Anime Recommendation System

Contents

About this project:

Building a recommendation system that will show relevant information of interest to users is more important to any business. In particular, for an online business portal, it is important to show the customers what they would probably be liking, to retain their business with you. The anime recommendation system is intended to show the top 15 anime of interest to an existing user based on their previous interests or based on other similar users and for a new user based on similar users interest. The channel with the better recommendation system will be more probable to become popular, thereby increasing the subscription count to the channel.

This project is aimed at recommending anime of interest to users, both existing and new, based on similar animes, similar user's favorite and using LightGBM Ranker.

Data

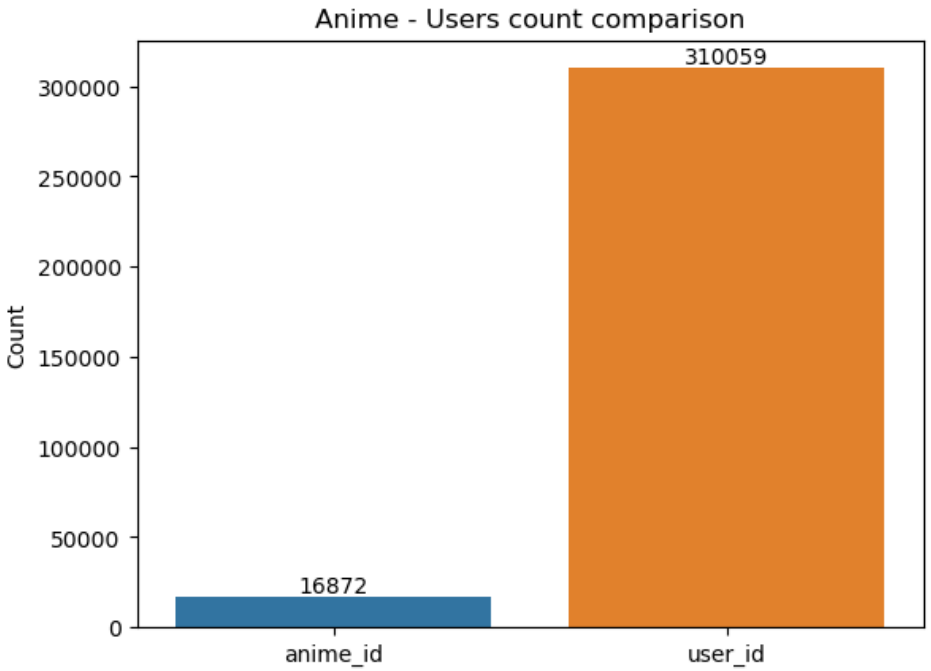
The dataset contains information about over 17000 anime and over 57M user's rating information. We have used the following files to develop the recommendation systems in this project.

Anime - Unique ID, Name of the anime in English and Japanese, Avg Score given by all the users, Genre, Type, Number of Episodes, Broadcasted date, PG rating info, its Popularity, Number of people Watched, Completed, Kept-onhold, and Dropped watching and some more.

