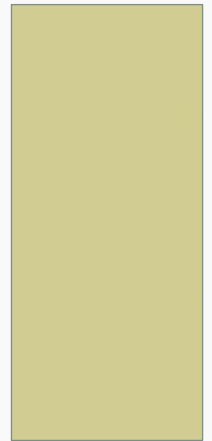


# SUCCESS INDICATORS OF CROWDFUNDING PROJECTS

LUKAS VLCEK



# CROWDFUNDING



Trogdor!! The Board Game  
1,141% funded



Smartibot: The world's first A.I. enabled cardboard robot  
661% funded



The Sons of Kirk: SciFi-Themed Rock Album and Music Video!  
290% funded



"Fuck Me" - A Photobook by Josh Kern  
195% funded

- Crowdfunding platforms
  - Aggregate funds from small investors
  - New technology, art, journalism, games, personal projects, charity
- Kickstarter
  - Biggest, oldest
  - 250,000 creative projects
  - 15 categories
  - International

# COMPETITION FOR FUNDING



- Projects only funded if goal is reached
- Goals from \$1 to \$100,000,000
- 150,000 funded
- 36% overall success rates
- How to increase the chances of success?

# INDICATORS OF FUNDING SUCCESS AND WHO CARES?

- Clients
  - Project proposers: individuals, startups, small companies
- Their goal
  - Maximize the chances of successful funding
- Service
  - Provide guidance in tailoring proposals
  - Payment: fixed amount or percentage of funds if proposal is successful
- How?
  - Predictive modeling based on historical data

# DATASETS: SOURCES AND AGGREGATION

- Online source:
  - Monthly updated repositories collected by Web Robots web scraping company
  - Available in **CSV** formats (30 files each month)
- Combined into a single data frame
- Removed duplicates
- 250, 000 records, 32 features, several in JSON strings
- Reduced from 43 GB to 90 MB of CSV data

# DATA WRANGLING

- Extracted information from JSON strings
- Formed a new feature of category type by aggregating subcategories
- Datastamp strings converted to datetime format
- Selected 18 features for further processing

# EXPLORATORY ANALYSIS

# CHANGES IN TIME

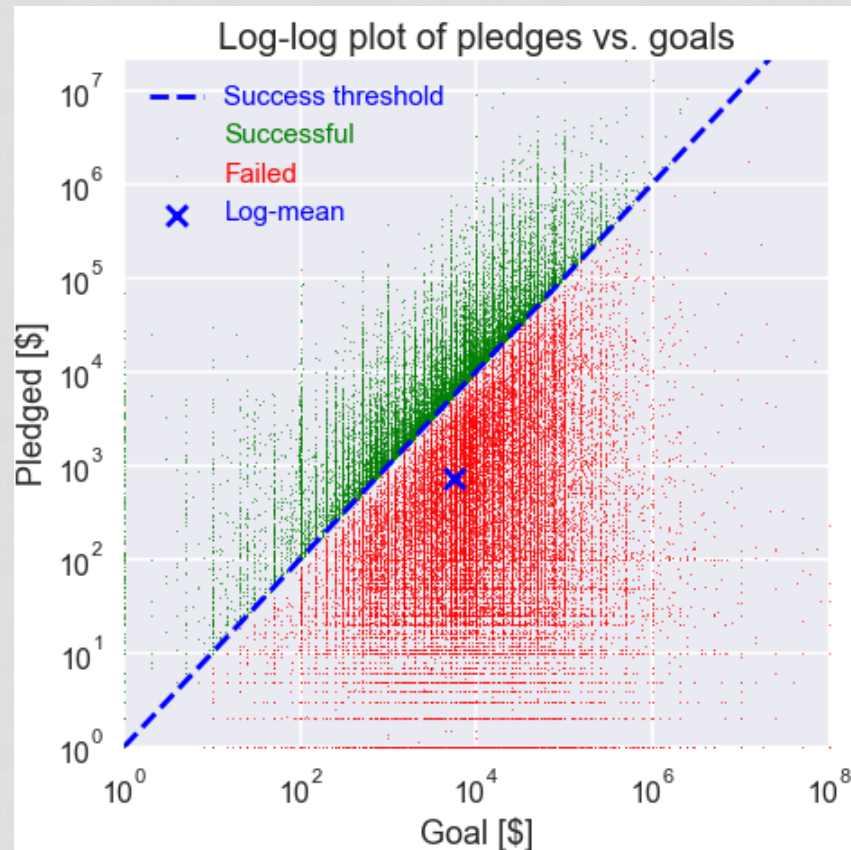


- Project numbers spiked in mid 2014 due to expansion to new countries
- Corresponding decrease in success rates
- Analysis limited to data after mid 2014
- Kickstarter keeps evolving – it pays to follow the trends and adapt the projects to reflect current conditions/demand



