



Tip Ranker: A M.L. Approach to Ranking Short Reviews



Enrique Cruz, Berk Kapicioglu

The Problem

Foursquare users have written over 87 million tips (short reviews) at millions of different venues and places around the world.

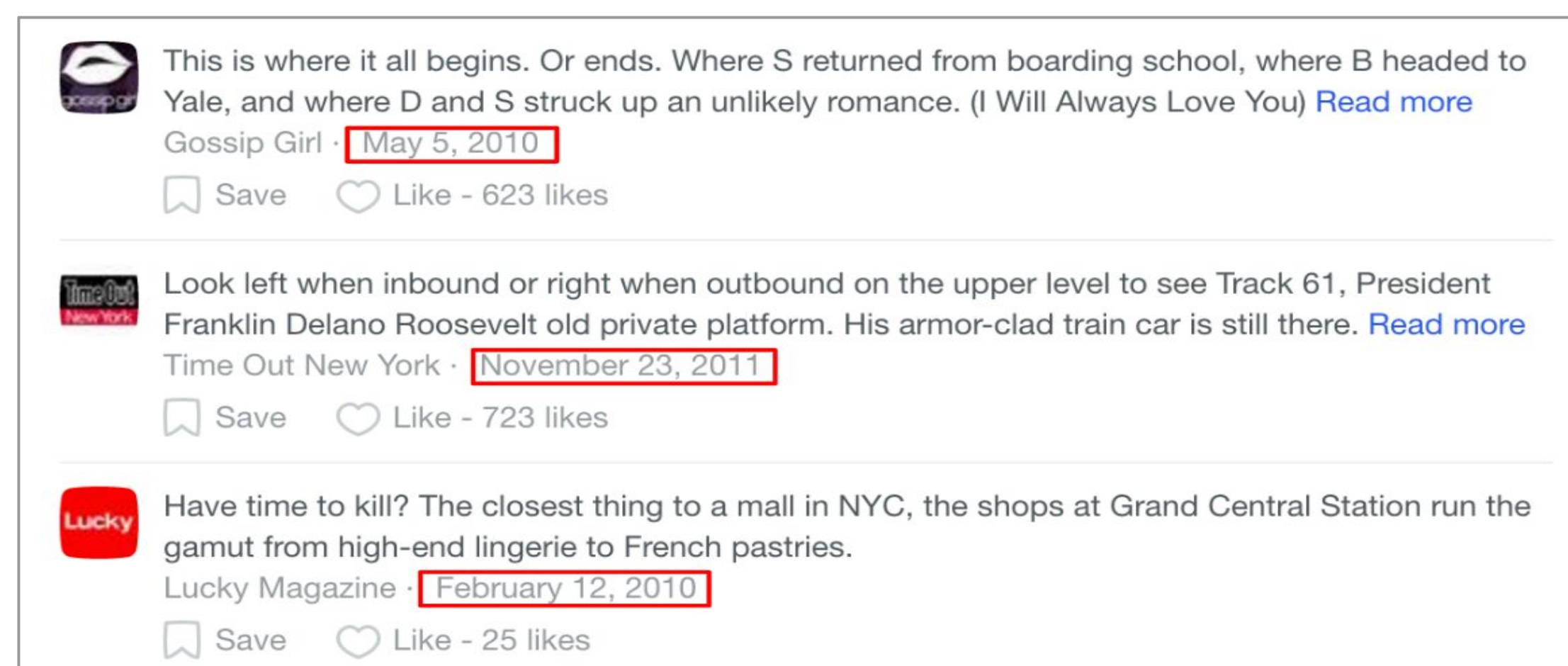
Whether to proactively ping a user with the best tip when he is at a given place, or to order the tips shown on a venue's page to maximize quality of the content - **How do we best rank tips?**

Background

- Foursquare has previously tried used a few different mechanisms to tackle tip selection.
- None of these were fully satisfactory on their own, each having its particular set of shortcomings.

