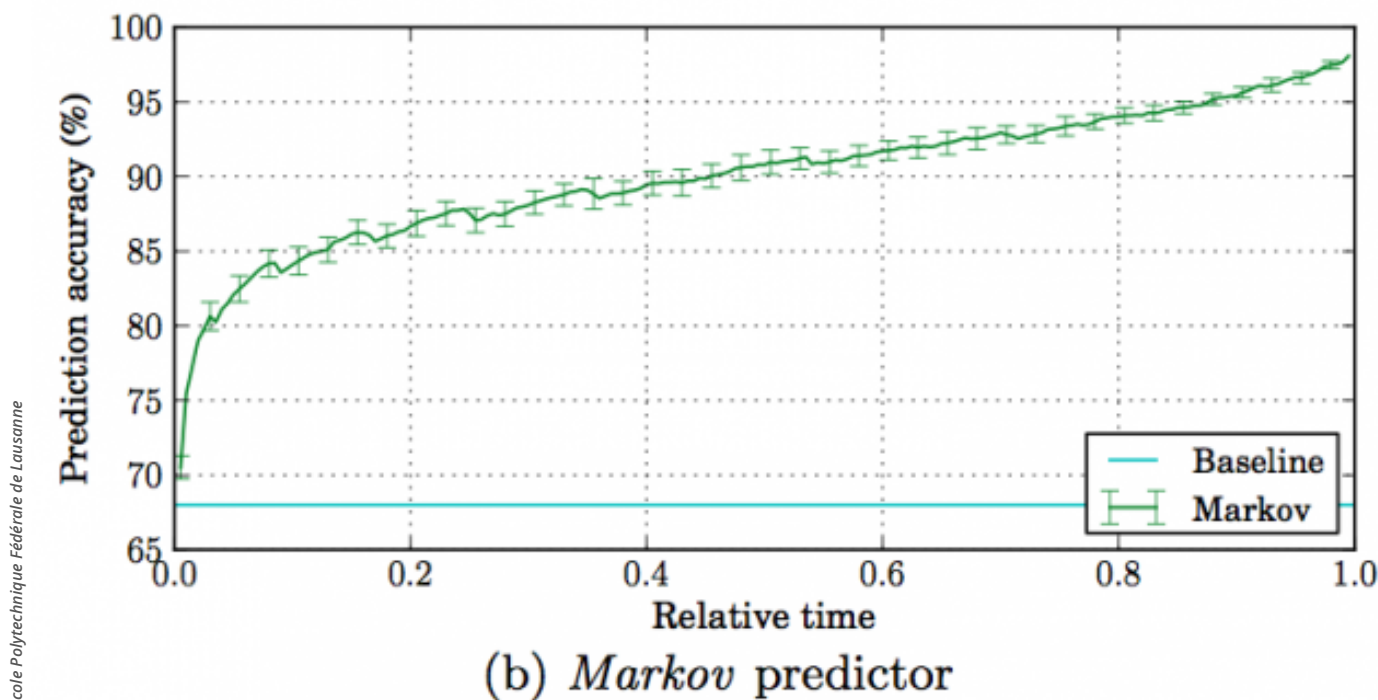


MINISTRY OF INNOVATION —

Statistical models can predict a Kickstarter's success within 4 hours

How people tweet about your Kickstarter matters a lot in this model.

CASEY JOHNSTON - 10/17/2013, 8:41 AM



École Polytechnique Fédérale de Lausanne

The Markov predictor's accuracy over time, using only the timing of financing and backer info.

Kickstarter has become the Internet's prime vector for Cinderella stories, catapulting pet projects to fame and burying would-be entrepreneurs in more logistics and minutiae than they were ready to handle. There are many different degrees of success on Kickstarter, but when broken down to a binary yes/no score, a group of scientists have found that **they can predict** with reasonable confidence whether a project will succeed or fail within the first four hours of its launch. Their method is based in part on its social media reception, according to a paper presented early in October.

Three researchers at the École Polytechnique Fédérale de Lausanne created statistical models fed with both funding data and discussions on Twitter. The data set was pulled from over 16,000 Kickstarter campaigns that had raised a collective total of \$158 million; approximately half of them failed.

