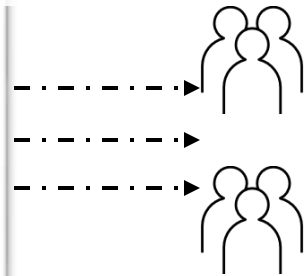
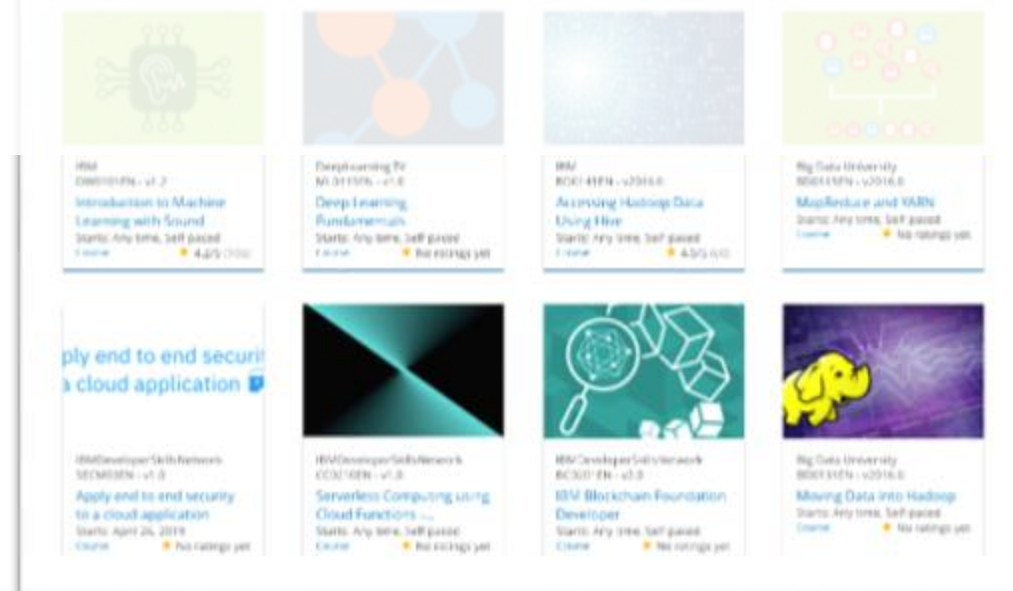


# Build a Personalized Online Course Recommender System with Machine Learning

Shahedul Karim  
26 July, 2025



# 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 (1)

---

- **Project background and context**

In today's fast-evolving digital learning environment, online learners often face challenges in discovering the most relevant courses that match their interests and learning goals. With hundreds or thousands of available courses, manually exploring and selecting the right content becomes overwhelming. This capstone project focuses on building a personalized **course recommender system** using real-world course data. The goal is to enhance learner experience by delivering customized course suggestions based on past behavior and preferences.



# Introduction (2)

---

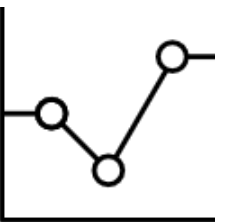
- **Problem states and hypotheses**

How can we design an intelligent recommendation system that effectively suggests relevant courses to users based on their profile, course content, and past interaction history

