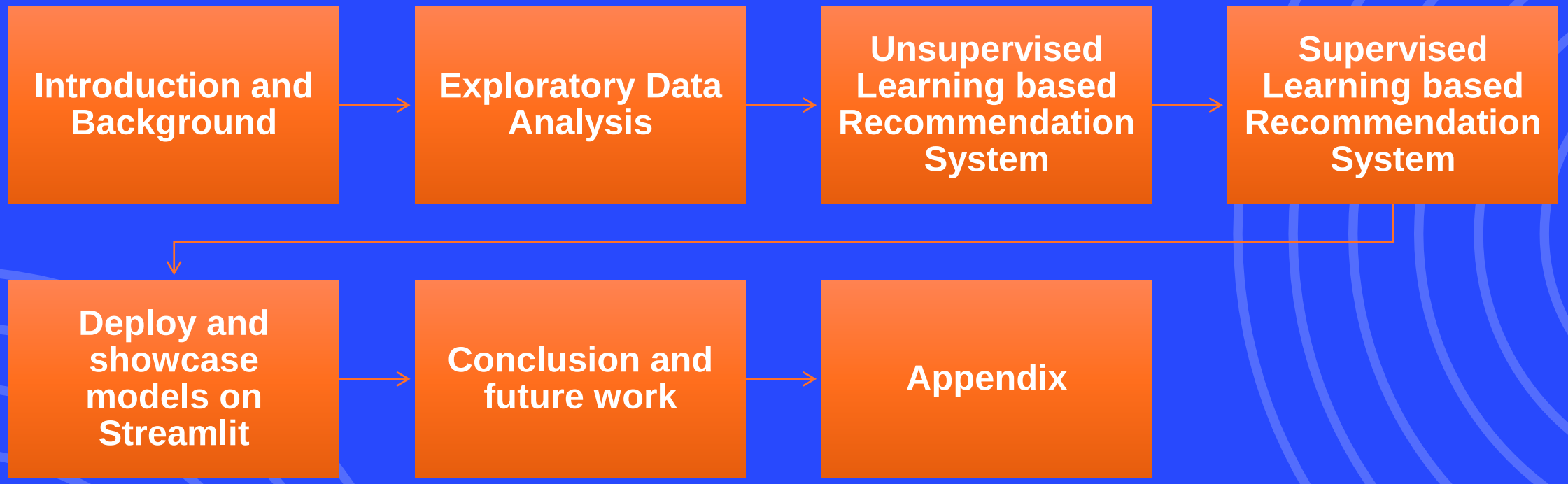
Final Report

Build A Personalized Online Course Recommender System with Machine Learning

Pham Quoc Nam
9/17/2024

Outline

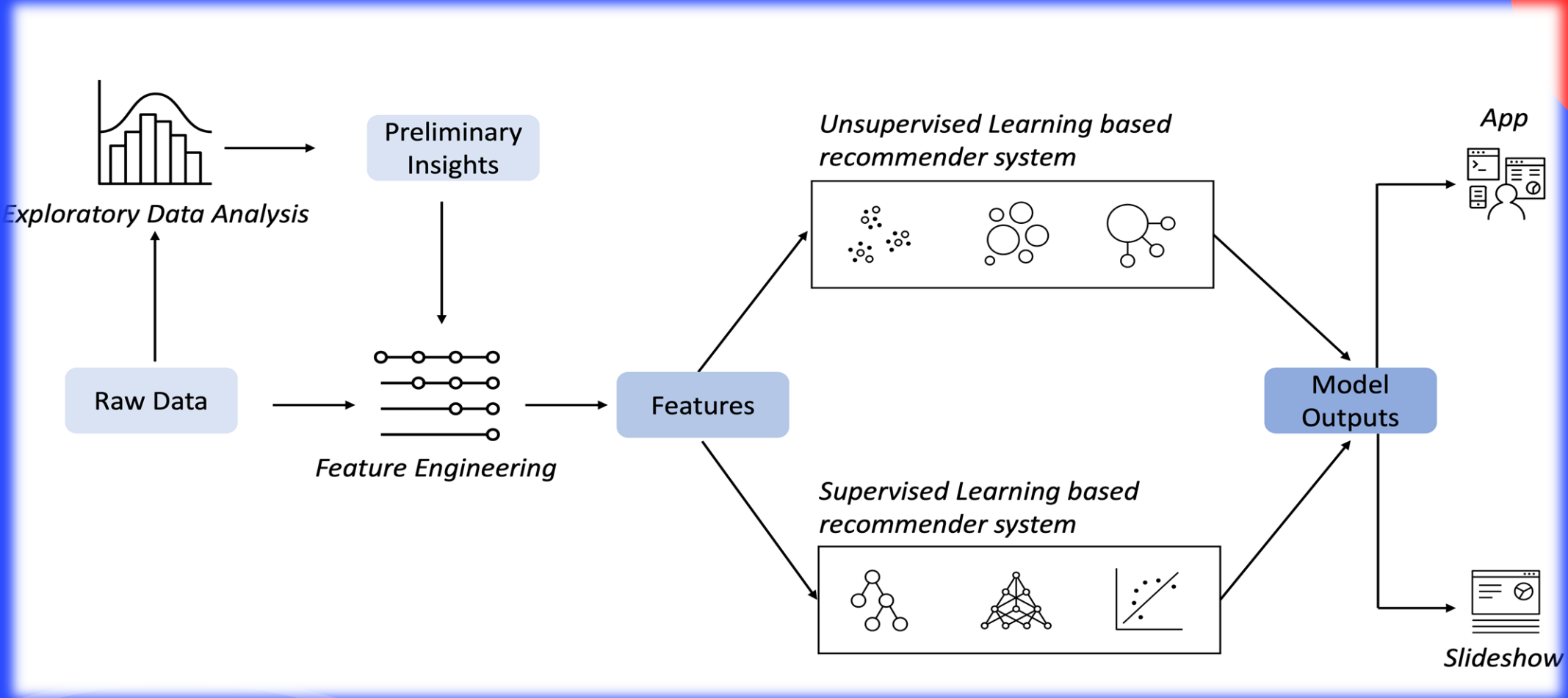

Introduction and Background
Exploratory Data Analysis
Unsupervised Learning based Recommendation System
Supervised Learning based Recommendation System
Deploy and showcase models on Streamlit
Conclusion and future work
Appendix

Introduction & Background

Introduction

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.

Mission


Exploratory Data Analysis
Preliminary Insights
Raw Data
Feature Engineering
Features
Unsupervised Learning based recommender system
Supervised Learning based recommender system
Model Outputs
App
Slideshow

