University of Zurich[UZH]

What Roles do social media Play in Hurricane Ian, Before, During and After the Event

GEO 511 Master's Thesis

Author
Mingyang Yuan
21-740-030

Supervised by
Prof. Dr. Ross Purves

Faculty representative
Prof. Dr. Ross Purves

21.09.2023
Department of Geography, University of Zurich

Universität Zürich[UZH]

What Roles do social media Play in Hurricane Ian, Before, During and After the Event

GEO 511 Master's Thesis

Author

Mingyang Yuan

21-740-030

Supervised by
Prof. Dr. Ross Purves

Faculty representative
Prof. Dr. Ross Purves

30.09.2023
Department of Geography, University of Zurich

Abstract

In recent years, natural disasters like wildfires, tsunamis, and floods have surged in both severity and frequency, causing widespread harm, including physical damage, loss of life, economic turmoil, and societal unrest. Among these disasters, hurricanes, defined by wind speeds surpassing 74 mph, pose a persistent threat, bringing hazards such as heavy rainfall and inland flooding. Hurricane Ian, one of the most significant in recent U.S. history, formed on September 23rd, hit Florida on September 28th, and dissipated on October 2nd, leaving widespread devastation. In the realm of disaster management, Location-Based Social Media (LBSM) has emerged as a crucial tool, aiding in early warnings, damage assessment, rescue coordination, and recovery evaluation. This thesis focuses on the analysis of English and Spanish tweets related to Hurricane Ian, covering the period from its formation to 50 days after its dissipation. The tweet datasets were divided into two categories: all tweets and the top 1% most shared tweets. Employing the Latent Dirichlet Allocation (LDA) model, the study unveiled prevalent themes within the tweets over different timeframes. Additionally, sentiment analysis was conducted on both English and Spanish tweet datasets, using the Valence Aware Dictionary and sEntimentReasoner (VADER) model for English tweets and Vader-multi for Spanish tweets. This aimed to capture the evolving sentiments of individuals and their emotional responses to various topics. The findings reveal Twitter's effectiveness as an early warning system and a valuable tool for risk assessment and recovery. Leading up to the hurricane's landfall, discussions mainly revolved around weather and disaster-related topics. During and after the hurricane, the focus shifted to disaster-related and situational topics. Sentiment analysis indicated a growing negativity as the storm approached, followed by a gradual return to less negative sentiments after the hurricane passed. This thesis emphasizes the significance of social media platforms as essential resources for rapid decision-making during crises, particularly when quick responses are imperative.

Keywords: Natural Disaster, LBSM, Hurricane Ian, Topic Modeling, Sentiment Analysis

Acknowledgements

I would like to express my gratitude to the one who helped and supported me during the journey of my master's degree and thesis.

- My supervisor, Prof. Dr. Ross Purves. Thank you for your supervision, regular meetings, the great ideas you provided as well as the care and support.
- My friends give me valuable feedback. Thanks to all my friends who I can share my ideas with and give me support in my work and personal life.
- My family for their support. I would like to thank my parents for their love and care. I would also thank my future family, my puppy, for giving me hope and expectation.

Table of Contents

List of Figures

List of Tables

1. Introduction

In this section, the motivation of this thesis will be interpreted, including the impact of current events of natural disasters. Furthermore, the reason for choosing hurricane as the specific type of natural disaster, Hurricane Ian as the focused event, and the reason of linking natural disaster to social media. Research questions and hypotheses will also be stated at the end of this section.

1.1 Research Motivation

Natural hazards, or natural disasters, are defined as "environmental phenomena that have the potential to impact societies and the human environment" (FEMA, n.d.). They are different from human-made hazards, which involve human activities in the cause of an event. Natural disasters are slightly different than natural hazards but related. Natural hazards are related to the threat of events that will likely cause a negative impact. Natural disasters are the negative impacts of natural hazards that already occurred and caused significant harm to society. 18 common natural hazards are listed by the Department of Homeland Security of the USA: "avalanche, coastal flooding, cold wave, drought, earthquake, hail, heat wave, hurricane, ice storm, landslide, lightning, riverine flooding, strong wind, tornado, tsunami, volcanic activity, wildfire, winter weather". Examples of some severe natural disasters in recent years are shown in *Figure 1.1*.

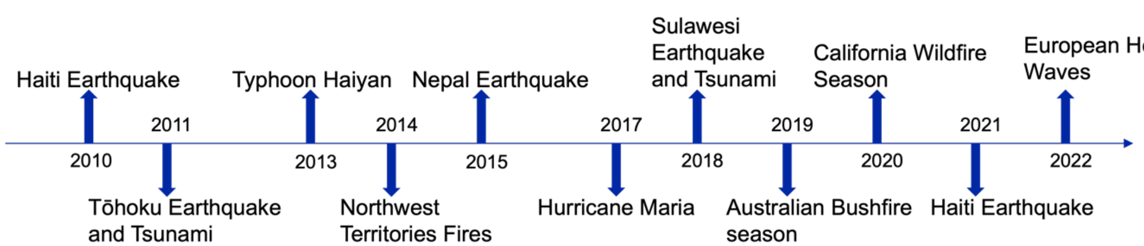


Figure 1.1. Examples of severe natural disasters from 2010 to 2022.

Natural disasters cause loss of life, physical damages, economic losses, and sometimes social unrest. According to the report of the United Nations Office of Disaster Risk Reduction, between 1998 and 2017, 1.3 million people were killed in natural disasters, and 4.4 billion people were affected, including injured, homeless, displaced, or in need of emergency assistance (Feng et al., 2022; Wallemacq & House, 2018). The direct economic losses of those countries hit by disasters were around US$ 2,908 billion and the loss reported has risen by 151% in these 20 years. Most of the life losses were due to geophysical disasters, such as earthquakes and tsunamis. However, over 90% of natural disasters are climate-related disasters, such as floods, storms, drought, etc., which are the major cause of economic losses. As extreme weather and natural disasters become more and more frequent because of climate change (Van Aalst, 2006), researchers and governments must find an efficient way to reduce the negative impacts of natural disasters.

1.1.1 Why Hurricane

A hurricane is a type of storm formed over tropical or subtropical waters, therefore it is called a tropical cyclone. It is a low-pressure weather system that would cause organized thunderstorms, but not a front, the boundary separating two air masses of different densities. A storm would be called a hurricane when its maximum sustained winds reach 74 mph (NOAA, 2021). As listed by the National Hurricane Center and Central Pacific Hurricane Center, the major hazards associated with hurricanes include storm surge, the abnormal rise of water caused by the winds of a storm; Storm tide, the rise of water caused by storm surge and the astronomical tide during a storm; Heavy rainfall and inland flooding, which relate to the geography of the area, as well as the speed and size of the storm; High winds, which could cause the damage of buildings; Rip currents, the waves formed by strong winds that break along the coast and flow away from shore; Tornadoes, relatively short and weak, yet threatening (NOAA, 2023). A hurricane risk map provided by the Federal Emergency Management Agency (FEMA) of the US is shown in *Figure 1.2*.

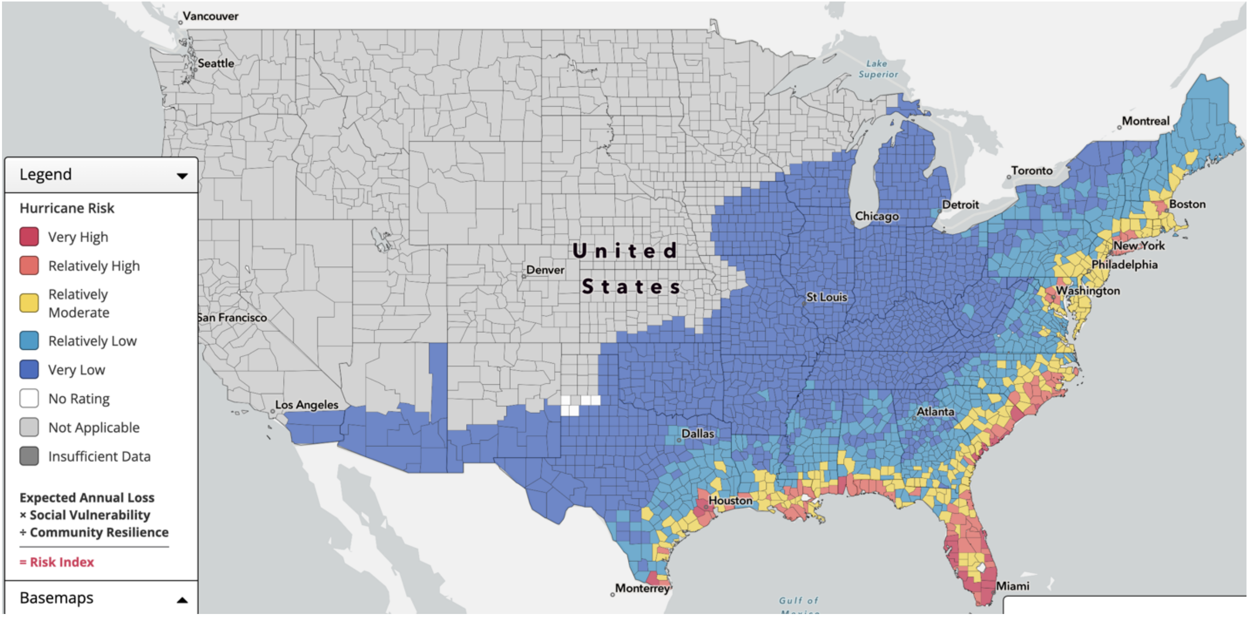


Figure 1.2.National Hurricane Risk from FEMA (https://hazards.fema.gov/nri/hurricane).

A hurricane has five categories based on its maximum sustained winds, and the higher number of categories indicates that this hurricane has a higher potential for property damages (as shown in *Table 1.1*). The hurricanes usually strike the coastal areas around the Pacific and Indian Oceans. In the United States, hurricanes cost around 9.5 billion dollars on average each year, mainly due to the Atlantic hurricane season, which occurs from June 1st to November 30th (Atlantic Oceanographic & Meteorological Laboratory, 2021). The East Coast of the United States, especially Florida, is the major part hit by hurricanes, which are in the Atlantic basin, with the maximum activities occurring in early to mid-September.

Table 1.1. The Categories of Hurricane Strength (Saffir-Simpson Scale).

Category	Miles Per Hour	Meters per Second
1	74 – 95	33 – 42
2	96 – 110	42 – 49
3	111 – 129	49 – 57
4	130 – 156	58 – 69
5	>= 157	> 70

1.1.2 Hurricane Ian

Hurricane Ian was a Category 4 (maximum strength reached Category 5) storm that formed on September 23rd, landing in the US, firstly in Florida on September 28th and dissipating on October 2nd, 2022 (Court et al., 2022). It was forecasted that Hurricane Ian would be one of the costliest hurricanes in the history of the US, and possibly the costliest one in Florida in 50 years with an estimated loss of US$67 billion (RMS, 2022). It is the most recent big hurricane event to hit the US, where the major language used is English, which makes it an ideal research area for this master thesis.

The report of Hurricane Ian from the National Hurricane Center (NHC) of the US was published on April 3rd, 2023. Ian produced damaging storm surges, destructive winds, and catastrophic flooding across central and northern Florida. It also affected the power system of western Cuba; besides, part of Georgia, North Carolina, and South Carolina had been influenced by the strong winds as well. According to the statistical records, Ian is the costliest hurricane in the history of Florida, and the third costliest in the history of the US. Ian was responsible for over US$112 billion worth of damage, and at least 156 direct and indirect deaths, 66 of which were considered direct deaths caused by the storm (Bucci et al., 2023). 90 casualties were caused by Ian indirectly, mostly in Florida. The major causes of death and casualties were limited access to timely medical treatment, various types of accidents, and cardiac events. Buildings, structures, roadways, and crops were destroyed or damaged because of the flooding and winds. Between September 28th and October 1st, an estimated 9.62 million people lost power in the United States.

1.2 Location-Based Social Media

It is important that the government warn the public about upcoming events, quickly respond to the disaster, and monitor the progress of post-disaster recovery. New tools should be involved in informing the public about potential danger, rescuing, and providing help for people in need. The location-based social media (LBSM) has been proven that it is reliable when analyzing spatial-temporal information and how it could help with rescue, risk assessment as well as community interaction and support (De Longueville et al., 2009; Page-Tan, 2021). Researchers also discovered that communities that actively use social media recover faster than those that are less active on social media (Page-Tan, 2021). It seems that the use of social media has larger potential in the analysis of disaster, whether in the prevention before the disaster, the rescue during the disaster, or the recovery after the disaster. During Hurricane Ian, the NHC was responsible for decision support services and public communication. NHC provided 16 live briefings on Ian, 12-hour activated media pool, and 13 live stream broadcasts via YouTube Live, Facebook, and Twitter.

1.2.1 Twitter

Twitter (now changed its name as “X” in July 2023. Nevertheless, in this master thesis the name “Twitter” refers to Twitter) is a real-time microblogging social media platform that currently has 396.5 million users (Iqbal, 2022), with 237.8 million average monetizable daily active usage (mDAU) according to the second quarter 2022 operational and financial report of Twitter (Twitter Inc., 2022a). There were 76.9 million users in the United States as of January 2022, which made up around 20% of the total active users (Dixon, 2022). The users could read and post messages, called “tweets”, which contain up to 280 characters, up to four photos, a GIF, or a video (Twitter Help Center, 2022). A tweet can be posted as a general tweet that is seen by other users, it could also be replied to, quoted, and retweeted. A user can mention other users by adding “@username” and use direct messages to communicate with each other. A Retweet is a tweet that the user shares with their followers publicly, which could add the user’s own comment or media when retweeting. Retweeting is known as a great way to share or spread information on Twitter (Twitter Help Center, 2023a).

Twitter could provide temporal and spatial data. For temporal information, each tweet has a time when it has been created. For each user, a timeline of tweets is organized and displayed in reverse chronological order. The accuracy of the created time of a tweet is one second. Spatial information is more difficult to identify. A user’s location provided in the profile could imply where this user lives, however, this location is not necessarily where the tweet was created. By default, the location of a tweet is off, and the users can always choose to hide their location. The precise location of a user could be detected once it has been enabled. GPS information (latitude and longitude) will be provided when the user chooses to attach a location to the tweet. The location, known as “geotag”, might be detected and suggested by Twitter using GPS, or manually typed and selected by the user (De Longueville et al., 2009; Twitter Help Center, 2023b).

Twitter Inc. provides Twitter API (Twitter Inc., 2022b) to researchers for retrieving and analyzing Twitter data programmatically, which makes Twitter a great resource for scientific social media textual analysis. However, the free full access to tweet archiving for academic researchers is no longer available from 2023. The free version of Twitter API only provides access to tweet creation, upload, and log in to the account. Higher-tier access to Twitter API, such as Twitter API Pro or Twitter API Premium which supports tweet pulling is fairly expensive.

1.3 Research Question

In this master thesis, the roles of social media, i.e., Twitter, played in Hurricane Ian, before, during, and after the hurricane will be analyzed and discussed. The hypotheses are made as following:

Hypothesis 1: In the event of Hurricane Ian, Twitter provided information about early warning, damage assessment, rescue scheduling, and recovery.

It has been researched that social media could contribute to the early warning system, immediate damage assessment, rescue scheduling, and analysis of the short- and long-term effects of natural disasters on people and communities (will be introduced in *Section 2*). In this thesis, topic modeling will be applied to the tweets of different periods to prove the hypothesis.

Hypothesis 2: The emotion or sentiment of people who tweeted about Hurricane Ian became more negative as the storm got close and recovered to normal (or less negative) after the hurricane dissipated.

Sentiment analysis will be applied to the tweets of different periods to prove the hypothesis. The changes of negative sentiments towards the event and the topics will be investigated.

2. Related Work

In this section, the previous studies related to the roles social media played in the events of natural disasters will be discussed, including early warning systems, risk assessment and rescue scheduling, and post-disaster recovery. Two main methods, topic modeling and sentiment analysis, that have been frequently used in the research of social media and natural disasters will also be introduced. Each section includes a summary or introduction of the role or method, followed by several case studies that have been examined in other research. Research gaps will be introduced in this section as well.

2.1 Early warning system

An early warning system (EWS) is a mechanism for informing people with related and punctual information so that people can be prepared for the upcoming event. Differing from traditional warnings that focus on the accuracy of the information, EWS aims to provide more punctual information to lead people to take action without hesitation (International Federation of Red Cross and Red Crescent Societies, 2020; Kitazawa & Hale, 2021). In natural disasters, EWS is important to the government and community for making timely decisions and taking actions in advance, thus, reducing the potential damages. EWS should be "end-to-end". The basic elements of EWS are risk knowledge, monitoring, response capability, and warning communication (Syed et al., 2021; UNDRR, 2007). To make a successful EWS, all elements should function properly and effectively. EWS should also be "people-centered", which requires public involvement. Social media could be a good source of warning communication, as it contains vast amounts of near real-time data and can be spread fast among people. Researchers are studying the use of social media in the EWS of different natural disasters, such as typhoons, tsunamis, and hurricanes (Bui, 2019; Jayasekara et al., 2021; Kitazawa & Hale, 2021; Wu & Cui, 2018). It is shown that social media has been used in communication before and during natural disasters, and the number of users is growing. The limitation of applying social media in the EWS might be the reliability or the issues of trust of the information shared by individual users, and how to translate the warning information into actions.

Twitter, as one of the most popular LBSNs, is frequently used as a source of spatial-temporal information in the analysis of natural disasters. The possible roles of Twitter in supporting emergency warning and planning, as well as risk and damage assessment were studied during an event of major forest fire in the South of France in July 2009 (De Longueville et al., 2009). In this study, 346 tweets were collected, comprising those posted more than one hour before the fire started and five hours after the fire was announced "under control" by the government. Several contents were extracted from the tweets, including the users' locations, geotagged placenames that were cited in the tweets, domains of the full URLs contained in the tweets, and more. The tweets had been studied in temporal, spatial, and social dynamics. In temporal dynamics, a local journalist had first posted a tweet related to fire, which maybe because the fire started in a less populated area and did not cause much threat to the public. As the fire became bigger and moved closer to the populated areas, more tweets were generated. The tweets peaked around the moment when the flames could be seen clearly, smoke and flying ashes severely influenced the nearby citizens. A lot of retweets were shared among citizens during the period of visible flames, as well as direct messages and mentions. Four main types of geographic information could be used in spatial analysis, including spatial terms, direct placenames, coded placenames, and location parting. The coded placenames were used in spatial analysis in this study. The area of cited places grew bigger as time went by. The path of fire was somewhat shown by the growing spatial information cited in the tweets; however, the areas of geo data and the actual fire were not identical. In social dynamics, the identity of users who tweeted was characterized into three categories (citizen, media, and the role as an aggregator). 64% of the users were identified as citizens, who contributed to around 55% of the total tweets. This result showed that the primary source of information on Twitter was from citizens. About 31% of tweets were published by aggregators. As secondary providers of information, aggregators did not provide timely information nor add too much more value to the spread of information. However, it is difficult to distinguish between primary and secondary providers of information.

It has been studied that the combination of social media and geo-location could provide timely information and help make the early warning system more efficient (Wu & Cui, 2018). Hurricane Sandy, the most destructive hurricane of 2012 in the US was chosen for this study. To find out which factors could contribute to the building of an EWS, both the volume and content of data were analyzed. Different levels of data had been studied, including national level, state level, and Zip Code Tabulation Areas (ZCTAs), to figure out whether the combination of location information and social media could make an ESW better. To make the geo-information of tweets identifiable, they were reversed to common geographic data (latitude and longitude). Google Map API and OpenStreetMap were used to reverse the geocode in Twitter. The result showed that the trends of the total number of tweets and the affected ZCTAs were similar. Both curves were steady before the disaster and suddenly increased to the peak value, then sharply decreased back to the average value. The trends were the same for any level of the places, which means national data and state-level data share the same pattern.

EWS is not solely about warning, it also involves educating people about how to react to the warnings. The Japan Meteorological Agency (JMA) operates a system that expects the public to respond fast, including Awareness, to pay attention to any disaster-related information; Preparation, to find safe places and be prepared for evacuation; and Action, to protect themselves immediately when facing a disaster (Kitazawa & Hale, 2021). In the case study of Typhoon Etau of Japan in 2015, the way people responded on social media to EWS was discussed. The geolocation of the collected tweets was extracted based on the content. Keywords in the text that related to the location were used to assign the tweets to Japanese states, cities, or towns. Google Map API was also used for location reverting. The number of disaster-related tweets and time had been examined. As the typhoon moved from the south to the northeastern part of Japan, the peaks of the number of tweets in each region also followed the pattern where the typhoon moved, which means the southern region reached its peak first, followed by the northern regions. Retweets were common in this event, which could be used to spread information about evacuation instructions. Around 60% of the tweets had been retweeted in most states, while the retweets of the states that suffered from more devastating damage were around 70% of the total number of tweets. In this study, public attention acted fast on social media, while preparation and action took more response time. Awareness of the disaster was talked about throughout the event, however, with a significant delay between a warning and its discussion.

In the case study of Hurricane Maria in Puerto Rico in 2017, the role that social media information played in EWS was examined (Bui, 2019). With the wide use of smartphones, the use of social media has grown significantly. Social media can spread information very fast among individuals, while it can also spread rumors. There were affordances and limitations of social media in a natural disaster event. Data were collected from official sources, such as official agencies, NGOs, and universities, as well as from individuals from the community. English and Spanish were used in the data-collecting process, as many of the interviewers were more comfortable with Spanish. In an EWS, rumors on social media could be one of the biggest challenges. To mitigate the effect of misinformation, the official agency, in this case the National Weather Service (NWS) would be responsive to the rumors on social media. NWS used both English and Spanish to denounce information about early warnings, while English was the primary language and the Spanish version usually was late for at least an hour. The study has found that EWS still highly relies on traditional communication technology, such as television and radio, but social media become more and more prevalent. People were communicating with their family and friends on social media when it was possible to do so. Around 94.5% of Puerto Ricans speak Spanish, and the native speakers were highly relying on Spanish-language media and social media, instead of the official source. This result may be due to the delayed translation of Spanish from the NWS. To make the EWS more efficient, social media should be used properly and the issue with rumor or misinformation should be considered.

2.2 Risk Assessment and Rescue Scheduling

After the application of EWS in the early phase of a natural disaster, risk assessment and rescue scheduling would be the next spotlights during a disaster. Fast damage assessment and response are required for the government and humanitarian organizations to understand the situation of the destruction and to plan rescue according to it (Nguyen et al., 2017). As people use social media for sharing disaster-related information, there would be useful data for crisis detection and damage assessment. In the research on Hurricane Sandy (Kryvasheyeu et al., 2016), the estimation of damage from official data from the Federal Emergency Management Agency (FEMA) and the data of insurance claims are analyzed against Twitter activity in the same region. The researchers examined the relationship between the activity of Twitter to the severity of damage and discovered a positive correlation. In this case, the number of tweets or the frequency of people's tweets could be a useful indicator of the damage assessment during a disaster. Retweet behaviors have been examined in the study of the 2009 Red River Valley flooding and the 2009 Oklahoma fires (Starbird & Palen, 2010). It has been found that compared to non-retweets, retweets were more likely to be related to a disaster. Formal information sources, such as local official agencies for emergency management were more reliable and had more valued sources for information that related to disaster. The users of Twitter who tweeted about a disaster were talking about distinguished information, related to the users' location proximity to the disaster.

Twitter, as an open platform that everyone has access to post information, contains a lot of informal information that has not been proven true. Although the official accounts have more valued sources of information as discussed above, in a natural disaster event, the eyewitness is also an important source of information. Identifying the accounts or messages with real or direct eyewitnesses could be useful in damage assessment. The characteristics of eyewitnesses for different types of disasters have been examined and the relative dataset and methodology were prepared for distinguishing whether a message is from an eyewitness in future research (Zahra et al., 2018). Tweets related to earthquakes, hurricanes, and floods were collected and manually analyzed to separate the accounts into three categories: direct eyewitnesses, indirect eyewitnesses, and vulnerable accounts. The characteristics of these three categories were also identified. Direct eyewitnesses were usually the people that suffered from the disaster, who described the severity of the situation and used various senses like "see", "feel", or "hear". Indirect eyewitnesses were those who shared information received from their families and friends who were direct eyewitnesses. The messages of indirect eyewitnesses often involved words that express their emotions, including thoughts and prayers. The vulnerable eyewitnesses in most of the situations shared warnings or alerts about the situation.

The spatial and temporal evolution of a natural disaster based on social media information was studied using Hurricane Sandy (Guan & Chen, 2014). "Degree of disaster", the Disaster Related Ratio (DRR) was introduced to assess the damage, which is calculated using the number of disaster-related tweets divided by the total number of tweets in the target area. In this study, the tweets were simply separated into two categories, disaster-related and non-disaster-related. The result has shown that in large urban areas, the number of disaster-related tweets is highly correlated with general tweets. Coastal flooding, growing population, and the increasing development of coastal areas make the coastline an important area in hurricane events. The results have shown that coastal areas, especially those closer to the coastline had higher DRRs, as they were more impacted by the hurricane. The DRRs declined as the distance between target areas and the coastline became larger, which was significant within five kilometers of the coastline. The pattern of power outages was consistent with the changes in DRRs, where the DRR dropped faster and had a faster recovery speed.

The damage of Hurricane Sandy in New York and New Jersey was studied using Twitter records (Wu & Cui, 2018). The correlations between the total damage losses, Twitter activities, and geoinformation were examined. The sum of disaster-related tweets was changing during the disaster, which was related to damage losses. In this study, the correlation coefficient between the disaster-related tweets and damage losses rose sharply three days before the hurricane landed, and slightly increased with time in the urban area of New York and New Jersey. The damages and Twitter activities also showed a high level when the users were close to the coastline, which means it is more likely to have higher losses in the places near the coastal region. This result is the same as discussed above in the study of Guan and Chen (2014).

Not only textual information can be used in natural disasters to assess the level of damage, but images from social media are also useful sources (Nguyen et al., 2017). This study is an extension of the Artificial Intelligence for Disaster Response (AIDR), which is a system combined with human and machine intelligence that was created in 2014 at Qatar Computing Research Institute (QCRI) (Imran et al., 2014). AIDR is a system collecting real-time tweets of a place suffering from a natural disaster and helps officials plan relief activities. In this study, imagery data from social media was tested whether it could be used in identifying the level of damage (severe, mild, or low) in a natural disaster. Traditional computer vision techniques, such as the model of Bag-Of-Visual-Words (BoVW) and the deep learning technique, Convolutional Neural Networks (CNN) were used for image classification. Images from several past natural disasters and Google images were used as labeled data. It is difficult to classify the level of damage as it is highly subjective. The result of the trained imagery classification model showed that the model could attain reasonable accuracy, however, performed less accurately on the mild level of damage. A better and clearer standard of “severe” and “mild” levels of damage should be further determined for different regions.

Scheduling algorithms are created to optimize the time and usage of resources under a limited situation, which aims to ensure fairness as well as maximize resource utilization. Another deep learning model based on Bi-directional Long Short-Term Memory (BLSTM) and CNN was built for disaster-related tweet classification and rescue scheduling (Kabir & Madria, 2019). To extract the location without location information, this study tried to collect the users’ profile data and the textual location information that appeared in the tweets. The priority of rescue was determined by the requests of a task, the number of rescue units needed (i.e., processors), arrival time, and time required to complete the task. The model could separate the rescue asking tweets into six classes and estimate the priority scores for each task. A multi-task hybrid scheduling algorithm was created for better efficiency of rescue if several tasks had different priorities but were around the same area.

2.3 Recovery from Natural Disasters

Natural disasters cause physical damage both to the properties of people and the infrastructure of society. They could also cause damage to human lives like physical and mental harm, economic damage, homelessness, etc. The recovery of people and communities after a natural disaster is another important stage of disaster management. According to the United Nations, recovery is “restoration and improvement, where appropriate, of facilities, livelihoods and living conditions of disaster-affected communities, including efforts to reduce disaster risk factor” (UNISDR, 2009). Research has been done on various types of natural disasters and different social media platforms (Jamali et al., 2019; Ogie et al., 2022). The findings include multiple aspects of the recovery from disaster, such as financial support, social cohesion, reconstruction and infrastructure services, social-economic and physical well-being, information support, mental health and emotional support, and economic activities.

With the increasing usage of social media worldwide, it has been an important resource for disaster analysis, including post-disaster recovery. A systematic literature review was done to analyze the social media platforms that were most frequently used, different patterns of use of social media for each type of disaster, and their temporal and spatial variations (Ogie et al., 2022). The studies related to social media used in post-disaster recovery before July 8^{th}, 2021, were collected. A total of 108 articles were included in the review, around 60 of them were about natural disasters in the US. Hurricanes, earthquakes, floods, and typhoons were the most frequent disasters. Twitter and Facebook were major resources of the social media data. Seven aspects of recovery were included in the review, including donations & financial support, business & economic activities, information support, mental health & emotional support, reconstruction & infrastructure services, socioeconomic & physical well-being, and solidarity & social cohesion. In the review, social media was useful in donation and financial support because of its fast and vast information. It supported social solidarity and cohesion because people usually seek emotional support on social media. Reconstruction was a long-term process that could be supported by social media data in the decision-making process. People tended to talk about their socioeconomic well-being on social media, while less frequently talking about physical wellbeing. Social media could help with information sharing by improving awareness of situations, enhancing communication, and providing information for the communities in decision-making. Economic recovery could be supported by social media as well, for example by posting job positions as there were often job losses after disaster.

To evaluate a post-disaster recovery, one of the three aspects would be considered: whether the environments are returned to the pre-disaster conditions, whether the community has built up to where it would have if there were no disasters, or the middle ground between the two conditions (Chang, 2010). A prior goal of recovery plans is to return an affected community to normal as fast as possible. A new method of analyzing the priorities of people who suffered from natural disasters through social media data for post-disaster recovery was introduced (Jamali et al., 2019). The model was built to identify people who experienced the disaster, predict their living locations, assess the topic they tweet about, evaluate their attributes to topics of discussion, and compare them with non-disaster experienced users. Tweets that related to Hurricane Sandy were collected, including the users' screen names. Although people tweeted about the hurricane, they may not have experienced the disaster themselves. The tweets that were sent by disaster-experienced users were filtered. The living quarters of the disaster-experienced users were estimated based on their geographical location information provided on social media and the census. The tweets were separated into four topics with selected words. A Dirichlet regression model was used to identify significant attributes and study their correlation for disaster-experienced users. The results showed that people who lived in the circle of disaster-affected areas had sent at least one tweet related to Hurricane Sandy, and they thought the hurricane influenced their daily lives. The disaster-experienced users tended to seek emotional support on social media. Different topics vary from each other in terms of disaster recovery behaviors, which will be introduced in *Section 4*.

In the case study of Hurricane Harvey, the first social media stress test as a disaster recovery mechanism, spatial analysis, and regression model were used to examine the use of social media in practical recovery (Page-Tan, 2021). The relationship between hyperlocal social media network ("Nextdoor" platform) activity and post-disaster recovery was investigated. A social network analysis (SNA) was used to obtain four major network structural properties: average degree, network diameter, path length, and clustering. The rate of recovery was calculated by spatially joining the data of 333 Nextdoor Neighborhoods. As a result, there were no significant differences in network structure during Hurricane Harvey. In other words, at the peak of the hurricane, social networks did not grow denser, clustered, or worked more efficiently around the area of Houston. For the regression model, a negative and significant relationship between the use of social media and the rate of recovery was found, which means as the use of social media increased, the number of days for the neighborhood to rebuild decreased.

2.4 Topic Modeling

Natural language processing (NLP) and machine learning are used in analyzing text and images when assessing the damages and recovery progress. Latent Dirichlet Allocation (LDA), naïve Bayes, convolutional neural networks (CNN), artificial neural networks (ANN), K-nearest neighborhood (KNN) are often used in the processing of texts and images data of social media, such as topic modeling models and sentiment analysis models (Jamali et al., 2019; Li et al., 2021; Ogie et al., 2022; Roy et al., 2020; Valentijn et al., 2020; Zahra et al., 2017). In this master thesis, topic modeling and sentiment analysis are used to separate tweets into different topics and calculate the sentiment score for each of the tweets.

Topic modeling is popular in text mining in different fields, such as bioinformatics, social-related studies, and statistics. It is an unsupervised machine-learning algorithm that could be used to discover hidden topics or structures from a collection of documents (Barde & Bainwad, 2017; Fu et al., 2022; Kherwa & Bansal, 2019; Vayansky & Kumar, 2020). A "*word*" is the fundamental unit of individual data, a "*document*" is the collection of words, and a "*corpus*" is the collection of documents, which is the entire dataset. A "*vocabulary*" is the entirety of the distinct words in a corpus, and a "*topic*" is the probability distribution given a vocabulary (Vayansky & Kumar, 2020). Topic modeling is an effective method, not just a classification or clustering tool, it could reflect meaningful results that help people better understand the collection of documents (Barde & Bainwad, 2017). There are many methods in topic modeling, including the Vector Space Model, Latent Semantic Indexing, Probabilistic Latent Semantic Indexing (PLSI), and Latent Dirichlet Allocation (LDA), which are popular among researchers (Barde & Bainwad, 2017; Kherwa & Bansal, 2019), an example of topic modeling classification hierarchy is shown in *Figure 2.1*. Topic modeling is also widely used in the analysis of geospatial data, such as evaluating geotagged Twitter data for resilience in a natural disaster (Vayansky et al., 2019), and dealing with the ambiguity of place names (Ju et al., 2016).

