Data Science Workflow

Recommender Systems Project

23.05.2024
Nicolas Hahn

Content

1. Dataset
2. Content-based
3. Collaborative filtering with neural network (matrix factorization)
4. Results

1. Dataset

1. Dataset

- Anime dataset (anime.csv and rating.csv)

  https://www.kaggle.com/datasets/CooperUnion/anime-recommendations-database?resource=download

	anime_id	name	genre	type	episodes	rating	members
0	32281	Kimi no Na wa.	Drama, Romance, School, Supernatural	Movie	1	9.37	200630
1	5114	Fullmetal Alchemist: Brotherhood	Action, Adventure, Drama, Fantasy, Magic, Military, Shounen	TV	64	9.26	793665
2	28977	Gintama°	Action, Comedy, Historical, Parody, Samurai, Sci-Fi, Shounen	TV	51	9.25	114262
3	9253	Steins;Gate	Sci-Fi, Thriller	TV	24	9.17	673572
4	9969	Gintama'	Action, Comedy, Historical, Parody, Samurai, Sci-Fi, Shounen	TV	51	9.16	151266

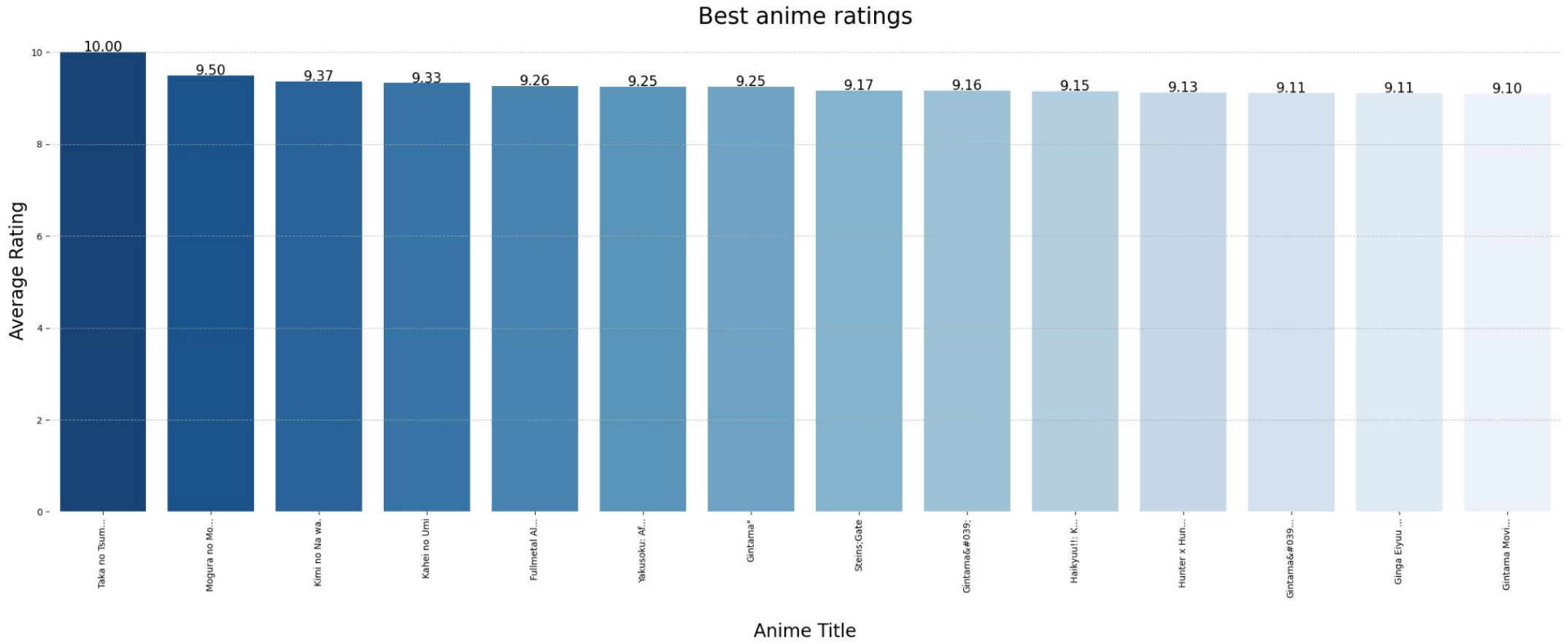
	user_id	anime_id	rating
0	1	20	-1
1	1	24	-1
2	1	79	-1
3	1	226	-1
4	1	241	-1

- After deleting rows with “NaN-Values” and “Unknown”:

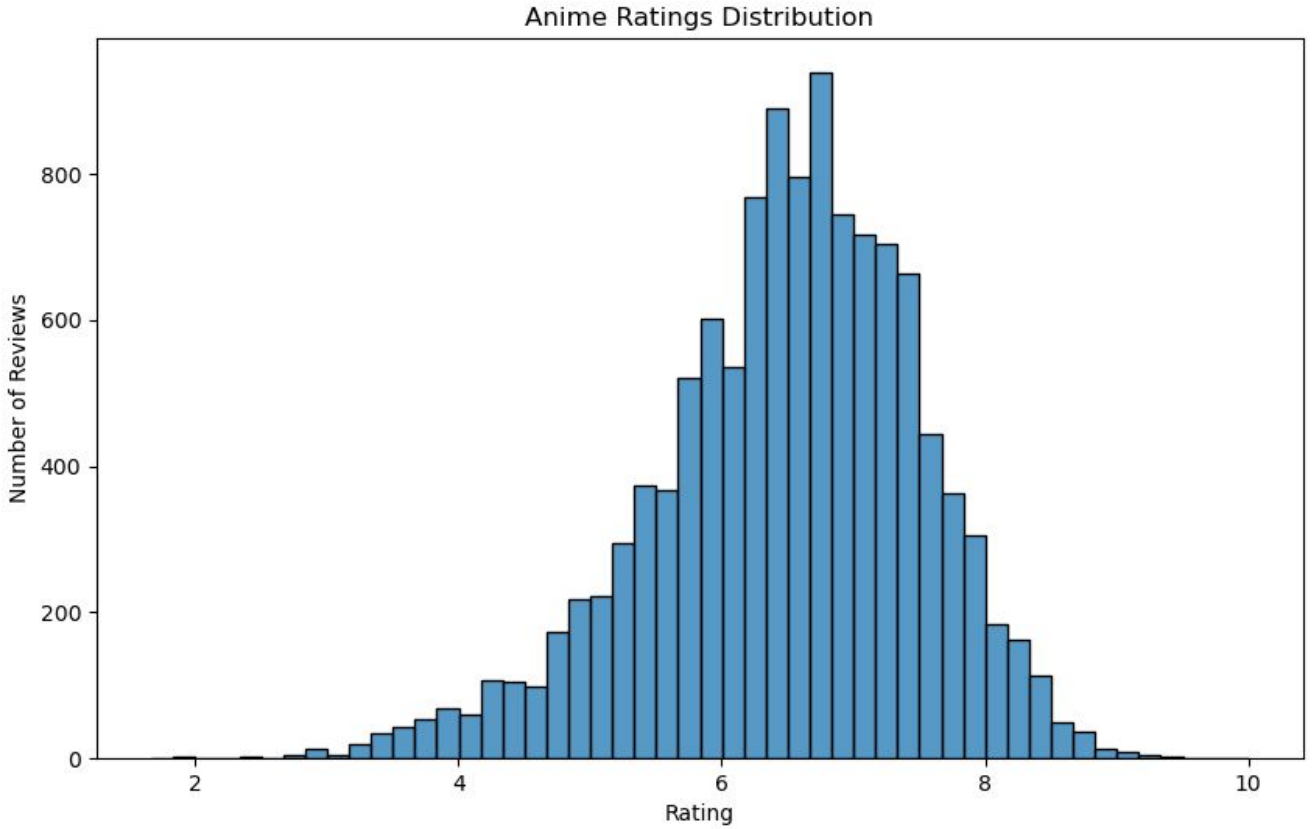
  anime data frame=11.830 rows

  ratings data frame= 7.813.737 rows

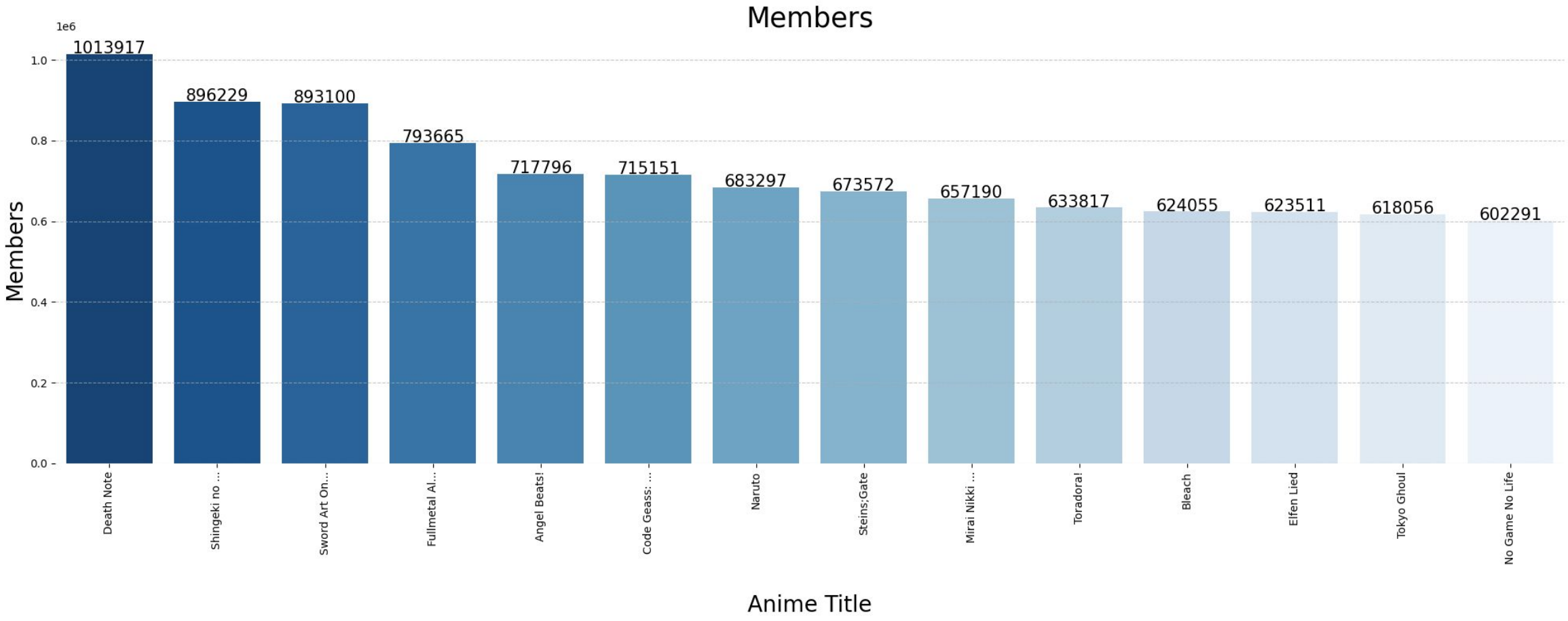
1. Dataset


Best anime ratings
Average Rating
Anime Title
10.00
9.50
9.37
9.33
9.26
9.25
9.25
9.17
9.16
9.15
9.13
9.11
9.11
9.10
Taka no Tsum...
Mogura no Mo...
Kimi no Na wa.
Kahei no Umi
Fullmetal Al...
Yakusoku: Af...
Gintama°
Steins;Gate
Gintama'
Haikyuu!!: K...
Hunter x Hun...
Gintama&#039...
Ginga Eiyuu ...
Gintama Movi...