Mission

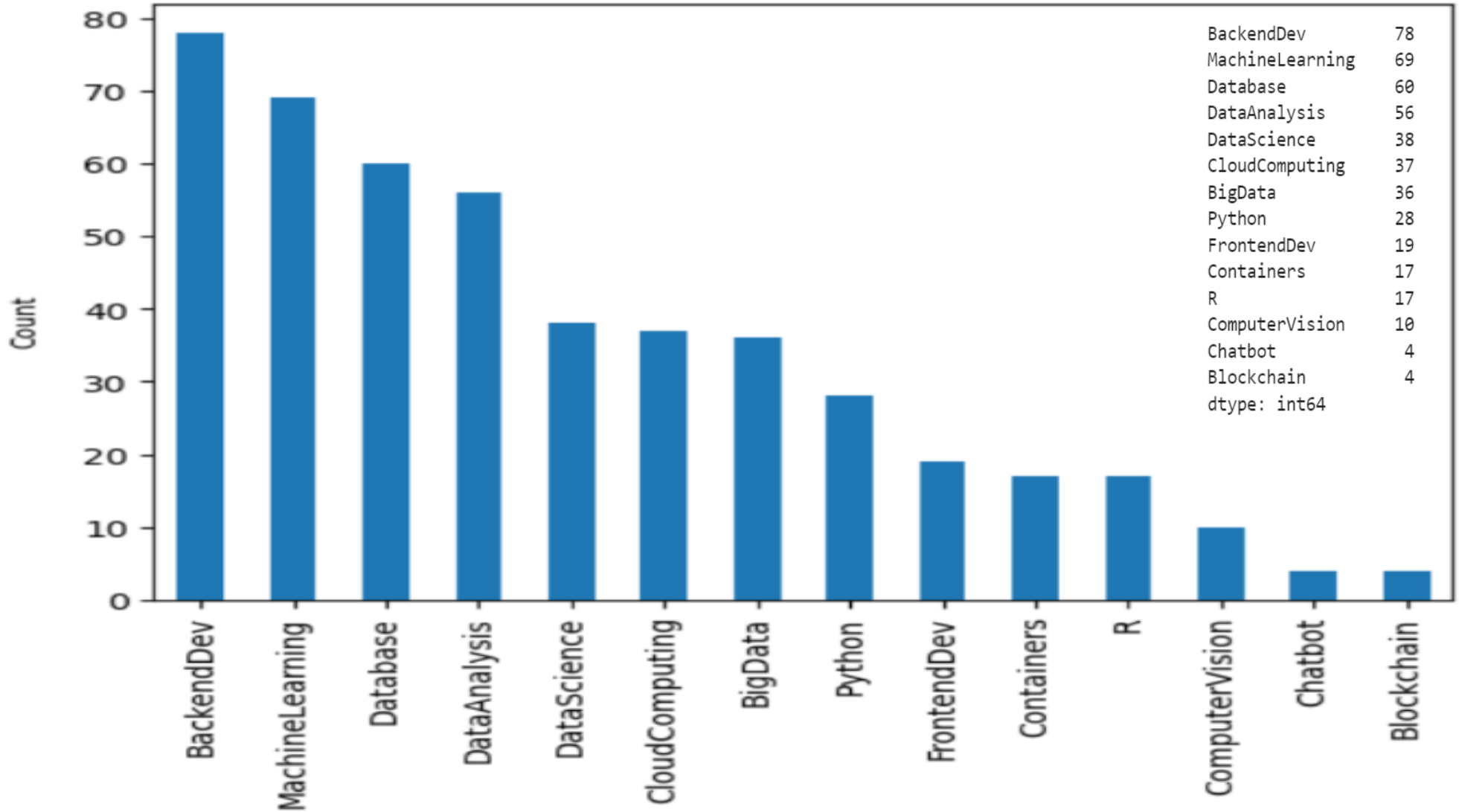
- Collecting and understanding data
- Performing exploratory data analysis on online course enrollments datasets
- Extracting Bag of Words (BoW) features from course textual content
- Calculating course similarity using BoW features
-Building content-based recommender systems using various unsupervised learning algorithms
- Building collaborative-filtering recommender systems using vari supervised learning algorithms
- Creating an insightful and informative slideshow and presenting to your peers

Exploratory Data Analysis

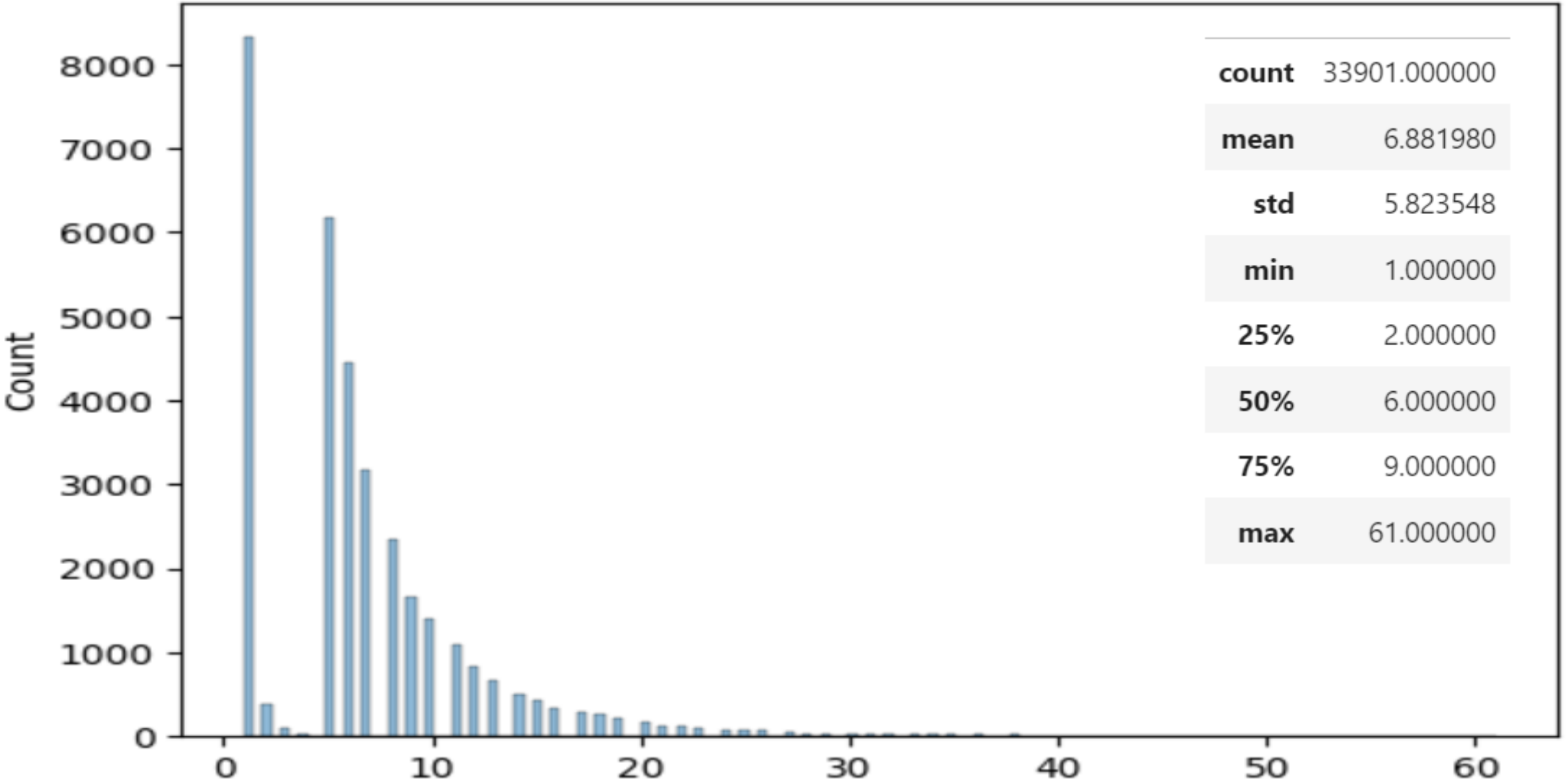
Target

- Identify keywords in course titles using a WordCloud
- Calculate the summary statistics and visualizations of the online course content dataset
- Determine popular course genres
- Calculate the summary statistics and create visualizations of the online course enrollment dataset
- Identify courses with the greatest number of enrolled students

