

Hollywood Hits of the Future

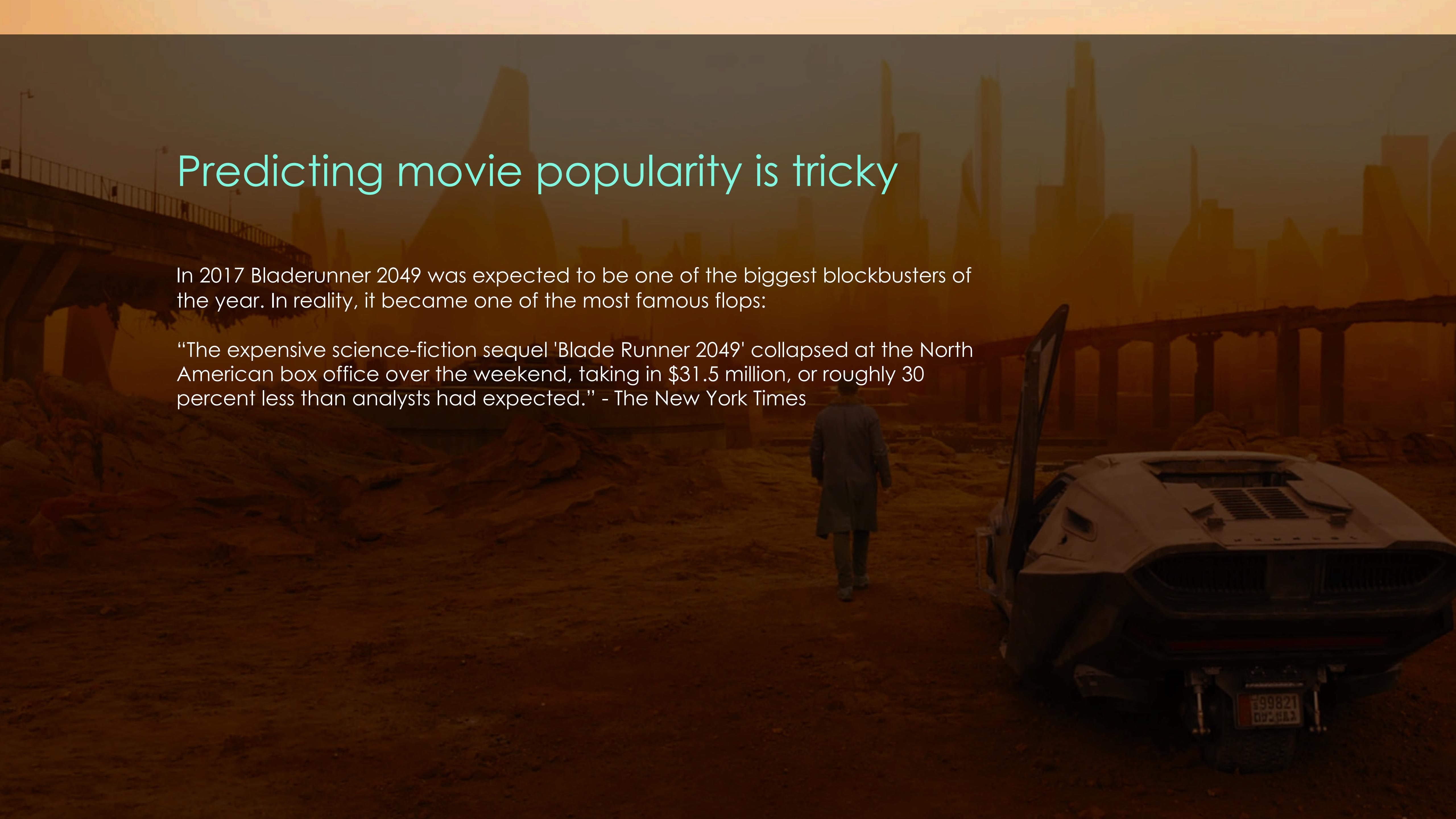
GT Data Bootcamp | Final Project | J-Cubed
Jennie Lee Gruber, Hong Leatheng, Justin Mahler, Jake Stanford



Predicting movie popularity is tricky

In 2017 Bladerunner 2049 was expected to be one of the biggest blockbusters of the year. In reality, it became one of the most famous flops:

“The expensive science-fiction sequel 'Blade Runner 2049' collapsed at the North American box office over the weekend, taking in \$31.5 million, or roughly 30 percent less than analysts had expected.” - The New York Times



Personalised
recommendations

Origin Story

To build our machine learning model we carefully considered what datasets we wanted to use for training. This extensive trial & error process included the following steps.

