Fall 2018

An Evaluation of Geotagged Twitter Data during Hurricane Irma using Sentiment Analysis and Topic Modeling for Disaster Resilience

Ike Robert Vayansky
Coastal Carolina University

AN EVALUATION OF GEOTAGGED TWITTER DATA DURING HURRICANE IRMA USING SENTIMENT ANALYSIS AND TOPIC MODELING FOR DISASTER RESILIENCE

By

Ike Vayansky

Submitted in Partial Fulfillment of the Requirements for the Degree of Master of Science in Information Systems Technology with a Concentration in Information Security and Data Analytics in the College of Science

2018

Dr. Sathish Kumar
Major Professor

Dr. Zhenglong Li
Committee Member

Dr. William Jones
Committee Member

Dr. Jean French
Chair, Department of CS

Dr. Michael Roberts
Dean, College of Science

Dedication

I would like to thank my girlfriend Allyson for putting up with me and my antics as I completed my Master's Thesis.

Acknowledgements

I would like to acknowledge my advisor Dr. Sathish Kumar for his support, patience, and guidance. Having known Dr. Kumar since my first introductory programming class when I first started in the Computer Science Department at. Coastal Carolina University. Dr. Kumar has been a major influence on my knowledge gained from the university. Next, I would thank Dr. Zhenglong Li from the University of South Carolina for helping with the dataset, figures, and edits included in the thesis. Dr. William Jones has been a major help since I joined the Computer Science Department and was always available to answer any questions, give guidance, and save me from many headaches. Dr. Jean French is a professor whom I interacted with mostly as an under-grad but all the classes I had with her helped me complete this thesis. Specifically, when she taught me to write scientific papers. Dr. Michael Roberts has been a major influence in my decision to move towards computing sciences rather than other sciences. I knew I enjoyed computing, robotics, and design ever since I created an underwater ROV (remotely operated vehicle) under his supervision. I cannot thank Dr. Michael Murphy for being a caring professor, guidance, and support since I have known him. That does not mean that other professors did not assist me, and I could probably list them all. I also thank Professor Cory Nance, Professor Clint Fuchs, Dr. Crystal Cox, Professor Casselman, and Dr. Stephen Sheel for being a teacher and helping me find my way through the lack of guidance I had. I would also like to thank Biraj Dahal specifically and members of the C-Surf REU program while I had the chance to be a part of it. Biraj was a major help in answering questions and everyone in the C-Surf program taught me something important.

Abstract

Disasters require quick response times, thought-out preparations, overall community, and government support. These efforts will ensure prevention of loss of life and reduce possible damages. The United States has been battered by multiple major hurricanes in the recent years and multiple avenues of disaster response efforts were being tested. Hurricane Irma can be recognized as the most popular hurricane in terms of social media attention. Irma made landfall in Florida as a Category 4 storm and preparation measures taken were intensive thus providing a good measure to evaluate in terms of efficacy. The effectiveness of the response methods utilized are evaluated using Twitter data that was collected from September 1st to September 16th, 2017. About 221,598 geotagged tweets were analyzed using sentiment analysis, text visualization, and exploratory analysis. The objective of this research is to establish an observable pattern regarding sentiment trends over the progression of the storm and produce a viable set of topic models for its totality. The study contributed to the literature by identifying which topics and keywords were most frequently used in tweets through sentiment analysis and topic modeling to determine what resources or concerns were most significant within a region during the hurricane Irma. The results from this study demonstrate that the sentiment analysis can measure people's emotions during the natural disaster, which the authorities can use to limit the damage and effectively recover from the disaster. In this work, we have also reviewed the related works from the text/sentiment analysis, social media analysis from hurricanes/disaster perspective. This research can be further improved by incorporating sentiment analysis methods for classifying emoticons and non-textual components such as videos or images.

Table of Contents

List of Figures

List of Symbols and/or Abbreviations

LDA	Latent Dirichlet Allocation
P	Probability (in Bayes' theorem)
H	Given hypothesis (in Bayes' theorem)
E	Observed evidence (in Bayes' theorem)
|	Given that, statistical probability
α	Parameter of the Dirichlet prior, per-document topic distribution. Also refers to the topic density distribution in a symmetric LDA distribution.
β	Parameter of the Dirichlet prior, per-topic word distribution
K	The number of topics, set value
V	The number of words in the vocabulary, determined by corpus (not displayed in plate notation)
M	The number of documents in the corpus
N	The total number of words over all documents
N_d	The number of words in a document, *d*
θ	A matrix consisting of rows derived from documents and columns derived from topics

	based on the corpus. Will have a dimension of M by K.
θ_{dk}	A single element of the θ matrix representing the probability of a topic k occurring within document d. This value can be no greater than 1.
φ	A matrix consisting of rows derived from topics and columns derived from words based on the corpus. Will have a dimension of K by N.
φ_{kw}	A single element of the φ matrix representing the probability of a word w occurring within a topic k. This value can be no greater than 1.
z	The identity of the topic of all words in all documents.
z_{dw}	The topic of word w in document d.
w	The identity of all words in all documents
w_{dw}	The identity of word w in document d.
BoW	Bag-of-Words
LSA	Latent Sentiment Analysis
PLSA	Probabilistic Latent Sentiment Analysis
NMF	Non-negative Matrix Factorization
ET-LDA	Event and Tweet Latent Dirichlet Analysis
API	Application Programming Interface
NLP	Natural Language Processing
NLTK	Natural Language Processing Toolkit
URL	Uniform Resource Locator. This represents the address of a specific webpage or document on the Interenet.
RT	Retweet, an outdated method of republishing a tweet posted by another twitter user.
analysis.sentiment.polarity	The call name for the calculated polarity value of a given document.

analysis.sentiment.subjectivity	The call name for the calculated subjectivity of a given document
c_v	Coherence value of a topic model.
AL	Alabama
FL	Florida
SC	South Carolina
GA	Georgia
NOAA	The National Ocean and Atmospheric Administration
NHC	National Hurricane Center

1.0 Introduction

Natural disasters have had profound effects on communities since the beginning of recorded history (Caragea et al., 2014). The geographical and topological effects of such events have been recorded through striking peaks, deep craters, islands, and layers of stone. The influence of these drastic occurrences on the lives of ancient humans is a little harder to decipher (Kozák, and V. Čermák, 2010). Preserving events like the eruption of Mount Vesuvius left detailed snapshots of ancient cities in peril, but more commonly encountered disasters left little evidence behind (Pescatore et al., 2001). More recent centuries have left detailed records of the impact and effects of earthquakes, hurricanes, tornadoes, and other disasters through recordkeeping and literature (R. Perry et al., 2001). However, much of the information about how people were impacted and what measures were taken during such events were lost even if they were recorded initially. The primary focus during these years was simply detailing the factual aspects of a disaster—where it happened, when it happened, and how many people died (R. Perry et al., 2001).

Since the technology revolution of the 1990's, collecting data and keeping records transformed from a time-consuming and inefficient task to a nearly impervious and limitless resource. Not only did traditional data become easy to gather, the collection of whole new varieties of data became possible as well. The emergence social networking and public forum sites started during the late 90's and have been rapidly gaining popularity since 2005, especially among younger individuals. Today, nearly every corner of the world is digitally connected by the Internet and social media. Social networking sites such as Facebook, Instagram, Snapchat and Twitter rose above the crowd and became leaders of their field, both as a platform to connect people and as a platform to discover things from others (J.

Phua et al., 2017). This infinite flow of virtual documentation and data from other social media sites presents scientists with endless opportunities to study how individuals and communities respond as a disaster unfolds (Bik and Goldtein, 2013). Around 500 million tweets have been posted every day since 2013, creating a massive database of real time reaction data for scientists to work with (Kirkorian, 2013). Twitter provides both data scientists and organizations with an invaluable source of real-time and historical social data directly from millions of individuals. Data includes the classification of topics using hash tags, user interaction data, and measures of public response to a tweet. The number of tweets posted each day is so massive that accurately gauging public response to highly discussed topics such as a hurricane or natural disaster is nearly impossible without the assistance of automated data analysis (S. Kumar et al., 2011).

Collected data can be used for a variety of purposes like marketing and product development. However, many researchers are using this data for more humanistic purposes (S. Kumar et al., 2011). Examples include understanding how people react to events and how disaster relief can be optimized to better serve victims of these events. One example of such research is the analysis of social media users' evacuation compliance during natural disasters, such as hurricanes (Martin, Li, Cutter, 2017). Unlike many other natural disasters such as earthquakes, hurricanes often have trajectories which can be accurately estimated several days ahead of actual landfall, making them excellent subjects for evacuation compliance because residents have significantly more notice time. Additionally, hurricanes often effect many people and have damaging effects which are evenly distributed over an area setting them apart from smaller more unpredictable events such as tornados. Many social media services now include geotags with user content which provides their precise geographical location when the post was made; these values can be used to map where users were before, during, and after a hurricane has come through.

Existing research shows that residents are cooperative with community and government preparation efforts, but this does not necessarily indicate that these preparations are efficient or pragmatic (P. Jaegar et al., 2007). To evaluate the functionality of disaster preparations and relief, it is essential to analyze the emotions or sentiments of users within the impact area during the storm (G. Beigi et al., 2016). These methods rely on implicit characteristics, such as phrasing, word choice, and punctuation, to calculate a measurable sentiment value from text-based posts. A high occurrence of positive scores would suggest public approval of relief actions, while a predominance of negative scores would indicate discontent with the handling of disaster efforts. Additionally, such a measurement could identify regions with abnormal spikes in emotional state, which could possibly assist rescue crews in locating areas which need more aid or medical attention. This research model was applied to twitter data collected during Hurricane Sandy, and although negative tweets were shown to consistently cluster in closer proximity to the storm, both positive and negative clustering tendencies were shown to increase at the points of Sandy's maximum impact and then disperse over the following days (C. Caragea et al., 2014; T. Shelton et al., 2014). The research did not reveal an overwhelming trend towards either positive or negative scores, but it did show that such a concept is feasible (K. Lachlan et al., 2014, H. Dong et al., 2013).

One of the challenges of conducting studies of social media data is that the data of the greatest interest is often text, which requires significant pre-processing to analyze properly. Working with such difficult data can be time and resource consuming, and similar work with text data has shown that it is markedly more problematic than analyzing numerical or categorical datasets (A. Bifet and E. Frank, 2010). These problems are only amplified when using text data from social media, which is additionally limited by character or word counts. The frequency of grammatical and spelling errors, ambiguous phrases or slang, non-text characters, and obtuse language concepts such as irony, sarcasm, and explicit language is much greater in social media users' posts than in more traditional text subjects such as literature and speech (A. Hassan et al., 2013). However, by proper pre-processing of the twitter data the effects of such complications can be greatly reduced. If properly organized and processed, social media data can provide an astounding amount of insight into public opinion, viral trends, and patterns which are usually difficult to evaluate.

The purpose of this study is to evaluate the twitter data from counties with high twitter activity during Hurricane Irma. Sentiment Analysis is used for studying how individual regions may have reacted differently compared to others and during each stages of the disaster. This data is evaluated to determine what issues were of greatest concern, whether response efforts generated a positive or negative sentiment, and what sentiment was most prevalent regarding both response efforts during the storm's progression. By determining the answers to these questions, the response efforts to future hurricanes and disasters could be optimized to better suit the needs of specific areas and overall improve the distribution of services during these events. We also seek to evaluate the underlying topics within the twitter data collected over this period to help form a concept of what issues were of the greatest concern for effected individuals and what topics attracted the most attention.

To summarize, the following are the contributions of our work: (a). Using twitter data collected from specific regions of the United States during the severe storm Hurricane

Irma, this research seeks to establish an observable pattern regarding sentiment trends over the progression of the storm. (b) The study also contributes by identifying which topics and keywords were most frequently used in tweets through sentiment analysis, and using this information to determine what resources or concerns were most significant within a region during the hurricane and (c) To demonstrate that the sentiment analysis can measure people's emotions during the natural disaster, which the authorities can use to limit the damage, effectively recover from the disaster, and improve future response plans. In addition, our work provides an in-depth topic modeling from the tweet data and contributes (d) an evaluation of the different topics of interest occurring over the course of the storm and what they indicate about users' concerns over this period. Our work is unique as it provides both sentiment analysis and topic modeling of social media data during a disaster-type event, as well as offering a detailed temporal and spatial analysis of these results. Although many previous works which study social media activity during hurricanes have focused on either sentiment analysis or topic modeling, no such research has been found which conducts both within the same paper. Furthermore, previous works which perform several types of social media analysis often do not delve into a deeper spatial-temporal analysis of their data.

Remainder of the manuscript is structured as follows. Section 2 describes the background for the research. Section 3 provides the review of the related work in the literatures. Section 3 describes the research objective. Section 4 describes the methodology of our work. Section 5 provides the discussion of the result. Section 7 outlines the lessons learned and future work recommendations. Finally, the conclusion is described in Section 8.

2.0 Background

2.1 Development, Impact, and Social Media Presence of Hurricane Irma

Hurricane Irma began as a tropical wave moving off and over the western coast Africa on August 27th and formed into an organized low-pressure region within the next few days. These factors made the-soon-to-develop Irma a textbook example of a Cape Verde hurricane, which is often the most intense and catastrophic hurricanes of any season. By early September 2017, Irma first pushed through the Atlantic islands before it made landfall in Florida as a Category 4 hurricane and continued inland into Alabama and Tennessee where it finally dissipated (A. Anandhi, 2017). Although the United States was spared from catastrophe, the island nations of the Atlantic were not as fortunate. The island of Saint Martin was hit particularly hard, resulting in 90% of the buildings being damaged and 60% left uninhabitable. Surrounding islands took moderate to minimal damage from quite sizable storm surges and rainfall. The hurricane arrived at Cuba as a Category 5 where Irma lost a great deal of its power but inflicted significant damage to many of Cuba's primary tourist cities. The hurricane left around 44 individuals dead in the Caribbean, mostly in Saint Martin and Cuba, and resulted in approximately 13 billion dollars in damages amongst all the affected islands (T. Zolnikov, 2018).

If the storm had not passed over Cuba, then it would have retained its strength as a Category 5 hurricane when it made landfall in Florida; therefore, many areas experienced less damage than forecasted. The Florida Keys experienced the most serious impact from Irma in the United States with buildings, roads, and communication networks suffering significant damage and severe flooding occurring on all the islands. States such as Florida, Georgia, and South Carolina had issued a state of emergency and evacuation orders prior to the storm's arrival, which most certainly resulted in reduced loss of life since impacts had been forecasted to be catastrophic (T. Zolnikov, 2018). Blackouts in Florida were widespread and affected nearly 12 million people, which was compounded by post-storm temperatures ranging from the mid-90's to the low 100's. Blackouts also caused cell and landline communication service to be down for some time (Y. Liu et al., 2017). Despite state regulations requiring that an elderly person be evacuated from a nursing home if the temperature cannot be regulated in the facility, eight elderly nursing home patients died during the heat wave due to a nonfunctional cooling system which had been struck by a tree during the storm (J. Shultz, and S. Galea, 2017); this resulted in a notable amount of social media attention. Overall, Hurricane Irma resulted in 88 deaths in the mainland—80 in Florida, five in South Carolina, and three in Georgia—and is estimated to have caused around 50 billion dollars of damages (S. Xian et al., 2018).

