Create a Custom Online Course Suggestion Tool Using Machine Learning

Eddie Tamir
03/31/2025

Outline

- Overview and Context
- Data Exploration
- Content-Driven Recommendation System with Unsupervised Methods
- User-Based Recommendation System with Supervised Techniques
- Summary and Insights
- Supplementary Material

Introduction

• Project Background and Context

Recommendation systems play a crucial role in tailoring and filtering content across platforms like e-commerce, streaming services, and social media. This project focuses on building a recommendation system by examining several techniques, including collaborative filtering, content-based methods, hybrid strategies, and deep learning. The objective is to enhance recommendation accuracy and relevance in alignment with user preferences.

• Problem Statement and Hypotheses

Problems:

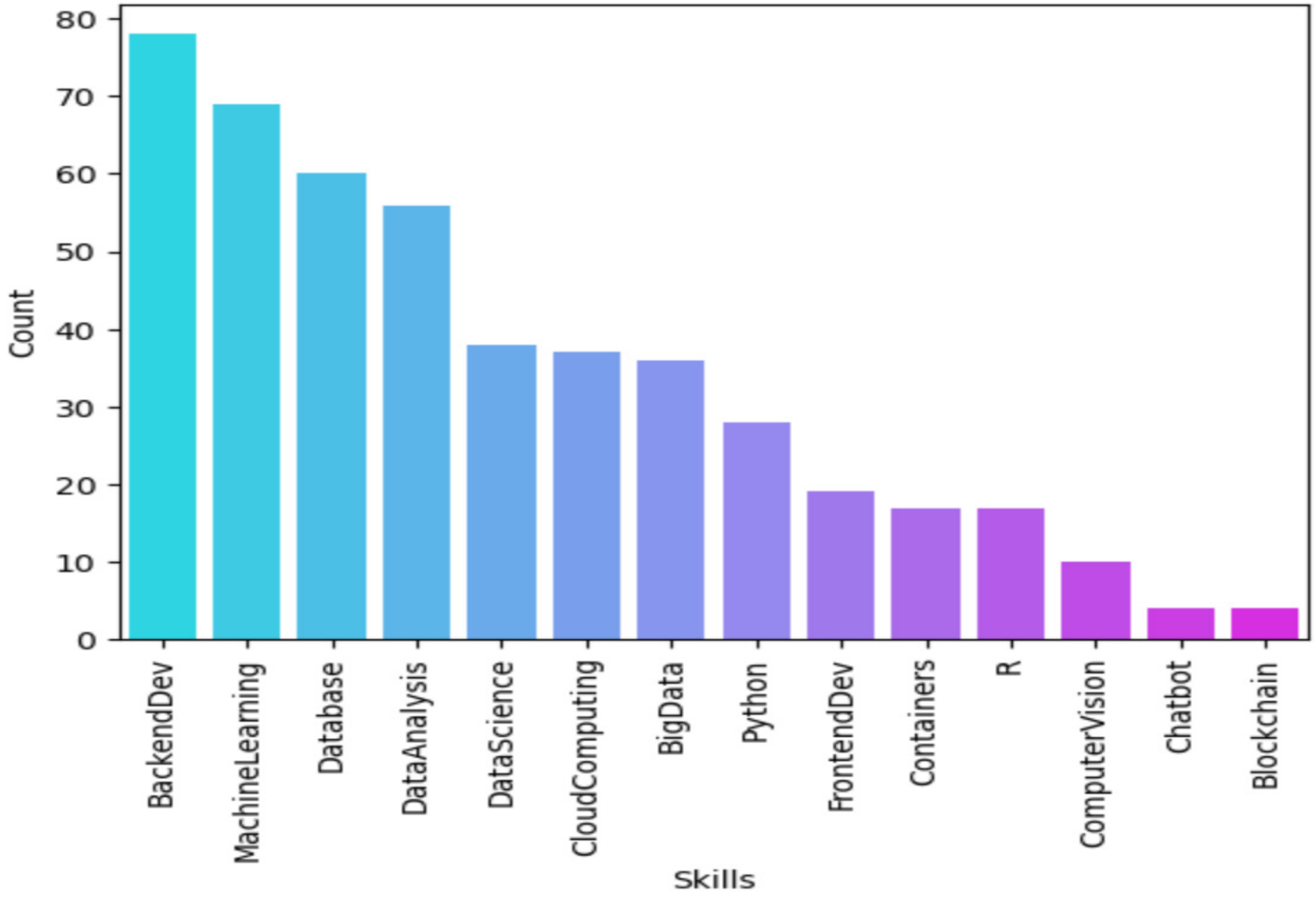
- Challenges with limited data and new user/item scenarios
- Handling growth and ensuring fast, real-time responses
- Recommendation bias and insufficient content variety