Utelly API

The Movie DB API

IMDb: Recently Releases data

Kaggle's Streaming Movies dataset

Initial input considerations:

Consumer ratings (popularity, vote average)

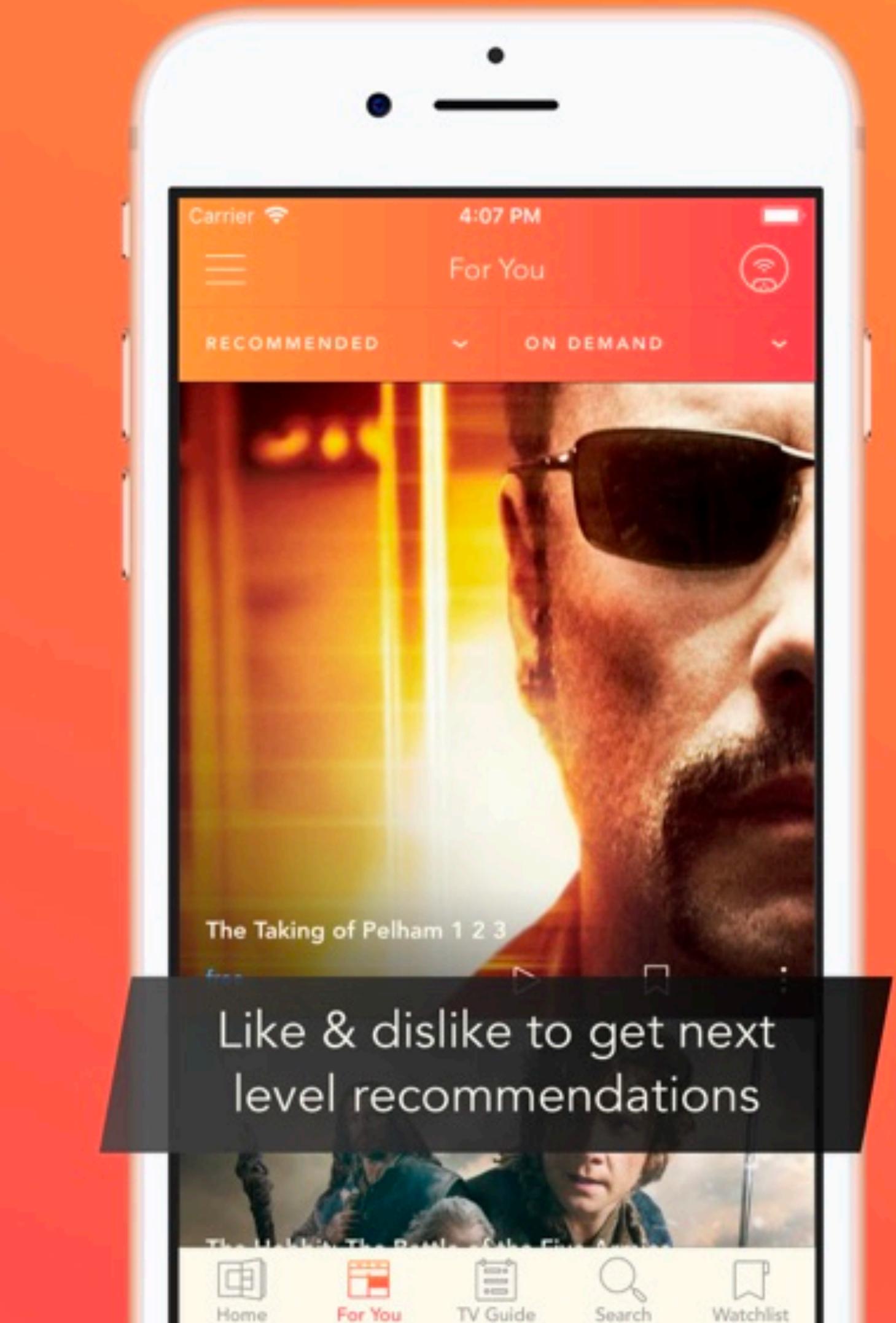
Original language

Release data

Streaming availability

Budget

Revenue



Deciding on our model's predictive powers

