

Kickwrangler

Predict Which Kickstarter Projects Won't Deliver

Alex McCauley

KICKSTARTER

\$1,768,328,118

pledged to Kickstarter projects over 6 years

It doesn't always work out...

FTC Takes First Action Against a Fraudulent Kickstarter Campaign



Derek Meier on May 8, 2014

Washington state attorney general is going after another kickstarter i backed. I will put in a complaint.

Kickwrangler predicts the projects most likely to not deliver.

KICKSTARTER

Data Sources

Project Features

Creator data:

- # past projects (experienced?)
- # they have backed

Backed 2 Created (1) Comments (11)

Project data:

- Category
- # backers
- \$ pledged vs. goal (overfunded?)

\$456,114

pledged of \$200,000 goal

Scraped from:



predict

Project Outcomes

No data/tracking of project delivery

- 500k tech comments
- Sentiment analysis → project status

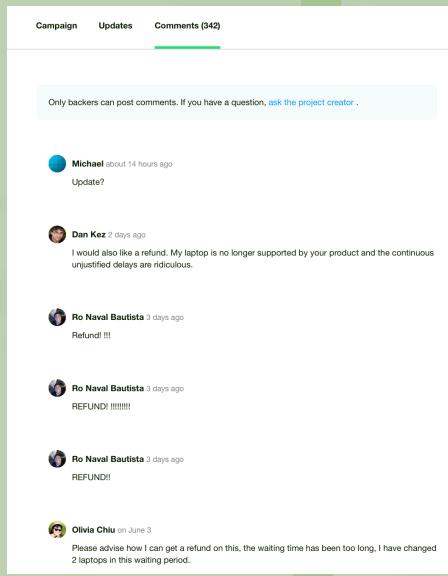
A screenshot of a Kickstarter campaign page. At the top, there are three tabs: "Campaign", "Updates", and "Comments (342)", with "Comments" being the active tab. Below the tabs, a message says "Only backers can post comments. If you have a question, [ask the project creator](#)." A large green arrow points from the "predict" text in the previous slide to this comments section. The comments themselves are as follows:

- Michael about 14 hours ago
Update?
- Dan Kez 2 days ago
I would also like a refund. My laptop is no longer supported by your product and the continuous unjustified delays are ridiculous.
- Ro Naval Bautista 3 days ago
Refund!!!
- Ro Naval Bautista 3 days ago
REFUND!!!!!!
- Ro Naval Bautista 3 days ago
REFUND!!
- Olivia Chiu on June 3
Please advise how I can get a refund on this, the waiting time has been too long, I have changed 2 laptops in this waiting period.

Sentiment Analysis

All comments

Small (1k) sample



Workers label sentiments



Labeled sentiments

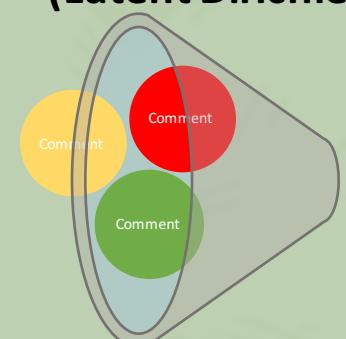
Linear regression:
topic – sentiment mapping:

$$S = \vec{M}^T \vec{T}$$

Sentiment
vs. time!