Hurricane Irma is an excellent model for this study due to the significant amount of Internet attention it gathered during its development all the way through its aftermath. This can be seen by comparing its online attention to that of major hurricanes both preceding and following Irma. Irma followed in the wake of Hurricane Harvey, which topped out as a Category 4 and was the first major hurricane to make landfall in the mainland United States since Hurricane Wilma in 2005. Harvey caused severe flooding in Texas and resulted in 82 deaths and an estimated 180 billion dollars' worth of damage (K. Bayardet al., 2017). Several hurricanes followed Irma, all of which made little impact until the development of Hurricane Maria. Maria also maxed out as a Category 5 storm but veered away from the mainland United States, instead primarily impacting the Dominican Republic and Puerto Rico. These areas had also been impacted by Irma and the combined damage sustained was catastrophic, wiping out more than half of both island's buildings and resulting significant numbers of deaths and damage costs which have still yet to be officially determined. Maria's impact on Puerto Rico is particularly noteworthy because the amount of public outcry that resulted from the government's reluctance to provide disaster relief to the U.S. territory (J. Roman, 2018). As shown in Figure 1, Google Trends search for the three storms by name shows that despite their similarity, Irma produced over twice as many web searches as Harvey and more than four times as many as Maria. Despite the controversy surrounding its impact on Puerto Rico, Maria had the least web attention. (Google Trends, 2017).

Figure 1. The Google Trends results for "Hurricane Harvey", "Hurricane Irma", and "Hurricane Maria" which are indicated by blue, red, and yellow respectively (Google Trends, 2017).

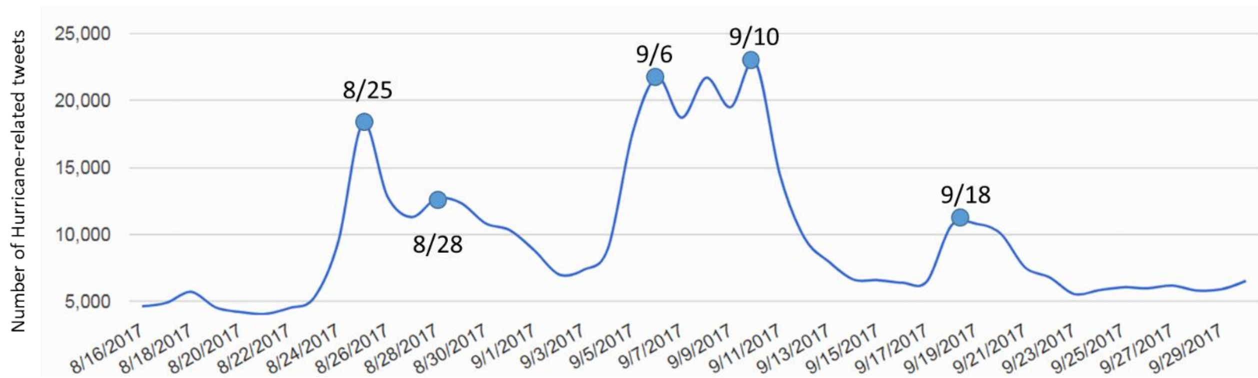


Figure 2. The frequency of tweets posted within the United States containing the defined keywords from August 16th to September 30th, 2017.

2.2 Sentiment Analysis Algorithms and Theory

One of the most important purposes of communication is to show emotion. This need to show emotion is why one of the most frequent text analysis techniques is sentiment analysis. Sentiment is often associated with terms such as belief, view, and opinion but is also reflective of the feelings an individual has to that opinion (J. Bollen et al., 2011). Sentiment is a characteristic of text that conveys unique emotional properties compared to other concepts such as topic and frequency. However, sentiment is much more difficult to accurately evaluate because it involves imprecise and often unquantifiable factors (B. Pang and L. Lee, 2008). This field of research is closely tied to studies like computational linguistics, natural language processing (NLP), and text mining. Sentiment analysis has seen a sudden surge of interest in recent years accompanying the rise of social media. It has a range of applications, from marketing to customer relationship management to government intelligence; however, it has found a niche in the analysis of political attitudes of both officials and individuals and was applied during both the 2008 and 2012 U.S. elections with great success (H. Wang et al., 2012, F. Wanner et al., 2009).

The first task which must be completed before performing a sentiment analysis is to process the data collected by correcting textual errors and cleaning, or removing, non-text characters using various methods. Cleaned data is separated into text data to be classified as either subjective or objective, which will indicate whether the item in question expresses an opinion or not. Non-opinions are considered 'neutral' and classified as an additional sentiment in this paper. The process then moves on to polarity classification, which is determined to be either negative or positive and is assigned a corresponding strength. This is the step where machine learning algorithms such as Naïve Bayes classifier, Support

Vector Machine, and Random Forest classifier can be used, but other more simplistic methods such as a reference library or lexicon are often used as well. The next step is to identify the target of the sentiment, which can vary in difficulty depending on the source domain, then collect and aggregate the resulting data. A more recent application of sentiment analysis has been in speech analysis, which relies either on acoustic factors such as tone, pitch, volume, intensity, and rate of speech or on linguistic factors such as pauses, hesitation, laughter, breathing, and words or phrases to provide a sentiment score (E. Turban et al., 2015).

The main arguments involved in sentiment calculations are credited to the pattern library in Python. The words within the lexicon are assigned a polarity, subjectivity, and intensity value based on the average values of these characteristics obtained for a specific word throughout the entire dataset of reviews. Negation within a phrase is calculated by multiplying the polarity of the associated word by –0.5 with no change to subjectivity. Modifier words used in association with other words multiply the word's polarity and subjectivity values by the intensity of the modifier. Negation and modification to a phrase is calculated by multiplying the polarity by –0.5 and the inverse intensity of the modifier word and multiplying subjectivity by the inverse intensity alone. Single letter words and words which are not registered in the lexicon are ignored and have no impact on overall sentiment. Finally, the total sentiment values are found by averaging the scores of recognized words and phrases within the subject text. Both polarity and subjectivity values for any subject text cannot exceed 1.0, and subjectivity is only positive whereas polarity can be negative.

2.3 Latent Dirichlet Allocation Topic Modeling Algorithms and Theory

Another important task in understanding communication and social media is understanding the relationships between topics which people discuss. This can be achieved using topic modeling, a statistical model in machine learning and data analytics applications used to discover abstract topics from a large collection of documents. These topic models can then be used to provide insight into otherwise unmanageably large or unstructured sets of text and discover themes within the data. Some of its most advanced work has been detecting instructive structures such as genetic information, images, or networks, but a significant amount of work has also been focused on its uses for social media data. There are some limitations, such as the small document size often encountered in social networking posts, but this has not stopped researchers from using topic modeling to visualize the direction of public discussion. For this paper we will be utilizing an algorithm called Latent Dirichlet allocation, or LDA, which is the simplest form of topic modeling. In this algorithm it is assumed that some number of topics, recognized as distributions over fixed vocabulary, exist for the entire collection of documents. The algorithm approaches its task if the topics generated the document, instead of the other way around. It assumes that those topics then generated the words which make up the document based on each topic's probability distribution (D. M. Blei, 2012).

Latent Dirichlet allocation is a Bayesian network, meaning it uses Bayes' theorem to make changes to probabilities as more information becomes available to the network and infer underlying probability distributions. This theorem computes posterior probability, or the probability of a random event considering relevant information, under the following premise:

$$P(H|E) = \frac{P(E|H)}{P(E)} \cdot P(H) \quad (1)$$

Where H stands for any hypothesis whose probability is contingent on some evidence E. The fraction component is best described as the impact of the evidence on the probability of the given hypothesis. As new evidence is discovered, this probability will update and change to reflect the changes of the likelihood of the new evidence being observed assuming the hypothesis, $P(E|H)$, and the probability of the hypothesis before considering the current evidence, $P(H)$.

This process involves many repeating variables for each individual item within a set, and therefore is often visualized with what is called plate notation to simplify these repeats. The plate notation for the LDA variation most frequently used for topic modeling today is shown in Figure 3.

In this notation, α is a vector parameterizing the continuous multivariate probability distribution called the Dirichlet prior representing the probability distributions of topics per document, and β is the same parameter only representing the probability distributions of words per topic. These values for each individual document are expressed as positive real numbers and all values for α and β over the dataset are vectors of K and V dimensions respectively, where K represents the number of topics and V represents the number of words in the vocabulary. Both values are generally less than one as we are working with sparse word distributions, or a limited number of topic words. In this form, we are modeling the topic-word distributions, so K is a preselected value based on optimizing coherence of topics. M in this notation represents the number of documents in the collection and N represents the total number of words over all the documents, or the sum of all N_d values for

M. The variables θ and φ are matrices representing the corpus of documents which are being used for modeling. The matrix θ consists of rows derived from documents and columns derived from topics, and the matrix φ consists of rows derived from topics and columns derived from words. Each individual element in these matrices represents an individual probability of an element, whether it be a word occurring within a topic (e.g. φ_{kw}) or a topic within a document (e.g. θ_{dk}). Since these are probabilities, the sum of each row in a matrix must be no greater than one, and each individual element in a row can be no greater than one. Finally, z is the identity of the topic of all words in all documents (with the topic of an individual word w in document d denoted as z_{dw}) and w is the identity of all words in all documents (with the individual word w in document d denoted as w_{dw}).

Given that we have no knowledge of how topics will be distributed over documents the best assumption is to approach the model using a symmetric LDA distribution, in which all the elements composing the vector α are given the same value. Since they all have the same value, α can be simplified to a single scalar value representing the topic density distribution. This results in a simplified topic density function which can be expressed in terms of α, seen below in equation 1.

$$f(x_1, \dots, x_{K-1}; \alpha) = \frac{\Gamma(\alpha K)}{\Gamma(\alpha)^K} \prod_{i=1}^{K} x_i^{\alpha-1} \quad (2)$$

For these scenarios, α directly correlates to the number of topics per document meaning a higher α corresponds to more topics per document. Similarly, the same can also be said regarding β in relation to the number of words per topic. Documents in this study are tweets selected by a very specific filtering method and are restricted to 140 characters per tweet, meaning that the number of words per document will be very small and all our documents will most likely center around a select number of topics. Therefore, when performing the topic modeling α and β should be very small numbers. K will be set as the iterations of topic modeling that we wish to complete, as the algorithm will return that number of models (D. M. Blei, 2003).

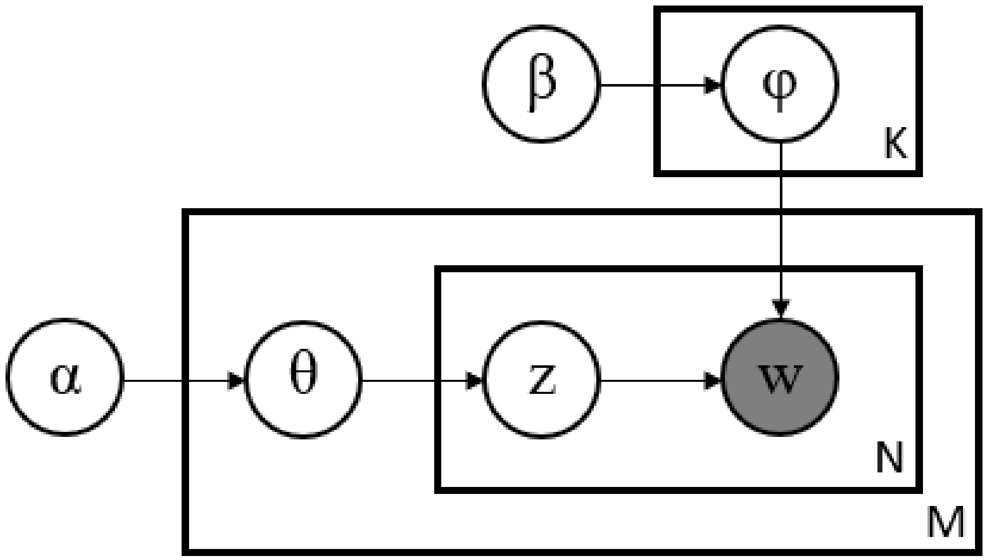


Figure 3. The plate notation for topic-distributed Latent Dirichlet allocation

3.0 Research Objectives

The objective of this research is to determine if Twitter data can be used to evaluate public perception of rescue and relief efforts in the aftermath of a natural disaster using sentiment analysis. From the collected sentiment data over the course of Hurricane Irma's progression from the regions of interest, we will attempt to establish an observable pattern regarding the storm's effects and the actions of the state and local governments. Additionally, through text analysis and topic modeling, we expect to be able to identify topics of concern or interest to those impacted by the storm. Finally, we shall complete a detailed spatial-temporal analysis of the data and visual present patterns in sentiment in the regions of interest over time.

4.0 Review of Related Studies and Literature

4.1 Text and Sentiment Analysis

Many of the related literature reviewed for this study involved research which utilized text or sentiment analysis, since both virtual communication and emotional response are a key part of social media. The techniques used in the related literature discussed here contributes to the field of text mining and sentiment analysis process, as well as natural language processing (NLP) methods and deeper emotional studies.

In a related study completed by J. Dodd, the author compares multiple machine learning algorithms' performance for the extraction and classification of an audience's sentiment related to a popular television program. The results found that the Random Forest classifier provided the most accurate results and concludes on some the best techniques in machine learning and natural language processing for this task (J. Dodd, 2014). Their study utilizes a tool which uses Native Bayes classifier to analyze sentiments, which was found to have the least accuracy, around 70%, when compared to other machine learning methods. Their description on how their text data, which was also retrieved from tweets, was prepared and cleaned for analysis is noteworthy. Many of the methods described, such as the removal of punctuation and hash tags, was utilized in the current work in the tweet cleaning process.

A twitter analysis performed by Dini and Bitar focused on analyzing users' emotional reactions in response to a product or event; however, unlike the previous work—which was focused on positive and negative—the authors here are concerned with more advanced emotions, such as anger, fear, and joy (2016). Since the expected response to a natural disaster is generally negative, a more complex emotional analysis would be better suited to paint an accurate picture of public sentiment. In this work, the methods of separating emotional and non-emotional tweets and classifying emotional tweets is performed by machine learning methods for two corpora (Dini and Bitar, 2016). The two corpora created for this research were called the "silver standard" and "gold standard" for evaluating emotional tweets, and if such resources were available for use then could most certainly improve the results of this analysis. The data presented by tweets is often limited due to reduced word use, therefore the analysis of posts on other social media such as Facebook could be even better suited to this approach of sentiment analysis. Additionally, the source which they designed their corpus collection method from used an even more intensive design which included the addition of emoticons and emoji in the consideration. This original source also used a different emotional model—Plutchik's "wheel of emotions"—which included a much more complex network of emotional states (J. Suttles and N. Ide, 2013).

