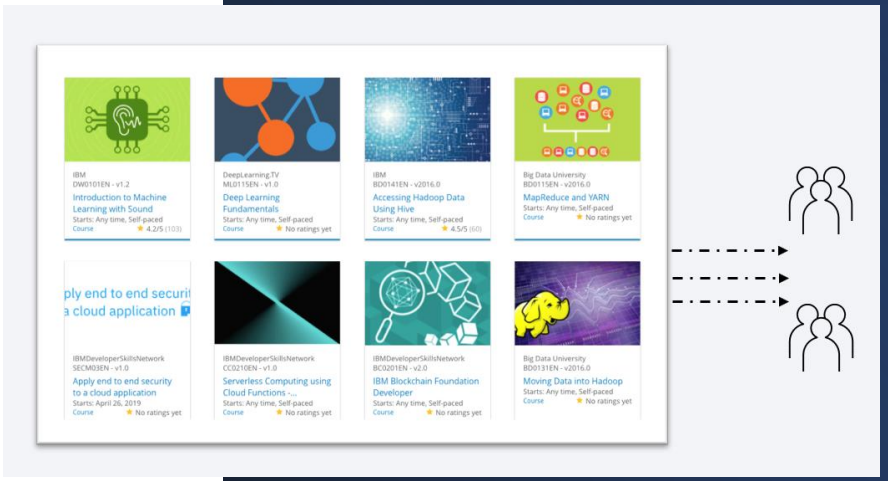
Build a Personalized Online Course Recommender System with Machine Learning