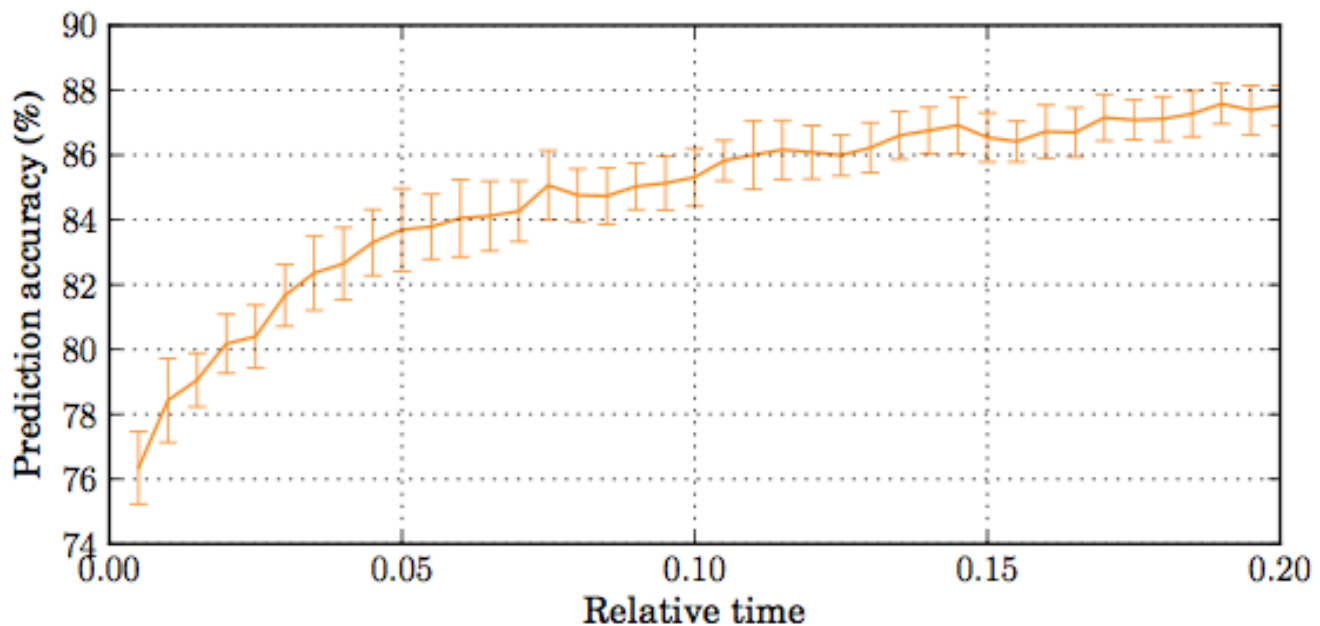
The scientists collected Twitter data by searching for the word "kickstarter," then matching tweets to the Kickstarter project using URLs included in the tweets. They also culled information from each project's "Backers" page to get a list of what users pledged money, and how much they had pledged collectively. The second step was time-intensive, so the authors only completed it every two days.

At first, they fed only the backers data into two models, a **k-nearest neighbor** classifier and a **Markov chain model**. For a control, the authors used a baseline static model that took into account factors like whether a project had a video or not, its category, and its financial goal.

The baseline model was able to predict the success of campaigns with a flat 68 percent rate. Both the nearest-neighbors and Markov chain models fared far better, even in the early hours of a project's lifetime.

At only 10 percent of the way through a Kickstarter's life, or about 3 days, both models could find the ultimate result with about 85 percent accuracy. The nearest-neighbor classifier starts off as a slightly better predictor and edges into the eightieth percentile more quickly, though the researchers note that model is significantly more computationally expensive compared to the Markov model.

Next they added the tweets. The authors augmented their models into "support vector machines" to process the twitter data they'd collected, which used tweets' timestamps as well as the number of replies and retweets. The authors found that a prediction model using tweets alone did not fare much better in the early stages of a Kickstarter campaign than the static control model did.



In a more granular view of the first fifth of a Kickstarter's lifetime, the prediction model gains significant accuracy very quickly, in a matter of hours.

However, when combined with financial data, the models performed the best in early days, shooting up to around 84 percent within the first day and a half and climbing to 87 percent at the end of the first six days.

Using the combined financial and social information, the final model that combined all this information was able to crack a 76 percent prediction rate within four hours of a campaign's launch—four percent better than the next best model, the researchers said. The highest prediction accuracy, though, would come at around 15 percent of the way through a campaign, or about four and a half days in. They peaked at 85 percent, using only fin

[SIGN IN](#)

Kickstarter who neglected to mention that it's a Kickstarter, for one. Likewise, URLs linking to that Kickstarter can be masked by different URL-shortening services, making it hard to track all tweets that are directing people to the same place.

Because the authors' prediction models are complex, they don't identify trends that any user could pick out to determine whether a particular project will crack its goal. There is a good answer inside Twitter's black box—or at least three-fourths of one—it's just not visible to the naked eye.

CASEY JOHNSTON

Casey Johnston is the former Culture Editor at Ars Technica, and now does the occasional freelance story. She graduated from Columbia University with a degree in Applied Physics.

TWITTER [@caseyjohnston](#)

READER COMMENTS

SHARE THIS STORY

[← PREVIOUS STORY](#)

[NEXT STORY →](#)

Related Stories

Lack of Twitter geotags can't stop researchers from getting location

The ugly afterlife of crowdfunding projects that never ship and never end

Internet users now have more and closer friends than those offline

Internet's cherished spambot turns out to be just another human

Today on Ars

[RSS FEEDS](#)

[VIEW MOBILE SITE](#)

[VISIT ARS TECHNICA UK](#)

[ABOUT US](#)

[CONTACT US](#)

[STAFF](#)

[ADVERTISE WITH US](#)

[REPRINTS](#)

WIRED Media Group

Use of this Site constitutes acceptance of our [User Agreement](#) (effective 1/2/14) and [Privacy Policy](#) (effective 1/2/14), and [Ars Technica Addendum](#) (effective 5/17/2012). Your California Privacy Rights. The material on this site may not be reproduced, distributed, transmitted, cached or otherwise used, except with the prior written permission of Condé Nast.

[SIGN IN](#)