

Natural language processing with screenplays

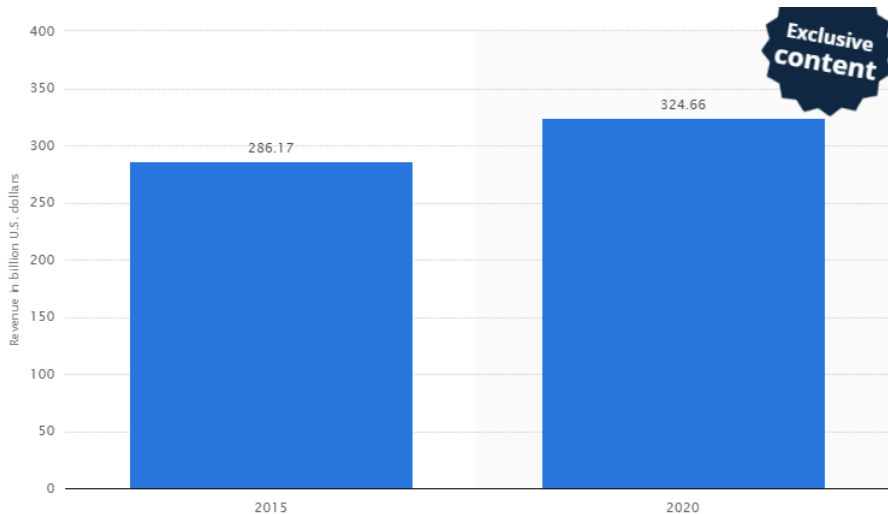
Is it possible to predict the success of a movie based only on it?

Pedro Argento

General Assembly - Data Science, 2016

The film and TV industry generated 287 Billion Dollars in 2016, and it is expected to generate 324B in 2020.

PWC expects a 13% growth in the next 5 years.



Although film industry is strong and profitable, a lot of the movies produced fail to break even.



Machine Learning models can help humans make better decisions in two ways:

- ▶ Find what "hidden" features makes a movie successful, and avoid unsuspected busts.
- ▶ Help filtering and prioritizing which scripts should be read by a human: An average producer gets at least 10,000 scripts a year, most go unread.

We can define a movie success in three categories: Financial, Public and Peers/critics

- ▶ Financial success will be measured on how much profitable the movie is (Revenue minus costs).
- ▶ Public success will be measured by IMDB ratings, voted by the general public.
- ▶ Peer success will be measured by if the film was nominated for an Oscar.

Concern: This work presents a sampling bias impossible to avoid

As all movies used to train the models were produced and approved by a human, we can not be sure if the results will be applicable for random scripts.