School of Computer Science

University of St Andrews

Automatic Evaluation of Geopolitical Risk

John Corcoran Burns

ID: 180013598

Thesis

PhD Programme

Supervisor: Dr. Tom Kelsey (School of Computer Science)

Co-Supervisor: Dr. Carl Donovan (School of Mathematics and Statistics)

Date: May 9th, 2024

Declaration:

I, John Corcoran Burns, hereby certify that this thesis, which is approximately 74,000 words in length, has been composed by me, that it is the record of work carried out by me and that it has not been submitted in any previous application for a degree. This project was conducted by me through University of St Andrews from December 2020 to November 2023 towards fulfillment of the requirements of the University of St Andrews for the degree of PhD in Computer Science under the supervision of Dr. Tom Kelsey and Dr. Carl Donovan.

John Corcoran Burns

Acknowledgments:

I would like to thank Dr. Tom Kelsey and Dr. Carl Donovan for all their support and advice throughout my PhD. They were fantastic advisors, who provided invaluable help and I will always be grateful to them. Also, I would like to thank my examiners, Dr. Rosa Filgueira, Dr. Peiran Jiao, and Dr. Alan Dearle, for their time and comments that helped greatly improve my thesis. Additionally, I would like to thank Irlanda Marquez Cruz, Yasmine Hassani, and Bruno Caneco for their Spanish, French, and Portuguese translation, and annotation work, respectively. I also thank the good people at Amazon Turk for their Arabic, Japanese, and Korean annotation work, as well as Gengo.com for their translation work. Finally, I thank all my family and friends for all their love and support during this process, I couldn't have done it without them.

Abstract:

This thesis aims to construct programs for automatically evaluating geopolitical risks. This project will use machine learning, specifically sentiment analysis, topic modeling and NER, to build new computer programs to evaluate and assess not just scheduled and predictable geopolitical events, but also unpredictable events. The gap in current literature this project intends to fill is the ability to respond to these risk events in real time. Thus, this project's objective is to build programs that can digest the vast quantities of data generated by Twitter / X, the data source chosen for this project, focusing on keywords indicating a potential geopolitical event or crisis.

With the use of Twitter / X, I was able to find that information appeared quicker through tweets than traditional news sources, thus I was able to identify emerging geopolitical topics, in some cases, hours or days before they became discussed in the mainstream media. I also achieved success with building a geopolitical risk index program with sentiment analysis to relate the index to the trends in the financial markets surrounding the start of the Ukraine War in 2022 at the daily level. Using Granger Causality, I found that the geopolitical risk index I created from the emotions gleaned from the sentiment analysis of the relevant tweets collected, contained predictive information of the movement of various financial assets over time. In addition, with the NER program, I was able to visualize the different geopolitical risks on a world map. While I managed to create a program that created a geopolitical risk index at the real time level, unfortunately, there was little relationship between the real time risk index and the change in financial markets. In combination, the output of these various programs allowed for the automatic evaluation of geopolitical risk.

Table of Contents

Chapter 1: Introduction

Geopolitical risk can be defined as "the risk associated with wars, terrorist acts, and tensions between states that affect the normal and peaceful course of international relations." (Caldara et al. 2018, Pg. 2) [105]. These risks arise from the outcome of international events that are either scheduled (such as a national election or a G7 Summit) or unscheduled (such as Ukraine War or the Taliban's takeover of Afghanistan). However, the roots of geopolitics and geopolitical risks, and how humanity responds to them lay much deeper. An example of this is from the great American historian, Barbara Tuchman, who wrote many books involving geopolitics. One of her most famous is *A Distant Mirror: The Calamitous 14th Century* [426], which centers on French nobleman Enguerrand VII de Coucy to examine the events of the 14$^{\text{th}}$ Century. Tuchman weaves massive geopolitical events through Coucy's life, such as the Hundred Years War between France and England, the Papal Schism, and the rise of the Ottoman Empire. Tuchman describes how the kingdoms of Europe navigated these troubled waters, both successfully and unsuccessfully, providing clear insight into how important geopolitics is, not only to the nobility of the time, but to everyone. This salient point is further reinforced in another groundbreaking book by Tuchman, *The Guns of August* [427]. This book focused on World War I, specifically, it provided the background to the War, how and why it started, and the battles of the first month of the War. Tuchman details how the War didn't appear from nothing, instead it was a multi-decade long failure of diplomacy which led to one of the most violent wars the world had ever seen. France's desire for revenge from the Franco-Prussian war in the 1870's, the German Empire's want to expand its border both locally in Europe and abroad, and the English King Edward VII maneuverings to encircle the German Empire to prevent this. While the War exploded in 1914 with a surprise attack from the Germans, it was clear to see the seeds of this massive geopolitical conflict were sown much earlier, leading to 40 million causalities by the war's end.

In today's world, the relentless pace of geopolitics has only increased. As Mark Twain said, "History never repeats itself, but it does often rhyme". Similar to how in the 14^{th} Century geopolitical tensions rose after the Black Death, after COVID-19 geopolitical tensions are growing as well. Russia and China's current rhetoric mirrors the German Empire's when they talk about encirclement by the United States, NATO, and their allies in East Asia. The rising geopolitical risk in the South China Sea, and the recent invasion of Ukraine by Russia could be the opening acts of a new Cold War or even another World War. Additionally, outside of these major issues, there are a few new, smaller conflicts that might arise. As Dodds describes in his book: *The New Border Wars: The Conflicts That Will Define Our Future* [161], there are plenty of other sources for potential geopolitical hotspots. With climate change felt around the world, areas previously inaccessible, such as the Himalayan glaciers, are increasing tensions between India and Pakistan, and India and China, as the three countries race to define their new borders as these glaciers receded. Building off this smaller scale, Marshall details in *Prisoners of Geography: Ten Maps That Explain Everything About the World* [298] how climate change is now opening new trade routes in the Arctic. In the near future the United States, Canada, Russia, and China will all compete for these routes further increasing geopolitical risk. Furthermore, Space is not only the final frontier for human exploration, but also geopolitics. In his follow-up book, *The Power of Geography: Ten Maps That Reveal the Future of Our World* [299], Marshall explains the geopolitics of Space and Space's growing militarization. With everything from potential Moon bases to anti-satellite weapons that could cut off communications for entire countries, geopolitics could leave the planet for the first time. While this might sound like the realm of science fiction, in fact, famous geopolitical thinker, George Friedman, describes a potential Space-based war. In his book *The Next 100 years: a forecast for the 21st century* [179] he details a potential conflict between the United States, Turkey, and Japan in space in the 2050s based on these countries current growth projections. With all these potential new conflicts emerging in the coming years, it is necessary to follow geopolitics as the world can utterly change on the actions of a few. Thus, getting information as early as possible is imperative. This is where my chosen data source for my research comes in: Twitter / X.

In our modern world, there are large amounts of data constantly generated making it virtually impossible to read and evaluate all the data at once. As Grimmer and Stewart describe: “Human-based methods for making these inferences are both time and resource intensive. Even after coding rules are developed and coders are trained, manual methods require that coders read each individual text. Automated methods can mitigate the cost of assigning documents to categories, by limiting the amount of classification humans perform” (Grimmer and Stewart, 2016, Pg. 8) [208]. Thus, developing automatic methods to evaluate geopolitics is vital for extending the research in this space into real time. Twitter / X provides that pathway to construct these methods. Founded in 2006, Twitter / X is a microblogging website that allows users to post messages, called tweets, of initially 140 characters, now up to 260 characters (Shephard, 2023) [391]. Below is an example of a tweet from US President Joe Biden on a recent missile attack launched by Iran towards Israel [85].

Figure 1: An example of a tweet, I used US President Biden's account as he is a public figure posting a public tweet which follows the ethical guidelines. Additionally, since he is the US President, he would mention geopolitics quite often and thus providing a good example of a geopolitical tweet that I would collect.

While a tweet doesn't contain an abundance of text, there is useful information contained within it. Outside of the text itself, it also gives the number of reposts (i.e. how many other people have retweeted this tweet), the number of quote tweets (the number of people who have posted this tweet, with their own comment added), the number of likes, and the number of bookmarks (the number of people who have saved this tweet), the number of views, and the username. While I will not be using these metrics in my research, they provide pertinent data for other types of research, such as network analysis and the spread of news [438]. However, for my research, outside of the text, I am most interested in the time the tweet is posted. Discussed in more detail later on in my thesis, having the time of the tweet will allow me to build timelines and indices out of the Twitter text data which forms the basis of my analyses.

What makes Twitter / X beneficial to use in my research, in addition to what has already been described, is that it has a global reach. Twitter / X has approximately 450 million monthly active users and out of that approximately 57% of those are daily active users (Shephard, 2023). Additionally, while the United States might have the highest number of active users, the next four countries (Japan, Brazil, India, and Indonesia) combined have approximate 60% more total users than the United States (Branka, 2023) [95]. This diversity in the userbase makes Twitter / X incredibly useful when trying to analyze geopolitical risk on a global scale. However, it is not just the reach that makes Twitter / X valuable, but it is the amount of data it produces which makes Twitter / X indispensable. On average, as of 2023, Twitter / X users produce more than 829 million tweets a day. This is an incredible amount of data, and of course not all of it would deal with geopolitics. Thus, Twitter / X has developed APIs such as the Filter Stream API [421], which not only allow for users to both obtain relevant tweets for their research, but also obtain these tweets in real time. Twitter also has up to date news, as Sankaranarayanan describes: "This is another reason that makes Twitter attractive for capturing breaking news, as there is very little lag between the time that an event happens or is first reported in the news media and the time at which it is the subject of a posting on Twitter" (Sankaranarayanan, 2009, Pg. 2) [377]. This multi-functionality of Twitter / X, plus its global reach, and ability to capture breaking news, made it the ideal data source for my research.

Currently, there are a few commercial enterprises that work on the evaluation of geopolitical risks similar to what I aim to do. BlackRock Geopolitical Dashboard[1] [9], Dow Jones Factiva[2] [6], and the GDELT project[3] [7], all focus on providing analysis of current geopolitical risk trends to their clients by extracting relevant geopolitical events out of the sea of news media that exists. The aim of my thesis is to take these risk analyses a step further and build programs that can react to geopolitical events in real time. The ability for governments and NGOs to understand and act on these events, both known and unknown,

[1] https://www.blackrock.com/corporate/insights/blackrock-investment-institute/interactive-charts/geopolitical-risk-dashboard

[2] https://www.dowjones.com/professional/resources/factiva-content/news-you-can-use.

[3] https://www.gdeltproject.org

instantaneously, would provide greater insight into the best way to manage the changing landscape.

While the geopolitical research area is large, there are a few potential outcomes I will investigate in this thesis:

- Evaluation of the current attempts to define and monitor geopolitical risk.
- Evaluation of data sources and NLP advances for an improved geopolitical risk monitoring tool.
- Development of topic models for the monitoring of geopolitical risk.
- Methods to monitor for emergent geopolitical topics.
- Methods for associating topic-based geopolitical risk with selected metrics for best predictive performance.

I aim to address these questions with a combination of sentiment analysis, topic modeling, and named entity recognition to create a geopolitical risk analysis profile that can automatically evaluate the data from Twitter / X.

The rest of this thesis is organized into the following chapters: Chapter 2 provides a literature review of the research that underpins my thesis from an interdisciplinary framework. Chapter 3 describes one of the major computer science methods I employed to build out my geopolitical risk evaluation: sentiment analysis. Chapters 4 through 6 are various case studies that employed programs and methods developed for analyzing geopolitical risk across various time intervals and subject matters, including topic modeling. Chapter 7 focuses on named entity recognition which was utilized for providing context and buttressing a few information gaps from the other methods. Chapter 8 is discussion of my research including limitations and future steps. Finally, Chapter 9 concludes.

The opening of *The March of Folly* [425], another of Tuchman's classics, begins with one of the first geopolitical conflicts, The Trojan War. While most of the Trojan War is within the mythical realm, what Tuchman expresses is that in geopolitics, information is powerful. The Trojan princess, Cassandra, was cursed with the ability of prophecy, but for no one to believe her. Thus, when she foresaw the danger of the Trojan Horse, she could not effectively pass along this information leading to the destruction of Troy. For my research, I aim to make sure that people have geopolitical risks evaluated quickly, to obtain information, so that they can stay ahead of potentially hazardous situations that can have not only local impacts or regional impacts, but global ones as well.

Chapter 2: Literature Review

This chapter focuses on a general literature review that describes the history of the research that relates to my thesis. I have decided to leave the literature review of bigger concepts such as sentiment analysis, named-entity recognition, and topic modeling as sections within their own subsequent chapters to leave this chapter as a more streamlined and focused discussion on how my project grew out of these previous works. To this end, Section 2.1 is a general record of the how major concepts of my thesis developed overtime. Section 2.2 takes a more interdisciplinary view of my project, looking into how different researchers attempted to answer similar questions to my own thesis through political science techniques. Section 2.3 brings the research back to computer science with different attempts using social media, mainly Twitter / X, to analyze data similar to how I do so in my research. Section 2.4 mentions research in sentiment analysis, while Section 2.5 describes the beginnings of topic modeling, and Section 2.6 details named entity recognition. Section 2.7 briefly discusses the commercial attempts to evaluate geopolitical risks. Finally, Section 2.8 concludes with an explanation of the research gaps that my project aims to fill.

Section 2.1: General Literature Review

The opening section of this chapter, I will describe some of the general concepts that have underpinned my project. One of the main research areas that buttresses my PhD thesis is Event Detection. Event Detection is a robust field with many practical uses, allowing for earlier awareness of everything from natural disasters to disease spreads, and, importantly, geopolitics [460]. As a concept, it has been studied cross-disciplinary, in political science (discussed in Section 2.2) and computer science, especially with the rise of social media in the recent decades (discussed here and more in depth in Section 2.3). The earliest major study involving event detection I came across was conducted by Defense Advanced Research Projects Agency ("DARPA") in the mid to late 1990's, Allan, et al [47]. Using a corpus of 16,000 news stories, the report focused on four main tasks: "Segmentation" – how to break a large amount of data into individual events; "Retrospective Event Detection" – how to find all the events that happened in a past time frame; "On – line New Event Detection" – how to identify when a new story emerges from incoming, continuous news articles; and "Tracking" – how to follow a story over a time frame (Allan, et al. pg. 2). Many early studies took these ideas further as with Yang, et al [456], who built directly off Allen, et al., investigating the best data clustering and non-clustering techniques at the time to do retrospective and on-line event detection (Yang, et al, 1998, Pg. 7). While Gabrilovich, et al. [184] used novelty detection (a way to determine if "a point is an outlier or not" (Roberts, 1999, Pg. 1) [365]) algorithms in their "Newsjunkie" system, to "analyze live newsfeeds and identify articles that carry most novel information given a model of what the user has read before" (Gabrilovich, et al. 2004, Pg. 7).

These early papers paved the way for further research. For example, Iacobelli, et al, [239] developed a news analysis system "Tell Me More" which built off Gabrilovich, et al. 2004, adding ways to show "what exactly is new information in the articles presented" to user of their system. Further, Karkali, et al, [258] included the inverse document frequency ("IDF") measure in their news event detection design. IDF is a measure of how rare a certain word is in the full corpus of documents, an uncommon or novel word will have a higher score and thus documents that have many novel words will have very high IDF scores, which would be an indication of a new event (Karkali, et al, 2013, Pg. 6). Going in a different direction from these studies, but still relevant to my research was the inclusion of spatiotemporal factors to better determine the location of these events and not just the events themselves. Ho, et al [226] developed a system to mine local news articles to determine their spatial attributes (place name, town name, etc.) and their temporal attributes (time of day, month, year, etc.) in combination with text analysis to provide early indicators to users about potential events, both positive and negative, in their surrounding area (Ho, et al. 2012, Pg. 1 – 2). These news-related event detection works were foundational to my research, but it was the inclusion of social media, specifically Twitter / X, into this research that showed how dynamic this field has become in recent years.

Atefeh and Khreich [62] state in their literature review paper on event detection in both traditional news media and Twitter / X, that what makes using Twitter / X beneficial is that: “unlike other media sources, Twitter messages provide timely and fine-grained information about any kind of event, reflecting, for instance, personal perspectives, social information, conversational aspects, emotional reactions, and controversial opinions” (Atefeh and Khreich, 2015, Pg. 2). Several studies have left their mark on the changing landscape in event detection. Metzler, et al [305] used Twitter / X to achieve retrospective event detection across different categories such as “Crime” or “Energy” (Metzler, et al, 2012, Pg. 6). Alsaedi, et al, [50], on the other hand, developed a way to detect on-line event detection using a real time Twitter / X Stream API to look for disruptive events which they define as: “An event that interferes in the achieving of the objective of an event or interrupts ordinary event routine. It may occur over the course of one or several days, causing disorder, destabilizing securities and may result in displacement or discontinuity” (Alsaedi, et al, 2015, Pg. 2). While Agarwal, et al, 2021 [39] utilized Twitter / X to investigate the correlation between local events they detected. In addition, Machine Learning techniques have become prominent with many studies using decision trees, support vector machines (SVM), logistic regressions, and Naïve Bayes to tackle the issue of "retrospective" and "specified" event detection [268, 270, 459]. However, despite the growth of research, the best way to capture emerging topics through the use of event detection and Twitter / X is still an open question.

There are numerous of benefits from using Twitter / X for event detection due to its large diversity of sources and speed, but as Atefeh and Khreich explain: "traditional text mining techniques are not suitable, because of the short length of tweets, the large number of spelling and grammatical errors, and the frequent use of informal and mixed language." (Atefeh and Khreich, 2015, Pg. 1). Thus, it is important to develop specific text data analysis systems when dealing with Twitter / X. Thus, while event detection itself is an important factor of my research, its combination with the concepts of text mining was equally vital. Kochut, et al, 2017 [268] define text mining as "the task of extracting meaningful information from text" (Kochut, et al, 2017. Pg. 1). This definition of text mining is incredibly broad and covers many concepts, however, for this project I only focus on a few: Sentiment Analysis, Natural Language Processing ("NLP"), and Topic Modeling. These concepts have their own literature review sections later in this chapter.

Finally, I briefly want to mention some of the most recent papers at the time of writing, that make use of these concepts like I do. In their 2022 paper, Smith and O'Hare [398] use sentiment analysis methods to analyze tweets from major accounts such as CEOs of large companies and heads of state to see if the changes in sentiment of the text of these accounts are correlated with changes in stock prices. Herrera, et al., [222] "employs the hybridization of deep learning models and investor sentiment to forecast the return and volatility of renewable energy stocks... we highlight the significant contribution of investor sentiment information for forecasting stock return and volatility of renewable energy companies" (Herrera, et al., 2022, Pg. 2), were they used Twitter / X sentiment as a proxy investor sentiment. Additionally, Lehkonen et al., 2022 [283] made use of Twitter / X and other news sources as well to develop a measure to compare the "Media Tone" (Lehkonen et al., 2022, Pg. 2) for comparing the changes in that tone to changes in the foreign exchange market which is one of the many financial markets that I investigate in my Case Study Chapters. A working paper by Adams, et al. [36] from 2023 also used Twitter / X to investigate how sentiment changes around Federal Open Market Committee ("FOMC") meetings. Finally, while not as recent as the others, Nisar and Yeung [324] made use of Twitter / X on a smaller scale local political event comparing reactions to local elections on Twitter / X in the UK to the changes in the UK Stock market, the FTSE 100. While my research delves more into global geopolitical events, this study provided insight into how to handle political events on Twitter / X.

With the general literature review, I wanted to give a brief overview of some of the concepts that will be appearing throughout my thesis. However, the rest of this chapter will dig down into the specific research fields that helped form the ideas for my project including political science, computer science, the text mining concepts, and commercial enterprises.

Section 2.2: Political Science Literature Review

In the political science field, there has been a great deal of interest in the analysis of geopolitical risk for decades. Schrodt and Gerner, 2012 [382] describe the development seven major datasets containing various geopolitical events, with one, the BCOW data set, reaching back to the early 19th Century (Schrodt and Gerner, 2012, Pg. 9 -13). In addition to these datasets, there is the International Crisis Behavior (ICB) Project [96, 245] started in 1975, which "undertook an inquiry into the sources, processes, and outcomes of all military-security crises since the end of World War I, within and outside protracted conflicts, and across all continents, cultures, and political and economic systems in the contemporary era." (James, et al. 2022). These datasets were foundational in various political science studies [67, 105] in their various inquires.

However, outside of these datasets, there have been multiple studies that have used various political risk indicators to make geopolitical inferences and predictions. For example, Aisen and Veiga, 2005 [42] use various within country "political, institutional, and economic variables" (Aisen and Veiga, 2005, Pg. 5) as metrics of political instability to examine if inflation is affected by the political instability. Another paper, Goldstone, et al., 2010 [202], built a model to examine what factors matter most in predicting political instability. They found that their "...new measure of regime type emerges as the most powerful predictor of instability onsets, leading us to conclude that political institutions, properly specified, and not economic conditions, demography, or geography, are the most important predictors of the onset of political instability." (Goldstone, et al., 2010, Pg. 1). In addition to political instability, predicting war and conflicts is a significant area of research in political science. Hegre et al., 2013 [221] developed a model for predicting armed conflict throughout the world from 2011 to 2050 based on "population size, infant mortality rates, demographic composition, education levels, oil dependence, ethnic cleavages, and neighborhood characteristics" (Hegre et al., 2013, Pg. 1). Using data from 1970 – 2009, in combination with these variables, Hegre et al, were able to produce stunningly accurate results with eight countries (Ethiopia, India, Philippines, Myanmar, Afghanistan, Sudan, Pakistan, Congo DRC) out of their predicted top ten countries that will experience armed conflict, experiencing at least a minor conflict by 2017 (defined as: "between 25 and 999 battle-related deaths per year" (Hegre et al., 2013, Pg. 2)). The remaining top ten, Thailand and Algeria, had a bloodless military coup in 2014, and had the president removed from office by protesters in 2019, respectively.

These papers provided a wider range of ideas about how to analyze geopolitical data outside of pure computer analysis. Their techniques are more statistical in nature using various data modeling structures such as gaussian mixture models (GMM) [42], conditional logistic regressions [202], and dynamic multinomial logit model [221]. However, out of the political science field, Chadefaux [111], Baker, et al [67], and Caldara and Iacoviello [105] provided a methodology that I have made the most use of, providing a foundation to expand my research.

While all three papers employ similar methods, I will start with Chadefaux, 2014 [111]. In his paper, Chadefaux gathered news articles at weekly intervals throughout the 20th century from 1902 to 2001, using keywords to identify countries and potential war words (Chadefaux, 2014, Pg. 3 – 4). Gathering these articles allowed Chadefaux to identify the trends and frequency in news reports that reference conflicts. In a simple count analysis (Chadefaux, 2014, Pg. 4 – 5), Chadefaux found that "a visible upward trend appears at least three to five years before large wars, and two to four years before minor wars. Unsurprisingly, I also find that the number of conflict-related news items is much higher within the year that precedes the outbreak of war than at other times, for wars of any scale or type" (Chadefaux, 2014, Pg. 5). Further, Chadefaux was able to show using multinomial logits, that using these conflict news stories provide some predictive capabilities when trying to assess when a war will break out, especially if the war was interstate (i.e. between countries) (Chadefaux, 2014, Pg. 9). This kind of retrospective study in war and geopolitics was a crucial building block for my research. While not automatic, this study was able to show this sort of event detection using keywords, can be applied to a geopolitical framework.

Building on this news frequency concept further, Baker, et al, 2016 [67] expand this research by building “a new index of economic policy uncertainty (EPU) based on newspaper coverage frequency” (Baker et al., 2016, Pg. 1). While more focused on economics and policy decisions rather than geopolitics, Baker et al., created their overall index, the EPU, by searching for articles since 1985 with at least one keyword each category of “uncertainty, the economy, and policy”. (Baker et al., 2016, Pg. 7). To further investigate specific categories, such as health care or national security, they added category specific keywords for the article to have in addition to the EPU keywords (Baker et al., 2016, Pg. 9 – 11). Using these categories, Baker et al. also constructed EPU indices for other countries to compare them to the United States (Baker et al., 2016, Pg. 12 – 13). One of their main analyses was investigating the relationship for the EPU index to trends in “firm-level stock price volatility, investment rates, and employment growth” which showed how uncertainty shocks, defined by the EPU index, are correlated to negative economic effects (Baker et al., 2016, Pg. 41). By expanding their research to cover multiple topics, Baker et al. laid the framework for my research which has a focus on analyzing multiple geopolitical topics. Additionally, Baker et al. showed that it is possible to connect changes in policy news articles identified with keywords to changes in different economic sectors. Additionally, Baker, et al, put forth a version of the EPU index that is updated daily (Baker et al., 2016, Pg. 9). This daily update, which was on a smaller time frame than Chadefaux’s weekly update, showed that it is possible to gain usable insights despite the shorter time frame.

Caldara and Iacoviello, 2022 [105] took the work of Baker et al. further still by creating a risk index similar to the EPU, but focusing on geopolitics which provided some of the most crucial insights on how to best construct my project. Caldara and Iacoviello constructed their index through reviewing newspapers articles like Chadefaux and Baker, et al (Caldara and Iacoviello, 2018, Pg. 7). However, the difference is that they investigated six different geopolitical categories: “Geopolitical Threats”, “Nuclear Threats”, “War Threats”, “Terrorist Threats”, “War Acts”, “Terrorist Acts” (Caldara and Iacoviello, 2018, Pg. 33). These six categories formed their overall geopolitical risk index (GPR), but Caldara and Iacoviello found that geopolitical threats and geopolitical acts have differing effects and decided to split their index into two separate indices, the geopolitical threats index (GPT) and the geopolitical acts index (GPA) (Caldara and Iacoviello, 2018, Pg. 9). Using these three indices, Caldara and Iacoviello were able to show that “exogenous changes in geopolitical risks depress economic activity and stock returns in advanced economies, most notably in the United States. Importantly, these adverse effects are sparked by heightened threats of adverse geopolitical events, rather than their realization” (Caldara and Iacoviello, 2018, Pg. 27).

Their GPR index captured the changes in geopolitical risks, and Caldara and Iacoviello were able to show how the increases in GPR index predicted lower stock returns (Caldara and Iacoviello, 2022). Specifically, they found that after the US stock market reopened after 9/11, that transportation stocks decreased by 13%, while precious metal commodities increased by 7.4% (Caldara and Iacoviello, 2022). For longer term affects, such as a major war, Caldara and Iacoviello showed through vector autoregressions that increases in their GPR index increased financial uncertainty and caused declines in the US stock market and Oil price lasting two years.

The results of the Caldara and Iacoviello paper provided clarity into how best to develop a wide-ranging geopolitical risk monitoring programs and the potential outcomes of the changes in geopolitical landscape. Furthermore, the genesis for the geopolitical topics generation described later in Section 2.5 came out of the findings of Caldara and Iacoviello.

These political science papers provided a solid theoretical background for my programs, but to achieve the automatic evaluation portion of my programs, I needed to explore how others had handled this concept in the computer science field.

Section 2.3: Computer Science Literature Review

Since the late 1990's, research into the tasks laid out in Allan, et al., [47] has greatly expanded. With the creation of Twitter / X in 2006, computer science researchers have moved from analyzing news articles to gathering millions of tweets that are constantly updated containing thousands of events. So much research has been done that one review paper written in 2019, Saeed, et al., on event detection techniques using Twitter / X had over 100 references [371]. Twitter / X studies have investigated a wide variety of topics over differing time scales. For example, Diakopoulos, et al, [156] examined tweets relating to one U.S. State of Union by President Barack Obama in 2010 determining how viewers felt about the different topics discussed in the address. While Bollen, et al, [91] used sentiment analysis to explore the relationship between public mood determined by tweets gathered over a four-month period in 2008 correlates to "cultural, social, economic, and political events" (Bollen, et al, 2011, Pg. 8) and economic markets such as "[U.S.] stock market and crude oil price indices" (Bollen, et al, 2011, Pg. 8). Furthermore, Beykikhoshk, et al, 2015, [83] harvested and assessed over 11 million tweets over a 13-month period to learn more about the Autism community (Beykikhoshk, et al, 2015, Pg. 1 and Pg. 4). However, one of the major areas of Twitter / X research, is health sciences which have been using Twitter / X for more than a decade.

Even before the start of COVID-19, many studies investigated the spread of diseases through tweets [113, 125, 140]. Culotta [140] studied the correlation between flu rates and posts on Twitter / X referencing flu symptoms. While

Chandrasekaran, et al, [113] studied the quality of information surrounding the Zika virus shared on social media, including Twitter / X. However, it was Chew and Eysenbach's [125] investigation into the spread of H1N1 (also known as swine flu) that yielded a valuable guide for my research. By gathering tweets that used keywords relating to swine flu (Chew and Eysenbach, 2010, Pg. 1), similar to the political science index studies [67, 105, 111]. Chew and Eysenbach found that "sharp increases in absolute H1N1-related tweet volume coincided with major H1N1 news event" (Chew and Eysenbach, 2010, Pg. 8). Using sentiment analysis in combination with the tweet trends, they were able to find insights into the type of messages being spread regarding swine flu (Chew and Eysenbach, 2010, Pg. 7 – 8). One of the most important takeaways from Chew and Esyenbach research was their determination "tweets can be used for near real-time content and sentiment analysis... allowing health authorities to become aware of and respond to real or perceived concerns raised by the public" (Chew and Eysenbach, 2010, Pg. 12). Thus, this paper provided an excellent example of how tweets could be used to evaluate real time data, making it possible to do the automatic evaluation required for my project.

This is just a small sample of the plethora of studies that make use of Twitter / X, however, there are a subsection of studies that contained valuable information regarding the methodology structure for the thesis. Many studies constructed various systems to help analyze the tweets to detect events more efficiently. Like the political science studies of Chadefaux [111], Baker, et al, [67], Caldara and Iacoviello [105], they focus on news. However, thanks to Twitter / X, these studies have access to a greater number of sources with greater speed. For example, the system developed by Sankaranarayanan, et al, TwitterStand [377] focused on using Twitter / X to detect breaking news events faster than using traditional news sources. In addition, they use "Natural

Language Processing (NLP), namely Part-Of-Speech (POS) tagging and Named-Entity Recognition (NER)" (Sankaranarayanan, et al, 2009, Pg. 8) to extract a location mentioned in the tweet text to help identify the location of an event. TwitterStand focused more on wide ranging news events such as the Iranian 2009 Election (Sankaranarayanan, et al, 2009, Pg. 9) to traffic accidents in Los Angeles (Sankaranarayanan, et al, 2009, Pg. 7). In contrast, EvenTweet [31] created by Abdelheq, et al., while followed a similar methodology to Sankaranarayanan, et al, but focused on event detection on what they define as "localized events (e.g., festivals, road jams, football matches)" (Abdelheq, et al., 2013, Pg. 2). To find these localized events, Abdelheq, et al, acquired various keywords that correspond to a local spatial area (Abdelheq, et al., 2013, Pg. 2). By including a temporal shifter to their system, Abeldheq et al. were able to track the emergence of these events and the popularity of them over time (Abdelheq, et al., 2013, Pg. 2). Additionally, the system developed by Valkanas, et al, TwInsight [434], included an emotional analysis feature to understand how people felt about the event (Valkanas, et al, 2013, Pg. 5). While all these systems examine different events and data, the core methodology is generally the same. The tweets are collected through a real-time Twitter / X API [31, 377, 434], looking for specific keywords, the tweets are then processed with various text mining techniques [31, 377] to find different inferences based on what the studies are looking for (i.e., sentiment analysis [308], location [31, 377, 434], or event described [31, 377, 434]). Finally, the results of the analyses are visualized.

While the studies above were some of the early studies that provided a proof of concept of event detection on Twitter / X, my thesis also relied on the research and systems that have arisen since then. While my project follows a similar methodology as these previous studies [31, 377, 434], there are a few additions as well. The program I created collects tweets through a real-time

Twitter / X API [31, 434]. Specifically, I implemented the Filter Stream API. This API allows the user to input key words and a language, such as English or Spanish, to collect tweets in real time that are in the stated language and contain the key words. This method of data collection was used by other studies such as Metzler, 2012 [305], and Culotta, 2010 [140] both use key words in their studies to gather tweets. My thesis will use this API find the tweets that used the geopolitical topic key words I am investigating. These tweets are then analyzed looking to capture a sentiment of a tweet like Rajput, et al [351] did for recent COVID tweets and how Rill, et al [362] did in their PoliTwi system, which focused on political tweets surrounding the German parliamentary election in 2013 (Rill, et al, 2014, Pg. 1). This will allow for to better understanding of the context of how the geopolitical events are discussed. Furthermore, my program will be able to geolocate where the events are taking place by using the words described in the tweets like Marcus, et al., 2011 [296] use in their TwitInfo system. While many of these studies use various data clustering techniques to identify the events discussed on Twitter / X, I will be including topic modeling over time (described in Chapter 5) to further improve upon these previous systems.

Lastly, my programs were also developed through the works that did not build data analysis systems, but I found were helpful in building out my own programs: Abouzahra and Tan [33], Rajput, et al [351], and Yu, W, et al [459]. Abouzahra and Tan, and Rajput, et al, both focus on one specific event, the 2015 Zika virus epidemic[33] and 2020 Coronavirus pandemic [351], respectively, using Twitter / X data to track the growth of the viruses. Abouzahra and Tan, used Twitter / X to track the mentions of Zika over time and across countries, they also use both Spanish and English tweets to track the Zika virus. This study provided a starting point for building out my research into multilingual analysis, which many of previous systems lacked. Differing from Abouzahra and Tan, Rajput, et al., applied sentiment analysis to their tweet data set which provided insight into how to best apply sentiment analysis with tweets. While my research focuses on the automatic evaluation of geopolitical risk, one of main the ideas that underpins this concept is the ability to identify emerging geopolitical topics. A paper to do so was Yu, W., et al., 2017 [459]. They used term frequencies in combination with graph theory to identify multiple emerging events discussed on Twitter / X. While this is ultimately different from how I plan to use topic modeling over time to identify the emerging topics, it was fundamental to my thinking into how to best identify the emerging geopolitical events and potential risk. Thus, currently, my programs can handle more topics than in Abouzahra and Tan, 2021, and Rajput, et al, 2020 while having more specificity than Yu, W., et al., 2017 by focusing on geopolitics.

Section 2.4: Sentiment Analysis Literature Review

One of the most important text mining concepts used in my research is sentiment analysis. Sentiment analysis, as described by Vohra and Teraiya [437], "is the automated mining of attitudes, opinions, and emotions from text, speech, and database sources through Natural Language Processing. Sentiment analysis involves classifying opinions in text into categories like "positive" or "negative" or "neutral". It is often referred to as subjectivity analysis, opinion mining, and appraisal extraction" (Vohra and Teraiya, 2013, Pg. 1). First used as a way to analyze customer responses on online reviews as it was "useful to product manufacturers because they want to know any new positive or negative comments on their products whenever they are available" (Hu and Bing, 2004, Pg. 9) [234]. However, since then, it has grown to analyze all sorts of text and research areas, such as Mitchell, et al, [308] employing sentiment analysis on tweets to determine happiness across the U.S. With the advent of review websites and social media, there was a flood of text data that no one person could possibly read and analyze to gain insight into what everyone was saying. With sentiment analysis, that became possible. Sentiment analysis is now used in everything from understanding how customers feel about different car companies [392], to applying sentiment analysis tweets to predict stock market changes [309]. Alsaedi, et al [50] implemented sentiment analysis as a metric in their event detection system. Some researchers even apply machine learning and deep learning techniques such as “Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and Long-Short Memory networks (LSTMs)” (Komban, 2022) [270].

There are three main ways to do sentiment analysis: 1) a dictionary / lexicon of words with values that is applied to the text to get a sentiment score; 2) using machine learning such as RNNs and Transformers which I will discuss later; 3) a hybrid of the two, an example of which can be found in Ng, 2022 [320]. The growth of sentiment analysis within text mining and its inclusion in my research produced fruitful results. I present a deeper discussion of these concepts in Chapter 3.

Section 2.5: Topic Modeling Literature Review

The next text mining concept is topic modeling, which Kochut, et al, defined as “a probabilistic model is used to determine a *soft* clustering, in which every document has a probability distribution over all the clusters as opposed to hard clustering of documents. In topic models each topic can be represented as probability distributions over words and each documents is expressed as probability distribution over topics” (Kochut, et al, 2017, Pg. 3). In plain English, topic modeling analyzes the documents of interest, in my case tweets, and builds topics based on words in the document corpus that have the greatest relationship to each other. Then it assigns each document differing probabilities for each topic. The highest probability associated with topic for a document is the topic the document most likely discusses. Thus, if you have a continuous data set over time, such as patient – psychologist meeting transcripts such as in Chaoua, et al [114], or a collection of tweets captured a specified time period as Saha, et al [372], it is possible to graph the changes in topics over time based on the new documents that enter my programs and the changing probabilities.

This section briefly describes the development of topic modeling, specifically looking into Topic Tracking over Time, Latent Dirichlet Allocation and Dynamic Topic Modeling. These ideas formed the backbone of my research into the development of tracking geopolitical topics and finding the emergence of topics which were two of my research aims. This chapter provides the background for topic modeling methods that I employ in Case Study 2 in Chapter 5, which contains the main analysis and application of the multilingual geopolitical topics that provide greater context for my evaluation of geopolitical risk.

Section 2.5.1: Topic Tracking over Time

The roots of topic tracking over time from news and social media lay in the mid to late 1990s with Allen, et al [47] funded study aimed at using various methods to " (1) segmenting a stream of data, especially recognized speech, into distinct stories; (2) identifying those news stories that are the first to discuss a new event occurring in the news; and (3) given a small number of sample news stories about an event, finding all following stories in the stream." (Allan, et al, 1998, Pg. 1). The idea is straight forward, with a time series of text data, for example, news articles, is it possible to identify emerging stories given the methods at the time? Partners in the study: CMU, UMASS, and Dragon Systems, tried to different ways based off clustering algorithms to group the different news articles into topics while looking into emerging and fading topics with success (Allan, et al, 1998, Pg. 11 - 13). However, since then, new methods for identifying topics with increased computational power have been applied to this research field [89, 138, 237, 445]. One of the most popular methods that have been employed, and the one my research uses, is topic modeling, specifically Latent Dirichlet Allocation ("LDA") [90].

Section 2.5.2: Latent Dirichlet Allocation

Topic modeling is similar process to clustering, but uses text data, where each document in the corpus (the set of documents) is sorted into groups of related documents which form the topics [274]. Topic modeling "can be considered as a fuzzy classification because it provides a soft degree of belonging of the documents to a specific topic" (Amara, 2021, Pg. 6) [54]. This means that one document can belong to multiple topics. LDA extends this concept. As defined by Blei, et al., "LDA is a three-level hierarchical Bayesian model, in which each item of a collection is modeled as a finite mixture over an underlying set of topics. Each topic is, in turn, modeled as an infinite mixture over an underlying set of topic probabilities. In the context of text modeling, the topic probabilities provide an explicit representation of a document." (Blei, et al., 2003, Pg. 1) [90]. In other words, LDA uses an expectation maximization ("EM") process with Dirichlet – Multinomial conjugate priors based off the algorithmic hyperparameters set for number of topics, and the alpha and beta of the Dirichlet representing the document – topic density and the word – topic density respectively [256].Using the hyperparameters as a start for the EM process, LDA updates the probabilities of each word in the corpus belonging to each topic and the probabilities each document in the corpus belonging to each topic. Once the EM process converges, the LDA will have optimal values for the probability of the words in each topic and the probability of which topic each document most belongs too, which allows for interpretation of the topics based of the words that make up each topic, thus "discovering the hidden themes in the collection [documents]" (Kulshrestha, 2019) [274]. Additionally, as discussed by Ranaei, et al. [352] "one of the prominent advantage[s] of LDA. . . is its focus on the context" (Ranaei, 2020, Pg. 24). Figure

2 from Blei, et al, 2003 displays graphically a hierarchical Bayes process which LDA is:

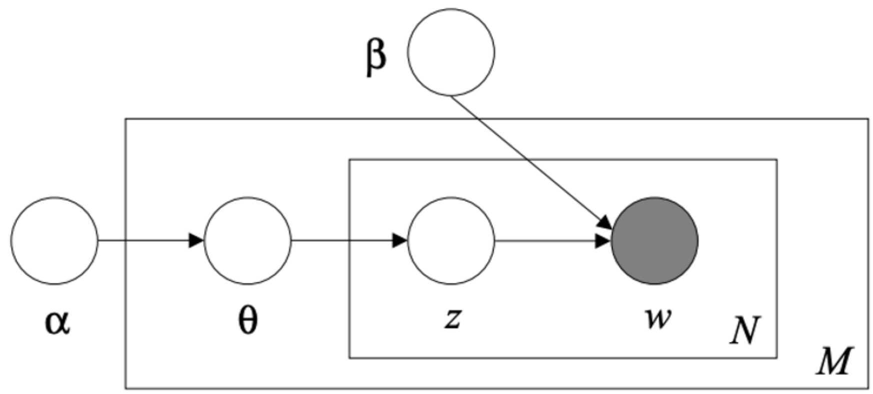


Figure 2: This is representation of the LDA model, $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are described above. $\boldsymbol{\theta}$ is the proportion per tweet (how much of a tweet is represented by each topic), **z** is the topic assignment per word (the probability of each word belongs to each topic), **w** is the observed word, **N** is the number of words, **M** is the number of tweets. $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are corpus level variables, $\boldsymbol{\theta}$ is a tweet level variable, and **z** and **w** are word level variables (Adapted from Blei, et al., 2003, Pg. 20, [90] and further description from SenthilPerumal [386])

While this is a powerful method to find topics in documents, the major drawback to LDA for my research is that it is "static" (Blei and Lafferty, 2006, Pg. 2) [89]. This means that it evaluates all the documents in corpus at once and not designed to evaluate the data chronologically, which posed a problem for my analysis. Luckily, they are several methods that extend LDA for this capability.

Section 2.5.3: Dynamic Topic Modeling

Dynamic Topic Modeling [89] is one such method which "developed sequential topic models for discrete data by using Gaussian time series on the natural parameters of the multinomial topics and logistic normal topic proportion models" (Blei and Lafferty, 2006, Pg. 6). By using a normal logistic distribution over the Dirichlet distribution, Dynamic Topic Modeling allows for developments of topics over time, and thus one can see when topics emerged. However, the issue with using the normal logistic distribution is that the normal logistic distribution is not conjugate with multinomial distribution. This non-conjunction prevents the use of the Expectation Maximization updating feature of the traditional LDA method, as developing a posterior inference distribution for the EM algorithm is now intractable, i.e., the distribution of the posterior inference cannot be evaluated with finite numerical operations, thus making it impossible to get stable probabilities. To get around this issue, Blei and Lafferty make use of variational approximation methods. As Blei and Lafferty describe: "The idea behind variational methods is to optimize the free parameters of a distribution over the latent variables so that the distribution is close in Kullback-Liebler (KL) divergence to the true posterior; this distribution can then be used as a substitute for the true posterior" (Blei and Lafferty, 2006, Pg. 3). This Kullback-Liebler divergence is a measure of the difference between two distributions, thus when this value is minimized between a solvable approximate distribution and the intractable true posterior distribution, it is possible to substitute the approximate distribution for the true distribution allowing for the Dynamic Topic Modeling algorithm to continue. This process is also shown in Figure 3, graphically:

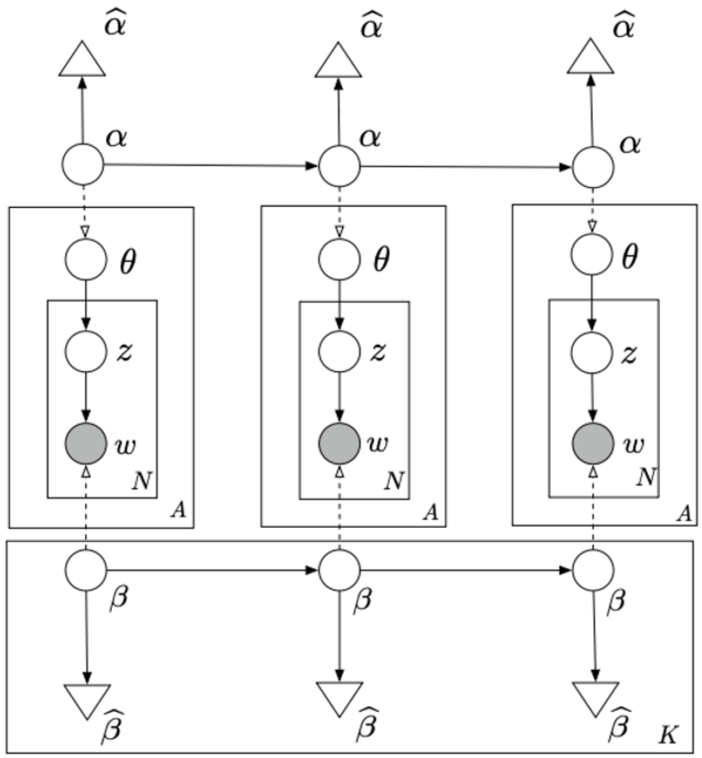


Figure 3: This is a graphical depiction of the variational version of the Dynamic Topic Model across three-time intervals. $\boldsymbol{\theta}$, **z**, **w**, and **N** are all the same as in Figure 2. **A** is now the number of tweets. **K** is number of topics hyperparameter. $\boldsymbol{\alpha}$ is now mean parameter of the logistic normal distribution at each time interval, while $\boldsymbol{\beta}$ is the multinomial distribution parameters for each topic, which are now able to evolve over time. $\hat{\boldsymbol{\alpha}}$ and $\hat{\boldsymbol{\beta}}$ are the variational approximation outputs (Adapted from Blei and Lafferty, 2006, Pgs. 2 – 3) [89].

While this is a fruitful method for evaluating topic modeling over time, it unfortunately did not merge as seamlessly and effectively with the rest of my programs as I found through exhaustive experimentation, thus I had to turn to additional methods. One method I tested was to run the LDA algorithm over each time frame, labeling topics that are created and comparing them to the previous time frames to track the emergence topics over time. While this manual application would work, it would mean that my program were not exactly automatic, thus I turned to a different process used by Besaw [82]. Besaw used weights for the different topics that are created and showing how the topics emerge and change over time. I incorporated this into my program and will go more in depth in the Methodology section of Case Study 2 in Chapter 5.

Section 2.6: Natural Language Processing Literature Review

Another key concept was NLP defined as a "sub-field of computer science, artificial intelligence and linguistics which aims at understanding of natural language using computers" (Kochut, et al, 2017, Pg. 2). There have been a great number of breakthroughs in the field over the last decade, including BERT [155] developed by Devlin, et al, which has become the gold standard in NLP tasks, including NER. Due to BERT's performance speeds and relative ease in fine-tuning for varying NLP tasks, it became one of the go-to processes for NLP. For example, it is now used in classifying multilingual texts [408] or Gonzalez, et al, who built their TWilBert system [204] off the BERT architecture for text classification of Spanish tweets.

Within NLP is the process known as NER, which Kochut, et al, explain as "the task of named entity recognition is to locate and classify named entities in free text into predefined categories such as person, organization, location, etc." (Kochut, et al, 2017, Pg. 2). For my research, I was able to find locations mentioned within tweets, adding more context to the geopolitical risks that were identified. I will go into more detail in Chapter 7 on NER and how it was used in my research, but here I will describe how a few studies used NER in a similar way to my research. Iacobelli, et al [239] used two different NER systems on their news articles to make the identification of named-entities more robust to give better indications of new information entering their system (Iacobelli, et al, 2010, Pg. 3 – 5). On the other hand, Alsaedi, et al [50] was more focused on getting the geographic information out of the tweets to help define the event identified by their system which creates data clusters out of tweets (Alsaedi, et al, 2015, Pg. 3 – 4).

While I did not make use of this system in my own research, a powerful NER model that should be mentioned in the Literature Review is the Edinburgh Geoparser [43, 44]. Issues with other NER systems, specifically with geoparsers is the issue of finding the right location. If there are multiple locations with the same name (for example, San Francisco exists in both the U.S. and Venezuela), a geoparser will only choose one location to return the coordinates for. This is detrimental for my research, as a geopolitical risk could be tagged in the wrong location on a map giving the wrong impression for the geopolitical risk. The Edinburgh Geoparser, however, can mitigate this issue with two methods. First, the user can specify a region that the text is based in, for example, if you set the geographic region to just the United States, then you will get all the San Franciscos in the United States rather than the whole world. The other method is setting a time for when the text was published or posted, this allows for easy verification as a quick search can verify if the text is referring to the right location based on the time that it happened. Additionally the Edinburgh Geoparser can process multiple text files, allowing for creation of maps with the full data displayed at once. Ultimately, I chose to use a different method, described in more detail in the Chapter 7, for a few reasons. First, the geoparser only works for English language texts, which would mean that to incorporate the Edinburgh geoparser into my program, I would first have to translate all the text data into English which could lead to translation errors which would decrease accuracy. Second, the translation would significantly slow down my programs, especially in the real time analysis. Thus while the Edinburgh Geoparser is useful, I needed to implement a different method.

Section 2.7: Commercial Literature Review

With a good understanding of geopolitical concepts through the political science research and ideas about how to best implement my risk evaluation programs through investigating research in computer science, I turned to attempts outside of academia. There were a few private institutions that have developed similar geopolitical risk indicators such as BlackRock, Dow Jones, and GDELT. BlackRock's Geopolitical Dashboard Report is comprehensive, containing a global risk indicator, top ten risks, and how these risks effect various markets such as bonds and currencies with brief breakdowns of a few of these risks, and a calendar of upcoming known geopolitical risks [9]. Dow Jones Factiva can provide reports that are tailored to their client's needs [6], "using more than 8,800 premium content sources and licensed third-party publishers to help businesses mine the universe of insights within the Dow Jones ecosystem" (dowjones.com). With their data sources Factiva states that the "content set can be read by text-mining or machine-learning systems for signals pertaining to risk or new revenue opportunities" (dowjones.com). GDELT, on the other hand, can process thousands of documents and news sources and search for hundreds of events, updating every 15 minutes [7]. GDELT has global coverage looking into news media from across the globe with translations and looking at historical data going back to the 1800s (gdeltproject.org). These three enterprises are incredibly powerful solutions to the issues surrounding the analysis of geopolitical risks however, they also leave gaps available for my project to fill.

Section 2.8: Gaps in Research

There are a few research areas I have identified that have been under-studied that my project aims to address. One sector that has room to grow is multilingual analysis. As Saeed, 2019 [371] explains:

> “We observed a research gap concerning multilingual content. Chierichetti et al. proposed an approach which is language independent. It focuses on non-textual features such as time and retweet. Other approaches, such as … model the tweets as a bag-of-words and rely on burstiness. Such methods ignore the multilingual semantics. Hence, there is a need for robust techniques that can effectively work for multiple languages.” (Saeed, 2019, Pg. 29).

There is currently an abundance of multilingual content analysis on Twitter / X [33, 139, 314, 446], and further explored in Chapter 5 in Case Study 1. However, there are multiple ways to do the multilingual analysis. Some studies use translations to English, such as Windsor, et al., 2019 [446]. Others use the original language of the tweets such as Mozetič, et al, 2016 [314]. Thus, there is still much debate around which way is better.

One gap that I have identified is the application towards geopolitics on Twitter / X. While many Twitter / X studies try to find all events that are occurring during a time frame such as the Yu et al RING technique [459]. In the computer science field in general, I found only one paper, Roberts et al, 2013 [364], that used the geopolitical example of the growth of China from the mid-1990’s to the mid- 2000’s to show off the functionally of their Structural Topic Model (“STM”) method that could track topics over time. Commercially, Dow Jones Factiva and GDELT mention that they make use of social media, but do not do so exclusively. Additionally, the political science studies that have looked into geopolitics [105, 111], have used data that comes from news articles exclusively, which may come out only daily, in this day and age is slow information.

The commercial solutions to analyzing geopolitical risk also provided fertile ground to build upon. With BlackRock Geopolitical Risk Indicator [9], it has monthly resolution, meaning that they update their analyses once a month. Using Twitter / X, my programs will be able to produce output on a daily, hourly, or even by minute timeframe and rapidly track emerging geopolitical risk. While Dow Jones Factiva [6] is incredibly powerful, it's use is tailored to companies and business sectors, thus the geopolitical risks that affected these companies specifically. However, my programs will be able to monitor and provide insight more generally. This will be beneficial since I am investigating large economic forces, such as stock and bond markets. My programs achieves similar potential benefits over GDELT [7] with Twitter / X, I can get near real-time updates to the events that are occurring, thus faster than GDELT can. Additionally, while GDELT uses machine translation, I analyze the text in the original language which has benefits discussed in Chapter 3.

I validated my results through a few different methods. While I go into more detail about the individual validation methods across Chapters 4, 5, and 6, the Case Study Chapters, I will briefly mention the main method here. In general, my program collected tweet data about general geopolitical risks with general keywords, thus to find out specific information I would look at news sites to see investigate the spikes in sentiment that occurred or the different topic words generated by topic modeling of the geopolitical tweets. While the results of the Case Study Chapters show the advantages of my program, I will also list them here. My programs can analyze geopolitical risk data in near real time, across multiple languages, while generating geopolitical risk indices that can be compared to financial market data, and emerging geopolitical risk topics.

These papers and enterprises were instrumental in shaping my thoughts on how to effectively build a novel way of analyzing geopolitical risk automatically. Through a thorough examination of a combination of political science research, computer science research, and commercial endeavors, my project was able to coalesce into what it is.

Chapter 3: Sentiment Analysis

This chapter focuses on sentiment analysis techniques that I employed for my research. Section 1 focuses on the Key Concepts of sentiment analysis and a breakdown of the different techniques, both lexiconic and machine learning based, that were used. Section 2 discusses the best policies of multilingual sentiment analysis, specifically whether it is better to translate the multilingual texts into English and use more powerful sentiment analysis algorithms developed for English, or instead analyze the texts in their original languages with sentiment analysis methods developed for each language. Section 3 takes this concept further, expanding into comparisons between different translation methods and different original language models on two classes of data: “Gold Standard Data” and “Full Data”.

Section 3.1: Key Concepts

There are a few concepts that my research uses that I believe should be detailed before I move on to the main analysis. These are Lexicon - Based Sentiment Analysis, Recurrent Neural Networks (“RNNs”), Transformers, and a non-sentiment analysis topic: Google Translate. I implemented all three methods to analyze the sentiment of the tweets I collected which contained the geopolitical keywords to obtain the sentiment of the geopolitical topics.

Section 3.1.1: Lexicon - Based Sentiment Analysis

For this section I’ll describe how the lexicon method works. Early lexicons were developed by Hu and Bing [234], which after tokenizing the text (i.e., breaking the text into individual words), one would apply the lexicon that contained words labeled positive or negative (1 or -1), when summed up would get the sentiment score for the text. Further lexicons were developed, such as AFINN [323], which applied degrees of negativity or positivity from -5 to

5, for certain words, to give a better understanding of the text. However, for my analyses, I used the Valance Aware Dictionary for sEntiment Reasoning or “VADER” lexicon developed by Hutto and Gilbert [238]. Developed for use on social media, it worked best for me as my data was short texts like social media posts. The VADER lexicon also includes the sentiment intensity from AFINN, it also included five “heuristics” frequently found in short text, social media posts to better capture the sentiment (Hutto and Gilbert, 2014, Pg. 6). These are “punctuation”, such as exclamation points; “capitalization”, for example if the text is all capitals, it would indicate stronger sentiment; “degree modifiers” before sentiment words like “extremely” or “marginally” would change the intensity of the sentiment word; “contrastive conjunction” words like “but” indicating a shift in sentiment throughout the text; and lastly they capture the trigram proceeding the sentiment word, as often these phrases can change the overall sentiment value of the word such as negations, for example, it will analyze "not happy" as negative sentiment instead of "not" and "happy" individually which would result in positive sentiment from the word "happy" (Hutto and Gilbert, 2014, Pg. 6). With these additions, the VADER lexicon was able to achieve higher accuracy in short text Twitter / X text sentiment than machine learning techniques such as Naïve Bayes and Support Vector Machines (Hutto and Gilbert, 2014, Pg. 9). That said, VADER does not include emoji and other non - textual analysis which are prevalent in social media posts, however, I have expanded on this limitation in Chapter 8.

Section 3.1.2: Recurrent Neural Networks

Recurrent neural networks are a form of neural networks created from layers of recurrent neurons. A recurrent neuron is a type of neuron that takes new input at each time step and its own output from the previous time step. Since the recurrent neuron can get its own output from each of the previous time steps, it can then remember short patterns and sequences for use as inputs and outputs, which can be used in a wide variety of tasks, such as sentiment analysis and translation (Géron, 2022, Pg. 500) [194]. For example, with sentiment analysis, it can be viewed as a type of sequence-to-sequence process were the sequence of the words of the text can be inputted into an RNN and that RNN can output a sequence of one aka a vector, which is the sentiment value of that text (Géron, 2022, Pg. 501). Many studies have applied RNNs to sentiment analysis with success, such as Sim, et al. [395] who tested various types of RNNs on getting the sentiment for Korean reviews of IKEA products, and Mishev, et al. [307] who applied the RNN's sentiment analysis capabilities to financial data. Similar to VADER, RNNs can also learn sentiment revolving around negation, modifying words, and emojis. As long as there is enough examples of these in the training data, these modifiers would not present a challenge for an RNN.

While RNNs are effective, there are a few drawbacks, namely unstable gradients and limited short-term memory (Géron, 2022, Pg. 501). However, there have been adaptations to RNNs to help overcome these issues. First to fix the gradients issue, it is possible to add dropout layers in the network, which bypasses some of the network's neurons, reducing complexity of the network over time and stabilizing the gradient. The second is the development of Long Short-Term Memory ("LSTM") and Gated Recurrent Unit ("GRU") layers. These layers allow the network to store longer sequences of data, but also forget data that is no longer needed (Géron, 2022, Pg. 514 – 520). As shown in Appendix E, in Tables 1 through 8, I used these improvements in the development of my own RNNs. These RNNs upgrades also have allowed RNNs to become effective language translator models.

By constructing two RNNs, one that is similar to the sentiment analysis sequence to vector model, which takes the text of one language and transforms it into a vector (this is known as the "encoder" network); the second is the reverse, called a vector to sequence model, which takes the same vector as an input, over and over, to create sequence (this is the "decoder" network") (Géron, 2022, Pg. 501). Combining these two networks allows for encoding of one language to decode it into another, thus translating it. An additional layer type helps this process is the Bidirectional layer. As described by Tracey, "sometimes we don't understand a sentence until an important word or phrase is provided at the end" (Tracey, 2019) [424], what the bidirectional layer does is it encodes the text both forwards and backwards, so the full context is transmitted through the network. Lee and Song [279] use this structure to create an effective English to Korean translator. And it is this dual network that underpins the Google Neural Machine Translation architecture which I will detail later [450].

