# Introduction

The chosen topic for the coursework is an anime recommendation system. Anime is hand-drawn or computer animation videos originating from Japan. Anime has been growing as a digital entertainment media not only in Japan but throughout the globe. According to a market analysis report by Grand View Research based on historical data from 2018 to 2019, the global anime market size was valued at USD 23.56 billion in 2020 and is expected to expand at a compound annual growth rate (CAGR) of 9.5% over the forecast period (2021-2028) (Grand View Research, 2021). Taking reference of the popular website named My Anime List keeping track of and rating most of the anime produced till date, there have been over 10,000 animes with multiple episodes, covering over 20 genres. However, that is not the end of anime production. As time moves forward, more and more anime will be added to the pool of choices.

As such, it can be quite a daunting experience to make a choice. For the service provider as well, having a group of consumers wasting their time trying make a choice instead of using that time to consume the services provided turns out to be costly. The problem would be resolved if there was a system in place to hand out personalized recommendations to each user based on the vast amount of user data collected by the service providers. Recommendation systems do just that. A recommendation system helps users to determine which item that might be interested to use. In the scenario described above, a recommendation system enables the consumers to make quick decisions thus minimizing the wasted time. This also means less operating expense for the service provider.

For this project, the dataset to use is extracted from the My Anime List website and available on the data science site Kaggle. The dataset contains data about the site users, their reviews as well as data about the animes. The users give a rating score to an anime from a scale of one to ten. For this project, we propose the use of a recommendation system using collaborative filtering. This approach is proposed to measure the similarity between shows, users and helps to predict which anime a user may enjoy.

# Background

Recommendation system was initially mentioned in a technical report by Jussi Karlgren at Columbia University as "digital bookshelf" in 1990. According to Jussi, the paper was rejected at the 1990 INTERACT due to user privacy and integrity issues (Karlgren, 2017). He later completed his work in 1994 when he worked at SICS (Karlgren, 1994). Since then, several types of recommendation systems have come to light. Some of the popular recommendation systems are:

1. Collaborative recommender systems
2. Content-based recommender systems
3. Hybrid recommender systems
4. Session based recommender systems

Considering the problem at hand and the available dataset, collaborative recommender system is the most suitable. Collaborative recommendation system works around a core assumption that users who liked something in the past will always like similar kinds of things. According to google, collaborative filtering uses similarities between users and items simultaneously to provide recommendations (Google, 2021).

According to Shuyu Luo, a collaborative recommender system can be either user based or item based. The user-based system calculates the similarity of a user with all other users. The top n number of similar users are selected and their weighted average is taken, where the user similarity is taken as the weight. While rating an item, some people tend to give high ratings in general, while some give rather lower ratings in general. In order to reduce this bias, each user’s average rating of all items is subtracted while calculating the weighted average, and added back for its target user. (Luo, 2018)

Some popular algorithms for calculating similarity are:

1. Euclidian Distance
2. Manhattan Distance
3. Pearson coefficient
4. Cosine Similarity

Related Works:

Many types of research about recommendations have been conducted using collaborative, content-based and hybrid both of them. Dev, in his research, uses a big data platform with MapReduce to develop a collaborative reconciliation system [3]. This system is able to reduce costs by removing the redundant computational process. Kumar [4] proposes a film recommendation system using collaborative filters that focus on the ratings given by users to provide recommendations. This allows the user to choose his choice from a given set of attributes and then recommends him to list the films based on the cumulative weights of the different attributes use algorithms k-means to recommend films based on the highest order. Shahjalal et al [5] implement a reconciliation system that combines both collaborative and userbased filtering approaches based on items. Researchers use the KUNN algorithm, a new algorithm for collaborative screening of one class, a setting that includes many applications

# Proposed Solution

The proposed solution to the problem is a collaborative recommendation System. The proposed solution can be broken down to:

## Problem formulation

To clearly understand the problem, a table of anime rating is extracted from the given dataset different anime names as the rows and their corresponding users as the columns. For example:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Animes | User A | User B | User C | User D | User E |
| X | 5 | ? | 6 | 7 | 10 |
| Y | 6 | 8 | ? | 5 | 5 |
| Z | ? | 6 | 1 | 2 | ? |

Table 1: Anime ratings by users

In the table above, user A gives ratings of 5 and 6 to animes X and Y. This means that the user enjoyed anime Y more than anime X. The more the rating number, the more the user has enjoyed the anime. The user does not have a rating for anime Z. This means that the user has not watched anime Z. The maximum rating a user can give is 10 and the minimum is 1.