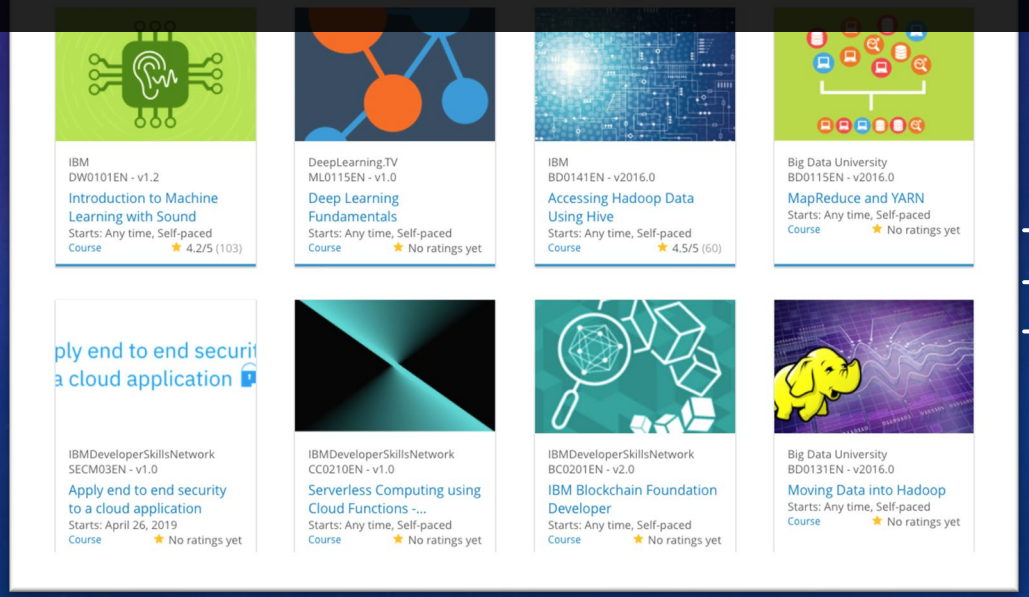
Build a Personalized Online Course Recommender System with Machine Learning

William Ruffu
11/14/2024



Outline

- Introduction and Background
- Exploratory Data Analysis
- Recommender Systems:
  - Content-based - Unsupervised Learning
  - Collaborative-filtering based - Supervised learning
- Conclusion
- Appendix

Introduction

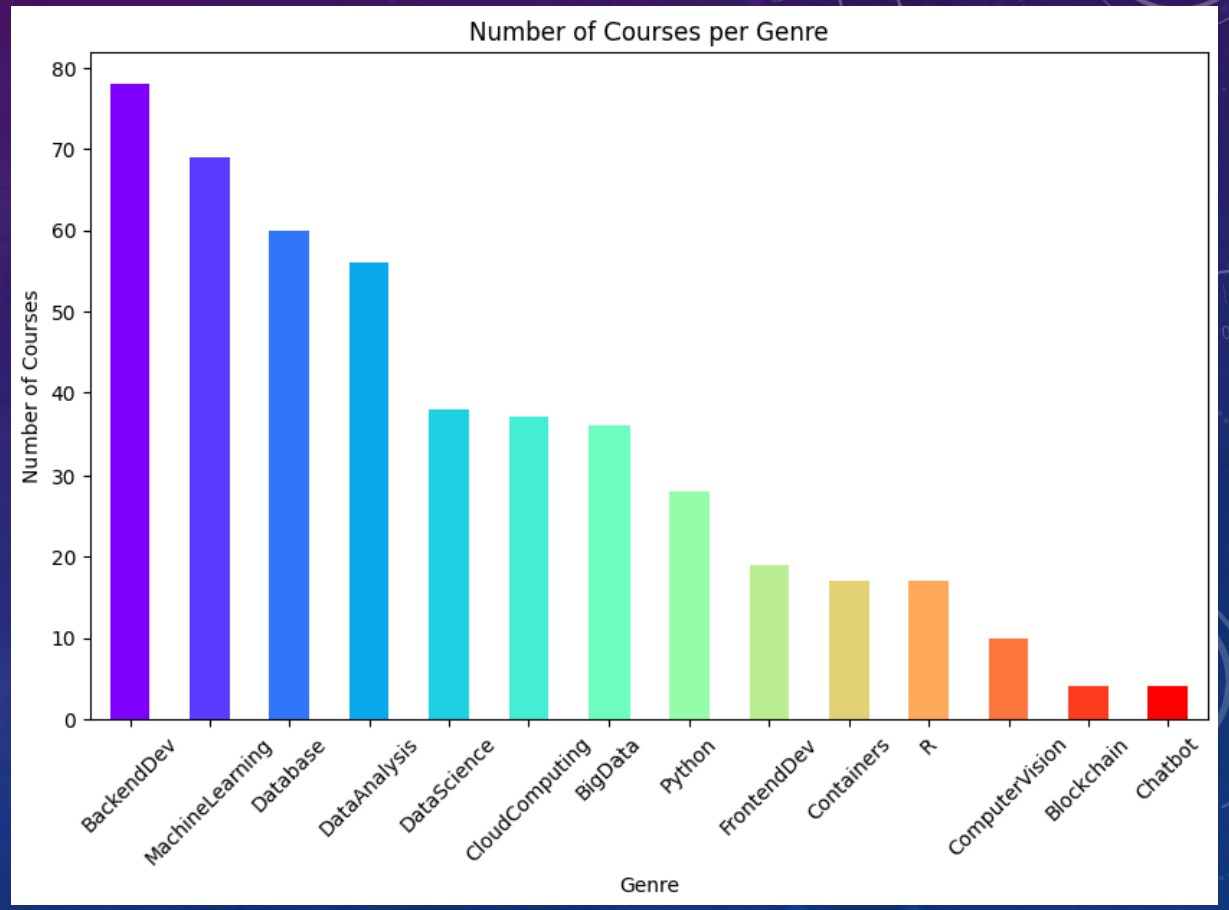
- Project Context: Build a ML based course recommender system to improve user satisfaction and retention for online learning platform.