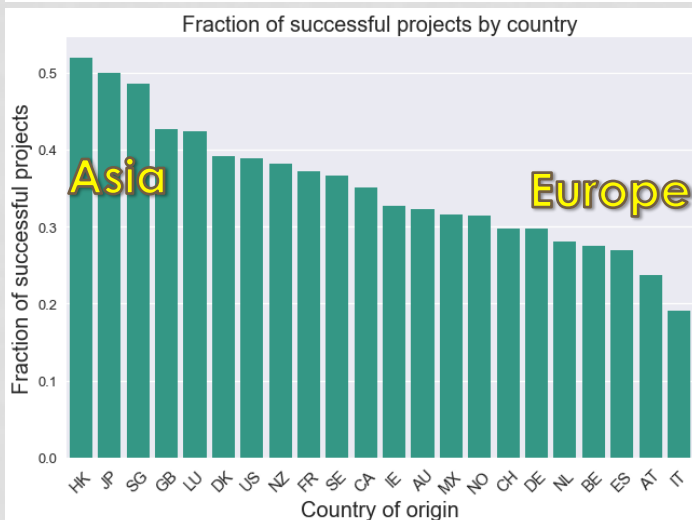
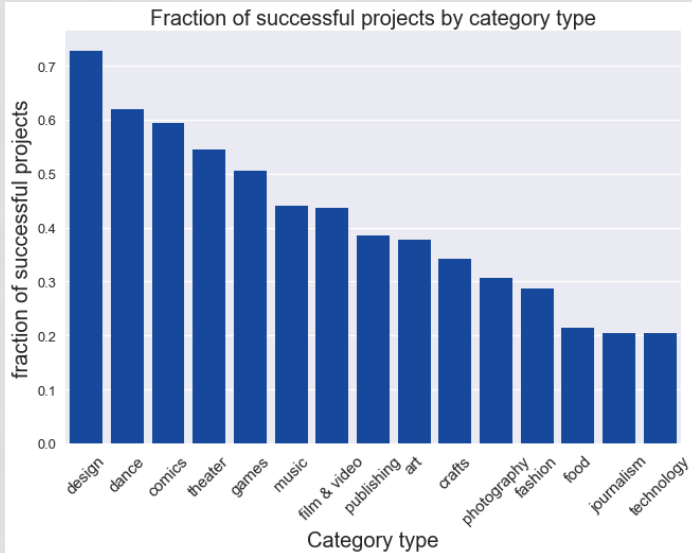


