Project PTO – Building a Computer Algorithm to **P**redict **t**he **O**scars

Overview

Method

Process

Takeaways

Results

Mode A: ML Algorithm

Mode B: Research Report

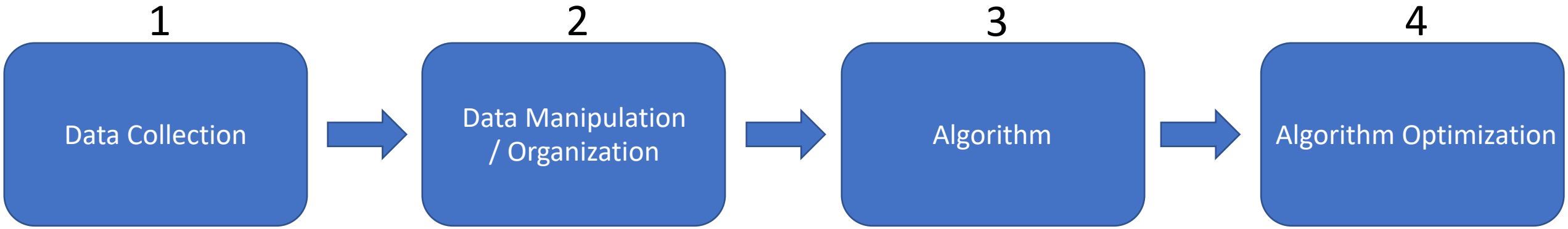
For the last 4 weeks, I developed a machine learning algorithm to predict the Oscars. Not just the winners, but also the nominees, in 6 categories: Best Picture, Leading Actor, Leading Actress, Supporting Actor, Supporting Actress, and Animated Film. I chose this project because I love movies and always watch the Oscars, and I'm pursuing CS major, so this project is the best of both worlds.

Data: Existing dataset + Web-scraping

Algorithm: Using Keras Functional API, MLP of hidden layer of 128 nodes w/ ReLU + 6 output layers of 3 nodes w/ Softmax

Results: 98.4% accuracy for the 2018 Oscars

Mentor: John Ciancutti, a Google director/scientist


1
Data Collection
2
Data Manipulation / Organization
3
Algorithm
4
Algorithm Optimization

Week 1: Data Collection

- Met w/ mentor and narrowed down project from movie recommendation system to Oscar prediction algorithm
- Found BIGML dataset
- Web-scraped list of movies from IMDB