Inputs:

Popularity

Vote count

Original language (en: y/n)

Release year

Release month

Release day

Revenue generation

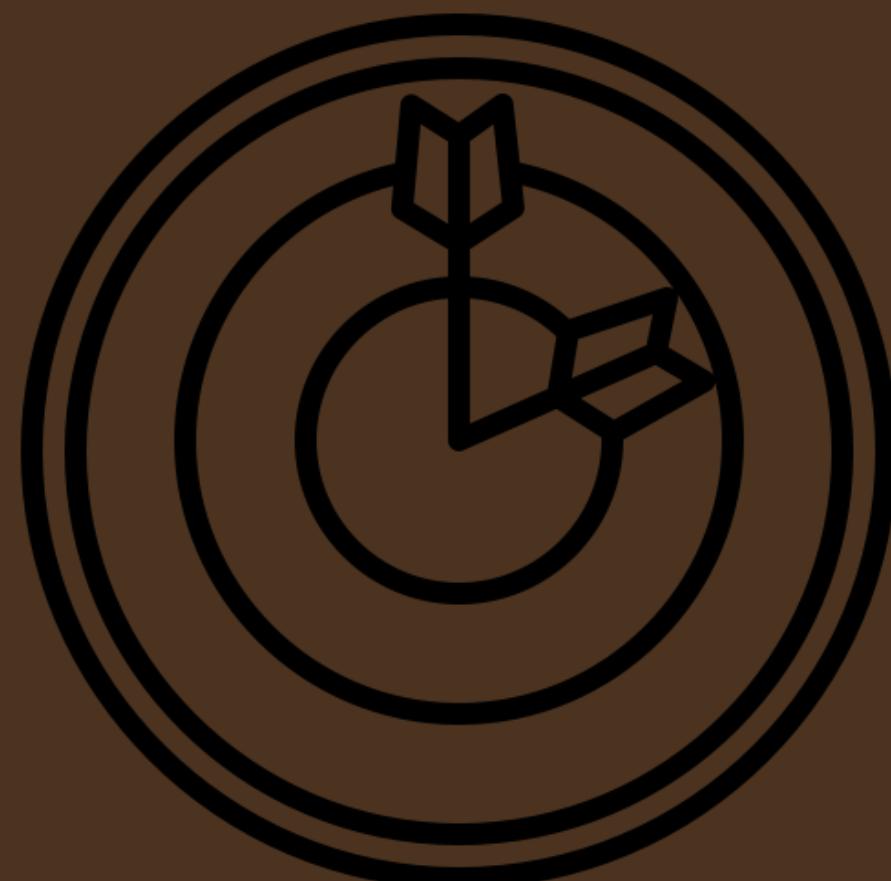
~~Accessibility on streaming site(s)~~

Popularity

Will a future move exceed a 6-point average consumer vote? (based on a 10pt scale. y: 1; n: 0)

