**Anime Recommendations - Data Visualization and Presentation**

Jessica Nguyen  - 1661857

Nathan Nguyen - 1629979

As someone who watches a lot of anime, there has been recently a lot of anime that have been released in the past decade and there is no way to find out if a show is worth watching besides to go into it yourself. What this project aims to do is to use your watched shows and use that data to recommend shows that the algorithm thinks that you will enjoy. As for the actual data, the algorithm gathers data from the myanimelist.com database which keeps track of what anime the users have watched and what rating was given. So using that data it can recommend shows based on how many people have given a rating to the show as well as the rating itself. So in order to use the predictor, all you would need to do is to input your MAL user id where the cell with the number id is in the modeling tab and it will use your user data to create recommendations. There are 2 separate models that were used to predict a show which uses 2 different statistical numbers that will determine whether or not a certain show is a good fit. SVD basically eliminates the variables that are deemed as unnecessary while Pearson's R uses the correlation, otherwise known as the relation between 2 shows and see how close they are to each other and determine whether or not it is a good fit based on that as well as the rating of the anime itself. The actual result will be a list of anime that is sorted by the specific statistic value which will check for if the rating is at least a 7 or above which is the MAL average rating.

Although both methods give you a list of shows to watch, SVD bases its predictions off of your actual data and the anime you watched, while the other one predicts based off of the show itself. For example, you would input your user id for SVD and you would get user based recommendations while pearson's R actually takes the anime's name as input. So when you input an anime like '91 Days', you would get shows that a lot of other users have also watched and enjoyed. You also get a correlation number to see how closely related, or how much percentage of people have actually watched it as well as the average rating of said anime. There is also another parameter that has a minimum number of ratings, so that you filter out the anime that have a high rating but less than a certain number of people that have actually watched. Since both of these methods use different ways to predict shows, the best use for these two would be to put them in tandem together in order to optimize predictions and searches. You could use the user predictions and use those predictions to make other

predictions and can put them all into a list just like you would Netflix or YouTube. The main use of these predictions would be to place them into a website and create a "You may like this" tab where you can see the shows that the website thinks you will enjoy.

```python
user_66789['score'] = user_66789['anime_id'].apply(lambda x: svd.predict(66789, x).est)

user_66789 = user_66789.drop(['anime_id', 'index'], axis = 1)

user_66789 = user_66789.sort_values('members', ascending = False)
user_66789.head()
```

| | name | genre | type | episodes | rating | members | rating_rounded | length | popularity | score |
|---|---|---|---|---|---|---|---|---|---|---|
| 38 | Death Note | Mystery, Police, Psychological, Supernatural, ... | TV | 37.0 | 8.71 | 1013917 | 9.0 | extra long | true | 5.0 |
| 82 | Shingeki no Kyojin | Action, Drama, Fantasy, Shounen, Super Power | TV | 25.0 | 8.54 | 896229 | 9.0 | long | true | 5.0 |
| 712 | Sword Art Online | Action, Adventure, Fantasy, Game, Romance | TV | 25.0 | 7.83 | 893100 | 8.0 | long | true | 5.0 |
| 1 | Fullmetal Alchemist: Brotherhood | Action, Adventure, Drama, Fantasy, Magic, Mili... | TV | 64.0 | 9.26 | 793665 | 9.0 | extra long | true | 5.0 |
| 154 | Angel Beats! | Action, Comedy, Drama, School, Supernatural | TV | 13.0 | 8.39 | 717796 | 8.0 | seasonal | true | 5.0 |

SVD will return this graph where you can see the predictions and you can sort them based on any specific column given that the value is changed within the code. If this were designed for a website or MAL itself, then you would be able to make the columns clickable to sort them automatically, though that is outside the scope of this report. Since the prediction cannot tell if the prediction it gave you was correct, all that can be done to contribute to the algorithm is for you to give the show that you watched a rating, and it will consider that into your prediction. Pearson's R method will select an anime and will create a list predicting anime that other people have also watched which is also popular with a minimum number of reviews and will generally have ratings around 7 or higher.

```
recommendation('Fullmetal Alchemist: Brotherhood', 1000)

C:\Users\nnath\AppData\Local\Programs\Python\Python310\lib\site-packages\numpy\lib\function_base.py:2683:
  c = cov(x, y, rowvar, dtype=dtype)
C:\Users\nnath\AppData\Local\Programs\Python\Python310\lib\site-packages\numpy\lib\function_base.py:2542:
  c *= np.true_divide(1, fact)

--Recommendations based on Fullmetal Alchemist: Brotherhood--
PearsonR                                                          name  count     mean
1.000000                             Fullmetal Alchemist: Brotherhood  13449 8.869433
0.561158    Haikyuu!!: Karasuno Koukou VS Shiratorizawa Gakuen Koukou   5790 8.439724
0.342475                                                     Haikyuu!!   7485 8.305945
0.329705                                         Ouran Koukou Host Club 13679 8.807077
0.322020          Code Geass: Boukoku no Akito 2 - Hikisakareshi Yokuryuu   1730 8.282081
0.315074            Detective Conan Movie 14: The Lost Ship in the Sky  11118 8.107483
0.314993                                                Bakemonogatari   1250 7.901600
0.311537                                            Hunter x Hunter (2011)   9385 8.419393
0.309657                                        Ping Pong The Animation   8296 8.145251
0.306819                               Layton Kyouju to Eien no Utahime   1135 8.186784
0.306501                                                     Soul Eater   1320 8.331818
0.297115                            Mahou Shoujo Lyrical Nanoha A&#039;s   3006 8.018297
0.296562                       Mobile Suit Gundam: Iron-Blooded Orphans   6664 8.634154
0.292194                              Nogizaka Haruka no Himitsu: Finale   5756 8.200313
0.290768 Sengoku Basara: Setonai no Gekitotsu! Hi wo Fuku Umi no Daiyousai - Fugaku!!   9710 8.660659
```

