Final Assignment

Build a Personalized Online Course Recommender System with Machine Learning

R.G.B.Dhananjaya

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

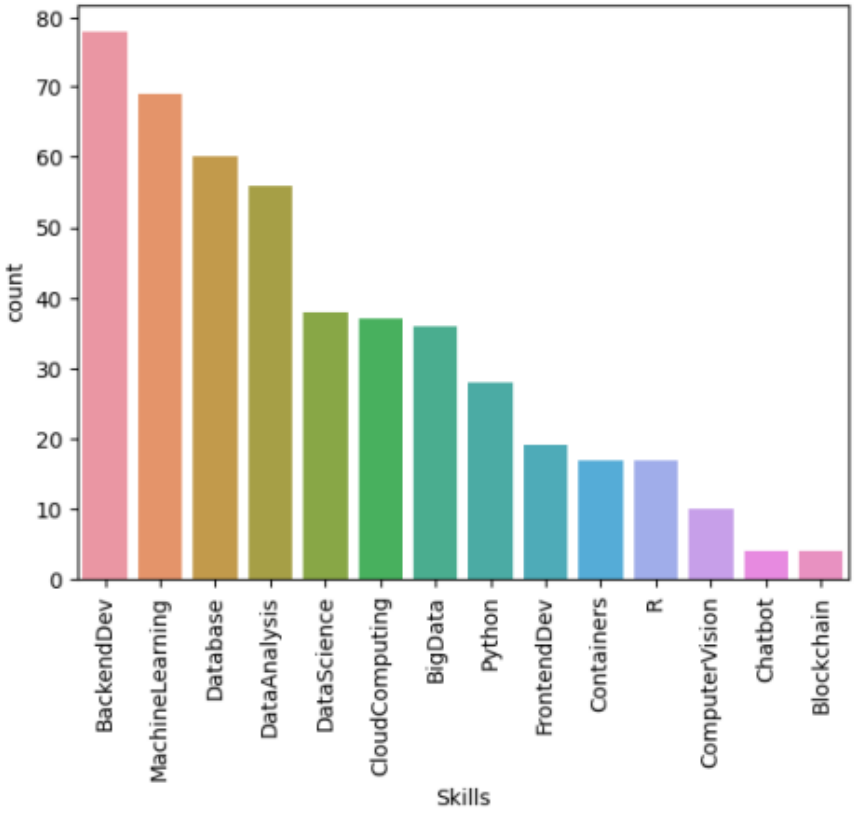
- The AI Training Room is a global platform where learners can explore a wide range of technologies, including Machine Learning, AI, Data Science, Cloud, and App development. As the platform experiences rapid growth, with an expanding array of courses and an increasing number of learners, there is a need for an effective recommender system. The goal of this project is to develop a recommender system that assists learners in discovering relevant courses aligned with their interests, facilitating a personalized and effective learning journey.

Problem states and hypotheses

- The central challenge addressed in this project is the growing complexity for learners to identify courses aligning with their interests and establish a personalized learning trajectory amid the expanding volume of courses and learners. Our hypothesis posits that the development of a personalized recommender system, leveraging both course content and learners' historical interactions with courses, can enhance the discovery of courses matching individual interests and streamline the learning path. To achieve this, we aim to investigate and compare the efficacy of different unsupervised and supervised machine learning models, seeking to identify the most effective model for this specific task.

