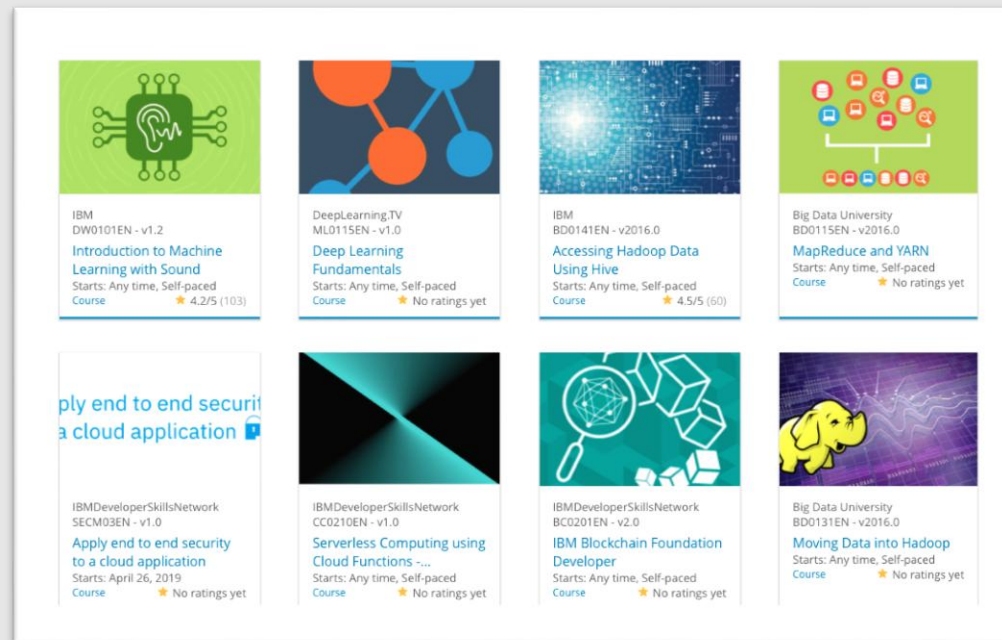# Build a Personalized Online Course Recommender System with Machine Learning

TING CHONG NA
26.05.2023

Outline

Introduction and Background

Exploratory Data Analysis

Content-based Recommender System
using Unsupervised Learning

Collaborative-filtering based Recommender System
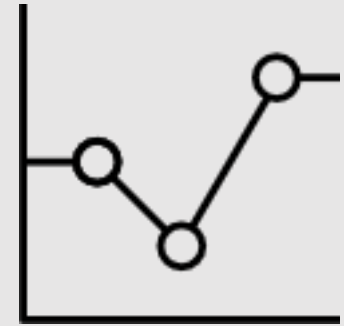using Supervised learning

