Project Proposal

Problem Statement

What kind of Recommendation Systems can I build using both traditional techniques and modern deep learning algorithms that have both high accuracy and make sense from a logical standpoint.

Context

This is my second capstone project for Springboard's Data Science course. Recommendation Systems are everywhere, and are an integral part of systems such as Google's YouTube videos and Amazon's Marketplace. For this project, I will create Recommendation Systems using the MovieLens dataset.

Criteria for Success

Create a model that can recommend movies and shows with accuracy greater than 50% over random chance, as well as make sense logically for a specific user.

Scope of Solution Space

Scope of solution space includes explanatory analysis of 7 variables that can be used for content recommendation. This includes a "description" feature which gives a short synopsis of the movie or show.

Constraints Within Solution Space

I am the mercy of the data. There could be missing values, mislabeled data, and even wrong information. This could highly influence my modeling and ultimately my results. I am also constrained by my computer. Certain Deep Learning algorithms need a lot of processing power and it may be time and cost prohibitive to use them for models.

Stakeholders to provide key insights

1. Kenneth Gil-Pascual, Data Scientist Consultant

Key Data Sources

1. Four CSV files that contain 100,836 ratings and 3,683 tag applications across 9742 movies. This dataset was generated from 610 users between March 29, 1996 and September 24, 2018.
2. The website of the group that published this dataset.

Deliverables

1. GitHub repository explaining my exploration process, analysis, and results.
2. A Project Report describing my process, conclusions, findings, and any other important details.
3. A PowerPoint presentation outlining and summarizing my results.