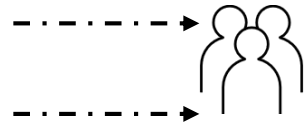
Building a Personalized Anime Recommender System with Machine Learning

vectorkoz

July 2024



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

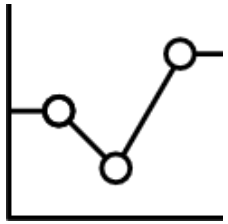
- This project uses the Anime Recommendations Dataset from Kaggle.
- This dataset contains information on user preference data from 73,516 users on 12,294 anime. Due to computational limitations, only the data on 66,974 users and 1000 anime was used.

- All the code from this project is available on my github page.

In this project, various types of recommender systems were constructed:

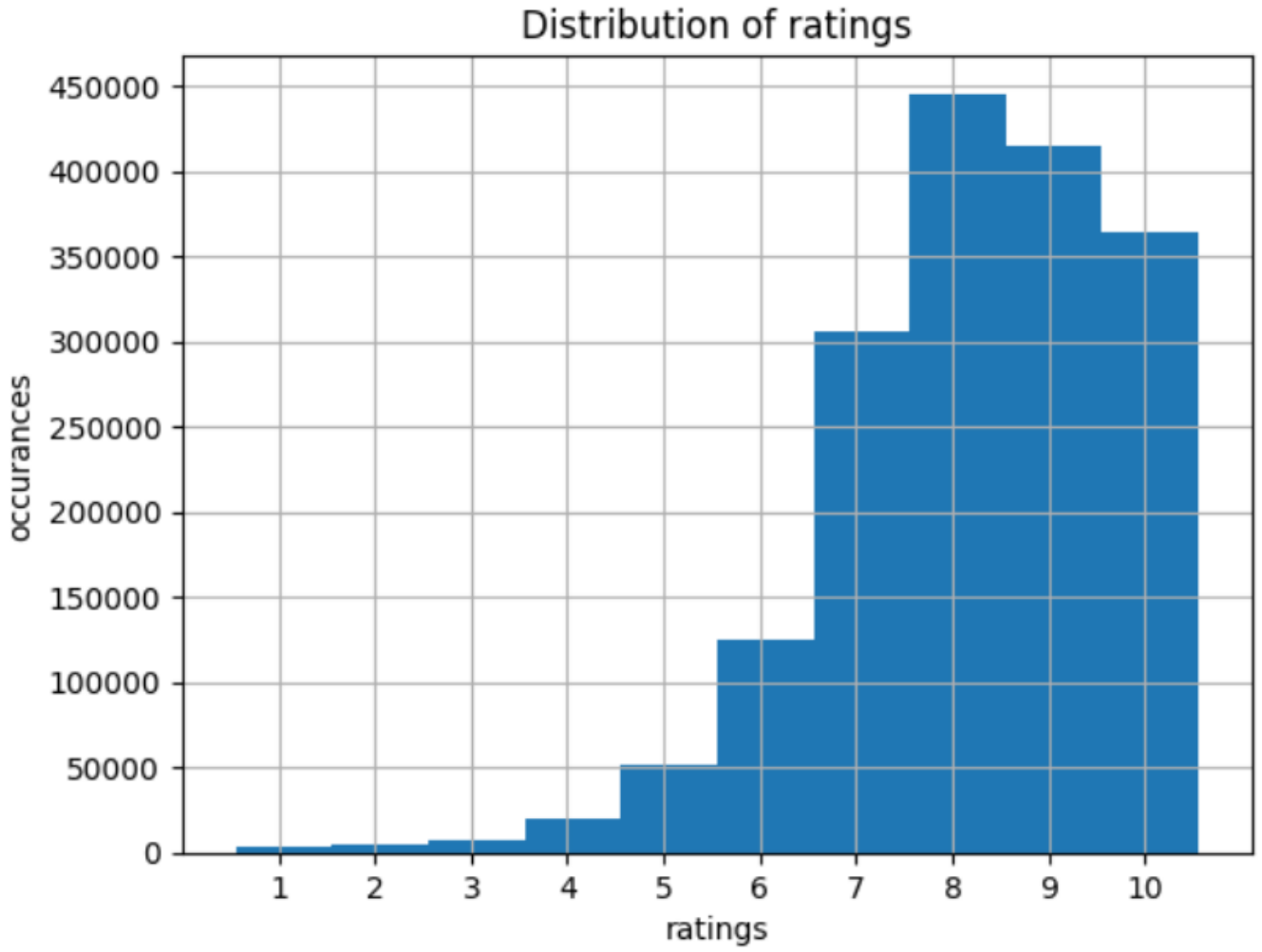
- Content-based using user profiles based on anime genres
- Content-based using computed anime similarity scores
- Content-based using clustering on user profiles
- KNN-based collaborative filtering
- NMF-based collaborative filtering
- Neural network-based collaborative filtering

