Misinformation & Natural Disasters?

By
Reem Mokhtar
Jared Delora-Ellefson
&
Daniel Gurzi

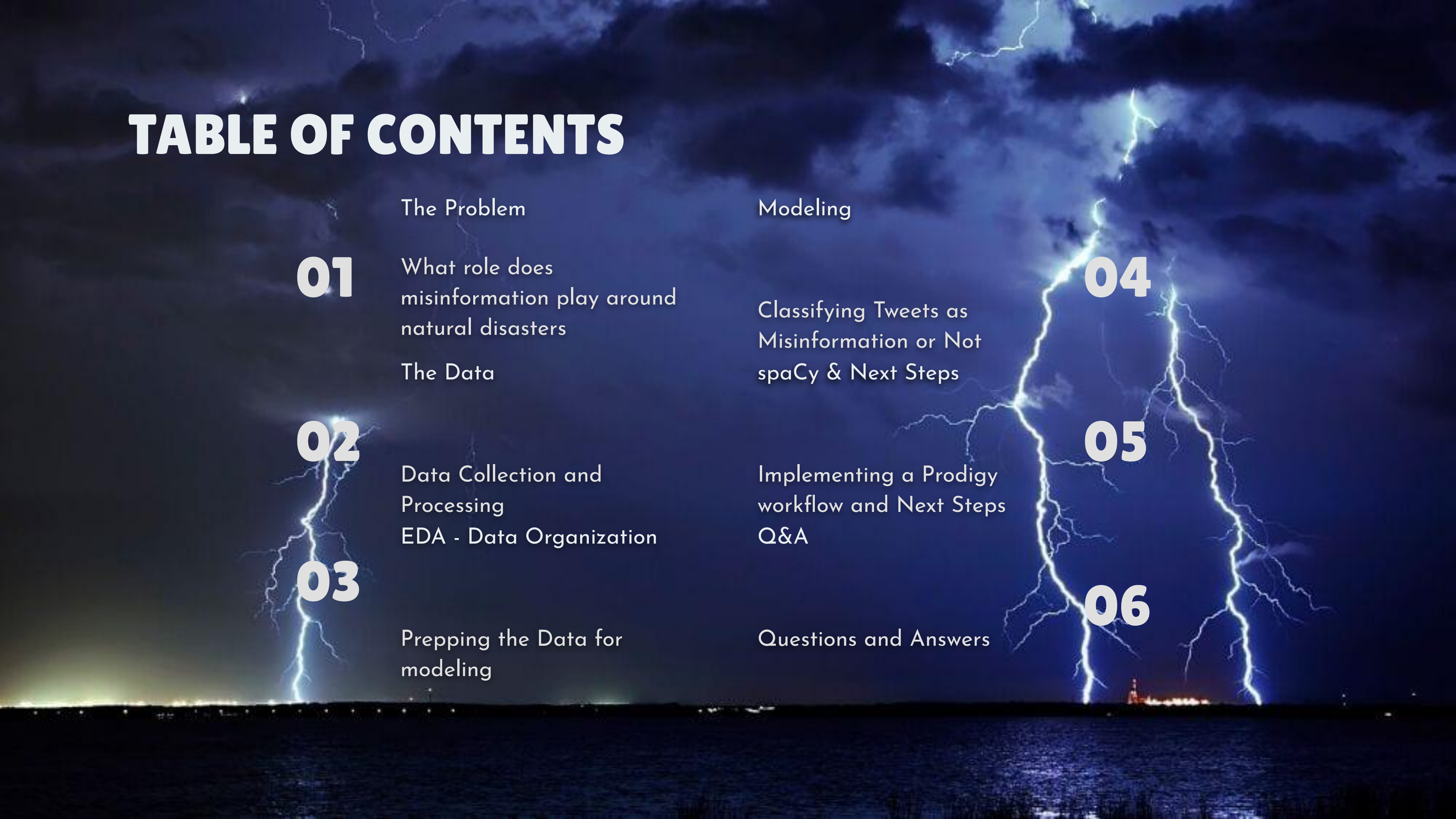


TABLE OF CONTENTS

01

The Problem

What role does misinformation play around natural disasters

02

The Data

Data Collection and Processing

03

EDA - Data Organization

Prepping the Data for modeling

04

Modeling

Classifying Tweets as Misinformation or Not

05

spaCy & Next Steps

