

*check* <sup>(that tweet)</sup> *yo self* 

## **Prioritizing Fact-Checking During Disasters**

Alina Ingel, Brittney Hosac, and Stephanie Caress  
May 2020



# What's the deal with “misinformation”?

- During times of crisis, people seek information to make the best possible response for their safety. **Relying on misinformation can put people in more danger** [[Red Cross](#)]
- During interviews with [FEMA](#), [CEDR](#), and [Florida Search & Rescue](#), we heard the same thing: **There's no getting around manual fact-checking**
- In 2018, MIT found falsehoods are 70% more likely to be retweeted and is mostly spread by humans... not bots! [[MIT](#)]



# Problem Statement

For emergency response agencies who spend a lot of time manually verifying information:

**How do we prioritize fact-checking tweets in order catch inaccuracies before they take off?**



# Process

1. Gather tweets via  
TwitterScrape about disaster  
(COVID-19)

2. Clean and explore the  
tweets gathered

3. Create a model that reflects  
the “reach” of a tweet

4. Prioritize tweets with the  
most potential for  
fact-checking efforts



# Roadblocks

1. **Gathering:** TwitterScrape requires two separate requests to get tweet AND user information. This makes it very slow!
2. **Volume:** There are 500 million tweets a day, it's a challenge to get the “complete” conversation
3. **Bias:** It was important to our group that we remove ourselves from the equation.



# Modeling Process

Since misinformation spreads more than truth, we set out to create a model that predicts the “reach” of a tweet. Tweets with largest potential audience are the priority for fact-checking.

- Supervised: Predicted total # of Retweets, Replies, Favorites (Target)
  - Linear Regression, Extra Trees, Bagging Regressor
  - Too much noise, too many outliers to train an accurate model
- Unsupervised: Grouping Users to see similarities in Target
  - KMeans and DBSCAN
  - **Final model: DBSCAN**
    - **5 clusters (excluding outliers) / .76 silhouette score**



# The Topic of Conversation

## Coronavirus Press Briefing, April 23rd 2020

“I see the disinfectant that knocks it out in a minute, one minute,” Trump said during Thursday’s [coronavirus press briefing](#). “And is there a way we can do something like that by injection inside, or almost a cleaning? Because you see it gets inside the lungs and it does a tremendous number on the lungs, so it would be interesting to check that.” - [Washington Post](#)

Top Words	Top Groups of Words
people	stop panic end total
trump	data stop panic end
just	way like injection inside
like	like injection inside cleaning
said	minute way like injection
drink	knocks minute minute way
know	minute minute way like
virus	cure wrote trump week
president	peddling cure wrote trump
think	group peddling cure wrote



# Common Words/Phrases

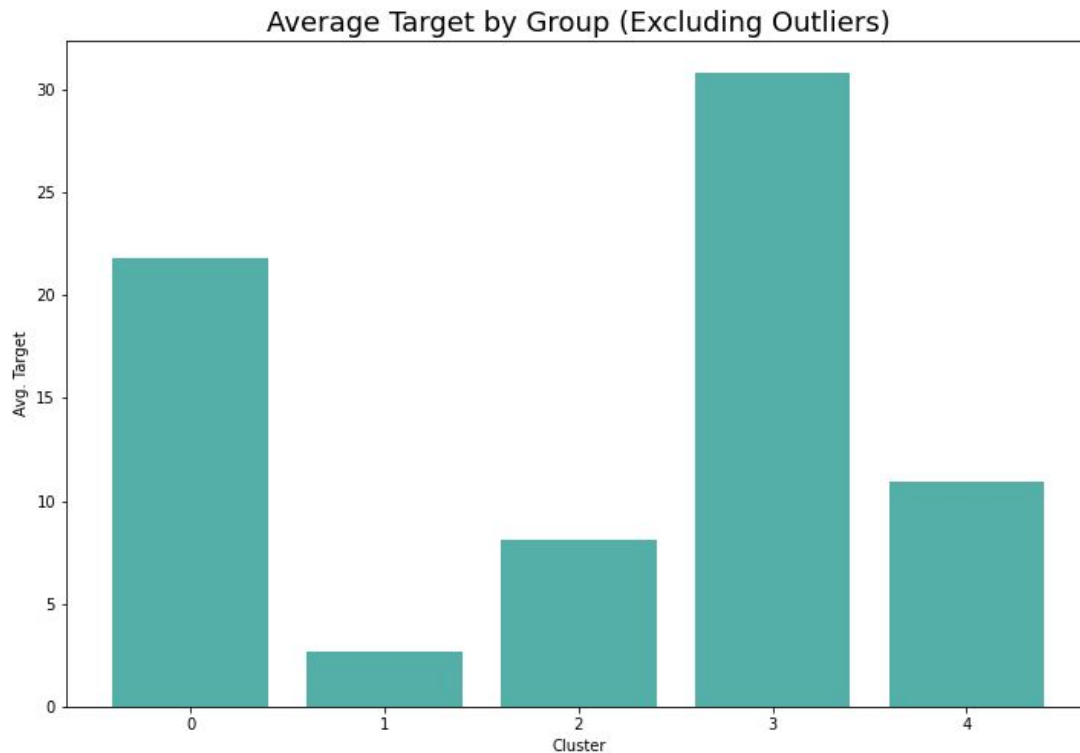
Because public conversation is centered around coronavirus and disinfectants, NLP is a weak feature to build predictive modeling on.