Hypotheses:

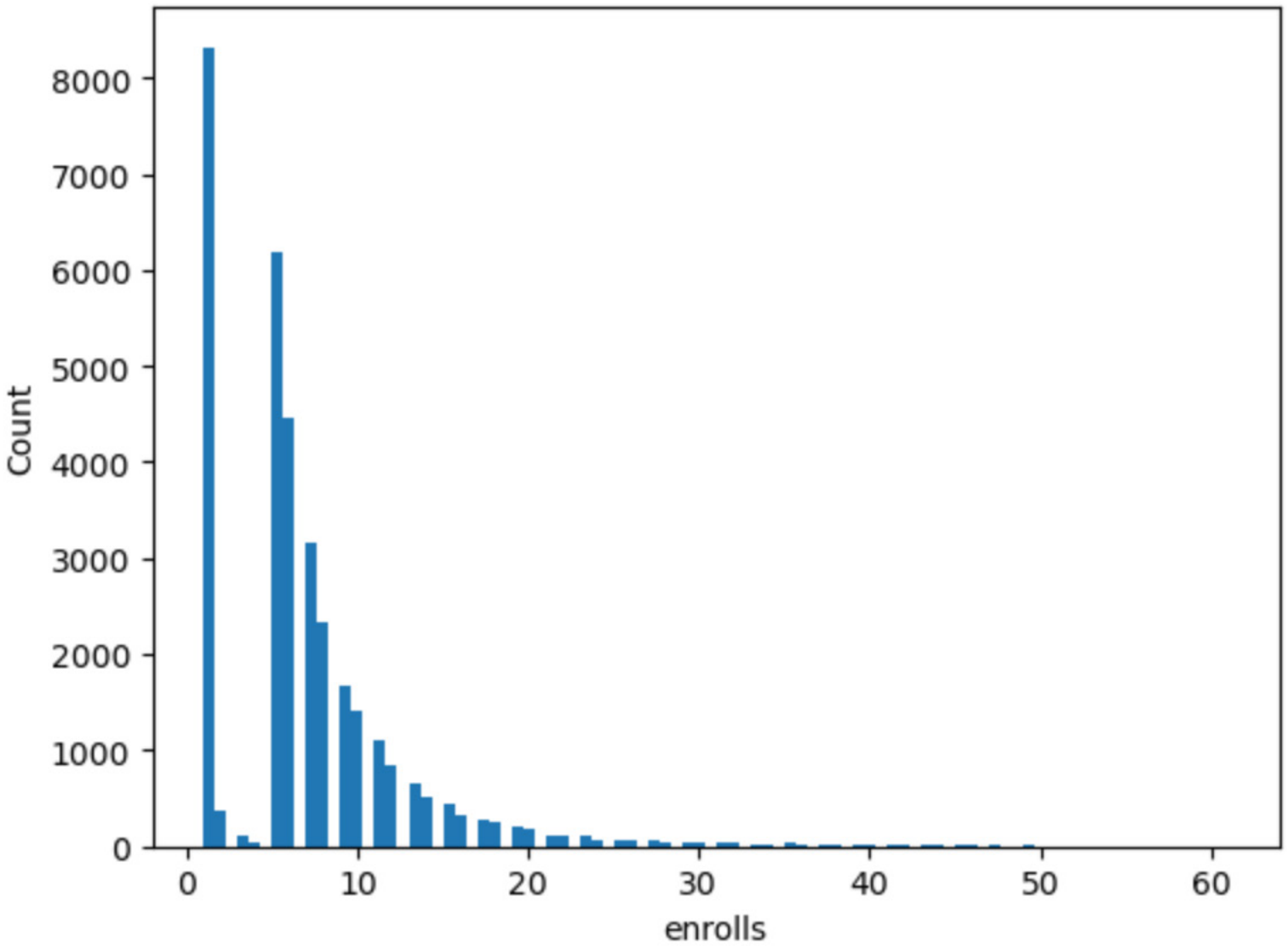
- Combining multiple methods boosts precision
- Deep learning leads to more effective recommendations
- Techniques like matrix factorization and graph models help address data sparsity
- Real-time personalization improves the user experience