Picking the appropriate model

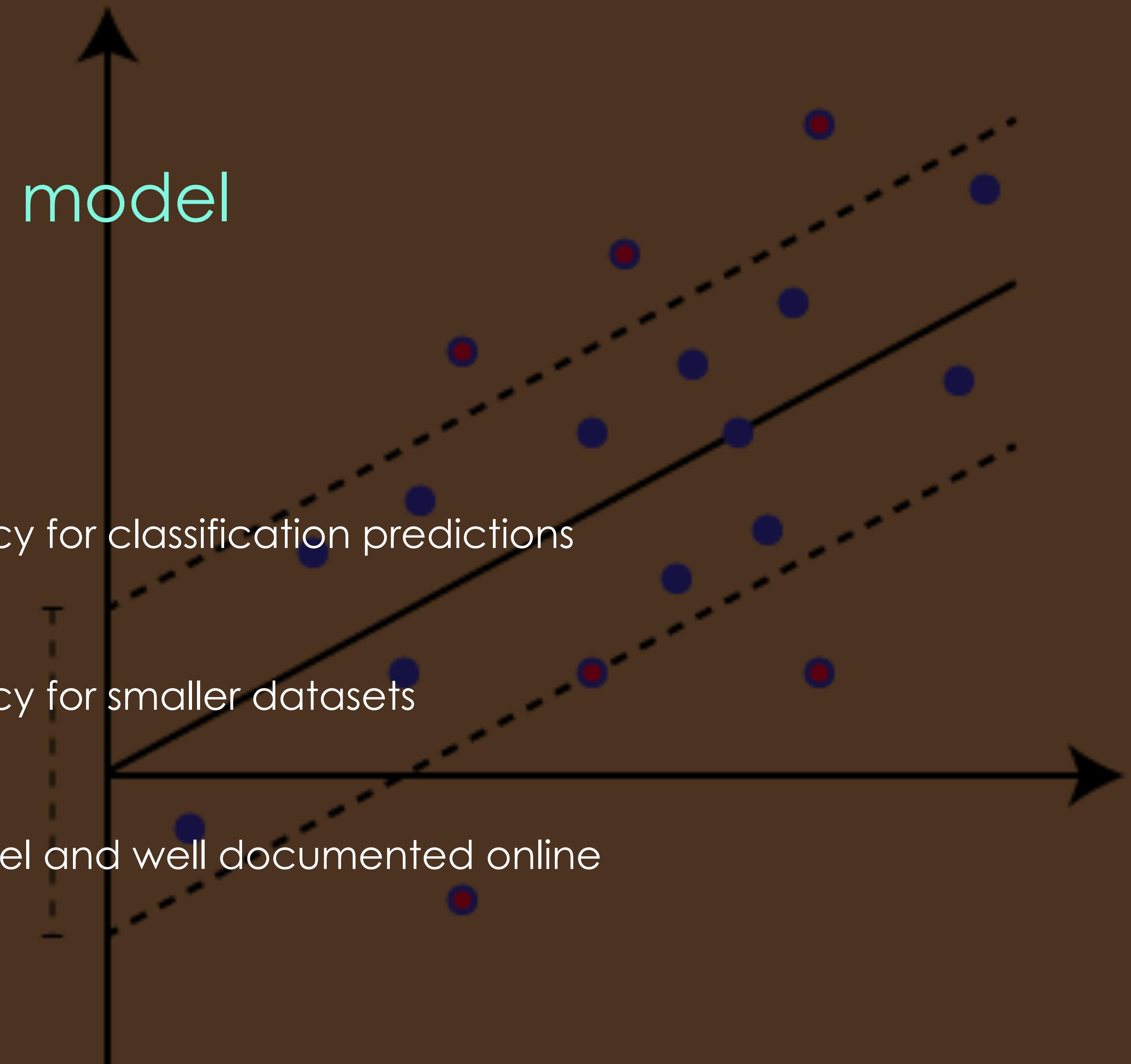
Why we chose to use a SVM model:



High accuracy for classification predictions

High accuracy for smaller datasets

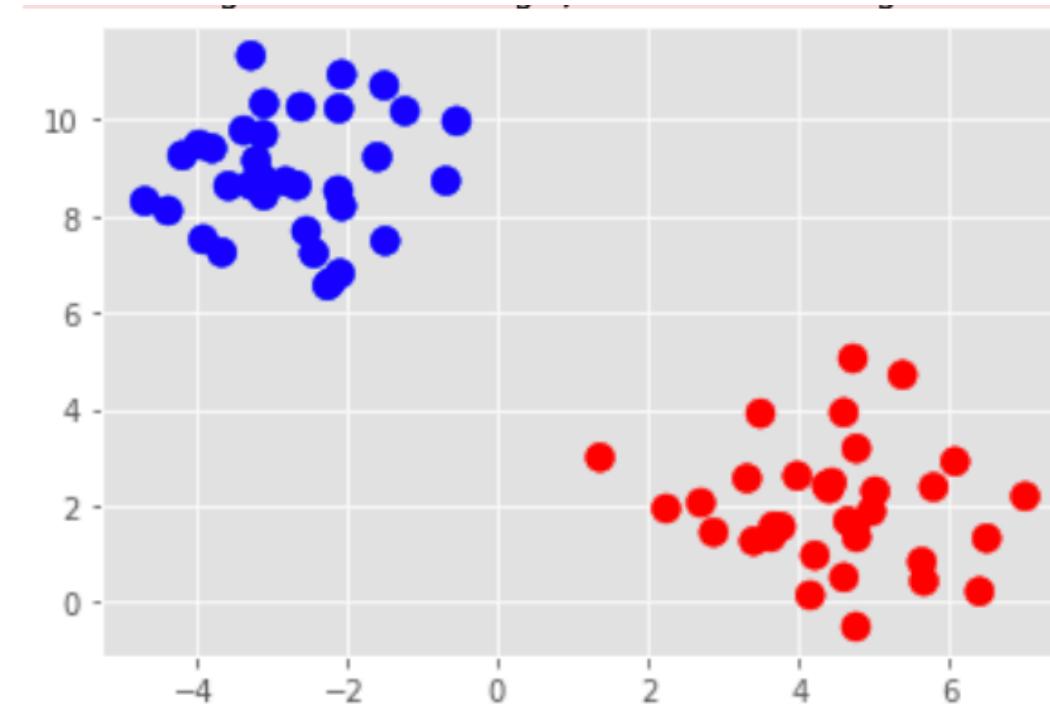
Popular model and well documented online



