Launch Hard or Go Home!

Predicting the Success of Kickstarter Campaigns

Vincent Etter, Matthias Grossglauser, Patrick Thiran
EPFL, Lausanne, Switzerland
What is Kickstarter?

• **Crowdfunding** website

• Raise money to launch creative projects

• Each campaign has:
  
  • a funding **goal**
  
  • a campaign **duration**

• **All-or-nothing** funding model
A Success Story

Since its launch in 2009:

- **50,000** projects successfully funded
- **44%** success rate
- **$811,000,000** raised
- **4,900,000** backers
Our Dataset

• Discover campaigns on the "Recently Launched" page

• Regularly get their status: number of backers, amount of money pledged

• Monitor Twitter in parallel for all tweets containing keyword kickstarter
Our Dataset

• From September 2012 to May 2013:
  • 16,042 campaigns
  • 48.24% of successful campaigns
  • 1,309,295 backers
  • $158,026,656 pledged
  • 737,398 "kickstarter" tweets
Our Predictors

- **Baseline**: campaigns' static attributes [1]
- Two sources of information:
  - **Money**: kNN, Markov
  - **Social**: tweets, project/backer graph

Graph Predictor

• Can we use **backers** as source of information?

• Consider the project/backer graph:
Graph Predictor

• Extract several features:
  • # of backers
  • #/prop. of first-time backers
  • # projects with co-backers
  • #/prop. of these that are successful

• Train a Support Vector Machine
Prediction Accuracy

The graph illustrates the prediction accuracy over campaign progress. The x-axis represents the campaign progress, ranging from 0.0 to 1.0, while the y-axis shows the prediction accuracy in percentage, ranging from 65% to 100%.

- **kNN**: A machine learning algorithm used for classification and regression. The graph shows a smooth increase in prediction accuracy over campaign progress.
- **Markov**: A probabilistic model used to model sequences of events. The graph indicates a steady rise in accuracy, slightly below the kNN trend.
- **Graph**: Refers to a structured data set that can be represented as a network of nodes and edges. The graph suggests a moderate increase in accuracy, comparable to kNN but not as steep.
- **Baseline**: Represents a simple prediction model, often used as a benchmark. The graph shows a flat line, indicating stable accuracy throughout the campaign progress.
- **Tweets**: Likely refers to social media interactions or textual data. The graph suggests a slightly elevated prediction accuracy compared to the baseline, but lower than the kNN and Markov models.
Combined Predictor

- Money-based results are quite good
- Can we use social predictors to help?
- Train a SVM to combine individual predictions
Combiner Results
Conclusion

• Combiner improves early predictions
• There is potential in social predictors
• Future work:
  • Improve graph predictor
  • Take dynamics on Twitter and project/backer graph into account
Thank you!

Data and real-time predictions on
http://sidekick.epfl.ch