Can you kick it?

Predicting what makes a Kickstarter campaign successful

Elizabeth Sia
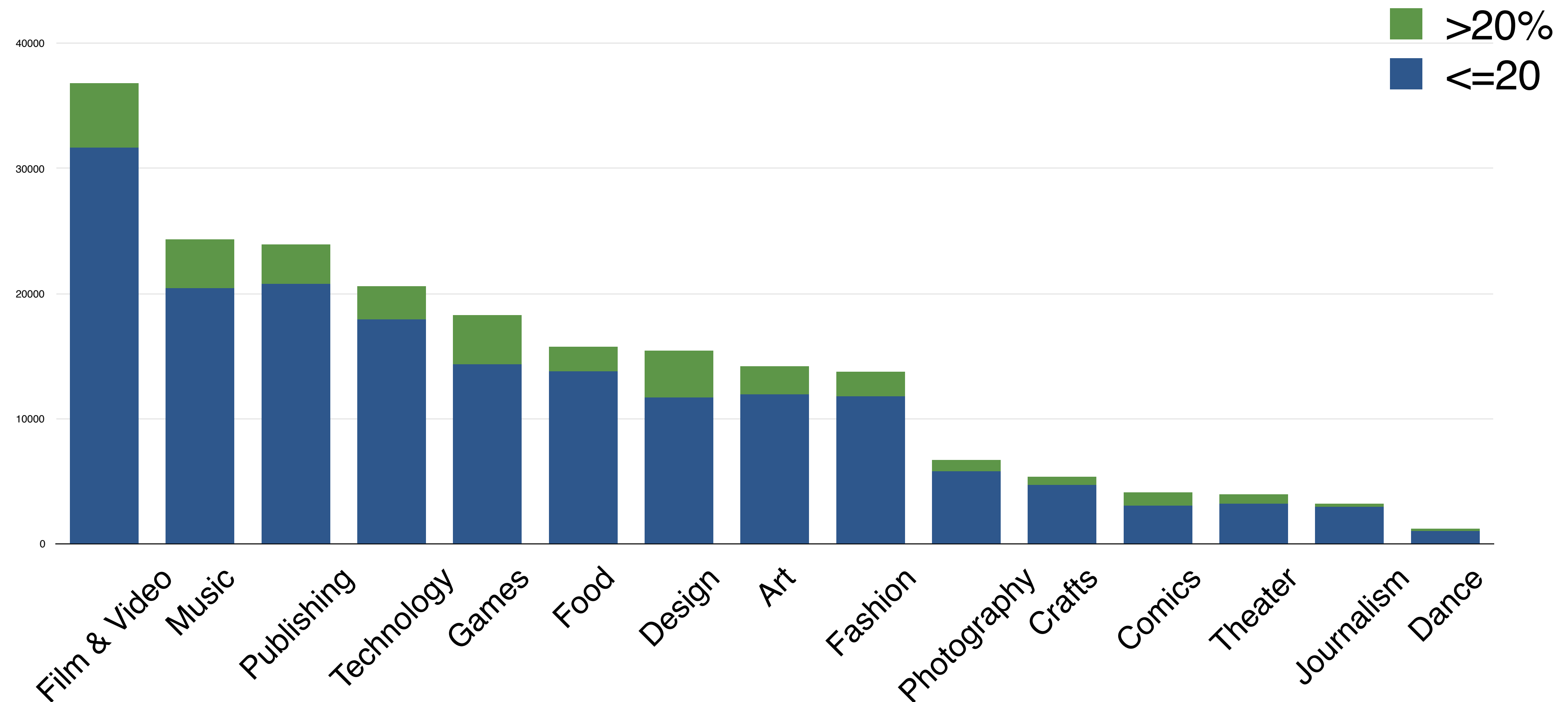
lizsia@gmail.com   **I**   https://www.linkedin.com/in/elizabeth-sia   **I**   https://github.com/lizbug/

KICKSTARTER

Only 35% of Kickstarter Projects are Successful
and when they fail, they fail BAD

80% of unsuccessful projects **ONLY ACHIEVE LESS THAN 20%** of funding goal

Solving this problem began with scraping Kickstarter campaign data

Solving this problem began with scraping Kickstarter campaign data

founder updates/ comments

Image/video content

Project reward structure

Campaign   Updates 15   Comments 71   Community

Share this project   ★

About this project

Suppo

[K]