Conclusion

Appendix

Introduction

IBM: MACHINE LEARNING CAPSTONE

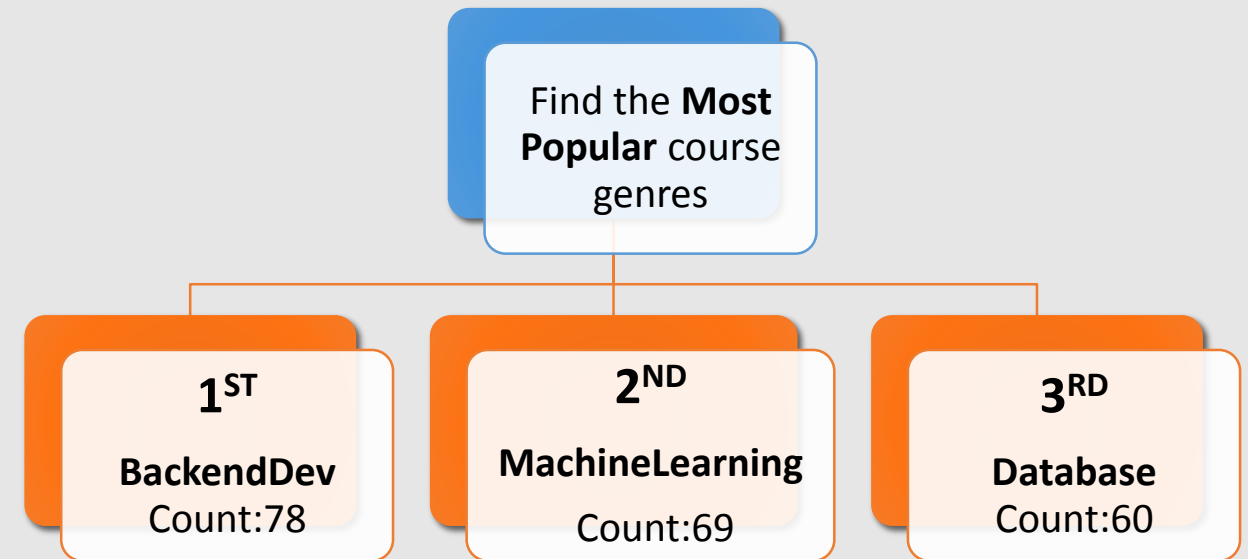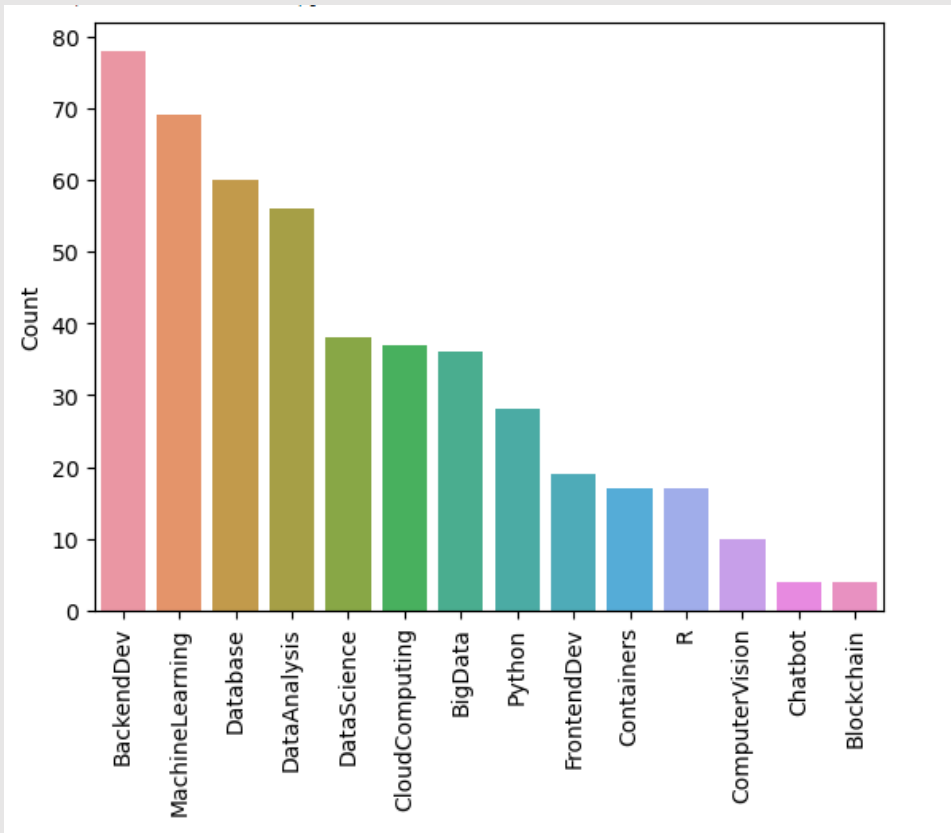**Recommender System**