- Goal: Increase Customer Lifespan
- Hypothesis: Personalized recommendations will enhance user engagement & business value.
- Approach: Use content-based collaborative filtering & clustering methods with neural networks to develop recommender system(s).

EXPLORATORY DATA ANALYSIS

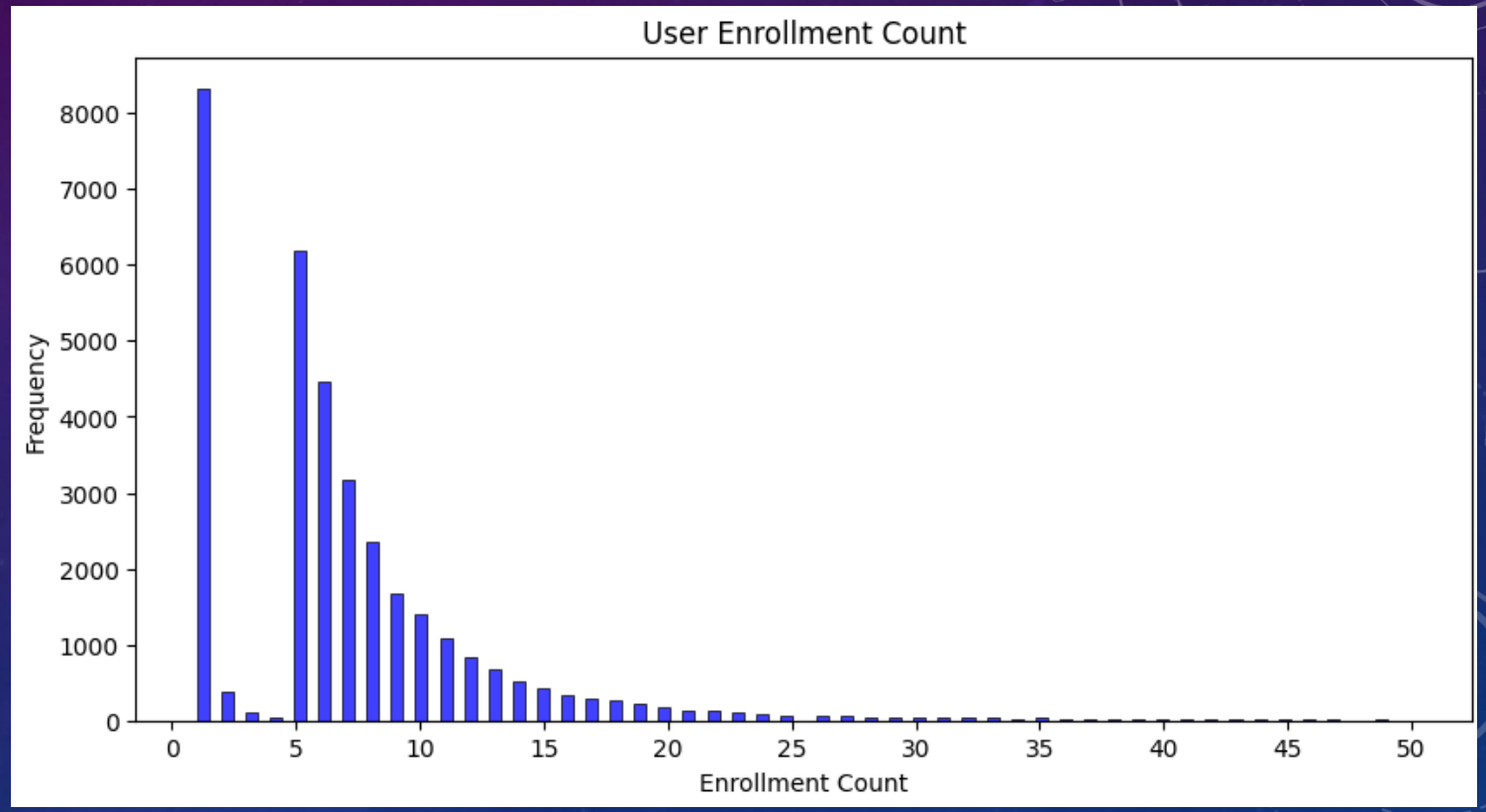
COURSE COUNTS PER GENRE

- Distribution of courses across genres
