KICKSTARTER Projects:

Is This Project Going to Reach its Crowdfunding Goal?

Lior Solomon, Eva Marciano, Omer Sarig and Yotam Aharony

TAU

Reminder

Project Problem Formulation

Determine whether a new Kickstarter project is going to raise his crowdfunding goal by Machine Learning Model based on dataset of about 250k past projects

What is KICKSTARTER?

Kickstarter is an American public-benefit corporation, that maintains a global crowdfunding platform focused on creativity and merchandising.

Changed our Dataset

We moved from our "comfortable" Kaggle Dataset (no NANs, no need for conversions) to a more detailed WebRobots Dataset.

Pros:

- **Golden** details such as URL, creators information, photos, blurb, kickstarter info (staff pick, spotlight and more).

Cons:

- Heavy, need to combine more then 200 CSVs
- CSVs are **not uniform**, lots of NANs and different columns
- Hard **Conversion** Working: JSON extraction, time and currency conversions and more
- Refactor our code!!!
- Different distribution – different model results

Our DataSet .

Over 250 projcets (entries)

ID internal kickstarter ID | Creator creator id

Name name of project | Blurb of the project

Category main and sub category | state current condition the project is

Initial, launched and deadline times | URL and Photos

Goal fundraising goal | Pledged amount pledged by "crowd" | Backers number of backers

Locationcountry and city of the project | Currency curency

Our DataSet .

We stayed with 274912 rows and 28 columns:

3]:

	id	photo	name	blurb	goal	pledged	state	slug	country	curren
0	6.644984e+08	{"small":"https://ksr-ugc.imgix.net/assets/011...	DreemBee - Sweet Dreams	An extremely user-friendly child's video playe...	40000	36479	failed	dreembee-sweet-dreams	GB	G
1	1.610438e+09	{"small":"https://ksr-ugc.imgix.net/assets/012...	NET Show (Next Entrepreneur Trade Show)	NET is a trade show exhibiting all that is new...	5000	76	failed	net-show-next-entrepreneur-trade-show	US	U
2	2.127141e+09	{"small":"https://ksr-ugc.imgix.net/assets/012...	Pure.Charger	Protect your mobile while charging in public s...	100000	10330	failed	purecharger	US	U
3	1.353155e+09	{"small":"https://ksr-ugc.imgix.net/assets/012...	LiteStem	LED headlight incorporated into the	1000	0	failed	litestem	US	U