# GOALS VS. PLEDGED AMOUNTS



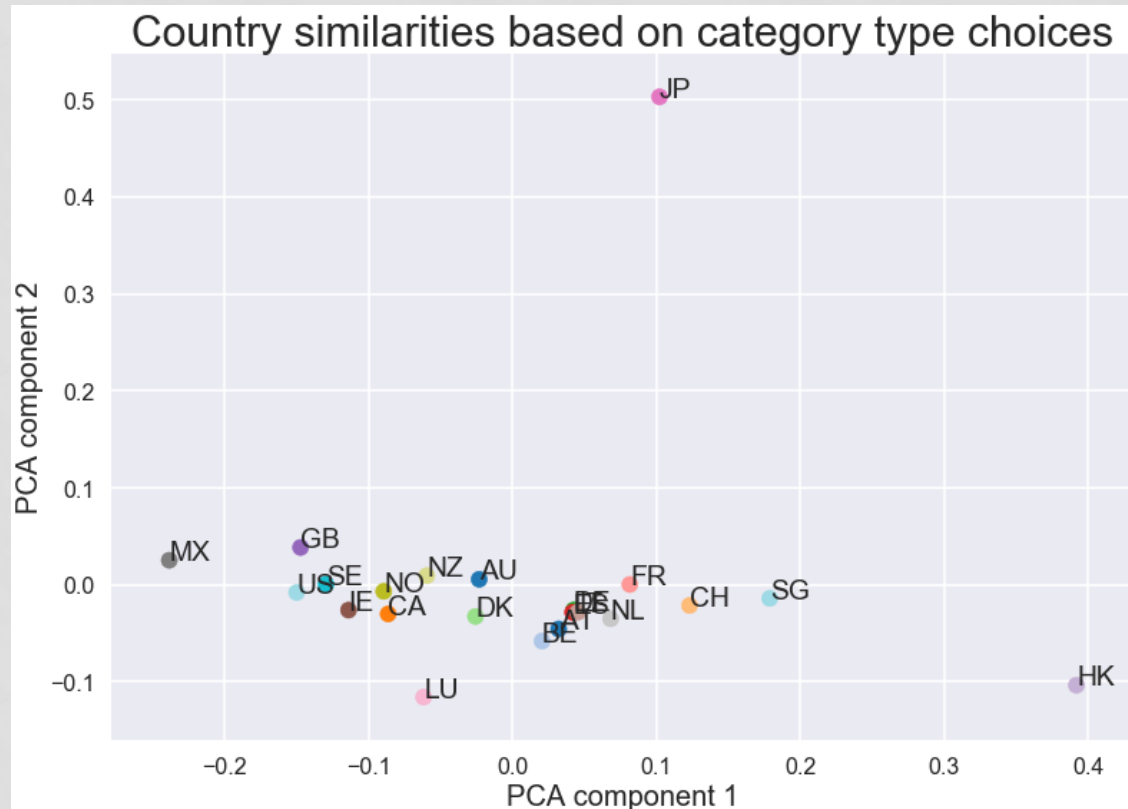
- Goal and pledged amount follow approximately log-normal distribution
- Smaller goals correlated with lower pledged amounts but higher success rates
- It pays to be modest – high goals can be counterproductive

# SUCCESS BASED ON CATEGORY AND COUNTRY OF ORIGIN



- Statistically significant differences between proposal success in different categories and for different countries of origin
- Are these differences correlated?
- Consider the specifics of your project category – there are big differences
- Also consider home bias of investors

# CATEGORY PREFERENCES BASED ON COUNTRY OF ORIGIN



Similarity based on distances in the space of conditional probabilities of category choice, given a country of origin

Reduced to 2 dimensions using the principal component analysis (PCA)

- Compare country-of-origin similarity based on category preferences
- Clear grouping of countries with similar geographic and cultural similarities:
- Europe vs. Asia vs. America
- Technologically advanced vs. developing

# PREDICTIVE MODELING

# APPROACH

- Supervised learning
- Binary classification:
  - Success (1) or Failure (0)
- Logistic regression
- Feature selection
- Tools
  - Pandas + Scikit-learn libraries

## 6 Informative features

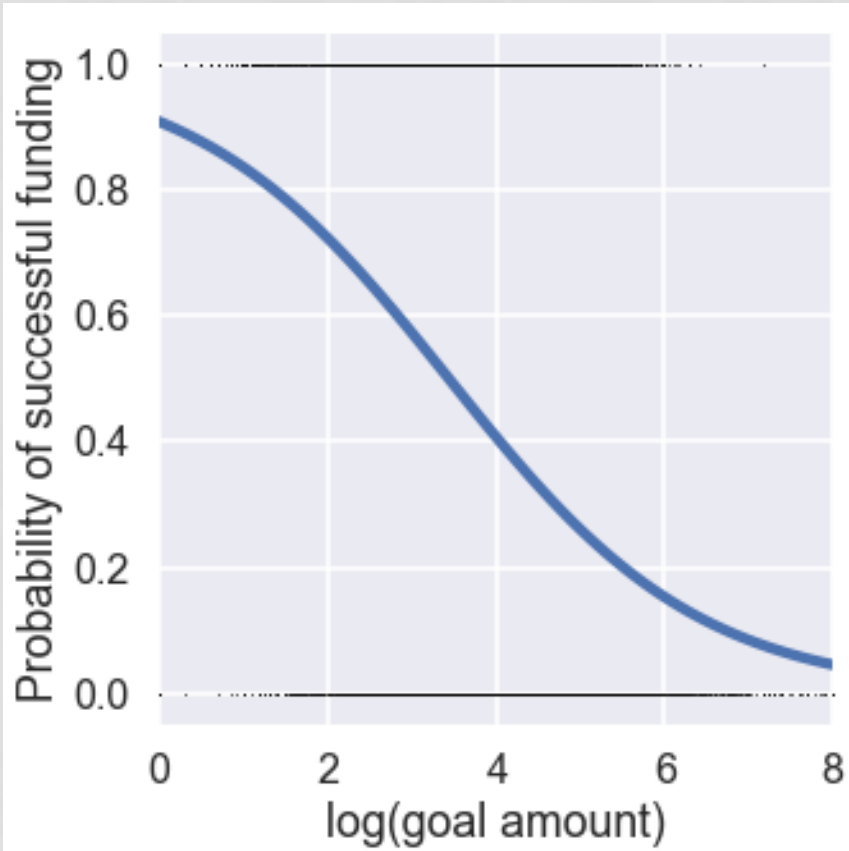
Feature	Type
Goal amount	Numerical
Category type	Categorical
Country of origin	Categorical
Project name	Text
Project description	Text
Project name length	Numerical

