Leveraging Social Media to Map Disasters

Kathryn Shupe
Michael Laccavole
Patrick McCaul

General Assembly
DSI 11 Boston

https://www.vox.com/the-highlight/2019/10/16/20908291/camp-fire-wildfire-california-paradise-survivors

Data Science Problem

How can social media posts be used to ***identify*** and ***locate*** people in urgent need of help during a disaster?

Expanding Existing Relationships

Social Media Partnerships

Facebook:

UNICEF

International Red Cross,

The World Food Program

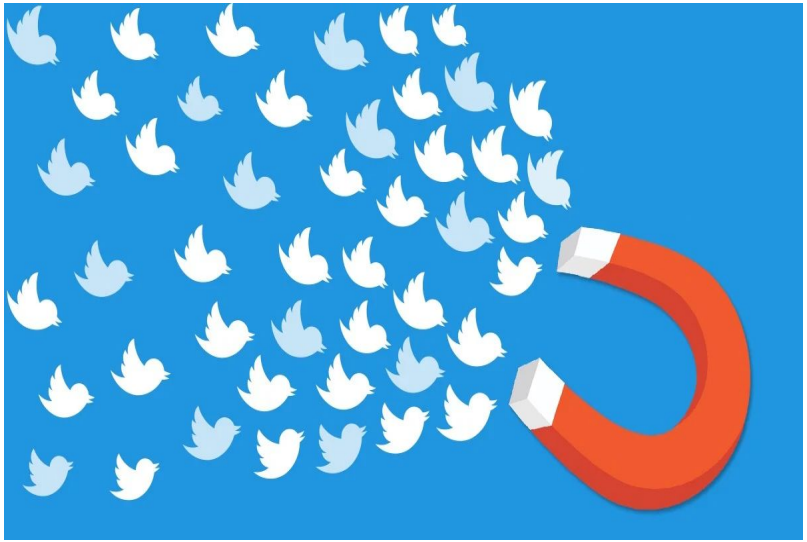
Real time data

Location & Movement

Dynamic tracking and mapping of those in need

Twitter

- Large user base
- Usage during disaster events
- A platform that can offer critical advantages to first responders searching disaster areas for those in need.