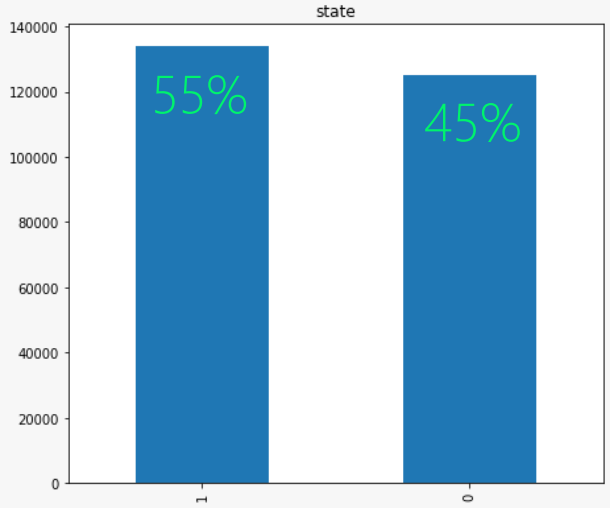
the path

In [4]:
```
raw_df.columns #Let's see which columns there are
```

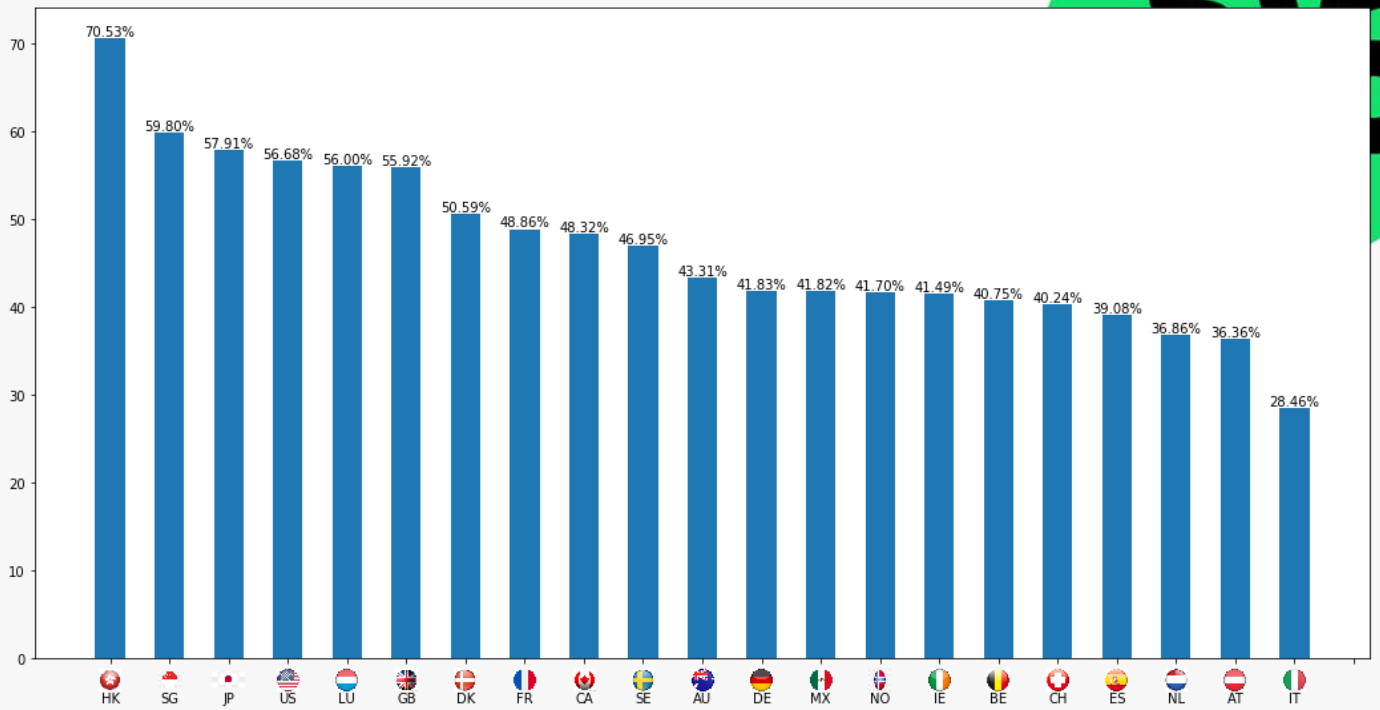
Out[4]:
```
Index(['id', 'photo', 'name', 'blurb', 'goal', 'pledged', 'state', 'slug',
       'country', 'currency', 'currency_symbol', 'currency_trailing_code',
       'deadline', 'state_changed_at', 'created_at', 'launched_at',
       'staff_pick', 'backers_count', 'static_usd_rate', 'usd_pledged',
       'creator', 'location', 'category', 'spotlight', 'urls',
       'current_currency', 'is_starrable', 'usd_type'],
      dtype='object')
```

Convert our data

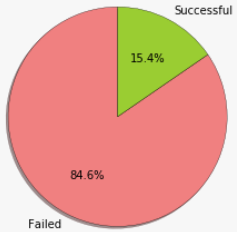
Initial Analysis



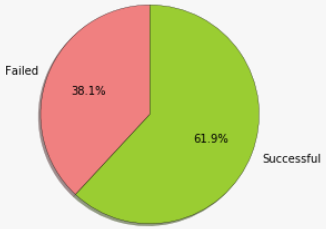
Percentile of successful projects according to country



