REVIEW-1

COURSE: Foundation of data analytics

SLOT: F2

FACULTY: VERGIN RAJA SAROBIN M

Project title: Anime recommendation

Team members:

VENKAT REDDY-19MIA1104

MOUNIKA SAI-19MIS1013

GAYATHRI-19MIS1035

SIRI MAHALAKSHMI-19MIS1044

MANASA-19MIS1063

MOHAN-19MIS1200

Paper Title:

Recommendation System for Anime Using Machine Learning Algorithms.

Problem Statement:

The existing model's potential is restricted due to the lack of content-based filtering technique. Our goal is to fill the gap in the existing research and construct a recommendation system that proposes anime to viewers based on their preferences and needs, resulting in a better streaming experience that increases income and time spent on a website.

Dataset Description:

We worked with **two** datasets.

The first is **Anime**, while the second is **Rating**.

These two datasets were collected from the website, Kaggle.

This data collection contains customer choices for 12,294 anime from a total of 73,516 users. The ratings in this data set are a collection of them.

The first anime dataset has 12294 rows and 7 columns. The second rating dataset has 7813737 rows and 3 columns. Each user may rate anime and add it to their finished list.

Anime.csv

anime_id - an anime's unique identifier.

name - the anime's entire name.

genre - list of anime genres separated with commas

episodes - number of episodes in a show (1 if it's a movie).

type - film, television etc.

members - the number of people in this anime's "group" from the community.

rating - anime's average rating out of ten.

Rating.csv

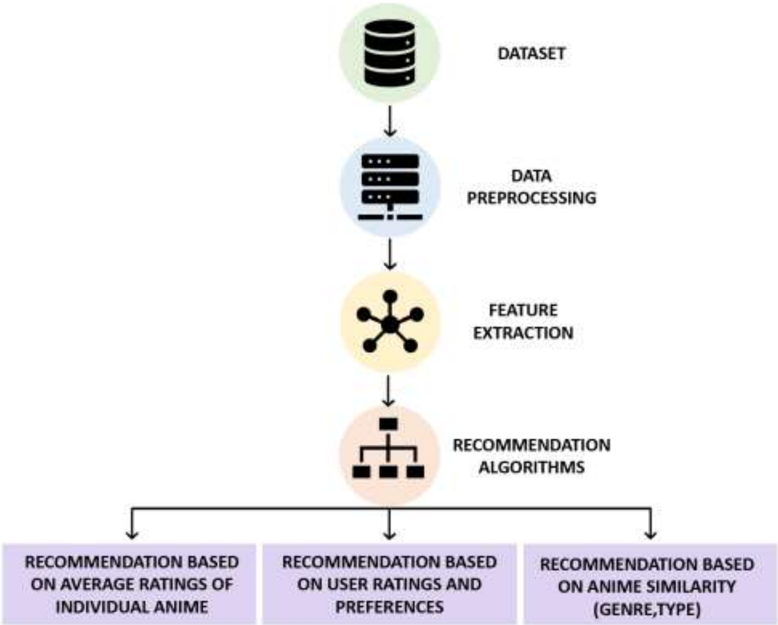
user_id - a randomly created non-identifiable user id.

anime_id - identifies which anime has been rated by a specific user.

rating - the user's rating (-1 if the user watched but didn't give it a rating).

Implementation Details:

The system can be implemented in multiple ways. We intend to address Content Based Filtering, Popularity Based Recommendation System, Collaborative Filtering using KNN and Collaborative Filtering using SVD approaches in our research work.