Exploratory Data Analysis



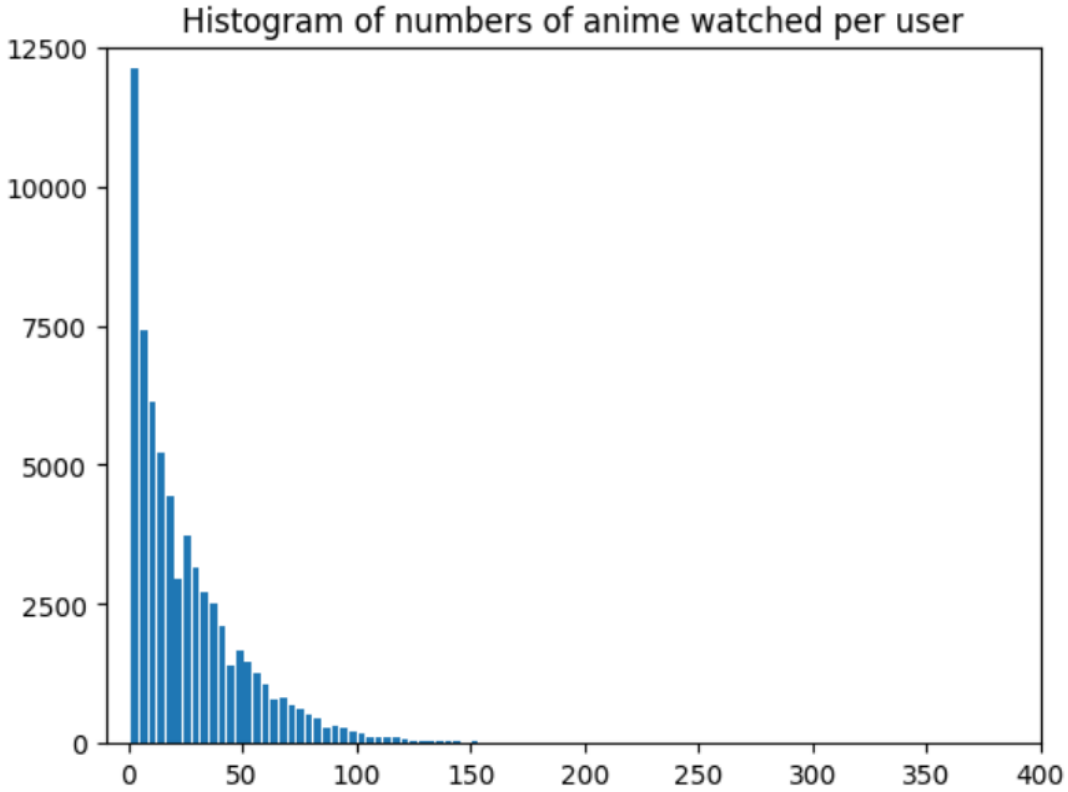
Anime rating distribution

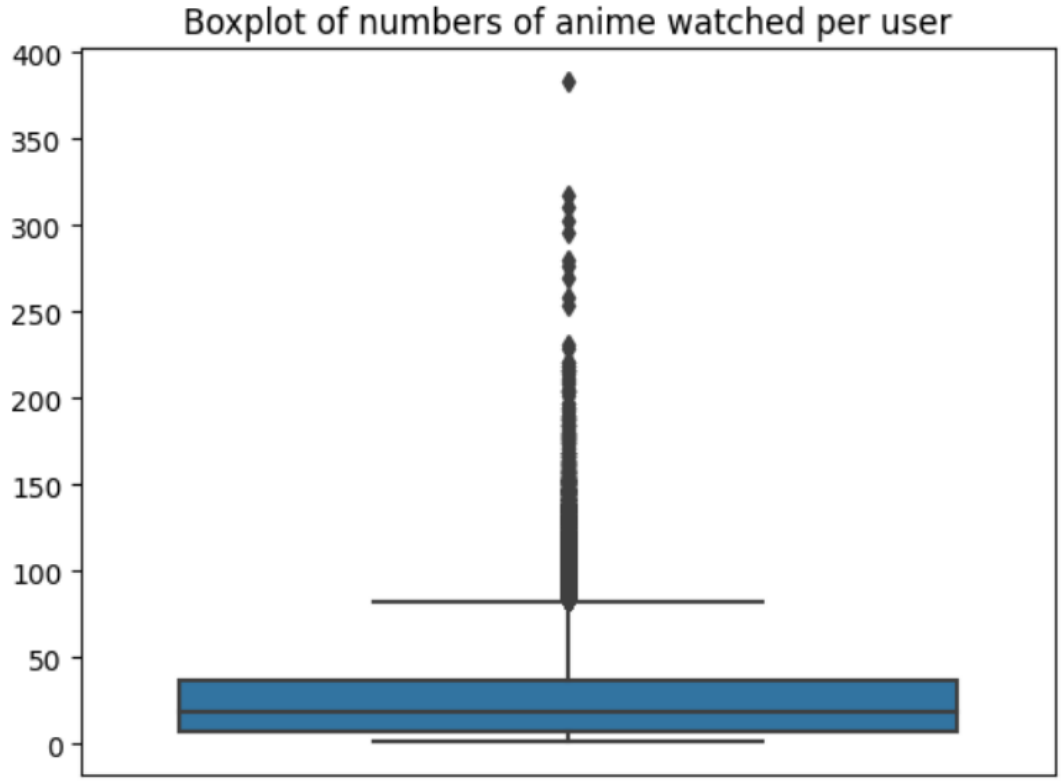
- When users bothered to rate anime, they usually gave a relatively high score
- There is only a small number of really low scores