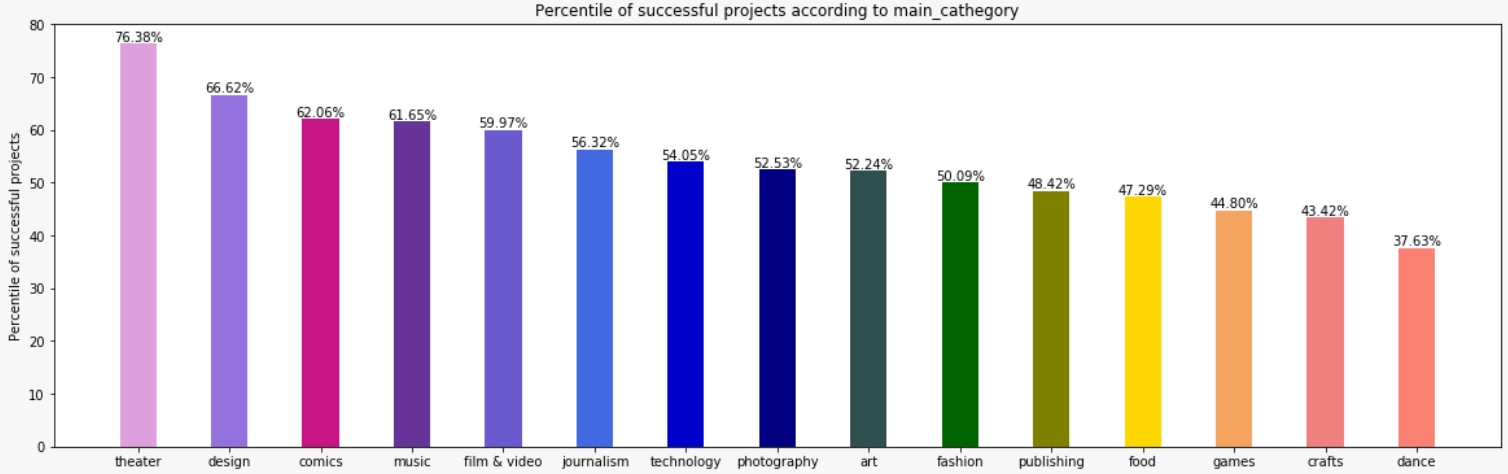
Percentile of success according to price goal between 100,000$-1,000,000$



Percentile of success according to price goal lower than 10,000$



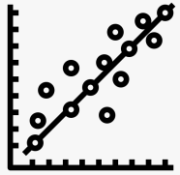
Percentile of successful projects according to main_cathegory



baseline MODEL

AUC: ~~55%~~ → 66%

based on goal, duration, country and main category



LinearSVC

Statistics	
Train Score	55.13%
Test Score	**55.87%**
Precision	55.78%
Recall	99.56%
fScore	71.50%

RandomForest

Statistics	
Train Score	73.95%
Test Score	**66.88%**
Precision	68.33%
Recall	72.09%
fScore	70.16%

KICKSTARTER

feature Extraction

Key words

Is the most "Successful" words in the name? in the blurb? the most "Failed" words?

GDP

Country's GDP

Happiness

Country's happiness state at the launched year

#Project past 7days

How many projects were launched at the same week?

#Project 7days category

How many projects were launched at the same week at the same category?

Time since last project in the category

In hours

Log(Goal)

Histogram looks better

Name and Blurb analysis

#Chars, #words, average word length, #words, #stopwords, #special chars

#Creator's Projects

Is this creator already launched a kickstarter project?

