Team5 – Project Proposal

DATA620 - Web Analytics - Summer 2018

Christopher Estevez, Meaghan Burke, Rickidon Singh, Ritesh Lohiya, Rose Koh

- **Overview**

  Twitter provides a social media platform for users to interact with short messages known as "tweets". Users can effectively build substantial networks by connecting with friends and following other users. For the final project, team 5 propose to employ some of Data 620 concepts to understand the overall tweet sentiment and social network influence of a popular Twitter user.

- **Data**

  While tweets are restricted to around 140 words, Twitter stores tweet and user metadata such as user name, ID, creation date, user location, the number of re-tweets, etc. This expansive data coverage will help team 5 understand the Twitter user's network and to measure his/her relative influence within the network.

  The tweet and tweet metadata can be accessed through Twitter's API service. Twitter allows a maximum download of 5,000 records. The below link is how team 5 intends to interact with Twitter's API.

- **Hypothesis**

  Through this project, team 5 intends to examine data pertaining to a popular Twitter user to understand the resounding sentiment of his/her tweets and understand the user's influence on his/her social network. To understand the sentiment, team 5 will use attempt to use sentiment analysis tools included in the NLTK library. To capture the user's influence, the team will use the built-in functionality of NetworkX.

- **Deliverables**

  Team 5 intends to deliver our findings in the form of .ipynb notebook and video presentation.

- **Workflow**
  1) Data preparation
  2) NLP / Network Analysis
  3) Evaluation
  4) Tuning and Optimization
  5) Finalize Model

- **Work Division**

Team 5 will distribute following workloads evenly and update our parts on the final notebook. Documentation, domain research, algorithm research and usage, finding basic stats and answers, research and experimentation of libraries, data preprocessing.

- **Christopher Estevez**
- **Meaghan Burke**
- **Rickidon Singh**
- **Ritesh Lohiya**
- **Rose Koh**

- **Resources**

  Library: NLTK, NetworkX

- **References**

  Interacting with Twitter's API: https://github.com/ptwobrussell/Mining-the-Social-Web-2nd-Edition/blob/master/ipynb/Chapter%201%20-%20Mining%20Twitter.ipynb

  Sentiment Analysis Library: https://www.nltk.org/_modules/nltk/sentiment/vader.html