Analyze course genres



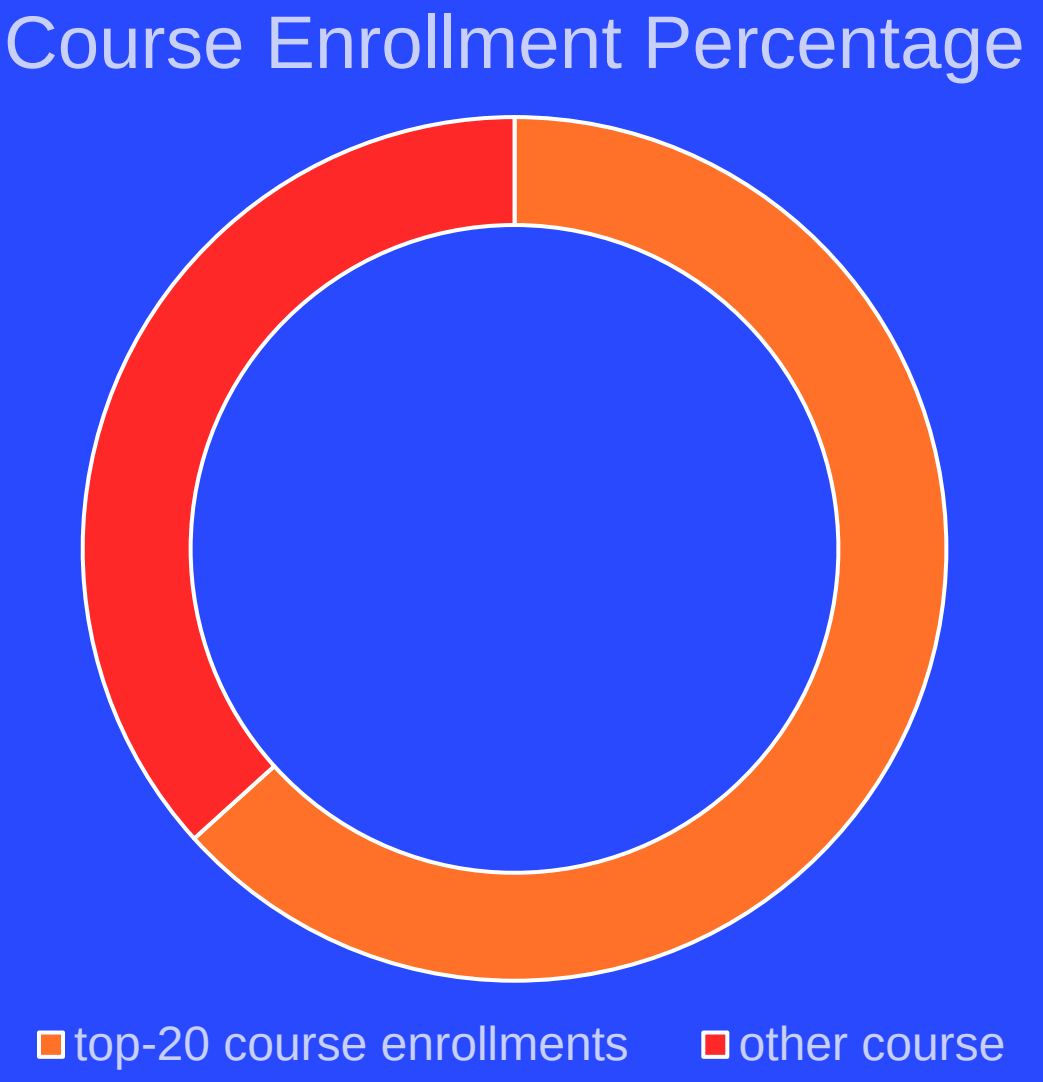
Analyze Course Enrollments



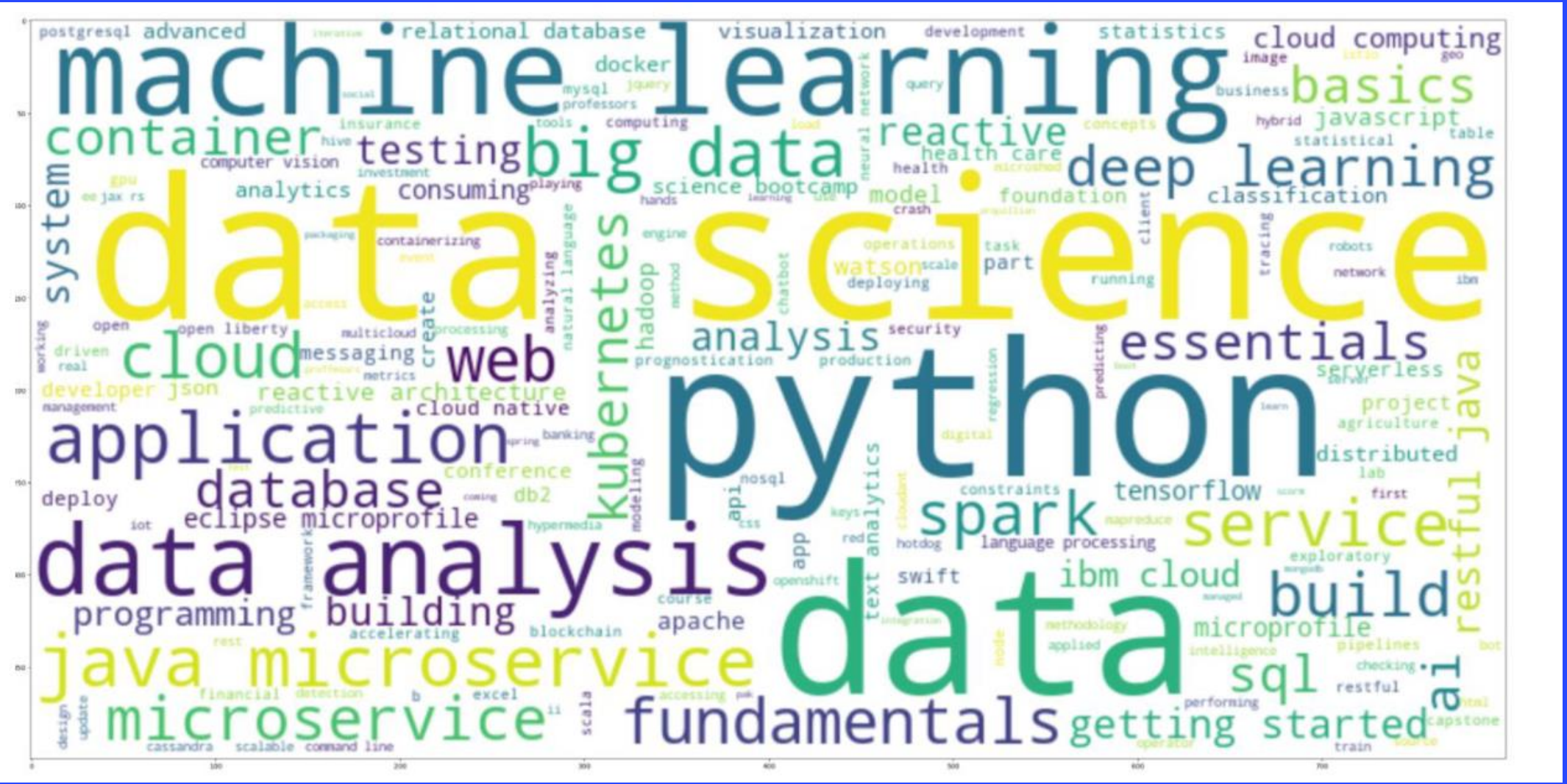
count	33901.000000
mean	6.881980
std	5.823548
min	1.000000
25%	2.000000
50%	6.000000
75%	9.000000
max	61.000000

