Kickstarter

Predicting Campaign Success

Business Case

1. Help Kickstarter inform their users how to better structure and set up their campaigns.
2. Identify Campaigns that have a high chance of succeeding. That way, these campaigns can be promoted to as many potential backers as possible.

Exploratory Data Analysis

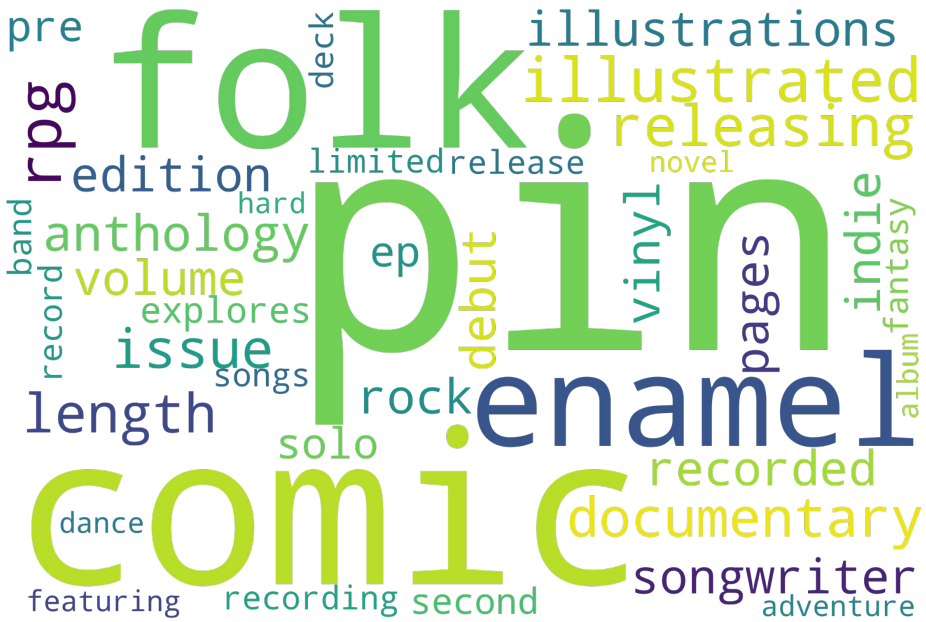
Extracting markers of success

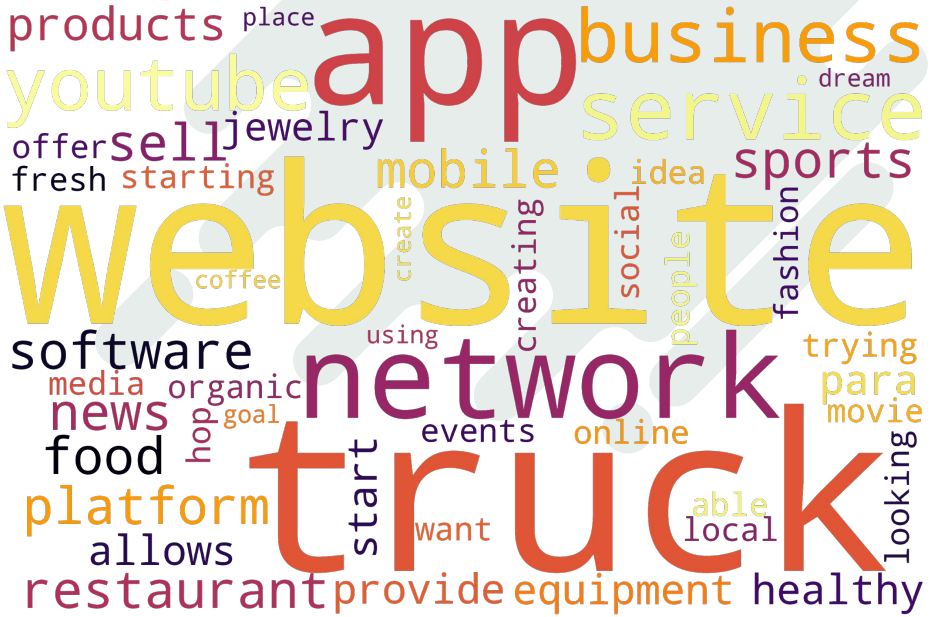
The language of success, extracted using multinomial Naive Bayes

“The Weekend Press is embarking on its first **adventure**: an ode to Oakland in 10 **limited-run** prints from 10 incredible Bay Area artists.”

“I'm **trying** to start a business **selling** coffee, smoothies, and food. I **want** to better my children's lives and follow my dream.”

