Mapping Situational Awareness

Project and Presentation by:
Alex Shon, Andy Zhang, Charles Spalding, Raymond Paller

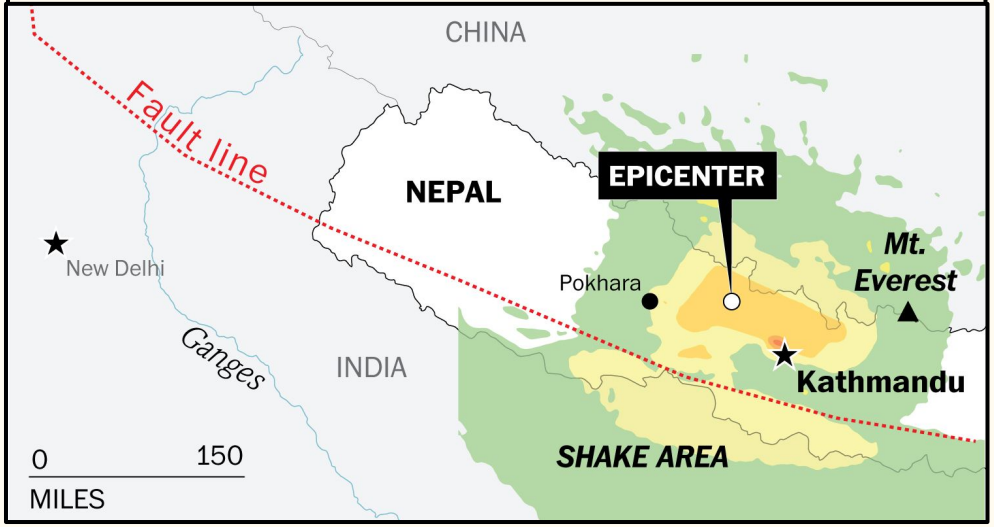
Social Media in Natural Disaster

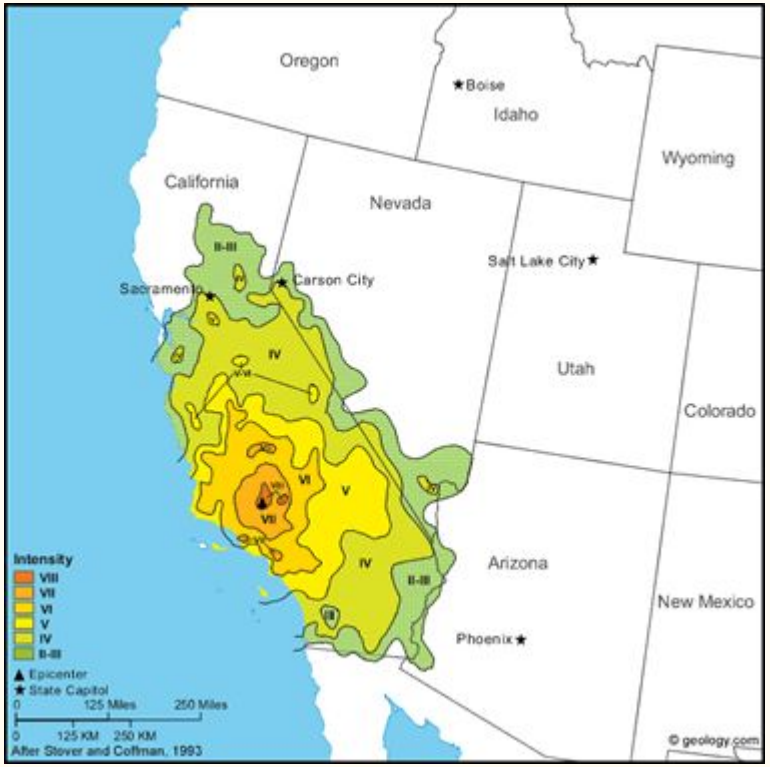
Ways social media is currently used in natural disaster response:

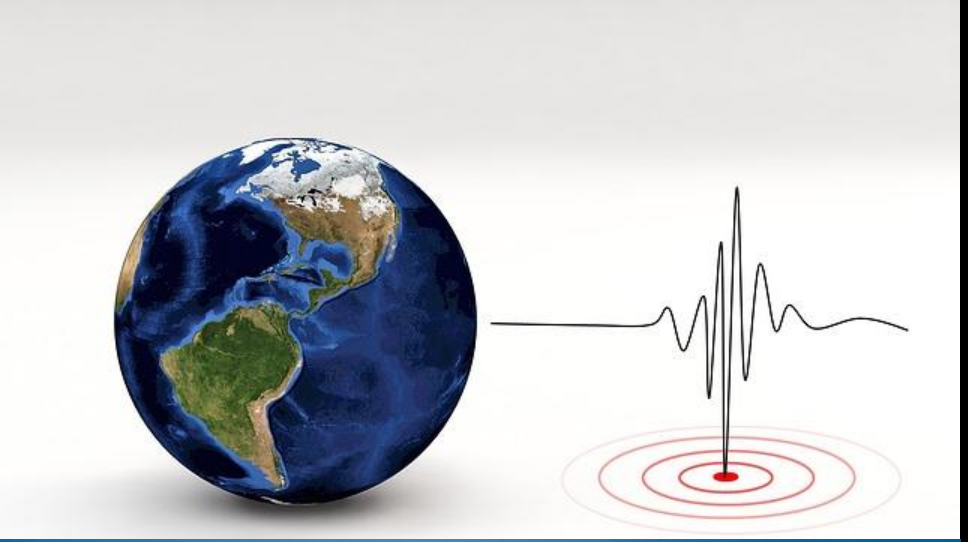
- Crowdsourcing funds
- Mobilize volunteers
- Identify if a person is safe or in crisis
- Counter misinformation quickly
- Replaces phone services in extremes

Identifying Crisis Areas

1. Identifying the threat or Impact
2. Initial Assessment
3. Response
4. Recovery

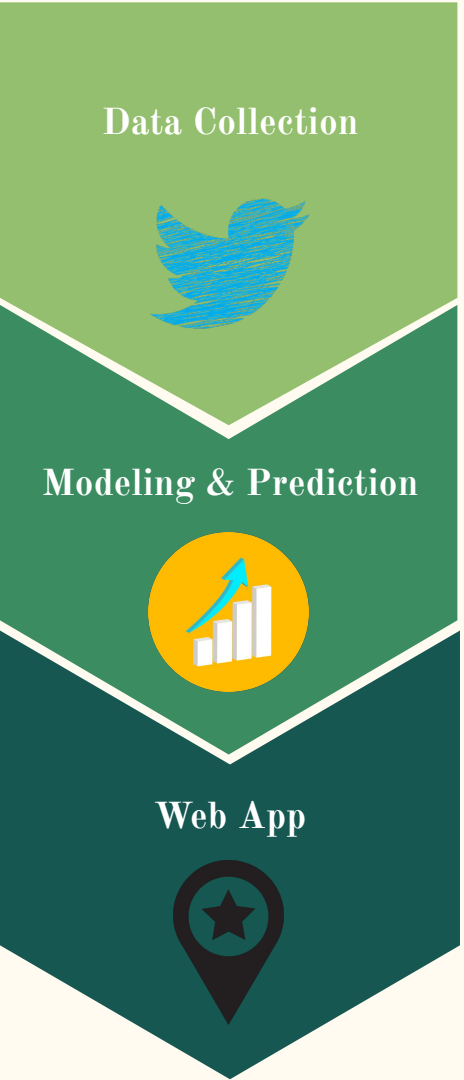






Problem Statement

Can social media posts be mapped to show the spread of situational awareness of an earthquake?