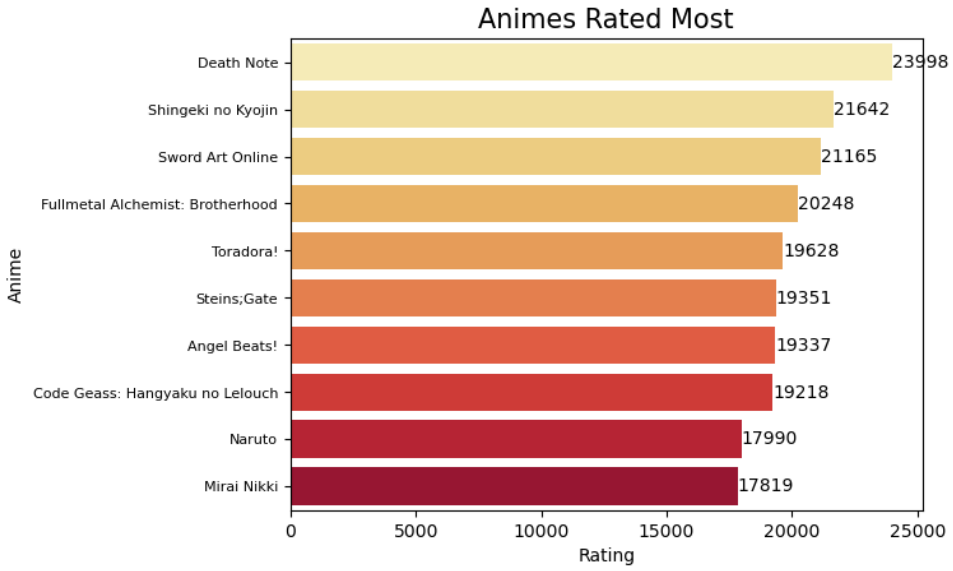
Rating - contains anime id, user id and the rating of the users who have completed watching the anime.

Data Analysis

The chart below shows the count of Anime vs Users in the database. The ratings given to these anime by the users are from 1 to 10.



The chart below gives a quick list of Anime to recommend for new users. These are the anime that have the most number of views in the database.



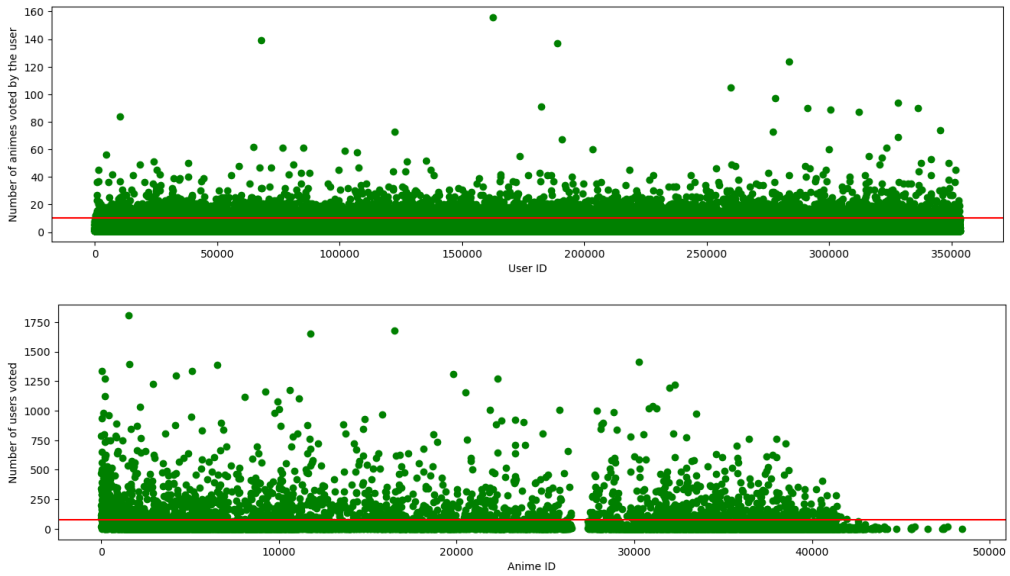
The image below shows the various genres of anime in the database. We can see that Comedy, Action and Fantasy are easily the most common genres of anime in this data.

Anime Genres - Word cloud

Model 1 - Item-based recommendation : NearestNeighbors

The NearestNeighbors algorithm is an unsupervised learner for implementing neighbor searches. The model has been built based on the cosine similarity.

We took a smaller sample to be able to keep the kernel running while creating a pivot for this dataframe. Also, we have added filters to the ratings dataset that the data points are mostly collected from very popular movies and highly engaged users. We wouldn't want movies that were rated by a small number of users because it's not credible enough. Similarly, users who have rated only a handful of movies need not be taken into account as well: To qualify a movie, atleast 75 users should have rated it. and to qualify a user's rating, the user should have rated atleast 10 movies.



