



KICKSTARTER Projects:  
Is This Project Going to Reach its  
Crowdfunding Goal?

Lior Solomon, Eva Marciano, Omer Sarig and Yotam Aharony

TAU

# Who **Are We**?



**Yotam Aharony**  
**EE & CS, 4<sup>th</sup> year**



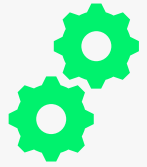
**Lior Solomon**  
**EE & CS, 4<sup>th</sup> year**



**Omer Sarig**  
**EE & CS, 4<sup>th</sup> year**



**Eva Marciano**  
**CS, 3<sup>rd</sup> year**



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 over 250k past projects

# 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 than 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!!!



# Our DataSet .



Over 250 projects (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



Location country and city of the project | Currency currency



# Our DataSet .

KICK  
STARTER  
STAFF  
PICK

We stayed with 274912 rows and 28 columns:

3]:

	id	photo	name	blurb	goal	pledged	state	slug	country	current
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:

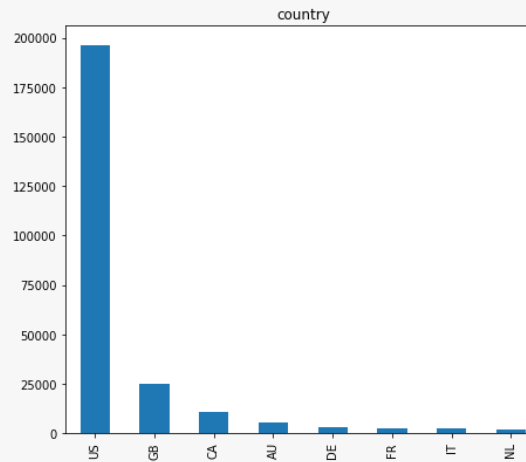
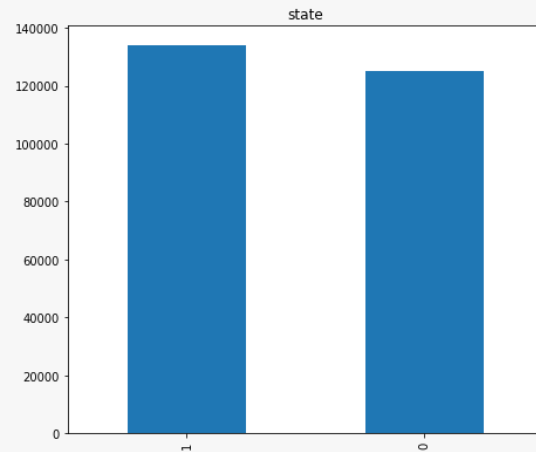
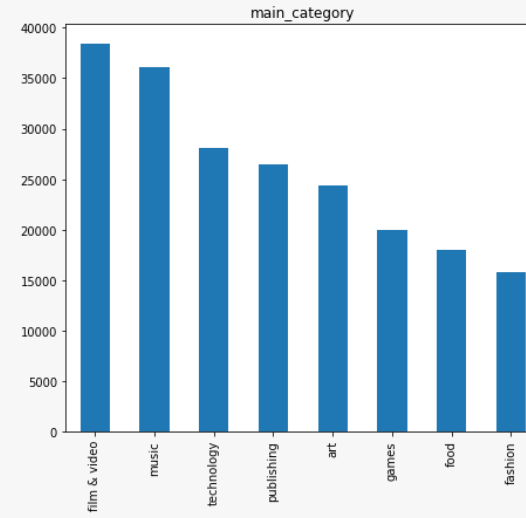
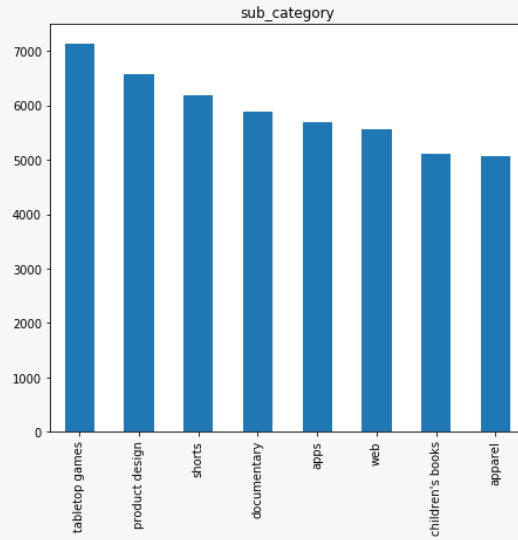
```
In [4]: raw_df.columns #Let's see which columns there are

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')
```