DateTime

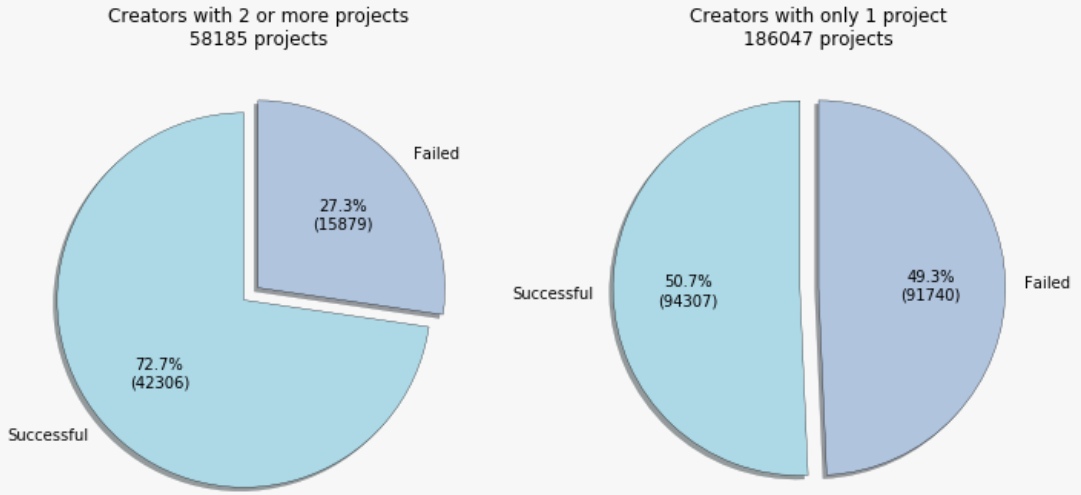
Week day, Month, Year

KICKSTARTER

feature Demonstration

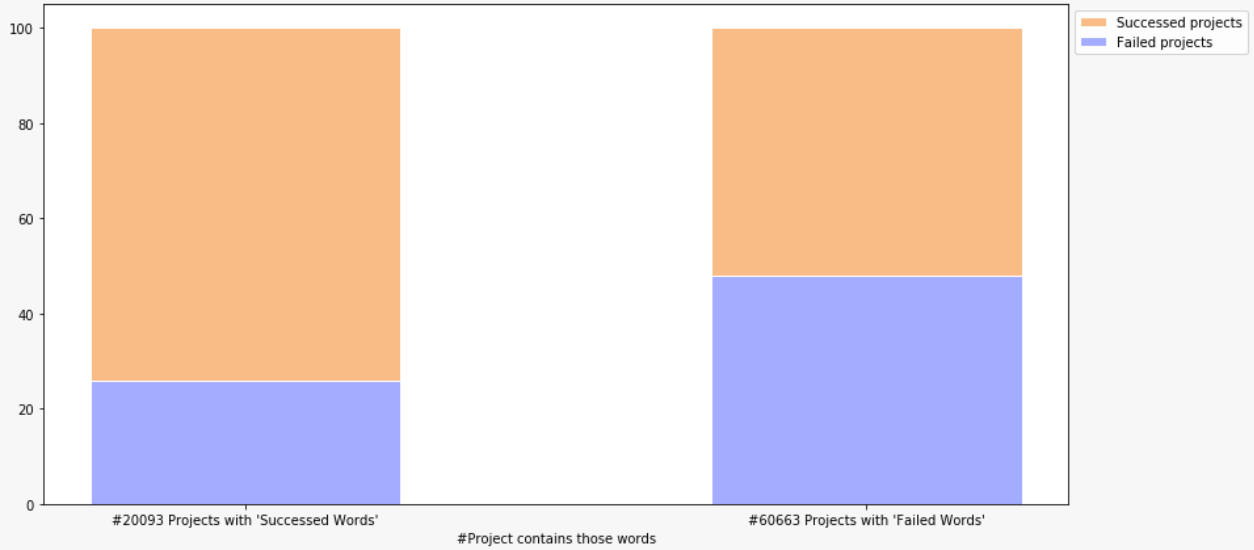
#Creator's Projects

Does Experience Matter?



Key words

Does successful / failed words influence?



KICKSTARTER

MODEL2.0

AUC: ~~55%~~ → ~~66%~~ → **77%**
#Better Performance

RandomForest
Many features

Improve of ~**11%** in Test Score (77% acc)

Statistics		
Train Score	86.43%	↑ 12%
Test Score	77.39%	↑ 11%
Precision	81.08%	↑ 13%
Recall	76.82%	↑ 4%
fScore	78.89%	↑ 8%

Feature ranking:

1) Sub_category (0.11)
2) Log_usd_goal (0.099)
3) Count_7_days (0.06)
4) Creator_Num_Projects (0.049)

...

7) Category_Count_7_days (0.04)

KICKSTARTER

Point of Focus

NLP and Statistics
Over ‘Name’ and ‘Blurb’