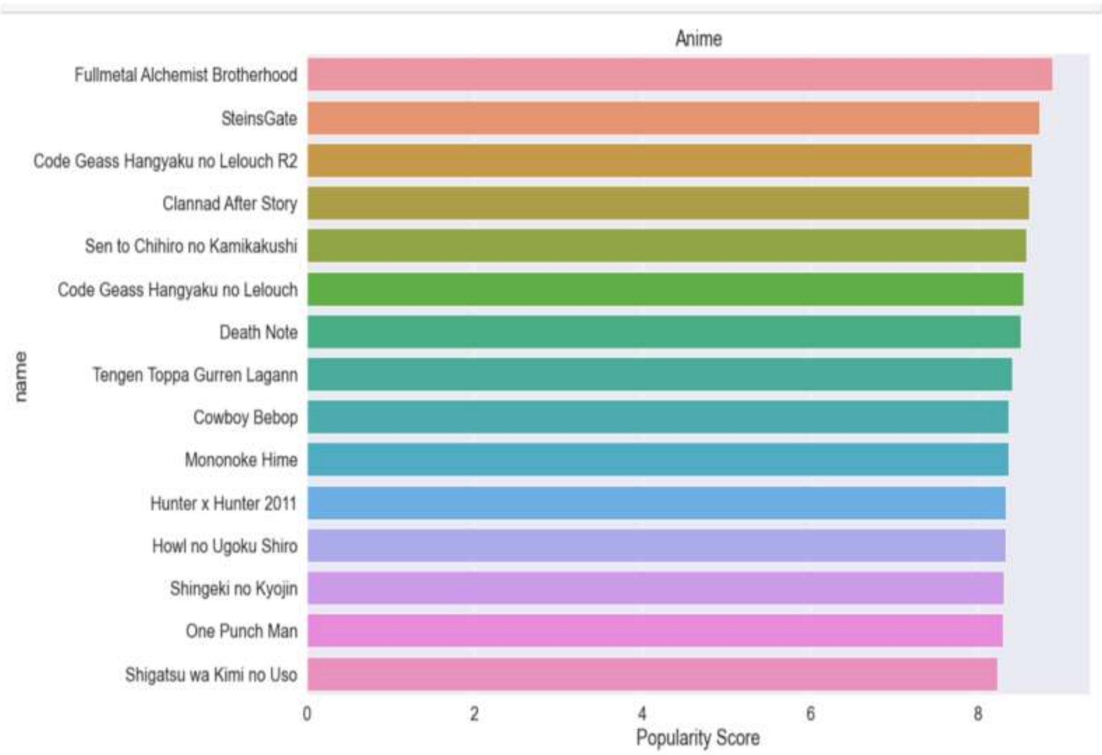


Results:

Popularity Based Algorithm

The picture below shows the top 15 most popular shows based on their average ratings given by the users. In the graph given below, the Y-axis represents different Animes and on the X-axis, the corresponding popularity score is given from 0-10.

Based on our recommendation system, we have obtained the top 15 most popular anime, according to people's ratings.



Collaborative Based Filtering:

We obtained the following result using the KNN recommender. The below table represents the top 5 recommended anime (similar genre/ratings), based on a user given Anime i.e. Anime 0. This is a really effective way to club Anime based on specific taste.

Table 3. Top 5 recommended anime based on Anime 0

Anime0	Anime1	Anime2	Anime3	Anime4	Anime5
Choisuji	Princess Tutu Recaps	Paizuri Cheerleader vs Sakunyuu Ouendan	PeroPero ☆Teacher	Please ♥♥♥♥ Me	Osaru no Sankichi Boukuusen
Gekiganger 3 The Movie	Ookiku Furikabutte Special	009 ReCyborg	Gundam Build Fighters Try Island Wars	Avengers Confidential Black Widow to Punisher	XMen
91	3x3 Eyes Seima Densetsu	3x3 Eyes	Maze☆Bakunetsu Jikuu TV	Hana yori Dango	Koihime† Musou OVA Omake
Candy Candy	Omae Umasou da na	Candy Candy Haru no Yobigoe	Okusama wa Joshikousei	Candy Candy Movie	Ashita no Joe Pilots
07Ghost	Vampire Knight Guilty	Pandora Hearts	Nabari no Ou	Vampire Knight	Dantalian no Shoka
11eyes	IS Infinite Stratos	Dragon Crisis	Mayo Chiki	IS Infinite Stratos 2	11eyes Momoiro Genmutan
11eyes Momoiro Genmutan	11eyes	Boku wa Tomodachi ga Sukunai Episode 0	IS Infinite Stratos	Asobi ni Iku yo	IS Infinite Stratos 2

PAGE TITLE: ANIME RECOMMENDATION SYSTEM

PROBLEM STATEMENT:

A hand-drawn computer animation with Japanese roots, anime has gained a cult following all over the world. With a comparable surge of generational recognition, anime and manga have long dominated Japanese culture. The popularity of anime and its comic book counterpart manga has increased dramatically in the UK during the past several years. And consequently the West. One of the reasons why anime has such a unique potential to grow with viewers is without a doubt. It has withstood the test of time and improved in quality everywhere. Although anime is popular in the UK and There are still a few people in Western countries who are unaware of what anime is. This essay seeks to create a Recommendation Offers recommendations for the most popular anime.

DATASET DESCRIPTION :

The information was acquired from Kaggle.com and includes ratings and information from 73,516 individuals. The range of scores or ratings is one to ten, with ten being the most efficient. A score of -1 means the user did not give the item a rating.

Anime.csv

anime_id - myanimelist.net's unique id identifying an anime.

name - full name of anime.

genre - comma separated list of genres for this anime.

type - movie, TV, OVA, etc.

episodes - how many episodes in this show. (1 if movie).

rating - average rating out of 10 for this anime.

members - number of community members that are in this anime's

"group".

