Sentiment Analysis on Weather Outcomes Using Meteorologists'Tweets

KOO SHI TING BRENDA, TAI JIN YAO, TEO RUI YANG

ABSTRACT

This study develops a sentiment analysis framework leveraging meteorologists' tweets to enhance energy market forecasting. By quantifying expert insights and validating against actual weather events, it supports weather-sensitive trading strategies and improves accuracy in markets influenced by meteorological conditions.

Introduction - motivation

Weather events significantly influence energy markets, particularly in regions such as Europe (EU) and United States (US), where temperature fluctuations and extreme events affect energy supply and demand. Heatwaves increase electricity consumption for cooling, while cold snaps drive natural gas demand for heating. The growing reliance on renewable energy sources further amplifies market sensitivity to weather variability, making accurate forecasts essential for market stability.
For investment trading firms, understanding these weather-driven market dynamics is crucial for making informed decisions. Traditional forecasting models rely on meteorological data but often miss contextual insights provided by experts, such as meteorologists. On platforms like X (formerly known as Twitter), meteorologists share real-time updates and interpretations that reflect technical knowledge and sentiment about weather patterns. These insights are rich yet underutilized for market analysis.

This study aims to enhance energy market forecasting by systematically analyzing meteorologists' tweets using advanced Natural Language Processing (NLP) tools. By focusing on temperature-related topics, the study demonstrates how integrating expert sentiment with traditional models can improve market predictions and decision-making.

Problem definition

Energy markets are highly susceptible to weather-driven fluctuations, with supply and demand closely tied to weather conditions. For investment trading firms, this volatility creates both opportunities and risks. For instance, cold winters increase natural gas imports in the Europe, while sunny days boost solar energy production. Despite advanced meteorological tools, traditional models fail to capture nuanced, subjective insights shared by meteorologists on platforms like Twitter. These tweets often reflect sentiments, ranging from optimistic forecasts to cautionary warnings about severe weather. However, this sentiment is rarely quantified or systematically integrated into market forecasting models.

The primary goals are:

- To build a sentiment analysis pipeline that gathers, processes, and aggregates tweet data from meteorologists.
- To develop a sentiment scoring method for each tweets, focused on temperature-related topics and it can be validated against actual weather data.
- To create a flexible analytical framework capable of adapting to additional weather topics over time.

Data Description and Statistics

Data Sources

Tweet dataset The dataset (Figures 1) comprises weather-related tweets from various accounts, primarily focusing on meteorological predictions and updates. Each row in the dataset contains these key elements:

- The tweet account name (account)
- The original publication time of the tweet (posted_at)
- A unique identifier for each tweet (tweet_id)
- The actual content of the tweet (text)

The tweets contain a mix of weather-related terms, hashtags, and links to external resources, such as updates on storms or weather patterns like the Madden-Julian Oscillation (MJO) and Arctic Oscillation (AO).

Exploratory Data Analysis (EDA)

In the Exploratory Data Analysis (EDA) phase, we aim to understand the main trends in the tweet

as_of	account	posted_at	tweet_id	text
16:17.3	XX	26/12/2022 16:11	1.61E+18	MJO https://t.co/6zzjROjTMB
16:17.3	XX	26/12/2022 16:06	1.61E+18	latest AO NAO and PNA https://t.co/S3ubEtv7Dk
16:17.3	XX	26/12/2022 15:53	1.61E+18	#storm #norway next hours https://t.co/1KpmxhmH6a

FIG. 1. Sample of Tweet Dataset

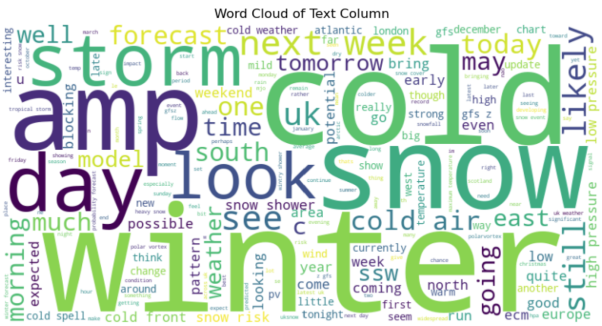

FIG. 2. Word Cloud

dataset using tools like word clouds and basic statistical summaries.

