



SPCA Retweet Predictor

...

Jason Montana
September 14, 2017

Question:

Can I predict the number of retweets
for SPCA organizations' tweets?



Twitter Tutorial

Hashtag

Text

Retweet

Reply

Like



This tweet has shown up in the feed of San Francisco SPCA's 19,600 followers. I decide to retweet it, so I clicked the retweet button, which means...

Twitter Tutorial

Retweeted

Now this tweet can
be seen by ALL my
followers



an additional 153 people (my wealth of followers) had this tweet appear in their feed. And they in turn can retweet it too, making it viewable to their followers and so on. Thus the viral nature of retweeting.

Why SPCA Organizations?



Meet
Kahlua!

In addition to 5 years of volunteering at the San Francisco SPCA, my wife and I adopted Kahlua at 8 months old back in April 2015.

Assumptions and Limitations

Photos

Assumption:

Every tweet these days includes a photo.

Limitation:

Photo characteristics are not within the scope of this project.

Retweets

Assumption:

The tweets have reached retweet maturity.

Limitation:

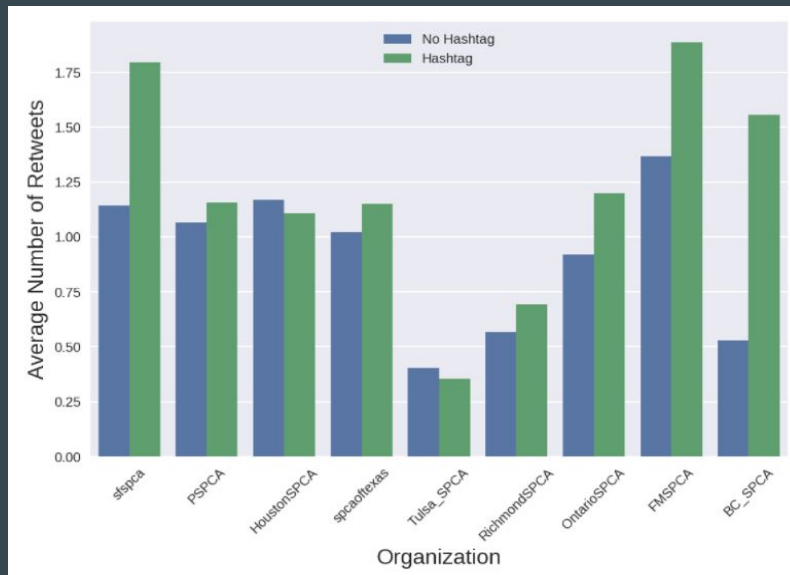
The alternate method for retweeting is not considered in this project.

Understanding the Data

- 1 million random tweets to build Twitter dictionary
- Over 92,000 tweets from 9 SPCA organizations in North America
 - San Francisco SPCA
 - Pennsylvania SPCA
 - Houston SPCA
 - SPCA of Texas
 - Richmond SPCA
 - Fort McMurray SPCA (Alberta)
 - Tulsa SPCA
 - Ontario SPCA
 - BC SPCA (British Columbia)
- Ultimately pared down to about 80,000 tweets

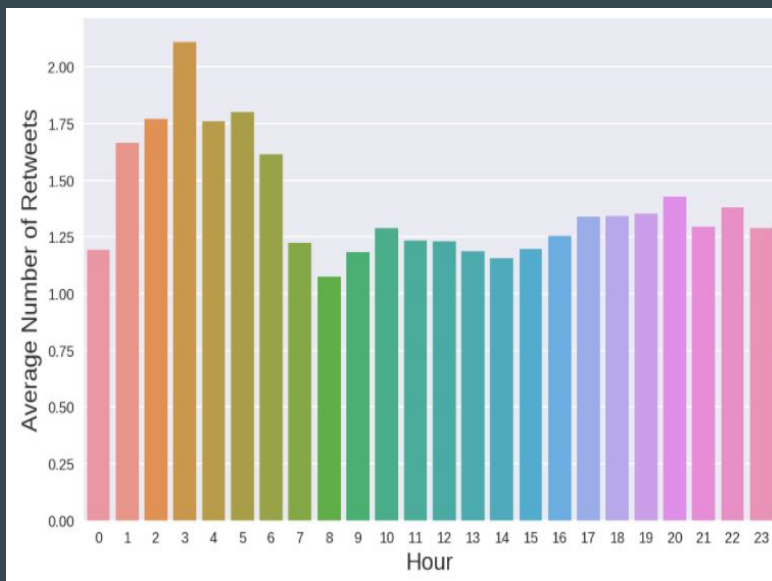
The 1 million random tweets came from a stream I set up one weekend to capture tweets from all over the US and Canada. This helped build the dictionary that was later used to analyze the text of the SPCA tweets.