Anime viewing distribution

- Most users have only watched up to a few dozen anime, however, there are a few outliers who watched more than 100





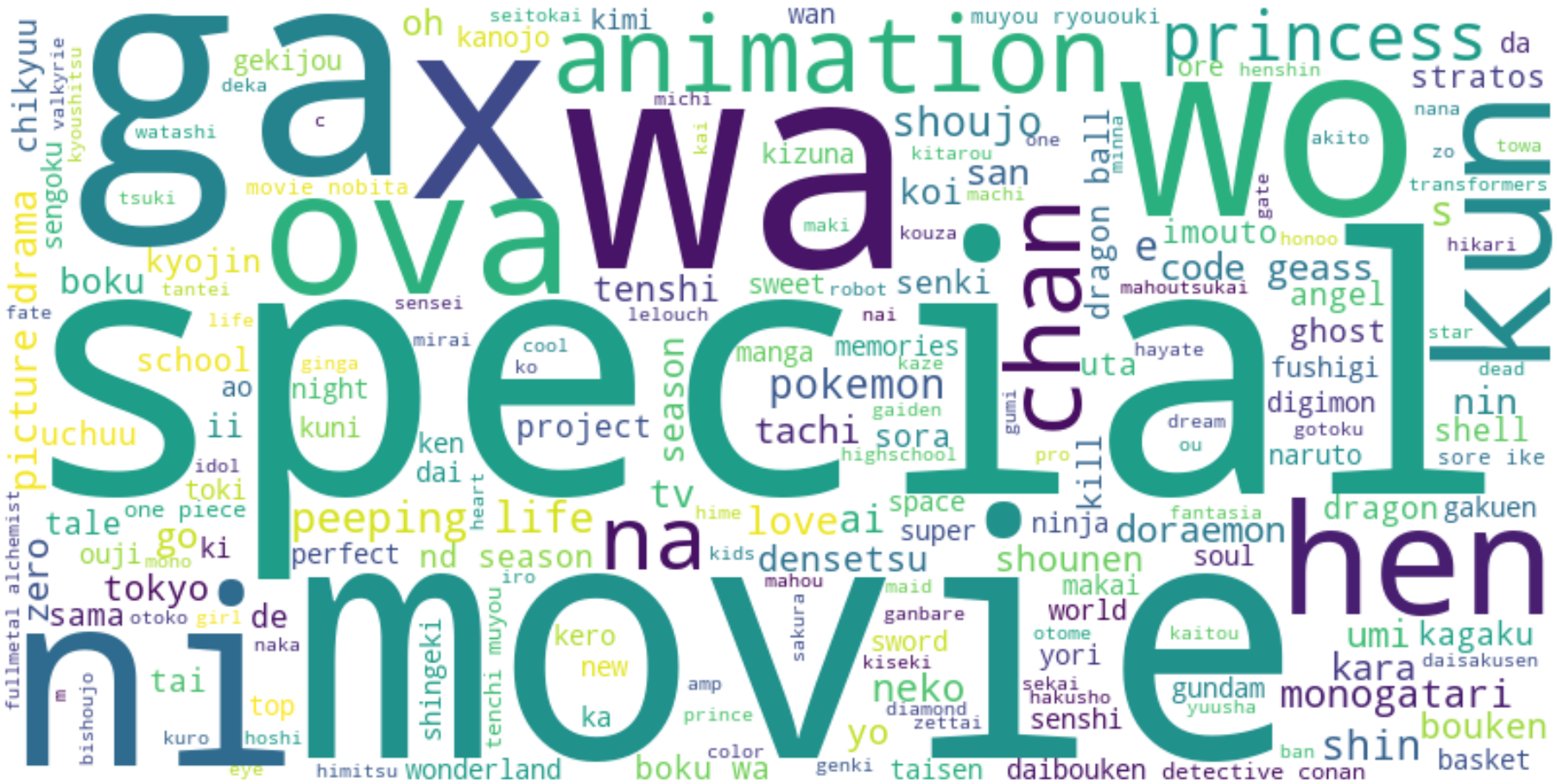
10 most popular anime

- Here are the 10 most-often rated anime in the dataset
- This list contains some famous titles

ratings	name	genre	rating
21124	Code Geass: Hangyaku no Lelouch R2	Action, Drama, Mecha, Military, Sci-Fi, Super ...	8.98
21332	Fullmetal Alchemist	Action, Adventure, Comedy, Drama, Fantasy, Mag...	8.33
21494	Fullmetal Alchemist: Brotherhood	Action, Adventure, Drama, Fantasy, Magic, Mili...	9.26
22071	Naruto	Action, Comedy, Martial Arts, Shounen, Super P...	7.81
23528	Elfen Lied	Action, Drama, Horror, Psychological, Romance,...	7.85
23565	Angel Beats!	Action, Comedy, Drama, School, Supernatural	8.39
24125	Code Geass: Hangyaku no Lelouch	Action, Mecha, Military, School, Sci-Fi, Super...	8.83
25289	Shingeki no Kyojin	Action, Drama, Fantasy, Shounen, Super Power	8.54
26309	Sword Art Online	Action, Adventure, Fantasy, Game, Romance	7.83
34226	Death Note	Mystery, Police, Psychological, Supernatural, ...	8.71