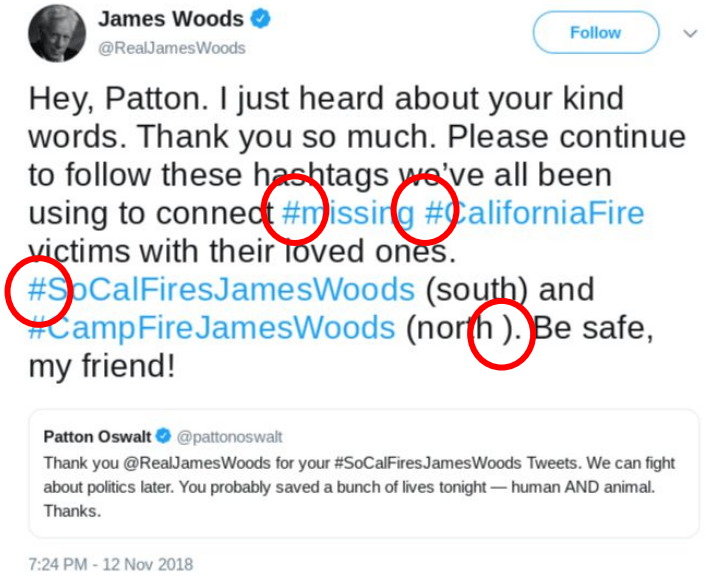
Generating Tweets

Web scraping

Query
Customization

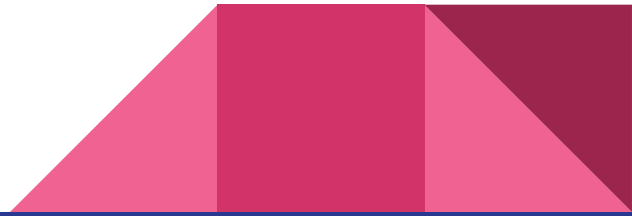
Cleaning

Removing
Unnecessary
information

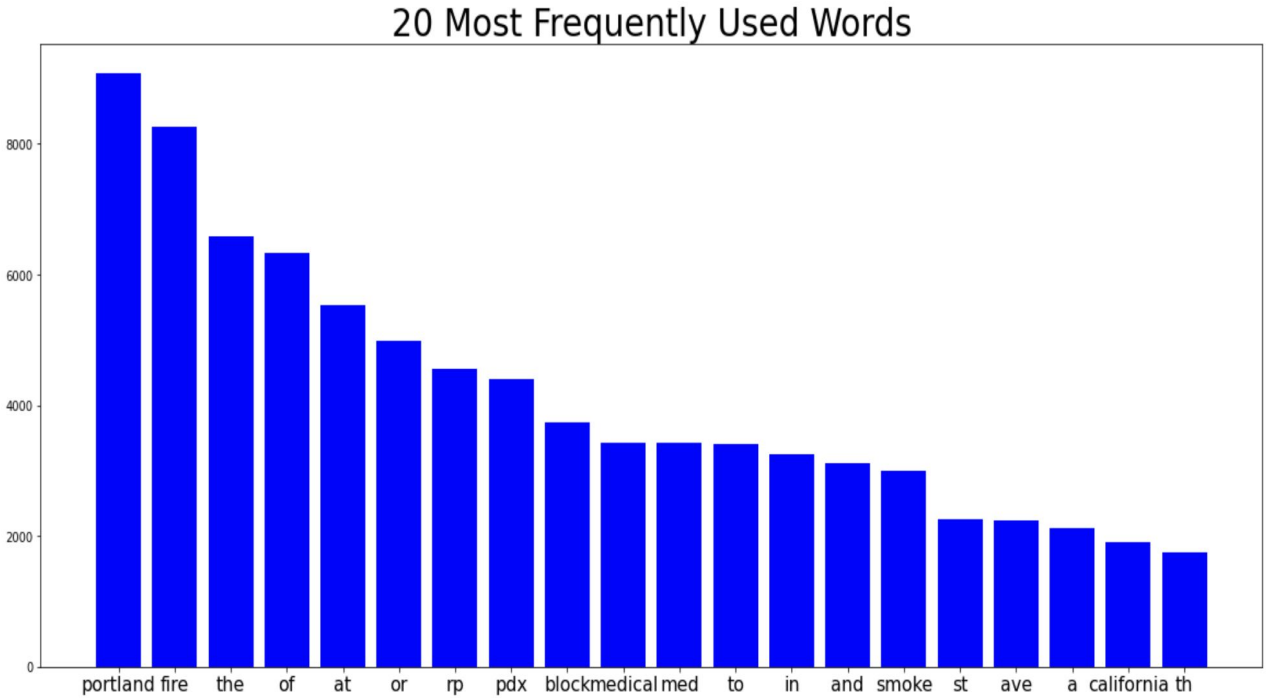


Next Steps: Prep for EDA and Modeling

- Tokenize the tweets for further processing
- Look at the words in the corpus: what are some of the most commonly used words?
- Lemmatize and Stem the tokens