Outlier	1st Priority	2nd Priority	3rd Priority	4th Priority	5th Priority
trump	people	people	people	people	people
people	trump	trump	trump	trump	trump
just	just	just	just	just	just
like	like	like	like	like	said
virus	time	said	drink	virus	like

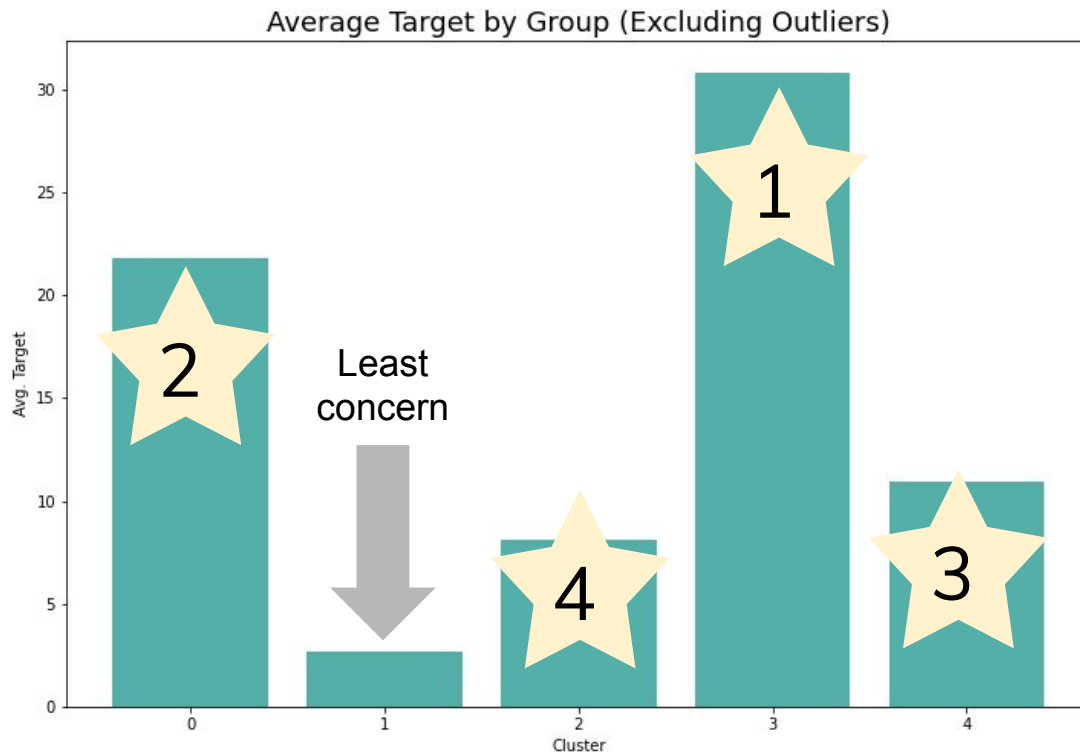
Outlier	1st Priority	2nd Priority	3rd Priority	4th Priority	5th Priority
stop panic end total	stop panic end total	like injection inside cleaning	way like injection inside	group peddling cure wrote	stop panic end total
data stop panic end	data stop panic end	way like injection inside	like injection inside cleaning	people actually exactly caution	way like injection inside
cure wrote trump week	cure wrote trump week	minute way like injection	minute way like injection	meant people actually exactly	data stop panic end
group peddling cure wrote	group peddling cure wrote	stop panic end total	minute minute way like	leader group peddling cure	like injection inside cleaning
leader group peddling cure	need personal mug check	data stop panic end	knocks minute minute way	hot cups hot coffee	minute way like injection