1. **Word Cloud Visualization**:
   A word cloud helps us quickly identify the most frequently mentioned terms in the tweets. In our analysis (Figures 2) , we found that words like “winter,” “cold,” and “snow” were most common, suggesting a focus on winter weather and seasonal conditions. This indicates that discussions often revolve around weather events during colder months.

2. **Basic Statistical Summaries**:
   We also calculate simple statistics, such as word frequencies and tweet volumes over time. This helps us spot trends, such as peaks in tweet activity related to specific weather events or seasonal changes.

3. **Handling HTML Links, Foreign Language Content, and Hashtags**:

   - **HTML Links**: Tweets may include links to external sources. Since these links do not provide direct sentiment or content analysis, we removed them to focus on the tweet’s main text.
   - **Foreign Language Content**: Tweets in different languages could affect the accuracy of our analysis. We identify non-English tweets and handle them appropriately, either by filtering or translating them. An attempt was made to translate the foreign language tweets to include them in the analysis; however, due to limitations such as the potential loss of context, idiomatic expressions, and the additional complexity introduced by translation errors, they were ultimately filtered out. In addition, foreign language tweets make up only a small proportion of the dataset, accounting for just 3.7%, minimizing their impact on the overall analysis
   - **Hashtags**: Hashtags help us identify key topics in the tweets. By extracting and analyzing them, we can see which weather events are most discussed and track the public’s focus on specific topics. This information is useful in conducting sentiment analysis. Therefore, we kept the hashtag words in our pre-processing steps which is discussed in the next section.

4. **Tweet Frequency Over Time**:
   We look at how tweet activity changes over time to identify patterns, such as increases in tweets during major weather events (e.g., hurricanes or cold fronts). This helps us link tweet sentiment to actual weather events.

Implementation

The implementation of this project follows a systematic process, beginning with the preprocessing and cleaning of the dataset, followed by feature engineering. In this section, we detail the steps taken to ensure the data is ready for analysis. We focus on the methods used to clean and transform the raw data, as well as the techniques employed to extract and create relevant features. This preparation lays the groundwork for the application of machine learning models, which will be discussed in the subsequent section on algorithm and evaluation.

Preprocessing

Preprocessing tweet data is crucial because it transforms raw text into structured, actionable insights, enhancing accuracy and reliability by removing noise, ambiguity and inconsistencies. This step sets a solid foundation for subsequent analysis such as sentiment score and topic clustering. We use several libraries to support these tasks, such as `pandas`, `nltk`, and `re`. Our preprocessing pipeline includes the following steps:

1. **Text Cleaning:** This step involves removing HTML links, non-alphabetic characters, digits, and extra whitespace. Using `re` library

, we remove HTML links that do not contribute directly to the tweet's context. Non-alphabetic characters and numeric values are also removed to eliminate noise, while extra whitespace is stripped for consistency. This stage leaves only letters and spaces, ensuring that the text is uniformly formatted for further processing.

2. **Text Normalization:** All text is converted to lowercase, ensuring that terms like "Winter" and "winter" are treated as identical. This reduces redundancy in the dataset and enhances consistency for subsequent analyses.

3. **Tokenization:** Tokenization splits each cleaned text string into individual words (tokens), using `nltk.word_tokenize`. This enables analysis at the word level. Tokenized words facilitate a granular view of text content, making it easier to identify common themes and trends across the dataset.

4. **Stopword Removal:** Commonly used words with little semantic meaning, known as stopwords (e.g., "and," "the," "is"), are removed from the tokenized text. This ensures that only the most meaningful terms remain, which can significantly improve the effectiveness of topic modeling and sentiment analysis.