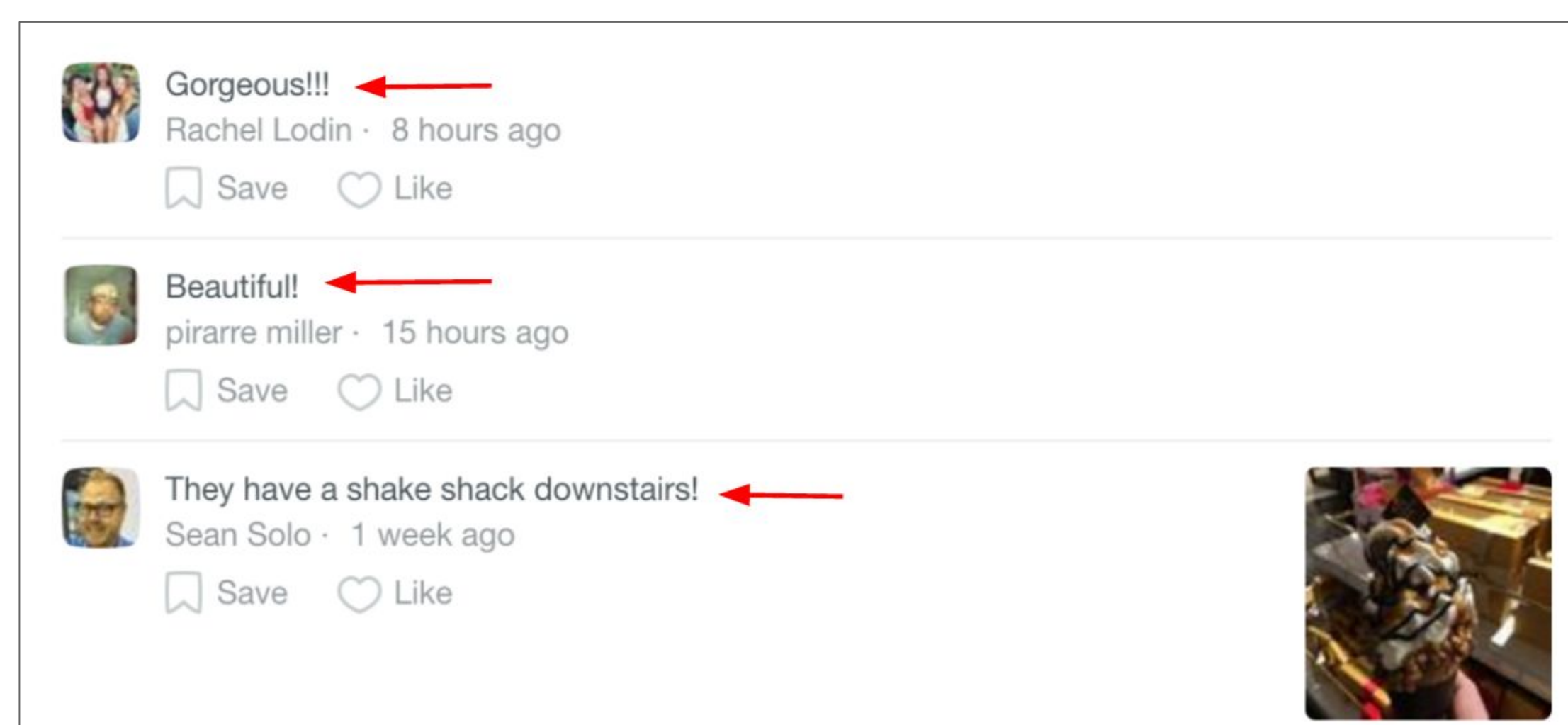
POPULARITY Measurements of social interactions on tip.

- ↳ **PRO** Showcases relevant content that is interesting and useful to users.
- ↳ **CON** Favors old & stale content. Leads to a feedback loop prominently exposing old content.



RECENCY Time that has passed since the tip was created.

- ↳ **PRO** Keeps the content fresh and showcases our vibrant and active community.
- ↳ **CON** Offers no guarantees of quality or relevance.