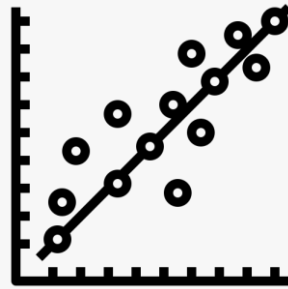
Convert our data

KICK  
STARTER

# Initial Analysis



# baseline MODEL



LinearSVC

based on goal, duration, country, main and sub\_category

Statistics	
Train Score	49.6%
Test Score	49.45%
Precision	76.87%
Recall	3.37%
fScore	6.46%

bad Recall and  
fScore!



# MODEL1.0

#Better Performance



RandomForest

based on goal, duration, country, main and sub\_category

Statistics	
Train Score	85.27%
Test Score	68.42%
Precision	84.05%
Recall	65.86%
fScore	73.17%

[much better model!](#)

# KICKSTARTER

# feature Extraction



## Key words

Is the most "Successful" words in the name? 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 analysis

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



## #Creator's Projects

Is this creator already launched a kickstarter project?

# MODEL2.0



RandomForest  
Many features

#Better Performance

Improve of ~7% in Test Score (75% acc)

Statistics		
Train Score	84.91%	↓ 0.3%
Test Score	74.94%	↑ 6%
Precision	73.03%	↓ 11%
Recall	79.49%	↑ 14%
fScore	74.94%	↑ 2%

## Feature ranking:

- 1) Sub\_category (0.191)
- 2) Log\_USD\_Goal (0.130)
- 3) Count\_7days (0.09)
- ...
- 8) Creator\_Num\_Projects (0.05)

KICKSTARTER

# What's NEXT?

With our detailed Dataset we have an access to the URL, blurb and Photos



## Enhance our Model

RandomForest was better than  
Adaboost, Linear Regression  
and more. Keep trying...



## Blurb

NLP and statistics over the blurb  
info



## Photos

Try analyzing the project's  
photos – Black/White, color,  
attraction and more  
May be difficult



# Rewards



When you pledge a project you can decide between **few options**.

- We built a **crawler** that extract the information from URL (amount to pledge + reward)
- Heavy and slow, still extracting it.
- TODO:
  - Min/Goal, Max/Goal
  - Avg
  - Rewards ↔ Amount pledged

\* (10\$=Digital copy, 25\$=Digital+Physical)

## 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 Swing" in the form of PAJIF Crooks water bottles and hoodies.

May bring some interesting conclusions!



# Epilogue

TOP 20 are equal!!!