Average Retweets Based on Hashtags



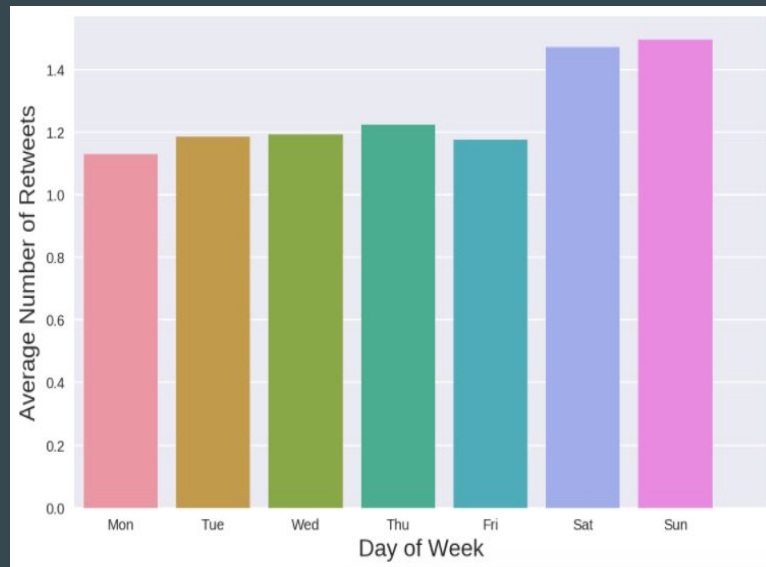
For most organizations, including a hashtag in their tweets resulted in a slight bump in retweets.

Average Retweets by Local Hour



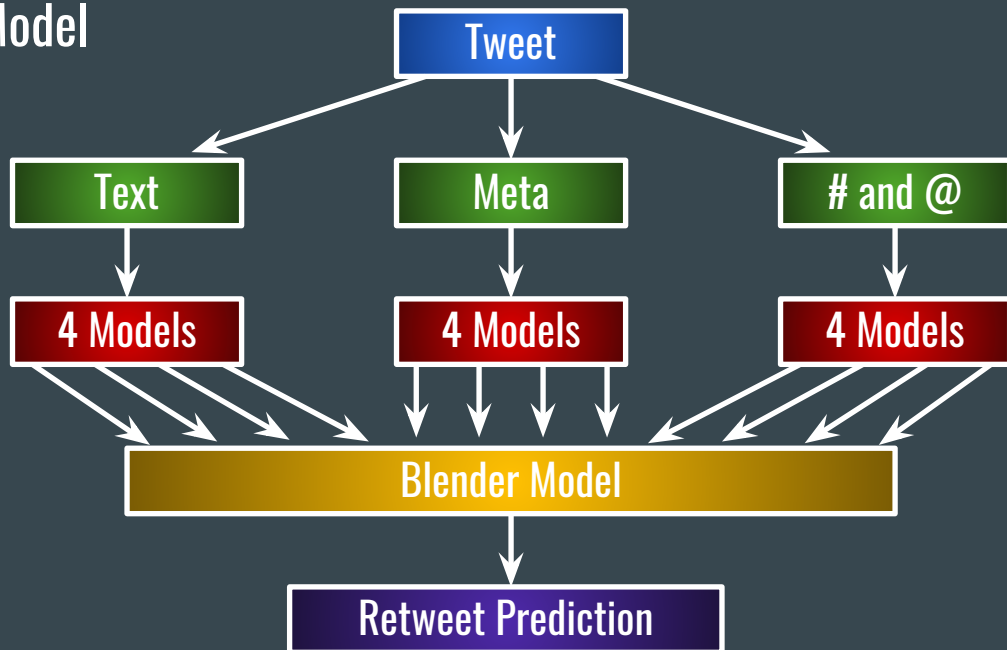
Tweeting in the early morning hours leads to more retweets. 96% of the tweets sent in the 3am hour were by the Fort McMurray SPCA in Alberta. Note: all times are local to the organization.

Average Retweets by Day of Week



Weekend tweets tend to get more retweets.

The Model

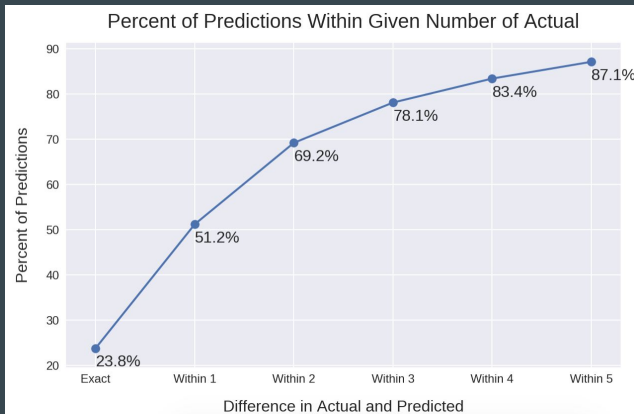


No individual model gave favorable results, so I decided to use all the mediocre individual models to make predictions. Then I fed those 12 predictions into a final model (blender) to arrive at a final prediction. Doing it this way allowed me to keep the specific strengths of each of the individual models.

Accuracy

Baseline: 0 retweets appeared most in the data - 24.9%

The model correctly predicted retweets for 23.8% of the tweets.



Though more than 51% of the tweets were predicted within 1 retweet of the actual number!

Guessing 0 retweets would have given a correct guess 24.9% of the time. The model was only correct 23.8% of the time; however, it was not off by much. An additional 27.4% of the predictions were off by only 1 retweet and another 18% were off by 2. To me, these were encouraging results.

Next Steps

Better Tune the Models

- Are all 12 Layer 1 models necessary?
- Should the Blender be modified?

User Interface

- Predict the retweets of a draft tweet
- Recommendations for edits to the tweet

Track Impact

- Connect to adoptions, donations, volunteers

Photos

- Add photo analysis to model

Since the incorrect predictions didn't miss by much, I think it's worth trying to sharpen up the model rather than embark upon an entirely new approach.