Exploratory Data Analysis

Course counts per genre
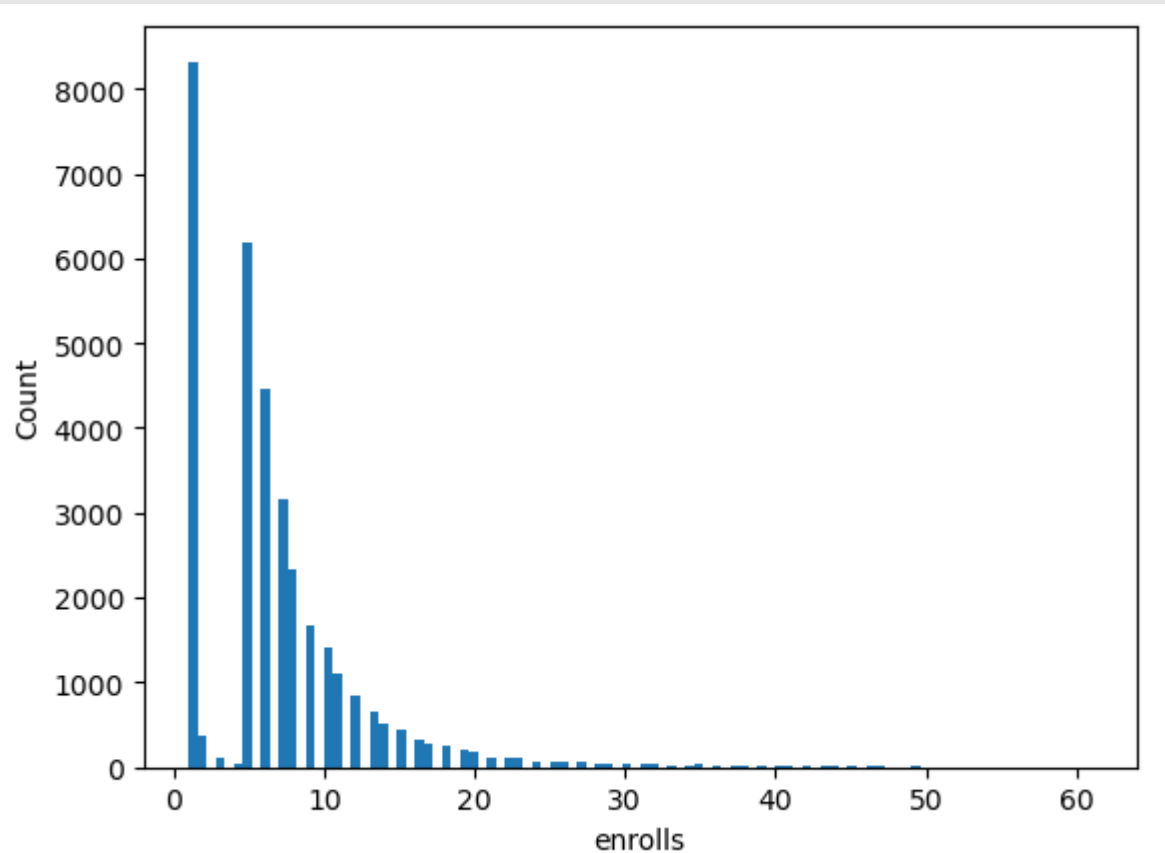
Course enrollment distribution

20 most popular courses

Word cloud of course titles

Course counts per genre

Course enrollment distribution

Histogram

Show distribution of course enrollment across all users

Min 1 course enrollment per user

(>8,000 users just enrolls 1 course)

Max 61 courses enrollment per user

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

20 Highest rated courses

1ST : Python for Data Science

2ND : Introduction to Data Science

3RD : Big Data 101

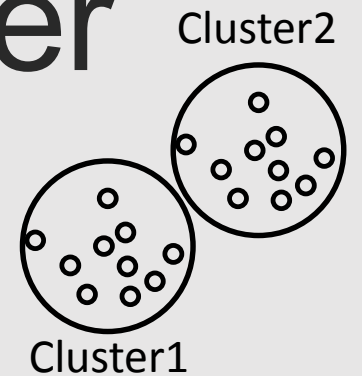Percentage of the top 20 course enrollments **63.3%**