Section 3.1.3: Transformers

Transformer models are used for all sorts of machine learning problems, but they came to prominence through translation work. Transformer translations differ from neural networks like RNNs due to the use of attention mechanisms and "multi-head self-attention" (Vaswani, et al., 2017, Pg. 7) in their implementation [66, 194, 436]. As described in Bahdanau, et al., 2014, [66] in the traditional RNN method for translation takes all the information in one language vector for the encoder and compresses it into a fixed − length vector which creates a "bottleneck" slowing down the training and translation process (Bahdanau, et al., 2014, Pg. 1). The attention mechanism instead encodes multiple vectors for the sentence and choses between them as it decodes the sentence into the other language. This difference speeds up translation and allows the network to work better on long sentences as it is no longer restricted to a single fixed vector (Bahdanau, et al., 2014, Pg. 2). The transformer expands on this by using "multi-headed self-attention" only, there are no RNNs or convolution layers found in other translation models (Vaswani, et al., 2017, Pg. 2) [436]. Breaking this down, what "self-attention" does is it takes a word in a sentence and calculates a score of how relevant that word is to other words in the sentence ("Glossary - Self Attention.", 2022) [153]. "Multi-head" applies the self-attention algorithm for each word in that sentence (Vaswani, et al., 2017, Pg. 4-5). Together, the "multi-head self-attention" form the base layer for the transformer architecture. Combining these "multi-head self-attention" layers for the encoder and decoder for a translation provides less computationally complexity, adds the ability to parallelize, and can analyze longer sequences than RNNs (Vaswani, et al., 2017, Pg. 6). The transformer translator I chose for my research is the OPUS-MT which "is based on a standard transformer setup with 6 self-attentive layers in both, the encoder and decoder network with 8 attention heads in each layer" (Tiedemann and Thottingal, Pg. 1) [417].

This architecture led to many developments in natural language processing, including the Bidirectional Encoder Representation from Transformers ("BERT") which expanded on the standard transformer architecture by implementing many transformer layers together with three embedding layers to determine the position of each word in the sentence [155, 204, 250]. With this basis, BERT also makes use of bidirectional masking, hence the name. Bidirectional, as described earlier, helps the model understand the sentence better, while masking hides one or more of the words in the sentence and has model try and predict the word. In addition to this training, BERT pre-trains on "next sentence prediction" where the model learns if one sentence follows the next (Géron, pg. 564). These two pre-training methods allows BERT adaptability for different NLP tasks and languages through fine-tuning, which is additional training of the BERT model with data sets for a language task, such as sentiment analysis. This architecture can also handle modifiers such as negation, capitalization, and emojis. This flexibility and strength made BERT models one of the most widely used architecture in NLP today. For example, HuggingFace.co[4], a popular repository for pre-trained machine learning models as over 9,000 models that reference "BERT" (roughly, 10% of the models in repository) and the most downloaded model is "bert-based-uncased" which has 24.4 million downloads [155]. More in line with my research, Pota, et al. [340] have also used BERT for sentiment analysis of tweets across English and Italian were able to achieve high levels of accuracy (Pota, et al., 2021, Pg. 9).

<u>Section 3.1.4: Google Translate</u>

First developed in 2006, Google Translate was initially a Phase-Based Machine Translation ("PBMT") method. As described by Quoc Le, a Google researcher, in the Washington Post: " 'With the previous PBMT model, when we translate a sentence from one language to another, we would translate one word or a phrase in the source sentence at a time, then re-order the words in the correct grammar of the target language' " (Turner, 2016) [428]. This led to mistranslations as the translation was not necessarily one for one. However, by 2016, Google had developed a new model based off the RNN architecture and attention mechanisms. The Google Neural Machine Translation ("GNMT") "consists of a deep LSTM network with 8 encoder and 8 decoder layers using residual connections as well as attention connections from the decoder network to the encoder" (Wu, 2016, Pg. 1) [450]. With an increased focus on "rare words", increasing parallelism, and reducing final translation time (Wu, et al., 2016, Pg. 1) in addition to treating " 'a whole sentence as a unit and translate [the words] in a group.' " (Turner, 2016), the Google research team found that "compared to the previous phrase-based production system, this GNMT system delivers roughly a 60% reduction in translation errors on several popular language pairs" (Wu,

[4] https://huggingface.co/

et al., 2016, Pg. 20). Several studies have made use of Google Translate for their research, such as Goel, et al. 2018 [199] who also used it for implementing multilingual sentiment analysis.

Section 3.2: Machine Vs Machine:

This section deals with how machine translation compares to machine learning regarding sentiment analysis for short texts across multiple languages.

Section 3.2.1: Introduction

Research into sentiment analysis across multiple language has been split by which techniques are best to use. They fall mainly into two camps: the first is using a machine translator service like Google Translate to make all the text English and analyze the translated text. The second is to create models and programs to work in the text's original language. Machine translation has improved significantly over the years and provided value for researchers especially as Pearse [334] describes: "if you have a very limited budget and need to translate large amounts of text very quickly" (Pearse, 2020). However, machine learning has made similar strides with vast improvements in neural networks, such as RNNs and transformer models, such as BERT. When working with multilingual data, many researchers will use a machine translator to translate all the data into English and apply a sentiment analysis classifier afterwards. In contrast, others have decided to apply machine learning algorithms directly to the untranslated text, usually using a different model for each language.

Since a best practice for analyzing multilingual data did not seem to exist, my questions became: which method is better? Is there a method that is overall superior? Or do certain techniques work better for different situations?

The rest of this section consists of a literature review detailing the competing methods in Section 2. The methodology and data description make up

Section 3. Section 4 presents the results, while Section 5 has a discussion of the implications.

Section 3.2.2: Literature Review

Many studies have used both translation and original language methods for their multilingual sentiment analysis needs. The idea for translation is explained by Cui et al. [139]: "An intuitive idea for multilingual sentiment analysis is to translate languages into a well-studied language (e.g., English); hence traditional methods can be applied" (Cui et al., 2011, Pg. 2). Several studies have made use of this method, for example, Goel, et al. [199] converted all their text data into English with the Google Translate API and then analyzed their datasets with RNNs and Naïve Bayes and compared accuracy with confusion matrices. Their RNN model was able to obtain an accuracy of 96.15% on their sentiment analysis dataset. Similarly, Arun and Srinagesh [59] translated their non-English data with the Google Neural Machine Translator and after testing their data against multiple machine learning algorithms, they found that a Support Vector Machine (SVM) obtained a 95% accuracy on their sentiment analysis tasks.

However, multiple studies have found flaws with using machine translation methods. While Bautin, et al. [75] state that they "believe that translated texts are sufficient to accurately capture sentiment" (Bautin, et al., 2008, Pg. 1), they also admit that "the loss of nuance incurred during the translation process. Even state-of-the-art language translation programs fail to translate substantial amounts of text, make serious errors on what they do translate, and reduce well-formed texts to sentence fragments." (Bautin, et al. 2008, Pg. 1). Ghorbel et al. [196] also found that while their translation from French to English was able to obtain a high accuracy for sentiment analysis, there were still a few intractable problems that arose from translating that reduced the accuracy potential of their research. Additionally, both Guo, et al., [209] and Erdmann, et al [168], only used machine translation on specific terms due to, as Guo, et al. describes, the "data noise arose from full-document machine translation often make a bad impact on the cross-lingual semantic association" (Guo, et al, 2010, Pg. 3).

The intuition behind using the original language is detailed by de Koning [149]: "If you are able to collect a big dataset in a native language, then the model trained on this dataset will probably give the best classification accuracy" (de Koning, 2020). Multiple studies have implemented this idea, such as, Cheng and Zhulyn [123] who created a naïve bayes classifiers boosted by a logistic regression and obtained an accuracy of over 90% in eight of the nine languages they analyzed. Attia, et al. [63] showed that Convolutional Neural Networks can obtain higher accuracy on sentiment analysis across multiple languages (English, German, and Arabic) than traditional classification methods like lexicons and machine learning algorithms like Naïve Bayes. Additionally, de Koning showed that that translation caused a 16% decrease in accuracy for sentiment analysis for Dutch.

While analyzing in the original language can provide high levels of accuracy, there are certain issues to consider. The main issue, outlined by McCormick [304]: "wouldn't it be better if Facebook just trained and published a separate model for each of these different languages? That would probably produce the most accurate models, yes–if only there was as much text available online in every language as there is English!" (McCormick, 2021). This is an important downside to working in the original language, that it is difficult to obtain adequate training and testing data, especially for sentiment analysis.

Section 3.2.3: Methodology and Data

Methodology:

For this section, I examined how well a sentiment analysis model developed in the text's original language compares to an English sentiment analysis model which evaluates the sentiment of the same translated text. The comparison will be across six languages: Spanish, French, Portuguese, Arabic, Japanese, and Korean. This is in line with the number of languages other studies [123, 340, 446], have used to evaluate their topics. These languages were chosen because, with English, they comprise roughly 80% - 85% of the languages used in tweets on Twitter / X where many of the short text data sources in multiple languages come from ("2018 Research on 100 Million Tweets", 2018) [1].

I constructed different sentiment analysis models for each language using different techniques to obtain a cross-section of analyses. For Spanish, French, Portuguese, and Japanese, I was able to obtain corpora large enough that effective RNNs were trained. These RNNs have good accuracy results and high processing speed for short text sentiment analysis which this experiment tests, thus when able, RNNs were the method I chose. For Arabic, a pre-trained Bidirectional Encoder Representation from Transformers ("BERT") model was used [155, 194, 240, 376] to evaluate the text. Inoue et al [240] were able to adapt BERT to create a model that does sentiment analysis for Arabic text. Their Arabic BERT model had higher accuracy than the RNN I created for Arabic, thus to both test a BERT model and to gain greater accuracy, I chose to use BERT for Arabic. For Korean, separate mixed model was used, with a multinomial naïve bayes model used to construct a sentiment lexicon to evaluate Korean text (Ng, 2022) [320]. Described in Ng, 2021, the multinomial naïve bayes breaks the labelled text data into word tokens calculating a score for each word based on the number of times the word appears in a positive or negative sentence. This score is then normalized and this word – score dataset can now be used as a lexicon for any text data that matches the language of the original data. This method works well if there is not enough data for training an RNN or BERT model. I chose the multinomial method for Korean for multiple reasons. Firstly, I wanted to evaluate the method's accuracy, thus when I could not locate an effective BERT model and the development of my Korean RNN encounter obstacles in training, I decided to pivot to the multinomial method. This pivot helped develop an sentiment analysis evaluation for Korean and both completeness for three sentiment analysis categories (Lexicon, Machine Learning, and Hybrid) . Lastly, for the English model, a lexicon-based model called VADER (Valence Aware Dictionary for sEntiment Reasoning), which reports a 96% accuracy for English tweets (Hutto and Gilbert, 2014, Pg. 9) [238], was utilized. Since my data is comprised of short texts like tweets, this makes VADER the best model to use for English. Thus, with its high accuracy and since it is a lexicon based sentiment analysis method, VADER was chosen for the experimental comparison.

To determine which method is better, I took samples of 1000 annotated short texts for each language and analyze them using their own sentiment analysis models. These samples are from separate data sets from what the sentiment analysis models were trained on, to evaluate the effectiveness of the models overall. I evaluated the models with confusion matrices and determined how accurately the labels produced by original language sentiment analysis models matched the original data. Afterwards, I translated each language sample to English, cleaned the data by removing punctuation, non-alphabet characters, and stop words, then used the VADER lexicon to label the samples, and used a confusion matrix to test for accuracy against the sample sentiment. Then I compared the output of the two confusion matrices to determine which was more accurate.

Finally, to determine a human ground truth, I used a combination of private human annotators and Amazon Turk workers to label the 1000 short texts as positive, negative, neutral, and ambiguous for all seven languages. I then reran the analysis for the data where the labels for positive and negative agreed between the original dataset and human annotators to determine if there is a difference in accuracy rates with this ground truth data.

Data:

The data I used comes from a few different sources. The English dataset for the baseline analysis was from Hammer [213] created for a Kaggle competition. For the creation of the sentiment analysis models, I used data from GitHub users, Darkmap [144], and Gamebusterz [187, 189] who provide Japanese review and French review corpora, respectively. As the Arabic model was pre-trained there was no need for Arabic training data. Park [331] provides a Korean corpus of movie reviews. The Portuguese corpus [64] came from a Kaggle competition created by Augustop. The Spanish corpus came from two sources, the first was the TASS 2020 datasets [406], the second was from another Kaggle competition [251].

For the data used in the 1000 samples, Bello [78] had a labeled corpus of random short Spanish texts. Gamebusterz [188, 190] had more short French movie review data sets that were not used in the model creation. The Portuguese sample came from another Kaggle competition created by Dias of various short product review [157]. Alomari [48] provided an Arabic corpus of labeled tweets. The chakki-works team on Github created a labeled dataset of Japanese product reviews in the “chABSA-dataset” [112]. Finally, Park [331] had another Korean movie review dataset that was not used in the creation of the Korean Lexicon.

Section 3.2.4: Results

Table 10 shows the initial accuracy of the language sentiment analysis models test data accuracy. For this section, the accuracy metric was based off the confusion matrix, with accuracy equal to the number of True Negatives plus the number of True Positives divided by the Sample Size (Acc = ((TN + TP) / (TN + FN + TP + FP)) where FN and FP are False Negative and False Positive respectively. The different methods had varying degrees of accuracy, the RNNs ranged from 74.2% to 94.1%. The Arabic BERT model had a 91.9% accuracy, and the Korean lexicon had a 78.7% accuracy.

Language	**Test Accuracy**	**Model Type**
Spanish	74.2%	RNN
French	82.1%	RNN
Portuguese	91.4%	RNN
Arabic	91.9%	BERT
Japanese	94.1%	RNN
Korean	78.7%	Hybrid - MNB

Table 10: Test Data Accuracy of the Sentiment Analysis Models

Table 11 displays the original text analysis with the cleaned English translated texts on the 1000 samples. The labels were compared to the labels from the dataset to check for accuracy with the confusion matrix. Unfortunately, my testing was not able to replicate the same results from the VADER paper for English. Overall, the models that analyzed the data in the original language produced higher accuracy rates than the translation / VADER model. There was a 20% difference in the average accuracy between the translation / VADER models and the original language models (54.0% and 74.5%, respectively).

Language	Data Translate	Data Original
English	N/A	63.0%
Spanish	67.6%	**68.4%**
French	51.5%	**81.3%**
Portuguese	53.4%	**68.9%**
Arabic	50.7%	**84.6%**
Japanese	48.6%	**68.0%**
Korean	52.4%	**75.5%**

Table 11: 1000 Sample Analysis with Translation Text Cleaning Results. Data Translate column is the sentiment analysis accuracy percentage for each language after the text was translated to English, cleaned, and evaluated with the VADER lexicon. Data Original column is the sentiment analysis accuracy percentage for each language evaluated with the original language model detailed in Table 10.

As for the ground truth based on human annotators shown in Table 12, the annotators found wide discrepancies between the original data sets' labels and what they believed to be an accurate assessment of the sentiment of the text. Four of the languages had a not-matched rate of over 60%. To combat this, I reran the analysis were the original data labels and the human annotators labels agreed. I found that, except for Spanish and Portuguese, the data followed the same pattern that the original language models produced a higher accuracy translation than the translation models.

Language	**Accuracy Translate**	**Accuracy Original**	**Percentage Not Matched**	**Percentage Matched**
English	N/A	65.4%	6.6%	93.4%
Spanish	**71.3%**	70.3%	3.5%	96.5%
French	81.9%	**94.6%**	62.9%	37.1%
Portuguese	**77.7%**	71.6%	5.3%	94.7%
Arabic	72.9%	**90.8%**	63.1%	36.9%
Japanese	75.3%	**76.4%**	62.7%	37.3%
Korean	68.8%	**73.1%**	64.7%	35.3%

Table 12: Percentage Accuracy with Human Evaluators on the 1000 Sample Results

Section 3.2.5: Discussion

These results show that original language models, in general, will produce more accurate sentiment analysis over the translated text models. There could be several reasons for this. One major reason is that the translation and subsequent cleaning of the English texts removes too much context from the original text, thus the VADER lexicon is not able to get an accurate evaluation, distorting the results. To test this, I reran the model removing the data cleaning step, and found that the accuracy only increased by 0.6%. Thus, the issue could lie with the VADER lexicon itself, as the lexicon might have a more difficult time classifying more ambiguous text data. This idea appears supported by the results of the human annotators.

While the original 1000 samples only labeled the data as positive or negative, the human annotators were given a greater degree of latitude and could label data as neutral or ambiguous as well. The latitude explains why many languages had an under 40% agreement between the original data labels and the human annotators. When I reran the models again with only the data where the labels agreed (either both positive or both negative), I saw a noticeable jump in the accuracy for the translation / VADER models to a 74.65% accuracy on average, a 20% increase over the original results. At the same time, the original language models' average accuracy increased to 79.5%, which was only a 5% improvement. This shows that with more clearly defined positive or negative data, developed from "human validated gold standard" (Hutto and Gilbert, 2014, Pg. 3) data, the translation / VADER models can obtain a sentiment analysis accuracy comparable to the original language models on more ambiguous data.

Overall, the original language models still have more accurate sentiment analysis results, except for Spanish and Portuguese, where the translation / VADER models did win out over the Spanish and Portuguese language RNN model. To make sure of this result, I also created lexicons for Spanish and Portuguese, similar to the Korean lexicon to see if there was a potential improvement in accuracy. However, I obtained worse results than the RNNs. Additionally, at the time of this research, individual Transformer or BERT models for Portuguese or Spanish for sentiment analysis did not have consistently reliable results, thus the RNN was the model type I chose for the analysis. The reason for these outliers in the overall pattern could arise from a few places. For example, it might be one of the issues identified earlier when working with non-English language data, that the was not enough data to build an RNN with a high transferability. Thus, the model can obtain a high accuracy on the test data, but when used on a different dataset, the accuracy decreases significantly.

When coupled with the increase in accuracy with the "gold standard" data for the translation / VADER model, the Spanish and Portuguese RNNs' performances were worse. However, I believe, in context of the other results, and that Spanish has nearly the same accuracy for both modelling techniques, that these outliers are most likely due to a modelling issue, rather than an indication of a general trend. Interestingly, Cheng and Zhulyn [123] also had issues with their Portuguese sentiment analysis classifier which arose from a lack of Portuguese documents (Cheng, et al, 2012, Pg. 11). Mozetič, et al., [314] also found that their Portuguese model obtained one of the lowest accuracies as well (Mozetič, et al., 2016, Pg. 19). Thus, perhaps there is something about Portuguese that causes issues with machine learning classifiers.

Section 3.3: Machine Vs Machine Vs Machine

This section explores how machine translation techniques compare to each other and to original languages machine learning methods for short text sentiment analysis. Like the previous section, this section was very important to my thesis as finding out the best way to do multilingual sentiment analysis was key to creating my geopolitical risk analysis program.

Section 3.3.1: Introduction

For this subsequent analysis, I expanded the evaluation cross section. Translation of text can be an effective way to understand text coming from multiple languages, however, which translation method to use is key to obtaining the best results. While Google Translate is convenient and reliable, in my previous research I did not have another translation system to benchmark the results of the translations against. Thus, I included a second translation technique for this study. Additionally, I constructed Recurrent Neural Networks ("RNNs") for each of the languages used and found BERT models trained for sentiment analysis for each language. Also, an additional BERT model (multilingual BERT model) was used that was trained on all the languages of this study.

My aim with this increased scope is to achieve an even better understanding of sentiment analysis across multiple languages. As these languages all have different words and grammar, and in some cases, even alphabets, thus I expect that different machine learning techniques might have different levels of accuracy for sentiment analysis.

The rest of this section is broken down into the following sections: the Methodology and the Data that was employed in this paper. Afterwards, I detail the Results with six tables breaking down the experiments. And finally describe the findings in the Discussion section.

Section 3.3.2: Methodology and Data

Methodology:

Building off the previous work in this section, I expanded the scope to investigate multiple avenues into potential improving the results for multilingual sentiment analysis. First, I wanted to determine how well the Google Neural Machine Translator works compared to a Transformer translator model. Would the GNMT produce a more accurate translation than separate Transformer models trained on each language? My second question involved the evaluation for the best English sentiment analysis method between the VADER lexicon, an RNN trained on English data, and a BERT model fined tuned for English Sentiment Analysis. Lastly, I assessed various sentiment analysis models trained on the original languages, and how these models compare to the translated English model results.

To appraise my models' effectiveness, I took a sample of 1000 short texts labeled positive or negative from each language, which I refer to as "Full Data". This data was separate from the training data used to develop to gauge the transferability of the models. Additionally, following Hutto and Gilbert [238], I also engaged human annotators to review these same 1000 short texts and label them positive, negative, neutral, or ambiguous. The data that where the original labels and the human annotators agree were kept and referred to as “Gold Data”. These two data sets are the same data as the previous section.

For the first two questions I ran four tests across different intersections between data and translators. I had the both the “Full Data” and “Gold Data” for each data translated using the GNMT and a transformer model developed for each language, the OPUS-MT [417] (the only exception to this is the Portuguese translation where the “Romance Language” version, which includes Portuguese, was used as there was no specific Portuguese transformer translation model). Each of the translations were evaluated using the VADER lexicon, the English RNN I created, and the English BERT sentiment analysis model [379]. In previous section, I found that the VADER lexicon produced good results however, as Grimmer and Stewart state: “when dictionaries are created in one substantive area and then applied to another, serious errors can occur” (Grimmer and Stewart, 2013, Pg. 8) [208]. While the VADER lexicon was developed for short text like my data is, there might exist other methods that would produce better accuracy for English sentiment analysis. Thus, I created an English RNN, detailed in Table 1 of Appendix E, which has showed to produce good results for sentiment analysis [309] and an English BERT model, which when fine-tuned for sentiment analysis produces high levels of accuracy (The English BERT specifications as well as the other language BERT model specifications are shown in Table 9 of Appendix E).

For the third question, I wanted to examine if I applied these same modelling methods to the untranslated data, would I produce accuracy results that surpassed the results from the translated English data. I created six additional RNN models, one for each language, with their specifications detailed in Tables 1 – 8 in Appendix E. Additionally, I found six BERT models for each language on HuggingFace.co fine-tuned for sentiment analysis [109, 155, 240, 261, 325, 457]. Lastly, McCormick [304] detailed how a BERT model trained on multiple languages at the same time could produce accuracy results that are comparable or even surpass those of models only trained on one language. Thus, I found another BERT model [71] trained on tweets from 30 languages and fine-tuned for sentiment analysis. I compared the results of these models for the “Full Data” and the “Gold Data” to the results from the translated English model results. With this, I can find the best method to use for each of the languages.

Data:

The data sources for my research come from a wide variety of previous research and models. I divided the sources into three sections: the data used to create the RNNs, the pretrained BERT models, and used the data used for the testing.

For the data used to create the RNNs, I made sure that all the data was balanced between positive and negative labeled data. The English data came from the IMDB Movie data [278] which had 50,000 data points. The Spanish RNN was the same RNN from the machine vs machine section and had combined had 5,764 data points. The French RNN was modified and trained on Gamebusterz data [187, 189] and the TheophileBlard [88] movie review data for a total of 415,702 data points. The Portuguese RNN was generated through the “No Theme” Tweets created for a Kaggle competition by Augustop [64] and the “utlc_movies” tweets created by Dias for a separate Kaggle competition [157], I used 100,000 data points from these combined datasets. The Arabic data for the RNN came from Motazsaad [313] which had 102,196 datapoints. The Japanese RNN was the same as the machine vs machine section and had 20,000 datapoints. Lastly, the Korean data came from the “testing data” created by Park [331] for the NSMC Korean movie review dataset which had 50,000 datapoints.

The pre-trained BERT models were found on HuggingFace.co. The English BERT Sentiment Analysis Model was created by sbcBI [379], while the Spanish BERT was created by Cañete, et al. [108] using the same TASS dataset that I used to build the Spanish RNN. For the French model, I chose the bert-base-multilingual-uncased-sentiment created by nlptown [325] where French was one of the six languages used to train the model, as I could not identify an individual French BERT model trained for sentiment. As for Portuguese, I used the FinBertPTBR model [109] created by turing-usp, this model was trained on Brazilian Portuguese financial data which was not exactly close to the base data, but it was the only Portuguese only BERT model trained to evaluate sentiment. For Arabic, the CamelBERT model developed by Inoue, et al. []240] was used again. For Japanese, I used the model created by ydaigo [457]. For Korean, like Portuguese, I chose a Korean BERT model trained on financial data created by snunlp [261], for similar reasons as the Portuguese choice. Lastly, I used the model developed by Barbieri, et al. [71] which was developed from tweets of multiple languages for the multilingual BERT tests.

Section 3.3.3: Results

Below, are the results from the Gold Data experiments with Table 13 and Table 14 showing the translators differences, and Table 15 displays the accuracy using the Original Language techniques. The bold percentages represent the best accuracy out of the table, the yellow highlights represent the best results for each data set (Gold Data and Full Data).

Language	**English VADER**	**English RNN**	**English BERT**
English	**77.3%**	70.3%	71.3%
Spanish	70.3%	80.4%	**80.5%**
French	**81.9%**	70.6%	80.3%
Portuguese	77.2%	73.1%	**77.8%**
Arabic	**71.5%**	68.0%	66.9%
Japanese	**68.9%**	63.0%	33.8%
Korean	62.9%	**65.2%**	55.8%
Non-English Average	**70.0%**	**70.0%**	65.9%

Table 13: Gold Data – Hugging Face Transformer Translator

Language	**English VADER**	**English RNN**	**English BERT**
English	**77.3%**	70.3%	71.3%
Spanish	71.5%	79.9%	**81.8%**
French	**81.6%**	74.7%	81.1%
Portuguese	**77.6%**	75.0%	76.3%
Arabic	**73.1%**	66.9%	**73.1%**
Japanese	**75.6%**	60.9%	32.7%
Korean	68.8%	**74.5%**	62.6%
Non-English Average	**74.7%**	71.3%	67.9%

Table 14: Gold Data – Google Neural Machine Translator

Language	Original RNN	Original BERT	Multilingual BERT
English	70.3%	71.3%	**92.5%**
Spanish	**73.9%**	63.4%	58.5%
French	**94.6%**	77.6%	**94.6%**
Portuguese	**75.5%**	24.0%	73.4%
Arabic	72.6%	**85.4%**	81.6%
Japanese	72.1%	**83.4%**	21.7%
Korean	**86.4%**	3.1%	76.8%
Non-English Average	**79.2%**	56.2%	67.8%

Table 15: Gold Data – Original Language Techniques

I found that across the four methods used to evaluate English accuracy, I found the Multilingual BERT produced the highest accuracy at 92.5%, followed by VADER, English-only BERT, and the English RNN. To answer the first research question, I found that for Gold Data, the GNMT produced more accurate results than the individual transformer translator models (average 4.7% more accurate for VADER, 1.3% for RNN, and 2.0% for BERT). That said, there were a few instances where the transformer translators did produce more accurate results, but the English RNN on the Japanese to English translation was the only category where the difference was more than 2.0%. Additionally, I discovered that with the GNMT, most languages had VADER as the most accurate, except for Spanish, which had the BERT with the highest accuracy, and Korean which had the RNN as the highest, and Arabic which tied between VADER and BERT. As for transformer translations, the results were similar, except that VADER is better for Arabic, and BERT is better for Portuguese.

For the second research question, I observed that across the four methods used to evaluate English accuracy, we found the Multilingual BERT produced the highest accuracy at 92.5%, followed by VADER, English-only BERT, and the English RNN.

Lastly, for the third question, I uncovered that for the Original RNN had the highest accuracy on average at 79.2%. However, the average for the one-language BERT models and the multilingual BERT model had significantly lower accuracy due to outlier results in Portuguese and Korean, and Japanese, respectively. Overall, the original language techniques had the highest number of most accurate techniques, with those for French, Arabic, Japanese, and Korean, while Spanish and Portuguese had the most accuracy when translated with the GNMT. Contrary to what McCormick [304] found, only English had the best accuracy with the multilingual model, while French multilingual model analysis tied the French RNN results. When looking at just the original language techniques for non-English languages, the RNNs produced both results with the lowest variability across languages. Except for Arabic and Japanese, the RNNs also had the highest accuracy for each language. The yellow highlights across Tables 13 – 15 indicate the highest accuracy for each category.

Tables 16 – 18 show the results for the experiments using the Full Data below.

Language	**English VADER**	**English RNN**	**English BERT**
English	**75.4%**	66.1%	68.5%
Spanish	68.5%	**79.1%**	78.8%
French	**64.4%**	58.7%	55.0%
Portuguese	74.4%	63.1%	**75.2%**
Arabic	**74.0%**	66.3%	66.9%
Japanese	**62.1%**	55.0%	52.4%
Korean	62.6%	**67.7%**	56.1%
Non-English Average	**67.7%**	65.0%	64.1%

Table 16: Full Data – Hugging Face Transformer Translator

Language	**English VADER**	**English RNN**	**English BERT**
English	**75.4%**	66.1%	68.5%
Spanish	69.6%	78.8%	**80.2%**
French	**64.8%**	58.4%	57.2%
Portuguese	**74.9%**	66.2%	73.7%
Arabic	**74.7%**	66.5%	72.8%
Japanese	**67.0%**	57.2%	53.3%
Korean	64.9%	**71.3%**	61.8%
Non-English Average	**69.3%**	66.4%	66.5%

Table 17: Full Data – Google Neural Machine Translator

Language	**Original RNN**	**Original BERT**	**Multilingual BERT**
English	66.1%	68.5%	**80.8%**
Spanish	**71.7%**	61.4%	56.6%
French	**81.2%**	58.0%	65.1%
Portuguese	**73.8%**	22.9%	70.1%
Arabic	72.6%	**84.6%**	79.1%
Japanese	66.8%	**68.9%**	22.0%
Korean	**82.9%**	4.0%	72.5%
Non-English Average	**74.8%**	50.0%	60.9%

Table 18: Full Data – Original Language Techniques

As the Full Data has more variability in it, the accuracy results were generally lower across all techniques and models. The GNMT again produced more accurate results on average just a lower gap in the difference (1.6% in VADER, 1.4% in RNN, 2.4% in BERT). For the "Full Data", only Portuguese had the transformer translator produce a most accurate result for Portuguese, across all the experiments, which was the only instance where the transformer translator method produced the most accurate results. Additionally, unlike with the Gold Data, the Portuguese to English BERT transformer translation is the only transformer translation that beats the GNMT by more than 0.5%. As it was with the Gold Data, VADER also produces the most accurate results on average for the Full Data, with the same deviations, except for Arabic in GNMT which VADER performed better than the BERT model. As for the original language models, the RNNs performed better than translated models. Again, the one-language BERT models and the multilingual BERT model average accuracy results were dragged down by the outliers.

Section 3.3.4: Discussion

This study reinforces the findings from the previous section into multilingual sentiment analysis, that analyzing the text in its original language produces more accurate results than translating to English then implementing an English based lexicon or model. On its face, an RNN model, on average, is far and ahead the best original language model to use, however, there are a few outliers to consider which greatly reduced the averages of the individual language BERT models and the multilingual BERT model. One issue was the Portuguese and Korean BERT models were trained on financial data, which can be more specific, and potentially vary from regular or short text Portuguese and Korean. Unfortunately, these were the only one-language BERT models I could find for these languages. Thus, once I removed these outliers, I found that the accuracy for the Gold Data for individual BERT models rose to 77.5%, while Full Data rose to 68.2%. Still not higher than the RNN's average but significantly closer.

This was also true for the multilingual BERT model. It was trained on over 30 languages based on Twitter / X data [71], creating an effective multilingual BERT model for similar data to what I was studying, however, there was still an outlier with Japanese. Although, the overall model was trained on Japanese, its sentiment analysis fine-tuning did not include it (Barbieri, et al, 2021, Pg. 5), thus potentially causing these results. However, once Japanese is removed, the average for multilingual BERT increased to 77.0% for Gold Data, and 68.9% for Full Data, similar to both the RNNs and the individual BERT models sans outliers.

The only outlier to this general trend was in the Portuguese experiment with the Full Data where the English BERT model with the OPUS-MT translation performed the best across both the translated and original language models. While it was only slightly better than the VADER lexicon with GNMT, it is still important to note. I believe that since for Portuguese, there was no single language Portuguese to English translation model, I used the multilingual romance language translation model which included 48 languages and dialects. This increase in language samples could have improved the translation results described by McCormick [304] leading to the out-performance of Portuguese in this case. I believe that future research into is required to fully answer this question, however.

While, on average, the original language models performed better, Spanish and Portuguese slightly underperformed the translated English models. There are a few potential reasons for this. For Spanish, there was a lack of high-quality sentiment analysis data to build the RNN model. While it was still effective for Spanish data, an English model trained on more data, such as the English BERT model, could explain the increased accuracy. As for Portuguese, the previous section identified quirks with the Portuguese analyses, thus, I was not surprised that a Portuguese to English translated model performed marginally better than an original language model.

Lastly, I wanted to examine why the RNNs performed more consistently than the BERT models. While Arabic and Japanese BERT models were the most accurate for those languages and the multilingual BERT model had the best results for English, the BERT models had widely varying accuracies across languages. I determined that this difference in consistency could potentially arise from the testing data and the BERT models themselves. The BERT models that performed better could have been trained on short text data, similar to the testing data, while the ones that performed worse, had data that was too different or did not include short texts. Additionally, the Arabic BERT model was fine-tuned on more data than the other base BERT models, which would lead to the increased accuracy. The RNNs, on the other hand, were all trained on short texts, like the testing data, leading to the more consistent results I found.

Section 3.3.5: Conclusion

126 experiments were created from the cross-section of the different sentiment analysis methods with both the Full Data and the Gold Data, the two different translation techniques and original language methods. I found that most languages had the best sentiment analysis results when an original language model was used. For both Full Data and Gold Data, these were original language RNNs for French and Korean, original one-language BERT models for Arabic and Japanese, and the multilingual BERT model for English. However, for the Gold Data, Spanish and Portuguese had the best results with the GNMT and an English based model (BERT and VADER, respectively). Also, for the Full Data, only Spanish had the best result for the GNMT with English-based BERT model, while for Portuguese, the OPUS-MT provided the best accuracy with the English-based BERT model as well.

Overall, my results in this section matched the Machine vs Machine section, the most accurate results for a multilingual short text sentiment analysis study would come from using original language machine learning techniques on Gold Data which the data that been reviewed by a human annotator. However, the same relative accuracy achieved with the Full Data, i.e., the more ambiguous data, on original language techniques as the Gold Data that has been translated using the GNMT and then analyzed with the VADER lexicon. Thus, a research study with high quality data can feel confident use the GNMT for sentiment analysis, but a study with lower quality data should use original language techniques to analyze sentiment to obtain the best results.

Overall, this chapter briefly outlines the benefits and disadvantages in working with either method for sentiment analysis. However, with this research, I aim to provide insight into this area as multilingual sentiment analysis grows across our ever-increasing interconnected world. Determining the best techniques to use will provide invaluable help to future researchers looking to further text analysis research. These results also proved invaluable in the construction of my geopolitical risk analysis program. As the main focal point to automatically evaluate geopolitical risk is based on sentiment analysis, it was important for me to exhaust all possible combinations of multilingual sentiment analysis to obtain the best possible methods to get the highest accuracy for my research.

Chapter 4: Case Study 1

Multilingual Twitter / X Sentiment Analysis of Geopolitical Risk using Granger Causality Analysis Focusing on the Ukraine War and its Effect on Financial Assets and Markets

For my thesis, I wanted to test the capabilities on the programs I created to evaluate geopolitical risk. Thus I decided on analyzing three case studies that would best reflect and aim to answer my initial research questions. While the previous chapters focused on evaluating the current methods that could be used to monitor and analyze geopolitical risk and the NLP advances, these case studies show the application of these methods. The first case study is a test to see if my programs could capture a major historical geopolitical event. As the Ukraine War was the biggest geopolitical event that occurred at the time of my research, I decided to collected six months of data from around the start of the War (this process is detailed further in this chapter). This was done to try and find the answer to my research question five about finding best predictive performance with geopolitical risk, thus I took larger time interval of the day level to build a geopolitical risk index similar to Caldara and Iacoviello [105]. I wanted to see if it was even possible to use generic geopolitical keywords to identify the events through using just Twitter / X alone. Additionally, with the multilingual sentiment analysis experiments conducted in Chapter 3, I applied these to the data analysis in my program for the geopolitical risk tweets to build my index.

Section 4.1: Introduction

On February 24th, 2022, Russia launched an invasion of Ukraine formally starting the Ukraine War. However, this invasion had been telegraphed months ahead of time, and contrary to Russian expectation of a short conflict, the

Ukraine War has continued up to the time of writing this thesis, more than a year later. This war represents one of the most, if not the most, important increase in geopolitical risk in our world today. But what is geopolitical risk? Restating the definition from Caldara and Iacoviello [105]: “the risk associated with wars, terrorist acts, and tensions between states that affect the normal and peaceful course of international relations. Geopolitical risk captures both the risk that these events materialize, and the new risks associated with an escalation of existing events” (Caldara and Iacoviello, 2022, Pg. 2). This definition effectively describes the Ukraine War, and as Caldara and Iacoviello show, geopolitical risks have an impact on various financial markets and assets (Caldara and Iacoviello, 2022, Pg. 1). However, getting up to date information of geopolitical events on a large scale can potentially be time consuming (Caldara and Iacoviello, 2022, Pg. 10-13). Thus, I wanted to test if there was a low cost, quicker way to evaluate geopolitical risk, thus I turned to social media, specifically Twitter / X.

For research purposes, Pak and Paroubek [328] best outlined why Twitter / X is an effective resource:

> “Microblogging platforms are used by different people to express their opinion about different topics, thus it is a valuable source of people’s opinions. Twitter contains an enormous number of text posts and it grows every day... Twitter’s audience varies... Therefore, it is possible to collect text posts of users from different social and interests groups. Twitter’s audience is represented by users from many countries. . . .it is possible to collect data in different languages” (Pak and Paroubek, 2010, Pg. 1).

Their last point was especially important for this case study as I was aiming to track worldwide sentiment, thus I needed text that came in multiple languages. Vicinitas states that English language tweets only comprise 30% of tweets posted [1], this means that I would miss out on a significant portion of all tweets posted if I only looked at English language tweets. However, if you include Japanese, Spanish, French, Portuguese, Arabic, and Korean in addition to English, I would obtain approximately 85% – 90% of all tweets posted to Twitter / X [1]. Thus, using all these languages gave me a larger corpus of tweets and a better sense of the sentiment surrounding the Ukraine War's associated geopolitical risks.

The rest of this case study is outlined as follows: Section 2 describes Key Concepts employed for these analyses. Section 3 provides a Related Work section of previous research into geopolitical risk, media, and social media, and how they can affect financial markets. Section 4 details the Methodology, while Section 5 displays the Results, and Section 6 discusses the findings.

Section 4.2: Key Concepts

There are three key concepts that make up the backbone of the research: the Goldstein Index, Sentiment Analysis, and Granger Causality. However, as Sentiment Analysis was defined earlier in Chapter 3, I will describe Goldstein Index and Granger Causality here.

Section 4.2.1: Goldstein Index

The "Goldstein Index" comes from the 1992 paper "A Conflict – Cooperation Scale for WEIS Events Data" by Goldstein [201]. The World Events Interaction Survey ("WEIS") data set was developed by McClelland [303], which is a "a record of the flow of action and response between countries (as well as non-governmental actors, e.g., NATO) reflected in public events reported daily in the New York Times from January 1966 through December 1978" (McClelland 2006) [303]). The individual WEIS events can be grouped into "61 event types" (Goldstein, 1992, Pg. 2). Goldstein then created a panel of eight International Relations faculty at the University of Southern California ("USC") to analyze and score the WEIS events (Goldstein, 1992, Pg. 6). This panel was individually given 61 cards with each WEIS event type and asked to "sort the cards into cooperative (friendly) actions and conflictual (hostile) ones" (Goldstein, 1992, Pg. 7) and rank them on a scale with -10 as the most conflictual and +10 as the most cooperative. The resulting rankings from each of the panel members was then averaged, creating what I refer to as the "Goldstein Index" which is a table of all 61 WEIS event types ordered from most conflictual to most cooperative. This table is the basis for my data gathering procedure further described in the Methodology section. One potential bias of the "Goldstein Index" to note is mentioned by Goldstein as at the time of his writing the table "seems to reflect the continuing emphasis placed on military affairs by international relations scholars" (Goldstein, 1992, Pg. 9). However, this bias is not a great concern as my research that made the most use of the "Goldstein Index" revolves around the Ukraine War, which is a military affair.

I used the "Goldstein Index" as the initial basis for my data gathering for this case study. Many studies that worked with Twitter / X [33, 68, 83] have made use of keywords to collect tweets through the Twitter / X API and so I decided to follow these methodologies. However, the "Goldstein Index" does not fit neatly into the Twitter / X API framework, as shown below in Table 19:

New Weights for WEIS Events		
Event Type	**Weight**	**SD**
223 Military attack; clash; assault	-10.0	0.0
211 Seize position or possessions	-9.2	0.7
222 Nonmilitary destruction / injury	-8.7	0.5
221 Noninjury destructive action	-8.3	0.6
182 Armed force mobilization, exercise, display; military buildup	-7.6	1.2
195 Break diplomatic relations	-7.0	1.3
173 Threat with force specified	-7.0	1.1
174 Ultimatum; threat with negative sanction and time limit	-6.9	1.4

Table 19: A recreation of portion from the Goldstein Paper [201] showing the table Goldstein created. As it can be seen, many of the phrases Goldstein uses would not work with the Twitter / X API as they are too long or awkward.

To get around this issue, I split the phrases in the index into single terms (such as “attack”, “clash”, “assault”) and bigrams (two term phrases such as “military attack”, “military clash”, “military assault”). These smaller phrases are more manageable for the Twitter / X API which allowed us to collect more data. Through experimentation, I found that while the single terms gathered more data, these tweets dealt with a wide variety of issues rather than the geopolitical tweets I was searching for. The bigrams, on the other hand, did provide a better corpus of tweets for geopolitics, just fewer of them. Thus, I chose to use bigrams for this research, as their increased precision over amount of data collected was more valuable for my research which can found in Appendix A.

Section 4.2.2: Granger Causality

First detailed by Granger in his 1969 paper [205], Granger causality aims to find the "the direction of causality between two related variables and also whether or not feedback is occurring" (Granger, 1969, Pg. 1). Since then, Granger causality tests have been used in a wide variety of studies including Thurman and Fisher study where they aimed to predict whether eggs Granger cause chickens or chickens Granger cause eggs [414]. However, it should be noted that causality in this case does not mean typical definition of causality, i.e., a change in one variable causes the change in another, but rather as Gilbert and Karahalios [197] put it: "Although the technique has the word "causal" in it, we clearly aren't testing true causation. We can only say whether one time series has information about another" (Gilbert and Karahalios, 2010, Pg. 4). And as Granger himself states, his definition of causality mentions that "1. The cause occurs before the effect; and 2. The cause contains information about the effect that that is unique, and is in no other variable. A consequence of these statements is that the causal variable can help forecast the effect variable after other data has first been used." (Granger, 2003, Pg. 6) [206]. Thus, the null hypothesis for Granger Causality Test is that the two time series are not related in anyway or provide any predictive information about each other. While the alternative hypothesis, which is accepted at a p-value less than 0.05, is that one tested time series does provided predictive information about the other time series. For this case study, I followed the lead of Bollen, et al., who used Granger causality to test "whether one time series has predictive information about the other or not" (Bollen, et al., 2011, Pg. 4) [91]. I chose to use Granger causality over traditional correlation to examine the relationship between the change in the sum of sentiment trend and the financial asset as traditional correlation tests for a linear relationship between the variables, in other words, it checks to see if the variables change together with a constant rate. Granger causality works better for my research as it is for testing if one series contains predictive information about the other, i.e. if one trend moves, does the other also move in the future. Since social media news reacts faster than the financial markets change their prices, the two time series will have a lag between them and not vary at the same time, thus Granger causality is the better statistical test for this case study.

Section 4.3: Related Work

Multitudes of studies make use of sentiment analysis especially with Twitter / X. For example, Pak and Paroubek showed in their study how to effectively use Twitter / X to construct a corpus of tweets and use sentiment analysis on those tweets to derive insights [328]. Additionally, Rajput, et al, used Twitter / X to analyze sentiment analysis around the Coronavirus pandemic [351]. Baker, et al. used Twitter / X to "construct a database of more than 14 million tweets that contain a keyword related to 'uncertainty'... from June 1st, 2011, and March 1st, 2021" (Baker, et al., 2021, Pgs. 2-3) [68]. They transformed the count of these tweets into a time series and used that time series as a measure of economic uncertainty in the US during their research period. The Baker study was important to my research as their use of keywords also provided a basis for my use of keywords to gather data for the analysis with the "Goldstein Index".

Many papers have also explored the relationship between news media and the effect on various financial markets through sentiment analysis. Using sentiment analysis to key in on anxiety related terms in a large online blog LiveJournal, Gilbert and Karahalios found through using Granger causality analysis that "increases in expressions of anxiety... predict downward pressure on the S&P 500 index" (Gilbert and Karahalios, 2010, Pg. 1) [197]. Uhl also showed that using a corpus of Reuters news articles that the sentiment analysis of those articles over time could "predict changes in stock returns better than macroeconomic factors." (Uhl, 2014, Pg. 1) [410]. Tetlock, et al., 2008, [431] took a more expansive approach to the returns of specific firms in the S&P 500 index by using sentiment analysis on articles from the *Wall Street Journal* and the *Dow Jones News Service* from 1980 to 2004 (Tetlock, et al., 2008 Pg. 2) to show that the number of negative words used in the articles about the firms can forecast lower earnings for the firms.

However, three papers were highly influential to my research: Bollen, et al., 2011 [91], Amen, 2020 [55], and Caldara and Iacoviello, 2022 [105]. Bollen, et al., provided a framework about how to work with Granger causality and Twitter / X. Caldara and Iacoviello, and Amen provided the theoretical basis about working with trends in geopolitical risk data and the assets I should investigate that might be affected by geopolitical events, such as the Ukraine War. Bollen et al., researched whether the change in moods and the change in the Dow Jones index are linked. To do so, they complied a corpus of tweets containing "author's mood states" (Bollen, et al., 2011, Pg. 2) and analyzed them through sentiment analysis programs to identify six moods: "*Calm*, *Alert*, *Sure*, *Vital*, *Kind*, and *Happy*" (Bollen, et al., 2011, Pg. 2). Creating a time series from the tweets' sentiment, Bollen, et al. then used Granger causality to find if the change in mood sentiment that predates a change in the Dow Jones index. They found that out of the six moods, only *Calm* passed the Granger causality test and had information that predicated the change in the Dow Jones from 2 – 6 lags (Bollen, et al., 2011, Pg. 4 -5).

Caldara and Iacoviello, and Amen, on the other hand, focused specifically on geopolitical risk. Caldara and Iacoviello built the Geopolitical Risk (GPR) Index, which used the count of news articles that mentioned their keyword indicators for geopolitical risks across 11 different English language newspapers starting from 1985 (Caldara and Iacoviello, 2022, Pg. 7). This GPR Index captured the changes in geopolitical risks, and Caldara and Iacoviello were able to show how the increases in GPR index predicted lower stock returns (Caldara and Iacoviello, 2022, Pg. 1). Lastly, Amen built the Thorfinn Sensitivity Index ("TSI"), which uses "over 30,000 daily feeds" (Amen, 2020, Pg. 2) to construct a daily index of the weighted average of 12 geopolitical risk groups which experts have scored based on the news feeds that have come in for that day (Amen, 2020, Pg. 2). Amen then compared the changes in the TSI to changes in various "safe havens" and "risky assets" (Amen, 2020, Pg. 6) to develop trading strategies. Caldara and Iacoviello, and Amen had a wide range of assets and markets that provided a starting point for my analyses. Appendix F contains Table 20 which shows the different financial assets I considered and their sources, while Table 21 regroups into the asset class categories I used.

My aim is to extend the literature by combining the Goldstein Index with Twitter / X to see if I can possibly capture large geopolitical events, such as the Ukraine War and see how the sentiment around a geopolitical event can affect different financial assets on an equivalent or smaller time scale than both Caldara and Iacoviello, whose index captures both daily and monthly data, and Amen, who's index is only for daily. Additionally, I intend to capture the global impact of a geopolitical event by using multiple languages. While many studies only look at English tweets [68, 328], or perhaps one additional language like Italian for Pota, et al. [340] or Dutch for Kleinnijenhuis [264], my goal is to capture a more expansive, worldwide sentiment using the seven languages I am studying.

Section 4.4: Methodology

The methodology for this case consists of three parts. The first is the data gathering, the second is the sentiment analysis, and the last is financial market analysis with Granger causality.

Section 4.4.1: Data Gathering

I gathered the tweets from December 1st, 2021, to April 30th, 2022, which was the timeframe around the start of the Ukraine War. To do so, I used "Twarc" [21], which collects and stores tweets from the Twitter / X API[5] from specific time periods that used keywords such as "Goldstein Index" bigrams whose creation was described in Section 5.2.1. I accessed this data through the Academic Track provided by X / Twitter and collected this data in January 2023, before X / Twitter's Academic Track Rule change which limited the number of tweets researchers could obtain. I chose to use the top ten most negatively and positively weighted bigrams that returned a non-zero number of tweets where the majority of tweets focused on geopolitics for my research. For example: while "call truce" ranked below "policy support" (2.9 and 4.5, respectively) many of the tweets I obtained for "policy support" focused more on internal politics rather than geopolitics than the "call truce" tweets, thus "call truce" was used. I also removed tweet duplicates by the "text" variable and the "created_at" variable generated from the Twitter / X API from my tweet data, as I viewed anything retweet within the same second after the original posting was most likely a bot. We chose this removal method over removing all retweets in general, because of what a retweet represents. With retweeting, the user does not write the original tweet, they only share the tweet. However, this sharing also represents the opinion of the retweeting user, thus I decided to include all retweets for this study. I only removed the tweets with the same text and time stamp as described earlier, as it would be impossible for a human user to retweet a tweet within the fraction of a second that the original was posted. However, this removal did not cause significant data loss. With this method, I collected over 3.6 million tweets for my research period. After collecting the tweets, I moved onto the sentiment analysis

[5] https://developer.twitter.com/en/docs/twitter-api/tweets/filtered-stream/introduction

of the tweets' text.

Section 4.4.2: Sentiment Analysis

I used three different methods for the sentiment analysis that had the best trade off for processing speed and accuracy. For English, I implemented the VADER Lexicon developed by Hutto and Gilbert [238]. VADER is a rules-based sentiment analysis lexicon that is highly accurate especially on short English texts (96%), such as tweets (Hutto and Gilbert, 2014, Pg. 9). For Arabic, I turned to CamelBERT developed by Inoue, et al. [240]. Based off the BERT model [155], a multi-layer Transformer based model used for various natural language processing tasks, CamelBERT achieves the same functionality as BERT, including accurate sentiment analysis results for Arabic. However, for Spanish, French, Portuguese, Japanese, and Korean, I could not find any pre-trained models or lexicons that suited my needs, thus I created Recurrent Neural Networks ("RNNs"), described in Chapter 3, to obtain the sentiment for the tweets [194]. RNNs were effective in this task because they can remember and analyze short sequences, such as a tweet, and return a sentiment analysis score. I was able to find sentiment analysis short text datasets for each language and thus was able to train effective RNNs.

Section 4.4.3: Financial Markets and Granger Causality

With the sentiment obtained for each tweet in the different languages, I complied the tweets into one data frame and grouped the tweets by day, attaining the sum of the sentiment for each day, thus creating a time series for the change in the sentiment by day. With this time series, I wanted to compare it to the various financial assets I mention in Appendix F. These assets were chosen as they were already proven to have sensitivity to changes in geopolitical risk [55, 105]. I obtained the daily price in USD of each financial asset I am studying for December 1st, 2021, to April 30th, 2022, through the Python package yfinance

[58] and various other websites that contained the relevant financial asset data [129, 200, 241, 242, 243, 291, 310]. However, these financial assets all traded a different times and certain regions observed certain holidays which closed the markets. Thus, I needed to alter the sum of sentiment time series so that each of the dates would line up correctly. Now with the sets of financial data time series, one must first check if the time series are stationary. Stationarity for a time series is defined as a series with no trends, it looks flat, has constant variance and autocorrelation structure, without seasonality fluctuations (National Institute of Standards, 2023) [401]. Stationary checks are important as Granger causality only works under the assumption of stationary time series [205]. Once I made the time series stationary through differencing,(where the current observation is subtracted from the previous observation, this differencing process makes a time series stationary. If differencing is not done, the Granger causality results obtained would be useless) and made sure I obtained the desired results, I was able to apply Granger causality to see if the sentiment analysis of the tweets obtained with the Goldstein Index bigrams can provide information about forecasting the change in the various financial assets. However, I did not create a Vector Autoregression (VAR) model, which focuses more trying to forecast the change of the model variables, as I was more interested in investigating the relationship between the sentiment on social media and various financial assets, thus Granger causality applied better in my case [276]. For the Granger causality model specifications I am testing to see if the financial asset is exogenous to the sum of sentiment, i.e if the sum of sentiment trend Granger causes the change in the price of the financial asset, which makes the financial asset price the endogenous variable and the change in the sum of sentiment the exogenous variable.

For the hourly time frame, the whole methodology is repeated with the only change is that for the time series, I summed the sentiment at the hourly level instead of the daily level. A diagram of the geopolitical risk work procedure described above is presented in Appendix B.

Section 4.5: Results

Below is the count of tweets for each bigram that had a negative weight ("Goldstein Negative") for each language:

Bigrams	English	Spanish	French	Portuguese	Arabic	Japanese	Korean	Total
Military Invasion	931,256	180,118	23,357	20,398	3,234	4,514	662	**1,163,539**
Military Attack	722,575	116,651	19,825	14,585	7,029	40,735	545	**921,945**
Military Clash	44,495	18,172	742	13,332	1,567	5,390	2,475	**86,173**
Military Assault	180,108	8,650	2,986	426	396	395	0	**192,961**
Seize Position	3,774	5	691	14,934	2,614	4	68	**22,090**
Seize Possession	1,460	27	4	782	478	0	1	**2,752**
Nonmilitary Destruction	7	5,813	38	1,641	66	0	0	**7,565**
Nonmilitary Injury	1	23	0	96	29	0	0	**149**
Force Mobilization	3,795	8,953	3,421	1,274	2,210	63	714	**20,430**
Force Exercise	50,138	50,117	1,650	1,263	3,959	4,339	366	**111,832**
Total	**1,937,609**	**388,529**	**52,714**	**68,731**	**21,582**	**55,440**	**4,831**	**2,529,436**

Table 22: The number of tweets the contained each Bigram from December 1st, 2021, to April 30th, 2022, for the "Goldstein Negative" Category.

For my research period, I was able to collect 2,529,436 tweets that used one of these bigrams across all 7 languages, of which 591,547 were non-English. It should be noted that I included "Military Invasion" as a bigram for this specific case study that was not included in the original Goldstein Index. While the original Goldstein Index would work in all other contexts for different events, since the Ukraine War was an invasion, leaving out the bigram synonym of Military Invasion would have missed over a million tweets from this Case Study, which would have been severely detrimental.

Table 23 shows the count of tweets for each bigram that had a positive weight ("Goldstein Positive") for each language:

Bigrams	English	Spanish	French	Portuguese	Arabic	Japanese	Korean	Total
Military Assistance	443,872	37,954	33,368	3,369	3,327	4,914	8,248	**535,052**
Economic Aid	111,168	137,250	9,997	6,552	390	55,645	2,388	**323,390**
Substantive Agreement	1,219	1,553	17,996	65	309	69	234	**21,445**
Suspend Sanctions	34,326	7,222	1,518	689	1	0	158	**43,914**
Diplomatic Recognition	16,062	4,500	493	199	183	11	202	**21,650**
Grant Privilege	9,608	1,572	204	519	1,055	3,421	4,420	**20,799**
Call Truce	55,011	1,477	233	1,347	716	7	1,899	**60,717**
Material Assistance	9,127	2,357	323	377	6,944	0	198	**19,326**
Endorse Position	7,278	6,907	1,457	81	503	3,243	15,881	**35,350**
Verbal Support	10,330	2,446	10,337	304	57	3	553	**24,030**
Total	**698,001**	**203,238**	**75,926**	**13,529**	**13,485**	**67,313**	**34,181**	**1,105,673**

Table 23: The number of tweets the contained each Bigram from December 1st, 2021, to April 30th, 2022, for the "Goldstein Positive" Category.

For the "Goldstein Positive" category I collected 1,105,673 tweets, of which 407,672 were non-English. This brings the total number of tweets during the study period to 3,635,109, (data size 1.45 GB) with nearly 70% of tweets coming in the "Goldstein Negative" Category and 27% of tweets non-English.

Figure 4 is the daily count of tweets captured by the Goldstein Index bigrams over time from December 1st, 2021, to April 30th, 2022. This represents a total of 151 days. Each day is the total count of tweets captured by each topic with both the Goldstein Negative and Goldstein Positive bigrams for that day. The count of Goldstein bigram tweets starts rising around two weeks before the start of the invasion and remained, on average, three times higher than before the War started, indicating increased discussion of geopolitical events on Twitter / X.

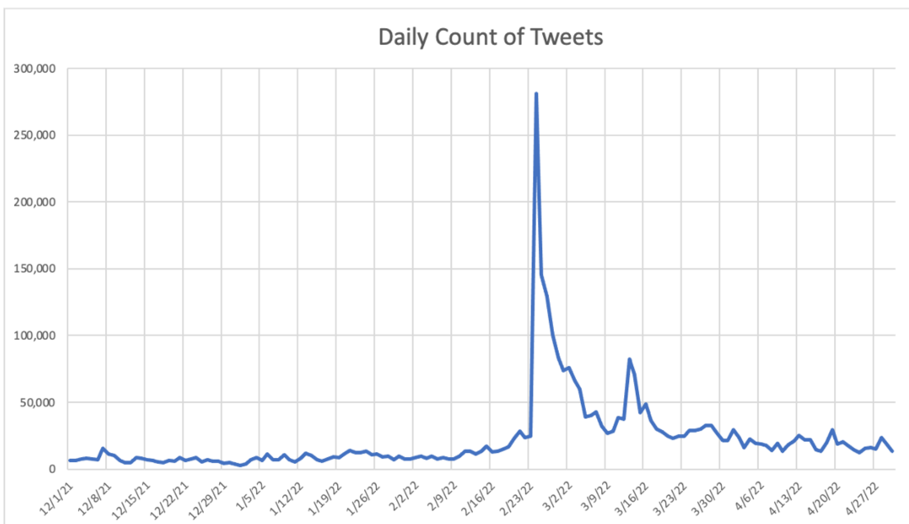


Figure 4: Daily Count of Tweets in both "Goldstein Positive" and "Goldstein Negative" Bigram Categories

Figure 5 displays the change in the Daily Sum of Sentiment, the start of the Ukraine War is on February 24th, 2022, which is evident by the large decrease in sentiment. The average sum of sentiment after the start of the War, like the count, was around three times more negative than before the start of the War. This indicates that, while more people were talking about geopolitics, they were talking about it in a more negative way than usual, which is unsurprising given the scale and the devastation the Ukraine War caused during this time.