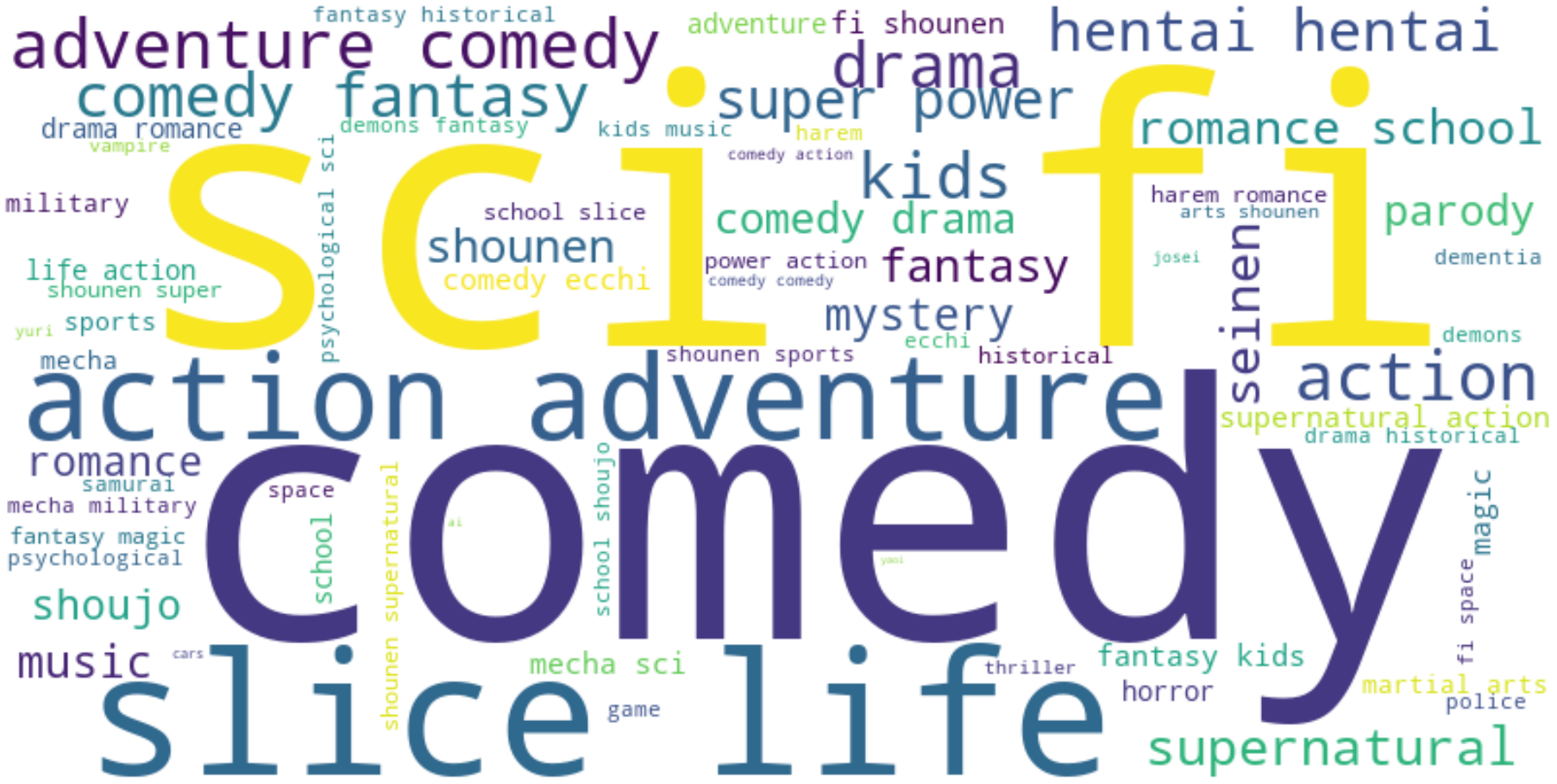
Word cloud of anime titles

- Most titles are in Japanese, however there are a few widely-used English words

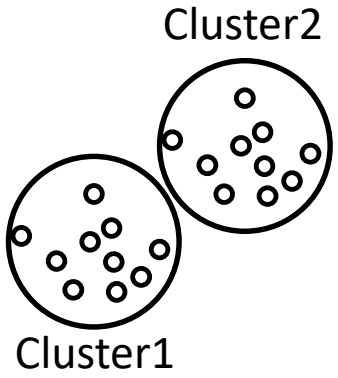


Word cloud of anime genres

- It’s clear which genres are the most popular. Some words in the cloud are duplicated.

