

# During a disaster, it's difficult to...

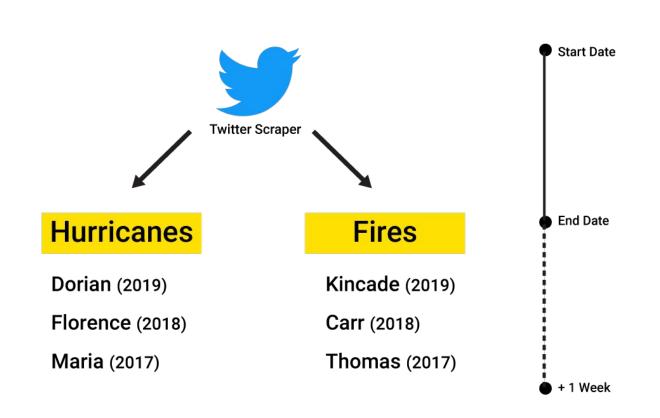
Locate all the needs

Triage and prioritize aid

Streamline communication

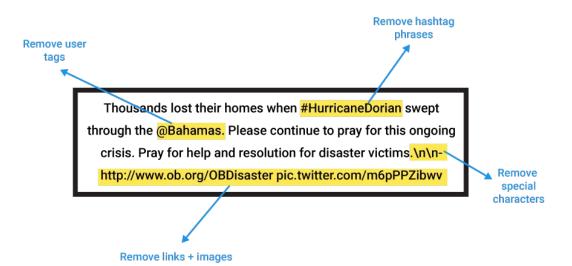
# need based on what they're tweeting?

How might we identify people in



## **Data Preparation**

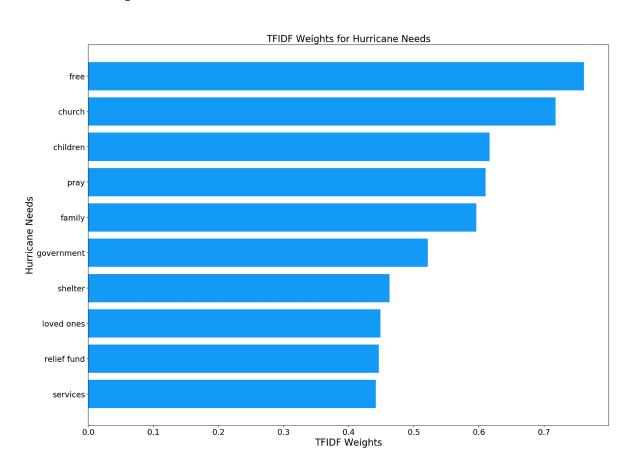
- 1. Clean Tweets with Regex
- 2. Filter for English Tweets
- 3. Manually label 1,000 Tweets to classify need



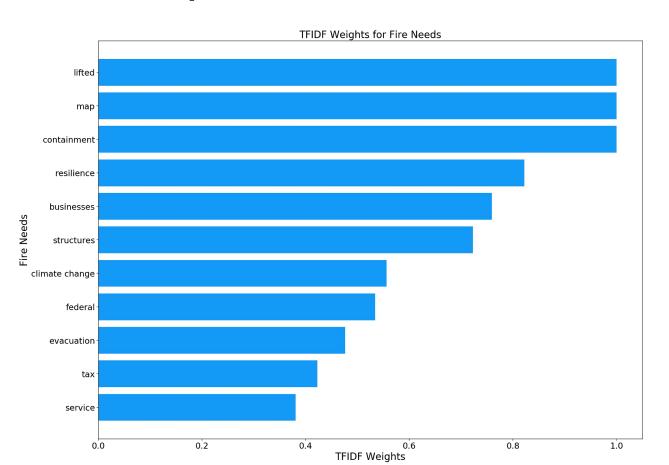
What kind of tweets are we

working with?