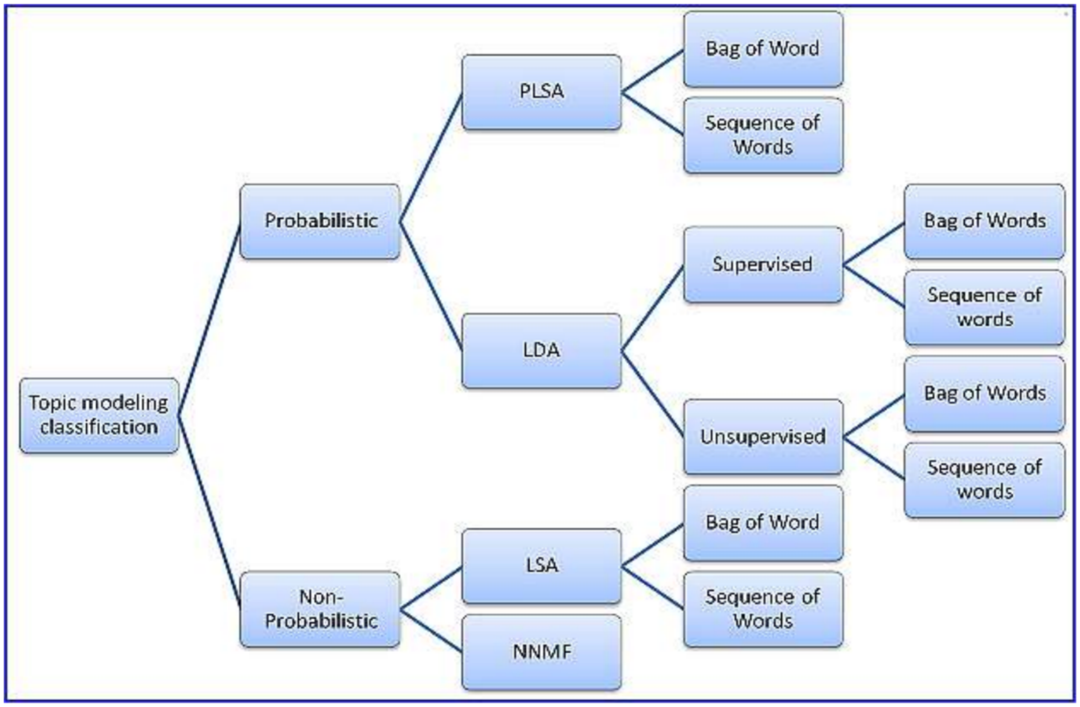


Figure 2.1. Example of Topic Modeling Classification Hierarchy from Topic Modeling: A Comprehensive Review (Kherwa & Bansal, 2019).

In the case of Typhoon Etau, it has been examined whether the EWS shifted from awareness to preparation, to action on social media (Kitazawa & Hale, 2021). Topic modeling was used to assign tweets to one of the five topic categories: awareness, preparation, action, impact, and others. Words that only appeared once and more than half of the tweets were removed, as well as the stopwords. After preprocessing, Latent Dirichlet Allocation (LDA) topic modeling was used to generate topic to tweets, and the researchers manually assigned these tweets to the categories mentioned above. In the awareness category, two main topics were generated, one discussing the intensity of typhoons, and another discussing the weather forecast. In the preparation category, two general warning-related topics and one landslide alert-related topic were generated. In the action category, a topic about evacuation was generated. In the impact category, topics about the damage or impact of the weather were generated. Less than 15% of tweets were assigned to awareness for all regions instead of Tokyo. People were mostly talking about the weather, such as the strong wind and dark sky. The preparation category has about 50% of the total tweets, especially the areas where emergency warnings were strongly needed. It showed a pattern that region with higher emergency warnings has more online activities. The action category did not contain much of the total tweets, while the impact category was high for all regions. The results indicate that people usually are aware of and prepared for a natural disaster event, although they do not take immediate action, such as evacuation.

In the study of Hurricane Sandy that focused on the use of social media in post-disaster recovery, topic modeling was used to classify the tweets into four classes (Jamali et al., 2019). Although LDA was a very popular method, this study needed a specific and accurate topic selection model, in this case, Dynamic Query Expansion (DQE), a frequency-based algorithm was used. The topics included financial, assets, community, and faith-based. A Dirichlet regression model was used to examine the relationship between the topics and attributes, such as age, income, mobility, mortgage, etc. The proportion of each topic of a user's tweet would be used as dependent variables for the Dirichlet regression model. Each topic played a major role in the process of recovery, for example, assets related topic was highly related to the users' jobs. Social interaction had a great impact on psychological recovery, people who were active on social media tended to recover faster. Disaster-experienced users tweeted less about community topics, as well as faith-based topics, compared to non-disaster-experienced users.

2.5 Sentiment Analysis

Sentiment analysis has been used to measure people's emotions or opinions toward an event. A score was calculated for each tweet to classify it as positive, neutral, or negative. The pattern of sentiment scores changing could show how people's opinions or attitudes towards an event change. Sentiment analysis has been used by researchers in multiple fields, such as finance (e.g., stock market), politics (e.g., people's opinion of an election) or disaster warnings (Khan et al., 2014; Schumaker et al., 2016; Wu & Cui, 2018). In Wu and Cui's study, the tweets that related to or contained information about Hurricane Sandy were classified as negative sentiments. The sentiment score reached the bottom when the number of tweets reached its peak. Both positive and negative tweets showed a similar pattern as the volume of tweets, however, the ratio of negative tweets was increasing as the hurricane grew stronger. This study showed that people use social media to share information about an event, which could reflect human activities and emotions. Both government and citizens who use social media could benefit from LBSNs for early warning in a natural disaster event.

There are multiple tools for sentiment analysis, such as the Python libraries Natural Language ToolKit (NLTK), Textblob, Valence Aware Dictionary, and sEntimentReasoner (VADER). These three lexicon-based tools for sentiment analysis were compared in a previous study of movie reviews (Bonta et al., 2019). NLTK used the library SentiWordNet to calculate the sentiment score. It gave each word a polarity score, including a negative score and a positive score ranging from 0 to 1. NLTK was widely used in different tasks, although it was not a “gold-standard” resource. Textblob was also popularly used in text processing. It provided consistent API access to different tasks of Natural Language Processing (NLP) including sentiment analysis (Loria, 2020). In Textblob, *polarity* and *subjectivity* were returned by the model. *Polarity* is a score ranging between -1 and +1, which indicates the negativity or positivity of the sentence. Subjectivity is a score ranging between 0 and 1, which indicates whether this sentence is a personal emotion, opinion, or judgment. VADER was a simple rule-based lexicon sentiment analysis tool (C. Hutto & Gilbert, 2014). The probabilities of the sentiment of positive, negative, and neutral of the words in the sentence were examined, a compound score of the sum of the sentiments and normalized to the range of -1 to +1. It had the quality of a “gold standard” which has also been proved by humans. In this study, 11,861 sentences of movie reviews from the website rotten-tomatoes were analyzed. The VADER model performed the best in precision, recall, F1 score, and accuracy. It was proved that VADER performed especially well in the domain of social media contexts.

Topic modeling and sentiment analysis are combined in some research to find out people’s attitudes toward different topics. In the study of Hurricane Irma, the combination of sentiment analysis and topic modeling was used to improve disaster relief and study how people in different regions reacted (Vayansky et al., 2019). Based on the polarity and subjectivity of a word, a score would be assigned to it. The sentiment score of a tweet was calculated based on the polarity and subjectivity of the words included. A “positive” tweet had a polarity larger than zero, and a “negative” tweet had a polarity less than zero. If the score was zero, this tweet would be assigned as “no emotion”, and all other results would be identified as “neutral”, which was nonpolar. The time series of daily sentiment scores of each state impacted by the hurricane were mapped, as well as the overall sentiment score for all records. The time series of the maximum wind speed of the hurricane was also mapped for comparison. The results showed that the relationship between sentiment scores and wind speed was negative, which means when the wind became stronger, the sentiment scores decreased, i.e., more negative. This trend was significant when the hurricane reached its strongest condition, and weaker for other conditions. After the disaster, areas with lower sentiment scores may need more time to recover from the hurricane.

The combination of topic modeling and sentiment analysis can be used in analyzing tweets related to global climate change (Dahal et al., 2019), overall research workflow is shown in *Figure 2.2*. Tweets with the keywords “climate”, “change”, “global”, and “warming” were collected using Twitter API. Sentiment analysis was applied to these tweets to discover the emotions or opinions of people about climate change. A positive number represented a “positive” sentiment, such as happiness. A negative number represented a “negative” sentiment, which was unhappy. A score of zero represented a “neutral” sentiment. The sentiment analysis model called Valence Aware Dictionary and sEntiment Reasoner (VADER) was used in this study. A total daily sentiment score was calculated by summing up the scores of all tweets. The results showed that the sentiment scores peaked when the temperature was high that day. On the day of Earth Hour in 2017, the positive sentiment score reached to top, because the responses were highly supportive or positive to the event.

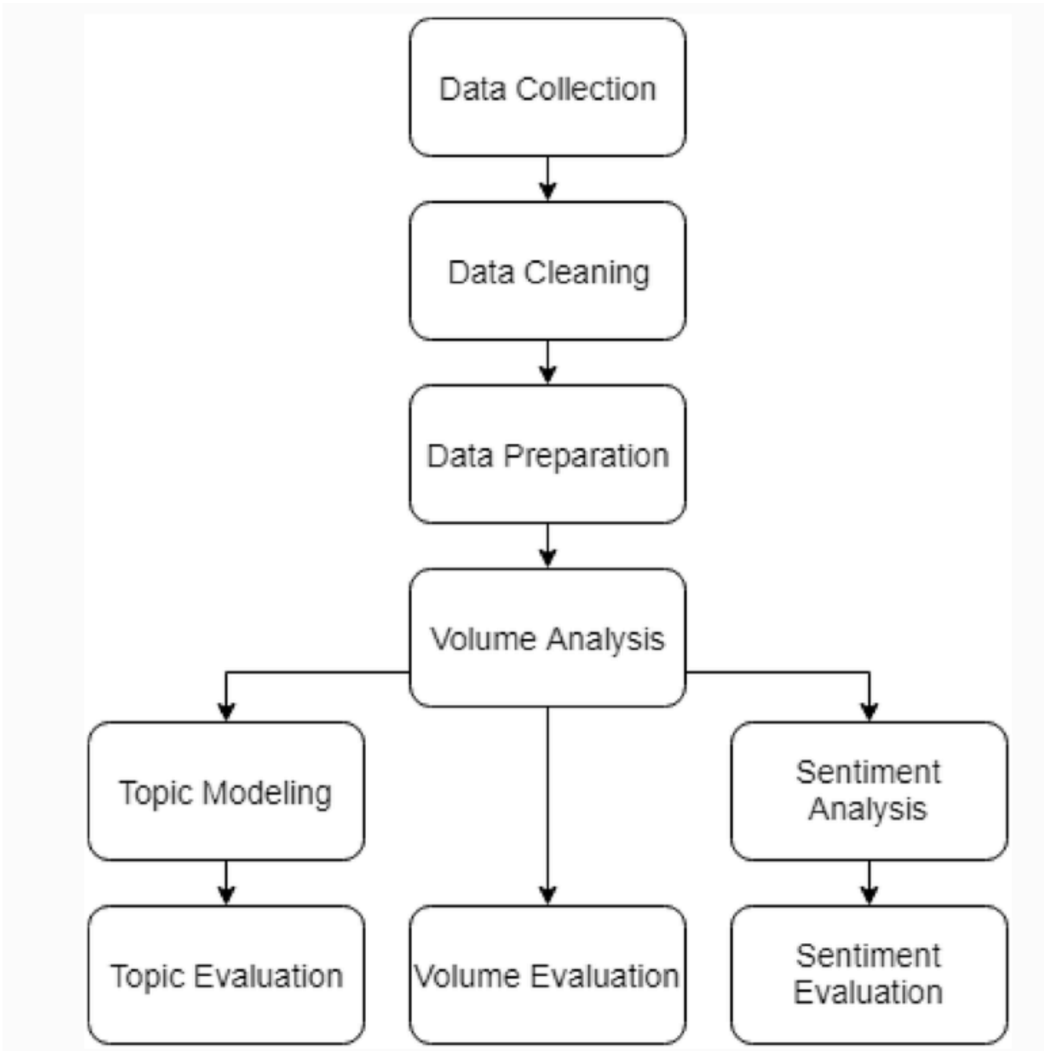


Figure 2.2. Example of Overall Workflow from Topic Modeling and Sentiment Analysis of Global Climate Change Tweets (Dahal et al., 2019).

2.6 Research Gap

It has been proved that social media could play the roles of early warning system, risk assessment, rescue scheduling and recovery in an event of natural disaster, separately. However, there is no research investigating all these roles in one disaster from the early preparing to the post-disaster recovery stages.

In this thesis, the following research gaps will be addressed:

1. The roles of social media in the whole time of a natural disaster event (i.e., Hurricane Ian), including the periods before the hurricane landed in Florida, during the hurricane struck Florida and 50-days after the hurricane dissipated.
2. The differences of the contents and emotions of people talked about on Twitter related to Hurricane Ian between the whole dataset of all tweets and the dataset of the top 1% of the tweets that had been mostly retweeted (i.e., the information that had been most widely shared among people).
3. The difference of the topics and reactions of people who used English and Spanish to communicate on social media about the event of Hurricane Ian.

English and Spanish tweets that related to Hurricane Ian from the day the hurricane formed to 50-days after the hurricane dissipated will be collected. The datasets will be separated into two parts, one includes all the tweets (the whole dataset), another includes the tweets that have been mostly retweeted (top 1% of the original tweets filtered by the number of their retweets). Topic modeling and sentiment analysis will be applied to these datasets. The differences of topics and emotions of people in each period of the event will be compared, including the difference between languages and two datasets.

3. Data

In this section, the study area will be introduced, including the places affected by Hurricane Ian and the languages used in those areas. Data collection, including the introduction of Twitter API, the process of data collecting, and the general description of data will be interpreted.

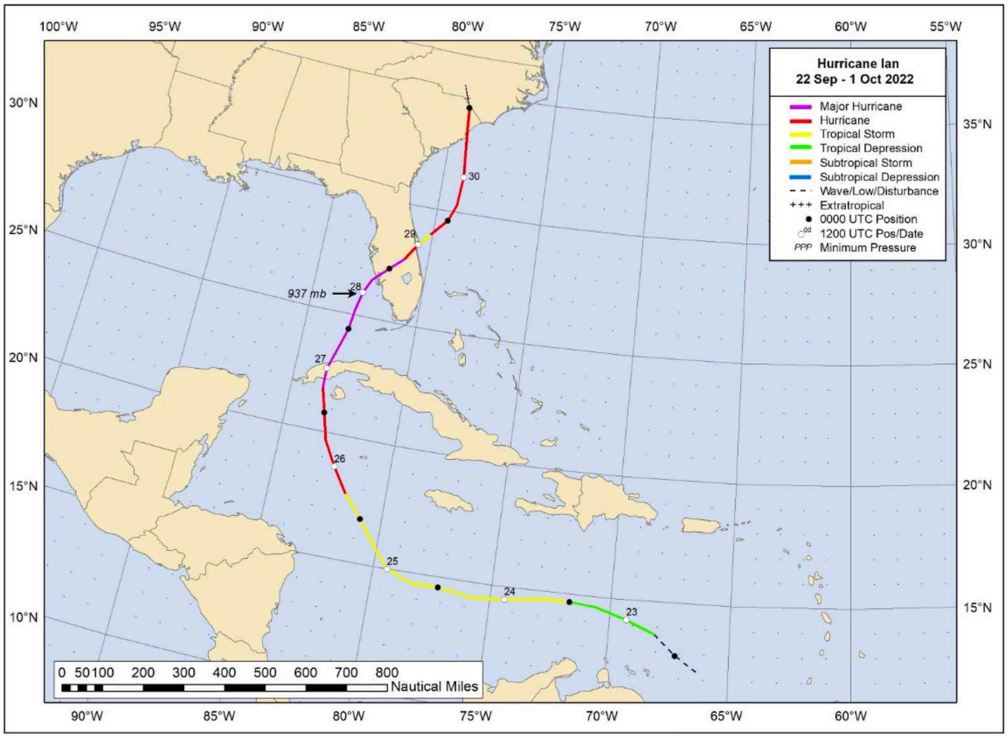
3.1 Study Area and Data

3.1.1 Study Area

Tracking back to the origin of Hurricane Ian, it was a tropical wave that formed from the west coast of Africa from September 14th to 15th. It slowly moved across the Atlantic and grew stronger when the wave was over the southeastern Caribbean. The hurricane was formed with sufficient wind speed and intensity near the east-northeast of Aruba around September 23rd. Ian kept intensify and moved past south of Jamaica, southwest of Grand Cayman Island, and headed northward to Cuba. On September 27th, Ian became a category 3 hurricane.

Ian slightly weakened when it passed over Cuba, but strengthened when it moved north towards the eastern United States. On September 28th, the eye of the hurricane made landfall on the barrier island of Cayo Costa, Florida, with a peak speed of 140 knots or 161 miles per hour (Category 5 hurricane). Ian moved across much of central and northern Florida, then directed to South Carolina on September 30th. It reached North Carolina and merged with a front, then dissipated after (Bucci et al., 2023). The best track of Hurricane Ian from September 23rd to 30th was provided in the report of the National Hurricane Center, as shown in *Figure 3.1*.

English is the official language in Florida, USA. According to the 2020 USA Census survey, 29.4% of the households in Florida used a non-English language as their primary language at home (Data USA, 2020). 74.1% of the non-English language spoken households used Spanish at home, which made Spanish the most common non-English language used in the households in Florida.



Best track positions for Hurricane Ian, 23–30 September 2022. Tracks during the extratropical stage are partially based on analyses from the NOAA Weather Prediction Center.

Figure 3.1. Screenshot of Figure 1. in the National Hurricane Center Tropical Cyclone Report of Hurricane Ian.

3.2 Twitter API

Twitter API is the official interface provided by Twitter Inc. that could give developers and researchers programmatic access to Twitter data, i.e., tweets, users, spaces, and more (Twitter Inc., 2022b). Twitter data could be retrieved and analyzed, which makes Twitter a great resource for scientific social media textual analysis. Twitter Inc. has been updating the version of Twitter API to scale the usage on their platform, and different versions are provided for various purposes of usage. It used to be free for academic researchers to retrieve Twitter data with a large amount once the users illustrated their research and how they would use the data. However, Twitter API will no longer be free for full access to academic researchers from 2023. The free version only provides access to tweet creation, upload, and login with Twitter. The basic version (Standard v1.1) was launched in 2012 and enables the user to post, interact, and retrieve data with a limited amount and access. The latest Pro version is Twitter API v2, which could support posting, pulling a large number of tweets, and access to full-archive search with a fee of $5000 per month. Premium and enterprise versions of Twitter API provided on the Twitter developer platform.

3.3 Data Collecting

In this thesis, the data was collected using Twitter API for academic researchers in 2022. Tweets with the hashtag "Hurricane Ian" of both English and Spanish versions (i.e., #HurricaneIan and #HuracánIan) were extracted. Those collected tweets with keywords include the original tweets created by users, as well as the tweets that have been retweeted, quoted, and replied to. The time periods of extracted tweets are from September 23rd, 2022 (the formation of Hurricane Ian) to November 22nd, 2022 (50 days after the hurricane dissipated).

The hurricane landed in Florida on September 28th, 2022. Therefore, the time periods have been divided by the steps of hurricane forming, landing, and dissipation. Tweets from September 23rd to September 27th, 2022, would be the "before hurricane" data. Tweets from September 28th to October 2nd, 2022, would be the "during hurricane" data. Tweets from October 3rd to November 22nd, 2022, would be the "after hurricane" data. Details are seen in *Table 3.1*.

Table 3.1. Description of Time Periods of Tweets Related to Hurricane Ian.

Time	23.09.2022 – 27.09.2022	28.09.2022 – 02.10.2022	03.10.2022 – 22.11.2022
Period	**Before** the hurricane	**During** the hurricane	**After** the hurricane

Various information about the tweets was extracted using query parameters including the referenced tweets, author ID, language, creation time, conversation ID, tweet texts, geographical information (geotag), id of the referenced tweet, number of retweets, number of replies, number of likes and the number of quotes. Details are seen in *Table 3.2*.

Table 3.2. Description of the parameters used in Twitter Data Extraction from Twitter Documentation.

Query parameter	Description
referenced_tweets	A list of Tweets this Tweet refers to (the type of this tweet, whether it is original, replied, quoted, or retweeted, and the id of the original referred Tweet)
author_id	The unique identifier of the user who posted this Tweet
id	The unique identifier of the requested Tweet
lang	Language detected by Twitter of the Tweet
created_at	Creation time of the Tweet
conversation_id	The Tweet ID of the original Tweet of the conversation
text	The actual UTF-8 text of the Tweet
geo	Location tagged (the unique id of place) in the Tweet
referenced_tweetid	The id of the original referred Tweet retrieved from the "id" of "referenced_tweets"
public_metrics	Public engagement metrics for the Tweet at the time of the request (number of retweets, replies, likes, and quotes)
retweet_count	Number retrieved from "retweet_count" of "public_metrics"
reply_count	Number retrieved from "reply_count" of "public_metrics"
like_count	Number retrieved from "like_count" of "public_metrics"
quote_count	Number retrieved from "quote_count" of "public_metrics"

In this thesis, both English and Spanish tweets related to Hurricane Ian were collected, as introduced in *Section 3.1.1*.

The tweets were extracted by the English and Spanish versions of "Hurricane Ian" as keywords (i.e., #HurricaneIan and #HuracánIan). However, it did not guarantee the language used in the tweet was English or Spanish. The parameter "lang" is the language used in the tweet detected by Twitter, in which "en" represents English, and "es" represents Spanish. The extracted tweets of different time periods were filtered based on the language detected by Twitter (i.e., filtered by "lang"). A total number of 919,698 tweets of hashtag Hurricane Ian (English version) before filtered by "lang" were collected, while the number of Spanish version hashtag was 559,679.

After being filtered by the language detected by Twitter, the number of tweets of Hurricane Ian (English version) was 866,058, and 551,409 tweets in the Spanish version. The number of different types of reference tweets, the original tweets, retweeted tweets, replied tweets, and quoted tweets, were calculated separately. Details are in *Table 3.3* and *Table 3.4*.

Table 3.3. Description of Twitter Data related to Hurricane Ian (English version).

	Before Hurricane	During Hurricane	After Hurricane
Total tweets	99,774	654,517	166,407
English tweets	94,415	614,957	156,686
Original English tweets	22,216	70,958	22,462
English retweeted tweets	66,535	522,595	128,033
English replied tweets	3,165	12,615	3,212
English quoted tweets	2,499	8,789	2,979

Table 3.4. Description of Twitter Data related to Hurricane Ian (Spanish version).

	Before Hurricane	During Hurricane	After Hurricane
Total tweets	60,439	284,832	214,408
Spanish tweets	59,918	278,350	213,141
Original Spanish tweets	13,515	37,834	20,210
Spanish retweeted tweets	45,363	236,861	190,308
Spanish replied tweets	611	2,062	1,824
Spanish quoted tweets	429	1,593	799

Twitter users could interact with each other through tweets, for example, retweeting (re-posting), replying to, liking, or quoting a tweet. Therefore, the number of retweets, replies, likes and quotes could imply public engagement of a tweet, which typically means this tweet has been shared with and seen by more users as the numbers get larger. To analyze the tweets that had been most shared, the top 1% of the most retweeted tweets were filtered and matched with their retweets. Details are seen in *Table 3.5*.

Table 3.5. Description of Twitter Data of the Top 1% retweeted tweets of English and Spanish versions related to Hurricane Ian.

	Before Hurricane	During Hurricane	After Hurricane
Top 1% English tweets	222	710	225
Matched English tweets	31,369	304,449	64,196
Top 1% Spanish tweets	135	378	202
Matched Spanish tweets	17,460	118,156	81,947

To examine the pattern of Twitter activities (i.e., how the number of tweets of different time periods changes), the relationship between the created time of tweets and the number of tweets was mapped. Maps containing all time periods for English and Spanish tweets were created, as well as for different periods (Before, During, and After the Hurricane).

4. Methodology

In this section, the methods used in this thesis will be interpreted, including the steps of detecting bot accounts, preprocessing of data, applying topic modeling, and sentiment analysis to the processed data. In each part, an introduction to used models or tools, and the precise steps of each process will be given. The workflow of this thesis is shown in *Figure 4.1*.

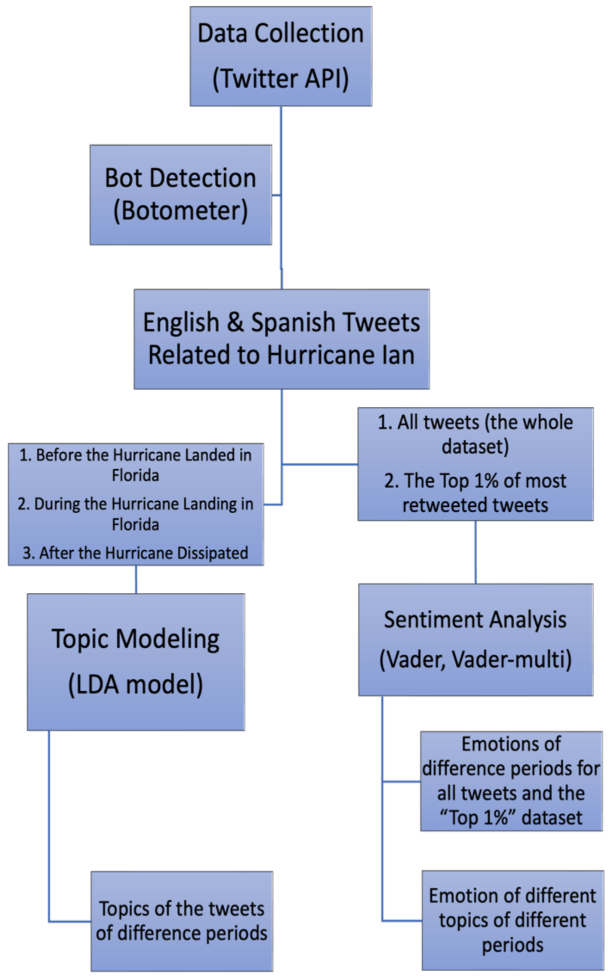


Figure 4.1. Workflow of this Thesis.

4.1 Bots' detection

4.1.1 Social Media Bot

No bot definition has been universally agreed upon because of its broad range of behaviors. Generally, a social media bot is a social media account that is automatically programmed to engage in social media, which is generating tweets in our case (i.e., Twitter Bot). These bots could mimic human activity, and some could be useful or harmless, but most of the bots are used in vicious and deceitful ways, such as Influence Bots and Spam Bots. Social media bots could be used for various harmful purposes. For example, they could manipulate the users with fake or biased news which might influence the financial market or politics. They could also be used to spread spam and amplify an account's popuarity or a movement by faking the number of followers or comments (Cloudflare Inc., 2023). Most social media bots act on triggers or a certain pattern, such as posting or retweeting all messages from a specific account (OSoMe Project, 2014). It is challenging to detect a bot since the software might be used to allow a single entity to control multiple accounts. When an account is accused suspicious, other accounts controlled by the entity will claim the authenticity. In this case, automatic and manual behaviors are mixed and difficult to distinguish, which makes it very tricky to prove an account is a social media bot.

As social media has become a vital data resource, concerns for the quality of the network content have been raised. Social media bots would affect the results of the analysis of data scientists and researchers as they highly depend on the quality of data. Low-quality data might lead to misleading trends or predictions of an event, resulting in an improper decision in the financial or political field, such as making a sensitive stock price fluctuate or directing a political movement to a dead end (Velayutham & Tiwari, 2017). To alleviate the influence of social media bots, multiple machine learning-based models are built for bots' detection. These models would identify the user's profile, compare a bot profile to a human profile, and analyze the pattern of tweeting, to calculate a score for each account to classify it as bot or human. In this thesis, the model "Botometer" from the Observatory on Social Media (OSoMe) of Indiana University was used for bot detection (Indiana University, 2023).

4.1.2 Botometer

Botometer, formerly named "BotOrNot" until 2017, is a project of OSoMe at Indiana University. OSoMe is a collaboration between the Network Science Institute, the Center for Complex Networks and Systems Research, and the Media School at Indiana University (OSoMe Project, 2014). Botometer is a trained machine learning algorithm to calculate a score for an account where higher scores indicate likely bot accounts and lower scores indicate likely human accounts. When an account has been checked, its public profile, public tweets, and mentions will be retrieved using Twitter API, and compare its activity to a large number of labeled examples. Botometer API would use the data to extract features of this account, to analyze and characterize the profile of the account, social network structure, friends' network, activity patterns, languages used, and sentiment. These features are used to calculate the score of the bot by various machine learning models. No personal or sensitive data would be retrieved or retained during the whole process (Yang et al., 2022).

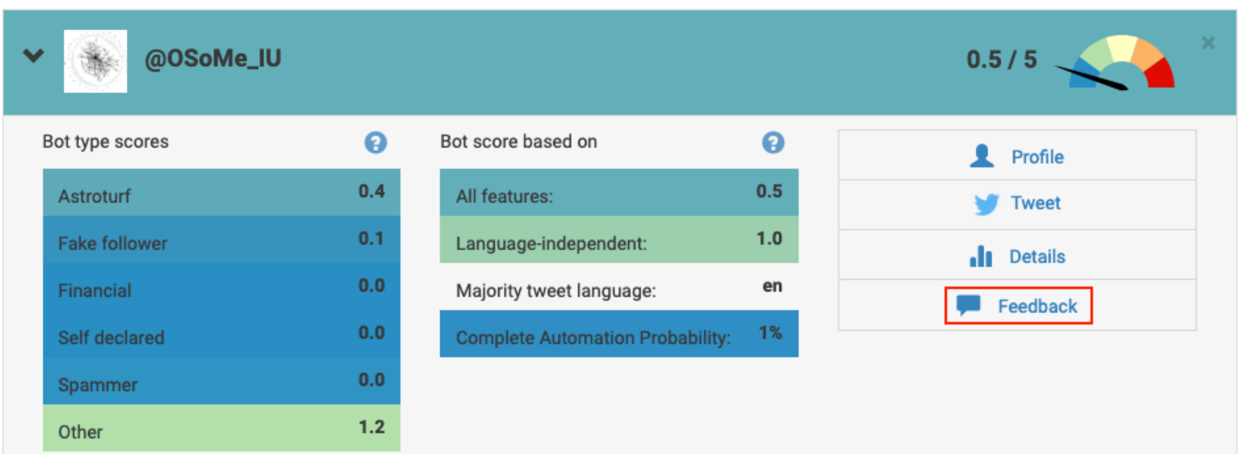


Figure 4.2. Example of Bot scores of Botometer from https://botometer.osome.iu.edu/faq.

Bot scores (the "overall score") are displayed on a 0-to-5 scale, shown in *Figure 4.2*. 0 means this account is the most human-like, and 5 means the most bot-like. However, there is no guarantee that the result is 100% correct. The model could identify an account wrongly because of the difference between human and machine cognition, and the limitation of algorithm. In this thesis, Botometer API was used to detect the accounts related to Hurricane Ian in both English and Spanish contents. There were 1,263 English accounts and 1,200 Spanish accounts for the original tweets, details seen in *Table 4.1*. Different threshold scores were set when classifying bots for both versions of tweets. Accounts with higher scores (such as those with a score of 4.8 or more) had been manually checked on the Twitter platform. Although an arbitrary threshold score could be problematic and misclassify accounts, it is necessary to set threshold scores for English and Spanish accounts as there were hundreds of accounts with scores larger than 4.0 (close to a bot-like end). The amount of English and Spanish accounts with higher scores was very different as shown in *Table 6*, therefore, the threshold scores should be set accordingly. The thresholds were set at 4.0 for English accounts and 4.6 for Spanish accounts, because the number of English and Spanish accounts detected as bots was around 10% of the total amount of accounts. After setting the threshold scores of both types of accounts, the accounts that had been detected as bots would be filtered out, thus, their tweets would be deleted from the data used in the next step, details seen in *Table 4.2*.

Table 4.1. Scores of Bot Detection of English and Spanish Accounts using Botometer API.

	English Accounts	Spanish Accounts
Total amount	1,263	1,200
Score > 4.0	141	682
Score > 4.5	42	283
Score > 4.6	24	135
Score > 4.7	23	95
Score > 4.8	9	22

Table 4.2. Description of English and Spanish Tweets after Bot Detection.

	English Tweets	Spanish Tweets
Total Tweets Before Bot Detection	115,636	71,559
Total Tweets	113,852	69,140
Before Hurricane	21,996	12,986
During Hurricane	70,160	36,431
After Hurricane	21,696	19,723

After June 30, 2023, Twitter updated the version of Twitter API, and the free endpoints were no longer available. Botometer website and API have stopped working since then because the endpoint that Botometer API relied on was deprecated. OSoMe has been working on building a new machine-learning model that works with Twitter's new paid API plans, however, with limited functionalities (OSoMe Project, 2023).