▶ PLAY

Project title, funding goal, # backers
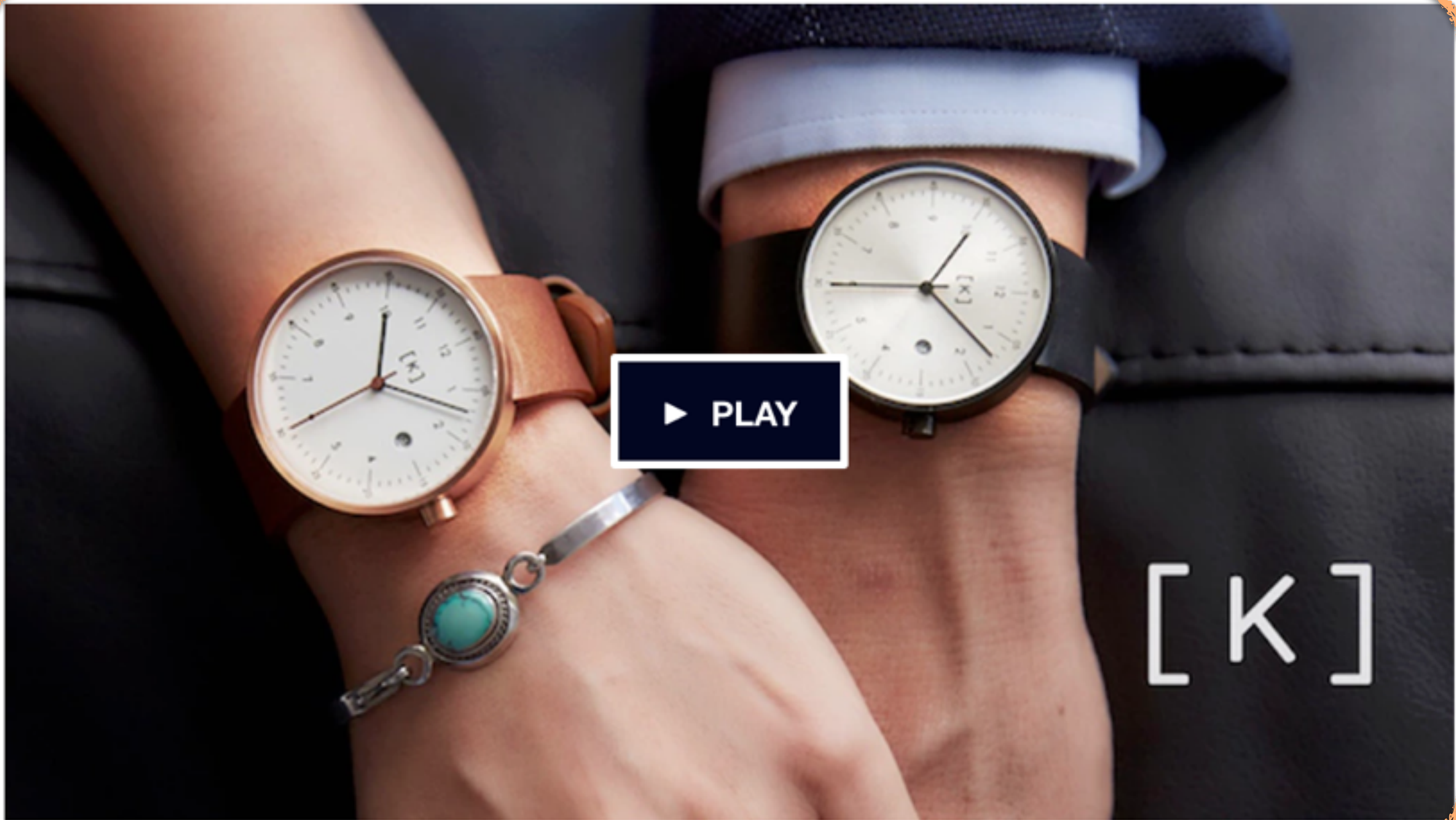
iKi 'A' Series Watch: Minimalist watch with Swiss movement

📍 Melbourne, AU   ✏ Product Design

AU$ 23,622
pledged of AU$ 15,000 goal

173
backers

Pledge AU$ 1 or more

A big thank you from the iKi Studio for your support!

ESTIMATED DELIVERY
Sep 2016

6 backers

Pledge AU$ 99 or more

Approximate 75 US/ 66 EUR
EARLY BIRD SPECIAL - 1 x 'A' Series iKi watch of your choice:

A01 - Rosegold casing with tan brown leather strap; OR
A02 - Gunmetal casing with black leather strap

