Twitter Recommendations based on text

Create suggestions on potential people/pages/ads a user would like to follow based on their tweet/post content.

1. What are the names and NetIDs of all your team members? Who is the captain? The captain will have more administrative duties than team members.

   Rohan Khanna - rohank2 (Team Captain)
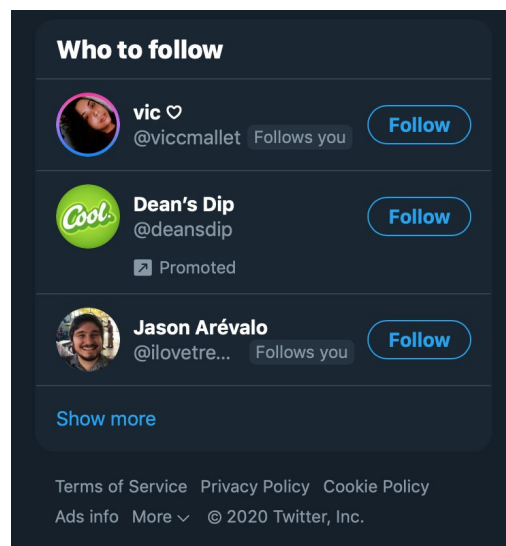   Cesia Bulnes - cbulnes2
   Tyler Wong - tylercw3

2. What is your free topic? Please give a detailed description.
   Our free topic is a recommendation algorithm that based on the text contents of tweets makes appropriate recommendations to the user. The recommendations can range from other tweets similar to a user's tweet, advertisements or even possible pages to follow. Consider a user who frequently tweets about a soccer club then our algorithm should recommend other followers of that club or advertisements specific to that club's products. Another example is if I frequently post about the new gaming console I should get recommendations to follow sellers such as gamestop,amazon etc on twitter as they are likely to be talking about console sale dates. These recommendation are all based on the text of a tweet.
   You may ask, doesn't Twitter have that already? Twitter's recommendation system currently is one page, where they recommend multiple sets of people who may be of different categories.

   Currently:

As you can see, the only relevant recommendation is the person who wrote it, or in many cases someone who was mentioned in the tweet. There is no recommendation of other people writing similar things. In the example below, you can see that Michelle obama

has a similar tweet, therefore she would be one of the top results returned since the ML/NLP algorithm recommends this content with a high % of similarity. We can then show the top tweets which can generate follows to different types of celebrities/businesses/ads etc.

Recommendation of Tweets:
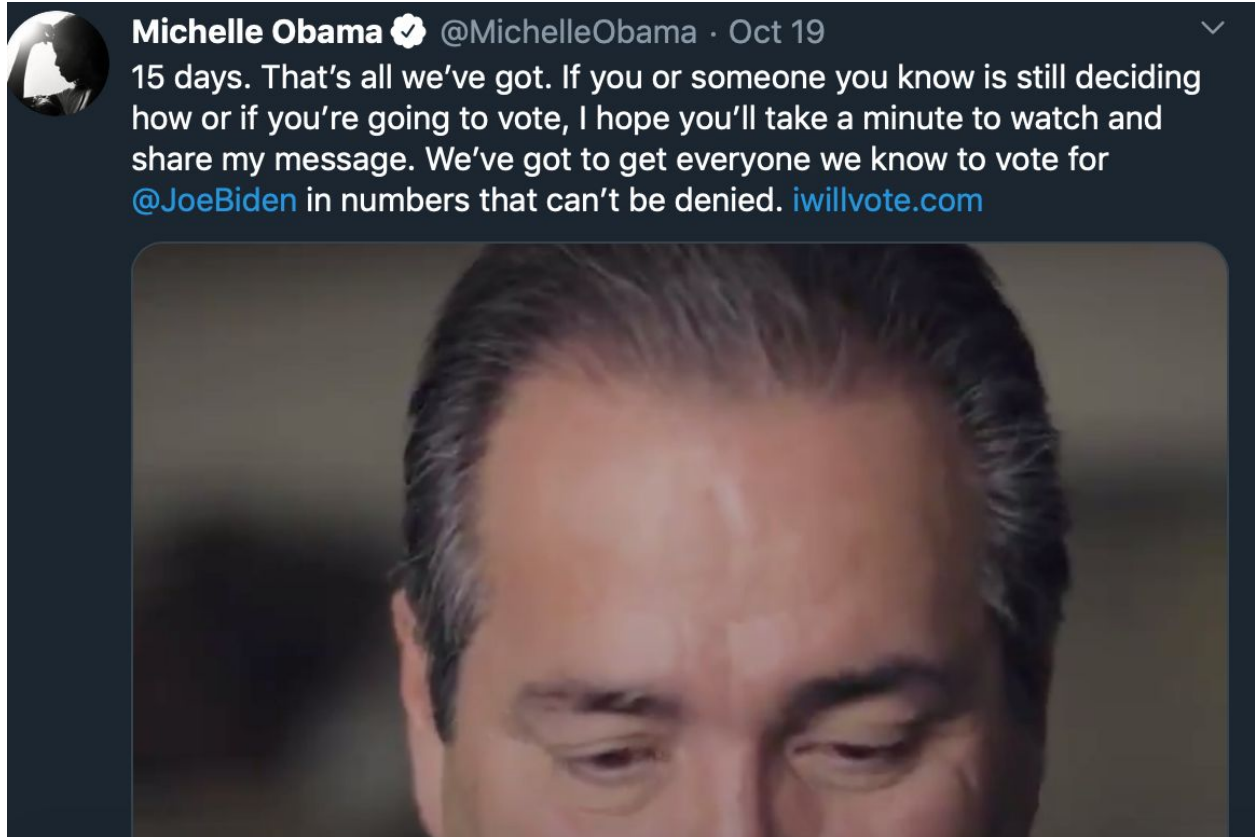
3. What is the task?

