

ORIE 4741 Project Proposal

Sean Sodha Anudeep Gavini Omar Abdul-Rahim

September 22, 2017

The data set we chose is called “IMDB 5000 Movie Dataset,” and it contains a wealth of information about over 5,000 movies featured on the International Movie Database. From basic features like title, genre, and director, to more niche and interesting characteristics like the number of Facebook Likes received by the lead actor, this data set has many features that can provide us interesting insights.

A topic of interest in any industry is the projected profitability of a project. Given the data that we have, we believe that we can predict with some accuracy the profitability of a movie when considering many of the movie’s characteristics. In other words, an example of what we are asking would be, “With actors X, Y, and Z who generate Q amount of social media buzz in a genre of % and a budget size of \$\$\$, what is the predicted net margin of this movie?” We believe that we can glean interesting insights using data features not often explored, and we hope to come out of the project with a better idea of the movie business.