ISTM 660 Team 3 Project Proposal

Proposal

In our project, the team aims to predict the values of movie ratings without focusing on the content in the movie. We aim to harness user-specific data such as their age, gender, and location coupled with other outside factors such as genre, year of release, and reviews from critics.

From an initial screening of the dataset, we conclude that the predictor variables will be variables that represent age, sex, and location of the user providing reviews, critics' influence on IMDb users, and using year of release to confirm how much of a role recency bias play.

The sections below elaborate on the variables used for this task and present a general idea of the dataset as well.

Dataset

Source: https://www.kaggle.com/stefanoleone992/imdb-extensive-dataset

We will primarily be using two datasets in the form of Excel files, namely *I*MDB movies.csv and *I*MDB ratings.csv. Both the files have 85,855 records. IMDB movies have 22 attributes such as movie description, average rating, number of votes, genre, etc. IMDB ratings have 49 attributes that give demographic information of the audience. We will be a few of the attributes from both the files as predictor variables.

Files:

Filename: IMDb movies.csv (85,855 movies with 22 attributes)

imdb_title_id: Unique identifier

year: Movie release year

genre: Genre of the movie

country: Country in which the movie was produced

language: Language in which the movie was made

votes: Number of votes received for the movie

reviews_from_users: Reviews received from the users

reviews_from_critics: Reviews received from the critics

Filename – IMDb ratings.csv (85,855 movies with 49 attributes)

imdb_title_id: Unique identifier

votes_10: Number of users voting 10/10 for the movie (and so on)

allgenders_0age_avg_vote: Specific group of all genders under the age of 18 (both number of votes and rating)

us_voters_rating: Votes and ratings for movies by users from United States

non_us_voters_rating: Votes and ratings for movies by users from countries other than United States