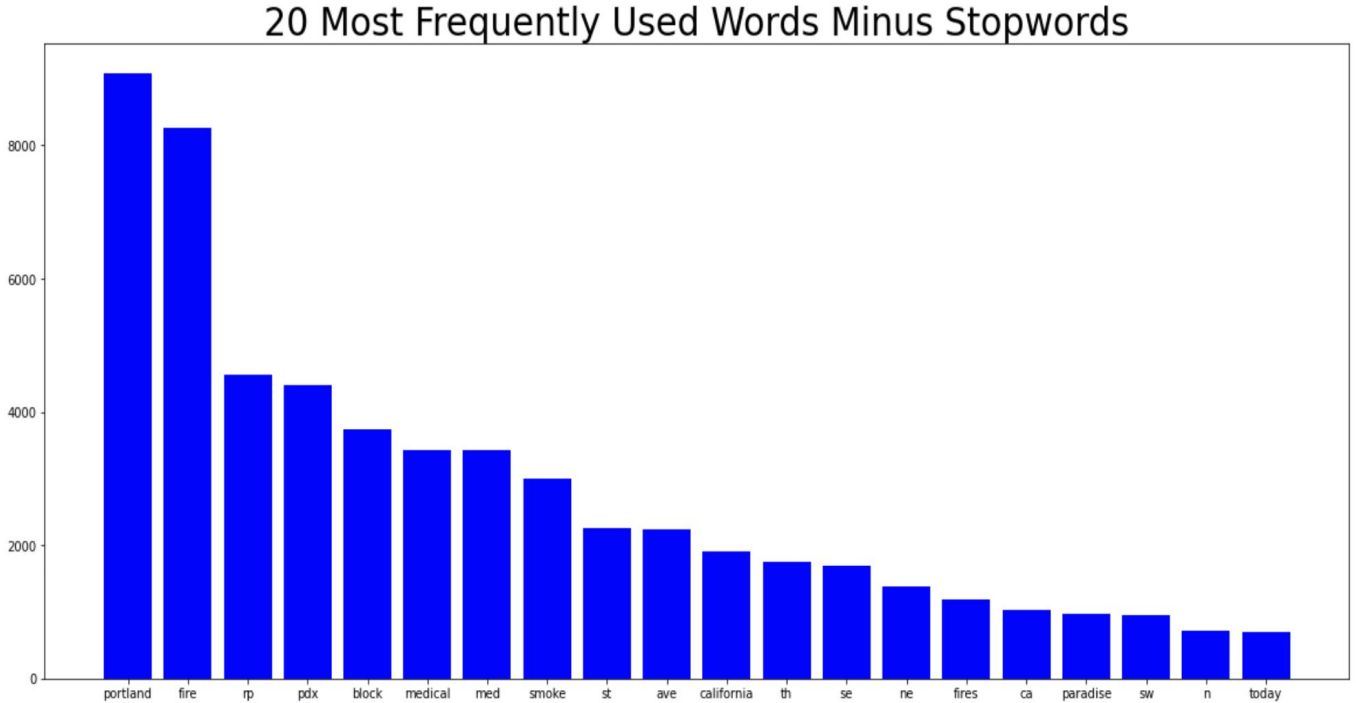


Most frequently-appearing words in tweets



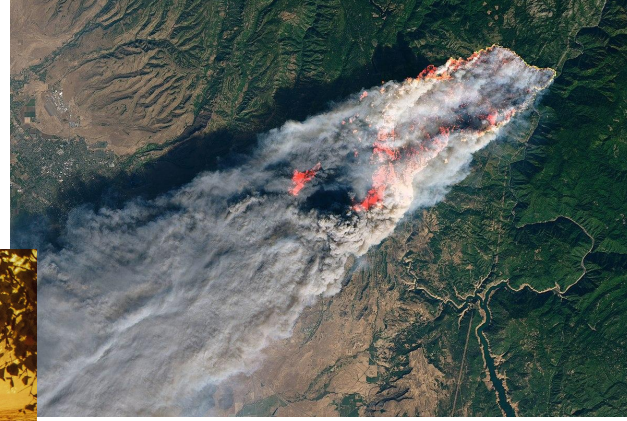
words	word_frequency	word_pct
portland	9081	0.039662
fire	8269	0.036116
the	6583	0.028752
of	6328	0.027638
at	5522	0.024118
or	4990	0.021794
rp	4557	0.019903
pdx	4405	0.019239
block	3733	0.016304
medical	3432	0.014990
med	3427	0.014968
to	3414	0.014911
in	3248	0.014186
and	3114	0.013601
smoke	2989	0.013055

Accounting for stop words



words	word_frequency	word_pct
portland	9081	0.057370
fire	8269	0.052240
rp	4557	0.028789
pdx	4405	0.027829
block	3733	0.023584
medical	3432	0.021682
med	3427	0.021650
smoke	2989	0.018883
st	2249	0.014208
ave	2236	0.014126
california	1910	0.012067
th	1743	0.011012
se	1701	0.010746
ne	1374	0.008680
fires	1188	0.007505
ca	1031	0.006513
paradise	961	0.006071
sw	944	0.005964
n	712	0.004498
today	703	0.004441

Camp Fire - Paradise, CA 2018



- Deadliest and most destructive wildfire in CA's history
- 153,336 acres
- At least 85 civilian fatalities
- Destroyed 18,804 structures

https://www.census.gov/topics/preparedness/events/wildfires/camp.html

How do we classify a tweet containing "more urgent" content for emergency responders?

Step 1: Begin to generate words for Bag of Words approach

> Logistic Regression on an open source Disaster Response Message database: A set of messages related to disaster response, covering multiple languages, suitable for text categorization and related natural language processing tasks.