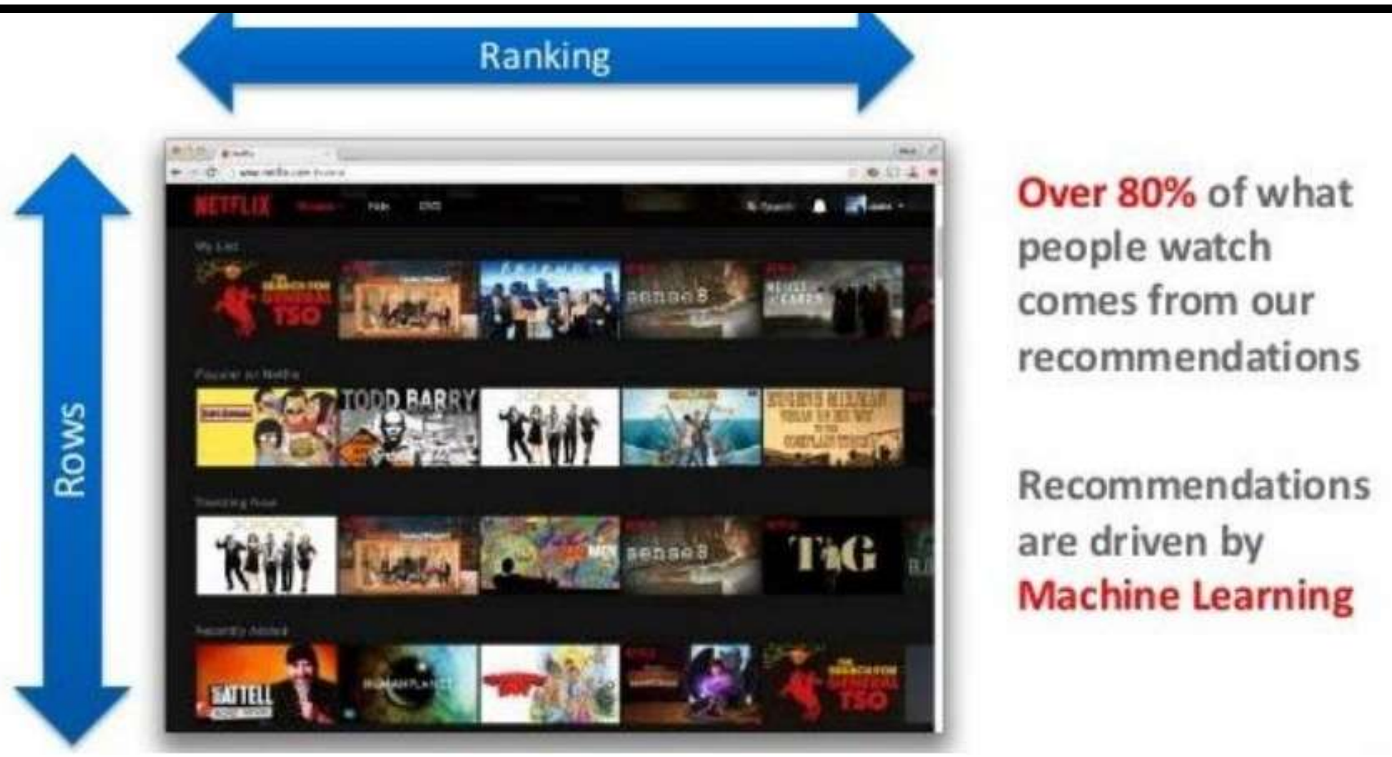
Rating.csv

user_id - non identifiable randomly generated user id.

anime_id - the anime that this user has rated.

rating - rating out of 10 this user has assigned (-1 if the user watched it but didn't assign a rating).

IMPLEMENTATION DETAILS :



Companies like Netflix, Amazon, and YouTube, among others, use the recommendation system frequently because it makes it easier for customers to find the movies they want. The recommendation system analyses the user's past viewing habits or viewing history and then recommends a movie to the user based on its findings. Numerous recommendation systems have been thoroughly investigated and shown to be beneficial to online businesses and customers. In fact, Netflix gave a developer team a $1 million reward in 2009 for coming up with a method that improved the company's recommendations by 15% in terms of accuracy. In this paper, the recommendation system aims to obtain an anime rating. It employs algorithms for collaborative filtering systems such To provide users with an accurate recommendation, KNN and SVD are used.

Collaborative Filtering

This method creates a model of the user based on past behavior. Users can have viewed movies, bought things, or provided ratings of previous things. This model predicts the product or a rating for the product in which a user may be interested. KNN and Singular value decomposition are employed as a collaborative filtering technique in based on an association between the user and the object. technique used in recommender systems.