4.2 Topic Modelling

4.2.1 Google Colaboratory

Google Colaboratory (Colab) is "a hosted Jupyter Notebook service that requires no setup to use and provides free access to computing resources, including GPUs and TPUs" (Google, 2022). Colab could be used in AI, machine learning, data analytics, cloud computing, data visualization, education purposes, and more. Google Colab also provides multiple resources for users, including solutions on Stack Overflow, machine learning packages and kits, such as TensorFlow, and datasets across various disciplines. It can also be connected to Google Drive, so the user can upload their data files to the drive and import them into the Colab notebook.

4.2.2 Latent Dirichlet Allocation

Topic modeling is an unsupervised machine learning method to extract topics from a set of documents. It is a probabilistic model that considers each document is talking about different topics, and each of these topics contains a few different words (Dhuriya, 2021). Topic modeling is a very useful technique in Natural Language Processing (NLP). It can be used in data mining or text mining, document clustering, discovering the relationships among data and text documents, retrieving information from a large corpus, and more (CR, 2020; Jelodar et al., 2019). Topic modeling has been applied in many fields, such as business (customer services), political science, linguistic science, etc. The results of topic modeling depend on the quality of data processing, the choice of the algorithm, and the number of topics selected in the algorithm. There are multiple algorithms for topic modeling and Latent Dirichlet Allocation (LDA) is one of the most popular used by researchers (Jelodar et al., 2019).

Latent Dirichlet allocation (Blei et al., 2003) was first introduced in 2003 by Blei, Ng, and Jordan. It is a generative probabilistic, three-level hierarchical Bayesian model that models the hidden topics to a finite mixture, and each topic is characterized by the probability distribution over a set of words. Given a corpus *D* consisting of *M* documents, with document *d* having N_d words (where $d \in 1, \ldots, M$), LDA models *D* according to the following generative process (Blei et al., 2003; Jelodar et al., 2019), description of parameters and equation are from Jelodar et al. (2019) (p. 15173):

a) Choose a multinomial distribution φ_t for topic *t* (where $t \in \{1, \ldots, \mathrm{T}\}$) from a Dirichlet distribution with parameter β

b) Choose a multinomial distribution θ_d for document *d* (where $d \in \{1, \ldots, \mathrm{M}\}$) from a Dirichlet distribution with parameter α

c) For a word w_n (where $n \in \{1, \ldots, \mathrm{Nd}\}$) in document d,
   a. Select a topic z_n from θ_d
   b. Select a word w_n from φ_{zn}

In the generative process above, words in documents are only observed variables, φ and θ are latent variables, α and β are hyperparameters. The probability of observed data *D* is computed and obtained from a corpus as follows:

$$p(D|\alpha,\beta) = \prod_{d=1}^{M} \int p(\theta_d|\alpha) \left(\prod_{n=1}^{Nd} \sum_{zdn} p(z_{dn}|\theta_d) p(w_{dn}|z_{dn},\beta) \right) d\theta_d$$

where α is the parameter of topic Dirichlet prior; β is the distribution of words over topics drawn from the Dirichlet distribution; *T* is the number of topics; *M* is the number of documents, *N* is the size of the vocabulary; θ_d variables are document-level variables, sampled when per document; z_{dn}, w_{dn} variables are word-level variables that sampled when for each word in each text document. (p. 15173)

4.2.3 Python Libraries and Data Loading

Python 3.10.12 was used in this thesis. There are many libraries that provide optimized LDA models for researchers, such as lda 2.0.0 (lda developers, 2020), sklearn (Pedregosa et al., 2011), and Gensim (Řehůřek, 2022). In this thesis, the library Gensim was used, details are seen in *Table 4.3*.

Table 4.3. Major Python Libraries Used in the Thesis.

Package Name	Version
nltk (Natural Language Toolkit)	3.8.1
pyLDAvis	3.4.1
re (Regular expression operations)	3.11.4
NumPy	1.22.4
pandas	1.5.3
Genism	4.3.1
matplotlib	3.7.1
spaCy (Industrial-Strength Natural Language Processing)	3.5.4

The Twitter data files were uploaded to Google Drive and imported to the Colab Notebook. Only the original tweets that the users detected as human-like accounts were used in the topic modeling process.

4.2.4 Pre-processing

After loading the data file, the duplicated tweets and tweets that were null were excluded. Texts that contain emails (“@xxx.com”), URLs (“https://xxx”), new line characters, and single quotes in Tweets were deleted.

Stop words, such as “the”, “a”, “and in”, should be ignored to increase the quality of data in topic modeling. The package of nltk called “stopwords” was imported for English and Spanish tweets. Besides, the words “hurricaneian”, “hurricane” and “ian” were added to the stop words list of

English tweets, the words “huracánian”, “huracán” and “ian” were added to the stop words list of Spanish tweets. These words were added because all the tweets were collected based on the keyword “hurricane ian” for both English and Spanish versions. “Hurricane” and “ian” must appear in the sentences, making them less useful as words in the topic. After applying topic modeling, the word “huracar” had appeared in every topic of Spanish version tweets. This may be due to the different spelling of the word “hurricane” in Spanish, in this case, “huracar” should also be added to the stopwords list for Spanish tweets. “rt”, represents retweeted tweets, and “amp”, represented the ampersand (“&”) should also be removed as stopwords, because they are meaningless.

The next step is tokenization, which separates the sentences into smaller pieces of text called tokens. The function of tokenization in library Gensim called “simple_preprocess” was used in this thesis. It is a function that converts a document into a list of tokens, making the token lowercase, and setting the parameter “deacc” to true to remove the punctuations in the sentences.

Human languages are very complicated, one word could have different meanings, and multiple words can be combined to form a phrase that has distinct meaning from the origin words. The issue of textual ambiguity should be considered to make the machine better understand natural language and give us better results (Mattingly, 2022). Grammar and syntax are the essential components of human languages, and in the case of textual ambiguity, syntax plays a more important role. A single word, such as “pencil”, is a unigram, which represents a single concept. Typically, a unigram would not cause much textual ambiguity. Because the combination of words could result in very distinct meanings, it is important for a model to understand and process the document in the way that humans understand. The common cases of textual ambiguity in natural language processing are bigrams and trigrams. Bigrams are two words occurring together in a document frequently that have distinct meanings than used separately. A good example of a bigram would be “New York”, a city in the United States. Trigrams are three words that have different meanings when used together. To eliminate the textual ambiguity, models of bigram and trigram were built using the function “Phrases” in the library of Gensim.

After building the bigram and trigram models, functions for removing the stop words, making bigrams and trigrams, and lemmatization were created. Lemmatization is the process of grouping and converting the different forms of the same word to a simple word that is understandable to humans. Lemmatization is better but takes more time than stemming because the results of lemmatization are more human-readable while stemming only get the base word. In this thesis, only nouns, adjectives, verbs, and adverbs were saved as the result of lemmatization.

4.2.5 LDA model

After pre-processing the data, a clean, lemmatized document was created. It would be used to create a dictionary and corpus as input for topic modeling. The function “Ldamodel” in the Gensim library was used to build the LDA model for topic modeling. The number of topics was set to ten, as there are thousands of tweets in the data file and the number of topics should manageable. The number of topics also affects the quality of the model by influencing the coherence score and perplexity, which are introduced later. To get a lower score of perplexity and a higher score of coherence, the number of 10 topics appeared to be a good choice (examples of the perplexity and coherence scores of different numbers of topics are shown in *Table 4.4*). Details of the parameters of the LDA model (Řehůřek, 2022) are seen in *Table 4.5*, only the parameters that were manually set were shown, and other parameters were set to the default value.

Table 4.4. Example of the Perplexity and Coherence Score of Different Choices of the Number of Topics.

Number of Topics	Perplexity	Coherence Score
6	-8.02	0.44
8	-8.25	0.39
10	-8.63	0.45
12	-9.32	0.37

Table 4.5. Description of LDA model parameters that were manually set.

Parameter	Description	Input
corpus	Topic density	Corpus created from data
id2word	Topic-word density	Dictionary created from data
num_topic	Number of topics	10
alpha	A prior belief in document	Auto (learn as asymmetric prior from the corpus)
per_word_topics	A list of topics sorted in descending order of most likely topic for each word	True

To evaluate the LDA model, coherence score and perplexity were calculated. The coherence score was calculated using the function "CoherenceModel" in Gensim's "models" package. Coherence measures how similar the topic words are to each other, and how interpretable those topics are to humans. A higher score of coherence means that the topics are consistent and relevant. Perplexity was calculated using the function "log_perplexity" in the Gensim library. It measures how well the topic model performs when predicting new data and, thus, how accurate the predictions are. A lower score of perplexity means that the model is more accurate.

After getting the result of topic modeling, each tweet was assigned to the most relevant topic. The number of tweets of each topic in the data files was counted. The topic distribution of the top 1% of tweets was counted as well, to compare with the whole dataset. The topic model was visualized using the library "pyLDAvis".

The topic modeling results would change slightly every time when restarting the runtime. Thus, the topics and words contained would not be the same. The model was rerun five times for each dataset to find stable topics. The distribution of topics, coherence, and perplexity for each runtime was recorded. The topics and words in the model that have the highest coherence and perplexity were selected for topic modeling analysis. The time when the tweets that have been retweeted the most (top 1%) were created and their topics were mapped, to examine when did the topic appear. For tweets before and during the hurricane event, the topics were analyzed hourly-wise, since these two periods only have five days each. For tweets after the hurricane, the topics were analyzed day-wise, because this period had 50-day data.

4.3 Sentiment Analysis

Sentiment analysis is a popular technique in NLP, which is used to measure the emotion or attitude of people toward an event. In this master thesis, sentiment analysis was applied to the tweets of different periods of hurricanes.

4.3.1 Valence Aware Dictionary for sEntiment Reasoning (VADER)

Vader was a simple rule-based sentiment analysis model introduced by Hutto and Gilbert (2014). They combined the qualitative and quantitative methods in the development of the "gold-standard" sentiment lexicon, which was especially attuned to social media (i.e., microblog-like) contexts. Some frequently used sentiment expressions in social media platforms were added to the list of lexicons, such as Western-style emoticons (such as "☺", "☹", smiley face and sad face), sentiment-related acronyms and initialisms (such as "LOL"), and common slang associated with sentiment value (such as "nah"). Human raters selected with strict standards were used in the process of validating the sentiment valence (*intensity*). Over 90,000 ratings were made by human workers. Valence scores or ratings indicated both the *polarity* (positive or negative) and the *intensity* of the lexicon (on a scale of -4 to +4, extremely negative to extremely positive). The lexical features (around 7,500) with non-zero mean rating and standard deviation less than 2.5 were kept. These lexical attributes were integrated with 5 universally applicable principles, the *heuristics* (including punctuation, capitalization, intensifiers (or degree adverbs), contrastive conjunction signals ("but"), and trigram) that encapsulate the grammatical and syntactical norms utilized by humans when conveying or accentuating the *intensity* of sentiment. . They used the average sentiment rating from 20 pretrained human raters and the classification statistical metrics of precision, recall, and F1 score to evaluate the results. VADER lexicons were compared with seven other sentiment analysis lexicons that have already been established, including Linguistic Inquiry Word Count (LIWC), General Inquirer (GI), and more. VADER lexicons performed exceptionally well in the domain of social media. At matching ground truth, the VADER model performed as well as human raters, which means the averages of sentiment intensity of each tweet from the model and individual human raters were around the same. At the accuracy of classification, VADER outperformed human raters, which means the model performed better when classifying the sentiment of tweets.

4.3.2 Python Library and Sentiment Analysis

The library "vaderSentiment" (VADER Sentiment Analysis) was used for English tweets (C. Hutto & Gilbert, 2014). For vaderSentiment, the polarity values were used for the sentiment score, there is a sentiment dictionary in the result of Vader, which includes a "neg" score, "neu" score, "pos" score, and "compound" score. The scores of "neg", "neu" and "pos" represent the percentage of this sentence that is rated as negative, neutral, or positive. Compound score is calculated by the sum of valence ratings of every word in the lexicon, then normalized to the score between 1 (most positive) and -1 (most negative). As the standard rule, sentences with a compound score larger than or equal to 0.05 are "Positive", those with a compound score less than or equal to -0.05 are "Negative", and the rest are "Neutral" emotions. Tweets with a sentiment score of 0.0 were excluded in this thesis, as it was not clear that the sentence was completely objective, or not allocated with sentiment score at all. In this thesis, the sentences with negative or positive scores were more important for the analysis. Therefore, tweets with a score of 0.0 would not be discussed to avoid ambiguity.

The library "vader-multi" (VADER Sentiment Analysis Multilanguage) was used for Spanish tweets (C. J. Hutto, n.d.). In this multilanguage version, the VADER model integrates with Google Translate API through the library "translatte". The scoring system of vader-multi is the same as VADER, as the language of the text will be automatically detected.

Examples of the number of negative, neutral, and positive tweets of English and Spanish tweets during the hurricane are shown in *Figures 4.3* and *4.4* (Figures before the hurricane and after the hurricane for both languages are shown in the Appendix). In the figures, bars on the left side represented negative sentiments, and bars on the right side represented positive sentiments. The score was more consistent with the standard threshold introduced above; therefore, the threshold of sentiments was used in this thesis.

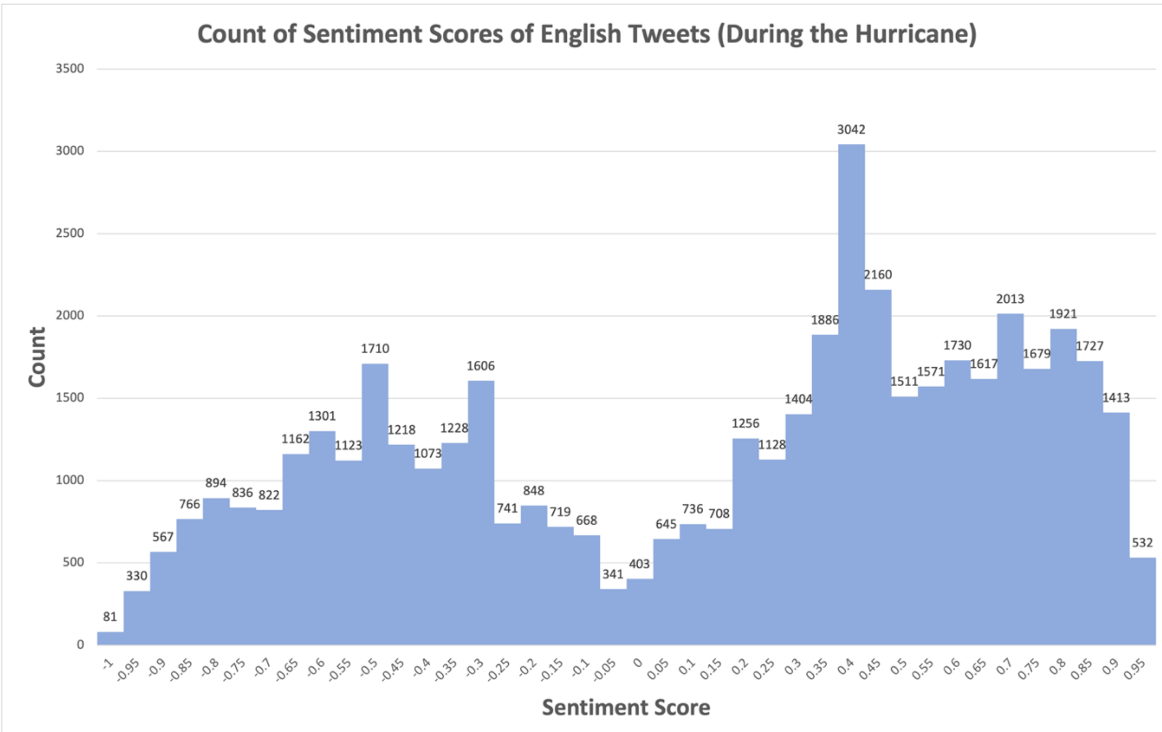


Figure 4.3. Count of Different Sentiment Scores of English Tweets During the Hurricane

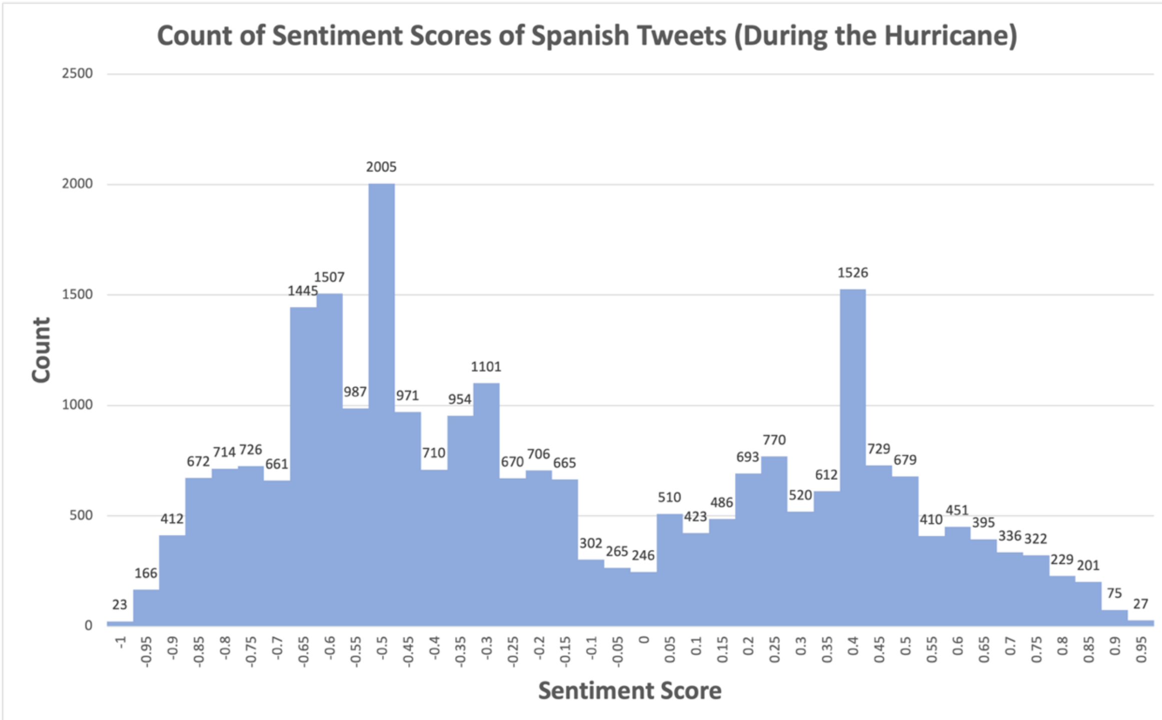


Figure 4.4. Count of Different Sentiment Scores of Spanish Tweets During the Hurricane.

The sentiment scores were calculated for all tweets, including the top 1% of the most retweeted tweets. The differences between the whole dataset of tweets and the top 1% of most retweeted tweets were compared. The number of tweets with different sentiment scores against time was mapped, to measure the changes in people's emotions during a natural disaster. The proportions of different sentiments for the period "before the hurricane", "during the hurricane" and "after the hurricane" were plotted as a bar chart for both languages, to compare the pattern of emotions of people who used different languages on social media. The sentiment score and topic were assigned for each tweet of all periods of both languages. Based on the sentiment, the changes in people's emotions towards different topics through time would be discussed.

5. Results

In this section, the results will be interpreted, including every step introduced in the methodology section. The results include data processing, detection of bot accounts, topic modeling, and sentiment analysis for the datasets of all tweets and the top 1% of most retweeted tweets for both languages. The trend of the number of tweets changed throughout the whole period will be mapped for both languages, to examine the Twitter activities of different stages of the hurricane. The topics and their proportions will be discussed for both languages, to examine the change of topics that people focused on in different periods of the hurricane event. The trend of the number of different sentiments ("Negative", "Positive" and "Neutral") will be mapped for the whole dataset and the top 1% of most retweeted tweets for both languages. This trend could show the growth or reduction of each emotion of people who use different languages on social media in each period of the hurricane. The proportion of different sentiments will be visualized as well, to compare the overall changes of sentiments (i.e., negative or positive) throughout the time. The proportions of negative sentiment for each topic will be mapped, to show the changes of emotion towards different topics. The proportions of topics and sentiments will be compared between all tweets and the top 1% of most retweeted tweets, to examine the difference between the whole dataset and the information that had been most widely shared among people.

5.1 Data Processing

Twitter data, Tweets, with the keyword "Hurricane Ian" were collected and separated by the time they were created. Both English and Spanish tweets related to Hurricane Ian have been collected using Twitter API, details are seen in *Figure 5.1*. Time periods of the data were divided based on the formation, landing, and dissipation of the hurricane, as discussed in *Section 3.3*, which are before the hurricane landed in Florida (September 23rd to September 27th, 2022), during the hurricane landing in Florida (September 28th to October 2nd, 2022), and after the hurricane dissipated (October 3rd to November 22nd, 2022). Both "before" and "during hurricane" tweets have data of five days, while "after hurricane" tweets have data of 50 days.

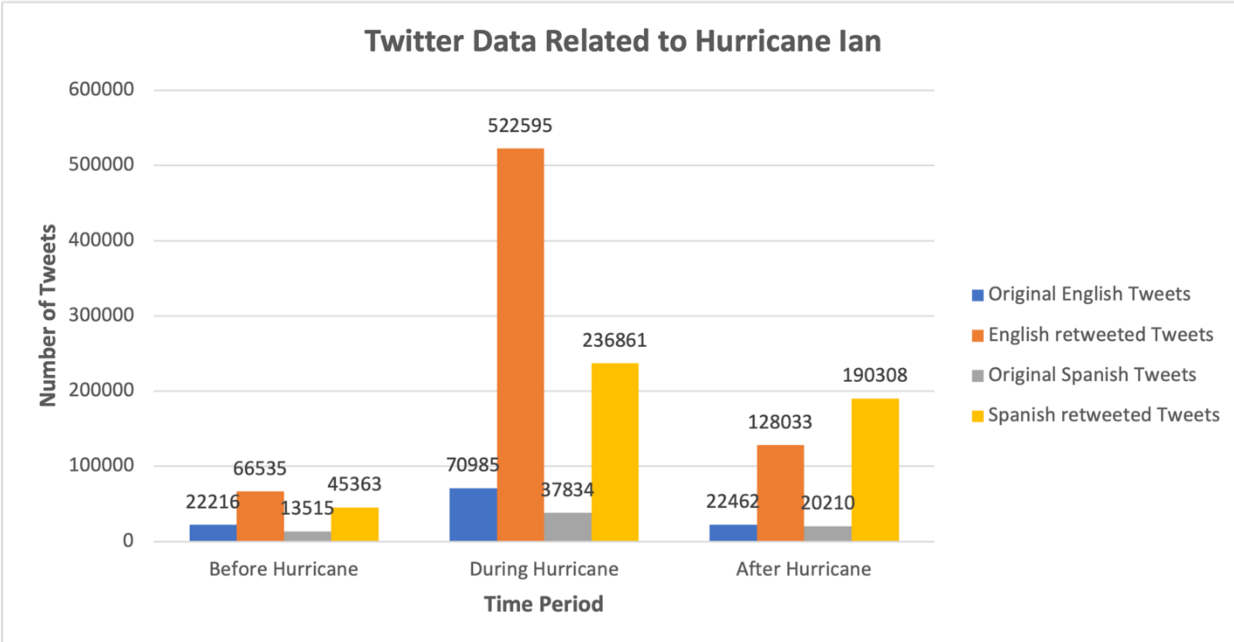


Figure 5.1. Description of Twitter Data Related to Hurricane Ian for English and Spanish Tweets.

The number of original tweets that people created by themselves could imply how much information had been created, and the number of retweeted tweets could imply how wide this information had been spread. For comparing original tweets across different periods, it was observed that there were approximately equal numbers of English tweets before and after the hurricane, both around 22,000. However, during the hurricane period, the tweet count was three times higher, at approximately 70,000. Regarding original Spanish tweets, during the hurricane period, they were roughly 2.8 times more numerous than those before the hurricane. In the aftermath of the hurricane, the count of tweets was 1.5 times that of the period before the hurricane. As for tweets that had been retweeted, there was a 7.85-fold increase in the number of English tweets during the hurricane period compared to before the hurricane. After the hurricane, the number of retweeted tweets was double the size of those before the hurricane. Regarding Spanish tweets that had been retweeted, the number of them during the hurricane period was more than five times those before the hurricane period. After the hurricane, retweeted Spanish tweets were more than 4.2 times those before the hurricane. While there was a notable decrease in retweeted English tweets after the hurricane compared to during the hurricane, this abrupt decrease wasn't observed in the case of Spanish retweeted tweets after the hurricane.

In comparing across languages, for both “before” and “during hurricane” periods, the quantity of original English tweets was twice that of Spanish tweets. However, in the period after the hurricane, the amounts of original English and Spanish tweets were roughly equivalent. When considering retweeted content, English tweets surpassed the count of Spanish tweets in both the "before" and "during hurricane" periods. However, in the "after hurricane" phase, retweeted Spanish tweets outnumbered their English counterparts. During the hurricane, both the original and retweeted English tweets were double the size of Spanish tweets. Based on the amount of data collected, English tweets have been created and spread wider than Spanish tweets before and during the hurricane, while Spanish tweets have been spread more than English tweets after the hurricane.

To analyze the tweets that had been most shared among users, the top 1% of the tweets that have been mostly retweeted were filtered out, as well as the count of their corresponding matched retweets for both languages. As illustrated in *Figure 5.2*, the number of matched English retweets for both the “before” and “during hurricane” periods were double the size of Spanish retweets, while the count of Spanish retweets was higher when it came to the period of “after hurricane”. To compare the dissemination of tweets across different languages, the percentages of the number of top 1% matched retweets relative to all tweets were calculated, as shown in *Figure 5.3*. The range of these percentages spanned from a minimum of 38% to a maximum of 58%. Notably, the English top 1% of tweets consistently held a larger proportion within the realm of retweeted content, indicating a more concentrated distribution pattern for English-language information.

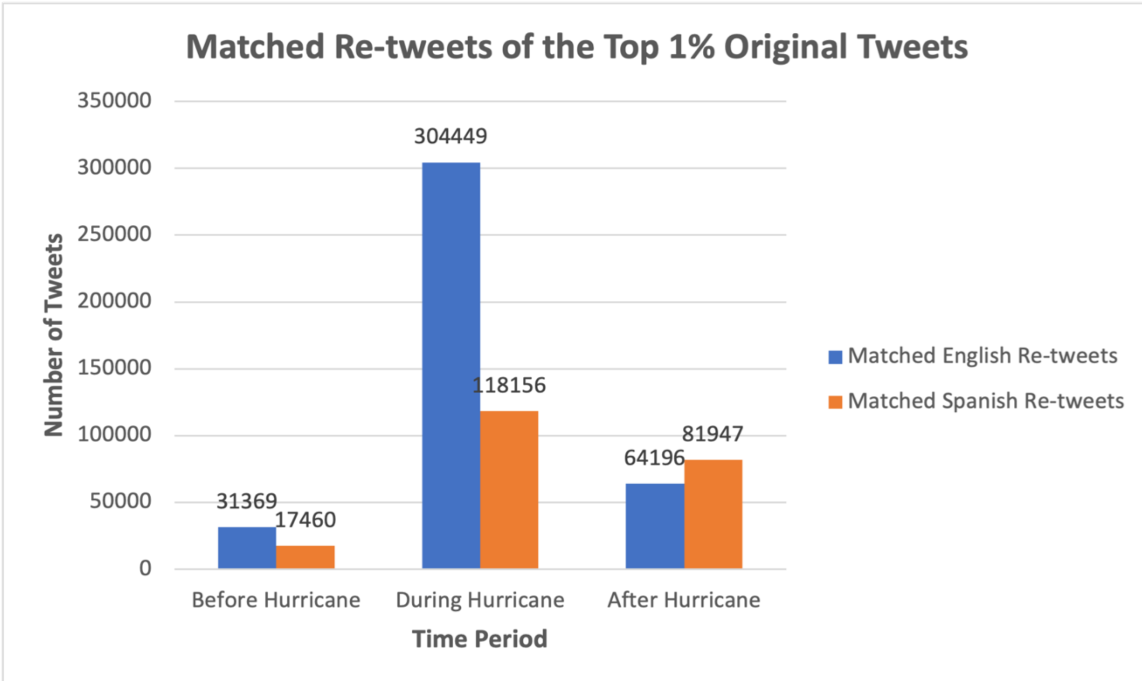


Figure 5.2. Matched Retweeted Tweets of the Top 1% Original Tweets of English and Spanish Tweets.

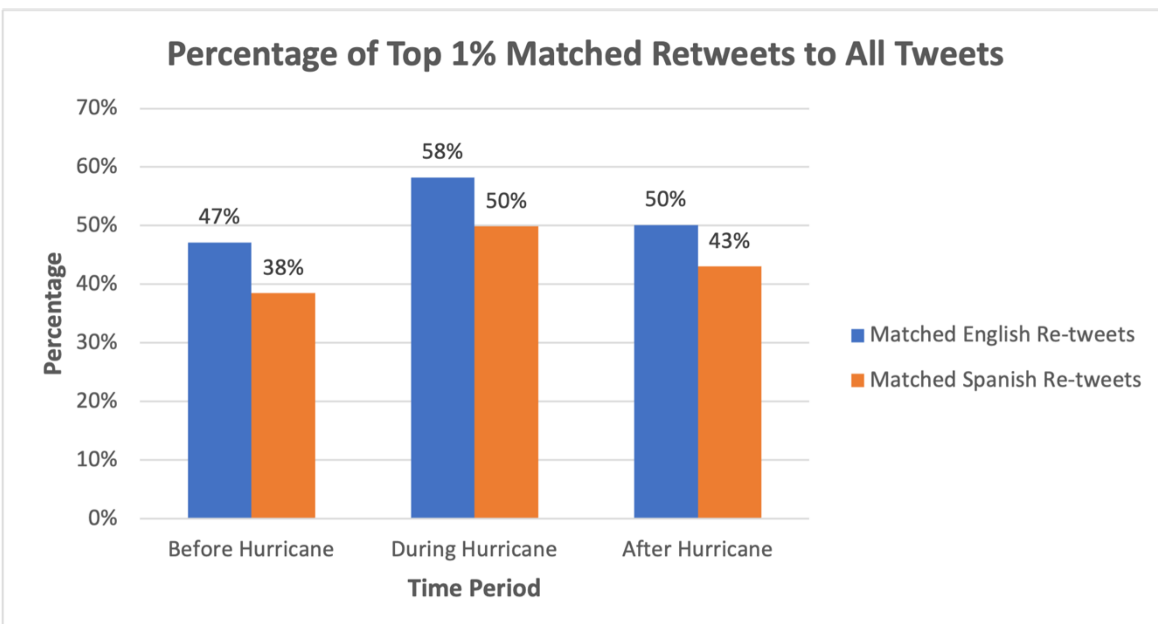


Figure 5.3. Percentage of the Top 1% Matched Retweets to All Tweets for English and Spanish Tweets.

The patterns of Twitter activities throughout the strike of Hurricane Ian were examined. The number of origin tweets was mapped with their created times for both English and Spanish tweets that related to Ian. In *Figure 5.4* and *5.5*, Twitter activities from the forming day of the hurricane to 50 days after it dissipated were shown. English and Spanish tweets have very similar patterns. They both increased slowly for the first two days and sharply increased on the 3rd and 4th days. English daily tweets grew to 1,115 on the second day, then rushed to 6,997 and 13,639 in the following two days. Spanish daily tweets shared the same pattern, but the number of tweets was around half of English tweets with 737 tweets on the second day, and 7,678 on the 4th day. The number of tweets peaked on September 28th, 2022, the day when Ian landed in Florida, with 30,128 English daily tweets and 13,572 Spanish tweets (both doubled the number of last day’s tweets). Activities on Twitter dropped vastly during the hurricane's landing in the US, while gradually dropping to their minimum through 50 days after the dissipation. English daily tweets dropped faster than Spanish tweets during the landing of the hurricane. English tweets almost decreased by 10,000 tweets each day for the first two days and halved in the next two days. Spanish tweets decreased at a pace of 3,000 tweets each day and only 1,000 on the 4th day. On October 2nd, the day before Ian dissipated, the number of tweets for English and Spanish was almost the same. After the hurricane dissipated, the number of tweets decreased sharply for the first five or six days, then slightly fluctuated for the rest of time. In the 50 days, the number of Spanish daily tweets was more than the number of English tweets for several days, which never happened in the periods before and during the hurricane.

