

The background features a complex network of thin grey lines connecting various points, some of which are solid black dots. Scattered throughout are several triangles of different sizes and orientations, some with solid outlines and others with dashed or dotted outlines. The overall aesthetic is technical and digital.

Using Social Media to Map Power Outages

Presented By: Bill Fu
Nathan Seither
Aidan Curley
David Lee

Meet the Team



Bill Fu

Data Scientist / Geophysicist

A motivated and creative data scientist, who is enthusiastic and committed to solve complex problems.



Aidan Curley

Data Scientist

Passionate about machine learning and artificial intelligence.



Nathan Seither

Data Scientist / Entrepreneur

Enjoy's bringing creative solutions to stakeholders. Has a solution for everything.



David Lee

Data Scientist / Devops

Loves to bring ideas to life. His business background works on refining production processes.

Overview

Existing methods of identification
Repercussions of slow response

01

Problem Statement

What we are trying to solve

02

Process

Data Collection
Exploratory Data Analysis
Success Metrics Selection

03

Modeling

Model Selections
Prediction Layering

04

Agenda

05

Implementation

Utilizing our findings into practice

06

Conclusions

Can we map power Outages and
Identify hotspots?

07

Next Steps

Road to production model.
Minimal Viable Product demo

08

QA

Questions or comments

Overview

- Lack of power = reduction in trust
- Lack of trust = added costs
- Technology can help bridge the gap.

[Click here](#) for reference to recent litigation against Eversource Power Company

Problem Statement

During a disaster, residential areas often experience massive power outages, that in many cases last for days. Traditional methods to map power outages include live feeds and data that is provided by major utility companies as well as on satellite data that capture the extent of light emitted at night.

Our method will utilize a Twitter bot, trained on historical data using NLP and classification models to detect tweets about power outages in real time and map the users location.