IV.KNN

Users' actions are used by collaborative filtering systems to suggest additional episodes of their favorite anime to In this paper, the recommendation system aims to obtain an anime rating. It employs algorithms for collaborative filtering systems such. To provide users with an accurate recommendation, KNN and SVD are used.

Determines the “distance” between the target anime with the highest rating and each anime in the database, ranks those distances, 4.1. KNN's operation

KNN makes no assumptions about the distribution of the underlying data based on item feature similarity. how it determines the "distance" between the target anime and every other anime in the database, ranks those distances, and finally selects the anime with the lowest distance. Offers recommendations for the most popular anime.

V.SVD

Single value decomposition (SVD) is used to extract characteristics and correlations from the user-item matrix. Singular Value Decomposition (SVD) would concentrate on the latent components for the Singular Value Decomposition in the instance of several anime categories.(SVD) strategy, for instance.

Working of the SVD

A linear algebraic technique called the singular value decomposition (SVD) is used in machine learning to reduce Dimensionality. The SVD technique reduces the space dimension of a dataset from N-dimension to K-dimension (where K*N). Decreases a dataset’s total number of features. To improve user experience, a collaborative filtering-based SVD recommender system Experience. In an array structure, users are represented by rows, and items by columns. The ratings that make up this matrix include Users

This matrix is factorized using the singular value decomposition method. It determines matrix factors from a factorization of a High-level matrix for user-item ratings.

The task of recommending anime uses singular value decomposition (SVD) based on collaborative filtering. Python implementation. The recommendation method makes use of an anime rating dataset. The selection of this dataset was made because As the primary focus of this study is on SVD and recommender systems, does not require any preprocessing. It trains the data and lowers the error between the user value and the top recommendations in order to predict recommendations with accuracy.

The recommendation system bases its calculations on an average rating method and determines an anime’s top rating using the formula above from user ratings. In response to user requests, it also lists the most well-known K-anime. The SVD’s output is also shown.

RESULTS :

KNN and SVD both have low RMSE values and are very effective, and their results are described in this section along with those of other techniques. Consequently, I recommend anime KNN and SVD are both used in the system’s implementation, along with other recommendation algorithms including Clustering and hybrid Systems that use recommendations, content, etc.,

Utilizing datasets and algorithms, recommender systems create new possibilities for finding personalized information. It also enables users to access goods and services that are not immediately available to users on the system, which helps to relieve the issue of information overload that is a very regular occurrence with information retrieval systems. The two were covered in this project. Strategies and outlined their advantages and disadvantages. Several learning methods were utilized to create the To assess the effectiveness and quality of the algorithms mentioned, evaluation measures and recommendation models were employed.

PAGE TITLE : A DEEP LEARNING RECOMMENDER SYSTEM FOR ANIME

PROBLEM STATEMENT :

This Recommender system can help to generate personalized and accurate anime recommendations for users on platforms. This research paper will help to build a recommender system with a Deep Learning collaborative filtering-based model which provides anime recommendations. This research paper Attempts to solve the problems in conventional Recommender Systems such as the data-sparsity and cold-start.

DATASET DESCRIPTION :

The anime dataset used in the research is obtained from the open-source platform Kaggle. This dataset in turn was obtained by making HTTP requests to the public APIs of the MyAnimeList website . MyAnimeList is a comprehensive and possibly the largest database for all the related information on the anime tv shows and movies available today. The dataset has five files anime.csv, anime with synopsis.csv, animelist.csv, rating complete.csv, watching status.csv with various attributes.

IMPLEMENTATION DETAILS :

Google collab was used in the research paper to implement the proposed model. The Tensor Processing Units (TPUs) present in Google Collab is used as the runtime. Python 3.7 is the programming language used for the implementation of the model.The proposed recommender system of Model-based Collaborative Filtering is evaluated on the anime dataset using two separate activation functions and a different number of epochs. This helps us better understand if they impact the quality of anime recommendations generated and if there is a change in the performance of the model indicative in its evaluation metrics

– MSE and MAE. The first experiment where the Sigmoid activation function is used with 20 epochs. The second experiment where the ReLU activation function with 50 epochs.

Dataset was split into Training and testing sets and do training and testing. Then, The proposed model in this research has been tested for the anime dataset by generating the anime recommendations by evaluation process.

RESULTS :

The first and second experiments results in MSE of 0.07 and MAE of 0.1909,MSE value of 0.0926 and MAE value of 0.2433 respectively. this research which provides anime recommendations that are highly relevant to the users' likes and interests. One of the main objectives while implementing the Model-based Collaborative Filtering using Deep Learning for recommending anime was to address the problem of data sparsity and cold-start.