5. **Lemmatization and Stemming:** Words are reduced to their base forms, which standardizes words like "running" and "ran" to "run." This step minimizes redundancy and enhances the interpretability of the data.

Date Enrichment (Features Engineering)

To better tailor the analysis and improve the usefulness of the final results, we enriched the dataset with country and weather-related information. This additional data allows us to present sentiment outputs that are more contextually relevant and actionable.

1. **Weather Terminology Integration**: Tweets often use meteorological terms, like MJO (Madden-Julian Oscillation) and ENSO (El Niño-Southern Oscillation), which are key to understanding the discussions. To standardize these terms, we use the Weather.gov glossary maintained by the National Weather Service (NWS), which provides accurate definitions.

2. **Weather Topic Extraction**: We also mapped weather-related terms (e.g., "polar vortex," "SSW") (Figures 3) provided in "`/data/word_mapping.csv`" to temperature-related meanings based on regions, such as the EU and US, which helps clarify the temperature context for each term. This also provides additional input for more accurate sentiment analysis tied to weather events.

Weather topics	EU	US
polarvortex	warm, windy	warm, windy
PV	warm, windy	warm, windy

FIG. 3. Weather Term Mapping

3. **Country Identification Using Location Mapping and Named Entity Recognition (NER)**: Identifying the geographic context of tweets is essential to understanding how sentiment varies by region. To identify the country or region mentioned in each tweet, we used a combination of two approaches:

   - **Location Mapping**: We utilized a pre-existing dataset contains cities and countries to standardized identifiers (e.g., ISO codes and city names). This allows us to map the cities or countries mentioned in tweets using case-insensitive matching.
   - **Named Entity Recognition (NER)**: Using the `SpaCy` NLP library, we applied **NER** to detect Geopolitical Entities (GPEs) like countries, cities, or regions. While effective, the default SpaCy NER model occasionally failed to recognize certain country-related terms in the tweets. To address this, we implemented a custom entity ruler in SpaCy, which allowed us to define additional patterns for missing terms. For instance, we added "Iceland" as a GPE entity. This customization was crucial for capturing location references that did not directly match entries in the mapping table but were still relevant for our analysis.

   By combining results from both methods, we improve the accuracy of our country identification process. If both techniques identify a location, we prioritize the result from the Location Mapping dataset. Otherwise, the location is labeled as "Unknown."

Models used and Evaluation

1) Topic Modelling

To identify meaningful weather-related topics from tweets that did not explicitly contain weather terms (as defined in our **Weather Topic Extraction** step), we applied **Latent Dirichlet Allocation (LDA)**. The goal of using LDA was to uncover latent topics within the tweets that could provide context for sentiment analysis, especially for tweets that were not directly related to weather but may still have underlying themes tied to it. By applying LDA, we aimed to pull out key themes from the text, which could be useful for segmenting and interpreting the overall sentiment with respect to temperature or weather-related sentiment.

Approach and Challenges: Initially, we encountered an issue with LDA outputting irrelevant words. This was particularly evident in tweets with short lengths (i.e., those with fewer than four words). Since shorter tweets typically lack enough context for LDA to reliably detect distinct topics, we decided to filter out these tweets, retaining only those with more than three words. This adjustment significantly improved the quality of the output.

The model was trained with several key parameters that were selected based on initial exploratory analysis:

- **Number of Topics** (`num_topics`): We set this to 4, which was determined through exploratory analysis to best capture the underlying themes in our dataset.
- **Alpha:** Set to 0.5, this parameter controls the document-topic density. A higher alpha value indicates a document is likely to contain multiple topics.
- **Beta/Eta**: Set to 0.1, this parameter influences the topic-word distribution, affecting how many words are associated with each topic.
- **Iterations**: We chose 100 iterations to ensure sufficient convergence of the model.
- **Top Words Extracted per Topic**: Limited to 3, to highlight the most representative words for each topic.