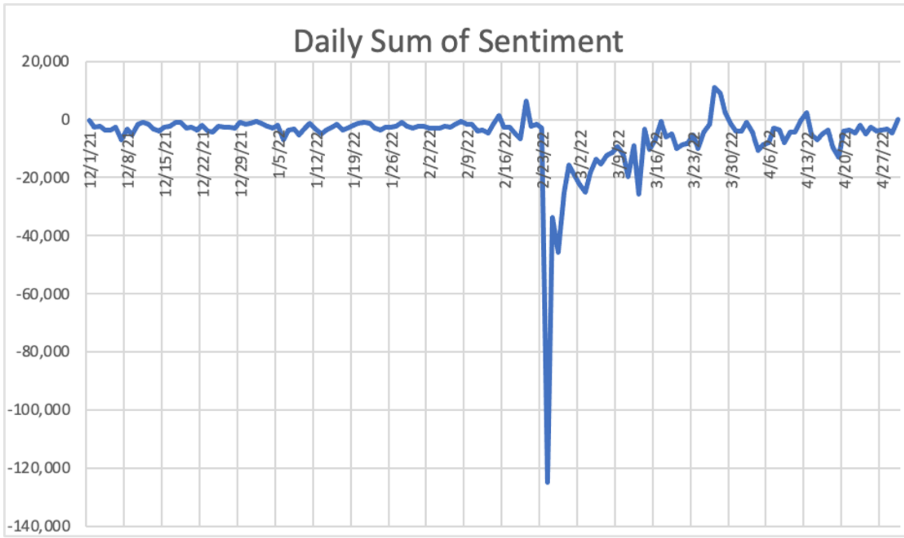


Figure 5: Daily Sum of Sentiment for all Tweets in both "Goldstein Positive" and "Goldstein Negative" Bigram Categories

Following a similar procedure to C. Pop, et al. [104], I chose to examine the Granger causality relationship between the daily sum of sentiment and each financial asset at different lagged values. In C. Pop, et al. they used "the Granger causality was considered for 1 lag, 5 lags (a typical trading week), and 20 lags (the average number of trading days within a month)" (C. Pop, et al.,2016, Pg. 132). Tables 24 – 26 show the results from the Granger Causality tests, were Table 24 shows the results of one lag (which represents one day), Table 25 is the results of five lags, and Table 26 is the results of 10 lags (which would be two weeks of trading). Note, that these tables only display the financial assets that I can say the sum of sentiment time series provides predictive information of the change in the financial asset at the lagged value. Any asset that doesn't appear in the table, but does in Appendix F, either did not have any Granger causality with the change in sum of the sentiment trend, or it did, but the "feedback"

(Granger, 1969, Pg. 5), denoted by "reverse" in my tables, was also Granger causal. To check this "feedback", I looked at if the financial asset could provide information about the change in the sum of sentiment and pass the Granger Causality test. Any test with a p-value less than 0.05 means that one can reject the Null Hypothesis of the Granger causality test and say that the daily sum of sentiment time series does have forecasting, predictive information for the financial asset at indicated lag. However, if the feedback test (which is when the variables of the original Granger causality test are reversed and tested) also passes, this would make the original Granger causality test meaningless as this would mean that the asset price trend and the change in the sum of sentiment trend would Granger cause each other, thus neither variable would contain predictive information about the other. Important to note here on the structure of the following tables, for organization I included all the passing results for each lag all in one table. However, I did not test all of the different Granger causality hypotheses for the change in sum of sentiment trend and the financial asset at the same time to avoid the issue of multiple hypothesis testing and the potentially erroneous p-values that multiple hypothesis testing can generate [293, 381]. Each entry in the tables below is the individual result for the Granger causality test between the change in the sum of sentiment and the financial asset.

Granger Causality	SentSum	PValue	Reverse Okay	Reverse PValue
Gold Price	Yes	0.004988	Yes	0.438339
Gold Futures	Yes	0.040277	Yes	0.53574
Wheat Futures	Yes	0.002161	Yes	0.599003
German 10y Bond	Yes	0.043168	Yes	0.618356
FTSE 100	Yes	0.003308	Yes	0.880743
10 Year US Treasury	Yes	0.035314	Yes	0.82779
10 Year US Futures	Yes	0.022427	Yes	0.909091
EUR	Yes	0.035038	Yes	0.08545
GBP	Yes	0.006035	Yes	0.080256
AUD	Yes	0.044733	Yes	0.154349
MXN	Yes	0.021253	Yes	0.149475

Table 24: Granger Causality Results for Lag 1 Against the Different Financial Assets Separated by Asset Class.

Granger Causality	SentSum	PValue	Reverse Okay	Reverse PValue
Oil Price	Yes	0.042714	Yes	0.904926
Oil Futures	Yes	0.03906709	Yes	0.928638
Wheat Futures	Yes	6.46596E-08	Yes	0.944131
German 10y Bond	Yes	0.005677	Yes	0.845183
FTSE 100	Yes	0.001391	Yes	0.860171
10 Year US Treasury	Yes	0.003009	Yes	0.714947
IG-ETF	Yes	0.044506	Yes	0.518171
10 Year US Futures	Yes	0.001307	Yes	0.708251
Bitcoin-Futures	Yes	0.013295	Yes	0.384176
2Y-Treasury Yield	Yes	0.030406	Yes	0.593837
GBP	Yes	0.005202	Yes	0.702139
MXN	Yes	0.04897	Yes	0.793235
RUB	Yes	0.000424	Yes	0.903236
Bitcoin	Yes	0.016982	Yes	0.333437

Table 25: Granger Causality Results for Lag 5 Against the Different Financial Assets Separated by Asset Class.

Granger Causality	**SentSum**	**PValue**	**Reverse Okay**	**Reverse PValue**
Gold Price	Yes	0.003813	Yes	0.489212
Oil Price	Yes	3.09826E-07	Yes	0.936389
Gold Futures	Yes	0.000121234	Yes	0.969781
Oil Futures	Yes	8.57092E-08	Yes	0.927111
Wheat Futures	Yes	1.3972E-14	Yes	0.608716
Nikkei 225	Yes	0.032658	Yes	0.680651
German 10y Bond	Yes	0.000445	Yes	0.826079
FTSE 100	Yes	0.001059	Yes	0.806077
10 Year US Treasury	Yes	0.000575	Yes	0.23006
Defense-ETF	Yes	0.005718	Yes	0.803637
Metals-ETF	Yes	0.039279	Yes	0.977376
10 Year US Futures	Yes	0.000117	Yes	0.212961
Bitcoin-Futures	Yes	0.010613	Yes	0.727189
EUR	Yes	0.000284	Yes	0.984743
GBP	Yes	0.000997	Yes	0.943754
MXN	Yes	0.000004	Yes	0.857881
RUB	Yes	0.000002	Yes	0.98621
Bitcoin	Yes	0.004504	Yes	0.214014

Table 26: Granger Causality Results for Lag 10 Against the Different Financial Assets Separated by Asset Class.

Many studies have looked at the Monthly [105] or the Daily level [55, 104], however I wanted to see if I could find forecasting on an even smaller time scale. Thus, I redivided the tweet data into the individual hours (for a total of 3,576-hour divisions) and reran the Granger causality tests on a smaller subset of financial assets exhibited in Table 27 below:

Granger Causality	SentSum	Number of Lags (Up to 24)	PValue	Reverse Okay	Reverse PValue
EUR	Yes	10	0.0196	Yes	0.0578
JPY	Yes	12	0.0262	Yes	0.7542
RUB	Yes	1	0.0019	Yes	0.3641
GBP	Yes	8	0.0362	Yes	0.1647
MXN	Yes	12	0.006	Yes	0.2507
EURGBP	Yes	5	0.0431	Yes	0.1953
AUD	Yes	10	0.018	Yes	0.35
ZAR	Yes	12	0.022	Yes	0.3978
BNB	Yes	1	0.0111	Yes	0.0822
Metals-ETF	Yes	12	0.0443	Yes	0.5334
CSI-300	Yes	8	0.0386	Yes	0.5287
Sensex	Yes	5	0.0019	Yes	0.6829
FTSE 100	Yes	4	0.0013	Yes	0.2928
Gold Futures	Yes	6	0.0005	Yes	0.2823
Oil Futures	Yes	21	0.0179	Yes	0.2311

Table 27: Granger Causality Results for the Hourly Sum of Sentiment Time Series Against the Different Financial Assets Separated by Asset Class

It should be noted that while the sum of sentiment was shown not to provide any predictive information for HY_ETF, the IG_ETF, and the Nikkei 225 within the 24 lags (representing at least one full day of data), it did have outside this limit, at 48 lags, 60 lags, and 30 lags respectively.

Section 4.6: Discussion

With the initial tweet gathering I found it unsurprising that more than twice the number of tweets were in the Goldstein Negative category given the nature of the Ukraine war. However, this shows that not only does the Goldstein Index find large geopolitical events, but when the sentiment analysis program is run on the captured tweets, the program returned an accurate sentiment result as shown by the large decrease in sentiment at the onset of the Ukraine War, followed by a sustained increase in negative sentiment relative to the before the War.

While I investigated 39 different financial assets time series, I found that only 11 assets were Granger causal with the sum of the sentiment from the Goldstein Index tweets at Lag 1, with the most immediate lag representing one day of trading. However, as I increased the number of lags, I found that more assets were Granger causal, i.e., the change in sum of sentiment provided predictive information for the change in the asset value (14 for Lag 5, a week of trading, and 19 for Lag 10, roughly two weeks of trading). The explanation for this, could come from how news can take time to disperse and effect the market, especially with assets that are "sticky", meaning that their prices don't move very quickly (Hayes, 2021) [218]. As Kleinnijenhuis, et al. [264] describe "news impact may not be limited to short-term effects, however. Long-term graphs showed that hope versus fear sentiments in financial news preceded actual economic developments." (Kleinnijenhuis, et al, 2013). What this means is that it may take time for certain financial assets prices to change in response to big geopolitical events such as the Ukraine War. Thus, as the number of lags increases, which represents the number of days after the change in the sum of Goldstein Index sentiment, it might have provided time for the changes in the finance assets' price to be realized, and increasing the number of financial assets that the Goldstein Index sentiment change is predictive of. For example, the change in sentiment was Granger causal to Steel Futures at 20 lags, nearly a month of trading after the change in sentiment. That said, there were some assets even when the maximum number of lags were used, the Goldstein Index sentiment never showed any predictive information, such as USD vs CNY Foreign Exchange Rate, which means that the Goldstein Index sentiment time series would not have any use in predicting the change in value of the asset.

As for the successful analyses, the findings matched Caldara and Iacoviello who "document that stock returns experience a short-lived but significant drop in response to higher geopolitical risk. The stock market response varies substantially across industries, with the defense sector experiencing positive excess returns, and with sectors exposed to the broader economy—for instance steelworks and mining—experiencing negative returns" (Caldara and Iacoviello, 2022, Pg. 3). This is shown by both the Defense ETF and the "Metals and Mining" ETF time series having a relationship to the "Goldstein Index" sentiment which is the proxy for geopolitical risk. Also, I discovered that both Oil Price and 2 Year US Treasury Bond Yield time series had a relationship with the "Goldstein Index" which Caldara and Iacoviello stated that their Geopolitical Risk Index had as well (Caldara and Iacoviello, 2022, Pg. 19). In addition, there was a mix of both the "risky" and the "safe haven" assets described by Amen appeared (Amen, 2020, Pg. 6). However, the FTSE 100 was only the "risky asset" (Amen, 2020, Pg. 6), to appear in all three different lag tests. While the US Treasury 10 Year Yield, was the only "safe haven asset" (Amen, 2020, Pg. 6) to appear in all three lag tests. For the assets I wanted to investigate, I found that both GBP/USD and USD/MXN appeared in all three lag tests.

After completing the Daily Analyses, I wanted to see if at a smaller time interval could I capture predictive information about the financial assets when I broke the sum of sentiment of "Goldstein Index" tweets down at the hourly level. Amen's Thorfinn Sensitivity Index [55] analyses on the daily level only, same as Bollen, et al [91], and Caldara and Iacoviello's GPR Index as well (Caldara and Iacoviello, 2022, Pg. 1) [105]. Rognone, et al. analyzed their data at 15-minute intervals (Rognone, et al., 2020, Pg. 3), so I wanted to examine a time frame below the daily level as well. My finding was, unlike the Daily Granger Causality analyses, I was able to find that many of the Forex assets, which trade 24 hours a day, were responding to changes in the sentiment within less than half a day. This result is in line with Rognone, et al. [367] who found that "Forex comoves and reacts homogeneously to news" (Pg. 1). And reinforces the findings of Nofsinger [326] which states that "financial markets adjust to changes in mood faster than real markets" (Nofsinger, 2005, Pg. 3). This discovery is important as it shows it is possible that the change in sentiment provides predictive information about the change in the Forex time series on a smaller time interval than other geopolitical risk indices and thus could inform different trading options in Forex markets.

The USD vs RUB exchange rate was important to look at as the Ruble is the Russian currency. I found that at the Daily Level, a change in the Goldstein Index sentiment had predictive power to the change in the USD vs RUB after five lags (roughly a week of trading). However, at the Hourly Level, the Goldstein Index sentiment change contained predictive information for the USD vs RUB change within one hour. This difference could potential be explained by the nature of the conflict and the different time scales. News about the Ukraine War updated frequently, especially at the onset, thus changes in the Goldstein Index information were very rapid. As mentioned above, Forex markets tend to move with news, so the low lag value at the Hourly Level was not surprising. However, at the Daily Level, the predictive information provided by the Goldstein Index could possibly be explained by amalgamation of the data at this level. At the Daily Level, the smaller changes would be evened out. While the Goldstein Index might have predictive information at a smaller time interval, at the Daily Level, the sum of all the smaller changes might eliminate the predictive information. However, over time at the Daily Level, the overall trends between the USD vs RUB and the Goldstein Index would become clearer, which would explain how the Goldstein Index has predictive power at the five and ten lags, but not at the first lag.

Lastly, there are three minor issues that I should make note off. The first is that I encountered the same issue as Bollen, et al., [91] who detailed: "we have no knowledge of the 'ground truth' for public mood states nor in fact for the particular subsample of the population represented by the community of Twitter.com users. This problem can only be addressed by increased research into direct assessments of public mood states vs. those derived from online communities such as Twitter" (Bollen, et al. Pg. 7). This is also related to a lack of a baseline econometric model for this data without any social media variables. Without the baseline model, the difference created by adding the change in the sum of sentiment time series as a variable to predict the change of the financial asset could not be found. While these issues were outside of the purview of this case study, as I was only investigating the relationship between them and not making creating a prediction of how the financial asset would change, I feel that by including a large lead time (nearly three months) before the start of the Ukraine War, this mitigates the effect described by Bollen, et al., [91] as I was able to develop a baseline "ground truth" for Twitter / X sentiment regarding geopolitical risk. The third issue is with Twitter / X itself. While Twitter / X's demographics have slightly balanced out over time, Twitter / X users are more often younger and male [397]. So, while I am capturing more sentiment worldwide, I may only capturing an uneven demographic which could have potentially skew my results.

Chapter 5: Case Study 2

Twitter / X and the Multilingual Analysis of Emerging Geopolitical Topics in Near Real Time

This second case study focuses more on research questions three and four revolving around identifying emerging geopolitical risk topic through multiple languages on Twitter / X. Unlike, the first case study, this case study deals with Twitter / X data not just at the daily level but at the near real time level as well. This was done to further test the capabilities of the programs, to see if the sentiment analysis and topic modeling could be done at this level. Specifically at this case study's aim was to investigate topic modeling at the both the daily level and the hourly level with my program. [6] By examining the data at the hourly level, it would be possible to find the emerging topics quicker than at a daily level, and we can also find topics that were only relevant within the day that might get lost if a larger time series was used. I only looked at three hour intervals, which is short, but I investigated over 40 days (the methodology expanded further in the chapter), which gave a larger opportunity for exploring topics generated both within the day and across the entire study period. Finally, the use of multilingual data in this study allowed me to investigate the topics generated across regions which allowed for comparison and contrasts between what geopolitical risk emerged in different parts of the world and which are more relevant to these regions.

Section 5.1: Introduction

Geopolitics is always significant in our world, however, in the last few years it has taken greater importance. From COVID-19 to the Ukraine War to

[6]The programs created and tested by me for this case study was DyAdF.py and Call1H_TM_DA_MP5.ipynb which can be found at https://github.com/jb370/Automatic-GR/blob/main/Python/Real_Time_Python_Code/foo9.zip, please see the Read_Me file for further details.

the increased tensions between the United States and China, especially with trade wars and the South China Sea. Developments in geopolitical events can occur rapidly and force governmental and non-government actors to modify their responses to them with little notice. Thus, it becomes vital to get as much lead-time as possible to find emerging geopolitical events to allow for the best course of action to materialize. However, some geopolitical events might not always have global implications; an occurrence may impact regional affairs, and be highly consequential to that area. Thus, is it essential that multiple languages are used to capture not only the widespread global events, but also the more local geopolitical events.

While it would be possible to use traditional media sources to track the emergence of geopolitical events, I decided to use Twitter / X for several reasons as described in Chapters 1 and 2. Twitter / X has data in multiple languages and can provide that data in real time, which made it a convenient way of gathering multilingual data for my study versus having to use multiple different traditional news media sources. This language diversity not only provided different regions where I found specific geopolitical events, but also gave a global reach for major worldwide geopolitical events that if I only used English, I might have missed those insights.

The rest of this case study breaks down as follows: Section 2 focuses on key concepts that supported this study, while Section 3 describes related work that help formed my methodology. Section 4 describes the methodology I developed for dynamically tracking emerging geopolitical risks. Section 5 details the results, and Section 6 provides discussion around the findings.

Section 5.2: Key Concepts

Section 5.2.1: Geopolitical Topic Generation

Initially, I only used terms from the Goldstein Index [201] from Chapter 4. However, I found through my initial testing, I did not gather enough tweets per hour for the programs I developed to create coherent topics for topic modeling in Case Study 2. Therefore, I decided to include more sources to expand my geopolitical topic gathering potential. Two sources I used were Klement [265] and Caldara and Iacoviello [105]. Klement wrote a textbook on Geo-economics where each chapter discusses factors that affect global economics that occur from different geopolitical events. Caldara and Iacoviello created an index for tracking and evaluating geopolitical events through looking at different English news articles for specific keywords. For Klement, I chose the key phrases based on my readings breaking the different chapters he discusses on geoeconomics into different topics. As for Caldara and Iacoviello, who focused exclusively on geopolitical risks, they had a table which broke down the different geopolitical topics they were searching for and the key phrases they used to search for them (Caldara and Iacoviello, 2022, Pg. 31). After determining my topics, I also gathered the most common key phrases from web scraping Wikipedia articles I could find that related to my topics (Terrorism[7], Oil Supply Shock[8], US – China Relations[9], Cyberwarfare[10], Nuclear Threats[11]). Combining these methods gave me a corpus of relevant bigrams for each topic, (the key bigrams themselves can be found in Table 28 in Appendix A), however, the next step was to translate these bigrams across the other languages I was investigating.

For the translation process, I opted to use human translation over machine

[7]Wikipedia contributors, "Terrorism."
[8]Wikipedia contributors, "1973 Oil Crisis."
[9]Wikipedia contributors, "China–United States Relations."
[10]Wikipedia contributors, "Cyberwarfare."
[11]Wikipedia contributors, "Nuclear Warfare."

translation, as Kravariti, 2016 [272] describes: "It's a translator's job to ensure the highest accuracy... Humans can interpret context and capture the same meaning, rather than simply translating words...Humans can spot pieces of content where literal translation isn't possible and find the most suitable alternative" (Kravariti, 2016). While not as fast as machine translation, the increased accuracy and the flexibility in translation that a human translator provides was vital for my data gathering needs as getting the correct keywords allows us to collect the most relevant tweets for my topics. Thus, I employed translators from Gengo.com[12] for French, Portuguese, Arabic, Japanese and Korean, while I had Spanish translated by private translator[13]. Unfortunately, the Twitter / X Filter Stream API rules have limits [420, 421], thus I could only choose a limited number of key bigrams for each topic when gathering the data. I decided to choose five bigrams each of the topics, except for the "Goldstein Index" topics which I doubled up the rules for these topics to ten bigrams each to try and gather more tweets that have a defined geopolitical focus. This gave me a total of 140 bigrams across the seven languages I studied.

Section 5.2.2: Google Trends

The second concept I need to discuss for this study is Google Trends [10]. As described by Choi and Varian [126]: "Google Trends provides *daily* and *weekly* reports on the volume of queries" (Choi and Varian, 2009, Pg. 3). Since 2009, Google has updated Google Trends functionality to include reports on changes in popularity of the search terms on Google to include reports at the minute level up to 4 hours before the time on the report generation. The change in popularity data is also detailed by Choi and Varian: "Google Trends data does not report the raw level of queries for a given search term. Rather, it reports a query index. The query index starts with the query share: the total query volume for search

[12]https://gengo.com/
[13]See Acknowledgments for Details

term in a given geographic region divided by the total number of queries in that region at a point in time. The query share numbers are then normalized so that they start at 0 in January 1, 2004. Numbers at later dates indicated the percentage deviation from the query share on January 1, 2004." (Choi and Varian, 2009, Pg. 4). Instead of January 1st, 2004, which is as far back as the Google Trends data goes that Choi and Varian used for their study, I generate Google Trends reports from February 4th, 2023, to March 23rd, 2023, at the daily level, and from June 1st, 2023, and June 6th, 2023, for the four-hour time frame at the minute level around the three – hour window which I collected the tweets. I chose to use Google Trends to compare with my emerging geopolitical topic analysis because Rill, et al, [362] showed that Google Trends provides an effective comparison to their PoliTwi System, which aimed to capturing emerging German political topic using Twitter / X. As I am studying similar concepts, I also decided to use Google Trends.

Section 5.3: Related Work

Section 5.3.1: Topic Modeling over Time

As detajled in Chapter 2, while Blei and Lafferty [89] developed dynamic topic modeling based off the LDA mathematical framework, other studies developed other methods to track topics over time. In a variation of LDA, Griffiths and Steyvers [207], replace the EM process of LDA with a Markov Chain Monte Carlo (MCMC) process known as Gibbs Sampling to obtain the topic and word probability distributions (Griffiths and Steyvers, 2004, Pg. 2 – 3). In addition, they also were able to track topics over time in what they called "Hot or Cold Topics" using Linear Trend Analysis to find the emergence of topics over time (Griffiths and Steyvers, 2004, Pg. 5 – 6). Wang and McCallum [441] took a different approach than the other two, they directly incorporate time into to the LDA algorithm via the Beta distribution. The Beta distribution creates a timestamp for each occurrence of a word in a document, thus this addition to the base LDA model allowed for "a continuous distribution over time associated with each topic, and topics are responsible for generating both observed timestamps as well as words" (Wang and McCallum, 2006, Pg. 1). This meant that Wang and McCallum's method could focus more specifically on the time of certain events. More studies built on the basis developed by this research including Ahmed and Xing's infinite Dynamic Topic Model [40] which employed a recurrent Chinese restaurant franchise process (a clustering algorithm similar to LDA topic modeling altered for time intervals) to allow the words to vary between topics dynamically and allow for new topics to emerge and other topics to die unrestricted by number of topics. Hurtado, et al. [237] extended the research by trying to forecast which topics will be popular in the near future. While Hida, et al. [223] combine the dynamic nature of Blei and Lafferty's Dynamic Topic Modeling to capture the emergence and changes of topics over time with the static nature of LDA to better understand how topics relate to one another. Importantly for this research, I found several studies [76, 138, 140] make use of the Twitter / X API temporal qualities to analyze topic identification and the growth of these topics over time. These studies proved that it was possible to track topics over time with using LDA variations, however, I wanted to explore if these methods could identify emerging geopolitical events on Twitter / X and tracking their changes in real time was possible.

Section 5.3.2: Multilingual Topic Modeling Literature Review

Another area that influenced my research was multilingual topic modeling. This is an important concern in topic modeling as Lind, et al., [290] states "automated methods of content analysis (such as topic modeling), are usually applied to text documents in just one language – mostly English" (Lind, et al., 2019, Pg. 2). Researchers have used various methods to develop topic modeling methods that can be applied in a multilingual framework. For example, Boyd-Graber and Blei [93] developed a topic model for multilingual text called MuTo. MuTo is based off the LDA algorithm, however instead of just words, MuTo uses matching pairs of words between the multiple languages to creating matching topics across languages. Boyd-Graber and Blei showed that MuTo worked well across similar text in English and German. In an updated study, Yang, et al. [455] showed that their multilingual topic model which "does not force one-to-one alignment across languages" which allows for better topic generation for languages that have small corpora (Yang, et al., 2019, Pg. 1). However, my research used language as a proxy for location to see what different geopolitical topics emerged in different regions. Thus, I did not want linkage between topics, fortunately many studies provided insight for me. Zheng et al [468] used separate LDA models for Japanese and Chinese to compare topics discussed in different online blogs. However, two multilingual papers that were incredibly helpful were Amara, et al. [54] and Sakamoto, et al. [374]. Both papers investigated trends (COVID-19 for Amara, and the US and Japanese legislatures for Sakamoto) across multiple languages using different LDA models for each language. They also tracked these trends over time, showing the topics' emergence and growth.

Section 5.3.3: Topic Modeling Contribution

I aim to advance research into this space by combining the multilingual topic modeling methods with the dynamic topic modeling methods and applying this combination to create a program to analyze geopolitical events found through the analysis of tweets. My goal is to develop a method that can apply these concepts on a smaller time frame than have been looked at previously, showing changes in geopolitics not just at the daily level, but also detailing the emergence of topics at the hourly level as well. Utilizing a variation of LDA for this aim will work because, as many other researchers have applied LDA to their research with success. Additionally, as Amara states: "The LDA can be used as predictive model to detect novel trends in the behaviour of the social media users' topics or to give a deeper analysis about a specific topic by detailing the document providing the target topic." (Amara, 2021, Pg. 6) [54]. Thus, I should find emerging geopolitical topics and develop geopolitical topics with a program that combines these methods.

Section 5.4: Methodology

The methodology is broken into four parts: the Data Gathering section, the Data Cleaning section, the Multilingual Emerging Geopolitical Topic Modeling section, and finally the Evaluation section.

Section 5.4.1: Data Gathering

Since, I used Twitter / X for the data source, I decided to implement the Twitter / X API, specifically the Filter Stream API[14]. This Twitter /X API allowed me to obtain real-time tweets that contain key phrases in the different languages I am investigating. I felt confident using this method of data collection as other studies such as Metzler, et al [305], and Culotta [140] both use key words in their studies to gather tweets. As described earlier, I used the "Goldstein Index" topics shown in Appendix A in Table 28 initially. However, I decided to expand the reach of potential geopolitical events and included topics from other sources, such as Klement [265], and Caldara and Iacoviello [105]. With the geopolitical topics in hand, the next step was to translate these bigrams across the six other languages I was investigating. For the key bigram translation process, I opted to use human translation over machine translation, as described in Section 2. I ran the Filter Stream API for an hour, (testing showed that a time interval under an hour would not produce viable topics), while storing all

[14]Twitter / X. "Filtered Stream."

the tweets that come in, into json files, and at the end of that hour, I took the json files in the storage folder and convert them into a Python data frame and then begin the data cleaning process.

Section 5.4.2: Data Cleaning

The first step in the data cleaning process was to label the languages of the tweets. While the Filter Stream API can have a language indicator as one of its filtering rules, the tweet output itself did not contain a language tag. Thus, it was important that I created this tag so I can separate out the languages for analysis later [143]. Once the tweets' languages are identified, I then split the data into individual language data frames and then sorted these data frames chronologically. I then tokenized (i.e., split the tweets' text into individual words) the tweets in the language data frames and then removed the language specific stop words from the tweets, while also including certain stop words that revolved around URLs and Twitter / X: (["http", "https", "co", "com", "app", "go", "amp", "RT", "rt"]). Afterwards, I grouped all the remaining words into to bigrams, i.e., two term phrases of words that occur next to each other [353], this process helped identify distinctive word combinations and build the corpus necessary to implement the LDA algorithm. Once the bigrams are constructed for each language, I then used lemmatization on the bigrams for English, Spanish, French, Portuguese, and Japanese [232]. Lemmatization removes the ends of words, to retain the word root, (for example: "bus" is retained from "buses"), so that these word roots can be grouped easier to create clearer topics later. However, I do not include Arabic or Korean in the lemmatization process for different reasons.

For Arabic, lemmatization generally involves removing of the diacritical marks, which are "short vowel symbols inscribed atop regular letters" that are apart of written Arabic (Hegazi, et. al, 2021, Pg. 1) [220]. However, "Arabic abjad or letter can be seen in several words which do not carry a similar meaning, so in order to remove that confusion that can even occur to native speakers, we use diacritical marks in a wide range of texts" (International House Cairo, 2023) [22]. Thus, I tested to see if removing the diacritical marks would have a large effect on the comprehension of the topics created. I found that the topics with diacritical marks did make more sense than those without, thus I decided to skip the lemmatization process for Arabic.

With Korean, however, I encountered a different problem. Through testing, I found that many of the out-of-the-box lemmatization programs for Korean lost much of the context of the tweet since I am working with such short texts. However, studies such as Lee and Song, 2020 [279], skipped the lemmatization step in their analysis step of Korean text data, so I felt safe doing the same. Once the data was cleaned, I continued to the first step of the LDA process which is creating the corpus and dictionary for each language out of the either the lemmatization bigrams data for English, Spanish, French, Portuguese, and Japanese, or the regular bigrams data for Arabic and Korean. With these created, I could move on to creating the emerging geopolitical topic models.

Section 5.4.3: Multilingual Emerging Geopolitical Topic Modelling

The emerging geopolitical topic modeling analysis program methodology comes mainly from two sources: Kapadia, 2019 [256], who provided the framework for building the dynamic selection of the hyperparameters for the LDA algorithm; and Besaw, 2020, [82] who developed a way to track the growth of topics over time that interfaced with the rest of the design.

The first part of my process involved the dynamic selection of hyperparameters for the LDA algorithm. As explained in Chapter 2, the LDA algorithm operates by first setting hyperparameters, which are parameters values determined before training the LDA model begins. While these can be picked at random, the best results are obtained when these hyperparameters are fine-tuned for the data analysis. Thus, for each of the seven languages I created hyperparameter tuning functions to dynamically set alpha, beta, and the number of topics. To do so, I relied on the concept of coherence. As defined by Roberts, et al. [364], coherence is “maximized when the most probable words in a given topic frequently co-occur together” (Roberts, et al., 2013, Pg. 10). This means that a topic can be considered to have high coherence if the bigrams of the topic appear together frequently, which also means that the topic is easily interpretable by a human reader, thus, higher coherence generally relates to better quality topics (Roberts, et al., 2013, Pg. 10). This means that finding the combination of hyperparameters that maximizes coherence was an important part of the LDA topic modeling process. To do so, I follow the procedure outlined by Kapadia, 2019 [256], who developed a method to loop through each combination of a set list for each hyperparameter (alpha, beta, and the number of topics) and record the coherence value from each pass of the LDA model with those hyperparameters in a data frame. I added to this by automatically finding the maximum coherence from the output data frame and imputing the hyperparameters that correspond to this value into my main LDA modeling function within my algorithm.

With the hyperparameters tuned, I could move on to implementing the main LDA model. I plugged in the hyperparameters, and I obtained the LDA model with the topics with the highest coherence. However, these topics did not describe how they changed over time or if any emerged, so I needed to manipulate the topic modeling results to work as a time series. First, I needed the topic proportions of each tweet for each topic, this is the proportion of words in a tweet that belong to a certain topic, also known as the weight (i.e., how much each topic holds weight in a document). Once I had each of topic proportions for each document (in my case, a tweet), I created a time series by breaking the sum of the individual topic proportions by hour. I divided that value by the count of tweets in each hour which gave the average topic weight by hour. This average topic weight allowed me to track the changes in the topic by hours and evaluate if certain topics are emerging or losing relevance [82, 445]. Finally, I used the first five topic words to label each of the topics. However, for all non-English languages, I used Google Translate [321] to translate the topic words to English before labeling the topic trends over time for easier comparison between the languages. The labels gave a better understanding of the trends which I visualized to track the emergence of topics. Lastly, this process was repeated every hour, where more data is gathered and reprocessed with the previous data collected to see how the topics have changed from the previous hour. This was done during each hour of the three-hour capture period. A diagram of the work process for the topic modeling program procedure is shown in Figure 6 in Appendix C. In Figures 7 – 9, below, I provide an example visualization of the process, these are the topics for English from June 6th, 2023, starting 2pm GMT (14:00):

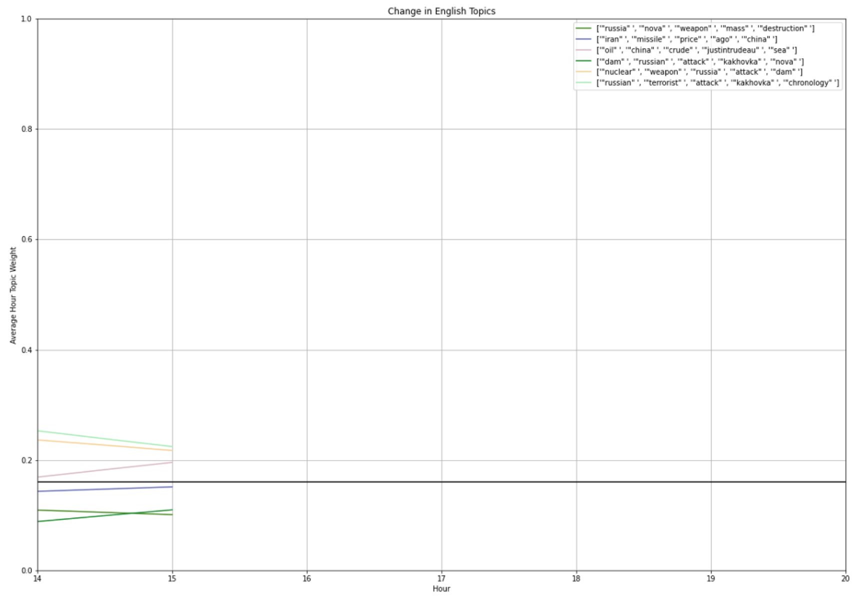


Figure 7: This is the first hour of data for the English topics for June 6th, 2023, starting at 2pm GMT (14:00). On this day, there was major news out of Ukraine with the destruction of the Nova Kakhovka Dam[15]. However, US also placed sanctions on Iran and China over Iran's hypersonic missile program[16]. The solid black line is the median value for the average weights of the topic proportions across the entire time period. Trends under this line are less relevant that the ones above it. Trends that cross it show an increase in relevance over time and could be considered emerging topics.

[15]Lakezina, Viktoriia. "War Zone Villagers Flee after Massive Ukraine Dam Destroyed."

[16]Psaledakis, Daphne. "Us Slaps Sanctions on Iranian, Chinese Targets over Tehran's Missile, Military Programs."

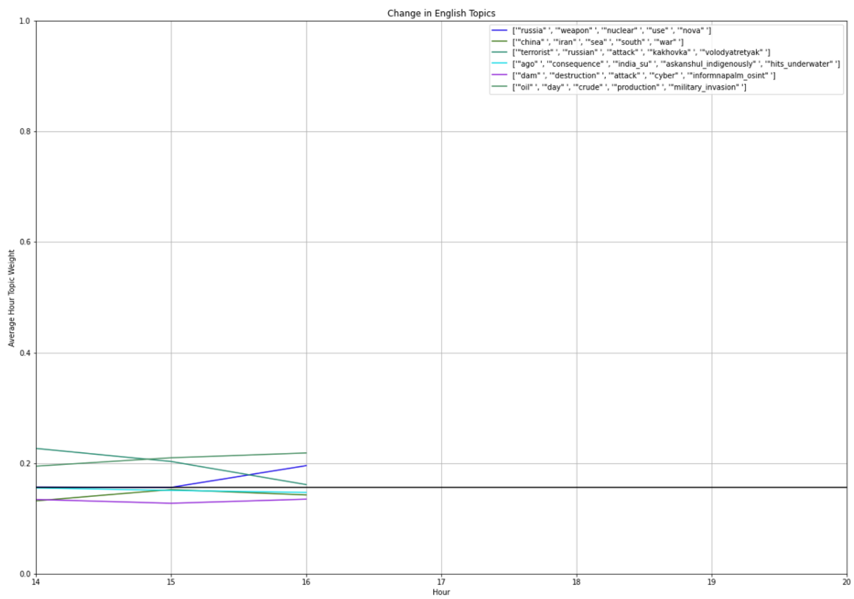


Figure 8: This is the second hour of data topics now added to the first hour, showing the changes in topics over the time, as the figure shows, the Nova Kakhovka Dam story is starting to fall in relevance, while the China and Iran sanction starts emerging more.

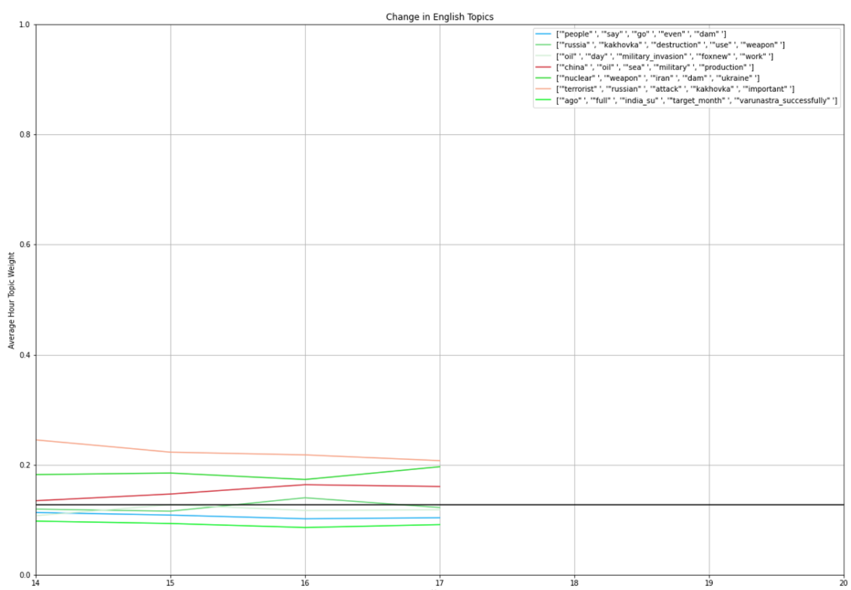


Figure 9: This is the third hour of data topics now added to the previous hours, the Nova Kakhovka Dam is still relevant but decreasing, while Iran and China have split into separate, but still relevant topics with the China topic leveling off, but the Iran topic is still growing.

For this case study at the daily level, I looked at the period from February 4^{th}, 2023, to March 23^{rd}, 2023. I took the final hour of data topics, like Figure 9, and got the count of number of topics that appeared for their labels. For example, in Figure 9, I would say that there were three topics related to the Dam, one for China, one for Iran, and one for India. This gave me a time series trend that I could evaluate using Google Trends described below.

Section 5.4.4: Evaluation and Validation

For both the Daily and the Hourly emerging geopolitical topics data trend, I follow the lead of Rill, et al, 2014 who evaluated German political topics on Twitter / X against Google Trends [10, 362]. I also wanted to see if my emerging geopolitical topics methods could also outperform Google Trends. For the hourly analysis, I collected the minute level data for Google Trends around a four-hour period during the Twitter / X model analysis time frame. I compared the Google Trends data to the trend lines generated through the Twitter / X topics change over time. This allows for me to evaluate if a topic found in the tweets emerged in the Google Trends first or through the Twitter / X topic models. For the daily level analysis, I used daily Google Trends data over the entire study period from February 4^{th}, 2023, to March 23^{rd},2023 and compared that to the number of topics generated for the day, (in this case would be the topics that still exist at the end of the model analysis time frame). This creates a trend in count over time for the Twitter / X topics which can be compared to the Google Trends for the same topic which allows for an evaluation about the emergence and relevance for each topic to see which method captures the topic first.

For the validation of the emerging topic models, after the completion of the analysis run, I would take the topic words generated and search them to see if there was any events that took place during the analysis time frame that would correspond to the emerging topics for each language. In the Results section and in Appendices G, H, and I, I have cited articles that I found related to the topics that were generated.

Section 5.5: Results

For the analyses, I obtained the Hourly Topic Results across two days: June 1st, 2023, and June 6th, 2023. As for the daily level, I looked at the period from February 4th, 2023, to March 23rd, 2023. I obtained many emerging geopolitical topics across my seven study languages, however, for the sake of clarity in the main body of this study, I only included one example that best exemplified each type of analyses. However, the remaining hourly results are in Appendix G, and the remaining daily results are in Appendix H. Additionally, further examples can be found in Appendix I.

Section 5.5.1: Hourly

Figure 26 below is the English Topics generated and tracked at the final hour of the three-hour time frame on June 1st, 2023, the time is set at GMT, as that is the time zone that Twitter / X uses when it records when a tweet is created at. As the chart shows, seven topics were monitored, while many topics were losing relevance during this period, as you can see the green colored topic of “Iran”, appears to emerge at around Hour 19. This emergence was related to the announcement of the U.S Department of State[17] announcing sanctions on Iranian operatives engaged in external plots, i.e., operations such as assassination attempts that took place outside of Iran. The NCRI[18] reported on these new sanctions at 20:30 GMT, which is just after the “Iran” topic emerged.

[17] Antony J. Blinken, Secretary of State. "Sanctioning Operatives Involved in Iranian External Plots."

[18] Shahrokhi, Sedighe. "Iran News in Brief – June 1, 2023."

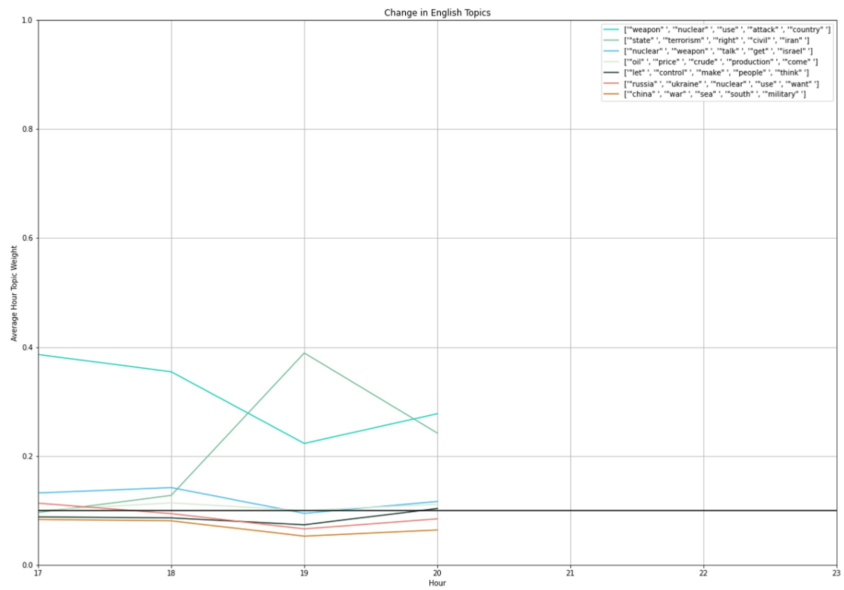


Figure 26: The English Topics that appeared and were tracked over time for June 1st, 2023

However, as Figure 27 below shows, while Iran was always a highly searched topic, the maximum searches for this four-hour time window only spiked after 22:00, which is nearly three hours after "Iran" as a topic emerged on Twitter / X. As the hourly Google Trends data can be rather noisy with potentially variations between minutes, I added a polynomial trendline in orange to provide more context to the changes in search trends over time.

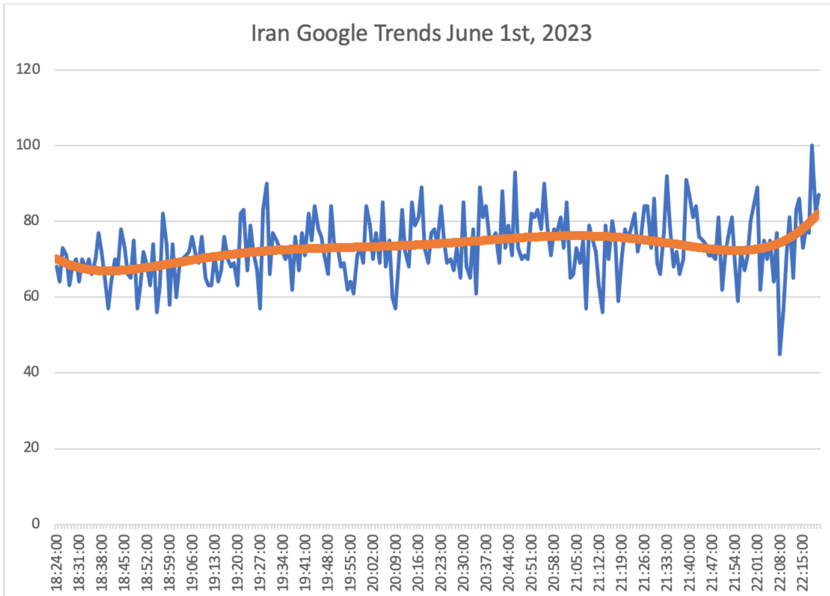


Figure 27: Google Trends tracking of "Iran" from 18:24 – 22:15 on June 1st, 2023

In Appendix G, I also show another "Iran" topic that emerged in French tweets on June 6th, 2023, and a topic on French President "Macron" that emerged in Japanese tweets also on June 6th, 2023.

Section 5.5.2: Daily

In contrast to the hourly analysis, I was able to compare the daily trends for both Twitter / X and Google on the same chart. However, as with the hourly analysis, I will only display the results from one geopolitical topic for English, but it is also a topic that emerged across all seven languages. The results of the other languages of this topic for the daily analysis are in Appendix H.

During the research period from February 4th, 2023, to March 23rd, 2023, North Korea launched multiple intercontinental ballistic missiles[19]. The first launch took place on February 18th, while two more missiles were launched

[19]Seo, Yoonjung; Brad Lendon; Junko Ogura. "North Korea Says It Tested Icbm in Surprise Drill."

on February 20th, landing near the waters of Japan[20]. Later, on March 9th, and March 19th, North Korea launched more missiles in response to US and South Korea joint military exercises[21],[22]. These missile launches caused a large geopolitical firestorm and were also the perfect test case for my Twitter / X multilingual emerging topics. Figure 28 presents how the “North Korea” topic emerged in English during my study period, and how it compares to the reaction time of Google Trends at the daily level.

North Korea emerged as a topic on Twitter / X corresponding to four of the missiles launches by North Korea. While this is not every missile launch from North Korea during this time period (the total count is ten [148]), these emerging Twitter / X topics corresponded to missile launches than Google Trends two peaks representing increased search interest on February 20th, and March 21st. Additionally, the Twitter / X topics emerged before the Google topics shown by the first and the third spike in the Twitter / X data, appearing almost two days before the increase in Google Trends.

[20] Choi, Soo-Hyang; Hyonhee Shin. "North Korea Fires Two More Missiles into Its Pacific 'Firing Range'."

[21] Yim, Hyunsu "North Korea Fires Ballistic Missile as U.S.-South Korean Drills Go On."

[22] Johnson, Jeese. "North Korea Fires Short-Range Missile Ahead of U.S.-South Korea Military Exercises."

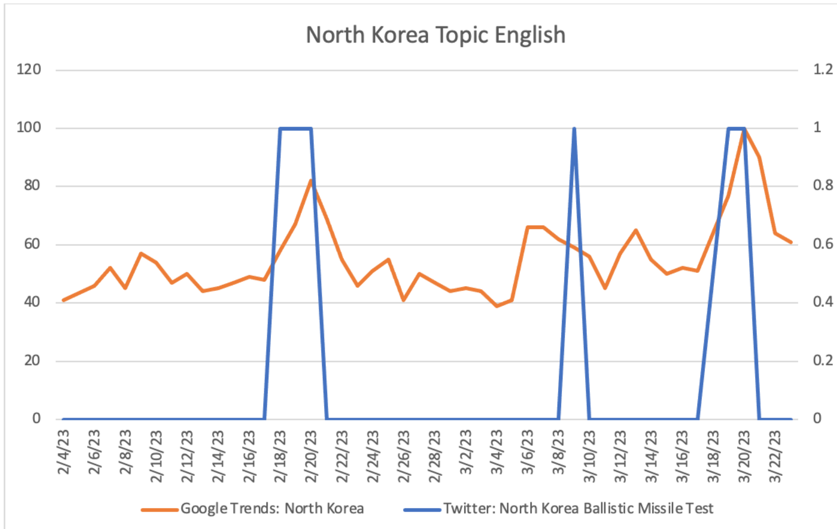


Figure 28: This is the comparison between the Twitter / X emerging geopolitical topics data for English and Google Trends search data for North Korea. In the legend the Google Trends label is the criteria used to gather the Google Trends data. The Twitter / X legend label is the topic label used for the emerging topic on Twitter / X

Lastly, I have the results from the daily analysis of all topics generated and their respective counts, shown in the Tables in Appendix J. English had 41 unique topics, Spanish had 46 topics, French had 42 topics, Portuguese had 42 topics, Arabic had 52 topics, Japanese had 12 topics, and Korean had 19 topics. Many countries use English and French as a lingua franca, thus there is an increase diversity from different parts of the world in the topics that emerged from during the study period. However, Spanish, Portuguese, and Arabic, had many topics that focused on regional issues. Japanese and Korean had less diversity in topics, but an increase on “Crude Oil” and “Nuclear Weapons”.

Section 5.6: Discussion

These results are in line with Rill, et. al, 2014 [362] who found “that new topics appearing in Twitter / X can be detected right after their occurrence. Moreover, we have compared my results to Google Trends. We observed that the topics emerged earlier in Twitter / X than in Google Trends." (Rill, et. al, 2014, Pg. 1). I also matched the results from Lee, et. al, 2017 [282], who showed that using Twitter / X produced more accurate and earlier results for influenza spread that Google Flu Trends. While I shared similar findings on the daily level for Rill, et al., I also went to a smaller time interval and showed that the Twitter / X emerging topics program can outperform Google Trends in capturing emerging topics at the hourly level. Additionally, I found that topics that emerged before the analysis's time frame started, would also appear in the analysis, such as the Nova Kakhovka Dam destruction on June 6th, 2023[23]. The explosion at the Dam occurred before the three – hour interval for that day, but I was able to show that topic was still popular on Twitter / X, as shown by Figure 29. This was also reflected Google Trends data where the search term remained high throughout the capture period as seen in Figure 30. There are a few potential reasons for this difference in response time to geopolitical events between Twitter / X and Google Trends. As described by Martin [300] in 2023: "Twitter is the most popular social platform for news and current events. Twitter's appeal has always been its short-form, real-time nature and that's still true today with 61.2% of people saying Twitter is where they go to stay up to date with news and events" (Martin, 2023). As I was tracking emerging geopolitical events, it's possible that people might first hear about the event on Twitter / X and since Twitter / X's micro-blogging nature doesn't provide a wealth of information, people might go to Google afterwards to find out more about the event. This would lead to my Twitter / X topic modeling program to outperform Google Trends as the information would get out on Twitter / X first, while Google Trends only captures search volume, which would only spike after people found out about the geopolitical event and started searching for more information.

[23] Faulconbridge, Guy. "Nova Kakhovka Dam Breach: What Do We Know So Far?"

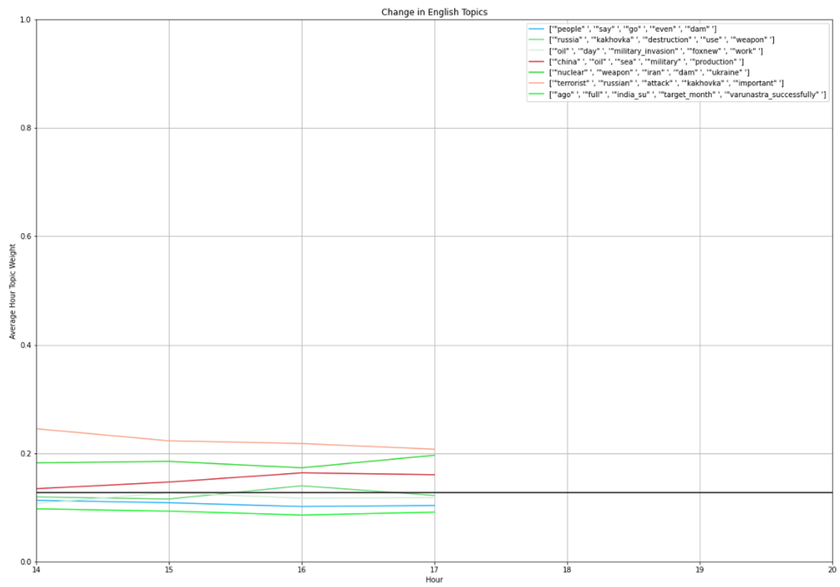


Figure 29: These are the English topics for June 6^{th}, 2023, from 14:00 to 17:00 GMT, the Kakhovka Dam topic is represented by the Orange-Peach line. This line is far above the black median average weight line showing that it still a very relevant topic.

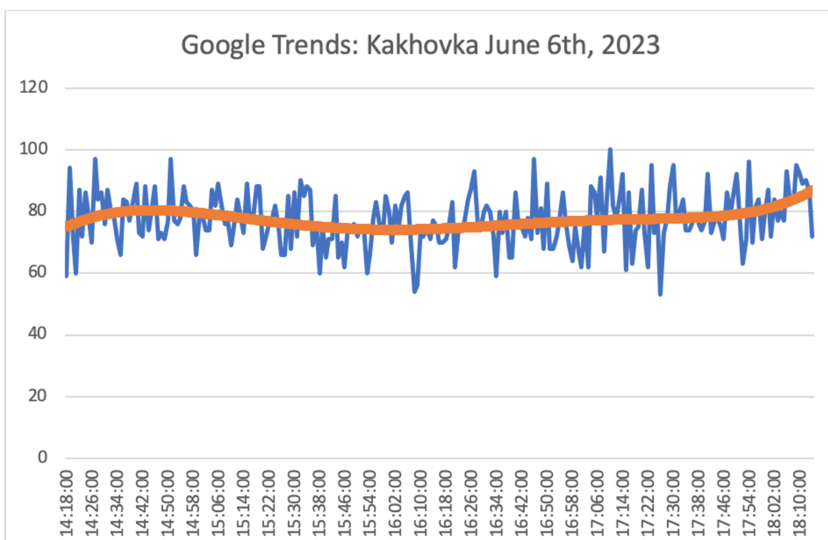


Figure 30: This is the Google Trend results for "Kakhovka" on June 6^{th}, 2023. The Google search results stay high, with the ratio remaining above 60 for the entire time frame, indicating increased interest in the geopolitical event

I was able to capture many geopolitical topics, but it was not an exhaustive list of geopolitical events. This is seen by the North Korea topics in the different languages had varying numbers of topics emerge. Additionally, the emerging geopolitical trends doesn't always capture the trend before Google Trends. For example, when Russia withdrew from the START treaty the Portuguese emerging geopolitical topics found it, but at the same day as Google Trends shown in Figure 31. That said, just because my Twitter / X topic modeling program can find topics faster than Google Trends does not mean that speed is the only measure of success. While it is important to have speed, Google Trends provides valuable information about the relevance of a topic that might be searched for more people over time as the information disburses, when compared to a social media like Twitter / X whose trending topics and hashtags last for around 11 minutes on average [317]. However, since I can capture some topics before they appear on Google Trends, this makes using topic modeling with Twitter / X to capture emerging geopolitical events a valuable source of information.

Figure 31: This is the Portuguese topic for Russia withdrawing from the START nuclear arms treaty with the U.S., as the chart shows, the Google trends spikes at the same time as the Twitter / X emerging topics

As Fukuhara, et al., [181] found "In the context of a cross-lingual concern analysis, finding common concerns across languages, and finding unique concerns in a specific language are important… For unique (domestic or monolingual) concerns, cultural events in each country, and domestic problems in a country might be included." (Fukuhara, et al, 2005, Pg. 3). Many geopolitical topics in the daily analysis appeared across all languages as "common concerns" such as the Ukraine War or North Korea. English, French, Spanish, and Portuguese had the Ukraine War as their most generated topic during the study period. This is not surprising that Ukraine remains one of the most impactful geopolitical events of the 21st century. However, other topics emerged in languages that were more regional. For example, with the English topics, "South China Sea" had the second most topics appear for English during the study period. This reflects anxiety in English speaking countries about how the South China Sea might become a potential geopolitical flashpoint in the near future as tension between China and U.S remain in flux[24]. Arabic had a wide variety of topics mainly focused on relationships between Arabic speaking countries and other major powers that are involved in the area. As the Arabic speaking region has been in geopolitical turmoil for decades and with many countries influencing outcomes within the region, the diversity in topics is also not surprising given how much geopolitics matters here. Crude oil was a major topic for both Korean and Japanese as this topic plays a major role in these regions. Japan imports nearly 97% of its oil[25] from the Middle East and South Korea imports nearly all its crude oil and natural gas[26]. Thus, it is understandable that most geopolitical based tweets would be focused on crude oil, a necessary commodity for a nation.

I also encountered a few notable data errors during these analyses. For example, with Korean, Jae, "재", which means re (as in a reply to a tweet), and yuga, "유가", which means oil price, occur frequently in the analysis. This was combined as jaeyuga, "재유가" which often appeared in the topics and should translate to "re: oil price". However, this does not happen as Google Translate leaves it as "jaeyuga", which is an error. I do not view this as not a major issue as I can interpret "jaeyuga" as a topic involving oil, but it is noteworthy as it shows that while Google Translate is very useful in these analyses, it can be inaccurate for some languages. One study showed that Google Translate had an 82.5% accuracy for Korean [407], which shows room for improvement. Another error comes from Japanese, with the terms "Prius" and "missile" that would co-occur in many Japanese topics, which was confusing. I found on Twitter / X that in Japan, the term "Prius missile" is a slang term used to describe lead-footed drivers that accidentally drive their cars into buildings while speeding as shown

[24]Hass, Ryan. "How Biden Could "Thaw" Us Relations with China.".

[25]"Total of 96.6% of Japan's March Imports of Oil Came from Arab Countries, Led by Saudi Crude."

[26]"South Korea: 2021 Primary Energy Data in Quadrillion Btu."

in Figure 32 [370, 471]. This error I also deemed not too detrimental as Rill, et al. explains: “with. . . search terms, some tweets without a political context have been collected. This occurs, whenever a term is used with polysemous meanings.” (Rill, et al., 2014, Pg. 5). Rill, et al., did not consider this a major issue, and neither did I, as it did not affect the analysis greatly.

