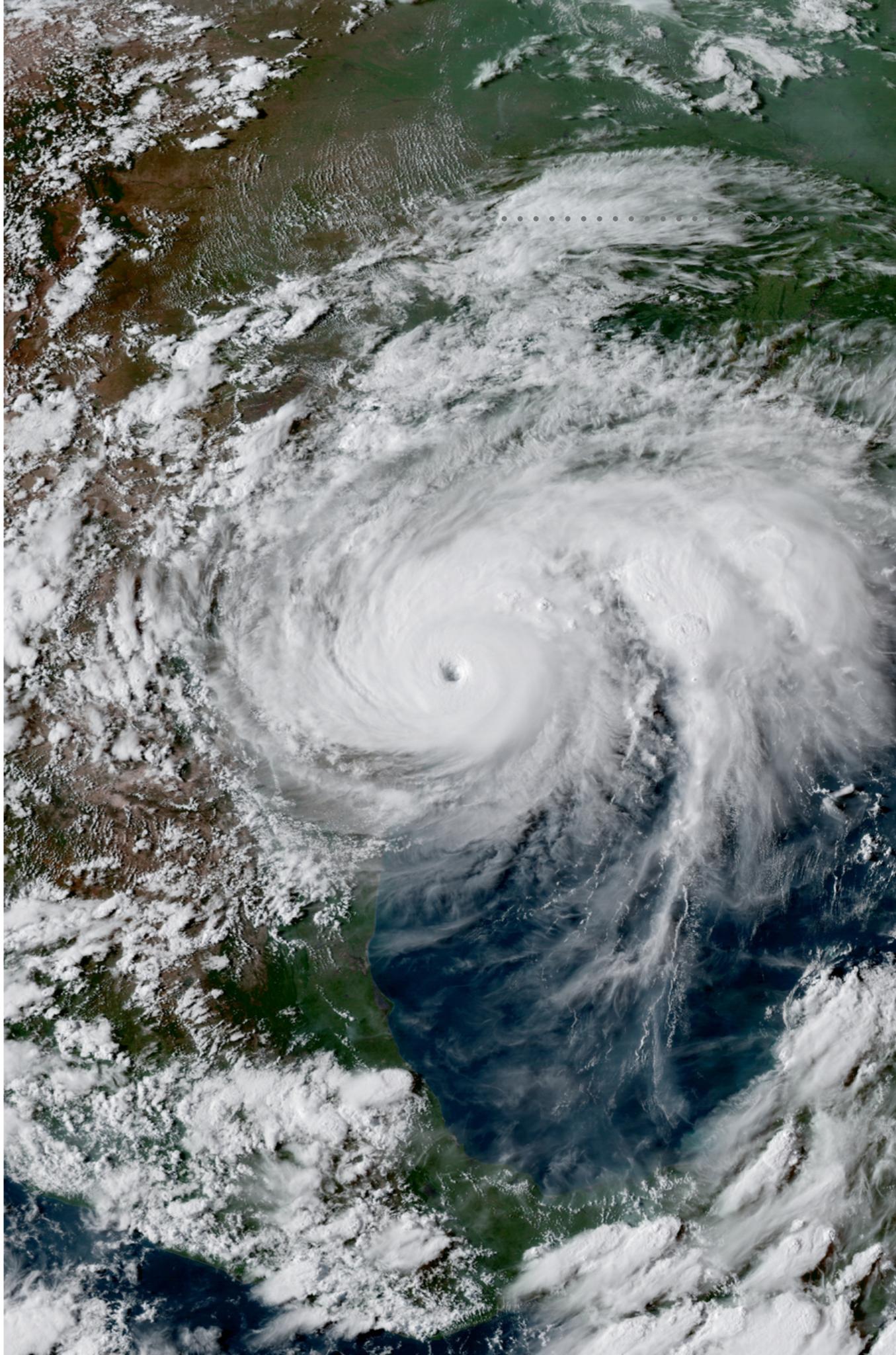


SOCIAL MEDIA EMERGENCY LANGUAGE PROCESSOR: **#DSI9PROJECT5**

Travis Darby, Doug Frey, Nina Lacson

HURRICANE HARVEY

- Category 4 hurricane
 - Texas and Louisiana
 - Landfall: Aug 25 - 29, 2017 (5 days)
 - \$190 billion damages
- Aftermath and casualties
 - 13 million people affected (TX, LA, MS, TN, KY)
 - ~40,000 displaced victims (68 deaths)
 - Heavy rainfall and flooding (max 60 inches)
 - Power outages (stranded)
- “First major natural disaster of the social media age.” - TIME