Given an anime name, we have used the NearestNeighbors to recommend 10 similar anime based on its score using Cosine similarity.

In [89]:
```
get_recommendation_for('Cowboy Bebop')
```

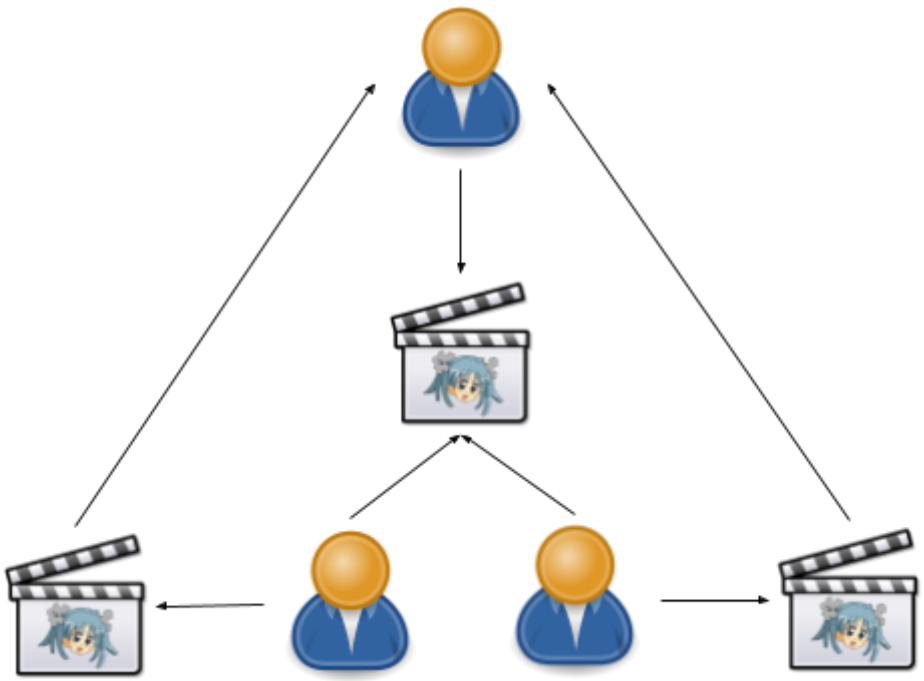
```
The overall score for the anime, Cowboy Bebop, is 8.78
```

Out[89]:

	Anime	Score	Distance
0	Diamond no Ace: Second Season	8.32	0.930311
1	Fullmetal Alchemist: Premium Collection	7.35	0.929910
2	Rozen Maiden	7.44	0.925376
3	Okusama ga Seitokaichou!+!	6.65	0.924135
4	Murder Princess	7.02	0.923775
5	ChäoS;Child	6.39	0.922198
6	Hidan no Aria	6.88	0.922051
7	Uchi no Ko no Tame naraba, Ore wa Moshikashita...	7.04	0.911138
8	Rewrite 2nd Season	7.26	0.909580

Model 2: User-based collaborative recommendation

In this model, for a given user, recommendation of top 10 anime has been done based on similar users' interests. This is a naive recommendation system made by considering the rating given by the user and the similar users for an anime.



In [91]:
```
recommendation_for_user = collab_recommendation(merged_df,6556)
```

```
Favorite animes of the user 6556 are:
Shingeki no Kyojin
Kidou Senkan Nadesico: The Prince of Darkness
Dragon Half

Recommended animes for 6556:
Shigatsu wa Kimi no Uso: Moments
Clannad
Hibike! Euphonium
Digimon Adventure
Soukyuu no Fafner: Right of Left - Single Program
Gotcha!
Guilty Crown
Nisekoi: OVA
Death Parade
Tokyo Ghoul √A
```

Model 3: LightGBMRanker

LightGBM, the Light Gradient Boosting Machine is a free and open source distributed gradient boosting framework developed by Microsoft. It is based on decision tree algorithms and used for ranking, classification and other machine learning tasks.