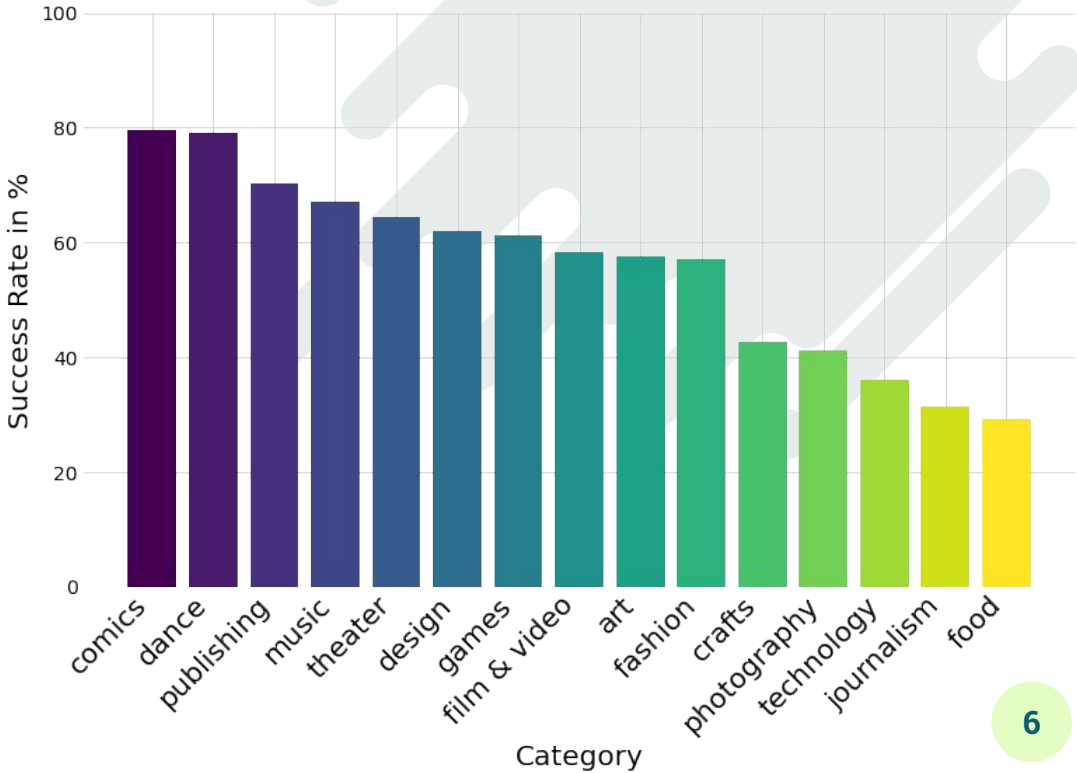
The language of success | | The language of failure





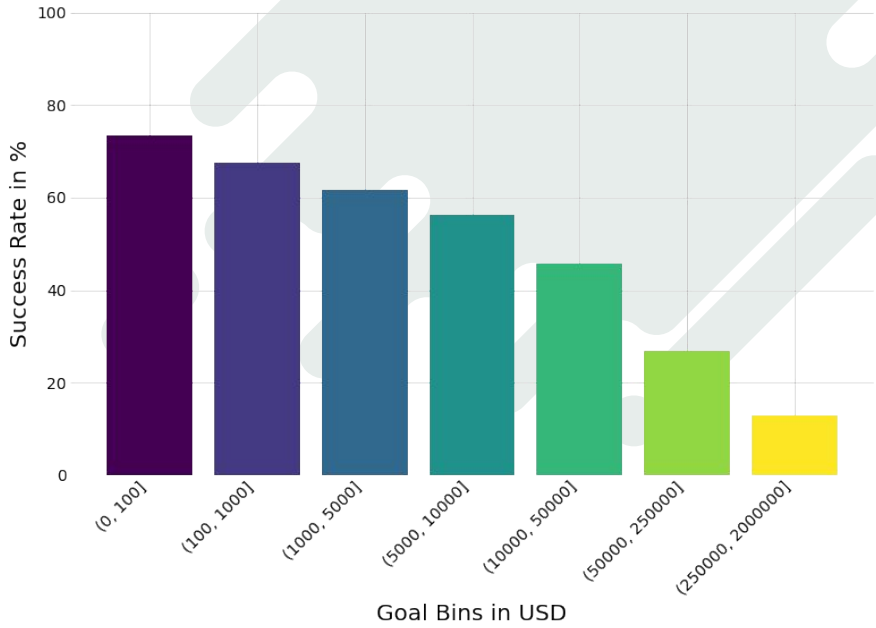
Success depends on the category

category	success_count	fail_count	percent_successful
comics	5057	1301	79.5
dance	2384	625	79.2
publishing	11942	5055	70.3
music	16538	8053	67.3
theater	3019	1667	64.4
design	3651	2232	62.1
games	6577	4158	61.3
film & video	13963	9926	58.4
art	10418	7643	57.7
fashion	5067	3802	57.1
crafts	2325	3123	42.7
photography	2348	3341	41.3
technology	6421	11361	36.1
journalism	1159	2525	31.5
food	3910	9388	29.4



