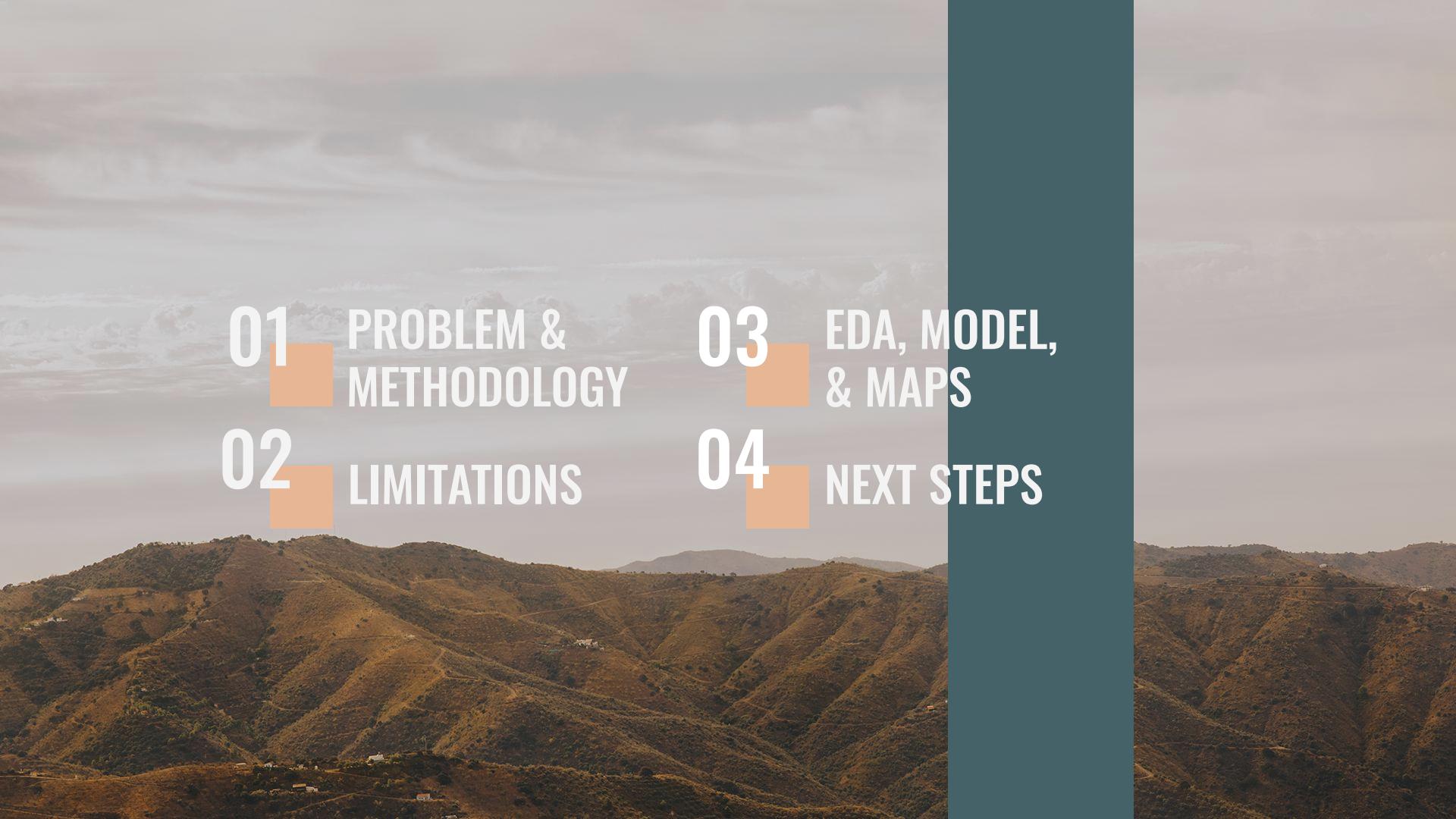


Mapping Disasters with Social Media



Presented by Femi Adenugba, Sue Moy & Teng Mao
from General Assembly in DSI-8-Boston