Rewards
Extract & Analyzing

Photo
Download & Quality Grade

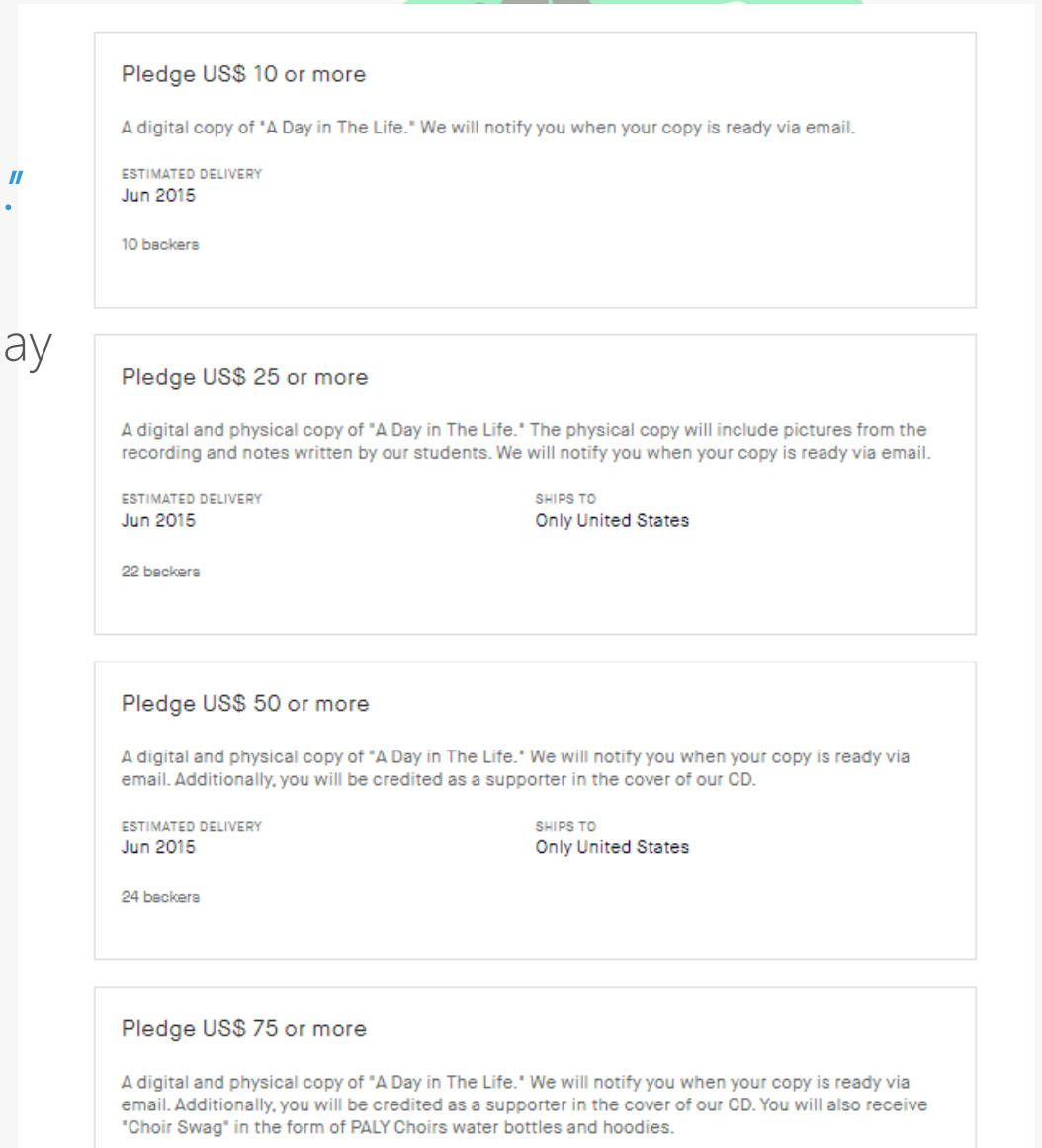
KICKSTARTER

Rewards

"Like it or not, almost no one does something for nothing."

Kickstarter lets the creator to create a series of rewards to say "thank you" to their backers.

It may give us a feature that will greatly affect success or failure of the project.



Pledge US$ 10 or more

A digital copy of "A Day in The Life." We will notify you when your copy is ready via email.

ESTIMATED DELIVERY
Jun 2015

10 backers

Pledge US$ 25 or more

A digital and physical copy of "A Day in The Life." The physical copy will include pictures from the recording and notes written by our students. We will notify you when your copy is ready via email.

ESTIMATED DELIVERY
Jun 2015

SHIPS TO
Only United States

22 backers

Pledge US$ 50 or more

A digital and physical copy of "A Day in The Life." We will notify you when your copy is ready via email. Additionally, you will be credited as a supporter in the cover of our CD.