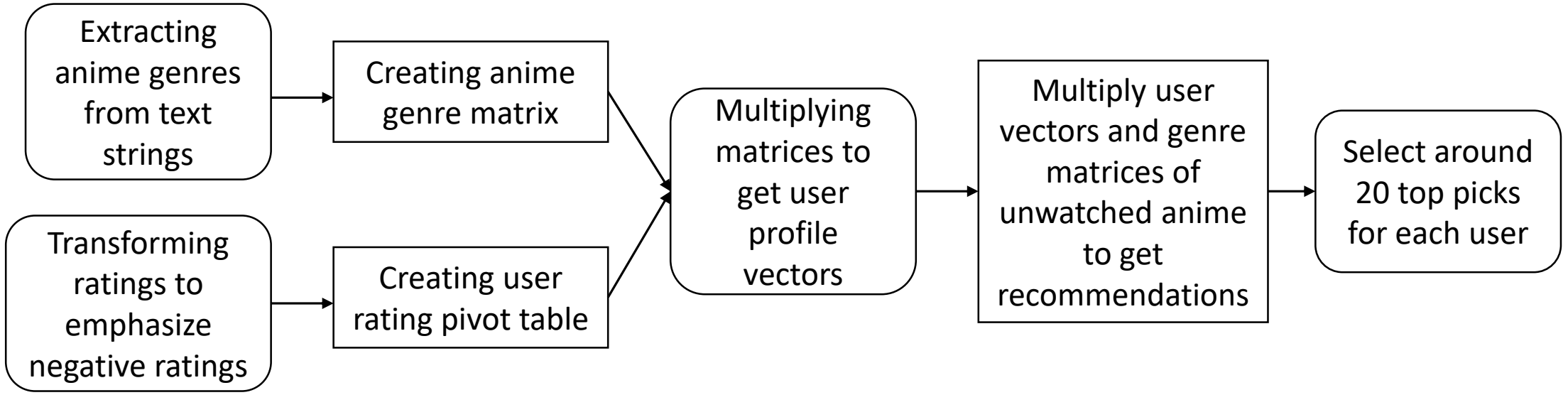


Content-based Recommender System using Unsupervised Learning



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

- All ratings lower than 5 were counted as negative to emphasize their significance
- Each user had a different threshold because of really different recommendation strengths
- The threshold value equaled the value of the 20th highest recommendation (so there could be >20 recommendations in case of a tie)



Evaluation results of user profile-based recommender system

- I evaluated the recommendations based on the ‘rating’ data that was available for every anime.
- I sampled top 10 anime by number of recommendations and sum of recommendation scores.
- Then, I calculated the average ‘rating’ value for these predictions, weighted by number of recommendations and sum of recommendation scores respectively.
- Average weighted rating of the top 10 most frequently recommended anime: **7.09**
- Average weighted rating of the top 10 anime by sum of recommendation scores: **7.06**
- All 1000 anime were recommended at least once, even though some anime in the dataset were not rated by any user.

Evaluation results of user profile-based recommender system

Top-10 anime by sum of recommendation scores across all users

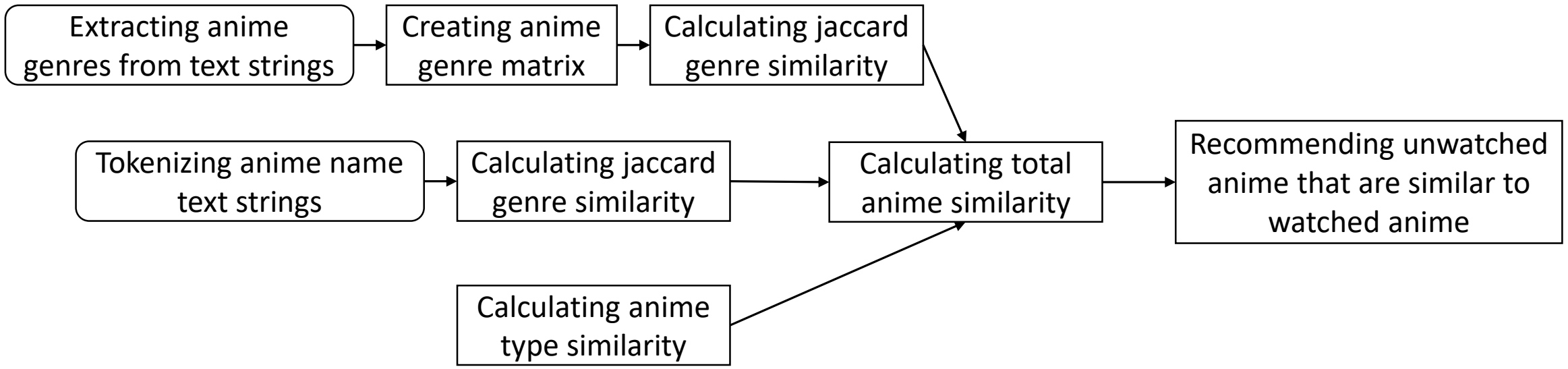
	sum_of_scores	name	rating
0	20219410.5	Konjiki no Gash Bell!!: Ougon no Chichi wo Mot...	6.63
1	19132357.5	Silent Mobius	6.95
2	19104047.5	Gakusen Toshi Asterisk	7.14
3	17840536.5	Dragon Drive	6.88
4	17462845.5	Kotetsu no Daibouken	6.03
5	16941529.0	InuYasha: Tenka Hadou no Ken	7.89
6	14189376.0	Dragon Ball GT	6.72
7	13822810.5	Dragon Ball Kai (2014)	8.01
8	13165799.0	Urusei Yatsura Movie 5: Final	7.72
9	12522940.5	One Piece Film: Gold Episode 0 - 711 ver.	7.04