- Raffaele Pane

- 2024-03-13

Outline

Introduction and Background

Exploratory Data Analysis

Content-based Recommender System using Unsupervised Learning

Collaborative-filtering based Recommender System using Supervised learning

Conclusion

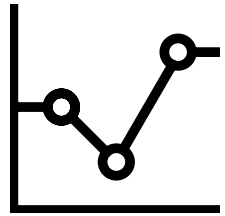
Appendix

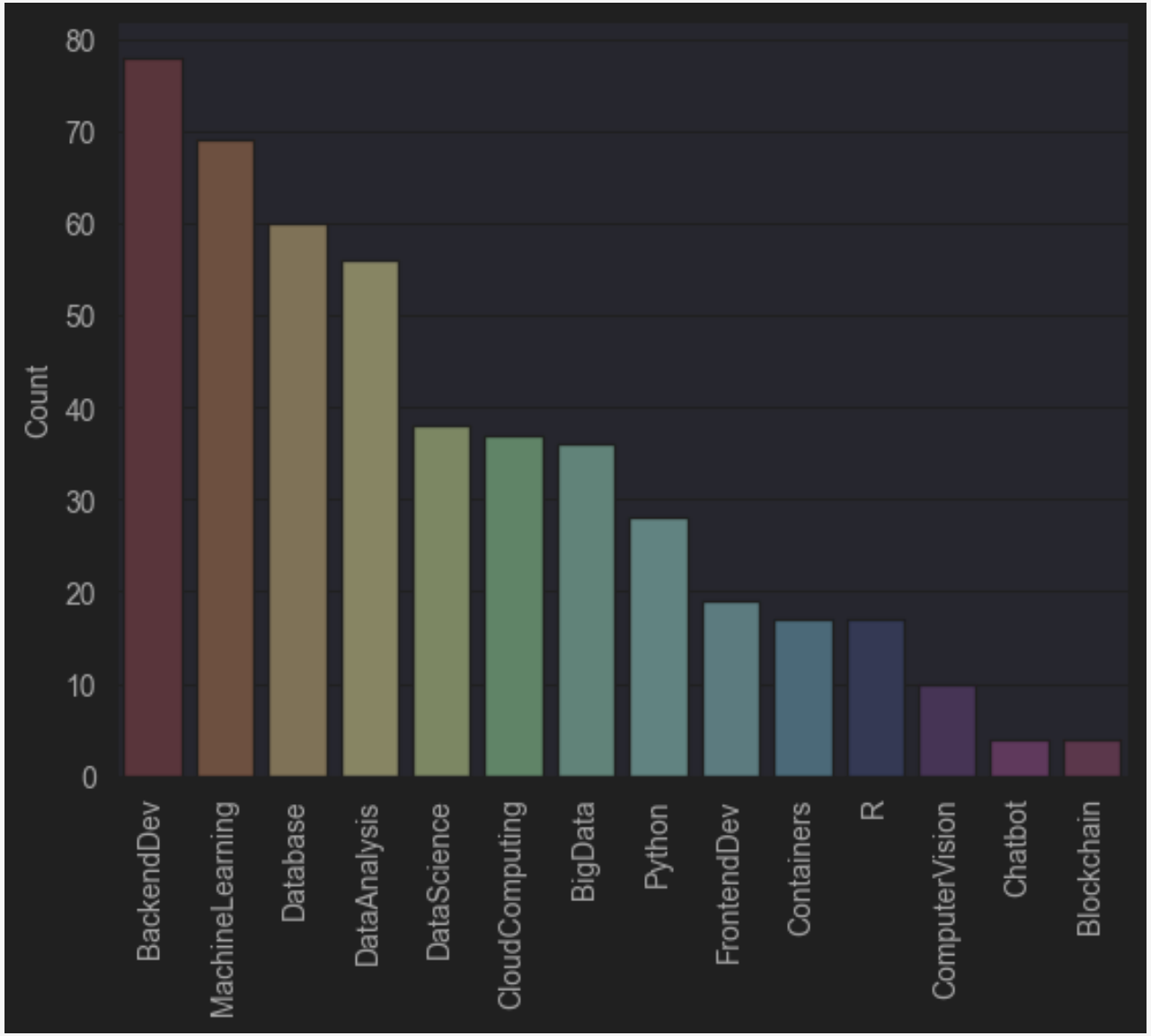
Introduction

INTRODUCTION TO RECOMMENDATION SYSTEMS, EXPLORING USER PREFERENCES

IDENTIFYING PATTERNS IN USER BEHAVIOR FOR ACCURATE PREDICTIONS

Exploratory Data Analysis

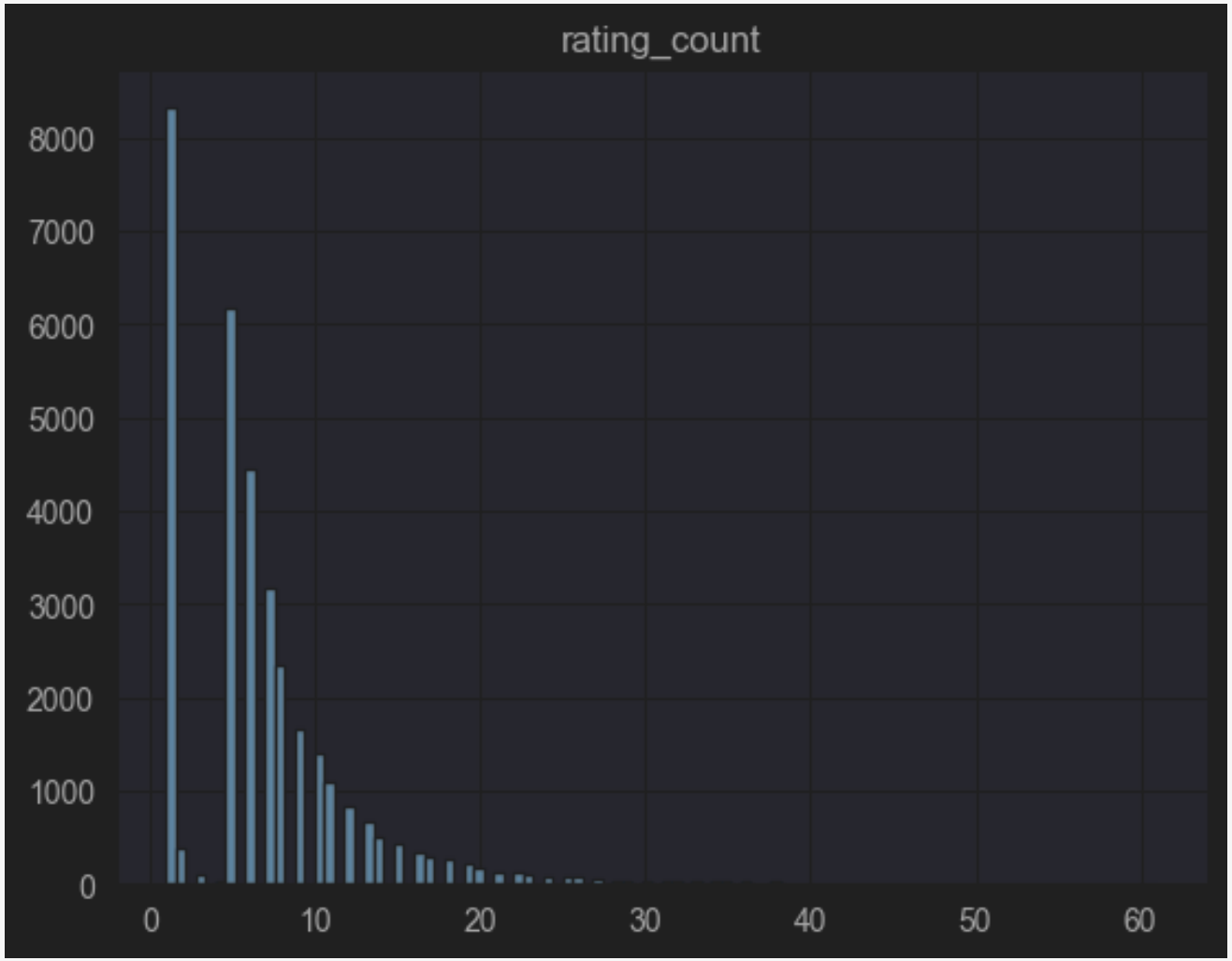




Course counts per genre

- In this image you can see a sorted course count per genre.

