Utilizing Twitter for Disaster Detection

By: Tofer Kim, Ritchie Kwan, Will Stecher

Problem Statement

- Traditional methods for alerting on disaster-related events like earthquakes and tsunamis rely on information derived from official sources (e.g. USGS).
- Our method utilizes social media, specifically Twitter activity, to identify these events and determine when an event first occurs.

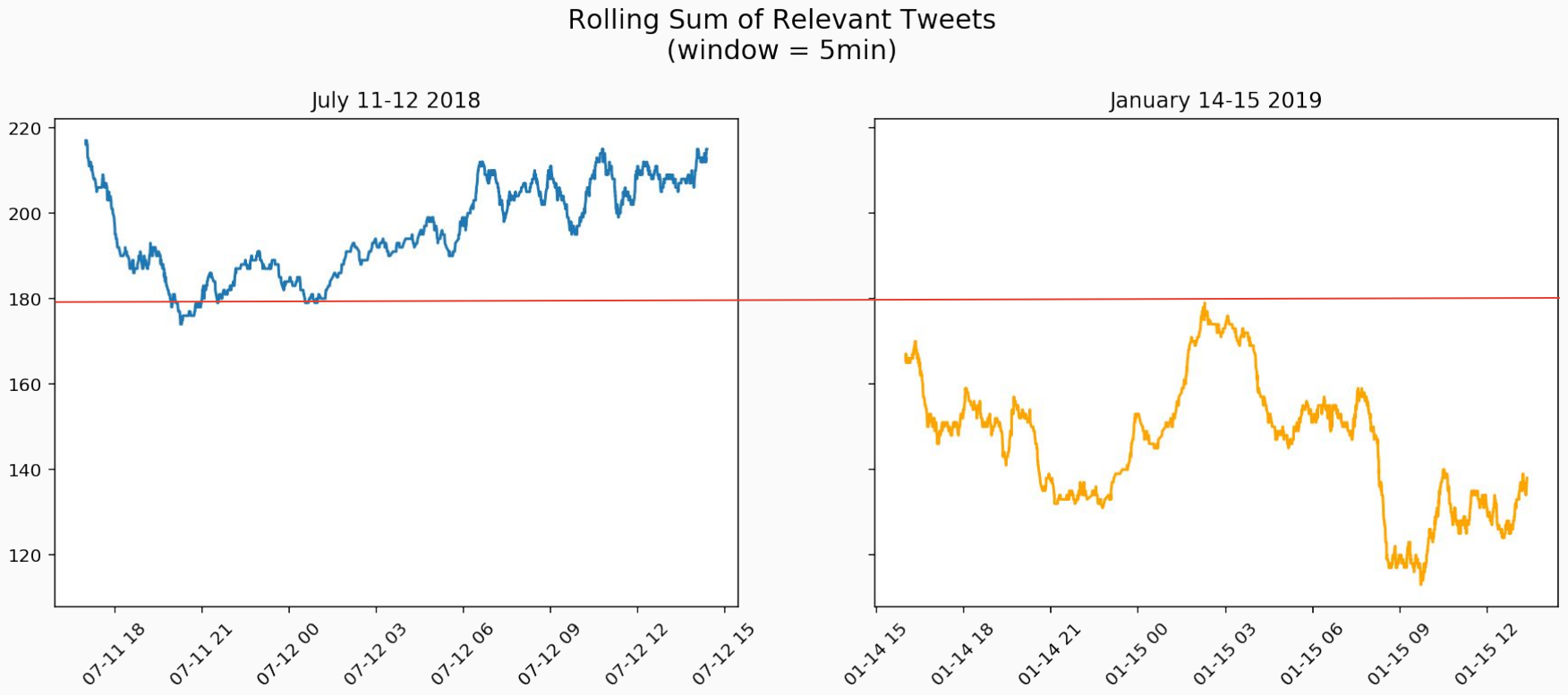
Investigating Fire Disasters

- Public dataset:
  - 10,877 disaster-related tweets, marked either “**Relevant**” or “**Not Relevant**”
- Collected tweets using keywords and date-ranges:
  - “**wildfire**” and “**forest fire**”
  - July 11-12, 2018
  - January 14-15, 2019

Classification Methods

- Determine if a tweet is “**relevant**” or “**not relevant**” to a disaster-related event
- Doc2Vec feature engineering
- Logistic Regression
- 57% baseline score

83.7% accuracy on unseen data



Threshold of ~180 relevant tweets in a five minute window differentiates our time periods.

Next Steps

- Build Doc2Vec vocabulary
  - Larger datasets, tune hyperparameters, rank urgency of tweets
- Identify different types of disasters & where they are occurring
  - Capture and implement geolocation data
  - Determine thresholds for distinct natural disasters and population sizes

Thank you!

Sources

1. Training dataset: https://www.figure-eight.com/data-for-everyone/
2. Doc2Vec and Classification procedures: https://towardsdatascience.com/multi-class-text-classification-with-doc2vec-logistic-regression-9da9947b43f4
3. Scraping Twitter: https://github.com/Jefferson-Henrique/GetOldTweets-python