Top-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

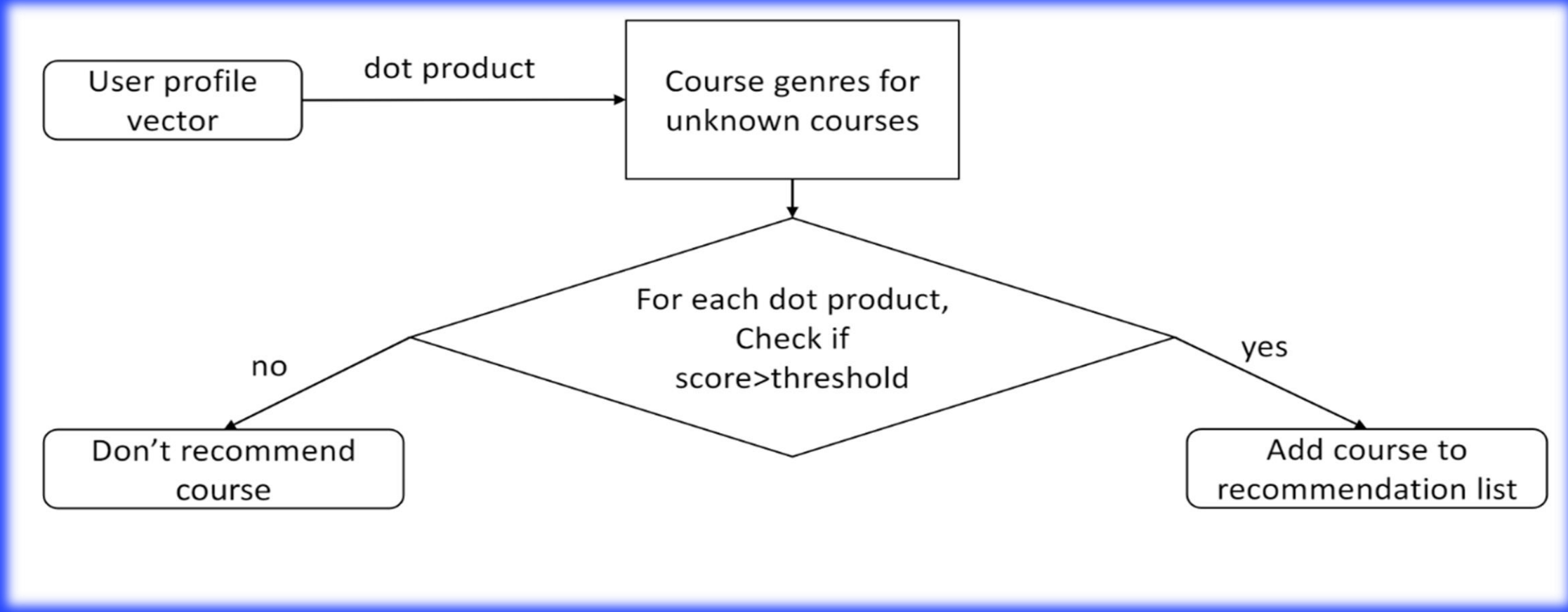


Unsupervised Learning based Recommendation System

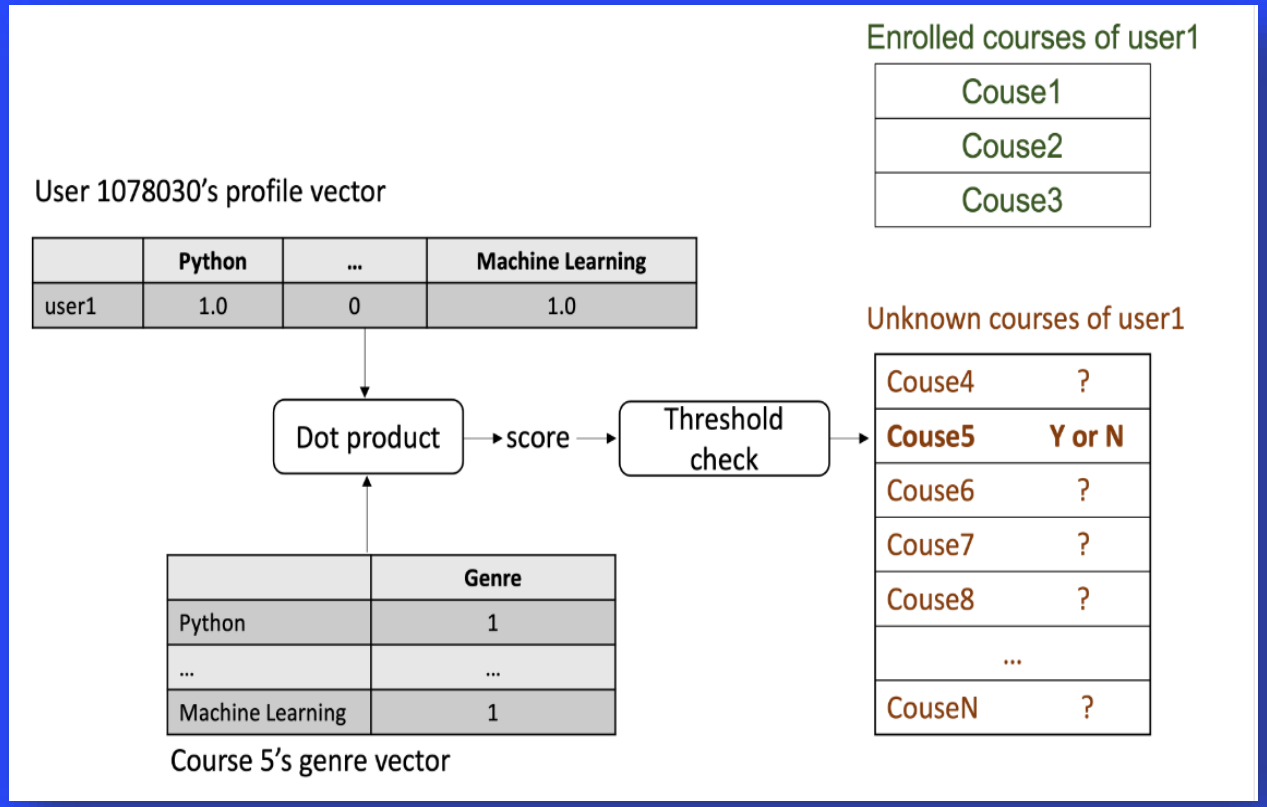
Outline

1. Content-based Course Recommender System using User Profile and Course
Genres
2. Content-based Course Recommender System using Course Similarities
3. Clustering based Course Recommender System