01 PROBLEM &
METHODOLOGY

02 LIMITATIONS

03 EDA, MODEL,
& MAPS

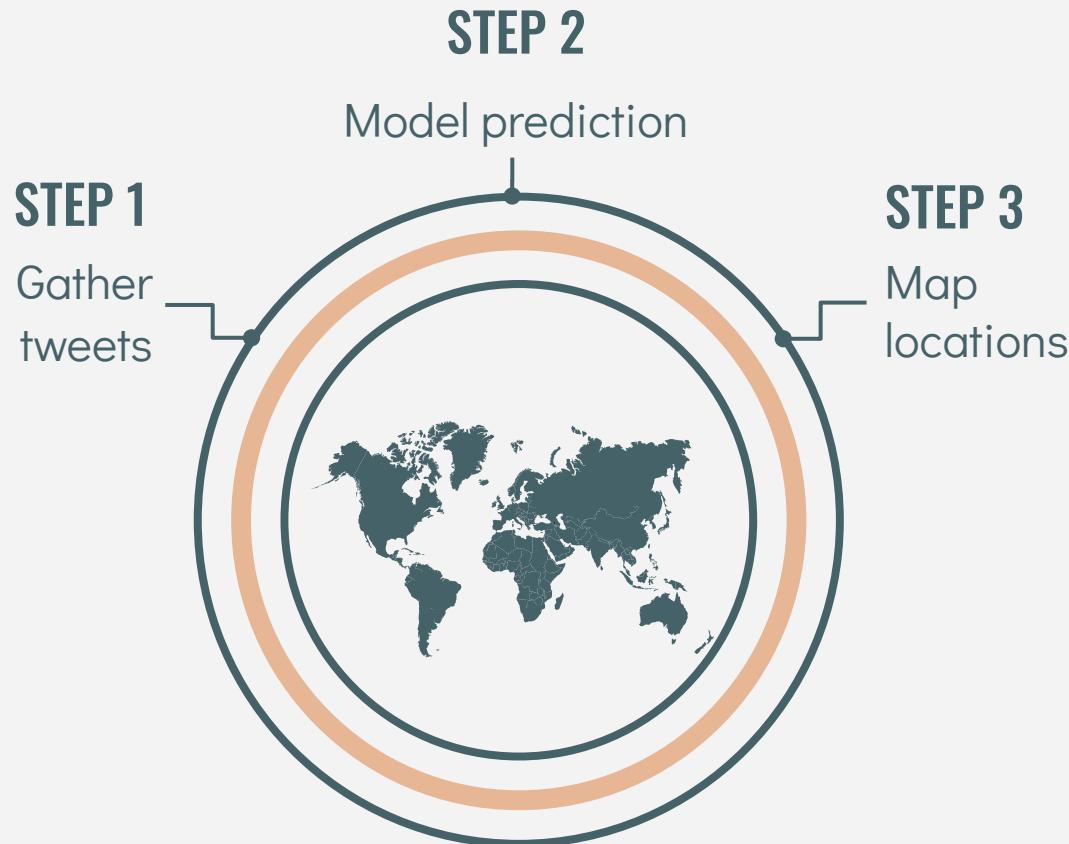
04 NEXT STEPS

PROBLEM STATEMENT

Can we turn to social media to determine areas that need immediate assistance after a natural disaster?