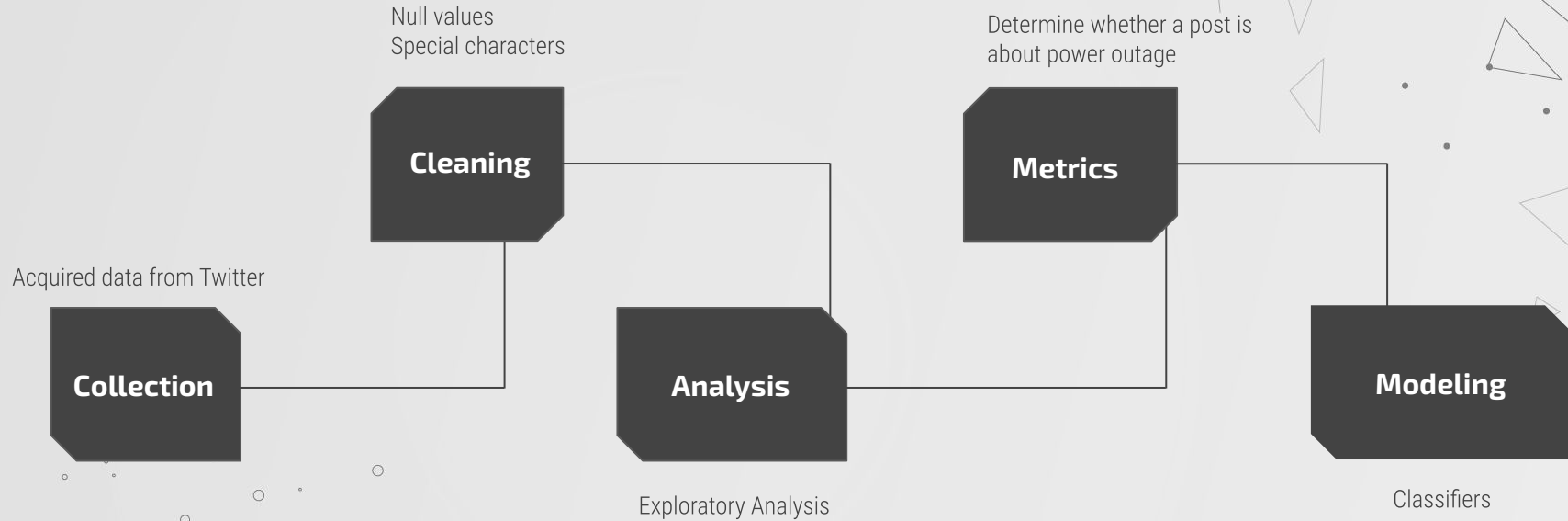


03

Process

Bill Fu ~ Data Scientist / Geophysicist

Process



Process: Data Collection

Posts scraped from Twitter between 1-1-2020 to 8-31-2020

190k

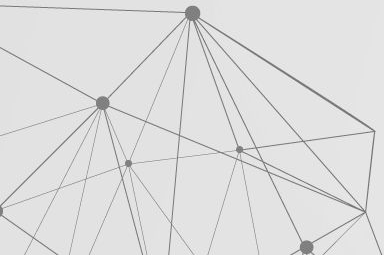
Power outage Tweet Criteria

90k

Non power outage Tweet Criteria

280k

Size of combined dataset



Process: Clean Data



Scraped Dataset

- Initial split scraping on phases
- Combined

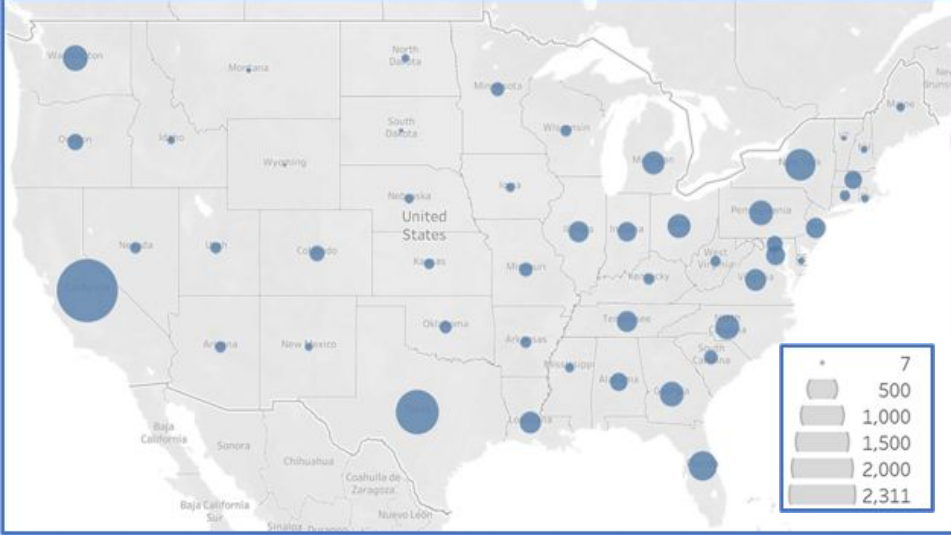
- Removed irrelevant locations
- Eliminated special characters
- Made labels

Cleaning Data

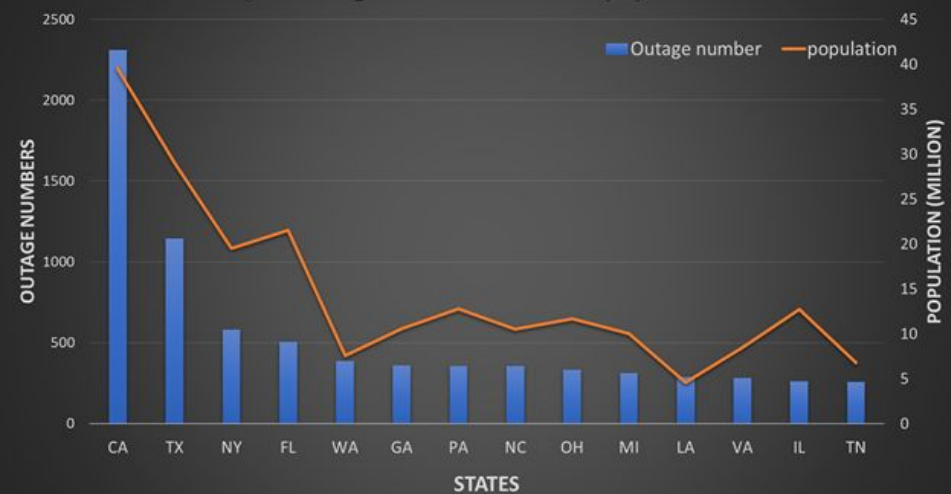


Ready for Analysis

The final data frame consisted of tweets by location that were either for a power outage or not a power outage.