# HANDLING TEXTUAL DATA



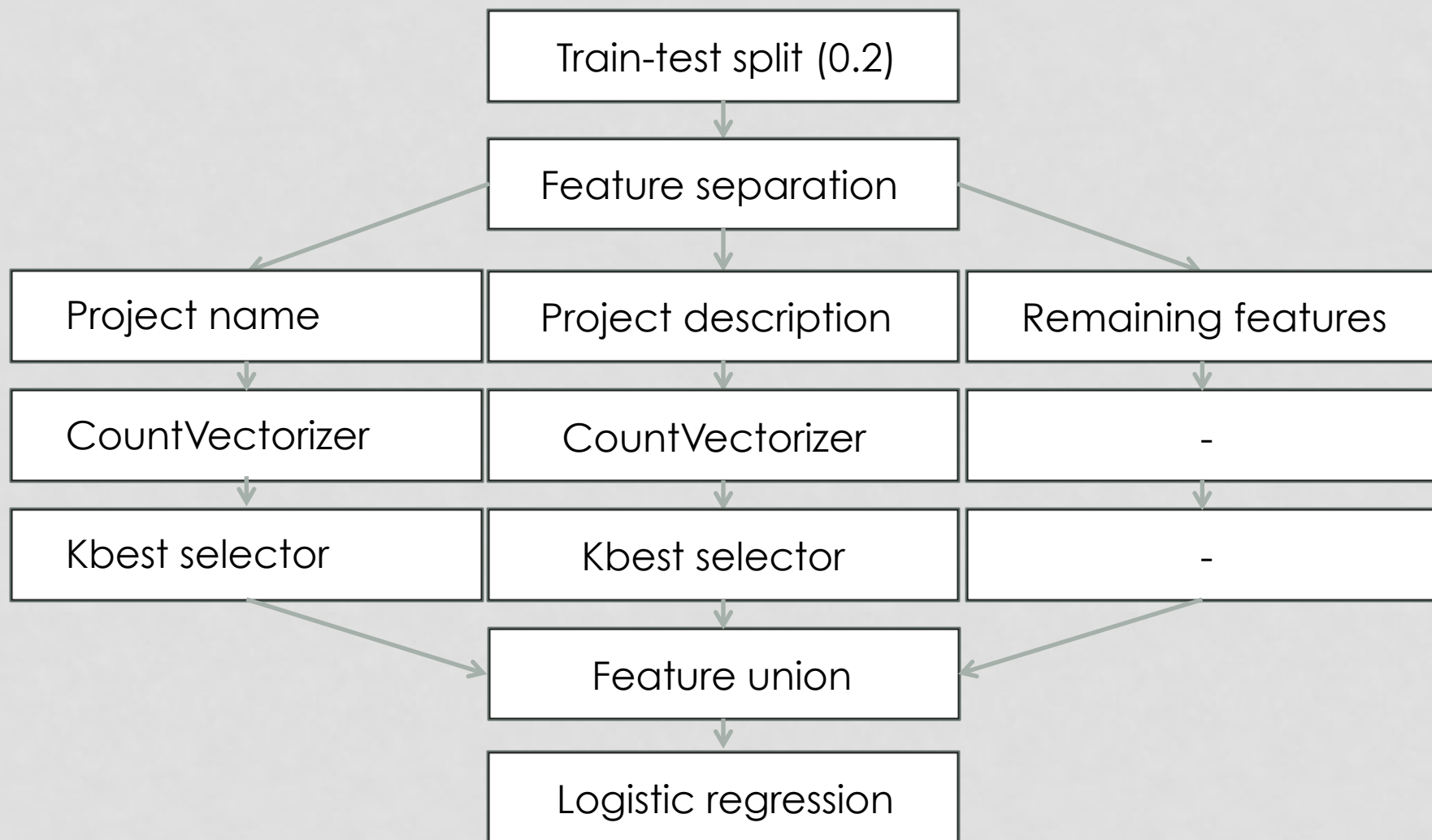
- Features: *project name, project description*
- ‘*Bag of words*’ approach
- Selection of most informative words
  - Chi-squared statistics
- Counts for selected words used in logistic regression
- Optimize project name and description by running it through the model