*is this message a direct call for help?

https://appen.com/datasets/combined-disaster-response-data/

Most correlated to Urgent*:

	word	coef	e^coef
15732	need	3.975140	53.257554
24916	us	3.909328	49.865451
8736	food	3.879859	48.417376
10393	help	3.861895	47.555405
10820	hungry	3.852896	47.129349
540	aid	3.464018	31.945077
23387	tents	3.457433	31.735403
23380	tent	3.442254	31.257317
20494	sandy	3.167935	23.758382
25568	water	3.164311	23.672426
10712	house	3.091979	22.020605
6989	dying	3.000751	20.100621
22289	starving	2.673388	14.488977
10818	hunger	2.633808	13.926696
1261	area	2.572726	13.101496
18957	rain	2.520618	12.436275
25187	victim	2.444524	11.525062
19217	received	2.400962	11.033786

Least correlated to Urgent*:

	word	coef	e^coef
12175	job	-3.166459	0.042153
11438	information	-2.393985	0.091265
20379	said	-2.215541	0.109094
11278	including	-1.900485	0.149496
11694	international	-1.862094	0.155347
16093	notes	-1.743793	0.174856
15844	news	-1.737536	0.175953
9632	government	-1.717573	0.179501
11287	incomplete	-1.700484	0.182595
5128	country	-1.635778	0.194801
23622	thousands	-1.527018	0.217182
19659	reported	-1.483239	0.226902
11439	informations	-1.386566	0.249932
14338	materials	-1.363976	0.255642
25992	work	-1.361786	0.256203
26233	year	-1.291173	0.274948
14819	million	-1.268201	0.281337
17177	passport	-1.240541	0.289228

Step 2: Apply industry knowledge to generate additional terms to add to Bag of Words vectors

Public Assistance Work can be defined as **1)Emergency** or **2)Permanent**

Emergency Work
Category A: Debris removal
Category B: Emergency protective measures

Permanent Work
Category C: Roads and bridges
Category D: Water control facilities
Category E: Public buildings and contents
Category F: Public utilities
Category G: Parks, recreational, and other facilities

Federal funding guidelines for each of these categories are listed in the *Public Assistance Program and Policy Guide*, which is located online at fema.gov/public-assistance-policy-and-guidance.

FEMA

Public Assistance Fact Sheet:
https://www.fema.gov/media-library-data/1572028370885-e5aabd18bf3145151a89919f82500613/PublicAssistanceFactSheetJune2017_Updated_Oct2019(FINAL).pdf

Bag of Words: Urgent vs. Non-Urgent

Urgent: help, tonight, today, fire, need, aid, removal, burn, tree, evacuation, smoke, unsafe, search, lost, victim, medical, med, urgent, home, haze, serious, fires, reporting, smoking, come, emergency, wildfire, send, valley, investigation, structure

Non-Urgent: job, news, government, country, materials, work, price, utilities, facility, irrosion, irrigation, restoration, shoulder, mud, silt, ditch, slip, pray, hope, thanks, thankful

Model:  word2vec

~Utilized a ¨pre-trained google news dataset (~100 billion words). The model contains 300-dimensional vectors for 3 million words and phrases"

https://code.google.com/archive/p/word2vec/

```
df['urgent'].value_counts(normalize=True)
```

```
1    0.64927
0    0.35073
Name: urgent, dtype: float64
```