Content-based Course Recommender System using User Profile and Course

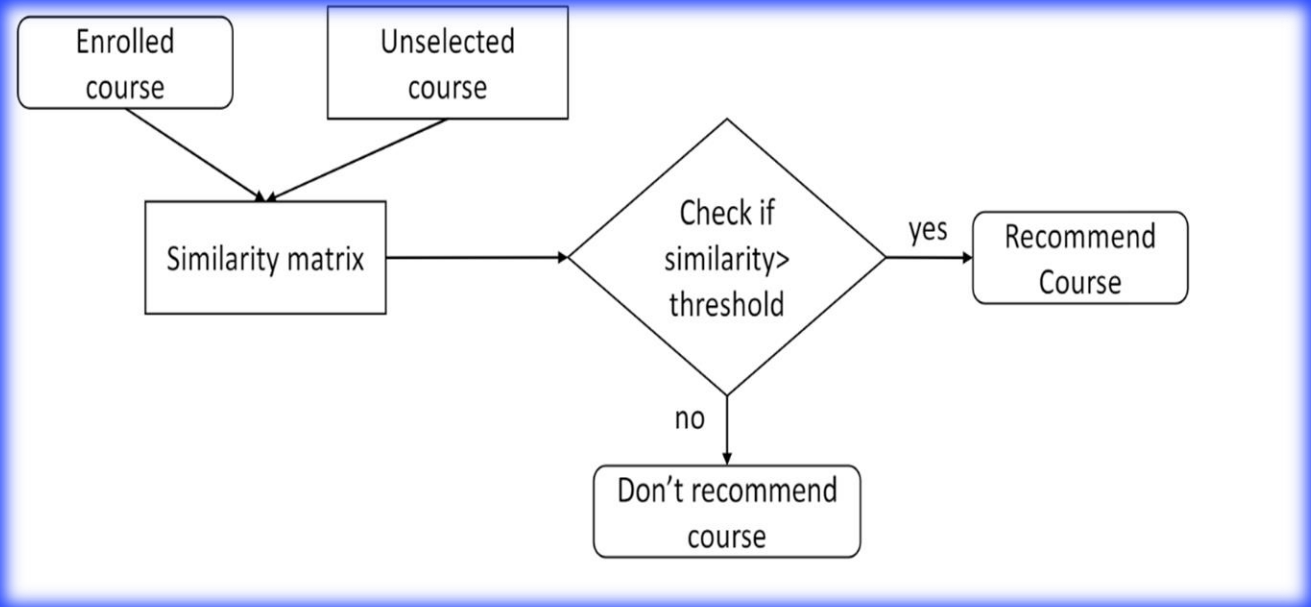


Results



	USER	COURSE_ID	SCORE
0	2	ML0201EN	43.0
1	2	GPXX0ZG0EN	43.0
2	2	GPXX0Z2PEN	37.0
3	2	DX0106EN	47.0
4	2	GPXX06RFEN	52.0
...	...	...	...
1500419	2102680	excourse62	15.0
1500420	2102680	excourse69	14.0
1500421	2102680	excourse77	14.0
1500422	2102680	excourse78	14.0
1500423	2102680	excourse79	14.0

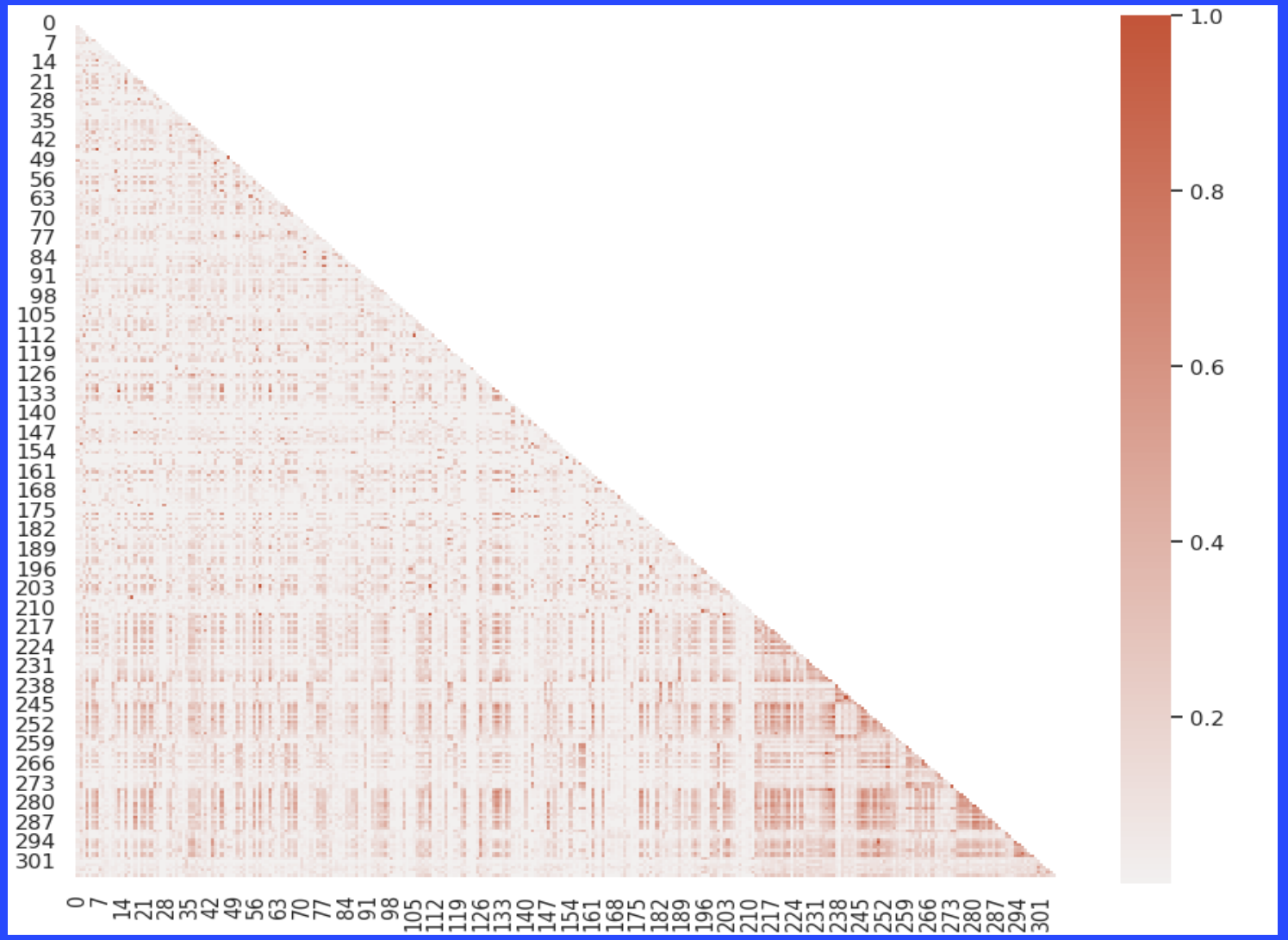
Content-based Course Recommender System using Course Similarities



1. the content-based recommender system is highly based on the similarity calculation among items
2. The similarity or closeness of items is measured based on the similarity in the content or features of those items.

Course similarity matrix:

	0	1	2	3	4	5	6
0	1.000000	0.088889	0.088475	0.065556	0.048810	0.104685	0.065202
1	0.088889	1.000000	0.055202	0.057264	0.012182	0.078379	0.032545
2	0.088475	0.055202	1.000000	0.026463	0.039406	0.000000	0.000000
3	0.065556	0.057264	0.026463	1.000000	0.000000	0.250490	0.390038
4	0.048810	0.012182	0.039406	0.000000	1.000000	0.000000	0.000000
...	...	...	...	...	...	...	...
302	0.033944	0.028239	0.018270	0.094759	0.060474	0.064851	0.053856
303	0.076825	0.063911	0.082698	0.030638	0.030415	0.000000	0.000000
304	0.072898	0.138270	0.133400	0.017443	0.129871	0.009285	0.000000
305	0.039276	0.031367	0.012684	0.018796	0.000000	0.015008	0.024926
306	0.121113	0.076940	0.000000	0.158073	0.000000	0.126211	0.157219



Evaluation results of user profile-based recommender system

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

Score_threshold=10

```
s = 0
for i in range(len(res_df['COURSE_ID'])):
    s+=len(res_df['COURSE_ID'].iloc[i])
avg = s/len(res_df['COURSE_ID'])
```

```
avg
```