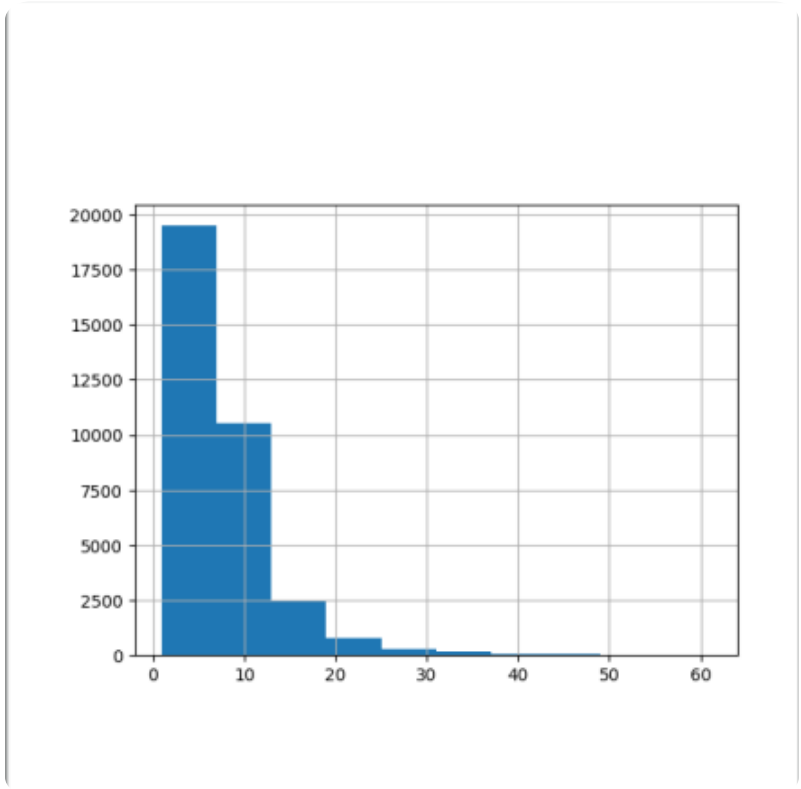
Exploratory Data Analysis

- Course counts per genre



Course enrollment distribution

- We have 233306 enrollments
- The histogram shows the enrolment distributions, e.g., how many users rated just 1 item or how many rated 10 items, etc.



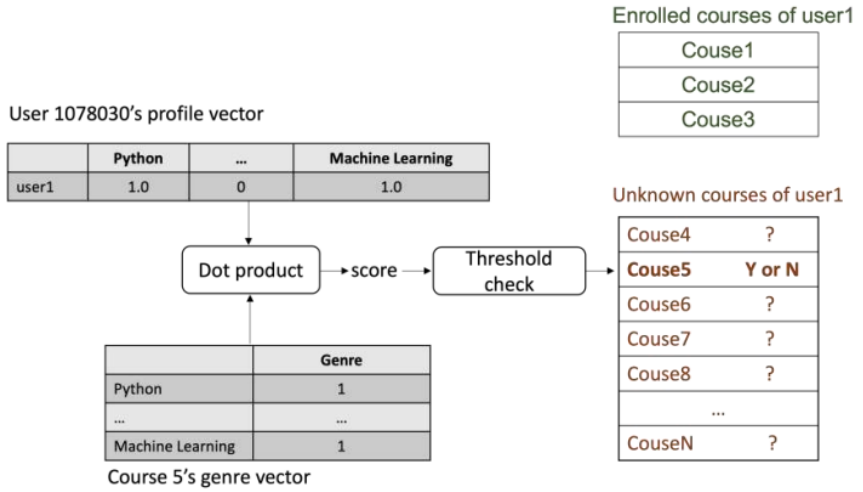
20 most popular courses

- Used Pandas groupby() and size() methods on the item column to aggregate the rating count for each item, then use the sort_values() method to sort the course enrollment count, and use the slice method to get the top 20 courses.

ID	ENROLLS	TITLE
DS0301EN	3624	data privacy fundamentals
BD0115EN	3670	mapreduce and yarn
DB0101EN	3697	sql and relational databases 101
CO0101EN	4480	docker essentials a developer introduction
CC0101EN	4983	introduction to cloud
ST0101EN	5015	statistics 101
RP0101EN	5237	r for data science
CB0103EN	5512	build your own chatbot
ML0115EN	6323	deep learning 101
DV0101EN	6709	data visualization with python
BC0101EN	6719	blockchain essentials
DS0105EN	7199	data science hands on with open source tools
BD0211EN	7551	spark fundamentals i
ML0101ENv3	7644	machine learning with python
DS0103EN	7719	data science methodology
DA0101EN	8303	data analysis with python
BD0111EN	10599	hadoop 101
BD0101EN	13291	big data 101
DS0101EN	14477	introduction to data science
PY0101EN	14936	python for data science