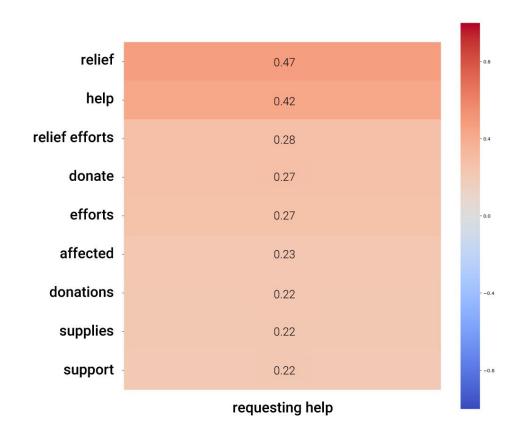
# **Specific Needs: Hurricane**



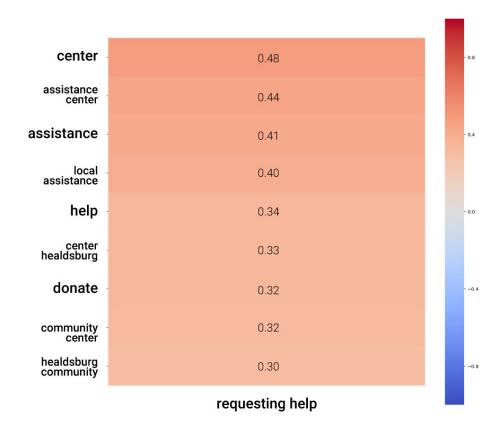
# **Specific Needs: Fire**



### **Correlations: Hurricane**



### **Correlations: Fire**



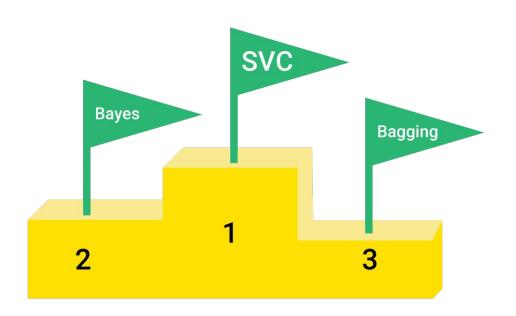
#### Limitations

- Location data
- Limited training on tweets
- Named disasters
- Internet trolls
- Twitter is not the go-to medium to request help
- Internet connectivity could be compromised

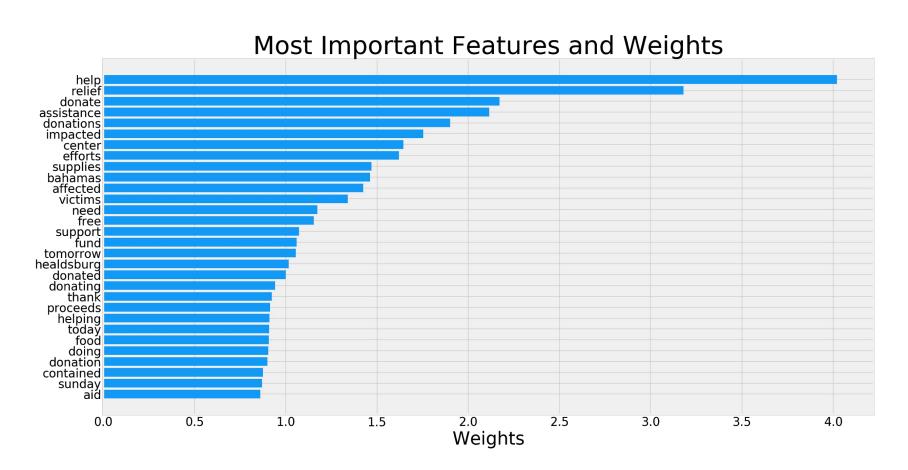
# Modeling

We noticed that there was a trade-off between adjusting an overfit model and optimizing for positive predictions.

For example, a bagging model with similar train & test scores returned a 0 (not requesting help) for nearly all our rows.



#### Insights from our best model: SVM & Count Vectorizer



## Insights from our best model:

Test Score 91.3 / Baseline 77

Support Vector Machine utilizing Count Vectorizer

SVM default kernel of RBF or Radial Basis Function, C=2

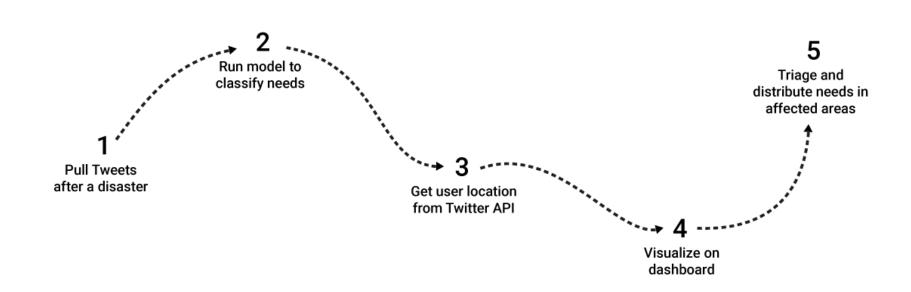
Maximum Features 2000

Min DF of 3

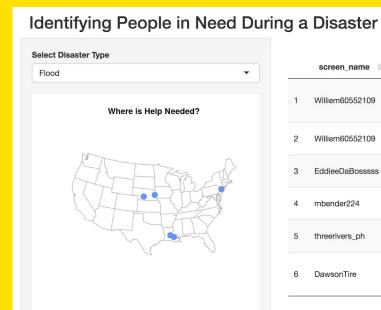
N-Gram range (1,1)

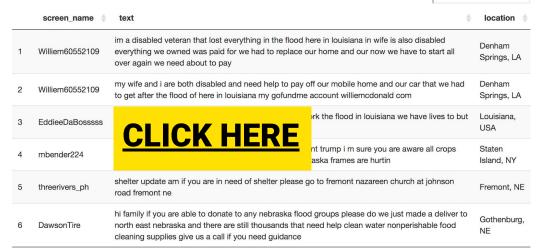
Stop Words: English

# **Creating a Relief Dashboard**



#### **Dashboard Demo**





Search:





Thank you!