Weather-Driven Sentiment Analysis using Machine Learning

Luke Williams, Nazmul Hossain, Shreyas Srinivas
lwilliams353@gatech.edu, nhossain32@gatech.edu, ssrinivas6@gatech.edu

Abstract—This paper outlines an approach to sentiment analysis of weather related discussions on Twitter, emphasizing meteorological terminology such as the North Atlantic Oscillation (NAO), Polar Vortex, and other common weather related terms. By analyzing tweets from expert meteorologists, the study aims to quantify sentiment and explore its potential applications in areas like energy demand forecasting and energy pricing. The analysis employs the lexicon based sentiment analysis tool, with a custom mapping of weather related terms to sentiment scores, to evaluate the sentiment of tweets on significant weather phenomena. The findings demonstrate the feasibility of sentiment driven insights into public perceptions of weather. This analysis provides investment firms with an opportunity to leverage meteorological sentiment data to make informed trading strategies, particularly in energy and commodities markets, by anticipating price movements influenced by weather events.

1 INTRODUCTION

Weather plays a crucial role in shaping electricity and gas markets, influencing both supply and demand. On the supply side, renewable energy production, such as wind and solar, is highly dependent on weather conditions, while on the demand side, extreme temperatures drive the need for heating and cooling. Consequently, accurate weather forecasts and insights can significantly impact market dynamics. Therefore understanding public sentiment, particularly that of meteorologists, regarding weather outcomes like teleconnections, hurricanes, and seasonal changes could offer valuable predictive insights for market trends.

Twitter has emerged as a rich platform for disseminating information, with meteorologists frequently sharing their perspectives on a wide range of weather related topics. From forecasts of major weather events such as hurricanes and tropical storms to discussions on longer-term seasonal outlooks, meteorologists' tweets provide timely and relevant insights. Given the substantial influence that weather has on energy markets, sentiment expressed in these tweets has the potential to predict market movements.

The central question this project seeks to address is whether statistical and sentiment analysis techniques can be applied to meteorologists' tweets to extract sentiment on specific weather outcomes. If successful, such a framework could serve as an important tool for energy traders, utility companies, and analysts by providing early warnings of market-relevant weather trends. Therefore, the objective of this project is to develop a robust and adaptable sentiment analysis framework that can analyze meteorological tweets across a variety of weather related topics. By harnessing the power of social media data, this framework aims to enhance decision making processes in markets highly sensitive to weather conditions.

2 DATA COLLECTION AND PREPROCESSING

The dataset for this project was provided in the form of a one-time manual CSV file containing tweets from various meteorologists collected over a span of several years. The data, provided by project stakeholder, was stored in a tabular format with relevant metadata accompanying the tweet text. After a thorough manual inspection of the dataset and preliminary analysis, a series of data cleaning steps were used to prepare the data for downstream tasks such as topic modeling and sentiment analysis. Given the complexity of Twitter data and the nuances of natural language processing (NLP), the cleaning process was iterative, requiring several attempts to refine the data before arriving at the final version.

2.1 Data Cleaning Process

To address the specific challenges posed by Twitter text data, the following steps were implemented:

1. **Removing Twitter-Specific Elements**: Since the focus was on analyzing meteorological sentiment, Twitter-specific items like URLs and mentions were deemed irrelevant for the analysis. Thus, all references to URLs (e.g., links starting with https://t.co) and mentions (e.g., @username) were removed to focus solely on the tweet content.
2. **Hashtag Expansion**: Hashtags can often convey critical meaning, especially in tweets related to specific weather events. To retain this information, hashtags were expanded into their constituent words. For example, #goodmorning was

expanded to "good morning," ensuring that key phrases were preserved and properly interpreted in both the topic modeling and sentiment analysis stages