Research paper title: RikoNet: A Novel Anime Recommendation Engine

Problem Statement: Today, anime is very popular, especially among the younger generations. There are numerous show genres available, which is drawing in a rising number of viewers. In this effort, they have created a novel hybrid recommendation system that might serve as both a way to discover new anime genres and titles as well as a mechanism for making recommendations. Deep autoencoders are used in the proposed technique to predict ratings and produce embeddings. The embeddings of the anime titles were then used to create clusters. These clusters make up the search space for anime with similarities and are employed to discover anime that are comparable to the ones the user likes and dislikes.

Dataset: The three categories of entities that make up the collected data are detailed below.

1. Anime Titles: This dataset includes a list of anime with both English and Japanese names, title synonyms, genre, studio, producer, duration, rating, score, airing date, episodes, source (manga, light novel, etc.), and other key information about each anime, providing enough detail on all the anime released to date.

2. Users: This dataset includes data on users that watch anime, such as username, join date (the date of registration), last seen date (the date of birth), gender, location, and a variety of aggregated variables from their anime lists (watch lists, also called User Lists).

3. User Lists: This dataset contains information on the individual anime watching experiences of distinct users. Each entry includes the user's username, the anime's ID, a rating out of 10, the status (watched, dropped, planned to watch, etc.), and the time the entry was last updated.

Implementation:

They implemented using hybrid filter, Because it integrates the outcomes of two distinct processes, namely collaborative filtering and content-based filtering, the Hybrid Filter is given its name. In our system, the Primary Autoencoder realises the former, and the Clustering process realises the latter. The filter's operation is outlined below. Let's call the group of anticipated ratings produced by the Primary Autoencoder for all the anime films that the user hasn't yet given a rating S. Depending on the anime titles' rating, the clusters may be their own or logically opposite ones. To create this distinction, we employ a threshold value. Let Union be C for them. The anime titles in the resulting set are arranged in decreasing order of their anticipated ratings using the intersection of the sets S and C. The top 10 of them are ranked and displayed in the similar anime category. We select the top 10 anime films from the final S C set and arrange them in decreasing order for the Anime You May Like category. The user is given recommendations for these 2 lists.

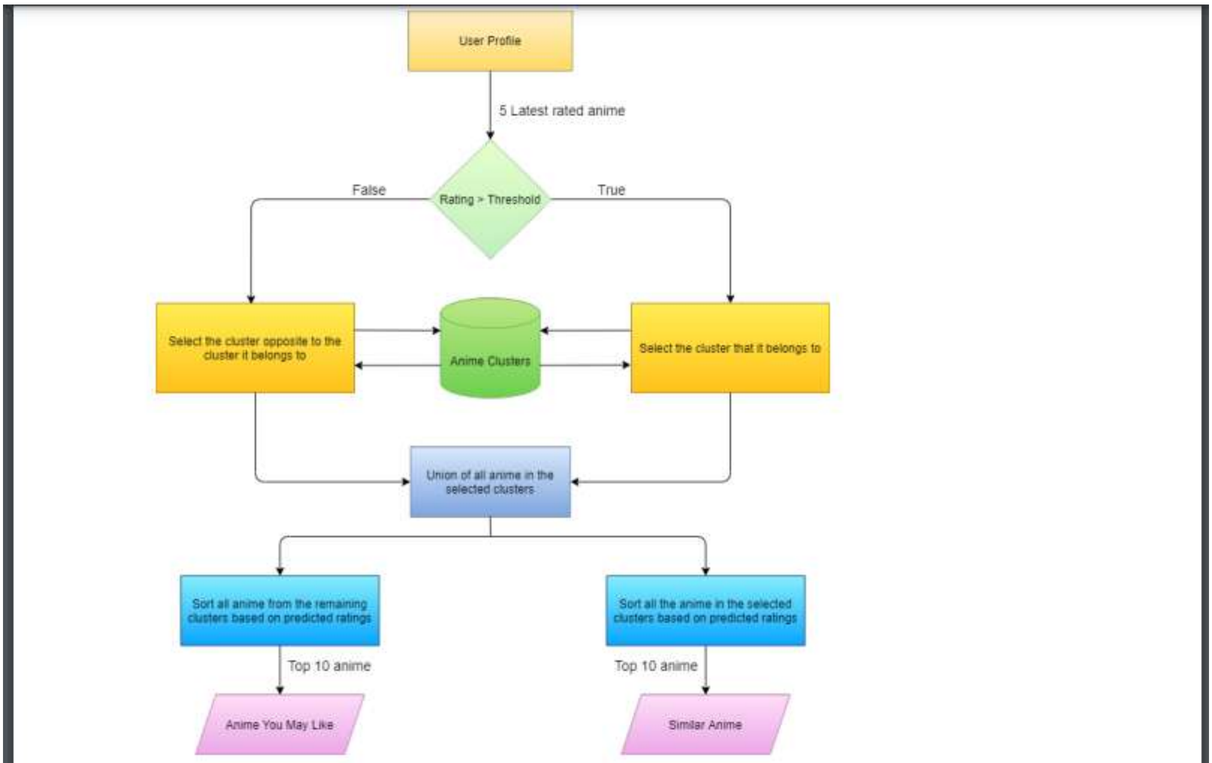
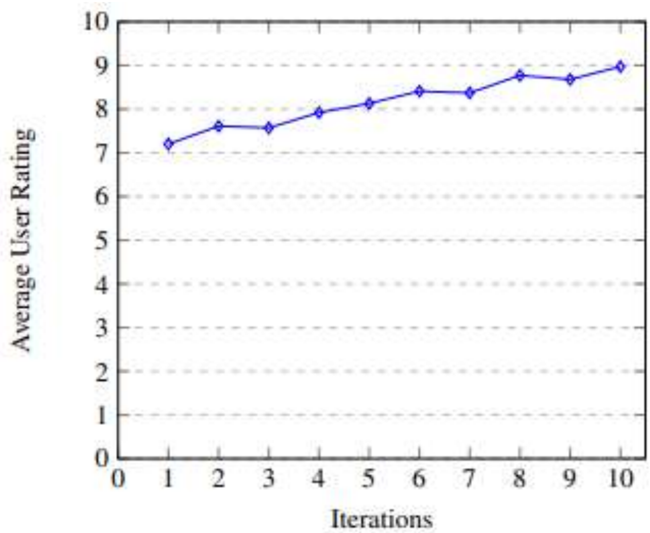


