

# Lunar Outpost: Executive Summary

William Reese, Charles D Ruggiero, Nathan Schley, Sheri Tamagawa

Every year approximately 6000 voting members of the Academy of Motion Picture Arts and Sciences are asked to nominate movies for an Oscar, which are finalized by a committee. Once these nominations are announced, interest skyrockets for every nominated film. Being able to predict these nominations before they are announced is important to many stakeholders. Two of these categories are as follows.

- ❖ Distributors (e.g. online streaming services and theater owners)

- Reasons for Interest:

- Securing licensing for movies before they are nominated
  - Dramatically increased viewership for key movies whose licenses are secured before other distributors can act first.

- Key Performance Indicators (KPIs):

- Decreased costs related to licensing fees
  - Increased revenue from sales

- ❖ Movie studios and movie industry professionals

- Reasons for Interest:

- Lots of planning needs to be done, and all of these stakeholders want to be successful in the industry.
    - Knowing movies for which to invest in For Your Consideration advertising
- KPIs:
- Increased annual income
  - Increased number of movies that are nominated (for movie studios)
  - Increased notariety

We decided to focus on the animated films category. Animated movies have had their own category for over two decades. We started our project by using Beautiful Soup and the Cinemagoer API to scrape movie data from IMDb's website.

The movie data we found came with many helpful features, such as production companies, producers, directors, genres, certificates, metascore, and the month in which the movie was released. There were other features we were not able to make use of, such as ratings, distributors, cast, language and runtimes.

Because we were creating a forecasting model, we performed a 5-fold time series split on the data. We then used a voting system for predictive modeling that incorporated  $k$  nearest neighbors, random forest, logistic regression and polynomial support vector classifier methods. We found weights that optimized our KPIs of precision and recall.

Our model can be used to predict which movies will be nominated for Best Animated Film Oscars shortly before the announcements are made. On our testing data, our model had an average precision of 100% and an average recall of 48.3%. Thus, our model is better suited for selecting Oscar nominees than excluding them.