   The task is to deliver working software code, software documentation, and a software usage presentation about our topic described above. The code will fulfil our primary use case so that as someone makes a tweet, a recommendation will show up based on that singular tweet. In addition, we want to be able to give a percentage of how much a tweet is related to another page, tweet, etc.

4. Why is it important or interesting?

   This project is important because we want to portray similarities between a person and other people around the world. It's a way to unite people based on one singular tweet, and for people to be aware that they are not alone in terms of a specific subject. We are

focusing on social aspects this year because of controversial topics, such as politics(Trump/Biden), the Black lives matter and All lives matter movements, LGBTQ, children in ICE Detention centers and many more. We hope to examine a tweet that contains keywords, and recommend other personalities/pages to follow when that tweet is put in to grow a person's network.

5. What is your planned approach?

We plan on taking an iterative approach throughout this software project so that we can quickly identify blockers and make consistent progress. With that being said, we will meet weekly to discuss our current progress and any blockers that we're experiencing. We will also be splitting up the work so that we can all work in parallel. In addition, we are consistently talking through in a group chat where we can quickly get feedback on an idea or a feature.

We do have a defined due date, so we will use that date as a target to deliver a minimal viable product which will deliver the major functionality. We will add on more work to that minimal viable product if we underestimated our time or if we have more capacity than expected.

6. What tools, systems or datasets are involved?

The first system that comes to mind is using Twitter Developer, since we could filter real time tweets, and cross examine them to make recommendations based on people tweeting similar content. I think most of the work needed to recommend would be coming from Twitter Developer's API's. We would also be using Pycharm to code in python to show our results through the terminal.
We would use the nltk tool kits to remove stop words and appropriately tag the data.
To store our test data and queries we will provision a non relational database which are available from major companies such as Microsoft,Mongo,Amazon etc.

7. What is the expected outcome?

The expected outcome is for someone to make a tweet, and for them to have a recommendation immediately based on that singular tweet. Currently, Twitter has a recommendation page, but not recommendations based on tweets. People get passionate about topics and I feel like this feature would increase a person's usage of Twitter.

8. How are you going to evaluate your work?
As discussed in class, evaluation of a text recognition system depends on its usefulness to the end users. We plan to create a google form where users can rate how useful the results were for a particular tweet. We would look to answer the following questions through the google form all on a scale of 0-10 :

1. How accurate are the results provided by our tool ?
2. Do the results relate to the category of your tweet ?
3. How likely are you to use this tool in a production environment ?
4. How satisfied are you with the speed of the system ?
5. Does the person/page recommended reflect your interests based on a tweet?
6. Would you use this feature if Twitter enabled it?
7. Do you think it is invasive or inappropriate?

The next important criteria is the speed. Google delivers its search results in approximately 0.67 seconds. We aim at a minimum to have our results show within 2 seconds. Optimistically we will target results being shown within one second.

9. Which programming language do you plan to use?

Python, JS, React

10. Please justify that the workload of your topic is at least 20*N hours, N being the total number of students in your team. You may list the main tasks to be completed, and the estimated time cost for each task.

Task	Estimated Hours
Set up Twitter API and get familiarized	15 (5 hours per teammate)
Create Github repo and create base project components	2 hours
Setting up Nonrelational DB	2 hours
Create software usage tutorial presentation	3 hours
Develop software documentation	12 (4 hours per teammate)
Develop nlp code that can process tweets and categorize the contents	8 hours
Based on the categories of the contents of the tweet scrape data from the twitter apis relevant to that content.	10 hours
Program a ranking algorithm that gives : 1. Top users to follow, 2. Top posts the user may like 3. Top retail sites based on their tweet. (Ads)	24 hours
Create google form for feedback evaluation	1 hour
(Optional) create a web app to present all	10 hours

results. Include setting up api and frontend	
Total	77 excluding optional