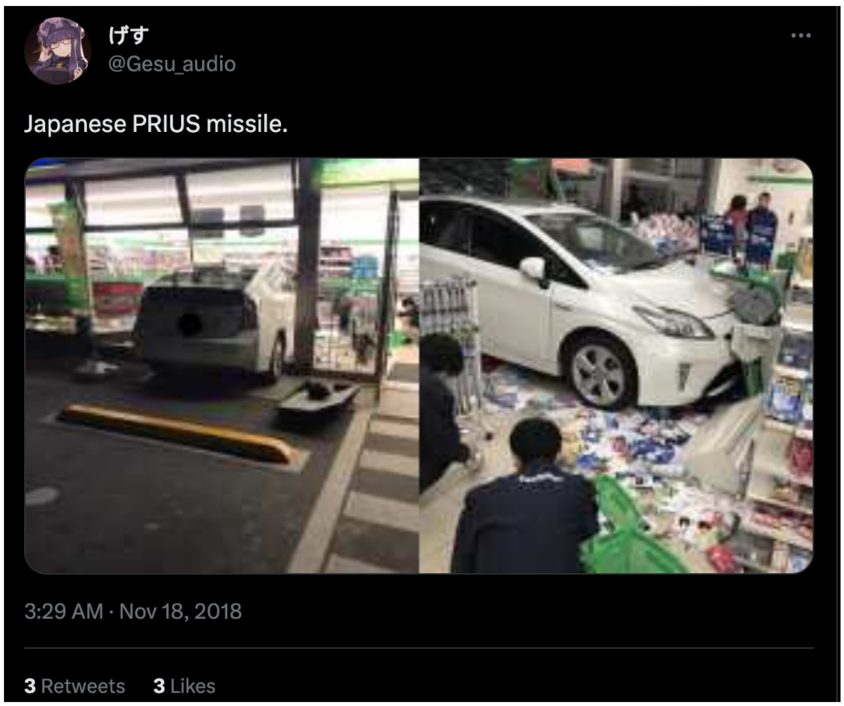


Figure 32: Tweet by Gesu_audio [471] about the Japanese Prius missile that cause the data collection error

Overall, this Twitter / X Emerging Geopolitical Topics program provides new insights into the growth of geopolitical events over time; however, I was not able to capture all geopolitical events for a region. This limitation comes from the capping the maximum number of topics that could be generated in the program. I did this because I was optimizing for speed of the dynamic hyperparameter selection process. I was aiming to make sure it would complete within an hour for the analysis's runs. However, I believe that more topics could potentially allow for a greater number of geopolitical events to be captured with the program.

Chapter 6: Case Study 3

Real-Time Multilingual Sentiment Analysis of Geopolitical Risk

This last case study is a combination of the first two case studies, where I investigate the real time changes in geopolitical risk sentiment across multiple languages using Twitter / X data in real time. This was a combination of all five research questions to get at the heart of my thesis to see if I could automatically analyze geopolitical risk. I chose to analyze this data in three hour intervals similar to the second case study because I would gather enough data per interval to run effective statistical analyses. However, I also collected roughly 40 days of data to make sure that any trend between the geopolitical risk index and the financial asset I compared would have a large enough sample size for me to say whether or not that trend was significant, and therefore could be generalized. This data was collected and analyzed in near real time through two programs. [27]

Section 6.1: Introduction

In recent years, geopolitics has become increasingly relevant with many high-profile events, such as the Ukraine War, the Syrian Civil War, the Ethiopian Civil War, increased tension between the US and China, and COVID-19, have caused major disruptions in global financial markets. Parallel to the growing geopolitical risk, social media has spread around the world. Twitter / X is one of the most popular social media websites with more than 285 million visits a month, with users sending out 9,596 tweets a second (Branka, truelist.co) [95]. This is a vast amount of text data generated with users discussing everything, including geopolitics. While studies into geopolitical used traditional news sources [105,

[27] The programs created and tested by me for this case study was Function_Filter_Test2.py and Twitter_API_JSON_File-MP5.ipynb which can be found at https://github.com/jb370/Automatic-GR/blob/main/Python/Real_Time_Python_Code/foo9.zip, please see the Read_Me file for further details.

338], I chose to use Twitter / X exclusively because it has been shown that Twitter / X is generally quicker than traditional news media [362]. In addition to this speed, I chose to implement sentiment analysis on this incoming tweet data to build an index to measure changes in geopolitical risk. As Kolasani and Assaf [269] describe "it is important to utilize social media data because events expressed through social media can significantly affect stock prices and trends due to the belief that prices change because of human behavior which can be reflected by social media" (Kolasani and Assaf, 2020, Pg. 2). Thus, I aim to use the geopolitical risk index developed from real time tweets to see if I can find any correlation between the change in geopolitical risk and the change in various financial markets.

However, as geopolitical events affect people worldwide, I aim to expand the research to cover multiple languages. Piskorski [338] puts the reasoning for this best by "extracting information on the same event from news in different languages potentially provides the opportunity to improve the overall coverage via deploying cross-lingual information fusion techniques" (Piskorski, et al., 2010, Pg. 4). Simply put, just investigating the sentiment for one language, such as English, to build my geopolitical risk index could lead to missing out on sentiment from non-English speakers, who might view the event quite differently. Thus, for this case study, I viewed it as imperative to make use of multilingual text real time tweets to build my geopolitical risk index.

The rest of this case study is as follows: in Section 2, I present a key concepts I incorporated and the related work to these concepts in Section 3. The methodology of my program which is outlined in Section 4. Section 5 details the results of the geopolitical risk index creation and the comparison between the geopolitical risk index and the various financial assets I investigated. Section 6 discusses these results.

Section 6.2: Key Concepts

Section 6.2.2: Real Time Analysis

The first concept is data analysis in real-time or near real-time specifically with Twitter / X. Research in this area is extensive with Concone, et al., [130] having developed "a system for real-time malware alerting using a set of tweets captured through the Twitter API's, and analyzed by means of a Bayes Naïve classifier" (Concone, et al., 2017, Pg. 1). By applying keywords representing malware terms to the streaming Twitter / X API, they can gather relevant tweets by labeling them with their Naïve Bayes classifier to check if there is an increase in malware related tweets in real time. Selvan and Moh [384] focused on creating a framework for large data real-time sentiment analysis by implementing Apache Hadoop and a sentiment lexicon to label tweets coming from the streaming Twitter / X API. They obtained 84% accuracy with their method which is quite high (Selvan and Moh, 2015, Pg. 1). Additionally, Prakruthi, et al., [344] used the Twitter / X API to gather tweets in real time and a sentiment lexicon to obtain a breakdown of tweet sentiment into positive, negative, or neutral. The use of the Twitter / X API and keywords in these studies was fundamental to the data gathering described in detail in the methodology section.

Section 6.2.3: Multilingual Analysis

The second concept that helped craft the study was multilingual analysis. Various studies applied different methods to handle multilingual data. For example, Agarwal, et al., [38] translated their multilingual tweet data into English and then applied sentiment analysis to those English-translated tweets. However, as shown in the Sentiment Analysis Chapter, this translation process can lose significant context, especially with short text, which can decrease the accuracy of the sentiment analysis. Two other studies did not translate their study languages however, such as Piskorski, et al., [338], who used various news sources to track events happening on the European border, as they describe: “since a significant fraction of relevant events are only reported in non-English, local news, with Italian, Spanish, Greek, French, Turkish, Russian, Portuguese, and Arabic being the most important ones at the moment” (Piskorski, et al., 2010, Pg. 2). The other study, Karageorgou, et al., [257] provided a basis for my multilingual sentiment analysis. In addition to using Twitter / X for their research, they also implemented different sentiment analysis methods for each of the three languages they were investigating (English, French, and Greek). Yet, instead of looking at data at the daily level or below as I aim to, they aggregated their data up to the monthly level to show the change in sentiment around various events (COVID-19 for Greek, the US 2020 Presidential election for English, and the Champions League Final for French).

Section 6.3: Related Work

Several studies have tackled the issues of sentiment analysis of tweets, the analysis of Twitter / X data in real-time, and multilingual sentiment analysis. Two studies used Twitter / X data in a similar way as I aim to do, i.e. using tweet sentiment to compare with changes in the stock market, were Kolasani and Assaf [269] and Chen and Lazer [120]. In Kolasani and Assaf [269], they collected a full year of tweet data related with the keywords “stock market” and “AAPL” (which is the Apple Inc name on the US stock market) (Kolasani and Assaf, 2020, Pg. 4). After collecting this tweet data, they applied sentiment analysis to create a time series data set at the daily level to predict the movement of the stock market and Apple Inc stock price with various machine learning algorithms. They found that “from the results of our work, it is seen that tweets do play a role in the prediction of stock market movement” (Kolasani and Assaf, 2020, Pg. 10). Chen and Lazer [120] applied a sentiment lexicon on their Twitter /

X data “to determine the probability of ‘happy’ and ‘sad’ for the entire tweet. These were then averaged per day to obtain a daily sentiment value.” (Chen and Lazer, 2011, Pg. 1). They implemented this daily sentiment value into a regression analysis on stock market movement and found “a correlation between Twitter sentiment data and stock market movement” (Chen and Lazer, 2011, Pg. 5). These results matched what I had found in the first case study, but I wanted to investigate this correlation on a smaller time scale, as close to real time as possible.

In terms of geopolitics, two studies were foundational to the development of this real time research. Caldara and Iacoviello [105] created geopolitical risk indices based off various print news media which they used in vector autoregressions to show how various markets and commodities were affected by changing geopolitical risk at the daily and monthly levels. Amen [55] used “30,000 daily feeds. . . including 1800 US think tanks, international peers for event risk, and thousands of academic & government publications” focused on geopolitics (Amen, 2020, Pg. 2). Amen took this text data and processed it with machine learning methods to create metrics of changes in different geopolitical risk. These metrics are combined into one geopolitical risk index and used to view how the change in geopolitical risks could affect the change in various financial assets across the world. Not only is the evaluation of geopolitical risk of academic interest, but several commercial entities have also produced methods for studying geopolitical risk. BlackRock [9], Dow Jones [6], and GDELT [7] have created various indices to track changes in geopolitical events and how they will affect their clients.

Section 6.4: Methodology

While this Case Study also investigates geopolitical risk with sentiment analysis, there are significant differences in methodology between a historical study as in the first case study in Chapter 5, and the real time analysis program presented here. There are three distinct parts to the methodology: the real time data gathering, the sentiment analysis of the multilingual data, and the comparison between the change in geopolitical risk sentiment and financial markets and assets.

Section 6.4.1: Real Time Data Gathering

Focusing my interest on financial assets and markets from around the world, I decided to run the analysis at random three-hour intervals to make sure that I would capture data from when all markets were open to see how change in sentiment could affect them. Additionally, as Vicinitas [1] shows, different languages have different hours of peak usage on Twitter / X (Vicinitas.com). Thus, by varying the collection time, I obtained results that contain these different peak time frames. However, I wanted to make sure that I was getting a random start time for each day, so before I began the data collection process, I created a random number generator from 0 to 23 for 30 days for my first month of data. Later, I determined that I needed more data for specific regions (such as the U.S Markets and the International Indices), and I forewent this randomization process and chose a time that would allow the most Twitter / X data collection possible where the markets were open and overlapped with each other.

As with Case Study 2, I used the Filter Stream Twitter / X API [421] to collect my data using the same key geopolitical bigrams I developed from Goldstein [201], Caldara and Iacoviello [105], and Klement [265] (see Table 28 in Appendix A). The translated bigrams allowed for the capture of data in Spanish, French, Portuguese, Arabic, Japanese, and Korean, in addition to English. It is important to note that this is the same data that was collected in Case Study 2 from February 4th, 2023, to March 23rd, 2023, as my programs were designed to handle the real time multilingual sentiment analysis and topic modeling at the same time. However, there are two differences in how I used this tweet data for sentiment analysis. The first difference is how I used the data gathered. In the Case Study 2, I collected all the data in the storage location at an hourly interval. However, the real time sentiment analysis program needed a significantly smaller time frame. Thus, I created a new JSON storage file from every 100 tweets captured within that time frame. My multilingual sentiment analysis program took in the most recent JSON file in the storage location and ran every minute. Depending on the time frame or if there was a major geopolitical event, the same JSON file can be processed multiple times, however, this isn't a major issue as duplicates are removed in the final time sentiment series. Second, the Filter Stream API allows for tagging of tweets, thus I was able to tag the tweets that come in with their topic labels, which was crucial for the analysis when I combined the sentiment across the seven languages.

However, I should mention that while I use the term "real-time", the data processing I use for this case study is actually batch processing on a small scale. The 100 tweet file batch processing was chosen for multiple reasons. First, the 100 tweet batches allowed me to build more stable trends in changes in sentiment across the different geopolitical topics. If I were to process the each tweet in true real time, I would only get a positive or negative value in one topic. Thus having 100 tweets allows for larger sentiment scores to be captured per topic. Additionally, the 100 tweet batch allows for better understanding of geopolitical topics as they happen. For example, if North Korea launched a missile during the testing interval, I would see the shift in the count of the tweets toward "Nuclear War" from other geopolitical topics. While real time processing would be able to show this as well, the batch processing makes it more clear. Finally, the batch processing was necessary to create sentiment trends per topic that were able to compare to the trends in the changes of the prices of the financial assets I tested. The financial asset price data was at the minute level, thus constructing a similar trend that was also at the minute level was necessary. A batch of 100 tweets were processed at 67.4 seconds, thus making the batch method a more attractive method.

Section 6.4.2: Sentiment Analysis of Multilingual Tweets

The first step in the analysis is to generate an additional label for language for each tweet in the most recent JSON file. Thus, I used the Python package langdetect [143] which created that label and allowed for the break down of the JSON into separate language data frames for processing. With these separate language data frames, I then processed each language with a variety of methods. As described by Selven and Moh, "At present, there is no algorithm that can provide a hundred percent accurate results for sentiment analysis." (Selven and Moh, 2015, Pg. 1). However, I followed the guidelines from Karageorgou, et al., who said that for their multilingual sentiment analysis models they "ultimately adopted [sentiment analysis]-model we select is the classifier that presents high accuracy and as short a training phase as possible." (Karageorgou, et al., 2020, Pg. 7). Thus, I chose methods that presented the highest accuracy and speed trade off, as I wanted high accuracy but in addition low processing speed as I was dealing with real time data. I found this with the VADER sentiment lexicon developed by Hutto and Gilbert [238] which had high accuracy and speed for English tweets. For Arabic, I relied on CamelBERT developed Inoue et al [240], based on the BERT system [155] which has high accuracy and speed for English Natural Language Processing. Inoue, et al.'s CamelBERT is based on the BERT model, and focuses on Arabic Natural Language Processing problems such as sentiment analysis. However, for Spanish, French, Portuguese, Japanese, and Korean, I was not able to locate a comparable pre-trained system or lexicon that had both high accuracy and high speed for sentiment analysis. Thus, I created my own Recurrent Neural Networks ("RNN") based on Géron's architecture

[194], who showed that RNNs provided both high speed and high accuracy for sentiment analysis. Once I had applied the different sentiment analysis techniques across the seven languages, I recombined the data frames into one. With this new data frame, I manipulated the data to get the count of tweets in each topic and the sum of the sentiment for each topic for this data frame. I then exported the count of tweets for each topic and the sum of the sentiment of the tweets for each topic into separate csv files (with the JSON file name and the time of the analysis as additional data variable tags for the data). I then re-imported these new csv files so I could automatically calculate the change in count and the change in sentiment comparing the most recent entry to the previous JSON entry. Lastly, I stored this change in two new additional csv files. This gave the base metrics (sentiment of tweets for each JSON file and change in sentiment for topic between files) for my geopolitical risk index.

Section 6.4.3: Comparison Between the Change in Geopolitical Risk Sentiment and Financial Markets

Once I had the sentiment and change in sentiment for each iteration of my algorithm, I shifted to obtaining the values of the different financial markets that are open during the 3-hour time window I was analyzing for that day. These are US Markets (S&P 500, MSCI, VIX, 2-Year US Treasury Bond Yields, 10-Year US Treasury Bond Yields, Defense ETF, Metals ETF), International Markets (FTSE 100 – UK, CSI 300 – China, Nikkei 225 – Japan, S&P Sensex – Indian), Foreign Exchange ("Forex") Markets (EUR-USD, USD-JPY, GBP-USD, USD-MXN, EUR-GBP), Commodities (Gold Spot Price, Crude Oil Spot Price, Gold Price Futures, Crude Oil Futures), Cryptocurrencies (Bitcoin, Ethereum, Chain Link, Algorand, Ripple). I also include, in Appendix K, a breakdown of what assets were active in six-hour intervals.

As mentioned earlier, since I am investigating the niche area of geopolitics, outside of major geopolitical events, sometimes I do not capture 100 tweets within the minute that between the algorithm iteration time. This creates duplicate values in my csv outputs. Through testing, I found that removing these duplicates was beneficial for the analysis, thus I removed them. Afterwards, I created three groups of topics in each of the four csv files: "All" which is all topics; "Goldstein", which are just tweets that were labeled either as the Goldstein Positive topic or Goldstein Negative topic, (I did this to provide a comparison to the first case study); and "Topic" which was all remaining non-Goldstein topics. I then compared the values of the relevant finance markets to the sum of the sentiment for each of the three topic groups based on the time stamps for both the financial data and the topic sentiment time series for the three-hour analysis window. With both sets of time series (financial markets and sentiment of topic), I ran correlation analysis and Granger causality analysis on all the different combination of topic groups and financial data to see if I obtained any relationship between them. I repeated this process for my study period to make sure that all financial markets had enough appearances in the 3-hour time window so I could calculate an average value of each of the correlation and Granger causality results.

Section 6.5: Results

I present the results from this case study of the sentiment analysis of geopolitical risk from Twitter / X at the real time / minute level from February 4^{th}, 2023, to March 23^{rd}, 2023, at three-hour intervals. I captured 345,759 tweets during this period that contained my key geopolitical bigrams. Each batch of 100 tweets was on average 35 KB. Processing time between batches was 67.4 seconds, however, I added a 30 second delay between batches to limit the number of duplicate files generated so the true processing time is roughly 37.4 seconds. On average 7,516 tweets collected per session, which is roughly

2.63 MB. Additionally, I was able to show why investigating geopolitical risk with a multilingual framework is important, as more than 50% of the tweets I collected were non-English. These findings support Karageorgou, et al., who stated: "Carrying out [sentiment analysis] in only one language increases the risk of missing essential information authored elsewhere." (Karageorgou, et al., 2020, Pg. 2). I also saw that different languages peaked at different capture time periods, as Vicinitas showed. The count of Japanese tweets was highest between 6pm and 6am US EST, while Portuguese was highest from 6am to 12pm US EST, and English was the highest between 12pm to 6pm US EST. The full results from the count of tweets are displayed in Table 36:

Time(US EST)	**Quarter**	**English**	**Spanish**	**French**	**Portuguese**	**Arabic**	**Japanese**	**Korean**	**Num-Tweets**
12am-6am	1	41,047	14,919	3,398	1,971	1,581	18,301	3,626	84,843
6am-12pm	2	37,743	15,291	2,893	6,631	2,385	10,120	2,936	77,999
12pm-6pm	3	62,226	26,645	5,454	3,543	2,917	10,057	2,297	113,139
6pm-12am	4	30,948	14,571	1,475	1,505	750	18,055	2,474	69,778
Total		171,964	71,426	13,220	13,650	7,633	56,533	11,333	345,759
Percentage		49.7%	20.7%	3.8%	3.9%	2.2%	16.4%	3.3%	100%

Table 36: This table shows the count of tweets I collected across the various languages I used in this study. I collected the most tweets for English, Spanish, and Japanese, and the least in Arabic and Korean

Additionally, using sentiment analysis across multiple languages I was able to get a picture of how the world was responding to geopolitical events. Figure 33 below displays the changes in the average sentiment for each day for the geopolitical risk topic groupings on Twitter / X.

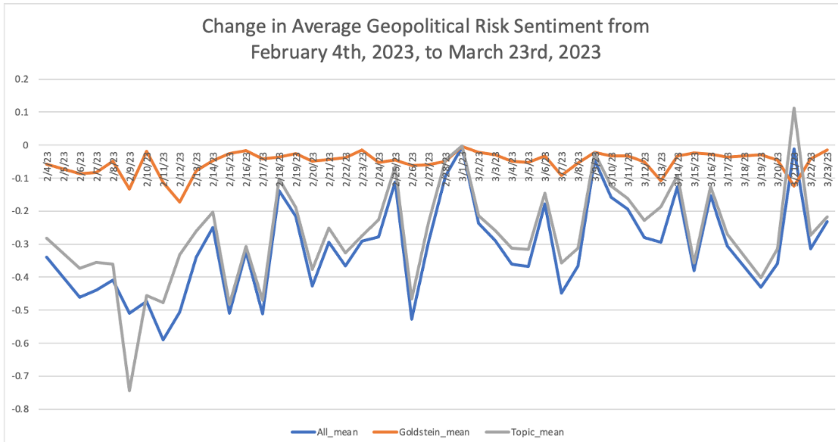


Figure 33: This chart shows the change in the average sentiment for the geopolitical risk groupings on Twitter / X, with the blue line is "All", the orange line is "Goldstein", and the silver line is "Topics".

As the above shows, the average sentiment was largely negative across all three grouping, which is in line with what both Schöne, et al. [380] and Acerbi, et al. [35] found that negativity spreads more on social media, especially Twitter / X, than positive news. Thus, I was unsurprised that average daily sentiment reached above zero only once with the "Topic" grouping on March 21st, 2023.

The comparison between my geopolitical risk sentiment index treads from Twitter / X and various financial markets and assets is explored below in Table 37. I chose assets that other studies (Caldara and Iacoviello [105] and Amen [55]) had investigated as well. In addition, I found a few of markets that I thought studying would potentially provide insights. Table 37 compares the average correlation of each day between the sentiment trends for each file for each topic grouping ("All", "Goldstein", and "Topic"), to each of the financial assets for the study period.

Market	All	Goldstein	Topics
SP500	-0.0446	0.0403	-0.0719
MSCI	-0.0381	0.0096	-0.0437
VIX	0.0459	0.0046	0.0471
Defense - ETF	0.0124	0.0378	0.0014
Metals - ETF	-0.012	0.0409	-0.0269
2Y - UST	0.1723	0.0817	0.1643
10 - UST	0.0581	-0.0348	0.07752
CSI-300	-0.0441	-0.0340	-0.0456
FTSE 100	0.0724	-0.0153	0.0847
Indian Sensex	0.0558	-0.0785	0.0847
Nikkei 225	0.0573	0.0236	-0.0326
EUR-USD	-0.0308	0.0030	-0.0330
USD-JPY	0.0215	-0.0017	0.0153
GBP-USD	-0.0353	-0.0162	0.0153
USD-MXN	-0.0179	-0.0011	-0.0187
EUR-GBP	0.0051	-0.0044	0.0114
Gold Futures	-0.0582	-0.0137	-0.0615
Gold Price	-0.0542	-0.0991	-0.0125
Crude Oil Price	-0.0043	0.0081	0.0033
Oil Futures	0.0235	0.0543	0.025
Bitcoin	-0.0446	-0.0568	-0.0176
Ethereum	-0.0496	-0.0796	-0.0335
ChainLink	-0.0092	-0.0242	-0.0012
XPR	-0.0218	-0.0373	-0.0245
Algorand	-0.0574	-0.0140	-0.0472

Table 37: This table shows the average of the correlation coefficients between the different markets and assets with the geopolitical risk index topic groupings of each day of the study period. Light blue is the U.S. markets, light green is the Forex markets, violet is commodities, gray is crypto currencies, while the multicolored grouping is various stock markets around the world.

For this case study, I followed Amen [55] who compared their geopolitical risk index against various financial assets using correlations. As the table above shows, unfortunately, I did not obtain strong correlations between the any of the three geopolitical risk index topic groupings and any of the financial markets shown on Table 37. The strongest correlation was with “All” and “2-Year US Treasuries” at around .17, which is still weak. I will discuss the differences between the methodologies in the discussion section for possible explanations of these findings. As for the change of sentiment over time (i.e., between the files) for each topic grouping, I found basically no correlation between these sentiment trends and the financial markets as shown by Table 38.

As Table 38 shows, I found that these changes in sentiment trends were basically uncorrelated with the trends of the financial assets as there was not a single average correlation existing outside of the range from 0.05 to -0.05.

Market	All	Goldstein	Topics
SP500	-0.0061	0.0197	-0.0194
MSCI	-0.0037	0.0018	-0.0058
VIX	-0.0036	-0.0021	-0.0012
Defense - ETF	-0.0233	0.0214	-0.0337
Metals - ETF	-0.012	0.0409	-0.0269
2Y - UST	-0.0032	0.0021	-0.0047
10 - UST	0.0105	0.0129	0.00772
CSI-300	-0.0051	-0.0322	-0.0032
FTSE 100	0.0076	0.0037	0.0066
Indian Sensex	0.0226	-0.0243	0.0240
Nikkei 225	-0.0103	0.0443	-0.0185
EUR-USD	-0.0026	-0.0188	0.0053
USD-JPY	0.0102	-0.0001	0.0095
GBP-USD	0.0228	0.0108	0.0221
USD-MXN	-0.0167	0.0120	-0.0171
EUR-GBP	0.0175	-0.0145	0.0183
Gold Futures	-0.0004	-0.0167	0.0044
Gold Price	0.0238	-0.0230	0.0272
Crude Oil Price	0.0162	0.0006	0.0211
Oil Futures	0.0077	0.0183	0.0065
Bitcoin	0.0176	-0.0137	0.0159
Ethereum	0.0184	-0.0274	0.0241
ChainLink	0.0129	-0.0274	0.0169
XPR	0.0126	-0.0247	0.0182
Algorand	-0.0058	0.0129	-0.0084

Table 38: This table depicts is the average correlation of change of sentiment for each topic grouping over time (between each file) compared to the financial assets I studied

While the correlation analysis did not produce any strong correlations at the real time / minute level when compared to the results from Amen [55] or Caldara and Iacoviello [105], I was able for obtain similar correlation signs for a majority of the financial assets as the two other papers studied. This is exhibited in Table 39.

Market	All	Goldstein	Topics
SP500	Yes	No	Yes
MSCI	Yes	No	Yes
VIX	Yes	Yes	Yes
Defense - ETF	Yes	Yes	Yes
Metals - ETF	Yes	No	Yes
2Y - UST	No	No	No
10 - UST	Yes	No	Yes
CSI-300	Yes	Yes	Yes
FTSE 100	No	Yes	No
Indian Sensex			
Nikkei 225	Yes	No	Yes
EUR-USD	No	Yes	No
USD-JPY	No	Yes	No
GBP-USD			
USD-MXN			
EUR-GBP			
Gold Futures	Yes	Yes	Yes
Gold Price	Yes	Yes	Yes
Crude Oil Price	Yes	No	No
Oil Futures	No	No	No
Bitcoin	Yes	Yes	Yes
Ethereum			
ChainLink			
XPR			
Algorand			
Count Yes	12	9	11
Total Count	17	17	17
Percent Matched	71%	53%	65%

Table 39: This table displays if the signs between my analysis and the analyses completed in Amen [55] and Caldara and Iacoviello [105] matched. At the bottom, the percentage of matching signs is shown. Any blank spots in the table are because they are the assets I chose to look into and do not have a comparison with another paper

I found that the “All” topic grouping had the largest number of same signs with the other papers. Interestingly, while the “Goldstein” grouping was focused on geopolitical events exclusively, it had the lowest matching percentage. Additionally, while it seems that the “All” and “Topic” topic grouping had very similar results, including the “Goldstein” topics in the “All” grouping, altered the results enough for the “All” grouping to match slightly better. “Goldstein” topics also matched the correlation signs for the Foreign Exchange markets, while the other two groupings did not.

Lastly, I wanted to see if I could use Granger causality analysis to see if there was any relationship between the geopolitical risk index trends and the financial assets at different lag values. Granger causality tests if one time series contains predictive information about the other, thus I utilized the test to see if a change in the geopolitical risk index trends occur before the change in the value of the financial market [55, 105]. I chose to explore five days toward the end of my case study window to see if I could find any values that could reject the null hypothesis of the Granger causality test and prove that the change in my index trends occurred before the change in the financial market displayed in Table 40.

Market-Topic-Combination	Day 37	Day 38	Day 39	Day 40	Day 41	AVG
Gold-Price-All	0.0056	0.1914	0.2342	0.2787	0.069	0.1558
USD-JPY-Alll	0.0259	0.6839	0.2172	0.4251	0.5965	0.3897
Gold-Futures-All	0.0088	0.6205	0.045	0.3464	0.163	0.2367
SP500-Goldstein	0.0348	0.3221	0.2821	0.0815	0.1393	0.1720
Metals-Goldstein	0.0261	0.5492	0.5868	0.0658	0.2641	0.2984
USD-MXN-Goldstein	0.041	0.1081	0.2149	0.3338	0.0365	0.1469
Gold-Future-Goldstein	0.0073	0.3658	0.0063	0.0065	0.0264	0.0823
Gold-Price-Goldstein	0.0151	0.0182	0.0152	0.1547	0.0629	0.0532
Oil-Price-Goldstein	0.1032	0.0437	0.0292	0.0106	0.2706	0.0915
Gold-Price-Topic	0.0438	0.0825	0.0792	0.4671	0.0949	0.1535
Oil-Price-Topic	0.0287	0.2666	0.6781	0.0273	0.1089	0.2219

Table 40: This table presents the results for the various Granger causality tests, testing whether a change in the different geopolitical risk trends occurs before a change in the financial markets. The blue highlight is the best results from the tests.

I used the time series of the sentiment for each file as the geopolitical risk index for the tests because as the above results showed, the change in the sentiment between the files did not produce any useable results. The table above shows the various combinations that passed the Granger causality test on Day 37 at the maximum allowable lag. I then tested these combinations on Day 38 – 41 to obtain an average for the result of the tests. I found that while some days the relationship between the geopolitical risk index and the financial markets passed the Granger causality test, there were no combinations where this was a consistent result. The only combination that was close on average was between the "Goldstein" geopolitical risk index trend and "Gold Spot Price", where the average result was only 0.00322 from rejecting the Granger Causality test's null hypothesis and thus having predictive information.

Section 6.6: Discussion

The methodology for analyzing changes in real time geopolitical risk was successful. I captured hundreds of thousands of tweets in this study in seven languages and was able to analyze them for their sentiment in real time and build a proxy for changing geopolitical risk. As the sentiment decreased the geopolitical risk would increase. An example is shown of my geopolitical risk index in Figure 34:

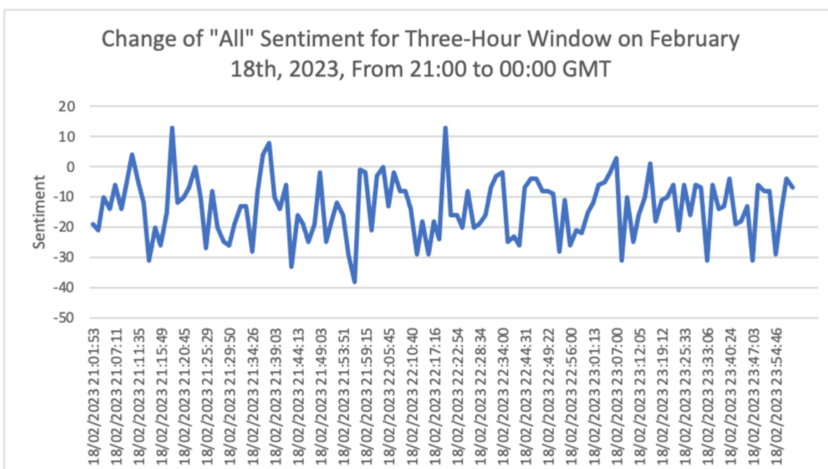


Figure 34: Change in the "All" topic groupings for February 18th, 2023, for the study period for the day, from 21:00 to 00:00 GMT

However, as Figure 34 shows, the change in the sentiment at the minute level is jagged and not consistent over time. This variation in sentiment could explain why the correlation analysis failed. Figure 35 compares the same sentiment data from Figure 34 with the change in Bitcoin prices for the same period.

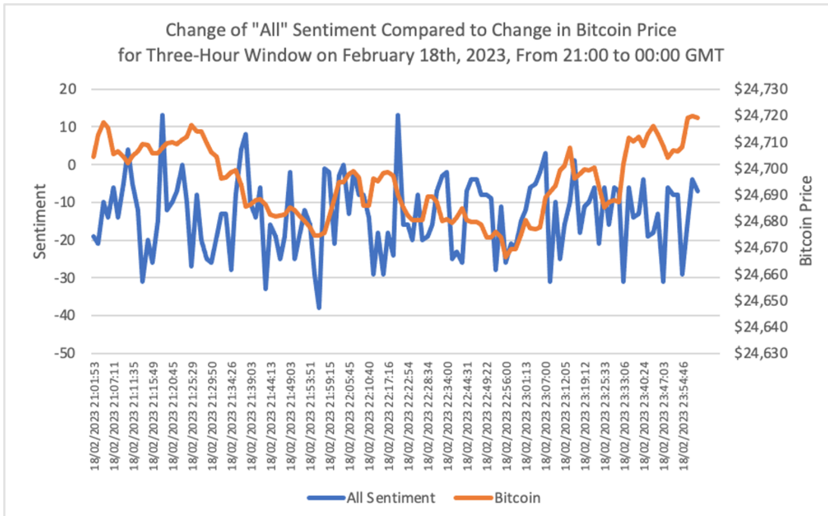


Figure 35: Comparison between the change in the "All" sentiment grouping and the change in the price of Bitcoin

Figure 35 displays how the Bitcoin prices are smoother over time than the change in sentiment, and the changes in sentiment don't correspond with the changes in prices over time either. In this example, I only obtained a correlation coefficient of 0.026, which is not much above zero, representing no correlation between the trends. As I repeated the correlation tests over each day, some days might have had a higher correlation incidentally (for example, I achieved a correlation coefficient of -0.45 for bitcoin price and "All" sentiment geopolitical risk index trend on February 10th), and the sign of the correlation might have changed. This led to an average correlation coefficient for this case study to remain low, and thus I could not conclude that there was a consistent strong correlation between any of the geopolitical risk index trend groupings and the financial markets. That said, I was surprised to see that the majority of the average of the correlation coefficients between sentiment groupings and the financial markets matched the direction of the correlations found in the other papers [55, 105]. This could mean that while I cannot say how much a financial market will vary with a change in geopolitical risk sentiment from Twitter / X, I can say that overall, the geopolitical risk indicator performs comparably to other methods in gauging the direction of the change in the financial market or asset.

Another interesting aspect for the real time research was that certain languages actually produced slightly higher correlations individually than when combined as they were in the main analysis. I collected three days of Spanish only data to test my hypothesis that perhaps certain languages' sentiment had greater predictive information for certain financial assets. Unfortunately, while certain days Spanish did have higher correlations and performed better on the Granger causality test for a few assets, for example the Eur-Gbp exchange rate, on average the Spanish language trends followed the combine language trends. I believe this is because of the factors that plagued the combined real time analysis. There could be a few solutions to this issue across all languages. One could be to aggregate the tweet sentiment data at ten minute intervals to create smoother sentiment trend lines that might correspond better to the financial asset tread lines. Additionally, for individual languages, one could only investigate financial assets and markets more relevant to the areas were the languages are most prevalent, for example look into Arabic and Oil price and Oil Futures. This closer relationship could also improve the average correlation and Granger causality results as there is already an underlying connection between the two trends.

As for the Granger Causality Analysis, I believe that there are a few reasons why I achieved these results. First, like the correlation analysis, the greater variation in the sentiment grouping time series over time for the day, and the difference in the time series between the days, could lead to vastly different results for the Granger Causality analysis when compared to financial markets and assets that generally have smoother trends over time. Another issue could be that the market does not respond as quickly to the changes in geopolitical risks as the changes I captured in real time. As seen in the Case Study 1, some assets needed hours, if not days of lags after the change in sentiment for have predictive information about the change in the asset. For example, both Amen, and Caldara and Iacoviello, analyzed their data at the monthly and daily levels, with Amen achieving their correlation results from their monthly data analysis (Amen, 2020, Pg. 3). This showed that it is possible that at the real time comparison, not enough time has passed for the change in sentiment to be reflected in the change in financial assets.

Chapter 7: Named Entity Recognition, Geocoding, and Mapping

Section 7.1: Introduction

So far in my thesis I have mainly described employing sentiment analysis and topic modeling techniques to analyze the data from Twitter / X. Case Studies 1, 2, and 3, all featured different ways that my programs can analyze geopolitical risks at varying time frames with these methods. However, there is one additional analysis type that rounded out the development of automatic evaluation of geopolitical risk. This last process is the combination of Named Entity Recognition ("NER"), Geocoding, and Mapping to build a visualization for changes in geopolitical risks.

Section 7.2: Key Concepts

As defined by Saldutti [375], NER is "a task of natural language processing, which identifies and tags entities within a text. These entities can be people, dates, companies, or . . . locations" (Saldutti, 2021). NER allows me to identify locations that are contained in the tweets which provide the first layer of the visualization. The next step is the geocoding, which Saldutti defines as "the process of finding spatial coordinates for locations" (Saldutti, 2021). This process is described in the methodology and gives the latitude and longitude for the locations I have identified in the geopolitical risk tweets. The final step is the mapping, by employing Folium [402] I make the map visualization for where the geopolitical risks are taking place colored with the type of geopolitical risk occurring there. This three-step process can also be referred to as geolocating.

Section 7.3: Related Work

Geolocating and NER have been used in several influential studies in combination with social media. For example, Sakaki, et al. [373] used the built-in geolocation data from tweets to build their real time Japanese Earthquake detector and Typhoon path predictor. They were able to show that through Twitter / X and geolocating, they could provide quicker earthquake detection than the U.S Geological Survey ("USGS"), a U.S. Government agency which tracks earthquakes. Hauff and Houben [217], on the other hand, combined Twitter / X and Flickr data to develop a geolocating system to increase the accuracy of the approximate location given by Flickr for an image posted on the site. Further, Fujisaka, et al. [180], made use of the Twitter / X geolocating data to track massive migrations of people around the world. These studies displayed the power of using built-in geolocated data, however, there is a significant downside for using the built-in geolocated data method for my research. First, as Sloan and Morgan [394] found "that approximately 0.85% of tweets are geotagged, meaning that the exact position of where the tweeter was when the tweet was posted is recorded using longitude and latitude measurements" (Sloan and Morgan, 2015, Pgs. 2 – 3). This means in Chapter 6, Case Study 3, for example, with an average collection of approximately 7,500 tweets per collection period, only 64 of those tweets would have a location included in the tweets' metadata. 64 tweets are far too few for effective data analysis, thus I turned to NER.

NER suited the needs of my research better for a number of reasons. First, NER is less invasive than using the direct geolocation metadata of the person tweeting. This Twitter / X metadata is attached to the account of the person tweeting, which could include the "precise location data in the form of latitude and longitude" (Sloan and Morgan, 2015, Pg. 2). Obtaining the exact location of a user's home would be hard to justify ethically if it is not needed for the research, and since I do not need or want that information, I did not gather that information. Additionally, the location a user tweets from might not be relevant for the research. For example, a person in Mexico might be tweeting about a nuclear threat in North Korea. With NER, this geopolitical risk would be correctly filed for North Korea. However, if the geolocation metadata from Twitter / X was used, this event would appear as a nuclear threat risk in Mexico, which is inaccurate. Thus, obtaining locations from the tweets' text directly using NER was the more appealing method. Other researchers have used this method as well, which provided a basis for me. Ritter, et al. [363] described a methodology for using NER with Twitter / X. Urbain [432] created their own NER algorithm for identifying the heart disease risk factors in text. Balasubramaniam [69] used a combination of NER and geolocating to get the proper names of the locations obtained through the geolocated Twitter / X user information. Finally, Saldutti [375] provided the beginnings of the methodology with NER, Geocoding, and Mapping that I used for my own research.

Section 7.4: Methodology

As shown in Chapters 4 – 6, I explored two different time frames for the data collection from Twitter / X. The first was historical time series data described in Chapter 5, Case Study 1, looking into the change in geopolitical risk surrounding the Ukraine War. The second was real time data collection and analysis described in Chapters 5 and 6: Case Study 2 and Case Study 3. However, for the purposes of NER, both types of data analyses followed the same methodology described here.

For my thesis, seven languages were used to analyze global geopolitical risk. The NER portion of the analysis was included to add location for more context for geopolitical events. This was done in two ways. One method involved checking the change in the count of the mentions of locations in the tweets which I obtained through NER methods. The other method involved mapping the locations of certain geopolitical events through geocoding to obtain location's coordinates. The program's procedure outlined in Chapter 6, details that the sentiment analysis and the NER analysis occur at the same time. Thus, once the tweets are captured for the various geopolitical topics, the tweets are broken into their separate language data frames, where the tweets' sentiment is first calculated. Once the sentiment is calculated, I applied NER on the tweets in various ways. For English, Spanish, French, Portuguese, and Japanese, I employed the Spacy system developed by Honnibal and Montani [238]. This system can break down text and label the parts of the text for various entities such as Organization, Person, Location, etc. However, there are differences between the labeling of the languages with Spacy. For English and Japanese, these languages had the Spacy label for "GPE" meaning "Geopolitical Entity", while Spanish, French, and Portuguese only had the label for "LOC" for "Location". The difference is that the "GPE" is more precise labeling of location than the "LOC" label, which is more expansive. However, this isn't a major issue, as the only problem is that "LOC" gives more data than necessary, which is removed later in the process.

For Arabic and Korean, at the time of my programs creation, Spacy did not have NER models that were effective for the needs of my programs, thus I had to look elsewhere. Luckily for Arabic, Hatmimoha [216] created a pre-trained NER model which contained a "location" tag that worked for the system. While the pre-trained model was slightly slower than Spacy, it was just as accurate and provided easy integration into my programs. However, for Korean, I was not able to locate a pre-trained NER model that integrated with my programs, so I constructed my own. First, I took the original Korean tweet text, and applied a Korean tokenizer to split the Korean text [178, 330] into individual words. This process created a new data frame for each tweet. Next, I used a list of countries in Korean obtained from Gabos [183], to cross-reference the Korean countries with each of the new tokenized data frames for the Korean tweets. If there was a match between them the country's name was added to the original Korean tweet analysis data frame in a new variable column for location. While my Korean NER system worked well, it was unfortunately limited to only country level NER labels as the data set I use to cross-reference the Korean text only contained countries' Korean names. This creates a limitation if, for example, a tweet only mentions capitals like London, the Korean system will not label a location for that tweet.

The next step was geocoding the data to obtain coordinates for the locations mentioned in the tweets. This process took place in the individual language data frames and was obtained through the Google Maps Geocoding API [312] with their geocoding feature. This process took place in the individual language data frames as the Google Maps API can work in multiple languages. The Geocoding API works by taking the location and returning the coordinates to the most precise degree possible. For example, if a country is inputted, it will return coordinates for the geographic center of that country. However, if a city or a district is inputted, it will return the more specific coordinates for that city or district. The latitude and longitude are collected for each location and appended to the main data frame for the language. Unfortunately, due to the data error between "LOC" and "GPE" discussed earlier, while "GPE" tagged words generally have coordinates, "LOC" tagged words frequently do not (for example, the common phrase "RT", which means "retweet", appears in tweets, and is tagged as a "LOC", but does not have any real-world location). In these situations, the coordinates are given as [0, 0], which is shown in the mapping examples in this chapter's Discussion Section. However, this created two limitations for my methodology with the NER analysis. Only English and Japanese have the more precise "GPE" labeling, while Spanish, French, and Portuguese, all have the "LOC" label. While this was not a major issue in the analysis, as the “LOC” labels mislabeling of certain text as location was dealt with, it is still a limitation that could not be fixed. Sometimes if there are multiple locations with the same name (for example, San Francisco exists in both the U.S. and Venezuela), the Google Maps API [312] will only choose one location to return the coordinates for, potentially putting the geopolitical risk in the wrong place. This only occurs in a small number of cases, so it doesn’t affect overall efficiency of the analysis, but it requires mention here.

After obtaining the coordinates, the separate language data frames are recombined. At this point, the sentiment analysis and NER analysis separate into two data frames. For the sentiment analysis, the duplicate tweets generated from the NER process are dropped, however, for the NER analysis, the duplicates are kept as many tweets can discuss more than one location. With the NER data, the next step was to translate the location’s name into English using Google Translate [321] for the mapping display. Once translated, the countries were grouped and a count of each country was obtained, which was subsequently exported for future analysis displayed in the Discussion section. Finally, the last step was to create a map using the Folium python package [402], which shows the geopolitical risk tag for each location mentioned in the tweets. However, sometimes if there are multiple locations with the same name (for example, San Francisco exists in both the U.S. and Venezuela), the Google Maps API [321] will only choose one location to return the coordinates for, potentially putting the geopolitical risk in the wrong place. This only occurs in a small number of cases, so it doesn’t affect overall efficiency of the analysis, however, it should be noted.

Section 7.5: Results

While the Methodology section describes the general pathway for the NER analysis involving geopolitical risk tweets, there are a few distinctions when it comes to the historical data analysis and the real-time data analysis.

Section 7.5.1: Historical Analysis

For the historical data, since I already had the entire dataset, the data could be processed all at once at the daily level. While the change in geopolitical risk sentiment for this period is discussed in greater detail in Case Study 1, I felt that Figure 36 below presents a different aspect of the results I was able to capture, adding further context to the results.

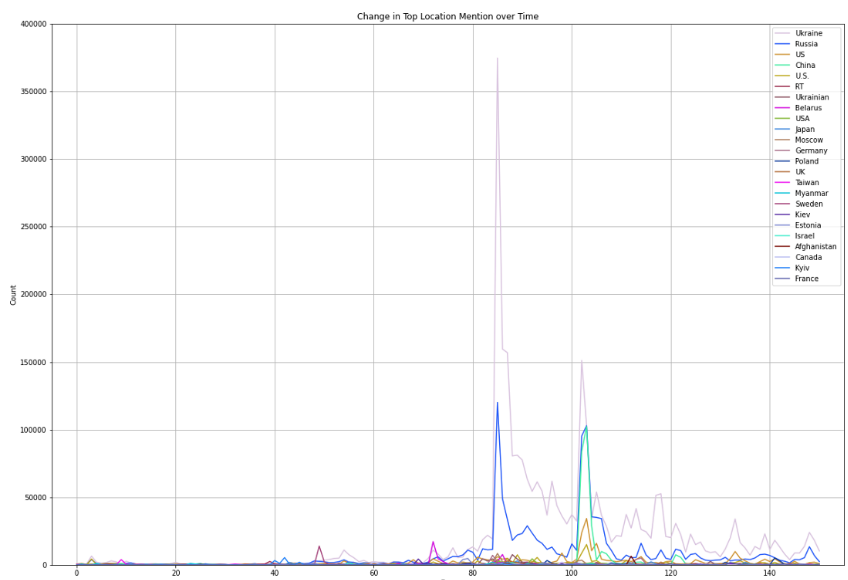


Figure 36: The Top 24 Locations by mention count in tweets gathered through NER analysis for the period from December 1st, 2021, to April 30th, 2022

As figure 36 shows, the Ukraine War is by far the most important geopolitical event that took place during the study period. However, compared to Case Study 1 in which the sentiment alone does not provide further details past, that there was a major geopolitical event that shifted sentiment. However, I do not know further information, for example, of where the event took place. Topic modeling could also provide similar insights on events with the greater context it provides, but one of the downsides of topic modeling is that some geopolitical risks can slip under the radar if there is a major occurrence which takes up all of the available topic space that can be generated. With this additional NER analysis, it is possible to gain additional understanding on smaller but still important events. As shown below in Figures 37 – 39, there was a major shift in Taiwanese diplomacy on December 10th, 2021, as Nicaragua stopped recognizing Taiwan [315] shown in Figure 37:

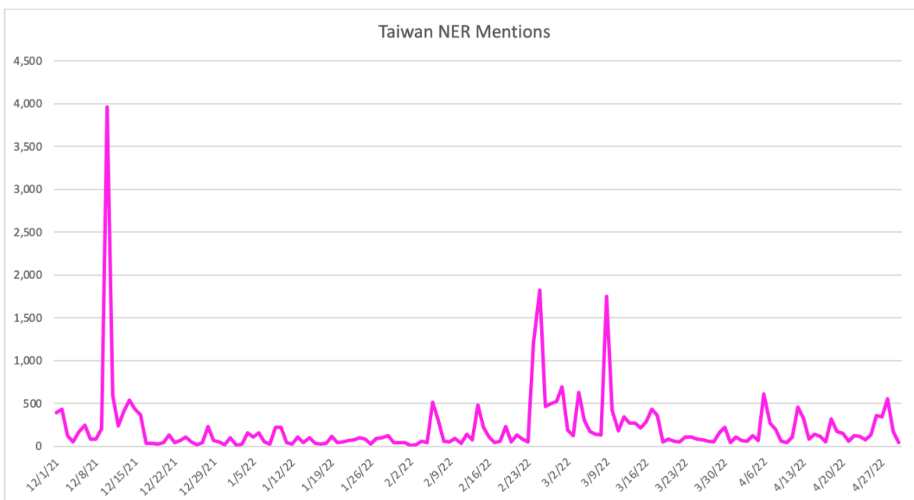


Figure 37: Taiwan NER Mentions in the historical data from Case Study 1 from December 1st, 2021, to April 30th, 2022. The color was chosen to match the color from Figure 36

In addition to Taiwan, major events were taking place in Afghanistan, shown in Figure 38, which at beginning of the study period for Case Study 1 had only been under Taliban control for four months.

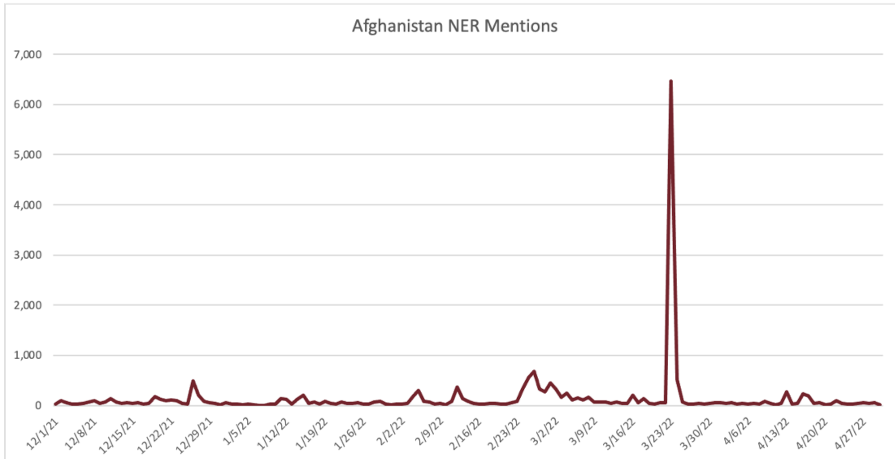


Figure 38: Afghanistan NER Mentions in the historical data from Case Study 1 from December 1st, 2021, to April 30th, 2022. The color was chosen to match the color from Figure 36

On March 23rd, 2022, the Taliban announced that they will close all girls' schools, thus not allowing girls above the age of 11 to attend school [247]. This caused an international outcry about the Taliban curtailing women's rights after they promised not to do so. These events are very important geopolitically, but when compared against the Ukraine War, shown in Figure 39, they are dwarfed by the conflict.

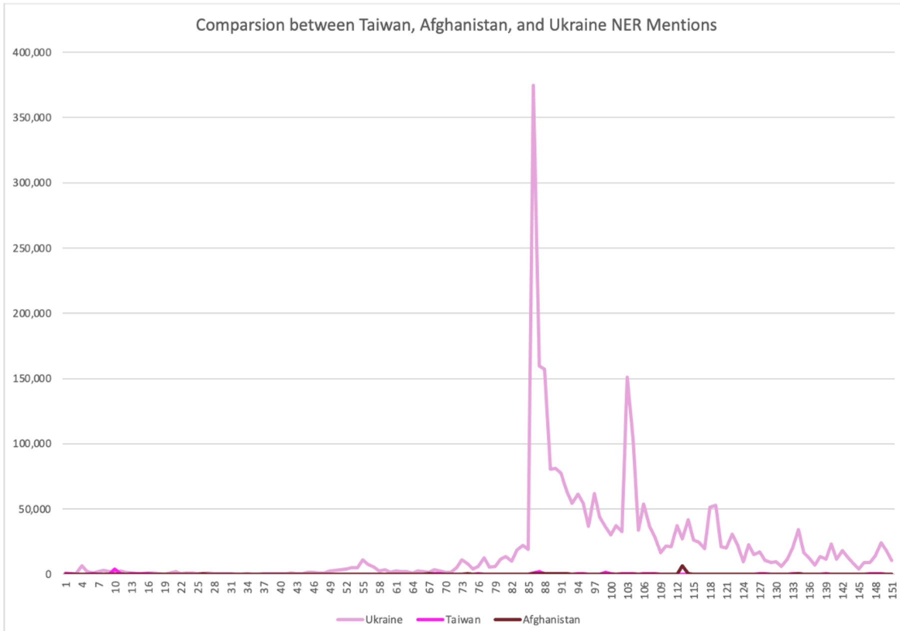


Figure 39: Taiwan, Afghanistan, and Ukraine NER Mentions in the historical data from Case Study 1 from December 1st, 2021, to April 30th, 2022. The color was chosen to match the color from Figure 36

With the advent of the Ukraine War, all other geopolitical events lost precedence. At the start of the Ukraine War, Ukraine received more than 350,000 mentions, while the peaks of Taiwan and Afghanistan reached almost 4,000 and 6,000, respectively. These signals would have been drowned out if only a topic modeling analysis was used. This discrepancy shows the importance of including NER in the geopolitical risk analysis as it can provide information that sentiment analysis and topic modeling might not be able to capture, thus providing a fuller picture of the events that could change risk profiles when looking at historical data surrounding major geopolitical risks.

Section 7.5.2: Real Time Analysis

For the Real Time Analysis discussed in Case Study 2 and Case Study 3, NER analysis also provides meaningful results, as shown in Case Study 2, given the constraints of collecting geopolitical risk data from Twitter / X in real time. For example, with topic modeling sometimes there is not enough data collected to develop coherent topics under an hour interval. However, as mentioned in the Historical Analysis section, while NER cannot provide the same level of information as topic modeling it can provide a decent stop-gap measure between analysis runs of the program's topic modeling algorithm. Following the same methodology described above, except with the smaller real time data files, I can calculate changes in location mentions through NER between these data files. For example, from data gathered on February 6$^{\text{th}}$, 2023, from 11:30am to 2:30pm US Eastern Standard Time (EST), Chile was one the countries mentioned the most in this time frame. The change in mentions is shown in Figure 40 below:

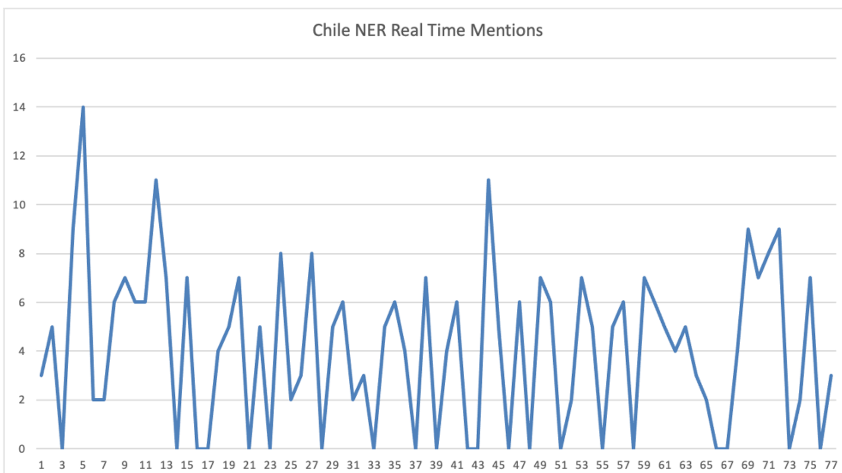


Figure 40: Chile NER Real Time Mentions, each data point represents the count of Chile mentions in the tweets in the data files of 100 tweets captured by the real time program described in Case Study 2 and Case Study 3

While in the real time analysis I can't know why Chile is popular without further investigation, it does provide insight into potential future potential topics that might develop during this period. This is shown by the Spanish Topics created on this day, which had several containing Chile:

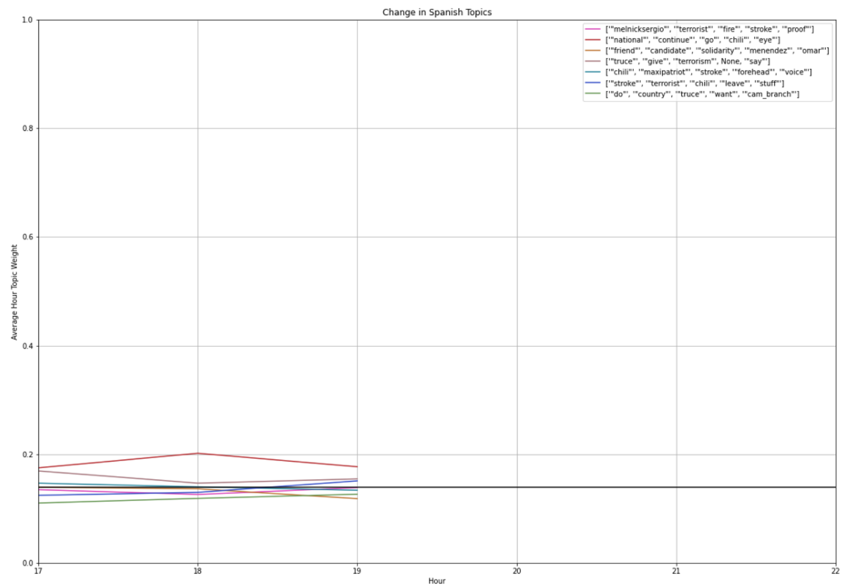


Figure 41: The change in the emergence and growth of the topics created from Spanish tweets on February 6th from 11:30am to 2:30pm US EST. It should be noted that the news involving Chile is in reference to massive wildfires taking place in Chile at this time and the international aid they are receiving [107]

In addition to this change in mention NER analysis, another real time analysis that was implemented was the mapping portion of the analysis. Figures 42 – 44 display the maps created during the real time analysis. These maps were created on March 17th, 2023, with the first created at the start of the analysis period at 10:20am US EST:

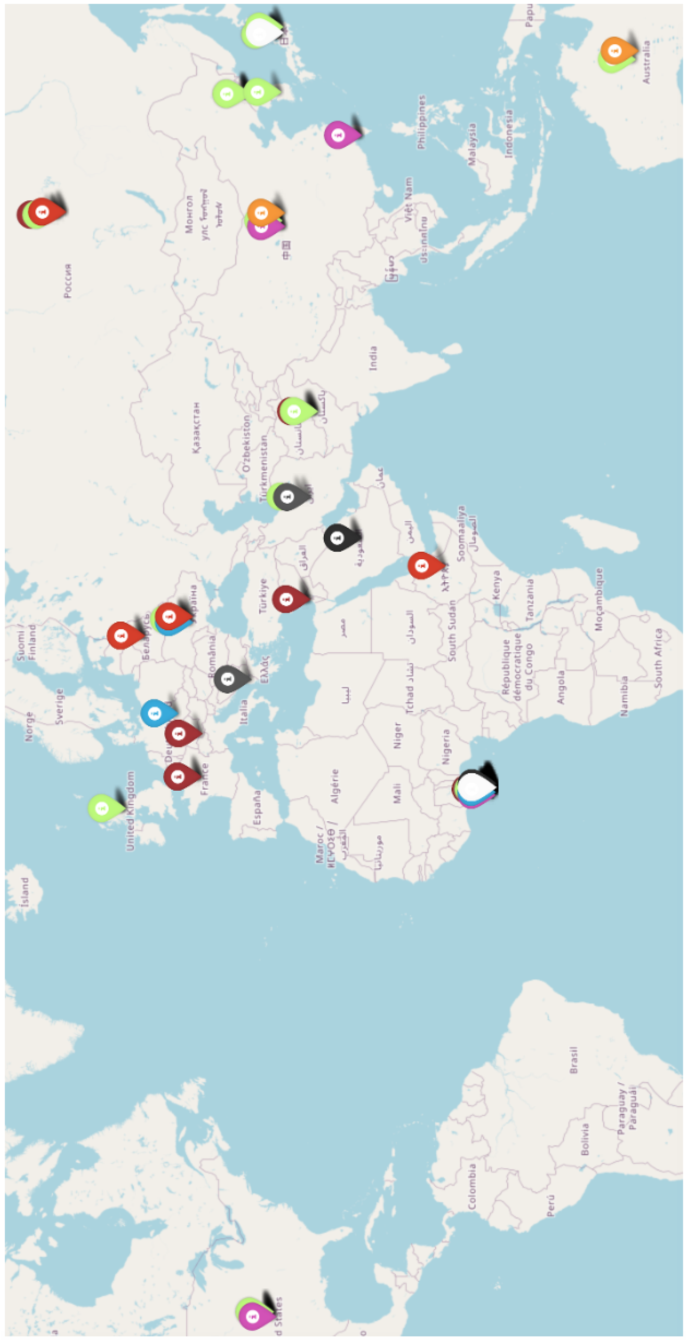


Figure 42: Map 1 created at 10:20am US EST, the start of the study period for March 17$^{\text{th}}$, 2023, this map was created with Folium [402]

The above map provides both the location of geopolitical risk events and the type of geopolitical risk event. An unfortunate drawback to Folium is that there is no way to add a legend to the map, so I have included a table of the colors and geopolitical risk category below for clarity. However, Folium does provide a fully interactive html output file, copies of which will be in the supplement materials where the labels display the geopolitical risk they represent and the ability to zone into specific regions of the world.