Although it was not applied to the work done in our study, there are already corpora available for training and analyzing the effectiveness of classification models. In this work by Calvo and Kim, a study was done using some of these corpora to create a 3-dimensional vector model for the emotional classification of text samples using an emotional thesaurus and a bag-of-words (BoW) model, with vectors of valence, arousal, and dominance (2013). Several different dimensionality reduction techniques were evaluated—Latent Semantic Analysis (LSA), Probabilistic Latent Semantic Analysis (PLSA), and the Non-Negative Matrix Factorization (NMF). Unlike the previous papers, this work focused on text documents such as news headlines, responses to questions, fairy tales, and course evaluations in their analysis. They also performed an affect detection analysis, which although like sentiment analysis is not the same. However, its basic approach is the same in that it seeks to attach emotional labels to text data and deals with similar limitations as this research encounters; the use of words against their contemporary meanings, such as “crying for joy” or “a good beating”, provided a significant source of confusion for the lexicon approach used in affect detection and could be similarly seen within the results of the analyzed twitter data. The NMF model of dimensionality reduction performed the best within the mentioned study, so if in future work such a method could be used to add additional levels of depth to the research (Calvo et al., 2013).

A review by R. Feldman (2013) presented some of the main research problems related to sentiment analysis, its applications, and few algorithms to solve these problems. The author also described some research challenges such as noisy texts, automated classification methods to identify sarcasm, and utilizing objective facts that our work tries to tackle. Given the small size and limited information presented in a tweet, it is unlikely we can utilize these algorithms at the time of this research. However, these are important consideration and may help to develop strategies for handling such complications in future work.

4.2 Social Media Analysis

Recently, many researchers have turned their attention to interpreting the latent data hidden within social networking and media content. Therefore, there is a vast number of resources from which to build a methodology for approaching this topic. Following are some of the works referred to for the social media analysis portion of this study.

The decision to use Twitter over social networks such as Facebook or Instagram was made primarily because of the company’s supportive stance on using their data for research purposes. However, there are many other considerations which made Twitter an optimal platform for evaluation. Twitter has recently emerged as a community information mechanism which helps provide up-to-the-minute reports before other news sources can publish anything, which has great benefit to crisis situations where communication or travel may be restricted. It is also imperative to mention that Twitter can just as easily perpetuate misinformation, unsubstantiated rumors, and propaganda under the guise of truth, and then quickly spread this information within a community before official reporting agencies can dispute the claims. A 2013 study reviewed twitter data over the course of several high-profile news events, such as terrorist attacks, shootings, and mass recalls, and found that much of rumor-mongering came not necessarily from ambiguous content but from ambiguous sources (Oh et al., 2013). In fact, source ambiguity was found to be more impacting than user anxiety, which had been a significant contributor in offline content research. This may present some problem to a broad twitter study such as our research work, since there is no vetting of tweet source, however the personal involvement of the sample area in the crisis may allow for more reliable information. It is also important to note one conclusion of the study which shows why studying disaster recovery is so important; in situations where many negative rumors were spreading, users begin to distrust leadership and showed signs of reduced morale (Oh et al., 2013). If such things were to occur during a disaster, the effects would surely compound and complicate the negative impacts already felt by the community.

In relation to an event such as a hurricane, the significance of the event is very different for individuals within the storm’s path compared to those outside its range. A cooperative and highly intensive report by Ardon et al. (2013) describes the geospatial analysis of events and topic popularity using over 10 million users and 196 million tweets, making it a perfect reference for considering geography in the evolution of an event's popularity. It also discusses the different stages of an event, which have been considered in this analysis as well. The study described in this paper considers many different topics in its analysis using popular hash tags; the different ways in which these topics were discussed on Twitter is seen through their visualizations. For example, the use of "#FollowFriday" showed a highly organized and inverse pattern of popularity and conductance, whereas the topic "#IranElection" showed very little correlation between popularity and conductance and did not demonstrate a repeating pattern (S. Ardon et al., 2013). Though only one topic is of value to this research, the understanding of how such topics fluctuate in popularity can help the understanding of similar patterns regarding Hurricane Irma.

The unfortunate limitation to researching emotions and social media is that in nearly all cases one user's sentiment is not independent from another's. Social media frequently influences users' emotions in a mob-type fashion in part because of the nature of viral content and subjects. Our paper delves deeper into this concept by measuring the influences that recently viewed content and communications may have on other users. This is a difficult concept to account for in a sentiment analysis, but its effects can be very significant. For instance, the deaths of several elderly nursing home patients quickly became viral and drew massive public outcry (J. Shultz, and S. Galea, 2017); such an event could cause a swell of negative sentiment data related to the causative event without being directly related to the preferred subject matter of the research. The information outlined in research by Ferrara and Yang (2015) could be very valuable when reviewing specific trends in sentiment, and the methods used could possibly be adapted for future work to add an additional layer of depth to the sentiment research.

4.3 Natural Disasters as Research Models

The primary objective of our work is to present a sentiment analysis of tweets during Hurricane Irma. Several other research papers have focused on hurricanes or similar disaster events, often with the use of social media data to gauge public response. In a related work (Martin et al., 2017) they evaluated twitter data from Hurricane Matthew in 2016 to evaluate evacuation compliance in affected regions. This metric of disaster response is a very important factor to consider when attempting to research public response to certain events, as state evacuation plans can often be judged based on the number of successfully evacuated citizens. Additionally, evacuation compliance can be a significant factor in understanding how a disaster is perceived by the public since many people will only evacuate when they believe that the dangers of remaining in the area outweigh the time and effort of leaving. This is supported by another related work which surveyed 3,272 households in the coastal region of South Carolina (Cutter et al., 2011). The survey found that 76.6% of respondents indicated that were somewhat to very likely to evacuate in the event of a major hurricane (Category 3 or higher), while only 21% would do so for a minor hurricane (Cutter et a., 2011). The Hurricane Matthew study was performed by evaluating geotagged tweets containing the keywords "matthew", "hurricane*", "evac*", "storm*" which were posted from October 2th to October 11th, 2016 in the areas of interest; although temporal analysis was performed at a state-level, the spatial analysis was more selective and only evaluated tweets with city-level and point-level accuracy. For each affected region tweets were grouped into three timeframes—pre-evacuation, evacuation, and post-evacuation—while additional tweets following the post-evacuation period were also collected as the 'return' period. Users were categorized as evacuated when they moved a certain distance away from the affected areas (Martin et al., 2017). This study would be a good complement to a sentiment analysis of Twitter data in these disaster scenarios, as specific areas with low or high evacuation rates could be studied to gain an understanding of what may have helped or hindered the process.

A related work (Leavitt and Clark, 2014) used a different social media site, Reddit, and a similar event, Hurricane Sandy, for their analysis and outlines the event-based news production processes on social media. One of the primary goals of their work was evaluating which posts were 'upvoted', or given precedence or popularity, over others and why. By dissecting the posts into distinct classes based on the information they contained, and then further categorizing them into subclasses and content types, the authors were able to create a very detailed plot of how each class trended over time. The topic classification method and the evaluation of post data used here is quite intensive and could be helpful in developing a similar method for classifying topics related to Hurricane studies, since many news sources often make use of twitter. Within the classification method, there is a class for photo-based posts, which also were determined to comprise a large portion of the tweet data used in the current study. Since it is focused on a less structured and more communication and interaction intensive social media site, this work could also provide better insight into the future evaluation of Facebook posts compared to the previous works, which are all primarily focused on Twitter. This work could be incorporated into the current line of study by including some analysis into the content of tweets or posts which receive the most attention in the form of likes, retweets, and shares. This would be of great use in a business intelligence format because the understanding of why posts or tweets gain popularity can assist in processes such as advertising, marketing, customer relations, and brand management (Leavitt and Clark, 2014).

In a combination of the geo-spatial approach (Ardon et al. 2013) and the subject of Hurricane Sandy seen in the related research (Leavitt and Clark, 2014), the work of Caragea et al. (2014) outlines the mapping of geographical location data in comparison to the position of the hurricane in question. This research would serve as the ideal design of future work using the Twitter API and uses an impressive visual scheme for demonstrating its results. The paper also utilizes Twitter data for sentiment analysis and outlines the way in which the authors approached concerns such as misspellings, emoticons, internet slang or acronyms, and punctuation to aid the sentiment evaluation process. Considering many tweets are no more than a dozen or so words, the consideration of elements such as emoticons, which are often highly associated to specific emotions, could greatly improve the ability to correctly assign sentiments. Additionally, the techniques and tools used by Caragea et al. are useful for future study, particularly the text processing, sentiment analysis method, and visual modeling used to demonstrate sentiments by region during Hurricane Sandy's progression up the eastern coast. The inclusion of the relative position of the storm in the geographic mapping of sentiments is highly inventive and better demonstrates how regions respond before, during, and after an event like a hurricane (Caragea et al., 2014). This work represents an example of how web scraping can be used for data science applications outside of business or marketing analytics. By studying something more humanistic, such as the response of individuals during a disaster situation, service organization and government agencies may be able to more clearly understand how these people can be helped and what mistakes may have been made.

A related work (Cheng and Cheng, 2011) used Social Network Analysis metrics and techniques to identify influential members of the online communities during 2009-2010 Australian flood for the disaster resilience perspective. However, they did not perform sentiment related analysis for the disaster recovery. Their methods do analyze the impact of different levels of influence in social media users and how these users' actions shape the online community around them. Regarding storm and disaster studies, influential users responding positively or negatively to response efforts could drastically shape the outcome of overall sentiment and direct the focus of topics discussed on these platforms. It therefore is a factor worth considering in this study and future work.

4.4 Topic Modeling Using Social Media Data

The first literature we will discuss for this portion of the study is "Empirical Study of Topic Modeling in Twitter" by Hong and Davidson (2010). This work focused on the advances made in topic modeling using twitter and other microblogging platforms as a data source. It also discusses the challenges faced by working with such limited data sources, as all twitter data is guaranteed to contain 140 words or less. This paper establishes that Latent Dirichlet Allocation (LDA) is considered the standard method of performing topic modeling for twitter data and describes some improvements and implementations that have been suggested throughout the academic community. The challenges of topic modeling in twitter have led some in the research community to claim that by aggregating the short tweets into longer messages, the results of the topic modeling can be improved. This is tested in this paper and is disputed by their results. The authors argue that there is evidence that LDA works much better unaltered than this proposed method. They also discuss several different topic modeling schemes and the concept of the Author-Topic model (Hong and

Davidson, 2010). Throughout the paper they compare several methods and discuss their performance overall. This paper was quite useful in determining what methodology to follow to complete topic modeling and introduced concerns which may be encountered during our study.

As mentioned in first literature discussed in this section, understanding the limitations of what can be done with topic modeling is important when developing a study. The work by Tang et al. (2014) is an in-depth study of the limiting factors of topic modeling in general using posterior contraction analysis. Although it does not specifically focus on social media uses, this paper is excellent source for understand what can and cannot be done with topic modeling and does describe social media in detail within the paper. The paper also has a specific focus on topic modeling using LDA, which makes it even more relevant for our study. Some of the specific questions addressed in this paper were the challenges related to changes in the number of words per document, changes in the number of documents in the corpus, and how they relate to each other. Their results showed that the performance of LDA topic modeling decreases when documents are particularly short, even if there are many documents in corpus. This was most certainly an obstacle in our study, as tweets are all restricted to 140 words or less. Additionally, the LDA model performed best when the underlying topics are well-separated, or if topics are concentrated around a small number of words (Tang et al., 2014). Since our dataset is so specific to one event, theoretically the results of the topic modeling should be improved over data taken over a non-descript period.

The next literature discussed once again focuses solely on the topic modeling of twitter data, however this study is even more relevant to our work because it focuses on aligning specific events and the following twitter response. Since we are working to evaluate the response of people to the hurricane and the following relief measures taken by the states, this paper provides a very useful approach for further improving this work. “ET-LDA: Joint Topic Modeling for Aligning Events and Their Twitter Feedback” by Hu et al. (2012) proposes a unique model called the joint Event and Tweets LDA, thus ET-LDA which would strive to model not just the event’s topics and their evolution though the use of event segmentation, but also model the associated tweets’ topics and evaluate the tweeting behaviors of the public. Events in this sense are assumed to formed from discrete, consecutive segments, which each discuss a set of topics. This method would be difficult to apply to a disaster-type event as it was originally based around the idea of publicly viewed events, such as the Super Bowl or Presidential speeches, which would then have a corresponding transcript of the event which can be analyzed. Although this is a drawback, if our work were to be continued in a more in-depth manner, transcripts from local and national televised news stations, written news articles, and other reputable forms of event documentation discussing the disaster could be pooled into a collection of documents with which a similar model could be used. The results of the model as discussed in this paper were quite good, as it out-performed several other similarly minded models in the tasks of event segmentation and topic modeling and showed significant improvements over the baseline methods for such tasks currently (Y. Hu, et al., 2012).

5.0 Methodology

5.1 Study Area

This research focuses on twitter data related to Hurricane Irma aggregated by county in the Southeastern states of Florida, Georgia, Alabama, and South Carolina as shown in Figure 3 below. All these states were directly impacted by Hurricane Irma, although Florida experienced the greatest storm strength as it is where the storm first made landfall. All four states border the Atlantic Ocean which means that storm surges are a possible concern with the most susceptible being Florida. The more inland states of Alabama and Georgia are likely to have less experience with cyclones and heavy flooding, meaning they will likely feel the most impact due to lack of preparation; conversely, these states could overprepare due to the areas' inexperience with such events. Wind and fallen debris should be a concern for all states, with the level of concern falling in relation with the gradual weakening of the storm. Irma's spatial distribution as it made landfall in the US is shown in figure 4. The position of the eye during its progression and the range of the storm's effects are indicted by the green dots and the pink border, respectively. The interior border represents the most damaging effects which circle the eye of the storm. It can be seen that the chosen states are certainly the areas which received the strongest effects of the storm, supporting their selection as areas of interest for this study.