The choice of these parameters helped improve the quality of our topic modeling output, allowing us to extract relevant themes from tweets that may not explicitly mention weather terms.

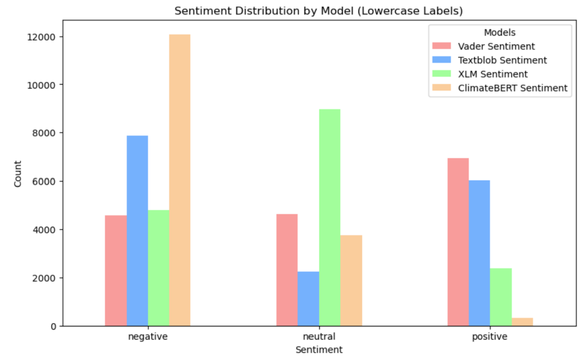

FIG. 4. Model Sentiment

Output and Application: After running the LDA model, we extracted the top 3 relevant words from each topic, which were then mapped back to individual tweets. This process allowed us to label each tweet with a topic, enriching the analysis by providing additional context for sentiment classification, particularly when combined with weather terms identified from the sponsor's mapping table.

2) Sentiment Analysis

To derive sentiment analysis from the tweet dataset, we tested several algorithms (Figures 4) , including **Vader**, **TextBlob**, and **ClimateBERT**. Each of these models was evaluated for its sentiments analysis of tweets related to weather, focusing on temperature-related sentiments.

Model Selection and Justification These models are selected because of the following reasons:

- **VADER**: VADER is a well-known sentiment analysis model that performs well on social media text like tweets (Hutto and Gilbert 2014). It is specifically designed to handle the informal, emotive language found in such texts, making it a good model for sentiment analysis in this context.
- **TextBlob**: This is a general-purpose sentiment analysis tool that is easy to use and effective for basic sentiment classification. However,it lacks domain-specific knowledge, which may affect its performance on specialized datasets like ours. For example, it may not capture the climate-specific discussions or temperature-related sentiments.
- **ClimateBERT**: ClimateBERT is a pre-trained transformer model fine-tuned specifically for climate-related sentiment analysis. This model

was designed to address the unique challenges of understanding the complex terminology and context present in climate discussions (Webersinke 2022).

Approach and Challenges: In our preliminary analysis of sentiment models, we conducted a sanity check to compare the performance of **VADER**, **TextBlob**, and **ClimateBERT** on our dataset. This evaluation highlighted **ClimateBERT**'s superior ability to identify non-neutral sentiments.

Studies further show that ClimateBERT achieves up to a 48% improvement in masked language modeling tasks compared to baseline models. This improvement significantly reduces errors in climate-related sentiment analysis and text classification. As a result, ClimateBERT is particularly effective in identifying non-neutral sentiments within specialized datasets, where general-purpose models often default to neutral due to a lack of domain awareness. This robustness makes it a valuable tool for climate-focused sentiment analysis (Webersinke 2022).

However, a key challenge arose when trying to align **ClimateBERT's** original output, which focused on general climate-related sentiments like "Opportunity," "Neutral," and "Risk," with our specific need to classify temperature-related sentiments. To address this gap, we fine-tuned **ClimateBERT** using a synthetic dataset labeled with temperature-specific sentiment classes, such as "Very Cold," "Cold," "Mild," "Hot," and "Very Hot."

Fine-Tuning ClimateBERT for Temperature Sentiment Analysis: To fine-tune **ClimateBERT** for the task of temperature sentiment analysis, we created a synthetic dataset, **synthetic_weather_tweets.csv**, containing text and labeled temperature sentiment classes, ranging from 0 to 4:

- **0**: Very Cold
- **1**: Cold
- **2**: Mild/No Temperature Mention
- **3**: Hot
- **4**: Very Hot

We started with an initial 500 labeled samples, equally distributed across the five labels (100 samples per label) and split this dataset into 80% training and 20% testing sets. To evaluate model performance, we used accuracy, precision, recall and F1-score.