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

Content-Based Recommendation System Leveraging Unsupervised Learning

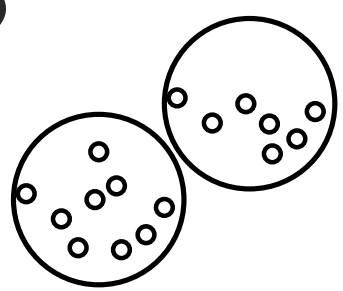
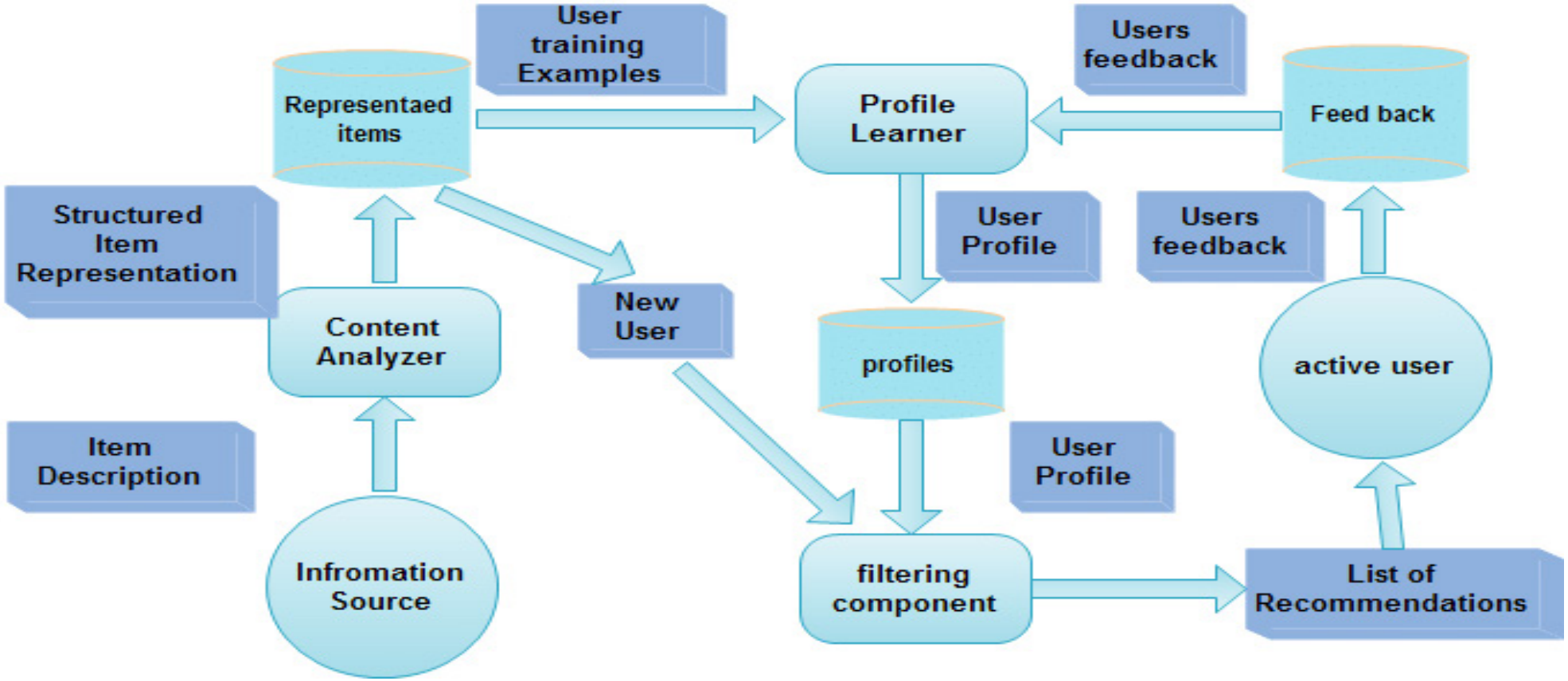


Diagram of a content-based recommender that matches user profiles with course genres