Fitting & testing our model

	Precision	Recall	F1-score	Support
0	0.00	0.00	0.00	4
1	0.76	1.00	0.87	13

	Accuracy			
Macro Average	0.38	0.50	0.76	17
Weighted Average	0.58	0.76	0.66	17



	actual_y	predicted_y
17	1	1
183	1	1
84	1	1
37	1	1
15	1	1
9	1	1
78	0	1

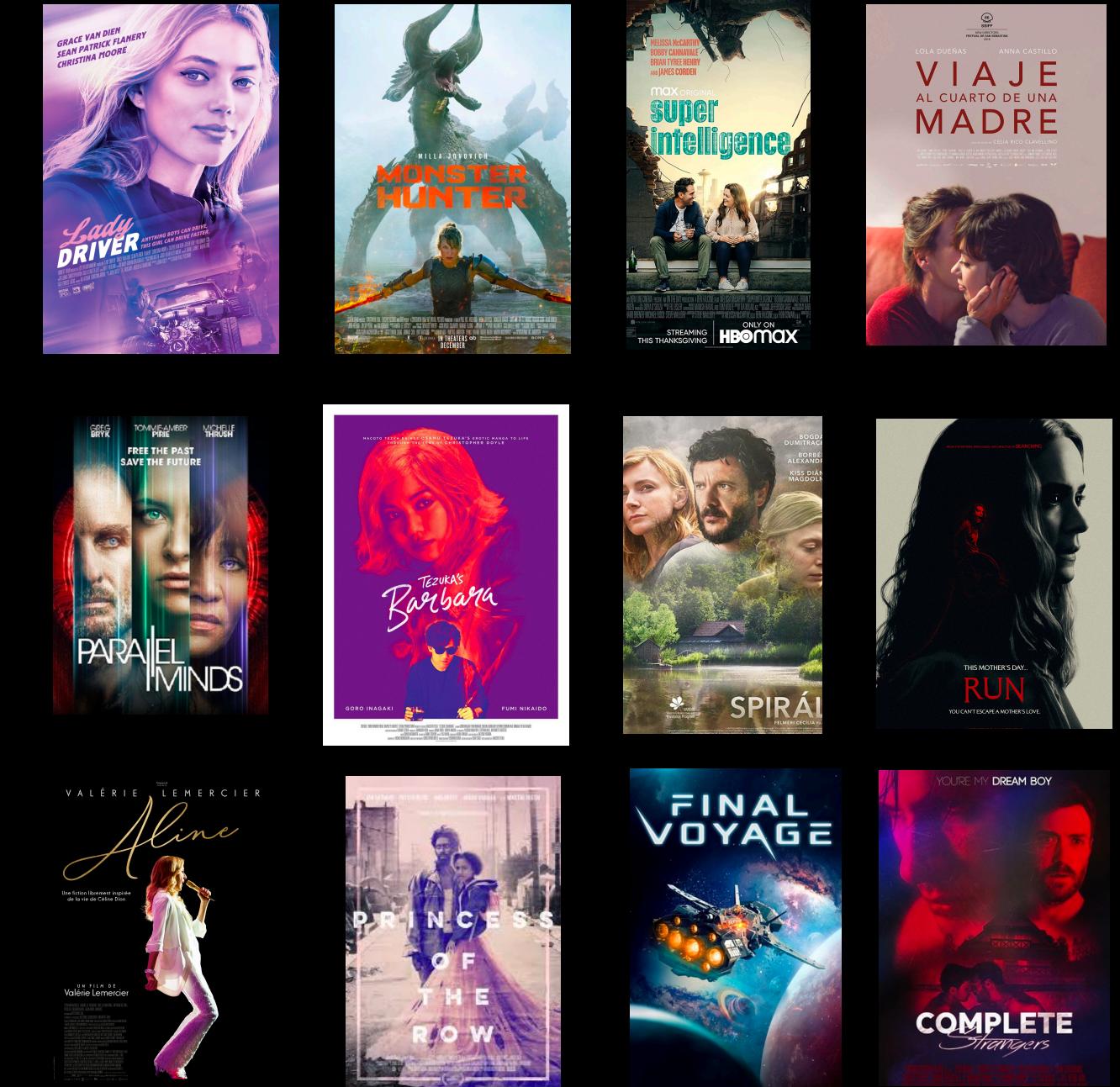
Future movie predictions

	id	title	popularity	en = lang	release_month	release_day	release_year	Output
6	591629	Lady Driver	267.579	1	12	3	2020	1
26	458576	Monster Hunter	53.043	1	12	3	2020	1
31	521007	Superintelligence	33.184	1	11	26	2020	1
49	314514	El Viaje de Ana	24.098	1	11	18	2020	1
58	502033	Sound of Metal	15.446	1	11	20	2020	1
63	426793	Parallel	16.473	1	12	11	2020	1
88	589456	Find Your Voice	10.972	0	11	13	2020	1
89	592777	Should the Wind Drop	15.249	1	11	18	2020	1
92	514466	Supercool	9.298	1	11	28	2020	1
94	353047	Zappa	14.826	1	11	27	2020	1
101	522772	Sk8r Grrl	12.022	1	11	22	2020	1
108	532961	Looking for Witch Apprentices	9.782	1	11	13	2020	1
109	581734	Nomadland	13.468	1	11	30	2020	1
125	572488	A Christmas Carol	10.626	1	12	4	2020	1
129	566135	Tezuka's Barbara	11.448	1	11	20	2020	1
130	583754	Young Blooded Wine	11.291	1	11	19	2020	1
131	583757	Princess of the Row	9.078	1	11	27	2020	1
137	530368	Spirál	10.769	1	11	19	2020	1
151	498402	The Power of Love (Famous)	5.632	1	11	11	2020	1
152	575072	Run	9.747	1	11	19	2020	1
155	570871	Final Voyage	8.181	0	12	3	2020	1
157	540023	Wander Darkly	9.446	1	12	11	2020	1
182	547565	The Night House	6.424	1	10	1	2020	1
194	583293	Complete Strangers	6.260	1	12	1	2020	1

Predicted Average Vote Scores

>6.0 6.0+

None





Trends & analysis >>>