Word cloud of course titles
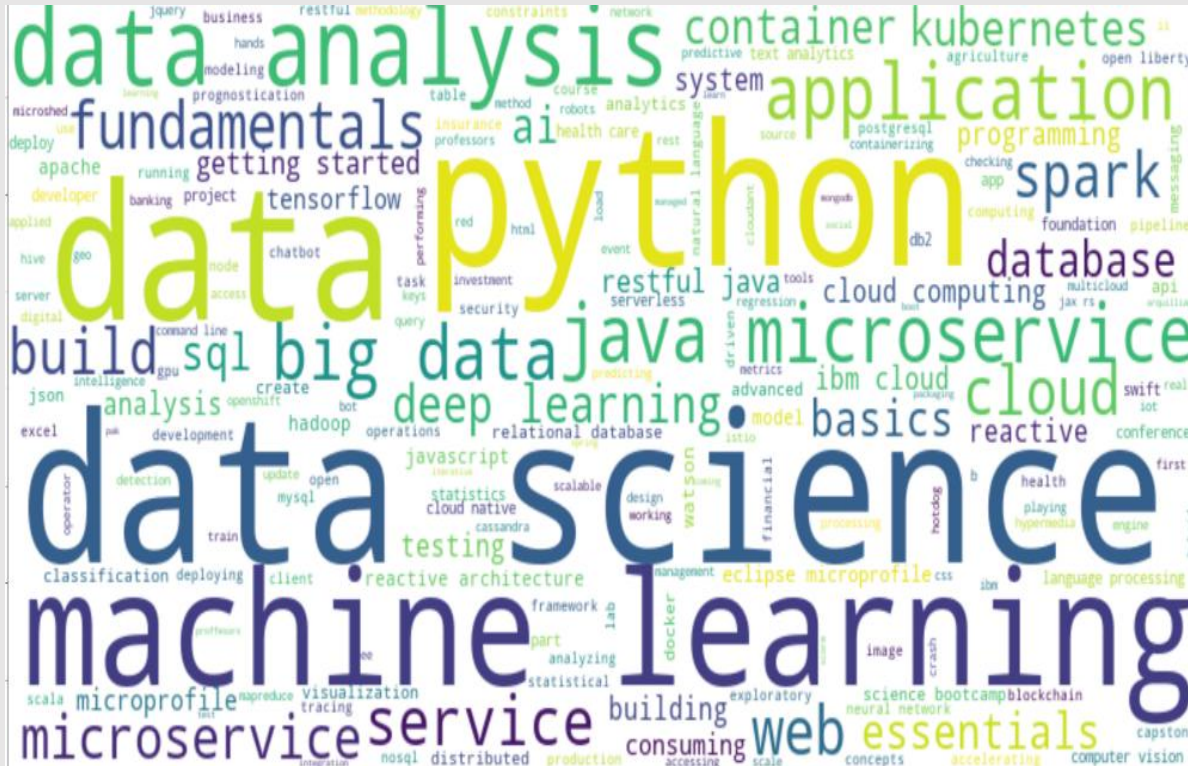
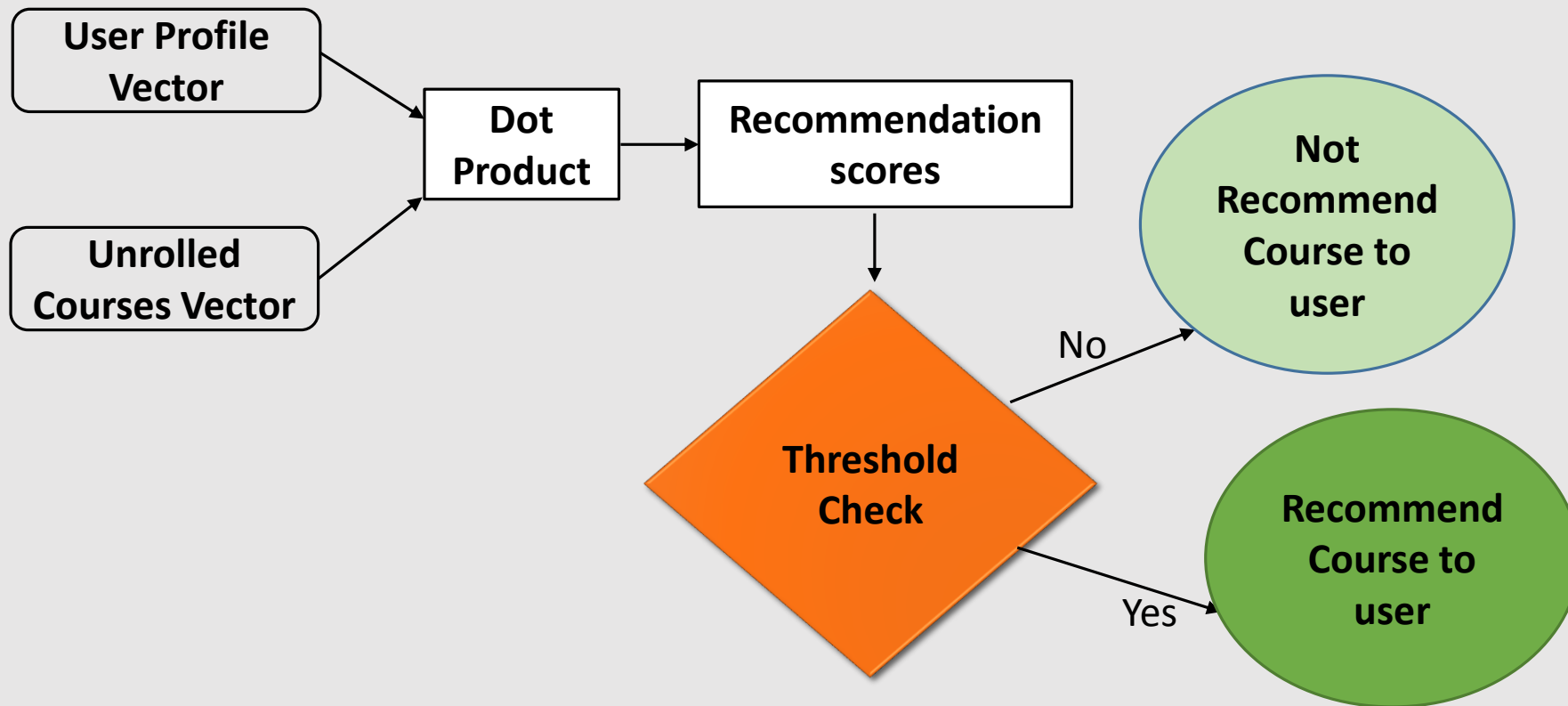Quick Visualization of the Popular learning topics across all the courses