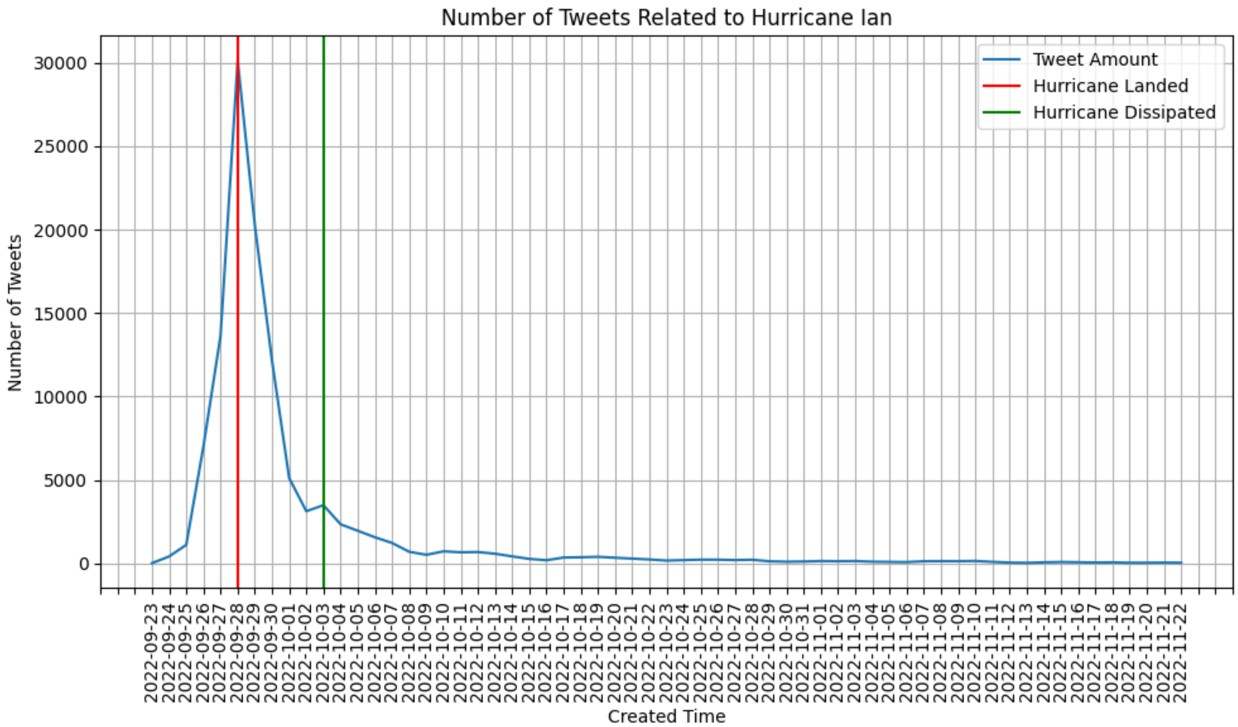


Figure 5.4. Twitter Activities Related to Hurricane Ian (English) Throughout the Whole Time Period.

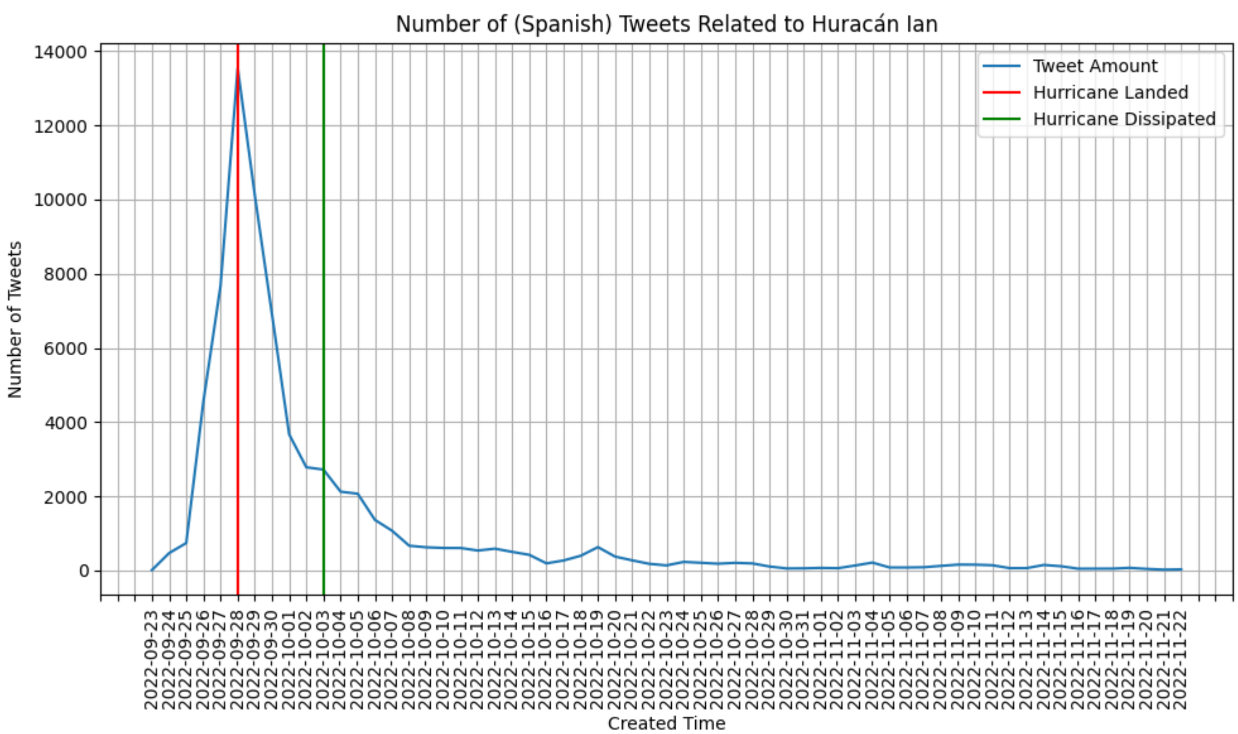


Figure 5.5. Twitter Activities Related to Hurricane Ian (Spanish) Throughout the Whole Time Period.

5.2 Bots Detection

Botometer was used to detect the bot accounts. To make the results convincible, bot-like accounts were deleted with different thresholds for English and Spanish. The thresholds were set manually based on how many accounts had been detected as bots of that score, as shown in *Section 4.1.2*. After testing different scores, the thresholds were set at 4.0 for English accounts, and 4.6 for Spanish accounts, when bots detected for both English and Spanish accounts was around 10% of their total amount of accounts, details are shown in *Figure 5.6*. Setting an arbitrary threshold for detecting bots is not ideal, but efficient for this thesis. After detecting the bots, those tweets created by the bot accounts were deleted. Bot is not an important factor in this thesis, as the number of tweets deleted was around 2,000, only 1.7% of all English tweets, and 2.7% of all Spanish tweets.

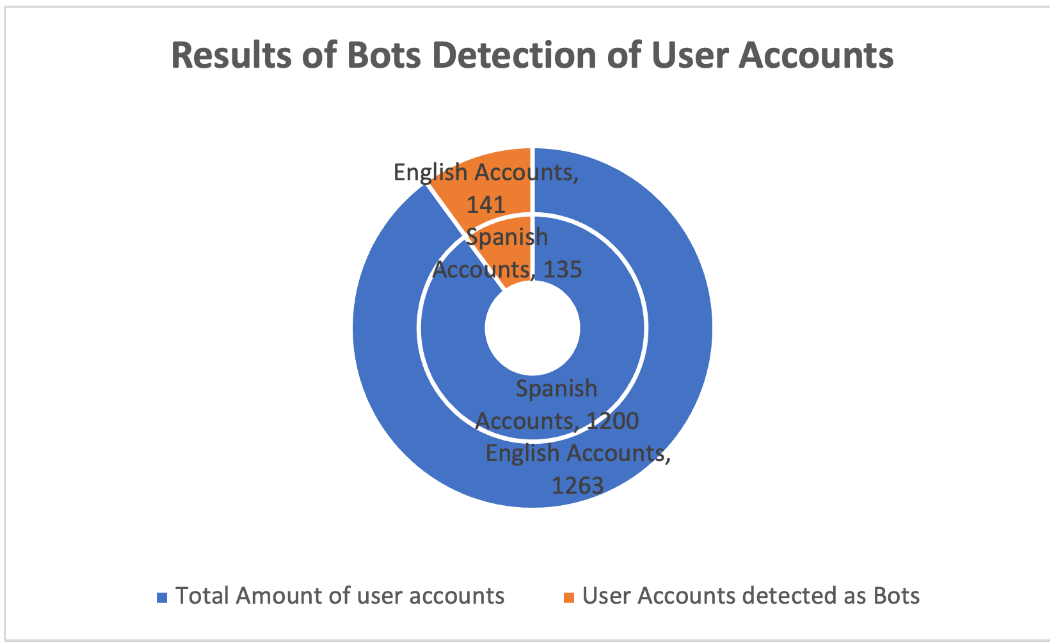


Figure 5.6. Results of Bots Detection of English and Spanish Accounts with Set Thresholds using Botometer API.

Although the scores have been calculated by the model, it cannot be guaranteed that the accounts with scores higher than the thresholds were bots. Some of the accounts detected as bots were manually checked by browsing the users' profiles and tweets. Users who create too many tweets in a short time, write about very similar content in several tweets, or send too many advertisements would be detected as bots. Even some official accounts, such as news or digital marketing could be detected as bots, examples shown in *Figure 5.7* and *Figure 5.8*.

Figure 5.7. Screenshot taken from Twitter of Account Detected as Bot which is a News Company Account.

Figure 5.8. Screenshot taken from Twitter of Account Detected as Bot which is a Newspaper Official Account.

5.3 Topic Modelling

The tweets were processed separately based on the language and time period of the tweet being created. After loading the file, duplicated and null data of the original tweets for both languages were excluded. The number of deleted tweets varies from 20 to thousands, as shown in *Table 5.1*. The duplicates and null values may not cause a serious problem in this thesis, because of the large size of data. However, the quality of the data, as well as the result, would be better if the duplicates and null values were deleted.

Table 5.1. Description of Original Data and Cleaned Data that Removed Duplicates and Null Values.

	Original File	Cleaned File
English Before Hurricane	21,996	21,973
Spanish Before Hurricane	12,986	12,913
English During Hurricane	70,160	70,002
Spanish During Hurricane	36,431	35,679
English After Hurricane	21,696	20,405
Spanish After Hurricane	19,723	19,032

After dropping the duplicates and null values, stopwords, email addresses, and URLs were removed as well. The cleaned data was tokenized and lemmatized, only nouns, adjectives, verbs, and adverbs were saved in the data. A dictionary and corpus for the topic modeling step were created based on the lemmatized data. The LDA model was built with ten topics for each document, and each topic includes 10 most probable words, example shown in *Figure 5.9*.

These words are divided into six categories according to the previous studies (David et al., 2016; Qu et al., 2011; Wu & Cui, 2018) including information related to the disaster, weather, emotion, action, situation, and time or location as shown in *Table 5.4, 5.7* and *5.10*. Disaster-related words are the words that directly describe a disaster, which are important in the sharing of information during the event and attract the attention of people in the early phase of a disaster. Weather-related words are words related to the change of climate due to natural disaster, which varies between the type of disaster. Emotion-related words are words expressing people's feelings or concerns, which are normally important during and after the event. Action-related words are the words that people use to attract the movement of government or government used to guide people. Situation-related words are words that describe the influence or update of the situation of disaster or people, which are important in the early warning and assistance during disasters. Spatial- or Temporal-related words are the words that directly state a certain time or location.

The topics and words changed slightly whenever restart the runtime, to find more stable topics, the model was run for 5 times for each dataset. For English topics, the results were more stable with similar and relatively better coherence and perplexity. For Spanish topics, the results were more random with relatively worse coherence and perplexity. Results for both languages were selected based on the coherence and perplexity of the model, as discussed in *Section 4.2.5*.

```
[(0,
  '0.031*"thought" + 0.019*"seize" + 0.015*"look" + 0.015*"website" + '
  '0.015*"provide" + 0.014*"response" + 0.013*"radar" + 0.013*"video" + '
  '0.013*"give" + 0.013*"take"'),
 (1,
  '0.029*"help" + 0.021*"shelter" + 0.018*"need" + 0.016*"emergency" + '
  '0.014*"prepare" + 0.011*"thank" + 0.011*"service" + 0.011*"support" + '
  '0.011*"resident" + 0.010*"community"'),
 (2,
  '0.024*"close" + 0.022*"due" + 0.018*"get" + 0.014*"state" + 0.014*"stay" + '
  '0.013*"safe" + 0.012*"cancel" + 0.012*"office" + 0.010*"go" + 0.009*"say"'),
 (3,
  '0.046*"storm" + 0.026*"watch" + 0.015*"evacuate" + 0.014*"zone" + '
  '0.014*"area" + 0.012*"prepare" + 0.012*"go" + 0.012*"surge" + '
  '0.011*"evacuation" + 0.011*"live"'),
 (4,
  '0.037*"storm" + 0.020*"make" + 0.019*"category" + 0.019*"tampa" + '
  '0.018*"landfall" + 0.014*"late" + 0.014*"update" + 0.014*"major" + '
  '0.014*"track" + 0.014*"expect"'),
 (5,
  '0.017*"power" + 0.016*"get" + 0.014*"wait" + 0.014*"ready" + 0.014*"news" + '
  '0.012*"come" + 0.011*"go" + 0.009*"water" + 0.008*"day" + 0.008*"date"'),
 (6,
  '0.088*"safe" + 0.070*"stay" + 0.025*"friend" + 0.023*"get" + 0.021*"path" + '
  '0.019*"pray" + 0.017*"prayer" + 0.017*"hope" + 0.015*"family" + '
  '0.013*"good"'),
 (7,
  '0.049*"live" + 0.023*"update" + 0.020*"get" + 0.015*"contact" + '
  '0.014*"coverage" + 0.014*"storm" + 0.014*"close" + 0.012*"stream" + '
  '0.012*"watch" + 0.011*"tomorrow"'),
 (8,
  '0.048*"wind" + 0.030*"rain" + 0.023*"mph" + 0.021*"update" + 0.021*"storm" '
  '+ 0.011*"impact" + 0.010*"expect" + 0.010*"continue" + 0.009*"start" + '
  '0.009*"still"'),
 (9,
  '0.037*"go" + 0.019*"know" + 0.018*"floridian" + 0.016*"m" + 0.016*"get" + '
  '0.012*"see" + 0.012*"right" + 0.011*"think" + 0.010*"need" + 0.009*"come"')]
```

Figure 5.9. Example of the Result of Topic Modeling for English Before Hurricane Tweets.

5.3.1 Before the Hurricane

The results of English and Spanish "Before Hurricane" tweets topic modeling are listed below (in *Table 5.2* and *5.3*); the order of topics and words was generated by the model automatically:

Table 5.2. Topics and Words of English Tweets (Before the Hurricane).

Topic	Words
0	safe, stay, path, pray, hope, prayer, family, thought, home, people.
1	storm, wind, mph, get, cat, update, move, category, track, continue.
2	impact, storm, rain, prepare, expect, wind, brace, day, area, flooding.
3	storm, hit, get, watch, surge, water, people, go, support, head.
4	need, help, contact, play, service, response, hour, set, governor, arrival.
5	get, go, safe, stay, friend, know, time, take, warning, floridian.
6	make, landfall, update, new, category, provide, affect, approach, track, late.
7	safe, stay, shelter, open, stream, need, wait, help, remain, leave.
8	live, emergency, update, watch, state, weather, want, find, visit, information.
9	close, due, tomorrow, see, school, update, get, cancel, office, today.

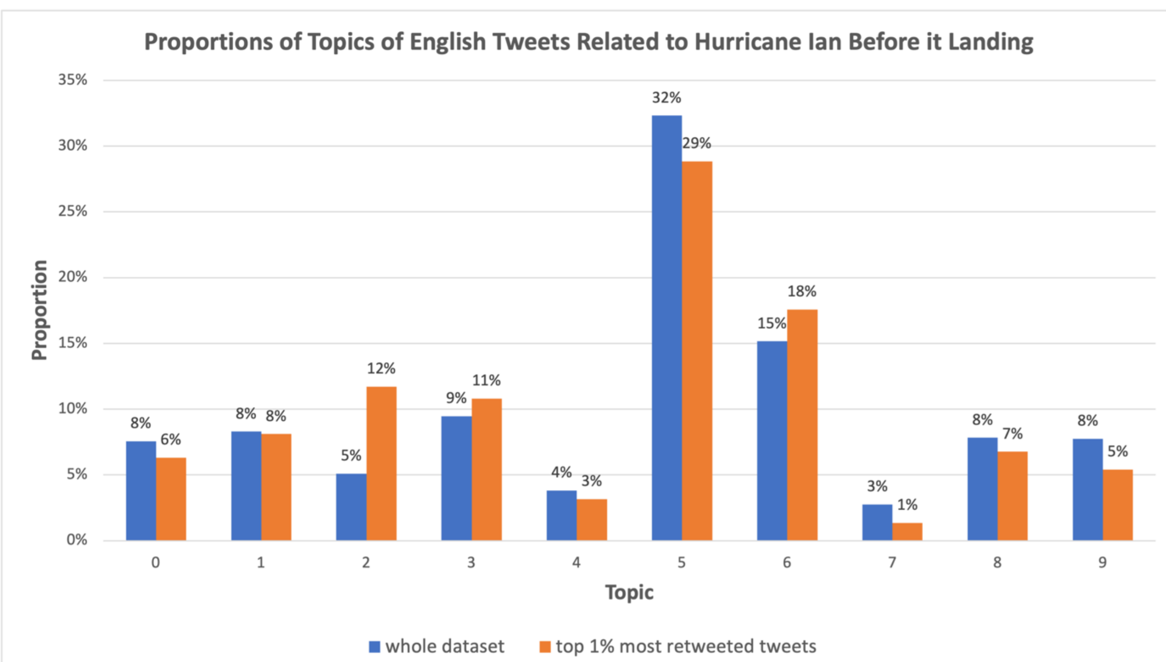


Figure 5.10. Proportion of Topics Related to English Tweets Before the Hurricane Landed.

Table 5.3. Topics and Words of Spanish Tweets (Before the Hurricane).

Topic	Words
0	feurza, impacto, alerta, azotar, mayor, viento, paso, mas, oeste, centro.
1	florido, pueblo, noticia, llegado, categorio, hora, acerca, dirigir, paso, llegar.
2	tierra, tocar, viento, categorio, categoria, fuerte, martes, madrugada, hora, lluvia.
3	ver, prepara, mas, internacional, inundaciones_apagón, septiembre, tomar, recibir, medida, miercol.
4	categoria, cuba, mas, convertir, occidental, azota, paso, alcanzar, costa, acercar él.
5	paso, dejar, categorio, provocar, lluvia, afectación, dano, destrozo, destruccion, apoyar.
6	pinar, rio, provincial, fuerzapinar, paso, mas, afectado, Cubano, occidental, huracaniar.
7	lluvia, tormenta, alertar, viento, fuerte, tropical, llegar, categorio, mas, campeche.
8	tormenta, vivo, mexico, tropical, trayectoria, golfo, avanzar, categorio, florido, ultimo.
9	paso, solidaridad, dano, hermano, lluvia, posible, inundación, estragon, territorio, afectar.

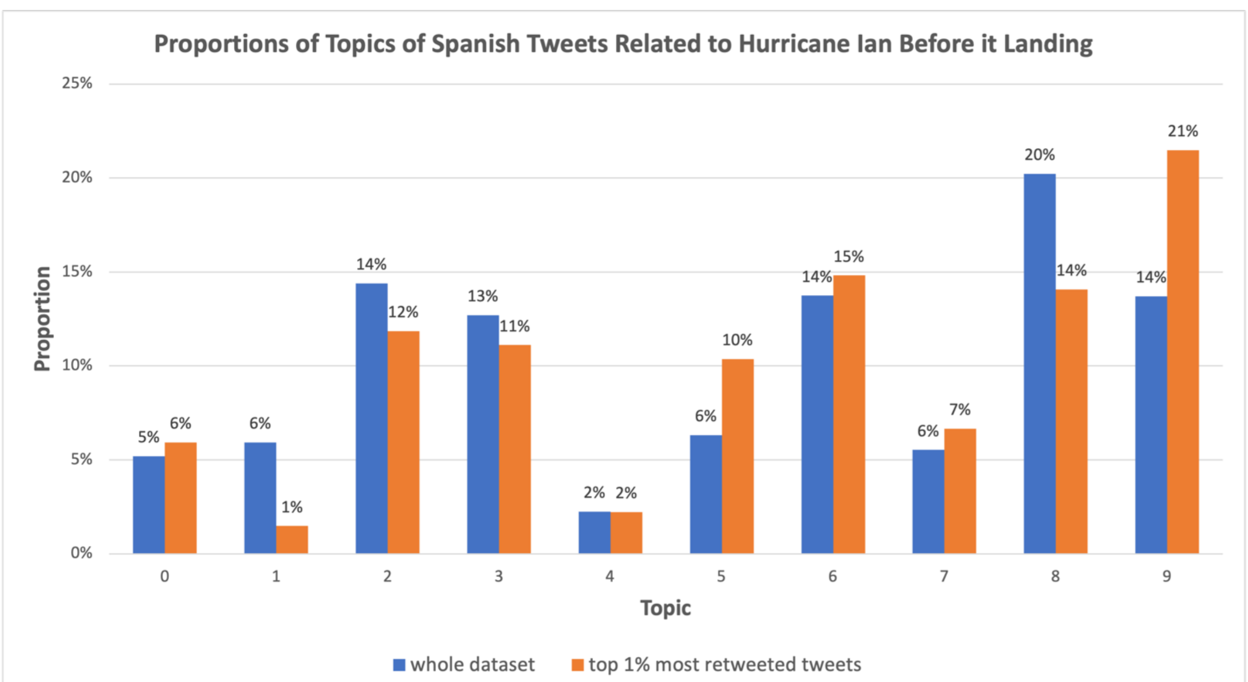


Figure 5.11. The proportion of Topics Related to Spanish Tweets Before the Hurricane Landed.

Table 5.4. Topics related to Early Warning of Twitter Data.

Topic	Related Words (English)	Related Words (Spanish)
Disaster-related	"storm", "landfall", "path", "surge", "category", "stream", "flooding", "approach", "arrival", "hit"	"fuerza", "paso", "categorio(a)", "fuerte", "acercar", "provocar", "tormenta", "trayectoria
Weather-related	"wind", "rain", "head", "close", "weather", "mph", "see"	"viento", "lluvia", "rio", "inundaciones_apagon", "alcanzar", "ver", "llegado", "avanzar"
Emotion-related	"prayer", "pray", "hope", "friend", "family", "people"	"solidaridad", "hermano", "mayor"
Action-related	"track", "move", "live", "emergency", "warning", "affect", "prepare", "contact", "ready", "provide", "find", "coverage", "shelter", "news", "cancel", "governor", "office", "state", "service", "response"	"alerta", "dirigir", "tocar", "prepara", "tomar", "dejar", "apoyar", "noticia"
Situation-related	"safe", "stay", "need", "update", "move", "start", "continue", "get", "help", "resource", "information"	"vivo", "dano", "destrozo", "destruccion",
Spatial/Temporal-related	"tomorrow", "school", "today", "day", "hour"	"oeste", "centro", "florido", "pueblo", "hora", "martes", "madrugada", "septiembre", "miercol", "coasta", "occidental", "cuba", "cuban", "tropical", "campeche", "golfo", "mexico"

In English tweets, there were fewer identical words generated. Words such as “storm”, “landfall”, “path” and “category” related to disaster appeared frequently in several topics. As well as words like “wind”, “rain”, “see” and “close” that relate to weather. These two topics indicated that people were aware of the change of climate or the weather, and the hurricane was approaching their territory. Emotion-related words were relatively limited. People mostly talked about praying and taking care of their family and friends. As for words related to action or response, “warning”, “emergency”, “prepare”, “store” and “ready” appeared in the early stage to notice people about the coming storms and be prepared for them. These words are usually used as an early warning, which could make people stay awake and get ready for possible evacuation and further activities. Words such as “track”, “coverage”, “provide”, “shelter” and “news” related to the action of the government or emergency response agencies that provide rescue or resources to people who suffer from the disaster. These words also play the role of educating people on how to be prepared for the upcoming event and where they could get information or help. Situation-related words often indicated the current situation of people or the community. “stay” and “safe” were two words that most frequently appeared in the topics, which expressed the desire of people to be safe. Words like “start”, “continue” and “move” indicated that the hurricane was approaching. Words like “need”, “update”, “get”, “help”, and “resource” related to what people needed in the current situation. In the period “before the hurricane”, words related to time or location were not informative. The 3 topics with the highest proportions for all tweets were Topics 5, 6, and 3. The 3 topics with the highest proportions for the top 1% of retweeted tweets were Topic 5, 6, and 2. In these topics, words related to the disaster and weather appeared most frequently, such as “landfall”, “storm” and “approach”. Words related to the situation of people like “stay” and “safe” also appeared multiple times. Generally in this phase, people tweeted mostly about the approaching storm, and the weather changes and informing others to stay safe.

In Spanish tweets, identical words, or words with the same meaning (such as “categorio” and “categoria”) were generated more than the words in the English topics. Disaster-related words, such as “paso” (passed), “fuerza” (force), “fuerte” (strong), “provocar” (provoke), “tormenta” (storm), and “trayectoria” (trajectory) also suggested the approaching of the hurricane, however, more intense. Weather-related words “viento” (wind), “lluvia” (rain), “apagon” (blackout), “ver” (see), “llegado” (arrived), and “alcanzar” (reach) were like words of the English tweets. In Spanish tweets, the disaster- and weather-related words seemed to be intenser than in English tweets. One possible reason would be the hurricane was formed in the South, where most people spoke Spanish. They experienced the storm earlier than English-speaking people, thus, the words in Spanish were more serious. Words like “hermano” (brother) and “mayor” (elderly) related to people’s emotions were similar to the English words “family” and “friends”. In Spanish tweets, there was no word related to prayer, but “solidaridad” (solidarity) appeared. Action-related words like “alerta” (alert), “prepara” (prepare), “dejar” (leave), “apoyar” (support), and “noricia” (news) were consistent with the English tweets. Similar to the disaster- and weather-related words, Spanish situation-related words were stronger and ahead than English ones, such as “vivo” (alive), “dano” (damage), “destrozo” (smashed), and “destruccion” (destruction). The spatial or temporal-related words in Spanish were more informative. Words that described direction or location, such as “oeste” (west), “centro” (center), “florido” (Florida), “pueblo” (town), “coasta” (coast), “occidental” (western), “cuba”, “tropical”, “Mexico”, “Campeche”, and “gulf” could help estimate the current affected area. There were also words related to time, such as “hora” (hour), “martes” (Tuesday), “madrugada” (early morning), “septiembre” (September), and “miercol” (Wednesday), which may indicate the time or frequency of the hurricane strike. The proportion of different topics for the whole dataset of Spanish tweets and the top 1% of most retweeted tweets were different. The highest topic was Topic 8 for the whole Spanish tweets, while Topic 9 for the top 1% of most retweeted tweets. Topics 2, 3, and 9 were the second-highest topics for the whole dataset with a proportion of 14%. For the top 1% of most retweeted tweets, Topics 6 and 8 were the second and third highest topics. Generally, people tweeted about the strength of the storm, alerting people to the disaster, as well as broad time and location.

5.3.2 During the Hurricane

The results of English and Spanish "During Hurricane" tweets topic modeling are listed below (in *Table 5.5* and *5.6*); the order of topics and words was generated by the model automatically:

Table 5.5. Topics and Words of English Tweets (During the Hurricane).

Topic	Words
0	depend, change, live, show, watch, update, news, hit, happen, coverage.
1	vote, help, need, affect, relief, disaster, donate, support, impact, community.
2	safe, stay, friend, hope, family, affect, thought, path, effect, pet.
3	top, power, damage, home, people, lose, hit, state, leave, line.
4	road, car, drive, climate, hurricane, beautiful, million, climatechange, year, ianhurricane.
5	insurance, victim, use, close, resource, open, due, assist, flood, water.
6	hit, prayer, see, flood, people, go, pray, water, damage, get.
7	month, person, next, life, happen, get, go, change, hit, power.
8	mostly, show, new, check, post, life, video, climate_change, assistance, suffer.
9	west_coast, flood, hit, storm, show, wind, make, rain, change, landfall.

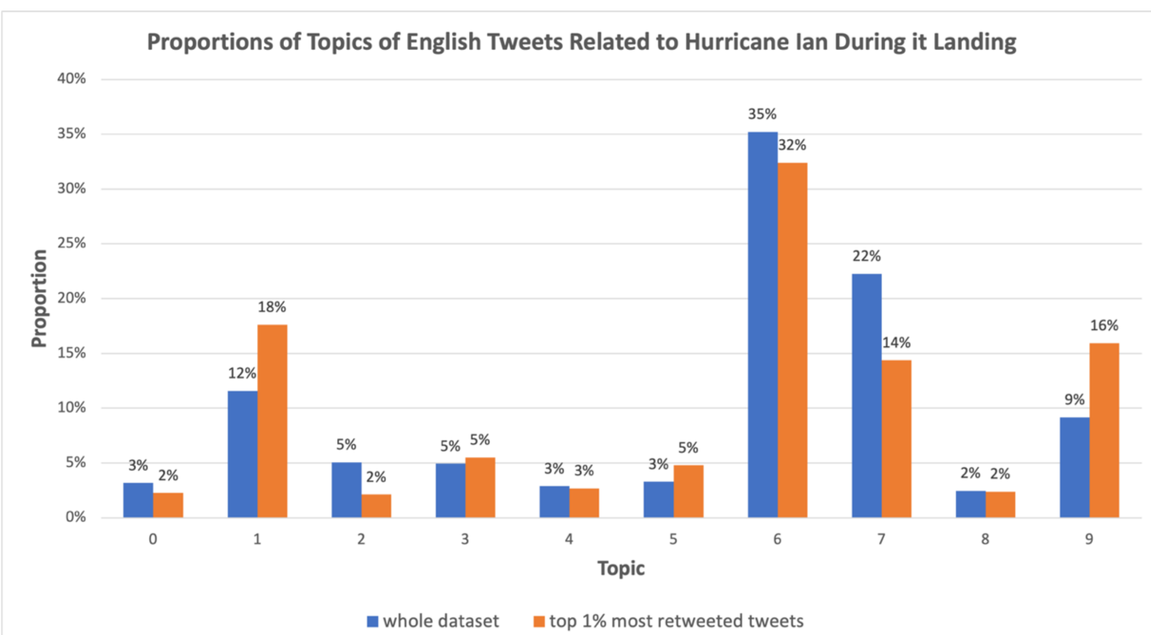


Figure 5.12. The proportion of Topics Related to English Tweets During the Hurricane Landing.

Table 5.6. Topics and Words of Spanish Tweets (During the Hurricane).

Topic	Words
0	inundación, florido, paso, dejar, fuerte, Cubano, mas, destruccion, viento, catastrofica.
1	pasar, ahora, mas, possible, emergencia, dano, ciudad, tormenta, splidario, mundo.
2	paso, muerto, dejar, dano, menos, persona, mas, florido, millón, electrico.
3	tierra, tocar, florido, categorio, viento, categoria, hora, sur, costa, carolín.
4	florido, paso, pedir, ayuda, familia, desastre, video, vivir, ir, consecuencia.
5	afectado, pinar, rio, paso, provincial, fuerzapinar, mas, zona, territorio, hoy.
6	imagen, florido, ver, asi, paso, llegar, bien, mostrar, destrozo, video.
7	pueblo, solidaridad, fuerzapinar, paso, mexico, calle, dano, apoyo, llevar, cubano.
8	vivo, florido, noticia, agua, seguir, electricidad, azotar, hogar, usar, paso.
9	mas, florido, president, recuperar, bidir, ayudar, podrio, historia, paso, joe_bidir.

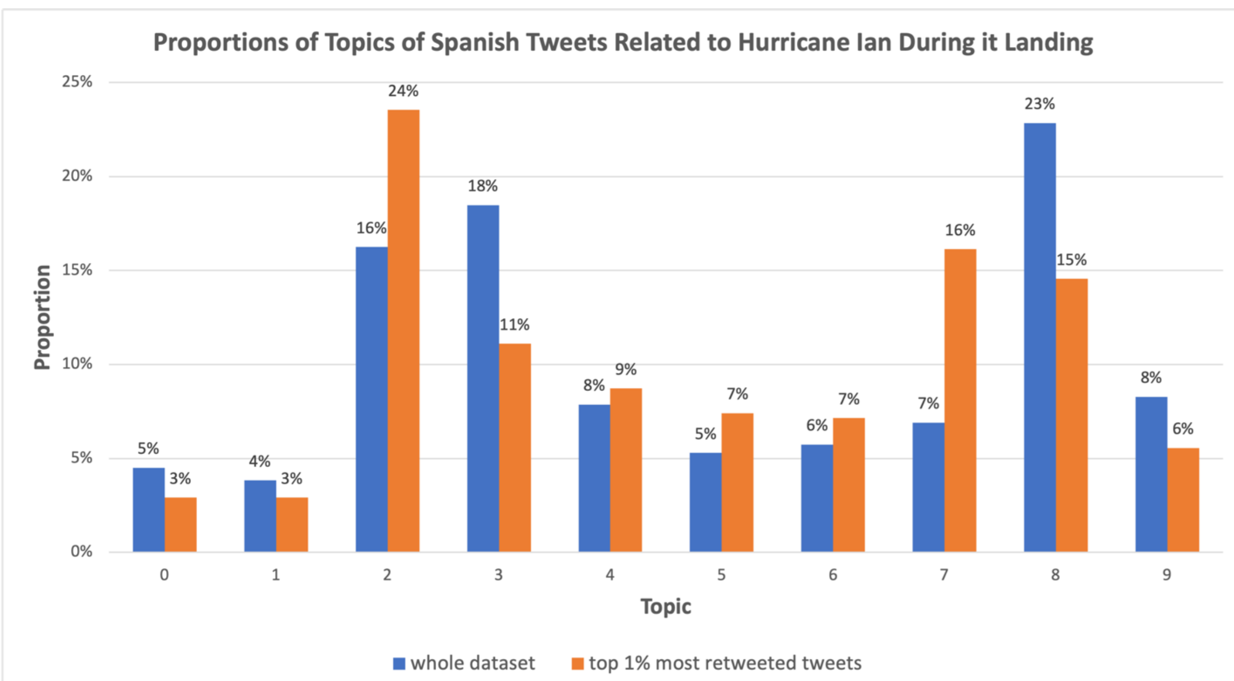


Figure 5.13. Proportion of Topics Related to Spanish Tweets During the Hurricane Landing.