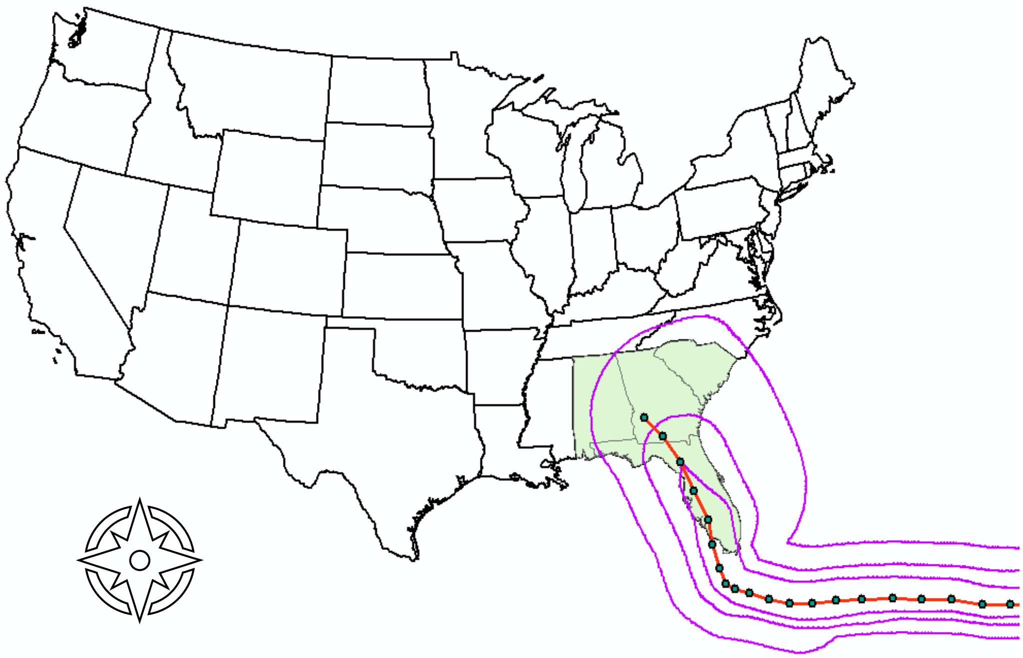


Figure 4. Map of Irma's path and the study area (zone of impact) from September 6th to its dissipation on September 13th.

5.2 Data Collection

The research approach used is based on different datasets, study areas, and timeframes. The data is collected over a period starting the day Hurricane Irma was named until its eventual dissipation, from August 30th to September 16th, 2017 within the continental U.S. using the Twitter Stream Application Programming Interface (API). These tweets are stored in a Hadoop (http://hadoop.apache.org) environment that served as tweet repository for this study. The repository is then queried with Apache Impala (https://impala.incubator.apache.org) using spatiotemporal criteria and keywords in the tweet message and hashtags, obtaining datasets covering different spatial regions, keywords, time periods, and spatial accuracies. The locational accuracy of a geotagged tweet depends on how a Twitter user shares his/her location when posting a tweet. The location can be shared in the format of place names (e.g. country, state, city) or the exact latitude and longitude (point-level, determined by the device's GPS or other signals such as cell tower).

As shown in the Figure 5, the search query based on the keywords related to Hurricane Irma is subjected to the Impala query to retrieve the tweets related to Hurricane Irma between August 30th and September 16th of 2017. Each of the 221,598 tweets extracted is then subjected to the text processing, sentiment analysis, and topic modeling as described in Section 4.4, 4.5, and 4.6. Each of the tweet records have the following data elements: tweetid, userid, tweet_postname, tweet_message, user location, source, latitude, longitude, county, and URLs. The sentiment analysis results in the temporal analysis, spatial analysis, word count and topic analysis as described in Section 6.1 and topic modeling results are discussed in Section 6.2

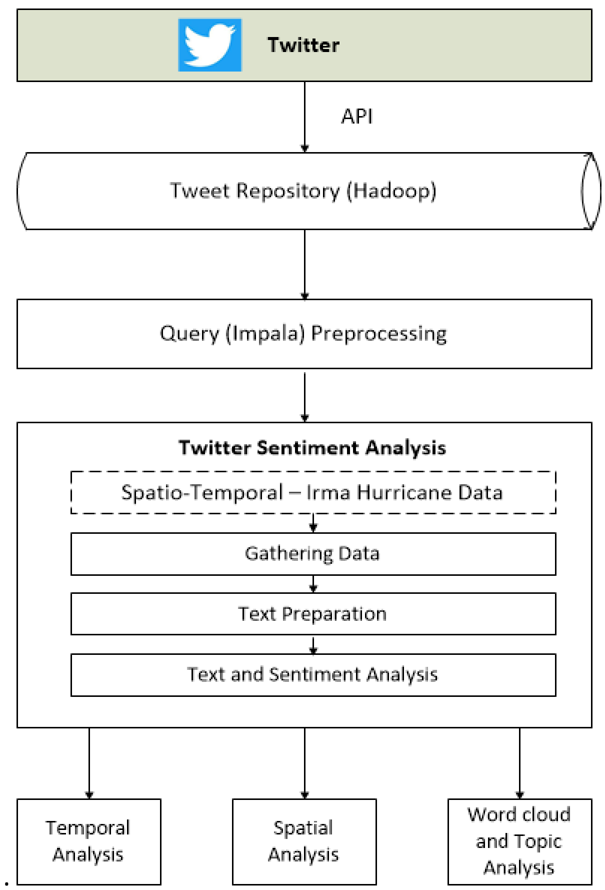


Figure 5: A diagram outlining the framework for the data collection and analytics in this study. This includes the storage and preprocessing steps

5.3 Platform and Libraries

The platform used for implementing this research is Python version 3.6, with the use of Anaconda 3 for library installation using the pip command. All coding is carried out within Jupyter notebook, assigning each region a unique notebook.

The libraries used within this work are *matplotlib.pyplot, pandas,* and *numpy*; these are basic data analysis and statistical tools which are included within the Anaconda package. Additional libraries that are installed for use in this work were *textblob*, *re*, *random, nltk, seaborn,* and *wordcloud*. The library *textblob* is utilized for text tools and sentiment analysis, *re* was installed for Perl-like regular expressions, *random* was used for random number generation, *nltk* stands for the natural language tool kit and was used for natural language processing (NLP), *seaborn* is used for implementing advanced plotting methods, and *wordcloud* is installed to produce word clouds for the tweet data. To successfully run *wordcloud,* the program C++ visual studio is also installed.

In order to complete the LDA topic modeling, the libraries *gensim*, *pyLDAvis*, and *spaCy* were installed as well.

5.4 Optimized Text Preparation and Alterations to Analysis Tools

We used Python to clean the information gathered from tweets posted by Twitter users during Hurricane Irma. Next, the tweet is cleaned by removing all alpha-numeric information and special characters that may interfere with sentiment analysis. Additionally, stopwords in both English and Spanish were removed, and the creation of a unique for each region, which removed words such as the state name, city name, and search parameter terms such “hurricaneirma”, “Irma”, and “hurricane”. Aside from one frequency analysis regarding hash tag topics, all “#”, “@”, “RT” or retweet tags, and URLs are removed as well. After cleaning, there are 221,598 tweets left.

The Python library *textblob* was utilized for text tools and sentiment analysis, however the code was first adjusted for the purposes of this study. A sentiment value is assigned with an assigned string value. String values assigned are positive, zero, negative, and neutral. These values depend on the sentiment polarity and subjectivity. Positive sentiment is defined by a sentiment polarity greater than 0, zero sentiment is defined by a sentiment polarity of 0 and subjectivity of 0, negative sentiment is defined by sentiment polarity less than zero, and neutral sentiment is defined as all other cases. The original code used by *textblob* had set all results with polarity equal to zero as “neutral”; however, this needs to be altered to account for incorrectly analyzed tweets, particularly those which were in Spanish from areas of Florida, and strictly factual tweets. This is done by instead setting all results which are found to have a polarity value equal to zero and a subjectivity value equal to zero to return a “zero” sentiment tweet. This is needed because tweets with these calculated parameters should be completely objective with no conveyed emotion and therefore will most likely be strictly informative tweets, which are often posted by government agencies and other service used for posting alerts and weather updates, Other tweets which may fall in this classification are tweets which could not be understood and therefore yielded no results, such as tweets written in a foreign language which are misspelled or contain slang. All other results not fitting any of these three classes, which would be nonpolar, subjective tweets, are set to return a class of “neutral”.

Each of the 221,598 tweets is scored using this method described above. Finally, the sentiment score of each tweet is aggregated at the county of level of the four states.

There is a total of 339 unique counties across four states; however, it should be noted that any counties which did not yield at least 5 tweets per day is removed from the analysis to avoid the occurrence of null values. In this study, only counties that experienced more than five geotagged tweets per day during the entire test period are used.

The initial text cleaning involves the removal of special characters and converting all letters to lowercase. The function is programmed using the following code as shown in the Listing 1:

```
def clean_tweet(tweet):

    tweet1 = tweet.lower()

return  ''.join(re.sub("(@[A-Za-z0-9]+)|([^0-9A-Za-z  \t])|(\w+:\/\/\S+)",
"", tweet1))
```

Listing 1 – Text Preparation Implementation in Python

These cleaned tweets are then aggregated into a separate dataframe, allowing for access to both and clean and raw data set. Initial sentiment analysis was performed at this point; however, additional text preparation occurred after this analysis which is significant for future sentiment and text analysis results. The tweets, both clean and raw sets, are then tokenized using an `x.split` function and the most common words are viewed; a similar process is used for positive and negative sentiment tweets. Stopwords, in both English and Spanish, were retrieved from the *nltk* corpus and removed; additional words were also removed by the author to clean up the frequency results. This was implemented by the following code in Listing 2.

```
negative_words = ' '.join(df_negative['text'])

neg_lower = [item.lower() for item in negative_words.split()]

neg_lower = ["".join( j for j iniif j not instring.punctuation) for iin
neg_lower] neg_lower

fneg_words = [word for word in neg_lower if word not
instopwords.words('spanish')

and word not in stopwords.words('english')

                and 'http' not in word

 and 'miami' not in word

and 'florida' not in word

and 'hurricane' not in word

and 'irma' not in word

and '…' not in word]
```

Listing 2 – Text Cleaning Implementation in Python

This function takes all words to lower case, removes punctuation and non-letter characters, URLs, the associated city and state of the data, and search variable such as "hurricane", "Irma", and "hurricaneirma".

5.5 Sentiment and Text Analysis

Next, *textblob* is used to calculate a numerical value for the polarity of each tweet and a string-based value as sentiment. As previously discussed, a tweet is returned as "positive" if the value for `analysis.sentiment.polarity` is greater than zero and is returned as "negative" when `analysis.sentiment.polarity` is less than zero. The original code used by textblob had set all results with `analysis.sentiment.polarity` equal to zero as "neutral"; however this is altered to account for incorrectly analyzed tweets, particularly those which were in Spanish from areas of Florida, and strictly factual tweets by instead setting all results with an `analysis.sentiment.polarity` value equal to zero and with an `analysis.sentiment.subjectivity` value equal to zero as a "zero" sentiment tweet, because the tweet was calculated to be completely subjective. All other results not fitting any of these three classes, which would be nonpolar, subjective tweets, return a class of "neutral". The code used for this function is listed in Listing 3:

```
def get_tweet_sentiment(tweet):

    analysis = TextBlob(clean_tweet(tweet))    if analysis.sentiment.polarity
> 0:

return 'positive'

elifanalysis.sentiment.polarity == 0 and analysis.sentiment.subjectivity == 0:

return 'zero'

elifanalysis.sentiment.polarity < 0:

return 'negative'

else:

return 'neutral'
```

Listing 2 – Sentiment Analysis Implementation in Python

The results are then iterated into a list of sentiment class by tweet and are stored into the dataframe as the column class "sentiment". The same process is repeated for the addition of sentiment polarity values to the dataframe as the column class "sentiment_value". The cleaned text data for each tweet was also added into the dataframe. The frequency of words in the corpus was visualized using both *wordcloud* and *seaborn*, viewing the 50 most common words. The sentiment of tweets are averaged by day for all tweets with a specific county, provided by each tweet's `county_id` found within data originally provided from Twitter; these averages were then assigned a color corresponding to the average sentiment of the county for that day, where blue to green indicated a negative sentiment, red to orange indicated a positive sentiment, and yellow was neutral sentiment. Each counties' assigned color was then plotted to a series of county-divided maps of the Southeastern US, each representing a day during the study period. A time series was then made by averaging the sentiments of tweets by state and plotting them by day, taking an overall sentiment average which was plotted for reference.

5.6 Topic Modeling Preparation and Analysis

The python library *gensim* was utilized for topic modeling in this study, in addition to *pyLDAvis* for visualizing the models to evaluate their performance. The library *spaCy* was used for tokenizing and lemmatizing the twitter documents to properly evaluate the data; this library is described as an industrial natural language processing tool for python. Libraries which were previously used for text preparation and sentiment analysis were also utilized for this task, which include *matplotlib* and *nltk* for plotting and stopword removal respectively.

The text preparation process for topic modeling was similar but slightly different from the previously implemented process used for sentiment analysis. First, all email addresses, new-line characters, single quotes, and URL or links were removed from the tweets. Then, all stopwords were removed using *nltk* English and Spanish stopword corpora as reference and a custom set of words were removed from the tweet documents; the custom words were decided at the discretion of the author based on early topic modeling results and include nonsense words (such as "amp" resulting from the character "&"), commonly found foreign words (such as "hurrican", spanish for hurricane), and words which were part of the tweet filtering process (such as "irma"). We then formed a bigram model at this point of the process; a bigram model allows for creation of topic which include two-word sequences but does not necessitate their occurrence. The tweet documents were then tokenized, or separated into words, and the words lemmatized, converted into their base form from past or active tense, using *spaCy*; this also included parts-of-speech parsing, during which only nouns, adjectives, verbs, and adverbs could remain.

To be able to analyze the twitter data, the text was transformed into a format more suitable for data analysis. We began this process by first assigning each word in bigram model with a numerical ID by creating an ID to word dictionary. This dictionary was then used to map a corpus where words where word ID and frequency within a document were expressed as a sparse vector, or ordered pair, in the form ([word_ID], [word_freq]). This resulted in a total of 71,222 unique tokens within the dataset. The final LDA model was produced using this corpus.

An important question to answer before we discuss the results of the topic modeling is how the number of topics was decided. For this purpose, we produced a coherence score graph expressing the coherence value in comparison the to the selected *K* value, or number of topics over the documents. The coherence value is a measure which distinguishes good topics from bad, or incoherent, topics in an LDA model; the higher the coherence value the better the topic results. This score does not consider repeating topics or similar topics in the results. Although this is not a complete method for judging the performance of topic models, it does provide a baseline metric against which different topic models can be compared. For this task, a function was made to continuously produce topic models with increasing *K* values, starting at a value of 2 and increasing by 3 up to a limit of 40 topics. The coherence values of these models were then compared using `c_v` values calculated through *pyLDAvis*. The following figure represents the results found from this procedure. At around 12 topics there is a notable dip in coherence of topics, and after this point the coherence of the model levels out despite increasing *K* values. Therefore 12 topics (coherence = 0.245) was chosen as the initial *K* value for this study, which would be further analyzed by personally checking this model for the occurrence repetitive or similar topic results. After running several iterations of the model using 12 topics, it was determined that many of the topics produced with this *K* value were unsatisfactory as they were very closely related and difficult to distinguish. A model using 8 topics was then ran, which was deemed to still be too repetitive; the results of this model are included in the results section of this paper as a point of reference to compare to the more optimized results. The final decision of 4 topics was reached upon observing that there were generally no more than 3 to 5 clear and unique topic subjects with any given *K* value for a model. The other parameters selected for this study was an α-auto value, which learns asymmetric prior values from the corpus, and assigning the per-word topics as true, which returns topics words in descending order of likelihood of occurrence within a given topic.