- The two most frequent genres are BackendDev and MachineLearning



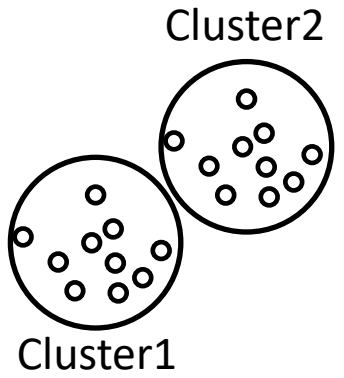
Course enrollment distribution

	TITLE	rating_count
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

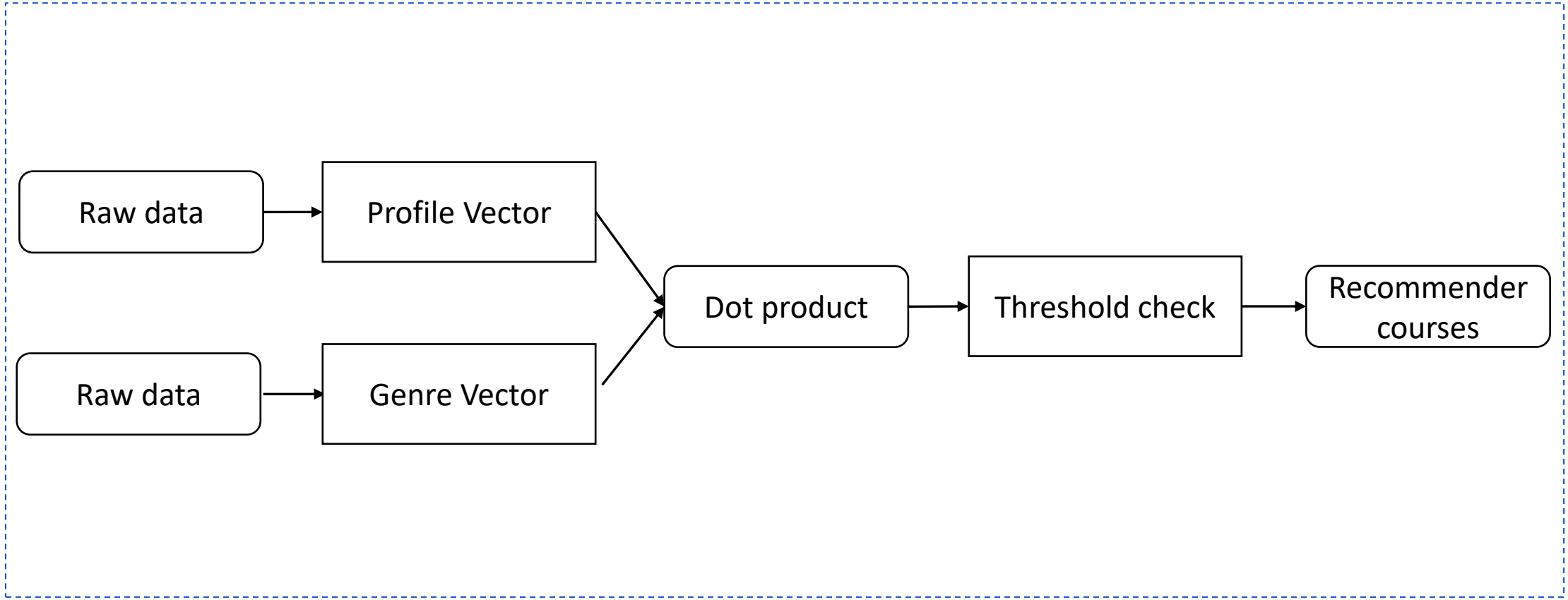
20 most popular courses

Word cloud of course titles

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 = 10.0

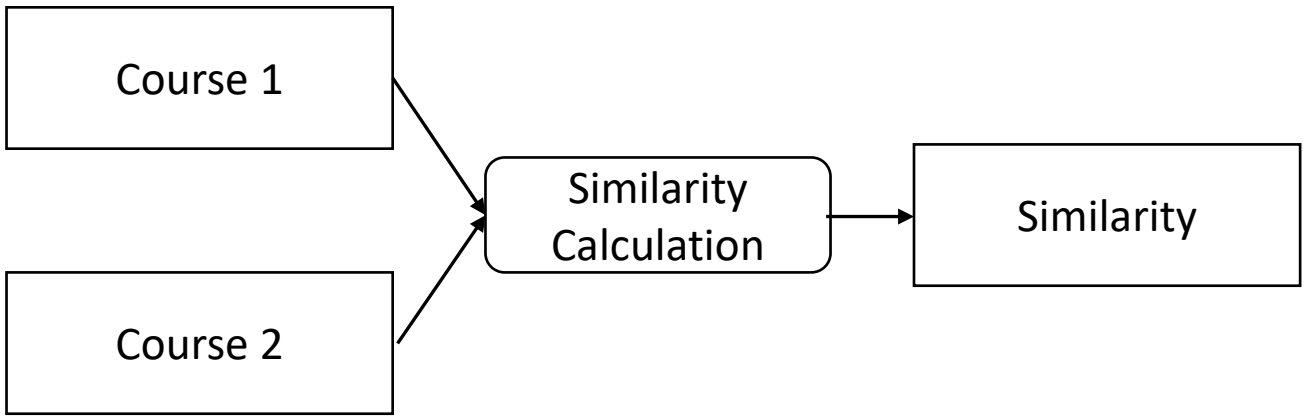
On average, how many new/unseen courses have been recommended per user (in the test user dataset)