ESTIMATED DELIVERY
Jun 2015

SHIPS TO
Only United States

24 backers

Pledge US$ 75 or more

A digital and physical copy of "A Day in The Life." We will notify you when your copy is ready via email. Additionally, you will be credited as a supporter in the cover of our CD. You will also receive "Choir Swag" in the form of PALY Choirs water bottles and hoodies.

Rewards Features

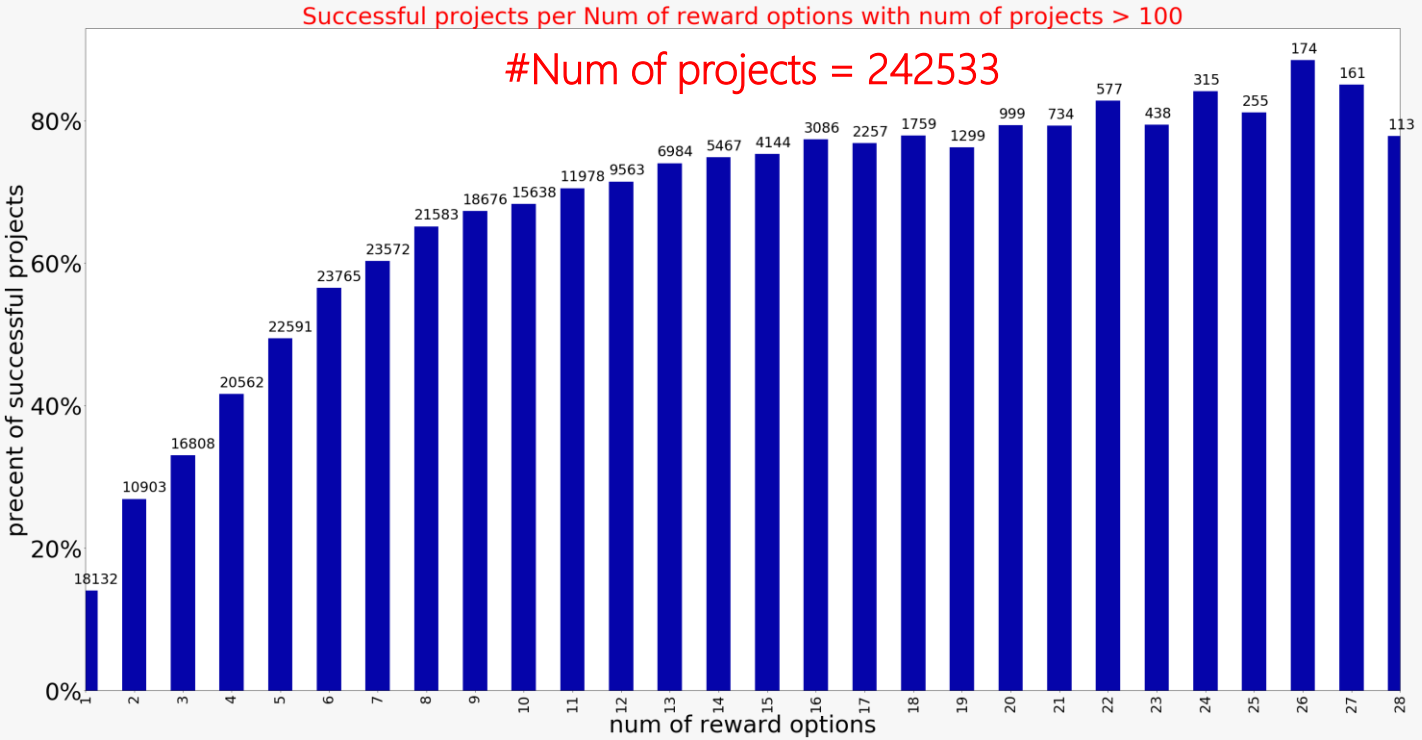
Built a Crawler

MinPledge/Goal

MaxPledge/Goal

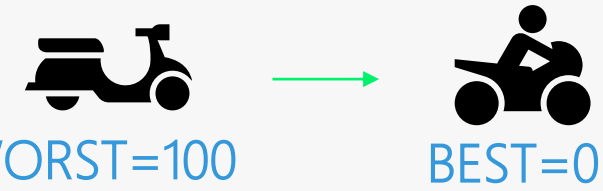
AvgPledge/Goal

Num of Rewards



Photos

- We downloaded each photo (res=1536x864) and scored
- We used BRISQUE (Blind/Referenceless Image Spatial Quality Evaluator) algorithm