3. **General Text Cleaning**: Several text normalization steps were applied to standardize the tweet data:
   - **Alphanumeric Removal**: All alphanumerical characters were removed.
   - **Punctuation and White space Removal**: All punctuation marks and unnecessary white spaces were stripped from the text to streamline the tokenization process.
   - **Lowercase Conversion**: The text was converted to lowercase to ensure uniformity during analysis, avoiding issues with case sensitivity.
4. **Contraction Expansion**: To maintain context in sentiment analysis, contractions (e.g., "isn't," "won't") were expanded into their full forms (e.g., "is not," "will not"). This ensured that the negations and other nuances were preserved, which is essential for accurate sentiment extraction. Although some models can handle contractions, the uncertainty regarding the final model choice required this preprocessing step to guarantee better consistency in the data.
5. **Customized Stop Words List**: A customized stop words list was employed to optimize the data for sentiment analysis. Unlike traditional stop word lists, this custom set ensured that important negations (e.g., "not," "never"), modifiers (e.g., "very," "barely"), verbs (e.g., "could," "should"), and conditionals (e.g., "but," "if") were retained. These words are crucial in capturing the context of sentiment, and their removal could distort the analysis. Other irrelevant words, however, were removed to declutter the vocabulary.
6. **Lemmatization**: To simplify the data for more accurate topic modeling, lemmatization was applied to reduce words to their base or dictionary form. For example, words like "running" and "ran" were reduced to "run." This process helped consolidate different word forms and improved the efficiency and accuracy of topic modeling.
7. **Removal of Temporal References**: For topic modeling, certain words related to time (e.g., days of the week, months, and general temporal indicators like "today" and "tomorrow") were removed. These terms were considered non-contributory to the identification of meaningful topics and could introduce noise into the model. Their removal ensured that the focus remained on core meteorological themes rather than irrelevant time-based references.
8. **Handling Empty or Blank Tweets**: During the cleaning process, tweets that became empty or blank after all preprocessing steps (e.g., after removing URLs,

mentions, hashtags, and irrelevant characters) were removed from the dataset. This step ensured that only tweets with meaningful content were retained for the analysis, preventing noise from interfering with both the topic modeling and sentiment analysis results.

The data cleaning process was iterative, involving multiple rounds of preprocessing, followed by evaluation. Each step was refined based on feedback from preliminary analyses, ensuring that the final version of the data was as clean and standardized as possible. This iterative approach allowed for flexibility in adapting the cleaning techniques to best fit the needs of both the topic modeling and sentiment analysis tasks, ultimately ensuring quality input for downstream processes.

By employing these data cleaning strategies, the dataset was transformed into a more structured and meaningful form, suitable for both extracting relevant topics and analyzing the sentiment within meteorologists' tweets.

3 DATA ANALYSIS

The analysis conducted to achieve the project's objectives was divided into three main sections. First, identifying prevalent topics of interest was essential, such as the impact of storms in the United States or the onset of winter conditions in Europe. To accomplish this, topic modeling was utilized to uncover the various themes present in each tweet, allowing the analysis to focus on specific weather events. Second, understanding the geographic location of these weather events was crucial for deriving actionable insights applicable to the relevant markets where such events were traded. Third, to inform the type of action to be taken, sentiment evaluation was performed using a lexicon-based approach, which helped distinguish between warm and cold conditions effectively. The following sections detail the development of each methodology and the subsequent results.

3.1 Topic Modeling

To effectively align tweets with relevant topics, multiple modeling approaches were explored, beginning with an unsupervised topic model, Latent Dirichlet Allocation (LDA). Although LDA provided initial insights, its results were mixed, as it struggled to capture nuanced weather related themes due to variability in tweet language. Recognizing these limitations, the decision was made to shift to a semi-supervised approach using the Correlation Explanation (CorEx) model, which allowed for guidance using key anchor words specific to weather topics.

The CorEx model's anchor-based tuning offered several advantages. Setting anchors allowed the model to focus on meaningful weather themes which ensured outputs align closely with the project's specific scope. An iterative process was employed to continually refine the anchor words, achieving an optimal balance between topic coherence and coverage. Additionally, incorporating bi-grams (e.g. "polar vortex", "negative NAO") improved topic accuracy, helping tweets align naturally with the intended categories. The table below lists the anchor words used in the final model, providing insight into the specific terms that guided the model's alignment of tweets to topics.

Table 1—anchor words used in training the Topic model.