NOTICE FROM U.S. COAST GUARD



To report a [#harvey](#) emergency you must call numbers below or 911 for assistance. If busy keep trying. Do not report distress on social media

The U.S. Coast Guard is conducting urban search and rescue in the city of Houston. If you are in need of rescue, call 911 or the U.S. Coast Guard Houston Command Center. Do not report your information on social media sites.

Sector Houston Command Center numbers:

281-464-4851

281-464-4852

281-464-4853

281-464-4854

281-464-4855

Stay calm, do not panic.

Do not go into the attic, rescuers from the air cannot see you.

Get to high ground immediately.

Mark the roof to be seen from the air. Wave sheets, towels, to be noticed from the air.

3,020 6:04 PM - Aug 27, 2017

- Lack of resources provided by the government
 - 6.6 million residents
 - Cannot meet demand
- Users turn to social media for aid
 - Tens of thousands called 911



USER RESPONSES IN TWITTER

Sparklepuff  @Sparklepuff2 · 11h
Replying to @USCG @Boba_Cup
Why doesn't @USCG monitor their Twitter for emergencies instead of tweeting that people cannot use it?
911 is busy & people can't recharge!

Roxy Rowins @roxyrowins · 9h
Agreed! They should be reading out tweets to help get people rescued in Texas.

* NightShiftNews!® *  @NightShiftNews · 6h
They need online logistics coordinating with incoming timeline with address areas.

Roxy Rowins @roxyrowins · 9h
Replying to @USCG
There are people trapped in a house with flood waters. Please help. Here is the address: 8507 Bluebird Rd. Houston, Texas.

Iua james @poptivist · 9h
just letting you know your info has been added to the #harveyrescue spreadsheet

➤ Volunteering efforts as a community



USER RESPONSES IN TWITTER

► Rescue & relief



indigosky42 @indigosky42 · 13h

Replying to @USCG

I will do the same. I know you don't have cell power. I'm actively helping others.
@uscg you rock, but some folks can't stay on dying phones

3 11 17



Laura Soto Salazar @lalasoto · 13h

S.O.S. Mentally disabled senior & wife stuck 2nd floor: 5735 South Braeswood Blvd, Houston, TX 77096. Please call the coast guard. Thank you

3 19 14



Nico @curemecfs · 12h

@ReaganBattalion are also trying to coordinate help to those reaching out online.
At least they were 6 hours ago.