64.93% Urgent Tweets
35.07 % Non-Urgent Tweets

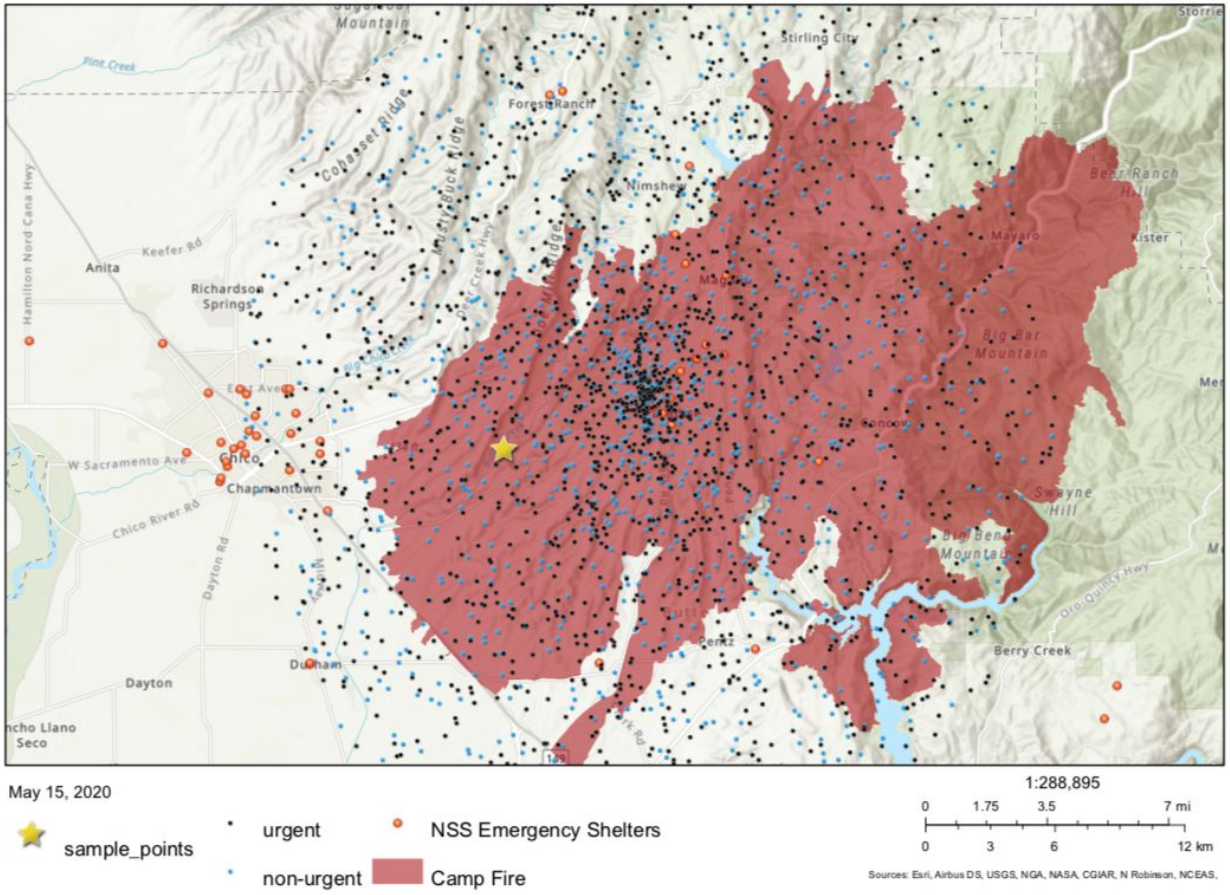
Urgent Tweets:

- 'Camp fire closure in #Oroville on Hwy 70 Both NB/SB between Pentz Rd and CA 89 #traffic http://bit.ly/Y31VAF\xa0'
- '!! sigalert !! the road is closed because of a brush fire. in #Grapevine on I-5 NB at Smokey Bear Rd, stopped traffic back to Templin Hwy'
- 'Really bad. So bad I called it in to the NWS (I'm a trained skywarn spotter) and they issued a Dense Smoke Advisory. @ Roseville, California https://www.instagram.com/p/BqBMDGjl2Njf9eO_zgiTRXZ7vkg61XF0PUxgeI0/?utm_source=ig_twitter_share&igshid=3vujyep2jbr1\xa0…'

Non-Urgent Tweets:

- 'An eerily hazy view down Market St at 2pm this Friday from the smoke and poor air quality due to the Wildfires burning 150+ miles north of the Bay Area.\n•\n#Smoke #Haze #CampFire #Wildfire'
- 'My morning weather report. Smoke? Wind has blown smoke from the horrible fires in Malibu, Thousand Oaks, Agoura Hills all over Los Angeles. Pray everyone…'
- 'My birthday gift redeemed, a visit with the Giant Sequoia survivors of Fire. #tbt #DaughteroftheSouth #forests #conservation #sentientbeings @ Sequoia & Kings Canyon National Parks

Camp Fire Tweets



https://kathrynshupe.maps.arcgis.com/apps/View/index.html?appid=5f59534f38a94081bd2d189806947915

https://www.usatoday.com/story/money/2018/11/14/california-wildfires-pge-pacific-gas-electric-camp-fire/1999338002/

In Summary

- We understand that social media is a valuable resource for citizens during emergency situations
- How can we utilize this resource to gain meaningful information to map areas/individuals in need during disaster response situations?
  - ArcGIS web app to map urgent vs. non-urgent tweets and attach their associated image/media for analysis by emergency responders
- Next Steps?
  - Establish partnerships with Twitter that are similar to that of facebook to access geolocation
  - Applying industry knowledge to generate more effective bag of words




Thank you

&

Questions?