50% off RRP of A$199

ESTIMATED DELIVERY   SHIPS TO
Sep 2016   Anywhere in the world

Metadata-based classifier model yields 80% precision
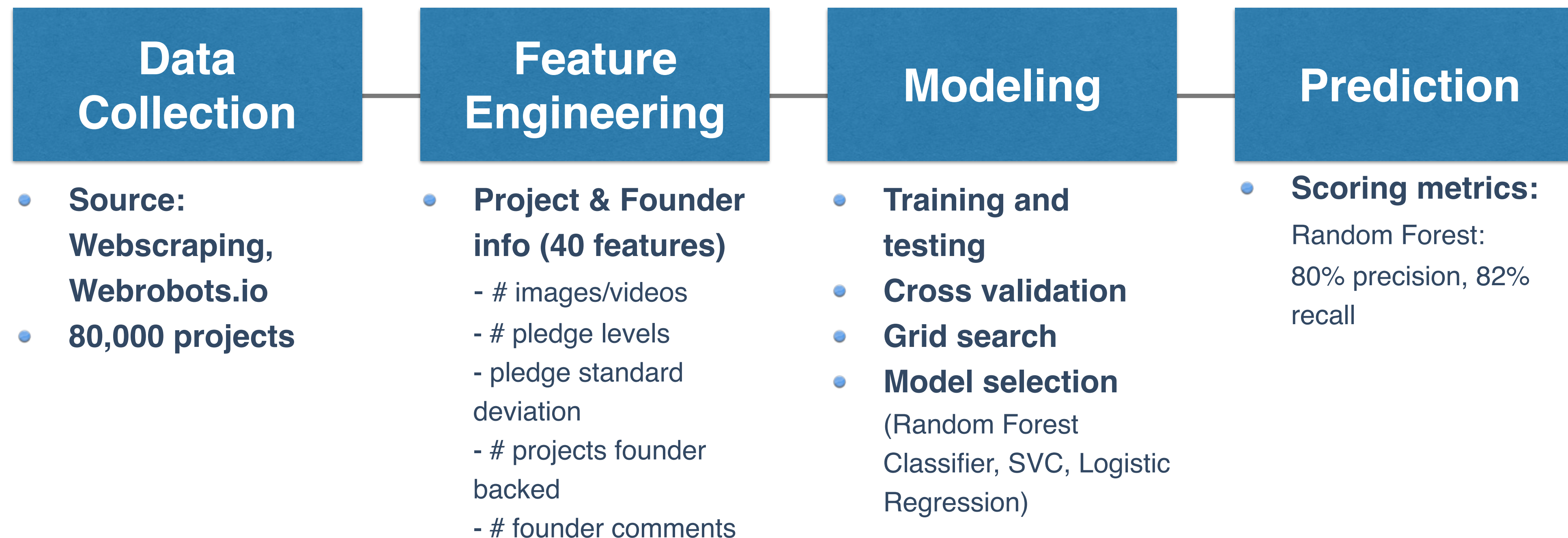
Data Collection

- **Source: Webscraping, Webrobots.io**
- **80,000 projects**

Feature Engineering

- **Project & Founder info (40 features)**
  - # images/videos
  - # pledge levels
  - pledge standard deviation
  - # projects founder backed
  - # founder comments

Modeling

- **Training and testing**
- **Cross validation**
- **Grid search**
- **Model selection**
  (Random Forest Classifier, SVC, Logistic Regression)

Prediction

- **Scoring metrics:**
  Random Forest:
  80% precision, 82% recall

The features that identified a successful project did not tell much about the project's story

