

Seeking Oscars

By: Andrew Wu

Using Machine Learning
models to predict the
success of a movie at
the Oscars



Goal: to explore characteristics that makes a film successful at the Oscars and predicting number of awards a film will win based on data



92nd Academy Awards

Photo by Trey P. / © A.M.A.S.



01

Methodology

Data collections, Tools & Models



Data Collection

- Webscraping using Selenium&Beautifulsoup
- **Oscar Nominee data**



Tools

- Jupyter Notebook
- sklearn,statsmodels,



Models

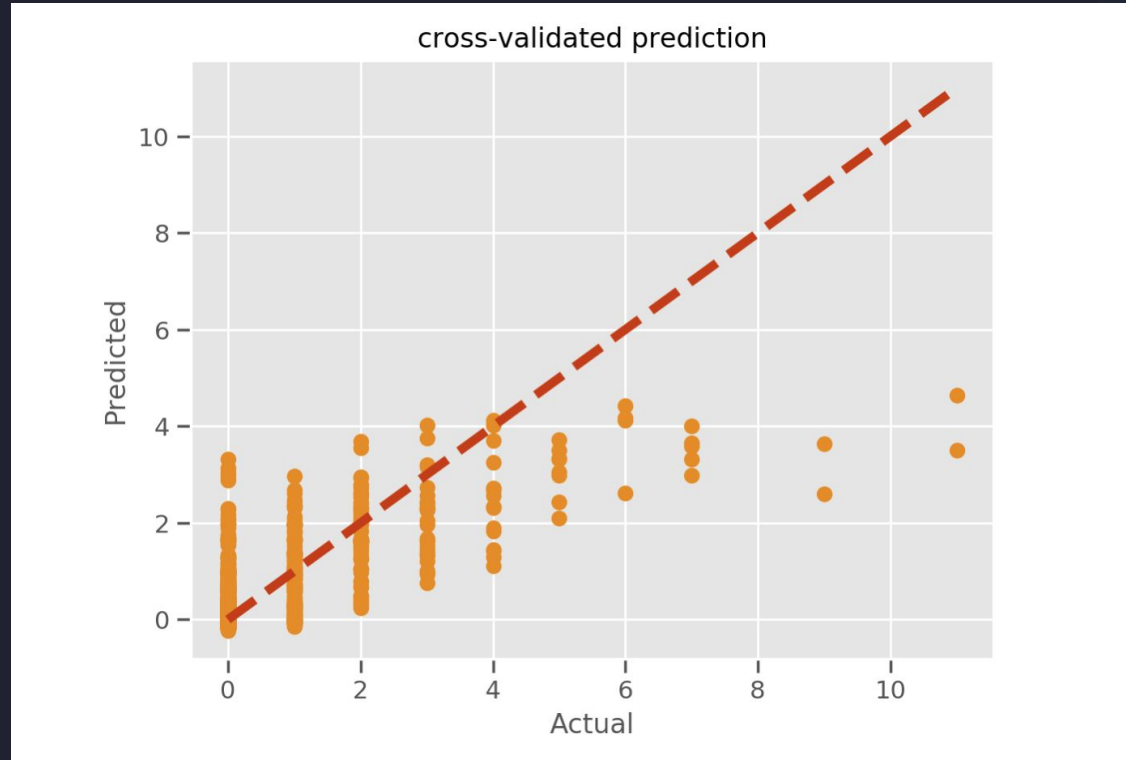
- Ordinary-Least-Square(OLS) regression
- LASSO regression
- Ridge regression