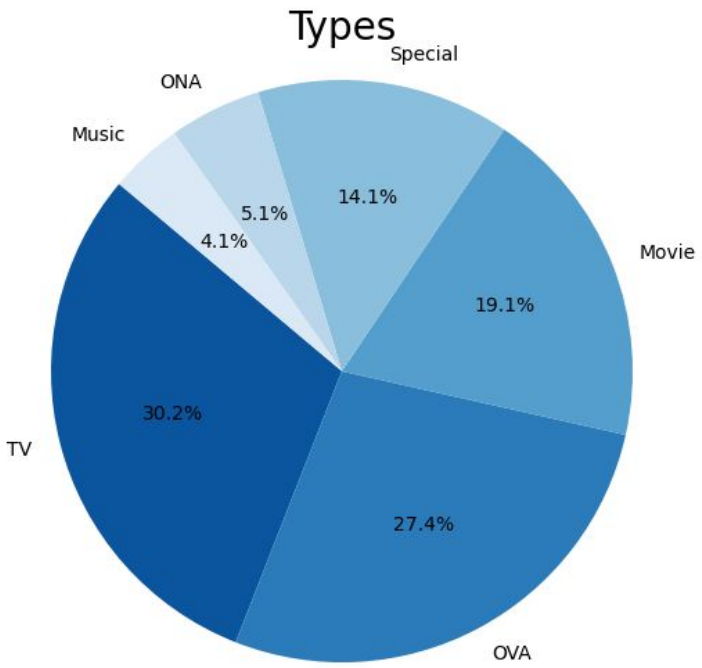
1. Dataset

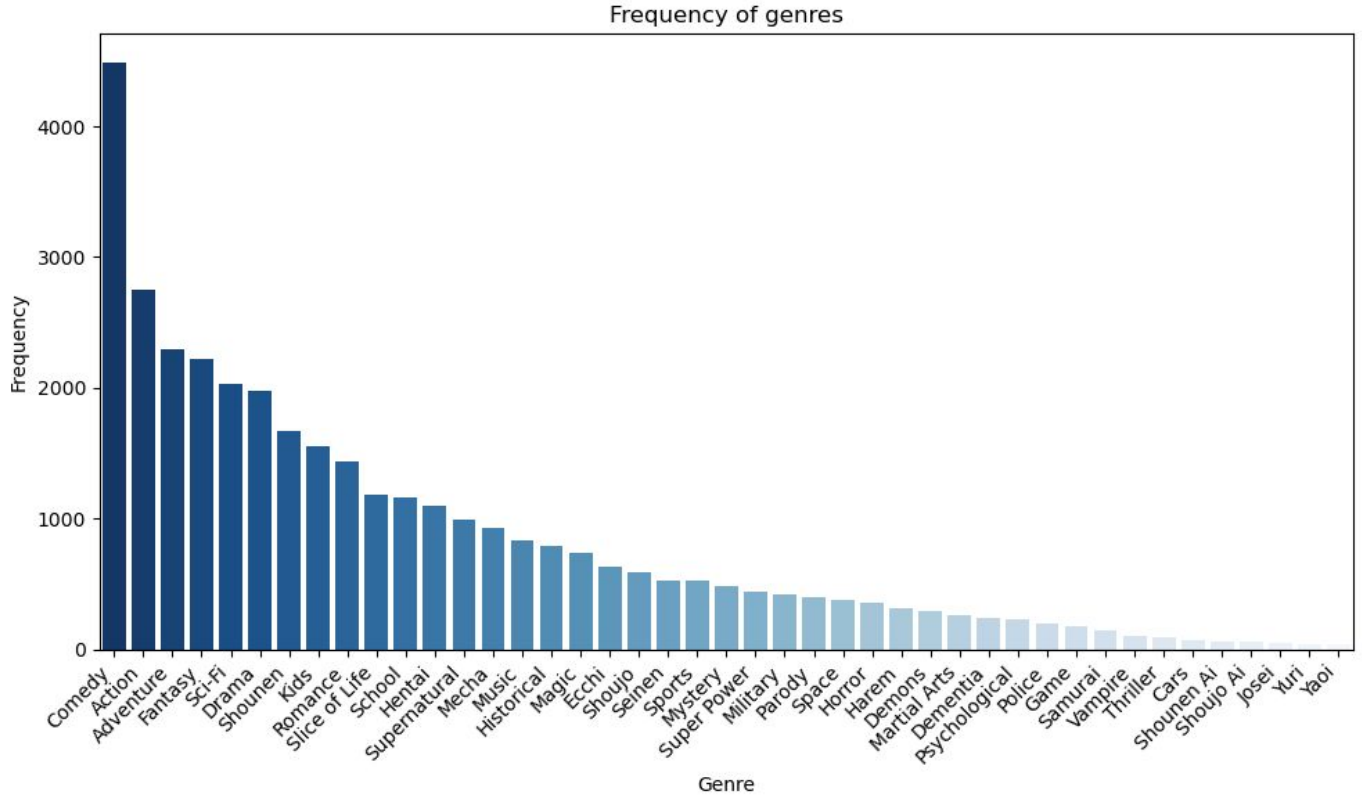

Anime Ratings Distribution
Number of Reviews
0
200
400
600
800
2
4
6
8
10
Rating

1. Dataset

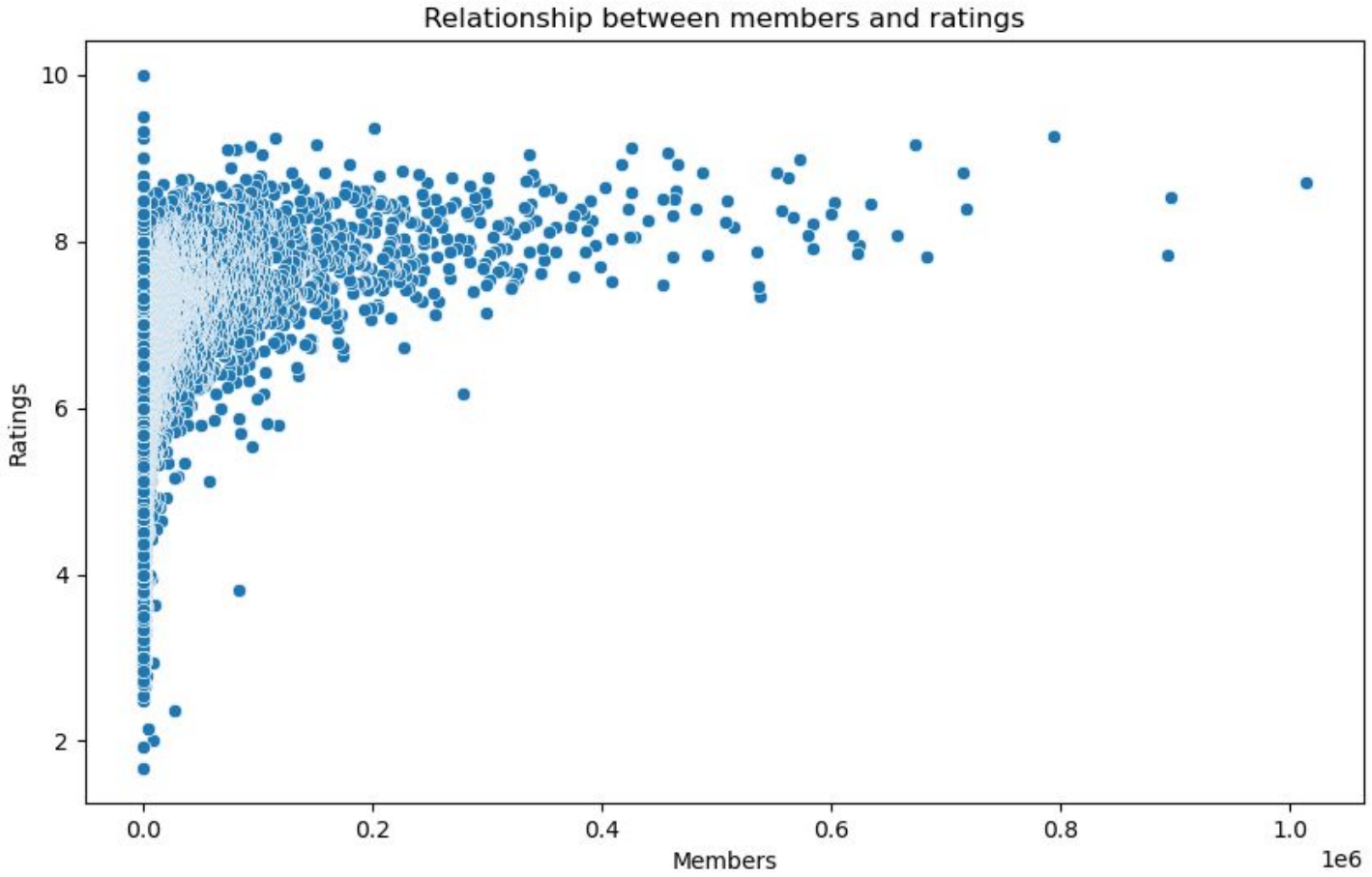

Members
1e6
1013917
896229
893100
793665
717796
715151
683297
673572
657190
633817
624055
623511
618056
602291
1.0
0.8
0.6
0.4
0.2
0.0
Members
Death Note
Shingeki no ...
Sword Art On...
Fullmetal Al...
Angel Beats!
Code Geass: ...
Naruto
Steins;Gate
Mirai Nikki ...
Toradora!
Bleach
Elfen Lied
Tokyo Ghoul
No Game No Life