Successful
Failed

Breakdown by # of projects a founder backed

- 0: 23%
- 1: 48%
- 2-4: 64%
- 5-15: 77%
- >15: 87%

Breakdown by project funding goal

- <=2k: 53%
- 2k-5k: 50%
- 5k-10k: 46%
- 10k-30k: 42%
- >30k: 24%

Breakdown by average number of backers required

- <=15: 58%
- 16-35: 56%
- 36-100: 45%
- >100: 22%
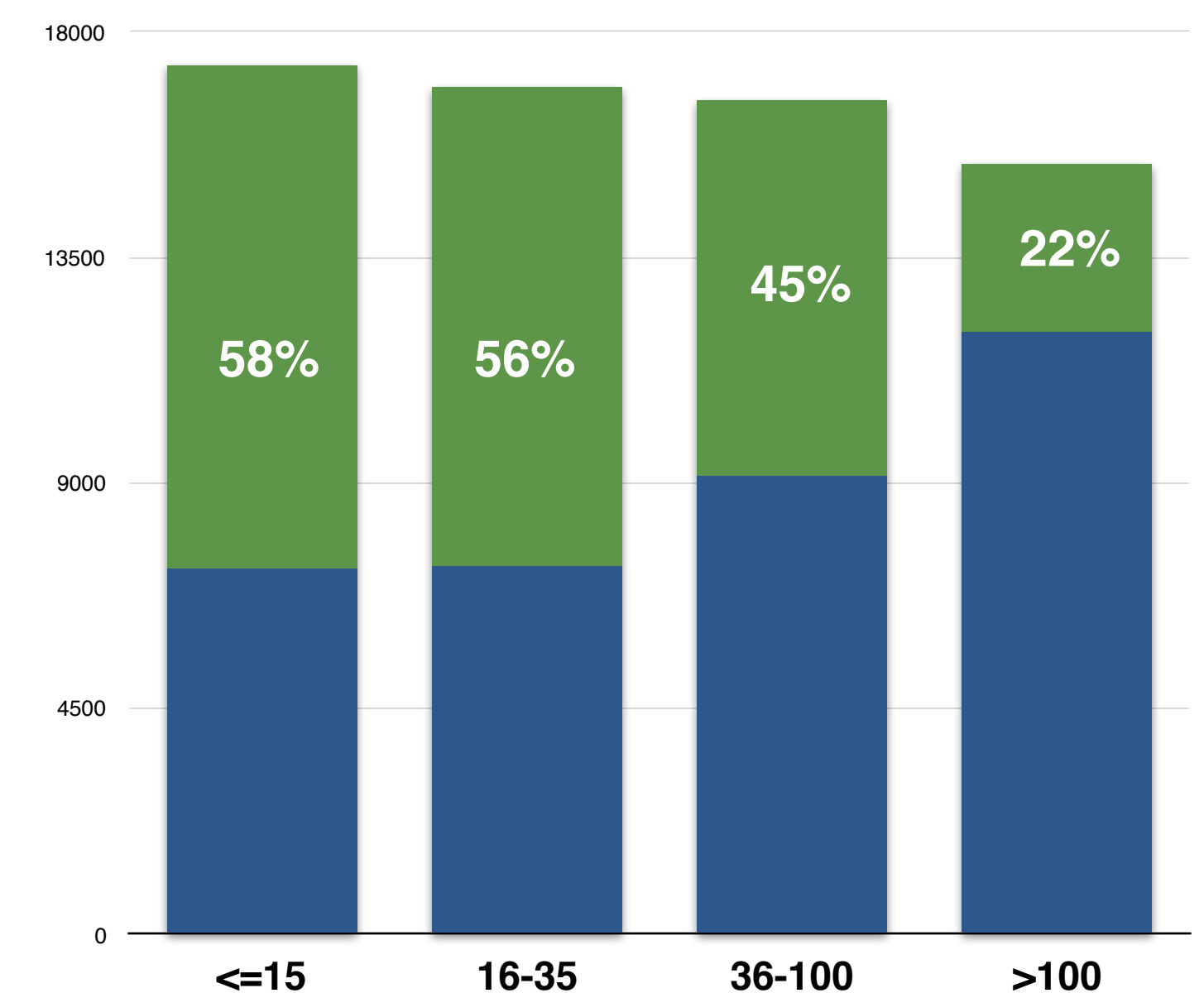
The features that identified a successful project did not tell much about the project's story
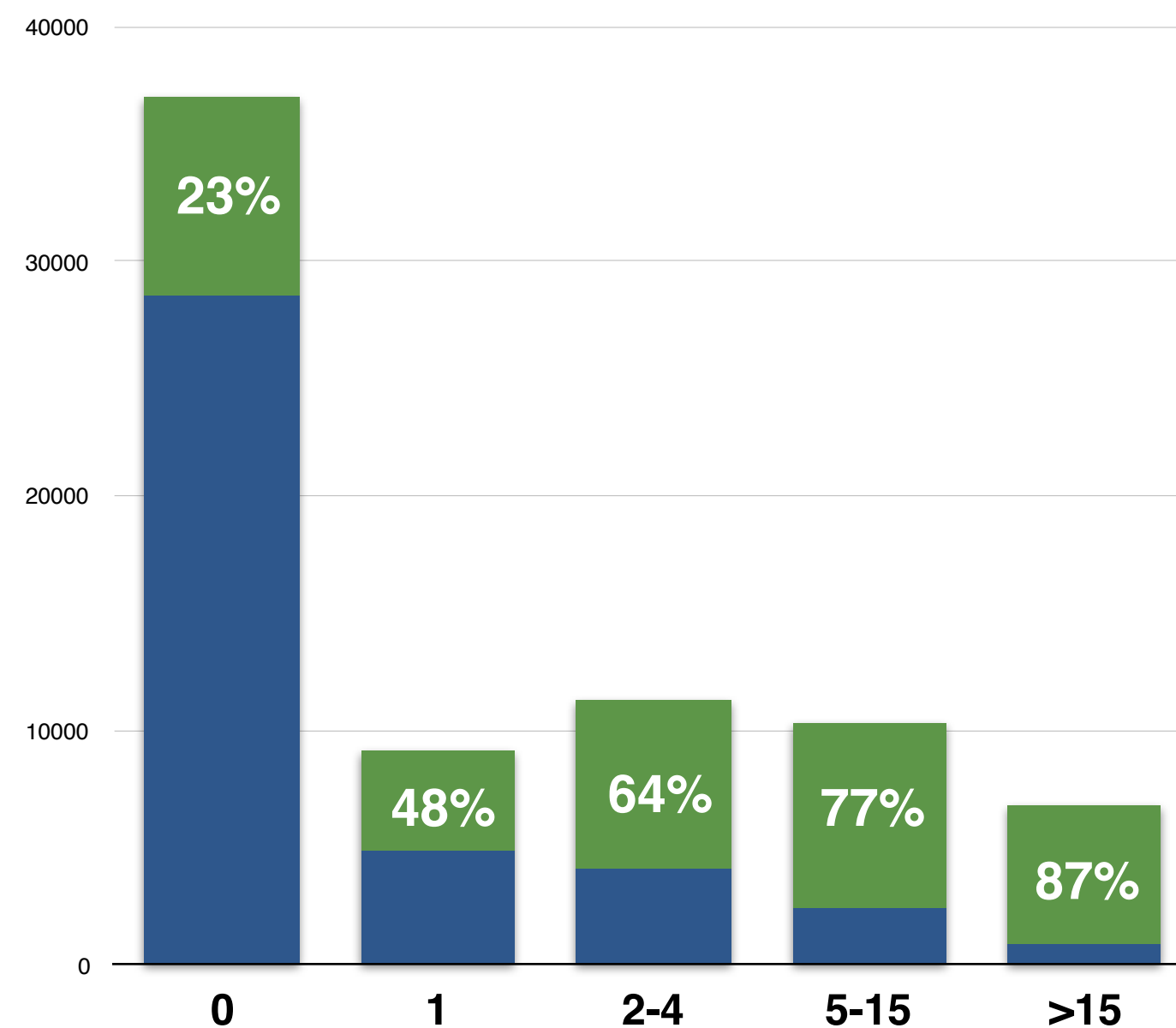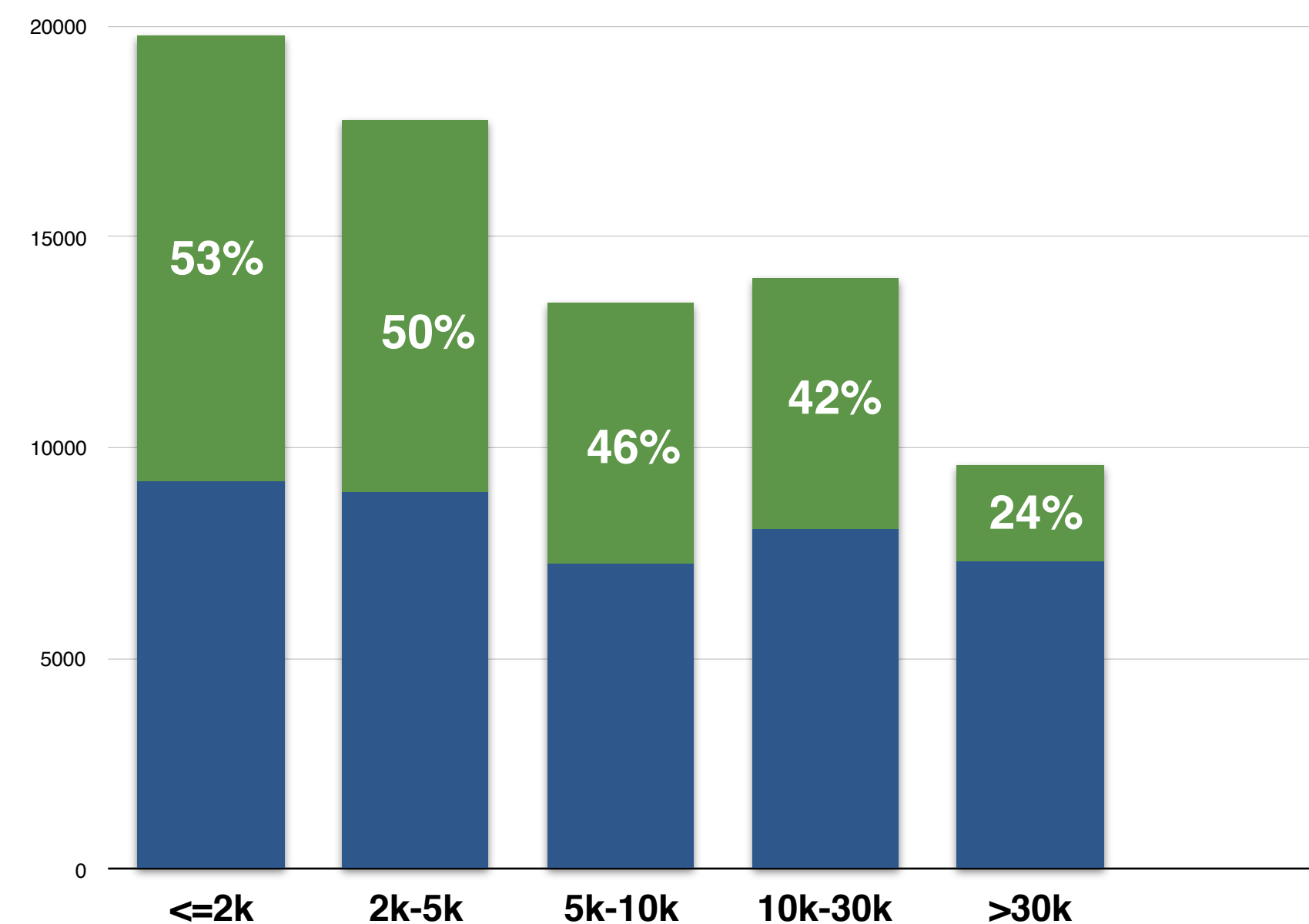