High goals are more difficult to realize

goal_bin	success_count	failed_count	percent_successful
(0, 100]	2802	1012	73.5
(100, 1000]	20964	10040	67.6
(1000, 5000]	34305	21321	61.7
(5000, 10000]	15343	11960	56.2
(10000, 50000]	16437	19437	45.8
(50000, 250000]	2402	6501	27.0
(250000, 2000000]	218	1461	13.0



New and improved features

- Time between campaign creation and launch
- Time between launch and deadline
- Name and blurb word count
- Language of success and failure




Modeling

Predicting Campaign Success

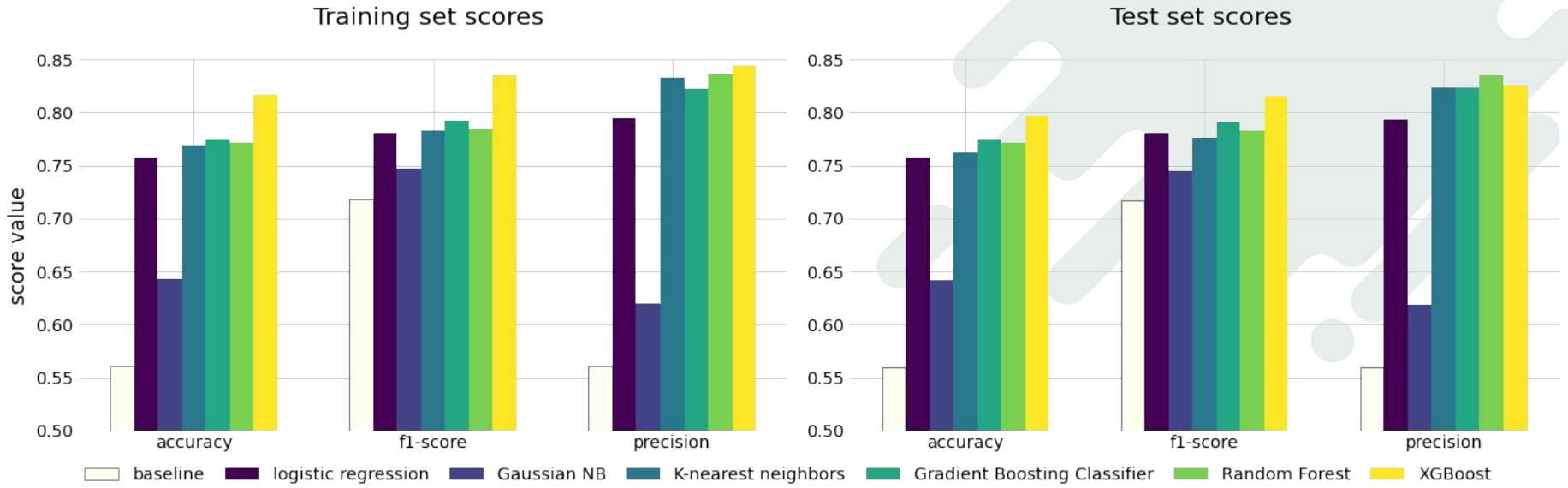
Predicting Campaign Success

- 6 machine learning models fitted
- baseline model as benchmark
- benchmark model always predicts most frequent outcome