Evaluation results of user profile-based recommender system

score_threshold = 10.0

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

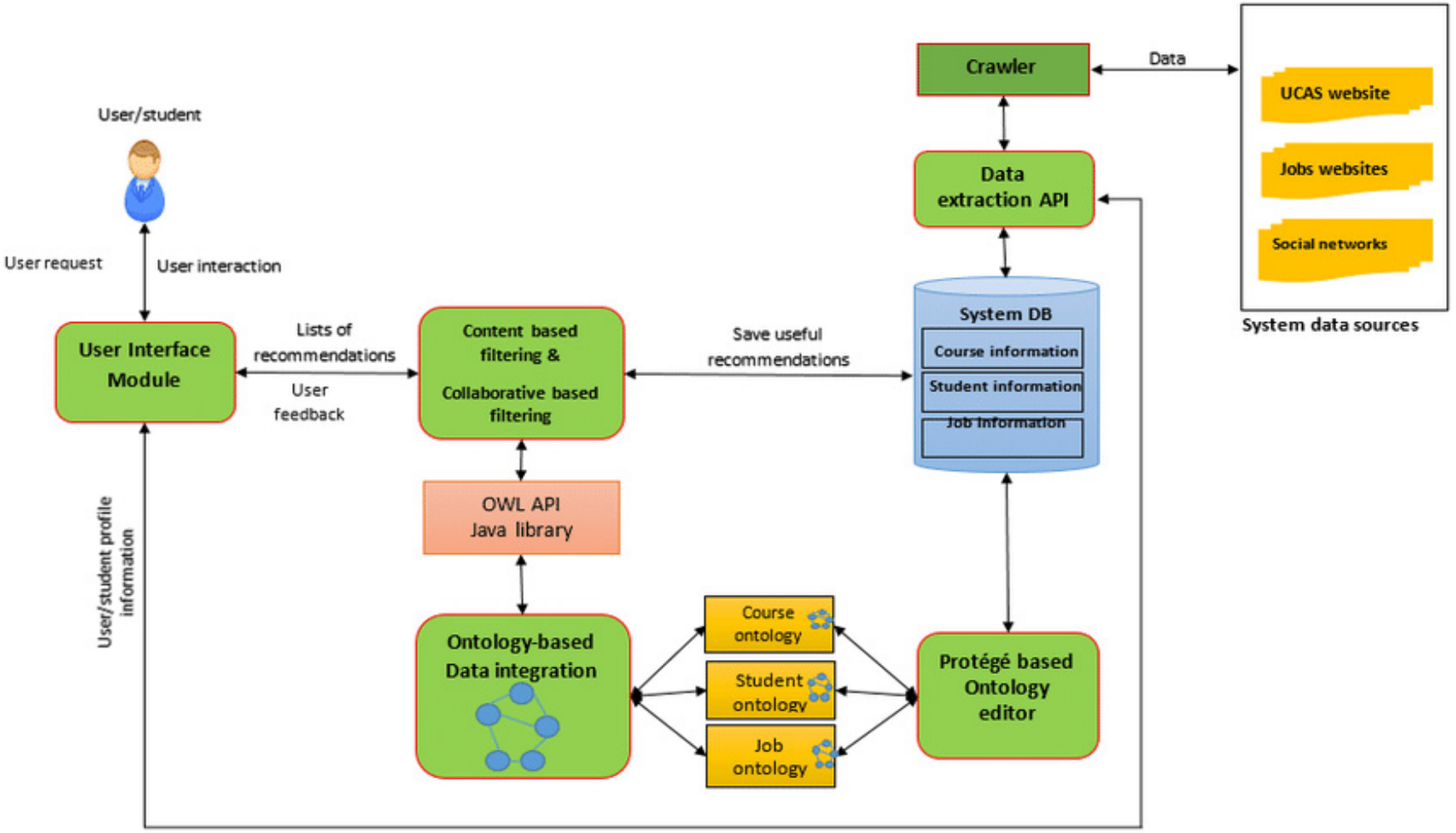
```
res_df['SCORE'].mean()
```

```
18.62679972290352
```

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

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

Diagram of a content-based recommender leveraging similarities between courses



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)

```
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'])
```