Implementing a Prodigy workflow and Next Steps

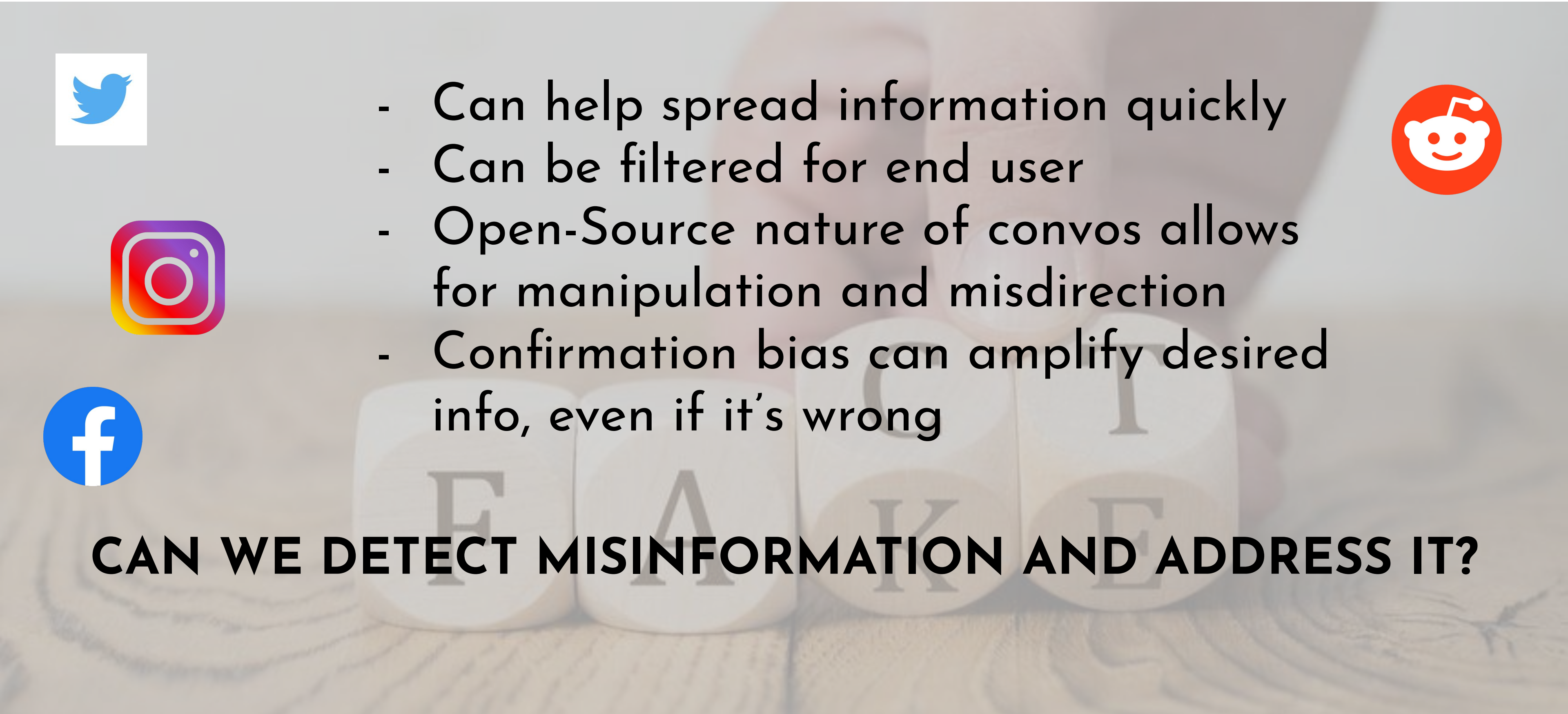
06

Q&A

Questions and Answers

THE PROBLEM

Finding Misinformation in Disaster Areas



- Can help spread information quickly
- Can be filtered for end user
- Open-Source nature of convos allows for manipulation and misdirection
- Confirmation bias can amplify desired info, even if it's wrong

CAN WE DETECT MISINFORMATION AND ADDRESS IT?

