

The Impact of film's features on IMDB Ratings: A logistic Regression approach

Yujie Tang, Jialu FU, Weiqing GUO, Bashiru Mukaila, Wanding Wang

2023-03-14

1 Conclusion

After a movie is released, it is important to pay attention to the audience's word of mouth, which is reflected in the ratings given by the general public on IMDB. We consider movies with ratings of 7 or higher as high-rated movies and explore which movie attributes affect the rating results.

The following conclusion was drawn by modelling binary logistic regression with one categorical variable as an explanatory variable and by mixed modeling with multiple numerical variables as explanatory variables:

- The *genres* of Action, Drama, and Romance have fewer high-rated movies, while Documentary and Short genres have more high-rated movies. This indicates that audiences prefer movies with logic and depth, rather than formulaic commercial films.
- The *genres* of Animation and Comedy receive mixed ratings, but overall, there are more high-rated movies, which is in line with our impression of popular trends among the general audience.
- In terms of numerical factors, the *year*, *length*, and *budget* are taken into account. However, the *votes*, representing the number of positive votes received by viewers, does not have a significant impact on the final rating of the movie, which may be due to its subjectivity.
- The release year and production budget of a movie have a positive impact on its rating, indicating that people prefer new movies that match their current tastes. Moreover, the budget invested in a movie can greatly influence its production quality, providing people with new and exciting experiences and thus gaining more popularity. However, the length of a movie has a negative impact on its rating, indicating that people no longer enjoy the long narrative of old movies, which is related to today's fast-paced culture.

Future work could include:

- In addition to the movie attributes mentioned in the study, other movie attributes such as box office, investment in art and music, etc., may also have an impact on movie ratings. Not only the movie's inherent attributes, but other cultural factors such as language and actors may also cause certain differences. If more variables can be included, it will be more effective in improving the accuracy of the classifier's functionality.
- In this study, the response variable was set to two categories, greater than 7 points or lower. If the response variable can be divided into more categories, such as high-rated movies, average-rated movies, and low-rated movies, it will be more discriminative.
- The distribution of genres in the sample used in the study is uneven, as there were no *Romance* films with ratings above 7, which has led to errors in the model. Obtaining a more comprehensive sample would result in better model performance.