# Cluster Profiles - Avg. Target (Reach) by Group

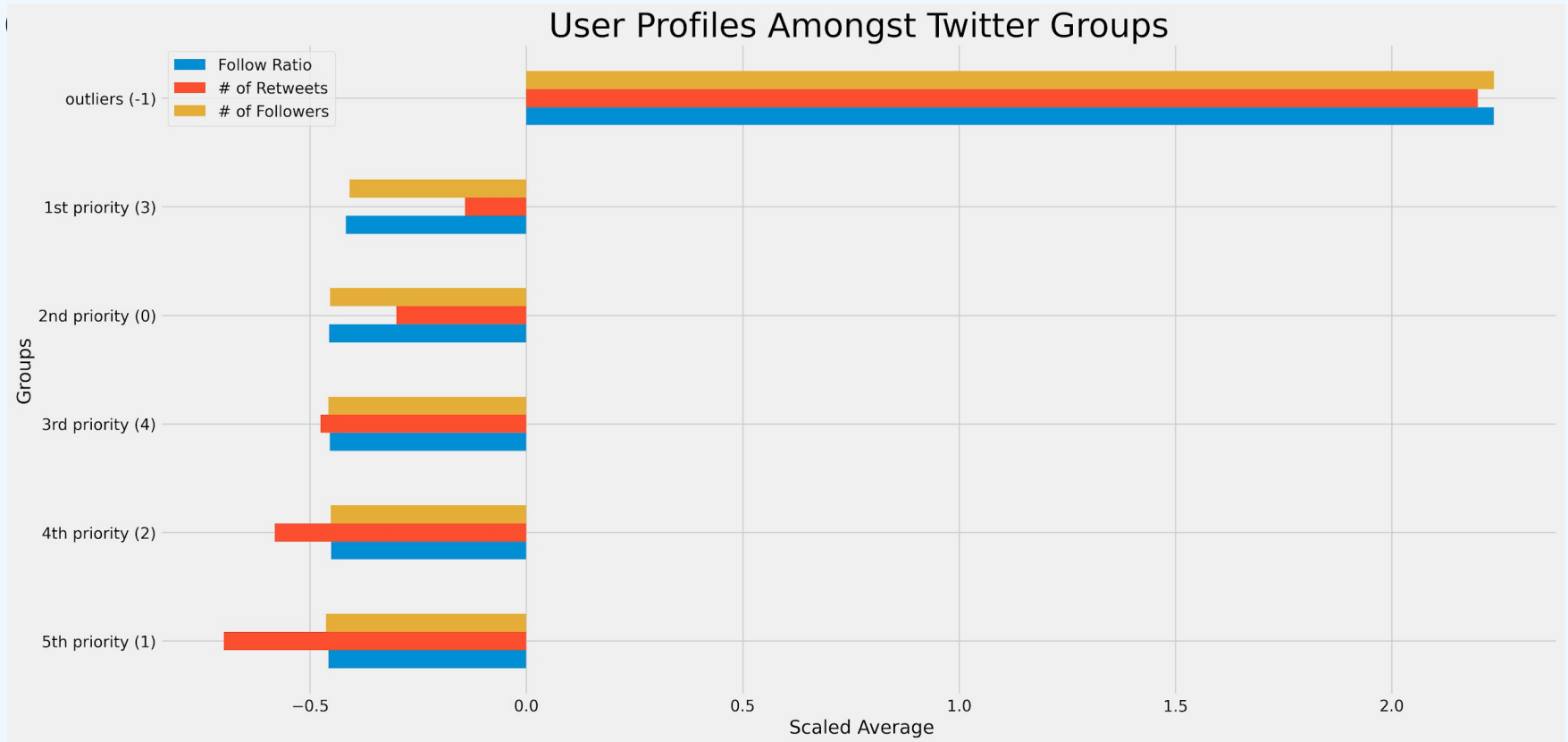


# Cluster Profiles - Avg. Target (Reach) by Group

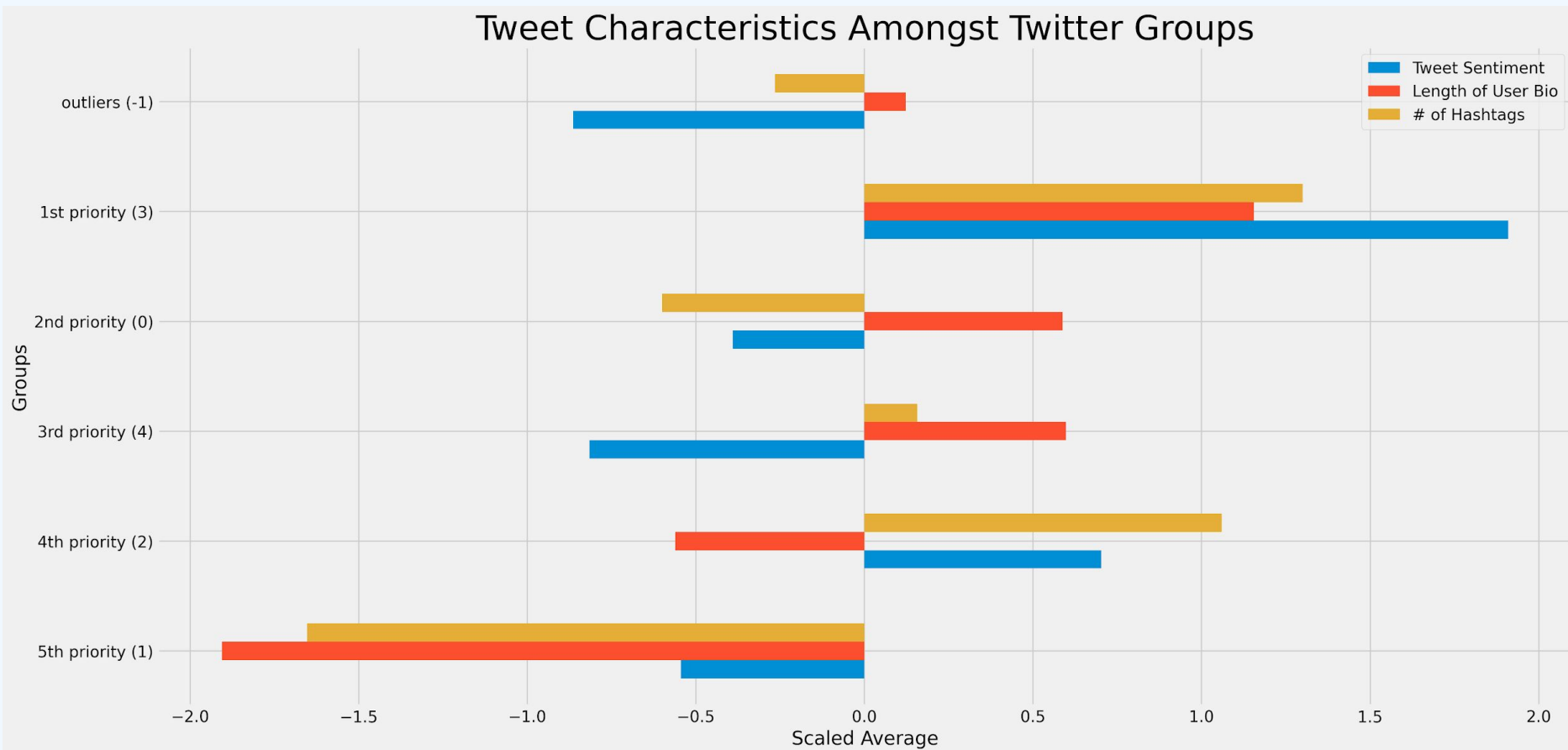




# Cluster Profiles - Scaled Avg. for User Info



# Cluster Profiles - Scaled Avg. for Tweet Info



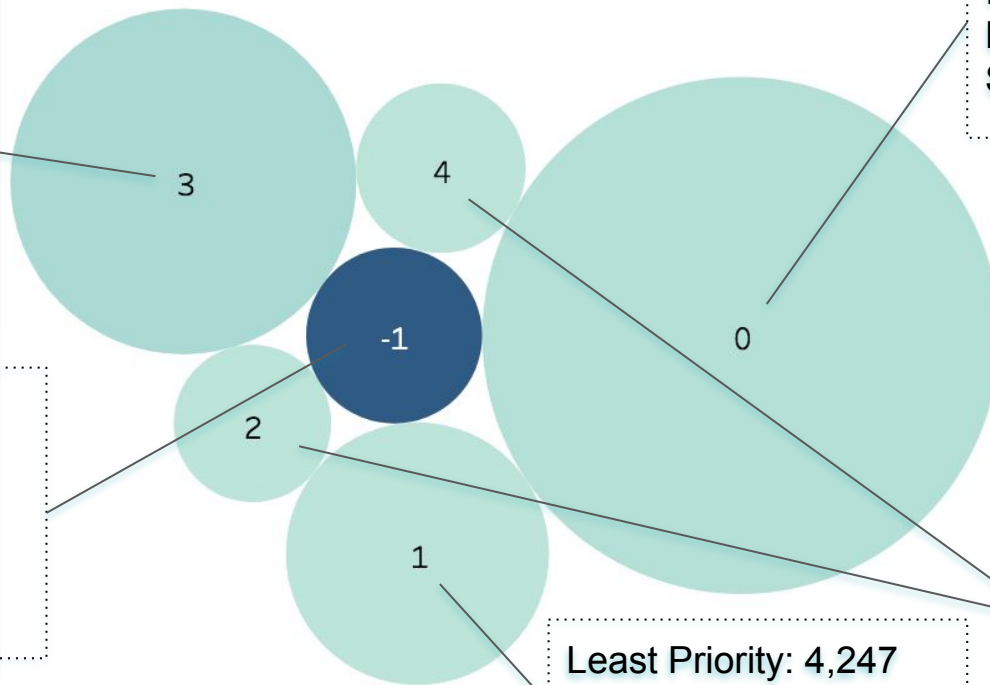
# Cluster Profiles

## DBSCAN Clusters



1st Priority: 7,342  
LAist  
Foreign Affairs  
TeamCanada  
500Startups  
Local Public Figures  
Potential high impact

Outliers: 1,903  
CNN  
EmmaWatson  
WSJ  
AP  
Media