Successful founders backed more projects

Breakdown by # of projects a founder backed

# projects	%
0	23%
1	48%
2-4	64%
5-15	77%
>15	87%

Lower funding goals were likely to be funded

Breakdown by project funding goal

goal	%
<=2k	53%
2k-5k	50%
5k-10k	46%
10k-30k	42%
>30k	24%

Successful projects have reward structures that require less # backers

Breakdown by average number of backers required
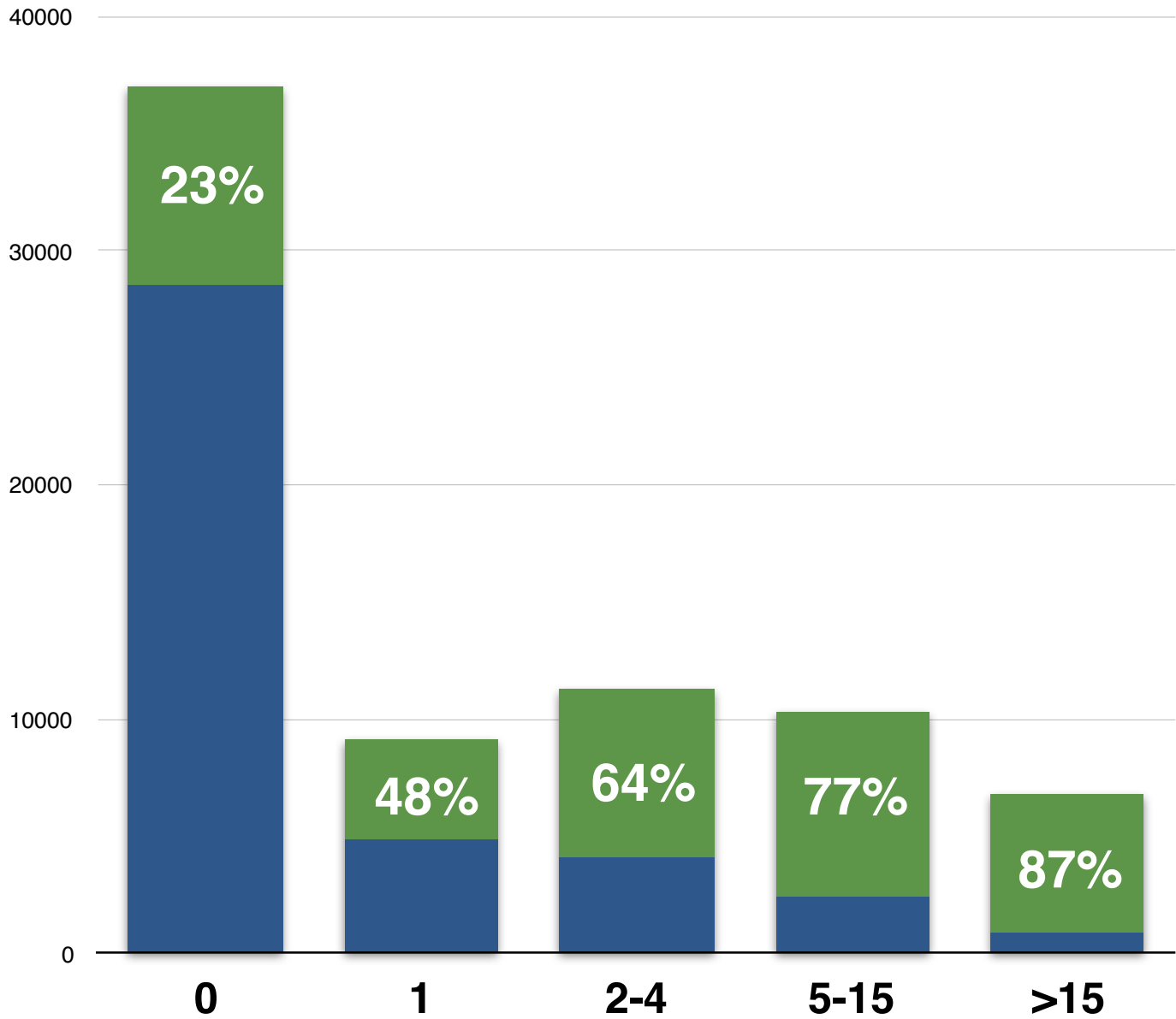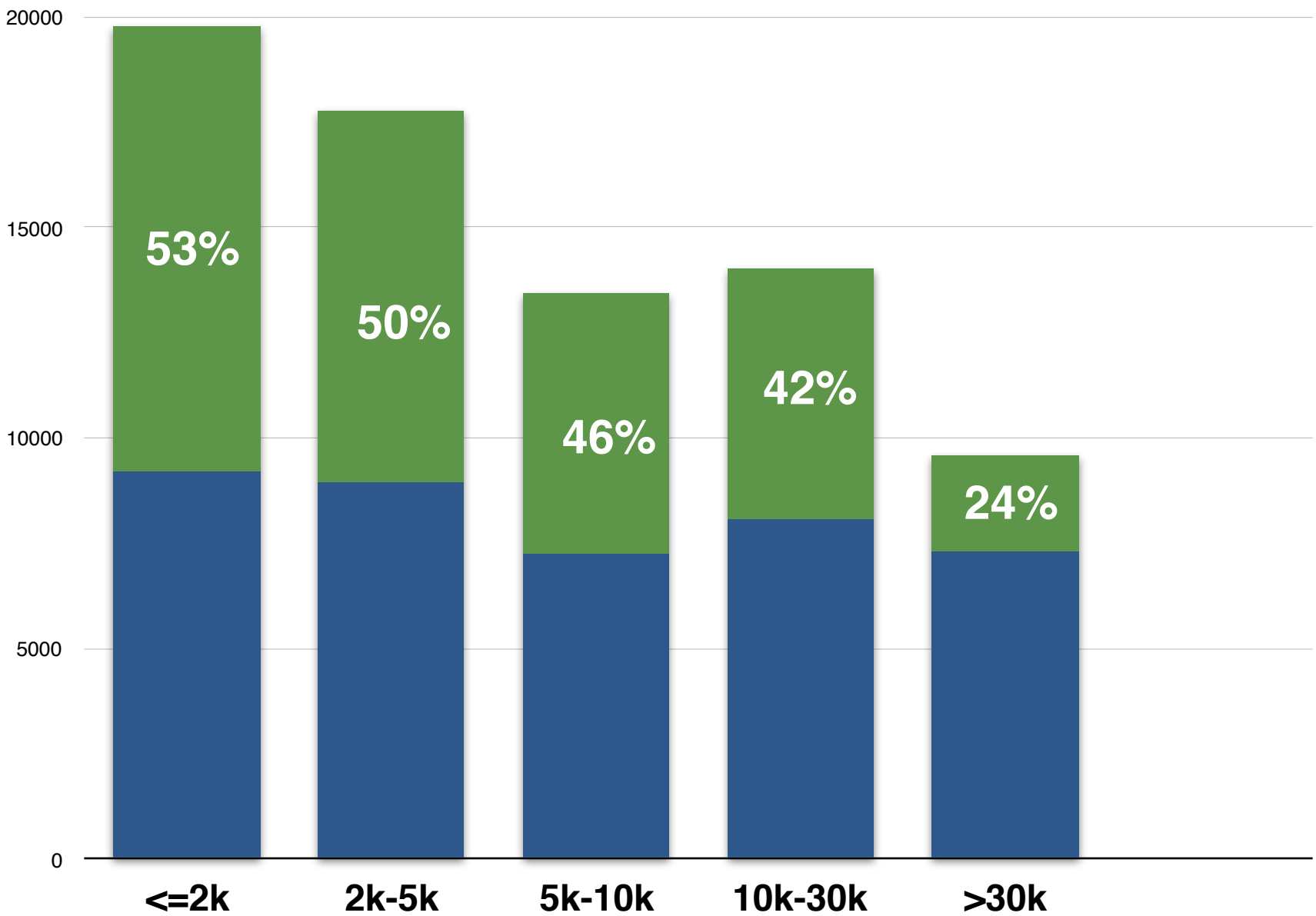
backers	%
<=15	58%
16-35	56%
36-100	45%
>100	22%