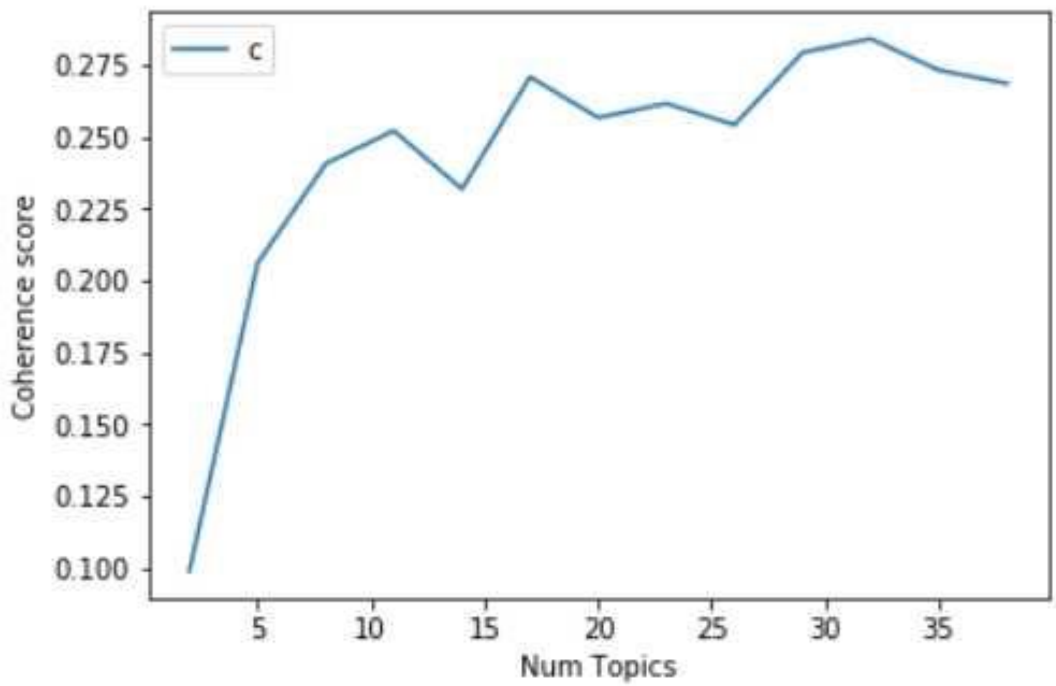


Figure 6. Coherence score in comparison to *K* value or number of topics.

6.0 Results and Discussion

Results of the performing sentiment analysis of 221,598 tweets has provided a potential analysis of the effects of Hurricane Irma from September 1^{st} to September 26^{th}. Time series graphs for most impacted areas were created after assigning a sentiment value to each cleaned tweet and averaging all tweets which were returned for each state. The spatial distribution map was formed by averaging the sentiment results by their county of origin over the course of the storm. The results of the text and sentiment analysis are described in the following order. First, we discuss the results of the word frequency analysis and their significance. Next, the temporal patterns seen in the sentiment data are visualized and described regarding the timeline of the hurricane. We then outline the spatial patterns seen in these results. Finally, we present and describe the results of the topic modeling and how it relates to the concerns of people impacted by Hurricane Irma.

6.1 Text Analysis

The tweets for this study were evaluated to attempt to determine what topics might have been of the most interest to users in the hurricanes impact zone. The top 50 most frequently used words are displayed in the following Figure 7 in wordcloud format. The size of the words in this format indicates frequency, with larger words being more frequent and smaller words being less frequently used. The orientation of words is not relevant to the data and is randomized. Wordcloud, is a model often used to display frequent words in an interesting manner, shows that many of the words seen in the tweets over this period were more positive. Phrases such as “stay safe” and “Puerto Rico” stand out from the other words, although other combinations such as “thank you”, “stay home”, and “people need” could be inferred. The occurrence of Miami in this listed is most likely a result of the high number of tweets from Florida, particularly the Miami-Dade County, compared to the other states, therefore overpowering other impacted areas such as Atlanta and Charleston. Many positive words like “good”, “love”, and “thank” stand out. Finally, the word “water” comes up, which could be related to water outages or flooding, making it a potential topic of concern. Although this format is visually interesting it can be more difficult to quantify data.

Since Figure 7 does not distinctly quantify the usage of words in comparison to others, a bar graph was made to better illustrate which words received the most attention over the course of the storm. We can see that there was a focus on the community and others from the words “us” and “people” which were the second and third most used words respectively; “everyone” and “family” also appear much further down on the chart. “Power” came in near the middle, indicating that it was a significant concern for many users. Additionally, “help”, “please”, and “need” were frequently used, which would suggest some level of distress or concern among users. Positive words like “safe” and “relief” occur frequently, indicating an optimism among users as well. It is interesting to note that the top 4 words make the phrase “get us people safe”, although this is likely coincidence; along the same line of speculation, many expected phrases appear if you put these frequent words together, such as “still don’t got power”, “need help”, “everyone please stay safe”, and “storm hit miami”. Co-occurrences of these words together may have drove up the word counts, however this is not a strong enough correlation to base conclusions on. Interestingly, the occurrence of “5” as a frequent word is most likely due to the anticipation of Irma making landfall as a Category 5 hurricane, which was barely avoided due to its passage over the island territories before reaching mainland United States.

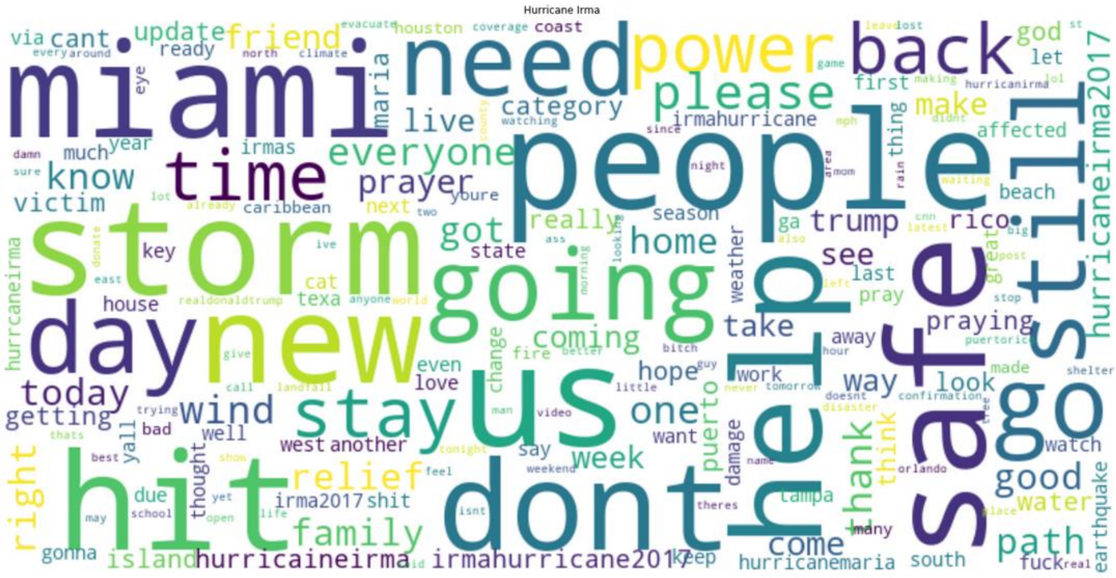


Figure 7. The 50 most frequently used words over the total timespan of the twitter data.

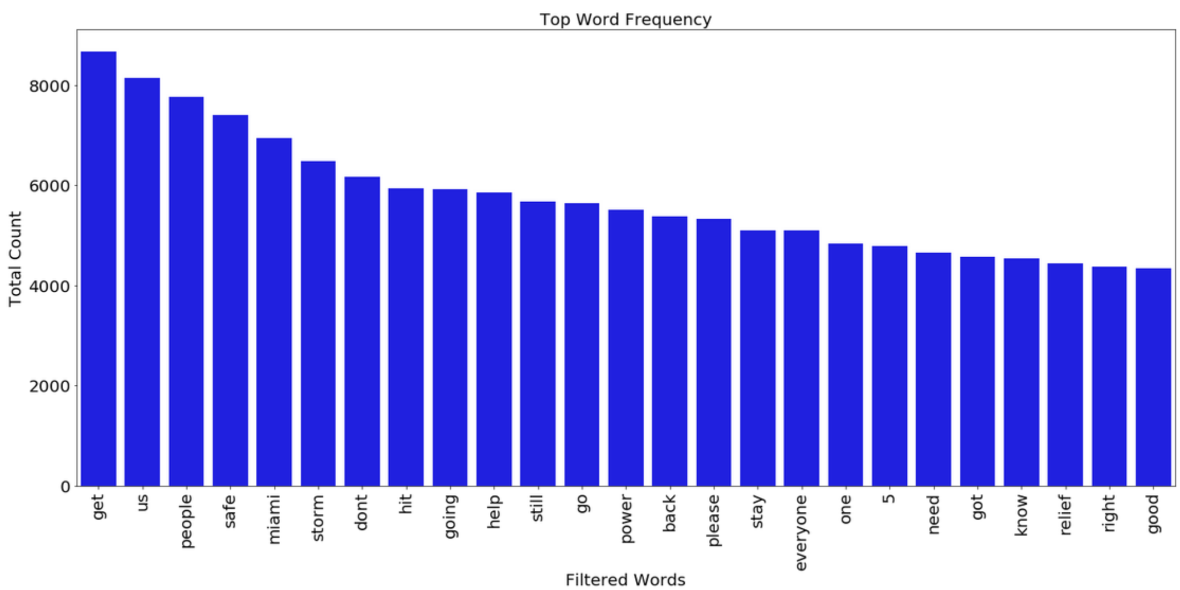


Figure 8. A bar graph ranking the top 25 words based on their frequency in tweets.

6.2 Sentiment Analysis

6.2.1 Temporal Patterns

Time series graphs for each state impacted by Hurricane Irma in this study were plotted as part of the analysis. The average sentiment scores for Alabama, Florida, South Carolina, and Georgia over the period of interest are shown in Figure 9, as well as the average sentiment scores of the all the twitter data overall. It should be assumed that an increase in sentiment is showing a more positive outlook overtime and a decrease in sentiment is showing a more negative outlook overtime. The other two figures, Figures 10 and 11, show the time series graph of Hurricane Irma's maximum recorded wind speed and the minimum recorded pressure over time respectively. The first graph extends until the 16th of September. The other two figures extend until September 13th, when Irma officially dissipated.

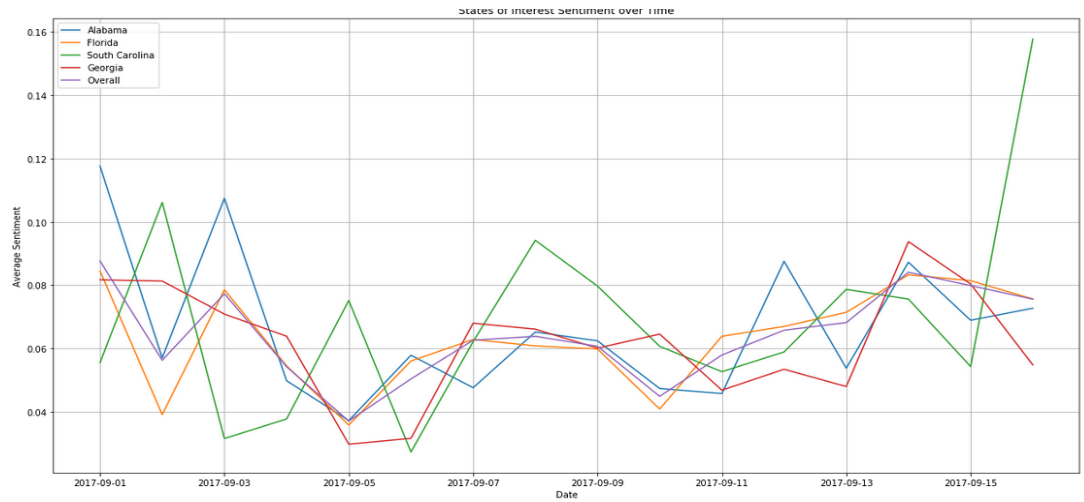


Figure 9. States of interest are shown in a time series progression. Average sentiment is plotted versus date.

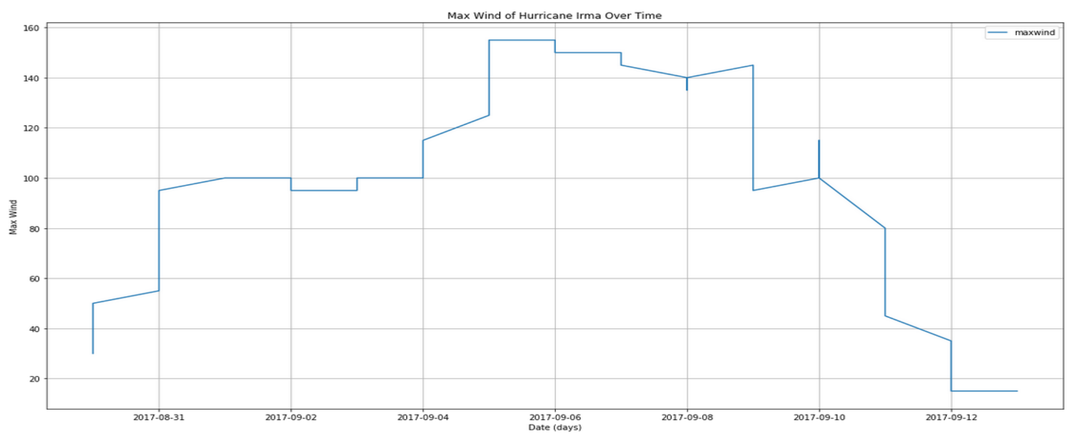


Figure 10. A time series showing the maximum wind speeds in knots of Hurricane Irma over time

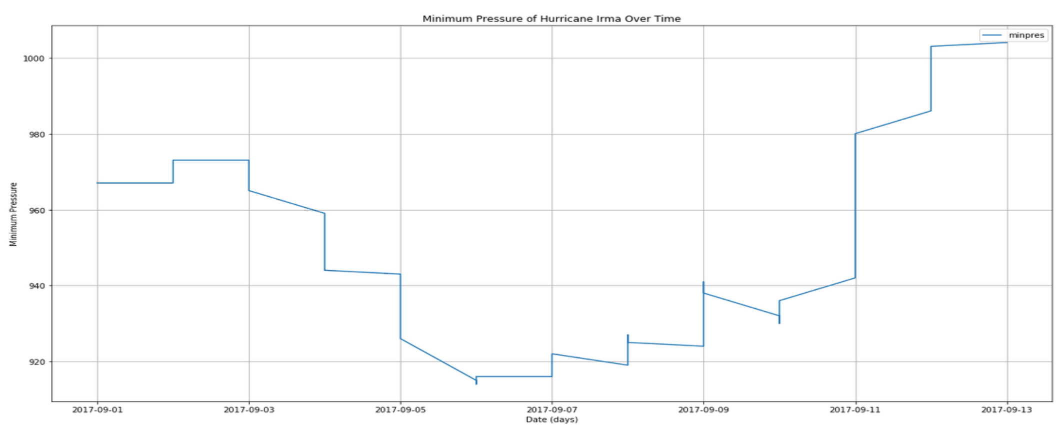


Figure 11. A time series showing the minimum atmospheric pressures in millibars over time during Hurricane Irma.