Top 15 outage number states and populations



Process: Analysis

~Initial location analysis on 28,000 tweets

Outages Across America

- 11,000 outage posts with location information across the US
- Most power outages located in:
 - Coastal States
 - Northeast and Central East States
- Power outages are highly correlated with population



SUCCESS METRIC: ACCURACY

04

Modeling

Aidan Curley ~ Data Scientist



Modeling Strategy

Naive Bayes

Performs well on categorical variables

LSTM RNN

Successfully used in language modeling

Multi Layered Prediction Model

To pass through a series of Model gates in order to be considered a valid power outage tweet

Logistic Regression

Great for predicting an outcome variable on categorical data

Random Forest

Ensemble learning model for classification using several decision trees.

Modeling: IN DEPTH

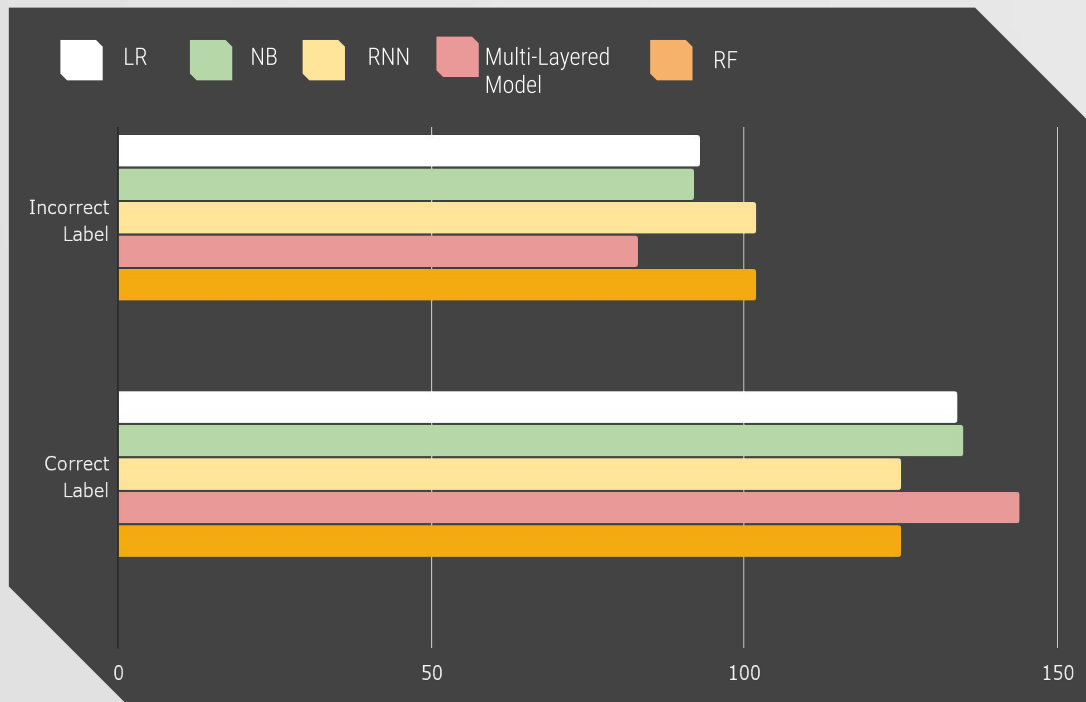
Why we chose this method:

In order to push our model to a production environment, we need to eliminate false positives, which would impact reliability.

Although some models perform well individually, having a layered solution will help in filtering out tweets that are not related to a power outage.



Model: Analysis



Multi-Layered Modeling

Our Multi-layered approach was able to achieve a higher accuracy rating than each model could've predicted alone.