Pre-process text

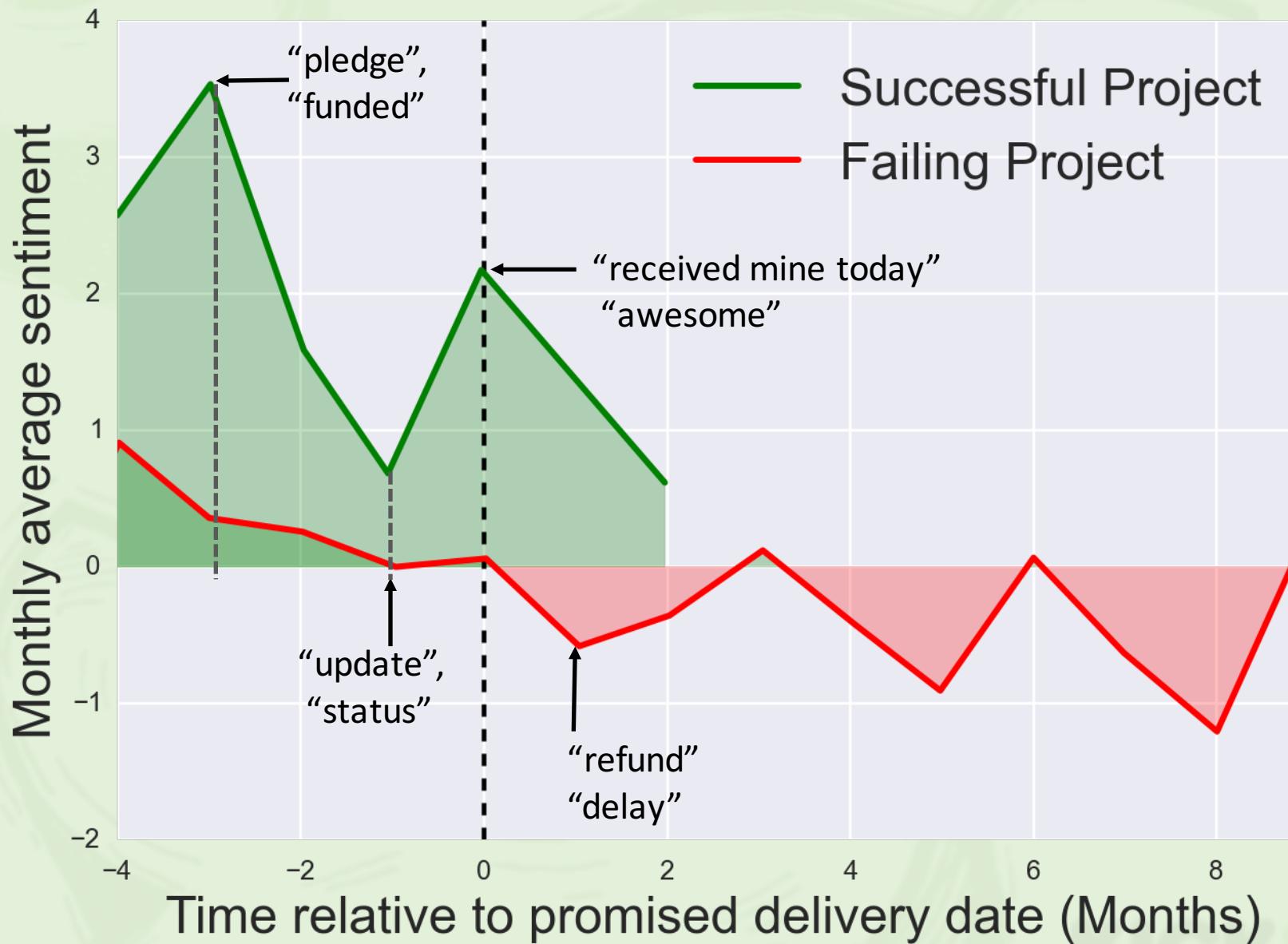
Topic modeling
(Latent Dirichlet Allocation)



