Deconstructing Data Science

Part 1 - Anubhav Gupta

Most of the features that I have enumerated here would be valid in predicting the winners of each of the six categories that have been mentioned. However, some features might be a better predictor of one categories than the others.

Features for the Best Picture Category

- **Oscar nominations from that movie in other categories like best actor, best actress and best director** - I think this is the single most important factor that might predict the chances of a film winning the Oscars . "Nominees for Best Picture often are represented by multiple nominees in other categories, and the chances of winning are generally though to increase the higher the total number of nominations. For example, the median number of nominations for winners of the Best Picture Oscar since its inception (1928–2004) is nine, whereas the median number of nominations for losing Best Picture nominees is six."[1]
- **Winner of the best Movie in Golden Globe award and the Directors Guild of America.**
- **Total Revenue generated by the movie** - I think the total revenue generated by the movie is a good indication of how popular the movie is which might be a good indicator of the probability of the movie winning the Oscars.
- **Rating of the movie on sites like IMDB, cotton tomatoes etc**. - These sites are pretty accurate on how good the movie is and might be a good indicator
- **Duration of the movie**
- **Budget of the movie** - Large budget films might have better directors and actors and hence might be more probable to win.
- **Duration it stays in theaters** - Good indication of popularity.
- **No of screens the movie is released in** - Good indication of popularity.
- **Opening day collections** - Good indication of popularity
- **Release date**
- **Whether its a sequel or a prequel** - Depending on how good the prequel or sequel was, it might win the Oscars.
- **User and Critic text reviews from sites like bottom tomatoes and movie blogs and columns about the movie** - Scraping text reviews and performing NLP to determine how a movie is perceived by the audience.

Features for the Best Director Category

- **Whether the movie directed is nominated for the Best Movie** - I think this is really an important predictor of whether the director might win the Best director because if

---

[1] https://www.amstat.org/newsroom/pdfs/JustHowPredictablearetheOscars.pdf

the movie wins the Oscar, its generally because the direction was very good. "Only three movies have won the Best Picture Oscar without also receiving a Best Director nomination ("Wings" in 1928, "Grand Hotel" in 1932, and "Driving Miss Daisy" in 1989)."[1]
- Oscar and other awards won by the director in previous years.
- **Total Oscar nominations for the movie (for which the director is nominated)**
- Total Revenue generated by the film
- Rating of the movie on sites like IMDB, cotton tomatoes etc.
- Duration it stays in theaters
- No of screens the movie is released in
- Opening day collections
- User and Critic text reviews from sites like bottom tomatoes and movie blogs and columns about the direction

Features for the Best Actor/Actress Category

- **Whether the movie acted in is nominated for the Best Movie.**
- **Total Oscar nominations for the movie (for which the actor/actress is nominated)**
- Total Revenue generated by the film
- Rating of the movie on sites like IMDB, cotton tomatoes etc.
- Duration it stays in theaters
- No of screens the movie is released in
- Opening day collections
- Oscar and other awards won by the actor/actress in the same or some other category in previous years
- User and Critic text reviews from sites like bottom tomatoes and movie blogs and columns about the Actor/Actress

Features for the Best Supporting Actor/Actress Category

- **Whether the movie acted in is nominated for the Best Movie.**
- **Total Oscar nominations for the movie (for which the actor/actress is nominated)**
- Total Revenue generated by the film
- Rating of the movie on sites like IMDB, cotton tomatoes etc.
- Duration it stays in theaters
- No of screens the movie is released in
- Opening day collections
- Oscar and other awards won by the actor/actress in the same or some other category in previous years
- User and Critic text reviews from sites like bottom tomatoes and movie blogs and columns about the Actor/Actress

The subset of features that which can be instantiated as data points from different sources are:

- **Total Revenue generated by the movie -** https://en.wikipedia.org/wiki/127_Hours
- **Oscar nominations from that movie in other categories like best actor, best actress and best director** - https://en.wikipedia.org/wiki/88th_Academy_Awards#Nominations
- Winner of the best Movie in Golden Globe award and the Directors Guild of America. - Wikipedia
- **Total Revenue generated by the movie -** https://en.wikipedia.org/wiki/127_Hours **Rating of the movie on sites like IMDB, cotton tomatoes etc**. - http://www.imdb.com/title/tt1542344/
- **Duration of the movie** - https://en.wikipedia.org/wiki/127_Hours
- **Budget of the movie** - https://en.wikipedia.org/wiki/127_Hours
- **Release date -** https://en.wikipedia.org/wiki/127_Hours
- **Whether its a sequel or a prequel** - https://en.wikipedia.org/wiki/127_Hours
- **User and Critic text reviews from sites like bottom tomatoes and movie blogs and columns about the movie** - http://www.imdb.com/title/tt1542344/