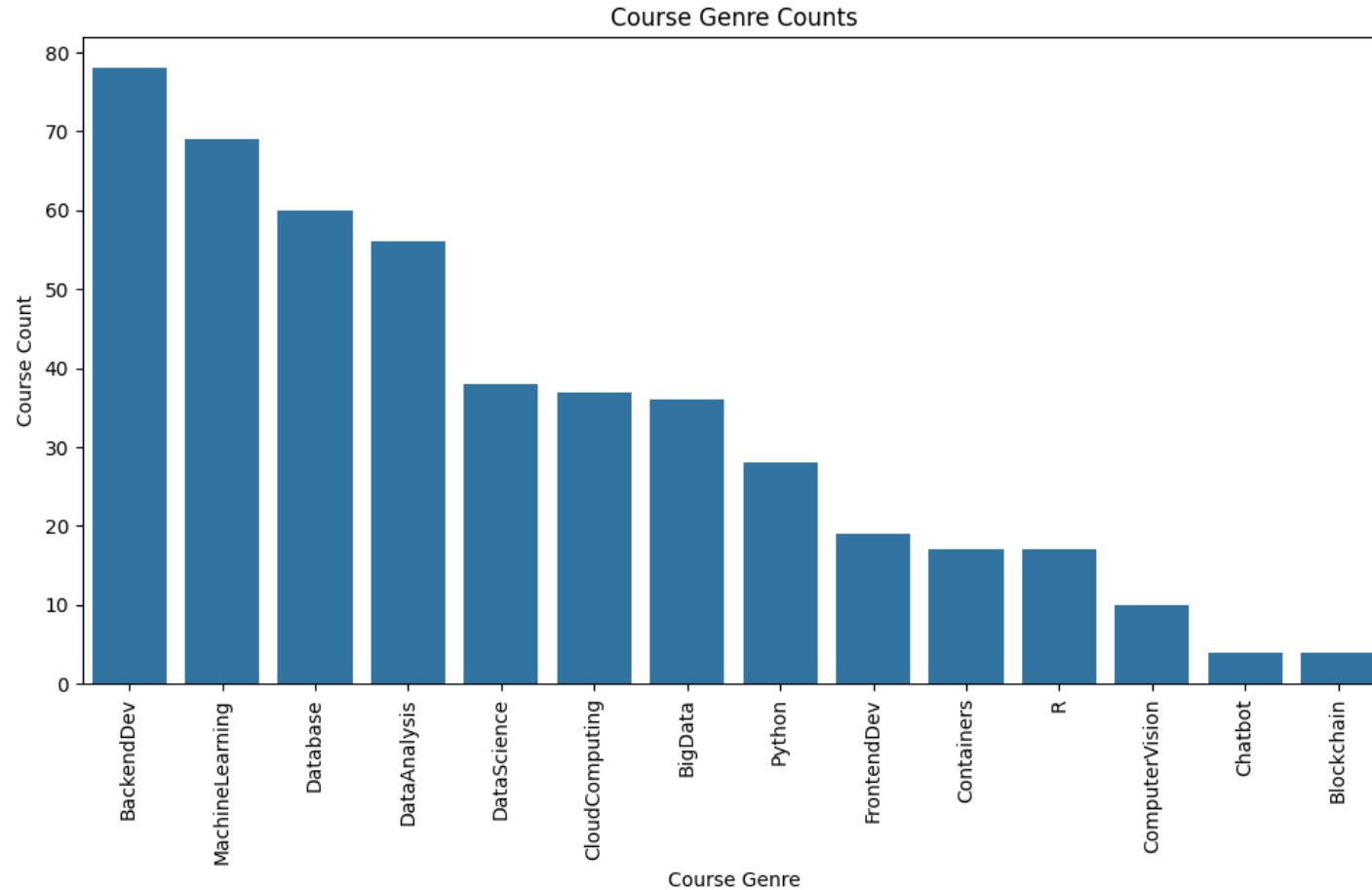
Hypotheses:

1. A content-based filtering approach using user preferences and course metadata (e.g., genre, title) will improve personalized course recommendations.
2. Collaborative filtering techniques such as KNN, NMF, and neural embeddings can discover latent user-course patterns and outperform content-based methods in accuracy.
3. Combining EDA insights with hybrid recommendation strategies will enhance course relevance and diversity in suggestions.