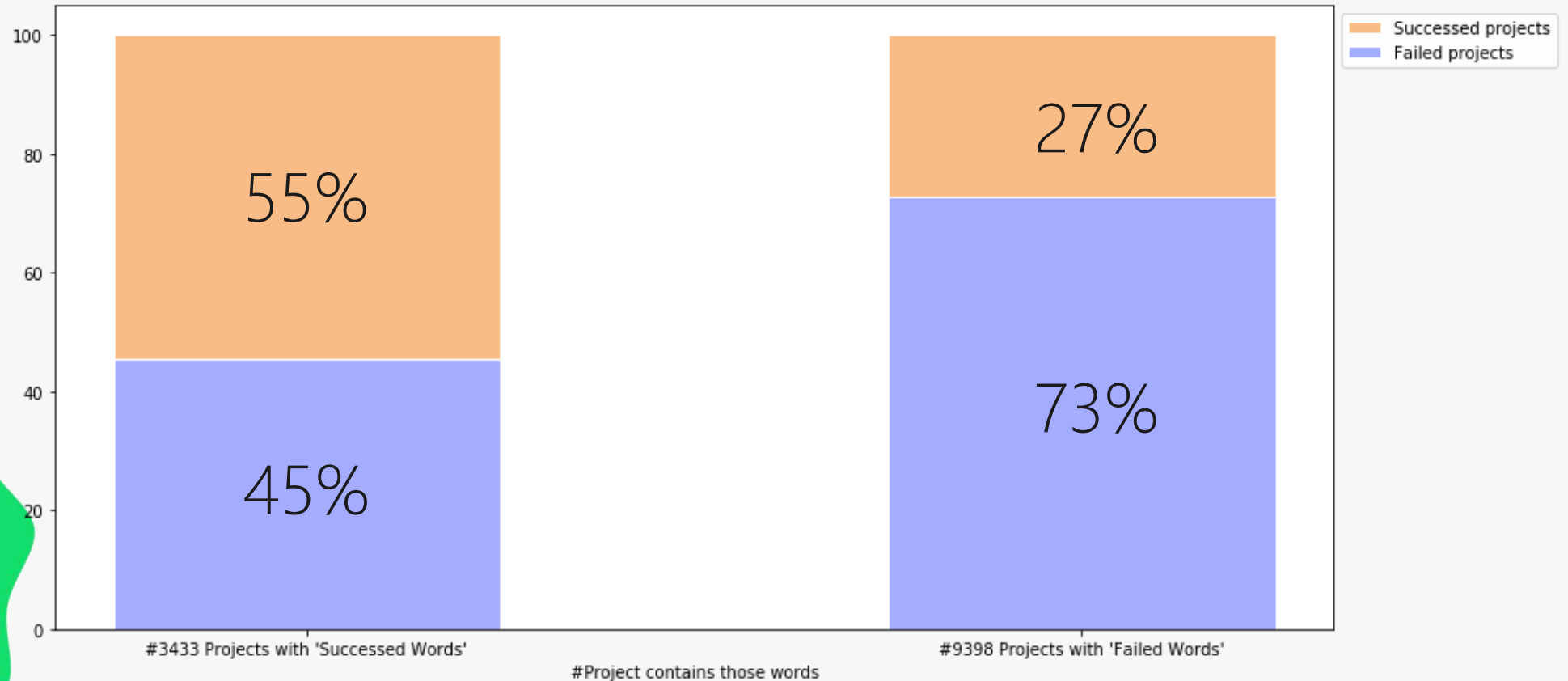
# KICKSTARTER



# Any CLUES?

Top50 but not Top100: ['edition', 'dance', 'release', 'presents', 'recording', 'full']

Top70 but not Top100: ['app', 'clothing', 'mobile', 'fashion', 'dream', 'design', 'apparel', 'journey', 'photography', 'system', 'way', 'online', 'social']