- Determined Genres Features Using the Course Catalog
- Used to Group Subclasses of Customers for Recommendation



COURSE ENROLLMENT DISTRIBUTION

- As the histogram shows, many customers only enroll in one course.
- An engaging recommendation can drive significant business value!



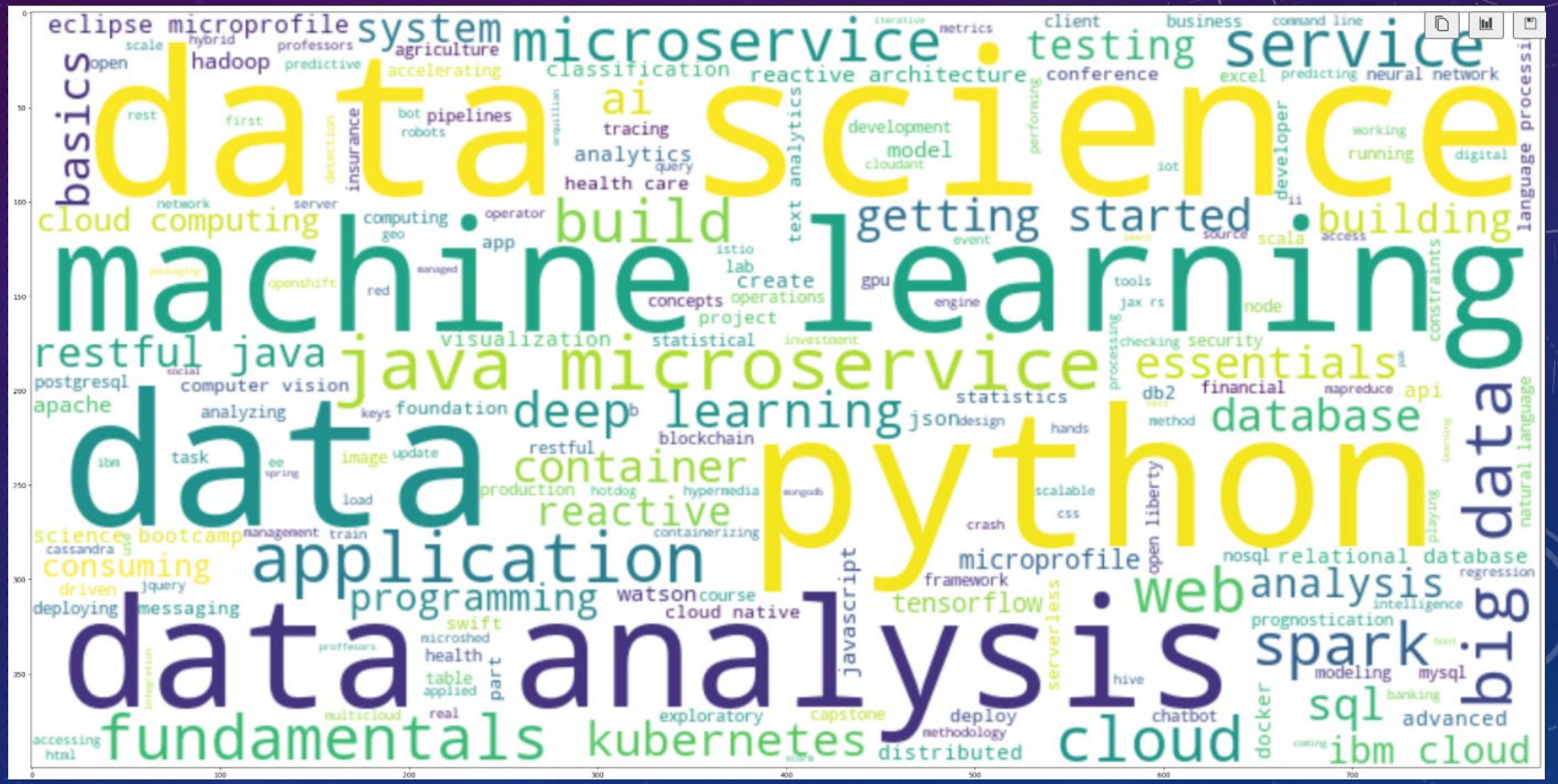
20 MOST POPULAR COURSE

- List the most popular 20 courses
- Python and Data Science are very popular courses!