Feature Components

 LANGUAGE IDENTIFICATION Prefer tips written in the user's language	 POPULARITY Measure of social interactions	 META-CONTENT RICHNESS Presence of photo, tastes and number of tokens
 AUTHOR TRUST User statistics (i.e. tenure and historical popularity of tips)	 RECENCY Time elapsed since tip creation	 GLOBAL QUALITY Statistical classifiers trained to identify traits like sentiment and spam

Data & Labels

1. Select the top 10K venues by user views
2. From each of these randomly sample 100 pairs of tips
3. Language filter and de-dupe → 50K pairs total
4. Collect labels via crowdsourcing
 - a. Show tip pairs in context and ask judge for ordering
 - b. Verify judge trust with honeypot
 - c. Collect judgements per pair and aggregate



Given this venue: **Black Tap**

A) Any of the burgers. I'm personally a fan of the All American, but I've never had a bad burger. Great craft beer selection too. Lots of amazing west-coast beers. - Nat Welch - 6/15/2015

B) Amazing All American Burger beats The Burger Joint at the Parker Meridian - Gareth Rees-John - 6/13/2015

Which text is more informative given the above venue?(required) **100% agreement**
TYPE: CML-RADIOS VALIDATORS: REQUIRED, SS-REQUIRED

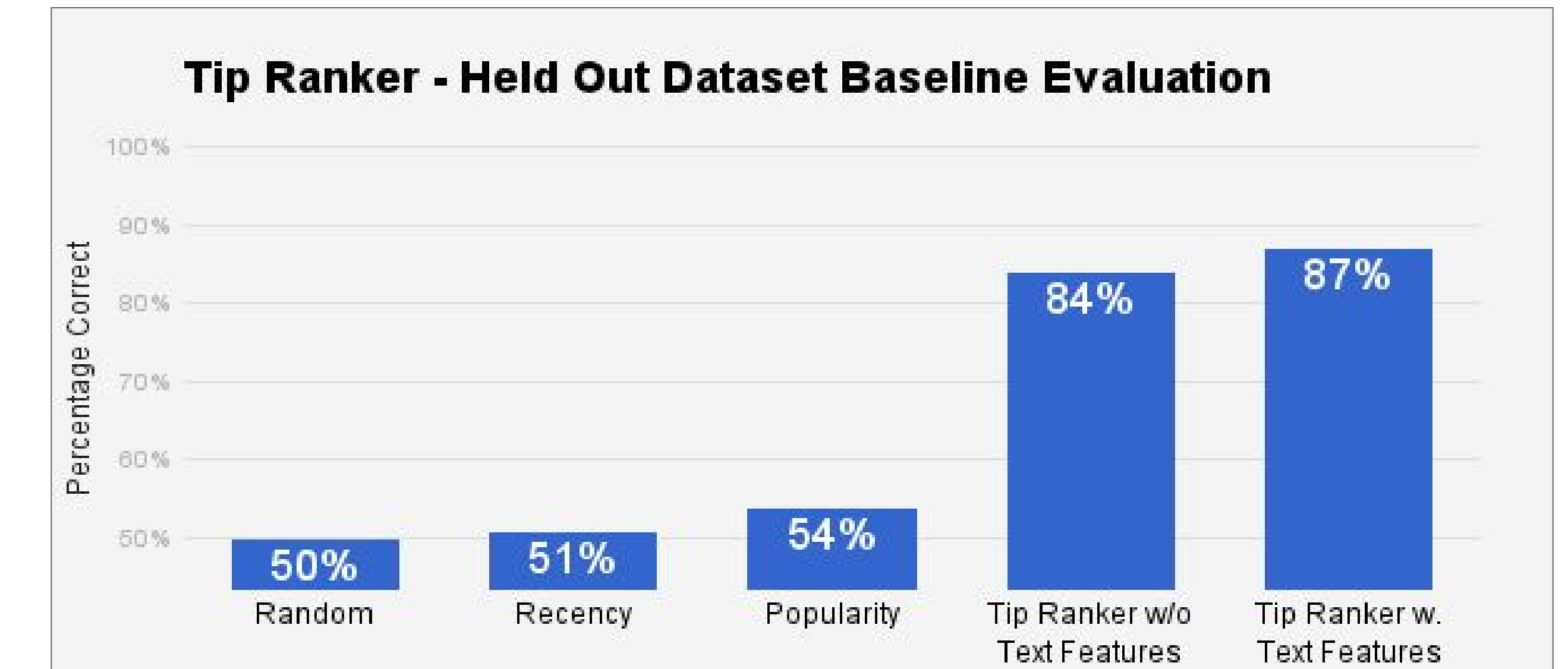
A is More Informative 100%

B is More Informative 0%

Both are Equally Informative or Equally Uninformative 0%

Training & Evaluation

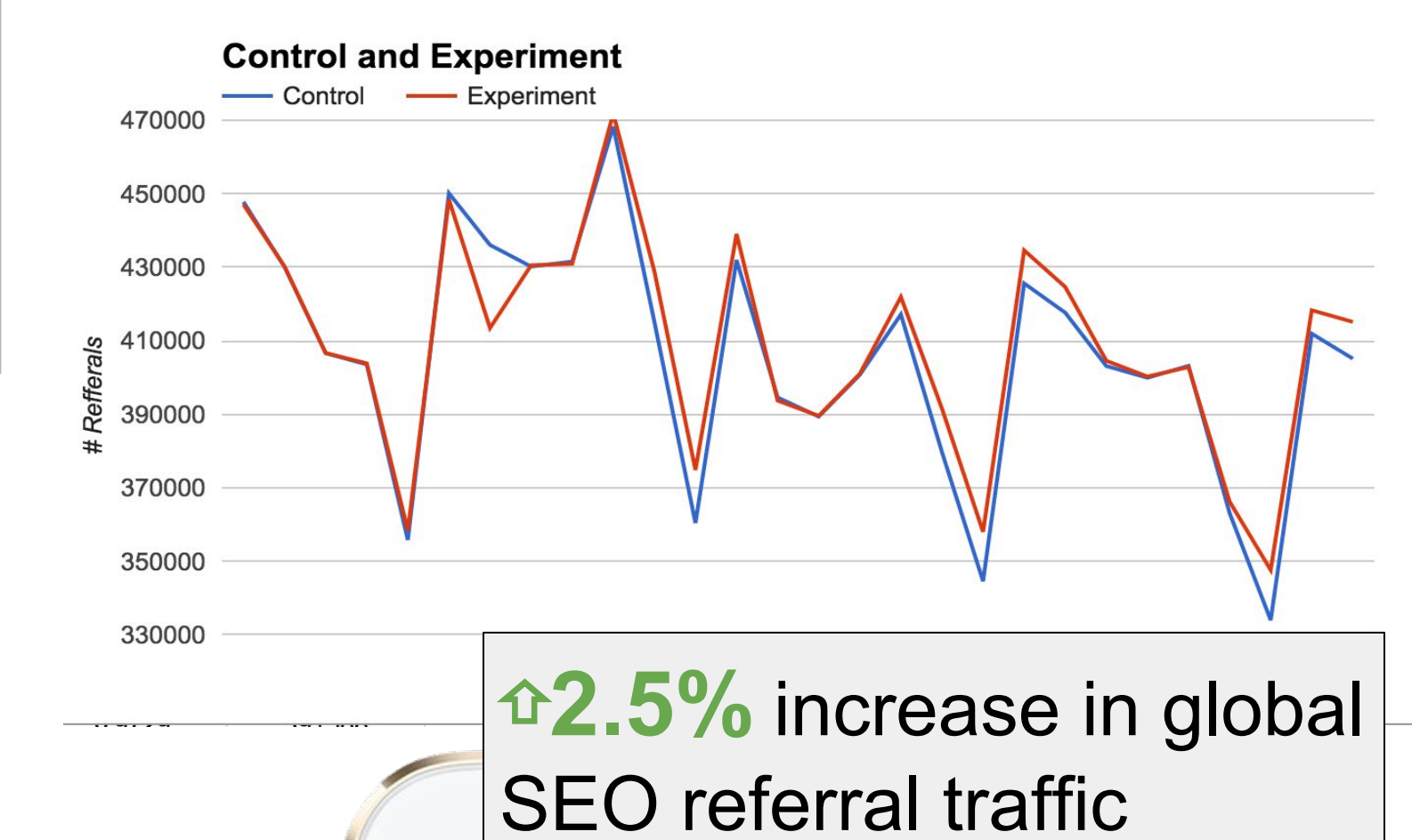
- Supervised learning algorithm: Support Vector Machine.
- Optimization metric: Minimizing the number of misordered pairs of tips with respect to labels.
- Evaluation: Measured the model's performance against a held out dataset and compared it against baseline metrics.



A/B Testing & Production Applications



- ↑5% more core app activity days
- ↑1.5% increase in ping CTR
- ↑32% more pings sent



- ↑1% Foursquare core active users from new sign-ups via Swarm