Fig. 2: Novel Hybrid Filter

Results:

As a user makes more and more use of the system, his output of ratings grows with time. The user's profile is updated with these ratings, which are then used to enhance the forecasts. As a result, it is anticipated that the recommendations will also be of higher quality. The system may also adjust to changes in the users' preferences, tailoring the recommendations to the specific user. The rating given by users to the recommendation lists ought to reflect this. The average of the total ratings from test takers is depicted in the graph up above. the vast majority of customers ran the system for ten iterations (successive recommendations). As more lists are worked through by the users, the score steadily rises from 7.2 to 8.97, demonstrating the validity of their system.



PAGE TITLE : AN INTELLIGENT ANALYTICAL FRAMEWORK FOR ANIME RECOMMENDATION AND PERSONALIZATION USING AI AND BIG DATA ANALYSIS

PROBLEM STATEMENT :

The popularity of anime has steadily increased in recent years. Like with other media, viewers frequently struggle with the urgent question of “What do I watch next?” The learning curve to efficiently use the present solution is too steep, according to our study, which looked carefully at the current approach to fixing this problem. We created a programme to guarantee simpler solutions to the problem. The program uses a Python-based machine learning algorithm from Scikit- Learn and data from My Animelist to create an accurate model that delivers what consumers want, good recommendations . We also conducted numerous iterations of various experiments to examine how the accuracy changed when other conditions were applied. Through these tests, we have successfully created a reliable Support vector machine

DATASET DESCRIPTION :

Dataset:We had to restructure the data so that each genre was distinct and assigned an integer value, as seen in the illustration below [6]. Multiple columns had to be created when we reorganised the data so that Excel could periodically check if it contained the string it was looking for.

If an anime has a romance tag, it would put a 1 and a 0 next to it. Data would appear

Like in the instance below.

Anime.csv

anime_id - myanimelist.net's unique id identifying an anime.

name - full name of anime.

genre - comma separated list of genres for this anime.

type - movie, TV, OVA, etc.

episodes - how many episodes in this show. (1 if movie).

rating - average rating out of 10 for this anime.

members - number of community members that are in this anime's

"group".

Rating.csv

user_id - non identifiable randomly generated user id.

anime_id - the anime that this user has rated.

rating - rating out of 10 this user has assigned (-1 if the user watched it but didn't assign a rating).