Week 2: Data Collection + Manipulation

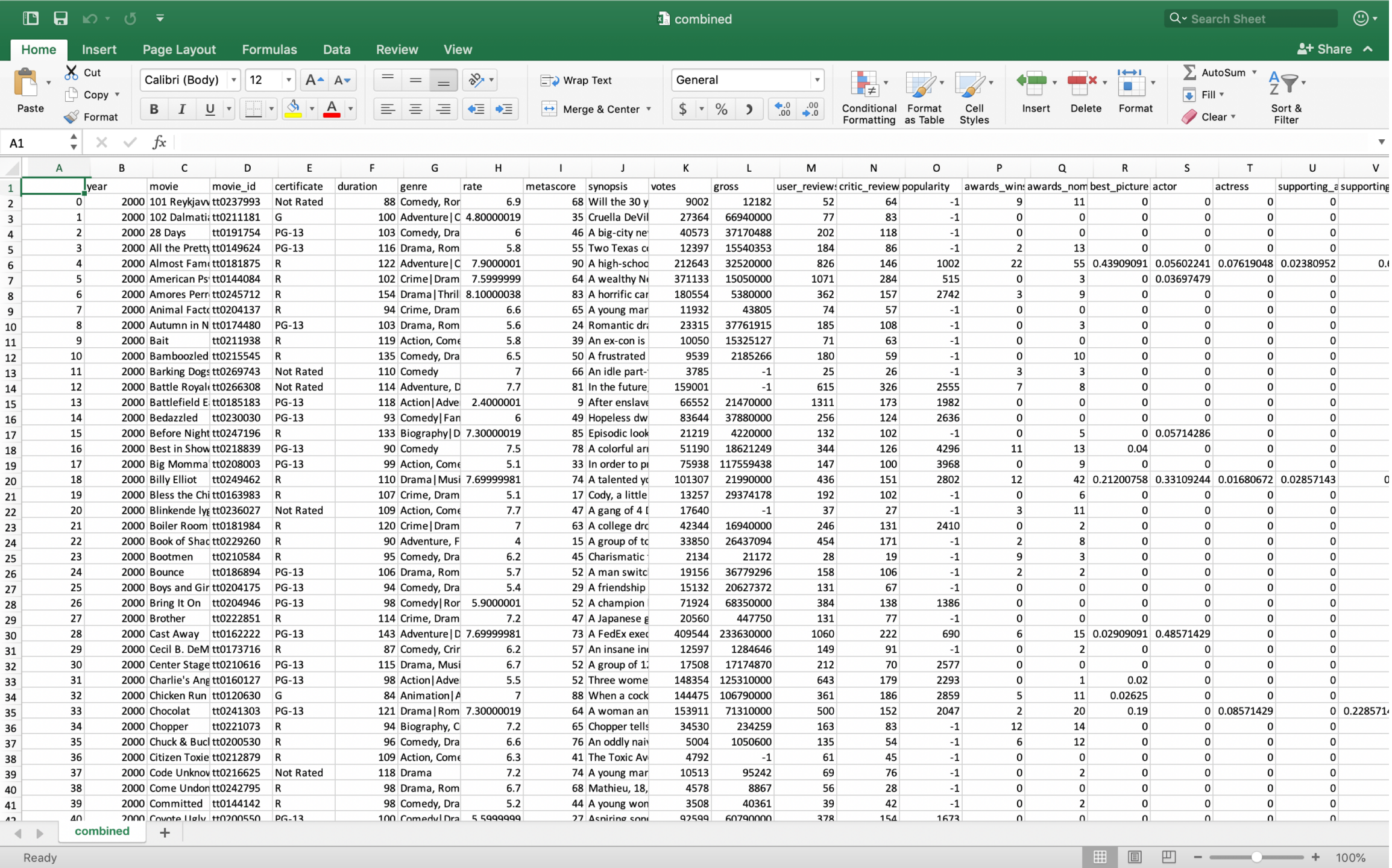
- Web-scraped input variables / details for all the movies collected (duration, certificate, IMDB rating, metascore, number of critic reviews, etc)
- Web-scraped results on how movies did at previous award ceremonies (Golden Globes, BAFTA, Producers Guild, OFTA, etc)
- Made point system for awards where winner = 1 point, nominee = 1/(number of nominees) point, nothing = 0 points
    - Incorporated into input feature vector

Week 3: Data Manipulation + Algorithm

- Incorporated/organized all the web-scraped data into a dataset
- Got rid of unnecessary data
- Tried various ML classifiers: SVM, Decision Trees, Random Forests
- Decided on neural network config

Week 4: Algorithm + Optimization

- Not working properly w/ predicting results for 24 Oscar categories → simple version: win, nominee, or nothing?
- Narrow down prediction to 6 Oscar categories due to lack of data
- Dealing with imbalanced dataset, trying different model architectures (hidden layer size, loss functions, class weights, activation functions, etc)