# LOGISTIC REGRESSION CLASSIFICATION



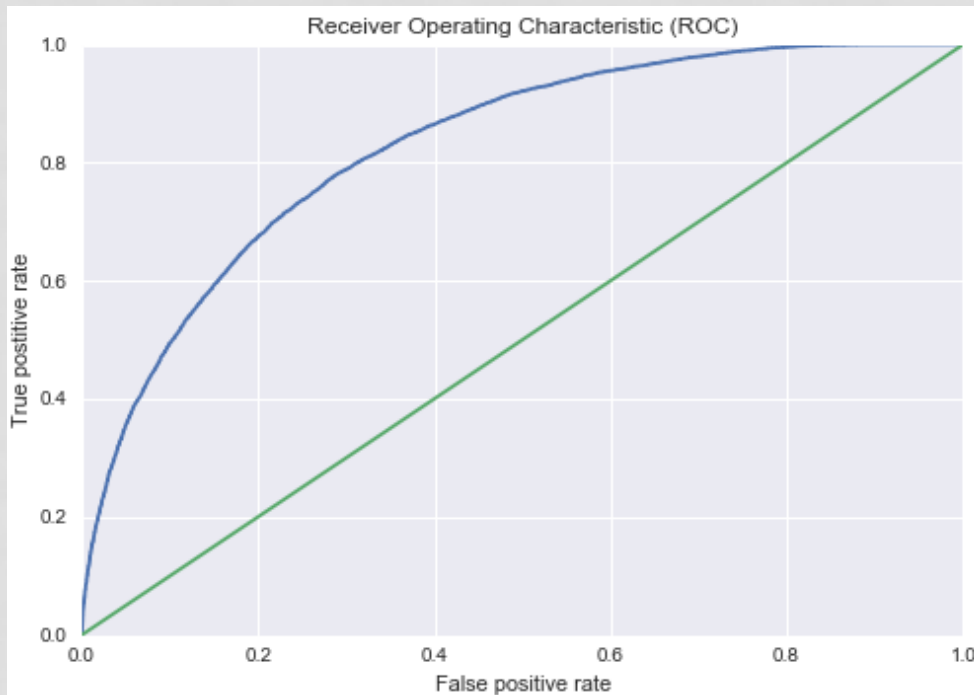
- Simple, fast, easy interpretation
- For textual data: direct mapping from Naïve Bayes classifier
- Consider the minimal viable goal amount for your project.
- The goal does not limit the eventual pledged amount but increases chances of actual funding

# MODELING PIPELINE





# MODEL EVALUATION



**Comparison with staff picks:**  
(10% acceptance threshold):

	Precision	Recall	F1 score
Model	0.86	0.23	0.36
Staff pick	0.86	0.21	0.34

Area under Receiver operating characteristic (ROC) curve: **0.83** ('good' predictor)

Accuracy: 0.75

Precision: 0.70

Recall: 0.60

**Model based on historical records can beat human intuition/assessment and provide guidance in proposal preparation, saving invested time and money.**

# CONCLUSIONS

## PREDICTING PROJECT SUCCESS

- Model allows tuning crowdfunding proposal parameters, such as goal amount, project description, and refine category choice
- Prediction metrics based on historical data can match or exceed human assessment (staff picks)
- Speed and accuracy of the model can be used for filtering out 'bad' ideas and increasing the chance of good ones
- Model can be further improved by more sophisticated natural language processing methods and including interactions between features (beyond linear model)