Roadmap & Roadblocks

Data Collection

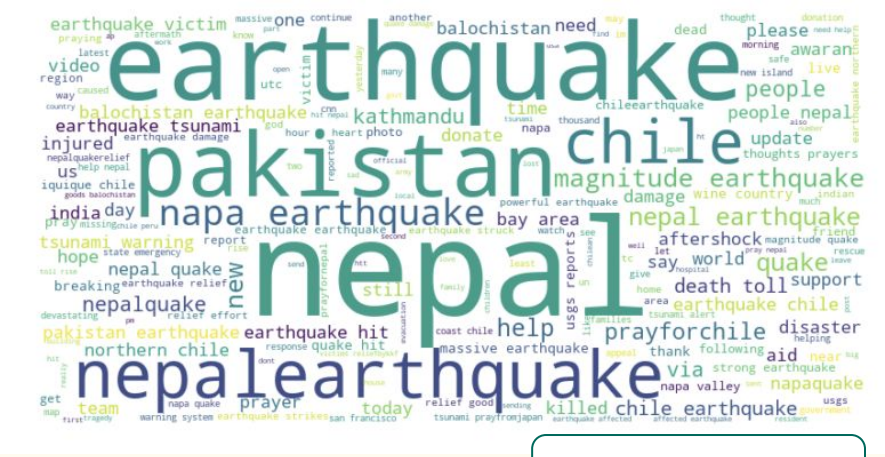
1. To manually label the training set is very time consuming not efficient. It was hard to find relevant data related to twitter and natural disaster.
2. According to our study, only 3% of tweets can be tracked from their geo-location, that makes the data collection process biased. But we still managed to get 600+ testing twitter.
3. Twitter API have a limit for 100 twitters per call and 180 call per 15 minutes. The speed of data collection is very limited

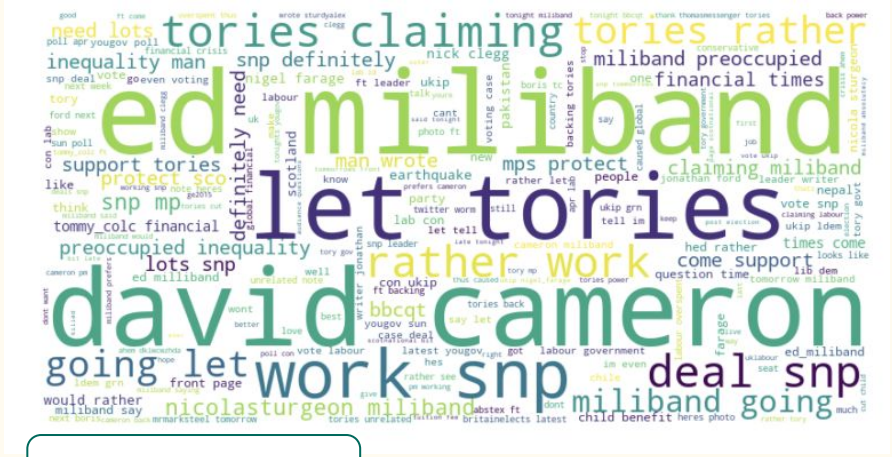
Modeling & Prediction

1. The data we collected are biased, working with biased data to get a solid result is a challenge

Result output

1. How to save data lively into SQL, and call data lively to the web application.
2. How to build interactive map to demonstrate our product