Topic	**Anchor Words**
North Atlantic Oscillation (NAO)	nao,north atlantic,positive nao,negative nao
Arctic Oscillation (AO)	ao,arctic oscillation,positive ao,negative ao
Madden-Julian Oscillation (MJO)	mjo,oscillation
Atmospheric Angular Momentum (AAM)	aam,angular,angular momentum
Polar Vortex	polar vortex,pv
Stratospheric Phenomena (SSW)	ssw,stratospheric warm
Blocking Patterns	block,scandiblock,beast east
Storm Systems	hurricane,cyclone,storm,tornado
Ocean-Atmosphere Interactions	sst,ssts,sea surface,sea temperature,niño,niña
Rainfall	rain,flood,drizzle,downpour,thunder,lightning
Snow	snow,blizzard,snowfall,frost,sleet
Air Pressure	aam,angular,angular momentum,air,pressure,high pressure,low pressure
General	summer,winter,spring,fall,autumn,weather,climate,temp,forecast,pattern,model

This combination of iterative data cleaning, semi-supervised learning, and anchor refinement helped build a model that captures the thematic structure of weather related tweets with greater precision. Ultimately, CorEx provided the flexibility to extract key themes, aligning each tweet with its most relevant topic.

3.2 Named Entity Recognition (NER)

NER is a technique within Natural Language Processing that identifies and categorizes various entities, such as people, organizations, and locations. The goal of the NER modeling was to extract locations mentioned in the tweets to aggregate weather sentiment for specific areas. The 'spacy' package was utilized to explore how Named Entity Recognition (NER) could be leveraged for identifying references to locations, with a primary focus on extracting Geopolitical Entities (GPE). GPEs included countries, cities, states, or other regions with defined boundaries or governance. Initially, the exploration extended to non-GPE items, such as mountain ranges and bodies of water, but these did not produce results that could be easily normalized. The GPE data was refined to include only countries which were then cross-referenced with an ISO database containing unique country identifiers (both full names and 2-digit codes). This hierarchical method ensured that the results were interpretable and actionable, while also simplifying the process of handling raw location data from the tweets.

3.3 Sentiment Analysis

For sentiment analysis, the technique employed was VADER analysis. VADER is an effective technique for analyzing social media text, due to its lexicon of words with predefined sentiment scores and its ability to handle stylistic elements like capitalization and punctuation. Preprocessing steps, such as removing timestamps and irrelevant text, ensured that the input data was cleaned and ready for analysis. This preparation allowed the model to effectively capture the nuanced sentiment of short informal text.

The choice of a semi-supervised methods was made because semi-supervised methods leverage predefined rules and lexicons, eliminating the need for labeled training data while still providing domain-specific insights. This makes them ideal for tasks involving dynamic datasets, such as tweets. In contrast, unsupervised approaches (e.g., clustering) may not sufficiently capture sentiment nuances, while fully supervised models require large labeled datasets and significant computational resources. By employing a semi-supervised method, the analysis achieved a balance between generalizability and efficiency.

To enhance the accuracy and domain specificity of the sentiment analysis, adjustments were made to the sentiment scoring mechanism of the VADER model. These modifications focused on incorporating weather related terminology, which is essential for accurately capturing the sentiment expressed in weather related tweets.

Domain-Specific Lexicon Mapping: A researched mapping of weather related terminology was developed to assign sentiment scores based on the intensity and emotional impact of specific words or phrases. For example:

- Positive high-impact terms, such as "scorching" and "perfect", were assigned scores of 4 and 3, respectively.
- Neutral or mild terms, like "breezy" and "sunny", were assigned scores ranging from 0.5 to 1.5.
- Negative terms that reflect cold or harsh conditions, such as "freezing" or "stormy winds", were assigned scores between -2 and -4.

Certain meteorological terms that hold significance in weather discussions were carefully integrated into the lexicon. These include terms such as:

- Beast from the East: A term referring to severe cold weather patterns originating from the east, scored at -4 because it often brings heavy snow, icy conditions, and sub-zero temperatures.
- SSW Polar: Refers to sudden stratospheric warming events in polar regions, often linked to disruptive weather patterns, scored -4 because it leads to cold air flowing south, resulting in extreme cold spells in North America and Europe
- Positive NAO: Represents the North Atlantic Oscillation in its positive phase, associated with milder winters, scored at 3.