In a business, anime is considerably less marketable than an actual domestic show or movie, but in the case of our model, it does not cover just anime. The model can also work on shows and movies from other websites given the data, which means that it can be applied in a multitude of websites and other backend assets. It does not even have to limit to anime either, because the concept itself can be used for other types of media like music or a book. For example, you can apply it to a website to recommend some books based on the ones you have already read. The only required data that is necessary is information on the book itself and a metric to measure one's personal enjoyment to the material, something like bookwalker. Another application would be food, as there is a similar system in place when it comes to apps like yelp and google reviews.

Even though these systems are viable and widely applied, there are also a number of pitfalls that could harm the system. The biggest thing that this model suffers from is negative bias which means that the entire model will work only if the actual person has given a rating, otherwise they will get recommendations that are not only inaccurate but will also continue to feed into the loop of not wanting to add data to the model through watched shows. A good example of this would be youtube where the algorithm will recommend videos that you do not want to watch when you decide not to give anything a like or a comment. They can only base it off of your watch time and the channels that you are subscribed to which is something that could easily fill up your home menu with junk which would not be very healthy. Another pitfall would be the data itself, where there are a lot of unnecessary variables as well as a lot of overlap that could result in completely different anime or other media that end up on your recommendations. For example, an anime might have the genres of music, drama, and comedy. These three genres are in very different fields and it is completely possible for an anime from each genre to end up in your recommendations. One way to prevent that would be to have a more concise genre tag for each anime to ensure that overlap does not happen. Another pitfall in the data would be the community based ratings. The average ratings on MAL are around 7 out of 10 which is not around the middle being 5 meaning that the majority of the community do not use the 10 point scale liberally, causing the average to be higher than 5. However, there is also a positive to that in

which if you find an anime that has a rating below 6, then you will know that that anime is significantly worse than normal.

Since the model and the concept of recommendations itself is completely based on the users and the anime itself. You cannot tell if the prediction will be good, so the only thing that can be done is to give the anime a rating, not even a review and it will add your user data to the database to be used in the algorithm. Although anime is mainly broadcasted on TV and sold via blu-ray, streaming services like netflix and crunchyroll have attempted to improve user content and preferences through things like surveys and implementing systems that let them know that users like certain anime which helps them create a better platform, hence why they are so popular. Some third party streaming websites do not even have that option, so a percentage of viewers are lost due to lack of ability to rate a show. Since there are so many places where you can stream anime from in the west, having a universal website like MAL is the first of its type, which is why everyone uses it as the main platform for rating anime.

Since anime is starting to become mainstream, that means that there are more and more platforms that are starting to invest in anime in the west. Netflix and crunchyroll now release simulcast, meaning the same broadcast from television to a different channel, on their platforms which adds a lot more potential data into the MAL database which will create a better system in the long run. Since the concept itself is easy to understand, it would be understood quickly whenever it gets put into work in the background where people will not notice. The method of collaborative filtering, the use of data from other people/objects, is an easy enough concept to implement, so each website has their own type of recommendation system but the main difference between each one is that they use a different algorithm for their own websites, so the strength cannot be measured. That ensures that their method does not get stolen or reproduced, but it also means that whether or not it will be widely used or appreciated will be up to the community that analyzes this type of thing.

Going back to anime itself, the actual list of anime on the MAL database is very massive, so there are a lot of data points that can affect the overall prediction of the model. Take for example you predict recommendations based on an old anime like One Piece or Legend of the Galactic Heroes. The result will be a mix of newer anime that

have a lot of decent ratings and there is also a chance that you will end up with older anime that may or may not match what genre/topic the recommendation is trying to present to you. Anime may also have multiple series with a number of sequels like Pokemon or Gundam which could possibly flood the predictions which is not a bad thing in itself, but if you are looking for variety or something new to watch, then it is troublesome and would require more processing to filter out properly. Cases like these are the reason that there is a trending tab that shows the currently popular anime or media at that time. A way to solve this problem would be to add a variable called age where it would keep track of the airing date of the anime and will have a sorted list that will filter out anime past a certain age. That way you wouldn't get recommended anime from the 90's based on something recent like 2016.

Lastly a recommendation system is good for a business due to the amount of user data that can be obtained from user reviews. They can obtain the actual number of users that contribute to their data as well as the actual content of the reviews itself. This is a viewable metric on how popular a business's platform can be and is a viable measure on how well a given anime can be doing. If there is an anime that is not doing well rating wise and has a low amount of people that gave it a rating, then they can remove the anime from the website and replace it with another one which would free up potential server space if needed. In a business model, having the optimal spending and costs will make the difference when it comes to turning for a profit or being able to stay afloat. A recommendation model is a good tool to view what the users think and can give a tell to the developers on what to improve with their website or software. This is not only applied to the number rating, but the word reviews on MAL that are on the anime profile page is an actual word from the community that can be used if applicable, but for the sake of software the numbers will provide most of the needed knowledge to be able to predict shows. Although there are pitfalls in the system itself that cannot be fixed with programming or processing alone, there are also a number of plusses that make the model robust and widely applicable. The users that make the effort to create an account on MAL are the ones that are more prideful and have a love for anime, or at least enough to give a rating and contribute to the data which makes the model constantly evolving which is something that will benefit a business over time.