IMPLEMENTATION DETAILS :

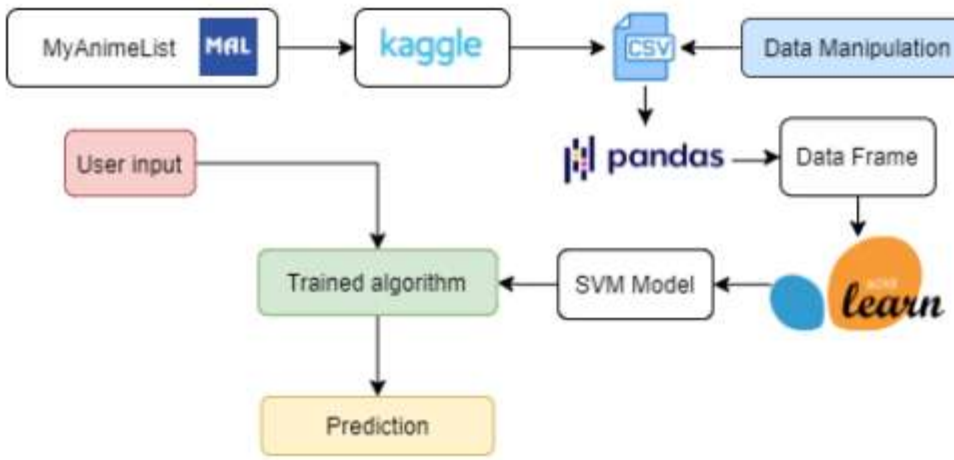


Figure 1. Overview of the project

We had to first gather data to train the machine learning algorithm before we could construct this application. Because My Anime List (MAL) has one of the largest databases and communities for anime, manga, and light novels, we choose to gather information from it. The machine learning system will fit the data based on this information and be able to develop a regression remedy. The user will be prompted to indicate whether they enjoy a particular genre. A point on the regression curve will be made based on the user's response, and that point will be used to forecast what anime to suggest.

Initially, we intended to obtain the data directly from MAL via built-in API calls, however doing so would have exceeded the scope of this project. Instead, we used a public database called Kaggle to obtain MAL data. We downloaded a CSV file that had information about anime id, name, genre, episodes, rating, and members. However, because the genre types were not isolated in separate columns but rather a collection of strings, this data was useless. A solution was required because Pandas, the data analyzer and sorter, does not support strings.

We had to restructure the data so that each genre was distinct and assigned an integer value, as shown in the illustration below [6]. Multiple columns had to be created when we reorganised the data so that Excel could periodically check if it contained the string it was looking for. For instance, it would indicate a 1 and a 0 for everything else if the anime has a romance tag. The data would resemble the illustration below.

By doing so, we can actually give Pandas these numerical values to generate data frames. The model is then trained using the data frames. For this project, we choose to employ

Scikit-learn, a machine learning technique built on Python, and, more specifically, Support Vector Machines for classification. The genre is our X axis and the titles are our Y axis. From there, we have a collection of user inputs that describe a particular parameter and, in this case, forecast or suggest an anime to the user using linear regression.

RESULTS :

To guarantee that the method is optimal and to discover potential elements that could affect it, we developed a machine learning classification scheme for anime. Two experiments were run. We initially tested if the training set's popularity had an impact on overall accuracy. Because of the nature of SVM and the popularity of the top 100 anime, we discovered that employing them resulted in poorer accuracy. In our second experiment we carried trials to determine whether common and uncommon genres had an effect on accuracy and the results showed that having uncommon genres will yield better accuracy. However it is best to use all the genres when modeling as the perforce gamin is not worth the 62.5% drop in.

PAPE TITLE :

Collaborative Recommendation System in Users of Anime Films

PROBLEM STATEMENT :

One way to determine the consumer's preferences is to use the recommendation system, which displays prospective objects. Additionally, this suggestion aids the buyer in obtaining the preferred item. The anime movie is one of the most well-liked items on the list. In this instance, we carry out research to suggest anime movies based on reviews of previously viewed movies. A method called collaborative filtering involves tallying up forecasts, recommendations, and similarities. This study uses the Kaggle dataset, which has 12,294 anime and 73,516 users. The alternating least squares (ALS) approach will be used to compare a user's history against the history of all users. On the basis of such findings, the anime will be suggested. Millions of users should find their ideal anime using this technique.

DATASET DESCRIPTION :

The film dataset used in our experiment was obtained from the anime recommendation database "Kaggle". Data provide two files namely anime data and rating on anime. This data consists of a relation between 73,516 users and 12,294 anime. Each user is able to add anime to their completed list and give it a rating and this data set is a compilation of those ratings.

Anime.csv

anime_id - myanimelist.net's unique id identifying an anime.

name - full name of anime.

genre - comma separated list of genres for this anime.

type - movie, TV, OVA, etc.

episodes - how many episodes in this show. (1 if movie).

rating - average rating out of 10 for this anime.

members - number of community members that are in this anime's

"group".

Rating.csv

user_id - non identifiable randomly generated user id.

anime_id - the anime that this user has rated.

rating - rating out of 10 this user has assigned (-1 if the user watched it but didn't assign a rating).