METHODOLOGY FOR SOLVING THE PROBLEM





LIMITATIONS



Twitter API - 7 day limit

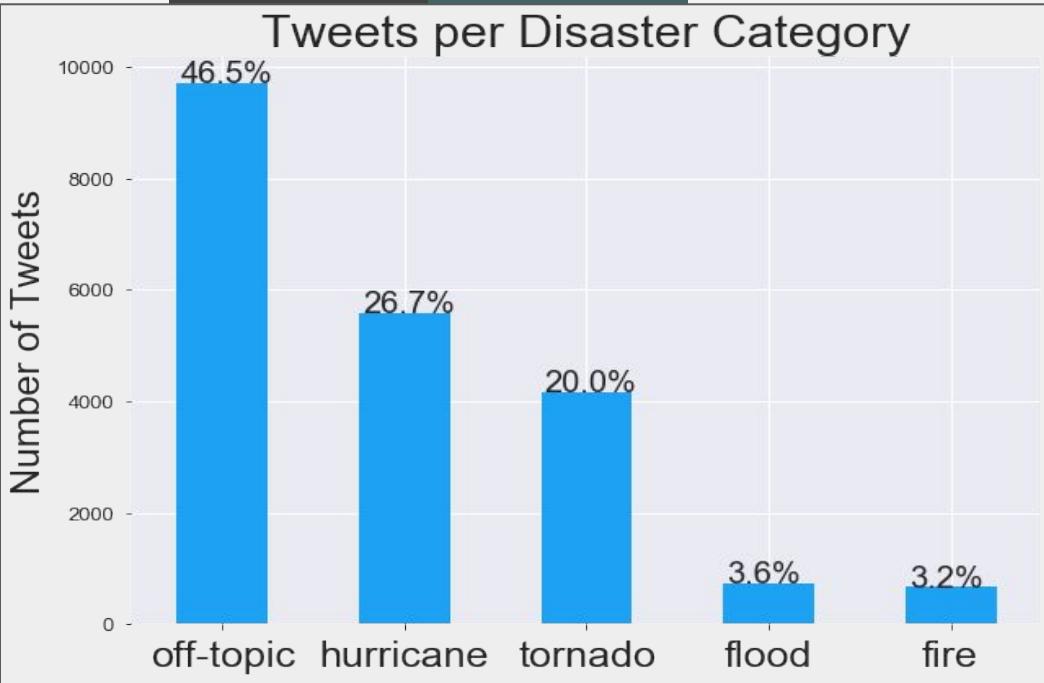


Lack of geolocation

CAPE COD TORNADO

Tweets with Geolocation

0.60%



TOTAL TRAINING TWEETS



Dataset of 20,873 total tweets from CrisisLex.org



5,579 tweets for Hurricane Sandy in 2012



4,168 tweets for the Oklahoma Tornado in 2013



743 tweets for the Colorado Floods in 2013



673 tweets for the Colorado Fires in 2012

MODEL CREATION

Created pipelines with CountVectorizer that tokenized, lemmatized, and removed stop words. The following models were used in each of the pipelines, and parameters were optimized using GridSearchCV.