Top-10 most commonly recommended anime across all users

	number_of_recommendations	name	rating
0	52922	Konjiki no Gash Bell!!: Ougon no Chichi wo Mot...	6.63
1	50389	Gakusen Toshi Asterisk	7.14
2	49825	Silent Mobius	6.95
3	46692	Dragon Drive	6.88
4	45930	InuYasha: Tenka Hadou no Ken	7.89
5	42905	Kotetsu no Daibouken	6.03
6	41604	Dragon Ball GT	6.72
7	37458	Dragon Ball Kai (2014)	8.01
8	35448	Zero no Tsukaima: Futatsuki no Kishi	7.73
9	34763	Digimon Frontier	7.25

Flowchart of content-based recommender system using anime similarity

- Anime similarity was calculated as a weighted sum of 3 different similarity measures
- Values were normalized so that the maximum similarity was 1
- A similarity threshold of 0.5 was used for predictions



Evaluation results of anime similarity based recommender system

- From the 1000 anime, only 238 were recommended at least once.
- Around 7.4% of users did not receive any recommendations.
- Among the users that did receive a recommendation, a user received around 13 recommendations on average.
- Average weighted rating of the top 10 most frequently recommended anime: **7.80**
- Average weighted rating of the top 10 anime by sum of recommendation scores: **7.64**

Evaluation results of anime similarity based recommender system

Top-10 anime by sum of recommendation scores across all users

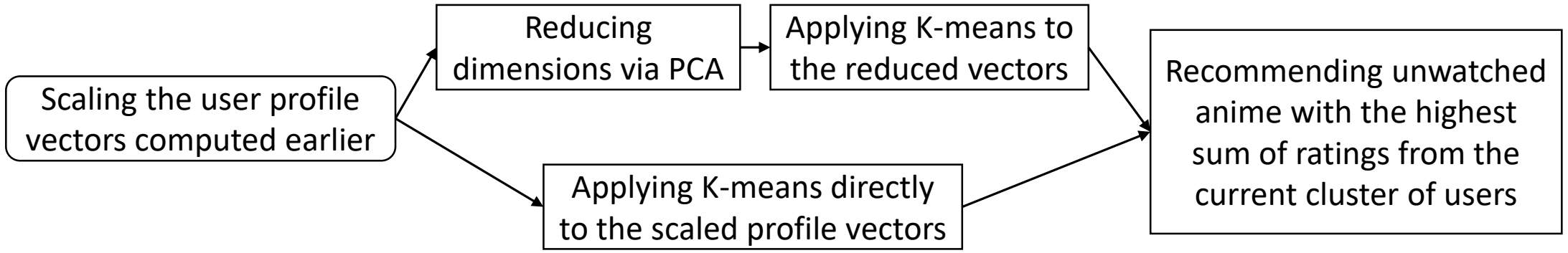
	sum_of_scores	name	rating
0	19273.417544	Dragon Ball Kai (2014)	8.01
1	17079.353383	I My Me! Strawberry Eggs	6.93
2	15280.375940	Shingeki no Kyojin OVA	7.88
3	14303.709273	Amagami SS+ Plus	7.61
4	13215.789474	Sword Art Online II	7.35
5	13214.025063	Bokura wa Minna Kawaisou	7.90
6	12614.406015	Sakurasou no Pet na Kanojo	8.40
7	10998.781955	Chuunibyou demo Koi ga Shitai!	7.95
8	10465.964912	Soul Eater NOT!	6.33
9	9784.000000	Ao no Exorcist Movie	7.88

Top-10 most commonly recommended anime across all users

	number_of_recommendations	name	rating
0	31310	I My Me! Strawberry Eggs	6.93
1	28421	Dragon Ball Kai (2014)	8.01
2	26055	Amagami SS+ Plus	7.61
3	24114	Bokura wa Minna Kawaisou	7.90
4	23578	Sakurasou no Pet na Kanojo	8.40
5	22581	Shingeki no Kyojin OVA	7.88
6	20669	Chuunibyou demo Koi ga Shitai!	7.95
7	17437	Tonari no Kaibutsu-kun	7.77
8	15726	Kimi to Boku. 2	8.11
9	15081	Kimi to Boku.	7.85

Flowchart of clustering-based recommender system

- The original user profile vectors have 45 values, corresponding to 45 genres in the dataset
- Many of the values are at least somewhat correlated with the others
- It was possible to reduce the number of dimensions to 17 with PCA while retaining 90% of the variance.
- More complicated dimensionality reduction and clustering techniques were not attempted due to computational limitations



Evaluation results of clustering-based recommender system

- From the 1000 anime, only 220 were recommended at least once.
- Dimensionality reduction with PCA changed very little: the recommendations were almost the same, the scores were exactly the same.
- Average weighted rating of the top 10 most frequently recommended anime: **8.64**
- Average weighted rating of the top 10 anime by sum of recommendation scores: **8.53**