Will adding text information from project descriptions build a stronger and more interpretable model?

Metadata

Data Collection

Feature Engineering

Modeling

Prediction

- **Source: Webscraping, Webrobots.io**
- **80,000 projects**

- **Project & Founder info (40 features)**
  - \# images/videos
  - \# pledge levels
  - pledge standard deviation
  - \# projects founder backed
  - \# founder comments

- **Training and testing**
- **Cross validation**
- **Grid search**
- **Model selection**
  (Random Forest Classifier, SVC, Logistic Regression)

- **Scoring metrics:**
  Random Forest: 80% precision, 82% recall

Will adding text information from project descriptions build a stronger and more interpretable model?

Metadata

| Data Collection | Feature Engineering | Modeling | Prediction |

Project descriptions

| Description Vectorization | Modeling | Prediction |
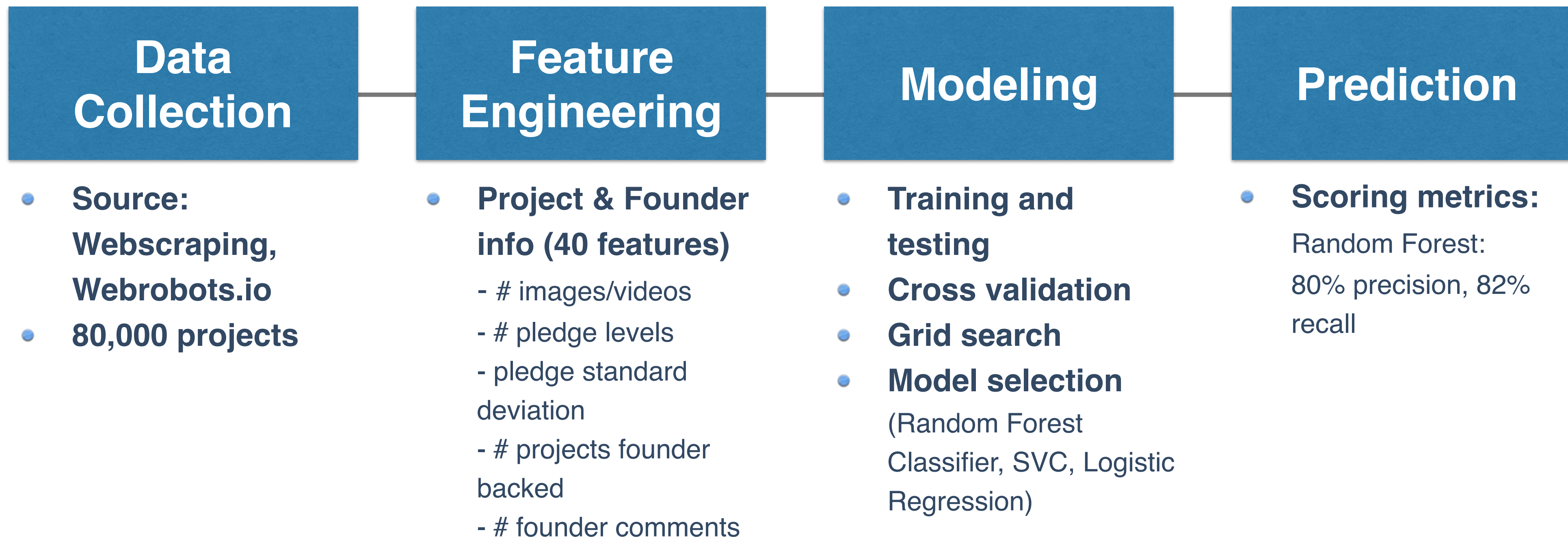
- **TF-IDF of ngrams**
  (n=1,2)