Content-based Recommender System using Unsupervised Learning

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



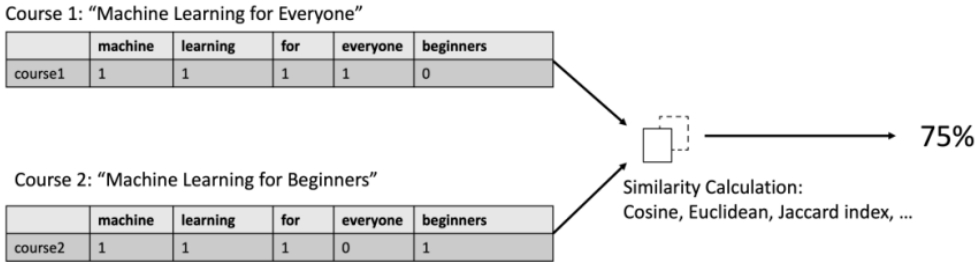
Evaluation results of user profile-based recommender system

Top-10 commonly recommended courses across all users

Course	Times recommended
TA0106EN	608
GPXX0IBEN	548
excourse22	547
excourse21	547
ML0122EN	544
excourse04	533
GPXX0TY1EN	533
excourse06	533
excourse31	524
excourse73	516

- # The threshold can be fine-tuned to adjust the size of generated recommendations score_threshold = 10.0

Flowchart of content-based recommender system using course similarity



Evaluation results of clustering-based recommender system

```
user in cluster 0 will be sugessted 3 courses as ['PY0101EN' 'CB0103EN' 'DA0101EN']
user in cluster 1 will be sugessted 3 courses as ['DS0101EN' 'BD0101EN' 'PY0101EN']
user in cluster 2 will be sugessted 3 courses as ['CO0301EN' 'CO0201EN' 'BC0101EN']
user in cluster 3 will be sugessted 3 courses as ['PY0101EN' 'ML0101ENv3' 'ML0115EN']
user in cluster 4 will be sugessted 3 courses as ['BD0111EN' 'BD0141EN' 'BD0115EN']
user in cluster 5 will be sugessted 3 courses as ['CB0103EN' 'DS0101EN' 'BD0101EN']
user in cluster 6 will be sugessted 3 courses as ['CO0101EN' 'CO0201EN' 'CO0301EN']
user in cluster 7 will be sugessted 3 courses as []
user in cluster 8 will be sugessted 3 courses as ['CO0101EN' 'PY0101EN' 'CC0101EN']
user in cluster 9 will be sugessted 3 courses as ['DS0101EN' 'RP0101EN' 'DS0103EN']
user in cluster 10 will be sugessted 3 courses as ['CO0101EN' 'LB0101ENv1' 'CO0401EN']
user in cluster 11 will be sugessted 3 courses as ['RP0101EN' 'DS0101EN' 'DS0103EN']
```

Collaborative-filtering Recommender System using Supervised Learning

Flowchart of KNN based recommend system

User-Item interaction matrix

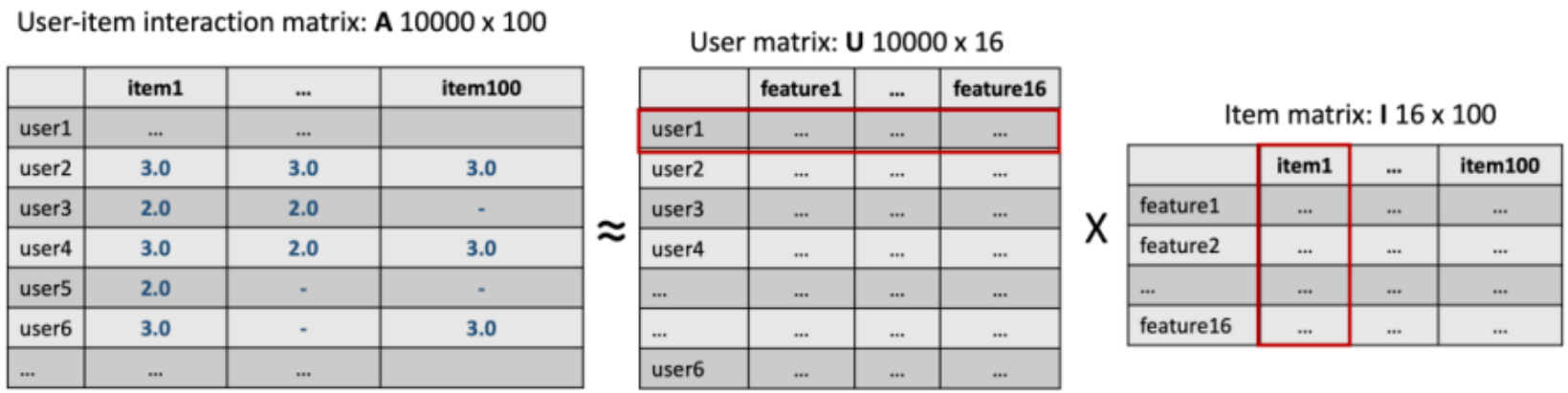
	Machine Learning With Python	Machine Learning 101	Machine Learning Capstone	SQL with Python	Python 101
...	...	...	...	...	
user2	3.0	3.0	3.0	3.0	3.0
user3	2.0	3.0	3.0	2.0	
user4	3.0	3.0	2.0	2.0	3.0
user5	2.0	3.0	3.0		
user6	3.0	3.0	?		3.0
...	...	...	...	...	