# Exploratory Data Analysis



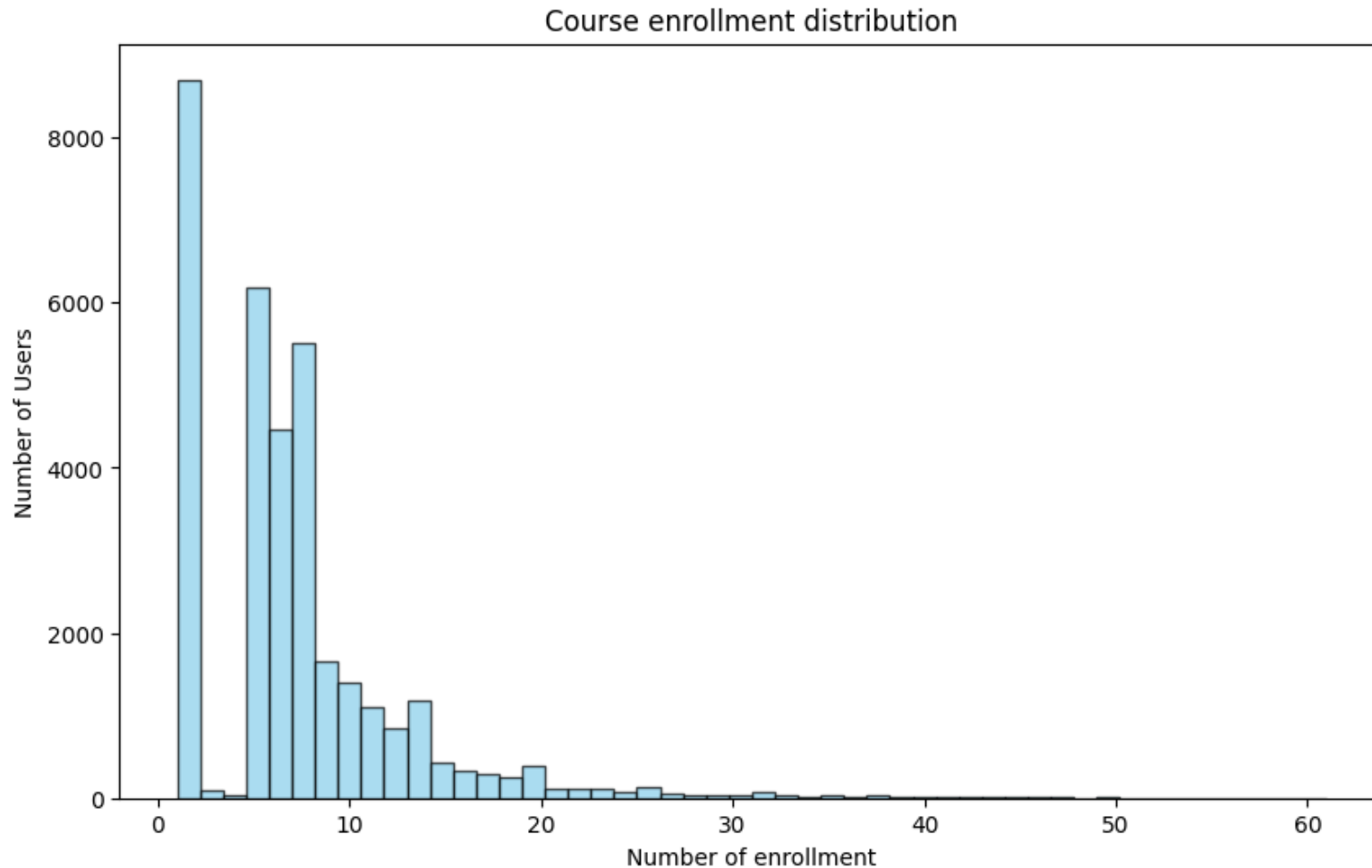
# Course counts per genre



## Key insight:

1. Backend dev has highest course count in the library
2. Blockchain has lowest course count
3. Trending topics like ML, DS has significant number of course