Color	**Geopolitical Risk Topic**
White	Error or No Location
Red	Goldstein Negative
Blue	Goldstein Positive
Light Green	Nuclear Threat
Gray	Cyber Warfare
Black	Oil Supply Shock
Purple	US-China Relations
Dark Red	Terrorism
Orange	War Threats
Dark Green	General Geopolitical Threats

Table 41: The Legend for the different maps generated with Folium, the colors are an approximation but are close to the colors displayed on the map

For the mapping process, some of the labels of location needed to be shifted if multiple geopolitical events occurred in the same location, such as "Cyber Warfare" and "Nuclear Threat" with Iran shown in Figure 42, the labels may obscure each other, thus the offset was needed. This isn't a major problem for countries, but for specific locations, like cities, this difference might cause the mapping to lose exactness. Lastly, the large number of labels at [0, 0], off the west coast of Africa, are the leftovers from the locations from the NER analysis that did not have coordinates and could not be removed.

Figure 43, below shows a map generated halfway through the study period for March 17$^{\text{th}}$, at 11:12am US EST:

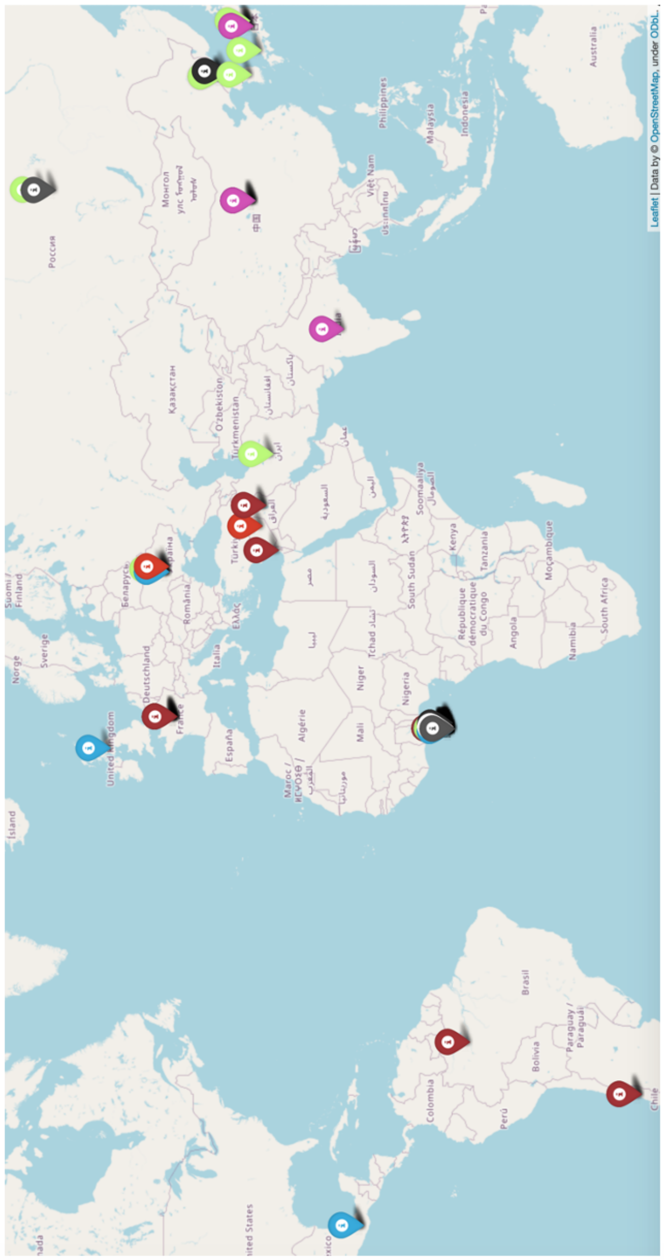


Figure 43: Map 2 created at 11:12 am US EST, the start of the study period for March 17$^{\text{th}}$, 2023, this map was created with Folium [402]

And Figure 44 of the final map created for the study period at 12:05pm:

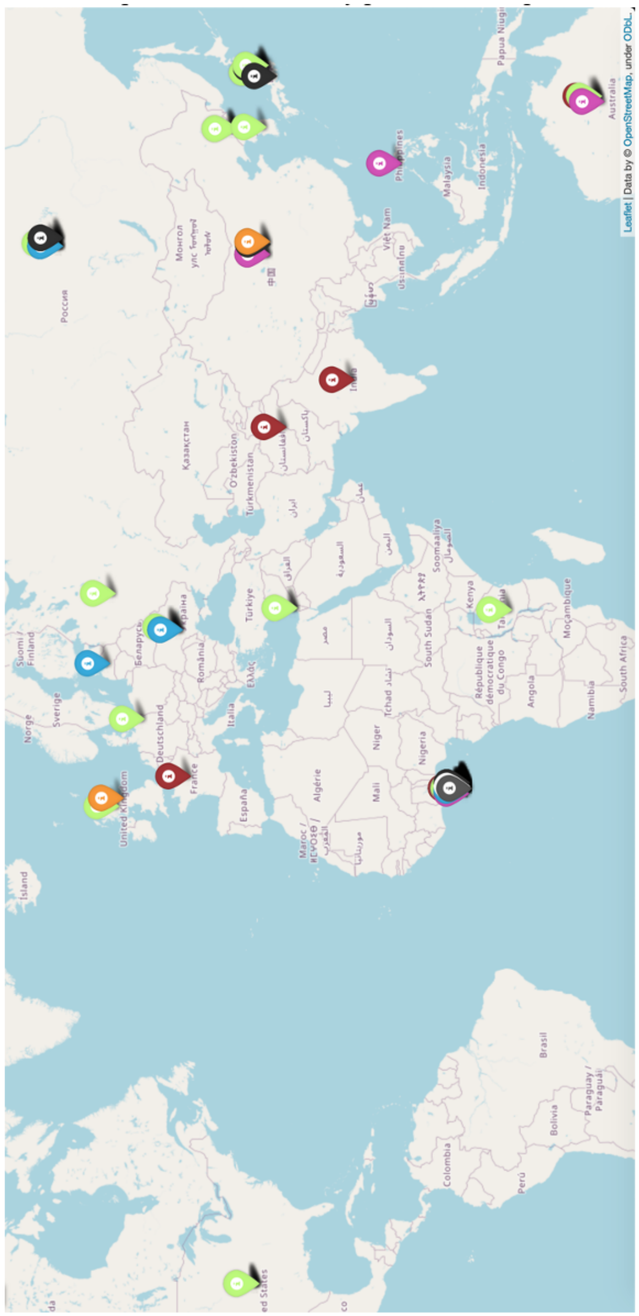


Figure 44: Map 3 created at 12:05pm US EST, the end of the study period for March 17th, 2023, this map was created with Folium [402]

While the previous charts and tables contained information about locations of geopolitical events, these maps provided much needed visualizations to show the changes in global geopolitical risks. For example, during the study period, North Korea was testing and firing Intercontinental Ballistic Missiles ("ICBMs") and other potential nuclear deployment weaponry [148]. While this is shown in Case Study 2 through the North Korea analysis and the Japanese and Korean Topic Models, one can also see how much of a potential geopolitical hot zone this region is with the number of nuclear threat labels appearing consistently on each of the maps in Japan, South Korea, and North Korea. Additionally, these maps reflect the ongoing Ukraine War and increased tension between China and its neighbors. With mapping providing the visualizations as the final piece in the automatic geopolitical risk analysis puzzle it is possible to gain a better understanding of the changes in geopolitical risk in real time.

Section 7.6: Discussion

While this NER analysis was not directly linked to previous aspects of the research of my thesis, it provides invaluable context and visualizations to help better understand the changes in geopolitical risk, and improves the evaluation of said risk when analyzed automatically. NER provided greater information than just sentiment analysis alone, and greater context than just topic modeling alone, as I can visualize the hotspots for geopolitical risk events differently. However, while mapping is a useful tool, it is potentially a very expensive one, as the geocoding portion of the Google Maps API will only allow 40,000 free geocoding requests per month [312], which for the real time data collection is only about four to five days of data on average. After the free period, if an average month of data collection has around 300,000 geocoding requests, that works out to approximately $1,100 USD a month or $13,200 USD a year. That said, while the mapping visualization might be too expensive, there can still be valuable knowledge gained from the changes in the NER counts displayed in Figure 36.

Chapter 8: Discussion

Section 8.1: Introduction

This chapter summarizes the findings of the various geopolitical risk analysis completed through my programs using sentiment analyses, topic modeling, and NER that comprise the automatic evaluation of geopolitical risk. It describes the context in literature, the key findings and the potential implications, some of the limitations faced in the research process, explanations for some unexpected results that were encountered, and finally some avenues for future work based off my research.

Section 8.2: Context in Literature

Section 8.2.1: Novelty

To the best of my knowledge, as of this writing, this is the first research to use social media exclusively to analyze geopolitical risk. Additionally, the real time analysis of Case Study 3 is the first to analyze changes in geopolitical risk at the real time level. The commercial enterprise, GDelt [7], updates every 15 minutes, my program can update in under a minute if needed. Also, the programs work with seven languages, and does not translate them into English, which is another novelty not present in other systems. Analysis in the original language was done to obtain a more accurate picture of sentiment regarding geopolitics and topics generated around the world. Lastly, my programs operate on their own and automatically: as soon as it is started it requires no additional input from the user to obtain a breakdown of changes in sentiment involving various geopolitical risks that create the index, emerging topics in various language regions around the world, and a count and geographic mapping of the countries involved in these risks.

Section 8.2.2: Key Findings and Implications

My key findings led to a few potential implications in reference to the existing literature. First, I found that with sentiment analysis, it is generally better to analyze the text in the original language, if possible, over a translation to English. These results match the research de Koning [149] performed concerning Dutch, which showed that there was a higher level of accuracy with an RNN model trained on Dutch versus translating the Dutch to English and then using an English model. While my analysis is not the final discussion on this topic, it does provide insight into the best practices for sentiment analysis across multiple languages. That said, if the data can reach the "gold standard" quality with the use of human annotators, then machine translation to English in combination with VADER will most likely provide decently accurate sentiment analysis results. Although if the data is of lower quality or not very definitively positive or negative, then it would be best to create a machine learning model for the required language as the model will capture the nuances in the text better than a translation based method can.

Another implication comes from the research into the changes in geopolitical risk surrounding the start of the Ukraine War. Building off the work of Bollen et al. [91] and Goldstein [201], through the use of Twitter / X, the Goldstein Index, which I developed from Goldstein's work, and sentiment analysis, and Granger Causality methodology that Bollen et al. employed, not only was I able to effectively identify the start of the Ukraine War but I was also able to show the increased negative sentiment that lasted months afterwards in terms of the increased geopolitical risk it caused. This negative sentiment was also reflected in various financial assets and markets. Some immediate effects, like with many of the Foreign Exchange Rate markets changed after change in sentiment after only a few hours. In contrast, other markets, such as the Nikkei 225 took two weeks of trading before the change in the sentiment provided predictive information. While this analysis took place on the daily level like that of Caldara and Iacoviello [105] and Amen [55], and I was able to show that looking exclusively into geopolitics like these studies do that my methodology also showed that the changes in geopolitical risks have predictive information for the changes in various markets.

Building off the work of Zheng et al. [468] and Sakamoto et al. [374], I was able to show how it is possible to track emerging geopolitical topics at the regional level using language as a proxy. Not only did my research into this area support the findings of Zheng et al. and Sakamoto et al., but expanded on them by allowing for the dynamic selection of the number of topics and the other hyperparameters for LDA to obtain better quality topics. I also showed by using the work of Besaw [82] as a basis that it is possible to do emerging topic modeling, at not just the yearly level which Besaw used but at the daily and hourly levels as well. This reduction in the time interval also implied that if more data can be gathered it would also be possible to investigate emerging geopolitical risk topics at smaller time frame.

Finally, my research provides new insight into the relationship between geopolitical risk and the financial markets at the real time level. Synthesizing the work of Selvan and Moh [384] and Karageorgou, et al. [257] together, I showed that it is possible to analyze changes in sentiment regarding geopolitical risk across multiple languages in real time. This change in sentiment was then compared against the changes in financial assets similar to the work of Caldara and Iacoviello [105] and Amen [55]. While the results I found did not have as strong of a correlation between the assets that both Caldara and Iacoviello and Amen found, I did find that the majority of the assets had the same direction of correlation as these studies found. This matched what I had found earlier in the daily level study in the Ukraine War, that changes in geopolitical risk might need a longer time horizon after they occur for the market to reflect the change. While I found the financial markets comparison results disappointing, I was pleased that methodology I had used was able to obtain multilingual tweets about geopolitics and I was able to apply the various sentiment analysis techniques to analyze these multilingual tweets in real time. This methodology developed the geopolitical risk index which created a way to analyze geopolitical risks in real time at a global level.

Section 8.3: Unexpected Results

There were a few unexpected results I found in my research that were very intriguing across the different research areas that I delved into.

Section 8.3.1: Multilingual Sentiment Analysis Unexpected Results

From the first research question for the multilingual sentiment analysis: would the Google Neural Machine Translator produce a more accurate translation than separate Transformer models trained on each language? I assumed that the transformer architecture of the OPUS-MT would provide translations that would have better captured the context of the data, as stated by Géron that the transformer "significantly improved the state of the art in NMT [Neural Machine Translation] without using any recurrent or convolutional layers, just attention mechanisms" (Géron, 2022, Pg. 554). Additionally, the OPUS-MT models are trained on the OPUS corpus, which is massive, "comprises 3.2 billion sentences with over 28 billion tokens in total" (Tiedemann, 2016, Pg. 9) [416]. The GNMT only had 36 million sentence pairs for the French to English dataset (Wu, 2016, Pg. 14) [450]. However, despite the difference in training data size and more advanced architecture, the GNMT outperformed the OPUS-MT in accuracy on both the "Gold Data" and "Full Data" across all three English sentiment analysis models. This may be due to the GNMT having a deeper training network than the OPUS-MT. The GNMT was eight layers deep in both the encoder and decoder with an attention connection between them. However, the OPUS-MT only six layers deep in the encoder and decoder. This increased depth of the network would have greatly improved the accuracy of the translator, in addition to Google constantly improving the GNMT and might have produced the results I saw.

The second question involved the evaluation for the best English sentiment analysis method between the VADER [238] lexicon, an RNN trained on English data, and a BERT model fined tuned for English Sentiment Analysis. I found that the VADER lexicon remained consistently the highest performing English sentiment analysis model for translated text. Despite the potential drawbacks described by Grimmer and Stewart [208] on using a dictionary-based sentiment analyzer, and the advancements in neural networks with RNNs and transformers with BERT, VADER did better on average. However, these outcomes make sense in context. VADER was specifically designed to evaluate short texts derived from social media. Thus, it might not be that surprising that it performed so well in this setting.

Section 8.3.2: Historical Analysis on the Ukraine War

When applying my geopolitical risk index based on sentiment analysis and compared it to the changes in different financial markets, I encountered a result that surprised me. In the Granger Causality analyses, only Bitcoin and Binance Coin ("BNB") appeared in any capacity out of the crypto currencies I investigated. As Baur, et al. find "Bitcoin is mainly used as a speculative investment" (Baur, et al., 2018, Pg. 2) [74], thus I assumed that Bitcoin and the other crypto currencies would experience a price change due to them being a "risky asset" (Amen, 2020, Pg. 6) with the massive change in sentiment generated by the start of the Ukraine War. While that was not the case, these results are in line with the findings of Rognone, et al. 2020 [367] which "suggest investor enthusiasm for Bitcoin irrespective of the sentiment of the news" (Rognone 2020, Pg. 1). This also matches the results from Abraham, et al. [34] who found that tweet volume was a better indicator of Bitcoin and Ethereum price changes than the tweet sentiment (Abraham, et al., 2018, Pg. 2). However, that said, their methodology collected tweets with keywords specifically for Bitcoin and Ethereum and only in English (Abraham, et al., 2018, Pgs. 8 - 9), so this difference in methodologies might explain the slight difference seen with the sum of sentiment for the "Goldstein Index" tweets time series having predictive information for the Bitcoin time series.

Section 8.4: Limitations

There were a few limitations I encountered in my research that require comment. These limitations break down into three groups: universal limitations, intransigent limitations, and limitations that may be overcome in the future.

Section 8.4.1: Universal Limitations

The first of the universal limitations comes from the demographics of Twitter / X itself. While I endeavored to collect sentiment and topics on a global scale for the geopolitical risk analysis, unfortunately the makeup of Twitter / X prevents that. As detailed by Shepherd [391], as of 2023, more than 60% of Twitter / X are below the age of 34, and more than 70% of them are male (Shepherd, 2023). Thus, for research like mine that focuses exclusively on social media like Twitter / X, the results skew towards the opinions of young men with internet connections, which does not reflect the world's population. However, this bias is encountered by all researchers working on social media so it is worth mentioning here. It is impossible to say whether a more diverse sample of age and gender would change the results I obtained or if the results based on this unbalanced demographic data are reflective of how the rest of the world also felt about the geopolitical risk event.

Another limitation with Twitter / X is the potential for the spread of misinformation that could cause errors in the data analysis. Vosoughi, et al. [438] found that false news has the potential to reach 100 times more people than the true news on Twitter / X. This issue could affect my geopolitical risk tweet gathering as people could tweet about a false event which would be gathered by my program and analyzed when they never happened. This could throw off the geopolitical risk sentiment analysis and emerging topic model programs as they would display false results based on false data. Unfortunately, the issue of misinformation and disinformation has become worse on Twitter / X under the ownership of Elon Musk. The EU has recently opened a probe into investigating the spread of disinformation [115], so there is the potential for this problem to alleviate in the future. While this is a problem, I was able to validate the emerging topics discovered through my program and the Case Study 1 showed that the Goldstein Index keywords collection of tweets could find true major geopolitical events. While this is a major issue, my programs still collected enough true data on a whole to produce meaningful results.

Lastly, there is a limitation to the Granger Causality analysis as well. While Granger Causality is powerful and can provide predictive information that one time series has on another, however, the extent of the information is that it is only predictive information, it can't tell you the direction of the change or how great the effect of the change of one time series will affect the other. A vector autoregression model could have detailed information about the change but one would need significantly more model variables to create a model that could predict the changes to that degree like Caldara and Iacoviello [105] do in their research. Also Granger Causality is also not true causality, thus I can not say for certain that the change in the geopolitical risk sentiment analysis indices I created caused the change in the financial assets, all I can say is that there is a relationship between them. However, identifying a method that shows a relationship between geopolitical risk sentiment on social media and financial markets will be useful for future research.

Section 8.4.2: Intransigent Limitations

I describe these following limitations as intransigent as there was no way to overcome or work around them. One example of this kind of limitation come from the NER analysis. Sometimes if there are multiple locations with the same name, the Google Maps API [312] will only choose one location to return the coordinates for, potentially putting the geopolitical risk in the wrong place. This only occurs in a small number of cases, so it doesn't affect overall efficiency of the analysis, but it requires mention here.

Another intransigent limitation comes from the methodological decision to use keywords, in this case key bigrams, to gather the data from the Twitter / X API for the research. While this method gave the most efficient way to identify various geopolitical risk tweets, I encountered a similar problem to the one Garcia and Berton [186] faced in their COVID-19 study: "One limitation of this work is the keywords employed to retrieve content related to COVID-19. It is possible that some relevant tweets were missed if they did not include the keywords." (Garcia and Berton, 2021, Pg. 11). There is the chance that while the Filter Stream API [431] or the Twarc system [21] gathered using the key bigrams did not encompass all tweets referencing the geopolitical event. For example, during the Ukraine War, many tweets would refer to Russian soldiers as "Orcs" [97], which was not one of my key bigrams, thus tweets referring to "Orcs" for the Ukraine War would not have been captured. That said, I believe that the key bigrams that were used were able to collect enough tweets about the different geopolitical risk for effective analysis. Conversely, the key bigrams also might pick up too much information, as detailed in Case Study 2, were the "Prius Missile" in Japan was a frequent slang term that appeared in the tweet data. However, both these issues were also relatively minor in the analysis process.

Section 8.4.3: Limitations Overcome in the Future

This last set of limitations I believe can be overcome in the future and are just limited by the technology or the data that is available at the time of writing. The first is mentioned in Case Study 2 about the number of topics generated through the topic modeling analysis. The greater number of topics generated, the longer the programs takes to process them, thus with more data and more topics, the programs might not be able to process the data within the hour required for the study. However, I believe as computing processing power keeps improving that this will be less of an issue, and more geopolitical topics will be able to be generated and tracked in near real time.

Three further limitations are in the translation, language identification and language lemmatization. The first was also mentioned before in Sentiment Analysis and Case Study 2 chapters on the Google translation of Korean text. As it stands in 2023, the Korean translation of specific words and terms to English is poor [407], however, I believe that Google will continue to work on this issue and in the future, there will be fewer translation errors. The language identification issue comes from the langdetect package [143], specifically with Spanish and Portuguese. There were a few tweets that were mislabeled as the wrong language between Spanish and Portuguese, causing small data errors in the topic modeling process. While this issue with langdetect I could not fix, I believe that the language differentiation process in langdetect will also improve over time. Finally, the lemmatization issue for Korean and Arabic described in

Case Study 2. While I did an exhaustive search for lemmatizers that could have worked for my program, for both languages, the linguistic eccentricities of each language make the development of lemmatizers for Korean and Arabic difficult, but research into NLP for these languages is always improving [195, 281].

Lastly and importantly, there is a limitation of the RNNs developed for my research. While I found the RNN models were effective, there is the potential for them to be even better as more sentiment analysis labeled data is released in the future. While English has a large amount of labeled data, the other languages I investigated in my research did not have as much as English does. With the advent of RNNs and transformer models like BERT [155], this labeled training data in other languages will only proliferate in the future, allowing for more powerful sentiment analysis models.

Section 8.4.4: Sentiment Analysis Non-Textual Limitations

In terms of my sentiment analysis methodology, there is a limitation that warrants discussion. Across the three different sentiment analysis methods I employ in my program: the VADER lexicon [238] , RNNs, and BERT [155] , do not analyze emojis and other non-textual elements for the sentiment analysis. For the RNNs, I constructed them from datasets that did not contain emojis in their text data, so the RNNs would not have an understanding of what an emoji was when one would be encountered. As for VADER and the Arabic BERT model [240], neither incorporate emoji / non-text analysis into their sentiment models. This lack of emoji analysis is a limitation to my work as emojis can express users emotions which is very beneficial to sentiment analysis of the overall tweet. By not including this emoji analysis into my program, there is the potential that I do not capture the full sentiment of the tweet which limits my work. That said, the lack of emoji analysis was not detrimental to my work, as analyses provided by Sayce [378] and Dixon [159] showed that there were 48.72 million tweets in January 2022 had one or more emojis [159] but there were roughly 15.5 billion tweets posted during the same month. This calculates to roughly 0.3% of tweets posted used an emoji. This low percentage of tweets with emojis in them provided me the justification to not include emoji analysis as the sentiment analysis methods such as VADER and BERT were already very accurate without the emojis and for the RNNs, as explained in Section 9.4.3, I already had trouble finding sentiment analysis datasets in non-English, much less finding one large enough that contained emojis to build an RNN with functionality to analyze emojis. However, this is still a limitation to my research and could be improve on in the future.

Section 8.5: Case Study Specific Future Research

For future research avenues, I have divided them into five different sections: three based on the Case Studies I have analyzed, one on the NER analysis extension. A final section for general further research will be included in the conclusion chapter.

Section 8.5.1: Historical Sentiment Analysis Future Research

For the Historical Sentiment Analysis Case Study surrounding the Ukraine War, there were a few potential directions for future research. First, adding Russian and Ukrainian to the languages I captured and analyzed, while not as popular on Twitter / X [1], these languages could potentially change the results overall. These additions maybe able to capture changes in financial markets and assets more specific to the Eastern European and Central Asian markets that were greatly affected by the Ukraine War. Another analysis examining the tweets captured by the “Goldstein Index” from May 2022 to the Present Day could prove interesting. Investigating how sentiment changed since the initial outset of the war and to see if the geopolitical risk sentiment still provides predictive information on the assets and markets that I found in the original case study. Additionally, I may see more Arabic tweets in this period as Iran gets more involved in the Ukraine War. Also, I could leave the Ukraine War entirely and focus on other big geopolitical risk events that happened previously in the Twitter / X Age, such as COVID-19 or the recent Hamas – Israel War to see if the "Goldstein Index" bigrams tweets produce the similar results. Lastly, a study into sunflower seed futures could yield interesting results as Ukraine was the largest producer of sunflower seeds before the 2022 War, thus sentiment around the Ukraine War might have predictive information about sunflower prices on a larger time scale [61].

Section 8.5.2: Real Time Sentiment Analysis Future Research

For the real time sentiment analysis into geopolitical risk, there were several potential changes to the methodology that might improve the correlation and Granger causality results for future studies. First, I could alter the capture period in two ways. Following from the first case study, I found the sentiment of geopolitical risk tweets had predictive information for several foreign exchange markets at the hourly level. Instead of a real time analysis, I could aggregate the sentiment at the hourly level and see if the same results occur. Additionally, I could extend the three-hour window to six-hours or more, perhaps even up to 24 hours. This way, I obtain a larger trend over time which could allow for better correlations to develop or allow more time for the effect of the geopolitical event to propagate through the market, thus showing that the change in the sentiment at the real time level could pass the Granger causality test. Lastly, I could try with more languages or a different combination of languages. I found Spanish on its own at slightly higher correlations, but not high enough to be called significant and, unfortunately, it also didn't produce consistent results over time. However, that does not rule out that with more data coming in from more languages, it's possible to get a different result. For example, Figure 47 (The Map of countries not included in my programs) of Appendix L shows a map of the countries that I did not include a spoken language for in that country. This map shows that there is a lot of potential growth for the geopolitical risk index.

Finally, while I chose to use batch processing for my thesis in general, I believe that true real time processing through Apache Spark [464] or other distributed computing methods similar to the frances system [173, 461]. These methods could reduce the average processing time for my sentiment analysis program and potentially lead to new insights and identify changing geopolitical risk events quicker.

Section 8.5.3: Emerging Geopolitical Topics Analysis Future Research

There are a few areas of interest that have room for expansion in this field. The first is increasing the number of languages I analyze to include German, Russian, Chinese, or Amharic. These languages can all be filtered by the Filter Stream Twitter / X API, which could provide insights into different regions of the world that I didn't investigate. Additionally, I identified other geopolitical topics that I chose not to include in this study, such as "Trade Agreements" from Klement [265] or "Monetary Policy" from Baker, et al. [67] that could provide insight into other geopolitical events that I did not find. Lastly, I could increase the number of key bigrams for each topic, which would allow for more tweets to be gathered for each topic which could allow for analysis below the hourly level for tracking topics over time to see how they emerge.

Additionally, I could try to improve the automatic labeling of the topics generated by the topic models. Right now, I use the top five most relevant words from each topic to develop a topic labelling. However, there are a two methods developed recently that could improve the topic labelling. First, the OntoLDA created by Allahyari and Kochut [46] includes a topic concept hidden layer into the base LDA model allow for more precise topic labelling (Allahyari and Kochut, 2015, Pgs. 1 – 2). The other, created by Bhatia, et al., [84] incorporates Wikipedia to cross referencing the topic words with Wikipedia articles to create topic labels (Bhatia, et al., 2016, Pg. 1).

Section 8.5.4: NER Analysis Future Research

As for the NER Analysis research, one potential project I have already tested was if changes in mentions in countries in the geopolitical risk tweets affect the change price of commodities if that country is a top exporter of the commodity. This is similar to the work done by Abdollahi [32] on changes with Oil price and Twitter / X news. While the historical data produced some good results for Russia and Oil Price, I could not replicate the results in the real time data at the daily level. Unfortunately, by the time of this analysis, Twitter / X CEO Elon Musk's changes to Twitter / X prevented the gathering of more data. Thus, I could not obtain more data to determine definitive results from this test. In addition, studies such as Abraham [34] who focused on Google Trends and Cryptocurrencies, and Janetzko [246] who also used Twitter / X tweet counts to model the Euro – Dollar exchange rate, provided insight into future NER based analysis as these are both assets I had analyzed before. Another area would be an analysis into Lithium and Cobalt which are metals used in renewable energies manufacturing and primarily mined in the Democratic Republic of Congo ("DRC") which has been a geopolitical flashpoint for decades. One potential analysis could involve looking into changes in mentions of the DRC with NER and comparing the trend to the change in the stock prices of renewable energy companies or electric vehicle companies like Tesla.

Chapter 9: Conclusion

Section 9.1: General Future Research

With the overall research, there were a few places that I identified would have rich research potential. If I had infinite time, I would have included Russian, German, Indonesian, and Turkish to the analysis languages as they have large user bases [175]. The map of Figure 47 in Appendix L shows that there is a lot of potential growth for the geopolitical risk index. Tangentially related to this is the lack of African Languages that make up the Filter Stream API. I obtained good coverage in Africa, shown by Figure 46 in Appendix L. According to Wikipedia [135] around 250 million people in Africa speak either English, French or Portuguese, which is about ¼ of the continent. And approximately 17% speak Arabic out of the 1 Billion total African population which is around 200 million and another roughly 50 million people speak Amharic. Altogether, the Filter Stream API language options covers only roughly half of the African population. This means in the fastest growing continent, I will only receive half the potential sentiment. In the future, Twitter / X could improve the languages incorporated in tweets into their APIs like Swahili (East Africa), Fulani (Senegal), or Yoruba (Nigeria, Benin, Togo) which millions speak and are currently overlooked.

One research area with room for growth could be an analysis of sentiment analysis of longer texts related to geopolitics. Perhaps the RNNs and VADER would perform worse when they are used on non-optimized text lengths for them, while the transformer-based models, like BERT would perform better given their ability to still function well on longer texts. There also exists room for improvement for the RNNs we constructed and the BERT models we implemented with increased data for creating and fine tuning these models, respectively. Also, an analysis that combined all the individual datasets together and applying the multilingual BERT model or training an RNN on multiple languages might have better results than just the individual datasets and models. Additionally, a study could be done with the more modern, more computationally heavy sentiment analysis and machine translation methods such as ChatGPT, XLNet, or T5. As Rodríguez-Ibánez, et al [366]. showed, these more recent methods outperformed our models in sentiment analysis accuracy. However, their implementation can be difficult [366] and GPT has been known to perform worse in non-English context (Dave, 2023) [147] . Thus, it would be interesting to see how these newer systems perform comparatively.

Another potential area is taking the geopolitical risk more locally, similar to Engelberg and Parsons [157] study into local media, who found that local media coverage strongly predicted local trading (Engelberg and Parsons, 2009, Pg. 1). A more specified study of geopolitical risks focusing on regions such as Latin America, or East Asia, and looking into financial markets and assets traded in these regions could yield more specific regional results for the combined geopolitical risk analysis.

Also applying the concept of network centrality to the tweets gathered could lead to valuable insights into the most influential Twitter / X accounts who discuss geopolitics. Network centrality aims to find the key and central nodes of a network [145] this process can be applied to social media networks like Twitter / X as well. As a Twitter / X user has the ability to retweet a post, it is possible to build out a network that links back to specific accounts. For example, using a network centrality process called k-shell decomposition analysis [263] , it would be possible to work backwards through the collected tweets and retweets finding their connections and allowing the discovery of the "super-spreader node" (Das, et al., 2018, Pg. 7) [145] of the geopolitical tweets. While this would be possible as the usernames and texts are collected, this network centrality analysis falls outside the realm of my research and would require a new ethics review to implement.

Outside of the geopolitical risk arena, my research can be applied to various other areas. The way my programs are designed it would be possible to monitor and evaluate other markets fairly easily. By replacing the keywords used to collect the data, one could collect different historical and real time tweets on anything from the spread of diseases to a comparison between different cars or the popularity of TV shows over time. The emerging topic modeling program could identify when a disease or disease symptoms start to be tweeted about. With the multilingual aspect of the programs, it would be also to be possible to identify the specific global regions were the disease was first spreading. Since my sentiment analysis program were not trained specifically on geopolitical data, it can be easily applied to other domains, for example, it would be possible to analyze the response to different weekly TV show and see how people on Twitter / X are responding each episode or even during the episode itself with the real time data analysis. With the ease of replacing the keywords with my program, the range of possible inquiries is quite vast.

Lastly, the advent of AI, especially transcription AI, potentially could incorporate audio and visual data into the data analysis process for geopolitical risk. Especially TikTok, that as of writing boasts 1.677 billion users [369], which provides mainly audio / visual discourses could be a potentially fruitful research space.

Section 9.2: Summary of Research

My thesis has met the aims set out during the Introduction and are summarized here. My first aim of the evaluation of the current attempts to define and monitor geopolitical risk were described through the Literature Review. Defining geopolitical risk and how to monitor it came from a combination of computer science techniques, such as Sentiment Analysis, Topic Modeling, and Named Entity Recognition, and the various theories and tests designed by the international relations and political science fields. The interdisciplinary nature of my research was the only way to obtain an accurate picture about how to monitor geopolitical risk. The second aim was the evaluation of data sources and NLP advances for an improved geopolitical risk monitoring tool. My thesis went through many iterations before I landed on using social media, specifically Twitter / X exclusively as the main data source. Initially, I had created a list of at least one news source from every country in the world and planned to create a program to scrape the articles from these daily news sources to build the measure of geopolitical risk at a global level, similar to Caldara and Iacoviello [105]. However, as the need to investigate at smaller time intervals grew, I needed a different approach. With the research of Sankaranarayanan [377], Marcus, et al [296], and Yu, W, et al [459], among others proving that Twitter / X is a viable alternative for research into real – time events. Also, as Rill et al. [362] showed, Twitter / X reacted sooner to events than traditional media or Google Trends. Additionally, with the Twitter / X Filter API, I could get data more specifically on geopolitical events and in multiple languages which allowed me to evaluate the advancements in NLP outside of English as well. I found that research into non-English NLP was robust, but for my research, I had to develop my own algorithms to employ sentiment analysis, topic modeling, and NER to get the best results I could for creating an improved geopolitical risk tool, one that could measure changes in risk at multiple time intervals, including real – time.

This brings me to aims three and four, which were the development of topic models for the monitoring of geopolitical risk, and methods to monitor for emergent geopolitical topics, respectively. Similar to the evaluation of the data sources, the development of topic models for the automatic evaluation of geopolitical risk took various directions before the method I settled on was chosen. While Latent Dirichlet Allocation ("LDA") was and still is the most popular form of topic modeling, I tested various other methods such Correlated Topic Model ("CTM") and Structural Topic Model ("STM") (Hill, 2020) [224]. However, for the overall program compatibility, as everything else was coded in Python [368], and LDA was the topic modeling method with the most interoperability with Python, it made the most sense to use it to develop the topic models to monitor geopolitical risk. This selection of LDA fed into aim four to develop emerging geopolitics monitoring in near real time. Using Besaw's [82] methodology as a basis, who used LDA to track the development of topics over time, I was able to develop a method to show how geopolitical topics emerge, grow, and disappear as well. Thus I was able to fulfilled the aims three and four with these methods.

Finally, the fifth aim of finding methods for associating topic-based geopolitical risk with selected metrics for best predictive performance, was achieved through the coalescence of all the research that I had done before. The research into geopolitics from the political science side, such as the "Goldstein Index" [201] allowed me to develop the geopolitical topic areas to investigate. Investigating the best ways to employ sentiment analysis with multilingual data gave the foundation for creating a metric to measure the changes in a geopolitical risk. This multilingual data expanded the purview of my analysis from just geopolitical risks that affect the English-speaking world to also include risks that effect different regions of the world that would never have been noticed if only English was used. Using this change in sentiment across multiple languages provided valuable information into changes in geopolitical risk which provided insight into the predictive performance of how geopolitics affects financial markets shown by my analysis of the Ukraine War in Case Study 1. Additionally, with the multilingual topic modeling research, I was able to get further context on the specific geopolitical risks that were at play in the various regions around the world, providing an awareness into factors that might affect the changes in geopolitical risk in the future. Lastly, with the NER, geocoding, and mapping research, I was able to provide visualizations to situate the geopolitical risks in the world and show where the hotspots are developing, which could affect the people and markets in these areas.

Unlike other commercial enterprises such as BlackRock's Geopolitical Risk Dashboard [9], Dow Jones Factiva [6], and GDelt [7], the methods I have developed through this research are easily accessible to anyone who would want more knowledge of geopolitics. With the current changes to the Twitter / X APIs under the new CEO, Elon Musk, this research process might become more difficult. However, with the rise of Meta's Threads app, and other microblogs such as the Mastodon app, it is possible to transfer this framework over to one of these other apps to continue this research. As our world becomes more perilous as tension grow between US and China, conflicts in the Middle East, coups in West Africa, and as Russia continues to invade Ukraine and threaten the wider East Europe, I believe the ability to receive timely information and create an automatic evaluation of geopolitical risk, is invaluable.

References:

1. "2018 Research on 100 Million Tweets: What It Means for Your Social Media Strategy for Twitter." Vicinitas https://www.vicinitas.io/blog/twitter-social-media-strategy-2018-research-100-million-tweets#language. Accessed November 5th, 2021.

2. "4 Economic Factors That Can Impact Your Currency Value." *Personal Finance*. Mint https://mint.intuit.com/blog/personal-finance/value-of-your-money/. Accessed April 21st, 2022

3. "6 Industries Most Vulnerable to Cyber Attacks." *Information Technology*. Western Governors University https://www.wgu.edu/blog/6-industries-most-vulnerable-cyber-attacks2108.html#close. Accessed March 10th, 2022.

4. "The Benefits of International Trade." *Trade Agreement*. U.S. Chamber of Commerce https://www.uschamber.com/international/trade-agreements/the-benefits-of-international-trade. Accessed March 10th, 2022.

5. "Country Main Exports." Nation Master, Accessed March 8th, 2022. https://www.nationmaster.com/country-info/stats/Economy/Exports/Main-exports.

6. "Dow Jones Factiva." Dow Jones https://www.dowjones.com/professional/resources/factiva-content/news-you-can-use. Accessed May 16th, 2022.

7. "The Gdelt Project." The GDELT Project https://www.gdeltproject.org/. Accessed May 16th, 2022.

8. "Geopolitical Risk – What Is It and How Can Companies Respond to It?" Kroll https://www.kroll.com/en/insights/publications/compliance-risk/companies-respond-geopolitical-risk. Accessed February 9th, 2021.

9. "Geopolitical Risk Dashboard." BlackRock https://www.blackrock.com/corporate/insights/blackrock-investment-institute/interactive-charts/geopolitical-risk-dashboard. Accessed May 16th, 2022.

10. "Google Trends." Google https://trends.google.com/trends/?geo=US. Accessed May 16th, 2022

11. "How to Get Current Date and Time in Python?" Programiz https://www.programiz.com/python-programming/datetime/current-datetime. Accessed May 6th, 2021.

12. "How to Get the Latest File in a Folder Using Python." Data to Fish https://datatofish.com/latest-file-python/. Accessed March 15th, 2022.

13. "How Much Does It Cost to Implement Real-Time Updates in Your App?" *Cost of Mobile Application Development*. NS804 https://www.ns804.com/blog/how-much-does-it-cost-to-implement-real-time-updates-in-your-app/. Accessed December 10th, 2020.

14. "Labour-Intensive Industries." *Economic and Social Commission for Western Asia*. United Nations https://archive.unescwa.org/labour-intensive-industries. Accessed March 10th, 2022.

15. "Processing Multiple Files and Writing Files." *Python Scripting for Computational Molecular Science*. Molssi https://education.molssi.org/python_scripting_cms/03-multiple_files/index.html. Accessed November 12th, 2022.

16. "Python | Merging Two Dictionaries." Geeks for Geeks https://www.geeksforgeeks.org/python-merging-two-dictionaries/. Accessed May 6th, 2022

17. "Python Script That Is Executed Every 5 Minutes." Geeks for Geeks https://www.geeksforgeeks.org/python-script-that-is-executed-every-5-minutes/. Accessed March 15th, 2022

18. "South Korea: 2021 Primary Energy Data in Quadrillion Btu." *International*, U.S Energy Information Administion, April 13th 2023. https://www.eia.gov/international/overview/country/KOR.

19. "The Top 10 Exporters of Bananas in the World." Opportimes https://www.opportimes.com/the-top-10-exporters-of-bananas-in-the-world/. Accessed March 8th, 2022.

20. "Total of 96.6% of Japan's March Imports of Oil Came from Arab Countries, Led by Saudi Crude." *Arab News*, Arab News, April 29th, 2023. https://www.arabnews.com/node/2294816/business-economy.

21. "Twarc." https://twarc-project.readthedocs.io/en/latest/. Accessed June 12th, 2023.

22. "The Use of Arabic Diacritics." International House Cairo https://ihcairoeg.com/arabic/the-use-of-arabic-diacritics/. Accessed June 20th, 2023.

23. "Vanilla | Imports and Exports | 2020." TradeEconomy https://trendeconomy.com/data/commodity_h2/0905. Accessed March 8th, 2022.

24. "What Is Forex Trading?" IG https://www.ig.com/us/forex/what-is-forex-and-how-does-it-work. Accessed April 20th, 2021.

25. "What's the Most Spoken Language in Africa? (It May Surprise You)." jumpspeak https://www.jumpspeak.com/blog/most-spoken-language-in-africa. Accessed December 16th, 2021

26. "Whole Clove." *Intelligence*. Tridge https://www.tridge.com/intelligences/clove/export. Accessed March 8th, 2022 .

27. "The World Factbook." Central Intelligence Agency, March 8th, 2022. https://www.cia.gov/the-world-factbook/countries/.

28. "World Map: Simple." Map Chart https://www.mapchart.net/world.html. Accessed May 17th, 2022.

29. Abadi, Martín; Ashish Agarwal; Paul Barham; Eugene Brevdo; et al. "Tensorflow: Large-Scale Machine Learning on Heterogeneous Systems." 2015. doi:10.5281/zenodo.4724125.

30. abarnet, f.ashouri;. "Extract Day of Year and Julian Day from a String Date." Stack Overflow https://stackoverflow.com/questions/13943062/extract-day-of-year-and-julian-day-from-a-string-date. Accessed March 19th, 2021.

31. Abdelhaq, Hamed et al. "Eventweet." *Proceedings of the VLDB Endowment*, vol. 6, no. 12, 2013, pp. 1326-29, doi:10.14778/2536274.2536307.

32. Abdollahi, Hooman. "Oil Price Volatility and New Evidence from News and Twitter." *Energy Economics*, vol. 122, 2023, doi:https://doi.org/10.1016/j.eneco.2023.106711.

33. Abouzahra, Mohamed; Joseph Tan. "Twitter Vs. Zika—the Role of Social Media in Epidemic Outbreaks Surveillance." *Health Policy and Technology*, vol. 10, no. 1, 2021, pp. 174-81, doi:https://doi.org/10.1016/j.hlpt.2020.10.014.

34. Abraham, Jethin; Higdon, Daniel; Nelson, John; and Ibarra, Juan. "Cryptocurrency Price Prediction Using Tweet Volumes and Sentiment Analysis." *SMU Data Science Review*, vol. 1, no. 1, 2018, https://scholar.smu.edu/datasciencereview/vol1/iss3/1.

35. Acerbi, Alberto et al. "Sentiment Analysis of the Twitter Response to Netflix's <I>Our Planet</I> Documentary." *Conservation Biology*, 2023, doi:10.1111/cobi.14060.

36. Adams, Travis; Andrea Ajello; Diego Silva; Francisco Vazquez-Grande. "More Than Words: Twitter Chatter and Financial Market Sentiment." *Finance and Economics Discussion Series 2023-034*, vol. 2023-034, 2023, doi:https://doi.org/10.17016/FEDS.2023.034.

37. Adewusi, David. "Everything You Need to Know About Google Translate." Scientifc Editing https://www.scientific-editing.info/blog/everything-you-need-to-know-about-google-translate/. Accessed November 21st, 2022.

38. Agarwal, Apoorv; Boyi Xie; Ilia Vovsha; Owen Rambow; Rebecca Passonneau. "Sentiment Analysis of Twitter Data." *Proceedings of the Workshop on Language in Social Media (LSM 2011)*, Association for Computational Linguistics, 2011, pp. 30--38. https://aclanthology.org/W11-0705.

39. Agarwal, Puneet; Rajgopal Vaithiyanathan; Saurabh Sharma; Gautam Shroff. "Catching the Long-Tail: Extracting Local News Events from Twitter." *Proceedings of the International AAAI Conference on Web and Social Media*, vol. 6, 2021, pp. 379-82. https://ojs.aaai.org/index.php/ICWSM/article/view/14317.

40. Ahmed, Amr; Eric P. Xing. "Timeline: A Dynamic Hierarchical Dirichlet Process Model for Recovering Birth/Death and Evolution of Topics in Text Stream." *Proceedings of the Twenty-Sixth Conference on Uncertainty in Artificial Intelligence (UAI2010)*, 2010. doi:10.48550/arxiv.1203.3463.

41. Ahmouda, Ahmed et al. "Analyzing the Effect of Earthquakes on Openstreetmap Contribution Patterns and Tweeting Activities." *Geo-spatial Information Science*, vol. 21, no. 3, 2018, pp. 195-212, doi:10.1080/10095020.2018.1498666.

42. Aisen, Ari; Francisco José Veiga. "Does Political Instability Lead to Higher Inflation? A Panel Data Analysis." *Monetary and Financial Systems Department*, edited by David S. Hoelscher, International Monetary Fund, 2005. https://www.imf.org/external/pubs/ft/wp/2005/wp0549.pdf.

43. Alex, Beatrice. "Geoparsing English-Language Text with the Edinburgh Geoparser." Programming Historian, vol. 6, 2017, https://doi.org/10.46430/phen0067.

44. Alex, Beatrice; Byrne, Kate; Grover, Claire; Tobin, Richard. "Edinburgh Geoparser " vol. 2024, LANGUAGE TECHNOLOGY GROUP, 2024. https://www.ltg.ed.ac.uk/software/geoparser/

45. Aljazeera. "Chinese Military Says Us Ship 'Warned' Away in South China Sea." Aljazeera, Aljazeera, March 23, 2023. https://www.aljazeera.com/news/2023/3/23/china-warns-away-a-us-warship-in-south-china-sea.

46. Allahyari, Mehdi; Krys Kochut. "Automatic Topic Labeling Using Ontology-Based Topic Models." *14th {IEEE} International Conference on Machine Learning and Applications, ICMLA 2015*, December 9-11, 2015, pp. 259–64. doi:10.1109/ICMLA.2015.88.

47. Allan, James; Jaime Carbonell; George Doddington; Jonathan Yamron; Yiming Yang "Topic Detection and Tracking Pilot Study Final Report." *Broadcast News Understanding and Transcription Workshop*, 1998, pp. 194-218. http://ciir.cs.umass.edu/pubfiles/ir-137.pdf.

48. Alomari, Khaled Arabic-twitter-corpus-AJGT, *Github.com*, 2016, https://github.com/komari6/Arabic-twitter-corpus-AJGT

49. Alostad, Hana; Hasan Davulcu. "Directional Prediction of Stock Prices Using Breaking News on Twitter." *Web Intelligence and Agent Systems: An International Journal*, vol. 5, 2016, https://www.public.asu.edu/%7Ehdavulcu/WIJ17.pdf.

50. Alsaedi, Nasser; Pete Burnap; Omer F. Rana. "A Combined Classification-Clustering Framework for Identifying Disruptive Events." *Proceedings of the 2014 ASE SocialCom Conference*, 2014, pp. 1– 10. doi:10.13140/2.1.3255.4242.

51. Alsaedi, Nasser; Pete Burnap; Omer F. Rana. "Identifying Disruptive Events from Social Media to Enhance Situational Awareness." *Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2015, ASONAM'15*, edited by Fabrizio Silvestri Jian Pei, Jie Tang, Association for Computing Machinery, New York, NY, United States, August 25 - 28, 2015, pp. 934–41. doi:https://doi.org/10.1145/2808797.2808879.

52. Alsumait, Loulwah; Daniel Barbara; Carlotta Domeniconi. "On-Line Lda: Adaptive Topic Models for Mining Text Streams with Applications to Topic Detection and Tracking." *Proceedings of the 8th IEEE International Conference on Data Mining (ICDM 2008)*, December 15-19, 2008. doi:10.1109/ICDM.2008.140.

53. Altig, David; Scott R. Baker; Jose Maria Barrero; Nicholas Bloom; Philip Bunn; Scarlet Chen; Steven J. Davis; Julia Leather; Brent H. Meyer; Emil Mihaylov; Paul Mizen; Nicholas B. Parker; Thomas Renault; Pawel Smietanka; Greg Thwaites. "Economic Uncertainty before and During the Covid-19 Pandemic." *Working Paper Series*, National Bureau of Economic Research, 2020. doi:10.3386/w27418.

54. Amara, Amina et al. "Multilingual Topic Modeling for Tracking Covid-19 Trends Based on Facebook Data Analysis." *Applied Intelligence*, vol. 51, no. 5, 2021, pp. 3052-73, doi:10.1007/s10489-020-02033-3.

55. Amen, Saeed. "Political Market Making: Trading Financial Markets Using Thorfinn Political Indices." *Data, Foreign Exchange, General*, vol. 2023, Cuemacro, 2020. https://www.cuemacro.com/2020/06/26/political-market-making/.

56. Antony J. Blinken, Secretary of State. "Sanctioning Operatives Involved in Iranian External Plots." U.S Department of State, June 1st, 2023. https://www.state.gov/sanctioning-operatives-involved-in-iranian-external-plots/.

57. Araújo, Adelson; WGierke. "Error While Loading Wiki40b Embeddings from Tensorflow Hub." Stack Overflow https://stackoverflow.com/questions/66852256/error-while-loading-wiki40b-embeddings-from-tensorflow-hub. Accessed May 6th, 2022.

58. Aroussi, Ran. "Reliably Download Historical Market Data from with Python." Ran Aroussi https://aroussi.com/post/python-yahoo-finance. Accessed May 26th, 2023.

59. Arun, K.; A. Srinagesh. "Multilingual Twitter Sentiment Analysis Using Machine Learning." *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 10, no. 6, 2020, p. 5992, doi:10.11591/ijece.v10i6.pp5992-6000.

60. Assis, Laura; Hector Huitzil; Jason I. McMann; Ross D. Schaap. "Political Risk and Emerging Market Equities: Applications in an Index Framework." S&P Global, April 2021. https://www.geoquant.com/papers6/3/2022.

61. Association, National Sunflower. "World Supply & Disappearance." National Sunflower Association https://www.sunflowernsa.com/stats/world-supply/. Accessed May 26th, 2023.

62. Atefeh, Farzindar; Wael Khreich. "A Survey of Techniques for Event Detection in Twitter." *Computational Intelligence*, vol. 31, no. 1, 2015, pp. 132–64, doi:https://doi.org/10.1111/coin.12017.

63. Attia, Mohammed; Younes Samih, Ali Elkahky, Laura Kallmeyer. "Multilingual Multi-Class Sentiment Classification Using Convolutional Neural Networks." *ACL Anthology*, vol. Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018), 2018, https://aclanthology.org/L18-1101.

64. Augustop. "Portuguese Tweets for Sentiment Analysis." Kaggle, 2019. https://www.kaggle.com/datasets/augustop/portuguese-tweets-for-sentiment-analysis?select=TweetsWithTheme.csv. Accessed: November 29th, 2021

65. Azar, Pablo D. and Andrew W. Lo. "The Wisdom of Twitter Crowds: <I>Predicting Stock Market Reactions to Fomc Meetings Via Twitter Feeds</I>." *The Journal of Portfolio Management*, vol. 42, no. 5, 2016, pp. 123-34, doi:10.3905/jpm.2016.42.5.123.

66. Bahdanau, Dzmitry; Kyunghyun Cho; Yoshua Bengio. "Neural Machine Translation by Jointly Learning to Align and Translate." *ICLR 2015*, 2014. doi:10.48550/arxiv.1409.0473.

67. Baker, Scott R. et al. "Measuring Economic Policy Uncertainty*." *The Quarterly Journal of Economics*, vol. 131, no. 4, 2016, pp. 1593-636, doi:10.1093/qje/qjw024.

68. Baker, Scott R. ; Nicholas Bloom; Steven J. Davisc; Thomas Renaultd. "Twitter-Derived Measures of Economic Uncertainty " *Twitter-based Uncertainty Indices*, Economic Policy Uncertainty, May 13th, 2021, pp. 1-14. general editor, Economic Policy Uncertainty, https://www.policyuncertainty.com/media/Twitter_Uncertainty_5_13_2021.pdf.

69. Balasubramaniam, Thirunavukarasu et al. "Understanding the Spatio-Temporal Topic Dynamics of Covid-19 Using Nonnegative Tensor Factorization: A Case Study." IEEE, 2020. doi:10.1109/ssci47803.2020.9308265.

70. Banton, Caroline. "How Are International Exchange Rates Set?" Investopedia https://www.investopedia.com/ask/answers/forex/how-forex-exchange-rates-set.asp. Accessed April 20th, 2022.

71. Barbieri, Francesco; Luis Espinosa Anke; Jose Camacho-Collados. "Xlm-T: Multilingual Language Models in Twitter for Sentiment Analysis and Beyond." *ArXiv Cornell University*, 2022, doi:10.48550/arxiv.2104.12250.

72. Barkoff, Sophia. "Cia Director: Iran's Nuclear Program Advancing at "Worrisome Pace." CBS News, CBS News, 2023. https://www.cbsnews.com/news/william-burns-cia-director-iran-nuclear-program-face-the-nation-interview/.

73. Baturo, Alexander et al. "Understanding State Preferences with Text as Data: Introducing the Un General Debate Corpus." *Research & Politics*, vol. 4, no. 2, 2017, doi:10.1177/2053168017712821.

74. Baur, Dirk G.; KiHoon Hong, Adrian D. Lee,. "Bitcoin: Medium of Exchange or Speculative Assets?" *Journal of International Financial Markets, Institutions and Money*, vol. 54, 2018, pp. 177-89, doi:https://doi.org/10.1016/j.intfin.2017.12.004.

75. Bautin, Mikhail; Vijayarenu, Lohit; Skiena, Steven. "International Sentiment Analysis for News and Blogs." *International AAAI Conference on Web and Social Media*, vol. 2, September 5th, 2021, pp. 19-26. https://ojs.aaai.org/index.php/ICWSM/article/view/18606.

76. Becker, Hila et al. "Beyond Trending Topics: Real-World Event Identification on Twitter." *Proceedings of the International AAAI Conference on Web and Social Media*, vol. 5, no. 1, 2021, pp. 438-41, doi:10.1609/icwsm.v5i1.14146.

77. Bekaert, Geert; Campbell R. Harvey; Christian T. Lundblad; Stephan Siegel. "Political Risk and International Valuation " *Columbia Business School Research Paper* vol. 15, no. 83, 2015, doi:http://dx.doi.org/10.2139/ssrn.2659257.

78. Bello, Hugo J. "large-bundle-clean, sentiment-analysis-model-neural-network 2020", Github.com, 2020, https://github.com/sentiment-analysis-spanish/sentiment-analysis-model-neural-network/blob/master/data/json_bundle_reviews/reviews-bundle.zip

79. Bello, Hugo J. "Sentiment-Spanish, sentiment-analysis-sentiment". *Github.com*, 2021, https://github.com/sentiment-analysis-spanish/sentiment-analysis-model-neural-network/blob/master/data/json_bundle_reviews/reviews-bundle.zip. Accessed November 9th, 2021.

80. Beri, Aditya. "Sentiment Analysis Using Vader." Towards Data Science https://towardsdatascience.com/sentimental-analysis-using-vader-a3415fef7664. Accessed February 21st, 2021.

81. Berkman, Henk et al. "Time-Varying Rare Disaster Risk and Stock Returns." *Journal of Financial Economics*, vol. 101, no. 2, 2011, pp. 313-32, doi:10.1016/j.jfineco.2011.02.019.

82. Besaw, Clayton. "Topic Modeling as Osint: Exploring Russian Presidential Speech Topics over Time." medium.com https://medium.com/the-die-is-forecast/topic-modeling-as-osint-exploring-russian-presidential-speech-topics-over-time-ad6018286d37. Accessed February 21st, 2021.

83. Beykikhoshk, Adham et al. "Using Twitter to Learn About the Autism Community." *Social Network Analysis and Mining*, vol. 5, no. 1, 2015, doi:10.1007/s13278-015-0261-5.

84. Bhatia, Shraey; Jey Han Lau; Timothy Baldwin. "Automatic Labelling of Topics with Neural Embeddings." *Proceedings of the 26th International Conference on Computational Linguistics (COLING 2016)*, 2016, pp. 953--63. doi:10.48550/arxiv.1612.05340.

85. Biden, President (@POTUS). "Iran Recently Launched One of… " 2024, April 18th.

86. Bird, Steven; Ewan Klein; Edward Loper. *Natural Language Processing with Python.* First ed., O'Reilly, 2019.

87. Bischke, Benjamin et al. "Contextual Enrichment of Remote-Sensed Events with Social Media Streams." *Proceedings of the 24th ACM international conference on Multimedia*, ACM, 2016. doi:10.1145/2964284.2984063.