Evaluation results of clustering-based recommender system

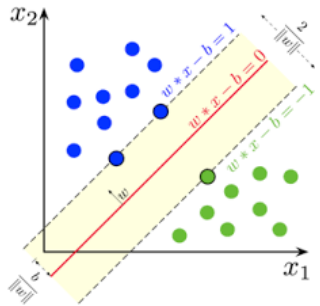
Top-10 anime by sum of recommendation scores across all users

	sum_of_scores	name	rating
0	859678506.0	Death Note	8.71
1	656205415.0	Shingeki no Kyojin	8.54
2	632255119.0	Sen to Chihiro no Kamikakushi	8.93
3	618850907.0	Fullmetal Alchemist: Brotherhood	9.26
4	565931397.0	Code Geass: Hangyaku no Lelouch	8.83
5	560949400.0	Sword Art Online	7.83
6	560721977.0	Angel Beats!	8.39
7	543285575.0	Ouran Koukou Host Club	8.39
8	543037361.0	Elfen Lied	7.85
9	534958016.0	Fullmetal Alchemist	8.33

Top-10 most commonly recommended anime across all users

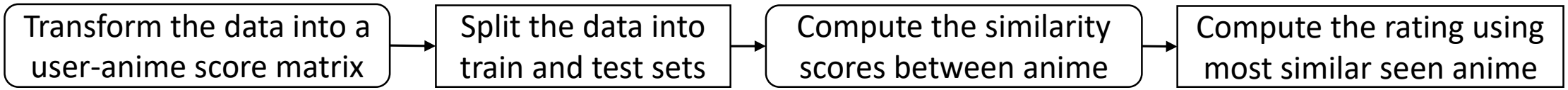
	number_of_recommendations	name	rating
0	45850	Code Geass: Hangyaku no Lelouch R2	8.98
1	45464	Fullmetal Alchemist: Brotherhood	9.26
2	44401	Sen to Chihiro no Kamikakushi	8.93
3	43443	Elfen Lied	7.85
4	43254	Fullmetal Alchemist	8.33
5	43207	Howl no Ugoku Shiro	8.74
6	42849	Code Geass: Hangyaku no Lelouch	8.83
7	41681	Toradora!	8.45
8	41649	Shingeki no Kyojin	8.54
9	41580	Angel Beats!	8.39

Collaborative-filtering Recommender System using Supervised Learning



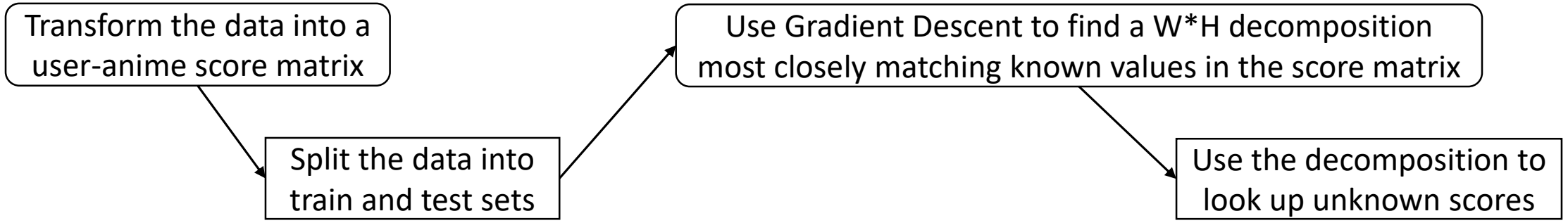
Flowchart of KNN based recommender system

- User-based KNN system was not attempted due to computational limitations
- Task was completed with the scikit-surprise package
- Different sets of hyperparameters were tried, none achieved performance improvements



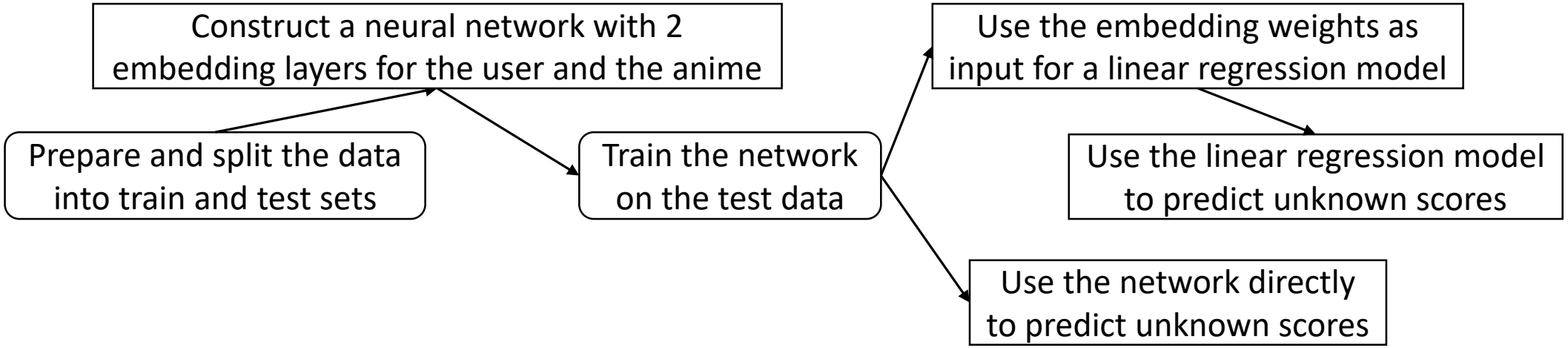
Flowchart of NMF based recommender system