Table 5.7. Topics Related to Risk Assessment and Rescue Scheduling of Twitter Data.

Topic	Related Words (English)	Related Words (Spanish)
Disaster-related	“hurricane”, “disaster”, “ianhurricane”, “flood”, “storm”, “landfall”	“inundación”, “paso”, “pasar”, “tormenta”, “desastre”
Weather-related	“climatechange”, “hit”, “change”, “wind”, “rain”	“viento”, “ver”, “llegar”, “rio”
Emotion-related	“friend”, “hope”, “family”, “pet”, “home”, “people”, “lose”, “prayer”, “suffer”	“sp(o)lidario”, “familia”, “hogar”, “solidaridad”
Action-related	“news”, “vote”, “help”, “donate”, “support”, “insurance”, “assist”, “leave”, “get”, “go”, “assistance”, “check”, “video”, “relief”	“dejar”, “tocar”, “pedir”, “ayuda”, “ir”, “apoyo”, “llevar”, “seguir”, “recuperar”, “ayuda”
Situation-related	“update”, “stay”, “safe”, “need”, “water”, “community”, “affect”, “impact”, “mostly”, “damage”, “power”, “road”, “car”, “life”, “top”	“destruccion”, “tierra”, “catastrofica”, “vivir”, “emergencia”, “dano”, “muerto”, “electrico”, “electricity”, “desastre”, “destrozo”, “agua”, “historia”, “consecuencia”
Spatial/Temporal-related	“year”, “month”, “west_coast”, “new”, “next”	“florido”, “cubano”, “ahora”, “ciudad”, “zona”, “hora”, “sur”, “costa”, “hoy”, “pueblo”, “calle”

In the period “During the hurricane landing”, English disaster- and weather-related words had not changed much. The list of emotion-related words added “hope”, “pet”, “lose” and “suffer”. People were experiencing damage and loss during the strike of the hurricane. Other than family and friends, pets were also worried about. In action-related words, “donate” and “insurance” were added to the list, which related to the financial system of the community. Words like “support”, “assist”, “check” and “relief” also appeared, which were related to rescuing. In the situation-related words, a resource like water was listed. Words such as “damage”, “power”, “road”, “car”, and “life” would help assess the risk or damages of the event. The spatial or temporal related words of English tweets were not informative again. Only the word "west_coast" indicated the location, however, too broad. Topic 6 had the highest proportion for both the whole dataset and the top 1% of retweeted tweets. The second and third-highest topics for the whole dataset were Topic 7 and 1, though the opposite for the top 1% of retweeted tweets. People tweeted about the current situation of the disaster (such as the damages from the storm), and the resources they needed (such as water and power shortage).

For Spanish tweets, "desastre" (disaster) was added to the disaster-related words, which were more serious than English words. Weather-related words for "during the hurricane landing" and "before the hurricane landed" were the same. Emotion-related words for English and Spanish tweets were very similar, as people cared about their families and homes. Action-related words like "dejar" (leave), and "apoyo" (support) appeared again. New words such as "pedir" (ask), "ayuda" (aid), "ir" (go), "llevar" (carry), "seguir" (continue), "recuperar" (recover), "ayuda" (help) were added. People could use social media networks to ask for assistance or seek help. In this period, recovery was also discussed, which was earlier than it was in English tweets. Situation-related words in Spanish were more intense in this period as well. The words "catastrofica" (catastrophic), "muerto" (dead), "desastre" (disaster), "destrozo" (smashed), "historia" (history), and "consecuencia" (consequence) indicated that the damages of the hurricane during its landing were severe. Spanish tweets provided information related to time and location, such as "ahora" (now), "hora" (hour), "hoy" (today), "sur" (south), "ciudad" (city), "costa" (coast), "pueblo" (town), and "calle" (street). The spatial- or temporal-related words were more precise than those in the period of "before the hurricane landed". The 3 topics with the highest proportions for the whole dataset were Topics 8, 3, and 2, while Topics 2, 7, and 8 were for the top 1% of retweeted tweets. People tweeted about the resource shortage (water and electricity), the places where most were damaged, and the loss of people (damages and deaths).

5.3.3 After the Hurricane

The results of English and Spanish "After Hurricane" tweets topic modeling are listed below (in *Table 5.8* and *5.9*); the order of topics and words was generated by the model automatically:

Table 5.8. Topics and Words in English Tweets (After the Hurricane).

Topic	Words
0	help, donation, animals_affect, need, support, food, claim, insurance, today, people.
1	hurricanenicole, storm, go, get, people, still, stay, story, keep, rise.
2	fortmyer, debris, solar, leave, clean, get, story, tree, need, fort_myer.
3	damage, power, day, thank, late, restore, storm, work, weather, hit.
4	damage, roof, lose, home, little, woman, get, take, house, recently.
5	help, disaster, recovery, need, community, support, relief, impact, effort, affect.
6	help, vote, student, disaster, day, people, time, election, make, recent.
7	flood, see, new, question, home, check, help, cause, video, share.
8	relief, help, donate, support, victim, family, thank, affect, friend, click.
9	insurance, get, still, many, flood, home, storm, hurricane, see, destroy.

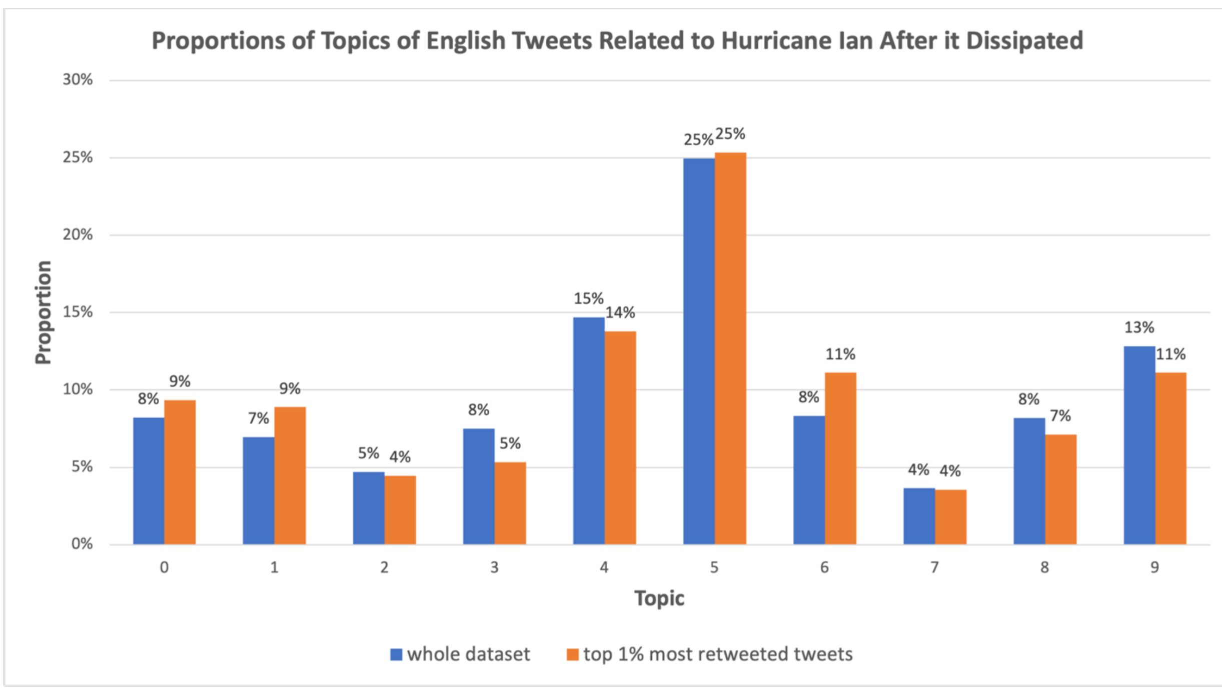


Figure 5.14. Proportion of Topics Related to English Tweets After the Hurricane Dissipated.

Table 5.9. Topics and Words in Spanish Tweets (After the Hurricane).

Topic	Words
0	paso, pinar, rio, bloqueo, cubano, pueblo, mas, diaz_canel, hermano, seguir.
1	millón, florido, mas, persona, unido, dolar, ayudar, despues, dejar, estrago.
2	paso, muerto, mas, dejar, florido, dano, aumentar, seguir, menos, autoridad.
3	president, afectado, zona, mas, dano, bidir, pueblo, florido, solidaridad, causado.
4	dano, cubano, agradecer, millón, damnificado, paso, solidaridad, ayuda, residente, enviar.
5	damidicado, cubaporlapaz, ayuda, paso, humanitario, cubano, agua, hecho, cuba, pinar.
6	paso, trabajador, pinar, afectado, rio, provincial, dano, pinardelrio, apoyo, fierzapinar.
7	donación, damnificado, paso, mexico, donar, mil_cocina, luego, octubre, mejorsinbloqueo, emergencia.
8	paso, florido, mas, damnificado, hacer, afectado, caso, noticia, sur, ver.
9	paso, pinar, afectado, rio, provincial, fuerzapinar, president, municipio, recuperacion, pinardelrio.

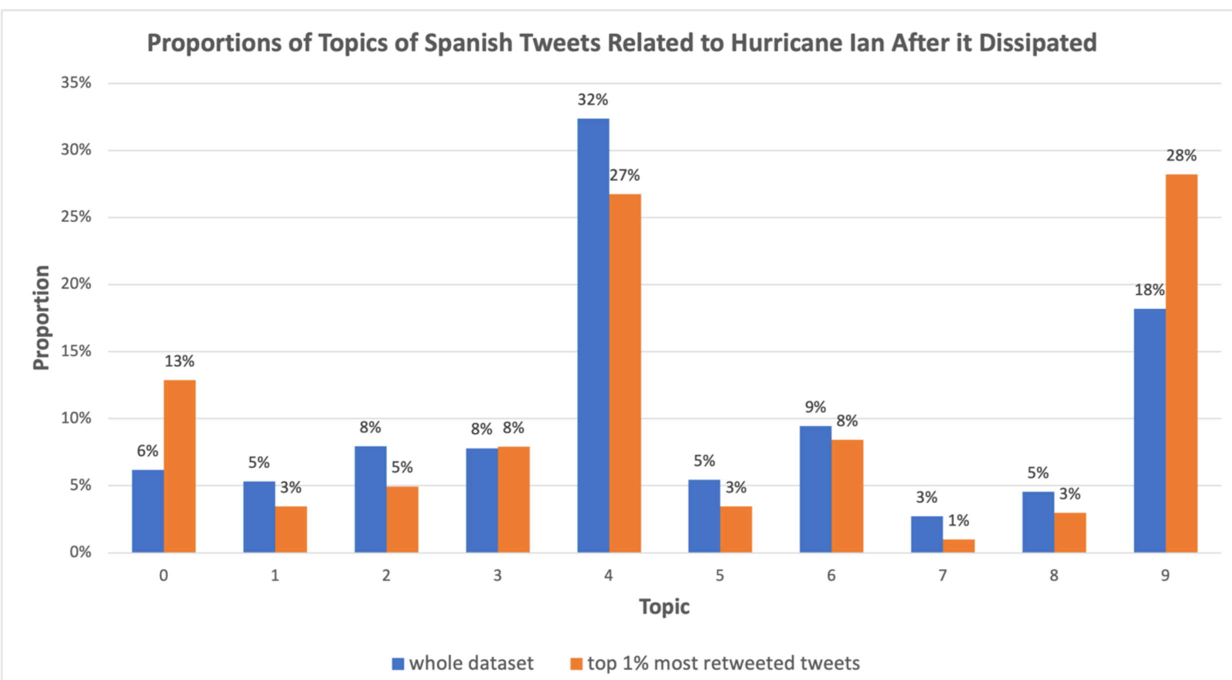


Figure 5.15. The proportion of Topics Related to Spanish Tweets After the Hurricane Dissipated.

Table 5.10. Topics Related to Recovery of Twitter Data.

Topic	Related Words (English)	Related Words (Spanish)
Disaster-related	"hurricanenicole", "storm", "flood", "hurricane"	"paso"
Weather-related	"weather", "hit", "see"	"rio", "ver"
Emotion-related	"thank", "lose", "home", "victim", "family", "friend"	"hermano", "unido", "solidaridad", "agradecer", "damnificado", "cuba por la paz"
Action-related	"help", "donate", "donation", "support", "claim", "insurance", "go", "get", "stay", "keep", "solar", "leave", "restore", "share", "work", "vote", "help", "support", "video", "relief", "clean"	"seguir", "ayudar", "dejar", "enviar", "hecho", "apoyo", "donación", "hacer", "noticia", "recuperacion", "municipio", "residente"
Situation-related	"need", "food", "rise", "debris", "tree", "power", "damage", "roof", "house", "still", "student", "disaster", "destroy", "recovery", "community"	"bloqueo", "major si bloqueo", "millón", "estragon", "muerto", "dano", "causado", "humanitario", "trabajador", "pinar del rio"
Spatial/Temporal-related	"today", "fortmyer", "day", "recently"	"pueblo", "Cubano", "diaz_canel", "despues", "luego", "octubre"

In the period "after the hurricane dissipated", disaster- and weather-related English words were very few and less informative. More positive terms like "thank" appeared in emotion-related words, however, "victim" and "lose" were still on the list, as people were mourning the loss. On the list of action-related words, "help", "donation" "support" and "insurance" still existed, as the community would need financial support after experiencing the disaster. Words like "restore", "relief", "clean" and "work" appeared in this period, as people were starting to enter the recovery phase. As for words related to time and location, "fortmyer" was listed, which was a city (Fort Myers) in Florida. The 3 topics with the highest proportions for the whole dataset and the top 1%

of retweeted tweets were the same, which were Topics 5, 4, and 9. People were discussing the damages, and how they could recover from the disaster.

For Spanish tweets, there were only 1 or 2 words related to disaster or weather. Emotion-related words were two-sided while mostly positive, only the word “damificado” (damaged) was negative. The rest words like “unido” (united), “solidaridad” (solidarity), “agradecer” (thank), and “cuba por la paz” (Cuba for peace) were all positive. Action-related words were very similar to those in the period of “during the hurricane landed”. Words “seguir” (continue), “ayudar” (help), “dejar” (leave), “apoyo” (support), “noticia” (news) appeared again. New words such as “enviar” (send), “hecho” (made), “donación” (donation), “hacer” (do), “municipio” (municipality), and “recuperacion” (recovery) were listed. The existence of new words indicated that the community or municipality would need financial support in the phase of recovery. For situation-related words, “bloqueo” (blocking), “estragon” (havoc), “muerto” (dead), “humanitario” (humanitarian), and “trabajador” (worker) were generated. As for spatial- or temporal-related words, “pueblo” (town), “diaz_canel”, “despues” (after), “luego” (then), and “octubre” (October) were listed. Like other periods, the Spanish words related to time and location were more precise than English ones. The 3 topics with the highest proportions for the whole dataset were Topics 4, 9, and 6, while the top 1% of retweeted tweets were Topics 9, 4, and 0. People were discussing how the damages affected people, and how they could be united to recover from the disaster.

Compared to English tweets, Spanish tweets had relatively more informative and precise words that related to time or location. However, the words in Spanish tweets related to disaster and weather were less than in English tweets. For emotion-related words, both English and Spanish tweets mentioned family, friends, home, and loss. Pets appeared in English tweets, while words such as united or solidarity appeared in Spanish tweets. The action- and situation-related words were very similar for English and Spanish tweets; however, Spanish words were more intense. The number of words related to disaster and weather decreased as time passed. Action- and situation-related words slightly increased as time passed. The spatial- or temporal-related words for English tweets were consistently less informative, while the number of those for Spanish tweets decreased as time passed. The precision of time- and location-related words for both

English and Spanish tweets were not high, while Spanish words were more precise than English ones.

5.4 Sentiment Analysis

Sentiment analysis was applied for the whole dataset, and datasets of "before", "during" and "after" periods of the hurricane for both English and Spanish tweets. The model from "vaderSentiment" was used for English tweets, in which a score smaller than or equal to -0.05 represented "Negative" emotion, a score larger than or equal to 0.05 represented "Positive" emotion, and the rest scores in between represented "Neutral". The model "vader-multi" was used for Spanish tweets, which has the same scoring system as the model used in English. For both English and Spanish tweets, three types of emotion were all detected. For both languages, sentences with a score of 0.0 were excluded because it was not clear whether these sentences were completely neutral, or they were not assigned with sentiment scores at all.

The datasets of tweets of the accounts detected as human were combined with their sentiment scores and the texts of the emotion ("Positive", "Negative" and "Neutral"). Three types of emotions in English and Spanish tweets were mapped with the creation time (on a daily basis) of tweets, as shown in *Figures 5.16*, *5.17, 5.18,* and *5.19*. The proportion of emotions of different periods was also mapped separately for English and Spanish tweets, as shown in *Figures 5.20, 5.21, 5.22,* and *5.24*. For each language, the trend and proportion of sentiments in each period were compared between the whole dataset (all of the tweets) and the dataset with the top 1% tweeted that had been mostly retweeted.

5.4.1 The Trend of Sentiments of English and Spanish Tweets in Different Periods

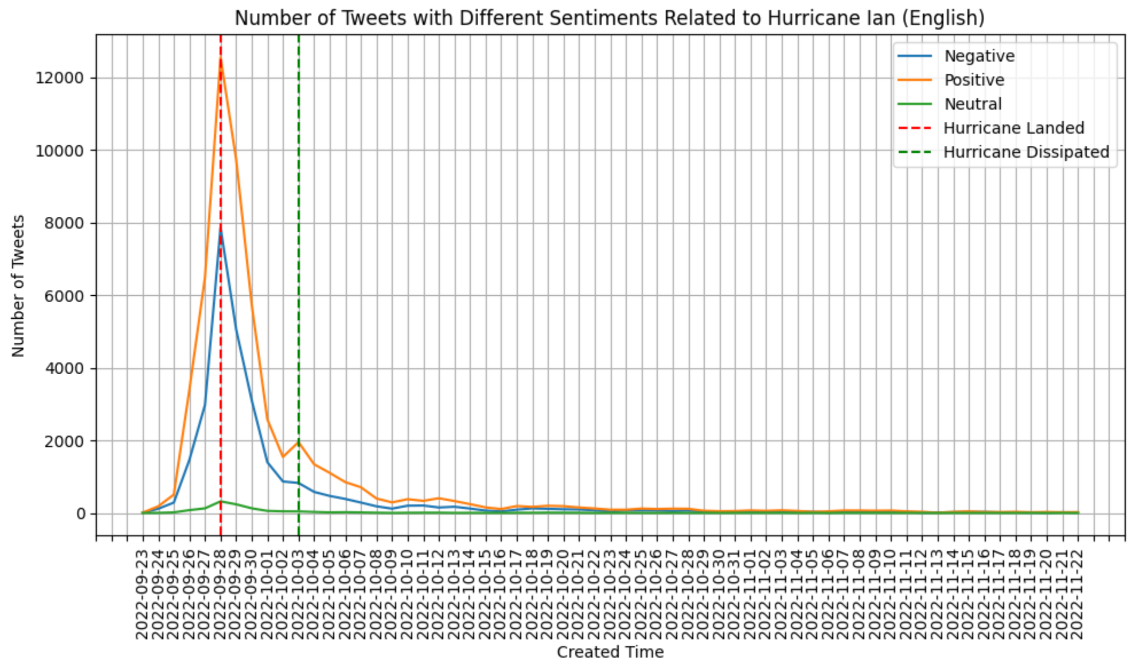


Figure 5.16. The Trend of Sentiments (Negative, Positive, and Neutral) of All Tweets Related to Hurricane Ian (English).

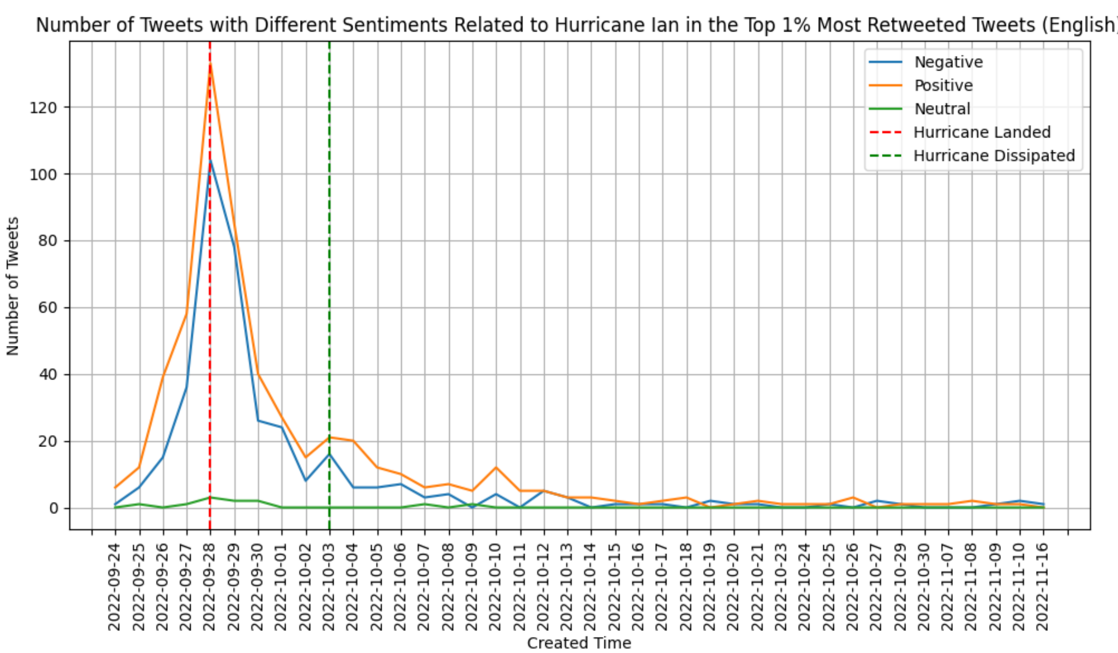


Figure 5.17. The Trend of Sentiments (Negative, Positive, and Neutral) of the Top 1% of Tweets that had been Most Retweeted Related to Hurricane Ian (English).

The trends of “Negative” and “Positive” emotions for all English tweets were consistent with the activities of Twitter (i.e., the number of tweets) throughout the hurricane event (as shown in *Figure 5.4*). The scores for all sentiments slightly grew in the first two days, and dramatically increased in the following three days, which peaked when the hurricane landed on Florida on September 28th, 2022. The scores dropped fast and vastly during the hurricane hitting Florida and slightly bounced up for positive tweets (but not for negative tweets) when the hurricane dissipated on October 3rd, 2022. The number of positive tweets was more than those of negative sentiment for the whole time. Even though the sentiments have the same trend, the overall sentiment of English tweets was positive.

The overall trend of “Negative” and “Positive” sentiments for the top 1% of English tweets that had been mostly retweeted was similar to the trend of all English tweets. The number of tweets with positive sentiment was higher than the negative ones most of the time, except for several days in the period of “after the hurricane”. In the dataset of most retweeted tweets, positive sentiment was obviously stronger than negative sentiment in the period “before the hurricane”, while the difference between positive and negative sentiments shrunk in the following periods.

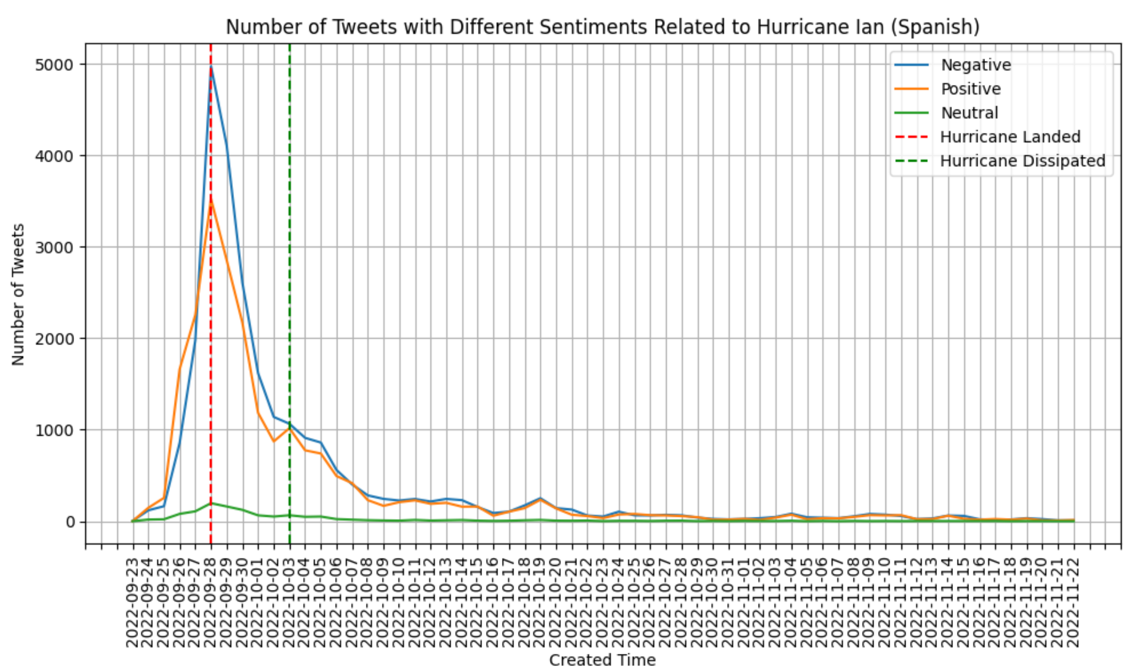


Figure 5.18. The Trend of Sentiment (Negative and Positive) of All Tweets Related to Hurricane Ian (Spanish).

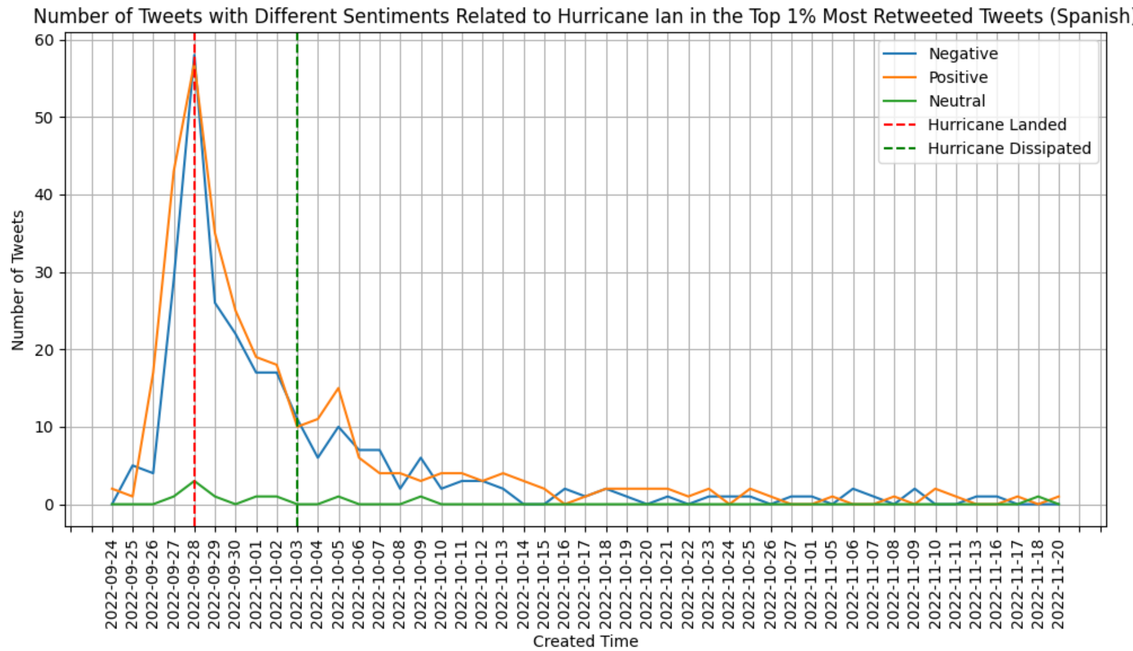


Figure 5.19. The Trend of Sentiments (Negative, Positive, and Neutral) of the Top 1% of Tweets that had been Most Retweeted Related to Hurricane Ian (Spanish).

The trend of "Negative" and "Positive" emotions for Spanish tweets was similar to the changes in Twitter activities. The number of tweets of both emotions increased slowly for the first two days and rushed up to its peak on September 28th, when the hurricane landed in Florida. Until the 27th of September, the number of positive tweets was more than the number of negative tweets. For the following three days, the number of tweets of both emotions reduced abruptly, then gradually decreased with small bounce ups and downs since then. On the day of the hurricane dissipated, positive tweets slightly increased, while the number of negative tweets continued to decrease. In Spanish tweets, the number of "Negative" tweets is more than the number of "Positive" tweets ever since the day before the hurricane landed in Florida.

The overall trend of "Negative" and "Positive" sentiments for the top 1% of Spanish tweets that had been mostly retweeted was different from the trend of all Spanish tweets. The number of tweets with positive sentiment was higher than those with negative sentiment most of the time, except for the second day before the hurricane landed, the day when the hurricane landing in Florida, and several days after the hurricane dissipated. Compared to all Spanish tweets, the overall sentiment of the top 1% of most retweeted Spanish tweets was more positive.

5.4.2 The Proportion of Sentiments of English and Spanish Tweets in Different Periods

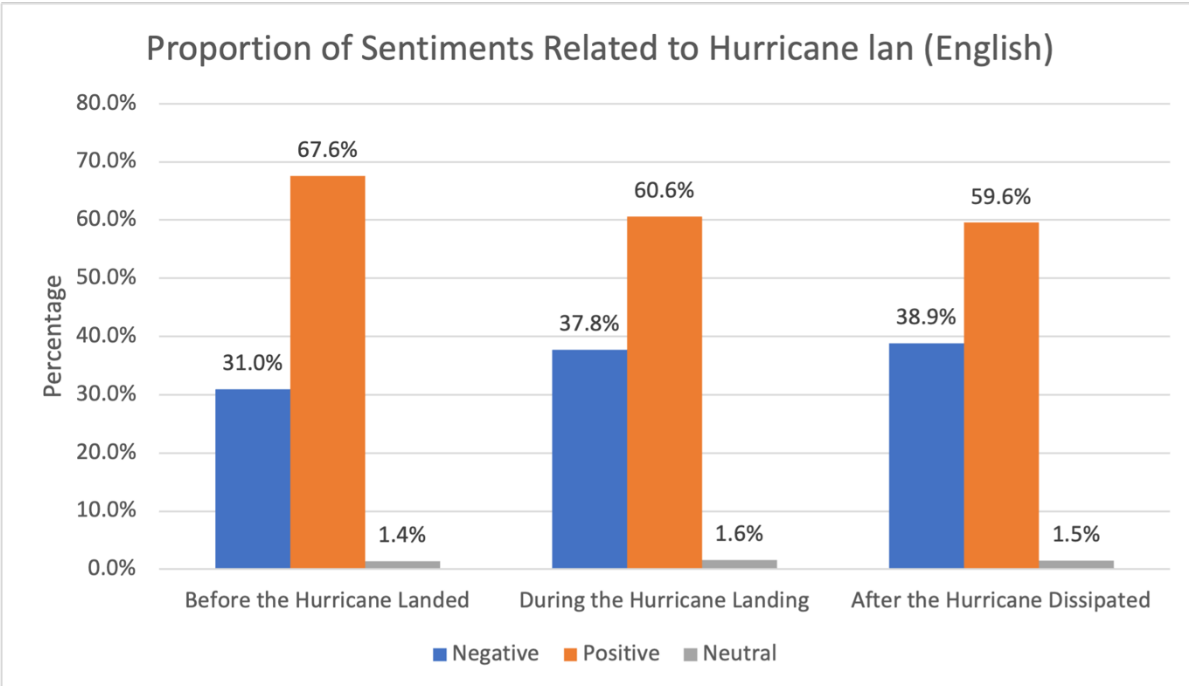


Figure 5.20. Proportions of Sentiments of All Tweets Related to Hurricane Ian (English) Before the Hurricane Landed, During the Hurricane Landing, and After the Hurricane Dissipated.