Similar users

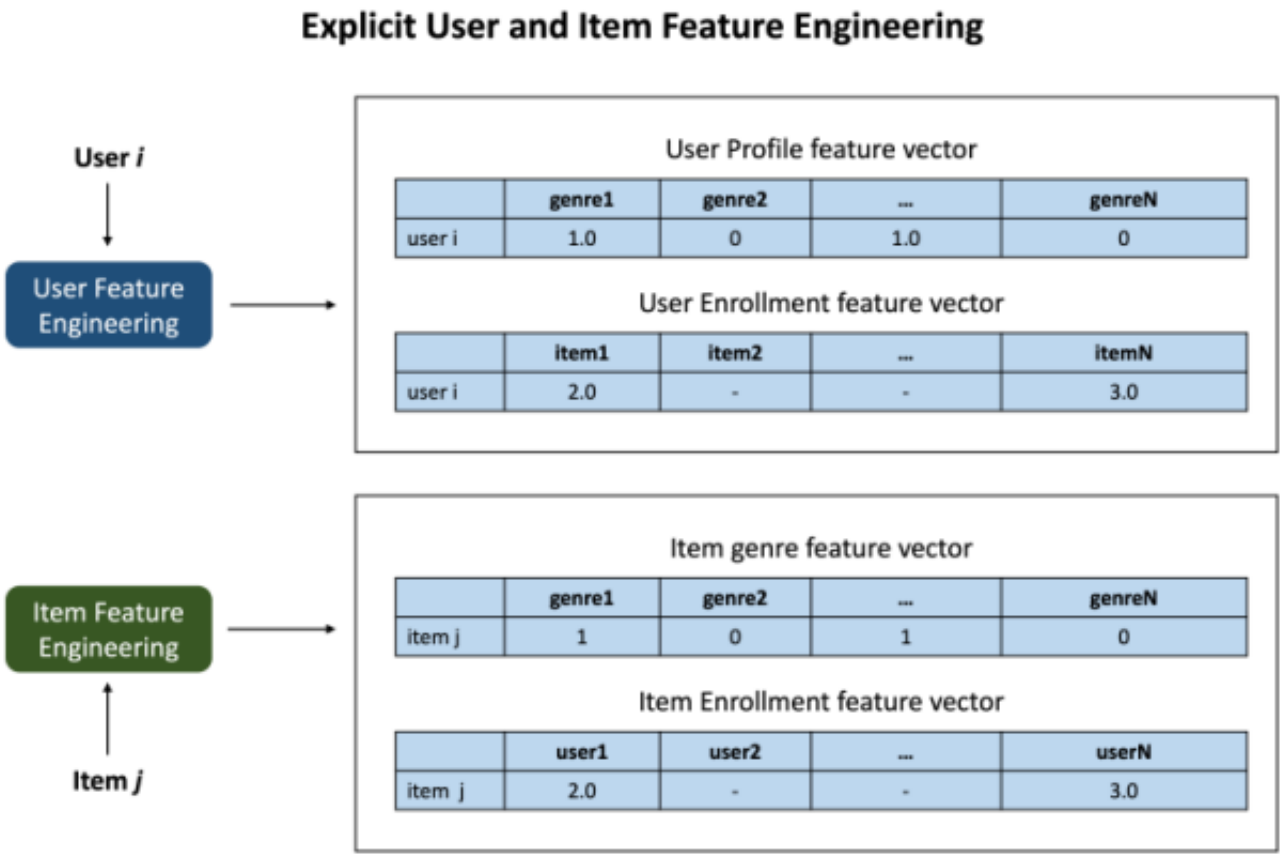
Predict the rating of user *user6* to item *Machine Learning Capstone*

