Andrey Ambartsumov

MAIS-202

Deliverable 1

1. Dataset:

I plan on making a Netflix recommendation algorithm for movies. I have found this dataset: Latest Netflix data with 26+ joined attributes (kaggle.com) which I believe will be enough for me to accomplish this task.

2. Data Pre-Processing:
   a. I found this dataset large enough for training and does not have enough noise. The most helpful columns are the title, the rotten tomatoes and IMDB ratings, series or a movie, since I plan on giving the user a choice of the 2, and then the run-time, as these parameters could all be important when choosing a movie.

   b. After talking with my TPM, I believe the best algorithm would be **Collaborative Filtering**, which although outside of the scope for this class, I am willing to do the research to implement it. However, **MLP**, which is part of the curriculum, is also a fitting algorithm, which I might implement instead as it will be seen in class.

   c. I believe the best algorithm for my task is Precision@K, which would calculate based on a user's input on how many of the recommended movies they found relevant out of all the recommended movies. Based on this metric the model can readjust. However, after doing some research, I could also try Normalized Discounted Cumulative Gain as an evaluation metric.

3. Application:
   I plan on creating a user interface with Flask, that will initially prompt the user to choose their favorite genres, and a certain number of favorite movies, like 10. Then, I will give an interface of proposing 5-7 movies, where you could click on each one and get a summary. Also, I will implement a shuffle button to get new results. Finally, I want to implement a search button in order to leave ratings to the movies that you've watched in order to make the algorithm more precise.