combined

Search Sheet

Home Insert Page Layout Formulas Data Review View Share

Paste Cut Copy Format

Calibri (Body) 12

Wrap Text Merge & Center

General

Conditional Formatting Format as Table Cell Styles

Insert Delete Format

AutoSum Fill Clear

Sort & Filter

A1

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V
1		year	movie	movie_id	certificate	duration	genre	rate	metascore	synopsis	votes	gross	user_review:	critic_review	popularity	awards_win:	awards_nom	best_picture	actor	actress	supporting_a	supporting
2	0	2000	101 Reykjavv	tt0237993	Not Rated	88	Comedy, Ror	6.9	68	Will the 30 y	9002	12182	52	64	-1	9	11	0	0	0	0	
3	1	2000	102 Dalmati	tt0211181	G	100	Adventure|C	4.80000019	35	Cruella DeVil	27364	66940000	77	83	-1	0	0	0	0	0	0	
4	2	2000	28 Days	tt0191754	PG-13	103	Comedy, Dra	6	46	A big-city ne	40573	37170488	202	118	-1	0	0	0	0	0	0	
5	3	2000	All the Pretty	tt0149624	PG-13	116	Drama, Rom	5.8	55	Two Texas c	12397	15540353	184	86	-1	2	13	0	0	0	0	
6	4	2000	Almost Fam	tt0181875	R	122	Adventure|C	7.9000001	90	A high-schoo	212643	32520000	826	146	1002	22	55	0.43909091	0.05602241	0.07619048	0.02380952	0.0
7	5	2000	American Ps	tt0144084	R	102	Crime|Dram	7.5999999	64	A wealthy Ne	371133	15050000	1071	284	515	0	3	0	0.03697479	0	0	
8	6	2000	Amores Perr	tt0245712	R	154	Drama|Thril	8.10000038	83	A horrific car	180554	5380000	362	157	2742	3	9	0	0	0	0	
9	7	2000	Animal Facto	tt0204137	R	94	Crime, Dram	6.6	65	A young mar	11932	43805	74	57	-1	0	0	0	0	0	0	
10	8	2000	Autumn in N	tt0174480	PG-13	103	Drama, Rom	5.6	24	Romantic dra	23315	37761915	185	108	-1	0	3	0	0	0	0	
11	9	2000	Bait	tt0211938	R	119	Action, Come	5.8	39	An ex-con is	10050	15325127	71	63	-1	0	0	0	0	0	0	
12	10	2000	Bamboozled	tt0215545	R	135	Comedy, Dra	6.5	50	A frustrated	9539	2185266	180	59	-1	0	10	0	0	0	0	
13	11	2000	Barking Dogs	tt0269743	Not Rated	110	Comedy	7	66	An idle part-	3785	-1	25	26	-1	3	3	0	0	0	0	
14	12	2000	Battle Royal	tt0266308	Not Rated	114	Adventure, D	7.7	81	In the future,	159001	-1	615	326	2555	7	8	0	0	0	0	
15	13	2000	Battlefield E	tt0185183	PG-13	118	Action|Adve	2.4000001	9	After enslave	66552	21470000	1311	173	1982	0	0	0	0	0	0	
16	14	2000	Bedazzled	tt0230030	PG-13	93	Comedy|Fan	6	49	Hopeless dw	83644	37880000	256	124	2636	0	0	0	0	0	0	
17	15	2000	Before Night	tt0247196	R	133	Biography|D	7.30000019	85	Episodic look	21219	4220000	132	102	-1	0	5	0	0.05714286	0	0	
18	16	2000	Best in Show	tt0218839	PG-13	90	Comedy	7.5	78	A colorful arı	51190	18621249	344	126	4296	11	13	0.04	0	0	0	
19	17	2000	Big Momma	tt0208003	PG-13	99	Action, Come	5.1	33	In order to pı	75938	117559438	147	100	3968	0	9	0	0	0	0	
20	18	2000	Billy Elliot	tt0249462	R	110	Drama|Musi	7.69999981	74	A talented yc	101307	21990000	436	151	2802	12	42	0.21200758	0.33109244	0.01680672	0.02857143	0
21	19	2000	Bless the Chi	tt0163983	R	107	Crime, Dram	5.1	17	Cody, a little	13257	29374178	192	102	-1	0	6	0	0	0	0	
22	20	2000	Blinkende lyg	tt0236027	Not Rated	109	Action, Come	7.7	47	A gang of 4 [	17640	-1	37	27	-1	3	11	0	0	0	0	
23	21	2000	Boiler Room	tt0181984	R	120	Crime|Dram	7	63	A college dro	42344	16940000	246	131	2410	0	2	0	0	0	0	
24	22	2000	Book of Shac	tt0229260	R	90	Adventure, F	4	15	A group of tc	33850	26437094	454	171	-1	2	8	0	0	0	0	
25	23	2000	Bootmen	tt0210584	R	95	Comedy, Dra	6.2	45	Charismatic	2134	21172	28	19	-1	9	3	0	0	0	0	
26	24	2000	Bounce	tt0186894	PG-13	106	Drama, Rom	5.7	52	A man switc	19156	36779296	158	106	-1	2	2	0	0	0	0	
27	25	2000	Boys and Gir	tt0204175	PG-13	94	Comedy, Dra	5.4	29	A friendship	15132	20627372	131	67	-1	0	0	0	0	0	0	
28	26	2000	Bring It On	tt0204946	PG-13	98	Comedy|Ror	5.9000001	52	A champion	71924	68350000	384	138	1386	0	0	0	0	0	0	
29	27	2000	Brother	tt0222851	R	114	Crime, Dram	7.2	47	A Japanese g	20560	447750	131	77	-1	0	0	0	0	0	0	
30	28	2000	Cast Away	tt0162222	PG-13	143	Adventure|D	7.69999981	73	A FedEx exec	409544	233630000	1060	222	690	6	15	0.02909091	0.48571429	0	0	
31	29	2000	Cecil B. DeM	tt0173716	R	87	Comedy, Crir	6.2	57	An insane in	12597	1284646	149	91	-1	0	2	0	0	0	0	
32	30	2000	Center Stage	tt0210616	PG-13	115	Drama, Musi	6.7	52	A group of 1	17508	17174870	212	70	2577	0	0	0	0	0	0	
33	31	2000	Charlie's Ang	tt0160127	PG-13	98	Action|Adve	5.5	52	Three wome	148354	125310000	643	179	2293	0	1	0.02	0	0	0	
34	32	2000	Chicken Run	tt0120630	G	84	Animation|A	7	88	When a cock	144475	106790000	361	186	2859	5	11	0.02625	0	0	0	
35	33	2000	Chocolat	tt0241303	PG-13	121	Drama|Rom	7.30000019	64	A woman an	153911	71310000	500	152	2047	2	20	0.19	0	0.08571429	0	0.228571
36	34	2000	Chopper	tt0221073	R	94	Biography, C	7.2	65	Chopper tells	34530	234259	163	83	-1	12	14	0	0	0	0	
37	35	2000	Chuck & Buc	tt0200530	R	96	Comedy, Dra	6.6	76	An oddly naiv	5004	1050600	135	54	-1	6	12	0	0	0	0	
38	36	2000	Citizen Toxie	tt0212879	R	109	Action, Come	6.3	41	The Toxic Av	4792	-1	61	45	-1	0	0	0	0	0	0	
39	37	2000	Code Unknov	tt0216625	Not Rated	118	Drama	7.2	74	A young mar	10513	95242	69	76	-1	0	2	0	0	0	0	
40	38	2000	Come Undon	tt0242795	R	98	Drama, Rom	6.7	68	Mathieu, 18,	4578	8867	56	28	-1	0	0	0	0	0	0	
41	39	2000	Committed	tt0144142	R	98	Comedy, Dra	5.2	44	A young wor	3508	40361	39	42	-1	0	2	0	0	0	0	
42	40	2000	Coyote Ugly	tt0200550	PG-13	100	Comedy|Dra	5.5999999	27	Aspiring son	92599	60790000	378	154	1673	0	0	0	0	0	0	

