

New Light Technologies: Using Social Media to Map Power Outages Across the United States



Matthew Malone, Danielle Medellin, Omar Smiley

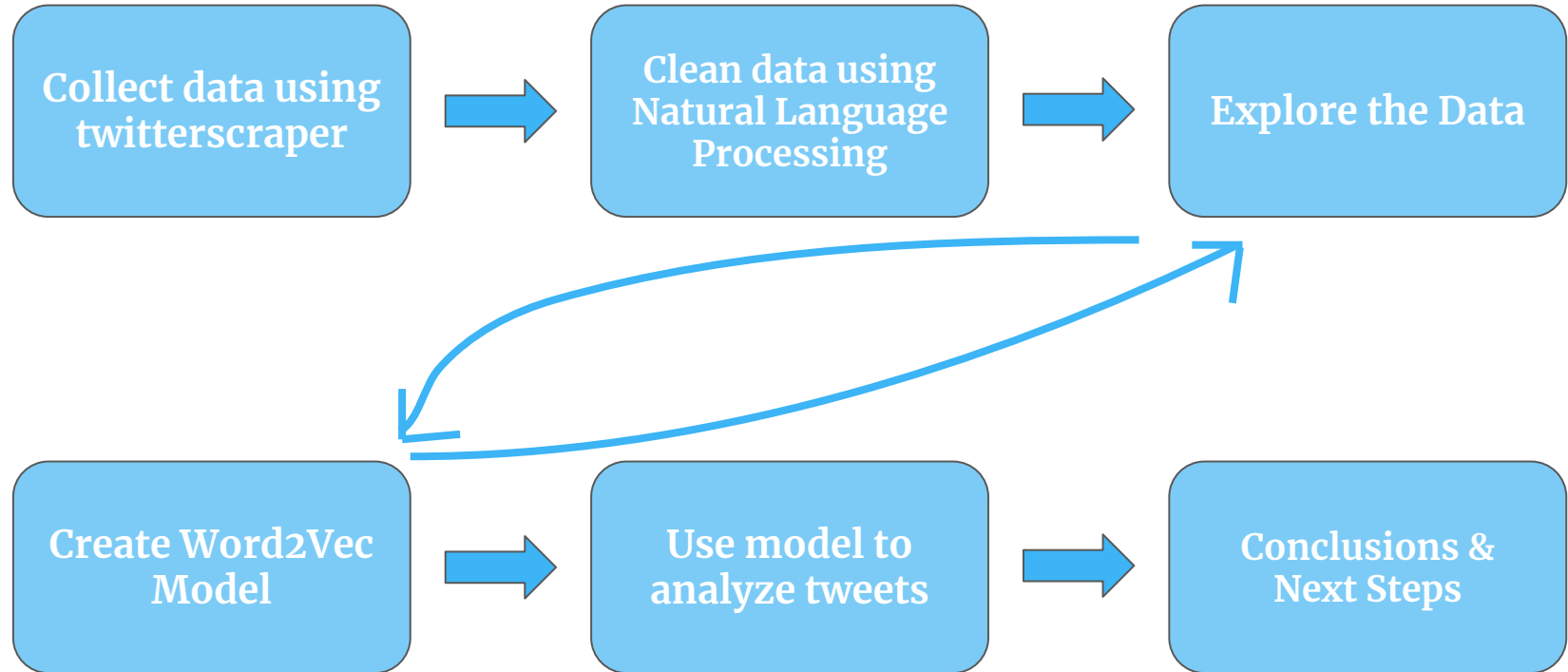
Problem Statement



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.

This tool will utilize posts on social media, specifically **Twitter**, to identify “hot spots” of concern and areas suffering from power outages. Following an event, the tool will scan tweets to identify areas likely to be suffering from power outage.