# Course enrollment distribution



## Key insight:

1. Most of the user enrolls in 0~10 courses.
2. Number of users enrolls in more than 20 courses is very less.

# 20 most popular courses

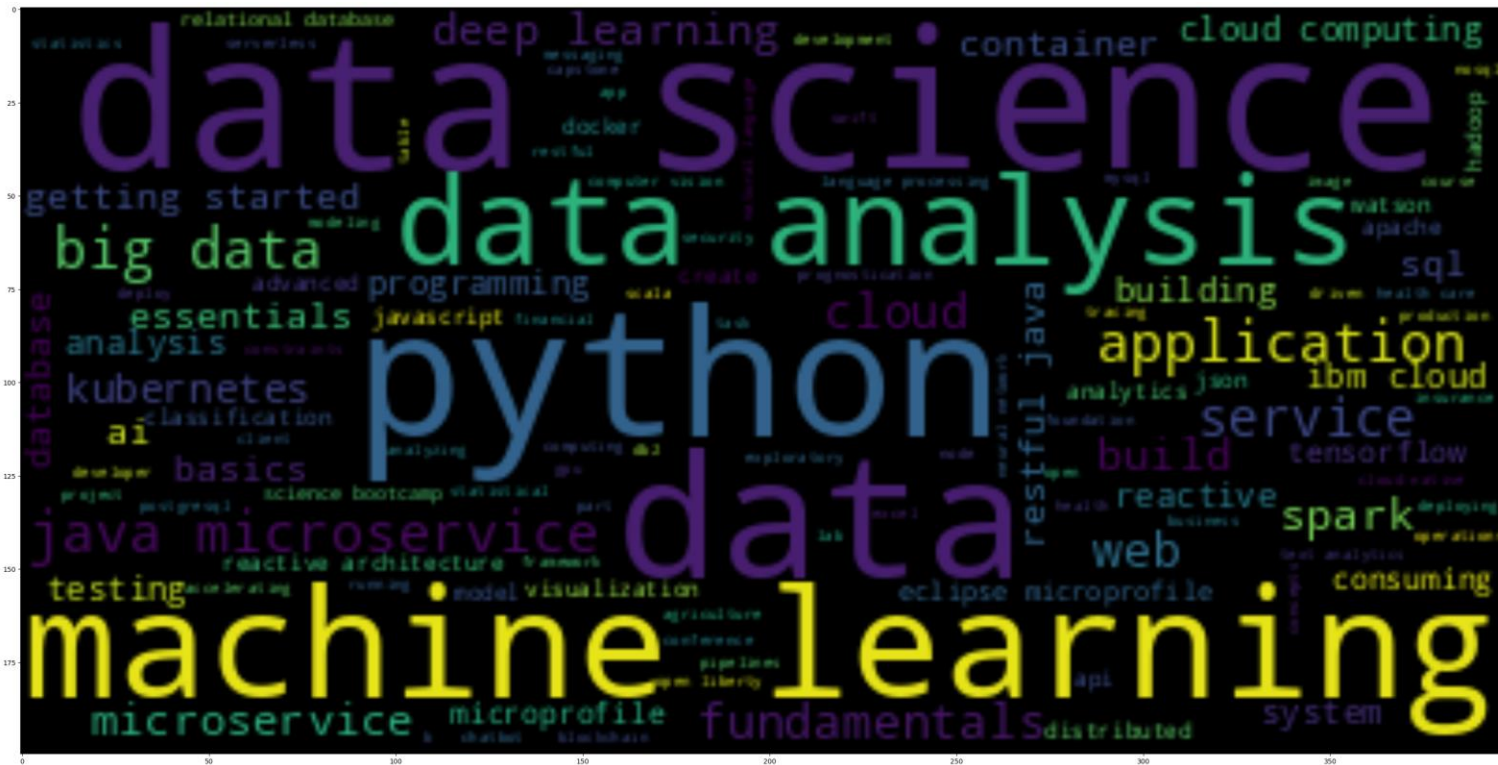
	📄 TITLE	# Enrolls
0	python for data science	14936
1	introduction to data science	14477
2	big data 101	13291
3	hadoop 101	10599
4	data analysis with python	8303
5	data science methodology	7719
6	machine learning with python	7644
7	spark fundamentals i	7551
8	data science hands on with open sou	7199
9	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 introdi	4480
17	sql and relational databases 101	3697
18	mapreduce and yarn	3670
19	data privacy fundamentals	3624

## Key insight:

1. Topics related to ML and DL are more trending among users.
2. Hadoop has more than 10k enrollments which suggest that users are also interested in distributed systems.



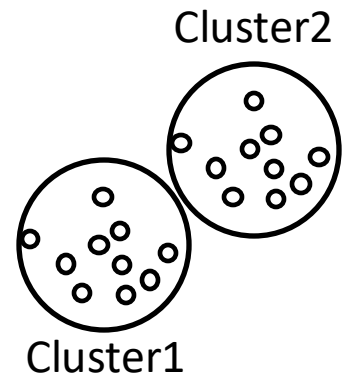
# Word cloud of course titles



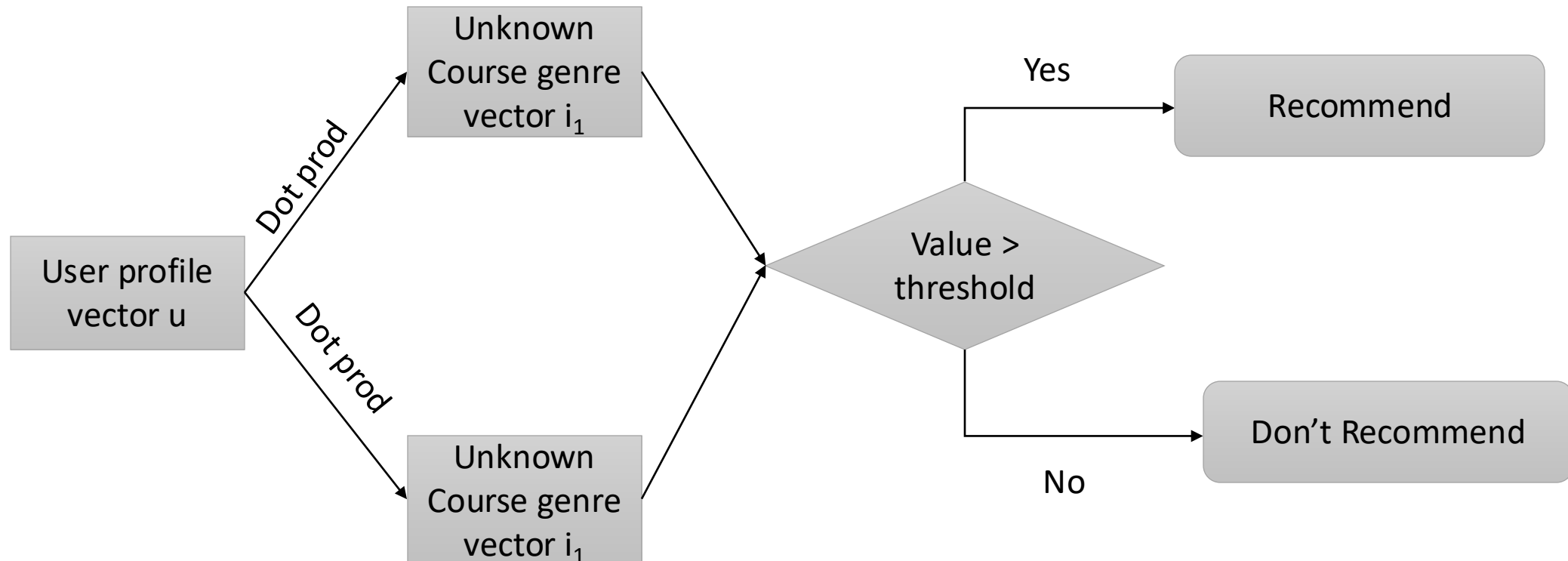
### Key insight:

1. Most frequently used words in course title is bigger in the illustration.
2. Words related to DS and ML has more frequency in course database.

# 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

score\_threshold = 9.0 is used for unknown course recommendation

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

```
#Average number of course recommendation for each user
reccomendation_count = len(res_df)
print(f"Recc. count : {reccomendation_count}")
user_count = len(test_users_df)
print(f"User count : {user_count}")

print(f"Avg {reccomendation_count/user_count}")
```

✓ 0.0s

Recc. count : 2189240  
User count : 233306  
Avg 9.38355635945925

What are the most frequently recommended courses

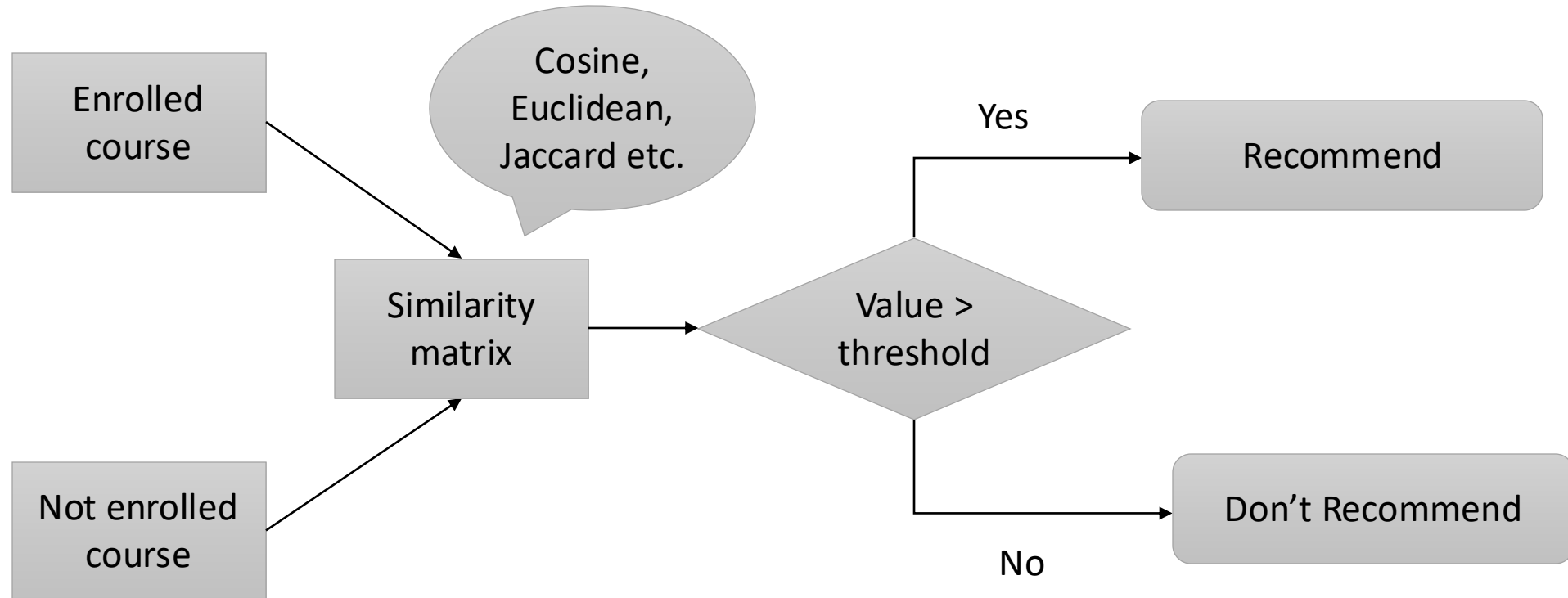
```
#Top 10 recommended courses across all users
res_df.groupby("COURSE_ID").size().head(10)
```

