2021 Academy Awards nominations

Sebastian Deręgowski

March 15, 2021

Contents

1 Introduction

Oscar nominations are usual the resultant of nominations for less prestigious awards (e.g. Golden Globes, BAFTA, Critics Choice, industry awards). Two models of machine learning (linear regression and logistic regression) were used to predict 2021 Academy Awards nominations based on 5 previous years' nominations.

Both models had access data about 5 last years, including nominations for Oscars, Golden Globes and so on (in total: 234 movies x 100 categories). For every movie there was a dedicated number (mostly 0 or 1, sometimes 2) that indicated how many nominations did the film achieve in a given category. That was a training data for both models. In this year, 61 movies were taken into account.

For each model there is a list of predicted nominees in every category. They are sorted by a probability of getting nomination within a model. Green cell indicates that model predicted the nomination correctly, the red one - wrong. Under the table there is info about accuracy and short explanation of results for given category.

In the last section there is a summary of both models including comparison of both models and conclusions.

2 Results

2.1 Best Picture

Logistic Regression Model	**Linear Regression Model**
Nomadland	
The Trial of the Chicago 7	
Promising. Young. Woman	
Mank	
Minari	
Ma Raine's Black Bottom	One Night in Miami
Sound of Metal	Ma Raine's Black Bottom
One Night in Miami	Sound of Metal

Accuracy of both models: **75%**. Skipped movies in both: ***Judas and the Black Messiah*** and ***The Father***. The first one was on the 10th place in both models, while the second one was on the 11th place.

2.2 Best Directing

Logistic Regression Model	Linear Regression Model
Chloe Zhao, *Nomadland*	
Lee Isaac Chung, *Minari*	
David Fincher, *Mank*	
Emerald Fennell, *Promising. Young. Woman*	
Aaron Sorkin, *The Trial of the Chicago 7*	

Accuracy of both models: **80%**. Academy gave the nomination to **Thomas Vinterberg, *Another Round*** instead of Aaron Sorkin. Thomas Vinterberg was on the 8th and 7th place in logistic and linear regression model respectively.

2.3 Best Leading Actor

Logistic Regression Model	Linear Regression Model
Chadwick Boseman, *Ma Raine's Black Bottom*	Riz Ahmed, *Sound of Metal*
Riz Ahmed, *Sound of Metal*	Anthony Hopkins, *The Father*
Anthony Hopkins, *The Father*	Chadwick Boseman, *Ma Raine's Black Bottom*
Gary Oldman, *Mank*	Tahar Rahim, *The Mauritanian*
Tahar Rahim, *The Mauritanian*	Gary Oldman, *Mank*

Accuracy of both models: **80%**. Academy rewarded **Steven Yeun, *Minari*** instead of Tahar Rahim. In logistic model, Steven Yeun was right after Rahim, on the 6th position. In linear model he was on 8th position.

2.4 Best Leading Actress

Logistic Regression Model	Linear Regression Model
Frances McDormand, *Nomadland*	
Vanessa Kirby, *Pieces of a Woman*	
Carey Mulligan, *Promising. Young. Woman.*	
Viola Davis, *Ma Raine's Black Bottom*	
Andra Day, *The United States vs. Billie Holiday*	

Accuracy of both models: **100%**.

2.5 Best Supporting Actor

Logistic Regression Model	Linear Regression Model
Daniel Kaluuya, *Judas and the Black Messiah*	
Leslie Odom Jr., *One Night in Miami*	Delroy Lindo, *Da 5 Bloods*
Delroy Lindo, *Da 5 Bloods*	Leslie Odom Jr., *One Night in Miami*
Paul Raci, *Sound of Metal*	
Sasha Baron Cohen, *The Trial of the Chicago 7*	Alan S. Kim, *Minari*

Accuracy of logistic regression model: **80%**. Accuracy of linear regression model: **60%**. Sacha Baron Cohen was on the 7th place in linear regression model. The actor that is missing in both models is **Lakeith Stanfield, Judas and the Black Messiah**. That movie didn't get 2 nominations in this category in any previous awards (they all nominated only Daniel Kaluuya) so it was impossible for both models to predict it.

2.6 Best Supporting Actress

Logistic Regression Model	Linear Regression Model
Maria Bakalova, *Borat Subsequent Moviefilm*	
Yuh-Jung Youn, *Minari*	
Glenn Close, *Hillbilly Elegy*	
Olivia Colman, *The Father*	
Amanda Seyfried, *Mank*	

Accuracy of both models: **100%**.