- **Training and testing**
- **Non-negative Matrix Factorization**
- **Model selection**
  (Logistic Regression, Multinomial Naive Bayes, Random Forest Classifier, SVC)

- **Scoring metrics:**
  Mult. Naive Bayes:
  76% precision
  72% recall

Ensemble model captures metadata and latent topics signals to increase score to 83%

Metadata

| Data Collection | Feature Engineering | Modeling | Prediction |

Ensemble model

Project descriptions

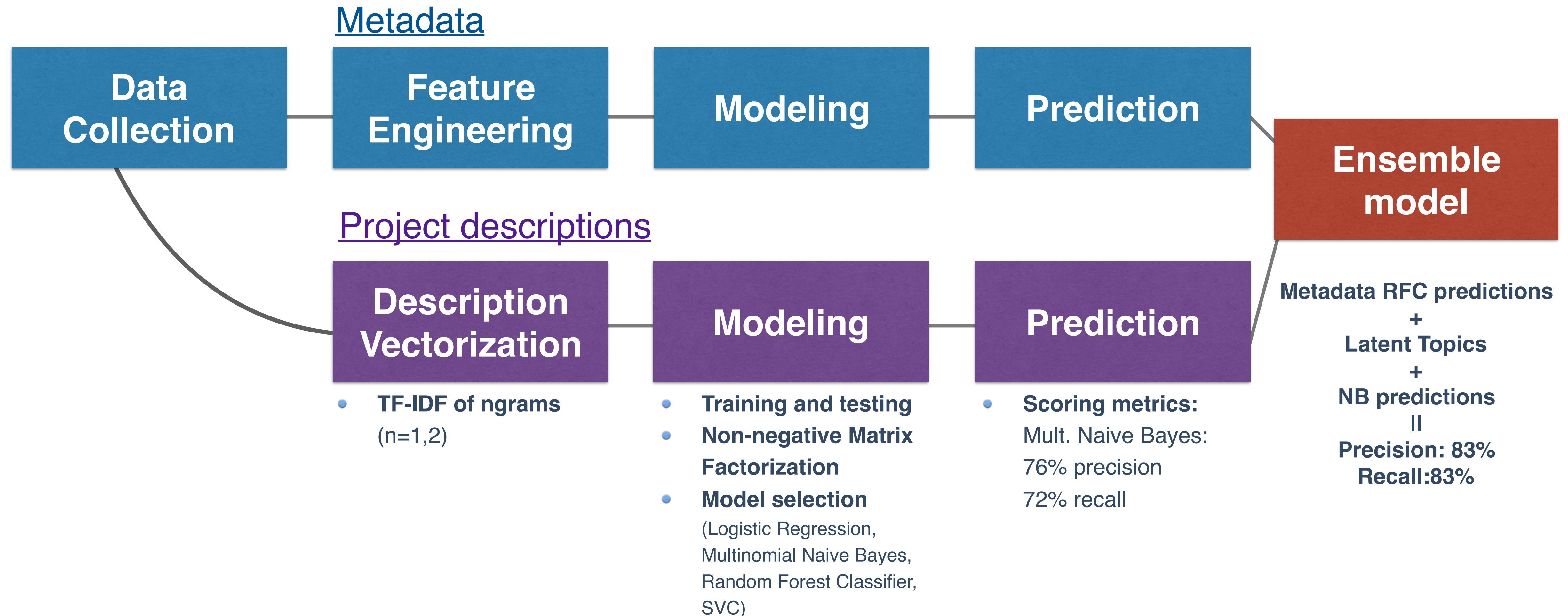| Description Vectorization | Modeling | Prediction |

- **TF-IDF of ngrams**
  (n=1,2)

- **Training and testing**
- **Non-negative Matrix Factorization**
- **Model selection**
  (Logistic Regression, Multinomial Naive Bayes, Random Forest Classifier, SVC)

- **Scoring metrics:**
  Mult. Naive Bayes:
  76% precision
  72% recall

Metadata RFC predictions
+
Latent Topics
+
NB predictions
||
Precision: 83%
Recall:83%

Successful projects followed certain underlying themes

Stacked barplot of successful & failed projects by themes

- successful
- failed

category average
success rate: 37%

founder's background	73%
genre: documentary	88%
new york	65%
genre: horror	32%
plot description	35%
genre: music video	35%
international exposure	80%

Successful projects followed certain underlying themes

Stacked barplot of successful & failed projects by themes

■ successful
■ failed

category average
success rate: 37%

'production', 'films',
'director', 'short', 'producer',
'university','working', 'school',
'years', 'feature',
'currently','television',
'crew', 'arts', 'cast'

'festival',
'film', 'film festival','short',
'best', 'award','feature',
'international', festivals',
'sundance','international film',
'independent','short film',

73%   88%   65%   32%   35%   35%   80%

founder's background · genre: documentary · new york · genre: horror · plot description · genre: music video · international exposure

Future work

- Optical character recognition to detect text within the images

- Explore additional features (sentiment analysis of comments and temporal features)

Thank you

Elizabeth Sia

lizsia@gmail.com

https://www.linkedin.com/in/elizabeth-sia

https://github.com/lizbug/capstone_project

Appendix

Feature importances from metadata Random Forest classifier

Feature importances