✓ 0.0s

COURSE_ID	#
AI0111EN	4377
BC0101EN	1409
BC0201EN	3589
BC0202EN	4452
BD0101EN	2615
BD0111EN	3285
BD0115EN	9716
BD0121EN	11747
BD0123EN	11770
BD0131EN	10492

10 rows x 1 cols 10 per page

# Flowchart of content-based recommender system using course similarity



# Evaluation results of course similarity based recommender system

similarity threshold = 6.5 is used for unknown course recommendation

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

Sum users: 33901

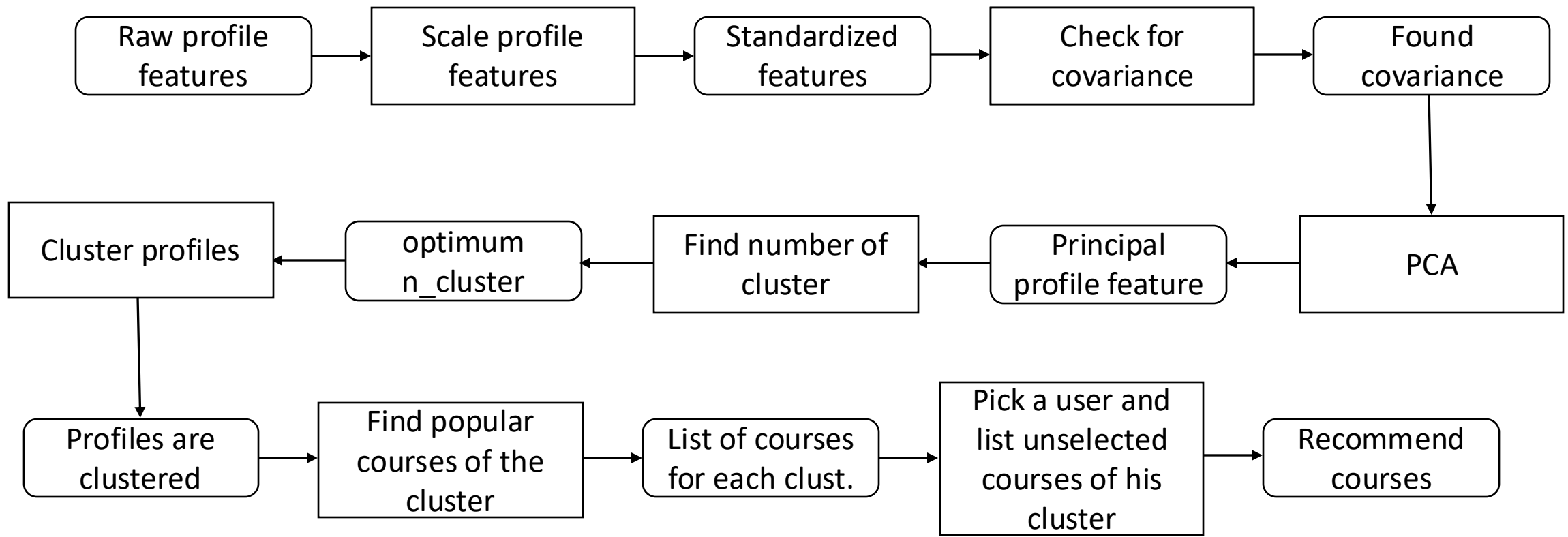
Sum recommendations: 171934

Average: 5.071649803840595

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

	Course	# Count
19	excourse63	14641
13	DS0110EN	14156
11	excourse72	13512
18	excourse74	13291
9	excourse67	13291
4	excourse23	8526
5	excourse36	8526
7	excourse32	8526
15	excourse33	8303
12	excourse38	8303