2.7 Best Original Screenplay

Logistic Regression Model	Linear Regression Model
Promising. Young. Woman	
The Trial of the Chicago 7	
Mank	
Sound of Metal	
Minari	

Accuracy of both models: **80%**. Skipped movie in both is ***Judas and the Black Messiah***. In both models it was placed on the 10th position so this nomination may be considered as a suprise.

2.8 Best Adapted Screenplay

Logistic Regression Model	Linear Regression Model
Nomadland	
Ma Raine's Black Bottom	
News of the World	
One Night in Miami	
The Father	

Accuracy of both models: **60%**. Skipped mobies in both: ***Borat Subsequent Moviefilm*** and ***The White Tiger***. The first one was on the 8th position in both models. The second one was right after the predicted nominees, on the 6th place, in both models as well.

2.9 Best Cinematography

Logistic Regression Model	Linear Regression Model
Nomadland	
News of the World	Mank
Mank	News of the World
The Trial of the Chicago 7	
Cherry	

Accuracy of both models: **80%**. Skipped movie in both is ***Da 5 Bloods*** which was on the 9th place in both models.

2.10 Best Original Score

Logistic Regression Model	Linear Regression Model
Soul	Mank
Mank	News of the World
News of the World	Soul
Minari	
Tenet	

Accuracy of both models: **80%**. Skipped movie in both is ***Da 5 Bloods***. This movie has not been placed in top 15 in any of two models.

2.11 Best Production Design

Logistic Regression Model	Linear Regression Model
Mank	
News of the World	
Ma Raine's Black Bottom	The Dig
Tenet	Rebecca
The Personal History of David Copperfield	The Father

The accuracy is **80%** and **60%** for logistic and linear regression model respectively. ***The Father***, which was nominated in linear model, was not even in top 15 in logistic model. ***Ma Raine's Black Bottom*** and ***Tenet*** which were nominated in logistic model, were right after predicted five, on the 6th and 7th place respectively.

2.12 Best Costumes

Logistic Regression Model	Linear Regression Model
Ma Raine's Black Bottom	
Emma	Mank
Mank	Emma
The Dig	
Ammonite	

Accuracy of both models: **60%**. Skipped movies in both: ***Mulan*** and ***Pinocchio***. ***Mulan*** was on the 6th and 7th place in logistic and linear regression model respectively. ***Pinocchio*** was on the 10th place in logistic regression model and was not in top 15 in linear one.

2.13 Best Make-Up and Hairstyling

Logistic Regression Model	Linear Regression Model
Ma Raine's Black Bottom	
Hillbilly Elegy	
Mank	
Pinocchio	
The Dig	

Accuracy of both models: **80%**. ***Emma*** was skipped in both models, it was on the 7th place in both.

2.14 Best Visual Effects

Logistic Regression Model	Linear Regression Model
Tenet	
Greyhound	
Mulan	
The Midnight Sky	
The One and Only Ivan	

Accuracy of both models: **80%**. ***Love and Monsters*** replaced ***Greyhound*** in both models. That movie hasn't been nominated in any previous awards and so that's why it wasn't even in consideration.

2.15 Best Sound

Logistic Regression Model	Linear Regression Model
Greyhound	
Sound of Metal	
News of the World	
Nomadland	
Soul	

Accuracy of both models: **80%**. ***Mank***, which was nominated in this category instead of ***Nomadland***, was quite far, on the 12th place in both models.

2.16 Best Film Editing

Logistic Regression Model	Linear Regression Model
Sound of Metal	
The Trial of the Chicago 7	
Nomadland	
Promising. Young. Woman	
Mank	The Father

Accuracy of logistic regression model: **80%**. Accuracy of linear regression model: **100%**. ***The Father*** was on 11th place in logistic model.

3 Summary

Summarized accuracy is **66/83 (79,52%)** for logistic regression model and **65/83 (78,31%)** for linear one. It means that both models are very similar. Logistic model made better predictions for **Best Supporting Actor** and **Best Production Design** categories, while linear model made it better for **Best Film Editing**.

In most cases both models predicted **4 out of 5** nominees. Taking a fix for suprises (e.g. movies that weren't nominated for given category in any previous awards) that is a very good result.

When it comes to misses, a nominee that wasn't predicted was on average on **8,46th** place in a logistic model and on **8,21th** place in a linear model. There were also 4 nominations that weren't even considered by logistic model and 5 by a linear one. It means that it's quite easy to predict the most favored nominees in each categories (excluding few suprises) and it's extremely hard to predict them all, mostly because of Oscars' independence and willingness to introduce new or unpredictable nominees.

There is also an interesting observation, that so few years (5 out of 93 years of Oscars history) are needed for an algoritm to find patterns between Oscars and another awards. It's even more suprising that adding data about another year (in the original version there was data from 4 not 5 last years) changed absolutely nothing in every category.