2nd Priority: 16,418  
Individual Accts  
Everyday Users  
Some Public Figures

Least Priority: 4,247  
Individual Accts with  
limited reach

3rd/4th Priority:  
1,768 / 1,521  
Individual Accts  
with potential for  
wider reach

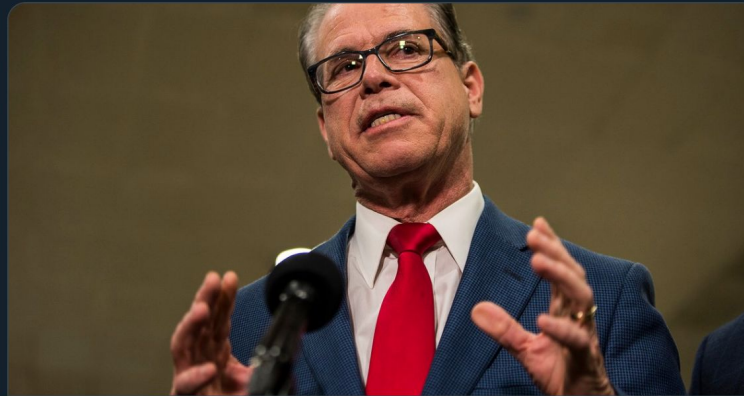
# Sample Tweets

Outlier (-1)



CNN  
@CNN

State leaders in Indiana are telling residents not to ingest disinfectants amid coronavirus concerns



Indiana state leaders tell residents not to ingest disinfectants

Indiana Department of Health Commissioner Kristina Box acknowledged that she has received questions about the wisdom of ingesting disinfectant and has told these ...

[cnn.com](https://www.cnn.com)

2:26 PM · Apr 24, 2020 · [SocialFlow](#)

148 Retweets 507 Likes

# Sample Tweets

1st Priority (3)



TinkerTweets45 ★★ ★  
@Tinktweets45

He's a genius.

"An international chorus of doctors and health experts has urged people not to drink or inject disinfectant"

Common vaccine ingredients:

Thimerosal - DISINFECTANT

Formaldehyde - DISINFECTANT

POTUS led Media and Doctors to admit that the vaccines are not safe.