# Flowchart of clustering-based recommender system



# Evaluation results of clustering-based recommender system

n\_clusters=17 is used for clustering

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

Sum users : 33901

Sum recommendation : 232752

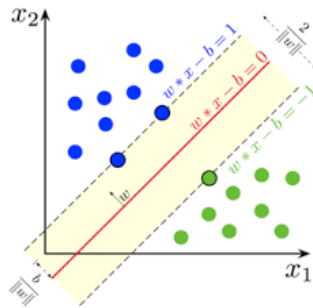
Average : 6.86563818176455

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

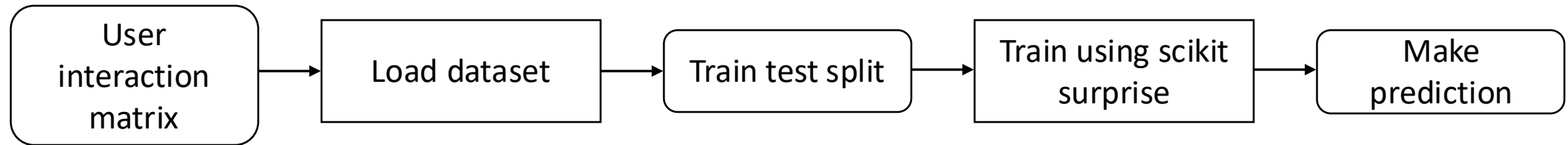
[('DS0103EN', 22079),  
('BD0101EN', 19664),  
('BD0111EN', 19606),  
('DS0101EN', 19424),  
('PY0101EN', 18965),  
('DS0105EN', 17488),  
('DA0101EN', 14703),  
('BD0211EN', 13588),  
('ML0101ENv3', 12272),  
('DVO101EN', 11368)]



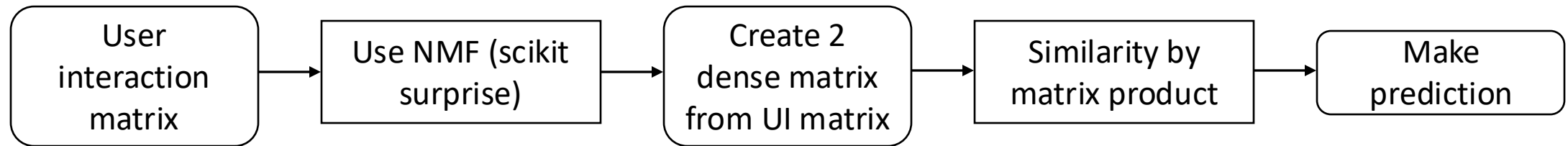
# Collaborative-filtering Recommender System using Supervised Learning



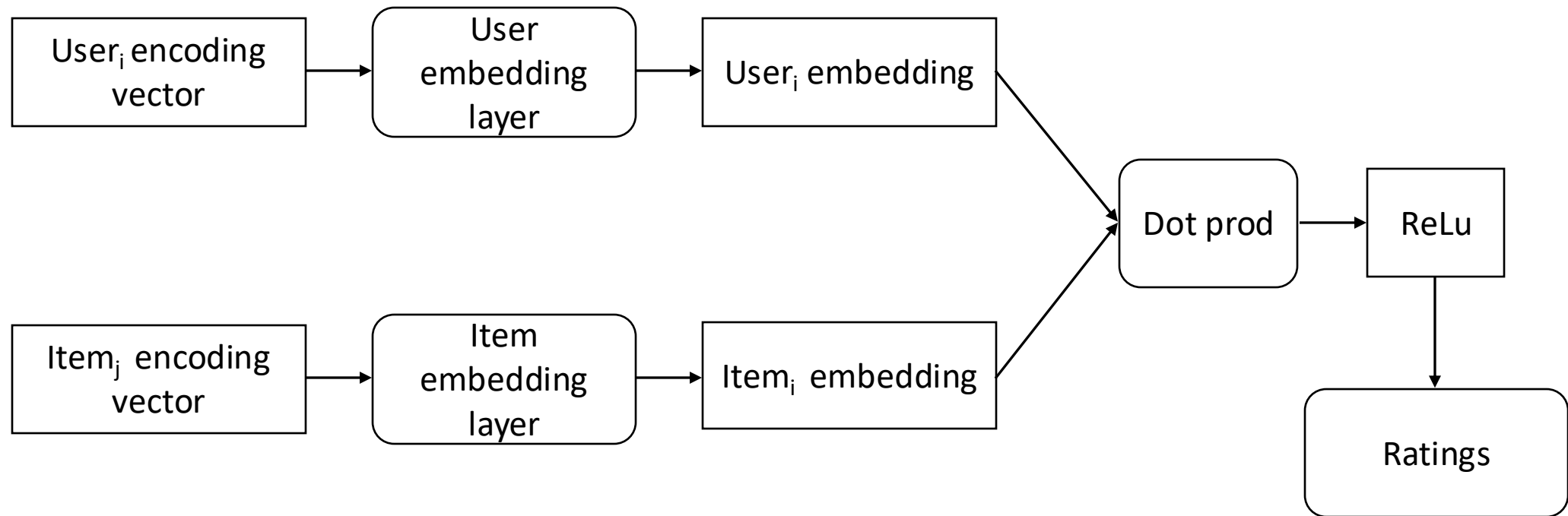
# Flowchart of KNN based recommender system