High Quality

State of Projects with photo_score less than 20

Low Quality

State of Projects with photo_score more than 80

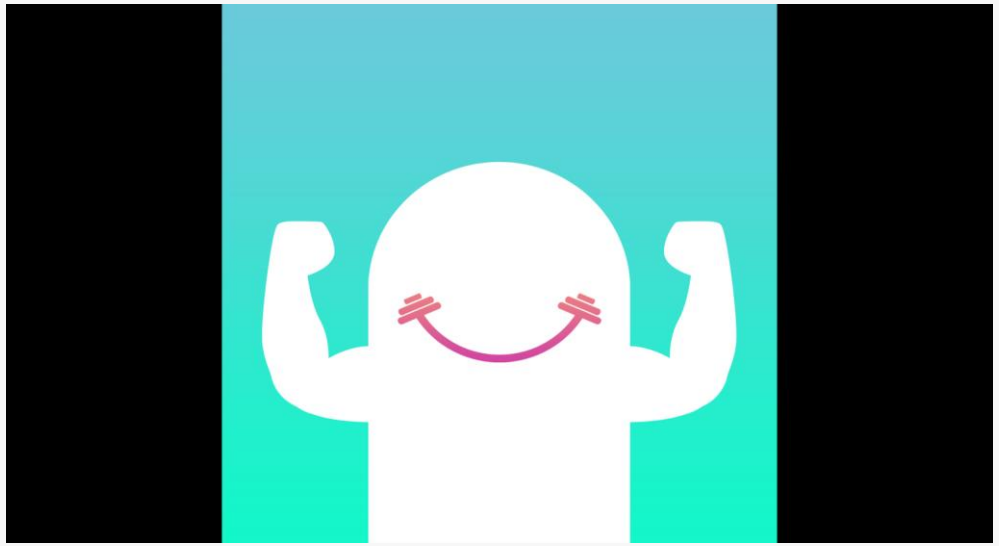
ExamplePhotoScore

“Good” Photo:

11 score

“Bad” Photo:

97 score



KICKSTARTER

NLP & Statistics

Proccessing

Lowercase, stopwords, stemming

Statistics

#Chars, #words, average word length, #stopwords, #special chars

Key words

Is the most “Successful” words in the name? in the blurb? the most “Failed” words?

‘Failed’ words

‘Successful’ words

KICKSTARTER

MODEL3.0

AUC: ~~55%~~ → ~~66%~~ → ~~77%~~ → **80%**

#Better Performance

RandomForest

Many features including rewards and photo_score

Improve of ~**3%** in Test Score (80% acc)

Statistics		
Train Score	89.51%	↑ 3%
Test Score	80.25%	↑ 3%
Precision	80.81%	↓ 0.3%
Recall	84.06%	↑ 8%
fScore	82.40%	↑ 3.5%

Feature ranking:

1) #num_rewards (0.100)
2) Avg_reward_price (0.09)
3) Log_usd_goal (0.07)

...

5) Count_7_days (0.048)

...

7) Creator_Num_Projects (0.04)
8) Photo_Score (0.033)