1 5 7



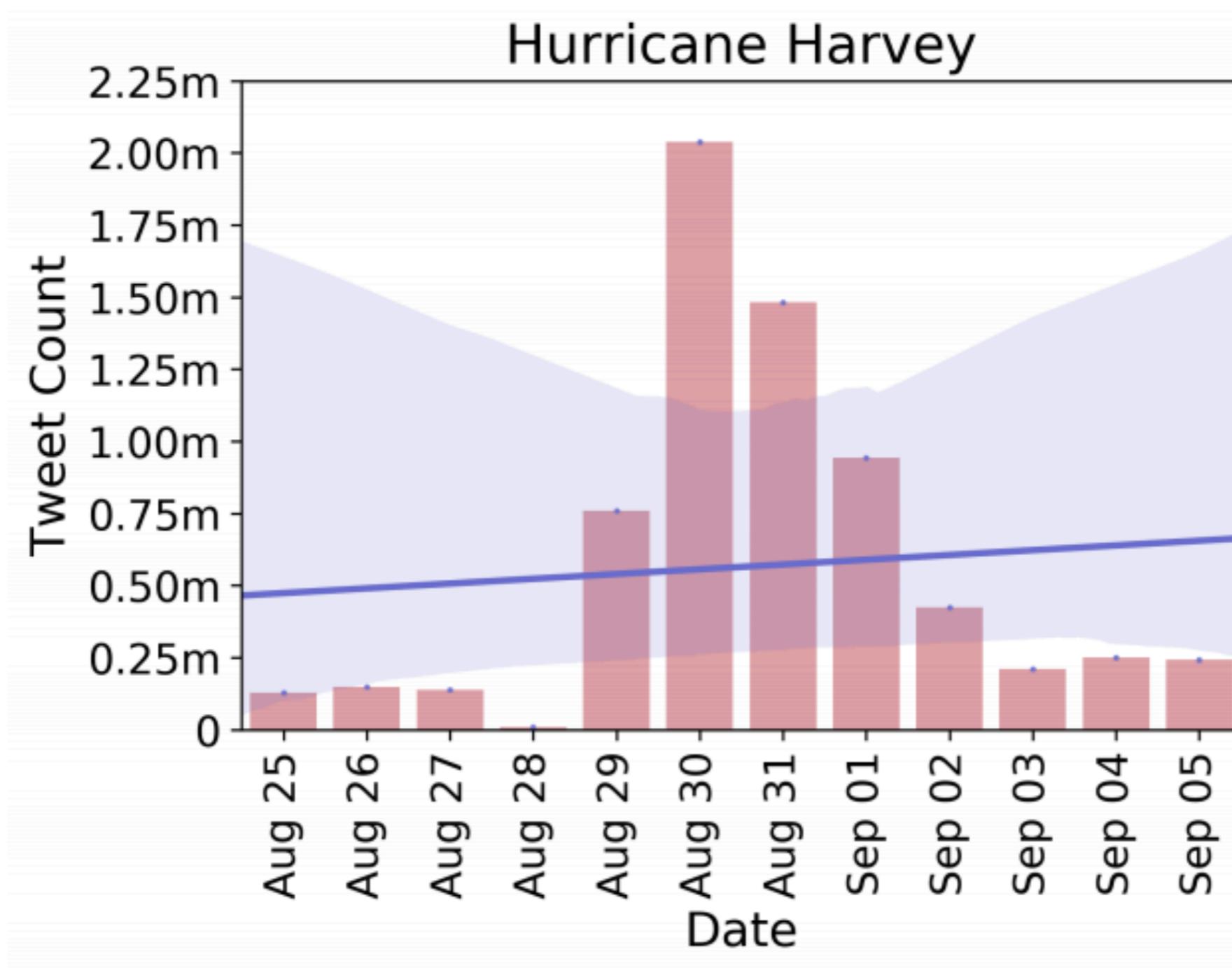
indigosky42 @indigosky42 · 12h

Are they taking addresses? Sorry I know they are doing the best they can but 7 lines for 7 million people is nuts

1 2 5



TWEET COUNTS REFERENCING HURRICANE HARVEY



Total number of tweets per day relevant to Hurricane Harvey with purple lines indicating trends in daily tweet data volume.

DATA SCIENCE PROBLEM

“Using trending hashtags and keywords, can a model read through social media posts and locate users in need of emergency assistance in a disaster?”

GATHERING DATA

- Training Data:
 - Figure Eight dataset
 - 10,000+ tweets human labeled, variety of disasters
- Testing data:
 - ~400,000 Hurricane Harvey tweets
 - From 8/25/17 - 8/29/17

