Christina Litas

PLSC 497.001

3/17/2021

Research Proposal

1. What research question you want to answers? Be specific!
   1. Do people(students) with mental illness self-identified, tweet more about mental health awareness during the school year then people without?
   2. What I'm trying to get out of this question is, do people who self-identified as having a mental illness tweet more about how their illness is affecting the flow of the semester than people without a mental illness.
2. What are two possible answers?
   1. The first possible answer is, yes. Meaning that there will be an uptick in tweets regarding their mental health during the semester and how they are overcoming their struggles.
   2. The second possible answer is, no. Meaning that there is not an uptick in tweets regarding their mental health during the semester and what they are doing to overcome their struggles.
   3. The third, which could be a confounding variable could be influencing the outcomes of either the first or the second possible outcome.
3. What are the documents of text data you will analyze?
   1. The documents of texts that I will be using is Tweets from the site and application twitter.
   2. I plan on getting tweets from the time period 08/2019 to 05/2020 and 5/2020 to 08/2020 and 08/2020 to 5/2020. I am using Penn State academic school year as my template.
4. How will you get those documents? (<https://gwu-libraries.github.io/sfm-ui/posts/2017-09-14-twitter-data>)
   1. There are four ways that I will use to retrieve twitter data. The first way that I will get this data is through the uses of a twitter api. The first step will be to write a python(Tweepy) program pulling tweets form the past seven to nine days and go up to the limit of data retrieval of 3,200 tweets.
   2. The second way is finding and retriveving historical tweets already in a dataset from DocNow Catalog(<https://catalog.docnow.io>) and

TweetSets(<https://tweetsets.library.gwu.edu>)

* 1. I will purchase tweets from Twitter using Historical Power Track (<https://developer.twitter.com/en/docs/twitter-api/enterprise/historical-powertrack-api/overview> )

1. What methods will you use to answer this question?
   1. I believe that I want to use Trees and Random Forests to analysis the data. I can use this to predicted if a student in the past who is a self-identified person with some form of mental illness will tweet again if they received a positive feedback form their tweet. I will use the decision to split up the tweet based on the words: mental health, mental illness, grades, overcoming task, student, college, university, class, classes, and recovery.