Stat-495 Initial Proposal

Emily Ye & Shukry Zablah

2019-10-27

Team Name: YZ Analytics

Title: The Life and Times of a Tweet

Purpose: As of 2019, Twitter has 330 million monthly active users, and more than 500 million tweets are sent each day (Ling, 2019). How, then, do certain tweets gain traction? The purpose of this project is to track how original tweets travel through retweets and reach their audience. This project would be useful for Twitter users trying to understand the impact they can have through the platform.

Data: We will be working with Twitter data, using the package `rtweet` (Kearney, 2019) to interact with Twitter's APIs using commands such as `search_tweets`.

Variables:

Identifying information of author:

- screen name
- user ID
- number of followers

Engagement of tweet:

- number of favorites
- number of retweets

Tweet content:

- the text of the tweet
- hashtags
- if the tweet has any images, texts, GIFs, videos or other media

Tweet metadata:

- location of tweet

Model: Our model aims to predict the number of retweets of a tweet. We are thinking of using a regression tree.

End Product: We aim to demo our analyses online with dynamic visualizations. We will display analyses of tweets from a set list of individuals. We also hope to create a web application for users to personalize our analysis to themselves. For example, such a web application could allow a user to type a tweet they want to send, input their follower count, and the app would display their predicted number of retweets.

References

Kearney, M. W. (2019). rtweet: Collecting Twitter Data. R package version 0.6.9, https://cran.r-project.org/package=rtweet.

Ling, Y. (2019). 10 Twitter Statistics Every Marketer Should Know in 2019. https://www.oberlo.com/blog/twitter-statistics

The app and idea sound good. I'm not familiar with the packages, but am concerned about the # of variables available. Can you do something else to get more vars? Like sentiment or some other analysis of tweet text? Can you make sure you can get data from the packages? ie. get a sample to make sure you know what vars you have, etc.