61.81

What are the 10 most frequently recommended courses?

COURSE_ID
TA0106EN
GPXX0IBEN
excourse22
excourse21
ML0122EN
excourse06
excourse04
GPXX0TY1EN
excourse31
excourse73

Flowchart of content-based recommender system using course similarity



Evaluation results of course similarity based recommender system

Threshold = 0.6

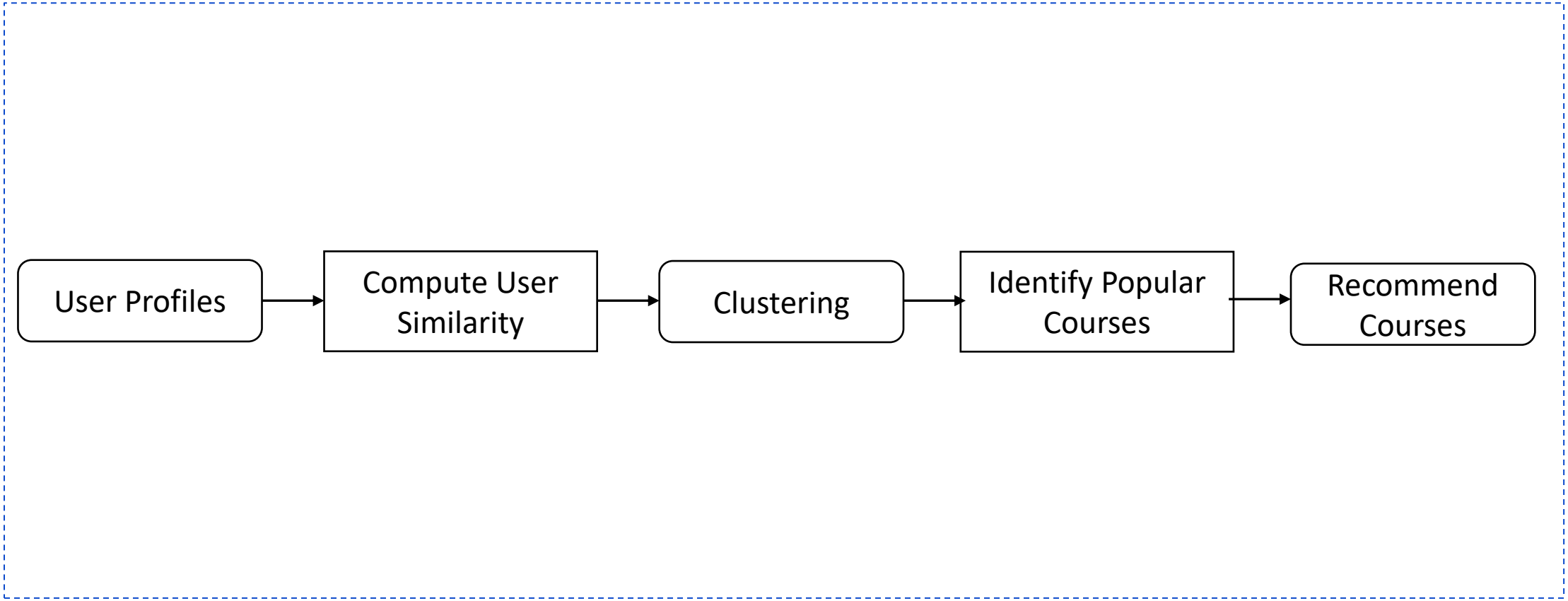
On average, how many new/unseen courses have been recommended per user (in the test user dataset)