88. Blard, Théophile. "French Sentiment Analysis with Bert." GitHub repository https://github.com/TheophileBlard/french-sentiment-analysis-with-bert. Accessed November 29th, 2021.

89. Blei, David M. and John D. Lafferty. "Dynamic Topic Models." ACM Press, 2006. doi:10.1145/1143844.1143859.

90. Blei, David M.; Andrew Y. Ng; Michael I. Jordan. "Latent Dirichlet Allocation." *Journal of Machine Learning Research*, vol. 3, 2003, pp. 993 - 1022, https://www.jmlr.org/papers/volume3/blei03a/blei03a.pdf.

91. Bollen, Johan et al. "Twitter Mood Predicts the Stock Market." *Journal of Computational Science*, vol. 2, no. 1, 2011, pp. 1-8, doi:10.1016/j.jocs.2010.12.007.

92. Bollen, Johan; Alberto Pepe; Huina Mao. "Modeling Public Mood and Emotion: Twitter Sentiment and Socio-Economic Phenomena." *International Conference on Weblogs and Social Media*, 2011, pp. 450–53. doi:10.48550/arxiv.0911.1583.

93. Boyd-Graber, Jordan; David Blei. "Multilingual Topic Models for Unaligned Text." *Proceedings of the Twenty-Fifth Conference on Uncertainty in Artificial Intelligence*, AUAI Press, June 18 - 21 2009. doi:10.48550/arxiv.1205.2657.

94. Brady, William J. et al. "Emotion Shapes the Diffusion of Moralized Content in Social Networks." *Proceedings of the National Academy of Sciences*, vol. 114, no. 28, 2017, pp. 7313-18, doi:10.1073/pnas.1618923114.

95. Branka. "Twitter Statistics - 2023." *Truelist* https://truelist.co/blog/twitter-statistics/#:%7E:text=Users%20send%20out%20an%20average,and%20829%2C094%2C400%20tweets%20per%20day. Accessed July 7th, 2023.

96. Brecher, Michael; Jonathan Wilkenfeld. *A Study of Crisis.* Ann Arbor: University of Michigan Press, 2000.

97. Breitig, Luka Karsten. "Ukrainian War Dictionary or Why Russians Are Called Orcs." *The Happy Beavers* https://www.thehappybeavers.com/post/ukrainian-war-dictionary-or-why-russians-are-called-orcs. Accessed August 3rd, 2023.

98. Bremmer, Ian; Cliff Kupchan. "Top Risk 2023." *Eurasia Group*, January 3rd, 2023, pp. 1-30. general editor, Eurasia Group, https://www.eurasiagroup.net/issues/top-risks-2023.

99. Brewer, Brendon J. *Stats 331: Introduction to Bayesian Statistics.* Department of Statisitcs, University of Auckland, 2016.

100. Brown, Ryan. "Topic: Geopolitical Risk." *Willis Tower Watson* https://www.wtwco.com/en-US/Insights/trending-topics/geopolitical-risk. Accessed February 8th, 2022.

101. Brownlee, Jason. "10 Clustering Algorithms with Python." *Python Machine Learning.* Machine Learning Mastery https://machinelearningmastery.com/clustering-algorithms-with-python/. Accessed April 22nd, 2021.

102. Brownlee, Jason. "How to Choose an Activation Function for Deep Learning." Machine Learning Mastery https://machinelearningmastery.com/choose-an-activation-function-for-deep-learning/. Accessed May 6th, 2022.

103. Brownlee, Jason. "Naive Bayes Classifier from Scratch in Python." *Code Algorithms From Scratch.* Machine Learning Mastery https://machinelearningmastery.com/naive-bayes-classifier-scratch-python/. Accessed November 1st, 2021.

104. C. Pop, D. Bozdog, A. Calugaru, M.A. Georgescu. "Chapter 7 - an Assessment of the Real Development Prospects of the Eu 28 Frontier Equity Markets." *Handbook of Frontier Markets, Academic Press*, 2016, pp. 117-46, doi:https://doi.org/10.1016/B978-0-12-803776-8.00007-0.

105. Caldara, Dario; Matteo Iacoviello. "Measuring Geopolitical Risk." *American Economic Review*, vol. 112, no. 4, 2022, pp. 1194-225, doi:10.1257/aer.20191823.

106. Cambria, Erik. "An Introduction to Concept-Level Sentiment Analysis." *Advances in Soft Computing and Its Applications. MICAI 2013*, edited by F. Castro, Gelbukh, A., González, M, vol. 8266, Springer, Berlin, Heidelberg, November 24 to 30, 2013. doi:https://doi.org/10.1007/978-3-642-45111-9_41.

107. Cambero, Fabian. "Chile Battles Deadliest Wildfires on Record as Heatwave Grips." *Reuters*, Reuters, February 6th, 2023. https://www.reuters.com/world/americas/chile-battles-deadliest-wildfires-record-heatwave-grips-2023-02-06/.

108. Cañete, José; Gabriel Chaperon; Rodrigo Fuentes; Jou-Hui Ho; Hojin Kang; Jorge Pérez. "Spanish Pre-Trained Bert Model and Evaluation Data." *CaneteCFP2020*, 2020. https://github.com/dccuchile/beto.

109. Carmo, Vinicius; Julia Pocciotti; Luísa Heise; Lucas Leme. "Finbertptbr " Hugging Face. https://huggingface.co/turing-usp/FinBertPTBR.

110. Carvalho, Joao P. et al. "Detecting Relevant Tweets in Very Large Tweet Collections: The London Riots Case Study." IEEE, 2017. doi:10.1109/fuzz-ieee.2017.8015635.

111. Chadefaux, Thomas. "Early Warning Signals for War in the News." *Journal of Peace Research*, vol. 51, no. 1, 2014, doi:https://doi.org/10.1177/002234 3313507302.

112. chakki-works; Kubo, Takahiro; Nakayama, Hiroki; Kamura, Junya "chABSA - dataset", *Github.com*, 2018, https://github.com/chakki-works/chABSA-dataset

113. Chandrasekaran, Neeraja et al. "The Utility of Social Media in Providing Information on Zika Virus." *Cureus*, 2017, doi:10.7759/cureus.1792.

114. Chaoua, Ilyas; Diego Reforgiato Recupero; Sergio Consoli; Aki Harm; Rim Helaoui. "Detecting and Tracking Ongoing Topics in Psychotherapeutic Conversations." *Joint Workshop on AI in Health*, edited by Isabelle Bichindaritz, vol. 2142, July 13-14, 2018, pp. 97-108. http://ceur-ws.org/Vol-2142/paper6.pdf.

115. Chee, Foo Yun; Dang, Sheila "Eu Opens Probe into X in Test of New Tech Rules, Pressure on Tiktok, Meta." Reuters, Reuters, October 13th, 2023. https://www.reuters.com/technology/eus-breton-gives-tiktok-ceo-24-hours-detail-disinformation-response-2023-10-12/.

116. Chen, James. "Currency Pair." *Investopedia* https://www.investopedia.com/terms/c/currencypair.asp. Accessed April 20th, 2022

117. Chen, James. "Forex Trading: A Beginner's Guide." *Investopedia* https://www.investopedia.com/articles/forex/11/why-trade-forex.asp. Accessed April 20th, 2022.

118. Chen, James. "Political Risk." *Investopedia* https://www.investopedia.com/terms/p/politicalrisk.asp. Accessed April 20th, 2022.

119. Chen, James. "Safe Haven." *Investopedia* https://www.investopedia.com/terms/s/safe-haven.asp. Accessed April 20th, 2022.

120. Chen, Ray; Marius Lazer. "Analysis of Twitter Feeds for the Prediction of Stock Market Movement." *Chen2011AnalysisOT*, 2011. https://www.semanticscholar.org/paper/Analysis-of-Twitter-Feeds-for-the-Prediction-of-Chen-Lazer/e3d60c77bfae3a6cfb5b15a61a0bce89cf44b13b.

121. Chen, Wen et al. "Neutral Bots Probe Political Bias on Social Media." *Nature Communications*, vol. 12, no. 1, 2021, doi:10.1038/s41467-021-25738-6.

122. Chen, Xieling et al. "A Comparative Quantitative Study of Utilizing Artificial Intelligence on Electronic Health Records in the USA and China During 2008–2017." *BMC Medical Informatics and Decision Making*, vol. 18, no. S5, 2018, doi:10.1186/s12911-018-0692-9.

123. Cheng, Alex; Oles Zhulyn. "A System for Multilingual Sentiment Learning on Large Data Sets." *ACL Anthology*, vol. Proceedings of COLING 2012, 2012, pp. 577–92, https://aclanthology.org/C12-1036.

124. Cheng, Tao; Thomas Wicks. "Event Detection Using Twitter: A Spatio-Temporal Approach." *PLoS One*, vol. 9, no. 6, 2014, p. e97807, doi:10.1371/journal.pone.0097807.

125. Chew, Cynthia; Gunther Eysenbach. "Pandemics in the Age of Twitter: Content Analysis of Tweets During the 2009 H1n1 Outbreak." *PLoS One*, vol. 5, no. 11, 2010, p. e14118, doi:10.1371/journal.pone.0014118.

126. Choi, Hyunyoung; Hal Varian. "Predicting the Present with Google Trends." Google Inc., 2009, pp. 1-23. https://static.googleusercontent.com/media/www.google.com/en//googleblogs/pdfs/google_predicting_the_present.pdf.

127. Choi, Soo-Hyang; Hyonhee Shin. "North Korea Fires Two More Missiles into Its Pacific 'Firing Range'." *Reuters*, Reuters, February 20th, 2023. https://www.reuters.com/world/asia-pacific/north-korea-fires-ballistic-missile-south-korea-military-2023-02-19/.

128. Clark, Kevin; Minh-Thang Luong; Quoc V. Le; Christopher D. Manning. "Electra: Pre-Training Text Encodersas Discriminators Rather Than Generators." *ICLR 2020*, 2020. doi:10.48550/arxiv.2003.10555.

129. CNBC. "U.S. 2 Year Treasury." CNBC https://www.cnbc.com/quotes/US2Y. Accessed May 26th, 2023.

130. Concone, Federico et al. "Twitter Analysis for Real-Time Malware Discovery." IEEE, 2017. doi:10.23919/aeit.2017.8240551.

131. Conneau, Alexis; et al. "Unsupervised Cross-Lingual Representation Learning at Scale." *Arxiv Labs*, 2019, doi:10.48550/arxiv.1911.02116.

132. contributors, Wikipedia. "1973 Oil Crisis." https://en.wikipedia.org/w/index.php?title=1973_oil_crisis&oldid=1161514686. Accessed June 30th, 2023.

133. contributors, Wikipedia. "China–United States Relations." https://en.wikipedia.org/w/index.php?title=China%E2%80%93United_States_relations&oldid=1162607472. Accessed June 30th, 2023.

134. contributors, Wikipedia. "Cyberwarfare." https://en.wikipedia.org/w/index.php?title=Cyberwarfare&oldid=1160321593. Accessed June 30th, 2023.

135. contributors, Wikipedia "Languages of Africa." Wikipedia https://en.wikipedia.org/w/index.php?title=Languages_of_Africa&oldid=1168537711. Accessed August 4th, 2023.

136. contributors, Wikipedia "Nuclear Warfare." https://en.wikipedia.org/w/index.php?title=Nuclear_warfare&oldid=1161701347. Accessed June 30th, 2023.

137. contributors, Wikipedia "Terrorism." https://en.wikipedia.org/w/index.php?title=Terrorism&oldid=1160928625. Accessed June 30th, 2023.

138. Cordeiro, Mário. "Twitter Event Detection: Combining Wavelet Analysis and Topic Inference Summarization." *DSIE'12, the Doctoral Symposium on Informatics Engineering*, 2012. https://www.researchgate.net/publication/281454053_Twitter_event_detection_combining_wavelet_analysis_and_topic_inference_summarization.

139. Cui, Anqi et al. "Emotion Tokens: Bridging the Gap among Multilingual Twitter Sentiment Analysis." Springer Berlin Heidelberg, 2011, pp. 238-49. https://dx.doi.org/10.1007/978-3-642-25631-8_22.

140. Culotta, Aron. "Towards Detecting Influenza Epidemics by Analyzing Twitter Messages." *Workshop on Social Media Analysis In conjunction with the International Conference on Knowledge Discovery & Data Mining (KDD 2010)*, July 25th, 2010. http://snap.stanford.edu/soma2010/papers/soma2010_16.pdf.

141. da Costa-Luis, Casper ; Stephen Karl Larroque; Kyle Altendorf; Hadrien Mary; richardsheridan; Mikhail Korobov; Noam Raphael; Ivan Ivanov; Marcel Bargull; Nishant Rodrigues; Guangshuo Chen; Antony Lee; Charles Newey; James; JC; Martin Zugnoni; Matthew D. Pagel; mjstevens777; Mikhail Dektyarev; Alex Rothberg; Alexander; Daniel Panteleit; Fabian Dill; FichteFoll; Gregor Sturm; HeoHeo; Hugo van Kemenade; Jack McCracken; MapleCCC. "Tqdm: A Fast, Extensible Progress Bar for Python and Cli." v4.64.0 ed., Zenodo, 2022. doi:https://doi.org/10.5281/zenodo.6412640.

142. Dai, Zihang; Zhilin Yang; Yiming Yang; Jaime Carbonell; Quoc V. Le; Ruslan Salakhutdinov. "Transformer-Xl: Attentive Language Models Beyond a Fixed-Length Context." 2019, doi:10.48550/arxiv.1901.02860.

143. Danilak, Michal Mimino. "Langdetect." *github.com* https://github.com/Mimino666/langdetect. 2021.

144. Darkmap "japanese_sentiment/data, *Github.com*, 2016, https://github.com/Darkmap/japanese_sentiment/tree/master/data

145. Das, Kousik et al. "Study on Centrality Measures in Social Networks: A Survey." Social Network Analysis and Mining, vol. 8, no. 1, 2018, p. 13, doi:10.1007/s13278-018-0493-2.

146. Dashtipour, Kia et al. "Multilingual Sentiment Analysis: State of the Art and Independent Comparison of Techniques." *Cognitive Computation*, vol. 8, no. 4, 2016, pp. 757-71, doi:10.1007/s12559-016-9415-7.

147. Dave, Paresh. "Chatgpt Is Cutting Non-English Languages out of the Ai Revolution." Wired, Wired, May 31st, 2023. https://www.wired.com/story/chatgpt-non-english-languages-ai-revolution/.

148. De Guzman, Chad. "North Korea Is Ramping up Its Missile Tests. How Worried Should We Be?" *World - North Korea*. Time https://time.com/6266737/north-korea-ballistic-missile-tests-2023/. Accessed June 22nd, 2023.

149. de Koning, Mattheus. "To Translate or Not to Translate, Best Practices in Non-English Sentiment Analysis." *Towards Data Science* https://towardsdatascience.com/to-translate-or-not-to-translate-best-practices-in-non-english-sentiment-analysis-144a53613913. Accessed December 2nd, 2020.

150. Dean, Brian. "How Many People Use Twitter in 2022? [New Twitter Stats]." *Backlinko* https://backlinko.com/twitter-users#users-by-country. Accessed May 3rd, 2022

151. Department, Statista Research. "Leading Countries Based on Number of Twitter Users as of January 2022." *Social Media & User-Generated Content*. Statista https://www.statista.com/statistics/242606/number-of-active-twitter-users-in-selected-countries/. Accessed May 3rd, 2022.

152. Desthuilliers, rh1990; Bruno. "Creating Path-Names with Variables in Python 2.7." Stack Overflow https://stackoverflow.com/questions/44432230/creating-path-names-with-variables-in-python-2-7. Accessed March 29th, 2022.

153. Developers, Google. "Glossary - Self Attention." *Machine Learning*. https://developers.google.com/machine-learning/glossary#self-attention.

154. Developers, Joblib. "Joblib: Running Python Functions as Pipeline Jobs." *Joblib*, 2021. https://joblib.readthedocs.io/en/latest/.

155. Devlin, Jacob; Ming-Wei Chang; Kenton Lee; Kristina Toutanova. "Bert: Pre-Training of Deep Bidirectional Transformers for Language Understanding." edited by Google AI Language, 2018. doi:10.48550/arxiv.1810.04805.

156. Diakopoulos, Nichola ; Mor Naaman; Funda Kivran-Swaine. "Diamonds in the Rough: Social Media Visual Analytics for Journalistic Inquiry." *2010 IEEE Symposium on Visual Analytics Science and Technology*, 25-26 Oct. 2010. doi:10.1109/VAST.2010.5652922.

157. Dias, Fred. "Brazilian Portuguese Sentiment Analysis Datasets". *kaggle.com*, 2021, https://www.kaggle.com/datasets/fredericods/ptbr-sentiment-analysis-datasets?select=olist.csv

158. Division, Department of Economic and Social Affairs: Statistics. "2020 International Trade Statistics Yearbook." vol. Volume II: Trade by Product, *United Nations*, 2021. https://comtrade.un.org/pb/downloads/2020/VolII2020.pdf.

159. Dixon, Stacy Jo; Statista. "Most Used Emojis on Twitter in January 2022, by Usage Rate." vol. 2024, Statista, 2022. https://www.statista.com/statistics/1367508/most-popular-emojis-twitter-usage-rate/.

160. Djatmiko, Fahim; Ridi Ferdiana; Muhammad Faris;. "A Review of Sentiment Analysis for Non-English Language." *2019 International Conference of Artifical Intelligence and Information Technology (ICAIIT)*, 2019, doi:10.1109/ICAIIT.2019.8834552.

161. Dodds, Klaus. *The New Border Wars : The Conflicts That Will Define Our Future.* First Diversion books ed ed., Diversion Books, 2021.

162. Dombrowski, Quinn. "Preparing Non-English Texts for Computational Analysis." *Modern Languages Open*, vol. 1, 2020, doi:10.3828/mlo.v0i0.294.

163. Duncalfe, Luke; jlewis91; Michal Škop; Nick Dickinson Wilde, " ISO-3166-Countries-with-Regional-Codes", Github.com, 2019, https://github.com/lukes/ISO-3166-Countries-with-Regional-Codes/blob/master/all/all.csv

164. EdChum, Sergei;. "How to Calculate Number of Words in a String in Dataframe? [Duplicate]." Stack Overflow https://stackoverflow.com/questions/37483470/how-to-calculate-number-of-words-in-a-string-in-dataframe. Accessed November 1st, 2021.

165. el Abdouli, Abdeljalil et al. "Sentiment Analysis of Moroccan Tweets Using Naive Bayes Algorithm." *International Journal of Computer Science and Information Security*, vol. 15, no. 12, 2017, https://www.researchgate.net/publication/324840130_Sentiment_Analysis_of_Moroccan_Tweets_using_Naive_Bayes_Algorithm.

166. Engelberg, Joseph; Christopher A. Parsons. "The Causal Impact of Media in Financial Markets." *Workshop in Behavorial Finance*, edited by Andrea Frazzini and Kent D. Daniel, translated by Fuller & Thaler Asset Management and LSV Asset Management Bracebridge Capital, Yale University, November 21st, 2009, pp. 1-44. http://www.econ.yale.edu/%7Eshiller/behfin/2009_11/engelberg-parsons.pdf.

167. Erb, Claude B.; Campbell R. Harvey; Tadas E. Viskanta. "Political Risk, Economic Risk, and Financial Risk." *Financial Analysts Journal*, vol. 52, no. 6, 1996, pp. 29-46, https://www.jstor.org/stable/4479958.

168. Erdmann, Maike et al. "Feature Based Sentiment Analysis of Tweets in Multiple Languages." Springer International Publishing, 2014, pp. 109-24. https://dx.doi.org/10.1007/978-3-319-11746-1_8.

169. Estevez, Juan; Skippy le Grand Gourou. "How to Read Multiple Json Files into Pandas Dataframe?" Stack Overflow https://stackoverflow.com/questions/57067551/how-to-read-multiple-json-files-into-pandas-dataframe. Accessed May 6th, 2022.

170. Faulconbridge, Guy. "Nova Kakhovka Dam Breach: What Do We Know So Far?" *Reuters*, Reuters, June 7th 2023. https://www.reuters.com/world/europe/what-is-kakhovka-dam-ukraine-what-happened-2023-06-07/.

171. fchollet. "English-to-Spanish Translation with a Sequence-to-Sequence Transformer." *Natural Language Processing* edited by Keras, vol. 2022, keras, 2021. https://keras.io/examples/nlp/neural_machine_translation_with_transformer/.

172. Fernando, Jason. "The Correlation Coefficient: What It Is, What It Tells Investors." *Investopedia* https://www.investopedia.com/terms/c/correlationcoefficient.asp. Accessed July 3rd, 2023.

173. Filgueira, Rosa. "Frances: A Deep Learning Nlp and Text Mining Web Tool to Unlock Historical Digital Collections: A Case Study on the Encyclopaedia Britannica,." 2022 IEEE 18th International Conference on e-Science (e-Science), 2022, pp. 246-55. doi:10.1109/eScience55777.2022.00038.

174. Finance, Yahoo. "Stock Market Live, Quotes, Business &Amp; Finance News." *Yahoo* https://finance.yahoo.com/. Accessed July 6th, 2023.

175. Fish, Tom. "These Countries Have the Most People on Twitter." *Tech & Science*. Newsweek https://www.newsweek.com/countries-most-people-twitter-social-media-us-japan-uk-1631479. Accessed November 5th, 2021.

176. Flint, Colin. *Introduction to Geopolitics.* Routledge, Taylor et Francis Group 2022, 2022.

177. Folorunsho, Dhikrullah. "Arabic Sentiment Analysis." *Towards Data Science* https://towardsdatascience.com/arabic-sentiment-analysis-5e21b77fb5ea. Accessed November 15th, 2021.

178. Foong, Ng Wai. "Korean Natural Language Processing in Python." *Towards Data Science* https://towardsdatascience.com/korean-natural-language-processing-in-python-cc3109cbbed8. Accessed November 15th, 2021.

179. Friedman, George. *The Next 100 Years : A Forecast for the 21st Century.* Black Inc., 2010.

180. Fujisaka, Tatsuya et al. "Discovery of User Behavior Patterns from Geo-Tagged Micro-Blogs." ACM, 2010. doi:10.1145/2108616.2108660.

181. Fukuhara, Tomohiro; Takehito Utsuro; Hiroshi Nakagawa. "Cross-Lingual Concern Analysis from Multilingual Weblog Articles." University of Tokyo, 2005. https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=25b75750796c32780c044a791dfafc5105e73d5b.

182. Fukumasu, Kosuke; Koji Eguchi; Eric Xing. "Symmetric Correspondence Topic Models for Multilingual Text Analysis." *Advances in Neural Information Processing Systems 25 (NIPS 2012)*, edited by F. Pereira and C.J. Burges and L. Bottou and K.Q. Weinberger, 2012. http://www.cs.cmu.edu/%7Eepxing/papers/2012/Fukumasu_Eguchi_Xing_NIPS12.pdf.

183. Gabos, Stefan. "Korean World Countries, world_countries" *Github.com*, 2022. https://stefangabos.github.io/world_countries/.

184. Gabrilovich, Evgeniy; Susan Dumais; Eric Horvitz. "Newsjunkie: Providing Personalized Newsfeeds Via Analysis of Information Novelty." *WWW2004*, May 17–22, 2004. http://erichorvitz.com/NewsJunkie.pdf.

185. Galarnyk, Michael. "Access Data from Twitter Api Using R and Python." Towards Data Science https://towardsdatascience.com/access-data-from-twitter-api-using-r-and-or-python-b8ac342d3efe. Accessed December 21st, 2020.

186. Garcia, Klaifer; Lilian Berton. "Topic Detection and Sentiment Analysis in Twitter Content Related to Covid-19 from Brazil and the USA." *Applied Soft Computing*, vol. 101, 2021, doi:https://doi.org/10.1016/j.asoc.2020.107057.

187. gamebusterz, "xac, French-Sentiment-Analysis-Dataset", *Github.com,* 2017, https://github.com/gamebusterz/French-Sentiment-Analysis-Dataset/blob/master/xac

188. gamebusterz, "xae, French-Sentiment-Analysis-Dataset", *Github.com,* 2017, https://github.com/gamebusterz/French-Sentiment-Analysis-Dataset/blob/master/xae

189. gamebusterz, "xaj, French-Sentiment-Analysis-Dataset", *Github.com,* 2017, https://github.com/gamebusterz/French-Sentiment-Analysis-Dataset/blob/master/xaj

190. gamebusterz, "xak, French-Sentiment-Analysis-Dataset", *Github.com,* 2017, https://github.com/gamebusterz/French-Sentiment-Analysis-Dataset/blob/master/xak

191. Gao, Xingyu et al. "A Novel Method for Geographical Social Event Detection in Social Media." ACM Press, 2013. doi:10.1145/2499788.2499819.

192. Garg, Gauri; Negi Babu. "Remove Usernames from Twitter Data Using Python." Stack Overflow https://stackoverflow.com/questions/50830214/remove-usernames-from-twitter-data-using-python. Accessed May 6th, 2022.

193. genediazjr, "stopwords-ko", Github.com, 2016, https://github.com/stopwords-iso/stopwords-ko

194. Géron, Aurélien. *Hands-on Machine Learning with Scikit-Learn and Tensorflow : Concepts, Tools, and Techniques to Build Intelligent Systems.* O'Reilly, 2022.

195. Ghallab, Abdullatif et al. "Arabic Sentiment Analysis: A Systematic Literature Review." *Applied Computational Intelligence and Soft Computing*, vol. 2020, 2020, pp. 1-21, doi:10.1155/2020/7403128.

196. Ghorbel, Hatem; David Jacot. "Further Experiments in Sentiment Analysis of French Movie Reviews." Springer Berlin Heidelberg, 2011, pp. 19-28. https://dx.doi.org/10.1007/978-3-642-18029-3_3.

197. Gilbert, Eric; Karrie Karahalios. "Widespread Worry and the Stock Market." *Proceedings of the International AAAI Conference on Web and Social Media*, vol. 4, no. 1, 2010, pp. 58-65, doi:10.1609/icwsm.v4i1.14023.

198. Gimenes, Manuel; Boris New. "Worldlex: Twitter and Blog Word Frequencies for 66 Languages." *Behavior Research Methods*, vol. 48, no. 3, 2016, pp. 963-72, doi:10.3758/s13428-015-0621-0.

199. Goel, Vikas; Amir Kr. Gupta; Narendra Kumar. "Sentiment Analysis of Multilingual Twitter Data Using Natural Language Processing." *2018 8th International Conference on Communication Systems and Network Technologies (CSNT)*, Nov. 24 - 26, 2018, pp. 208-12. doi:10.1109/CSNT.2018.8820254.

200. GoldHub. "Gold Spot Prices." *GoldHub* https://www.gold.org/goldhub/data/gold-prices. Accessed May 26th, 2023.

201. Goldstein, Joshua S. " A Conflict-Cooperation Scale for Weis Events Data." *The Journal of Conflict Resolution*, vol. 36, no. 2, 1992, pp. 369-85, https://www.jstor.org/stable/174480.

202. Goldstone, Jack A. et al. "A Global Model for Forecasting Political Instability." *American Journal of Political Science*, vol. 54, no. 1, 2010, pp. 190-208, doi:10.1111/j.1540-5907.2009.00426.x.

203. Gonalves, Teresa; Paulo Quaresma. "Multilingual Text Classification through Combination of Monolingual Classifiers." *Legal Ontologies and Artificial Intelligence Techniques*, vol. Proceedings of the 4th Workshop, 2010, pp. 29-38, http://ceur-ws.org/Vol-605/paper3.pdf.

204. González, José Ángel et al. "Twilbert: Pre-Trained Deep Bidirectional Transformers for Spanish Twitter." *Neurocomputing*, vol. 426, 2021, pp. 58-69, doi:https://doi.org/10.1016/j.neucom.2020.09.078.

205. Granger, C. W. J. "Investigating Causal Relations by Econometric Models and Cross-Spectral Methods." *Econometrica*, vol. 37, no. 3, 1969, p. 424, doi:10.2307/1912791.

206. Granger, Clive W.J. "Time Series Analysis, Cointegration, and Applications." Nobel Prize, 2003. https://www.nobelprize.org/uploads/2018/06/granger-lecture.pdf.

207. Griffiths, T. L.; M. Steyvers. "Finding Scientific Topics." *Proceedings of the National Academy of Sciences*, vol. 101, no. Supplement 1, 2004, pp. 5228-35, doi:10.1073/pnas.0307752101.

208. Grimmer, Justin; Brandon M. Stewart. "Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts." *Political Analysis*, vol. 21, no. 3, 2013, pp. 267-97, doi:10.1093/pan/mps028.

209. Guo, Honglei; Huijia Zhu; Zhili Guo; Xiaoxun Zhang; Zhong Su. "Opinionit: A Text Mining System for Cross-Lingual Opinion Analysis." *CIKM '10: Proceedings of the 19th ACM international conference on Information and knowledge management*, Association for Computing Machinery, New York, NY, United States, October 2010, pp. 1199–208. doi:https://doi.org/10.1145/1871437.1871589.

210. Guo, Mandy; Zihang Dai; Denny Vrandečić; Rami Al-Rfou. "Wiki-40B: Multilingual Language Model Dataset", *research.google,* 2020, https://research.google/pubs/pub49029/

211. Gupta, Yashi. "Explained: What Is the Economic Impact of Geopolitical Unrest?" *Economy News*. CNBC TV 18 https://www.cnbctv18.com/economy/explained-what-is-the-economic-impact-of-geopolitical-unrest-9985541.htm. Accessed September 21st, 2021

212. Hafezi, Parisa "Iran Presents Its First Hypersonic Ballistic Missile, State Media Reports." *Reuters*, Reuters, June 6th, 2023. https://www.reuters.com/world/middle-east/iran-unveils-its-first-hypersonic-ballistic-missile-state-media-reports-2023-06-06/.

213. Hammer, Ben. "crowdflower-airline-twitter-sentiment, Twitter US Airline Sentiment Distributor", kaggle.com, 2016, https://www.kaggle.com/datasets/crowdflower/twitter-airline-sentiment

214. Harris, Charles R. et al. "Array Programming with Numpy." *Nature*, vol. 585, no. 7825, 2020, pp. 357-62, doi:10.1038/s41586-020-2649-2.

215. Hass, Ryan. "How Biden Could “Thaw” Us Relations with China." Brookings, May 23rd, 2023. https://www.brookings.edu/articles/how-biden-could-thaw-us-relations-with-china/#:%7E:text=The%20year%202023%20was%20supposed,in%20the%20fall%20of%202023.

216. hatmimoha. "Named Entity Recognition System for Arabic." GitHub, 2019. https://github.com/hatmimoha/arabic-ner.

217. Hauff, Claudia; Geert-Jan Houben. "Placing Images on the World Map." ACM, 2012. doi:10.1145/2348283.2348376.

218. Hayes, Adam. "What Is Price Stickiness? Definition, Triggers, and Example." *ECONOMICS*. Investopedia https://www.investopedia.com/terms/p/price_stickiness.asp#:%7E:text=%22Sticky%22%20is%20a%20general%20economics,that%20is%20resistant%20to%20change. Accessed June 13th, 2023.

219. Hecht, Brent; Lichan Hong; Bongwon Suh; Ed H. Chi. "Tweets from Justin Bieber's Heart: The Dynamics of the Location Field in User Profiles." *CHI '11: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, Association for Computing Machinery, New York, NY, United States, May 7th, 2011, pp. 237–46. doi:https://doi.org/10.1145/1978942.1978976.

220. Hegazi, Mohamed Osman et al. "Preprocessing Arabic Text on Social Media." *Heliyon*, vol. 7, no. 2, 2021, doi:10.1016/j.heliyon.2021.e06191.

221. Hegre, Håvard et al. "Predicting Armed Conflict, 2010-2050¹." *International Studies Quarterly*, vol. 57, no. 2, 2013, pp. 250-70, doi:10.1111/isqu.12007.

222. Herrera, Gabriel Paes; Michel Constantino; Jen-Je Su; Athula Naranpanawa. "Renewable Energy Stocks Forecast Using Twitter Investor Sentiment and Deep Learning." *Energy Economics*, vol. 114, 2022, doi:https://doi.org/10.1016/j.eneco.2022.106285.

223. Hida, Rem; Naoya Takeishi; Takehisa Yairi; Koichi Hori. "Dynamic and Static Topic Model for Analyzing Time-Series Document Collections." *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics*, vol. Volume 2: Short Papers, Association for Computational Linguistics, 2018, pp. 516–20. doi:10.18653/v1/P18-2082.

224. Hill, Chelsey. "Topic Models." *Rpubs* https://rpubs.com/chelseyhill/672546. Accessed August 8th, 2023.

225. Hisano, Ryohei et al. "High Quality Topic Extraction from Business News Explains Abnormal Financial Market Volatility." *PLoS One*, vol. 8, no. 6, 2013, p. e64846, doi:10.1371/journal.pone.0064846.

226. Ho, Shen-Shyang et al. "Mining Future Spatiotemporal Events and Their Sentiment from Online News Articles for Location-Aware Recommendation System." *Proceedings of the First ACM SIGSPATIAL International Workshop on Mobile Geographic Information Systems - MobiGIS '12*, ACM Press, 2012. doi:10.1145/2442810.2442816.

227. Hoang, misguided; Quang. "Pandas Merge and Keep Only Non-Matching Records [Duplicate]." Stack Overflow https://stackoverflow.com/questions/63184392/pandas-merge-and-keep-only-non-matching-records. Accessed May 6th, 2022.

228. Hodge, Victoria J.; Jim Austin. "A Survey of Outlier Detection Methodologies." *Artificial Intelligence Review*, 2004, pp. 85-126, doi:https://doi.org/10.1023/B:AIRE.0000045502.10941.a9.

229. Holmes, Frank "Update: Top 10 Gold Producing Countries." *Markets*. Forbes https://www.forbes.com/sites/greatspeculations/2021/06/23/updated-top-10-gold-producing-countries/?sh=714e6aa02ce2. Accessed March 8th, 2021.

230. Hong, Liangjie; Brian D. Davison. "Empirical Study of Topic Modeling in Twitter." *Proceedings of the First Workshop on Social Media Analytics - SOMA '10*, ACM Press, July 25, 2010. doi:10.1145/1964858.1964870.

231. Hong, Lichan; Gregorio Convertino; Ed H. Chi. "Language Matters in Twitter: A Large Scale Study " *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media*, Association for the Advancement of Artificial Intelligence, July 17 – 21 2011. https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.309.5773&rep=rep1&type=pdf.

232. Honnibal, Matthew; Ines Montani. "Spacy 2: Natural Language Understanding with Bloom Embeddings, Convolutional Neural Networks and Incremental Parsing." Explosion https://spacy.io/usage. Accessed August 25th, 2021.

233. Hoque, Mohammad Enamul and Mohd Azlan Shah Zaidi. "Global and Country-Specific Geopolitical Risk Uncertainty and Stock Return of Fragile Emerging Economies." *Borsa Istanbul Review*, vol. 20, no. 3, 2020, pp. 197-213, doi:10.1016/j.bir.2020.05.001.

234. Hu, Minqing; Bing Liu. "Mining and Summarizing Customer Reviews." *Proceedings of the 2004 ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '04*, ACM Press, 2004. doi:10.1145/1014052.1014073.

235. Huang, Tao; Fei Wu; Jing Yu; Bohui Zhang. "International Political Risk and Government Bond Pricing." *Journal of Banking and Finance, Forthcoming, 26th Australasian Finance and Banking Conference 2013*, 2014, doi:http://dx.doi.org/10.2139/ssrn.2312188.

236. Hunter, J. D. "Matplotlib: A 2d Graphics Environment." *Computing in Science & Engineering*, vol. 9, no. 3, 2007, pp. 90 - 95, doi:10.1109/MCSE.2007.55.

237. Hurtado, Jose L. et al. "Topic Discovery and Future Trend Forecasting for Texts." *Journal of Big Data*, vol. 3, no. 2, 2016, doi:10.1186/s40537-016-0039-2.

238. Hutto, C.J.; Eric Gilbert. "Vader: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text." *International AAAI Conference on Weblogs and Social Media (ICWSM)*, 2014, http://eegilbert.org/papers/icwsm14.vader.hutto.pdf.

239. Iacobelli, Francisco; Larry Birnbaum; Kristian J. Hammond. "Tell Me More, Not Just "More of the Same"." *IUI '10: Proceedings of the 15th international conference on Intelligent user interfaces*, edited by Qiang Yang Charles Rich, Marc Cavazza and Michelle Zhou, Association for Computing Machinery, New York, NY, United States, February 7 - 10, 2010, pp. Pages 81–90. doi:https://doi.org/10.1145/1719970.1719982.

240. Inoue, Go; Bashar Alhafni; Nurpeiis Baimukan; Houda Bouamor; Nizar Habash. "The Interplay of Variant, Size, and Task Type in Arabic Pre-Trained Language. Models." *Proceedings of the Sixth Arabic Natural Language Processing Workshop*, Association for Computational Linguistics, April 2021. https://huggingface.co/CAMeL-Lab/bert-base-arabic-camelbert-mix-sentiment.

241. Investing.com. "Crude Oil WTI Futures Historical Data" *Investing.com* https://www.investing.com/commodities/crude-oil-historical-data Accessed May 26th, 2023

242. Investing.com. "Germany 10-Year Bond Yield." *Investing.com* https://www.investing.com/rates-bonds/germany-10-year-bond-yield-historical-data. Accessed May 26th, 2023.

243. Investing.com. "United States 2-Year Bond Yield." *Investing.com* https://www.investing.com/rates-bonds/u.s.-2-year-bond-yield-historical-data. Accessed May 26th, 2023.

244. Irish, John. "Israel: 'All Possible Means on the Table' to Prevent Iran Getting Nuclear Weapon." Reuters, Reuters, 2023. https://www.reuters.com/world/middle-east/israel-all-possible-means-table-prevent-iran-getting-nuclear-weapon-2023-02-17/.

245. James, Patrick; Kyle Beardsley; Jonathan Wilkenfeld; Michael Brecher "ICB Project: International Crisis Behavior", duke.edu, https://sites.duke.edu/icbdata/project-info/

246. Janetzko, Dietmar. "Using Twitter to Model the Eur/Usd Exchange Rate." *arXiv pre-print server*, 2014, https://arxiv.org/abs/1402.1624

247. Jazeera, Al. "The Taliban Closes Afghan Girls' Schools Hours after Reopening." Al Jazeera, March 23rd, 2022. https://www.aljazeera.com/news/2022/3/23/taliban-orders-girls-schools-shut-hours-after-reopening.

248. Jiao, Peiran et al. "Social Media, News Media and the Stock Market." *Journal of Economic Behavior & Organization*, vol. 176, 2020, pp. 63-90, doi:10.1016/j.jebo.2020.03.002.

249. Johnson, Jeese. "North Korea Fires Short-Range Missile Ahead of U.S.-South Korea Military Exercises." *The Japan Times*, The Japan Times, March 9th, 2023. https://www.japantimes.co.jp/news/2023/03/09/asia-pacific/north-korea-missile-launch-march-9/.

250. Joshi, Prateek. "How Do Transformers Work in Nlp? A Guide to the Latest State-of-the-Art Models." vol. 2022, Analytics Vidhya, 2019. https://www.analyticsvidhya.com/blog/2019/06/understanding-transformers-nlp-state-of-the-art-models/?utm_source=blog&utm_medium=demystifying-bert-groundbreaking-nlp-framework.

251. Kaggle. " Spanish Airlines Tweets Sentiment Analysis", *Kaggle.com*, 2018, https://www.kaggle.com/competitions/spanish-arilines-tweets-sentiment-analysis/overview/description

252. Kaity, Mohammed and Vimala Balakrishnan. "An Automatic Non-English Sentiment Lexicon Builder Using Unannotated Corpus." *The Journal of Supercomputing*, vol. 75, no. 4, 2019, pp. 2243-68, doi:10.1007/s11227-019-02755-3.

253. Kamel, Mahmoud. "Machine Translation and Multilinguality in Text Classification." Medium https://medium.com/@mahmoud.kamel104/machine-translation-and-multilinguality-in-text-classification-6e20ef9dbce8. Accessed December 18th, 2020.

254. Kana, Michel. "Natural Language Understanding with Sequence to Sequence Models." Towards Data Science https://towardsdatascience.com/natural-language-understanding-with-sequence-to-sequence-models-e87d41ad258b. Accessed 2021.

255. Kaneko, Takamu; Keiji Yanai. "Visual Event Mining from Geo-Tweet Photos." *2013 IEEE International Conference on Multimedia and Expo Workshops (ICMEW)*, 2013, pp. 1–6. doi:10.1109/ICMEW.2013.6618407.

256. Kapadia, Shashank. "Evaluate Topic Models: Latent Dirichlet Allocation (Lda)." Towards Data Science https://towardsdatascience.com/evaluate-topic-model-in-python-latent-dirichlet-allocation-lda-7d57484bb5d0. Accessed March 9th, 2021.

257. Karageorgou, Ioanna et al. "A Sentiment Analysis Service Platform for Streamed Multilingual Tweets." IEEE, 2020.

258. Karkali, Margarita; Francois Rousseau; Alexandros Ntoulas; Michalis Vazirgiannis. "Efficient Online Novelty Detection in News Streams." *WISE 2013: Web Information Systems Engineering – WISE 2013*, edited by Yannis Manolopoulos Xuemin Lin, Divesh Srivastava, Guangyan Huang, vol. 8180, Springer, Berlin, Heidelberg, 2013. doi:https://doi.org/10.1007/978-3-642-41230-1_5.

259. Karim, Asma Abdel; Gheith Abandah. "On the Training of Deep Neural Networks for Automatic Arabic-Text Diacritization." *(IJACSA) International Journal of Advanced Computer Science and Applications*, vol. 12, no. 8, 2021, pp. 276-86, https://thesai.org/Downloads/Volume12No8/Paper_32-On_the_Training_of_Deep_Neural_Networks.pdf.

260. Kenton, Will. "Nominal Effective Exchange Rate (Neer) Definition." Investopedia https://www.investopedia.com/terms/n/neer.asp. Accessed April 22nd, 2022.

261. Kim, Eunhee; Hyopil Shin. "Kr-Finbert: Fine-Tuning Kr-Finbert for Sentiment Analysis." *Github.com*, 2022. https://huggingface.co/snunlp/KR-FinBert-SC.

262. Kilicoglu, Halil; Sabine Bergler. "Syntactic Dependency Based Heuristics for Biological Event Extraction." *Proceedings of the BioNLP 2009 Workshop Companion Volume for Shared Task*, Association for Computational Linguistics, June 2009, pp. 119-27. https://aclanthology.org/W09-1418.

263. Kitsak, Maksim et al. "Identification of Influential Spreaders in Complex Networks." Nature Physics, vol. 6, no. 11, 2010, pp. 888-93, doi:10.1038/nphys1746.

264. Kleinnijenhuis, Jan et al. "Financial News and Market Panics in the Age of High-Frequency Sentiment Trading Algorithms." *Journalism*, vol. 14, no. 2, 2013, pp. 271-91, doi:10.1177/1464884912468375.

265. Klement, Joachim *Geo-Economics: The Interplay between Geopolitics, Economics, and Investments.* CFA Institute, 2021. *Research Foundation Books.*

266. KM_83, mohsen.21;. "Sample of Filtered Twitter Api Stream." Stack Overflow https://stackoverflow.com/questions/64160766/sample-of-filtered-twitter-api-stream. Accessed December 10th, 2021.

267. Koch, Korbinian. "A Friendly Introduction to Text Clustering." Towards Data Science https://towardsdatascience.com/a-friendly-introduction-to-text-clustering-fa996bcefd04. Accessed April 22nd, 2021.

268. Kochut, Krys; Allahyari Mehdi; Seyedamin Pouriyeh; Mehdi Assefi; Saeid Safaei; Elizabeth D. Trippe; Juan B. Gutierrez. "A Brief Survey of Text Mining: Classification, Clustering and Extraction Techniques." *KDD 2017*, 2017. http://cobweb.cs.uga.edu/%7Esafaei/A%20Brief%20Survey%20of%20Text%20Mining%20Classification,%20Clustering%20and%20Extraction%20Techniques.pdf.

269. Kolasani, Sai Vikram and Rida Assaf. "Predicting Stock Movement Using Sentiment Analysis of Twitter Feed with Neural Networks." *Journal of Data Analysis and Information Processing*, vol. 08, no. 04, 2020, pp. 309-19, doi:10.4236/jdaip.2020.84018.

270. Komban, Stanley Jose; Raghavendra Prasad Narayan; Giorgio Orsi "Deep Learning Models for Sentiment Analysis." underthehood.meltwater.com https://underthehood.meltwater.com/blog/2019/08/22/deep-learning-models-for-sentiment-analysis/. Accessed January 13th, 2022.

271. Koren, Vitaliy. "Korean Tokenization & Lemmatization." Medium https://korenv20.medium.com/korean-tokenization-lemmatization-a741fc9939cc. Accessed January 30th, 2022..

272. Kravariti, Alexandra. "Machine Vs Human Translation: The Pros, Cons and When to Use Each." *Blog*. translate plus https://www.translateplus.com/blog/machine-vs-human-translation-pros-cons-use/. Accessed June 20th, 2023.

273. Krupenin, Arseniy; jezrael. "Count Unique Values Per Groups with Pandas [Duplicate]." Stack overflow.com https://stackoverflow.com/questions/38309729/count-unique-values-per-groups-with-pandas. Accessed May 6th, 2022.

274. Kulshrestha, Ria. "A Beginner's Guide to Latent Dirichlet Allocation (Lda)." Towards Data Science https://towardsdatascience.com/latent-dirichlet-allocation-lda-9d1cd064ffa2. Accessed May 13th, 2022.

275. Kumar, Shubham; Shashank Kumar; Sunanda Dixit; Piyush. "Classification of Tweets into Various Categories Using Classification Methods." *International Journal of Advance Research, Ideas and Innovation in Technology*, vol. 4, no. 3, 2018, pp. 937-39, https://www.ijariit.com/manuscripts/v4i3/V4I3-1505.pdf.

276. L., Davide. "How Does Granger Causality Relate to Vector Autoregression?" Stats, vol. 2024, Stack Exchange, 2019. https://stats.stackexchange.com/questions/438822/how-does-granger-causality-relate-to-vector-autoregression.

277. Lakezina, Viktoriia. "War Zone Villagers Flee after Massive Ukraine Dam Destroyed." *Reuters*, reuters, June 6th, 2023. https://www.reuters.com/world/europe/russia-says-it-thwarts-another-major-ukrainian-offensive-donetsk-2023-06-05/.

278. Lakshmipathi, N. "Imdb Dataset of 50k Movie Reviews." *Kaggle.com*, 2018. https://www.kaggle.com/datasets/lakshmi25npathi/imdb-dataset-of-50k-movie-reviews?resource=download.

279. Lee, Hagyeong; Jongwoo Song. "Understanding Recurrent Neural Network for Texts Using English-Korean Corpora." *Communications for Statistical Applications and Methods*, vol. 27, no. 3, 2020, pp. 313-26, doi:https://doi.org/10.29220/CSAM.2020.27.3.313.

280. Lee, Hakyeon; Hanbin Seo; Youngjung Geum. "Uncovering the Topic Landscape of Product-Service System Research: From Sustainability to Value Creation." *Sustainability*, vol. 10, no. 4, 2018, doi:https://doi.org/10.3390/su10040911.

281. Lee, Jae Jun; Suhn Beom Kwon; Sung Mahn Ahn. "Sentiment Analysis Using Deep Learning Model Based on Phoneme-Level Korean." *Journal of Information Technology Services*, vol. 17, no. 1, 2018, pp. 79-89, doi:https://doi.org/10.9716/KITS.2018.17.1.079.

282. Lee, Kathy et al. "Forecasting Influenza Levels Using Real-Time Social Media Streams." IEEE, 2017 2017. doi:10.1109/ichi.2017.68.

283. Lehkonen, Heikki; Kari Heimonen; Kuntara Pukthuanthong. "Is Media Tone Just a Tone? Time-Series and Cross-Sectional Evidence from the Currency Market " *SSRN*, 2022, doi:http://dx.doi.org/10.2139/ssrn.4283275.

284. Leschinski, Tamer; Eric. "How Can I Import Other Python Files?" Stack Overflow https://stackoverflow.com/questions/2349991/how-can-i-import-other-python-files. Accessed March 29th, 2022.

285. Lhessani, Sajid. "Python: How to Get Live Market Data (Less Than 0.1-Second Lag)." Medium https://towardsdatascience.com/python-how-to-get-live-market-data-less-than-0-1-second-lag-c85ee280ed93. Accessed May 26th, 2023.

286. Li, Fen et al. "Does Geopolitics Have an Impact on Energy Trade? Empirical Research on Emerging Countries." *Sustainability*, vol. 13, no. 9, 2021, doi:10.3390/su13095199.

287. Li, Susan. "Named Entity Recognition with Nltk and Spacy." Towards Data Science https://towardsdatascience.com/named-entity-recognition-with-nltk-and-spacy-8c4a7d88e7da. Accessed August 4th, 2021.

288. Lien, Kathy. "The Major Central Banks." *Investopedia* https://www.investopedia.com/articles/forex/06/centralbanks.asp. Accessed April 21st, 2022

289. Liew, Lucas. "Sentiment Analysis with Python – a Beginner's Guide." *Algo Trading 101 Blog*, vol. 2021, Algo Trading 101. https://algotrading101.com/learn/sentiment-analysis-python-guide/. Accessed: December 3rd, 2020

290. Lind, Fabienne; Jakob-Moritz Eberl; Sebastian Galyga; Tobias Heidenreich; Hajo G. Boomgaarden; Beatriz Herrero Jiménez; Rosa Berganza. "A Bridge over the Language Gap: Topic Modelling for Text Analyses across Languages for Country Comparative Research." Working Paper, Reminder, November 2019. https://www.reminder-project.eu/wp-content/uploads/2019/11/D8.7.pdf.

291. LiveCharts.co.uk. "Live Charts - Crude Oil Chart." LiveCharts https://www.livecharts.co.uk/MarketCharts/crude.php. Accessed May 26th, 2023.

292. Lofaro, Lucas. "Working with Json Data in Python." Real Python https://realpython.com/python-json/. Accessed December 21st, 2020.

293. Lydersen, Stian. "Adjustment of P-Values for Multiple Hypotheses." Medicine and Numbers, vol. 2024, Tidsskriftet, 2021. https://tidsskriftet.no/en/2021/09/medicine-and-numbers/adjustment-p-values-multiple-hypotheses#:~:text=The%20simplest%20method%20is%20a,comparing%20with%20the%20significance%20level.

294. Maas, Andrew L.; Raymond E. Daly; Peter T. Pham; Dan Huang; Andrew Y. Ng; Christopher Potts. "Learning Word Vectors for Sentiment Analysis." *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, Association for Computational Linguistics, 2011, pp. 142--50. http://www.aclweb.org/anthology/P11-1015.

295. Mankiw, N. Gregory. *Macroeconomics.* Worth Publishers, 2012.

296. Marcus, Adam et al. "Twitinfo: Aggregating and Visualizing Microblogs for Event Exploration." *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11)*, ACM, New York, NY, USA, 2011, pp. 227-36. doi:10.1145/1978942.1978975.

297. Marshall, Christopher. "What Is Named Entity Recognition (Ner) and How Can I Use It?" Medium https://medium.com/mysuperai/what-is-named-entity-recognition-ner-and-how-can-i-use-it-2b68cf6f545d. Accessed July 28th, 2023.

298. Marshall, Tim. *Prisoners of Geography : Ten Maps That Explain Everything About the World.* First Scribner paperback ed ed., Scribner, 2016.

299. Marshall, Tim. *The Power of Geography : Ten Maps That Reveal the Future of Our World.* First Scribner trade paperback ed. ed., Scribner an Imprint of Simon & Schuster, 2022.

300. Martin, Michelle. "29 Twitter Stats That Matter to Marketers in 2023." *Strategy.* Hootsuite https://blog.hootsuite.com/twitter-statistics/. Accessed June 23rd, 2023.

301. Massicotte, Philippe; Dirk Eddelbuettel. "Gtrendsr: Perform and Display Google Trends Queries." 1.5.1 ed., *CRAN,* 2022. https://github.com/PMassicotte/gtrendsR.

302. Matthes, Eric. *Python Crash Course : A Hands-on, Project-Based Introduction to Programming.* No Starch Press, 2019.

303. McClelland, Charles. "World Event/Interaction Survey (WEIS) Project, 1966-1978", Inter-university Consortium for Political and Social Research, 2006, doi: 10.3886/ICPSR05211.v3

304. McCormick, Chris. "How to Apply Bert to Arabic and Other Languages." Towards Data Science https://towardsdatascience.com/how-to-apply-bert-to-arabic-and-other-languages-5c3410ddd787. Accessed November 30th, 2021.

305. Metzler, Donald ; Congxing Cai; Eduard Hovy. "Structured Event Retrieval over Microblog Archives." *Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Association for Computational Linguistics, 2012, pp. 646–55.

306. Middleton, Stuart E. et al. "Real-Time Crisis Mapping of Natural Disasters Using Social Media." *IEEE Intelligent Systems*, vol. 29, no. 2, 2014, pp. 9-17, doi:10.1109/mis.2013.126.

307. Mishev, Kostadin et al. "Evaluation of Sentiment Analysis in Finance: From Lexicons to Transformers." *IEEE Access*, vol. 8, 2020, pp. 131662-82, doi:10.1109/access.2020.3009626.

308. Mitchell, Lewis et al. "The Geography of Happiness: Connecting Twitter Sentiment and Expression, Demographics, and Objective Characteristics of Place." *PLoS One*, vol. 8, no. 5, 2013, p. e64417, doi:10.1371/journal.pone.0064417.

309. Mittal, Anshul; Goel, Arpit. "Stock Prediction Using Twitter Sentiment Analysis." Stanford University, 2017. http://cs229.stanford.edu/proj2011/GoelMittal-StockMarketPredictionUsingTwitterSentimentAnalysis.pdf.

310. Monitor, Markets. "Metals & Mining Overview." ETF.com https://www.etf.com/topics/metals-mining. Accessed May 26th, 2023.

311. Montani, Ines. "Advanced Nlp with Spacy." Explosion https://course.spacy.io/en/. Accessed August 25th, 2021.

312. Morgan, Brett; Chris Broadfoot; Dave Holmes; Luke Mahe; Mark McDonald; Sam Thorogood; Sean Wohltman; Stephen McDonald. "Google-Maps-Services-Python." Google Inc., 2015. https://github.com/googlemaps/google-maps-services-python.

313. motazsaad, "arabic-sentiment-analysis", *Github.com*, 2019, https://github.com/motazsaad/arabic-sentiment-analysis

314. Mozetič, Igor et al. "Multilingual Twitter Sentiment Classification: The Role of Human Annotators." *PLoS One*, vol. 11, no. 5, 2016, p. e0155036, doi:10.1371/journal.pone.0155036.

315. Myers, Steven Lee. "Taiwan Loses Nicaragua as Ally as Tensions with China Rise." *New York Times*, New York Times, December 10th, 2021. https://www.nytimes.com/2021/12/10/world/asia/taiwan-nicaragua-china.html.

316. Neill, Daniel B. "New Directions in Artificial Intelligence for Public Health Surveillance." *IEEE Intelligent Systems*, vol. 27, no. 1, 2012, pp. 56-59, https://www.cs.cmu.edu/%7Eneill/papers/ieee-is2012.pdf.

317. Nelyon, Kaye. "How Many Tweets Does It Take to Trend?" News, vol. 2024, Jungle, 2023. https://jungle.marketing/news/how-many-tweets-does-it-take-to-trend/#:~:text=Trending%20topics%20on%20Twitter%20may,based%20on%20your%20location%20settings.

318. Ness, Eric. "Converting a Pandas Dataframe into a Tensorflow Dataset." When I Work Data https://medium.com/when-i-work-data/converting-a-pandas-dataframe-into-a-tensorflow-dataset-752f3783c168. Accessed May 6th, 2022.

319. Neufeld, Dorothy. "Mapped: The Top Export in Every Country." *Markets*. Visual Capitalist https://www.visualcapitalist.com/mapped-top-export-in-every-country/. Accessed March 8th, 2022

320. Ng, Ritchie. "Vectorization, Multinomial Naive Bayes Classifier and Evaluation." *Ritchie Ng*, vol. 2022, Ritchie Ng, 2022. https://www.ritchieng.com/machine-learning-multinomial-naive-bayes-vectorization/.

321. Nidhaloff. "Deep-Translator." GitHub, 2020. https://github.com/nidhaloff/deep-translator.

322. Nidhaloff. "How to Translate Text with Python." Analytics Vidhya https://medium.com/analytics-vidhya/how-to-translate-text-with-python-9d203139dcf5. Accessed March 10th, 2022.

323. Nielsen, Finn Årup. "A New Anew: Evaluation of a Word List for Sentiment Analysis in Microblogs. ." *ESWC 2011. Workshop on 'Making Sense of Microposts': Big things come in small packages 718 in {CEUR}*, May 29, 2011 - Jun 4, 2011. doi:10.48550/arxiv.1103.2903.

324. Nisar, Tahir M.; Man Yeung. "Twitter as a Tool for Forecasting Stock Market Movements: A short-Window Event Study." *The Journal of Finance and Data Science*, vol. 4, no. 2, 2018, pp. 101-19, doi:https://doi.org/10.1016/j.jfds.2017.11.002.

325. nlptown. "Bert-Base-Multilingual-Uncased-Sentiment " Hugging Face. https://huggingface.co/nlptown/bert-base-multilingual-uncased-sentiment?text=I+like+you.+I+love+you.

326. Nofsinger, John R. "Social Mood and Financial Economics." *Journal of Behavioral Finance*, vol. 6, no. 3, 2005, pp. 144-60.

327. Owoputi, Olutobi; Brendan O'Connor; Chris Dyer; Kevin Gimpel; Nathan Schneider; Noah A. Smith. "Improved Part-of-Speech Tagging for Online Conversational Text with Word Clusters." *NAACL 2013*, 2013. http://www.cs.cmu.edu/%7Eark/TweetNLP/owoputi+etal.naacl13.pdf.

328. Pak, Alexander; Patrick Paroubek. "Twitter as a Corpus for Sentiment Analysis and Opinion Mining." *Seventh International Conference on Language Resources and Evaluation ({LREC}'10)*, European Language Resources Association (ELRA), 2010. http://www.lrec-conf.org/proceedings/lrec2010/pdf/385_Paper.pdf.

329. Pandey, Parul. "Simplifying Sentiment Analysis Using Vader in Python (on Social Media Text)." Medium.com https://medium.com/analytics-vidhya/simplifying-social-media-sentiment-analysis-using-vader-in-python-f9e6ec6fc52f. Accessed January 18th, 2021.

330. Park, Eunjeong L.; Sungzoon Cho. "Konlpy: Korean Natural Language Processing in Python"." *Proceedings of the 26th Annual Conference on Human & Cognitive Language Technology*, 2014. https://konlpy.org/en/latest/.