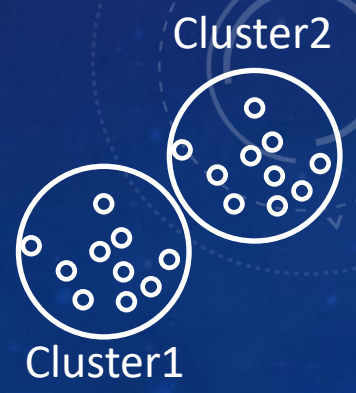
	TITLE	Enrollment
83	python for data science	14936
64	introduction to data science	14477
9	big data 101	13291
56	hadoop 101	10599
69	machine learning with python	9394
27	data analysis with python	8303
36	data science methodology	7719
99	spark fundamentals i	7551
35	data science hands on with open source tools	7199
11	blockchain essentials	6719
39	data visualization with python	6709
45	deep learning 101	6323
15	build your own chatbot	5512
85	r for data science	5237
105	statistics 101	5015
62	introduction to cloud	4983
49	docker essentials a developer introduction	4480
46	deep learning with tensorflow	3914
104	sql and relational databases 101	3697
71	mapreduce and yarn	3670

WORD CLOUD OF COURSE TITLES

- WordCloud Visualization of Our Course Offerings



CONTENT-BASED RECOMMENDER SYSTEM USING UNSUPERVISED LEARNING



FLOWCHART OF CONTENT-BASED RECOMMENDER SYSTEM USING USER PROFILE AND COURSE GENRES

- User profiles matched with course genres
- Using Provided csv Data: Course Genre, User Profile, Ratings

Sample Process:

EVALUATION RESULTS OF USER PROFILE-BASED RECOMMENDER SYSTEM

Recommendation Score Minimum Set to 20

Average of 14 new courses recommended per user based on the currently enrolled course.

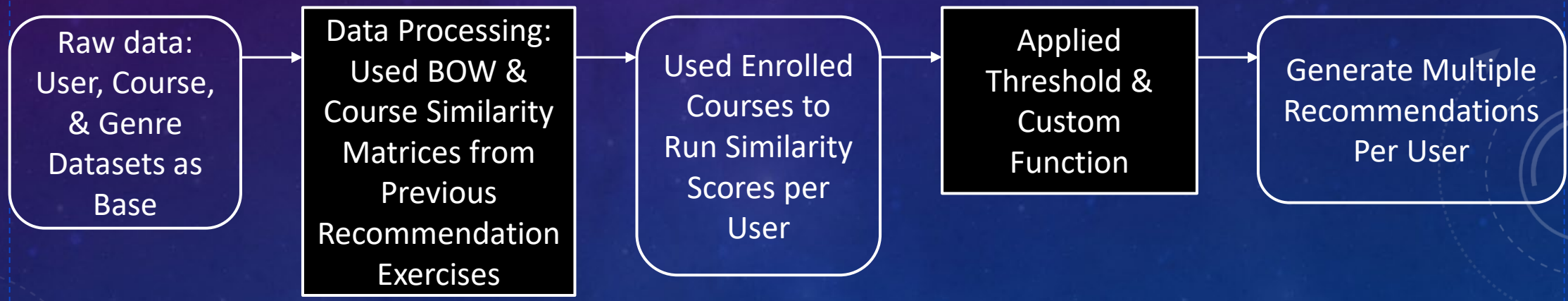
Top-10 commonly recommended courses across all users:

	COURSE_ID	count	TITLE
0	excourse72	9138	foundations for big data analysis with sql
1	excourse73	9138	analyzing big data with sql
2	TMP0105EN	8954	getting started with the data apache spark ma...
3	RP0105EN	8769	analyzing big data in r using apache spark
4	SC0103EN	7970	spark overview for scala analytics
5	excourse31	7853	cloud computing applications part 2 big data...
6	excourse21	7671	applied machine learning in python
7	excourse22	7671	introduction to data science in python
8	ML0122EN	7633	accelerating deep learning with gpu
9	BD0212EN	7203	spark fundamentals ii

FLOWCHART OF CONTENT-BASED RECOMMENDER SYSTEM USING COURSE SIMILARITY

- From looking at similar courses from the users enrolled course we can make different predictions.
- User profiles matched with course genres.