1.81

What are the 10 most frequently recommended courses?

COURSE_ID
excourse62
excourse22
WA0103EN
TA0105
DS0110EN
excourse47
excourse46
excourse65
excourse63
TMP0101EN

Flowchart of clustering-based recommender system



Evaluation results of clustering-based recommender system

Number of Clusters: 18

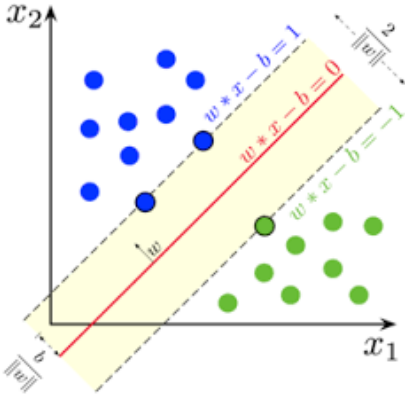
On average, how many new/unseen courses have been recommended per user (in the test user dataset)

1.29

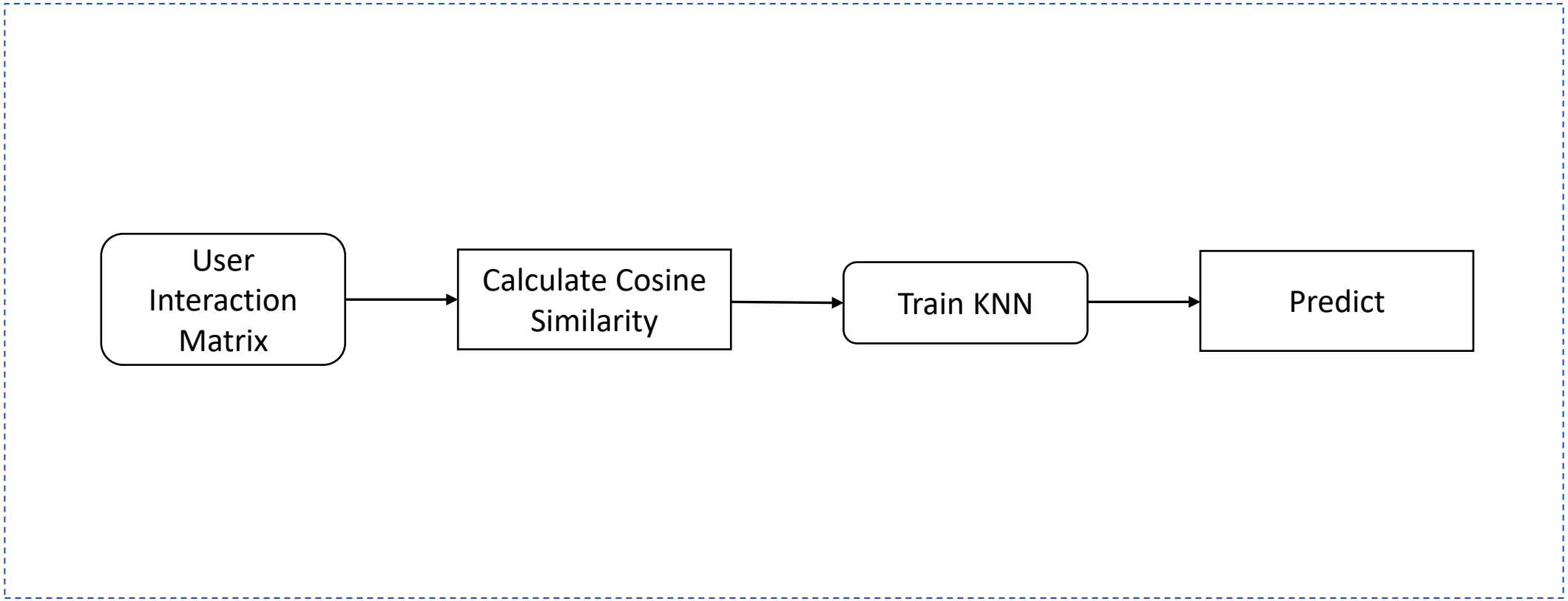
What are the most frequently recommended courses?