The time series shown in figure 9 indicates the average sentiment value of tweets within each state on an individual day. It shows that in each of the four states, initially the tweets from September 1^{st} were more positive, apart from South Carolina which had the lowest sentiment at the beginning of the graph. By the following day, the average tweet sentiment decreased by approximately 50% for both Alabama and Florida, while South Carolina's sentiment increased significantly, and Georgia's sentiment remained approximately the same. The changes are reverse on the 3^{rd}, with Alabama and Florida nearly doubling and South Carolina decreasing sharply; Georgia's sentiment decreased as well but less than any changes seen in the other three states. Between the 4^{th} and the 5^{th}, the states have begun to converge around a similar sentiment value, approximately 0.05. During this period, all states are reaching or will soon reach their lowest sentiment values; Alabama, Florida and Georgia all experience their minimum daily average sentiment on the 5^{th}, while South Carolina experiences its lowest value on the 6^{th}. After this day, all four states begin to experience an upward trend, although showing significant fluctuation between days. There are notable drops in the average sentiment of tweets in each state during this time; the day of the 10^{th} in Florida, from the 10^{th} to the 11^{th} in Alabama, from the 10^{th} to the 12^{th} in South Carolina, and from the 11^{th} to the 13^{th} in Georgia. Alabama and South Carolina both experienced another notable drop on the 13^{th} and 15^{th} respectively. Only South Carolina ended the period of the graph with a sentiment value higher than it had started with, a significantly high 0.16. Variation between days may be due to limited tweet information from that day, or possible misclassified tweets and possible sentiment change overtime is always varying. It should be noted that although the averages drop significantly at certain days, the overall average sentiment of any day in any state is positive.

Figures 9, 10, and 11 can be compared to observe the relationship between the average sentiments of each state and the extreme weather conditions being produced by Hurricane Irma. As previously stated, the first figure relates to the average sentiment over time, while the next two figures represent the maximum sustained wind speed in knots and minimum pressure in millibars within the storm respectively. The data for these two figures was retrieved from the National Oceanic and Atmospheric Administration (NOAA), specifically from the National Hurricane Center's revised Atlantic hurricane database (C. Landsea etal., 2015). The maximum wind speeds reached by Hurricane Irma were around 160 knots which was seen starting September 5th until the 6th, and a minimum pressure of around 915 millibars on the 6th. This corresponds with the date that Irma reached Category 5 conditions, September 5th. Comparing these figures to the sentiment time series, three of the subject states (AL, FL, GA) experienced their minimum average sentiment values on the 5th and the final state (SC) experiences its minimum value on the 6th. These reductions in sentiment relates directly to the extreme strength of the storm currently, with a negative correlation between sentiment and maximum wind speed and a positive correlation between sentiment and minimum pressure. This relationship is much weaker along the other temporal regions of the time series, indicating that this specific event had the most significant impact on sentiment.

The average trend for the main areas impacted by Hurricane Irma show clear signs of fluctuating in respect to the storm pattern according to the sentiment data. All areas show an average sentiment that is greater initially than the days leading up to the impact and the actual impact day of Hurricane Irma. The minimum sentiment scores for the states coincide with the day that Hurricane Irma reached its peak strength, with wind speeds of approximately 180 miles per hour. As stated, the previous forecasts were catastrophic for the US, so this day was most likely due to anticipation of this catastrophe. The storm hit Cuba however, and lost a great deal of strength, which most likely corresponds with the slight increase in sentiment before each states' impact. This is consistent with the actual days of impact for most states. The hurricane first struck the Florida Keys on the 10^{th}, entered Georgia on the 11^{th}, and ended in Alabama where it dissipated by the 16^{th}. It should be noted that for Florida and Georgia, the minimum average sentiment was experienced starting the day of impact for each state. In between the days of the forecast and the impact, the average sentiment can be seen increasing slightly and then decreasing right before and during the impact of the Hurricane. This is normal because people will generally be more positive towards the outlook of the hurricane right before it impacts the area. An average negative sentiment is experienced during impact, often the result of the storm's damage and yet as a response to any relief efforts by the country or state authorities. The increase back to a more positive sentiment is shown in these charts after the hurricane is no longer an issue for the area. However, two states—Georgia and Alabama—ended with an average sentiment which was significantly lower than their starting sentiment on the 1^{st} of September. Florida's final sentiment value was approximately the same as its initial sentiment and only South Carolina ended with a significant increase, nearly tripling from the first sentiment value. It is likely that Georgia and Alabama residents were less content with relief effort than residents from other states based on the lower sentiment score at the end of the time series, however they were also the last states to be impacted and therefore it could possibly be contributed to residual negativity.

6.2.2 Spatiotemporal Analysis

Figure 12 shows the spatial distribution of the collected Hurricane Irma-related tweets in a heat map indicating tweet density. It is apparent from this figure that many of the tweets were from the areas surrounding Miami, Orlando, and Jacksonville, Florida. There were also a notable number of tweets gathered from Atlanta, Georgia, Houston, Texas, and the island of Puerto Rico. Comparing to Figure 4 in Section 5.1, a clear correlation between the path of the hurricane and the density of tweets can be observed. Additionally, the collected tweets clearly condense around major city areas, even if they were not directly impacted by the storm. The most striking examples of this is the levels of activity seen in Los Angeles, New York City (the area just above and to the right of Philadelphia), and Washington D.C. (just below and to the left of Philadelphia). Spatial query and keyword filtering were used to extract the Hurricane Irma-related tweets covering the study area (the four states highlighted in green on Figure 4—AL, GA, SC, and FL). The particularly high density observed in Houston is likely the result of the impact of Hurricane Harvey in the previous month. Cuba is seen to be notably dark in this figure; as it is not a territory of the United States, twitter activity from users living in this area is not included in the study. Additionally, travel to Cuban by U.S. citizens is restricted and therefore it is unlikely that many U.S. users were present at the time of the storm. This is contrasted by the slight activity seen in Mexico City, the Turks and Caicos Islands, Toronto, and Ottawa, which are areas where citizens can travel to with few restrictions aside from a U.S. passport.

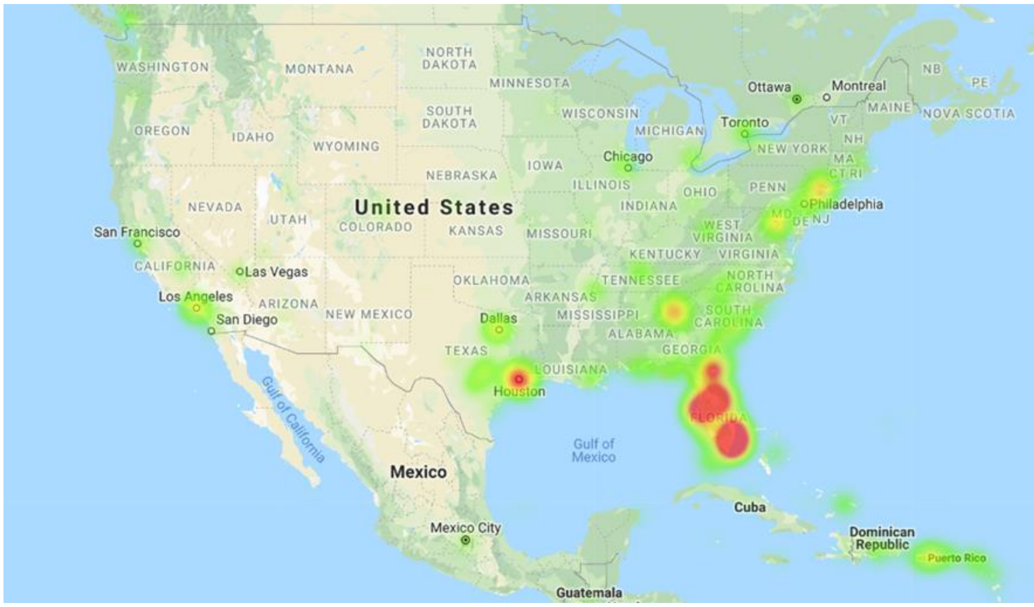


Figure 12. Heat map showing the spatial distribution of the collected Hurricane Irma-related tweets

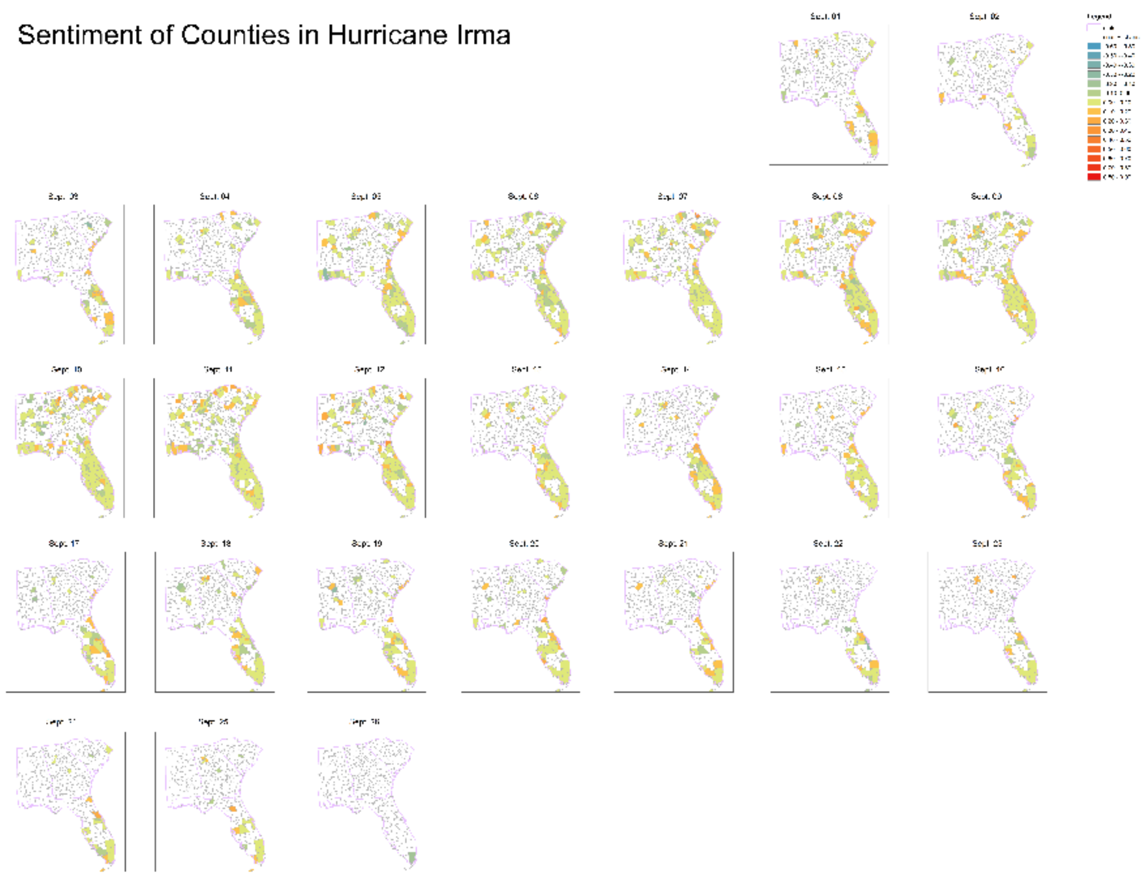


Figure 13. A geospatial map indicating sentiment over time by county. The red indicates more positive sentiment, whereas blue indicates more negative. White counties indicate areas which did not meet the necessary number of tweets for analysis.

As shown in Figure 13, the map of sentiments by county over the course of the storm shows an overall slightly positive trend, with most counties exhibiting an average sentiment value of 0.0-0.10 and 0.10-0.20. A surge in tweets about Irma on September 4th begins a little less than a week before it made landfall in Florida, and does not subside until around the 21th, a little more than two weeks after its initial impact. The greatest number of active counties is seen on the day of impact, the 10th. Interestingly, the most activity is seen in South Carolina, Alabama, and Georgia counties in the days prior to impact and then sharply drops off several days after the storm made landfall, but Florida maintains steady levels of activity over a period of about two weeks and does not see a significant decrease in active counties until the final week in the figure. This is potentially due to widespread power outages and heat wave that followed the hurricane in Florida, which complicated recovery effort and exacerbated the negative impacts of the storm. Florida also experienced the most intense conditions out of the mainland states as Irma made landfall as a Category 4 hurricane. The most negative tweets were seen several days before and the several days during Irma's passage through the US. This indicates some level of discontent with the disaster relief in Florida, while the other states were neutral in their response as tweet activity subsided quickly. Since the forecasts for the more inland states was worse than the actual impact which occurred, negative sentiments seen during the pre-impact period were most likely the result of worry and apprehension. The least positive counties were also more often located inland rather than along the coast over many of the days, which runs counter the expectation that coastal regions would be more negative both before and after the hurricane as they often feel the effects before the storm makes landfall and experience the most intense storm surges and flooding. Overall, the counties mostly expressed near neutral to positive sentiments, indicated by the green-yellow color on the map, or slightly positive sentiments, indicated by the yellow-orange colored areas. The lack of overwhelming negative sentiments and greatly reduced twitter activity relating to hurricane Irma after the storm suggests that the residents in these states were not significantly dissatisfied with the disaster relief efforts of their states and counties. However, many of the counties which did remain active after the storm in Alabama and Georgia reported an average sentiment which was negative, and there was a great deal of variance between the sentiments of active counties in Florida. This suggests that some areas did not receive the attention needed to recover from the hurricane quickly. Some examples of this was Miami-Dade, Leon, and Bay county of Florida, which consistently display a neutral or slightly negative sentiment in the two weeks following the storm; Miami-Dade is the most populous county in Florida and received some of the most extensive damage in the country.

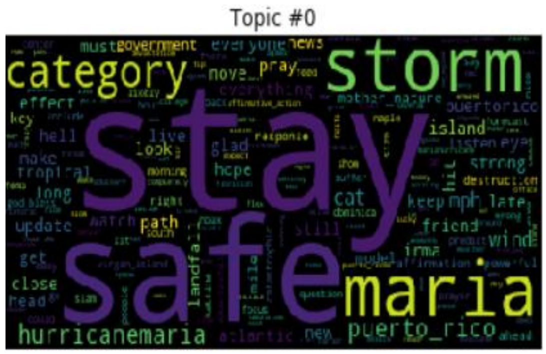
6.3 Topic Modeling

In the following subsections we discuss the results of the LDA topic model analysis of the Irma twitter data. The first results provided were produced from a topic model with an assigned K value of 8 and without the removal of custom selected words; this is to demonstrate how topics may be repetitive within a model, despite having a higher coherence value. It also shows how the filtering method used to select tweets for this study resulted in an inherent bias toward specific topics. The second topic model results show how this problem was overcome by selective word removal and reducing K to a value which produced the best results while still maintaining decent coherence.