combined

Ready

100%

- Always start with a simpler problem
- Importance of data analysis
- Ask as many questions as possible
- Should keep good documentation
- Chances are, someone already did it

INPUT: 21-dimensional feature vector

certificate, duration, IMDB rating, metascore, number of votes, USA gross, # of user/critic reviews, # of non-Oscar wins/nominations, and award points for the 6 categories

OUTPUT: 6 classes for each category

Best Motion Picture of the Year

WINNER

Green Book
Jim Burke, Charles B. Wessler, Brian Hayes Currie, Peter Farrelly, Nick Vallelonga

A Star Is Born
Bill Gerber, Bradley Cooper, Lynette Howell Taylor

BlacKkKlansman
Sean McKittrick, Jason Blum, Raymond Mansfield, Jordan Peele, Spike Lee

Black Panther
Kevin Feige

Bohemian Rhapsody
Graham King

Roma
Gabriela Rodriguez, Alfonso Cuarón

The Favourite
Ceci Dempsey, Ed Guiney, Lee Magiday, Yorgos Lanthimos

Vice
Dede Gardner, Jeremy Kleiner, Adam McKay, Kevin J. Messick

Best Performance by an Actor in a Leading Role

WINNER

Rami Malek
Bohemian Rhapsody

Bradley Cooper
A Star Is Born

Christian Bale
Vice

Viggo Mortensen
Green Book

Willem Dafoe
At Eternity's Gate

Best Performance by an Actress in a Leading Role

WINNER

Olivia Colman
The Favourite

Glenn Close
The Wife

Lady Gaga
A Star Is Born

Melissa McCarthy
Can You Ever Forgive Me?

Yalitza Aparicio
Roma

Best Performance by an Actor in a Supporting Role

WINNER

Mahershala Ali
Green Book

Adam Driver
BlacKkKlansman

Richard E. Grant
Can You Ever Forgive Me?

Sam Elliott
A Star Is Born

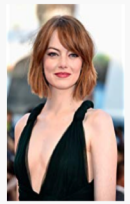
Sam Rockwell
Vice

Best Performance by an Actress in a Supporting Role

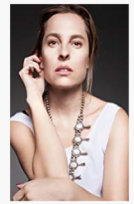
WINNER

Regina King
If Beale Street Could Talk

Amy Adams
Vice



Emma Stone
The Favourite



Marina de Tavi
Roma

Rachel Weisz
The Favourite

Best Animated Feature Film

WINNER

Spider-Man: Into the Spider-Verse
Bob Persichetti, Peter Ramsey, Rodney Rothman, Phil Lord, Christopher Miller

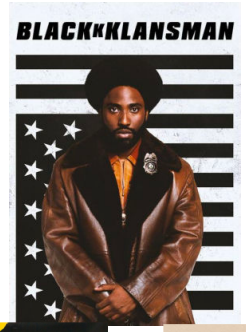
Incredibles 2
Brad Bird, John V
Paradis Grindle