4:14 PM · Apr 25, 2020 · [Twitter for iPhone](#)

61 Retweets 94 Likes

# Sample Tweets

2nd Priority (0)



Patrick Howley  
@HowleyReporter

WOW: In 2019, before Coronavirus outbreak, the American Journal of Epidemiology found that social isolation makes black people more than TWICE as likely to die from any cause.

Have Fauci and Birx seen this???



PRE-CORONAVIRUS STUDY: Social Isolation Makes Black People More Likely To D...  
Research conducted shortly before the Coronavirus outbreak shows that social isolation dramatically increases the risk of death from any cause, particularly ...  
[nationalfile.com](#)

10:51 AM · Apr 25, 2020 · [Twitter Web App](#)

34 Retweets 41 Likes

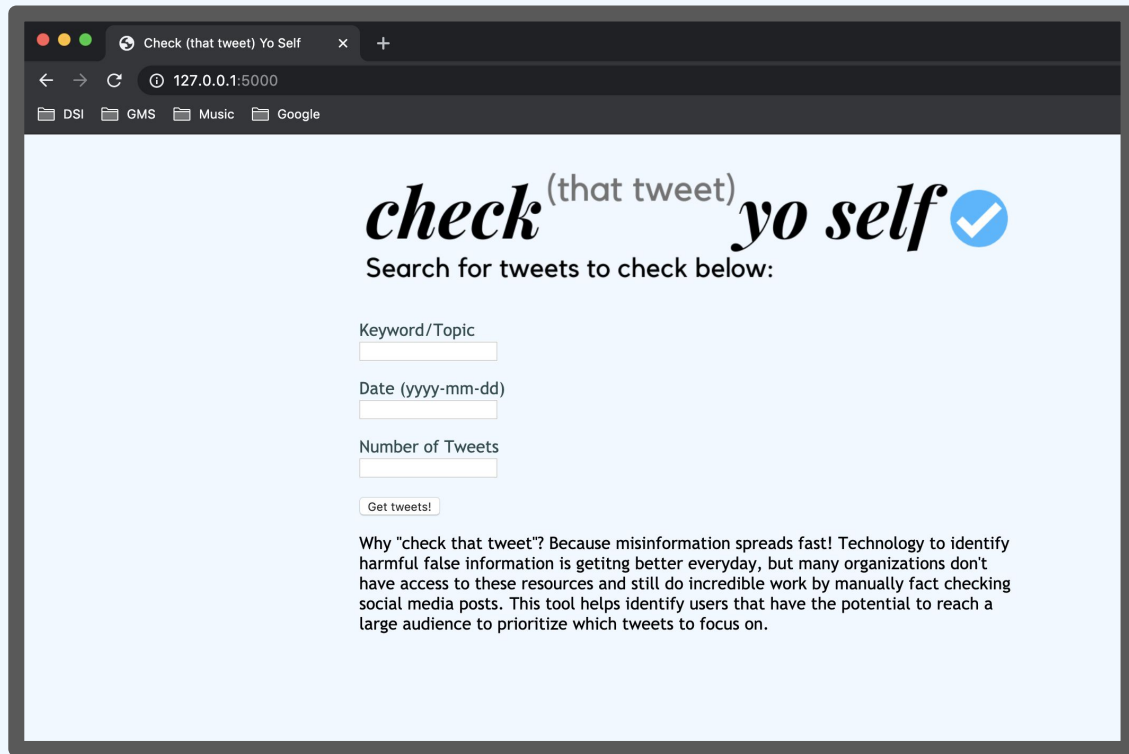


Our model in action...

*check* <sup>(that tweet)</sup> *yo self* 

# How it Works

1. Users input a topic, date, and number of tweets
2. The tweets are pulled and categorized based on user info
3. Results of the search returned, noting which tweets to focus in



Note: DBSCAN is not built to predict. The model behind the app is K Nearest Neighbors, trained on the clusters with 99% train and test accuracy.