The Data

Twitter

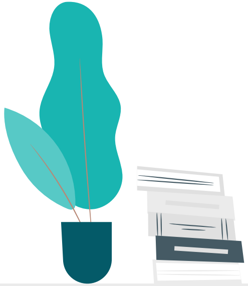
- Efficient Use of communication
- Easy to use API through GOT3

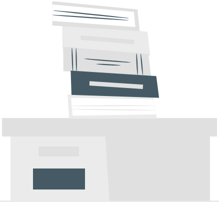
Facebook

- More Users
- Less Moderation
- Extremely Difficult to get access

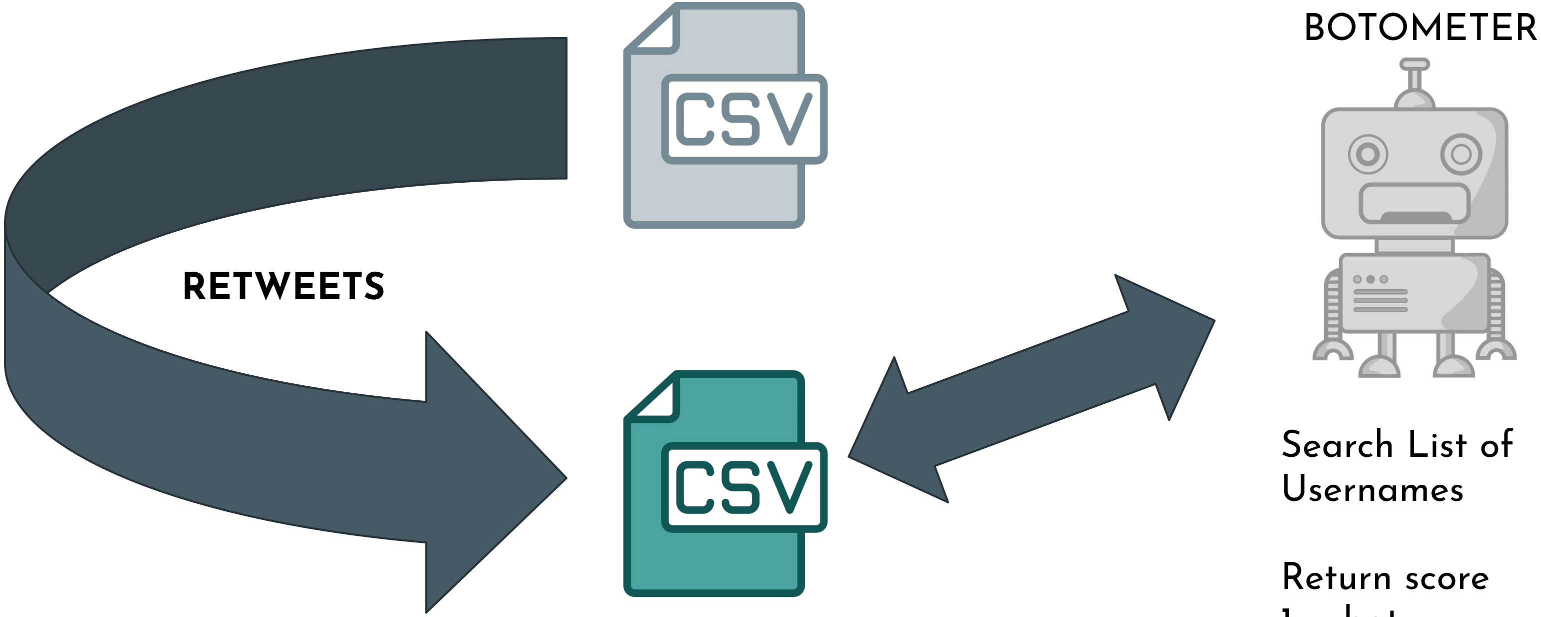
Reddit

- Bigger dataset with more context words per post
- Broader scope than an individual disaster