Logistic Regression

Accuracy score of 95.8%
on the train data and
92.5% on the test data



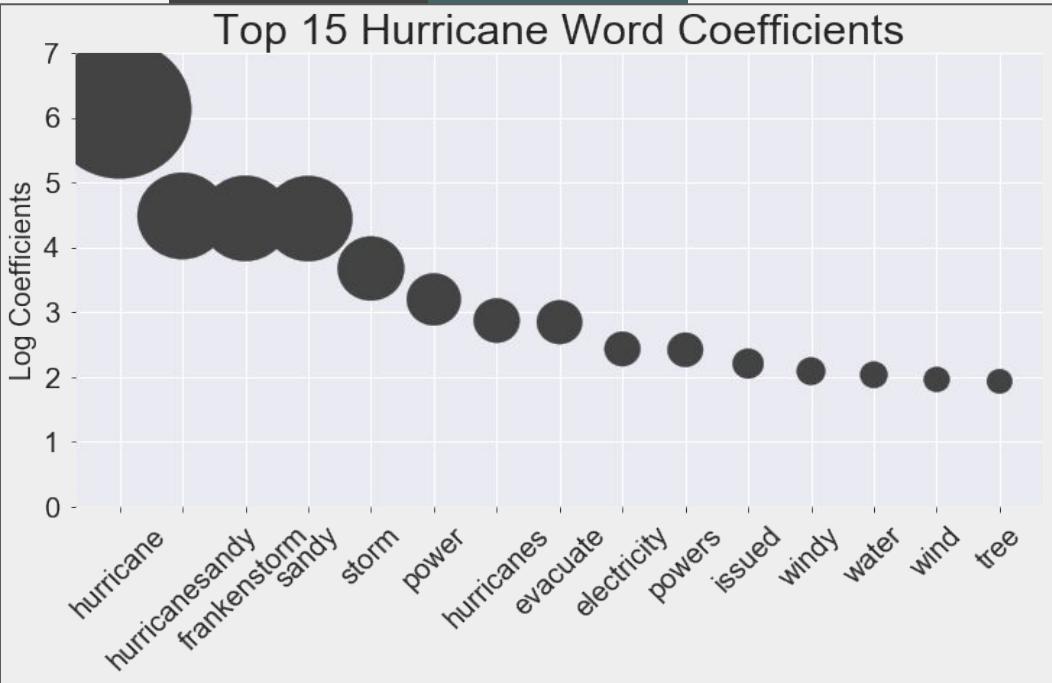
Support Vector Machines

Accuracy score of 89.2% on
the train data and 88.6% on
the test data



Random Forests

Accuracy score of 91.9% on
the train data and 91.1% on
the test data



TOP HURRICANE RELATED WORDS

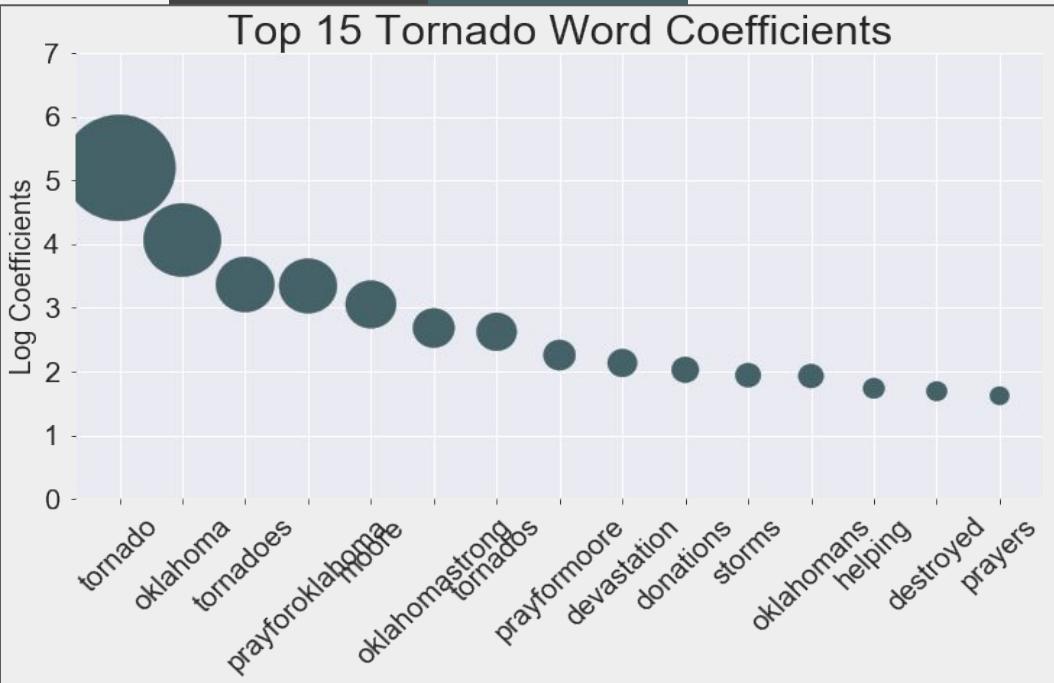


"Hurricanesandy", "frankenstorm", "power", "evacuate", "windy", and "water" are words that jump out for Hurricane Sandy.



Based on the log coefficient of 4.45 for "frankenstorm", the appearance of that word in a tweet would indicate an 98.8% chance that the tweet is about the hurricane.

TOP TORNADO RELATED WORDS



For tornado, "prayforoklahoma", "moore", "oklahomastrong", "devastation", and "donations" are all words related to the tornado that blew through Moore, Oklahoma.



Based on the log coefficient of 4.06 for "oklahoma", the appearance of that word in a post would indicate an 98.3% chance that the post is about the tornado.

FINAL MODEL

CountVectorizer Parameters:

- max_features = 1000
- min_df = 2
- ngram_range = (1,3)

Random Forest Parameters:

- max_depth = 24
- min_samples_leaf = 4
- min_samples_split = 3
- n_estimators = 100

	Predicted Hurricane	Predicted Off-Topic	Predicted Tornado
Actual Hurricane	1,237	157	1
Actual Off-Topic	58	2,286	84
Actual Tornado	3	129	910