Anime Title

1. Dataset

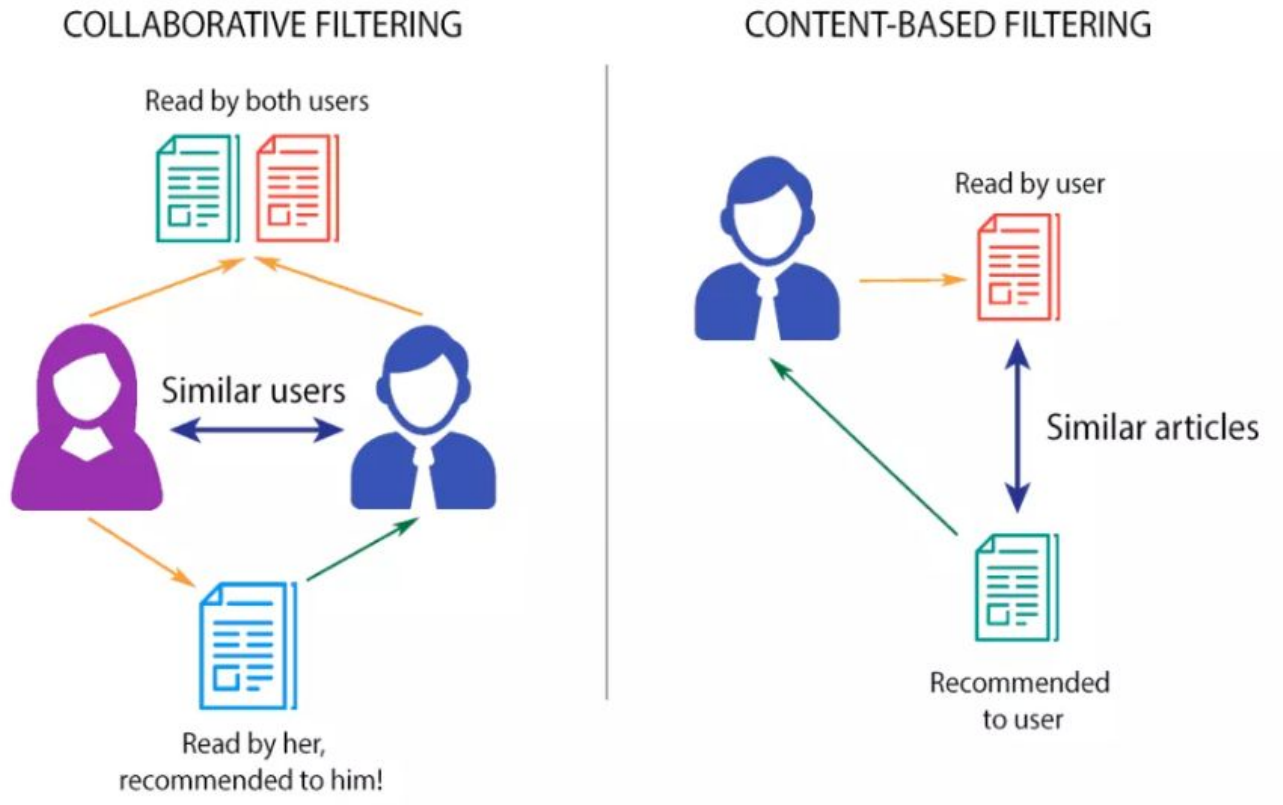

Types
Special
ONA
Music
Movie
TV
OVA
14.1%
5.1%
4.1%
19.1%
30.2%
27.4%


Frequency of genres
Frequency
0
1000
2000
3000
4000
Comedy
Action
Adventure
Fantasy
Sci-Fi
Drama
Shounen
Kids
Romance
Slice of Life
School
Hentai
Supernatural
Mecha
Music
Historical
Magic
Ecchi
Shoujo
Seinen
Sports
Mystery
Super Power
Military
Parody
Space
Horror
Harem
Demons
Martial Arts
Dementia
Psychological
Police
Game
Samurai
Vampire
Thriller
Cars
Shounen Ai
Shoujo Ai
Josei
Yuri
Yaoi
Genre