KICKSTARTER

FINALMODEL

AUC: ... → ~~77%~~ → ~~80%~~ → **84%**
#Best Performance

LightGBM

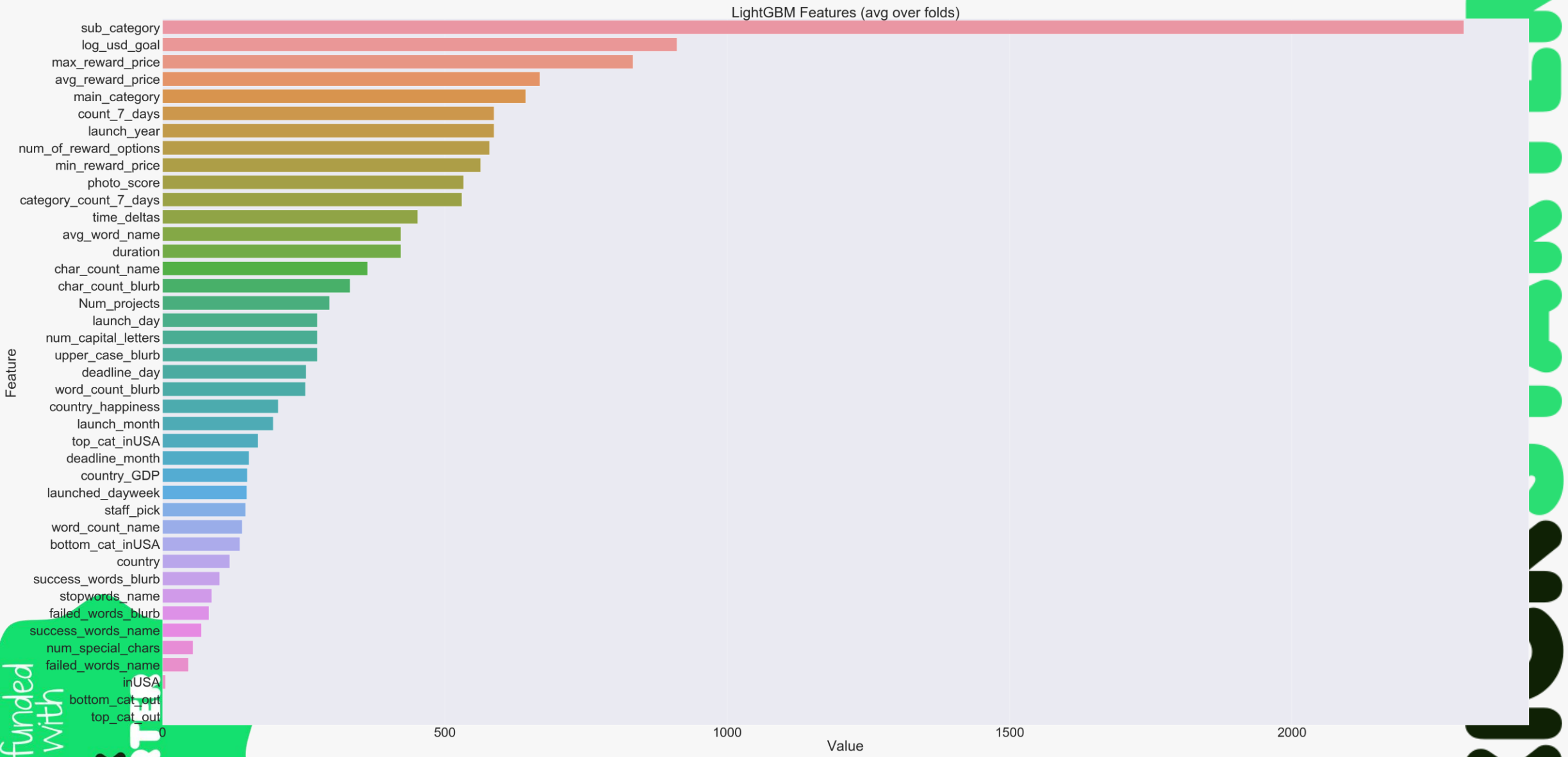
Same features as RandomForest3.0

Improve of ~3.5% in Test Score (84% acc)

Statistics		
Test Score	83.89%	↑ 3.5%
Precision	85.16%	↑ 4.5%
Recall	85.67%	↑ 1.5%
fScore	85.43%	↑ 3%

Confusion		Pred	
		No	Yes
Act	No	8927	1998
	Yes	1918	11468

KICKSTARTER


LightGBM Features (avg over folds)
sub_category
log_usd_goal
max_reward_price
avg_reward_price
main_category
count_7_days
launch_year
num_of_reward_options
min_reward_price
photo_score
category_count_7_days
time_deltas
avg_word_name
duration
char_count_name
char_count_blurb
Num_projects
launch_day
num_capital_letters
upper_case_blurb
deadline_day
word_count_blurb
country_happiness
launch_month
top_cat_inUSA
deadline_month
country_GDP
launched_dayweek
staff_pick
word_count_name
bottom_cat_inUSA
country
success_words_blurb
stopwords_name
failed_words_blurb
success_words_name
num_special_chars
failed_words_name
inUSA
bottom_cat_out
top_cat_out
Feature
0
500
1000
1500
2000
Value

What's Next?

- Fine tuning
- Documentation
- Upload our dataset to Kaggle
- Upload our project to Kaggle

KICKSTARTER