Animendations

Team members

- Sravan Sriram - 730000571
- Vijay Anand Raghava Kanakagiri - 933004894
- Anushka Garg - 532008694
- Millennium Bismay - 633002191

1. **What exactly is the function of your tool? That is, what will it do?**
   We will develop an Anime recommendation system, which will take into account the many key anime features such as user ratings, anime genres, anime popularity, anime posters, anime synopsis and also respective user preferences. We will use these recommended anime to recommend user anime in accordance with their preference.

2. **Why would we need such a tool and who would you expect to use it and benefit from it?**
   There are a large number of anime watchers around the world who have their own unique tastes. Instead of wasting time having to research a particular anime to find out whether it would suit their tastes, our recommendation tool will use both user demographic and profile information as well as anime characteristics to personalize anime recommendations based on inferred user preferences from past watched history and ratings.

3. **Does this kind of tool already exist? If similar tools exist, how is your tool different from them? -**
   Most of the tools that recommend anime have considered content-based recommendations, user-user collaborative filtering and item-item collaborative filtering as well as SVD based recommendations (approaches from anime recommenders in past CSCE 670 classes and notebooks sourced from Kaggle). However, these tools did not consider hybrid based approaches which we seek to include in our tool.

   We will use content boosted collaborative filtering to combine a hybrid content-based and collaborative filtering approach as well as integrate VBPR to analyze anime covers to match user visual preferences and anime visual characteristics in a certain style space.

   We aim to use a weighted approach of several content features like genre, number of episodes, keywords of the anime synopsis to power our hybrid model.

Would people care about the difference?
We hope! The expectation is that an ensemble hybrid model will lead to better recommendations. Right now though there are platforms like “Myanimelist”, however, we do not believe there are any platforms which specifically recommend anime using an ensemble hybrid model and a VBPR based visual preference approach.

How hard is it to build such a tool? What is the challenge?
The biggest challenge in building our tool is the sparsity associated with data. The raw dataset we will be using contains about 99.9% sparsity in the user-item matrix. We believe that even after preprocessing and removing users and items with few ratings from our dataset, there will still be significant sparsity involved. Collaborative filtering approaches do not do that well when the dataset is extremely sparse. Our challenge is to both address this issue of sparsity using a hybrid model and integrate VBPR into our final recommendation model. Using VBPR to understand the style space preferences and also incorporating them in this domain is an unexplored area. Hence we expect significant challenges in integrating these parts. There are also lots of aspects in our dataset such as anime genres, user reviews, ratings, anime synopsis, anime episodes, anime air date, anime rank, anime popularity, anime score, thus exploring each of the categories and then finding a proper weighted balance to handle all our previously mentioned data is challenging.

4. **How do you plan to build it? You should mention the data you will use and the core algorithm that you will implement?**
   The core algorithm we will be implementing is content boosted collaborative filtering. This algorithm mitigates the effects of sparsity of the data by using a content-based similarity approach to assign a predictive score to all items for each user in our matrix. We can then apply traditional collaborative filtering algorithms on our data. We will be using k-fold cross validation instead of a randomized train-test split to evaluate our model to avoid the issue of irregular distribution of data that would arise in the test dataset. We aim to supplement this core algorithm with VPBR on the anime covers to identify user visual preferences in this area. We will be using the Anime Dataset with Reviews - MyAnimeList dataset for this project.

5. **What existing resources can you use?**
   We would be looking for APIs for extracting images for our dataset. We might additionally explore the use of language based keyword analysis using deep learning models for understanding keyword synopsis extraction and look out for some pre-trained convolutional neural networks that we can apply for our VBPR analysis.

6. **How will you demonstrate the usefulness of your tool?**
   We aim to validate our model using a k-fold cross validation approach. We plan to use RMSE as our evaluation metric to minimize our loss.

7. **A rough timeline to show when you expect to finish what. List a couple of milestones.**

Work Item	03/20	03/25	04/01	04/08	04/15	04/22	04/29	05/01
Preprocessing and Feature Engineering								
Synopsis Keyword Extraction and Embedding								
Weighted Content Based Similarity								
Content Boosted Collaborative Filtering								
VBPR								
NLPBR + Model Evaluation								
Website Implementation								