Predicting IMDb Movie Rating using Machine Learning

Ananya Jain
ananya19408@iiitd.ac.in

Manasvi Singh
manasvi19369@iiitd.ac.in

Pritish Wadhwa
pritish19440@iiitd.ac.in

Yash Bhargava
yash19289@iiitd.ac.in

1. Motivation

Is it possible to predict the rating of a movie prior to its release or production? Every year countless movies are released worldwide. IMDb rating is the singlemost influential factor in deciding consumer's opinion and inherently the blockbuster success of a movie.

With the machine learning techniques at our disposal, we aim to predict the seeming unpredictable IMDb rating of any movie before its theatrical release. Successfully predicting IMDb rating is beneficial to both producers (from a financial standpoint) and consumers (from an entertainment standpoint) alike.

2. Related Work

A. Oghina et. al in their paper[1] predicted IMDb movie ratings by considering two sets of features: surface and textual. For the latter, data from multiple channels that are linked to a particular movie was collected. Textual features from each channel were used in the prediction model.

B. Quader et. al in their paper[2] predict rating of a movie on platforms like IMDb and Roten Tomatoes based on pre-released and post-released features of a movie. SVM and neural networks were utilized in this paper.

C. Çizmeci et. al in their article[3] explored the IMDb ratings by exploring matrix decomposition, regression analysis and factorization machines on social media data.

3. Timeline

Week 1-2 : Analysis of research papers and data collection
Week 3 : Curating Dataset, Preprocessing and Explanatory Data Analysis (EDA)
Week 4 : Feature analysis, Feature selection, Correlation, Heat maps
Week 5 : Classification, Regression models
Week 6,7,8 : Decision Tree, Random Forest, SVM, Basic Neural Networks
Week 9-10 : Comparative analysis of various machine learning models, Selecting Best Model
Week 11: Final Report

4. Individual Tasks

Each member would be responsible for all the tasks pertaining to this project. Tasks will be divided equally, and every member will carry out the assigned tasks for holistic learning of the entire project.

5. Final Outcome

Our expectations from the project are to predict the ratings of any movie using features that are available prior to its release and analyse the discriminatory power of features to analyse how different parameters impact the rating for any movie using various machine learning models.
Predicting movie ratings has been an active area of research but predicting movie ratings solely on the basis of features present prior to release has not been explored in depth. Most of the research focuses on predicting movie rating using social media but that is done after a movie releases. This project will help us identify important features which affect the ratings of any movie, prior to its release. This can find key applications among movie enthusiasts, film producers and big movie houses in assisting them with decisions before a movie releases.

References

[1] Andrei Oghina, Mathias Breuss, Manos Tsagkias, and Maarten de Rijke. Predicting IMDB Movie Ratings Using Social Media. *Advances in Infor-mation Retrieval. ECIR 2012. vol 7224. Springer, Berlin, Heidelberg*, pages 503–507, 2012.

[2] Nahid Quader, Md. Osman Gani, Dipankar Chaki, and Md. Haider Ali. A machine learning approach to predict movie box-office success. *2017 20th International Conference of Computer and Information Technology (ICCIT)*, pages 1–7, 2017.

[3] Chuan Sun. Predict Movie Rating, 2020.