# check<sup>(that tweet)</sup> yo self

Tweet results below:



	Time	Author	Text	Reply Count	Favorite Count	Retweet Count	Tweet Url	User Followers	User Following	Groups	To Check
1	2020-05-12 16:59:59	FireRokita	Hold Trump accountable for blowing up our global pandemic response team, and for believing Xi and touting his transparency for months. That would be more useful than Lindsey's bill, and all the GOP xenophobia you're parroting.	1	13	2	<a href="https://twitter.com/FireRokita/status/1260359096596865026">https://twitter.com/FireRokita/status/1260359096596865026</a>	328	427	2	Check this tweet!
2	2020-05-12 16:59:52	jcvolt	If it were a real pandemic there would be no need for this.	0	0	0	<a href="https://twitter.com/jcvolt/status/1260359065575788544">https://twitter.com/jcvolt/status/1260359065575788544</a>	5242	4754	2	Check this tweet!
3	2020-05-12 16:59:52	RaquelEvita	"Ramadan has never been empty bellies, but about full hearts." Imam Khalid Latif on the revelations available to us this #Ramadan in pandemic <a href="https://twitter.com/BkwBFE8v5i">pic.twitter.com/BkwBFE8v5i</a>	1	5	0	<a href="https://twitter.com/RaquelEvita/status/1260359066121052160">https://twitter.com/RaquelEvita/status/1260359066121052160</a>	18646	1883	2	Check this tweet!
4	2020-05-12 16:59:52	suburbanwater	Suburban is proud to announce that we have donated a total of \$3,000 to Meals On Wheels La Mirada, Meals on Wheels Industry, and Tzu Chi USA who are providing food and other essentials to our most vulnerable in our communities during this pandemic. #SuburbanCarespic <a href="https://twitter.com/MTA04N8NgH">twitter.com/MTA04N8NgH</a>	0	3	1	<a href="https://twitter.com/suburbanwater/status/1260359066175459329">https://twitter.com/suburbanwater/status/1260359066175459329</a>	20	56	2	Check this tweet!
5	2020-05-12 16:59:52	linda_hazlett	"Tam said that before allowing more people to cross Canada's international border, lawmakers must, among other things, keep track of where other countries are in their fight against the Covid-19 pandemic." <a href="https://www.politico.com/news/2020/05/12/canada-us-border-reopening-251965">https://www.politico.com/news/2020/05/12/canada-us-border-reopening-251965</a> ...	0	0	0	<a href="https://twitter.com/linda_hazlett/status/1260359067496833026">https://twitter.com/linda_hazlett/status/1260359067496833026</a>	119	75	2	Check this tweet!
6	2020-05-12 16:59:54	jikemannino	Ed Markey needs to be supported this election cycle. Please make sure to support him against his more conservative opponent, A Fucking Kennedy	2	6	2	<a href="https://twitter.com/jikemannino/status/1260359074354475009">https://twitter.com/jikemannino/status/1260359074354475009</a>	256	174	2	Check this tweet!
7	2020-05-12 16:59:54	Sydneyquid	.@senatemaJldr your leader has been in office for 3 years, he could have not trashed the pandemic group Moscow Mitch. He could have started action as soon as he heard of the virus. He didn't.	0	0	0	<a href="https://twitter.com/Sydneyquid/status/1260359076271325184">https://twitter.com/Sydneyquid/status/1260359076271325184</a>	64	5	2	Check this tweet!
8	2020-05-12 16:59:55	WestIslandProud	We do. But the regions have decided they hate Montreal. Not a great way to get through a pandemic together.	0	0	0	<a href="https://twitter.com/WestIslandProud/status/1260359077286289410">https://twitter.com/WestIslandProud/status/1260359077286289410</a>	96	694	2	Check this tweet!
9	2020-05-12 16:59:55	AttorneyJosh	"Got that pandemic!"	0	0	0	<a href="https://twitter.com/AttorneyJosh/status/1260359078171348994">https://twitter.com/AttorneyJosh/status/1260359078171348994</a>	27	294	2	Check this tweet!

# *check*<sup>(that tweet)</sup>*yo self* ✓

Search for tweets to check below:

Keyword/Topic

Date (yyyy-mm-dd)

Number of Tweets

Get tweets!

Why "check that tweet"? Because misinformation spreads fast! Technology to identify harmful false information is getting better everyday, but many organizations don't have access to these resources and still do incredible work by manually fact checking social media posts. This tool helps identify users that have the potential to reach a large audience to prioritize which tweets to focus on.



# Next Steps

- Additional cluster modeling to strengthen tweet categorization and refine the output
- Pulling a larger dataset to get a more representative look at the broader conversation
- More interviews with organizations, how can this tool directly address their challenges
- Adding thorough descriptions of the groups to the app webpage

When in doubt...

*check* (that tweet) *yo self* 