331. Park, Lucy. "Nsmc." GitHub https://github.com/e9t/nsmc. Accessed November 21st, 2021.

332. Parker, Tim. "The Basics of Currency Trading." *Investopedia* https://www.investopedia.com/financial-edge/0412/the-basics-of-currency-trading.aspx. Accessed April 20th, 2021.

333. Pástor, Ľuboš and Pietro Veronesi. "Political Uncertainty and Risk Premia." *Journal of Financial Economics*, vol. 110, no. 3, 2013, pp. 520-45, doi:10.1016/j.jfineco.2013.08.007.

334. Pearse, Becky. "Human and Machine Translation: Both Alive and Kicking — and Here to Stay." *Translation & Localization Blog.* Smart Cat https://www.smartcat.com/blog/human-and-machine-translation-both-alive-and-kicking-and-here-to-stay/. Accessed November 15th, 2021.

335. Pedregosa, Fabian; Gaël Varoquaux; Alexandre Gramfort; Vincent Michel; Bertrand Thirion; Olivier Grisel; Mathieu Blondel; Peter Prettenhofer; Ron Weiss; Vincent Dubourg; Jake Vanderplas; Alexandre Passos; David Cournapeau; Matthieu Brucher; Matthieu Perrot; Édouard Duchesnay. "Scikit-Learn: Machine Learning in Python." *Journal of Machine Learning Research*, vol. 12, no. 85, 2011, p. 2825 - 30, https://www.jmlr.org/papers/v12/pedregosa11a.html.

336. Peersman, Gert; Frank Smets. "The Industry Effects of Monetary Policy in Euro Area." *Working Paper Series*, edited by European Central Bank, 2002. https://www.ecb.europa.eu/pub/pdf/scpwps/ecbwp165.pdf.

337. Pelgrim, Richard. "Arabic Nlp: Unique Challenges and Their Solutions." *Medium* https://towardsdatascience.com/arabic-nlp-unique-challenges-and-their-solutions-d99e8a87893d. Accessed June, 20th, 2023.

338. Piskorski, Jakub et al. "Real-Time Text Mining in Multilingual News for the Creation of a Pre-Frontier Intelligence Picture." ACM, 2010. doi:10.1145/1938606.1938610.

339. Pokharel, Rashmi et al. "Analysing Emerging Topics across Multiple Social Media Platforms." *Proceedings of the Australasian Computer Science Week Multiconference*, ACM, 2019. doi:10.1145/3290688.3290720.

340. Pota, Marco et al. "Multilingual Evaluation of Pre-Processing for Bert-Based Sentiment Analysis of Tweets." *Expert Systems with Applications*, vol. 181, 2021, p. 115119, doi:https://doi.org/10.1016/j.eswa.2021.115119.

341. Prabhakaran, Selva. "Topic Modeling Visualization – How to Present the Results of Lda Models?" *Machine Learning Plus,* December 4th, 2018, https://www.machinelearningplus.com/nlp/topic-modeling-visualization-how-to-present-results-lda-models/. Accessed March, 3rd, 2021

342. Prabhakaran, Selva. "Granger Causality Test in Python." *Machine Learning Plus* https://www.machinelearningplus.com/time-series/granger-causality-test-in-python/. Accessed May 26th, 2023.

343. Prabhakaran, Selva. "Topic Modeling with Gensim (Python)." *Machine Learning Plus* March 26th, 2018, https://www.machinelearningplus.com/nlp/topic-modeling-gensim-python/#9createbigramandtrigrammodels. Accessed March 3rd, 2021.

344. Prakruthi, V.; Sindhu, D.; Anupama Kumar, Dr. S. "Real Time Sentiment Analysis of Twitter Posts." *2018 3rd International Conference on Computational Systems and Information Technology for Sustainable Solutions (CSITSS)*, 2018, pp. 29-34. doi:10.1109/CSITSS.2018.8768774.

345. Preis, Tobias et al. "Quantifying Trading Behavior in Financial Markets Using Google Trends." *Scientific Reports*, vol. 3, no. 1, 2013, doi:10.1038/srep01684.

346. Press, The Associated. "The Philippines Accuses a Chinese Ship of Aiming a Laser at a Filipino Boat Crew." NPR, NPR, February 13th, 2023. https://www.npr.org/2023/02/13/1156475789/the-philippines-accuses-a-chinese-ship-of-aiming-a-laser-at-a-filipino-boat-crew.

347. Primusa, Dookoto_Sea;. "What Is the Difference between the Terms Accuracy and Validation Accuracy." Stack Overflow https://stackoverflow.com/questions/51344839/what-is-the-difference-between-the-terms-accuracy-and-validation-accuracy. Accessed May 22nd, 2021.

348. Psaledakis, Daphne. "Us Slaps Sanctions on Iranian, Chinese Targets over Tehran's Missile, Military Programs." *Reuters*, Reuters, June 6th, 2023. https://www.reuters.com/world/middle-east/us-slaps-sanctions-iranian-chinese-targets-action-over-tehrans-missile-military-2023-06-06/.

349. PV8, user9079492;. "Train_Test_Split Crashing Ram When Feeding with (Big) Numpy Arrays." Stack Overflow https://stackoverflow.com/questions/57493754/train-test-split-crashing-ram-when-feeding-with-big-numpy-arrays. Accessed May 6th, 2022.

350. Rabash, Rita;. "Python: How to Determine the Language?" Stack Overflow https://stackoverflow.com/questions/39142778/python-how-to-determine-the-language. Accessed February 7th, 2021.

351. Rajput, Nikhil Kumar; Bhavya Ahuja Grover; Vipin Kumar Rathi. "Word Frequency and Sentiment Analysis of Twitter Messages During Coronavirus Pandemic." *Computing Research Repository (CoRR)*, vol. abs/2004.03925, 2020, doi:10.48550/arxiv.2004.03925.

352. Ranaei, Samira; Arho, Suominen; Alan Porter; Stephen Carley. "Evaluating Technological Emergence Using Text Analytics: Two Case Technologies and Three Approaches." *Scientometrics*, vol. 122, no. 1, 2020, pp. 215-47, doi:10.1007/s11192-019-03275-w.

353. Rehurek, Radim; Petr Sojka. "Software Framework for Topic Modelling with Large Corpora." *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*, ELRA, 5/22/2010. https://pypi.org/project/gensim/.

354. Reback, Jeff ; Wes McKinney; jbrockmendel; Joris Van den Bossche; Tom Augspurger; Phillip Cloud; gfyoung; Sinhrks; Adam Klein; Matthew Roeschke; Simon Hawkins; Jeff Tratner; Chang She; William Ayd; Terji Petersen; Marc Garcia; Jeremy Schendel; Andy Hayden; MomIsBestFriend; Vytautas Jancauskas; Pietro Battiston; Skipper Seabold; chris-b1; h-vetinari; Stephan Hoyer; Wouter Overmeire; alimcmaster1; Kaiqi Dong; Christopher Whelan; Mortada Mehyar. "Pandas-Dev/Pandas: Pandas." latest ed., Zenodo, 2020. doi:10.5281/zenodo.3509134.

355. Reiff, Nathan. "The Top 3 Industries Affected by the Trade War with China." *Government News*. Investopedia https://www.investopedia.com/industries-most-likely-to-be-impacted-by-trade-disputes-with-china-in-2019-4580508#:%7E:text=Chip%20makers%20and%20electronics%20manufacturers,in%20a%20trade%20war%20scenario. Accessed March 10th, 2022.

356. Reitz, Kenneth; Requests Team. "Python Request Package." Python Request https://docs.python-requests.org/en/latest/user/quickstart/. Accessed January 27th, 2021.

357. Resch, Bernd et al. "Combining Machine-Learning Topic Models and Spatiotemporal Analysis of Social Media Data for Disaster Footprint and Damage Assessment." *Cartography and Geographic Information Science*, vol. 45, no. 4, 2018, pp. 362-76, doi:10.1080/15230406.2017.1356242.

358. Reuters. "France Plans Major Police Presence for June 6 Day of Protest." *Reuters*, Reuters, June 4th, 2023. https://www.reuters.com/world/europe/france-plans-major-police-presence-june-6-day-protest-2023-06-04/.

359. Reuters. "Russia Says U.S.-Built F-16s Could 'Accommodate' Nuclear Weapons If Sent to Ukraine." Reuters, Reuters, June 6th, 2023. https://www.reuters.com/world/russia-says-us-built-f-16s-could-accommodate-nuclear-weapons-if-sent-ukraine-2023-06-06/.

360. Revert, Félix. "An Overview of Topics Extraction in Python with Lda." Towards Data Science https://towardsdatascience.com/the-complete-guide-for-topics-extraction-in-python-a6aaa6cedbbc. Accessed Febuary 23rd, 2021.

361. Riberio, Marco Tulio; Sameer Singh; Carlos Guestrin. "Local Interpretable Model-Agnostic Explanations (Lime): An Introduction." https://www.oreilly.com/content/introduction-to-local-interpretable-model-agnostic-explanations-lime/. Accessed December 1st, 2020.

362. Rill, Sven; Dirk Reinel; Jörg Scheidt; Roberto V. Zicari. "Politwi: Early Detection of Emerging Political Topics on Twitter and the Impact on Concept-Level Sentiment Analysis." *Knowledge-Based Systems*, vol. 69, 2014, pp. 24-33, doi:https://doi.org/10.1016/j.knosys.2014.05.008.

363. Ritter, Alan; Sam Clark, Mausam; Oren Etzioni. "Named Entity Recognition in Tweets: An Experimental Study." *2011 Conference on Empirical Methods in Natural Language Processing*, Association for Computational Linguistics, July 27–31, 2011, pp. 1524–34. https://aclanthology.org/D11-1141.pdf.

364. Roberts, Margaret E.; Brandon M. Stewart; Dustin Tingley; Edoardo M. Airoldi. "The Structural Topic Model and Applied Social Science." *Neural Information Processing Society*, 2013, https://scholar.princeton.edu/files/bstewart/files/stmnips2013.pdf.

365. Roberts, Stephen J. "Novelty Detection Using Extreme Value Statistics." *IEE Proceedings on Vision, Image and Signal Processing*, vol. 146, 1999, pp. 124-29. http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.49.1338.

366. Rodríguez-Ibánez, Margarita; et al. "A Review on Sentiment Analysis from Social Media Platforms." Expert Systems with Applications, vol. 223, 2023, https://doi.org/10.1016/j.eswa.2023.119862.

367. Rognone, Lavinia ; Stuart S. Hyde; Sarah Zhang,. "News Sentiment in the Cryptocurrency Market: An Empirical Comparison with Forex." *International Review of Financial Analysis*, vol. 69, 2020, doi:https://doi.org/10.1016/j.irfa.2020.101462.

368. Rossum, Van; Guido & Drake; Fred L. *Python 3 Reference Manual.* CreateSpace, 2009.

369. Ruby, Daniel "45+ Tiktok Statistics for Every Marketer (August 2023)." *Demand Sage* https://www.demandsage.com/tiktok-user-statistics/#:%7E:text=TikTok%20has%20over%201.677%20billion,active%20users%20as%20of%202023. Accessed August 4th, 2023.

370. Ryall, Julian. "Older Drivers in Japan Are Mixing up Their Accelerate and Brake Pedals with Fatal Results." *East Asia.* South China Morning Press https://www.scmp.com/news/asia/east-asia/article/2050472/older-drivers-japan-are-mistaking-stop-and-go-fatal-results. Accessed May 26th, 2023.

371. Saeed, Zafar et al. "What's Happening around the World? A Survey and Framework on Event Detection Techniques on Twitter." *Journal of Grid Computing*, vol. 17, no. 2, 2019, pp. 279-312, doi:10.1007/s10723-019-09482-2.

372. Saha, Ankan; Vikas Sindhwani. "Learning Evolving and Emerging Topics in Social Media." ACM Press, 2012. doi:10.1145/2124295.2124376.

373. Sakaki, Takeshi et al. "Earthquake Shakes Twitter Users." ACM Press, 2010. doi:10.1145/1772690.1772777.

374. Sakamoto, Takuto and Hiroki Takikawa. "Cross-National Measurement of Polarization in Political Discourse: Analyzing Floor Debate in the U.S. The Japanese Legislatures." *2017 IEEE International Conference on Big Data (Big Data)*, IEEE, 2017. doi:10.1109/bigdata.2017.8258285.

375. Saldutti, Gregg. "Named Entity Recognition and Geocoding with R." *Towards Data Science*. Medium https://towardsdatascience.com/quick-guide-to-entity-recognition-and-geocoding-with-r-c0a915932895. Accessed July 31st, 2023.

376. Sanad, M. "Demystifying Bert: A Comprehensive Guide to the Groundbreaking Nlp Framework." vol. 2022, Analytics Vidhya, 2019. https://www.analyticsvidhya.com/blog/2019/09/demystifying-bert-groundbreaking-nlp-framework/.

377. Sankaranarayanan, Jagan et al. "Twitterstand: News in Tweets." *Proceedings of the 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems - GIS '09*, ACM Press, 2009. doi:10.1145/1653771.1653781.

378. Sayce, David. "The Number of Tweets Per Day 2022." vol. 2024, David Sayce, 2022. https://www.dsayce.com/social-media/tweets-day/.

379. sbcBI. "Sentiment_Analysis_Model." Hugging Face. https://huggingface.co/sbcBI/sentiment_analysis_model.

380. Schöne, Jonas Paul et al. "Negativity Spreads More Than Positivity on Twitter after Both Positive and Negative Political Situations." *Affective Science*, vol. 2, no. 4, 2021, pp. 379-90, doi:10.1007/s42761-021-00057-7.

381. Schmit, Sven. "Multiple Hypothesis Testing." vol. 2024, MultiTreaded, 2015. https://multithreaded.stitchfix.com/blog/2015/10/15/multiple-hypothesis-testing/.

382. Schrodt, Philip A..; Deborah J. Gerner. "Analyzing International Event Data: A Handbook of Computer-Based Techniques." *KEDS Project*, edited by Penn State University, March 22, 2012, pp. 1 - 42. https://parusanalytics.com/eventdata/papers.dir/AIED.2012.ch1.pdf12/1/2020.

383. Seif, George. "The 5 Clustering Algorithms Data Scientists Need to Know." Towards Data Science https://towardsdatascience.com/the-5-clustering-algorithms-data-scientists-need-to-know-a36d136ef68. Accessed April 22nd, 2021.

384. Selvan, Lokmanyathilak Govindan Sankar; Teng-Sheng Moh. "A Framework for Fast-Feedback Opinion Mining on Twitter Data Streams." IEEE, 2015. doi:10.1109/cts.2015.7210440.

385. Selvaraju, Sendhilkumar; Nachiyar S Nandhini; G.S. Mahalakshmi. "Novelty Detection Via Topic Modeling in Research Articles." *Third International Conference on Computer Science, Engineering & Applications*, edited by David C. Wyld, 2013, pp. 401-10. doi:10.5121/csit.2013.3542.

386. SenthilPerumal, Chidam. "The Intuition Behind Latent Dirichlet Allocation (Lda)." *Analytics Vidhya* https://medium.com/analytics-vidhya/the-intuition-behind-latent-dirichlet-allocation-lda-fb1e1fb01543. Accessed May 13th, 2023.

387. Seo, Yoonjung; Brad Lendon; Junko Ogura. "North Korea Says It Tested Icbm in Surprise Drill." *CNN*, CNN, February 18th, 2023. https://www.cnn.com/2023/02/18/asia/north-korea-missile-launch-intl-hnk-ml/index.html.

388. Shahid, Nauman. "How to Explore Twitter Streaming Data Using Python and Spark." *Analytics Vidhya* https://medium.com/analytics-vidhya/exploring-twitter-streaming-data-using-python-and-spark-3f4f189ec660. Accessed November 22nd, 2021.

389. Shahrokhi, Sedighe. "Iran News in Brief – June 1, 2023." *National Council of Resistance of Iran* https://www.ncr-iran.org/en/news/iran-news-in-brief-june-1-2023/. Accessed June 30th, 2023.

390. Shaikh, JalFaizy. "Essentials of Deep Learning – Sequence to Sequence Modelling with Attention (Using Python)." vol. 2022, Analytics Vudhya, 2018. https://www.analyticsvidhya.com/blog/2018/03/essentials-of-deep-learning-sequence-to-sequence-modelling-with-attention-part-i/?utm_source=blog&utm_medium=understanding-transformers-nlp-state-of-the-art-models.

391. Shepherd, Jack. "23 Essential Twitter Statistics You Need to Know in 2023." *The Social Shepherd* https://thesocialshepherd.com/blog/twitter-statistics#:%7E:text=Here's%20an%20overview%20of%20Twitter's,year%20olds%3A%2038.5%25%20of%20users. Accessed August 3rd, 2023.

392. Shukri, Sarah E. et al. "Twitter Sentiment Analysis: A Case Study in the Automotive Industry." *2015 IEEE Jordan Conference on Applied Electrical Engineering and Computing Technologies (AEECT)*, IEEE, 2015. doi:10.1109/aeect.2015.7360594.

393. Silge, Julia Robinson David. *Text Mining with R : A Tidy Approach.* O'Reilly Media, 2017.

394. Sloan, Luke; Jeffrey Morgan. "Who Tweets with Their Location? Understanding the Relationship between Demographic Characteristics and the Use of Geoservices and Geotagging on Twitter." *PLoS One*, vol. 10, no. 11, 2015, doi:10.1371/journal.pone.0142209.

395. Sim, YuJeong, Dai Yeol Yun; Chi-gon Hwang; Seok-Jae Moon. "Korean Sentiment Analysis Using Natural Network: Based on Ikea Review Data." *International Journal of Internet, Broadcasting and Communication*, vol. 13, no. 2, 2021, pp. 173-78 doi:http://dx.doi.org/10.7236/IJIBC.2021.13.2.173.

396. Smeaton, Josh; ars. "What Is the Best Way to Call a Script from Another Script? [Closed]." Stack Overflow https://stackoverflow.com/questions/1186789/what-is-the-best-way-to-call-a-script-from-another-script. Accessed March 15th, 2023.

397. Smith, Aubree. "23 Essential Twitter Statistics to Guide Your Strategy in 2023." *Sprout Blog*. sproutsocial https://sproutsocial.com/insights/twitter-statistics/. Accessed June 15th, 2023.

398. Smith, Stephen; Anthony O'Hare. "Comparing Traditional News and Social Media with Stock Price Movements; Which Comes First, the News or the Price Change?" *Journal of Big Data*, vol. 9, no. 1, 2022, doi:10.1186/s40537-022-00591-6.

399. Smith, Tim. "Currency Depreciation." Investopedia https://www.investopedia.com/terms/c/currency-depreciation.asp#:%7E:text=Countries%20with%20weak%20economic%20fundamentals,its%20trade%20deficit%20over%20time. Accessed April 22nd, 2022.

400. Sohangir, Sahar et al. "Financial Sentiment Lexicon Analysis." IEEE, 2018. doi:10.1109/icsc.2018.00052.

401. Standards, National Institute of. "Engineering Statistics Handbook: Stationarity." *Process or Product Monitoring and Control*. National Institute of Standards https://www.itl.nist.gov/div898/handbook/pmc/section4/pmc442.htm#:%7E:text=Stationarity%20can%20be%20defined%20in,no%20periodic%20fluctuations%20(seasonality). Accessed June 14th, 2023.

402. Story, Rob. "Folium." https://python-visualization.github.io/folium/index.html#. Accessed August 24th, 2021.

403. Strycharz, Joanna; Nadine Strauss; Damian Trilling. "The Role of Media Coverage in Explaining Stock Market Fluctuations: Insights for Strategic Financial Communication." *International Journal of Strategic Communication*, vol. 12, 2017, pp. 1-19, doi:10.1080/1553118X.2017.1378220.

404. Surowiecki, James "What Social Media Is Doing to Finance." *Ideas*. The Atlantic https://www.theatlantic.com/ideas/archive/2023/03/silicon-valley-bank-run-social-media-financial-crisis/673375/. Accessed May 26th, 2023.

405. Szymanowski, Joshua. "Creating a Custom Corpus." *Python in Plain English* https://python.plainenglish.io/creating-a-custom-corpus-928789c089e7. Accessed December 17th, 2020

406. TASS 2020. "Workshop on Semantic Analysis at Sepln 2020." 2020. http://tass.sepln.org/2020/.

407. Taira, Breena R. MD, ; Vanessa Kreger, MD; Aristides Orue; Lisa C. Diamond, MD. "A Pragmatic Assessment of Google Translate for Emergency Department Instructions." *Journal of General Internal Medicine*, vol. 36, no. 11, 2021, pp. 3361-5, doi: 10.1007/s11606-021-06666-z.

408. Tanskanen, Aapo. "How to Classify Text in 100 Languages with a Single Nlp Model." Gofore https://gofore.com/en/how-to-classify-text-in-100-languages-with-a-single-nlp-model/. Accessed December 8$^{\text{th}}$, 2020.

409. Team, Data Science Learner. "How to Use Custom Stopwords Python Nlp ? Lets Add Stopwords Python." *Data Science Learner* https://www.datasciencelearner.com/custom-stopwords-python-nlp/. 2021.

410. TensorFlow. "Neural Machine Translation with a Transformer and Keras." *Text Generation*, TensorFlow, 2022. https://www.tensorflow.org/text/tutorials/transformer.

411. TensorFlow. "Text Classification with an Rnn." *Text Generation*, TensorFlow, 2022. https://www.tensorflow.org/text/tutorials/text_classification_rnn.

412. Tetlock, Paul C. "Giving Content to Investor Sentiment: The Role of Media in the Stock Market." *The Journal of Finance*, vol. 62, no. 3, 2007, pp. 1139-68, doi:10.1111/j.1540-6261.2007.01232.x.

413. Tetlock, Paul C. et al. "More Than Words: Quantifying Language to Measure Firms' Fundamentals." *The Journal of Finance*, vol. 63, no. 3, 2008, pp. 1437-67, doi:10.1111/j.1540-6261.2008.01362.x.

414. Thurman, Walter; Mark E. Fisher. "Chickens, Eggs, and Causality, or Which Came First?" *American Journal of Agricultural Economics*, vol. 70, no. 2, 1988, pp. 237-38, https://webdoc.agsci.colostate.edu/koontz/arec-econ535/papers/thurman%20fisher%20(ajae%201988).pdf.

415. Tibshirani, Robert et al. "Estimating the Number of Clusters in a Data Set Via the Gap Statistic." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 63, no. 2, 2001, pp. 411-23, doi:10.1111/1467-9868.00293.

416. Tiedemann, Jörg. "Opus – Parallel Corpora for Everyone." *Proceedings of the 19th annual conference of the European Association for Machine Translation EAMT)*, 2016. https://www.bjmc.lu.lv/fileadmin/user_upload/lu_portal/projekti/bjmc/Contents/4_2_28_Products.pdf.

417. Tiedemann, Jörg; Santhosh Thottingal. "{Opus-Mt} — {B}Uilding Open Translation Services for the {W}Orld." *Proceedings of the 22nd Annual Conferenec of the European Association for Machine Translation (EAMT)*, 2020. https://github.com/Helsinki-NLP/Opus-MT.

418. TREASURY, U.S. DEPARTMENT OF THE. "Treasury Targets Iran's International Uav Procurement Network." March 9th, 2023. https://home.treasury.gov/news/press-releases/jy1331.

419. Twin, Alexandra. "6 Factors That Influence Exchange Rates." Investopedia https://www.investopedia.com/trading/factors-influence-exchange-rates/. Accessed April 21st, 2022.

420. Twitter / X. "Building High-Quality Filters for Getting Twitter Data." *Tutorials*. developer.twitter.com https://developer.twitter.com/en/docs/tutorials/building-high-quality-filters. Accessed May 16th, 2022

421. Twitter / X. "Filtered Stream." Twitter https://developer.twitter.com/en/docs/twitter-api/tweets/filtered-stream/integrate/build-a-rule. Accessed May 16th, 2022

422. Twitter / X. "Stream Tweets in Real-Time." Developer Platform https://developer.twitter.com/en/docs/tutorials/stream-tweets-in-real-time. Accessed May 16th, 2022

423. Twitter / X. "Twitter Api V2: Early Access." Developer Platform https://developer.twitter.com/en/docs/twitter-api/tutorials. Accessed May 16th, 2022

424. Tracey, Thomas. "Language Translation with Rnns." vol. 2022, Towards Data Science, 2019. https://towardsdatascience.com/language-translation-with-rnns-d84d43b40571.

425. Tuchman, Barbara W. *The March of Folly : From Troy to Vietnam.* Ballantine Books, 2014.

426. Tuchman, Barbara W. *A Distant Mirror : The Calamitous 14th Century.* Penguin Books, 2017.

427. Tuchman, Barbara W. *The Guns of August.* Ballantine Books, 2014.

428. Turner, Karen. "Google Translate Is Getting Really, Really Accurate." *Washington Post*, Washington Post, 2016. https://www.washingtonpost.com/news/innovations/wp/2016/10/03/google-translate-is-getting-really-really-accurate/?variant=d846d44221ecba95.

429. Tvarožek, Tomlein; Jozef Matúš. "Method for Novelty Recommendation Using Topic Modelling." *UMAP 2014 Extended Proceedings*, edited by Iván Cantador; Min Chi; Rosta Farzan; Robert Jäschke, July 7-11, 2014. http://ceur-ws.org/Vol-1181/nra2014_paper_04.pdf.

430. Ughelli, Daniel; John Demos; Ethan James Wright; Aurelia Specket. "Twitter-Api-V2-Sample-Code/Filtered-Stream/Filtered_Stream.Py." *twitterdev*. Twitter https://github.com/twitterdev/Twitter-API-v2-sample-code/blob/main/Filtered-Stream/filtered_stream.py. Accessed August 12th, 2021

431. Uhl, Matthias W. "Reuters Sentiment and Stock Returns." *Journal of Behavioral Finance*, vol. 15, no. 4, 2014, pp. 287-98.

432. Urbain, Jay "Mining Heart Disease Risk Factors in Clinical Text with Named Entity Recognition and Distributional Semantic Models." *Journal of Biomedical Informatics*, vol. 58, 2015, pp. S143-S49, doi:https://doi.org/10.1016/j.jbi.2015.08.009.

433. Urquhart, Andrew. "What Causes the Attention of Bitcoin?" *Economics Letters*, vol. 166, 2018, pp. 40-44, doi:10.1016/j.econlet.2018.02.017.

434. Valkanas, George and Dimitrios Gunopulos. "Event Detection from Social Media Data." *IEEE Data Eng. Bull.*, vol. 36, no. 3, 2013, pp. 51-58, http://sites.computer.org/debull/A13sept/p51.pdf.

435. Varun. "Python: How to Append a New Row to an Existing Csv File?" *thisPointer Programming Tutorial*s https://thispointer.com/python-how-to-append-a-new-row-to-an-existing-csv-file/. Accessed February 3rd, 2021.

436. Vaswani, Ashish, et al. "Attention Is All You Need." *31st Conference on Neural Information Processing Systems (NIPS 2017)*, 2017. doi:10.48550/arxiv.1706.03762.

437. Vohra, Saifee; Jay Teraiya. "Applications and Challenges for Sentiment Analysis : A Survey." *International Journal of Engineering Research & Technology (IJERT)*, vol. 2, no. 2, 2013, https://www.ijert.org/research/applications-and-challenges-for-sentiment-analysis-a-survey-IJERTV2IS2331.pdf.

438. Vosoughi, Soroush et al. "The Spread of True and False News Online." Science, vol. 359, no. 6380, 2018, pp. 1146-51, doi:10.1126/science.aap9559.

439. Waddell, Melanie. "Svb Collapse Is 'First Twitter-Fueled Bank Run': Lawmaker." ALM Think Advisor https://www.thinkadvisor.com/2023/03/13/svb-collapse-a-twitter-fueled-bank-run/. Accessed May 26th, 2023.

440. Wallach, Hanna M. et al. "Evaluation Methods for Topic Models." *Proceedings of the 26th Annual International Conference on Machine Learning - ICML '09*, ACM Press, 2009. doi:10.1145/1553374.1553515.

441. Wang, Xuerui; Andrew McCallum. "Topics over Time: A Non-Markov Continuous-Time Model of Topical Trends." *Conference on Knowledge Discovery and Data Mining (KDD)*, edited by ACM, ACM, 2006. https://people.cs.umass.edu/%7Emccallum/papers/tot-kdd06s.pdf.

442. warmspringwinds. "Sentiment Analysis of Non-English Texts." https://stackoverflow.com/questions/29169732/sentiment-analysis-of-non-english-texts%20-. Accessed December 2nd, 2020.

443. Westgate, Martin J. ; Philip S. Barton; Jennifer C. Pierson; David B. Lindenmayer. "Text Analysis Tools for Identification of Emerging Topics and Research Gaps in Conservation Science." *Conservation Biology*, vol. 29, no. 6, 2015, pp. 1606-14, doi:10.1111/cobi.12605.

444. Wieringa, Jeri. "Intro to Beautiful Soup." http://programminghistorian.org/en/lessons/retired/intro-to-beautiful-soup. Accessed January 8th, 2021.

445. Wieringa, Jeri. "Ways to Compute Topics over Time, Part 1." https://jeriwieringa.com/2017/06/21/Calculating-and-Visualizing-Topic-Significance-over-Time-Part-1/. Accessed June 20th, 2023.

446. Windsor, L. C. et al. "Automated Content Analysis across Six Languages." *PLoS One*, vol. 14, no. 11, 2019, p. e0224425.

447. Wold, Charlotte. "Top 3 Defense Etfs (Ppa, Xar)." Investopedia https://www.investopedia.com/news/top-3-defense-etfs-ppa-xar/. Accessed May 26th, 2023.

448. Wolf, Thomas; Lysandre Debut; Victor Sanh; Julien Chaumond; Clement Delangue; Anthony Moi; Pierric Cistac; Tim Rault; Rémi Louf; Morgan Funtowicz; Joe Davison; Sam Shleifer; Patrick von Platen; Clara Ma; Yacine Jernite; Julien Plu; Canwen Xu; Teven Le Scao; Sylvain Gugger; Mariama Drame; Quentin Lhoest; Alexander M. Rush. " Transformers: State-of-the-Art Natural Language Processing." *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, Association for Computational Linguistics, 2020, pp. 38–45. doi:10.18653/v1/2020.emnlp-demos.6.

449. Workman, Daniel. "Top Water and Ice Exporters." *World's Top Exports*. World's Top Exports https://www.worldstopexports.com/ice-exporters/. Accessed March 8th, 2022

450. Wu, Yonghui, et al. "Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation." *CoRR*, 2016, doi:10.48550/arxiv.1609.08144.

451. Yadav, Kajal. "Scraping 1000's of News Articles Using 10 Simple Steps." *Towards Data Science* https://towardsdatascience.com/scraping-1000s-of-news-articles-using-10-simple-steps-d57636a49755. Accessed January 19th, 2021.

452. Yalçın, Orhan G. "Sentiment Analysis in 10 Minutes with Bert and Tensorflow." *Natural Language Processing*. Towards Data Science https://towardsdatascience.com/sentiment-analysis-in-10-minutes-with-bert-and-hugging-face-294e8a04b671. Accessed November 9th, 2021

453. Yamada, Ikuya; Akari Asai; Jin Sakuma; Hiroyuki Shindo; Hideaki Takeda; Yoshiyasu Takefuji; Yuji Matsumoto. "Wikipedia2vec: An Efficient Toolkit for Learning and Visualizing the Embeddings of Words and Entities from Wikipedia." *The 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP 2020)*, 16th – 20th November 2020. doi:10.48550/arxiv.1812.06280.

454. Yang, Jingchao et al. "A Twitter Data Credibility Framework—Hurricane Harvey as a Use Case." *ISPRS International Journal of Geo-Information*, vol. 8, no. 3, 2019, p. 111, doi:10.3390/ijgi8030111.

455. Yang, Weiwei; Jordan Boyd-Graber, Philip Resnik. "A Multilingual Topic Model for Learning Weighted Topic Links across Corpora with Low Comparability." *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, Association for Computational Linguistics, 2019, pp. 1243–48. doi:10.18653/v1/d19.

456. Yang, Yiming et al. "A Study of Retrospective and on-Line Event Detection." *Proceedings of the 21st annual international ACM SIGIR conference on Research and development in information retrieval*, 1998, pp. 28-36.

457. ydaigo. "Bert-Base-Japanese-Sentiment." Hugging Face, 2020. https://huggingface.co/daigo/bert-base-japanese-sentiment.

458. Yim, Hyunsu "North Korea Fires Ballistic Missile as U.S.-South Korean Drills Go On." *Reuters*, Reuters, March 19th, 2023. https://www.reuters.com/world/asia-pacific/north-korea-fired-ballistic-missile-towards-sea-off-east-coast-yonhap-2023-03-19/.

459. Yu, Weiren et al. "Ring: Real-Time Emerging Anomaly Monitoring System over Text Streams." *IEEE Transactions on Big Data*, vol. 5, no. 4, 2019, pp. 506-19, doi:10.1109/tbdata.2017.2672672.

460. Yu, Manzhu et al. "Spatiotemporal Event Detection: A Review." *International Journal of Digital Earth*, vol. 13, no. 12, 2020, pp. 1339-65, doi:10.1080/17538947.2020.1738569.

461. Yu, Lilin; et al. "Frances: Cloud-Based Historical Text Mining with Deep Learning and Parallel Processing." 2023 IEEE 19th International Conference on e-Science (e-Science), 2023, pp. 1-10. doi:10.1109/e-Science58273.2023.10254798.

462. Yuning, Chen; Liu Lianzhong. "Development and Research of Topic Detection and Tracking." IEEE, 2016. doi:10.1109/icsess.2016.7883041.

463. Zach. "How to Perform a Granger-Causality Test in Python." statology https://www.statology.org/granger-causality-test-in-python/. Accessed May 26th, 2023.

464. Zaharia, Matei; et al. "Apache Spark: A Unified Engine for Big Data Processing." Communications of the ACM, vol. 59, no. 11, 2016, pp. 56–65, https://doi.org/10.1145/2934664.

465. Zemek, Petr. "Restarting a Python Script within Itself." *Blog of a Software Engineer*, vol. 2021, 2014. https://blog.petrzemek.net/2014/03/23/restarting-a-python-script-within-itself/.

466. Zhang, Amy J. et al. "Using Twitter for Data Collection with Health-Care Consumers." *International Journal of Qualitative Methods*, vol. 17, no. 1, 2018, p. 160940691775078, doi:10.1177/1609406917750782.

467. Zhang, Xiaoming et al. "Event Detection and Popularity Prediction in Microblogging." *Neurocomputing*, vol. 149, 2015, pp. 1469-80, doi:https://doi.org/10.1016/j.neucom.2014.08.045.

468. Zheng, Liyi et al. "Comparative Topic Analysis of Japanese and Chinese Bloggers." *2014 28th International Conference on Advanced Information Networking and Applications Workshops*, IEEE, 13-16 May 2014. doi:10.1109/waina.2014.107.

469. Zielinsk, Andrea ; Stuart E. Middleton, Öcal Necmioglu,; Martin Hammitzsch. "Spatio-Temporal Decision Support System for Natural Crisis Management with Tweetcomp1." *Decision Support Systems III-Impact of Decision Support Systems for Global Environments*, May 29–31, 2013 2013. doi:10.1007/978-3-319-11364-7_2.

470. Zucchi, Kristina. "Why the 10-Year U.S. Treasury Yield Matters." Investopedia https://www.investopedia.com/articles/investing/100814/why-10-year-us-treasury-rates-matter.asp#:%7E:text=Because%20they%20are%20backed%20by,while%20rising%20prices%20lower%20yields. Accessed March 22nd, 2022.

471. "Japanese Prius Missile." edited by @Gesu_audio, Twitter, November 18th, 2018. https://twitter.com/Gesu_audio/status/1064073144355287041.

Appendix A: The Geopolitical Key Bigrams

This is the Appendix for the Geopolitical Topics and Key bigrams table in English.

Topics	**Source**	**Key Bigrams**
Goldstein Negative	Goldstein Paper	"military invasion", "military attack", "military clash", "military assault", "seize position"
Goldstein Negative	Goldstein Paper	"seize position", "nonmilitary destruction", " nonmilitary injury", "force mobilitization", "force exercise"
Goldstein Positive	Goldstein Paper	"diplomatic recognition", "substantive agreement", "economic aid", "military assistance", "grant privilege"
Goldstein Positive	Goldstein Paper	"suspend sanctions", "call truce", "material assistance", "endorse position", "verbal support"
Cyber Warfare	Geo-Econ	"cyber attack", "cyber capabilities", "cyber defense", "cyber warfare", "cyber terrorism"
Oil Supply Shock	Geo-Econ	"oil crisis", "oil price", "petroleum exporting", "crude oil", "oil production"
US-China Relations	Geo-Econ	"China sea", "China relations", "China trade", "Thucydides Trap", "Taiwan sales"
Terrorism	Caldara Paper	"state terrorism", "counter terrorism", "terrorist attack", "political violence", "global terrorism"
Geopolitical Risk - General	Caldara Paper	"geopolitical risk", "geopolitical concerns", "geopolitical tension", "geopolitical uncertainty"
War Threats	Caldara Paper	"war risk", "war fear", "military threat"
Nuclear Threat	Caldara Paper	"nuclear weapons", "nuclear weapon", "ballistic missile", "nuclear conflict", "nuclear attack"

Table 28: This table contains the Topic, the source of the Key bigrams, and the bigrams themselves. Note that “Goldstein Positive” and “Goldstein Negative” have ten key bigrams to capture more potential emerging geopolitical events. “War Threats” and “Geopolitical Risks” have less than the typical five key bigrams as Caldara and Iacoviello [105], only included this amount terms in their topics. These key terms were translated for Spanish, French, Portuguese, Arabic, Japanese, and Korean.

Appendix B: Data Analysis Procedure Diagram For Geopolitical Risk Change Over Time Program

The procedure of the Real – Time Twitter / X Analysis program

Step 1: Twitter / X Filter API activated with the different Topics and Keywords in the 7 languages, Filtered Tweets start coming in and stored in json files, as of now, they are stored in 100 tweet batches

Step 2: The Call5min file is activated, and the storage csv files are created. The functions for the analysis are called by this file and the tweet json files start to be processed.

Step 3: The most recent json file is taken and converted to a Pandas data frame. Each tweet in the json file is first labeled with a language and a date-time variable is added.

Step 4: The remaining tweets are then separated into the different language data frames.

Step 5: Each data frame is then processed by the language analysis functions for each language. The sentiment analysis, and the coordinates of the locations mentioned in the tweets are added to the language data frames.

Step 6: The separate data frames and then brought back together into one complete data frame.

Step 7: Any tweets that have the same id number in the complete data frame are removed, creating the "no dups" data frame

Step 8.1: First the "no dups" dataframe is processed

Step 8.2: The tweets in the same topic are grouped together and counted, creating a topic count data frame

Step 8.3: The tweets in the same topic are grouped together and the sentiments per topic are calculated, creating a topic sentiment data frame

Step 8.4: Topic Count data frame and Topic Sentiment data frame are outputted to the output csv files

Step 8.5: The output files are then imported back into the program and the change between the previous row and the current row is calculated. The changes of topic count and topic sentiment are then also outputted to the change output csv files.

Step 9: The full complete data frame from Step 6 is then mapped using the folium python package and a HTML map is output with the topics color coded and the locations geotagged on the map

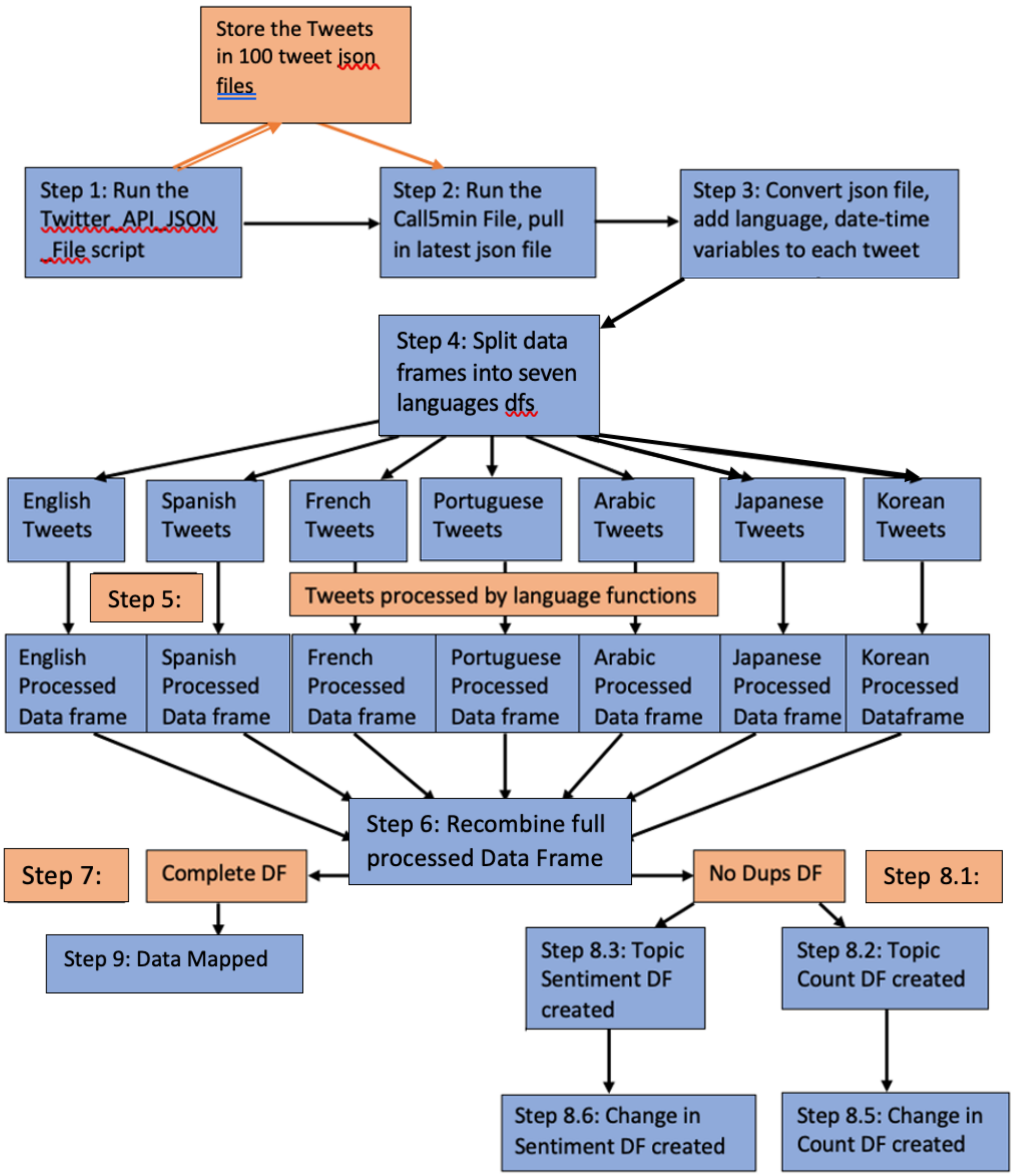


Figure 48: Work Process Chart for the Language Analysis Section of the Automatic Evaluation of Geopolitical Risk

Appendix C: Data Analysis Procedure Diagram for Topics over Time

Step 1: Run the Filter Twitter API and create the tweet json files

Step 2: Based on a chosen time scale, import all the files in the tweet json folder and create a full dataset out of them

Step 3: Create the Year, Month, Day, Hour, Minute, and Second variables so that change over time can occur

Step 4: Split the data frame by language into the seven separate data frames

Step 5: Remove the stop words from each of the tweets in each language

Step 6: Send each language data frame through the LDA topic modeling functions, which will dynamically get the optimal number of topics for the data, and then dynamically get the optimal parameters and run the topic models for each language and get the topics over time based on the time scale chosen.

Step 7: Take the results from the topics over time from each language and send them through the different language visualization functions

Step 8: Rerun this process for the next time scale using the sleep function in the implementation function in the CallTM program so the next topics are created and compare the visualizations over each time-frame to see what topics have changed

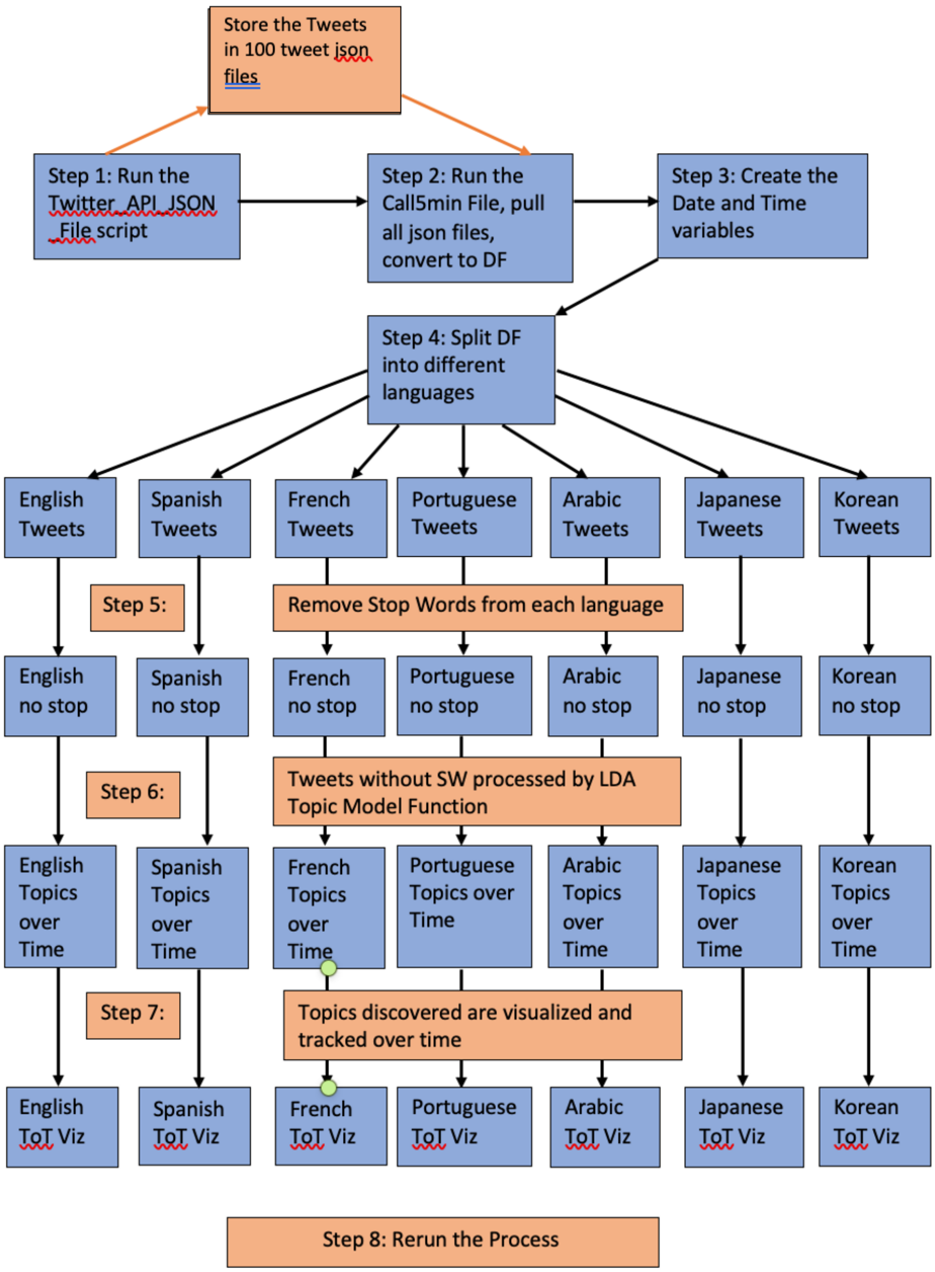


Figure 6: Work Process Chart for the Topic Modeling over Time Section of the Automatic Evaluation of Geopolitical Risk

Appendix D: Computational Resources and Data Set Sizes

This appendix covers the computational resources and the data set sizes used in the training of the RNNs. Table 42 details the computer I used to train the RNNs and run the program collecting and analyzing tweets. Table 43 contains the sizes of the datasets used in the training of the RNN.

Computer System	**Description**
Computer	2019 MacBook Air, Retina, 13-inch
Processor	1.6 GHz Dual-Core Intel Core i5
Graphics	Intel UHD Graphics 617 1536 MB
Memory	8 GB 2133 MHz LPDDR3

Table 42: Computational Resources used for training and data processing.

Language	**Name**	**Size MB**
English	IMDB Movie	66.2
French	Gamebusterz / TheophileBlard	139.1
Arabic	Motazsaad	19.2
Japanese	Darkmap	7.36
Korean	Park [316]	4.67
Portuguese	Augustop / Dias	20.3
Spanish	TASS 2020 / Kaggle	2.2

Table 43: RNN Training Data Sizes for Each Language

Appendix E: The Neural Network Structures of the Sentiment RNNs for the Analyses

These tables are the various neural network architectures I used for the RNNs I created, in addition Table 9 contains the architectures for the BERT models I implemented as well. The first layer is the wiki40b [142, 210] embeddings from the TensorFlow Hub for each language. As stated by Géron [194], using pre-trained embeddings can greatly improve the effectiveness of neural networks as the word embeddings used to build the network would have hundreds of millions of parameters as opposed to an embedding built off the training data which would have significantly smaller number. These were constructed in Python through the TensorFlow package which each layer shown below. These RNNs were designed to only return a positive or negative result, with no neutral option. Through testing we found, due to the lack of verbiage of short text, the RNNs defaulted to a neutral option unless the text was extremely positive or negative. This resulted in us missing a large number of potentially significant sentiment scores which were incorrectly labeled as neutral. By removing the neutral category, we forced the RNNs to choose if a tweet was positive or negative, thus achieving a greater number of responses which allowed us to make the most out of our data.

Layers	**Output Shape**	**Params**
wiki40b-lm-en	(None, None, 768)	116,768,000
Dense	(None, None, 128)	98,432
Dropout (10%)	(None, None, 128)	0
GRU	(None, 128)	99,072
Dropout (10%)	(None, 128)	0
Dense	(None, 1)	129

Table 1: English RNN Architecture

Layers	**Output Shape**	**Params**
wiki40b-lm-fr	(None, None, 768)	116,768,000
Dense	(None, None, 128)	98,432
Dropout (10%)	(None, None, 128)	0
GRU	(None, 128)	99,072
Dropout (10%)	(None, 128)	0
Dense	(None, 1)	129

Table 2: French RNN Architecture

Layers	Output Shape	Params
wiki40b-lm-ar	(None, None, 768)	116,768,000
Dense	(None, None, 128)	98,432
GRU	(None, None, 128)	99,072
Dropout (25%)	(None, None, 128)	0
GRU	(None, None, 64)	37,248
Dropout (20%)	(None, None, 64)	0
GRU	(None, 32)	9,408
Dropout (15%)	(None, 32)	0
Dense	(None, 1)	33

Table 3: Arabic RNN Architecture

Layers	Output Shape	Params
wiki40b-lm-ja	(None, None, 768)	116,768,000
Dense	(None, None, 256)	196,864
Dropout (10%)	(None, None, 256)	0
Bidirectional GRU	(None, None, 256)	296,448
Dropout (10%)	(None, None, 256)	0
Bidirectional GRU	(None, 256)	296,448
Dropout (10%)	(None, 256)	0
Dense	(None, 1)	257

Table 4: Japanese RNN Architecture

Layers	Output Shape	Params
wiki40b-lm-ko	(None, None, 768)	116,768,000
Dense	(None, None, 128)	98,432
GRU	(None, None, 128)	99,072
Dropout (25%)	(None, None, 128)	0
GRU	(None, None, 64)	37,248
Dropout (20%)	(None, None, 64)	0
GRU	(None, 32)	9,408
Dropout (15%)	(None, 32)	0
Dense	(None, 1)	33

Table 5: Korean RNN Architecture

Layers	**Output Shape**	**Params**
wiki40b-lm-pt	(None, None, 768)	116,768,000
Dense	(None, None, 128)	98,432
Dropout (30%)	(None, None, 128)	0
GRU	(None, None, 64)	37,248
Dropout (25%)	(None, None, 64)	0
GRU	(None, None, 64)	24,960
Dropout (25%)	(None, None, 64)	0
GRU	(None, 32)	9,408
Dropout (20%)	(None, 32)	0
Dense	(None, 1)	33

Table 6: Portuguese RNN Architecture

Layers	**Output Shape**	**Params**
wiki40b-lm-es	(None, None, 768)	116,768,000
Dense	(None, None, 128)	98,432
Dropout (25%)	(None, None, 128)	0
GRU	(None, None, 64)	37,248
Dropout (20%)	(None, None, 64)	0
GRU	(None, 32)	9,408
Dropout (15%)	(None, 32)	0
Dense	(None, 1)	33

Table 7: Spanish RNN Architecture

Language	Total Params	Data Count	Layers	Batch Size	Epochs	Data Balance
English	116,965,633	50,000	4	4	5	50%
Spanish	116,913,121	5,764	5	2	7	50%
French	116,965,633	415,702	4	128	7	50%
Portuguese	116,938,081	200,000	6	4	5	50%
Arabic	117,012,193	102,196	6	4	5	50%
Japanese	117,558,017	20,000	5	5	7	50%
Korean	117,012,193	150,000	6	5	5	50%

Table 8: These are the hyperparameters for the RNNs models that I constructed. The Total Params are the number of parameters from wiki40b-lm layers plus the additional parameters specified by each RNN. The Data Count is the number of data points used to train the RNN models, Layers is how many neural network layers were used to train the models, Batch Size is the number of samples sent through the model during training at one time, Epochs is the number of times the data was cycled through in training the model, and Data Balance is the percent of positive sentiment values (in this case a 1) compared to the total data set. While the data size varies for each of the RNNs models this was done to make sure that most amount of data that could be used to create the most accurate RNN model with a balanced data set.

Language	Total Params	Vocab	Fine Tuned Count	Batch Size	Learning Rate	Layers
English	110M	30k	11.8k	128	3e-5	12
Spanish	110M	32k	8.4k	2,048	0.0001	12
French	110M	30k	140k	128	3e-5	12
Portuguese	110M	29.8k	.5k	16	2e-5	12
Arabic	110M	30k	23.3k	32	3e-5	12
Japanese	110M	30k	?	128	3e-5	12
Korean	102M	20k	50k	32	5e-5	12
Multilingual	550M	150k	24.3k	8,192	0.0001	24

Table 9: These are the hyperparameters for the BERT models that I implemented for my research in Chapter 3. Total Params is the number of parameters each BERT model was constructed with, Vocab is the size of the vocabulary for each BERT model that is used to tokenize the text. Fine Tuned Count is the number of sentiment analysis data points used to fine tune the BERT models for sentiment analysis, Batch Size is the number of samples sent through the model during training at one time, Learning Rate is the rate at which the models' weights are updated during the training process, and Layers is how many layers were used in the BERT model construction. The question mark under Fine Tuned Count for Japanese is due to the creator of the model not specifying how much data was used fined tune the Japanese sentiment analysis.

Appendix F: Financial Assets and Markets Investigated in Case Study 1

Table 20 below lists the various financial assets and markets I analyzed for this study. The **Caldara and Iacoviello**, and the **Amen** financial assets or markets come directly from their respective papers. **My Own** assets or markets come from a few different sources. I was interested in expanding on the assets listed in the other papers (such as Gold Futures and Oil Futures), I also wanted to look at smaller international markets or emerging markets (such as the Sensex or the USD-MXN FX Rate). In addition, I wanted to see if different crypto currencies outside of bitcoin reacted differently to geopolitical events. Lastly, since the study involved the Ukraine War, I wanted to see how the Natural Gas markets and Wheat Market responded to the crisis, as both Russian and Ukraine are two of the world's largest producers of Wheat, and Russia is the main source of Natural Gas for Europe. Table 21 is a reordering of the assets based on asset class.

Source	**Financial Asset or Market**
Caldara and Iacoviello (5)	Defense ETF, Metals and Mining ETF, Crude Oil Price, 2 Year US Treasury Yield, Steel Futures
Amen (19)	S&P 500 Index (US Stock Exchange), Morgan Stanley Capital International Index ("MSCI"), CSI 300 Index (Chinese Stock Exchange), FTSE 100 Index (UK Stock Exchange), Nikkei 225 (Japanese Stock Exchange), Bitcoin, USD vs. EUR, JPY, AUD, CNY, RUB, and ZAR FX Rates, VIX Index (Volatility Index), MSCI Futures, Bitcoin Futures, US High Yield (HY) ETF, US Investment Grade (IG) ETF, Gold Price, 10 Year US Treasury Yield
My Own (15)	Gold Futures, Crude Oil Futures, 10 Year US Treasury Yield Futures, S&P BSE Sensex (Indian Stock Exchange), 10 Year German Bond Yields, USD vs. GBP, MXN FX Rates, EUR-GBP FX Rate, Ethereum (ETH), ChainLink (LINK), Ripple (XPR), Binance Coin (BNB), Algorand (ALGO), Wheat Futures, Natural Gas Futures

Table 20: The Financial Assets and Markets I analyzed and their sources.

Asset Class	Asset or Market
Commodity (7)	Gold Price, Crude Oil Price, Gold Futures, Crude Oil Futures, Steel Futures, Wheat Futures, Natural Gas Futures
International Markets and Assets (5)	CSI 300, Nikkei 225, BSE Sensex, FTSE 100, 10 Year German Bond Yield
U.S. Based Markets and Assets (12)	S&P 500, MSCI, VIX, 2 Year US Treasury Yield, 10 Year US Treasury Yield, Defense ETF, Metals and Mining ETF, US HY ETF, US IG ETF, 10 Year US Treasury Yield Futures, Bitcoin Futures, MSCI Futures
Foreign Exchange Markets (9)	USD vs. EUR, JPY, GBP, AUD, MXN, ZAR, RUB, CNY, and EUR-GBP
Crypto Currencies (6)	Bitcoin, ETH, Link, XPR, BNB, ALGO

Table 21: The Financial Assets and Markets I analyzed grouped by Asset Class

Appendix G: Hourly Emerging Topic Modeling Analysis for Case Study 2

Figures 10 and 11 display the second example, while Figures 12 and 13 display the third example for the Hourly Emerging Topic Modelling Analysis.

Figure 10 depicts the topics created and monitored from French tweets in the three-hour window on June 6th, 2023. The lime-green line shows the emergence of another “Iran” topic, this time around the debut of its hypersonic missile[28]. The Reuters article was published at around 18:00 GMT, which I can see the “Iran” topic emerging before this article was published.