IMPLEMENTATION DETAILS :

Mean averaging

Because it hasn't been recorded due to the supported actions, newly registered users won't have preferences for the movie. So, it is quite difficult to use the required film. The debate's opening segment is referred to as a "cold start problem." To get around this, the average value of each film is determined using the film rating. In order to standardize the film rating, movies with average ratings of 0 will be used. This will significantly improve efficiency.

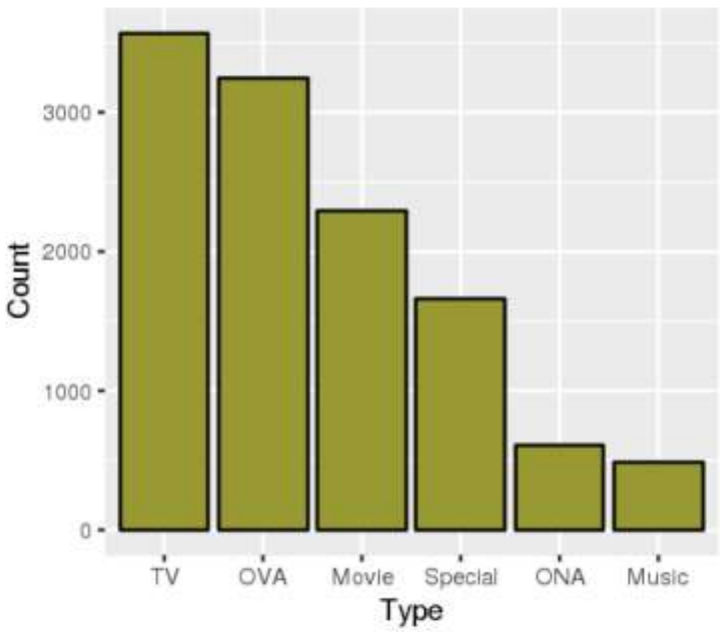
Similarity matrices

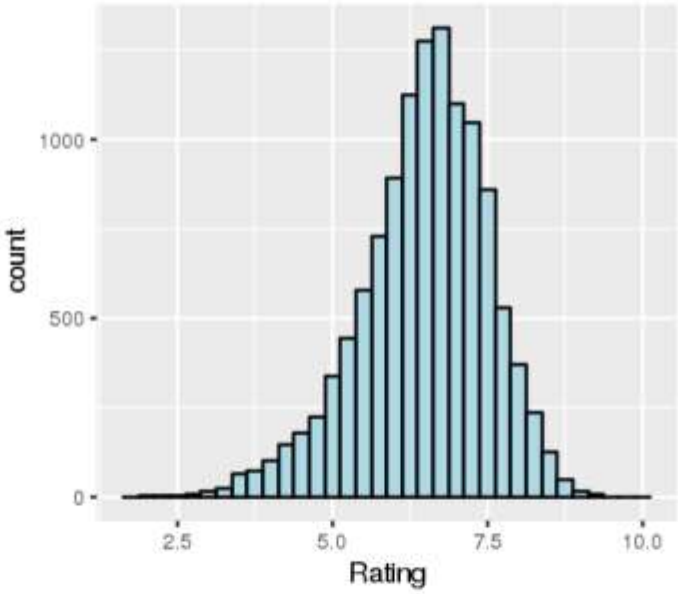
A matrix rating that may be altered expresses user and movie interaction. Users and items are represented by the node, and an edge connects the two. so that structural stages in the bipartite graph can be used to first identify the environment and then find the target user's closest neighbour. SimRank is a well-known tool that has been shown to be successful in minimising sparsity issues. In this study, we incorporate SimRank into the collaborative filtering algorithm to determine how similar users are.

Recommend top movies

The formula TX is the outcome of minimising the X matrix and user ratings using the parameter and movies with the X feature. Calculating the differences in features between films is the strategy. Users will be advised to utilise it for minor differences. For instance, the fact that films I and j have minor variances (Xi - Xj) indicates that they are similar. Based on the friendliness of the anime, from largest to largest difference, Top Anime will be suggested in this study.

RESULTS :





The collaborative filtering method judges the quality of the user's data based on how closely their preferences match those of other users. Alternating Least Squares (ALS) is used to build the recommendation system using the training data.

a superior anime recommendation engine that just considers viewer history. The resemblance between performances and users may be measured, and a basic recommendation system can also predict whether or not a person will like a particular anime. However, 100k data sets were chosen because it was difficult to calculate really big data sets. If an experiment with a million data set could be conducted on a machine that is significantly more powerful. This study presents a straightforward yet effective recommendation system.