The lexicon covered a wide range of terms, from general weather expressions (sunshine or rain) to technical meteorological phenomena (polar vortex or sudden stratospheric warming). This mapping ensured that the sentiment analysis was sensitive to the context and magnitude of weather related conditions described in the tweets.

Incorporating this custom lexicon allowed the sentiment analysis to more accurately capture the connotations of weather related discussions and produce more precise sentiment scores for the VADER model. This enhancement improved the model's ability to differentiate between extreme positive, neutral, and extreme negative weather sentiments. It also accounted for domain-specific language, including meteorological terms, and enhanced the interpretability of results by aligning sentiment scores with different weather phenomena. This process underscores the importance of adapting sentiment analysis tools to meet the requirements of a dataset which significantly improves the model's effectiveness.

4 RESULTS

4.1 Topic Modeling

The use of anchor words proved effective in identifying distinct topics within the dataset. However, the initial strategy of aligning each tweet to a single topic presented significant challenges. Tweets often spanned multiple topics (e.g., 'Latest AO, NAO, and PNA forecast') and therefore making a strict, single-topic classification impractical. To address this, a soft topic assignment approach was employed, allowing for the capture of the degree of association between each tweet and various topics. This method provided a weighted representation of topic membership rather than restricting each tweet to a single topic. Given the frequent instances where multiple topics exhibited similar probabilities this approach was the most effective for representing the complexity of the data.

The topic modeling results effectively captured various themes within each category. The image displaying the topic counts shows that each category is well represented. It is expected that the "General" topic has the highest number of aligned tweets, as many tweets often consist of simple phrases like "winter is here". Additionally, the notable volume of tweets associated with the "SST" topic was an intriguing finding. This observation was less about the actual prevalence of tweets discussing SST and more about a limitation within the model. The model assigns each tweet a probability score for belonging to a specific topic, and in many instances where SST was incorrectly aligned, the assigned probability was low. These cases of low probability often revealed misalignment when the tweet content was examined alongside the topic selected by the model.

An additional analysis focused on evaluating the co-occurrence of topics to understand which ones frequently appeared together. This was achieved by measuring the overlap between topics through shared terms and assessing their thematic similarity. A term-document matrix was created for the topics, representing each topic by its top 25 words and weighted scores (e.g., TF-IDF). Pairwise cosine similarity was then computed between topic vectors. The analysis revealed that the "AO & MJO" category had the highest similarity score, closely followed by the "Rainfall & Snow" category. Despite the high similarity score, the "Rainfall & Snow" category had 950 tweets mapped to it, compared to just 97 tweets mapped to "AO & MJO".