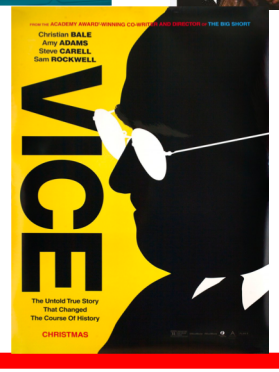
Isle of Dogs
Wes Anderson, Scott Rudin, Steven Rales, Jeremy Dawson

Mirai
Mirai no Mirai (origi
Mamoru Hosoda,

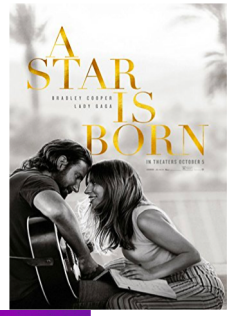
Ralph Breaks the Internet
Rich Moore, Phil Johnston, Clark Spencer

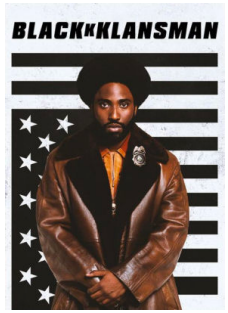
Best Picture

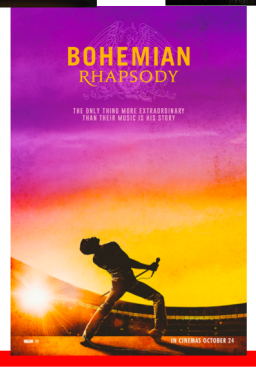




Best Leading Actor

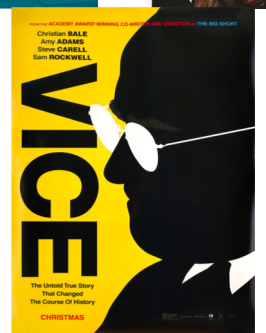




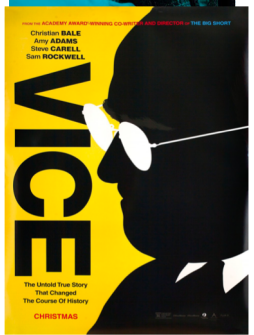


Best Leading Actress

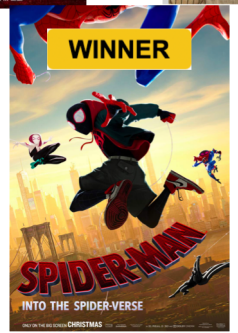
Best Supporting Actor



Best Supporting Actress



Best Animated Film



https://github.com/j0hnk1m/predict-the-oscars		https://github.com/j0hnk1m/predict-the-oscars
https://github.com/j0hnk1m/predict-the-oscars		https://github.com/j0hnk1m/predict-the-oscars
https://github.com/j0hnk1m/predict-the-oscars		https://github.com/j0hnk1m/predict-the-oscars
https://github.com/j0hnk1m/predict-the-oscars		https://github.com/j0hnk1m/predict-the-oscars
https://github.com/j0hnk1m/predict-the-oscars		https://github.com/j0hnk1m/predict-the-oscars
https://github.com/j0hnk1m/predict-the-oscars		https://github.com/j0hnk1m/predict-the-oscars
https://github.com/j0hnk1m/predict-the-oscars		https://github.com/j0hnk1m/predict-the-oscars
https://github.com/j0hnk1m/predict-the-oscars		https://github.com/j0hnk1m/predict-the-oscars
https://github.com/j0hnk1m/predict-the-oscars		https://github.com/j0hnk1m/predict-the-oscars
https://github.com/j0hnk1m/predict-the-oscars		https://github.com/j0hnk1m/predict-the-oscars
https://github.com/j0hnk1m/predict-the-oscars		https://github.com/j0hnk1m/predict-the-oscars
https://github.com/j0hnk1m/predict-the-oscars		https://github.com/j0hnk1m/predict-the-oscars
https://github.com/j0hnk1m/predict-the-oscars		https://github.com/j0hnk1m/predict-the-oscars
https://github.com/j0hnk1m/predict-the-oscars		https://github.com/j0hnk1m/predict-the-oscars
https://github.com/j0hnk1m/predict-the-oscars		https://github.com/j0hnk1m/predict-the-oscars
https://github.com/j0hnk1m/predict-the-oscars		https://github.com/j0hnk1m/predict-the-oscars
https://github.com/j0hnk1m/predict-the-oscars		https://github.com/j0hnk1m/predict-the-oscars