Example Process:



EVALUATION RESULTS OF COURSE SIMILARITY BASED RECOMMENDER SYSTEM

Similarity Threshold = .6 or 60% Similarity Score

On average, this method produced 10 new/unseen courses recommendations per user (in the test user dataset).

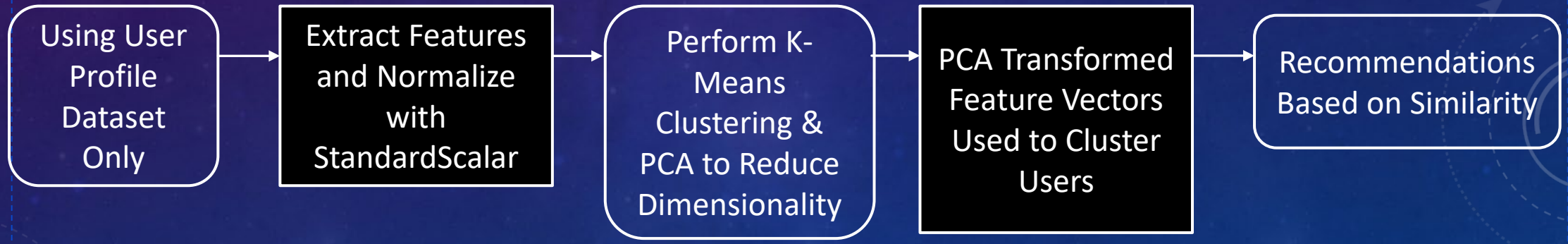
Top-10 commonly recommended courses:

	COURSE_ID	count	TITLE
0	DS0110EN	15003	data science with open data
1	excourse22	14937	introduction to data science in python
2	excourse62	14937	introduction to data science in python
3	excourse63	14641	a crash course in data science
4	excourse65	14641	data science fundamentals for data analysts
5	excourse68	13551	big data modeling and management systems
6	excourse72	13512	foundations for big data analysis with sql
7	excourse74	13291	fundamentals of big data
8	excourse67	13291	introduction to big data
9	BD0145EN	12497	sql access for hadoop

FLOWCHART OF CLUSTERING-BASED RECOMMENDER SYSTEM

- User segments created for targeted recommendations.
- K-Means clustering with PCA for dimensionality reduction.

General Process:



EVALUATION RESULTS OF CLUSTERING-BASED RECOMMENDER SYSTEM

PCA Similarity Threshold of .9

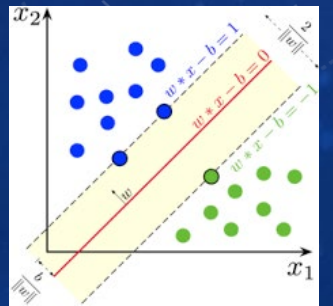
Enrollment Threshold = 1500

Using a enrollment threshold of 1500 in the test_users group, we can recommend 16 courses on average per user.

Top-10 Commonly Recommended Courses

recommended_course	
ST0101EN	25176
CC0101EN	25162
RP0101EN	24762
ML0115EN	24123
BC0101EN	24089
CNSC02EN	23691
DV0101EN	23529
DS0301EN	23512
BD0115EN	23471
DB0101EN	23152

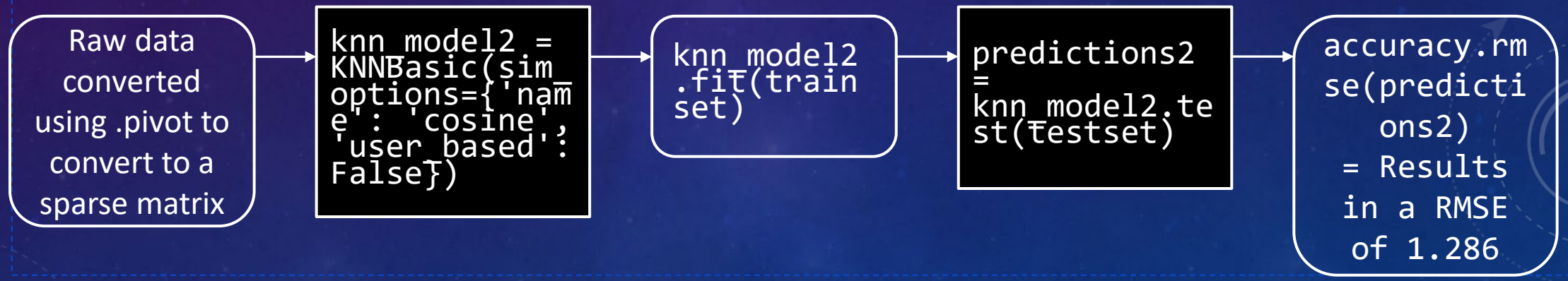
COLLABORATIVE-FILTERING RECOMMENDER SYSTEM USING SUPERVISED LEARNING



FLOWCHART OF KNN BASED RECOMMENDER SYSTEM

- Enrollment History & User Ratings csv Data
- User-Item Interaction Matrix Used
- Scikit-learn Surprise Library