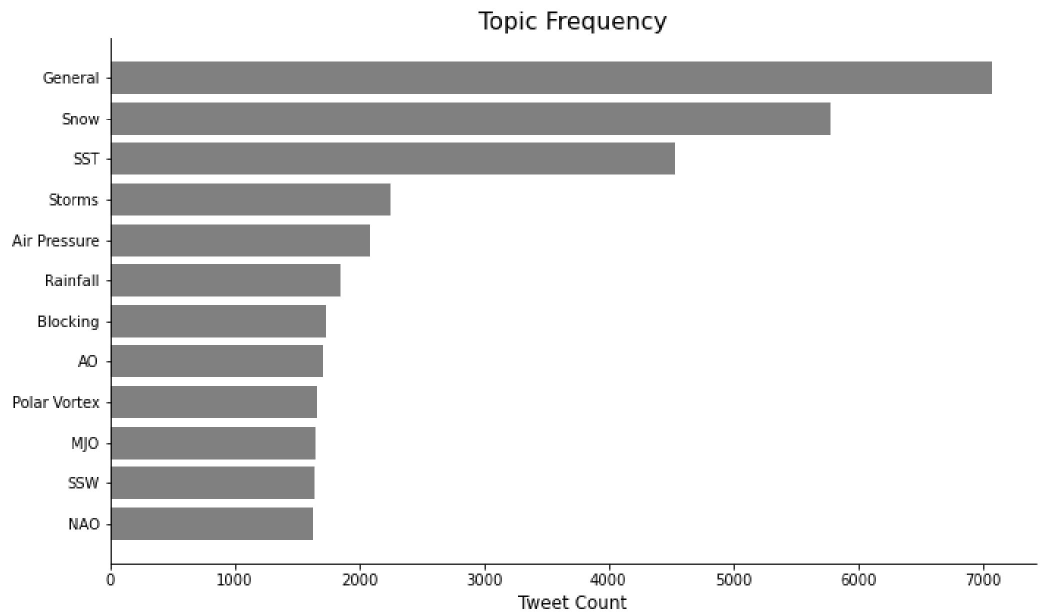

Figure 1—Number of tweets associated with each weather related topics

4.2 Named Entity Recognition

NER performed reasonably well, however, not all tweets contained references to a location. In cases where a GPE was detected, the model successfully identified and mapped the locations accordingly. Out of 16,144 tweets, a country was identifiable in 2,466 instances, representing approximately 15% of the data. A total of 72 countries were mapped across 17 regions, with the top three being the United Kingdom (1,436 instances), Greenland (228 instances), and the United States (217 instances). The accuracy of the results was influenced by the way tweets were originally written. For example, NER often failed to recognize countries when they were not properly capitalized (e.g., "norway" instead of "Norway"). Additionally, certain references, such as the use of 2-digit codes (e.g., "UK"), led to misinterpretations. This occasionally caused errors where NER mapped a code like "LA" to the wrong location (e.g., Laos instead of Los Angeles). Overall, while NER demonstrated effective performance in many cases, the accuracy of location mapping was impacted by the original format and context of the tweets.

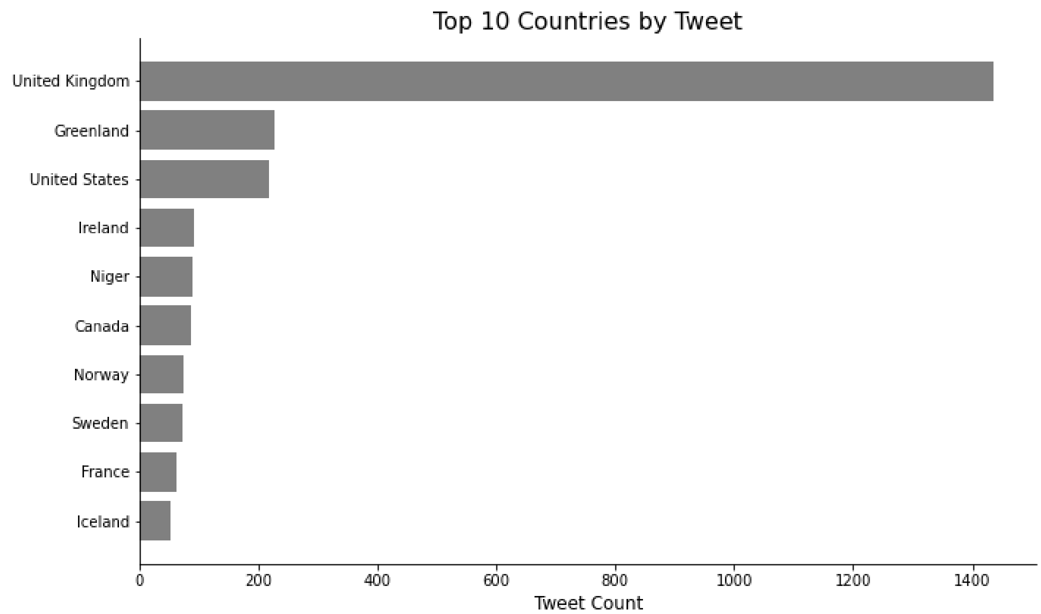

Figure 2—Number of tweets associated with each country

4.3 Sentiment Analysis

For each tweet, the VADER compound score was calculated. This score is determined by summing the individual sentiment scores of each word, and normalizing the total to a value between -1 and 1 using a standard formula. Scores closer to -1 indicate negative sentiment, while scores closer to +1 indicate positive sentiment. The average sentiment across all tweets in the dataset was -0.48, suggesting that the overall sentiment leaned toward negative weather events, such as cold weather and storms. Additionally, average monthly sentiment was plotted over the entire time horizon to assess changes in sentiment trends.