item	123
BD0211EN	29
BD0101EN	22
BD0111EN	15

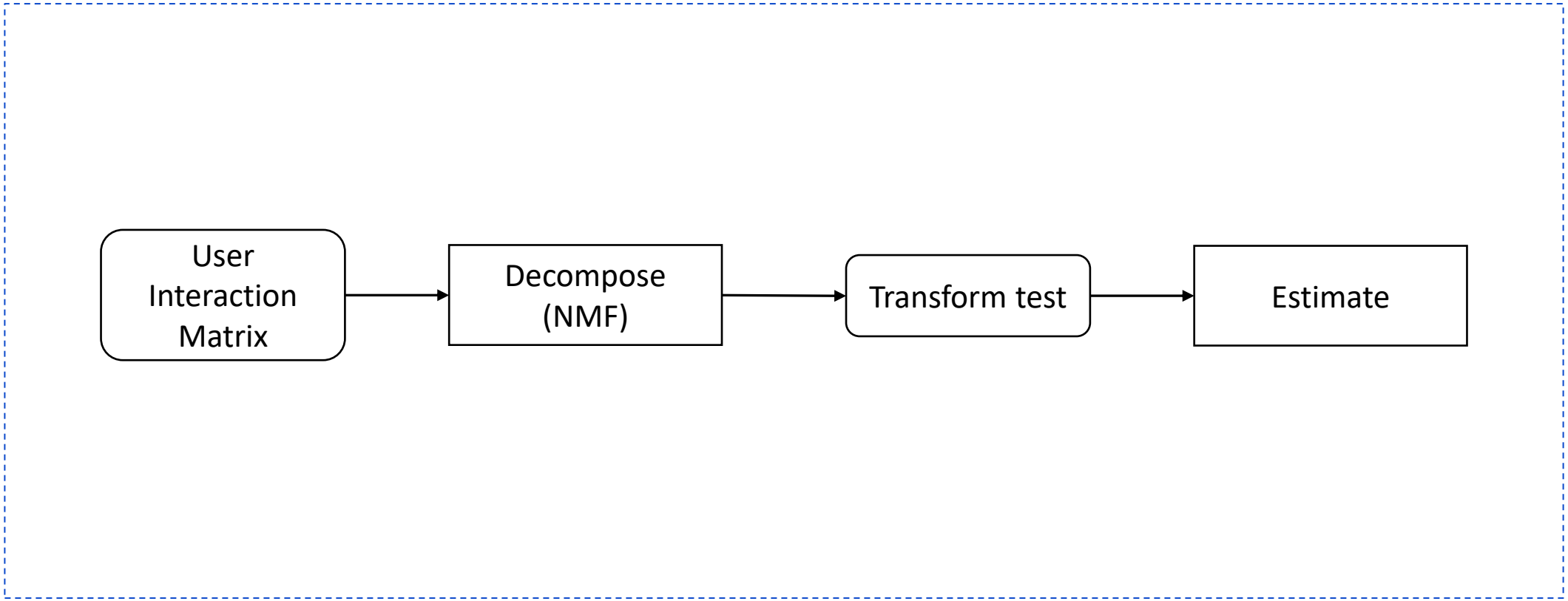
Collaborative-filtering Recommender System using Supervised Learning