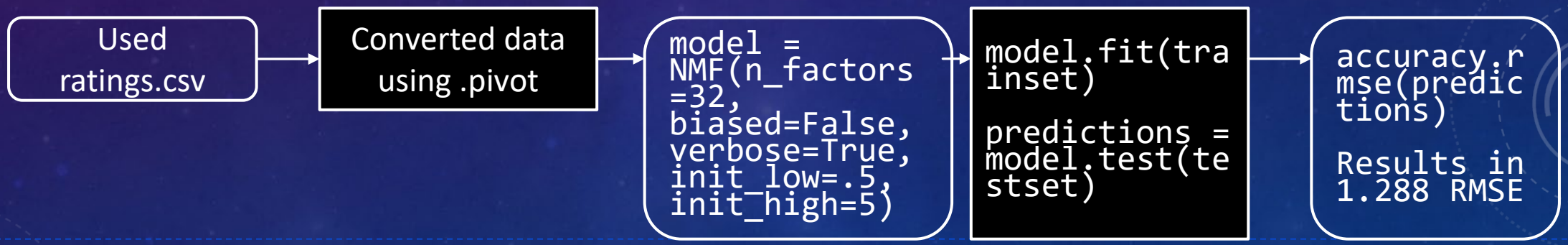
General Process:



FLOWCHART OF NMF BASED RECOMMENDER SYSTEM

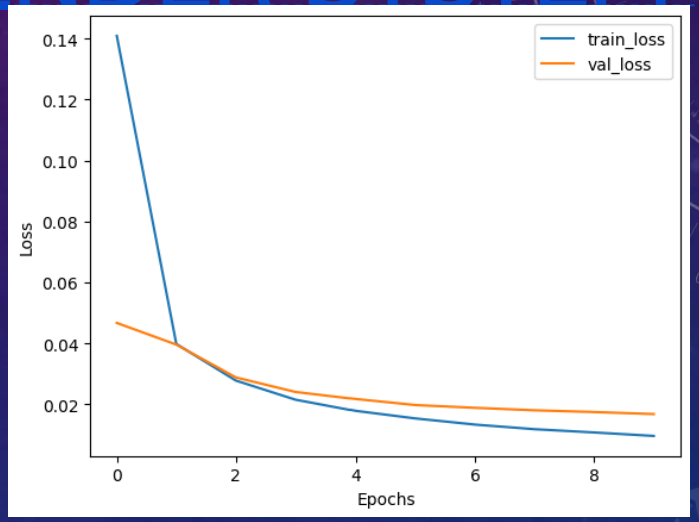
- Non-Negative Matrix Factorization
- Surprise Library

General Process:



FLOWCHART OF NEURAL NETWORK EMBEDDING BASED RECOMMENDER SYSTEM

- TensorFlow Keras Model SuperClass RecommenderNet
- One-Hot Encoded w/ Adam Optimizer = .001
- Rating Dataset