For hurricane, the model generated a 95% precision and 89% sensitivity score

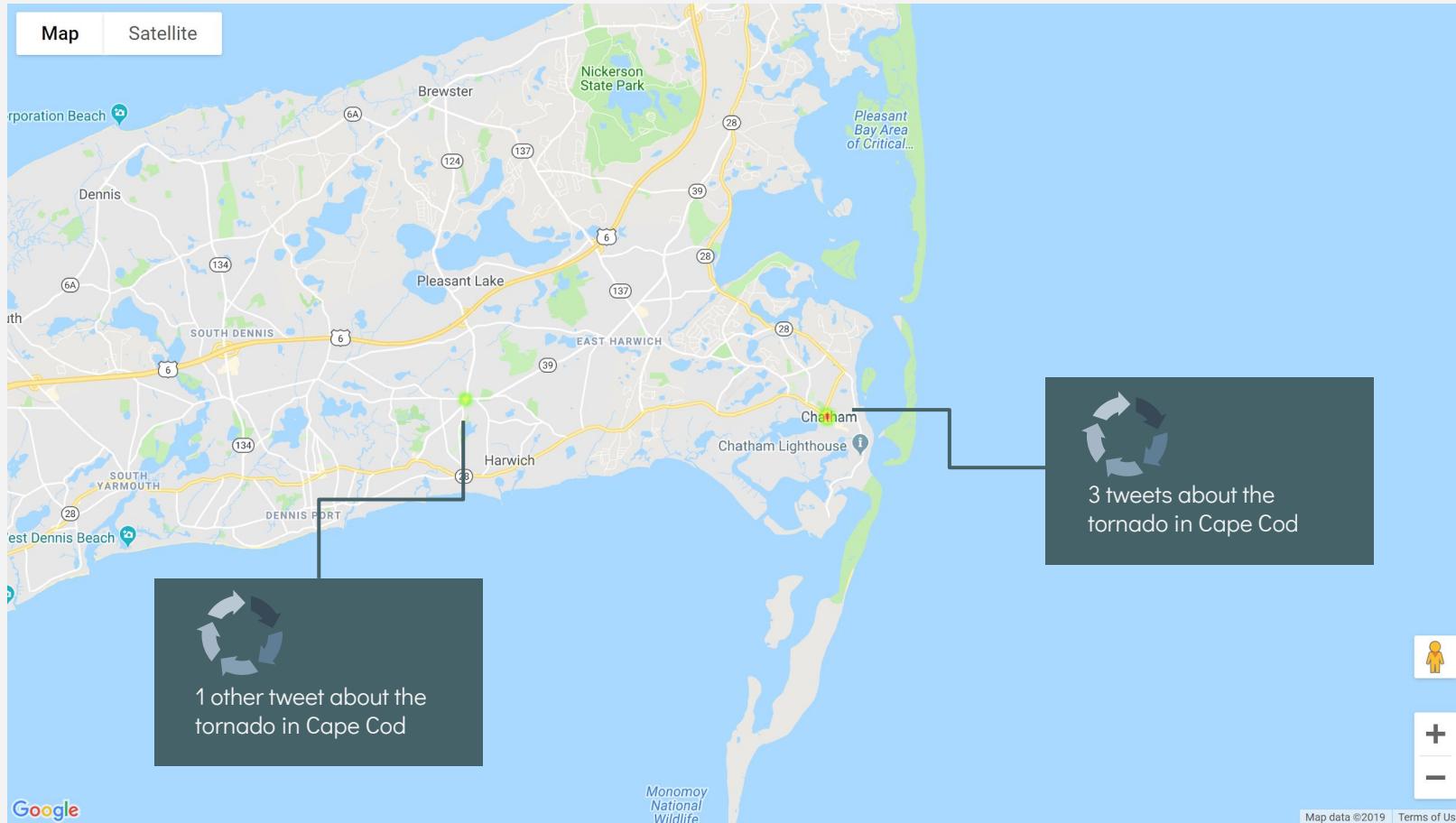


For tornado, the model generated a 91% precision and 87% sensitivity score

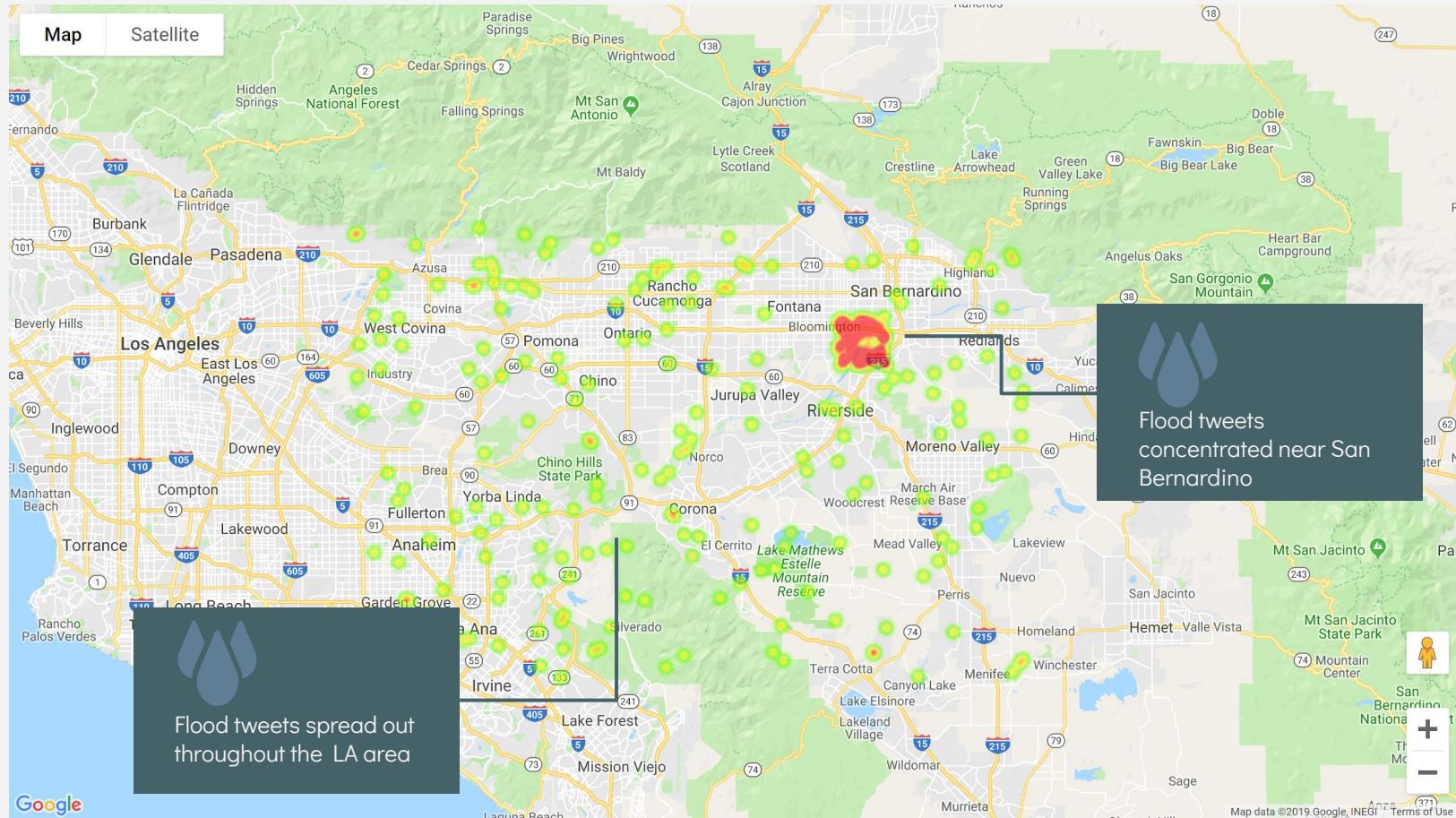
TWEET MAPPING - TORNADO



TWEET MAPPING - CAPE COD TORNADO



TWEET MAPPING - LA FLOODS



CONCLUSION

With our model, we'll be able to recognize disasters as they occur and provide location data for the public's use.