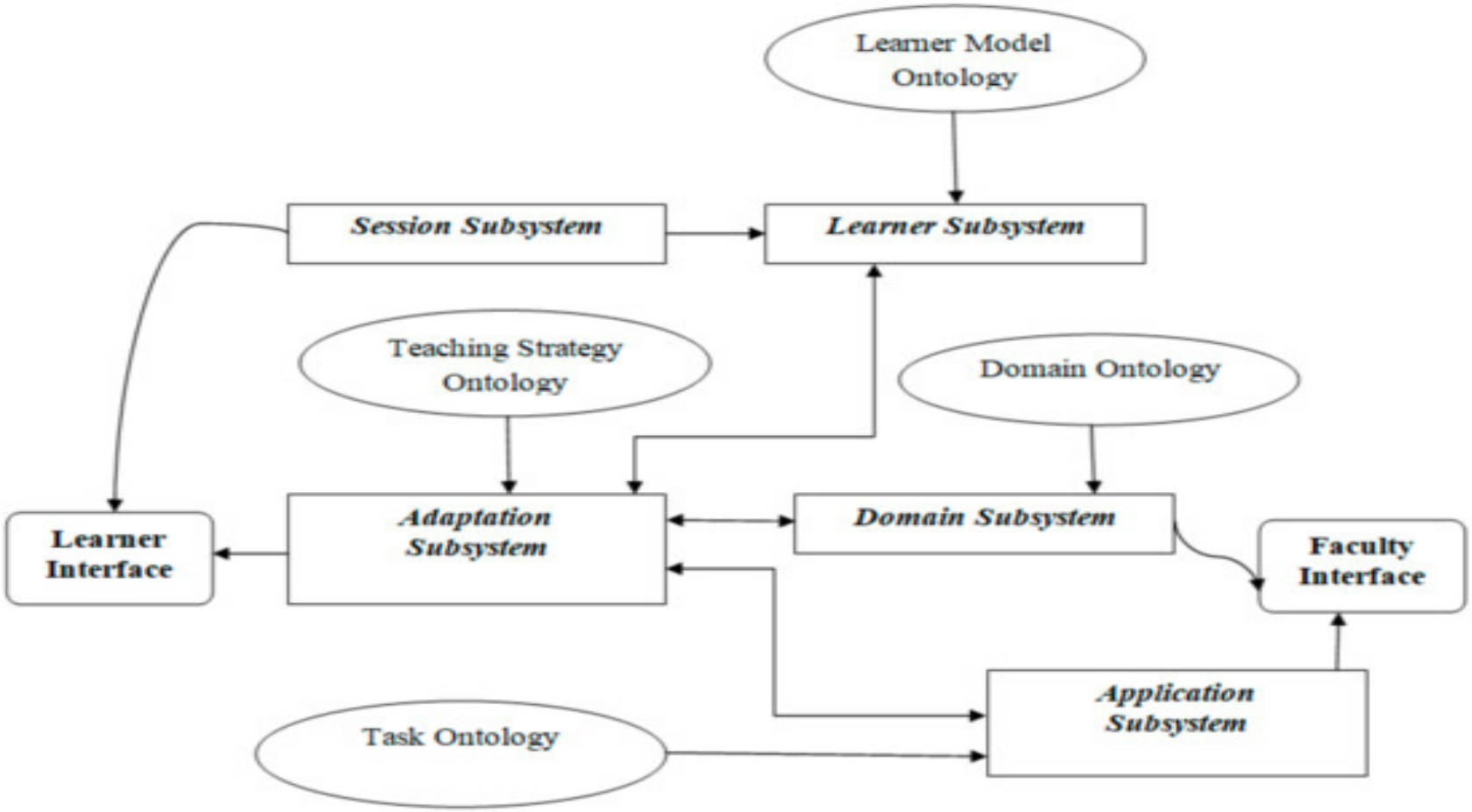
```
avg
```

```
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

Diagram of a clustering-driven recommendation 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)

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

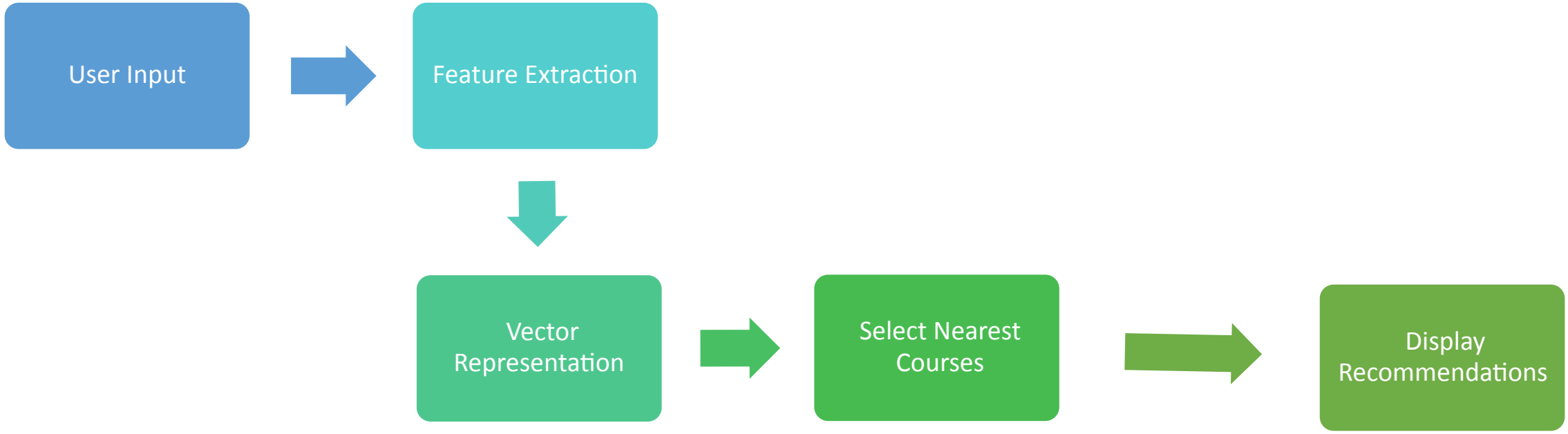
```
5.733
```

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

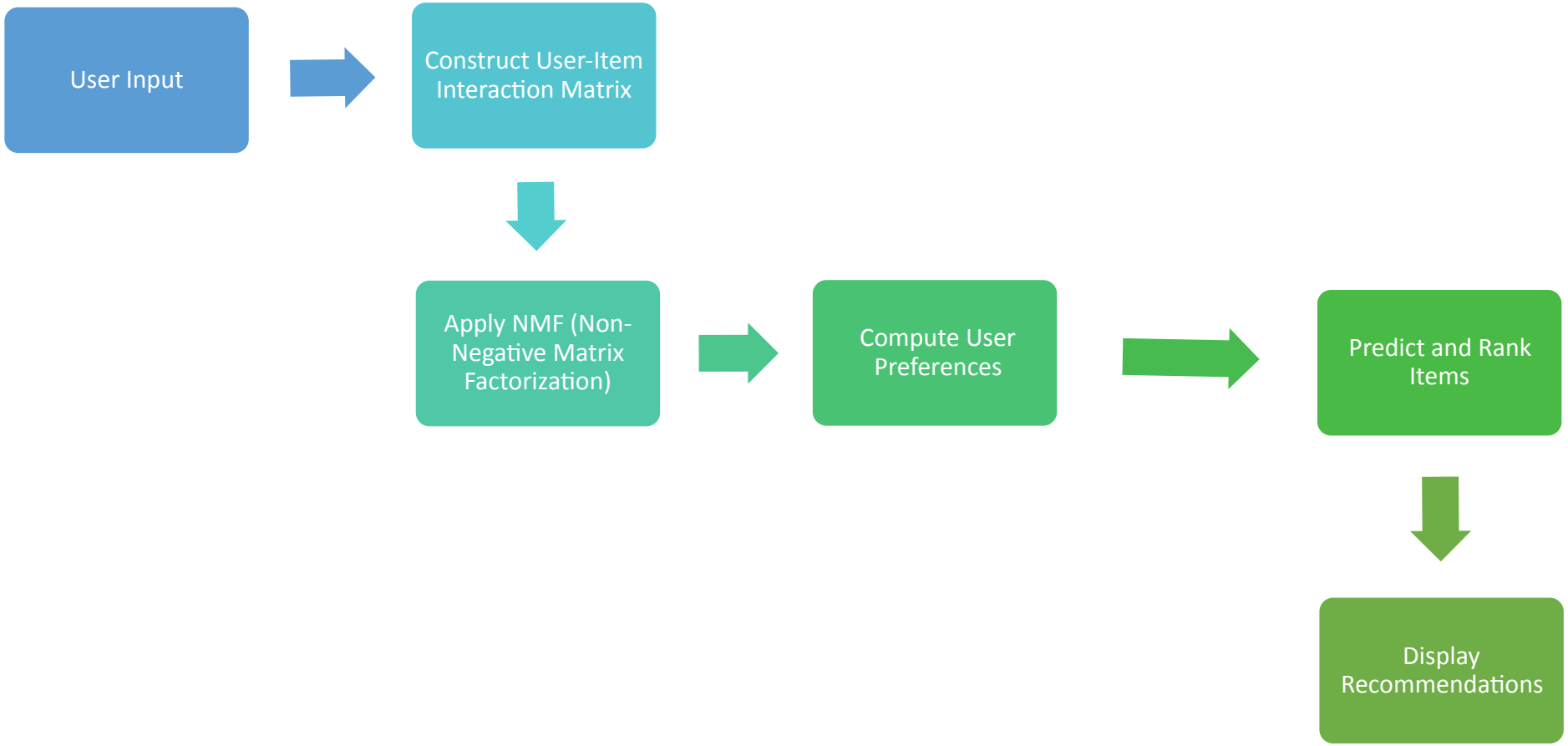