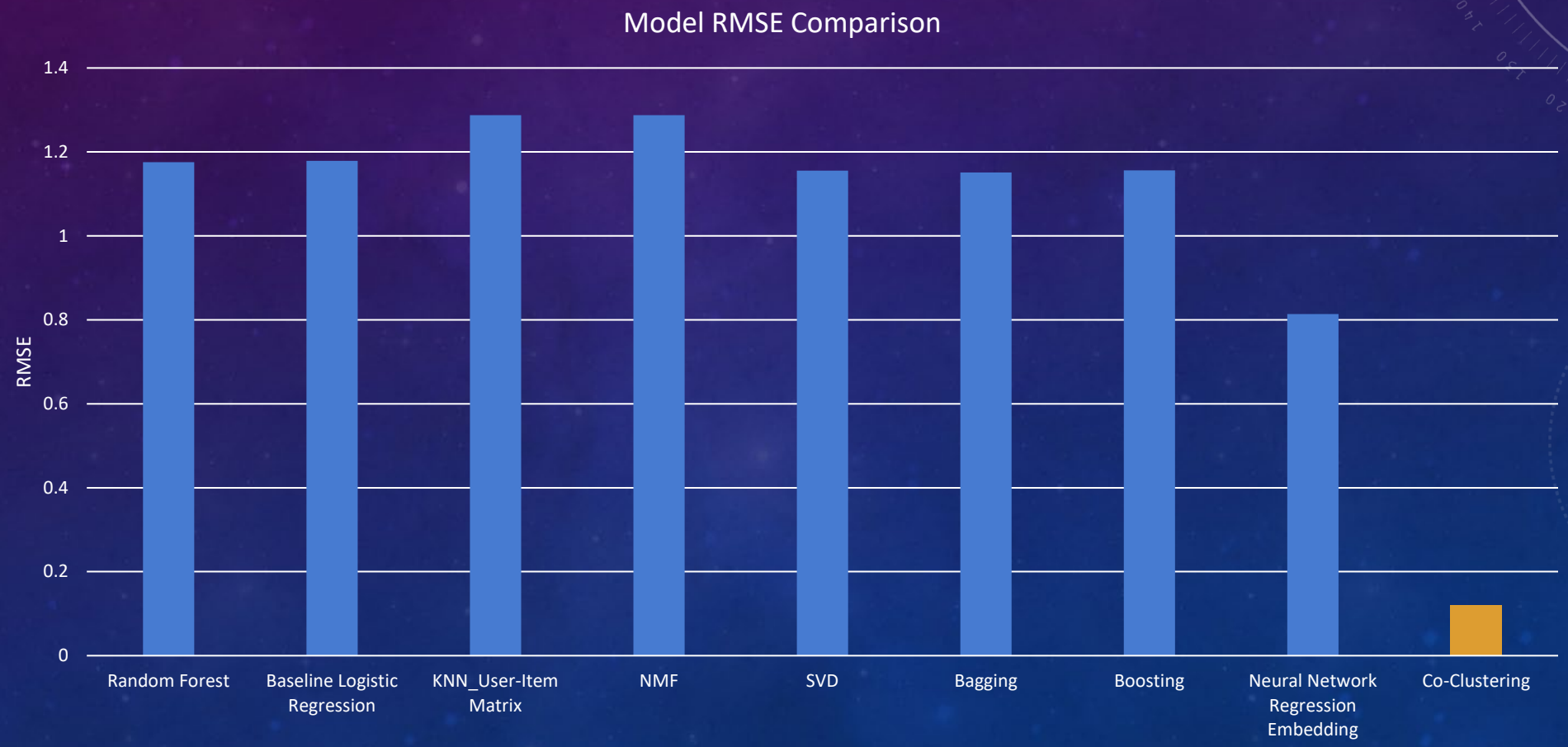


General Process:

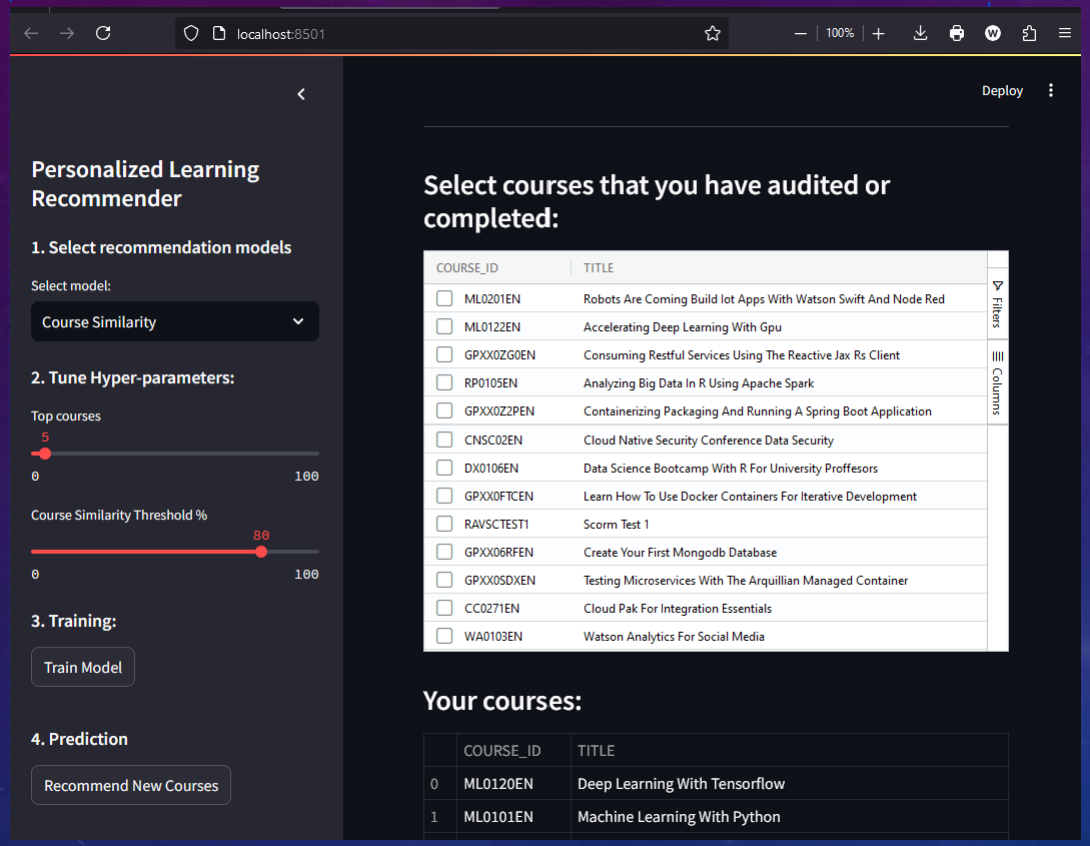
Raw data split into item and user sets. → One-Hot Encoding of each Set → Embedding Layer(s) → Smaller Embedded matrice(s) of 16 x 1 → Rating Result of Dot Product