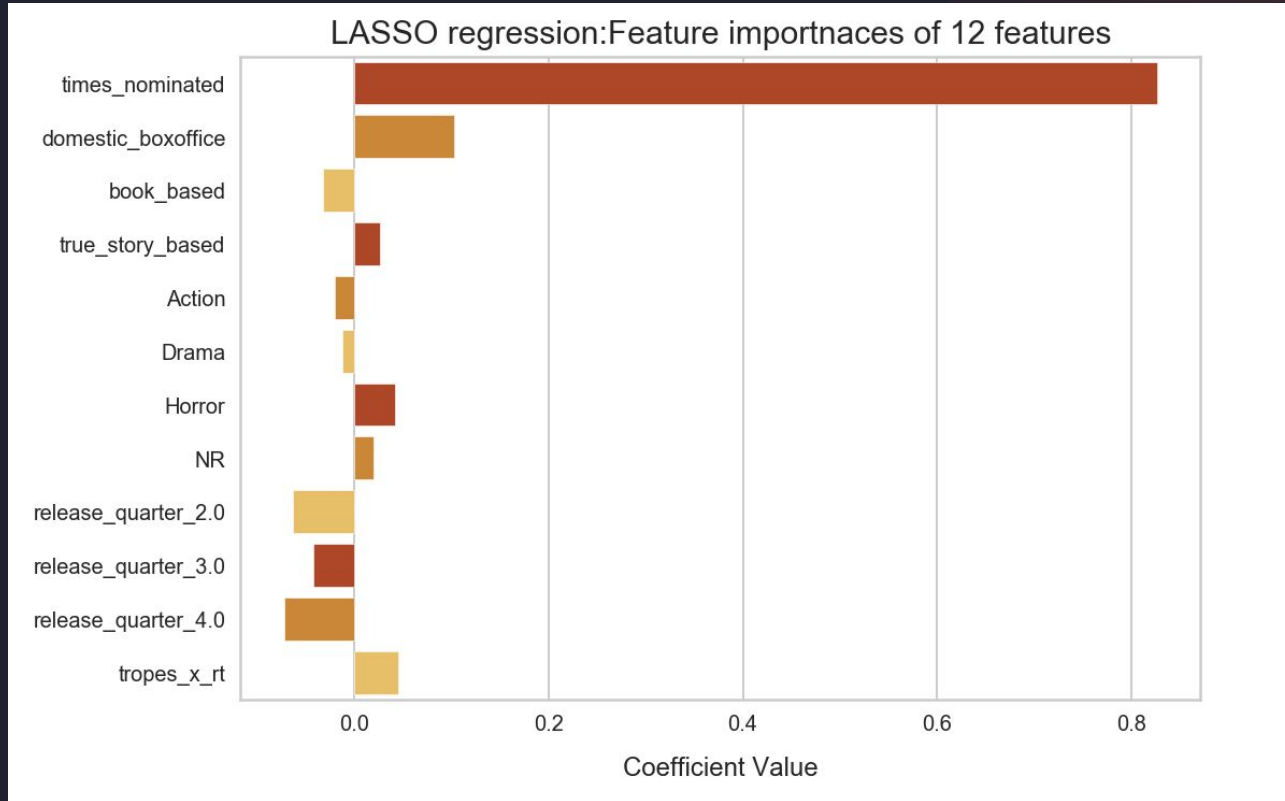
02

Results

Cross-validated prediction: Linear model is accurate until a certain number of awards won



LASSO regression key Features: defying norms



Conclusions: The model shows promise but could be further optimized



Final model accuracy with cross
validation test: 0.459

Alpha value of: 0.0109

Cond no: 1.53 e+09

Explanations:

- Dynamic judging process
- High multicollinearity
- Oscars are political

03

Future improvements



What could be done in the future to improve the model?

Different
model

01



More feature
engineering

02



Introduce
Time-series

03



Introduce
classification

04



The background image shows the interior of a grand, ornate theater. The seating area is filled with rows of red upholstered chairs with dark wooden frames. The walls are decorated with intricate carvings and murals, and the ceiling is high with ornate lighting fixtures. A large, semi-transparent blue circle is centered over the image, containing the word "Appendix" in white text.

Appendix

Adjusted $R^2 = 0.61$

