



Kickstarter Success Predictor

...

Omar Qusous

What can you do to increase the project's chances of success?

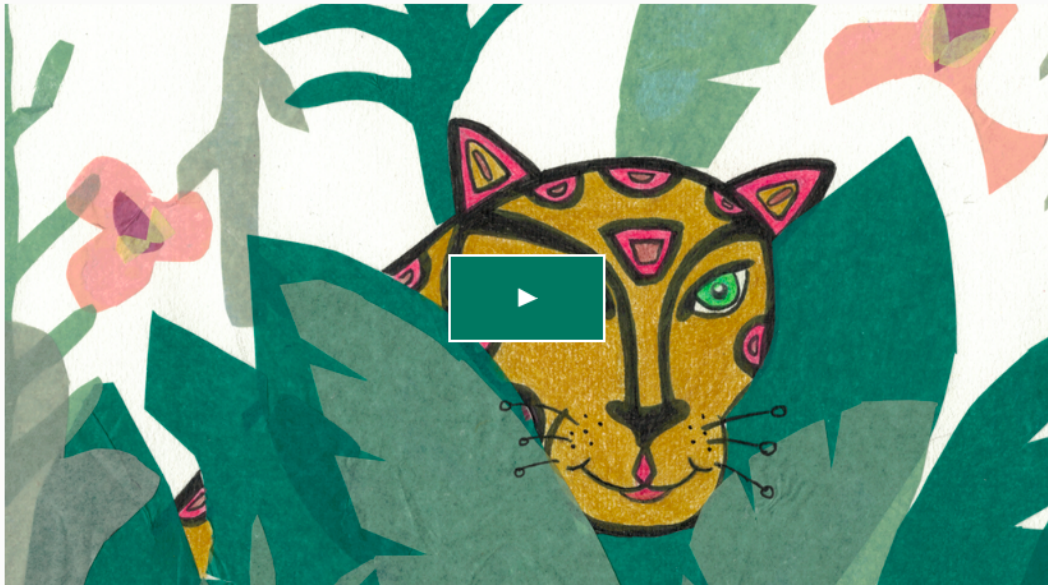
[Explore](#) [Start a project](#)

KICKSTARTER

[Search](#) [Log in](#)

Omar, el Jaguar: A book that celebrates diversity

Learning and teaching Spanish just jumped to the next level with this beautifully illustrated picture book and its resources.



\$1,257
pledged of \$3,000 goal

37
backers

21
days to go

Back this project

[♥ Remind me](#) [f](#) [t](#) [✉](#) [</>](#)

All or nothing. This project will only be funded if it reaches its goal by Thu, October 10 2019 12:58 PM EDT.

Data Source 1

- <https://webrobots.io/kickstarter-datasets/>
- Total of 9338 points. Aim to keep 5000 or more (duplicates, Nans, etc...)

Data Source 2

Feature Selection:

Repeated information:

Data leakage:

Not useful (logic):

Key:

All NAN:

['backers_count', 'blurb', 'category', 'converted_pledged_amount', 'country', 'created_at', 'creator', 'currency', 'currency_symbol', 'currency_trailing_code', 'current_currency', 'deadline', 'disable_communication', 'friends', 'fx_rate', 'goal', 'id', 'is_backing', 'is_starred', 'launched_at', 'location', 'name', 'permissions', 'photo', 'pledged', 'profile', 'slug', 'source_url', 'spotlight', 'staff_pick', 'state', 'static_usd_rate', 'urls', 'usd_pledged', 'usd_type']

Final Features:

- category: Film, Music, Fashion etc.
- name: projects name
- staff_pick: projects highlighted on homepage
- goal: desired amount of money to succeed
- created_at: data of starting the campaign
- deadline: deadline set for achieving the desired amount of money