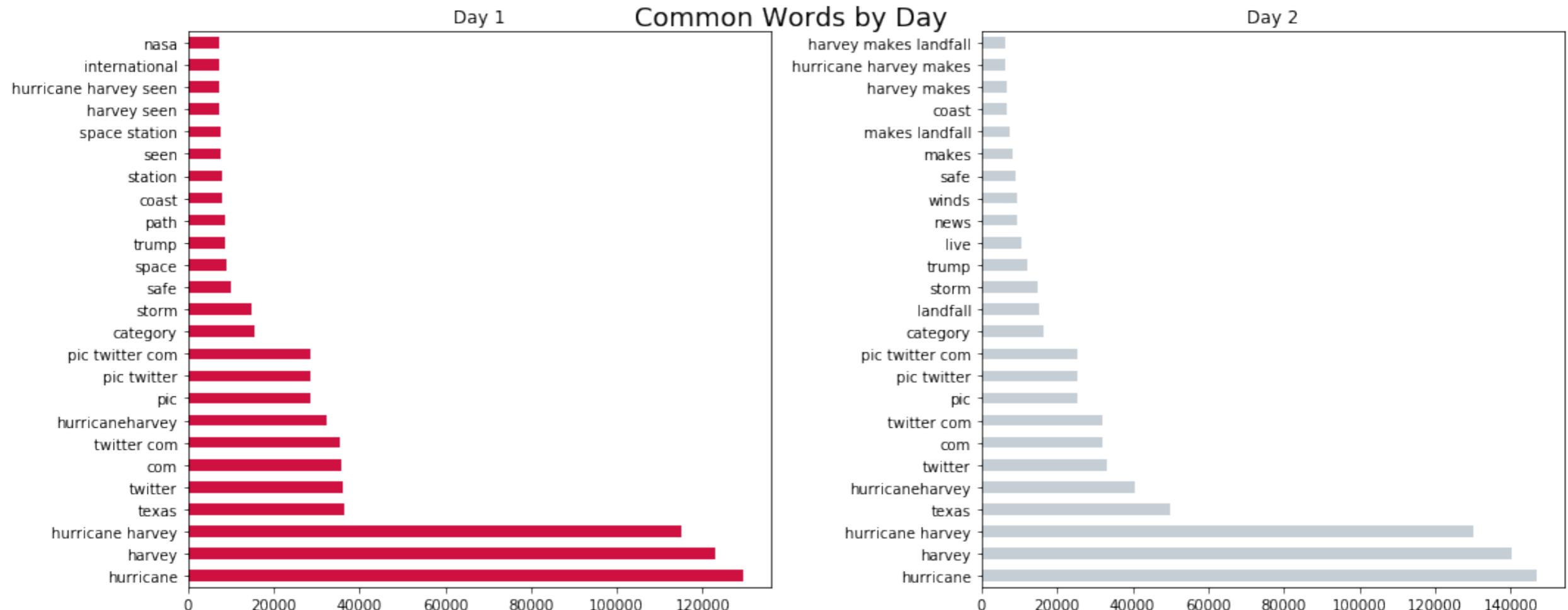
EXPLORATORY ANALYSIS

► Figure Eight Dataset:



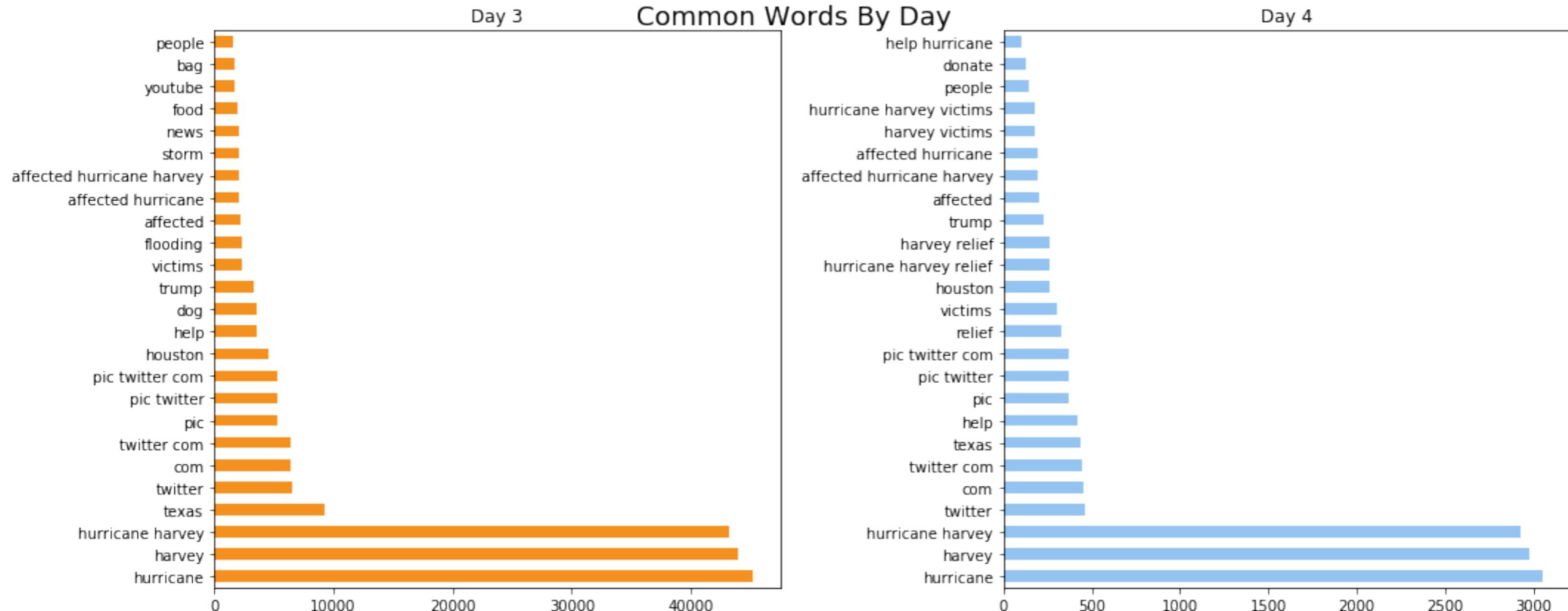
EXPLORATORY ANALYSIS

► Harvey Dataset



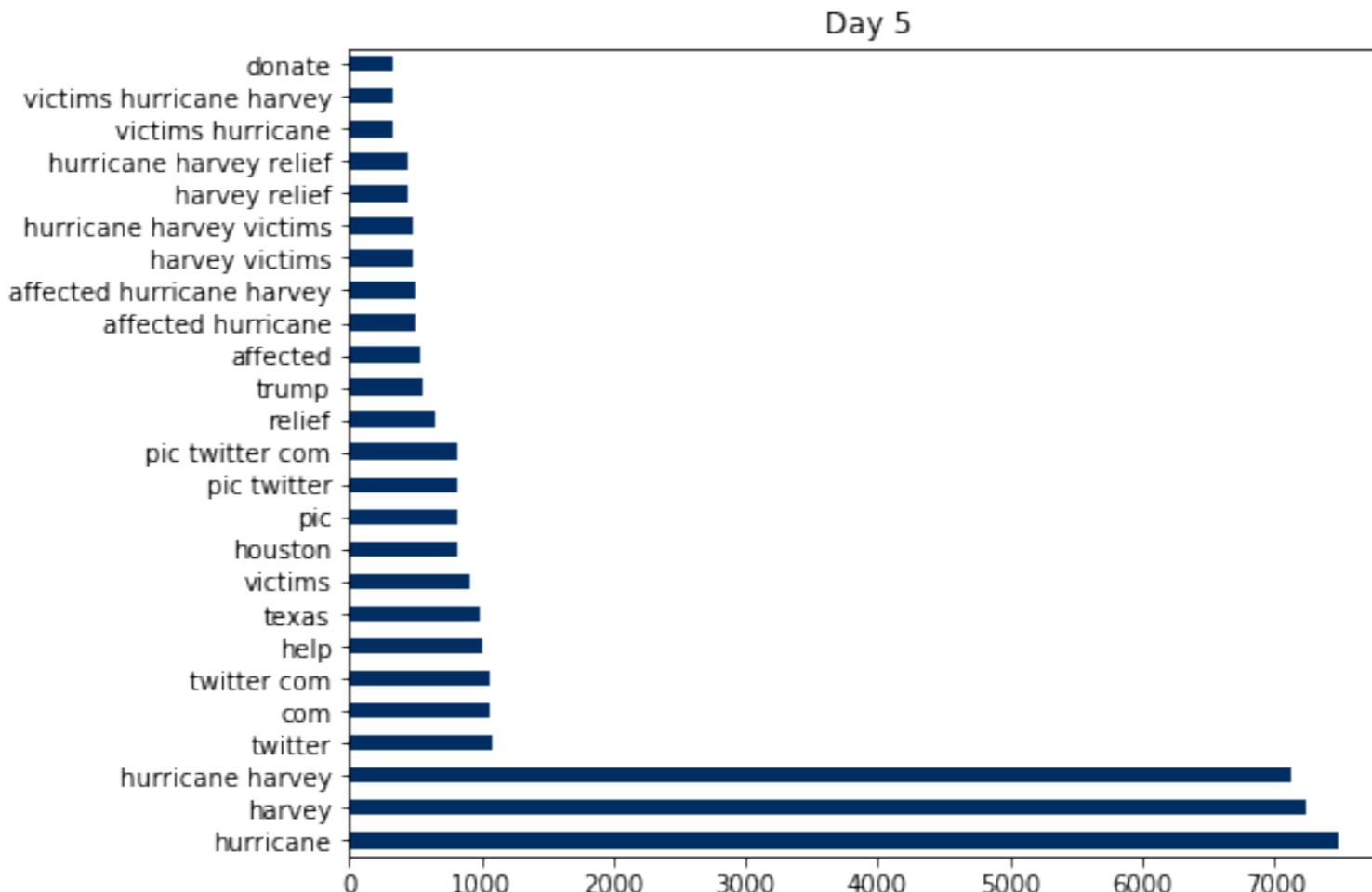
EXPLORATORY ANALYSIS

► Harvey Dataset



EXPLORATORY ANALYSIS

► Harvey Dataset

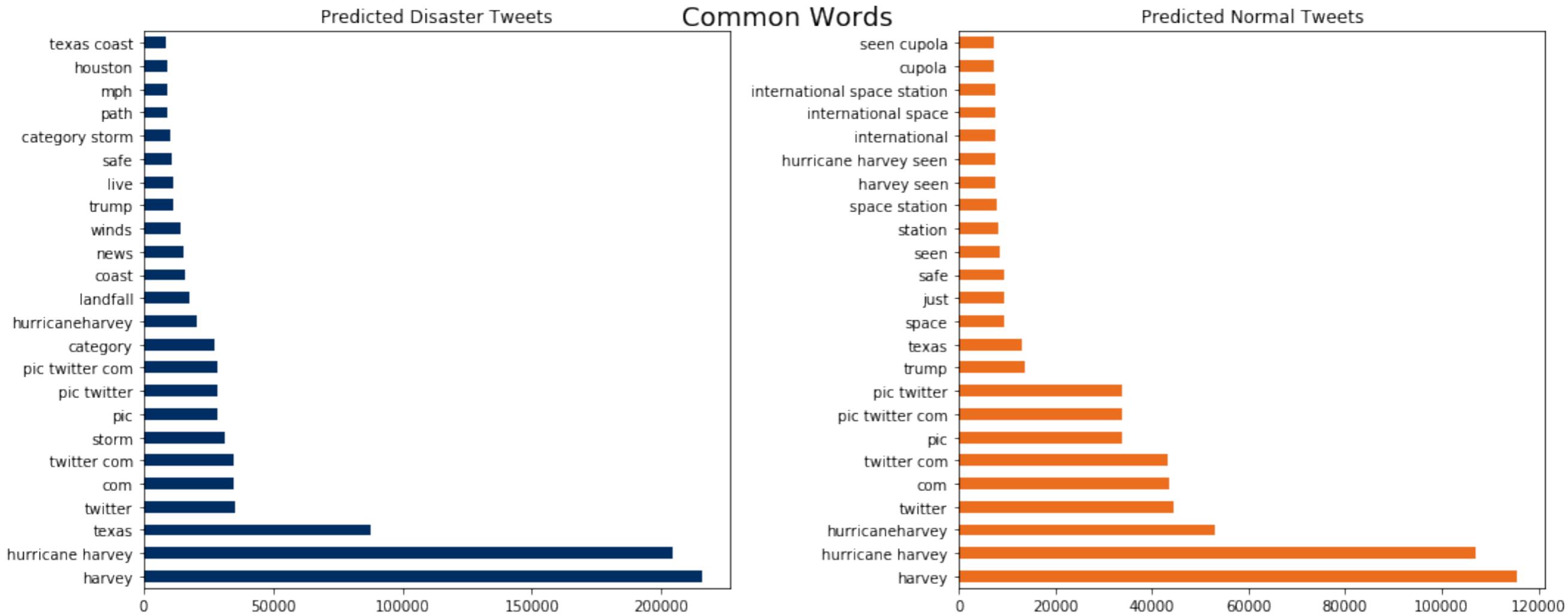


GRIDSEARCH COMPARISON

	CountVectorizer Parameters	Model Parameters	Training Score	Testing Score
Logisitic Regression	min_df: 2, n-gram range (1,2), stop_words - ENGLISH	C: 0.5, Penalty: L2	.9135	.8025
Decision Tree	min_df: 2, ngram: (1,1), stop_words: english	max_depth: 5, min_samples_leaf: 10, min_samples_split: 4	.6250	.6253
AdaBoost (DecisionTree)	min_df: 6, ngram: (1,1), stop_words: english	n_estimators: 50	.8295	.7913
Random Forest	min_df: 6, ngram: (1,1), stop_words: english	max_depth: 75, max_features: 5, min_samples_leaf: 5, min_samples_split: 2, n_estimators: 250	.7452	.7303
Naive Bayes	min_df: 2, ngram: (1,2), stop_words: english	alpha: 1	.8668	.8041
SVM	min_df: 4, ngram: (1,2), stop_words: english	C: 1	.9129	.8038
XBG	min_df: 2, stop_words: english	max_depth: 5, min_samples_split: 2, n_estimators: 1500	.8760	.7919