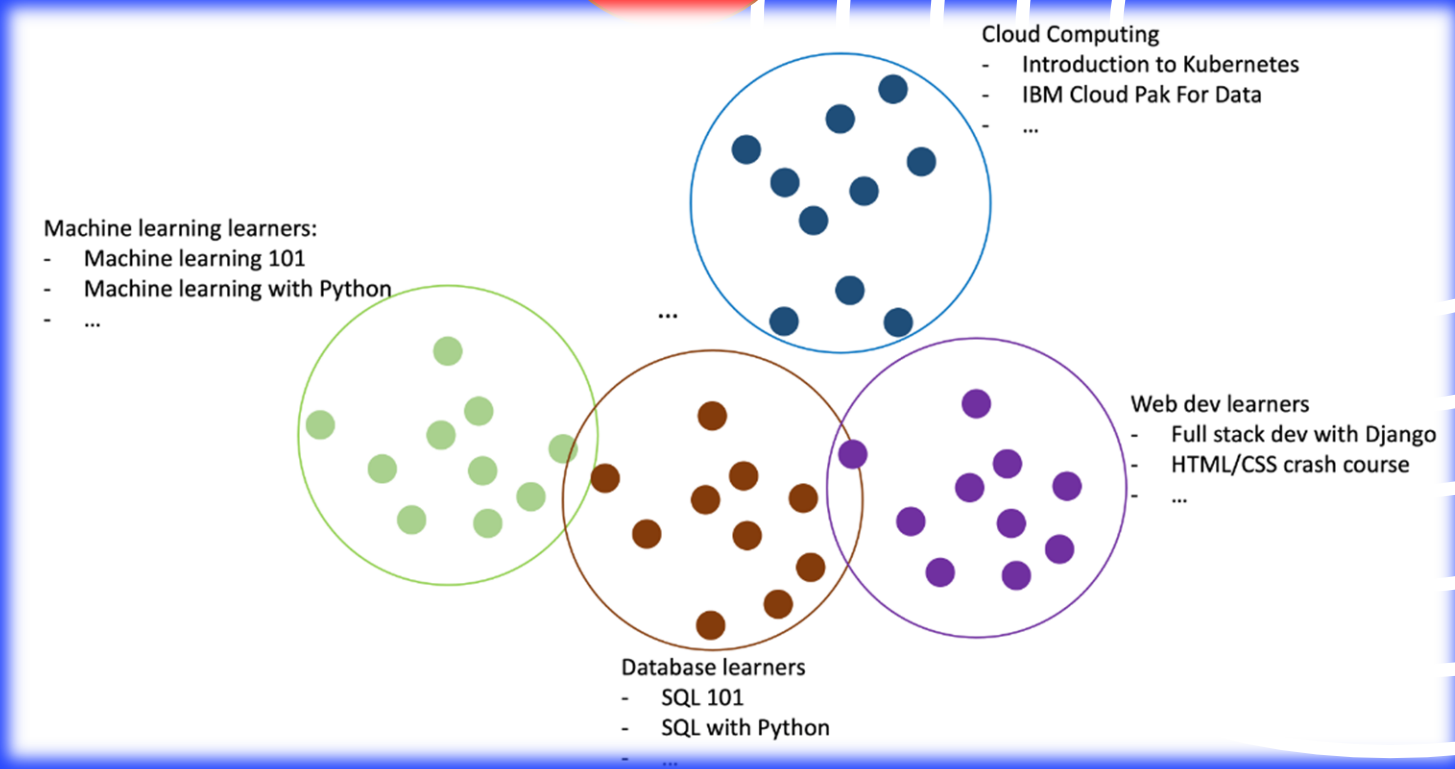
```
11.377
```

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

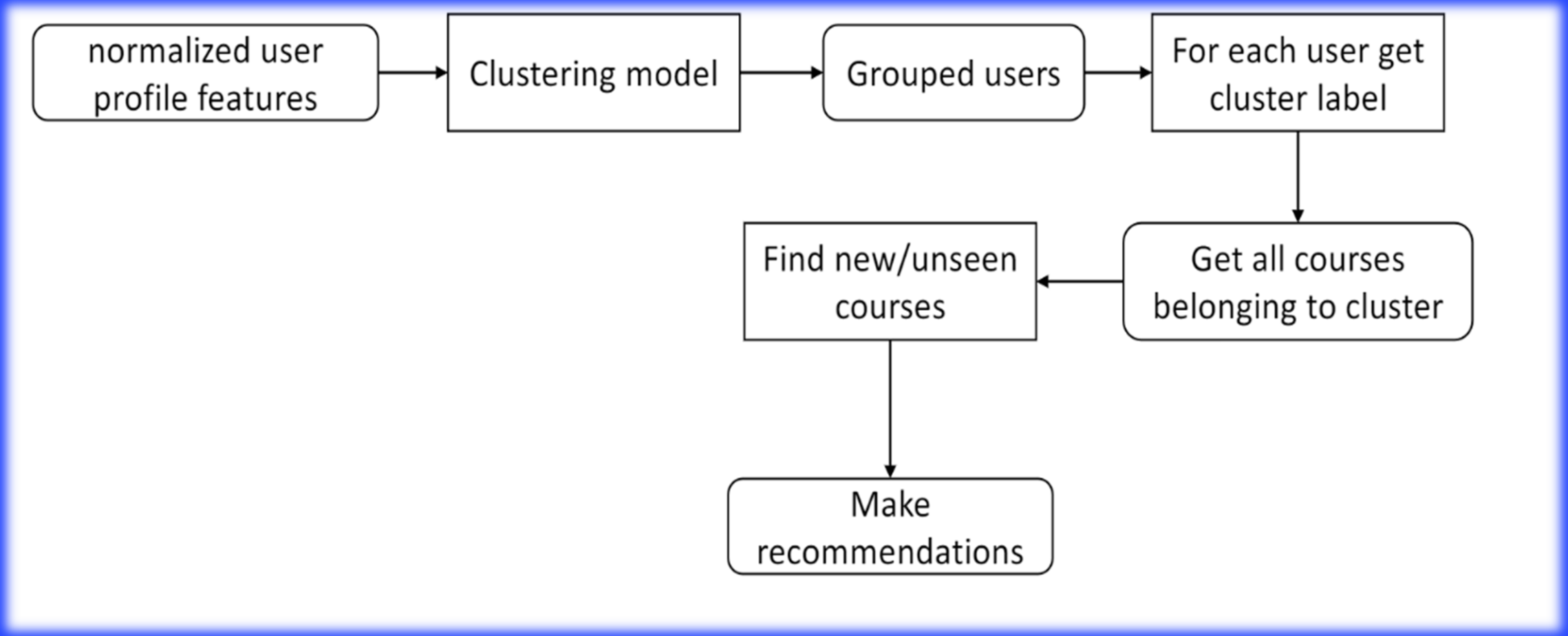
excourse22	579
excourse62	579
DS0110EN	562
excourse65	555
excourse63	555
excourse72	551
excourse68	550
excourse67	539
excourse74	539
BD0145EN	506

Clustering based Course Recommender System

We could perform clustering algorithms such as K-means or DBSCAN to group users with similar learning interests. For example, in the below user clusters, we have user clusters whom have learned courses related to machine learning, cloud
computing, databases



Flow chart


normalized user profile features
Clustering model
Grouped users
For each user get cluster label
Get all courses belonging to cluster
Find new/unseen courses
Make recommendations

Evaluation

Number of clusters = 20

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

```
s = 0
for r in user_recommendations.value
    s+=r[1:].sum()
avg=s/len(user_recommendations)
print(avg)
```