\wedge



$+$



$+$



$+$



)

=



RECOMMENDATIONS

GEO-LOCATION

Reach out to Twitter to add the option of disaster geo-location



EARLY WARNING NOTIFICATION

Based on population density and tweet threshold

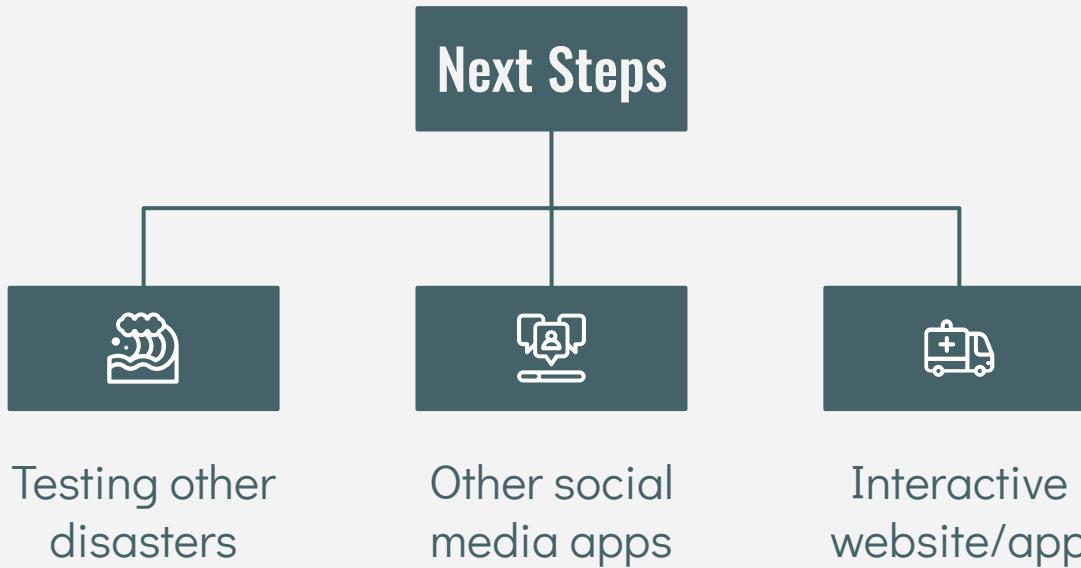


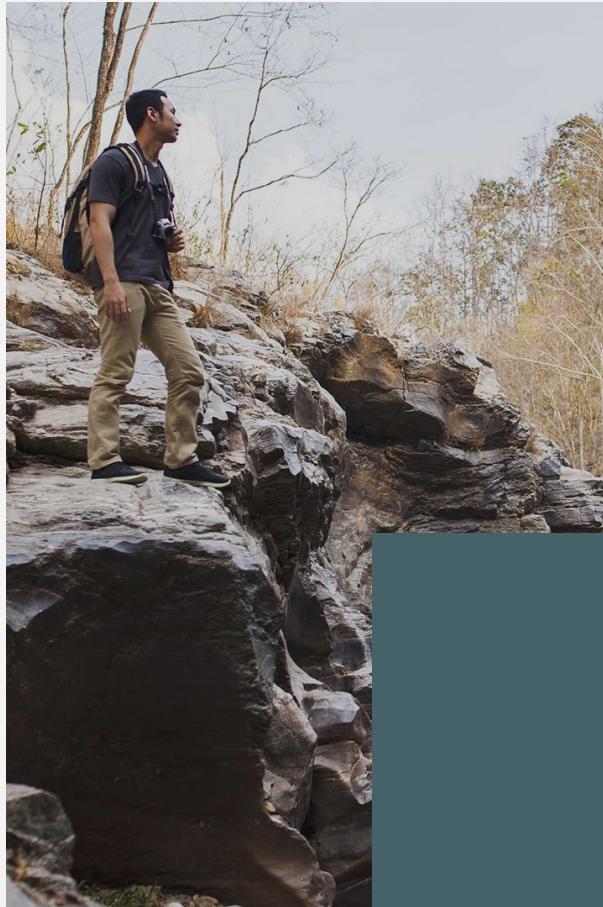
INFO POP-UPS

Load tweets and pictures onto google maps output based on likes and RT's



FUTURE EXPLORATION





THANKS!

Does anyone have any questions?



CREDITS

This presentation was completed as part the curriculum for General Assembly's Data Science Immersive by the following:

- ◀ Femi Adenugba
- ◀ Sue Moy
- ◀ Teng Mao

Slides template by:

- ◀ Presentation template by [Slidesgo](#)
- ◀ Icons by [Flaticon](#)
- ◀ Infographics by [Freepik](#)
- ◀ Author introduction slide photo created by Freepik
- ◀ Text & Image slide photo created by [Freepik.com](#)