are Coming Build Iot Apps With Watson Swift And Node Red
1	GPXX0ZG0EN	Consuming Restful Services Using The Reactive Jax Rs Client
2	DX0106EN	Data Science Bootcamp With R For University Professors

Recommendations generated!

	USER	COURSE_ID	TITLE
0	2103075	ML0122EN	Accelerating Deep Learning With Gpu
1	2103075	RP0105EN	Analyzing Big Data In R Using Apache Spark
2	2103075	GPXX0Z2PEN	Containerizing Packaging And Running A Spring Boot App
3	2103075	CNSC02EN	Cloud Native Security Conference Data Security
4	2103075	GPXX0FTCEN	Learn How To Use Docker Containers For Iterative Develo
5	2103075	RAVSCTEST1	Scorm Test 1
6	2103075	GPXX06RFEN	Create Your First MongoDB Database
7	2103075	GPXX0SDXEN	Testing Microservices With The Arquillian Managed Conta
8	2103075	CC0271EN	Cloud Pak For Integration Essentials
9	2103075	WA0103EN	Watson Analytics For Social Media

Conclusions

Key Insights:

- The recommender system successfully identified courses aligned with user preferences using both content-based and collaborative filtering approaches.
- The content-based approach proved effective in recommending unseen courses through user profiles and course similarities.
- Collaborative filtering models, including KNN and neural network embeddings, demonstrated strong performance in predicting user preferences.

Strengths of the Approach:

- The combination of multiple methodologies (unsupervised and supervised learning) improved recommendation accuracy.
- Flexibility in threshold adjustments allowed for tailored recommendations

Impact:

- This system has the potential to improve user engagement and satisfaction by delivering personalized course suggestions tailored to individual interests and learning paths.

Appendix

- GitHub: <https://github.com/RSennaM>
- LinkedIn: <https://www.linkedin.com/in/sennarodrigo/>