

Determining IMDb Movie Ratings using Machine Learning

Dayton Wu and Upama Ferdousi

Why did you choose this topic?

This project is performed to find if IMDb movie ratings are directly affected by the various presented categories in our dataset.

What are the sources of data?

We extracted movie data from *Kaggle* and also performed web scraping on the *Reelgood* site. We pulled movie listings between 1902 to 2020. It includes:

- Movie titles
- Year of release
- Age group
- Platforms
- Directors
- IMDb ratings
- Genre
- Runtime

and more...

Questions the team hopes to answer...

Through our analysis, we would like to dig deeper and find out which feature had significant impact on the IMDb ratings.

Steps to Data Analysis...

- I. Data Acquisition and Cleaning
- II. Features and models
- III. Testing and training
- IV. The Results

Method of Data Analysis

- Data extraction
- Web scraping (requests, HTML, BeautifulSoup)
- EDA (pandas, numpy)
- Quantitative Classification of Data (Random Forest)
- Linear regression (scikit-learn)
- Data visualization (seaborn, matplotlib)