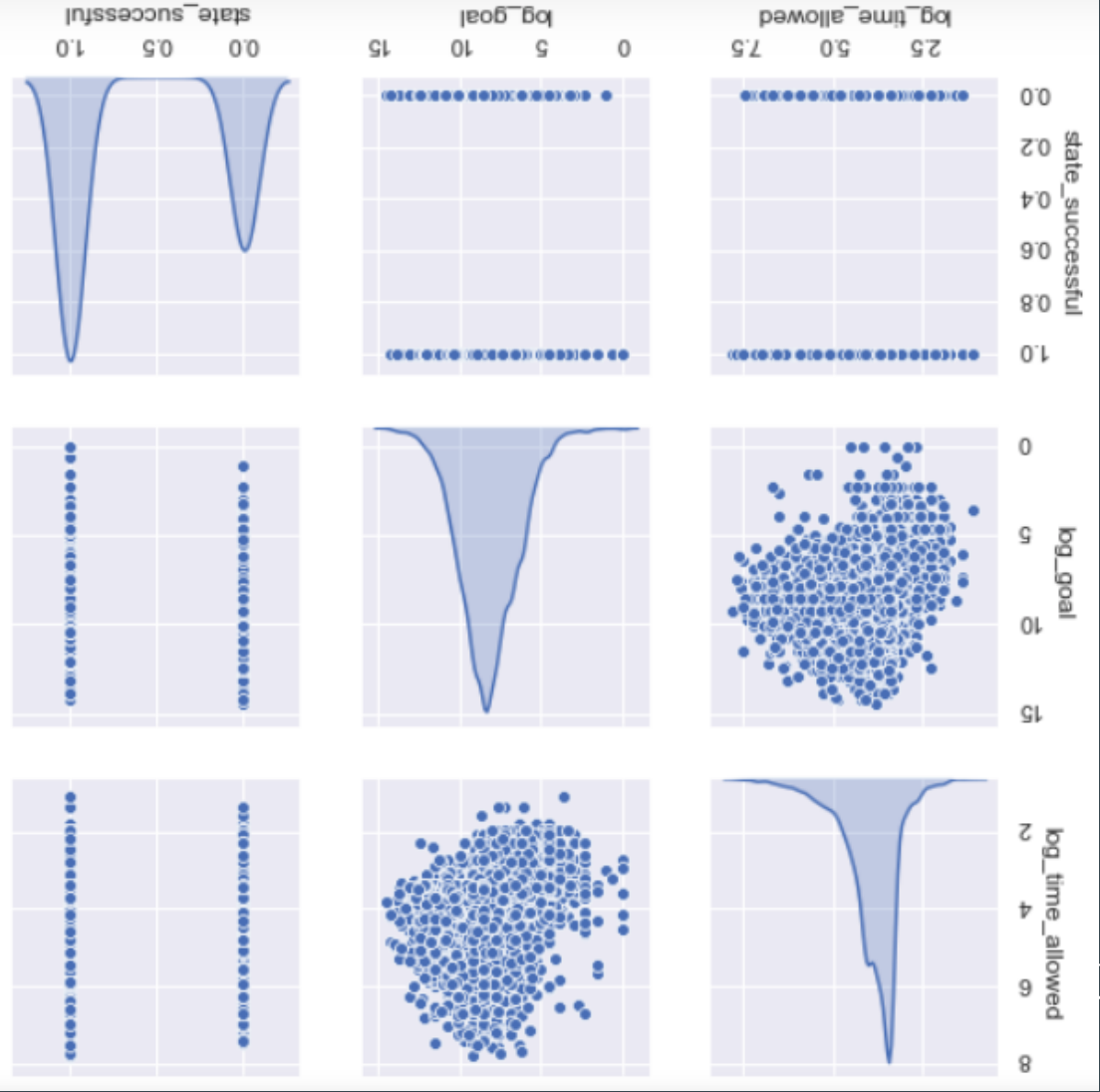
EDA and Feature Engineering 1

- Continuous Features:
 - Log of time allowed (deadline-created_at) and Log of 'goal'
 - Data was filtered so that time allowed is less than 5000 days and goal less than 2 million USD.
- Categorical Features:
 - 'category', 'staff_pick', 'country' were all converted to dummies

Total of 168 columns in X

-

EDA and Feature Engineering 2



Target

Target:

- 'State' : successful, failed, suspended, live, cancelled
- Eliminated live, cancelled and suspended