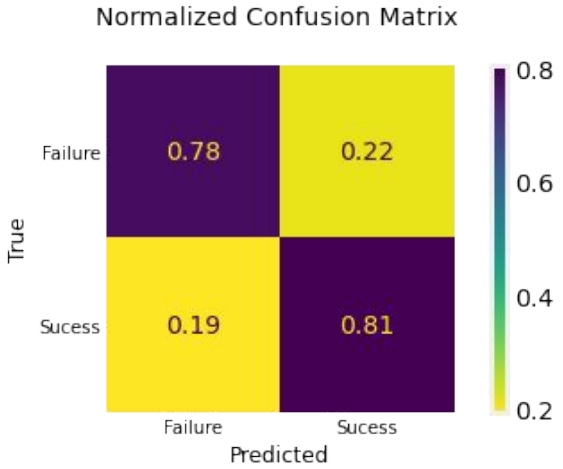
How we chose the right model

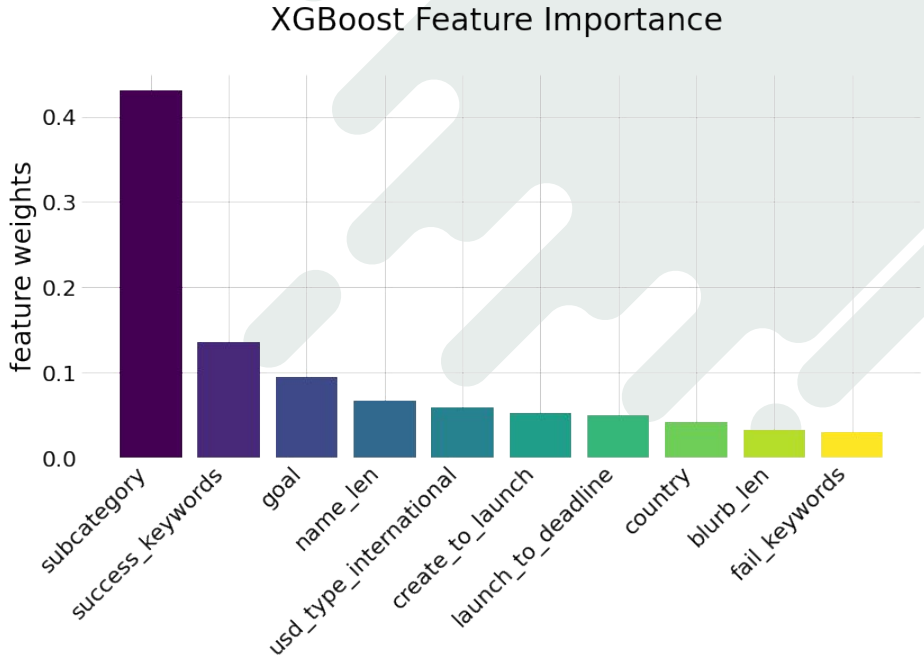
- if the model falsely predicts success, Kickstarter will have wasted resources promoting the campaign
- maximize rate of true positives to false positives
- evaluation metric: **precision**

Model Comparison

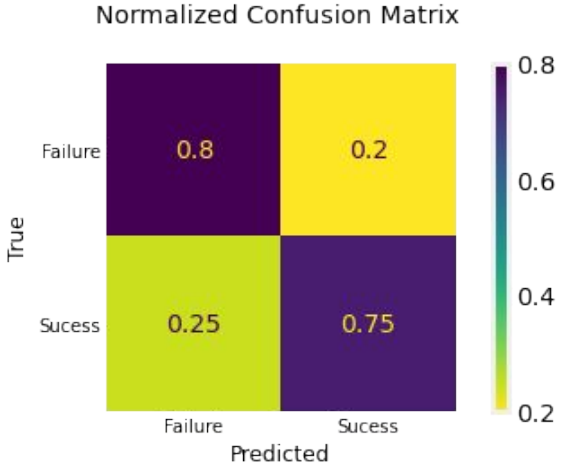


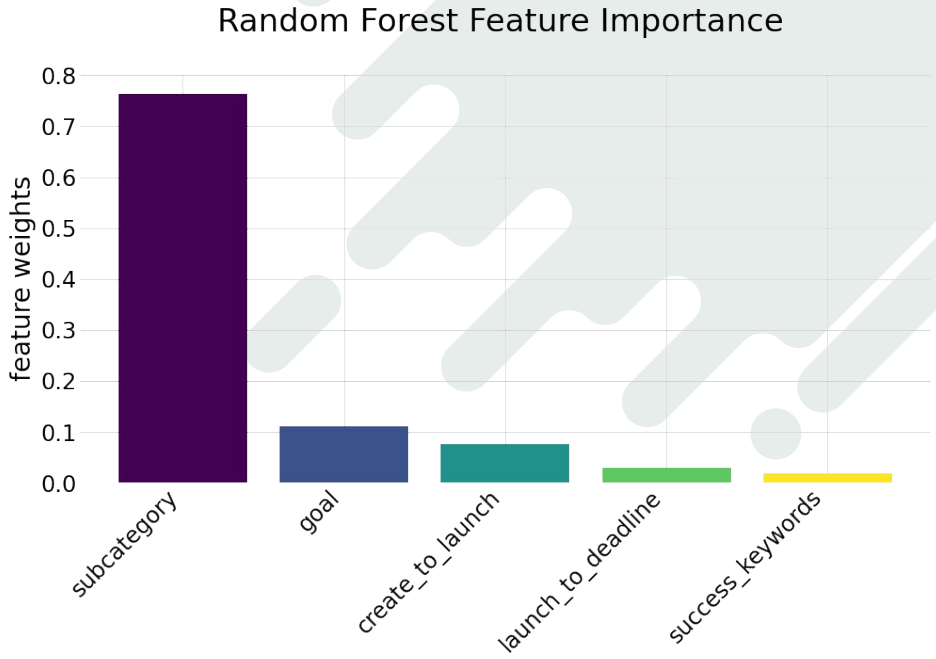
Model Summary (XGBoost)





Model Summary (Random Forest)





Thank you for your attention!

google slides design template from slidescarnival.com
(https://www.slidescarnival.com/minola-free-presentation-template/11829)