Workflow & Methodology



Data Collection & Cleaning



Data Collection:

- Utilized TwitterScraper to scrape tweets from Twitter
- Used 34 keywords associated with ***power outages*** to pull tweets dating back to 2016
- Targeted the five states with the most power outages (CA, TX, OH, MI, NY)
- Top three most populous cities were selected from each state

Data Cleaning:

- Pulled tweet, username, date/time, location
- Cleaned tweets & usernames of non-alphabetic characters, links, emojis
- Concatenated usernames & tweets into one string
- During analysis, we used stopwords to more easily capture the sentiment of tweets

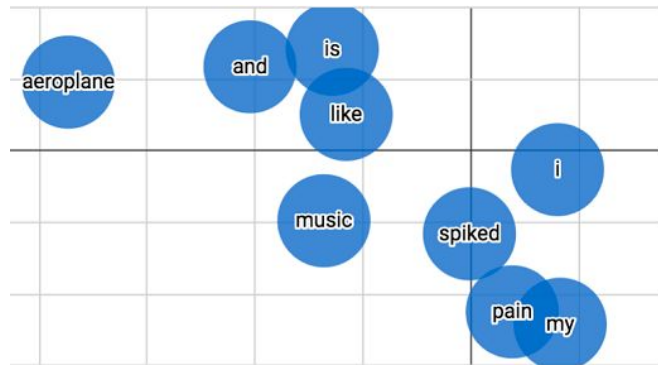
Word2Vec



Word2Vec is a two-layer neural net that processes text by “vectorizing” words. Its input is a text corpus and its output is a set of vectors: feature vectors that represent words in that corpus.

The model looks at the *inherent structure* for how people talk about power outages.

- Corpus created from the cleaned tweets found through our search terms
- Word2Vec was utilized to identify the words most associated with our Key Search Terms
- The Model Returns a *cosine similarity* for each pair of words which indicates how similar or associated a pair of words is.
 - $1 \rightarrow$ words completely the same
 - $0 \rightarrow$ word gives no information about other word
 - $-1 \rightarrow$ words exact opposites



Word2Vec Results



Positive Association

Outages - Customers 0.789, Crews 0.767, Restoration 0.763, Areas 0.761, Without 0.758

Interruption - Centerpoint 0.932, Report 0.928, Customer 0.923, Consumers 0.919, Utilities 0.918

ConED - Affect 0.975, Warnings 0.962, Capacity 0.981, Thousand 0.980, Disabled 0.977

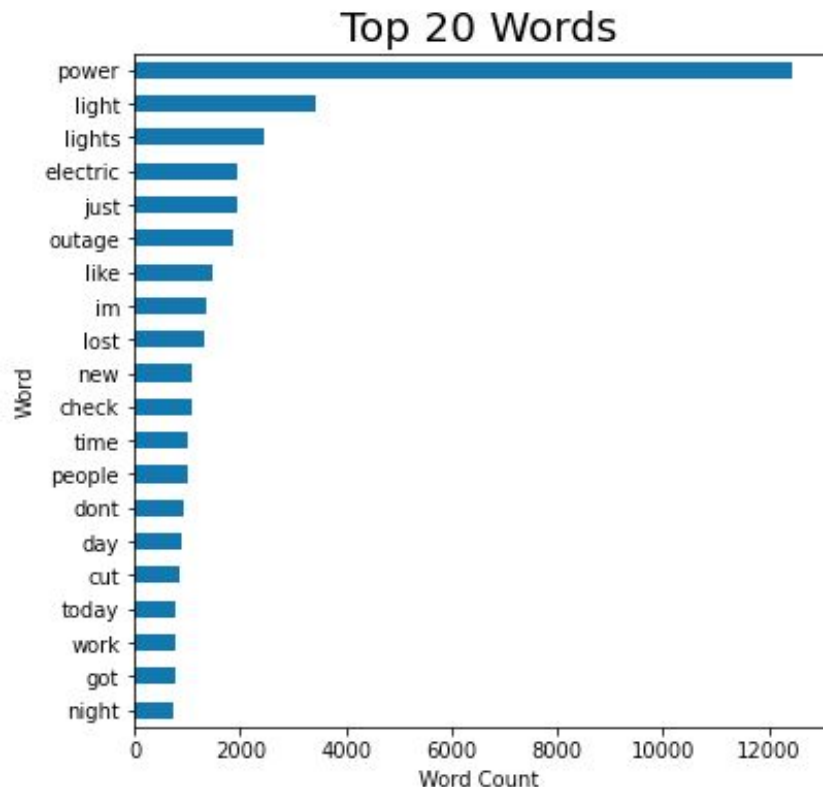
Neutral Association

Outages - Cuts 0.098, Family 0.086, Due 0.082, Fire 0.078

Interruption - Looks 0.181, id 0.178, lit 0.177, air 0.182

ConED - Kelly 0.173, Going 0.164, Ok 0.158, Keith 0.155, massage 0.127

Exploratory Data Analysis



- CountVectorizer Used to Identify The Most Frequently Occuring Words In Our Search Results
- Analyzed tweets for potential misclassification (i.e. Blackout, Light, Electric, etc.)
- Analyzed The Distribution of Word Sentiments to Establish A Sentiment Threshold For Evaluation