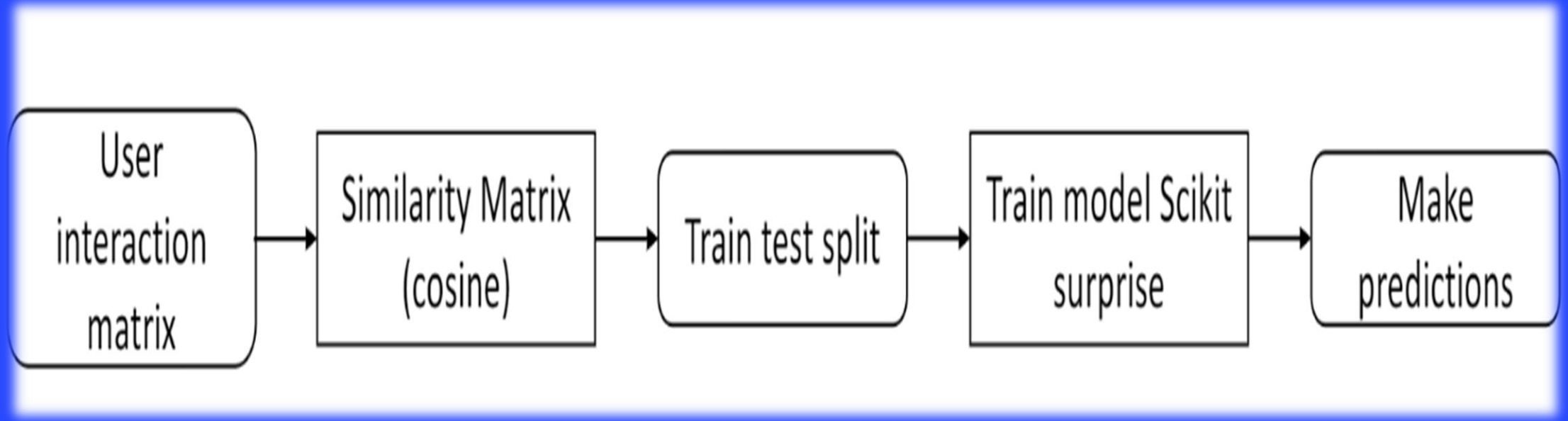
```
5.733
```

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

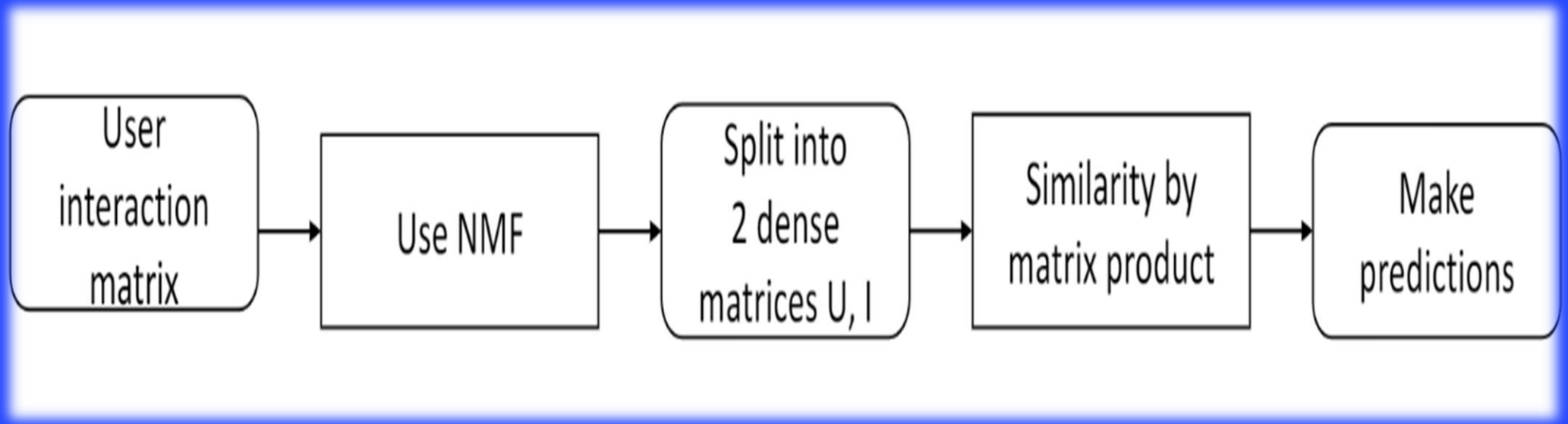
DS0103EN	579
DA0101EN	532
BD0111EN	456
DS0101EN	444
BD0101EN	428
PY0101EN	386
DS0105EN	319
ML0101ENv3	299
BC0101EN	296
ML0115EN	286

Supervised Learning based Recommendation System

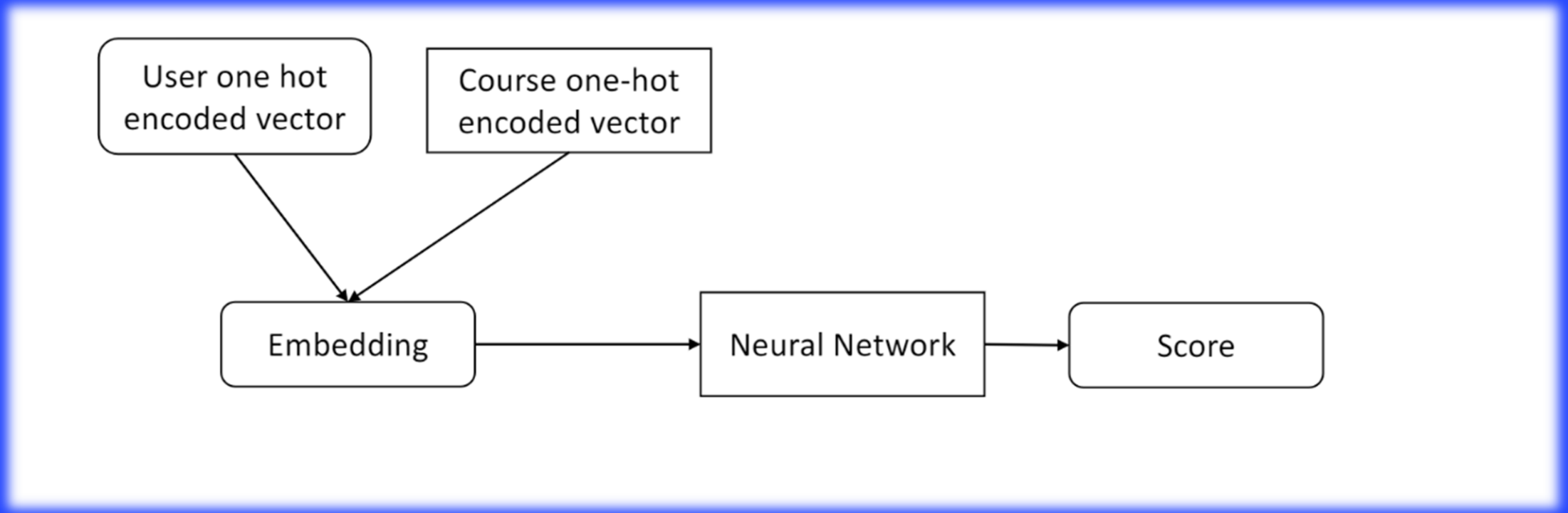
Flow chart of KNN based recommender system