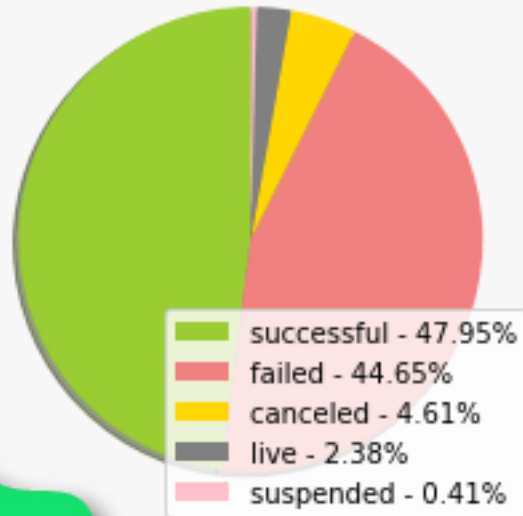


KICKSTARTER  
funded with

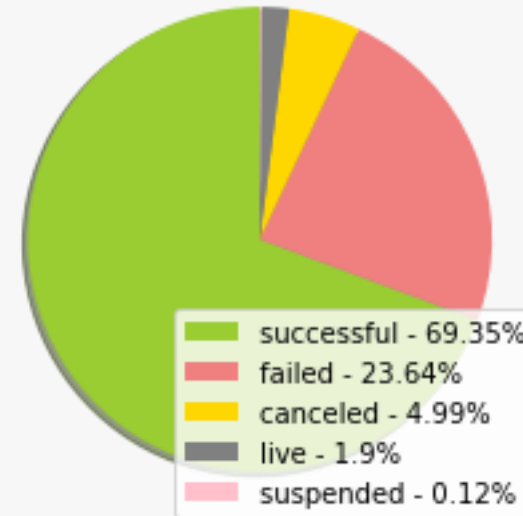


# Creators Info

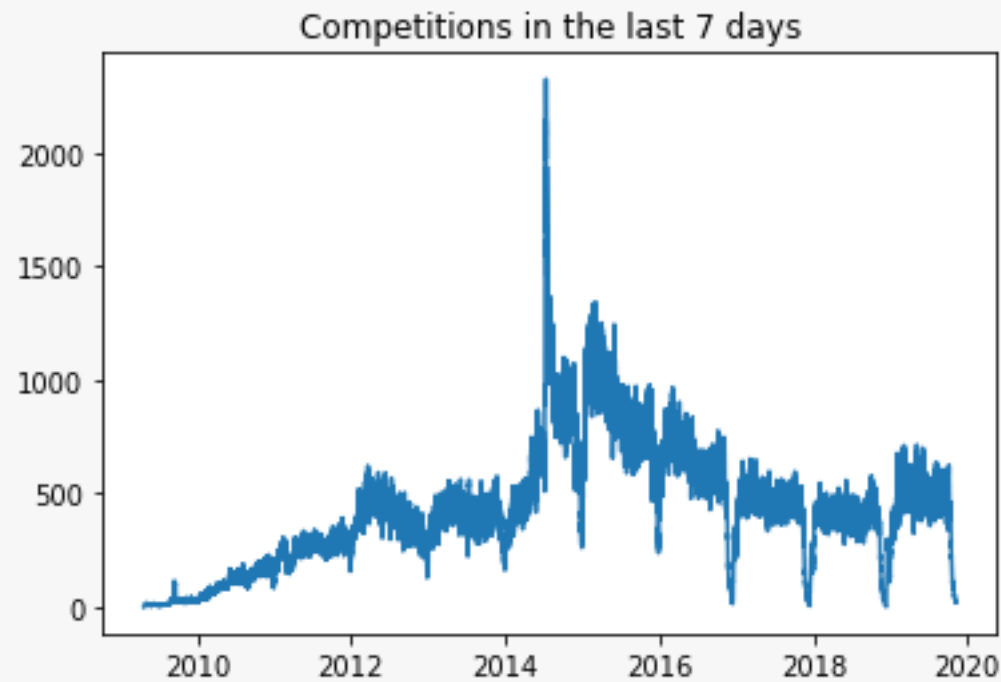
State of projects - Creators with exactly one project



State of projects - Creators with two or more projects



How many projects in  
the last 7 days?



KICKSTARTER

# References

- Main Kickstarter projects DATASET
  - <https://www.kaggle.com/kemical/kickstarter-projects>
- Country / Gross Domestic Product DATASET
  - <https://www.kaggle.com/stieranka/predicting-gdp-world-countries/data>
- World Happiness Report DATASET
  - <https://www.kaggle.com/unsdsn/world-happiness#2019.csv>
- Previous Works:
  - <https://www.kaggle.com/ratatman/data-cleaning-challenge-scale-and-normalize-data>
  - <https://www.kaggle.com/shivamb/an-insightful-story-of-crowdfunding-projects>