Of all time, the proportion of “Negative” sentiment was around half of the “Positive” sentiment in the dataset of all English tweets. However, the proportion of “Negative” tweets changed slightly during different periods. Before the hurricane landed in Florida, the proportion of “Negative” tweets was 31.0%, and the proportion of “Positive” tweets was around 67.6%. During the hurricane hitting Florida, the proportion of “Negative” tweets increased by 6.8%, and the proportion of “Positive” tweets decreased by 7%. After the hurricane dissipated, the proportion of “Negative” tweets was 38.9%, about 1% higher than of the “during the hurricane landing” period. The proportion of “Positive” tweets was 59.6%, which was lower than the proportion of the previous periods. Even though the overall emotion of English tweets was “Positive”, the proportion of “Negative” tweets did increase as the hurricane striking Florida. The trend of “Negative” sentiment in tweets is partly different from the hypothesis, that people’s emotions become negative as the hurricane comes close, however, the negative emotions did not drop after the hurricane dissipated but continued to increase for English-speaking social media users.

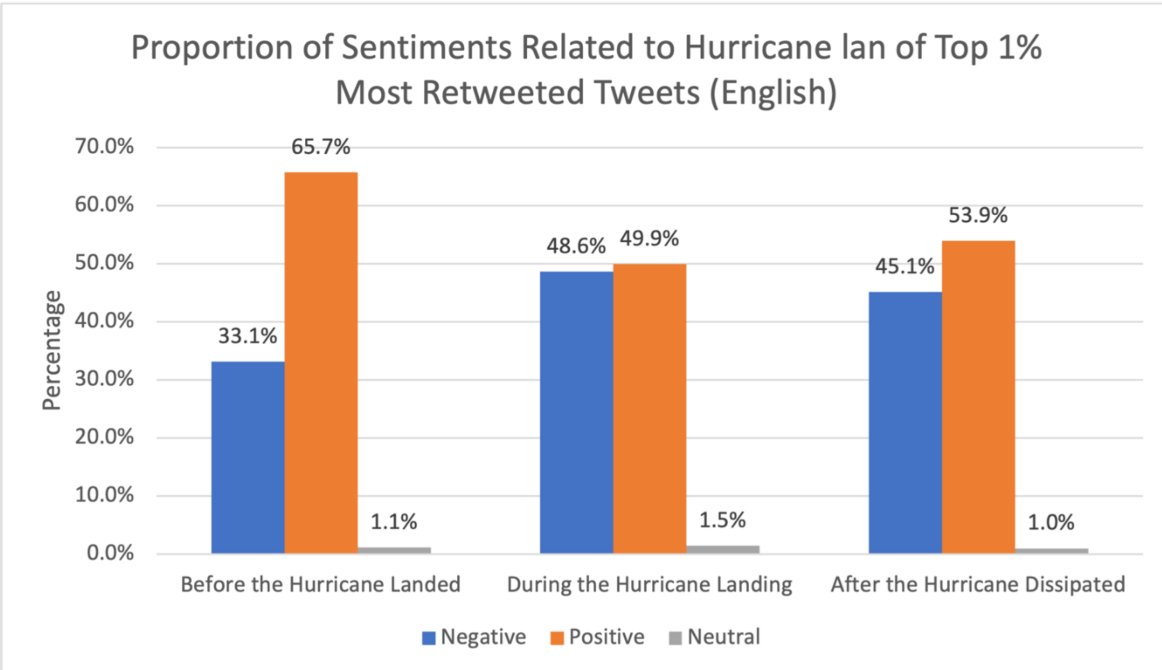


Figure 5.21. Proportions of Sentiments of the Top 1% Tweets that had been Mostly Retweeted Related to Hurricane Ian (English) Before the Hurricane Landed, During the Hurricane Landing, and After the Hurricane Dissipated.

The proportion of positive sentiment was higher than that of negative sentiment of all time in the dataset of the top 1% of English tweets that had been mostly retweeted. Although the overall trend was the same for the whole dataset and the most retweeted tweets, the proportions of different sentiments in each period changed. Compared to the whole dataset of English tweets, the negative and positive sentiments were similar in the period "before the hurricane", with 33.1% and 65.7%, respectively. However, in the following periods, negative sentiment increased. During the hurricane landing in Florida, negative tweets contained 48.6% of the retweeted tweets dataset, which was almost equal to the proportion of positive tweets (49.9%). In the period "after the hurricane dissipated", the proportion of negative sentiment decreased slightly, while still higher than it of all English tweets. In the dataset of the top 1% of most retweeted tweets, negative emotions grew when the storm approached then slightly dropped after the hurricane dissipated, which is the same as stated in the hypothesis.

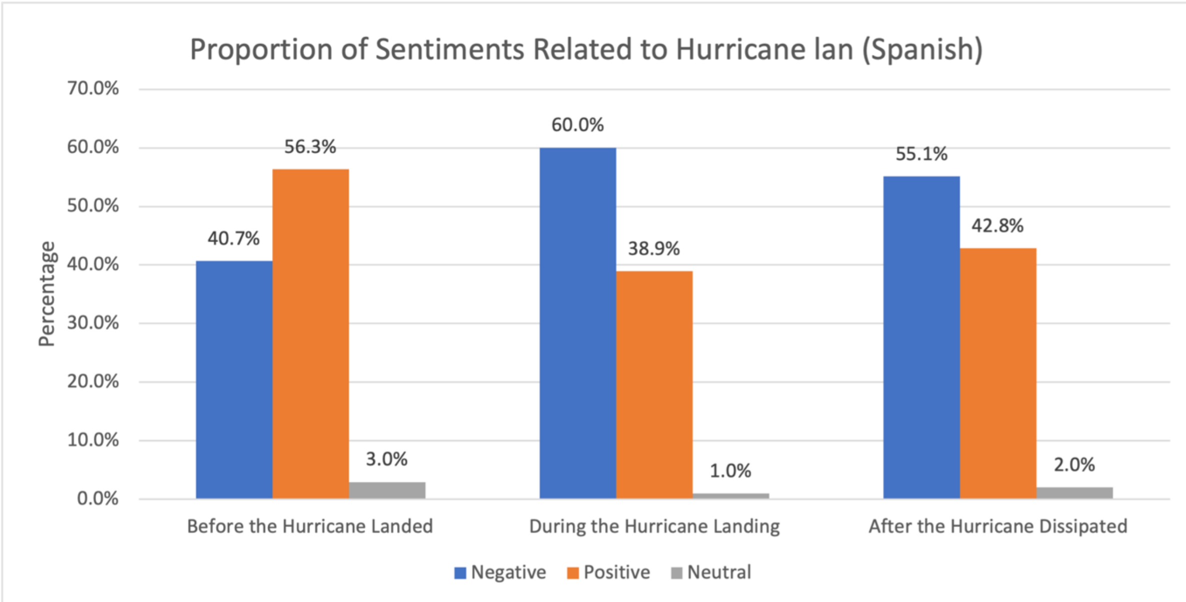


Figure 5.22. Proportions of Sentiments of All Tweets Related to Hurricane Ian (Spanish) Before the Hurricane Landed, During the Hurricane Landing, and After the Hurricane Dissipated.

For the whole dataset of Spanish tweets, the proportion of "Negative" emotions was lower than the "Positive" emotions in the period before the hurricane but higher in the following periods. Before the hurricane landed in Florida, the proportion of "Negative" tweets was 40.7%, and the proportion of "Positive" tweets was around 56.3%. During the hurricane hitting Florida, the proportion of "Negative" tweets increased by 20%, and the proportion of "Positive" tweets decreased by 17%. After the hurricane dissipated, the proportion of "Negative" tweets was 55.1%, about 4% lower than of the "during the hurricane landing" period. The proportion of "Positive" tweets was 42.8%, which was higher than the proportion of the previous period, but lower than the first period. The trend of the negative tweets shown above proved the hypothesis that people tweeted about negative content during the hurricane, while the negative emotions decreased after the hurricane dissipated.

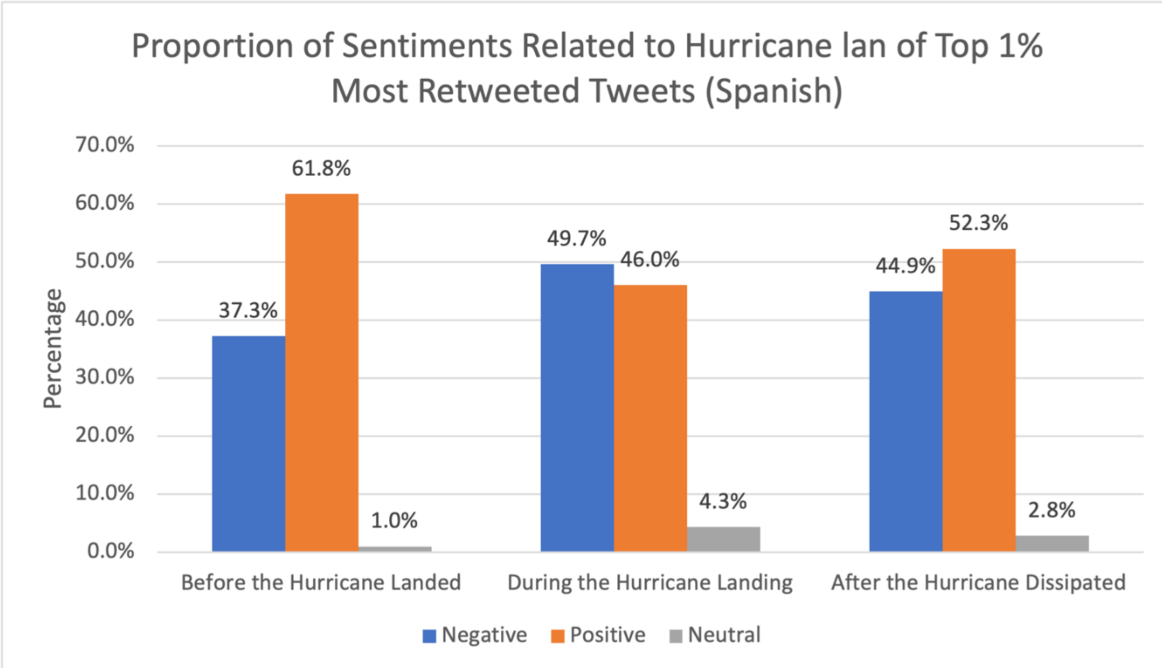


Figure 5.23. Proportions of Sentiments of the Top 1% Tweets that had been Mostly Retweeted Related to Hurricane Ian (Spanish) Before the Hurricane Landed, During the Hurricane Landing, and After the Hurricane Dissipated.

The proportion of negative and positive sentiment in the dataset of the Top 1% of most retweeted Spanish tweets was similar to the whole dataset in the first two periods, while different in the last period. The overall positive sentiment of the concentrated dataset was higher than that of the whole dataset. Before the hurricane landed in Florida, the proportion of negative sentiment was 37.3% in this smaller dataset, which is lower than it was in the whole dataset. During the hurricane landing in Florida, the proportion of negative sentiment (49.7%) beat the positive ones, while still lower when compared to the whole dataset. In the period “after the hurricane”, the proportion of negative tweets was lower than positive ones, which is the opposite in the whole dataset. In the tweets that had been mostly retweeted, people tended to be more positive in the after period. The trend of negative tweets also proved the hypothesis that people felt more negative when the hurricane struck and turned to more positive feelings after it dissipated.

5.4.3 The Sentiments of Each Topic of English and Spanish Tweets

5.4.3.1 Before the Hurricane Landed

In this period of "before the hurricane", both English and Spanish tweets had more positive sentiment than negative sentiment. This trend was the same for both the dataset of all tweets and the dataset of the top 1% of tweets that had been retweeted.

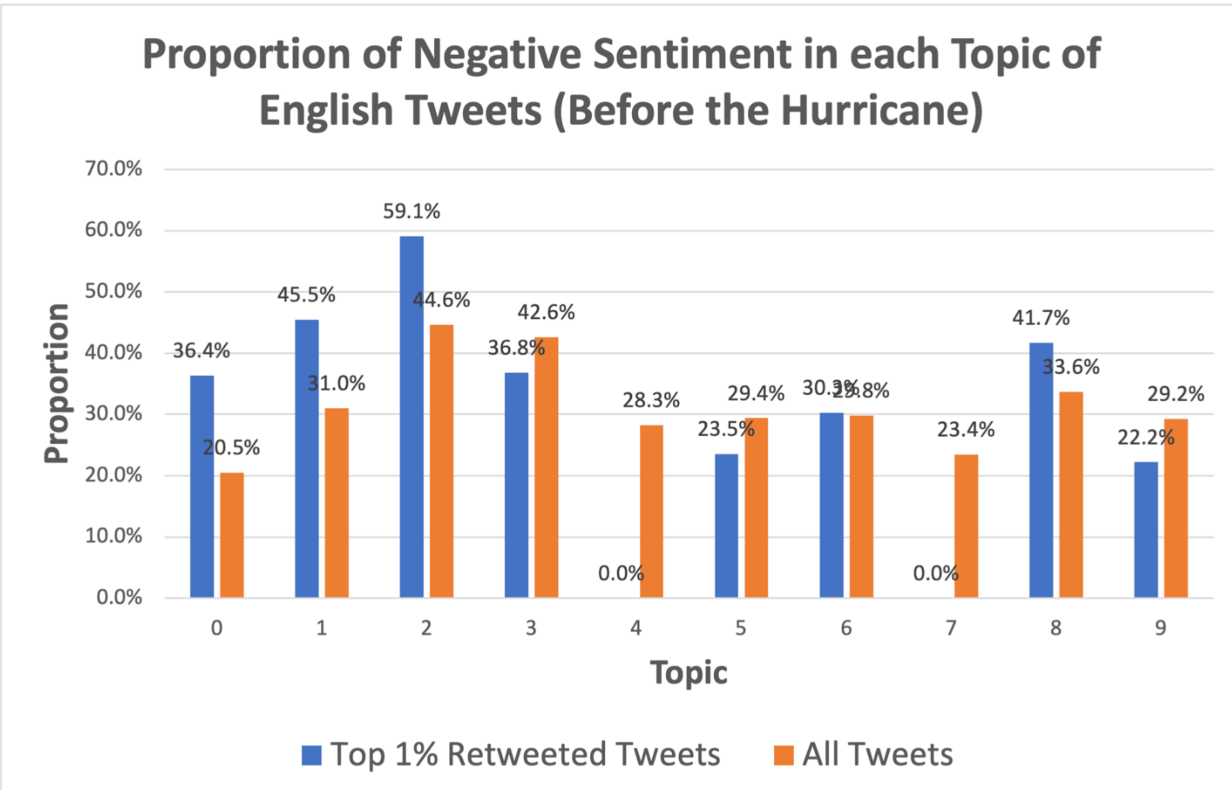


Figure 5.24. The proportion of Negative Sentiment Tweets in each Topic of English Tweets (Before the Hurricane).

As shown in *Figure 5.24*, the proportions of negative sentiment of each topic in the top 1% of tweets that have been mostly retweeted and all English tweets were different. In most of the topics, the number of tweets detected as "positive" was higher than it detected as "negative". Only topic 2 in the top 1% of most retweeted tweets had higher negative sentiment than positive sentiment. For topics 0, 1, 2, 6, and 8, the proportions of negative emotion in the top 1% of most retweeted tweets were higher than those in all tweets. In these topics, words related to disaster and weather appeared the most. The information about the changes in weather and the approaching of storms was more negative among people who actively talked about Hurricane Ian at this stage. As shown in *Figure 5.10*, the number of tweets that fell into topic 5 was the highest. In the top 1% of most retweeted tweets, the emotion of topic 5 was less negative than it was in all tweets. Topic 5 contained words related to situation and emotion, which expressed the expectation of people to stay safe and take care of their family and friends. In the period before the hurricane landed in Florida, the English tweets that had been most widely shared among people appeared more negative on topics related to disaster and weather, but more positive on topics related to the situation.

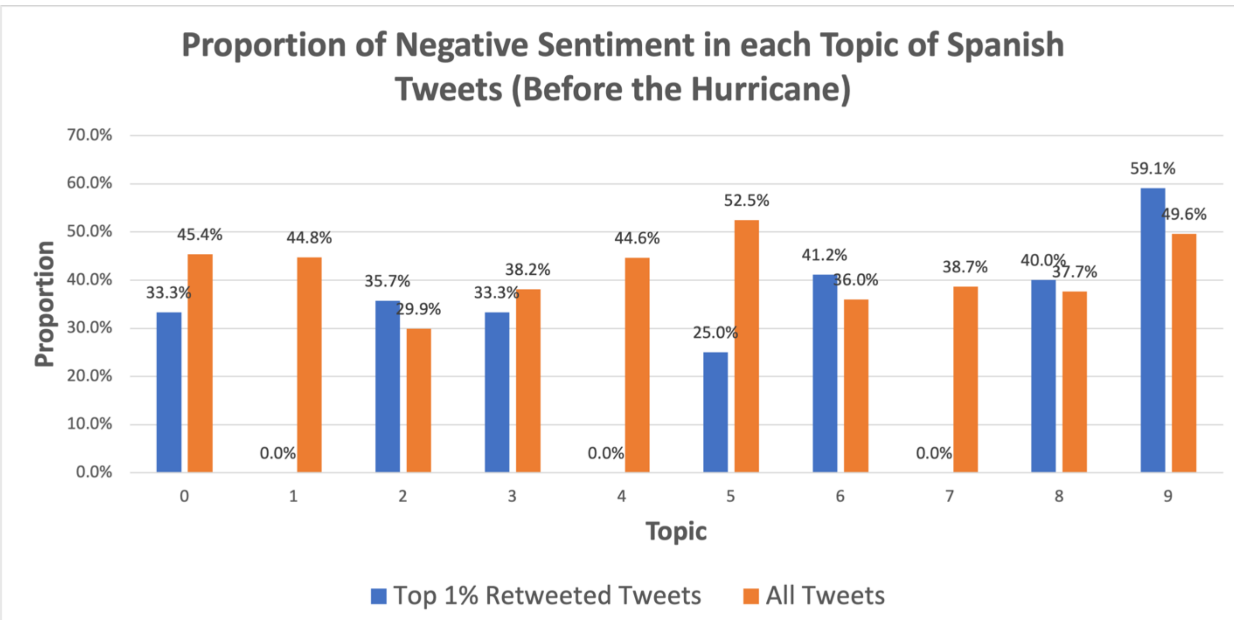


Figure 5.25. The proportion of Negative Sentiment Tweets in each Topic of Spanish Tweets (Before the Hurricane).

As shown in *Figure 5.25*, the number of tweets that had been detected as “negative” was less than the “positive” ones. Topic 9 had more negative sentiment than positive in the top 1% of most retweeted tweets, and topic 5 in all tweets had more negative sentiment than positive. Topics 2, 6, 8, and 9 were more negative in the top 1% of most retweeted tweets than in all tweets. These topics also contained high proportions in Spanish tweets at this stage. In these topics, words related to disaster, weather, and situation appeared most frequently. People felt more negative when they shared information about storms, rain, wind, and damages, although the overall sentiment was positive. As discussed in *Section 5.3.1*, Spanish words that described the disaster and weather were intenser, but the words related to emotion were also stronger, such as “solidarity”. In the period before the hurricane landed in Florida, the Spanish tweets that had been most widely shared among people appeared relatively more negative on topics related to disaster, weather, and situation.

5.4.3.2 During the Hurricane Landing

In this period of “during the hurricane”, English tweets had more positive sentiment than negative sentiment, however, Spanish tweets were detected as more negative. This trend was the same for both the dataset of all tweets and the dataset of the top 1% of tweets that had been retweeted.

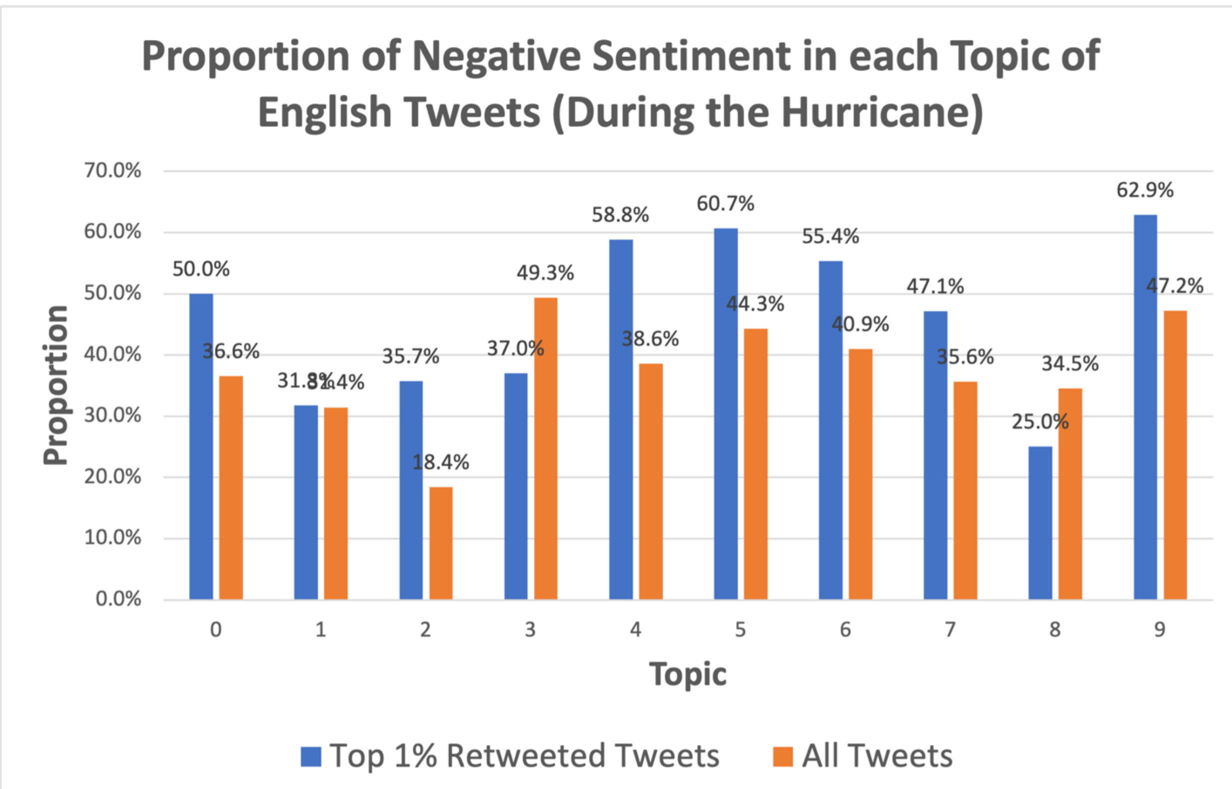


Figure 5.26. The proportion of Negative Sentiment Tweets in each Topic of English Tweets (During the Hurricane).

As shown in *Figure 5.26*, the proportion of negative sentiment in each topic increased compared to the period “before the hurricane”. Especially in the top 1% of most retweeted tweets, the proportions of negative sentiment were detected more than 50% in topics 0, 4, 5, 6, and 9. Words related to action and situation appeared most frequently. Disaster- and weather-related words were less than those in the previous period, however, the sentiment was more negative. In this period, the storms had already landed in Florida and caused damages. During a natural disaster event, people would seek support or assistance, for example, food, evacuation, and shelter. People would also talk about the situation (usually damages) of their properties or even their safety. The information shared could be used in assessing the current risk or situation of the disaster. The shortage of resources, such as food, drinking water, and electricity was also shown in the content of people's tweets. It could help the government and the relevant agencies know the situation of people who were suffering from the disaster and provide timely help to them. In this period during the hurricane landing in Florida, the English tweets shifted to situation- and action-related topics, and the overall negative sentiment grew compared to the "before the hurricane" period.

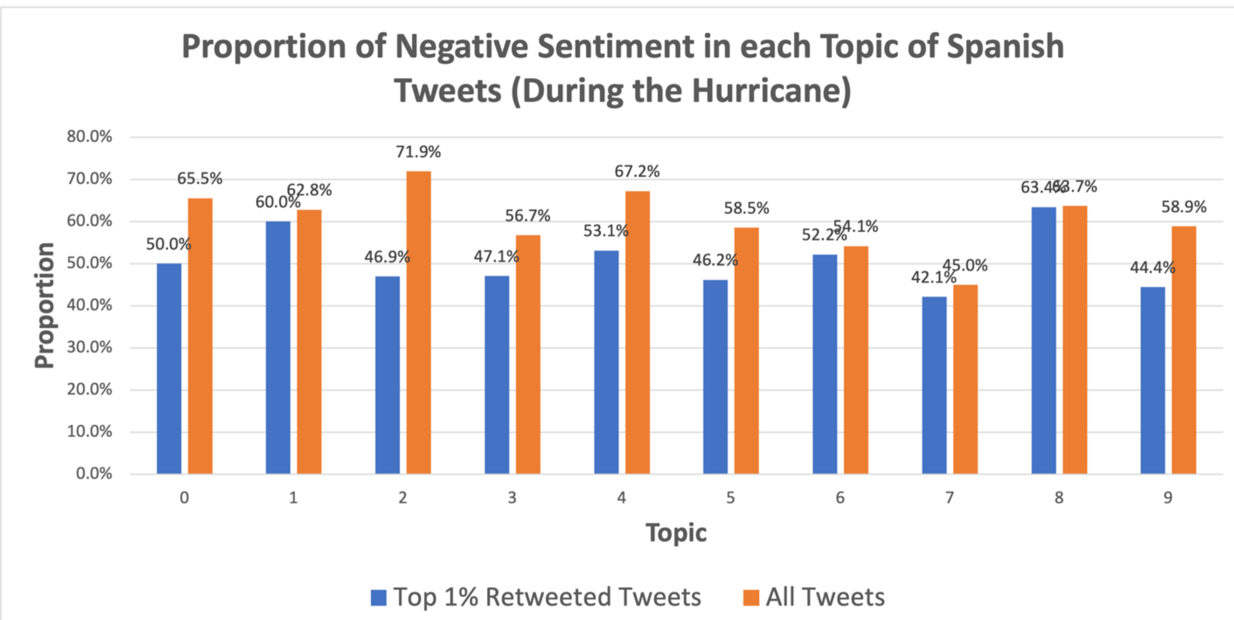


Figure 5.27. The proportion of Negative Sentiment Tweets in each Topic of Spanish Tweets (During the Hurricane).

As shown in *Figure 5.27*, the proportion of negative sentiment in each topic increased for both the whole dataset of Spanish tweets and the top 1% of most retweeted tweets compared to the former period. Negative emotions contained more than half of all the tweets for all topics in the period of "during the hurricane". The topics with the highest proportions of negative sentiment were topics 0, 1, 2, 4, and 8. In these topics, words that related to disaster, action, and situation appeared most frequently. Compared to English tweets, Spanish tweets still had a negative emotion when the content included disaster-related words. It became more negative in this period because the words used were more serious. As for action-related topics, people still focused on seeking support, help, or assistance, which is the same in English tweets. For situation-related topics, the words used in Spanish tweets were much more negative than those in English tweets. Resources like water and electricity were mentioned, which is the same for both languages. In Spanish tweets, people described the event as a catastrophe, and the words “smashed” and “dead” were used when they talked about the situation. As the negative sentiment was stronger in Spanish tweets, the area where Spanish-speaking people live might have been damaged more severely. As the Spanish tweets shifted to action- and situation-related topics, it could help the government and agencies to assess the damages and schedule the rescue, based on the frequency and severity of the words used in this period.

5.4.3.3 After the Hurricane Dissipated

In the period “after the hurricane dissipated”, the overall sentiment was positive for the top 1% of tweets that had been mostly retweeted for both languages. For the whole dataset, the overall sentiment of English tweets was positive, while negative for Spanish tweets.

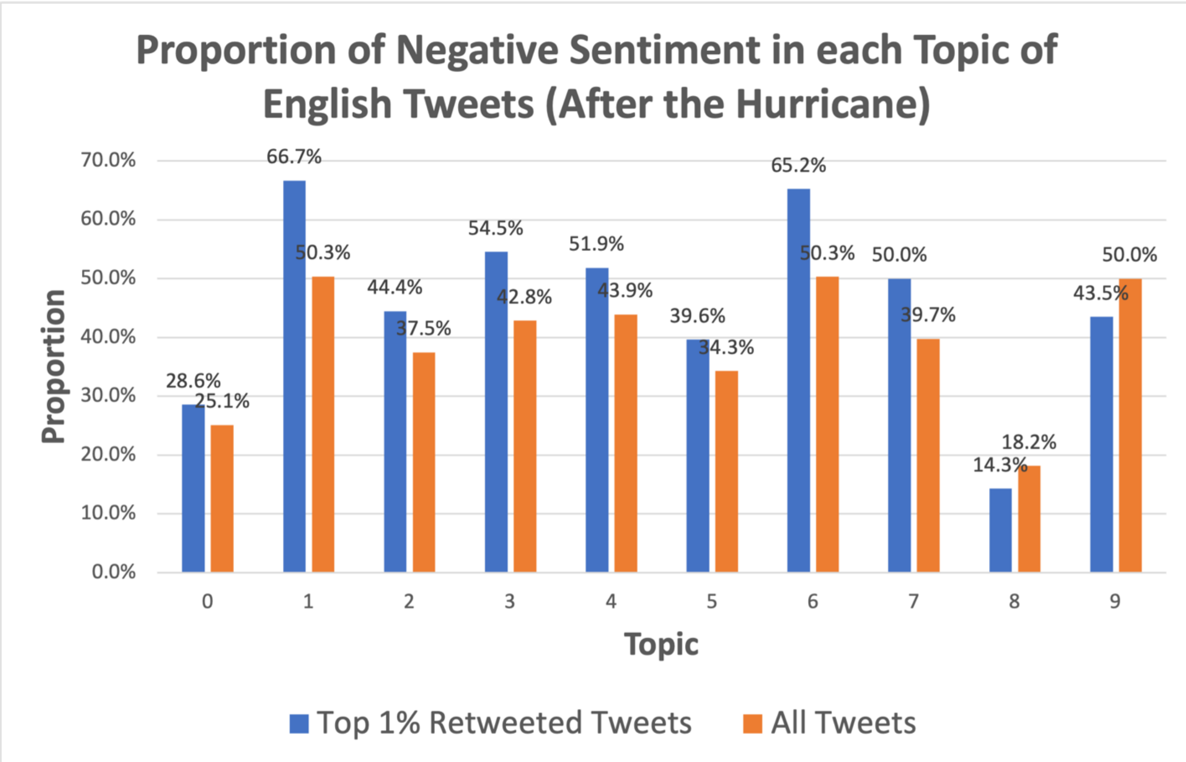


Figure 5.28. The proportion of Negative Sentiment Tweets in each Topic of English Tweets (After the Hurricane).

As shown in *Figure 5.28*, the proportion of negative sentiment in each topic was still high. Half of the topics in the dataset of the top 1% retweeted tweets and three of the topics in the whole dataset had more than 50% negative sentiment. For the smaller dataset, topics 1, 2, 3, 4, and 6 had the highest proportions of negative sentiment. Although the negative proportions were higher, only topic 4 contained a higher proportion (around 15%) in the number of tweets. Most of the words were related to action and situation. The content people focused on in this period is not the same as before. As the hurricane dissipated, the community needed to recover from the disaster, which would require financial and mental support. In topic 5 (25% of all tweets dropped on this topic), the sentiment was more positive, as people were talking about recovering their community, to provide or seek help and relief. In the period after the hurricane dissipated, English tweets shared the information about support and relief for the post-disaster recovery and people were more positive in this stage.

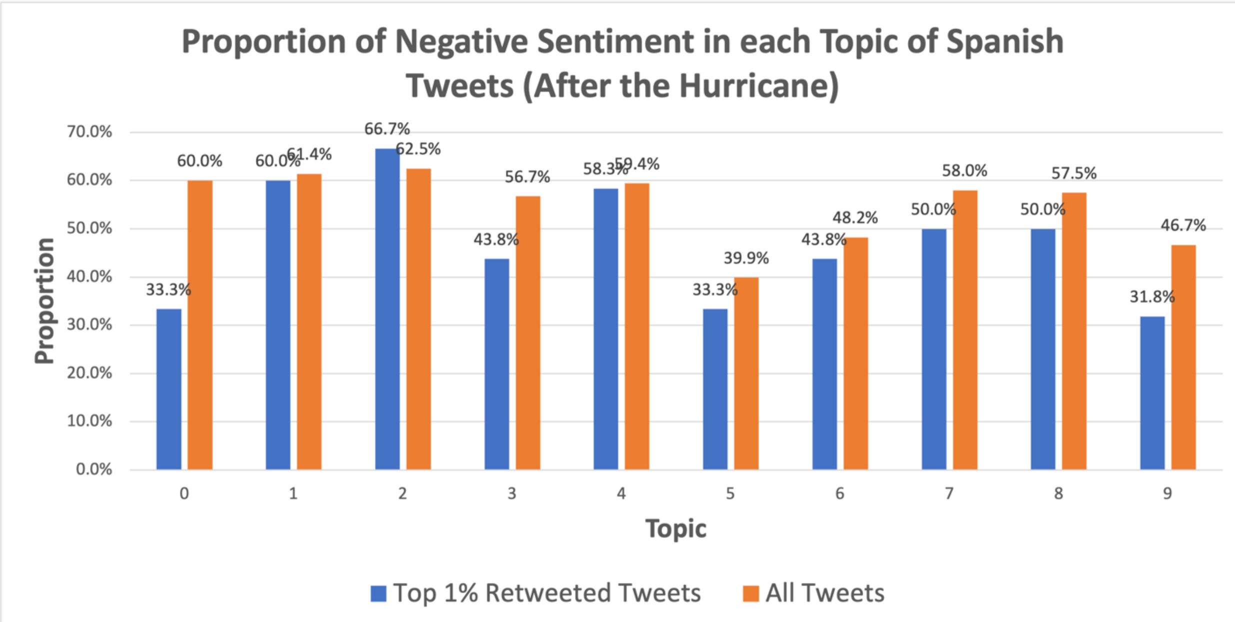


Figure 5.29. The proportion of Negative Sentiment Tweets in each Topic of Spanish Tweets (After the Hurricane).

