

Data Science Project Proposal

1. Principal Investigator

Matthew DiDomizio (didomizm1@newpaltz.edu)

Maranda Dominguez (domingum6@newpaltz.edu)

2. Title of Project

Anime Recommendation System

3. Mentoring

Professor Min Chen, Department of Computer Science, SUNY-New Paltz

chenm@newpaltz.edu

4. Objective of Project

The primary goal of this recommendation system is to be able to reliably suggest new series that a given user would like to watch based upon their interests and prior viewing history. Ideally, the associations between specific series which other users gravitate towards en masse would sufficiently implicate what any particular individual with some of the same preferences would similarly like to watch.

5. Background/History of the Study

A recommender system is a process that seeks to predict user preferences. They ultimately help users discover products and content they might not have found on their own. Recommender systems are trained to understand the preferences, previous decisions, and characteristics of people and products using data gathered about their interactions. This can be done using various techniques, including collaborative filtering and content-based filtering. A content-based filtering system uses the attributes or features of an item. This type of recommender system is based on similarity of item and user features -- given information about a user and items they have interacted with, it can predict future interactions. With content-based filtering, it is highly probable that a user will get recommended items with the same attributes or features as their past preferences. On the other hand, a collaborative filtering system uses similarity of user preference behavior based on past interactions in order to predict future

interactions. The concept of collaborative filtering is that if users have made similar decisions in the past, like an anime choice, it will recommend a similar user that same anime choice. Machine learning algorithms are used to build such recommendation systems based on users' previous behavior or interests. Machine learning algorithms learn from users' behaviors and categorize them into groups which contain users of similar behaviors. Various types of machine learning algorithms have been used, such as clustering algorithms, deep learning models, classification algorithms, and matrix factorization. Though, the type of algorithm you use fully depends on whether your data is labeled or not. If the data is labeled, you must use a supervised machine learning algorithm such as naïve Bayesian classifier. Conversely, if it is unlabeled, an unsupervised machine learning algorithm like K-means clustering must be used.

6. Approach to the Study

In order to ensure that associations are made between users with similar preferences, we will utilize the K-Means clustering algorithm along with our unlabeled dataset of user ratings. When a user indicates they have a high affinity towards a certain series, they will get instant feedback using this collaborative filtering method based on the large volume of previously acquired data, wherein they will be recommended other series which related users enjoyed. As opposed to the content-based approach, it is not necessary to know the specifications of each series, as this is only relevant at the user level; the system need only have knowledge of ratings which are linked to their respective users. In addition, a long history of feedback is not needed for a new user to get accurate results since recommendations are based upon an existing dataset. While the collaborative method traditionally can suffer from the “cold-start” problem of not having enough reference data, this will be easily avoided considering the robust size of our initial dataset.

7. Initial timeline for implementation

1 Week	Literature review
1 Week	Data Collection
1 Week	Data Preparation
2 Weeks	Exploratory Data Analysis
1 Week	PC Setup
3 Weeks	Programming Algorithm

3 Weeks	Machine Learning
2 Weeks	Visualization

8. References

- Asad, Syed Muhammad. "AI Movies Recommendation System with Clustering Based K-Means Algorithm." *Medium*, Medium, 19 Aug. 2020, asdkazmi.medium.com/ai-movies-recommendation-system-with-clustering-based-k-means-algorithm-f04467e02fcd.
- Dey, Victor. "Collaborative Filtering vs Content-Based Filtering for Recommender Systems." *Analytics India Magazine*, 25 Aug. 2021, analyticsindiamag.com/collaborative-filtering-vs-content-based-filtering-for-recommender-systems/.
- Dwivedi, Rohit. "What Are Recommendation Systems in Machine Learning? | Analytics Steps." *Analytics Steps*, 16 Apr. 2020, www.analyticssteps.com/blogs/what-are-recommendation-systems-machine-learning.
- Grover, Prince. "Various Implementations of Collaborative Filtering." *Towards Data Science*, Towards Data Science, 28 Dec. 2017, towardsdatascience.com/various-implementations-of-collaborative-filtering-100385c6dfe0.
- Rocca, Baptiste. "Introduction to Recommender Systems." *Medium*, Towards Data Science, 2 June 2019, towardsdatascience.com/introduction-to-recommender-systems-6c66cf15ada.
- Roy, Abhijit. "Introduction to Recommender Systems- 1: Content-Based Filtering and Collaborative Filtering." *Medium*, 31 July 2020,

towardsdatascience.com/introduction-to-recommender-systems-1-971bd274f421.