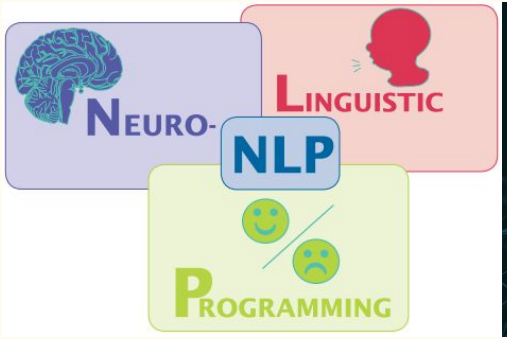




Relevant

Not Relevant

Collected 8517 training set labeled by volunteer from CrisisNLP*

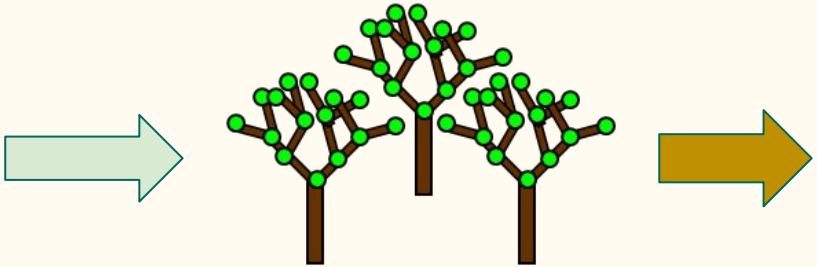


Preprocessing & Building Model

* Training set provided by: Muhammad Imran, Prasenjit Mitra, and Carlos Castillo: **Twitter as a Lifeline: Human-annotated Twitter Corpora for NLP of Crisis-related Messages.** In Proceedings of the 10th Language Resources and Evaluation Conference (LREC), pp. 1638-1643. May 2016, Portorož, Slovenia.

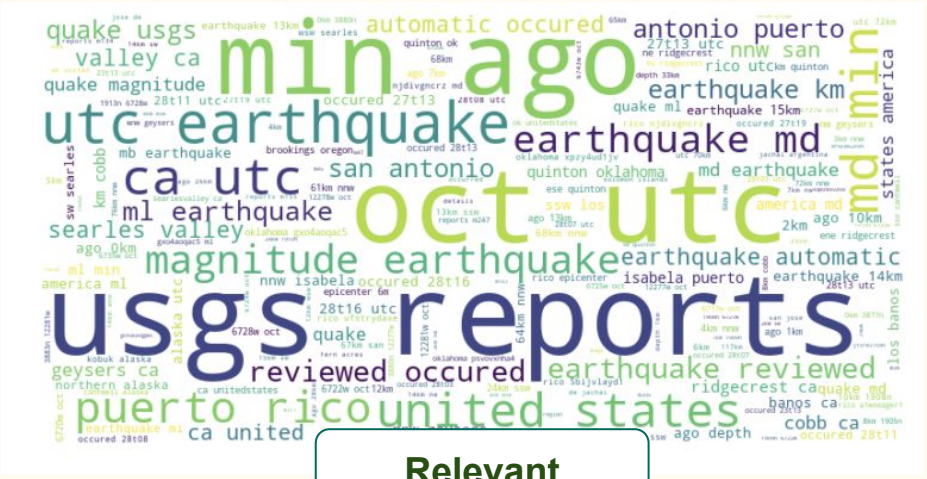
Model Performance

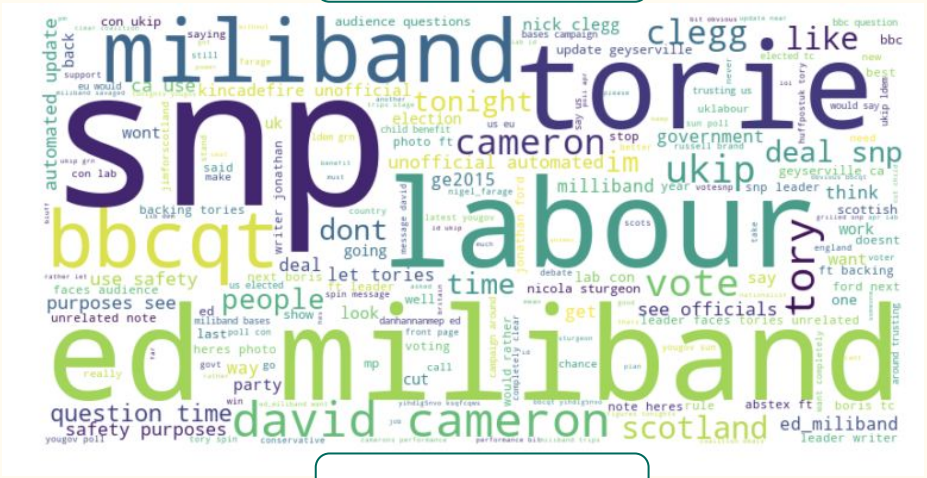
	Training	Testing		Training	Testing
Neural Network	96.44%	97.96%	Logistic Regression	96.44%	97.96%
Multinomial N.B	95.04%	97.37%	Random Forests	99.92%	98.55%
Gaussian N.B	93.85%	93.89%	Extra Trees	99.92%	97.88%
Gradient Boosting	94.70%	98.89%	Ada Boost	95.25%	98.81%



