

Project Luther (Movie Project Cheat Sheet)

Subject

- The Backstory

First project: Introduction to Linear Regression, how can we predict movie information?

- Client

Movie Studio, looking to make more popular movies

- Question?

The movie is being released, all the critics ratings have come in, what will people's opinion be of the film? If we have a movie script, which directors or actors should we cast in order to make it a popular one?

- Sample

Scraped all movies from metacritic to get aggregate movie ratings. Then connected data set with IMDb to get IMDb Ratings. Overall 8500 movies were present

- Model

Ridge Regression with all directors and actors who appeared more than 10 times in dataset. (~50 Directors, ~500 Actors, all genres (25)).

Started out by putting in individuals which I thought would make a difference, in the end used Ridge Regression, tuned the alpha based on cross validation to minimize mean squared error.

R-squared went from .471 to .54

-Outcomes:

Overall leading men seem to matter much more than women for IMDb Ratings,

Tyler Perry disliked by people who vote on IMDb Ratings, horror disliked more by people, documentaries enjoyed more.

Even with genres, directors, actors and critic ratings there is still much variability in critic ratings.