Learning to rank is the application of machine learning, typically supervised, semi-supervised or reinforcement learning. Training data typically consists of lists of items with some partial order specified between items in each list. This order is typically induced by giving a numerical or ordinal score for each item. The goal of constructing the ranking model is to rank new unseen lists in a similar way to rankings in the training data.

Ranking is the central part of many information retrieval systems, such as document retrieval, collaborative filtering, sentiment analysis, and online advertising.

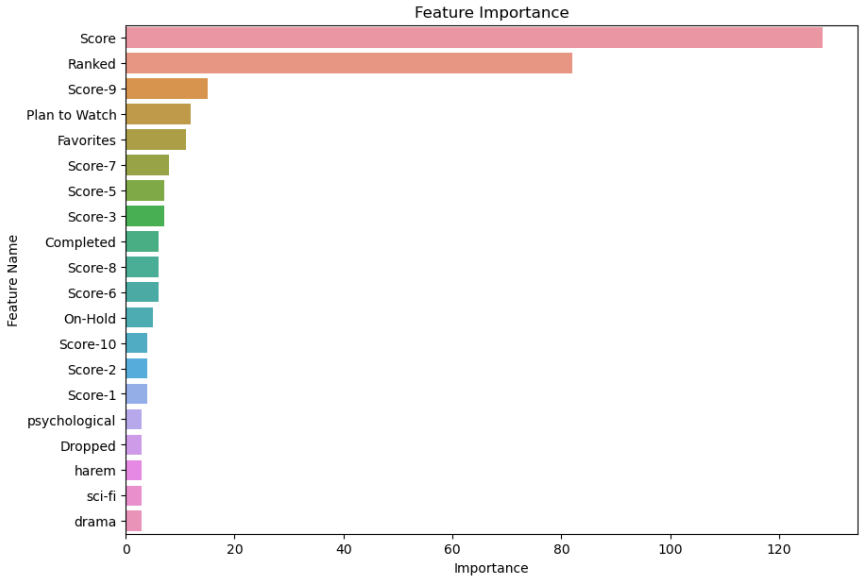
We have used Optuna, an open source hyperparameter optimization framework to automate hyperparameter search.

Features used in this model:

```
features = ['Popularity', 'Score','Ranked',
      'Members', 'Favorites', 'Watching', 'Completed', 'On-Hold', 'Dropped',
      'Plan to Watch', 'Score-10', 'Score-9', 'Score-8', 'Score-7', 'Score-6',
      'Score-5', 'Score-4', 'Score-3', 'Score-2', 'Score-1', 'action',
      'adventure', 'comedy', 'drama', 'sci-fi', 'space', 'mystery', 'shounen',
      'police', 'supernatural', 'magic', 'fantasy', 'sports', 'josei',
      'romance', 'slice of life', 'cars', 'seinen', 'horror', 'psychological',
      'thriller', 'super power', 'martial arts', 'school', 'ecchi', 'vampire',
      'military', 'historical', 'dementia', 'mecha', 'demons', 'samurai',
      'game', 'shoujo', 'harem', 'music', 'kids',
      'hentai', 'parody', 'yuri', 'yaoi','rating_count', 'rating_mean']
user_col = 'user_id'
item_col = 'anime_id'
target_col = 'rating'
```

With the best parameter values: {'lambda_l1': 3.001700605238165, 'lambda_l2': 2.2085252160803295, 'max_depth': 3, 'num_leaves': 135, 'colsample_bytree': 0.4747326915651492, 'min_child_samples': 9}, the recommendation system can rank the anime with the Normalized Discounted Cumulative Gain score of 0.9862.

Below are the Feature Importance plot where Score of an Anime, which is the average of all the user ratings, and Ranked, which is the rank given to an anime based on the Score. The features that states the watching status of the anime are also among the important features.