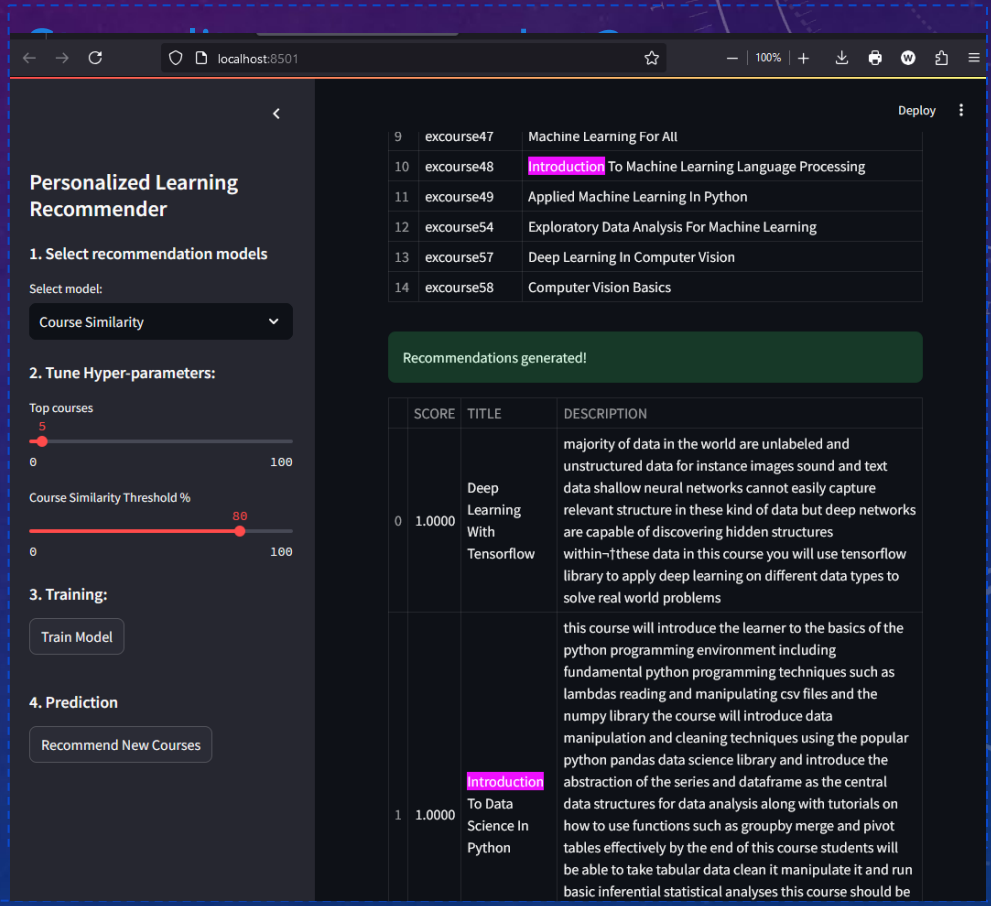
COMPARE THE PERFORMANCE OF COLLABORATIVE-FILTERING MODELS

- Summary Stats of RMSE Performance Metrics Across Collaborative-filtering Models



DEPLOYED STREAMLIT APP





Conclusions

- Main Insights:
  - Machine learning significantly enhances accuracy in recommendations.
  - Neural network models outperform other methods in accuracy and scalability.
  - Clustering-based approaches effectively segment users for targeted suggestions.

- Strategic Value:
  - Improved user engagement drives platform growth and retention.
  - Tailored recommendations increase trust and customer satisfaction.
  - Scalable systems provide adaptability for future datasets and user bases