- Task was completed with the scikit-surprise package
- Different sizes of W and H matrices were tried, none achieved significant performance improvements



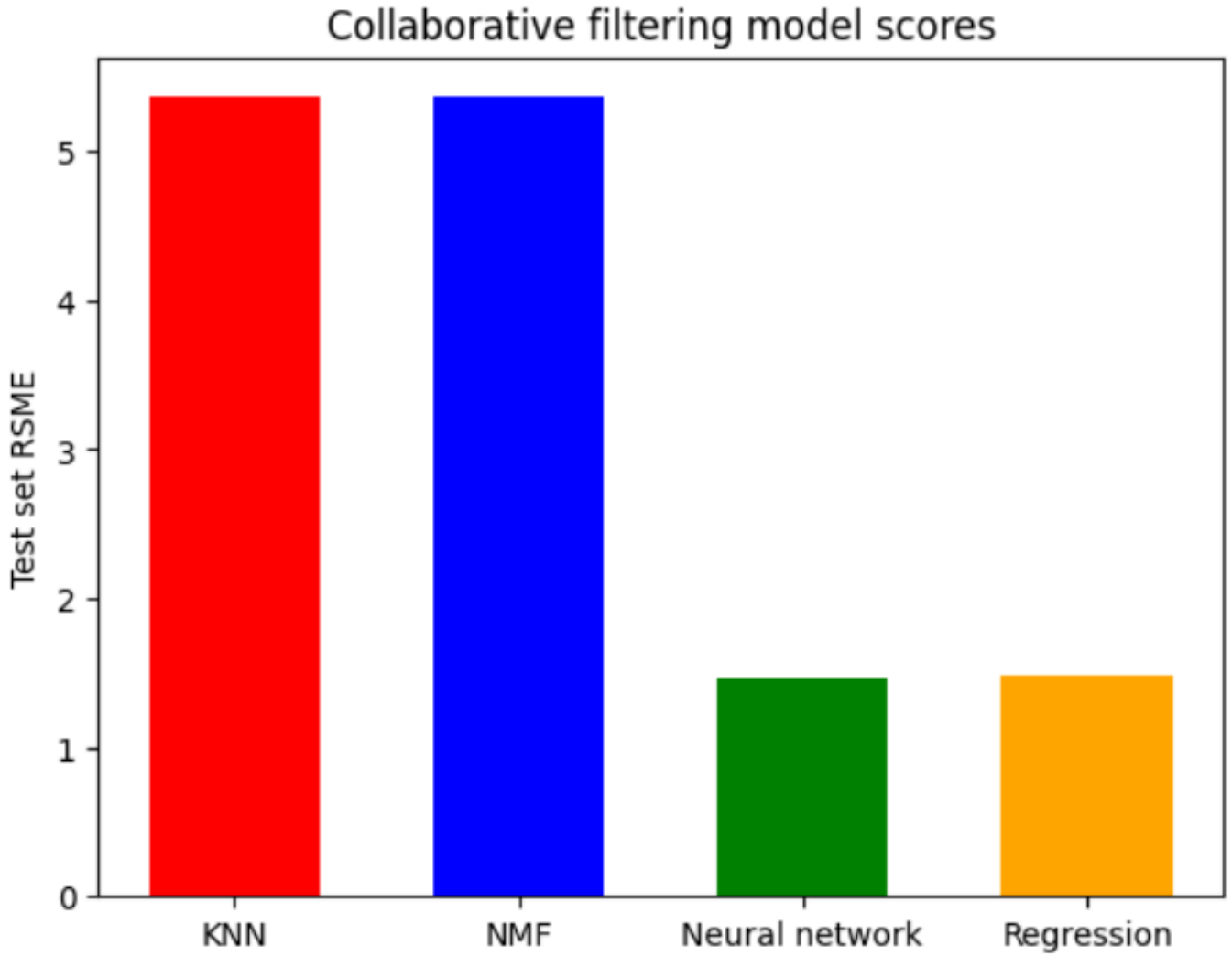
Flowchart of Neural Network Embedding based recommender system

- Embeddings of users and anime had 32 values each
- Early stopping was used while training the network, optimal performance was achieved only after 4 epochs
- Classification on the embedding vectors was not attempted due to a large number of potential classes (10)



Compare the performance of collaborative-filtering models

- All KNN and NMF-based models do poorly and shouldn't be used for this problem
- The neural network model performed much better
- Curiously, the regression model based on the embedding vectors from the neural network performed at about the same level
- Overall, RMSE of 1.5 for neural network and regression models is a fine result because the scores were ranging from 1-10



Conclusions

- All content-based systems performed relatively well
- The user clustering-based system tended to recommend only popular anime, while the system using user profiles and course genres was the only one to recommend anime that wasn’t rated previously
- In practice, probably a combination of those approaches would be best to ensure that the recommendations contain both ‘good’ and ‘new’ anime
- Collaborative filtering systems based on the neural network have also performed well, and it would be interesting to see what recommendations they come up with for all the users. But they must be used in conjunction with other approaches since they won’t be able to recommend ‘new’ anime.

Appendix

- Heatmap of anime similarity scores
- Github link:

https://github.com/vectorkoz/anime-recommender-system