What should we watch tonight?: A Movie Recommender System

By: Ben Sturm

MVP Summary

Domain

For my final project at Metis, I have chosen to build a movie recommender system. One of my motivations to work on this project is that I dislike with a passion spending too much time surfing on Netflix or Amazon Video for movies to watch. If instead I could build my own recommender system that could take into account my preferences as well as the preferences of anybody else I'm watching with, such as my better half, that could save us both time and frustration.

Data

I will be utilizing the MovieLens 20M Dataset to build my recommender system. This data set consists of 20 Million movie ratings for 27,000 different movies by 138,000 users. The movies were all released between January 9, 1995 and March 31, 2015. A summary of the data provided in this dataset is given below.

Name	Type	Description
movieId	Int	e.g. 1,2,3…
Title	String	e.g. "Grumpier Old Men"
Genres	String	e.g. "Adventure, Fantasy"
userId	Int	e.g. 1,2,3…
Rating	Float	0.5 – 5.0
Timestamp	Int	Seconds since Epoch

Known Unknowns

There are a few additional files included in this dataset, including tags.csv, genome-scores.csv, and genome-tags.csv. I haven't yet studied what additional data these files contain, but I may want to use this additional metadata. I also don't know if this will be a purely collaborative based recommender system or a hybrid between collaborative and content based system.