Popular IT keywords :

Python, Data Science, Machine Learning, Big Data...

Content-based Recommender System using Unsupervised Learning

Cluster2

Cluster1

Flowchart

Average recommended courses per user

Top 10 most frequently recommended courses

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

Evaluation results of user profile-based recommender system

score_threshold = 10.0

```python
#On average, how many new courses have been recommended per test user?
res_df['COURSE_ID'].shape[0] / res_df['USER'].unique().shape[0]
```

61.81828703703704

```python
#What are the most frequently recommended courses?
res_df['COURSE_ID'].value_counts()[:10]
```
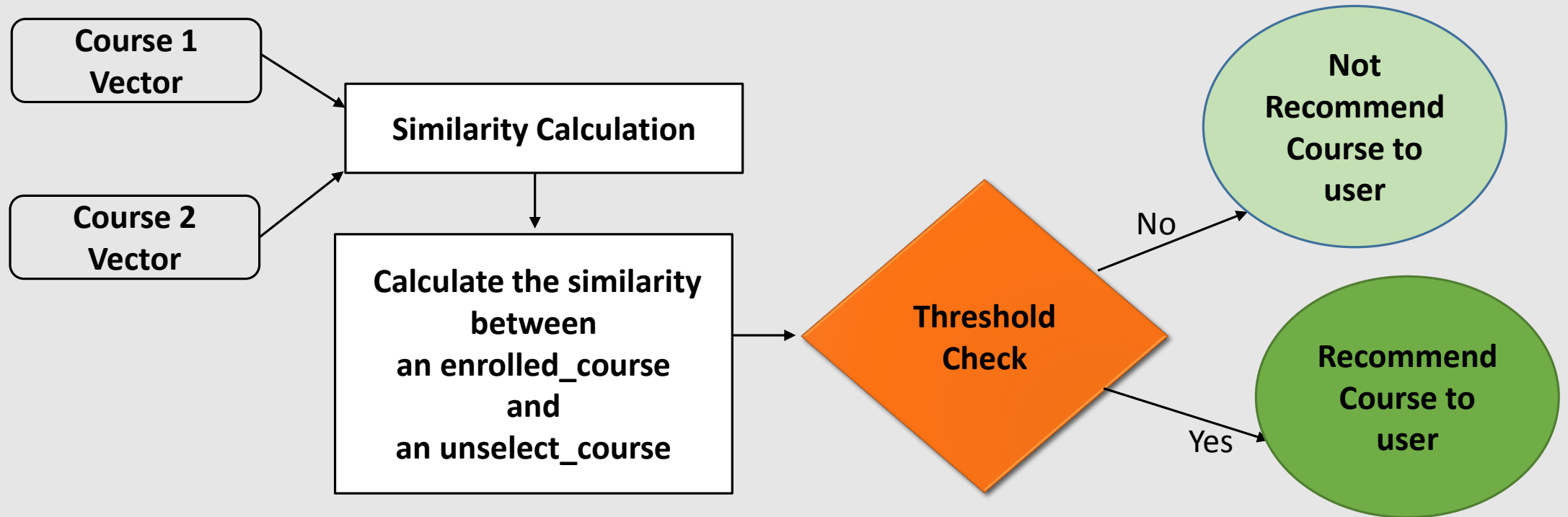
```
TA0106EN      608
GPXX0IBEN     548
excourse22    547
excourse21    547
ML0122EN      544
GPXX0TY1EN    533
excourse04    533
excourse06    533
excourse31    524
excourse73    516
Name: COURSE_ID, dtype: int64
```