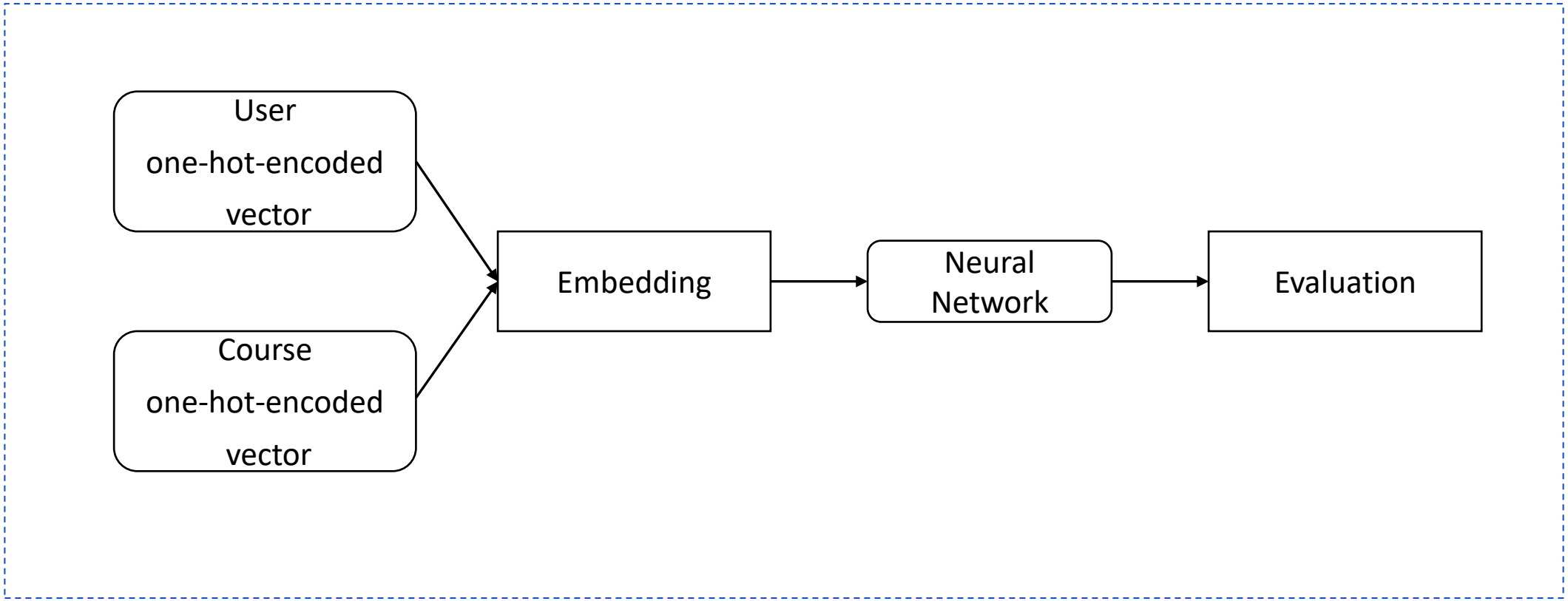
Flowchart of KNN based recommender system



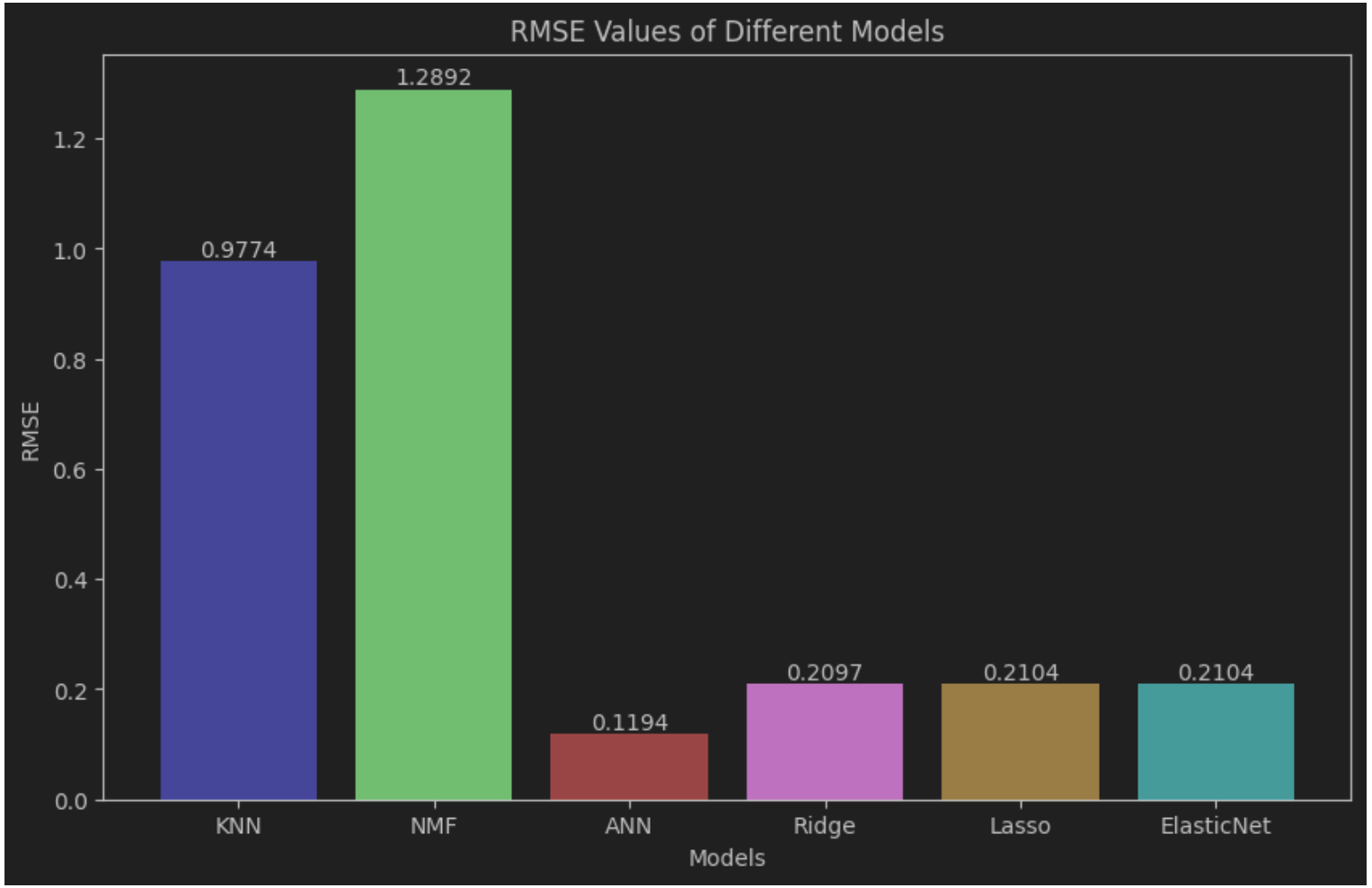
Flowchart of NMF based recommender system