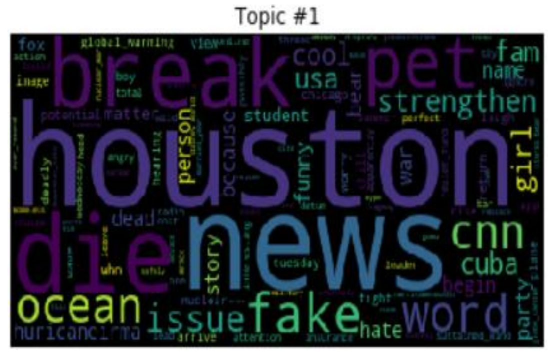
6.3.1 High K LDA Model with Unaltered Text Preparation

The following results were obtained using a K value of 8 and including filtering words such as “hurricane” and “irma”, as well the omittance of the further text preparation for topic modeling. The results as well as the top five words returned for each topic subject can be seen in Figure 15 and 16.

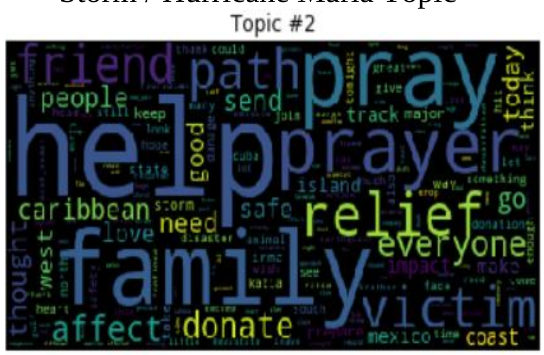
It is clear from the results that the topic subject “Hurricane Irma” is the most frequent topic overall even if it has various subtopics subjects, and these topics often contain the same top words. Other hurricanes, such as Maria, and regions previously impacted by hurricanes, such as Houston, also occur within the results. These results represent topics which we are not necessarily interested in, as they involve storms and regions other than those which are of interest for this study. There are still several topics which are quite unique, such as topics 2 and 6, which focus around subjects which are related to Irma but distinct from the storm itself. However, these are only two results out of 8 total topics and other results within the set are incoherent or unclear, such as topics 5 and 7. These results demonstrate that despite the $K = 8$ model having a coherence value of nearly 0.25, these results are not necessarily better than those with less topics. Additionally, it shows how results can be greatly improved by the addition of more rigorous text preparation as many of the topics from this model include words which are either unrelated, vague, irrelevant, or just nonsense. A key example is the top word for topic 6, which is simply “S”, or the appearance of multiple tenses of one word within a single topic



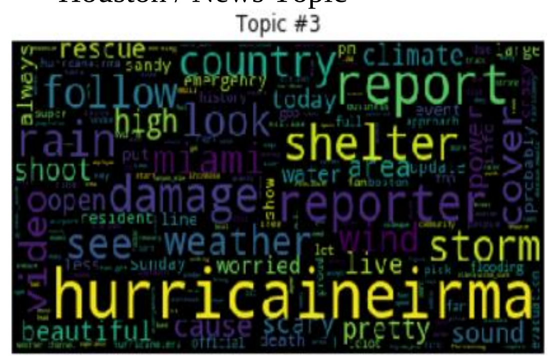
Storm / Hurricane Maria Topic



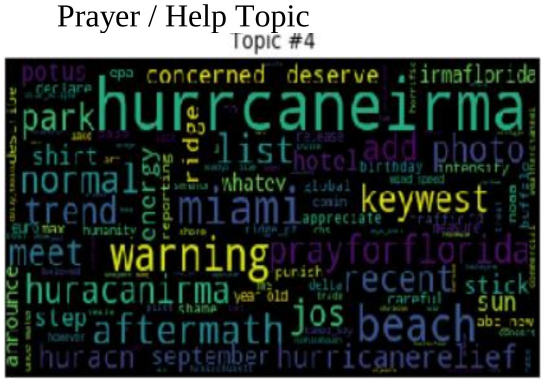
Houston / News Topic



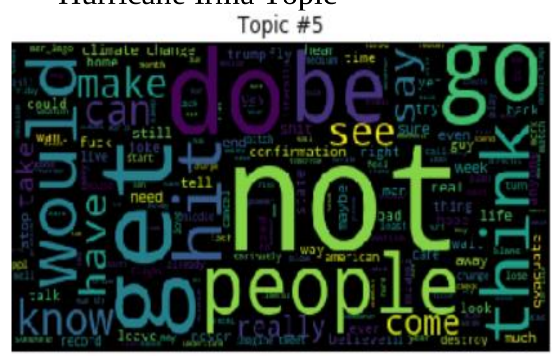
Prayer / Help Topic



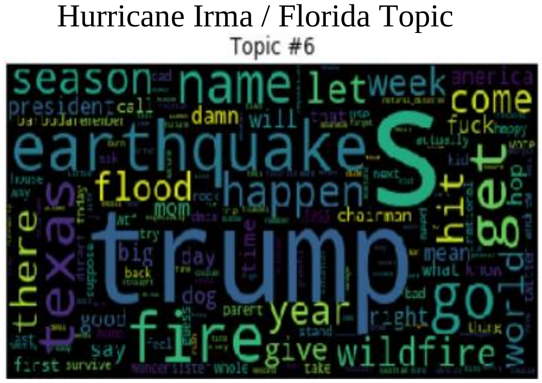
Hurricane Irma Topic



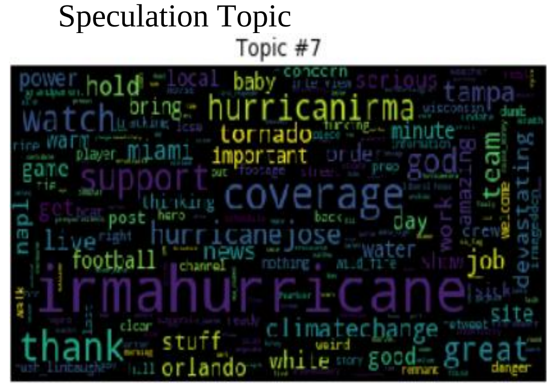
Hurricane Irma / Florida Topic



Speculation Topic



Trump / Disasters Topic



Hurricane Irma / Junk (Unsorted)

Figure 15: Word clouds representing the LDA topic model results using a value of K = 8 and excluding additional text cleaning processes. Topics are designated from 0 to 7 and labelled according to their overall topic subject based on the resulting words and phrases returned in the results

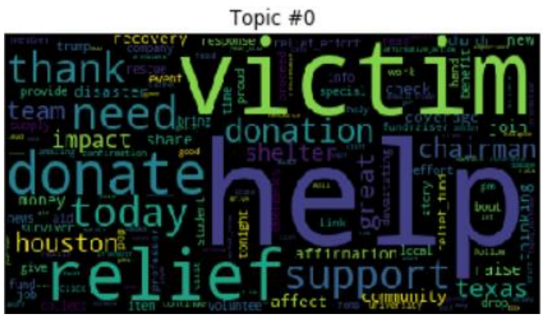
Top Five Topic Words in Each Distribution	
0. Storm / Hurricane Maria – Stay, Safe. Storm, Maria, Category	**1. Houston / News** – Houston, News, Die, Break, Pet
2. Prayer / Help – Help, Family, Pray, Prayer, Relief	**3. Hurricane Irma** – Hurricaineirma, Report, Reporter, Damage, Shelter
4. Hurricane Irma / Florida – Hurrcaneirma, Miami, Warning, Keywest, Aftermath	**5. Speculation Topic** – Not, Get, Do, Go, People
6. Trump / Disasters Topic – S, Trump, Earthquake, Fire, Get	**7. Hurricane Irma Junk** – irmahurricane, coverage, hurricanirma, thank, watch

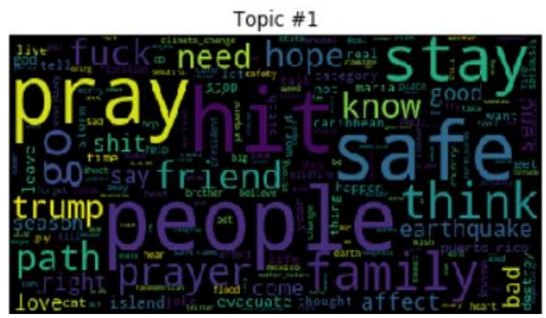
Figure 16. LDA topic model results using $K = 8$ and excluding text cleaning. The top five words within each topic result are listed.

.

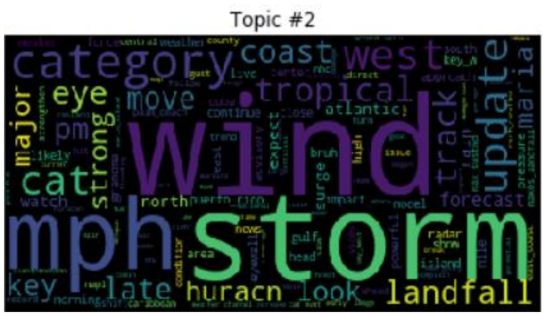
6.3.2 Low K LDA Model with Optimized Text Preparation

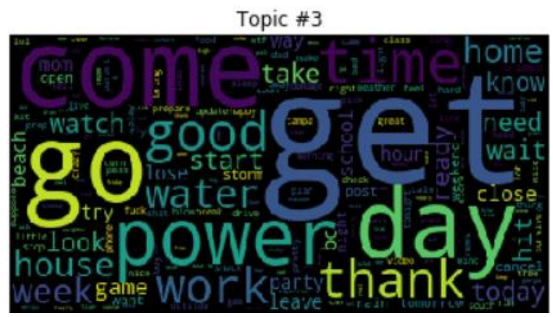
To achieve a more focused topic model, there were several changes made to the parameters of the model. As previously mentioned, the word “hurricane” and the names of the most recent set of storms during this data collection period—Matthew, Harvey, Irma, and Maria—were removed from the set as possible topics. This was done because the previous model showed a preference for these topics and since the primary focus of this research was topics related to the occurrence of Hurricane Irma they were considered redundant or irrelevant. Additionally, the *K* value was reduced to 4, which gave a coherence value of 0.212; despite having a marginally lower coherence score than the previous model it was predicted that this model would produce better results as most test models produced only between 3 to 5 clear and unique topics. The α value was assigned as ‘α-auto’ as this showed improvement in the model. The following sets of topic subjects were produced from this focused algorithm, seen in Figure 17. These were the models considered to have both strong coherence within topics and be most relevant to the purpose of this study. The most notable words seen as topic results for each model is further identified in Figure 18.





Relief Topic **Safety Topic**





Hurricane Topic **Speculation Topic**

Figure 17. Topic model results found from the focused method. Topic models are designated topic 0 through topic 3 and labeled with an identifying overall topic focus.

Top Five Topic Words in Each Distribution	
0. Relief – Help, Victim, Relief, Donate, Today	**1. Safety** – Hit, Stay, Safe, People, Pray
2. Hurricane – Wind, Storm, MPH, Category, Update	**3. Speculation** – Get, Come, Go, Day, Power

Figure 18. The top five topic words per topic model seen in the $K = 4$ modeling results seen in Figure 17.

Considering both the overall results seen in Figure 15 and the top results in Figure 16, each topic model (0-3) was assigned with an identifying overall topic label, Relief, Safety, Hurricane, and Speculation respectively. The Speculation topic is included in this discussion to illustrate the limitations of LDA topic models and describe how such vague results may be able to be interpreted.

We can see in Topic 0 that there is a clear focus on help and relief during the storm, with top words being “help”, “victim”, “relief”, “donate”, and “today”. Other mentionable topics which are seen in this model were “support”, “thank”, and “need”. Thus Topic 0 was assigned as the Relief Topic. It shows that there was some level of online relief effort being carried out due to the topics on donation, however it is notable that this specific model also has a slight focus on Texas as well, seen by the topics “Texas” and “Houston” in the results. This may be a result of greater discussion on the recovery of Texas from Hurricane Matthew than on Irma and its impacted states since recovery from Matthew was most likely still underway during this time. Additionally, the relief efforts for Irma were likely not at the point of organized donation campaigns until several weeks after its impact, which is not considered in this study. These topics do show that twitter is used as a platform for relief efforts after other hurricanes, and it can be assumed it will likely be used as a platform for Irma as well.

Topic 1 was identified as the Safety Topic, due to its apparent focus on more humanitarian and people-related issues. Its top topic words were identified as “hit”, “stay”, “safe”, “people”, and “pray”, but also contained many other notable topics such as “friend”, “family”, and “hope”. These are groups which most people would be most concerned about regarding safety during the storm, and the pair “stay safe” is once again seen together as it was in word frequency calculations in Figure 7, Section 5.1.1. An odd occurrence with this model was topics words such as profanity, which have been marked out with red, and “trump”, likely about President Donald Trump. These topics do not appear to be directly related to the rest of the results, and the reason for their presence is unknown.

The results seen in Topic 2 are the most coherent of the four models presented, all containing a strong association with physical characteristics and weather patterns related to Irma, and therefore was labeled the Hurricane Topic. The top results were “wind”, “storm”, “mph”, “category”, and “update” which are usually the key talking points when discussing a storm’s strength and location. Additional topics were “tropical”, “track”, “landfall”, “coast”, “eye”, “major”, and “strong” which all are commonly used in discussing storm features and path. There are no obvious outliers within these results compared to the other models, which could suggest that this topic focus was discussed the most in the twitter data compared to the other focuses. Since many news outlets and weather services will post updates and weather conditions via social media, this would not necessarily be surprising. It suggests that more focus on twitter may be put on real-time updates and alerts than on discussion and reporting of aftermath conditions.

Finally, Topic 3 is identified as the Speculation Topic because it appears to have a notable pattern, but it is not clear what the overall focus is. Its top words are “get”, “come”, “go”, “day”, and “power” and contains words such as “time”, “work”, “thank”, “water”, “home”, “house”, and “lock”. Many of these seem to be material belongings that people would be worried about and access to basic utilities. The rest of the results seem to relate to acting in some way and movement, possibly a result of evacuation attempts, although evacuation itself does not appear in the results. As LDA applies a certain level of randomness to its selection of results, the final topic models can vary highly in topic and quality. This topic model demonstrates that variability and serves as an example of the challenges of topic modeling using small documents sizes.