DS0103EN	579
DA0101EN	532
BD0111EN	456
DS0101EN	444
BD0101EN	428
PY0101EN	386
DS0105EN	319
ML0101ENv3	299
BC0101EN	296
ML0115EN	286

Supervised Learning-Based Collaborative Filtering Recommender

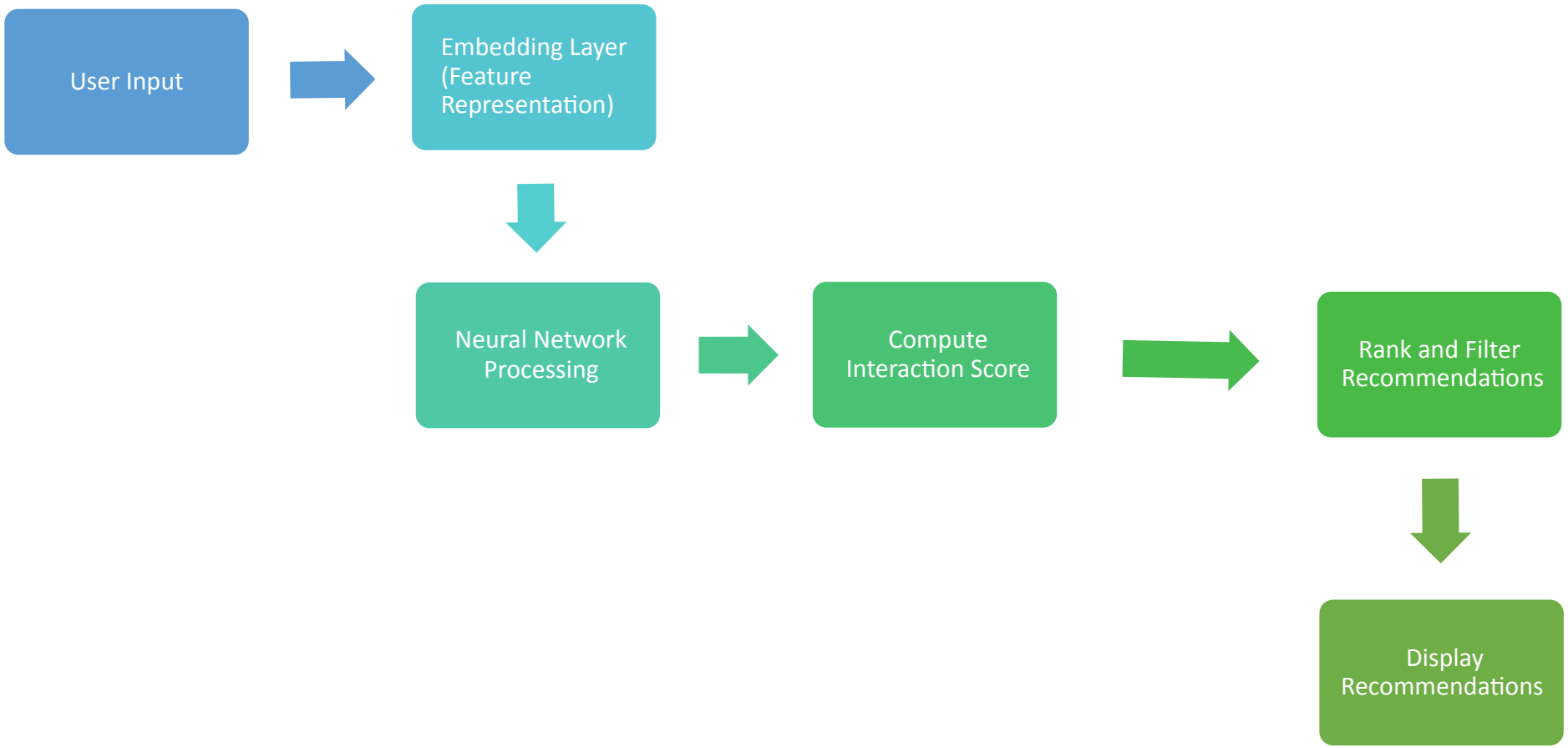
Recommendation Engine Based on Supervised Collaborative Filtering



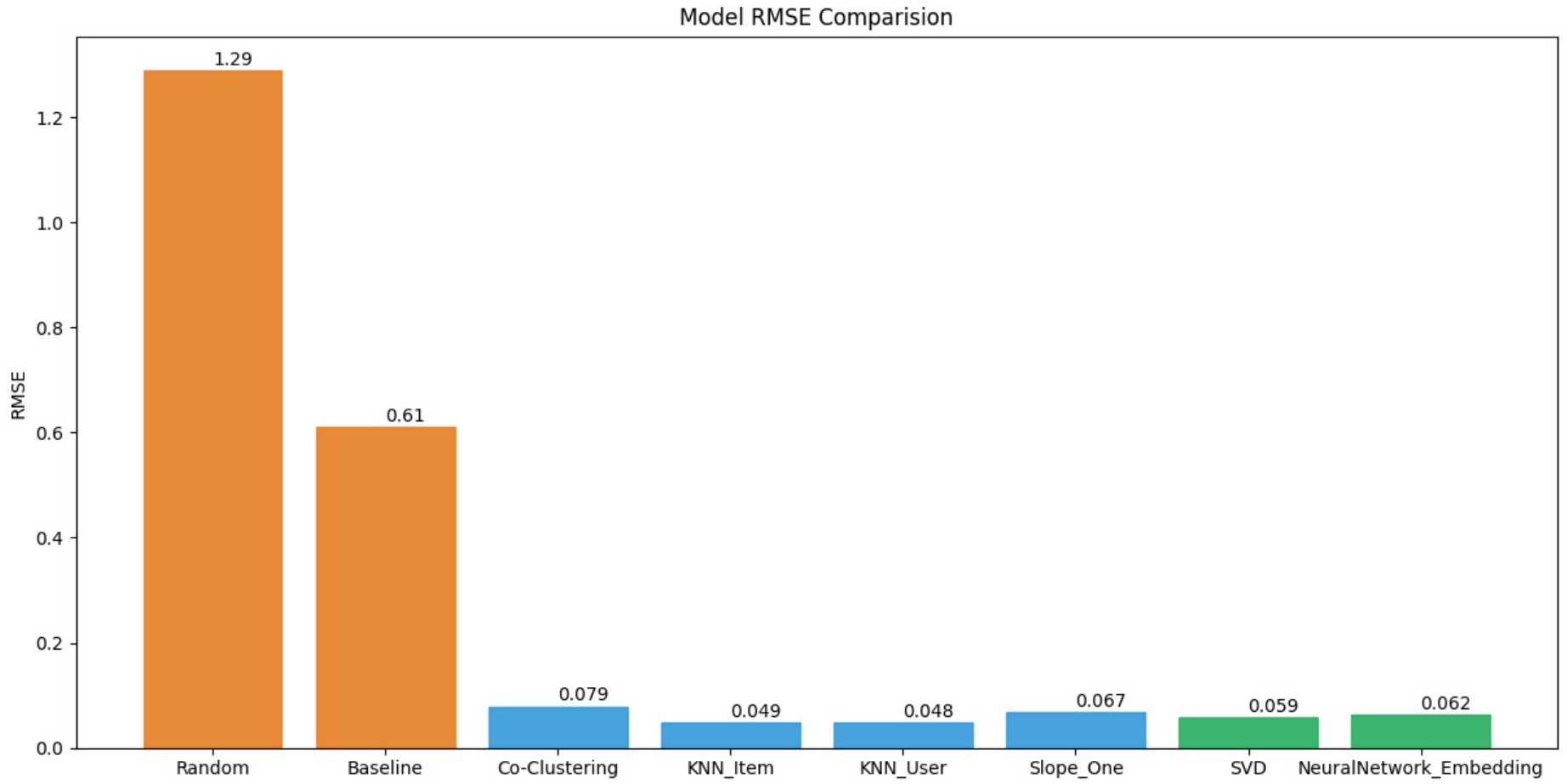
Visual representation of a recommender system leveraging NMF



Flowchart illustrating a neural network-based embedding recommender system



Evaluate the performance of different collaborative filtering models



Conclusions

- **Models show comparable levels of performance** [21]

- **The user profile-based approach generates the highest number of recommendations**
- **The stacking classifier achieves the best overall results**
- **Similarity matrix methods are computationally intensive**
- **NMF offers an effective alternative to address complexity**

Appendix

A Neural Network Embedding-based Recommendation System represents users and items as dense vectors to uncover latent relationships. These embeddings are processed through a neural network, which estimates similarities or predicts user-item ratings. Based on these predicted scores, the system ranks the items and suggests the most relevant ones to the user.