Average New Course Recommended per user

Top 10 Most Frequently Recommended Courses

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 to each user
courses_count/len_user
```
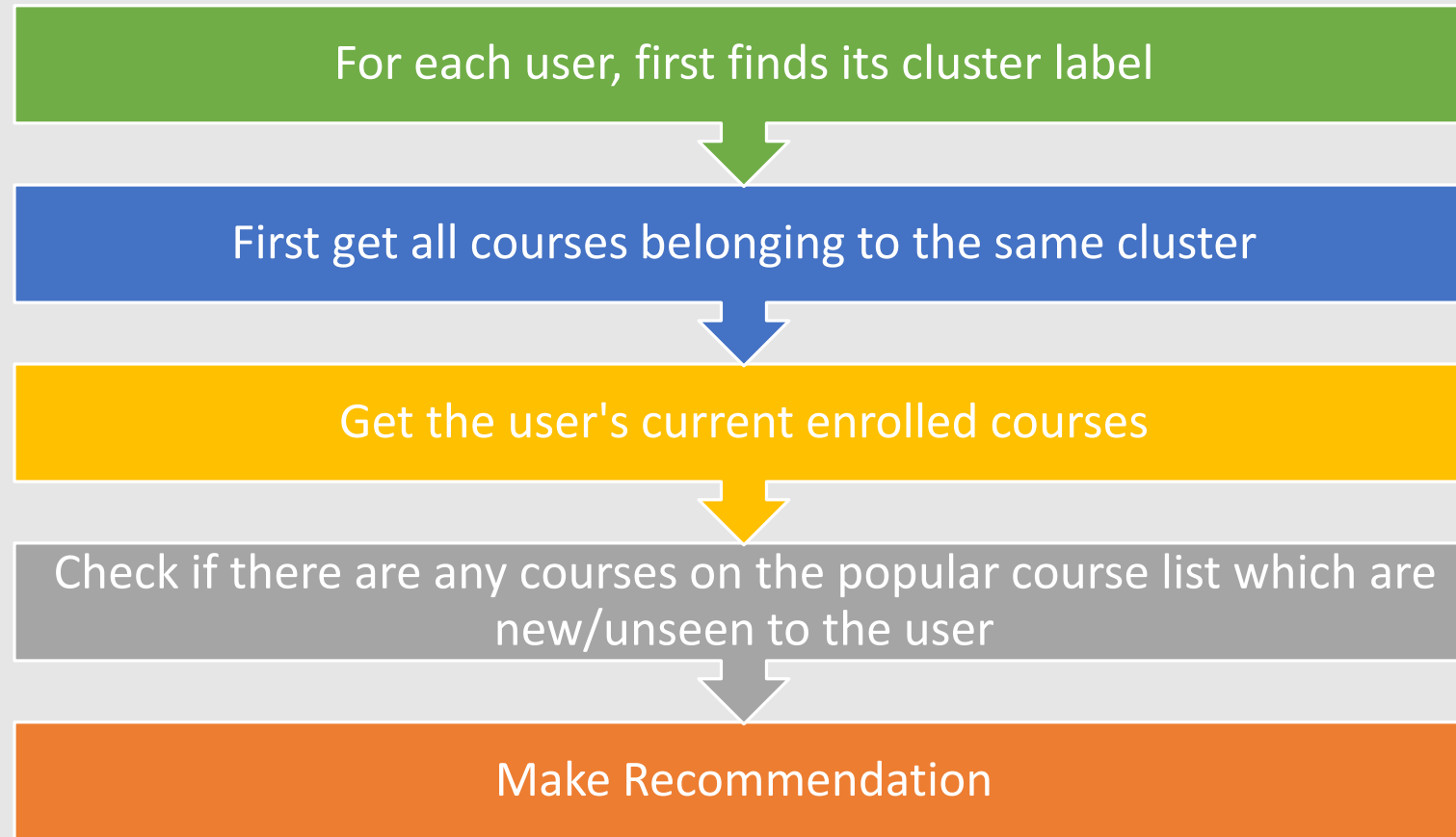
```
0.987
```

```
#What are the most frequently recommended courses?
courses_list=[]
for c in courses:
    for name in c:
        courses_list.append(name)
pd.Index(courses_list).value_counts()[:10]
```