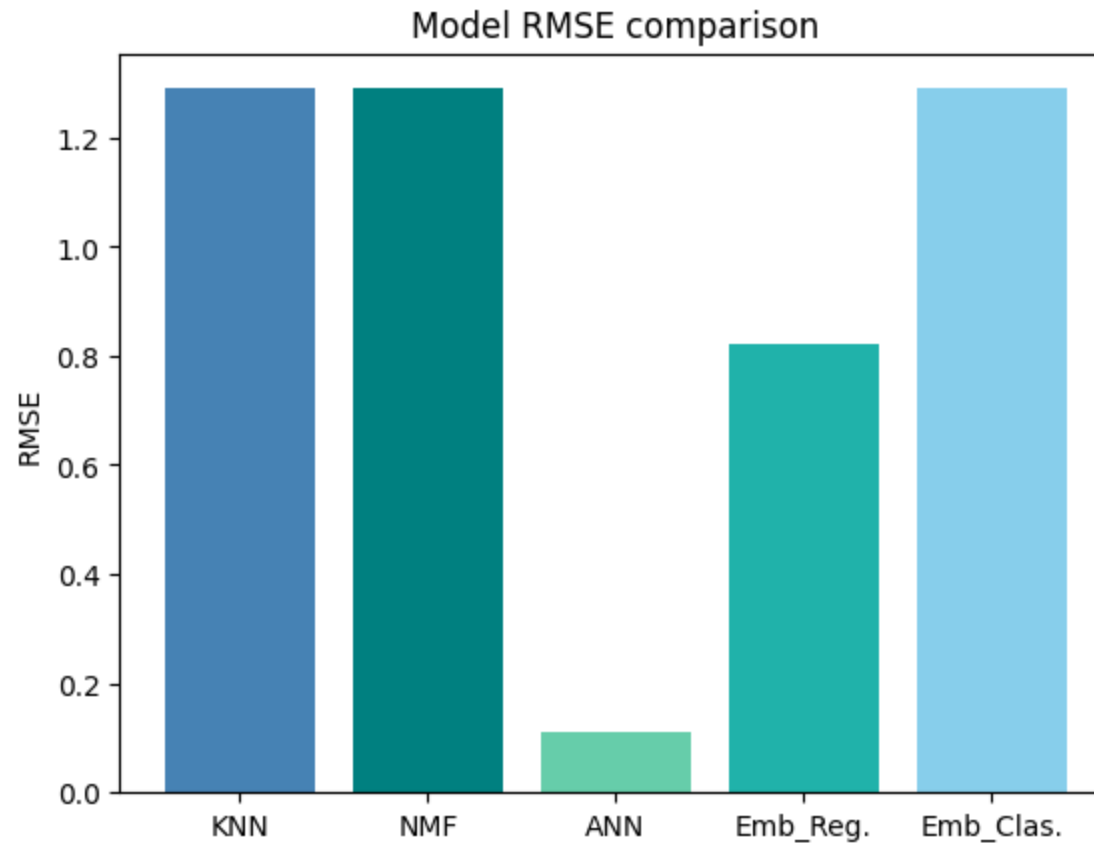
# Flowchart of NMF based recommender system



# Flowchart of Neural Network Embedding based recommender system



# Compare the performance of collaborative-filtering models



# Conclusions

---

- **EDA Insights** revealed key patterns in user-course interactions and popular genres, guiding our model design.
- **Content-Based Recommenders** provided personalized suggestions using user profiles, course similarity, and clustering.
- **Collaborative Filtering Models** (KNN, NMF, Neural Embeddings) successfully captured latent user preferences and improved recommendation accuracy.
- **Model Evaluation** showed that neural embedding models outperformed others in terms of recommendation precision and diversity.
- **Future Enhancements** include integrating hybrid models, real-time feedback loops, and deployment in a production environment for continuous improvement.

# Appendix

---

- [Coursera : Machine Learning Capstone](#)
- [Github repo for capstone materials](#)