Outage Sentiment



Outage Sentiment Range: 0 - 1



Carmen Garsia
@CarmenGarsia

Here at the movies to see [#lightsout](#) They weren't kidding. Where are... [instagram.com/p/BIInvFbDBj5Rm](https://www.instagram.com/p/BIInvFbDBj5Rm)



KENS 5 
@KENS5

STORM UPDATE: CPS reports there are currently 629 power outages affecting 39,666 customers in the San Antonio area. [on.kens5.com/2lcirn0](https://www.kens5.com/2lcirn0)

Not Associated with
Power Outages

0.0

Associated with
Power Outages

1.0



Outage Sentiment

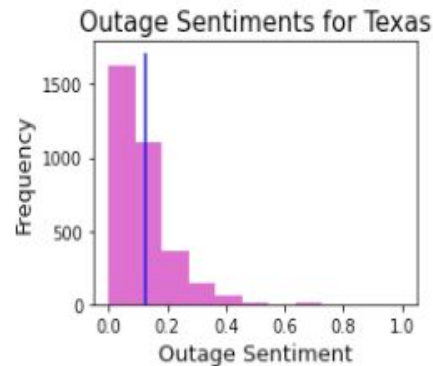
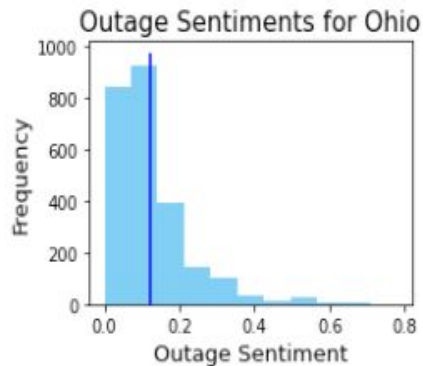
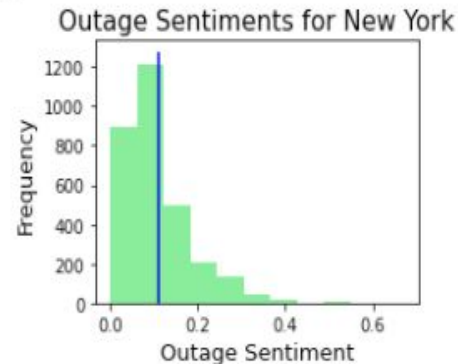
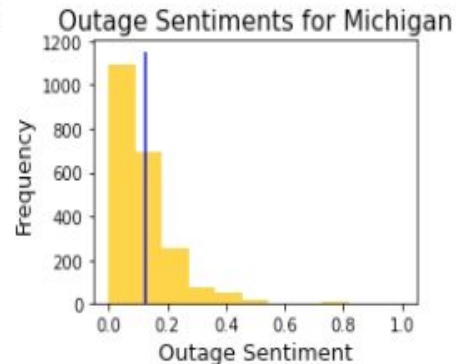
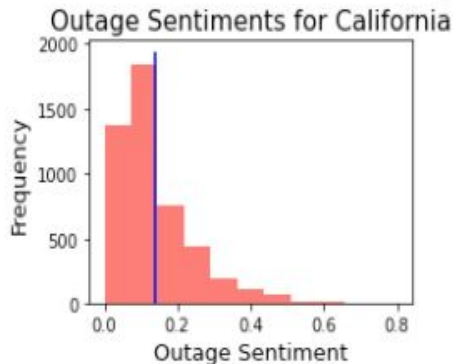