```
excourse62      257
excourse22      257
WA0103EN        101
TA0105           41
DS0110EN         38
excourse47       24
excourse46       24
excourse63       23
excourse65       23
ML0151EN         17
dtype: int64
```

Average New Course Recommended per user

Top 10 Most Frequently Recommended Courses

Flowchart of clustering-based recommender system

For each user, first finds its cluster label

First get all courses belonging to the same cluster

Get the user's current enrolled courses

Check if there are any courses on the popular course list which are new/unseen to the user

Make Recommendation

Evaluation results of clustering-based recommender system

```
#On average, how many new/unseen courses have been recommended to each user?
len(courses)/len(set(users))
```
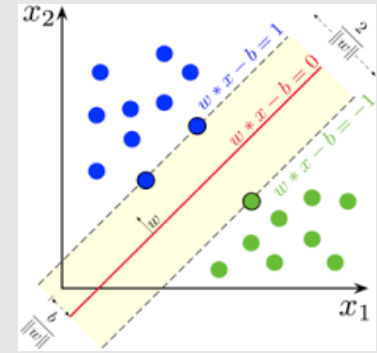
```
4.71
```

item	count
BC0101EN	277
DS0105EN	281
ML0101ENv3	299
BD0211EN	303
DS0103EN	320
DA0101EN	321
BD0111EN	420
BD0101EN	539
DS0101EN	551
PY0101EN	579

Average New Course Recommended per user