DATA PROCESSING



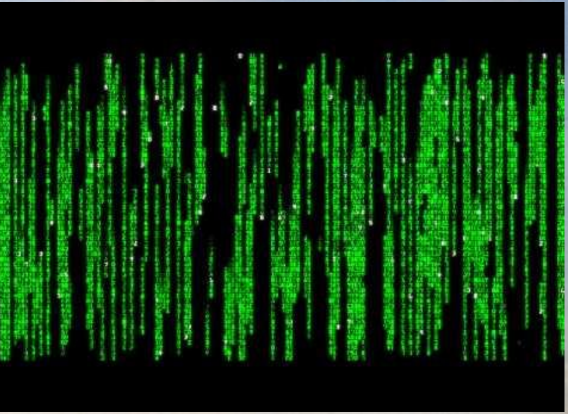
RETWEETS

BOTOMETER

Search List of Usernames

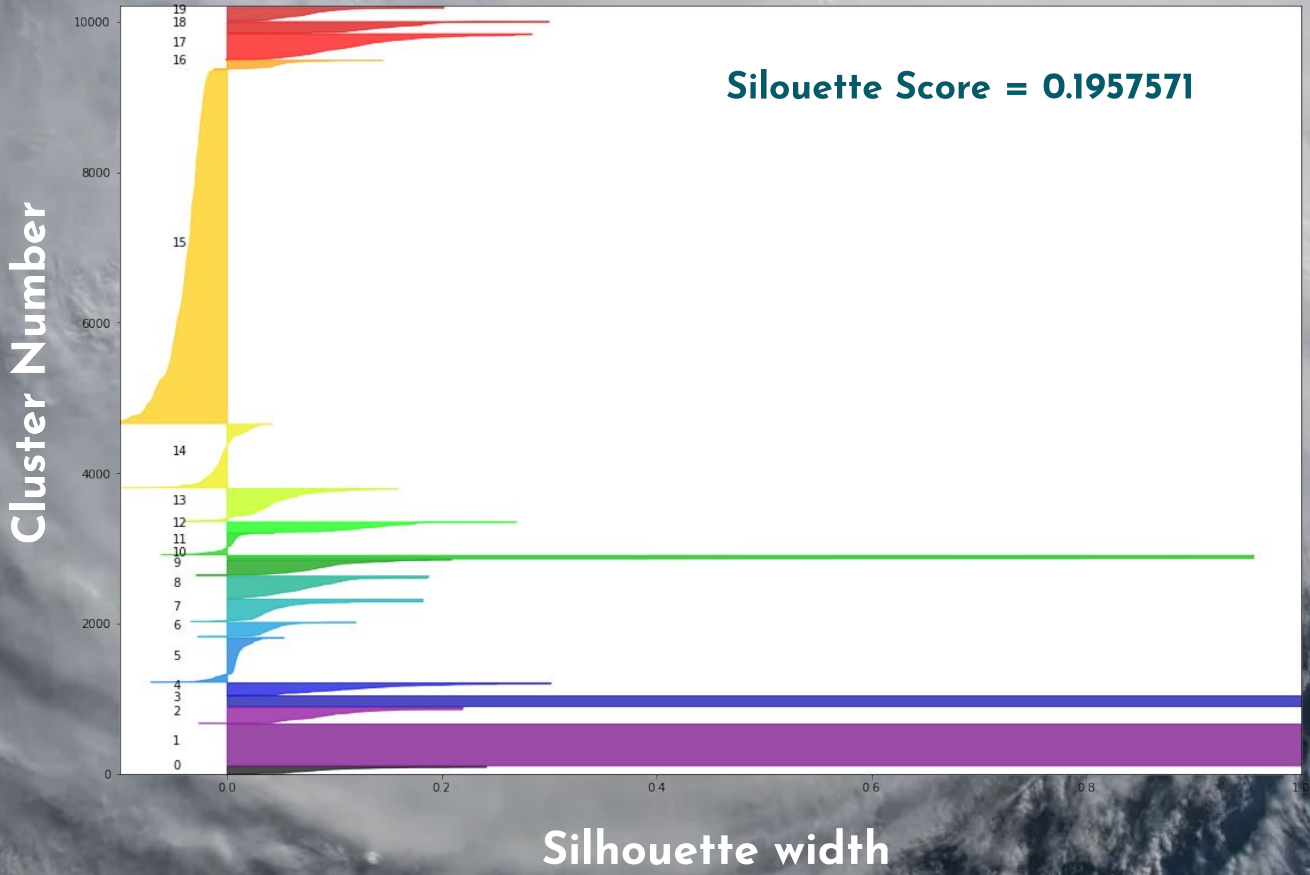
Return score
1 = bot
0 = human

DATA PROCESSING



- Clean Tweets using REGEX
    - No Special Characters
    - No URLs
- Lemmatize Tweets
    - Group word stems
    - (Flood, Flooding, Flooded = Flood)
- Vectorize Tweets
    - CountVectorizer & TFDIF
    - Word2Vec