Topics

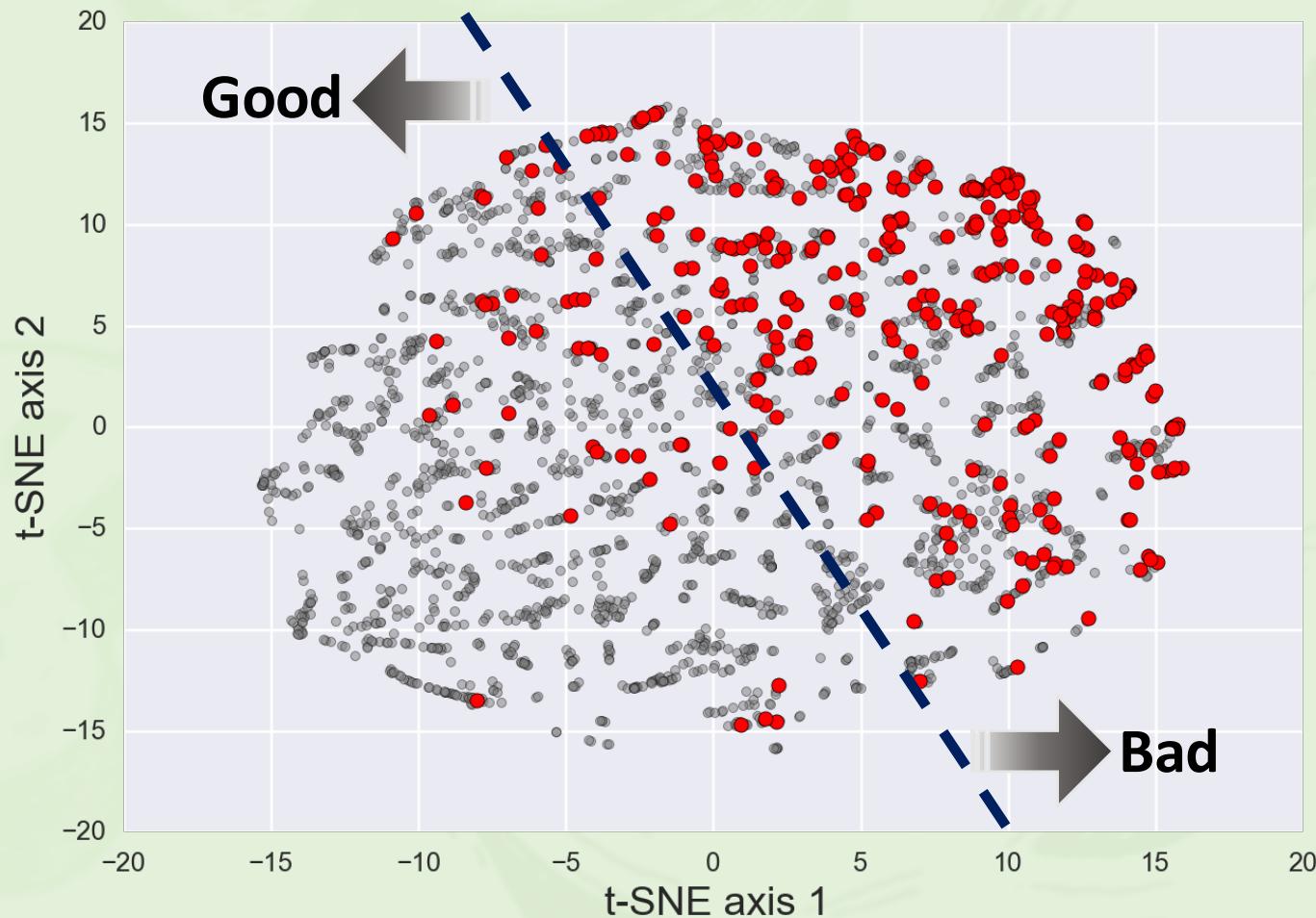
Curate: keep only 7
most relevant topics

Timeline of Project Sentiments



Classification Algorithm

Build intuition: 2-D reduction



Classification results

- Random Forest classifier

Accuracy	0.88
Precision	0.6
Recall	0.3
F1	0.43 (0.24 = random)

Most important feature:

- \$ pledged / \$ goal – overfunded campaigns are more risky

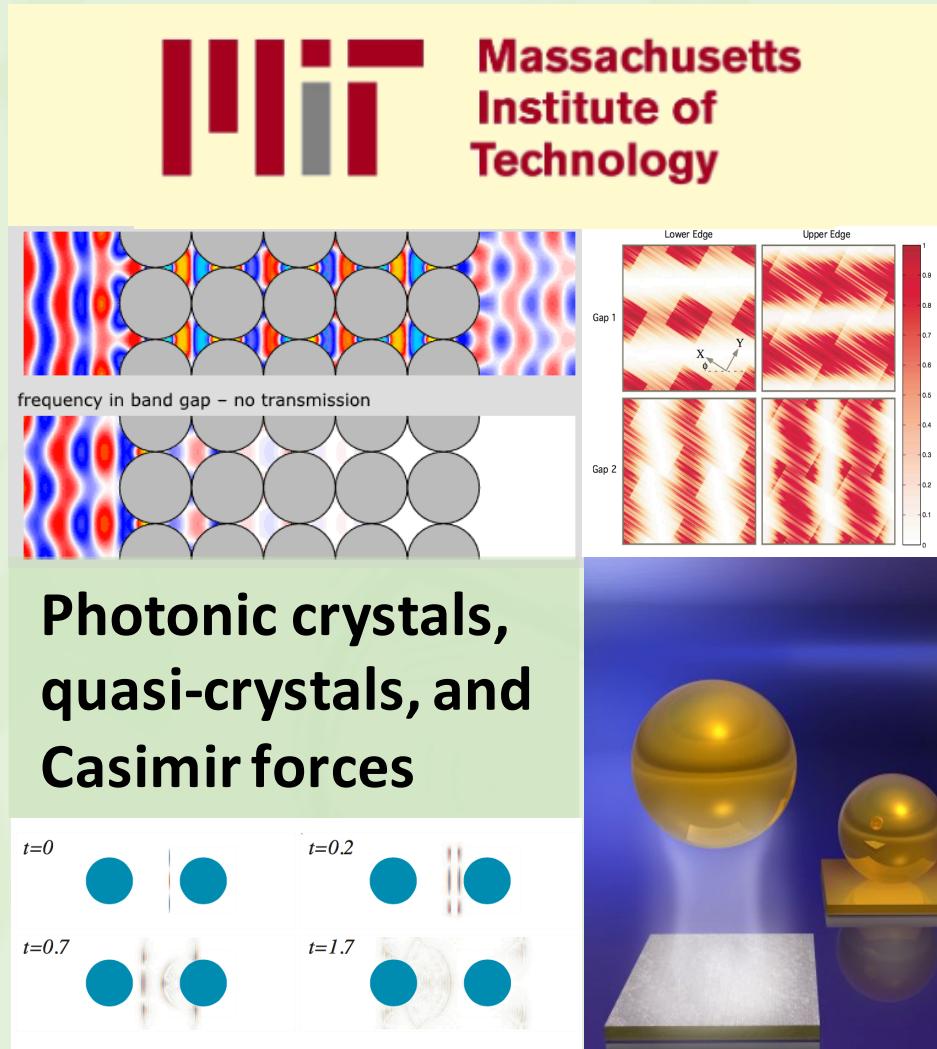
Surprisingly not important:

- Creator history (# previous projects)

Include initial project sentiment:

- Improves predictive power
- Biased (not independent of labels)

About Me



Wireless
power!



Validation

Mechanical Turk ratings

- Remove ratings with high scatter
- Inspect subset

Topic-sentiment map

- Train linear regression on labeled comments
- Test on holdout set: $r^2 = 0.7$

Project status

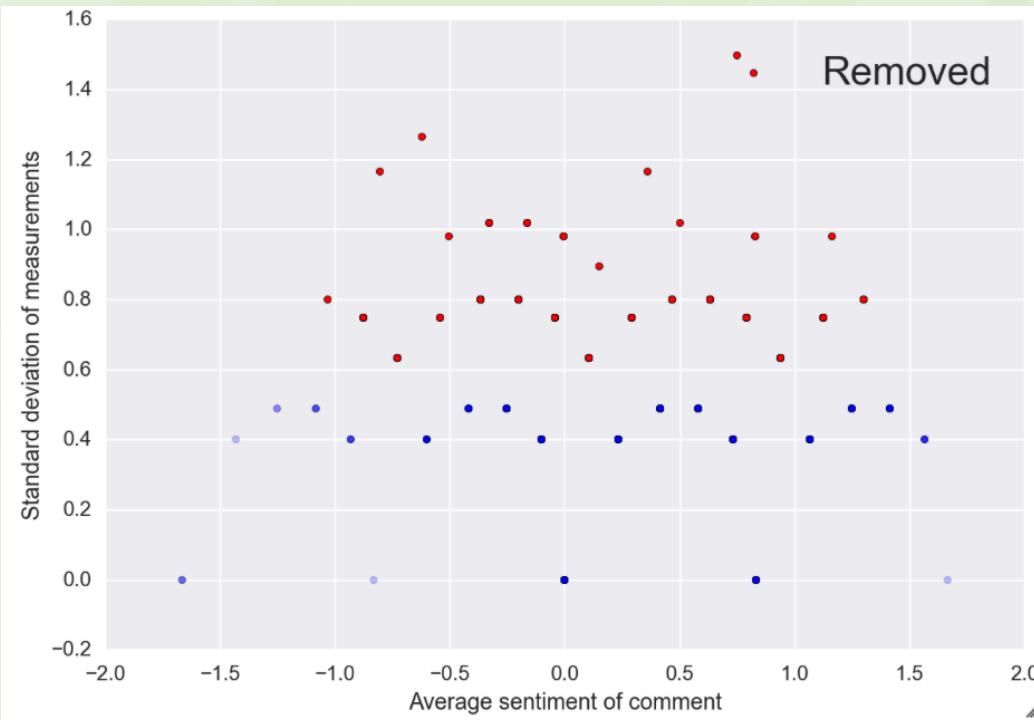
- Recall: 10/10 on “top 10 failed projects” list
- Sample 20 projects randomly:

	N	P
N	12	0
P	3	5

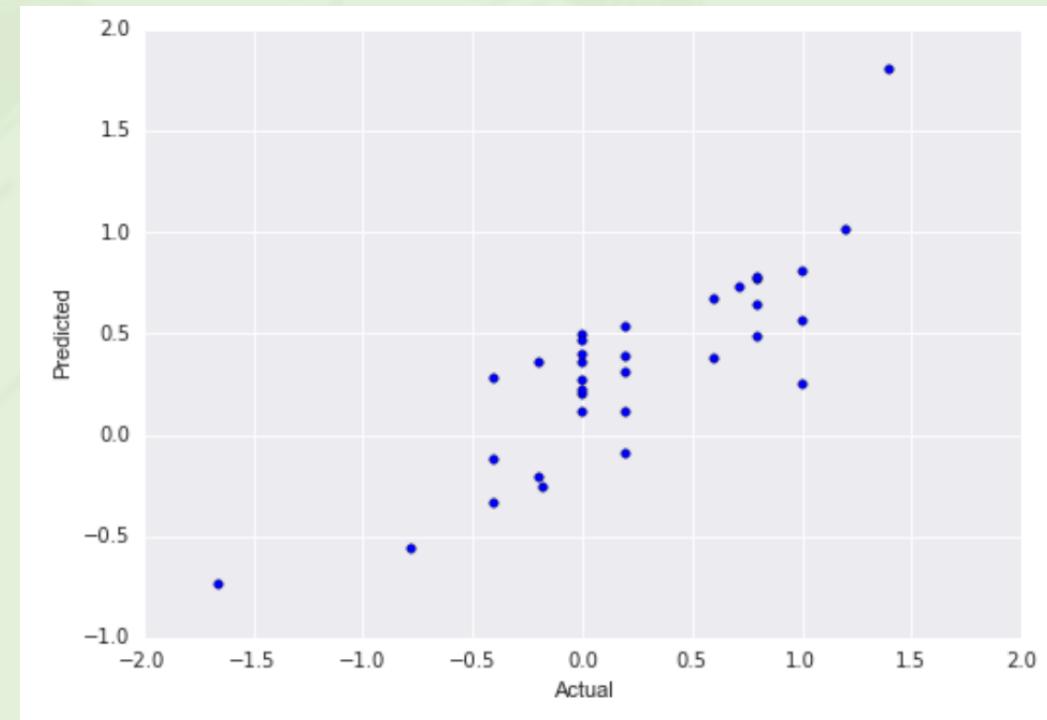
KICKSTARTER

Validation: sentiment analysis

Mechanical Turk ratings



Topic-sentiment map



- Remove comments with standard deviation > 0.5 (S.D. = 1.3 is a random guess)
- Train linear regression on labeled comments, mapping topic content to sentiments
- Test on holdout set: $r^2 = 0.7$

Validation: project labeling

- Test on 10 well-known failure cases, gets 10/10
- Test validity on random sample of 100 projects (TODO)