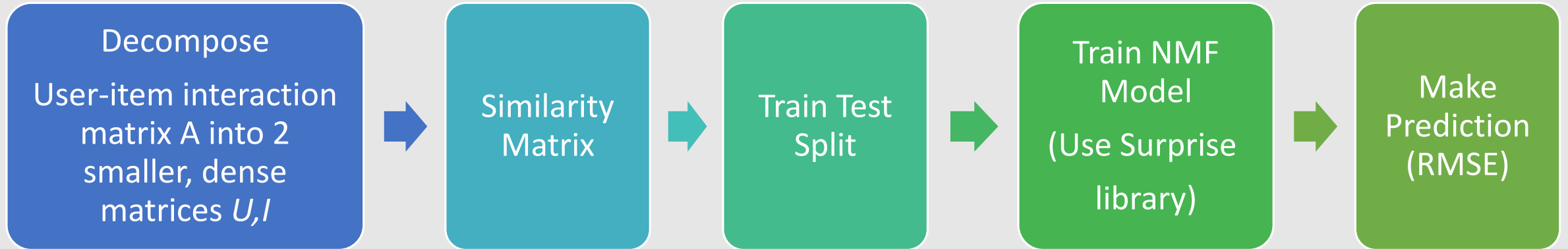**Top 10 Most Frequently Recommended Courses**

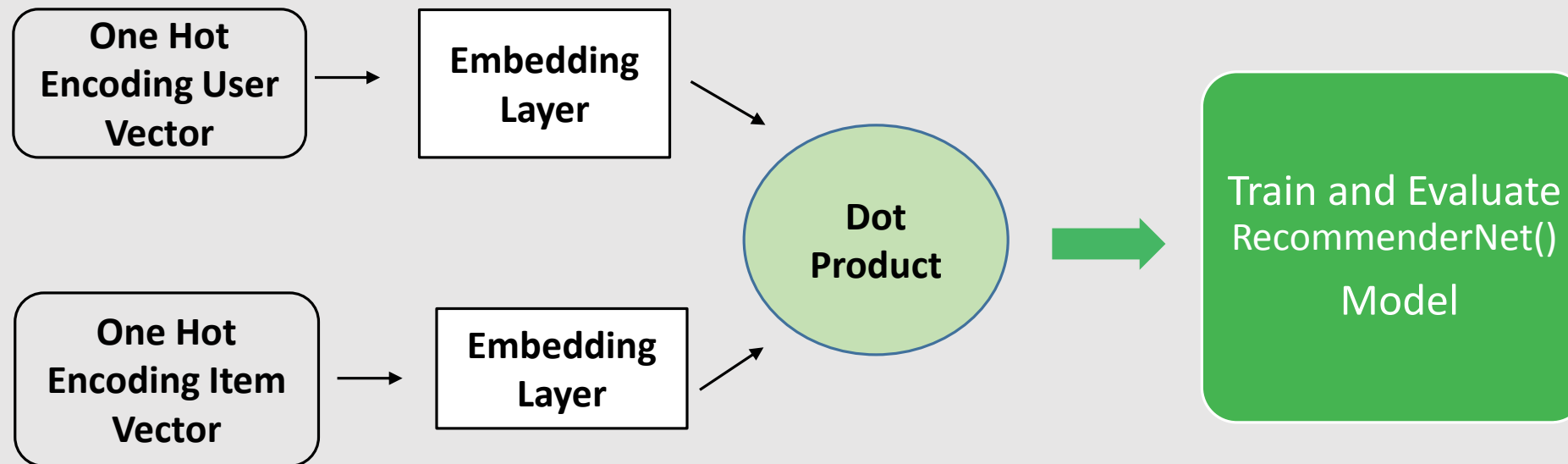# Collaborative-filtering Recommender System using Supervised Learning

# Flowchart of KNN based recommender system

Flowchart of NMF based recommender system

Decompose
User-item interaction matrix A into 2 smaller, dense matrices *U,I*

Similarity Matrix

Train Test Split

Train NMF Model
(Use Surprise library)

Make Prediction
(RMSE)

Flowchart of Neural Network Embedding based recommender system

Compare the performance of collaborative-filtering models
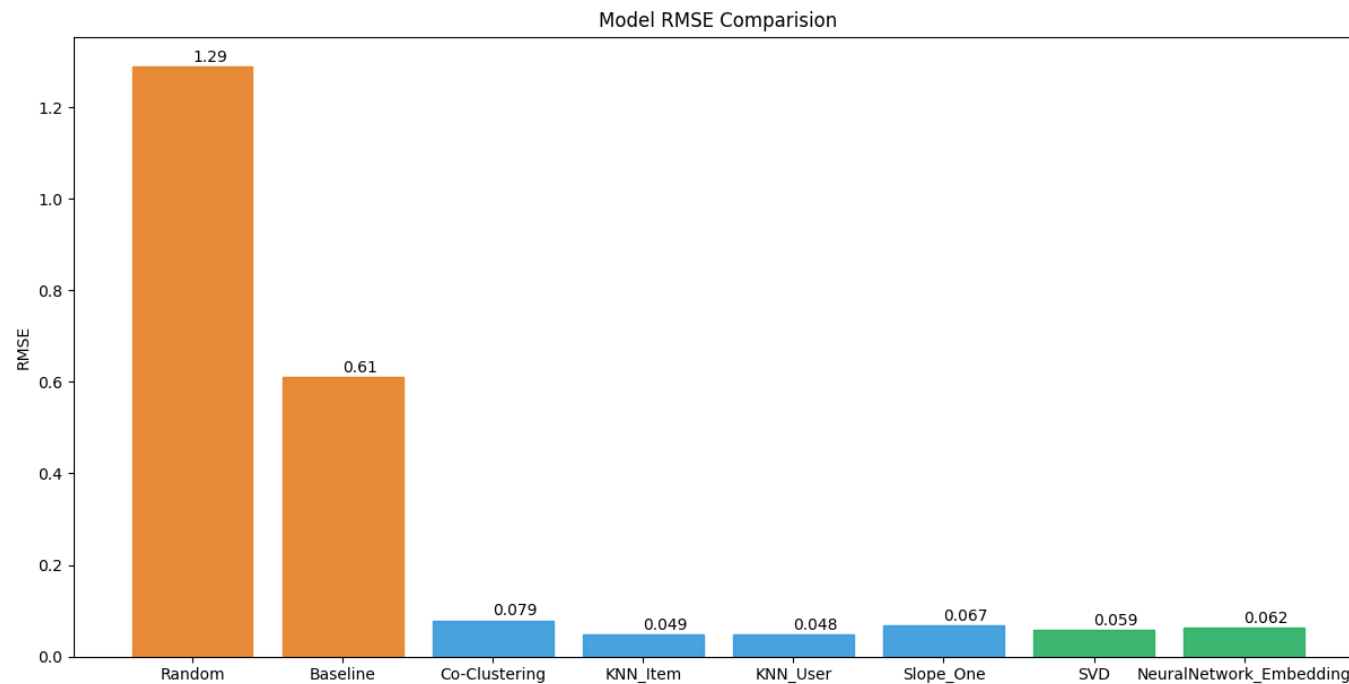
Conclusions

Unsupervised-Learning Based Recommender System	Supervised-Learning Based Recommender System	Predictive Model Based Recommender System
• KMeans • PCA	• K Nearest Neighbor (KNN) • Non-negative Matrix Factorization (NMF) • Neural Network	• Regression • Classification

Appendix

Github

- https://github.com/chongna95/IBM-Maching-Learning/tree/main