1. Dataset


Relationship between members and ratings
Ratings
10
8
6
4
2
0.0
0.2
0.4
0.6
0.8
1.0
Members
1e6

Overview


COLLABORATIVE FILTERING
Read by both users
Similar users
Read by her,
recommended to him!
CONTENT-BASED FILTERING
Read by user
Similar articles
Recommended
to user

[10]

2. Content-based recommendation

2. Content-based recommendation

- Generates recommendations for users based on the attributes (features) of items and the preferences of the users:
    1. Extracting features of items
    2. Creating a user profile
    3. Similarity calculation
    4. Generating recommendations

- Different ways to calculate the similarity:
    1. Dot product: weighted sum of overlapping features
    2. Cosine similarity: measures the angle between two vectors

2. Content-based recommendation

One-Hot-Encoding for categorical columns *genre* and *type*:

anime_id	name	genre	type	episodes	rating	members	Shoujo Ai	Police	Thriller	...	Demons	Shounen Ai	Fantasy	Comedy	Action	OVA	TV	ONA	Special	Movie
32281	Kimi no Na wa.	[Drama, Romance, School, Supernatural]	[Movie]	1	9.37	200630	0	0	0	...	0	0	0	0	0	0	0	0	0	1
5114	Fullmetal Alchemist: Brotherhood	[Action, Adventure, Drama, Fantasy, Magic, Military, Shounen]	[TV]	64	9.26	793665	0	0	0	...	0	0	1	0	1	0	1	0	0	0
28977	Gintama°	[Action, Comedy, Historical, Parody, Samurai, Sci-Fi, Shounen]	[TV]	51	9.25	114262	0	0	0	...	0	0	0	1	1	0	1	0	0	0
9253	Steins;Gate	[Sci-Fi, Thriller]	[TV]	24	9.17	673572	0	0	1	...	0	0	0	0	0	0	1	0	0	0
9969	Gintama'	[Action, Comedy, Historical, Parody, Samurai, Sci-Fi, Shounen]	[TV]	51	9.16	151266	0	0	0	...	0	0	0	1	1	0	1	0	0	0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
9316	Toushindai My Lover: Minami tai Mecha-Minami	[Hentai]	[OVA]	1	4.15	211	0	0	0	...	0	0	0	0	0	1	0	0	0	0
5543	Under World	[Hentai]	[OVA]	1	4.28	183	0	0	0	...	0	0	0	0	0	1	0	0	0	0
5621	Violence Gekiga David no Hoshi	[Hentai]	[OVA]	4	4.88	219	0	0	0	...	0	0	0	0	0	1	0	0	0	0
6133	Violence Gekiga Shin David no Hoshi: Inma Densetsu	[Hentai]	[OVA]	1	4.98	175	0	0	0	...	0	0	0	0	0	1	0	0	0	0
26081	Yasuji no Pornorama: Yacchimae!!	[Hentai]	[Movie]	1	5.46	142	0	0	0	...	0	0	0	0	0	0	0	0	0	1

2. Content-based recommendation

Item-Feature matrix:

anime_id	Samurai	Action	Dementia	Sports	Demons	Ecchi	Seinen	Super Power	Music	Shoujo Ai	...	Mecha	Parody	School	Psychological	Martial Arts	ONA	Movie	Special	OVA	TV
32281	0	0	0	0	0	0	0	0	0	0	...	0	0	1	0	0	0	1	0	0	0
5114	0	1	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	1
28977	1	1	0	0	0	0	0	0	0	0	...	0	1	0	0	0	0	0	0	0	1
9253	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	1
9969	1	1	0	0	0	0	0	0	0	0	...	0	1	0	0	0	0	0	0	0	1
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...

User-Feature matrix

user_id	rating	Samurai	Action	Dementia	Sports	Demons	Ecchi	Seinen	Super Power	...	Mecha	Parody	School	Psychological	Martial Arts	ONA	Movie	Special	OVA	TV
52969	9	0	0	0	0	0	0	0	0	...	0	0	0	1	0	0	1	0	0	0
46409	7	0	0	0	0	0	0	0	0	...	0	0	0	1	0	0	1	0	0	0
51573	9	0	0	0	0	0	0	0	0	...	0	0	0	1	0	0	1	0	0	0
5820	10	0	0	0	0	0	0	0	0	...	0	0	0	1	0	0	1	0	0	0
57932	9	0	0	0	0	0	0	0	0	...	0	0	0	1	0	0	1	0	0	0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...

2. Content-based recommendation

- Multiply the scores of features by rating, normalize per user and similarity with example for user=8:

Pick Top-10 Recommendations

```
recommended.sort_values('sim',ascending=False).head(10).merge(anime_df, left_index=True, right_index=True).T
```

✓ 0.0s Python

anime_id	231	25157	30544	573	157	6811	249	98	1195	1840
sim	0.773469	0.7	0.646939	0.644898	0.638776	0.630612	0.630612	0.626531	0.62449	0.62449
name	Asagiri no Miko	Trinity Seven	Gakusen Toshi Asterisk	Saber Marionette J	Mahou Sensei Negima!	InuYasha: Kanketsu-hen	InuYasha	Mai-HiME	Zero no Tsukaima	Zero no Tsukaima: Futatsuki no Kishi
genre	[Action, Comedy, Drama, Fantasy, Magic, School, Sci-Fi, Shounen, Slice of Life, Super Power, Supernatural]	[Action, Comedy, Ecchi, Fantasy, Harem, Magic, Romance, School, Shounen, Supernatural]	[Action, Comedy, Ecchi, Fantasy, Harem, Romance, School, Sci-Fi, Supernatural]	[Action, Adventure, Comedy, Drama, Harem, Martial Arts, Mecha, Romance, Sci-Fi, Shounen]	[Comedy, Ecchi, Fantasy, Harem, Magic, Romance, School, Shounen, Super Power, Supernatural]	[Action, Adventure, Comedy, Demons, Fantasy, Magic, Romance, Shounen, Supernatural]	[Action, Adventure, Comedy, Demons, Fantasy, Magic, Romance, Shounen, Supernatural]	[Action, Comedy, Drama, Fantasy, Magic, Mecha, Romance, School, Shoujo Ai]	[Action, Adventure, Comedy, Ecchi, Fantasy, Harem, Magic, Romance, School]	[Action, Adventure, Comedy, Ecchi, Fantasy, Harem, Magic, Romance, School]
type	[TV]	[TV]	[TV]	[TV]	[TV]	[TV]	[TV]	[TV]	[TV]	[TV]