As shown in *Figure 5.29*, the proportion of negative sentiment in each topic was lower compared to the previous period yet was still high overall. Seven of the topics in the whole dataset and five in the top 1% most retweeted tweets dataset had more than 50% negative sentiment. Action- and situation-related topics appeared most frequently. The overall sentiment for the whole dataset in this period was negative, while positive for the top 1% most retweeted tweet dataset. Topic 4 had the highest proportion (around 30%) of all topics in both datasets, and the negative sentiment was more than 50%. It might be caused by the disaster-related words in this topic. Topic 9 had the second highest proportion in both datasets, in which more tweets were related to action and positive. Same as English tweets, people talked about recovery and support in various factors, including the community, work, and donation. Topic 0 had the third highest proportion in the whole dataset, and 60% of them were detected as negative. Topic 6 had the third highest proportion in the concentrated dataset, with more than 50% of positive sentiment. In this period after the hurricane dissipated, the sentiment of Spanish tweets was different for all the tweets and the top 1% of tweets that had been retweeted.

6. Discussion

In this section, the findings of the roles that Twitter played in the event of Hurricane Ian will be interpreted. Hypothesis 1 and 2 introduced in *Section 1.4* will be discussed. The main methodology, which includes topic modeling and sentiment analysis, will be discussed in separate sections and compared with current studies. The implications and limitations of this thesis will be explained in the following sections.

6.1 Twitter Activities

Twitter activity analysis, i.e., the number of tweets, and topic modeling were applied to test this hypothesis. As expected, Twitter activities became increasingly active as the hurricane approached and returned to normal stage after the hurricane dissipated. The keywords of the topics of tweets indicated the information people mostly discussed and wanted to share. These words and the distribution of topics changed in different periods as hypothesized. This demonstrates that Twitter played varying roles as an information provider during different stages of the disaster.

6.1.1 Interpretation of Twitter Activities

The activities on social media, in this case, Twitter, from the formation date of Hurricane Ian to 50 days after it dissipated, were examined. For both English and Spanish tweets, the pattern was similar. The number of tweets increased abruptly in the "before the hurricane landed" period, the first 5 days, and peaked when the storm landed in Florida. Twitter activity decreased rapidly in the following 5 days ("during the hurricane landing" period), then gradually dropped to zero in the period "after the hurricane dissipated". This finding was similar to the studies of De Longueville et al. (2009) and Wu and Cui (2018), in which the trend of the number of tweets was relatively steady in the beginning and suddenly increased to its peak during the strike of the hurricane, then sharply dropped back to the value before the formation of Hurricane. This trend was consistent with common sense: People became concerned and wanted to share information with others during a disaster, then became less and less focused on information related to this specific event.

6.2 Topic Modeling

Hypothesis 1: During the event of Hurricane Ian, Twitter served as a platform for disseminating information related to early warnings, damage assessment, rescue scheduling, and recovery.

The LDA model from the Gensim library was applied to the English and Spanish Twitter datasets covering the periods "before the hurricane landed", "during the hurricane landing", and "after the hurricane dissipated". The topics were generated for the entire dataset of these three periods, as well as for the top 1% of retweeted tweets. Six categories were manually defined for the topics, including disaster-related topic, weather-related topic, emotion-related topic, action-related topic, situation-related topic, and spatial- or temporal-related topic. The differences in the generated topics were compared across different periods and the languages.

6.2.1 Interpretation of the Results

In the period "before the hurricane landed", the number of action-related words was the highest in English tweets, while in Spanish tweets, the highest was spatial- or temporal-related words. In both English and Spanish tweets, disaster-related and weather-related words such as "storm", "wind", and "rain" were prevalent, which indicated the awareness of people of the changing climate. Among the action-related words like, "warning", "alert", "ready", and "prepare" were commonly used in both languages, typically during the early stage of a disaster event. The generated topics revealed that people mostly tweeted about warnings regarding the approaching storms, the current situation, and to remind people to be prepared for the disaster. This finding supports the hypothesis that people used Twitter as an early warning system in the period before Hurricane Ian landed in Florida.

In the period "during the hurricane landing", the number of situation-related words was the highest in both English and Spanish tweets. Words like "damage", "disaster", and "destruction" were generated, which helped in assessing the extent of damages or risks during this event. Words like "power", "need", "water" and "electricity" were associated with the required resources during the event, which contributed to the rescue process. Action-related words, such as “help”, “support”, “assist”, and “aid” appeared in this period, indicating that people used Twitter to seek assistance. In English tweets, there were relatively few words related to location, with the only one being “west coast”, which remained somewhat vague. In Spanish tweets, more location-related words were evident including “city”, “coast”, “town” and “street”. Although precise location identification for those in need of help was challenging based on the results, this finding provided some evidence that people used Twitter to share information about resource shortages and to request aid. In this case, social media could prove valuable for government and disaster-responding agencies in the process of risk assessment and rescue coordination.

In the period “after the hurricane dissipated”, words related to action and situation for both languages featured a high volume. Words like “restore”, “relief”, and “recovery” appeared during this time, which directly indicating that people or the community had entered the recovery stage. Words like “debris”, “damage”, “house”, “roof”, and “destroy” conveyed the extend of the damages suffered by local residents after the disaster. Words like “donation”, “student”, “worker”, “municipality” and “humanitarian” indicated the community’s need for financial support, the groups most affected by the disaster, and the agencies that could help in the recovery process. Based on the results, people tweeted about how they were affected by the disaster, and how they could recover from it. This finding supports the hypothesis that Twitter played a role in the recovery process following the event of Hurricane Ian.

6.2.2 Connection with Current Studies

Words related to disaster and weather appeared more frequently in the period “before the hurricane landed”, which is consistent with previous studies on of Hurricane Sandy and forest fires in the South of France (De Longueville et al., 2009; Wu & Cui, 2018). These words could remind people of the changing weather as well as the approaching storms, essentially functioning as early warnings. Situation- and action-related words may indicate to the current extent of damage, shortages in resources, and the assistance people required. These words are usually linked with risk assessment and rescue planning for the government and related agencies, as studied in the case of Hurricane Sandy and the extension of AIDR (Kryvasheyeu et al., 2016; Nguyen et al., 2017). Different levels of the intensity of words reflected the severity of the situation. English tweets used milder words to describe the disaster, while Spanish tweets employed more intense language. Not only the choice of words but also the timing of word usage in Spanish tweets was earlier than in English tweets. One possible reason could be that the hurricane first landed in the Southern part of the North American continent around the Caribbean, where Spanish is commonly spoken in local communities, such as Cuba. In the period “after the hurricane dissipated”, the topics indicated that people were transitioning into the recovery phase. Words related to the support in financial, infrastructure and reconstruction services appeared, as discussed in a literature review on the use of social media platforms in post-disaster recovery (Ogie et al., 2022). However, in this thesis, there is no clear evidence of the used of social media to evaluate the recovery process, as analyzed in the case of Hurricane Harvey (Page-Tan, 2021).

6.3 Sentiment Analysis

Hypothesis 2: The emotion or sentiment of people who tweeted about Hurricane Ian became more negative as the storm got close and recovered to normal (or less negative) after the hurricane dissipated.

Sentiment analysis was employed in this thesis to measure the change in people’s attitude toward the event of Hurricane Ian. This hypothesis was formulated based on people’s tendencies to be worried about themselves, their family and friends when the disaster approached and recovered to normal after the disaster. Furthermore, as the event of a hurricane could cause severe damage and life losses, the emotions of people were expected to be more negative than positive throughout the entire period.

The model VADER from “vaderSentiment” was applied to English tweets, and “vader-multi” to Spanish tweets, which is essentially VADER combined with the Google Translate API. Polarity values were calculated resulting in a sentiment dictionary, including scores for “neg”, “pos”, “neu”, and “compound”. The compound score is the normalization of the sum of valence ratings, which is used to identify the sentiment of a sentence. According to the standard rule of VADER is that sentences with a compound score less than or equal to -0.05 are considered negative, those with a compound score larger than or equal to 0.05 are considered positive, and the rest are considered neutral. In this thesis, the tweets with a compound score of 0.0 were excluded to avoid the ambiguity of being completely neutral, and no sentiment score was assigned to such tweets. The models were applied to English and Spanish datasets including the periods “before the hurricane landed”, “during the hurricane landing” and “after the hurricane dissipated”. The number of tweets that were classified as “Negative”, “Positive” and “Neutral” were mapped with the time on a daily basis. The proportions of different sentiments of each period and the whole period of the hurricane event were compared for both English and Spanish tweets. Additionally, the proportion of negativity of each topic in every period of both languages was compared. Not only the difference between the two languages was examined, but also the difference between the whole dataset (all tweets) and the tweets that received the most shares (the top 1% of most retweeted tweets).

6.3.1 Interpretation of the Results

In the case of all English tweets, the trend of negative emotions among those who tweeted about Hurricane Ian continued to increase as the hurricane made landfall in Florida and even after the hurricane dissipated. However, the proportion of positive emotions consistently remained twice that of negative emotions throughout all periods. Consequently, the hypothesis failed to be substantiated since the negative sentiment continued to grow throughout the entire duration. Regarding the English tweets that were most widely shared, the overall sentiment was positive across all periods, although negative sentiment increased as the hurricane landed in Florida and decreased slightly after it dissipated. In the last two periods, the proportions of negative and positive sentiments were almost equal. In this case, the hypothesis was supported, as the negative sentiment increased as the hurricane approached and decreased after it dissipated. However, the overall sentiment remained positive, even though in the last two periods, negative sentiment accounted for almost the same proportion as positive sentiment.

In the case of all Spanish tweets, the overall sentiment was positive during the period before the hurricane, but turned negative in the following two periods. Negative emotions increased as the hurricane landed in Florida, then decreased after it dissipated, thus confirming the hypothesis.

The overall sentiment remained negative for the last two periods, with negative emotion showing an increase followed by a decrease over time. For Spanish tweets that were mostly retweeted, the overall sentiment was positive during the “before” and “after the hurricane” periods. Only during the period when the hurricane made landfall in Florida, the overall sentiment turned negative. The pattern of negative emotion aligns with the hypothesized trend, as it increased as the storm got close and decreased after its dissipation.

Combined with the topics generated in each period for English and Spanish tweets, the changes in the sentiment also revealed people’s emotions towards different topics during the hurricane event. Before the hurricane landed in Florida, disaster- and weather-related words appeared most frequently in both languages. People talked about the change in weather and the approaching storms. In this period, the overall sentiment was positive, as people were preparing to face the hurricane, and expressing their expectation of staying safe. During the hurricane landing in Florida, words related to action and situation appeared most frequently. The topics were related to the damage to the community, people’s properties, as well as resource shortage. At this stage, the storms had already struck the area causing damages. People were seeking help and available resources, which contributed to a more negative sentiment. After the hurricane dissipated, action- and situation-related words continued to appear frequently, although the content was differed from the previous period. People were talking more about support and recovery after the disaster, such as financial assistance, donations, and job opportunities. These topics led to a less negative sentiment, although the sentiment was influenced by the extent of post-disaster recovery.

6.3.2 Connection with Current Studies

The relationship between retweet behavior and a disaster has been substantiated in the research on Red River Valley flooding and Oklahoma fires in 2009 (Starbird & Palen, 2010). In comparison to non-retweets, the retweeted tweets had a higher probability of being related to a disaster. In this thesis, the tweets that had been most widely shared or retweeted (top 1%) were isolated and compared with the whole dataset. Negative sentiment increased as the hurricane approached and decreased after it dissipated, both for English and Spanish in the dataset of most retweeted tweets, as well as the dataset of all Spanish tweets. Only the entire dataset of English tweets exhibited a continual increase in negative sentiment. Therefore, based on the results of the most retweeted tweets, which were more likely to be linked to the disaster, the hypothesis was validated. VADER has been demonstrated to perform well in the sentiment analysis of social media (or blog-like) contexts in the analysis of movie reviews and global climate change (Bonta et al., 2019; Dahal et al., 2019; Hutto & Gilbert, 2014). The sentiment scores became more negative as the hurricane grew stronger or drew nearer, which is the same as discussed in the study of Vayansky et al. (2019). In their research, the relationship between the sentiment scores and the wind speed was negative, which is consistent with the findings presented in this thesis.

6.4 Implications

This thesis examined the topics and sentiments of English and Spanish tweets related to Hurricane Ian in the periods before the hurricane landed, during the hurricane's landing, and after the hurricane dissipated. The findings confirm that Twitter played the roles of an early warning system, risk assessment, rescue scheduling, and post-disaster recovery. Social media can serve as a tool to assist the government and relevant agencies in the event of a natural disaster. Government bodies or agencies can use social media before a disaster to inform and encourage people to prepare for the coming event. During the disaster, social media could provide timely information and assist in assessing the risks or damages. After a disaster, the content shared on social media can offer insights into community recovery. People using various languages on social media may also react slightly differently based on the extent of damage, or the cultural differences. This thesis serves as a reminder researchers and rescue agencies that they should consider how individuals from diverse backgrounds may exhibit varying responses to disaster. Tweets that are been most widely shared among users could be a useful source when analyzing the situation of the disaster or gauging the public reactions. This finding contributes to the timely analysis of future events, not only in natural disasters but also in other fields, such as the analysis of social and political developments. In a situation when a quick decision is required to be made, the social media contexts that have been most widely shared among people could provide important information.

6.5 Limitations

6.5.1 Data Collection: Twitter API

Twitter API is the official and legal tool for accessing, retrieving, and analyzing tweets. It is provided by Twitter, the company, and is suitable for use in both academic research and enterprise. Access to academic researchers was free before; however, Twitter changed the policy in 2023. The free version of Twitter API only provides access to create or upload tweets and log in to the account. The higher-tier access, Twitter API Pro or Twitter API Premium which support pulling large amounts of tweets at high frequency is expensive.

Other than the availability of Twitter API, the quality and retrievable data from tweets is also debatable. Theoretically, all information related to tweets could be retrieved using API. However, in practice, the access to the number and frequency of tweets that can be retrieved are limited, and the researcher would not know whether they have retrieved all the desired information or tweets.

6.5.2 Limited Geolocation Information

Geoinformation in tweets is typically consists of the geotags that Twitter users manually select when posting a tweet. However, the number of geotagged tweets is very limited. In the extracted data during Hurricane Ian from September 23rd to 27th, geotagged tweets accounted for only about 7.3% of the total tweets. In the data from September 28th to October 2nd, only about 2.3% of them had geotags. In the data from October 3rd to November 22nd, roughly 7.2% of the tweets contained geotags. There is an option to use a bounding box or limit the region/location of the tweets when extracting Twitter data. By using the geolocation information, certain tweets within the bounding box or location would be extracted accordingly. However, the option is only applicable for geotagged tweets, which make up less than 10% of the total tweets. In such cases, geotags are not an ideal source for determing the geolocation of a tweet.

The location of a user in their profile is also a source of geoinformation, which is provided by the user when they create the account and can be changed. However, not all users provide their location or provide accurate information. Consequently, the location in a user's profile is not a reliable source for pinpointing a tweet's location. Besides, Twitter uses its special code of place ID, which is different from the normal coordination of a place. There is no list available for converting the place ID to place names or coordinates.

Conclusion

This thesis aimed to investigate the roles of social media (i.e., Twitter) in the event of Hurricane Ian throughout the entire period, which includes the period before the hurricane landed in Florida, during the hurricane's impact on Florida, and up to 50 days after the hurricane dissipated (the recovery stage). Twitter is a real-time microblogging platform where people can share information and interact with others about specific events or topics. English and Spanish tweets related to Hurricane Ian were collected and separated into two datasets: one included all tweets, and the other included the top 1% of most retweeted tweets. Topic modeling and sentiment analysis were applied to both datasets. The LDA model and VADER model were used in this thesis. The differences in the contexts people discussed (different topics) were investigated, as well as the changes in their emotions throughout all periods.

Main Findings

- In the period before the hurricane landed in Florida, the topics of tweets were mostly relevant to weather and disaster. People discussed the changes in weather and the approaching storms. Twitter could function as an early warning system during this stage.

- In the period when the hurricane struck Florida, the topics of tweets were mostly relevant to the actions and situation of individuals and the community. People discussed their current situations, the damages to their properties and the community, and the resources they needed. Twitter could serve as a tool for risk assessment during this stage, while the information provided might be limited for timely rescue scheduling.

- In the period after the hurricane dissipated, the topics of tweets were mostly relevant to the action and situation related to the post-recovery progress. People talked about the support they needed or were receiving, such as donations and job opportunities. Words like "reopen" also indicated that the community was in the recovery stage. The information on Twitter could suggest that people were in the process of recovering from the disaster; however, the extent of recovery remained unclear.

- The overall sentiment of English tweets was positive. In the dataset of all tweets, negative sentiment continued to increase for all periods. In the dataset of the tweets that had been most widely shared, negative sentiment increased when the hurricane landed in Florida and slightly decreased after it dissipated.

- The overall sentiment of Spanish tweets was positive in the period before the hurricane landed and changed to negative in the following periods. For both datasets, negative sentiment increased when the hurricane landed in Florida and decreased after it dissipated.

This thesis has addressed the gap in understanding the roles of social media throughout the entire duration of a natural disaster event. It has been demonstrated that social media can serve as both an early warning system and a tool for assessing risks during the event and in post-disaster recovery. Two datasets were compared: One including all tweets and one containing filtered tweets that had been most widely shared. The findings confirmed that the retweets were very informative and could be used when requiring quick decisions. Furthermore, the sentiment changes of English and Spanish tweets were compared. The results showed that people with different cultural backgrounds reacted differently in the same event, including the intensity of words used and the proportion of negative emotion.

Future Work

First, it has been demonstrated that social media can serve as an early warning system. Future work could focus on how the government or relevant agencies use social media to inform people about the potential danger and educate them to take timely precautions. Second, the information of Twitter data is not so helpful in the scheduling of rescue, because the location information is limited. Future research could investigate the source of useful geolocation data or the methods of converting existing geo-information to mappable coordinates. Third, future research could investigate the reaction of people using different languages in an event. In a disaster, people may react differently based on their cultural backgrounds, which is also an interesting point to discuss.

References

Atlantic Oceanographic & Meteorological Laboratory. (2021, June 1). Hurricane FAQ. *NOAA's Atlantic Oceanographic and Meteorological Laboratory: Hurricanes FAQ*. https://www.aoml.noaa.gov/hrd-faq/

Barde, B. V., & Bainwad, A. M. (2017). An overview of topic modeling methods and tools. *2017 International Conference on Intelligent Computing and Control Systems (ICICCS)*, 745–750. https://doi.org/10.1109/ICCONS.2017.8250563

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet Allocation. *Journal of Machine Learning Research*, *3*.

Bonta, V., Kumaresh, N., & Naulegari, J. (2019). A Comprehensive Study on Lexicon Based Approaches for Sentiment Analysis. *Asian Journal of Computer Science and Technology*, *8*, 1–6. https://doi.org/10.51983/ajcst-2019.8.S2.2037

Bucci, L., Alaka, L., Hagen, A., Delgado, S., & Beven, J. (2023). *National Hurricane Center Tropical Cyclone Report: Hurricane Ian* (AL092022; p. 72). National Hurricane Center. https://www.nhc.noaa.gov/data/tcr/AL092022_Ian.pdf

Bui, Li. (2019). Social Media, Rumors, and Hurricane Warning Systems in Puerto Rico. *Hawaii International Conference on System Sciences 2019 (HICSS-52)*, *4*, 2667–2676. https://aisel.aisnet.org/hicss-52/dsm/social_is/4

Chang, S. E. (2010). Urban disaster recovery: A measurement framework and its application to the 1995 Kobe earthquake. *Disasters*, *34*(2), 303–327. https://doi.org/10.1111/j.1467-7717.2009.01130.x

Cloudflare Inc. (2023). *What is a social media bot? | Social media bot definition*. https://www.cloudflare.com/learning/bots/what-is-a-social-media-bot/

Court, C. D., Qiao, X., Saha, B. B., He, F., & McDaid, K. (2022). *Preliminary assessment of agricultural losses resulting from Hurricane Ian*. University of Florida-Institute of Food and Agricultural Sciences. https://fred.ifas.ufl.edu/extension/economic-impact-analysis-program/disaster-impact-analysis/hurricane-ian-damage-assessments-2022/

CR, A. (2020, July 26). Topic Modeling using Gensim-LDA in Python. *Analytics Vidhya*. https://medium.com/analytics-vidhya/topic-modeling-using-gensim-lda-in-python-48eaa2344920

Dahal, B., Kumar, S. A. P., & Li, Z. (2019). Topic modeling and sentiment analysis of global climate change tweets. *Social Network Analysis and Mining*, *9*(1), 24. https://doi.org/10.1007/s13278-019-0568-8

Data USA. (2020). *Florida for Florida*. https://datausa.io/profile/geo/florida/demographics/languages

De Longueville, B., Smith, R. S., & Luraschi, G. (2009). "OMG, from here, I can see the flames!": A use case of mining location based social networks to acquire spatio-temporal data on forest fires. *Proceedings of the 2009 International Workshop on Location Based Social Networks*, 73–80. https://doi.org/10.1145/1629890.1629907

Dhuriya, A. (2021, February 1). What is Topic Modeling? *Analytics Vidhya*. https://medium.com/analytics-vidhya/what-is-topic-modeling-161a76143cae

Dixon, S. (2022, November 22). *Countries with most Twitter users 2022*. https://www.statista.com/statistics/242606/number-of-active-twitter-users-in-selected-countries/

FEMA. (n.d.). *National Risk Index: Natural Hazards*. Official Website of the Department of Homeland Secuirity. Retrieved August 2, 2023, from https://hazards.fema.gov/nri/natural-hazards

Feng, Y., Huang, X., & Sester, M. (2022). Extraction and analysis of natural disaster-related VGI from social media: Review, opportunities and challenges. *International Journal of Geographical Information Science*, *36*(7), 1275–1316. https://doi.org/10.1080/13658816.2022.2048835

Fu, X., Li, C., & Zhai, W. (2022). Using Natural Language Processing to Read Plans. *Journal of the American Planning Association*, *0*(0), 1–12. https://doi.org/10.1080/01944363.2022.2038659

Google. (2022, January 5). *Colab.google*. Colab.Google. http://0.0.0.0:8080/

Guan, X., & Chen, C. (2014). Using social media data to understand and assess disasters. *Natural Hazards*, *74*(2), 837–850. https://doi.org/10.1007/s11069-014-1217-1

Hutto, C., & Gilbert, E. (2014). VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text. *Proceedings of the International AAAI Conference on Web and Social Media*, *8*(1), Article 1. https://doi.org/10.1609/icwsm.v8i1.14550

Hutto, C. J. (n.d.). *vader-multi: VADER Sentiment Analysis. VADER (Valence Aware Dictionary and sEntiment Reasoner) is a lexicon and rule-based sentiment analysis tool that is specifically attuned to sentiments expressed in social media, and works well on texts from other domains.* (3.2.2.1) [Python]. Retrieved September 4, 2023, from https://github.com/brunneis/vader-multi

Imran, M., Castillo, C., Lucas, J., Meier, P., & Vieweg, S. (2014). AIDR: Artificial intelligence for disaster response. *Proceedings of the 23rd International Conference on World Wide Web*, 159–162. https://doi.org/10.1145/2567948.2577034

Indiana University. (2023). *Observatory on Social Media*. Indiana University Observatory on Social Media. https://osome.iu.edu/

International Federation of Red Cross and Red Crescent Societies. (2020, September 19). *Community Early Warning Systems: Guiding Principles*. https://www.ifrc.org/document/community-early-warning-systems-guiding-principles

Iqbal, M. (2022, November 4). *Twitter Revenue and Usage Statistics (2022)*. Business of Apps. https://www.businessofapps.com/data/twitter-statistics/

Jamali, M., Nejat, A., Ghosh, S., Jin, F., & Cao, G. (2019). Social media data and post-disaster recovery. *International Journal of Information Management*, *44*, 25–37. https://doi.org/10.1016/j.ijinfomgt.2018.09.005

Jayasekara, R. U., Jayathilaka, G. S., Siriwardana, C., Amaratunga, D., Haigh, R., Bandara, C., & Dissanayake, R. (2021). Identifying gaps in early warning mechanisms and evacuation procedures for tsunamis in Sri Lanka, with a special focus on the use of social media. *International Journal of Disaster Resilience in the Built Environment*, *ahead-of-print*(ahead-of-print). https://doi.org/10.1108/IJDRBE-02-2021-0012

Jelodar, H., Wang, Y., Yuan, C., Feng, X., Jiang, X., Li, Y., & Zhao, L. (2019). Latent Dirichlet allocation (LDA) and topic modeling: Models, applications, a survey. *Multimedia Tools and Applications*, *78*(11), 15169–15211. https://doi.org/10.1007/s11042-018-6894-4

Ju, Y., Adams, B., Janowicz, K., Hu, Y., Yan, B., & McKenzie, G. (2016). Things and Strings: Improving Place Name Disambiguation from Short Texts by Combining Entity Co-

Occurrence with Topic Modeling. In E. Blomqvist, P. Ciancarini, F. Poggi, & F. Vitali (Eds.), *Knowledge Engineering and Knowledge Management* (pp. 353–367). Springer International Publishing. https://doi.org/10.1007/978-3-319-49004-5_23

Kabir, M. Y., & Madria, S. (2019). *A Deep Learning Approach for Tweet Classification and Rescue Scheduling for Effective Disaster Management*. 269–278. https://doi.org/10.1145/3347146.3359097

Khan, F. H., Bashir, S., & Qamar, U. (2014). TOM: Twitter opinion mining framework using hybrid classification scheme. *Decision Support Systems*, *57*, 245–257. https://doi.org/10.1016/j.dss.2013.09.004

Kherwa, P., & Bansal, P. (2019). Topic Modeling: A Comprehensive Review. *EAI Endorsed Transactions on Scalable Information Systems*, *7*(24). https://eudl.eu/doi/10.4108/eai.13-7-2018.159623

Kitazawa, K., & Hale, S. A. (2021). Social media and early warning systems for natural disasters: A case study of Typhoon Etau in Japan. *International Journal of Disaster Risk Reduction*, *52*, 101926. https://doi.org/10.1016/j.ijdrr.2020.101926

Kryvasheyeu, Y., Chen, H., Obradovich, N., Moro, E., Van Hentenryck, P., Fowler, J., & Cebrian, M. (2016). Rapid assessment of disaster damage using social media activity. *Science Advances*, *2*(3), e1500779. https://doi.org/10.1126/sciadv.1500779

lda developers. (2020). *lda: Topic modeling with latent Dirichlet allocation* (2.0.0) [C, Cython, Python; MacOS, Microsoft :: Windows, POSIX, Unix].

Li, H., Caragea, D., & Caragea, C. (2021). *Combining Self-training with Deep Learning for Disaster Tweet Classification*. 719–730.

Loria, S. (2020). *Textblob*. Tutorial: Quickstart — TextBlob 0.16.0 Documentation. https://textblob.readthedocs.io/en/latest/quickstart.html#quickstart

Mattingly, W. J. B. (2022). Introduction to Python for Humanists. In *1.3. Bigrams and Trigrams—Introduction to Python for Humanists*. CRC Press. python-textbook.pythonhumanities.com

Nguyen, D. T., Ofli, F., Imran, M., & Mitra, P. (2017). Damage Assessment from Social Media Imagery Data During Disasters. *2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, 569–576.

NOAA. (2021, June 14). *What is a hurricane?* National Ocean Service. https://oceanservice.noaa.gov/facts/hurricane.html

NOAA. (2023). *Hurricane Preparedness—Hazards* [National Hurricane Center and Central Pacific Hurricane Center]. https://www.nhc.noaa.gov/prepare/hazards.php

Ogie, R. I., James, S., Moore, A., Dilworth, T., Amirghasemi, M., & Whittaker, J. (2022). Social media use in disaster recovery: A systematic literature review. *International Journal of Disaster Risk Reduction*, *70*. https://doi.org/10.1016/j.ijdrr.2022.102783

OSoMe Project. (2014, May 1). *Botometer FAQ*. https://botometer.osome.iu.edu/faq

OSoMe Project. (2023, June 30). *Botometer by OSoMe*. https://botometer.iuni.iu.edu

Page-Tan, C. (2021). The Role of Social Media in Disaster Recovery Following Hurricane Harvey. *Journal of Homeland Security and Emergency Management*, *18*(no.1), 93–123. https://doi.org/10.1515/jhsem-2018-0054

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, *12*, 2825–2830.

Řehůřek, R. (2022, December 21). *Gensim: Topic modelling for humans*. https://radimrehurek.com/gensim/models/ldamodel.html#

RMS. (2022, October 7). *RMS Estimates US$67 Billion in Insured Losses from Hurricane Ian | RMS*. RMS-a Moody's Analytics Company. https://www.rms.com/newsroom/press-releases/press-detail/2022-10-07/rms-estimates-us67-billion-in-insured-losses-from-hurricane-ian

Roy, K. C., Hasan, S., & Mozumder, P. (2020). A multilabel classification approach to identify hurricane-induced infrastructure disruptions using social media data. *Computer-Aided Civil and Infrastructure Engineering*, *35*(12), 1387–1402.

Schumaker, R. P., Jarmoszko, A. T., & Labedz, C. S. (2016). Predicting wins and spread in the Premier League using a sentiment analysis of twitter. *Decision Support Systems*, *88*, 76–84. https://doi.org/10.1016/j.dss.2016.05.010

Starbird, K., & Palen, L. (2010). Pass It On?: Retweeting in Mass Emergency. *Pass It On?: Retweeting in Mass Emergency*. Conference on Information Systems for Crisis Response and Management (ISCRAM 2010), Seattle, WA.

Syed, M. A., Shrestha, M. S., & Khadgi, V. (2021). Chapter 33—Last mile communication of multihazard early warning—A case study on Bangladesh. In I. Pal, R. Shaw, R. Djalante, & S. Shrestha (Eds.), *Disaster Resilience and Sustainability* (pp. 725–765). Elsevier. https://doi.org/10.1016/B978-0-323-85195-4.00002-0

Twitter Help Center. (2022). *How to Tweet – what is a Tweet, keyboard shortcuts, and sources*. https://help.twitter.com/en/using-twitter/how-to-tweet

Twitter Help Center. (2023a). *How to Retweet*. https://help.twitter.com/en/using-twitter/how-to-retweet

Twitter Help Center. (2023b). *Tweet location FAQs | Twitter Help*. https://help.twitter.com/en/safety-and-security/tweet-location-settings

Twitter Inc. (2022a). *Twitter Announces Second Quarter 2022 Results* [Financial Quarterly Result].