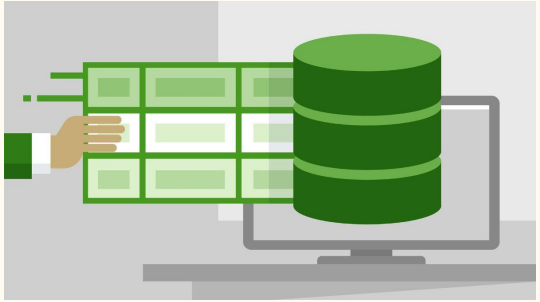
Random Forest Model

Using USGS live earthquake data to collect coordinates location. Searching twitter through twitter-python api* using the live coordinates for our testing data. Final size is 657





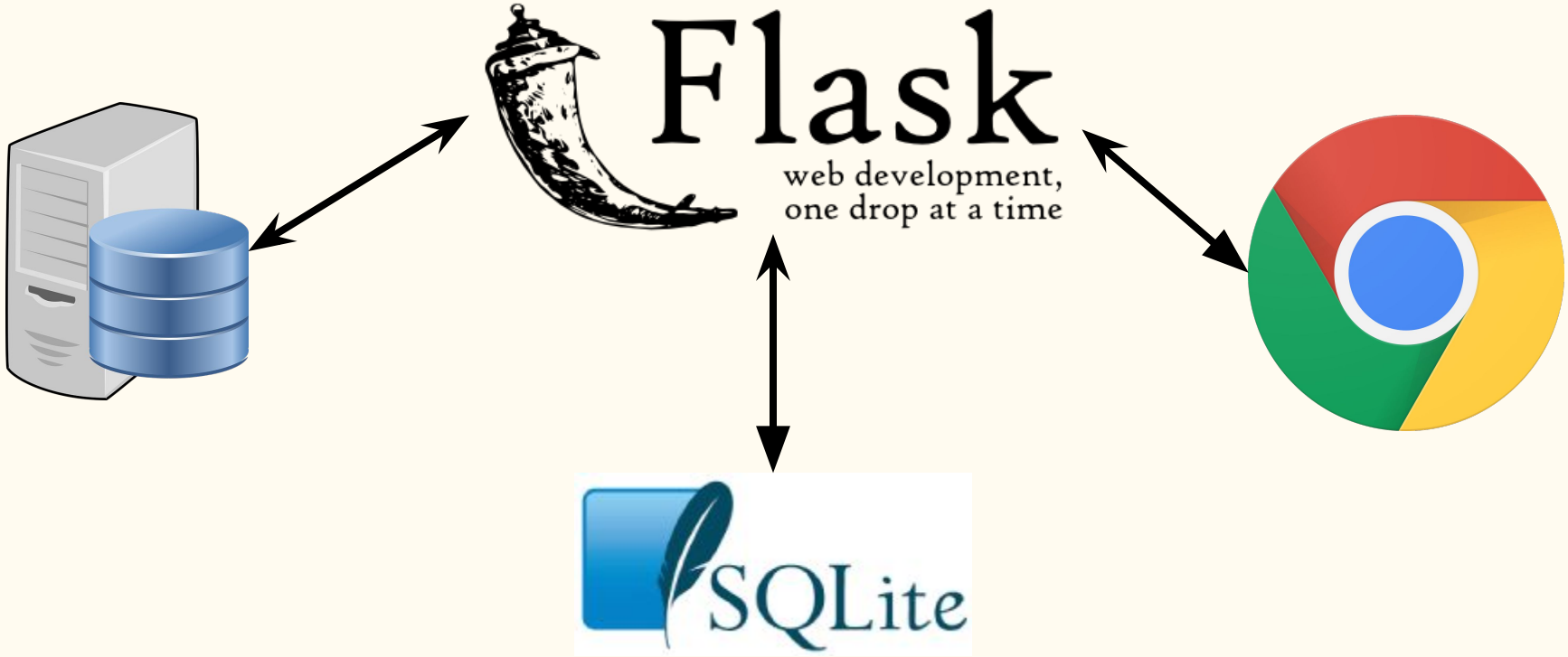
Prediction



Interact with SQL

Web Application

Backend of Web Application

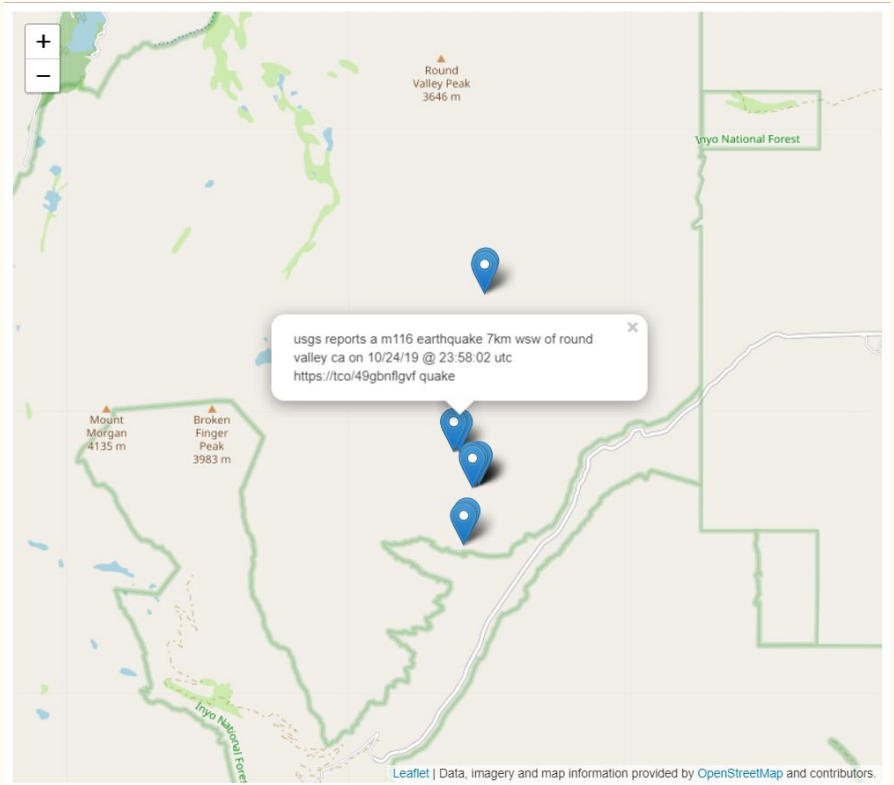


Web Application Target Audiences

Government & Volunteer Agencies

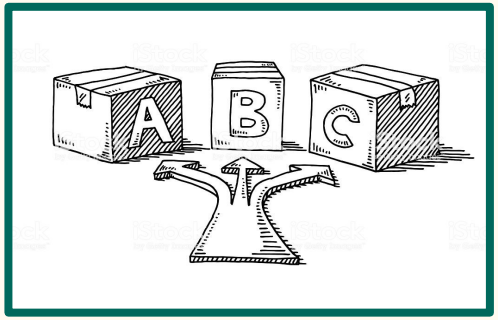
Private Companies

General Public



Conclusion

- Mapping situational awareness from social media posts can be done efficiently when the affected area is known
- Reducing the categories to classify allows for high accuracy of the model
- Successfully mapping social media is dependent on the involvement of the users of the platform



Next Steps

1. Multiclass Classification
   - Classification between urgent and not urgent would be critical for volunteers and rescue teams

2. Latency Reduction
   - Live tweet collection was solely based on government reports of an earthquake and updates about every 2.5 days

3. Filtering Misinformation
   - Misinformation is a major problem in natural disaster responses. Identifying malicious links and incorrect information spread should always be in consideration

Recommendations

1. Changes in platform popularity
   - A model should adjust to the popularity of a social media platform in the affected area

2. Non-text based Platforms
   - As technology and communication platforms develop and improve, there is an increasing presence on video based communications

3. Crisis Push Notifications
   - Social media Goliaths, like Twitter, ought to consider the option of casting geolocation easier in the event of a crisis