Voting Classifier	min_df:2 ngram: (1,3)	Logreg: Penalty: L2, C: 0.5 XGB: n_estimators: 1500, min_samples_split: 2, max_depth: 5 SVC: gamma: scale, kernel: rbf, C: 1, probability = True NB: aplha:1 RF: max_depth: 75, min_samples_split: 4, n_estimators: 250 voting = soft	.8914	.8086
<u>Selected Model:</u> Voting Classifier	min_df:2 ngram: (1,3)	Logreg: Penalty: L2, C: 0.5 XGB: n_estimators: 1500, min_samples_split: 2, max_depth: 5 NB: aplha:1 voting = soft	.8611	.8701

MODEL PREDICTIONS

- Voting classifier:
 - Predicted 58.3% of the Harvey dataset was a disaster tweet



MODEL PREDICTIONS

- Filtered the disaster predicted tweets through a “Go” list
 - Words that might be used only in the case of a real emergency
 - Cut the number of tweets down to 3.25% of the total dataset, or 13,000.
 - Was not fruitful in helping identifying “urgent” tweets

SYSTEM IN ACTION

- Step 1:
 - We would use our Twitter scraper to collect tweets using event specific keywords and hashtags.
- Step 2:
 - Run the collected twitter posts through our model to determine which posts are relevant.
- Step 3:
 - Filter those relevant tweets into urgent or non urgent categories.
- Step 4:
 - Flag and map those urgent tweets for organizations to search and send assistance.



By SC National Guard - 170831-Z-AH923-081, Public Domain, <https://commons.wikimedia.org/w/index.php?curid=62096178>

FUTURE USE

- Potential organizations that can make use of our system.
 - FEMA
 - Local emergency respondents
 - Third party volunteer organizations
- Our system can be applied to any other disasters and future events
 - Trending hashtags and keywords can be used for future disaster searches.
- Eventually with access to users location data from the urgent tweets.
 - Development of heat-maps, plotting, and other visual aids to assist organizations in identifying areas to allocate more resources to.

FUTURE IMPROVEMENTS

- Improve datasets:
 - Collect and validate more tweets for our model to learn from.
- Improve filter keywords:
 - Expand upon our “Go” list
- Unsupervised model learning
 - Cluster and Identify similar tweets containing urgency keywords.
- Language Support:
 - Train models on multiple languages
- Potential to handle photos & videos:
 - Image processing could be another step in the future.
- Support from Twitter:
 - ACCESS to twitter users location data.