Model Metrics

Model Accuracy:

Logistic Regression Accuracy: 58.26%

Naive Bayes Accuracy: 59.57%

LSTM RNN: 55.65%

Random Forest: 55.65%

Multi-Layered Model: **63.91%**

Confusion Matrix for Layered Prediction

	Predicted (1)	Predicted (0)
Actual (1)	58	40
Actual (0)	43	89

The Accuracy score is: 63.91%

The Missclassification rate is: 36.1%

The Sensitivity is: 57.43%

The Specificity is: 68.99%

The Precision is: 59.18%



05

Implementation

Implementation: ML Twitter Bot



New Tweets

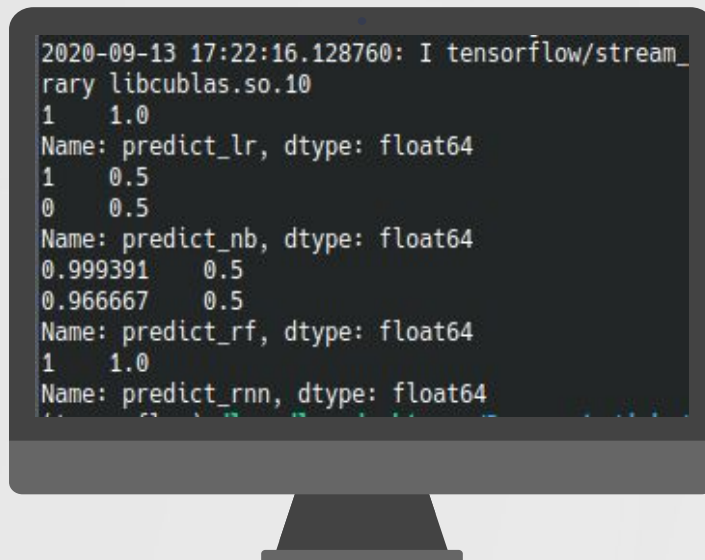
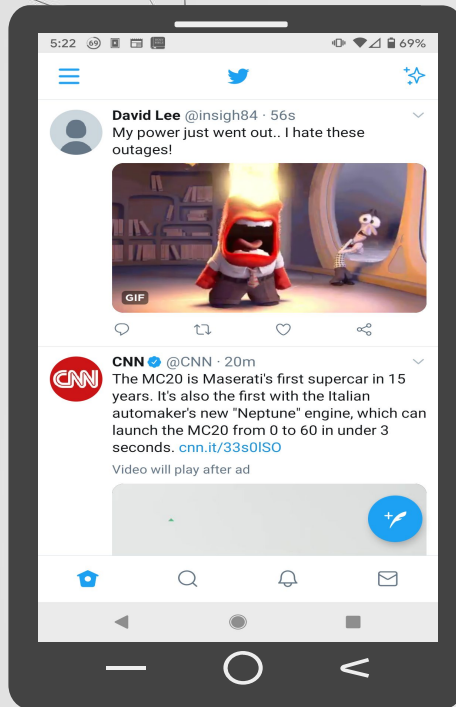
As tweets are coming in live we run those tweets through preprocessing to prepare for modeling. We then run that tweet through our 3 models and get a predicted output label



Advantages

Having a live bot will enable us to map user location in real time.