7.0 Possible Future Research Applications and Improvements

Adapting to a shift from working with mostly numerical, quantifiable data to text-based data is not easy. It was also challenging to seek out descriptive and engaging analytic methods which involve predictive modeling, as it was unnecessary for this study. The ability to be able to produce results from such unwilling data is a valuable skill, however, and developing such talents is well worth the trouble encountered. It was important to appreciate the various ways a user will communicate online that is not easily understood by a simple sentiment analysis. The use of double negatives, conjunctions, spelling errors, slang, emoticons, and other versions of shorthand must be considered as most basic sentiment analysis methods will not understand these concepts, potentially resulting in erroneous results. The key is to translate the data as much as possible while preserving the intended meaning before beginning the analysis. Or conversely approaching the monumental effort of training the underlying machine learning algorithms to be able to comprehend these complications.

The study requires further research to conclusively state that the response effort during this storm was favorable, however the data currently shows that no region was particularly dissatisfied with the performance of relief, rescue, and response services. The analysis in this study show that the response to Irma was primarily positive in the affected regions, although the average sentiment ended lower for two states—Alabama and Georgia. In these states, both of which are less commonly impacted by hurricanes, it is quite possible that relief and recovery efforts were not as well coordinated as they were in states which are more frequently impacted by such disasters, such as Florida and South Carolina. More research would need to be done into specific factors for these areas, such as what damage these states experienced, how long certain utilities and roads were nonfunctional, and if there were any storm-related events which gained criticism, in order to evaluate if the lower sentiment was indeed a result of public discontent with recovery efforts.

This study was limited by translating tweets to English, stopword removal, and the method of analysis used. A portion of the tweets collected during the data collection were non-English and included various languages. Most of the problems with language when the tweets were cleaned of any tweets not in the region of the United States. Other languages could be translated to English, but most translators do not accurately represent what was said in the other language. Translations provide issues with sentiment analysis because the nature of how people speak is important in the analysis and through translation this may be lost. Another issue with translation of tweets is the way people communicate online. Online communication is an issue because people use slang, emoticons, and other shorthand to communicate especially on Twitter. Not translating these for sentiment analysis limits the understanding of the method used for sentiment analysis. Stopword removal is key to sentiment analysis because it removes all the most common words used in the specific language being analyzed. Removal of stopwords allows for correct representations of frequent words in the text. The method of analysis is important to the study and was limited here because the training to twitter data was difficult to achieve. Expanding on the method of analysis and incorporating machine learning algorithms would increase the accuracy of the results found.

More advanced methods can be used in the future to allow for greater assistance in sentiment analysis. A variety of software libraries are currently in development for text and sentiment analysis and are improving with each day. Text data contained in tweets is poorly suited for basic level sentiment analysis, as they often contain few words, slang or phrases, videos and images, and emoticons. As previously mentioned, concepts like sarcasm, double negatives, negation, and expressive profanity are not understood by the current model and may have resulted in many incorrectly assigned sentiment values. The implementation of a more advanced method for determining sentiments would greatly improve the results such machine learning algorithms like Random Forest classifier, which was determined to be the most accurate algorithm in the sentiment analysis study by Dodd (2014). Additionally, incorporating sentiment analysis methods for classifying emoticons and non-text characters would be highly influential as many users primarily convey their emotion state through emojis and other characters. A unique sentiment library could also be compiled using a lexicon which is format towards a disaster type event, therefore putting negative connotations on phrase or words like "no power", "flooding", and "no electricity". Since many tweets contain a video or image with the text, the utilization of image analysis for sentiments would also be of great help in improving sentiment classification. Twitter data requires extensive cleaning and more effective methods should be utilized in future work.

8.0 Summary and Conclusions

Overall, the study shows that sentiment data can be representative of people's emotions during a natural disaster. On average, people tend to have a more positive outlook. Results show that days leading up to the disaster will experience a fall in average sentiment and this sentiment will fall even further during the actual disaster affecting the individual. After a disaster the average sentiment is increasing because of relief that it has ended and other positive thoughts that usually go with a disaster passing. The potential of Twitter data is endless, however accessing and interpreting the data has shown itself it be difficult. This data can greatly aid the improvement and optimization of relief efforts here and in other countries by evaluating the concerns voiced by individuals affected and the sentiments of users in response to current response efforts on a real-time basis. Using twitter data from cities affected by Hurricane Irma, the sentiments of individuals over the course of the disaster were determined and found to be mostly positive. The topics which were most often seen within tweets during this time centered around help and safety of others, as well as physical characteristics of the storm. These results were not unexpected but do show more focus on community than on self among twitter users. We had originally speculated that there would be more concern for users' own personal possessions, which was not seen in the topic modeling results.

The contributions of this work are as follows; (a)Developing a method for disaster relief analysis using twitter data recovered from specific regions of the United States during the severe storm Hurricane Irma by establishing an observable pattern regarding sentiment trends over the progression of the storm. (b)The identification of keywords and phrases that were most frequently used in tweets through sentiment analysis, to determine what resources or issues were most significant within a region during the hurricane. (c)Demonstrating that the sentiment analysis can measure people's emotions during the natural disaster, which may in the future allow authorities to limit the damage, effectively recover from the disaster, and improve future disaster response. (d)Evaluating related topics of interest and overall topic groupings occurring during the storm using topic modeling and studying what the topic results indicate about users' concerns.

Hurricane Irma was one of the more intense hurricanes in the recent history that hit the United States and studying the way the people responded to this storm can give insight into how the public responds to a future disaster, and how the government and federal services can limit damage and effectively recover.

9.0 Works Cited

[1] A. Anandhi, "How do extreme streamflow due to hurricane IRMA compare during 1938-2017 in South Eastern US?." In AGU Fall Meeting Abstracts. 2017.

[2] T.Zolnikov, "A Humanitarian Crisis: Lessons Learned From Hurricane Irma." , 108(1), pp. 27–28, American Journal of Public Health (AJPH), January 2018

[3] Y. Liu, R. H. Weisberg, J. Chen, C. R. Merz, J. Law, and L. Zheng. "West Florida Shelf Response to Hurricane Irma." In AGU Fall Meeting Abstracts. 2017

[4] J. Shultz, and S. Galea. "Preparing for the next Harvey, Irma, or Maria—addressing research gaps." New England Journal of Medicine 377, no. 19 (2017): 1804-1806.

[5] S. Xian, K. Feng, N. Lin, R. Marsooli, D. Chavas, J.Chen, and A. Hatzikyriakou, "Rapid Assessment of Damaged Homes in the Florida Keys after Hurricane Irma." arXiv preprint arXiv:1801.06596 (2018).

[6] K. Bayard, R. Decker, and C. Gilbert, "Natural Disasters and the Measurement of Industrial Production: Hurricane Harvey, a Case Study", No. 2017-10-11. Board of Governors of the Federal Reserve System (US), 2017.

[7] J. Roman, "Hurricane Maria: A Preventable Humanitarian and Health Care Crisis Unveiling the Puerto Rican Dilemma." Annals of the American Thoracic Society 15, no. 3 (2018): 293-295.

[8] "Google Trends," Google Trends. [Online]. Available: /trends/explore. [Accessed: 08-Oct-2017].

[9] E. Turban et al., Business intelligence and analytics: systems for decision support. Boston: Pearson, 2015.

[10] J. Dodd, "Twitter Sentiment Analysis," M.S. thesis, Data Analytics, Nat. College of Ireland, Dublin, Leinster, Ireland, 2014.

[11] L. Dini and A. Bittar, "Emotional analysis on twitter: the hidden challenge" in *LREC 2016:Proc. of 10th Int. Conf. on Language, Resources and Eval., May 2016, Slovenia*, Grenoble: Holmes Semantic Solutions, 2016.

[12] J. Suttles and N. Ide, "Distant Supervision for Emotion Classification with Discrete Binary Values," in *Computational Linguistics and Intelligent Text Processing*, 2013, pp. 121–136.

[13] Calvo, Rafael Alejandro and Sunghwan Mac Kim. "Emotions in Text: Dimensional and Categorical Models." *Computational Intelligence,* 29, pp. 527-543. 2013.

[14] Onook Oh, Manish Agrawal, and H. Raghav Rao. "Community intelligence and social media services: A rumor theoretic analysis of tweets during social crises." *Mis Quarterly*, 37.2 (2013).

[15] S. Ardon, A. Bagchi, A. Mahanti, A. Ruhela, A. Seth, R. M. Tripathy, and S. Triukose, "Spatio-temporal and events based analysis of topic popularity in twitter." In *ACM 2013:*

Pro. of the 22nd ACM int. conf. on Information & Knowledge Management, Oct 27-Nov 1 2013, San Francisco, CA, USA, pp. 219-228. ACM, 2013.

[16] E. Ferrara and Z. Yang, “Measuring Emotional Contagion in Social Media,” *Plos One*, vol. 10, no. 11, Jun. 2015.

[17] Y. Martín, Z. Li, and S. L. Cutter, “Leveraging Twitter to gauge evacuation compliance: Spatiotemporal analysis of Hurricane Matthew,” *Plos One*, vol. 12, no. 7, 2017

[18] Cutter, S. L.; Emrich, C. T.; Bowser, G. C.; Angelo, D.; Mitchell, J. T. South Carolina hurricane evacuation behavioral study. 2011.

[19] A. Leavitt and J. A. Clark. "Upvoting hurricane Sandy: event-based news production processes on a social news site." In the proceedings of *ACM 2014: Proc. of the 32nd ann. ACM conf. on Human factors in computing systems, Sept. 9-12, 2014, Southport, UK*, pp. 1495-1504. ACM, 2014.

[20] C. Caragea, A. C. Squicciarini, S. Stehle, K. Neppalli, and A. H. Tapia, “Mapping moods: Geo-mapped sentiment analysis during hurricane sandy.,” in *ISCRAM*, 2014.

[21] J. Kozák, and V.Čermák. The illustrated history of natural disasters. Dordrecht: Springer, 2010.

[22] T. Pescatore, M. Senatore, G. Capretto, and G. Lerro. "Holocene coastal environments near Pompeii before the AD 79 eruption of Mount Vesuvius, Italy”, Quaternary Research 55, no. 1 (2001): 77-85.

[23] R. Perry, M. K. Lindell, and K. J. Tierney, eds. "Facing the unexpected: Disaster preparedness and response in the United States", Joseph Henry Press, 2001.

[24] J. Phua, S.V.Jin, and J. Kim. "Uses and gratifications of social networking sites for bridging and bonding social capital: A comparison of Facebook, Twitter, Instagram, and Snapchat." Computers in Human Behavior 72 (2017): 115-122.

[25] H. Bik, and M. C. Goldstein. "An introduction to social media for scientists." PLoS biology 11, no. 4 (2013): e1001535.

[26] S. Kumar, G. Barbier, M. Abbasi, and H. Liu. "TweetTracker: An Analysis Tool for Humanitarian and Disaster Relief.", In ICWSM. 2011.

[27] P. Jaeger, B. Shneiderman, K. Fleischmann, J. Preece, Y. Qu, and P. Wu. "Community response grids: E-government, social networks, and effective emergency management." Telecommunications Policy 31, no. 10-11 (2007): 592-604.

[28] G. Beigi, X. Hu, R. Maciejewski, and H. Liu. "An overview of sentiment analysis in social media and its applications in disaster relief." In Sentiment analysis and ontology engineering, pp. 313-340. Springer, Cham, 2016.

[29] C. Caragea, A. Squicciarini, S. Stehle, K. Neppalli, and A. Tapia. "Mapping moods: Geo-mapped sentiment analysis during hurricane sandy." In ISCRAM. 2014.

[30] T. Shelton, A. Poorthuis, M. Graham, and M. Zook. "Mapping the data shadows of Hurricane Sandy: Uncovering the sociospatial dimensions of ‘big data’." Geoforum 52 (2014): 167-179.

[31] K. Lachlan, P. Spence, X. Lin, and M. Del Greco. "Screaming into the wind: Examining the volume and content of tweets associated with Hurricane Sandy."Communication Studies 65, no. 5 (2014): 500-518.

[32] H. Dong, M. Halem, and S. Zhou. "Social media data analytics applied to hurricane sandy." In Social computing (SocialCom), 2013 international conference on, pp. 963-966. IEEE, 2013.

[33] A. Bifet, and E. Frank. "Sentiment knowledge discovery in twitter streaming data." In International conference on discovery science, pp. 1-15. Springer, Berlin, Heidelberg, 2010.

[34] A. Hassan, A. Abbasi, and D. Zeng. "Twitter sentiment analysis: A bootstrap ensemble framework." In Social Computing (SocialCom), 2013 International Conference on, pp. 357-364. IEEE, 2013.

[35] J. Bollen, H. Mao, and A. Pepe, "Modeling public mood and emotion: Twitter sentiment and socio-economic phenomena." Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media 11 (2011): 450-453.

[36] B. Pang, and L. Lee. "Opinion mining and sentiment analysis." Foundations and Trends in Information Retrieval 2, no. 1–2 (2008): 1-135.

[37] H. Wang, D. Can, A. Kazemzadeh, F. Bar, and S. Narayanan. "A system for real-time twitter sentiment analysis of 2012 us presidential election cycle." In Proceedings of the ACL 2012 System Demonstrations, pp. 115-120. Association for Computational Linguistics, 2012.

[38] F. Wanner, C. Rohrdantz, F. Mansmann, D. Oelke, and D. Keim. "Visual sentiment analysis of rss news feeds featuring the us presidential election in 2008." In VISSW. 2009.

[39] R. Feldman, "Techniques and applications for sentiment analysis." Communications of the ACM 56, no. 4 (2013): 82-89.

[40] F. Cheong, and C. Cheong. "Social Media Data Mining: A Social Network Analysis Of Tweets During The 2010-2011 Australian Floods." PACIS 11 (2011): 46-46.

[41] C. Landsea, J. Franklin, and J. Beven. The revised Atlantic hurricane dataset (HURDAT2). NOAA National Hurricane Center (2015).

[42] D. M. Blei, “Probabilistic topic models,” *Communications of the ACM*, vol. 55, no. 4, p. 77, Apr. 2012.

[43] D. M. Blei, “Latent Dirichlet Allocation,” *Journal of Machine Learning Research*, vol. 3, pp. 993–1022, 2003.

[44] L. Hong and B. D. Davison, “Empirical study of topic modeling in Twitter,” in *Proceedings of the First Workshop on Social Media Analytics - SOMA ’10*, Washington D.C., District of Columbia, 2010, pp. 80–88.

[45] J. Tang, Z. Meng, X. Nguyen, Q. Mei, and M. Zhang, “Understanding the Limiting Factors of Topic Modeling via Posterior Contraction Analysis,” p. 9.

[46] Y. Hu, A. John, F. Wang, and S. Kambhampati, “ET-LDA: Joint Topic Modeling for Aligning Events and their Twitter Feedback,” *Proceedings of the National Conference on Artificial Intelligence*, vol. 1, Nov. 2012.

[47] Kirkorian, R. (2013). New Tweets per second record, and how!. [Blog] *Twitter-Engineering*. Available at: https://blog.twitter.com/engineering/en_us/a/2013/new-tweets-per-second-record-and-how.html [Accessed 6 Sep. 2018].