Here is the recommendation of 15 movies for a user taking a sample of 10000 movies.

```
n [65]: r = recommend_for_user(266729,model_with_best_params,15,10000)

---------- Recommend ----------
1:Fullmetal Alchemist:Brotherhood : 鋼の錬金術師 FULLMETAL ALCHEMIST : Action, Military, Adventure, Comedy, Drama, Magic, Fan
tasy, Shounen : 9.19
2:Hunter x Hunter : HUNTER×HUNTER（ハンター×ハンター） : Action, Adventure, Fantasy, Shounen, Super Power : 9.1
3:Gintama Season 4 : 銀魂° : Action, Comedy, Historical, Parody, Samurai, Sci-Fi, Shounen : 9.1
4:Gintama Season 2 : 銀魂' : Action, Sci-Fi, Comedy, Historical, Parody, Samurai, Shounen : 9.08
5:Attack on Titan Season 3 Part 2 : 進撃の巨人 Season3 Part.2 : Action, Drama, Fantasy, Military, Mystery, Shounen, Super Po
wer : 9.1
6:March Comes In Like A Lion 2nd Season : 3月のライオン 第2シリーズ : Drama, Game, Seinen, Slice of Life : 9.0
7:A Silent Voice : 聲の形 : Drama, School, Shounen : 9.0
8:Gintama Season 5 : 銀魂。 : Action, Comedy, Historical, Parody, Samurai, Sci-Fi, Shounen : 8.99
9:Gintama.:Silver Soul Arc - Second Half War : 銀魂. 銀／魂篇 後半戦 : Action, Sci-Fi, Comedy, Historical, Parody, Samurai, S
hounen : 8.86
10:Cowboy Bebop : カウボーイビバップ : Action, Adventure, Comedy, Drama, Sci-Fi, Space : 8.78
11:Mob Psycho 100 II : モブサイコ100 II : Action, Slice of Life, Comedy, Supernatural : 8.84
12:Descending Stories:Showa Genroku Rakugo Shinju : 昭和元禄落語心中～助六再び篇～ : Drama, Historical, Josei : 8.79
13:Made in Abyss : メイドインアビス : Sci-Fi, Adventure, Mystery, Drama, Fantasy : 8.74
14:Samurai X:Trust and Betrayal : るろうに剣心−明治剣客浪漫譚−追憶編 : Action, Historical, Drama, Romance, Martial Arts, Samur
ai, Shounen : 8.73
15:Monogatari Series:Second Season : 〈物語〉シリーズ セカンドシーズン : Mystery, Comedy, Supernatural, Romance, Vampire : 8.78
---------- Actual ----------
My Hero Academia 2 : Action, Comedy, Super Power, School, Shounen : 8.33
Perfect Blue : Dementia, Drama, Horror, Psychological : 8.48
Puella Magi Madoka Magica the Movie Part 2:Eternal : Drama, Magic, Psychological, Thriller : 8.39
Everyday Life with Monster Girls OVA : Comedy, Ecchi, Fantasy, Harem, Romance, Seinen : 7.19
Yurikuma Arashi : Psychological, Drama, Fantasy, School, Seinen, Shoujo Ai : 7.08
Unknown : Mystery, Thriller, School : 7.58
Promare : Action, Mecha, Sci-Fi, Super Power : 8.08
Angel's Egg : Fantasy, Dementia, Drama : 7.69
Ghost in the Shell : Action, Mecha, Police, Psychological, Sci-Fi, Seinen : 8.29
Puella Magi Madoka Magica the Movie Part 1:Beginnings : Psychological, Drama, Magic, Thriller : 8.23
```

Further Study

Using the Synopsis data which contains a brief description of each anime in the database and the other text features such as age-rating, producers, licensors and the source of anime, NLP can be applied along withTF-IDF for ranking the anime.

Data Citation

The data has been downloaded from the Kaggle website. Link to Download.
The data files have been scraped from the MyAnimeList website. Link to Data Source.