Mean: 0.124

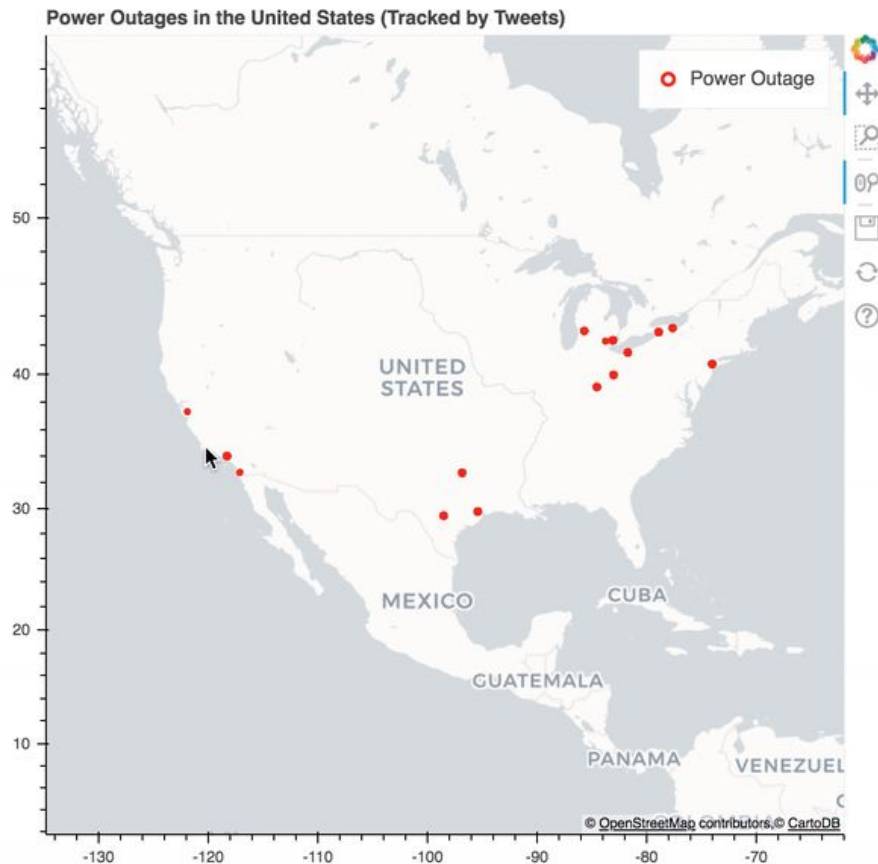
St. Dev: 0.097

Threshold: 0.3

Histograms of Outage Sentiment by State



Geovisualization - Bokeh



Conclusions & Next Step



First step towards an application/tool that will:

- Screen tweets in real time
- Determine outage sentiment
- Visualize where potential power outages are on an interactive map

Next Steps:

- Use Twitter API for a cleaner pull of tweets and ***accurate longitude & latitude***
- Confirm power outages by cross referencing with either: weather data or historical power outage data
- Decrease bias in our model introduced by our search terms
- Insert a time slider into visual to show historical power outage

Resources



- Top 5 U.S. States For Power Outages, Generac (Feb, 2018),
<https://www.generac.com/be-prepared/power-outages/top-5-states-where-power-outage-occur>
- Gensim Word2Vec Tutorial, Kavita Ganesan, <https://kavita-ganesan.com/gensim-word2vec-tutorial-starter-code/#.Xr3yDBNKhTa>
- A Beginner's Guide to Word2Vec and Neural Word Embeddings, <https://pathmind.com/wiki/word2vec>
- Lesson-NLP-i, Matt Brems <https://git.generalassemb.ly/DSI-US-11/5.03-lesson-nlp-i>
- Lesson Word Vectors, Matt Brems, <https://git.generalassemb.ly/DSI-US-11/8.07-lesson-word-vectors>
- Can We Use Social Media to Locate Legitimate Power Outages?, Jen Hill,(Aprl, 2017),
<https://towardsdatascience.com/can-we-use-social-media-to-locate-legitimate-power-outages-7b7409708447>
- Adam Cohen, Twitter Scraper <https://github.com/Adam395/Twitter-Scraper>
- Elliptical (true) Mercator Projection, https://wiki.openstreetmap.org/wiki/Mercator#Elliptical_.28true.29_Mercator_Projection