The lowest average monthly sentiment score observed was -0.735, recorded in August 2016. During this month, 49 tweets were analyzed, with 37 of them focusing on storm-related topics. Examples of these tweets included discussions surrounding the anticipated arrival of a hurricane:

- *Tropical Storm #Earl has formed in the Caribbean. Max winds of 45 mph. Earl will continue to move westward today.*
- *Tropical Storm Fiona forms, the 6th of the season and 2835 miles from West Palm Beach. Check the track.*
- *Gulf Depression upgraded to Tropical Storm Hermine with top winds 40 mph.*

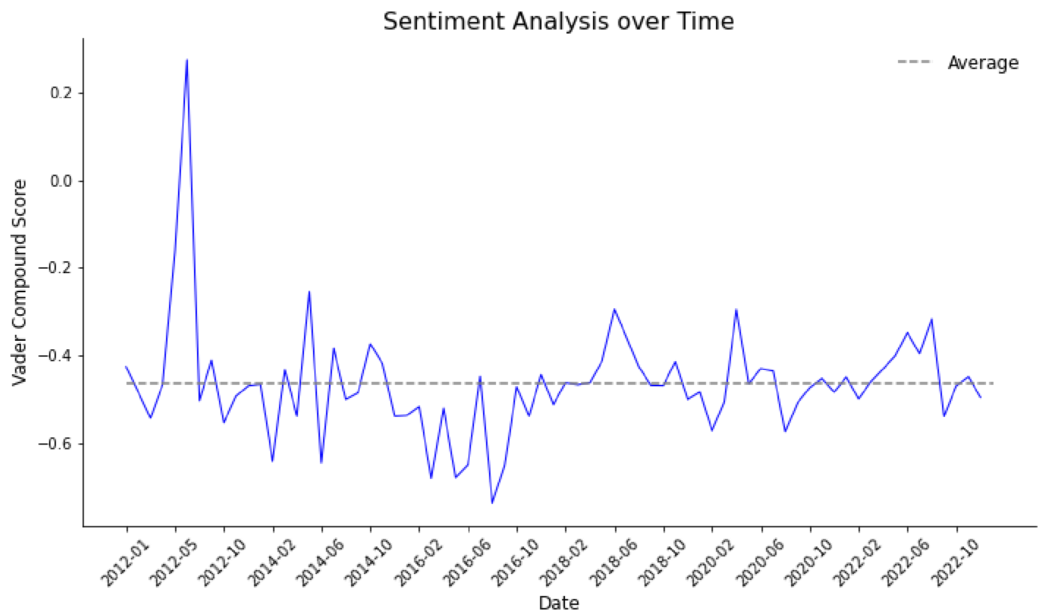

Figure 3—Average monthly sentiment across all tweets

Overall, the sentiment analyzer performed reasonably well, however, due to the brevity of many tweets after data cleaning, accurately gauging the 'true' sentiment expressed in those messages proved challenging.

5 EVALUATION AND DISCUSSION

5.1 Model Performance

The sentiment analysis model showed reasonable effectiveness in capturing the general sentiment of weather-related tweets from meteorologists. By leveraging the VADER compound score, the model successfully normalized sentiment values within a range of -1 to 1, allowing for a comprehensive evaluation of tweet sentiment. The overall average sentiment score of -0.48 indicated a prevailing focus on negative weather events, highlighting the inherently negative bias often present in meteorological discussions.

The model performed notably well in identifying sentiment trends over time, as illustrated by the monthly average sentiment analysis. It effectively detected the sentiment linked to specific events, such as tropical storms and hurricanes, which carried negative connotations due to their associations with strong winds and potential damage.

5.2 Validation

Validating the sentiment and topic classification models posed unique challenges, largely due to the absence of corresponding weather data that could confirm the accuracy of the meteorologists' predictions. Without a direct link between tweets and actual weather outcomes, it was not feasible to conduct a thorough comparison between the predicted sentiments and real-world events to evaluate the models' accuracy rigorously.