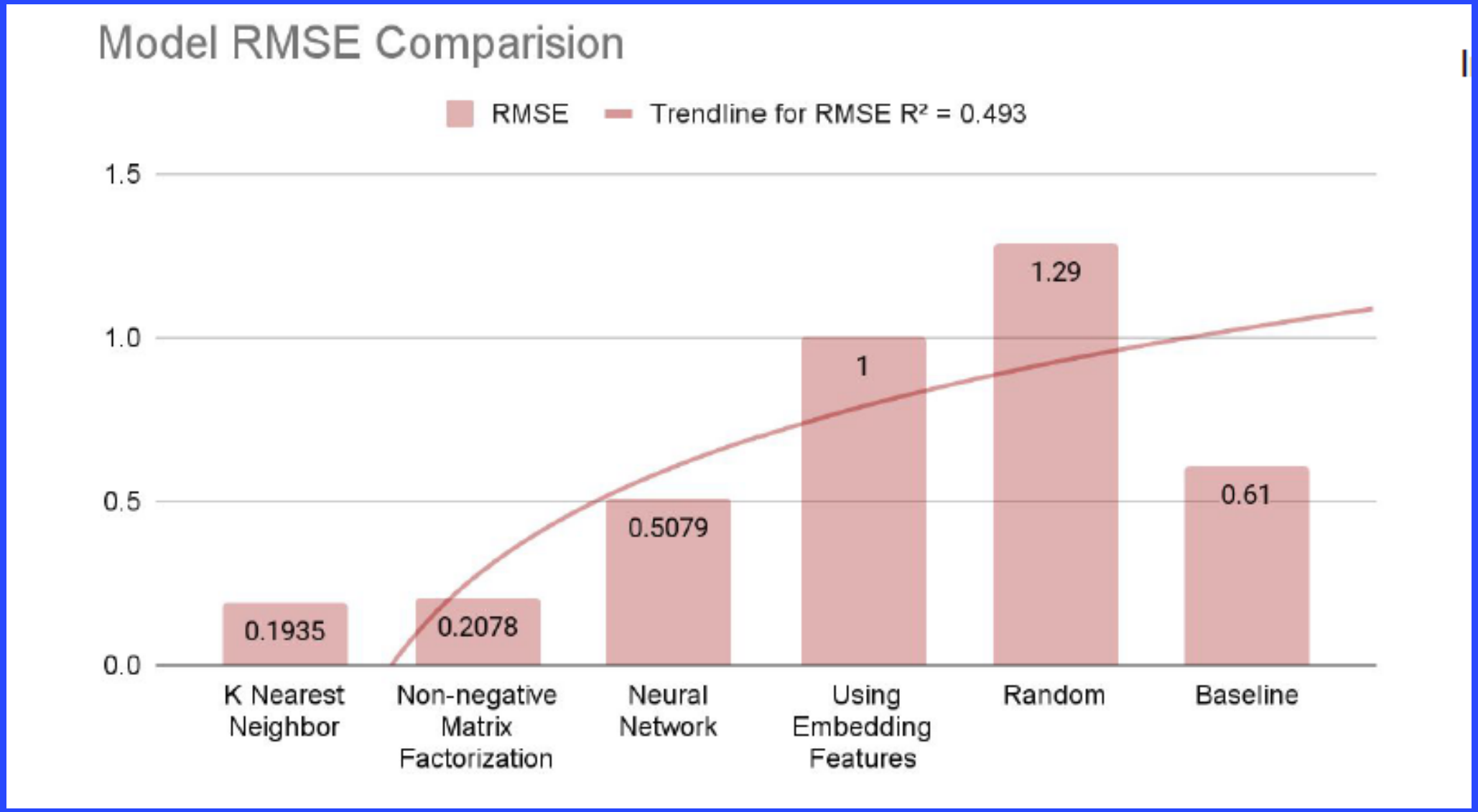
Flow chart of NMF based recommender system



Flow char of Neural Network Embedding based recommender system



Compare the performance of models



Deploy and showcase models on Streamlit

st.sidebar()

Personalized Learning Recommender

st.title()

1. Select recommendation models

Select model:

Course Similarity

st.selectbox()

2. Tune Hyper-parameters:

Top courses

10

0 100

Course Similarity Threshold %

50

0 100

st.slider()

3. Training:

Train Model

st.button()

4. Prediction

Recommend New Courses

Select courses that you have audited or completed:

st_aggrid()

COURSE_ID	TITLE	DESCRIPTION
DW0101EN	Introduction To Machine Learning With Sound	get hands on experience creating a
ML0111EN	Machine Learning With Apache Systemml	apache systemml is a declarative st
GPXX0ZMZEN	Data Science In Health Care Advanced Machine Learning Classification	learn to apply an advanced analysis
ML0101EN	Machine Learning With Python	are the phrases , it is certain , yes y
ML0109EN	Machine Learning Dimensionality Reduction	machine learning dimensionality re
ML0101ENv3	Machine Learning With Python	machine learning can be an incredi
ML0151EN	Machine Learning With R	this machine learning with r course
excourse21	Applied Machine Learning In Python	this course will introduce the learn
excourse40	Exploratory Data Analysis For Machine Learning	this first course in the ibm machine
excourse46	Machine Learning	machine learning is the science of
excourse47	Machine Learning For All	machine learning often called artifi
excourse48	Introduction To Machine Learning Language Processing	t s no secret that machine learning

Search...

COURSE_ID

TITLE

machine learning

(Select All)

Applied Machine Learning In

Data Science In Health Care

Deploying Machine Learning

Exploratory Data Analysis Fo

Introduction To Machine Lea

DESCRIPTION

Filters

Columns

Your courses:

st_table()

	COURSE_ID	TITLE
0	excourse46	Machine Learning
1	excourse47	Machine Learning For All

Recommendations generated!

	SCORE	TITLE	DESCRIPTION
0	0.6893	Machine Learning With R	this machine learning with r course dives into the basics of machine learning using an approachable and well known programming language you ll learn about supervised vs unsupervised learning look into how statistical modeling relates to machine learning and do a comparison of each

Personalized Learning Recommender

1. Select recommendation models

Select model:

Course Similarity

2. Tune Hyper-parameters:

Top courses

10

0 100

Course Similarity Threshold %

50

0 100

3. Training:

Train Model

4. Prediction

Recommend New Courses

Your courses:

	COURSE_ID	TITLE
0	ML0201EN	Robots Are Coming Build Iot Apps With Watson Swift And Node Red
1	GPXX0Z2PEN	Containerizing Packaging And Running A Spring Boot Application
2	DX0106EN	Data Science Bootcamp With R For University Proffesors
3	RAVSCTEST1	Scorm Test 1

Recommendations generated!

	SCORE	TITLE	DESCRIPTION
0	0.9476	Data Science Bootcamp	a multi day intensive in person data science bootcamp offered by big data university
1	0.6823	Data Science Bootcamp With Python For University Professors	data science bootcamp with python for university professors
2	0.6685	Data Science Bootcamp With Python For University Professors Advance	data science bootcamp with python for university professors advance
3	0.6499	Data Science Bootcamp With Python	data science bootcamp with python
4	0.6065	Data Science With Open Data	data science with open data

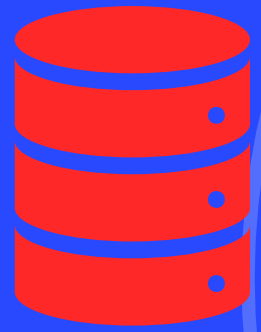
Outline

1. Content-based Course Recommender System using User Profile and Course
Genres
2. Content-based Course Recommender System using Course Similarities
3. Clustering based Course Recommender System

Conclusion

- User profile based model has the highest number of recommendation
- Similar matrix’s high complexity
- NMF, KNN as a solution

Appendix



Reporter: Pham Quoc Nam
(github, linkedin)

Github repository: IBM ML

Thank you

Pham Quoc Nam
0763117015
pikkerpham168@gmail.con