Twitter Inc. (2022b). *Twitter API Documentation*. https://developer.twitter.com/en/docs/twitter-api

UNDRR. (2007). *Early warning system*. Uniter Nations Office for Disaster Risk Reduction. https://www.undrr.org/terminology/early-warning-system

UNISDR. (2009). *Terminology on Disastere Risk Reduction*. 35. https://www.unisdr.org/files/7817_UNISDRTerminologyEnglish.pdf

Valentijn, T., Margutti, J., van den Homberg, M., & Laaksonen, J. (2020). Multi-Hazard and Spatial Transferability of a CNN for Automated Building Damage Assessment. *Remote Sensing*, *12*(17), Article 17. https://doi.org/10.3390/rs12172839

Van Aalst, M. K. (2006). The impacts of climate change on the risk of natural disasters. *Disasters*, *30*(1), 5–18. https://doi.org/10.1111/j.1467-9523.2006.00303.x

Vayansky, I., & Kumar, S. A. P. (2020). A review of topic modeling methods. *Information Systems*, *94*, 101582. https://doi.org/10.1016/j.is.2020.101582

Vayansky, I., Kumar, S. A. P., & Li, Z. (2019). An Evaluation of Geotagged Twitter Data during Hurricane Irma Using Sentiment Analysis and Topic Modeling for Disaster Resilience. *2019 IEEE International Symposium on Technology and Society (ISTAS)*, 1–6. https://doi.org/10.1109/ISTAS48451.2019.8937859

Velayutham, T., & Tiwari, P. K. (2017). Bot identification: Helping analysts for right data in twitter. *2017 3rd International Conference on Advances in Computing,Communication & Automation (ICACCA) (Fall)*, 1–5. https://doi.org/10.1109/ICACCAF.2017.8344722

Wallemacq, P., & House, R. (2018). *Economic losses, poverty & disasters: 1998-2017* (p. 31). Centre for Research on the Epidemiology of Disaster, United Nations Office for Disaster Risk Reduction. https://www.undrr.org/publication/economic-losses-poverty-disasters-1998-2017

Wu, D., & Cui, Y. (2018). Disaster early warning and damage assessment analysis using social media data and geo-location information. *Decision Support Systems*, *111*, 48–59. https://doi.org/10.1016/j.dss.2018.04.005

Yang, K.-C., Ferrara, E., & Menczer, F. (2022). Botometer 101: Social bot practicum for computational social scientists. *Journal of Computational Social Science*, *5*(2), 1511–1528. https://doi.org/10.1007/s42001-022-00177-5

Zahra, K., Imran, M., & Ostermann, F. O. (2018). Understanding eyewitness reports on Twitter during disasters. In K. Boersma & B. Tomaszewski (Eds.), *Zahra, Kiran; Imran, Muhammad; Ostermann, Frank O (2018). Understanding eyewitness reports on Twitter during disasters. In: 15th International Conference on Information Systems for Crisis Response and Management ISCRAM 2018, Rochester (USA), 21 May 2018—23 May 2018. ISCRAM, 687-695.* (pp. 687–695). ISCRAM. https://doi.org/10.5167/uzh-161922

Zahra, K., Ostermann, F. O., & Purves, R. S. (2017). Geographic variability of Twitter usage characteristics during disaster events. *Geo-Spatial Information Science*, *20*(3), Article 3. https://doi.org/10.1080/10095020.2017.1371903

Appendix

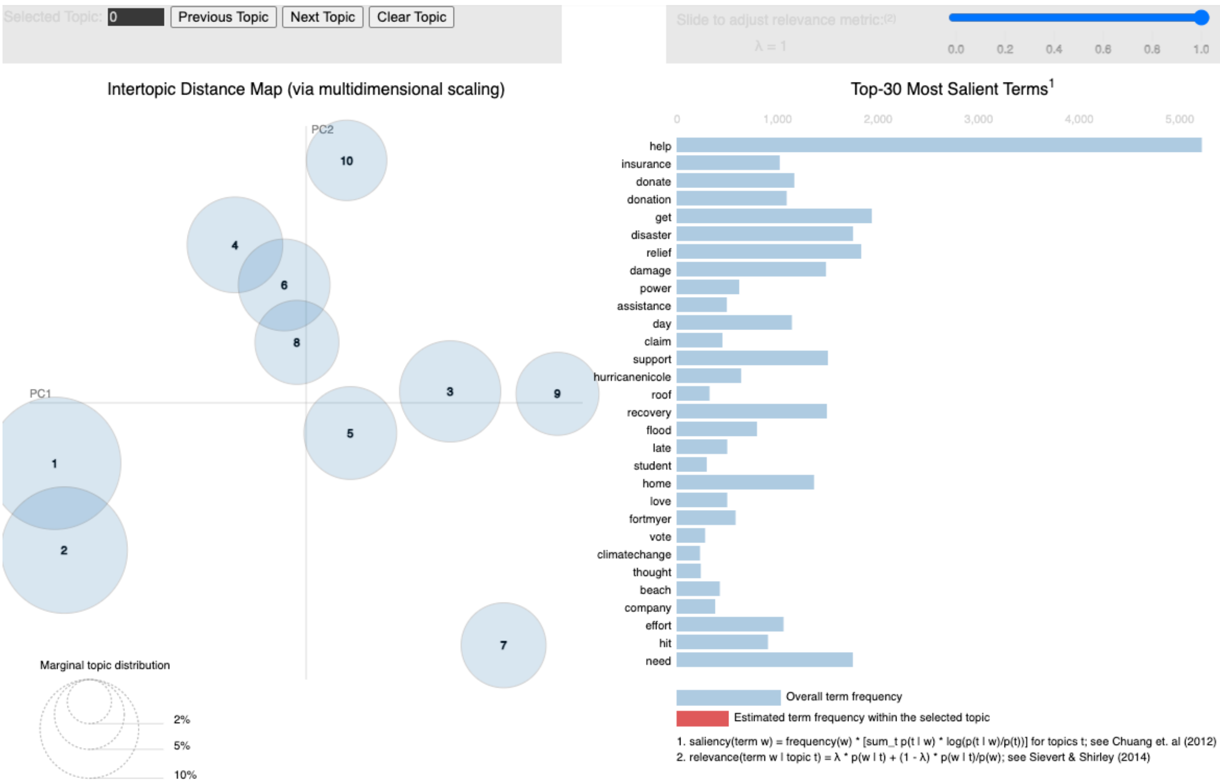


Figure A. 1. Example of Visualization of Topic Modeling by pyLDAvis.

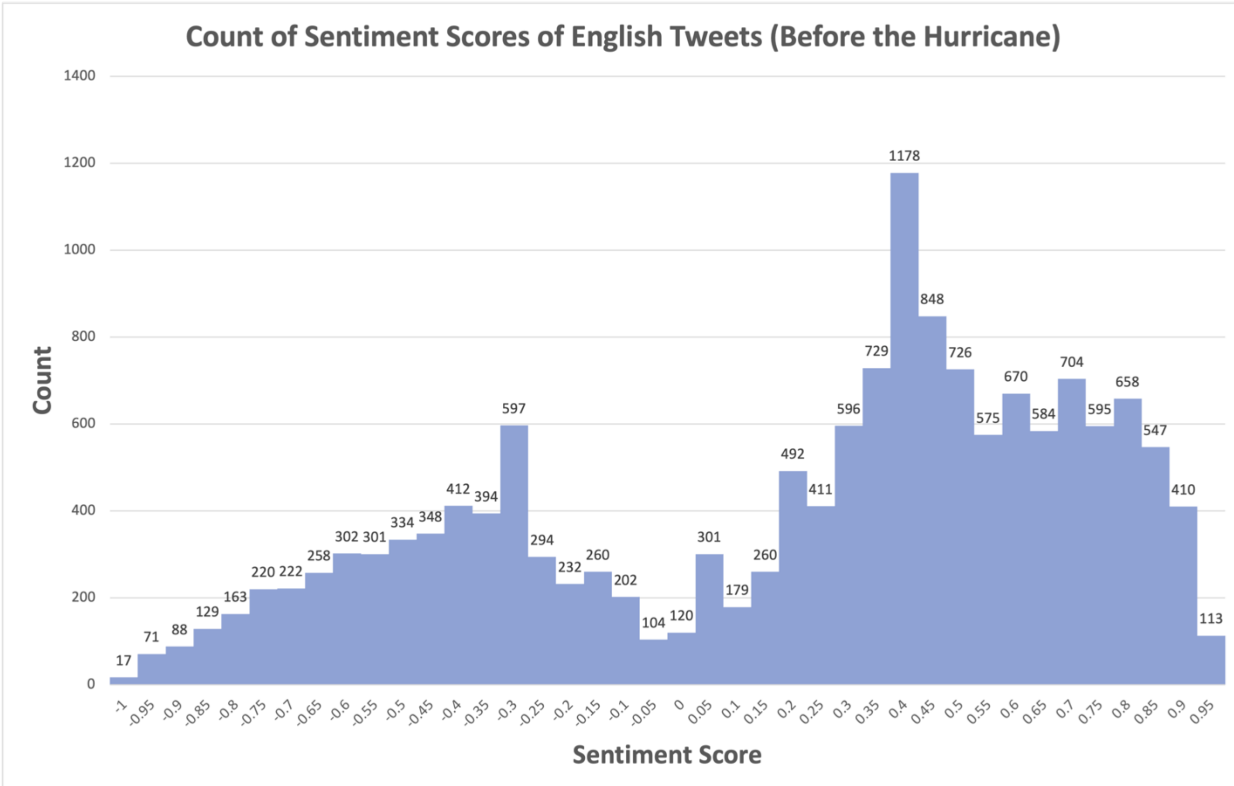


Figure A. 2. Count of Sentiment Scores of English Tweets (Before the Hurricane).

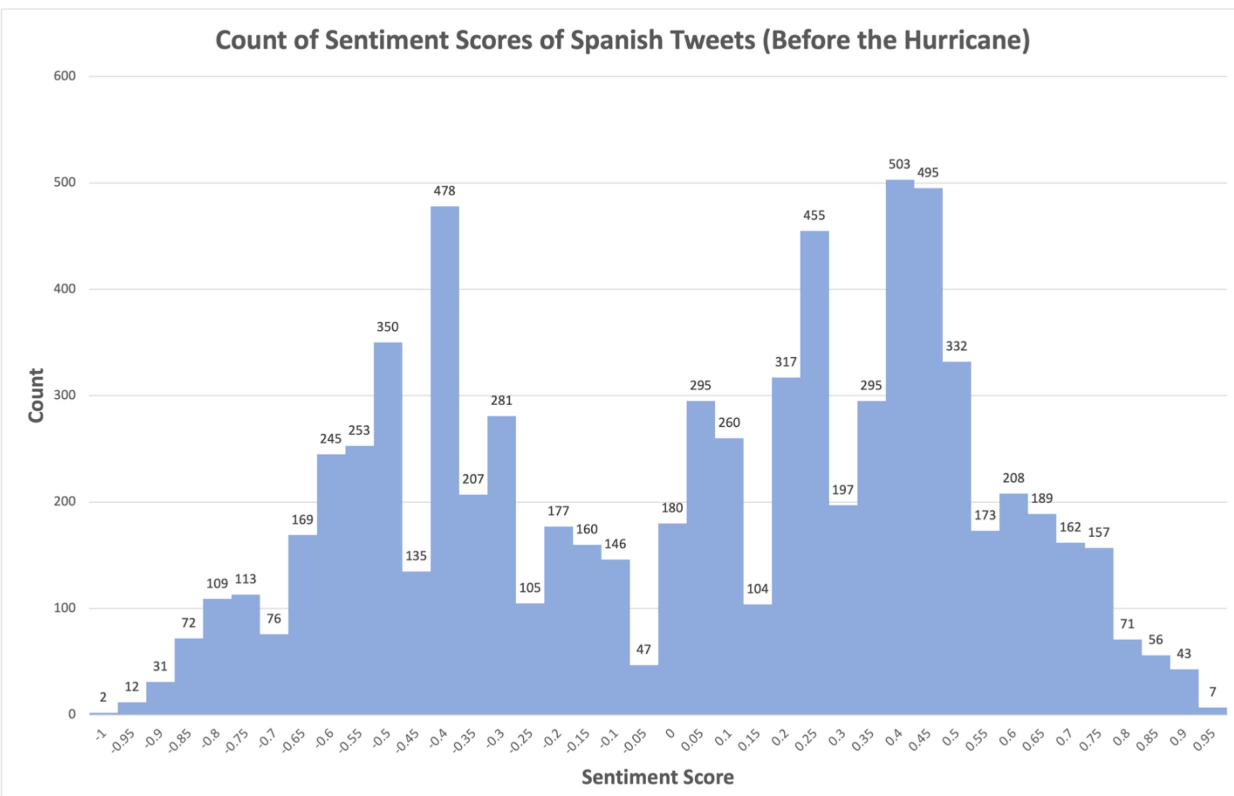


Figure A. 3. Count of Sentiment Scores of Spanish Tweets (Before the Hurricane).

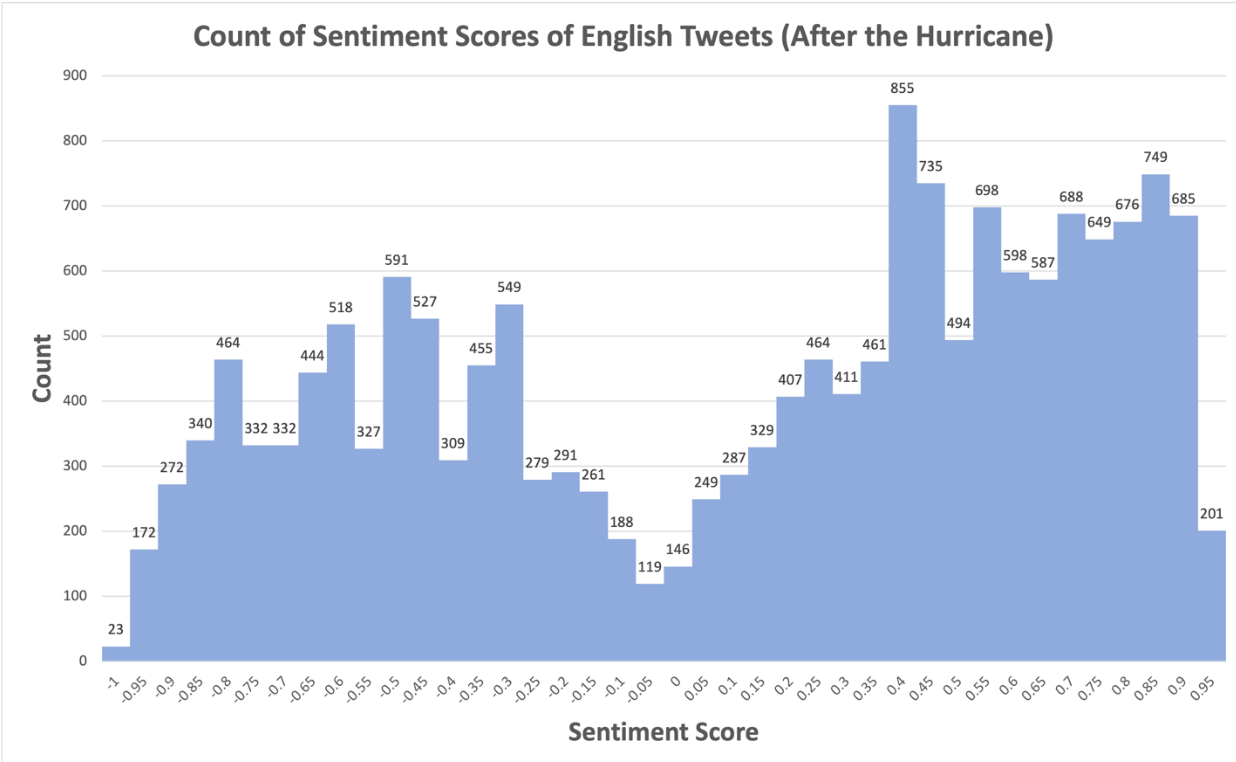


Figure A. 4. Count of Sentiment Scores of English Tweets (After the Hurricane).

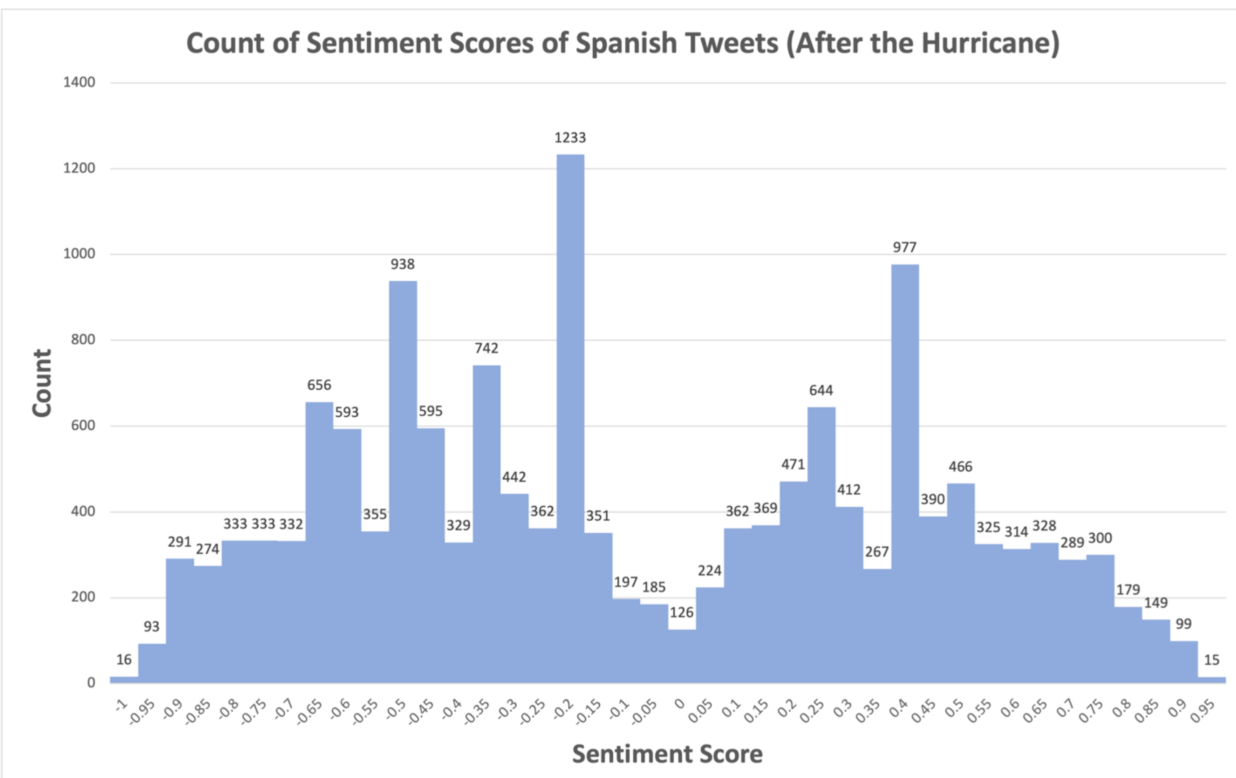


Figure A. 5. Count of Sentiment Scores of Spanish Tweets (After the Hurricane).

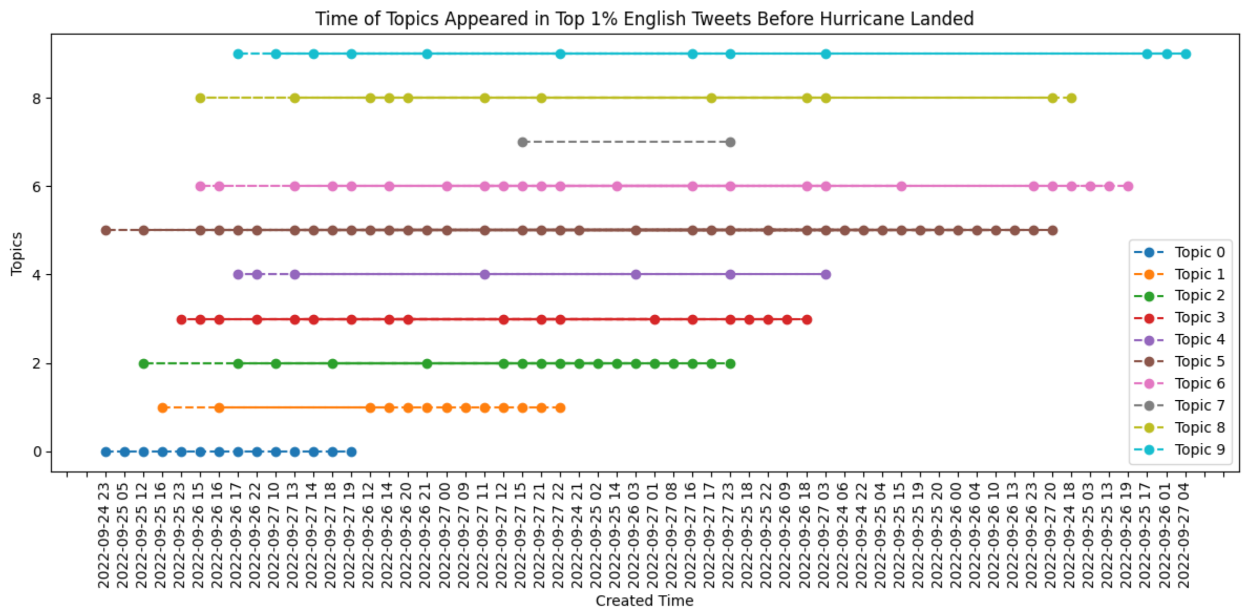


Figure A. 6. Time of Topics Appeared in the Dataset of the Top 1% most Retweeted English Tweet Before the Hurricane Landed.

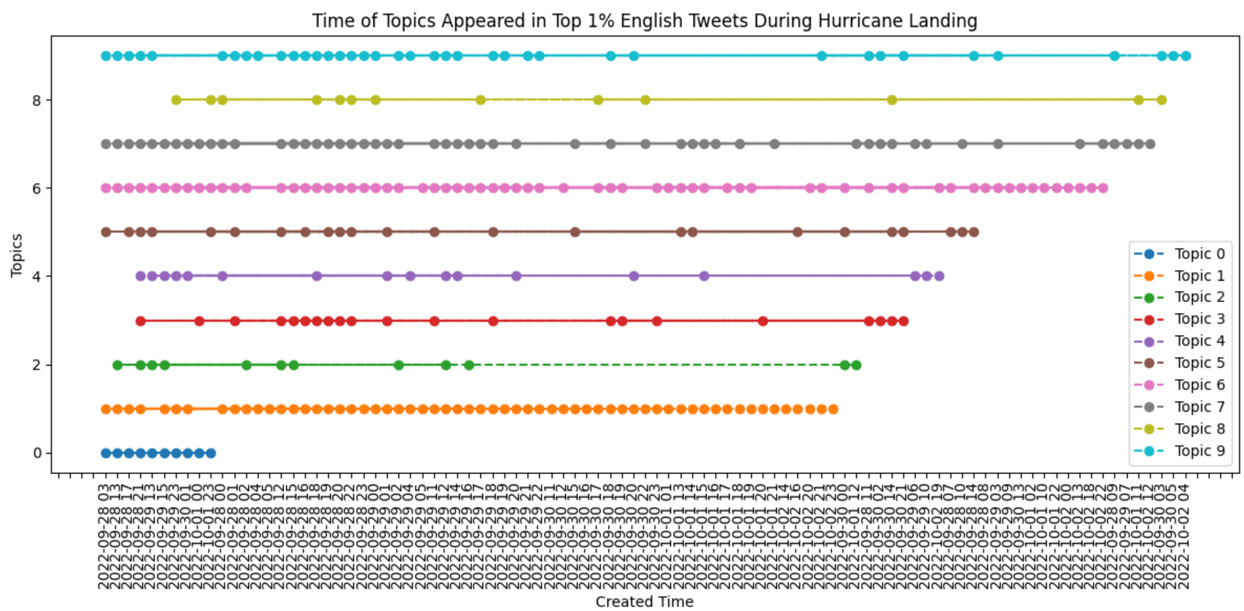


Figure A. 7. Time of Topics Appeared in the Dataset of the Top 1% most Retweeted English Tweet During the Hurricane Landed.

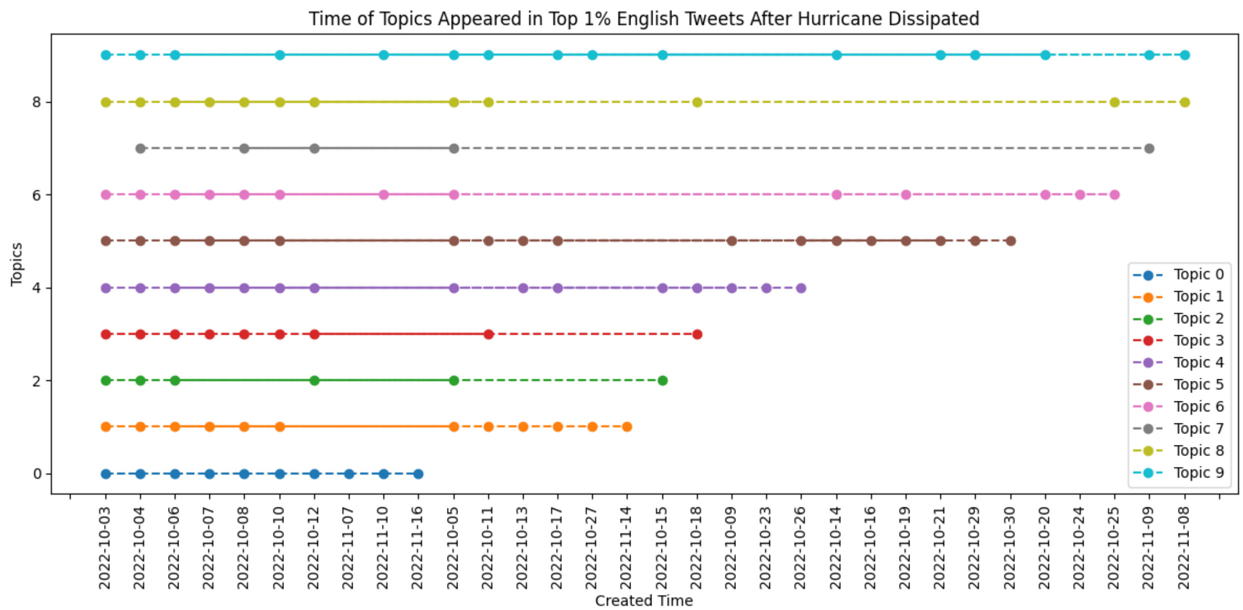


Figure A. 8. Time of Topics Appeared in the Dataset of the Top 1% most Retweeted English Tweet After the Hurricane Dissipated.

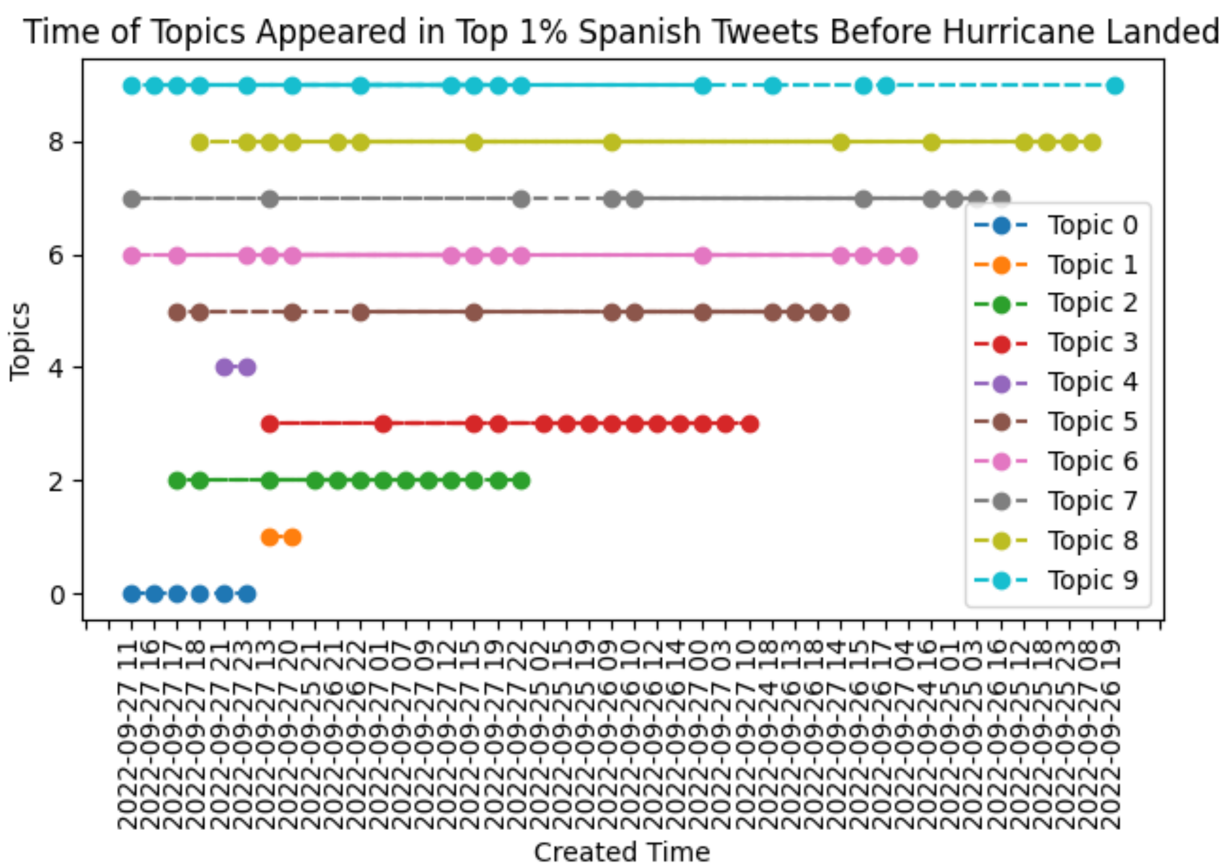


Figure A. 9. Time of Topics Appeared in the Dataset of the Top 1% most Retweeted Spanish Tweet Before the Hurricane Landed.

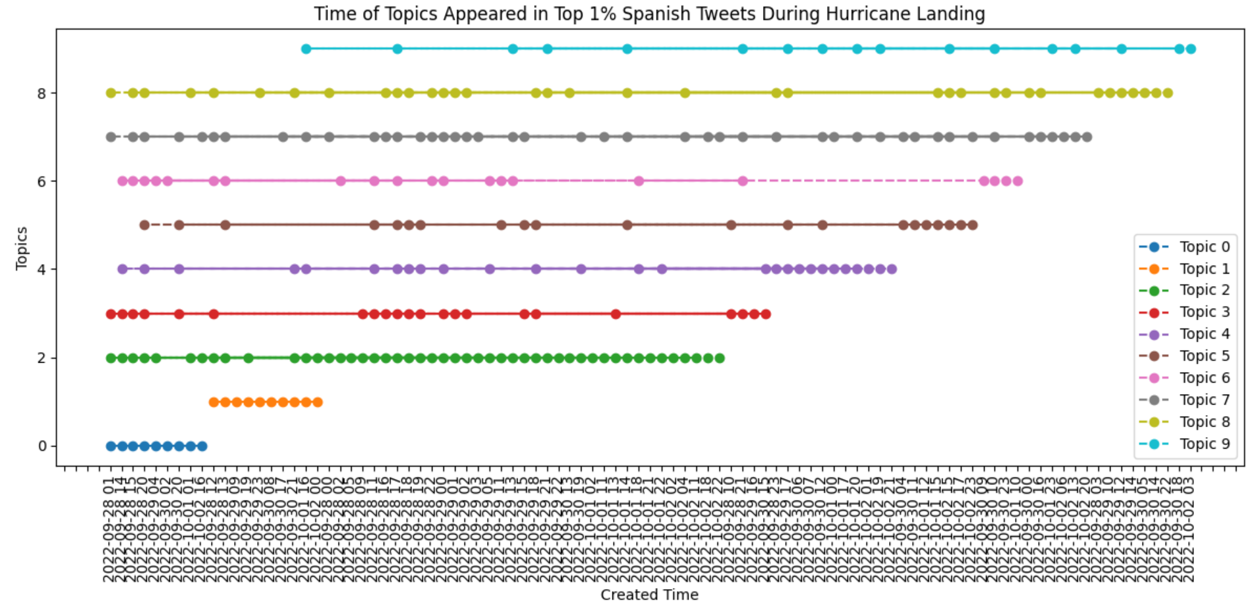


Figure A. 10. Time of Topics Appeared in the Dataset of the Top 1% most Retweeted Spanish Tweet During the Hurricane Landed.

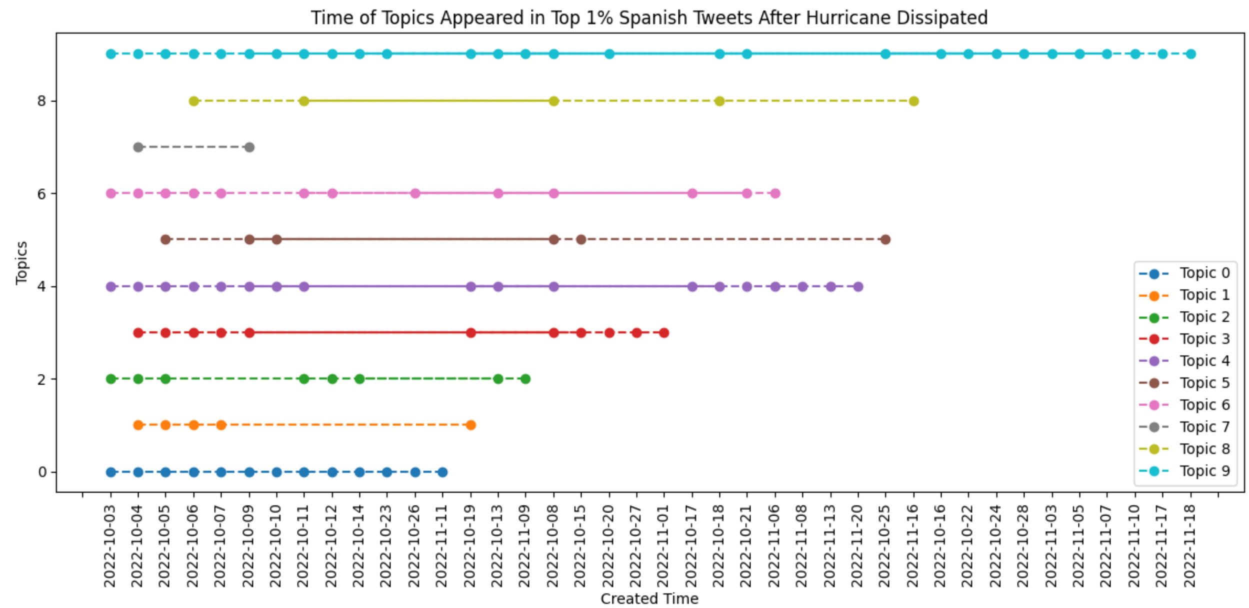


Figure A. 11. Time of Topics Appeared in the Dataset of the Top 1% most Retweeted Spanish Tweet After the Hurricane Dissipated.

Personal Declaration

I hereby declare that the submitted thesis is the result of my own, independent work. All external sources are explicitly acknowledged in this thesis.

Mingyang Yuan, September 2023