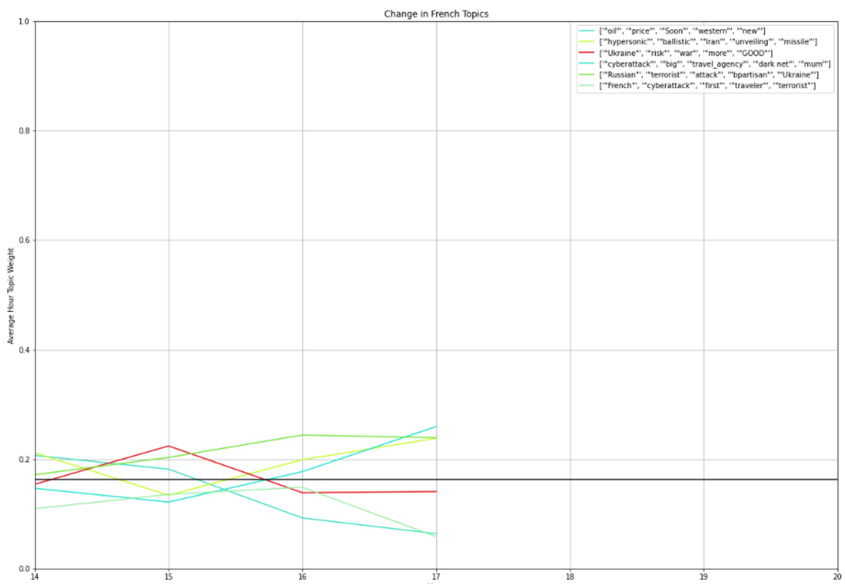


Figure 10: French Topics generated and tracked for June 6th, 2023, from 14:00 to 17:00 GMT

Below, Figure 11 shows the Google Trends for Iran on June 6th, 2023. As the orange trend line shows, interest in Iran as a search term starts increasing at 17:00 GMT and spikes after 18:25 GMT, which is after the topic had emerged in the French tweets.

[28]Hafezi, Parisa, "Iran Presents Its First Hypersonic Ballistic Missile, State Media Reports."

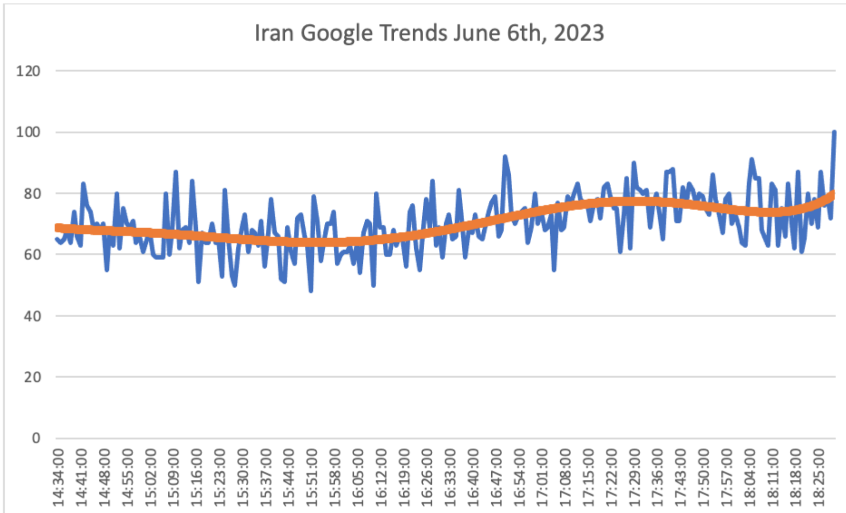


Figure 11: Google Trends tracking of "Iran" from 14:34 – 18:25 on June 1st, 2023

Figure 12 exhibits the topics created from Japanese tweets for June 6th, 2023, from 14:00 to 17:00 GMT. There are no major spikes in the any of the topics, which are generally clustered around the median value for average weights of the topics. This indicates that these topics were generally discussed equally throughout the time period. That said, the green line topic representing "Macron" stood out from the rest around 15:00 GMT. This small spike is in reference to the French protests that were occurring in this period, and an increased police presence for protest scheduled on June 6th, 2023[29].

[29]Reuters. "France Plans Major Police Presence for June 6 Day of Protest."

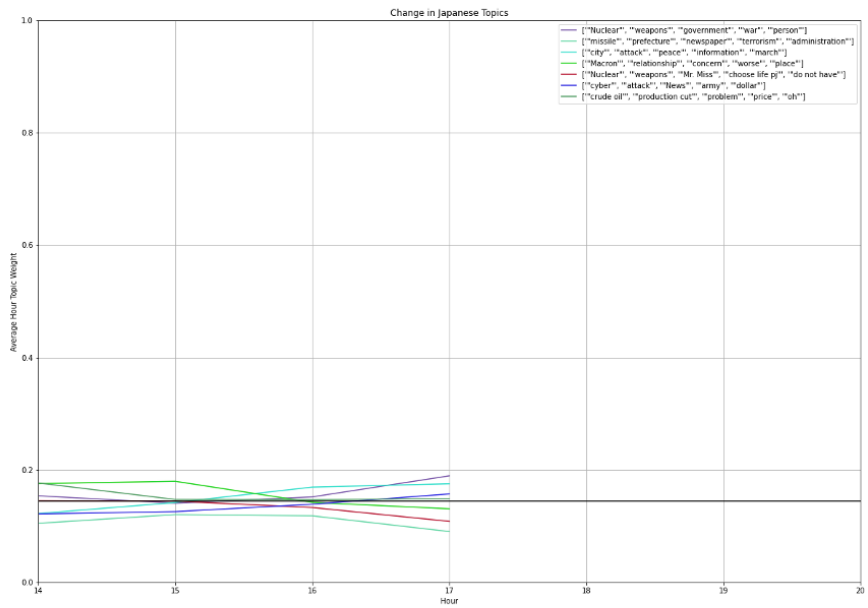


Figure 12: Japanese Topics generated and tracked for June 6th, 2023, from 14:00 to 17:00 GMT

Figure 13 presents the change in Google Trends for "Macron" from 14:36 to 18:31 GMT. As the figure shows, the increase in search activity about Marcon did not peak until around 17:00 GMT, which is nearly two hours after it had peaked on Twitter.

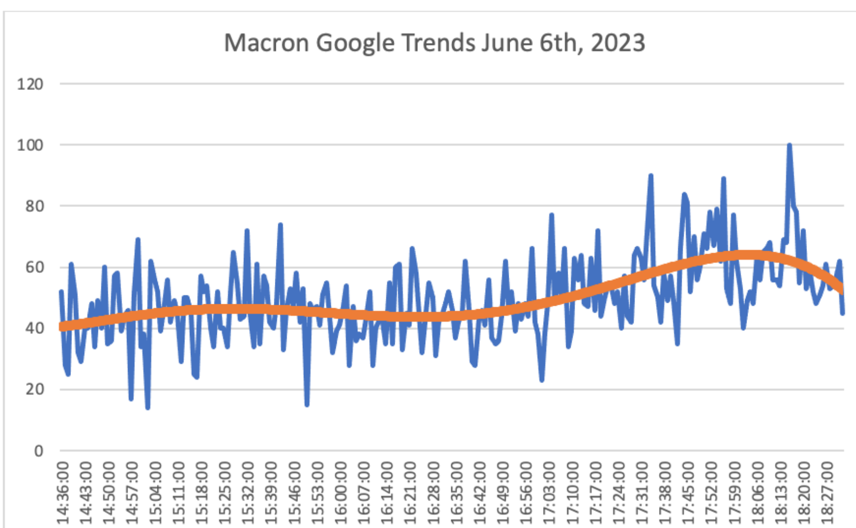


Figure 13: Google Trends tracking of "Macron" from for June 6th, 2023, from 14:36 to 18:27 GMT

Appendix H: Daily Topic Modeling Analyses for Case Study 2

This Appendix contains the other Daily Analysis charts

Below Figures 14 – 19, detail the emergence of "North Korea" as a topic for the other six languages (Spanish, French, Portuguese, Arabic, Japanese, and Korean).

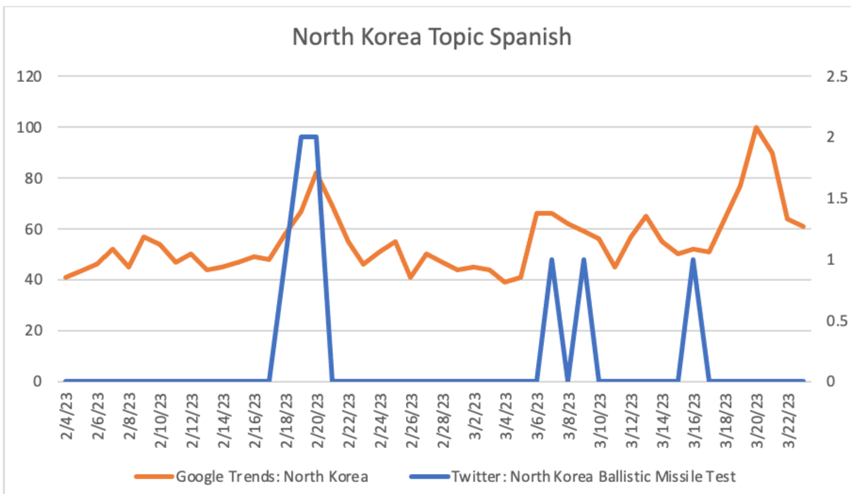


Figure 14: Spanish North Korea topic emergence comparison

In Spanish, I saw a similar trend as I saw for English, with North Korea emerging as a topic around the launch of missiles on Twitter / X, and sometimes before they became trending topic on Google or even picking up on some tests that English did not have such as the March 16th test [148]. However, unlike English and Google, a topic for North Korea did not emerge for the March 19th tests.

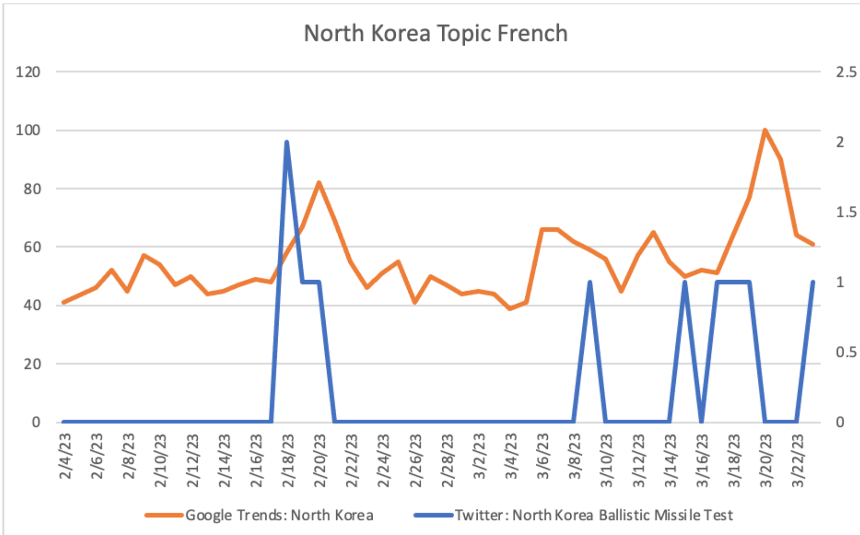


Figure 15: French North Korea topic emergence comparison

French also followed similar trends as English and Spanish, as North Korea also emerges as a topic on Twitter usually a day before it does on Google. In addition, it also picked up on the March 23rd test [148], which the other languages and Google did not.

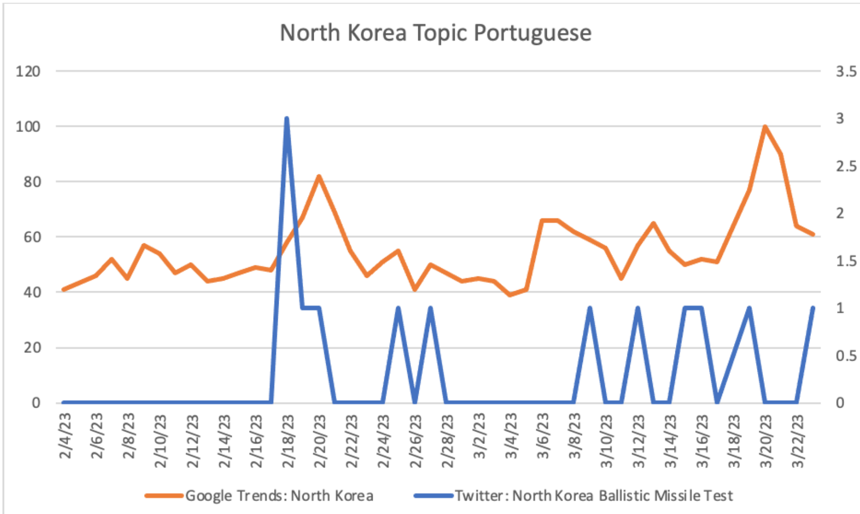


Figure 16: Portuguese North Korea topic emergence comparison

With Portuguese, the Twitter / X emerging geopolitical topic modeling program did very well, capturing seven out of ten missile launches in the research period [148]. In line with the other languages, Portuguese tweets captured these emerging topics roughly a day before they were trending on Google.

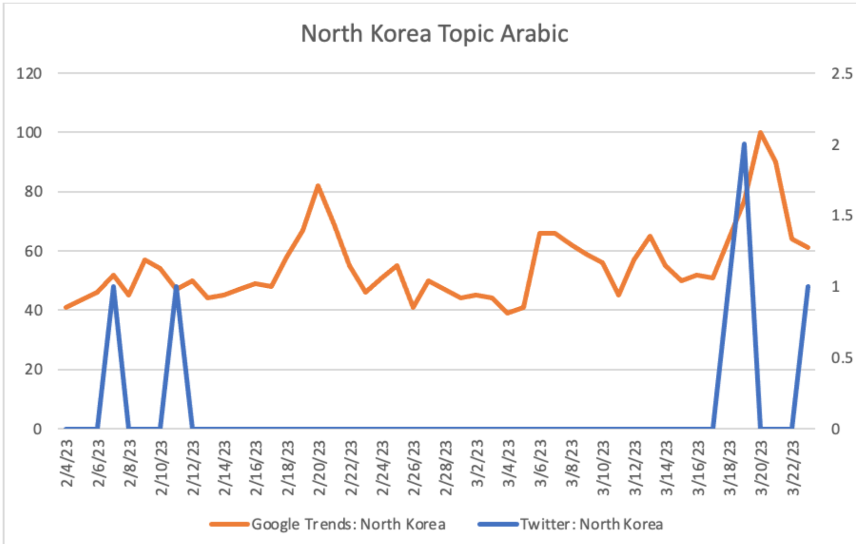


Figure 17: Arabic North Korea topic emergence comparison

However, with Arabic, only two missile tests emerged as a topic through Twitter during the research period. While the ones that did still proceed the Google Trends data, there are significantly less topics created for North Korean than in other languages. I believe this is due to an increased focus in Arabic tweets on regional geopolitical issues and the greater diversity of geopolitical issues discussed as detailed in Appendix H.

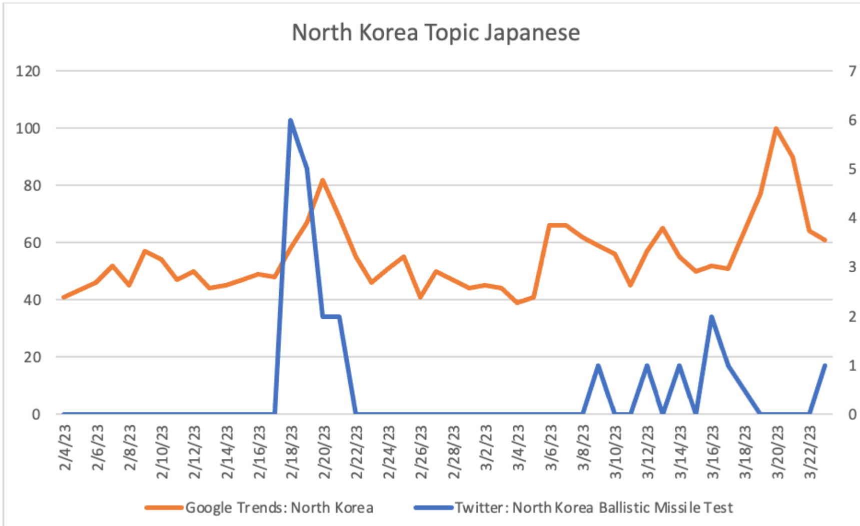


Figure 18: Japanese North Korea topic emergence comparison

Unsurprisingly, geopolitical topics that emerged from Japanese tweets had a large focus on North Korea. As seen by the large spike on the missile launch on February 18th, nearly every topic that was created for that day of Japanese tweets related to North Korea, as the missile landed in the waters close to the Japanese Mainland. Like other languages, North Korea also emerged as a topic in Japanese tweets before Google Trends.

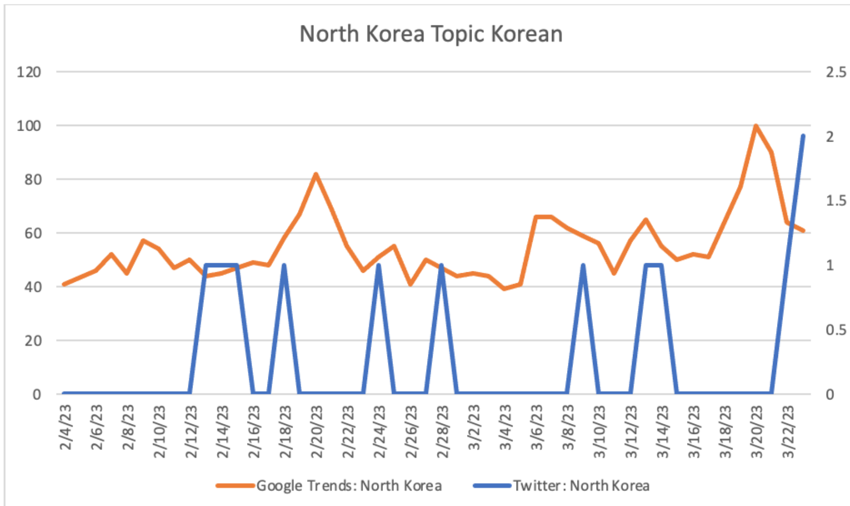


Figure 19: Korean North Korea topic emergence comparison

As explained in the Discussion section, unfortunately, topics generated with Korean tweets were less easily interpretable than in other languages. While I speculated that this could be from machine translation error or the complexities of the Korean language, I can be certain that North Korea does emerge as a geopolitical topic at around the time of missile launches and generally predated the Google trends.

Appendix I: Additional Topic Modeling Analyses Examples for Case Study 2

Daily Topics Examples:

I wanted to also include some further examples of identified topics at the different time scales to illustrate the effectiveness of the Twitter / X Emerging Topic Modeling program. While I will not go into as much detail as the main examples from Chapter 6, Appendix G, and Appendix H, these cases display a wider range of topics that were investigated during the study periods. The first is an example of combining all the count of the topics generated by each language together, unlike the North Korean missile launches example were I kept the languages separate. Like the North Korean missile launches, the Ukraine War was discussed in all seven languages especially since the study period happened to include the first anniversary of the beginning of the invasion.

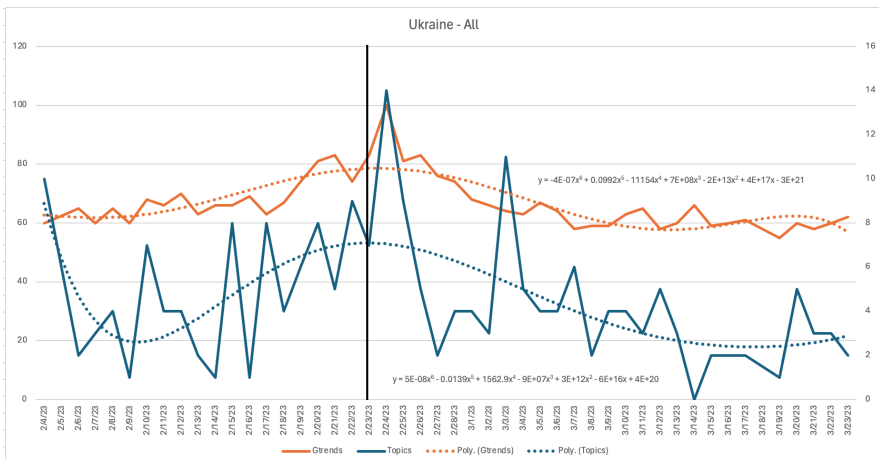


Figure 20: Ukraine War - All Language topic model comparison

While the individual languages had different days and times for when the Ukraine War emerged as a topic, when put together, the overall Twitter / X trend (in blue) matches the Google Trends (in orange). Also the combined language Twitter / X trend line also spikes roughly of March 23rd, 2023 (the black line), which is slightly before the peak of the trend line of Google Trends, indicating that Twitter / X emerging topics trends occur faster than the emergence on Google Trends.

I also wanted to show a few individual daily topics for certain regions. Below are the trends for South China Sea for English. This was the second most generated emerging topic in English, so it warrants study. The South China Sea has been a geopolitical hot spot for most of the last decade as China started to project more of its power over the region. This has created worries for the United States and its allies, especially Taiwan and the Philippines.

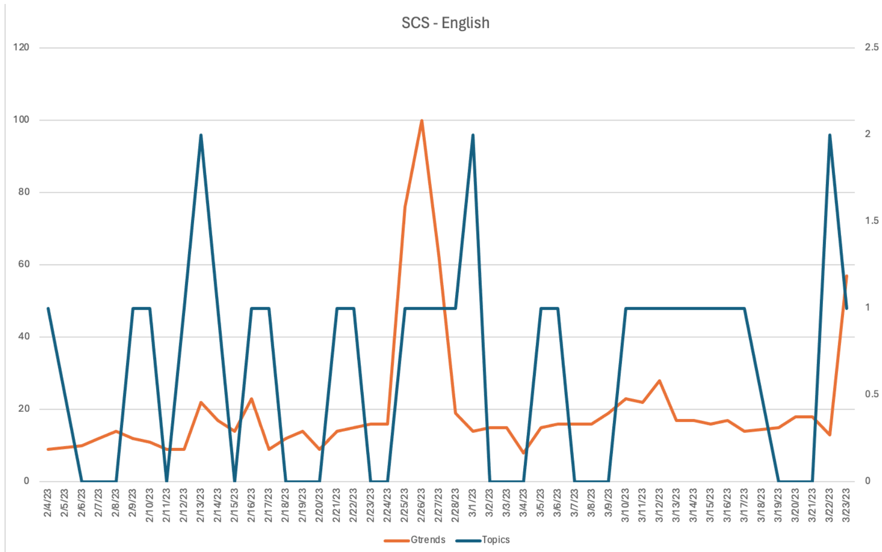


Figure 21: South China Sea - English topic model comparison

As the chart above presents, there were three major spikes in the Twitter / X Emerging topic model trend and two in the Google Trends. These spikes coincide with minor conflicts occurring in the South China Sea. The first spike was on February 13th, 2023, the Philippines accused a Chinese Coast Guard ship of harassing one of their Coast Guard ships[346]. This event created the emerging South China Sea Twitter / X topic for that date and was the major spike without a corresponding spike in the Google Trends data indicating that Twitter / X Emerging Topic programs can pick up more information than Google Trends alone. The second spike relates around February 26th, 2023 seems to also relate to this incident, which was the highest spike for Google Trends, but there was a similar spike in the Twitter / X data. Finally, there was the report of the Chinese naval command warning off a US Navy ship in the South China Sea on March 23rd[45]. While this event caused a spike in both the Twitter / X trends and Google Trends, the Twitter / X spike occurred a day before the Google Trends, further illustrating my previous findings in Case Study 2.

Lastly, another major geopolitical risk that was discussed during this time was the Iranian Nuclear Program. While I discussed this issue in Chapter 5 for the Hourly results which were in June of 2023, however, this is a major geopolitical worry for the Arabic speaking world and emerged as a topic quite frequently in both the Twitter / X trends and Google Trends.

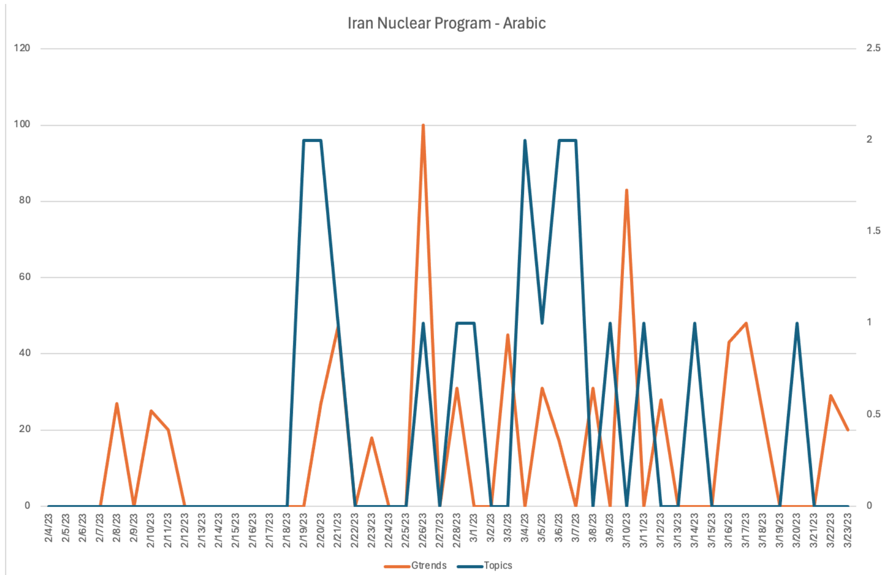


Figure 22: Iran Nuclear Program - Arabic topic model comparison

Similar to the South China Sea, the emerging topics correspond to events surrounding the Iranian Nuclear Program. For example, on February 17th, Israel announced that "all possible means on the table" (Irish, 2023) to stop the Iranian Nuclear Program [244] which caused multiple topics to emerging through the Twitter / X trends program, while a corresponding spike was since on Google Trends a few days later. A contemporaneous spike in both Twitter / X trends and Google Trends on February 26th relating to the CIA director discussing the worries of the growth of Iranian Nuclear program [72]. Lastly the bunch of small spikes around March 8th - March 12th, is related to the US Treasury Department putting sanctions on Iranian's UAV program and its collaboration with China [418].

These additional three examples show the Twitter / X Emerging Topic model program works in discovery emerging topics either quicker or at the same time as Google Trends, and potentially more emerging instances than Google Trends, which makes the Twitter / X program an effective tool in monitoring geopolitical risk.

Hourly Topics Examples:

In addition to the Daily Topic examples, I wanted to include a few further examples of the hourly emerging topics, displaying that on specific topics that the Twitter / X trends spike a few hours before they appear through Google Trends.

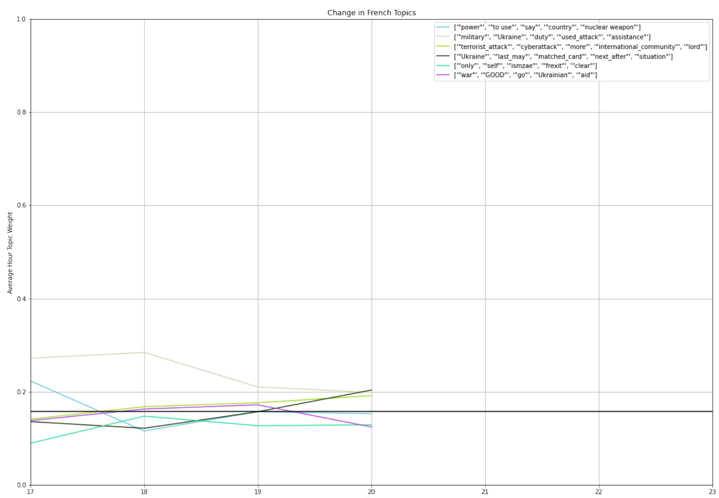


Figure 23: French Twitter / X Emerging Topics for June 1st, 2023

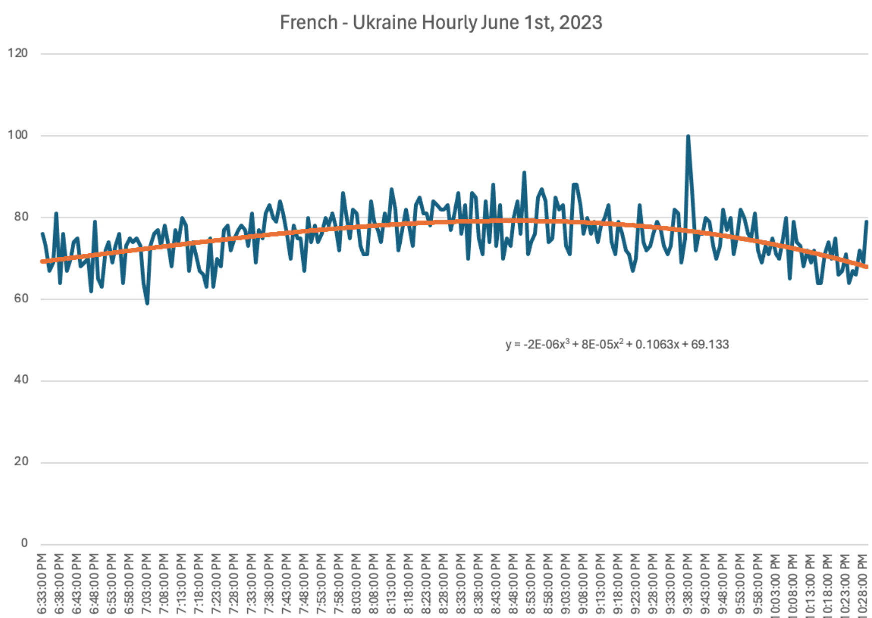


Figure 24: Ukraine - French Google Trends for June 1st, 2023

Like the other examples, Ukraine as a topic in French Twitter / X program emerged and spiked as a topic roughly three hours before the topic spikes on Google Trends, as indicated by the orange trend line in Figure 24, with the most search volume occurring for this period roughly four hours after Ukraine had peaked as an emerging topic on Twitter / X.

Another hourly example is the Portuguese Twitter / X program emerging topics on Russian Nuclear fear.

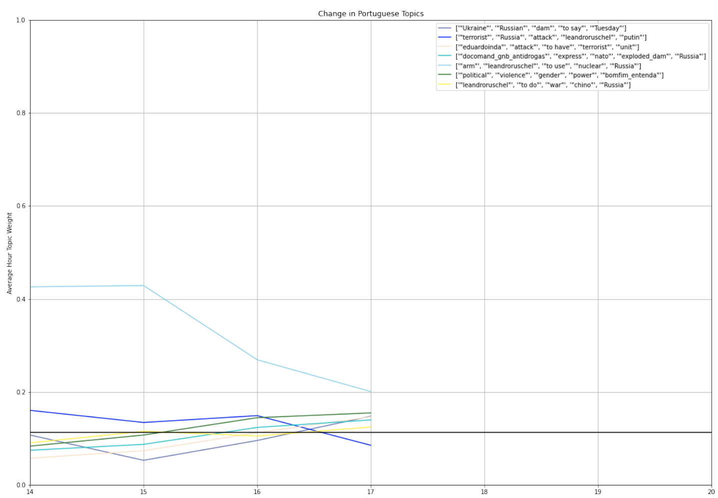


Figure 25: Portuguese Twitter / X Emerging Topics for June 6th, 2023

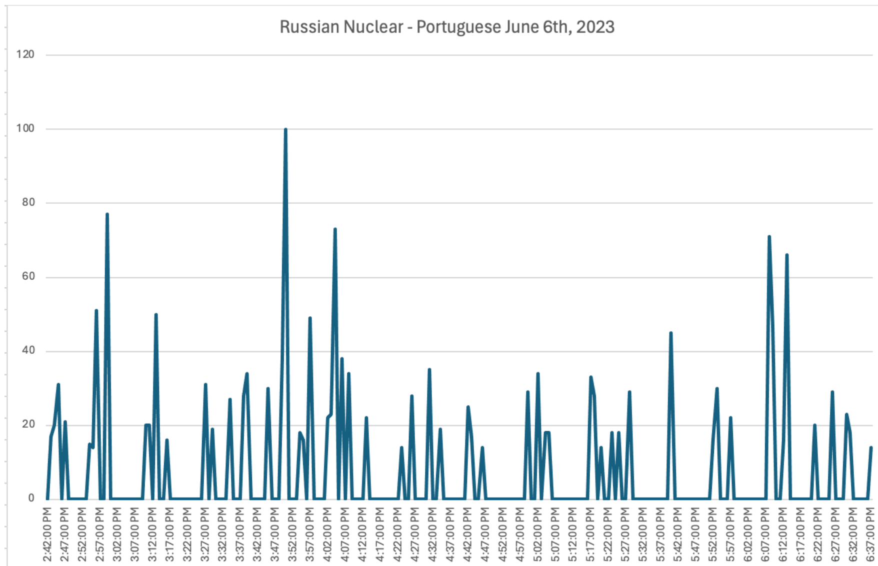


Figure 26: Russian Nuclear - Portuguese Google Trends for June 1st, 2023

The spike in search volume is closer to the spike in the emerging topic of Russia Nuclear on the Portuguese Twitter / X program, however, the topic has seemed to emerged before the collection period which would explain the closer time difference than the average. Nevertheless, the Twitter / X program still spiked before the Google Trends search volume. This topics seems to have been generated by Russia claiming that U.S F-16s could be equipped with nuclear weapons if they are sent to Ukraine [359].

Appendix J: The Topics Generated for Each Language in Case Study 2

The Tables 29 – 35 display the topic labels and counts across all 7 languages for the Daily Analysis.

English	
Topics	Count of Topics
Ukraine War	53
South China Sea	30
China	13
Russian Oil	10
Pakistan	8
North Korea Ballistic Missile Test	6
Iran Nuclear Program	6
Pemex Mexico Oil Fires	4
Nordstream Pipeline	3
Jan 6th Riot	3
East African Oil Pipeline (EACOP)	3
Israel	3
Terrorist Attack on Karachi Police Station	3
China - Russia Relations	3
Israeli Terrorist Attack	3
Russia Nuclear Weapons	3
Biden	2
Chinese Balloon	2
UN's Antonio Guterras	2
Nigerian Election Violence	2
Iran	2
Pulwama Terrorist Attack Anniversary	2
Cyber Attack on US Marshalls	2
Russia Supplying Enriched Uranium to China	2
India - Russia Relations	1
Italy Cyber Attack	1
Gun Violence	1
Covid Vaccine	1
Japan	1
Keystone Pipeline	1
Norway Spots Russian Ship Carrying Nuclear Weapons	1
Wieambilla Shooting in Australia	1
Russia Withdrawing from the START Treaty with the US	1
Germany	1
Uganda in the International Criminal Court	1
Maoist in India	1
US Food Aid for Kenya	1
Atlanta "Cop City" Protests	1
Niger Delta Oil Spill	1
Transnistria	1
US - China Relations	1

Table 29: Count of English Topics that Emerged from February 4th, 2023, to March 23rd, 2023

Spanish	
Topics	Count of Topics
Ukraine War	32
Violence against Women	17
Terrorism	15
Nuclear Weapons	10
Spanish Minister Irene Montero	9
North Korea Ballistic Missile Test	8
United States	4
Peru	3
China	3
Ecuador and the Albanian Mafia	3
Norway Spots Russian Ship Carrying Nuclear Weapons	2
Ecuador's President Lasso	2
Iran Nuclear Program	2
Slovakia Designates Russia a State Sponsor of Terror	2
Israel Terrorist Attack	2
Barcelona Hospital Cyber Attack	2
US targetting Mexican Cartels	2
Argentina	2
El Salvador	2
Israel	2
Mexico Supreme Court	2
Omar Menendez Assassination	1
Chinese Spy Balloon	1
Bolivarian National Guard in Venezuela	1
Colombia and Venezuela Trade Deal	1
Nordstream Pipeline	1
Chile Wildfires	1
Terrorist Attack on Karachi Police Station	1
Afghanistan - US Marines Scandal	1
ELN Rebels Blow Up Bridge in Bogota	1
Mexico Voting Rights Protests	1
Cuba	1
North Korea - US Relations	1
Israel - Palestine Relations	1
Colombia Oil Protests	1
Mexico - Russia Relations	1
Israel - Iran Relations	1
Russia Nuclear Weapons	1
Ecuador	1
South Korea	1
Russia - Syria Relations	1
Former Mexican President Vicente Fox	1
Colombia Military Helicopter Crash	1
Cuba - United States Relations	1
Mexico President	1
Russian Nuclear Submarines Near US	1

Table 30: Count of Spanish Topics that Emerged from February 4th, 2023, to March 23rd, 2023

French	
Topics	Count of Topics
Ukraine War	79
North Korea Ballistic Missile Test	9
Cyber Attacks	8
Israel	8
China	7
Norway Spots Russian Ship Carrying Nuclear Weapons	6
US - China Relations	6
Pemex Mexico Oil Fires	5
Iran Nuclear Program	5
American Military	4
Italy Cyber Attack	3
Turkey Earthquake	3
South China Sea	3
Russia Withdrawing from the START Treaty with the US	3
Macron	3
Saudi Arabia	3
Russia Nuclear Weapons	3
Chinese Spy Balloon	2
Macron"s Speech in Munich on Ukraine War	2
Russia's Failed Test of ICBM	2
COVID	2
Russian Oil	2
Iran - Israel Relations	1
Moldova	1
American Nuclear Weapons	1
US Navy	1
Russia - China Relations	1
Chinese Military	1
Slovakia Designates Russia a State Sponsor of Terror	1
Nord Stream Pipeline	1
Iraq	1
Algeria - Morocco Relations	1
Potential Attack through Belarus	1
Palestine	1
Hamburg Terrorist Attack	1
Russia - United States Relations	1
Australia - France Relations	1
Canada	1
US in West Africa	1
Czech Republic	1
Burkina Faso	1
United States	1

Table 31: Count of French Topics that Emerged from February 4$^{\text{th}}$, 2023, to March 23$^{\text{rd}}$, 2023

Portuguese	
Topics	Count of Topics
Ukraine War	25
Brazil Oil Export Tax	18
Nuclear Attack	15
North Korea Ballistic Missile Test	13
Terrorism	10
Russia Withdrawing from the START Treaty with the US	9
Brazilian Congress Member admits to employing a hacker	6
Bolsonaro Jewelry Scandal	6
Russia - Brazil Relations	4
China - US Relations	4
China - Russia Relations	4
Israel	4
Cyber Attacks	4
Neo Nazi Attack at Sao Paolo School	3
Norway Spots Russian Ship Carrying Nuclear Weapons	3
Bolsonaro	3
Russia	3
Venezuela	2
China	2
Peru Protests	2
Ecuador's President Lasso	2
Pemex Mexico Oil Fires	2
Former Mexican President Vicente Fox	2
Korean Oil Trade	1
Turkey Earthquake	1
Jan 8th Brazilan Riots	1
Spanish Minister Irene Montero	1
Germany	1
Mozambique	1
Talca, Chile Earthquake	1
Slovakia Designates Russia a State Sponsor of Terror	1
Brazil Nuclear Weapons	1
United States	1
Lula Prosection Scandal	1
Sandra Cuevas Suspension from Office in Cuahtemoc	1
Japan	1
Taiwan - China Relations	1
Brazil Coffee Exports	1
Mexico's Claudia Sheinbaum	1
Korea	1
China - Brazil Relations	1
Chile	1

Table 32: Count of Portuguese Topics that Emerged from February 4th, 2023, to March 23rd, 2023

Arabic	
Topics	**Count of Topics**
United States	19
Iran Nuclear Program	19
Ukraine War	13
Yemen Civil War	12
Russia	12
Iraq	10
South China Sea	8
China - US Relations	6
North Korea	5
China's Saudi - Iran Deal	5
China	5
Israel	5
OPEC	4
Saudi Arabia	4
Russia Withdrawing from the START Treaty with the US	4
Turkey Earthquake	3
Syria	3
Algeria	3
Turkey	2
Saudi - US Relations	2
UK Army	2
UAE PM Maktoum	2
Iraqi Counter - Terrorism Raid	2
Tunisia Opposition Crack Down	2
UAE	2
Russian Nuclear Submarines Near US	2
Taiwan	1
Libya	1
UAE's Lana Zaki Nusseibeh appoitment to Chair of UN Security Council's Counter Terrorism Committee	1
Palestine	1
China - Turkey Relations	1
China's Laser Ship in South China Sea	1
Russia - US Relations	1
Norway Spots Russian Ship Carrying Nuclear Weapons	1
Chinese Spy Balloons	1
ISIS	1
UK - Russia Relations	1
Pemex Mexico Oil Fires	1
Ecuador Bridge and Oil Pipeline Collapse	1
Egypt	1
China - Kuwait Relations	1
China - Russia Relations	1
China - Saudi Relations	1
Russia - Iran Relations	1
China - Iran Relations	1
United States - Iran Relations	1
Israel - United States Relations	1
Israel - Qatar Relations	1
Israel - Iran Relations	1
Saudi - Kuwait Relations	1
Kuwait	1
US - Pakistan Relations	1

Table 33: Count of Arabic Topics that Emerged from February 4th, 2023, to March 23rd, 2023

Japanese	
Topics	Count of Topics
Nuclear Weapons	103
Crude Oil	54
North Korea Ballistic Missile Test	22
Self - Defense Force	18
Nord Stream Pipeline	3
Hyper Sonic Missiles	3
Japan - US Relations	2
PM's Cabinet Meeting	2
Ukraine War	2
China Balloon	1
Imperial Family	1
Natural Gas	1

Table 34: Count of Japanese Topics that Emerged from February 4th, 2023, to March 23rd, 2023

Korean	
Topics	Count of Topics
Crude Oil	128
North Korea	12
China	6
United States	5
Nuclear Weapons	3
US Fed	3
Arrest Warrant for Oppostion Leader Lee Jae-myung	3
South Korean PMI Measure	3
Ukraine War	2
Japan and South Korea	2
Cyber Attack	1
Turkey	1
North Korean Ballistic Missiles	1
Coal	1
North Korea Blaming US for Ukraine War	1
China's Baltic Sea Reinforcements	1
Iranian Oil	1
Japan	1
Iran	1

Table 35: Count of Korean Topics that Emerged from February 4th, 2023, to March 23rd, 2023

Appendix K: The Financial Markets and Assets Analyzed for Case Study 3

These are when the markets assets I evaluated in this case study were active. The Two-Year Treasure Yield data came from CNBC [129], the Gold Spot Price came from Goldhub.com [200], the Crude Oil Spot Price came from LiveCharts.co.uk [291]. All other financial data came from Yahoo Finance[30] [174] through the Python package yfinance developed by Aroussi [58]. Table 44 below shows when each market was active. Note: only Cryptocurrencies were actively throughout the entire time frame including weekends, all other markets had periods where they were closed.

Time (US Eastern Standard Time)	Market Active
12 am – 6am	Commodities, Forex, FTSE 100, CSI 300, Nikkei 225, Indian Sensex, FTSE 100, Crypto
6 am – 12pm	Commodities, US Markets, Forex, FTSE 100, Crypto
12pm – 6pm	Commodities, US Markets, Forex, Crypto
6 pm – 12am	Commodities, Forex, CSI 300, Nikkei 225, Indian Sensex, Crypto

Table 44: When each market in the study was active

[30] https://finance.yahoo.com/

Appendix L: World Map Figures

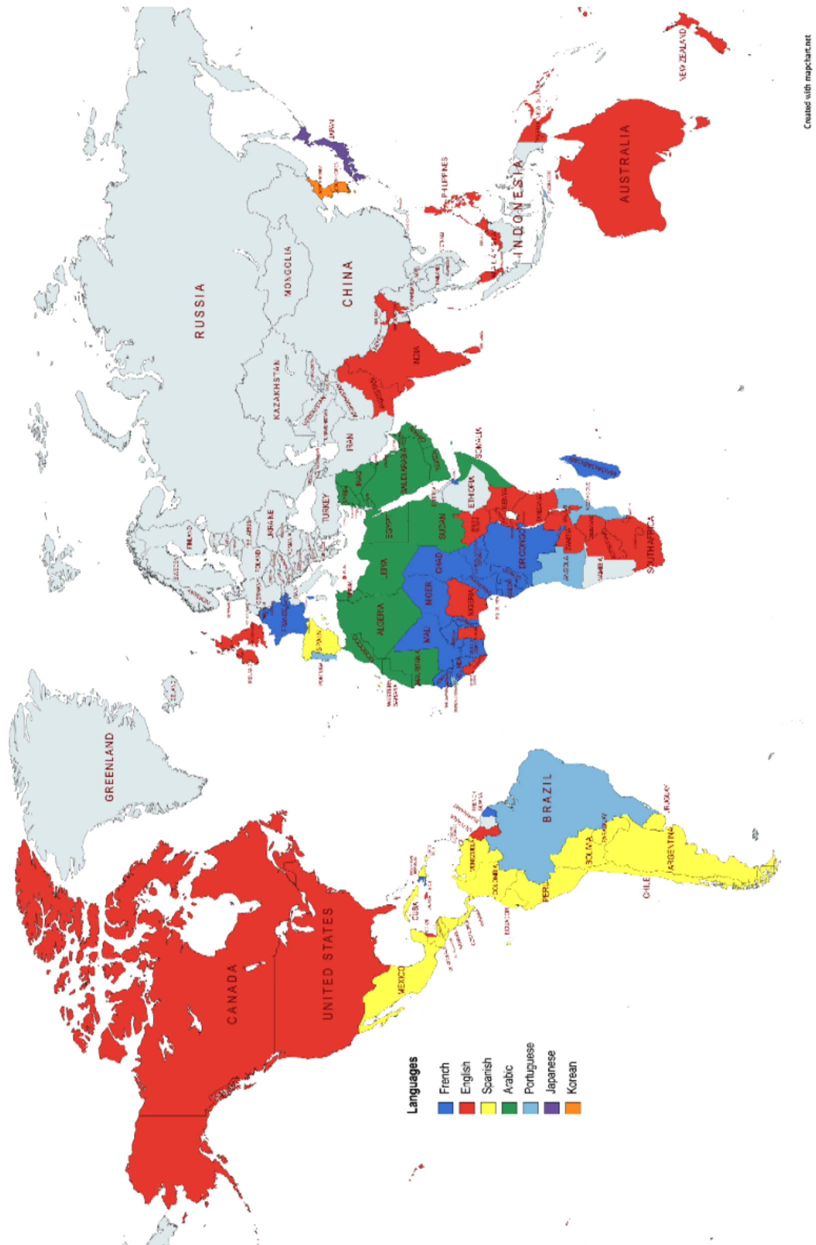


Figure 46: Map of countries that my programs can get tweets from in one of the national languages used by that country.

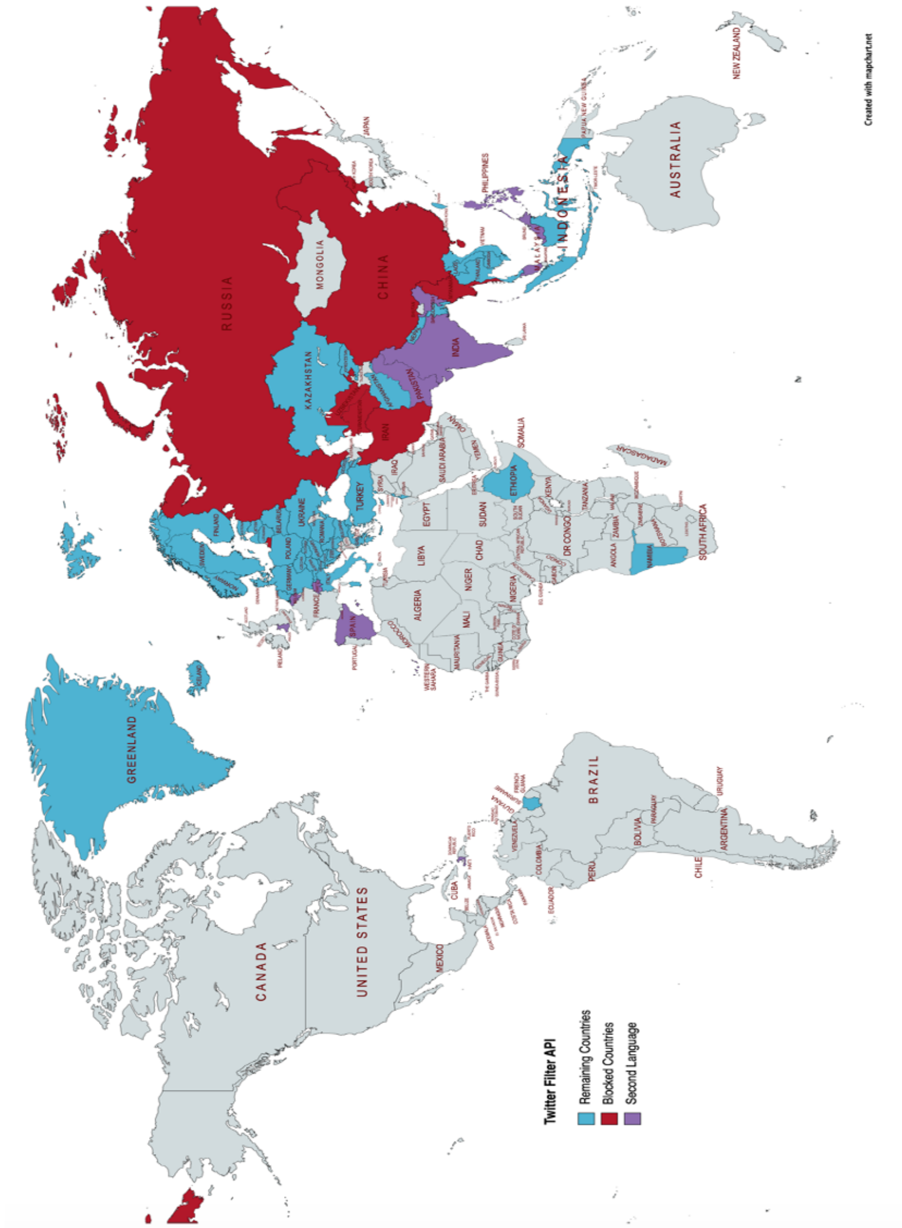


Figure 47: Map of countries: A) Blue: The country's language is available in Twitter / X Filter API selection but not used in my programs. B) Red: where Twitter / X is blocked. C) Purple: where a second language is available in the Twitter / X Filter API, but the country is already covered by my program. Note: A country that is gray on both Figures 46 and 47, for example, Mongolia, means that country allows Twitter / X, but doesn't officially speak any language covered by the Filter API

Appendix M: Ethical and Legal Considerations

This appendix covers the ethical and legal considerations I undertook in conducting my research. As this research revolves around social media, I have implemented safe guards to protect the identities of the posters whose data I have collected which I detail below.

Participants:

The participants were any Twitter / X user that post publicly using the keywords I've defined for the research, and I will not be interacting with them. As the University of St Andrews Social Media Research page states, it is impractical to obtain informed consent for this research. As this is passive social media research, the research will be conducted completely online.

Data Categories Collected:

As defined in the School of Computer Science Ethics Application Form, personal data is information related to a person which can be directly or indirectly connected back to them. Special category data is personal data relating to a person's race, ethnicity, politics, religion, etc. My research does not collect special category data, but it collects the bare minimum in personal data. Later on, I will describe how I handle this personal data.

Ethical Considerations:

According to the University of St Andrews Social Media Research Ethics Concerns page, the three Ethical Considerations for passive research are: 1. Ensuring the rights of the social media users, 2. What Social Media users might think about repurposing their tweets, 3. Protecting the identities of the social media users. While the data I collect contains the tweet text and the tweet ID initially, I will be anonymizing the data by removing these identifying variables. I will also only be analyzing public tweets, which are public posts online, not private messages. The users know that these public tweets can be seen by anyone, and according to Reuter et al. 2019, most Twitter / X users don't find monitoring Twitter / X an "inappropriate surveillance or a violation of privacy (Pg. 2) and thus my research will not violate their privacy, satisfying Considerations 1, 2, and 3.

Citation: Reuter, Katja; Zhu, Yifan; Angyan, Praveen; Le, NamQuyen; Merchant, Akil A.; and Zimmer, Michael, "Public Concern About Monitoring Twitter Users and Their Conversations to Recruit for Clinical Trials: Survey Study" (2019). *Computer Science Faculty Research and Publications.* 29. https://epublications.marquette.edu/comp_fac/29

Data Lifecycle: Collection

This section describes how I will collect the data required for my research, how I make the initial data I collect pseudonymized (i.e. meaning only indirectly linked to a person), and finally how that pseudonymized data becomes full anonymized.

Initially, I will be collecting Twitter / X data including the tweet's text, the time it was posted, and the tweet's internal Twitter ID. However, I have modified my code so that the tweet ID is removed and replaced with a generic "ID" tag from the incoming tweets json file. Also, in accordance with the "Passive" section of "Types of social media research: active vs passive", I have pseudonymized the tweet text by replacing the text with another generic "textpseudo" variable. These two variables must be pseudonymized as it is possible to use them to trace the tweet back to the original poster, thus revealing the user and violating the ethical standards of the University of St Andrews.

This new data will only have the time the tweet was created, a label for the Geopolitical Topic the tweet relates to, and the two new pseudonymized variables for the ID and Text. This pseudonymization process will take place on the University's One Drive. These new data files will also be the data that I will store and disseminate, while the original data files will be deleted once the analysis is complete or any related paper that requires the raw data to be sent to the publisher is complete. With the deletion of the original files, there will be no way to track the pseudonymized data back to the original poster, making the new datasets fully anonymous.

Before deletion, I will be transferring and storing these data files on the University's One Drive to safely secure the pseudonymized data.

Data Lifecycle: Sharing and Publication

Many journals require the raw data as well as the manuscripts for paper publications. However, I will not publish any of the raw data and only disseminate the data that had been pseudonymized, and without the original data files, the pseudonymized data files become fully anonymized. Any publication that requires a deviation from this plan will first be discussed with my supervisors, and the Ethics Committee before going forward. This will be in line with the Ethical Considerations described on the University of St Andrews Social Media Research page.

Data Lifecycle: Retention and Destruction:

As described in the University Guidance, the data will be retained by me on the University's One Drive to secure it in the pseudonymized format, at least until I complete my PhD. If I stay in academia, I will retain the data for the standard 10 years and its deletion will be subject to periodic review every year. If I leave academia, I will transfer the data to my supervisor, Dr. Tom Kelsey.

Data Access Statement:

All code and data related to my thesis that can be made publicly available within the guidelines of the University of St Andrews School of Computer Science Research Ethics Board can be found on GitHub, through this URL: https://github.com/jb370/Automatic-GR/tree/main

In the Github page, only the anonymous Twitter / X data is stored, however, access to the full data will require the approval of the Research Ethics Board. Please contact me at jb370@st-andrews.ac.uk or burnsjack45@gmail.com if you would like copies of the full data subject to approval.

Additional Questions from the Ethics Committee:

Question 1: "Is the use of social media the most appropriate method?"

Answer 1: Yes, social media and the data it generates are the source for my research, there is no alternative.

Question 2: "Can you obtain informed consent? If not, are you safeguarding people's rights? Are you inferring or using special category data?"

Answer 2: No, I am not obtaining informed consent as this is passive research and as stated under the "Passive" subsection of "Types of social media research: active versus passive" that "if there is no interaction with participants, then it may not be practical to obtain informed consent". However, I am safeguarding people's rights, addressed in the Data Lifecycle: Collection subsection, and any issues with special category data are addressed below.

Question 3: "How will you protect the identity of your participants during data collection, analysis and dissemination?"

Answer 3: Addressed in the Data Lifecycle: Collection subsection. In addition, I do not collect any Twitter / X geotagged data (i.e., location data that the user has authorized to attach to their tweets) further protecting their identity.

Question 4: "Are you making assumptions about whether people would be willing to let you use their data?"

Answer 4: After reviewing Reuter et al. 2019, who find that most Twitter users don't find monitoring Twitter / X "inappropriate surveillance or a violation of privacy" (Reuter et al. 2019, Pg. 2), thus I believe that users would be willing to use their data for my research. The Reuter paper is regarding clinical trial recruitment, while my research is focused on the user's public opinion expressed in tweets. Thus my research is less invasive than monitoring the users to recruit them for health studies, which the majority of users expressed they were already okay with.

Question 5: “Are you following any subject-specific codes or guidance?”

Answer 5: In addition to the University of St Andrews Ethical Guidance, I also reviewed the Internet Research: Ethical Guidelines 3.0 by the Association of Internet Researchers (AoIR 2019) to make sure that I am following an overarching ethics guide, not just my own opinions. For example, I made sure that the existing datasets that I use for the machine learning aspects of my project address all the questions with the Existing Datasets section (AoIR 2019, Pg. 39), likewise for the new data that is collected in the New Data Collection section.

Citation: franzke, aline shakti, Bechmann, Anja, Zimmer, Michael, Ess, Charles and the Association of Internet Researchers (2020). *Internet Research: Ethical Guidelines 3.0.* https://aoir.org/reports/ethics3.pdf

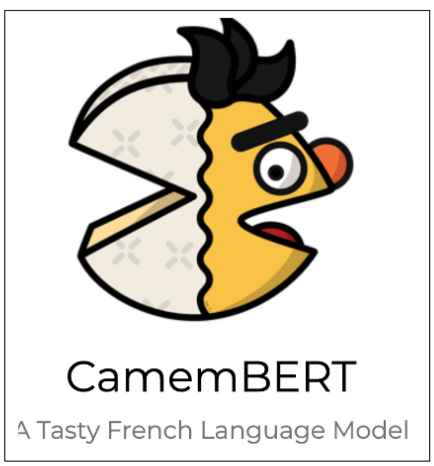
Issues with Special Categories:
Unfortunately, there is no way for me to know if the data is coming from people in special categories. The data coming from the Twitter / X API has only the createdat variable (i.e., when the tweet was published), the tweetid variable (i.e., the ID of the tweet) and the text variable (i.e., the text of the tweet). I have addressed the issue of the tweetid and tweet text in Data Lifecycle: Collection subsection, which will remove any possible identifying characteristics from the data. This will ensure that nobody will be able to identify any user, including those of special categories.

Additional Legal Questions:

Data Usage Agreements:
The Twitter / X data was collected through the Twitter / X Filter API and the Twarc package both of which used the Twitter Academic Track v2 which I obtained from Twitter / X. All data was collected before the changes to the Twitter / X API came into affect which removed the free API system in 2023. There was no illegal data scraping of Twitter / X conducted for this research.

Appendix N: Collection of BERTs

Many researchers into the BERT algorithm [155] have created version of BERT for different languages. Some of these researchers have taken the Muppet character, Bert, from Sesame Street as their mascots. I deem it necessary, nay, vital that I include these mascots here:



References:

Top Left: AlBERTo

marcopoli. "Alberto the First Italian Bert Model for Twitter Languange Understanding." Github https://github.com/marcopoli/AlBERTo-it. Accessed August 17th 2023.

Top Right: BERTimbau

Souza, Fabio; Rodrigo Nogueira; Roberto Lotufo. "Bertimbau: Pretrained Bert Models for Brazilian Portuguese." *9th Brazilian Conference on Intelligent Systems, BRACIS*, October 20-23, 2020. https://huggingface.co/neuralmind/bert-base-portuguese-cased.

Bottom Left: CamemBERT

Martin, Louis; Benjamin Muller; Ortiz Suarez; Pedro Javier; Yoann Dupont; Laurent Romary; Eric Villemonte de la Clergerie; Djame Seddah; Benoit Sagot;. "Camembert: A Tasty French Language Model." *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 2020. https://camembert-model.fr/.

Bottom Right: BERTurk

Schweter, Stefan. "Berturk - Bert Models for Turkish." Zenodo https://github.com/stefan-it/turkish-bert.