Implementation: Sneak Peek



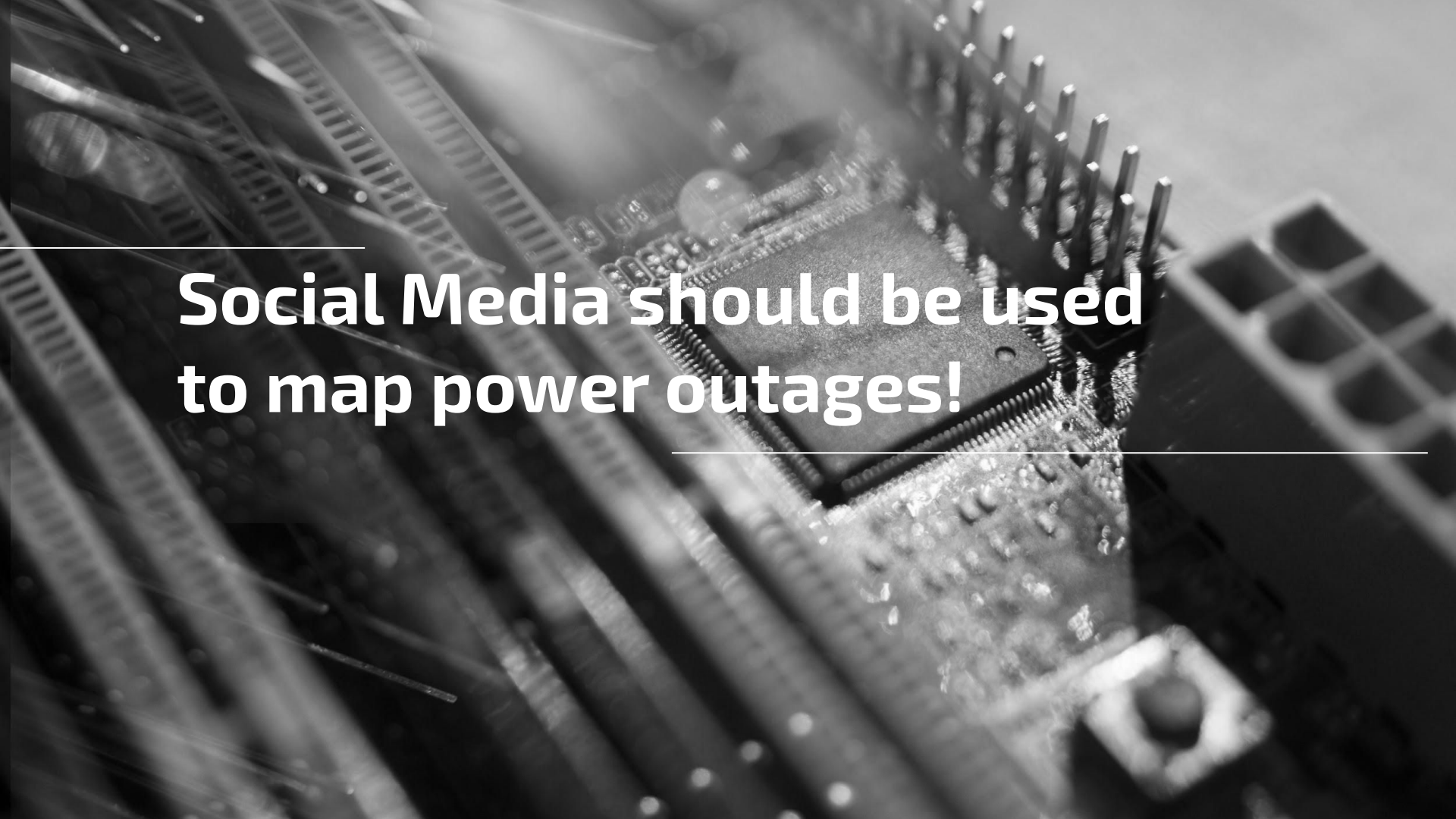
user	is_retweet	tweet	location	state	predict_lr	predict_nb	predict_rf	predict_rnn	labels
insigh84	False	my power just went out i hate these outages https t co rhpkpjvno	CT		1	1	1	0.9994	1
	False	due to a power outage we are closed this evening we			1	0	1	0.9667	0



06

Conclusions

David Lee ~ Data Scientist / Devops



**Social Media should be used
to map power outages!**

Conclusions: Findings

- Our initial modeling reveals that predicting tweets being a power outage is possible.
- Analysis on our model's performance identify optimization areas.
- Potential partnerships with Twitter may allow for better location information
- More scraped data can help improve our analysis

