# Movie IMDb Rating Predictor

# EECS 349 Project Proposal

## Group Member

Zhilin Chen, zcj456

## Tasks

Movies, originally invented in the 1990s for entertainment, has now became an indispensable part of human culture. There are multiple criteria to define a “good” movie: if a movie profitable, popular or if it introduces new techniques in filmmaking. In my tasks, I’m going to adapt the rating of IMDb(Internet Movie Database) to define if a movie is “good” because it fairly reflects how public evaluate this movie.

My task is **to determine what and how attributes decide the IMDb rating of a movie**. For example, how movie’s genre, director, stars or production corporation effect its IMDb ratings. This would be interesting and meaningful. A businessman would want to use it because he can know will a movie be popular and profitable beforehand. What’s else, this predictor could also provide us and relevant scholars how public preferences in movies change over time.

## Data and Approach

Basically, I would retrieve most of my data from [**www.imdb.com**](http://www.imdb.com) via its **API** or a **Python crawler** write by myself.

The **features** for my tasks would first include the most basic information of a movie: **genre**, **director**, **writers**, **stars**, **production corporation** and **those which play important roles in filmmaking**. I would add any features if I found it necessary to the movie rating during the work. For example, I would try to use public reviews as a feature if I find it possible to do some semantics analytics on reviews.

First, I would try to define my task a **classification problem**: splitting rating into few classes and then use some suitable algorithms (**such as KNN**) to predict how “good” a movie will be. Then, if possible, I would try to turn my task into a **regression problem** therefore I could predict the movie’s rating more precisely. Of course, I would try to do some **Natural Language Processing** if I am going to consider public reviews as an important feature.