→ user like animes from genre action, comedy, drama etc. (whole frame can be viewed in the notebook with the corresponding features)

3. Collaborative filtering with neural network (matrix factorization)

3. Collaborative filtering with neural network (matrix factorization)

- Matrix factorization is a class of collaborative filtering models, predict the reaction of a user on a specific item based on reactions of “similar” users
- Training data of the model:

	user_id	anime_id	rating
0	1	20	-1
1	1	24	-1
2	1	79	-1
3	1	226	-1
4	1	241	-1

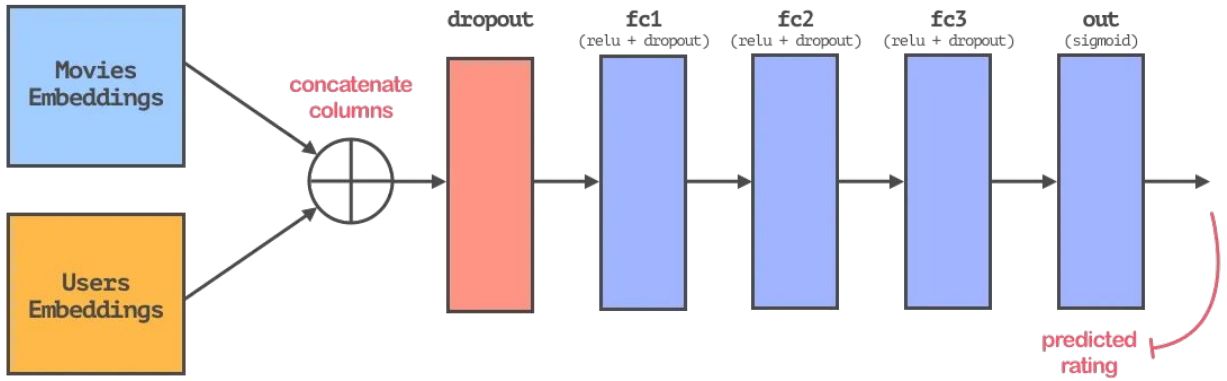
anime_id	name	genre	type	episodes	rating	members
32281	Kimi no Na wa.	Drama, Romance, School, Supernatural	Movie	1	9.37	200630
5114	Fullmetal Alchemist: Brotherhood	Action, Adventure, Drama, Fantasy, Magic, Mili...	TV	64	9.26	793665
28977	Gintama°	Action, Comedy, Historical, Parody, Samurai, S...	TV	51	9.25	114262
9253	Steins;Gate	Sci-Fi, Thriller	TV	24	9.17	673572
9969	Gintama'	Action, Comedy, Historical, Parody, Samurai, S...	TV	51	9.16	151266

```
Number of ratings: 7813737
Number of unique anime_id's: 11200
Number of unique users: 73515
```

- One hot encoding does not work because of the amount of users and animes: Embedding N-dimensional vectors

3. Collaborative filtering with neural network (matrix factorization)

- The matrix factorization model used to predict ratings that a user might give to an item (*Koren et al., 2009*, [12]):



[1]

- Architecture: 1. Embeddings matrices convert IDs into arrays
  2. Fully-connected layers with dropouts
  3. Output: predicted ratings

3. Collaborative filtering with neural network (matrix factorization)

- Loss (root-mean-square erro

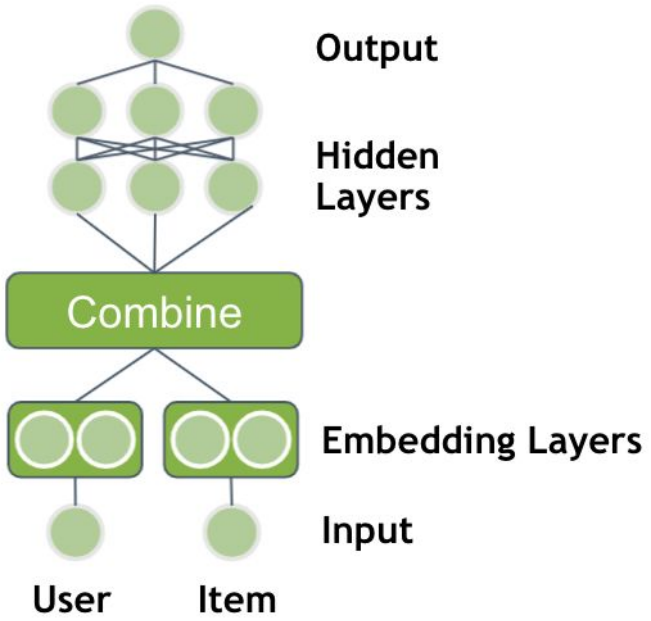
$$\mathrm{RMSE} = \sqrt{\frac{1}{|\mathcal{T}|} \sum_{(u,i)\in\mathcal{T}} (\mathbf{R}_{ui} - \hat{\mathbf{R}}_{ui})^2}$$

[12]

- Architecture with data set specific input embeddings:

not the unique number of users
and items → max value

```
Model:  RecommenderNet(
  (user_embedding): Embedding(73515, 100)
  (item_embedding): Embedding(34367, 100)
  (hl1): Linear(in_features=200, out_features=128, bias=True)
  (hl2): Linear(in_features=128, out_features=64, bias=True)
  (hl3): Linear(in_features=64, out_features=1, bias=True)
  (dropout): Dropout(p=0.1, inplace=False)
)
Number of parameters: 10822249
```



[14]

3. Collaborative filtering with neural network (matrix factorization)

- Recommend-method to get to get information about the predictions:

UserID Number of recommendation

```
recommendation = model.recommend_user_animes(2000, 10)
```

- Recommend-method to pick animes with top k predictions for a user:

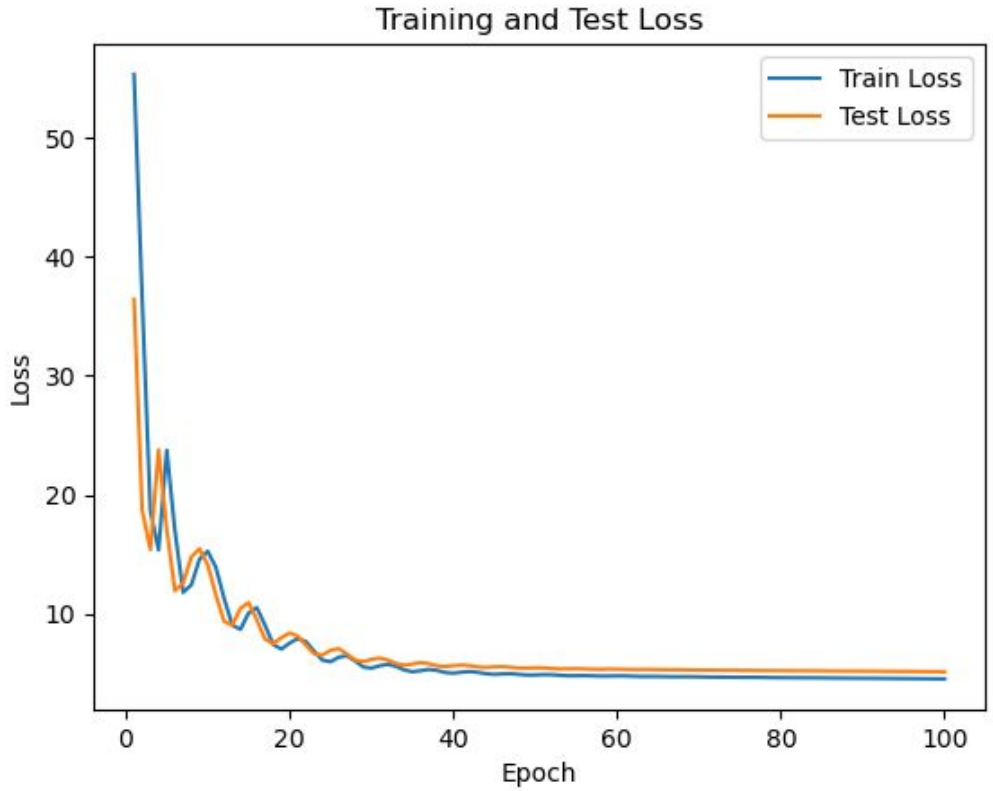
Score for the anime for this user

	anime_id	name	genre	type	episodes	rating	members	Recommender_score
411	1689	Byousoku 5 Centimeter	Drama, Romance, Slice of Life	Movie	3	8.10	324035	1.000000
605	1505	Detective Conan Movie 09: Strategy Above the D...	Adventure, Comedy, Mystery, Police, Shounen	Movie	1	7.94	20347	0.990223
1449	5228	Initial D Extra Stage 2	Action, Cars, Drama, Seinen, Sports	OVA	1	7.54	11468	0.981912
1954	944	The Third: Aoi Hitomi no Shoujo	Action, Adventure, Sci-Fi, Seinen	TV	24	7.40	14304	0.970310
2852	1492	Project ARMS	Action, Martial Arts, Super Power	TV	26	7.15	6903	0.957009
3254	580	Kogepan	Comedy	TV	10	7.04	4436	0.955027
3750	1127	UFO Princess Valkyrie	Comedy, Ecchi, Magic, Sci-Fi	TV	12	6.90	11794	0.954029
4282	2911	Kyoushirou to Towa no Sora Specials	Drama, Magic, Romance, Shounen	Special	6	6.76	3946	0.953592
4536	23	Ring ni Kakero 1	Action, Shounen, Sports	TV	12	6.70	2224	0.951459
7884	1765	Vampire Sensou	Super Power, Supernatural, Vampire	OVA	1	4.76	2002	0.950869

4. Results

4. Results

- The network learns and improves:

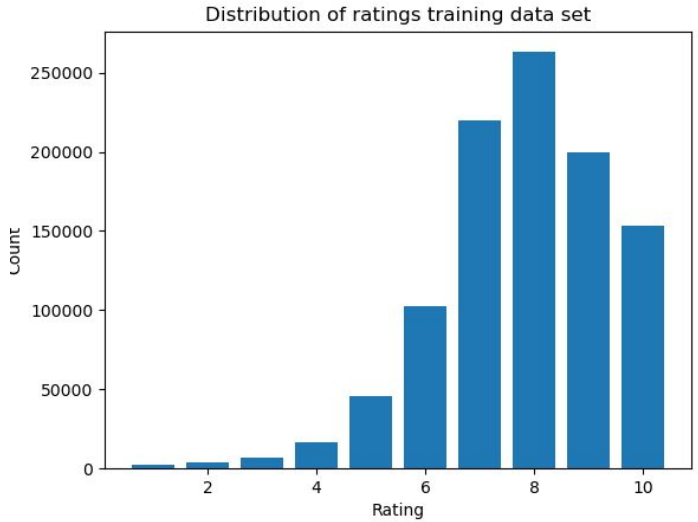


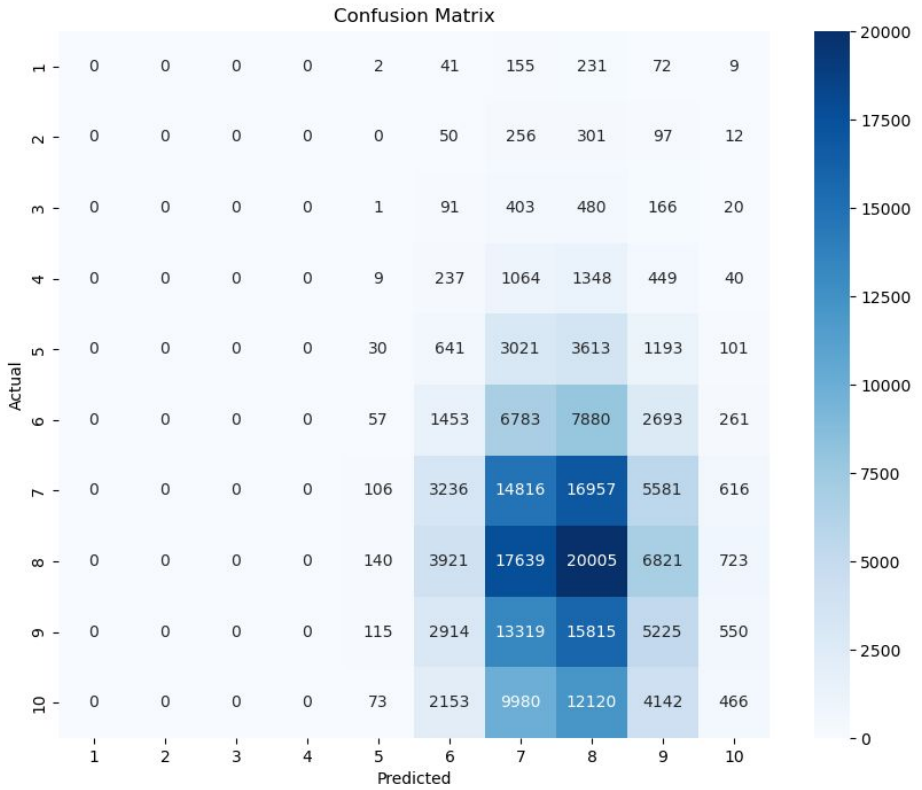
4. Results

- Confusion matrix and distribution of training data:

  Ratings 6 to 9: still close to the actual ratings

  Ratings 1 to 4: poor performance in predicting lower ratings



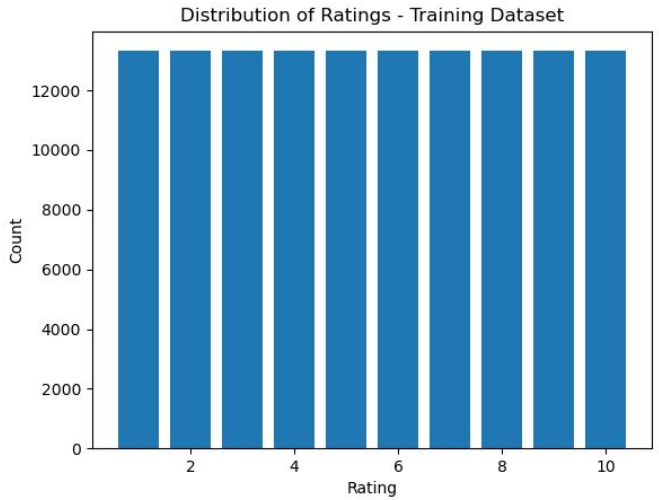


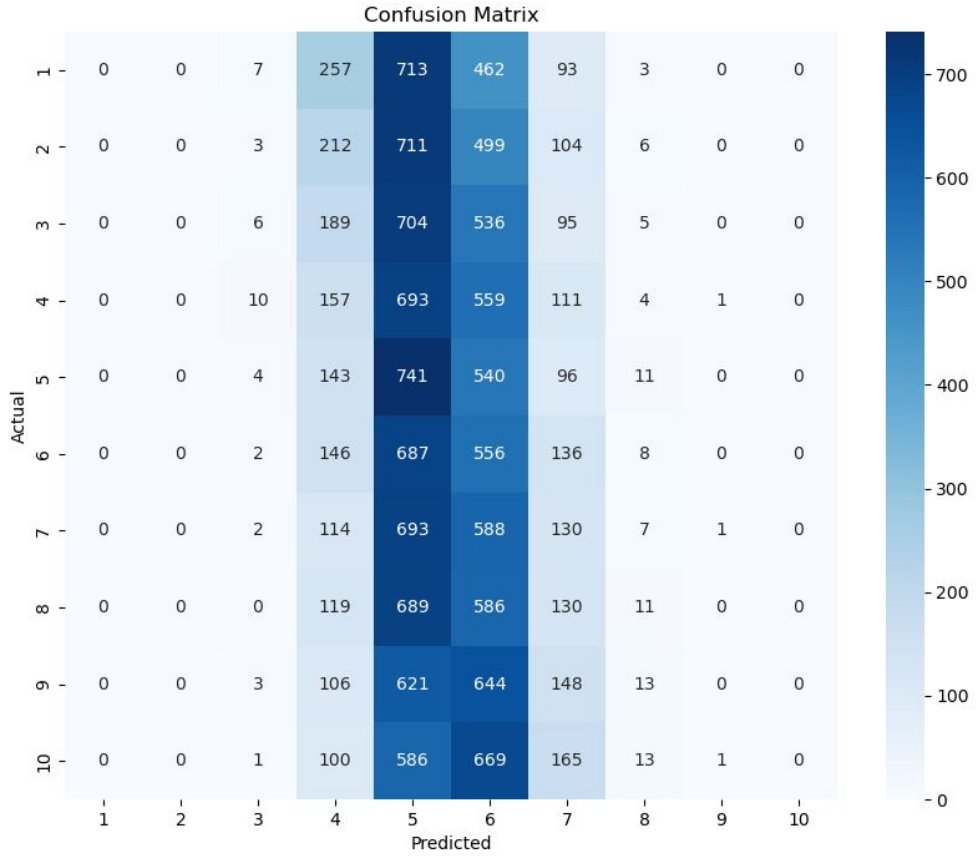
4. Results

- Try with a smaller data set where all classes are equally distribut

→ No improvement, model appears to have a strong bias towards predicting ratings in the range of 5 and 6

→ Model is not effectively distinguishing between different rating levels and tends to converge towards the middle ratings.





Source

1. https://medium.com/coinmonks/how-to-implement-a-recommendation-system-with-deep-learning-and-pytorch-2d40476590f9
2. https://course.fast.ai/
3. https://www.kaggle.com/code/hasibalmuzdadid/anime-ratings-analysis-recommender-system
4. https://www.kaggle.com/code/mejbahahammad/movie-recommender-systems-using-neural-network/notebook
5. https://github.com/ufkunmndrs/Neural-Book-Recommender-Systems-in-PyTorch
6. https://github.com/Sajid030/anime-recommendation-system/tree/master
7. https://github.com/DragoKami/Anime-Recommendation-System-Amime2022-dataset/tree/main
8. https://www.d2l.ai/chapter_recommender-systems/recsys-intro.html
9. https://pureai.substack.com/p/recommender-systems-with-pytorch
10. https://datasolut.com/was-ist-ein-empfehlungsdienst/
11. https://github.com/ufkunmndrs/Neural-Book-Recommender-Systems-in-PyTorch/blob/main/Evaluation/NeuralNetworkRecommender.ipynb
12. https://www.d2l.ai/chapter_recommender-systems/mf.html
13. https://datajobs.com/data-science-repo/Recommender-Systems-%5BNetflix%5D.pdf
14. https://developer.nvidia.com/blog/using-neural-networks-for-your-recommender-system/