Conclusions: Challenges

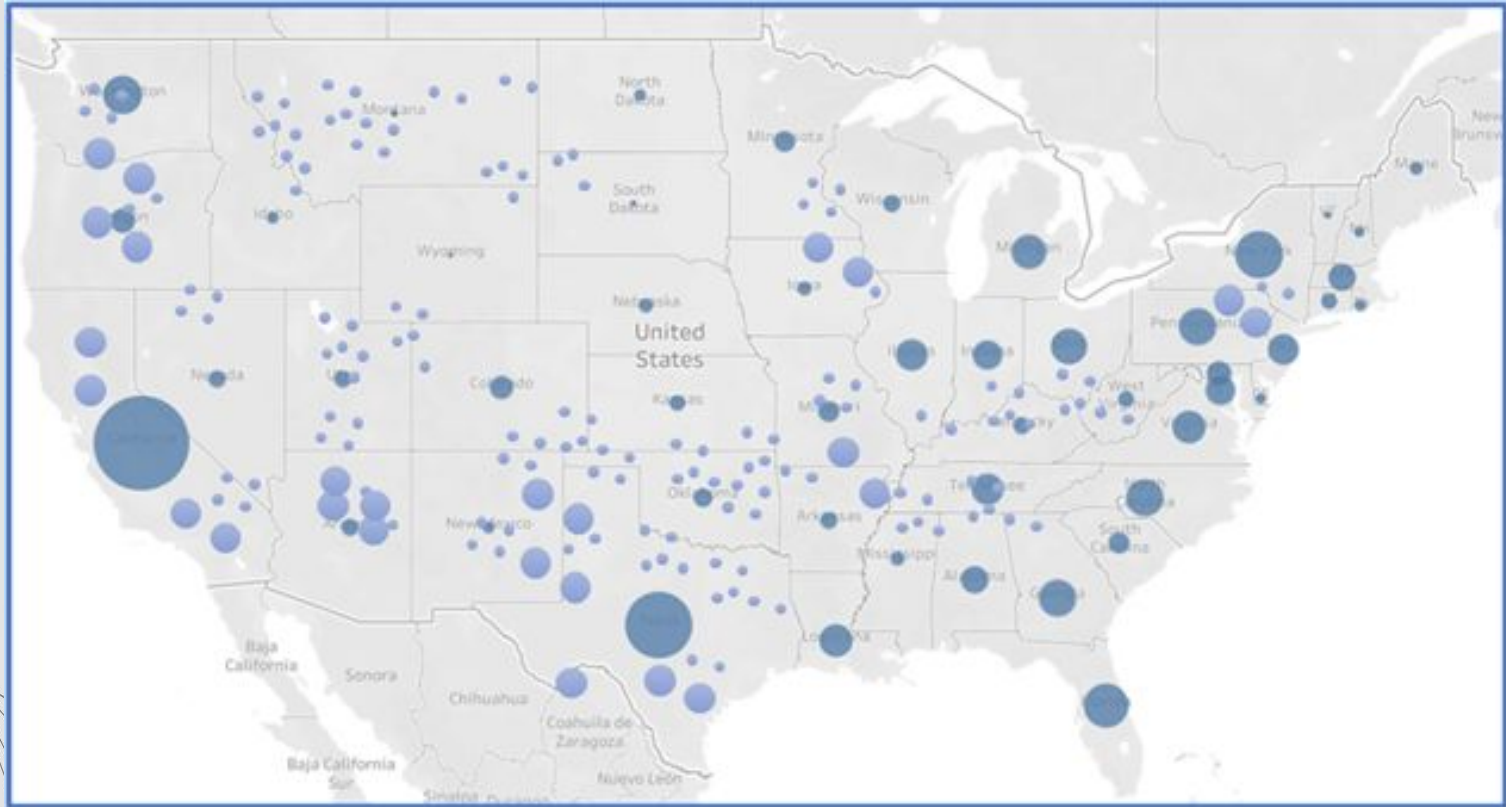
- Due to how Twitter's API is set up, we were not able to pull streams from certain locations or specific queries
- Initial scrapes were on specific phrases
- Model labels were not given a confidence parameter
- Criteria on scraped data for Model training can be improved.

07

Next Steps!



Next Steps: Mapping



Next Steps: Scalability

Refined Modeling Accuracy

To broaden scope and identify hot zones



Containerize with Docker and Deploy

Environment setup that can be replicated on all Systems to Scale up.