Despite these validation constraints, the trends observed in the dataset offered some level of reassurance. For instance, the predominance of negative sentiment during periods characterized by frequent storm activity aligned with general expectations for discussions related to adverse weather events. This alignment suggested that while direct validation was limited, the models were effective in capturing overarching patterns reflective of meteorological conditions.

5.3 Challenges

The sentiment analysis process faced several challenges, primarily due to the nature of Twitter data and the intricacies of interpreting weather related discussions. One significant issue was the ambiguity of sentiment context. Many tweets featured images with minimal or no accompanying text, making it difficult for the model to accurately infer sentiment. For instance, references to "NAO" were common but often lacked details indicating whether it referred to a positive or negative phase. This differentiation is crucial, as the sentiment associated with a positive NAO (indicating milder weather) is very different from that of a negative NAO (implying severe cold). This lack of specificity introduced uncertainty in the sentiment scoring, impacting the overall accuracy of the analysis.

Additionally, tweets often depicted snowy landscapes, graphs, or forecasts with simple captions like "NAO," offering limited context for the model to determine sentiment. The reliance on visual elements in tweets posed a further limitation. Many users conveyed sentiment through weather-related images, such as maps, memes, or photographs, which the current model could not process or interpret. As a result, valuable sentiment embedded in these visuals was not captured, reducing the model's effectiveness, especially in cases where visual context was essential for accurate sentiment interpretation.

6 CONCLUSION

6.1 Summary of Findings

The analysis revealed several key insights into the relationship between meteorological tweets and weather sentiment. Negative weather events, such as storms, hurricanes, and cold winter conditions, were more prevalent, reflected in the overall average sentiment score of -0.48. Notably, August 2016 stood out as the month with the lowest sentiment, averaging -0.735, driven by an increase in discussions surrounding tropical storms like Earl, Fiona, and Hermine.

Certain weather phenomena emerged as significant topics within the dataset. Severe weather terms, including "hurricane," "storms," and "polar vortex," were frequently mentioned and were strongly associated with negative sentiment. In contrast, positive sentiment, while less frequent, was linked to phrases such as "perfect weather" and "mild conditions."

Sentiment trends consistently showed more negative sentiment during periods of severe weather activity. These findings underscore the potential of using meteorological tweets as indicators of weather-related sentiment, which could be leveraged for forecasting market impacts, particularly in sectors such as energy and commodities.

6.2 Implications

The findings from this sentiment analysis have practical applications across various domains. One significant implication is its potential use in energy and gas demand forecasting. By assessing weather related sentiments, organizations can better anticipate changes in energy consumption, such as increased heating demand during severe cold or heightened cooling demand during extreme heat. Additionally, the analysis can enhance emergency preparedness by providing insights into public sentiment during adverse weather events. For instance, identifying heightened negative sentiment linked to terms like "hurricane" or "storm surge" could act as an early warning for resource deployment and signal potential increases in energy costs due to higher demand. Ultimately, applying and further developing this analysis can support more accurate predictions of future energy prices, enabling stakeholders to stay ahead in the energy market.

7 FUTURE WORK

This project presents several opportunities for future improvements and expansion to enhance its applicability and accuracy. One significant area for development is the integration of real-time weather data. By incorporating live weather information, the analysis could provide more immediate and precise sentiment assessments, enabling the tracking of public reactions to current weather events. This would facilitate the creation of dynamic and predictive models, where sentiment shifts in real time as weather conditions evolve.

Additionally, future enhancements could include developing multifaceted sentiment analysis capabilities to process not only text but also images and videos shared in tweets. This improvement would address the current limitation of excluding visual content from analysis, potentially boosting the accuracy of sentiment scores, particularly when images provide critical context for weather-related discussions.

Another advancement would be to reduce the manual steps currently involved in the analysis, such as selecting anchor words for topic modeling and assigning sentiment scores to weather-specific keywords. Automating these processes would streamline the end-to-end workflow and improve scalability.

Finally, the model's ability to handle ambiguous meteorological terms could be strengthened by implementing advanced natural language processing techniques and creating a graphical analyzer. This would enable more precise sentiment classification, especially for terms like NAO and other weather phenomena with varied meanings.