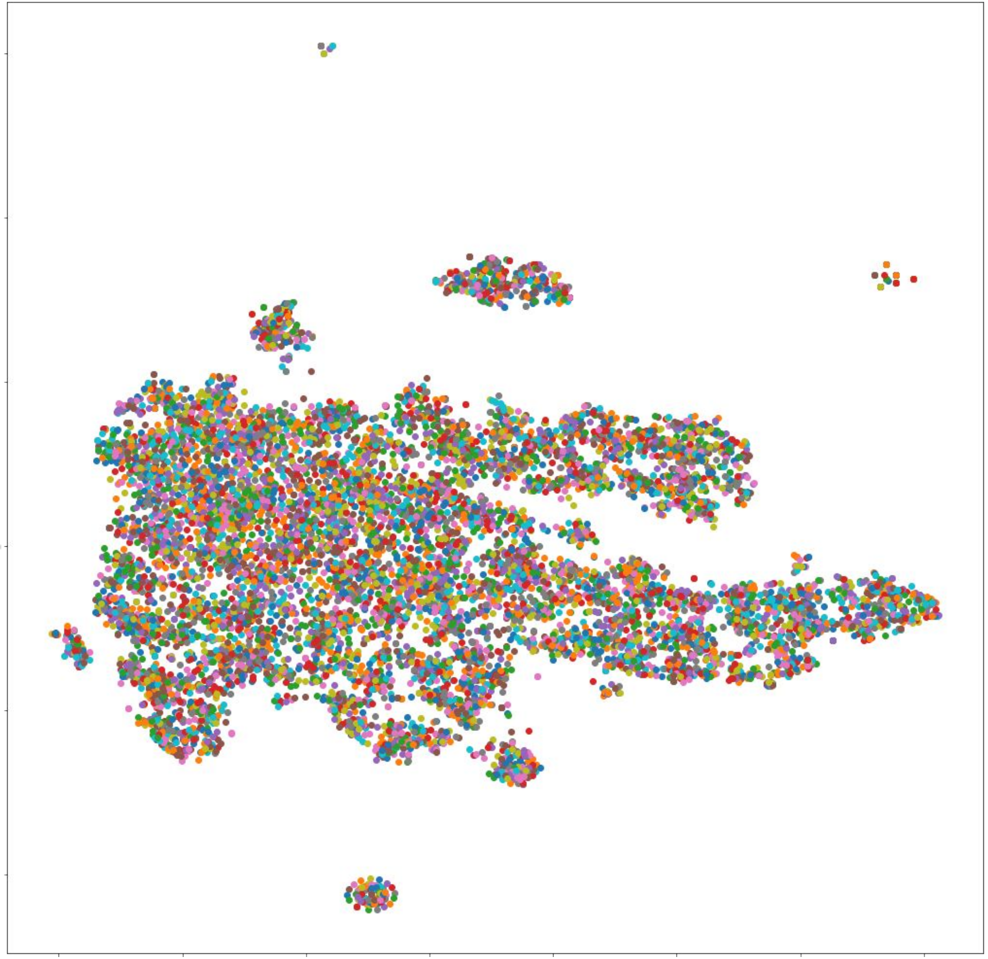
KMEANS MODEL



Silouette Score = 0.1957571

TSNE PLOT

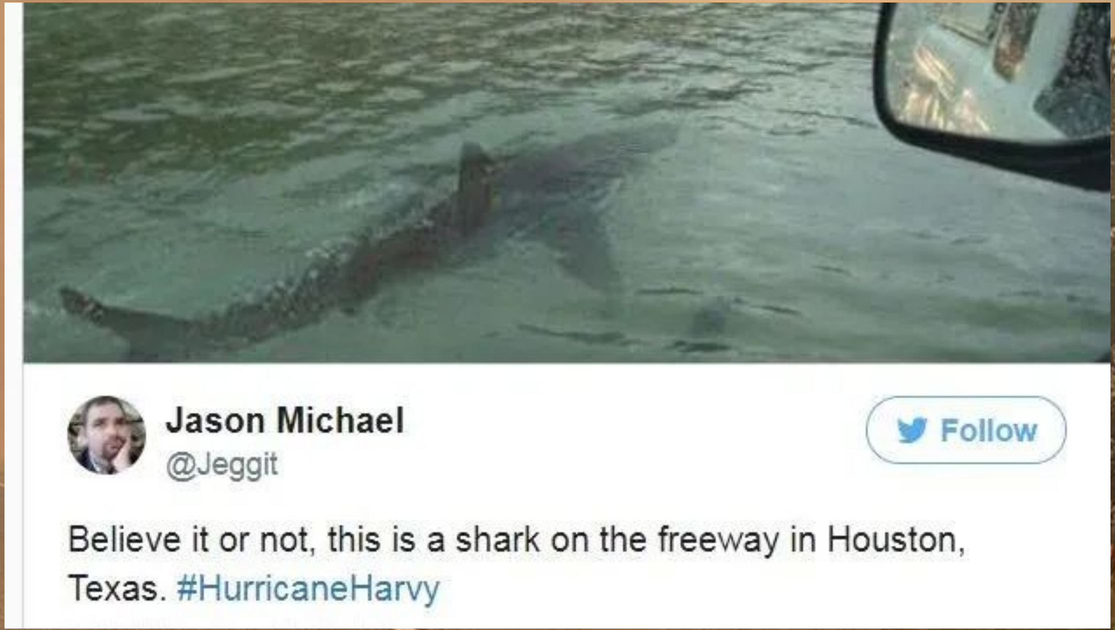
High level visualization showing clustering by tweet



WHAT'S HAPPENING

NEED MORE BAD DATA

LESS DISASTER SPECIFIC

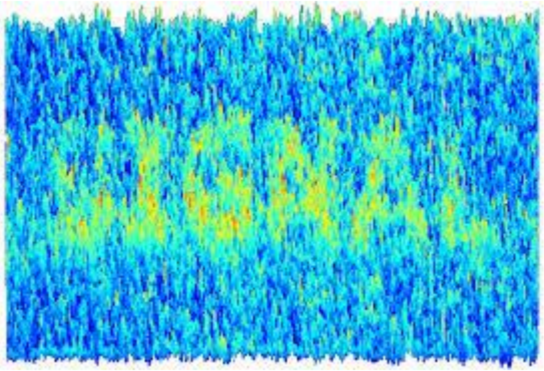


Jason Michael
@Jeggit

Believe it or not, this is a shark on the freeway in Houston, Texas. #HurricaneHarvy

What is Spacy? Natural Language Processing

Social Media



Our Signal - Misinformation

- Twitter
- Reddit

Tweet or Reddit posts contain misinformation, spaCy can be trained to detect this signal.

Misinformation Detected!

The model that were used for this project did the following:

- Detect specific example words
- Detect patterns of speech

Amplifying the Signal

The Noise in Hurricane Harvey tweets drowns out the signal of misinformation

- Added language examples from r/Conspiracy to amplify the signal
- Randomly shuffle misinformation examples into the Harvey Tweets
- Train spaCy to detect misinformation

Known Misinformation

Misinformation from r/Conspiracy

Social Media Tweets During Disaster

spaCy is trained with a shuffle of data from r/Conspiracy and Tweets from Hurricane Harvey

Results

Model Performance	Train	Harvey Tweets	Irma Tweets
Entity Detection	~40%	~40%	~10% Accuracy
Entity Detection + Text Classification	~94%	-	-

Recommendations & Future Work

1. Our text classification is overfit for conspiracies.
2. Collect more specific misinformation examples from natural disasters to train on.
    a. Misinformation such as power shutdowns, evacuation orders, area clearing, spread of disease, etc
    b. Disinformation (fake images during Sandy).
3. Expand model from hurricanes to natural disasters
4. Manual annotation for subset of disasters to generalize

Recommendations & Future Work

Paper: *Atodiresei, C. S., Tănăselea, A., & Iftene, A. (2018). Identifying fake news and fake users on Twitter. Procedia Computer Science, 126, 451-461*

1. Credibility scores (credit to Daniel)
2. NER
3. Text classification
4. Emoji sentiment
5. Hashtag sentiment
6. Botometer scores
7. Image recognition/classification
8. Geolocation weighting (weights decrease outwards in radial increments from center of disaster).

QUESTIONS?