Recommender System for Movies

# **Hanh Tran Spring 2021 GitHub Portfolio URL: https://github.com/htran017/DSC680projects.git**

# Which Domain?

What domain is this data going to come from? Please list 10 references (with a brief annotation) to use to make sense of what you’re doing with these data.

I currently work in the movie industry on the digital distribution operations end, assisting with iTunes movies and Google play movies. This domain will be in the movie and film industry. I am using Spark and Python to build a recommender system.

References:

<https://movielens.org/>

<https://www.kaggle.com/grouplens/movielens-20m-dataset>

<https://itunesconnect.apple.com/login> - (extra sales info)

# Which Data?

What is the dataset you’ll be examining? Please provide a codebook if there is one or a link to the dataset as well as a detailed description.

Movie lens dataset

<https://www.kaggle.com/grouplens/movielens-20m-dataset>

# Research Questions? Benefits? Why analyze these data?

How are you proposing to analyze this dataset? This is about your approach. Here, you’ll be proposing your research questions as well as justifications for why you’d offer these data in this way.

What movie should we offer to customers on our platform?

# What Method?

What methods will you be using? What will those methods provide in terms of analysis? How is this useful?

Content-Based and Collaborative Filtering (CF). Collaborative filtering is based on the idea of the “wisdom of the crowd” to recommend items whereas Content-based is focused on the attributes of the items and give a recommendation based on the similarity between them.

Spark.ml – supports model-based collaborative filtering

ALS – Matrix Factorization approach to implement a recommendation algorithm to decompose item/user matrix into lower dimensional user factors and item factors

# Potential Issues?

What challenges do you anticipate having? What could cause this project to go off schedule?

I just know that I’ll have a small typo somewhere and not see it until hours of trying different troubleshooting. I think tunnel vision can cause this project to go off schedule but I’m going to utilize the Teams board for help from other classmates.

# Concluding Remarks

Tie it all together. Think of this section as your final report’s abstract.