Flowchart of NMF based recommender system

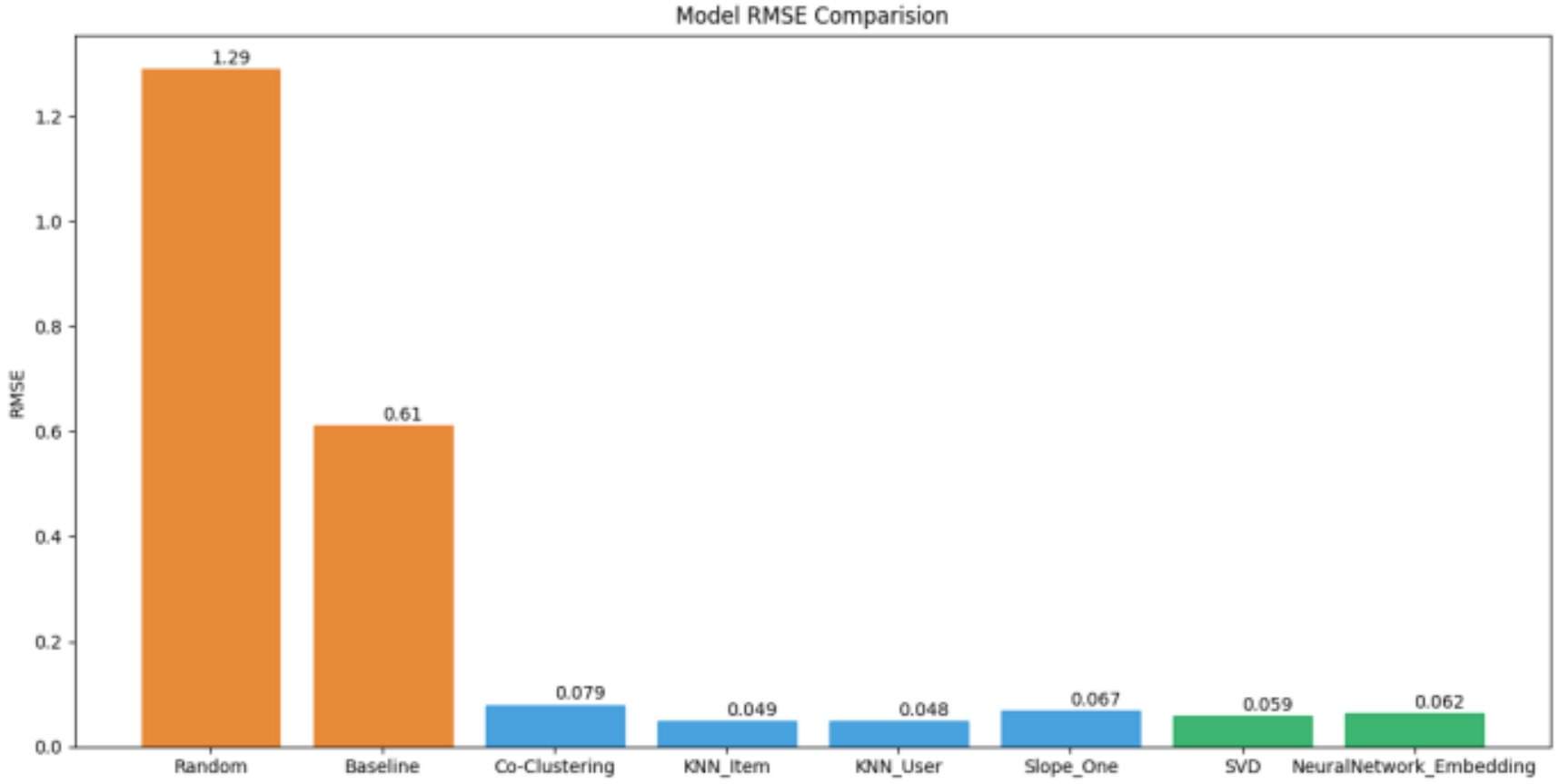
Non-negative Matrix Factorization



Flowchart of Neural Network Embedding based recommender system



Visualization of the performance metric (such as RMSE) of different collaborative-filtering models built so far



Thank You