

Predicting Ratings from Scripts

Context

XYZ television station is an up and coming NBC affiliate. It is new to its market. It has the stable of NBC shows behind it, but it is looking for other ways to improve its ratings in its market. Management has decided that broadcasting movies may help. They feel that well-rated movies will improve ratings.

As an NBC affiliate, XYZ has access to some money. From this limited pot of money, they can create their own movies. Management thinks that producing movies will not only improve their ratings, but improve their standing with NBC.

XYZ has access to scripts for the movies, but the movies do not have ratings from customers. It wants to know which movies to produce, because the money from NBC is limited. That is, XYZ wants to **predict** how well a movie will rate based on the script so they can choose good movies for production and broadcasting.

Project

Using previously rated movies, I will create a rating model that will predict ratings given scripts. XYZ can then run their new scripts through the model and use it to help them decide whether to produce the movie.

Constraints

The model will only use scripts, not cast or crew. XYZ needs a way to decide which scripts should be produced before they offer them out to production teams.

Data Sources

I will use available movie scripts (<https://www.simplyscripts.com/movie-scripts.html>) and IMDb ratings (<https://www.kaggle.com/datasets/ashirwadsangwan/imdb-dataset>; IMDb now charges) to create a model that predicts IMDb ratings.