Flowchart of Neural Network Embedding based recommender system

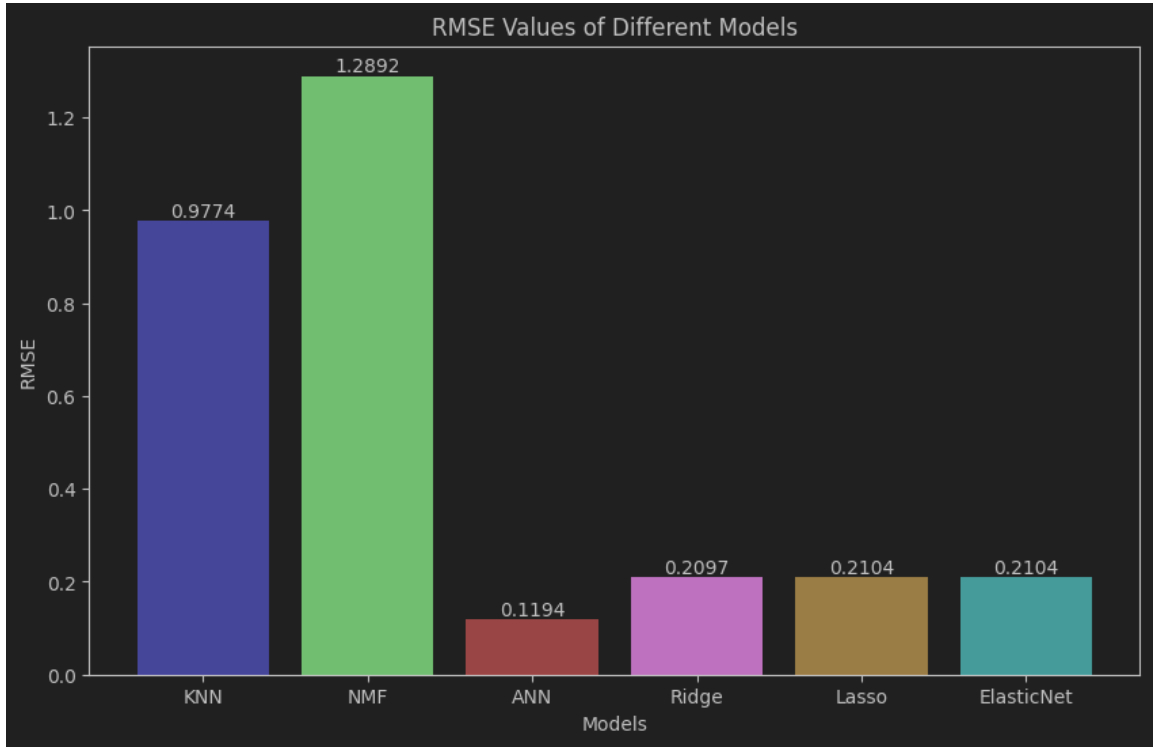


Compare the performance of collaborative-filtering models



Conclusions

- The ANN model has the lowest RMSE value, indicating that it has the best performance among all the models.
- Viceversa, NMF has the worst performance.
- The Ridge, Lasso, and ElasticNet models have similar performance.



Appendix

- https://www.coursera.org/professional-certificates/ibm-machine-learning