- 'Binary distribution (1 : success, 0 : failure)

Quick Interesting Stats 1

Most successful and unsuccessful project categories

state		cat_slug	
successful	failed	149.0	30.0
music/hip-hop		101.0	27.0
crafts/diy		74.0	59.0
technology/wearables		73.0	9.0
games/mobile games		70.0	27.0
technology/software			

state		cat_slug	
successful	failed	239.0	220.0
publishing/fiction		NaN	NaN
music/indie rock		23.0	219.0
fashion/accessories		27.0	155.0
film & video/narrative film		NaN	135.0
design/product design			

Quick Interesting Stats 2

Percentage Success/Failure by Continent

state	failed	successful	country
Aisa	28.0	72.0	NAmerica
	37.0	63.0	Euro
	41.0	59.0	Aus
	46.0	54.0	SAmerica
	51.0	49.0	

Average 'Goal' for successful and failed projects

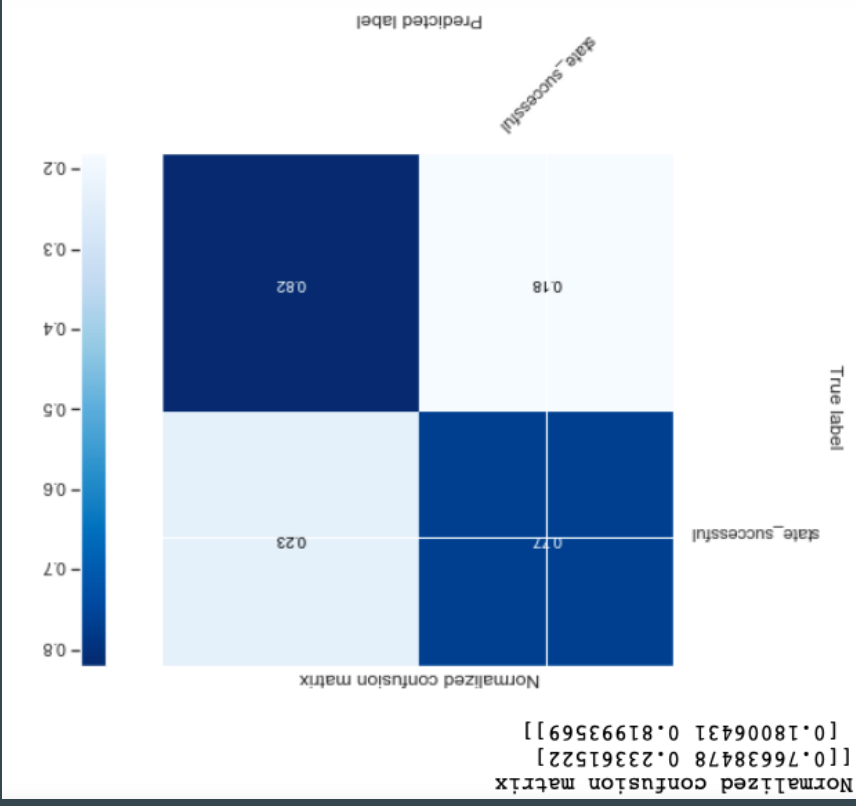
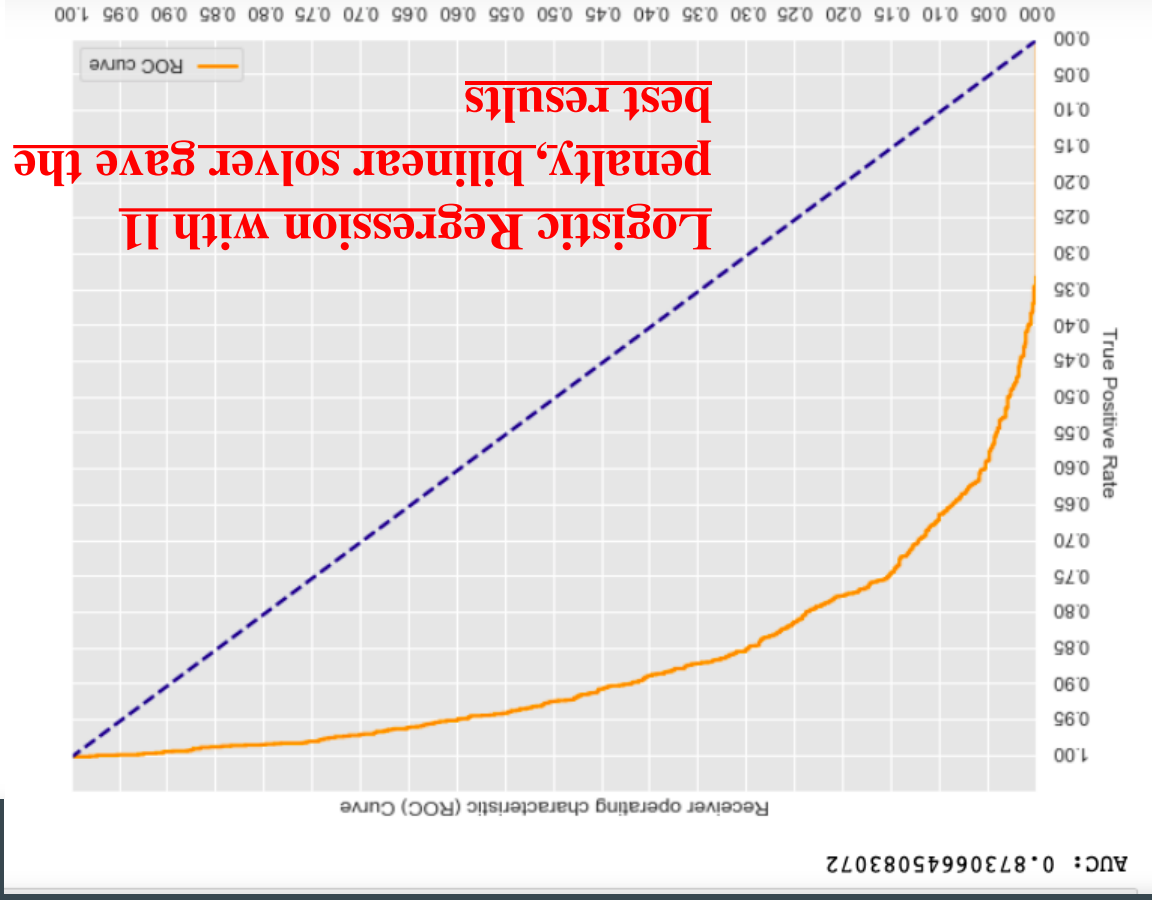
state	failed	successful	country
Aisa	82395.0	108596.0	Aus
	43713.0	9458.0	Euro
	26291.0	9551.0	NAmerica
	31571.0	9234.0	SAmerica
	142798.0	66359.0	

Models 1

Ran Three Models:

- Baseline model of dummyclassifier used which gave 63% accuracy
- Random Forest with hyperparameter tuning using iterations and AUC vs Parameter range plots
- Logistic Regression with hyperparameter tuning in solver type, C parameter and penalty
- XGBoost with and without Gridsearch.
- All models were validated first with training data and then tested with the testing data.

Models 2 - Final Model: Optimized Logistic



Models 3 - Comparisons

Model	Confusion Matrix TN, FP, FN, TN	Accuracy	Recall	Precision	F1
Random Forest Tuned	[[681 265] [314 1241]]	77.2%	85.9%	79.5%	82.6%
LogReg	[[712 234] [277 1278]]	79.6%	82.2%	84.5%	83.3%
LogReg Tuned (Final Model)	[[725 221] [280 1275]]	80.0%	82.0%	85.2%	83.6%
XGBoost with GridSearch	[[593 353] [187 1368]]	78.4%	88.0%	79.5%	83.5%

Models 4 - Improvements

- Quantify quality of the project's presentation through recognising the use of videos, images and rewards.
- Monitor updates from project founders and number of backers/amount of pledges for first 10-20 days and quantify it as a feature.
- Work on better classifying project categories and make them more uniform

Thank you