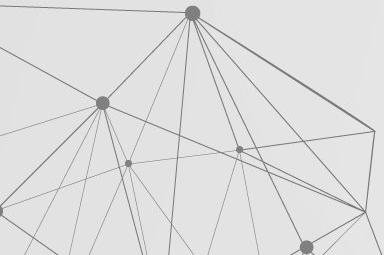
Geo Location Identification

To improve assessments by region

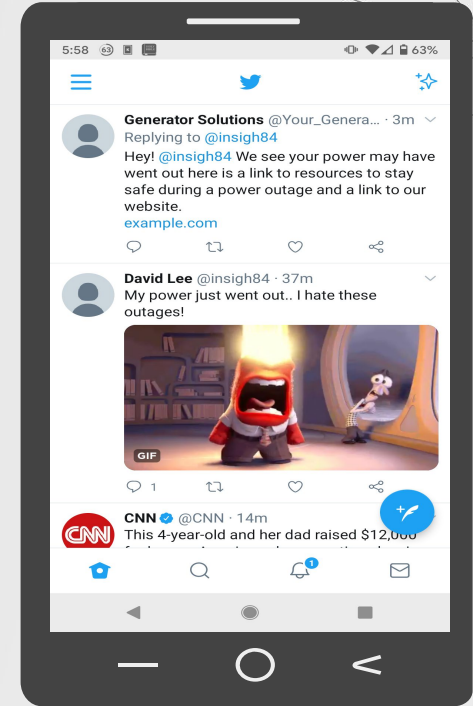
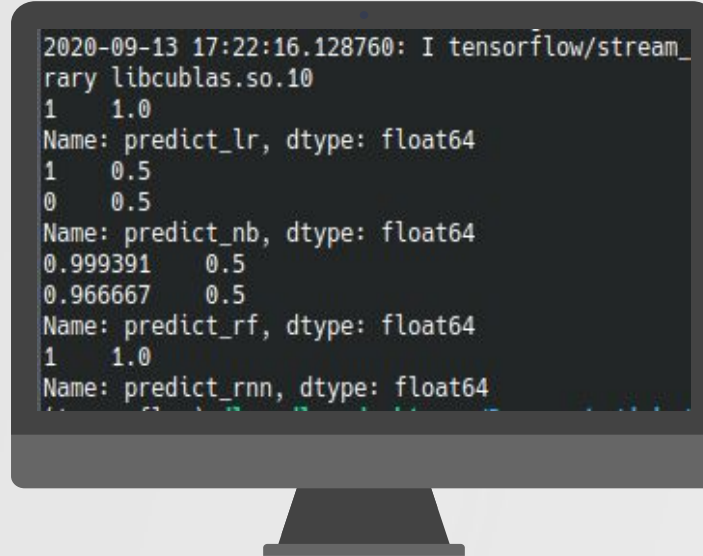
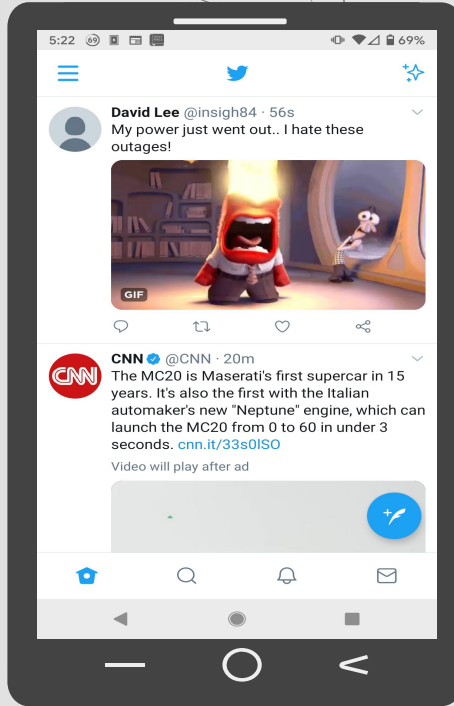


Multiple Twitter bots

Adding multiple services, and as backups



Next Steps: Services



Expanding reach, reassuring clients, and access to safety best practices.

user	is_retweet	tweet	location	state	predict_lr	predict_nb	predict_rf	predict_rnn	labels
insigh84	False	my power just went out i hate these outages https://t.co/rhpkpivno	CT		1	1	1	0.9994	1
	False	due to a power outage we are closed this evening we			1	0	1	0.9667	0



Next Steps: A two pronged vision

An early detection system based on user Tweets used for assessing necessary manpower and deploy faster restoration.

- **Energy Provider**

People who lose power can trust that their energy providers do their part in ensuring safety and a quick recovery.

- **Consumer Safety & Satisfaction**
- 



08

Q & A / THANKS

Does anyone have any questions?

<https://www.linkedin.com/in/nseither/>

<https://www.linkedin.com/in/billfu2020/>

<https://www.linkedin.com/in/aidancurley/>

<https://www.linkedin.com/in/daviddaeshinlee/>

CREDITS: This presentation template was created by **Slidesgo**, including icons by **Flaticon**, and infographics & images by **Freepik**.

Please keep this slide for attribution.