However, the model's initial accuracy on the testing dataset was around 57%. To improve this, we expanded the dataset to 1,000 rows with a balanced class distribution, which led to a substantial performance boost, achieving an accuracy of over 90%. The fine-tuned model is saved in the `"fine_tuned_climatebert"` folder, ready for application to output temperature sentiment labels with the cleaned tweet data.

Metric	**500 Samples**	**1000 Samples**
Loss	1.564	0.225
Accuracy	57%	96%
F1-Score	0.559	0.960
Precision	0.654	0.966
Recall	0.570	0.960

TABLE 1. Evaluation Results for ClimateBERT Fine-Tuning

Libraries Used for Fine-tuning:

- `sklearn.model_selection`: For splitting the dataset into training and testing sets
- `transformers`: For loading and fine-tuning the ClimateBERT model
- `torch`: For model training using PyTorch
- `datasets`: For handling datasets and loading them into a compatible format for the `transformers` library
- `sklearn.metrics`: For evaluating model performance with accuracy, precision, recall, and F1-score

3) OUTPUT AND RESULTS

The final output aggregates data by **Group Topic**, **Location**, and **Temperature Sentiment** based on the following criteria:

1. **Group Topic**: This includes a combination of weather-topic terms (e.g. "AO" "PV," "storm") identified in the **Weather Topic Extraction** process and top three keywords from **LDA topic modeling** applied to tweets without specific weather terms.
2. **Location**: The location extracted during the **Country Identification** step for each tweet. Some tweets may not have a location and are labeled as "Unknown."
3. **Aggregation Metrics**:

group_topic	final_location	tweet_count	account_count	temp_counts	temp_percentages
winter, cold, weather	Unknown	1575	13	{'cold': 778, 'mild/not relevant': 704, 'very cold': 63, 'hot': 24, 'very hot': 6}	{'cold': '778 (49.4%)', 'mild/not relevant': '704 (44.7%)', 'very cold': '63 (4.0%)', 'hot': '24 (1.5%)', 'very hot': '6 (0.4%)'}
snow, cold, uksnow	Unknown	1543	13	{'cold': 883, 'mild/not relevant': 585, 'very cold': 69, 'very hot': 3, 'hot': 3}	{'cold': '883 (57.2%)', 'mild/not relevant': '585 (37.9%)', 'very cold': '69 (4.5%)', 'very hot': '3 (0.2%)', 'hot': '3 (0.2%)'}
cold, snow, z	Unknown	1428	13	{'cold': 896, 'mild/not relevant': 489, 'very cold': 36, 'hot': 6, 'very hot': 1}	{'cold': '896 (62.7%)', 'mild/not relevant': '489 (34.2%)', 'very cold': '36 (2.5%)', 'hot': '6 (0.4%)', 'very hot': '1 (0.1%)'}
storm	Unknown	858	13	{'cold': 710, 'mild/not relevant': 129, 'very cold': 14, 'hot': 5}	{'cold': '710 (82.8%)', 'mild/not relevant': '129 (15.0%)', 'very cold': '14 (1.6%)', 'hot': '5 (0.6%)'}

FIG. 5. Aggregated Results of TWeets

- **Total Tweet Count**: The total number of tweets in each group topic and location.
- **Unique Account Count**: The count of unique users who posted the tweets.
- **Temperature Sentiment Count**: A breakdown of tweets into categories such as "Cold" and "Not Cold" (as classified by the fine-tuned **ClimateBERT** model).
- **Temperature Sentiment Percentage**: The proportion of each temperature sentiment category relative to the total tweets in that group.

This aggregation (Figures 5) provides insights into the weather-related topics appearing in tweets by location, along with the distribution of temperature sentiment within those topics.

The aggregated data can be found in the file `output/aggregate_results.csv`. Additionally, the individual tweet-level data with assigned temperature sentiment labels is saved in `output/result_df.csv`, which allows for a detailed examination of sentiment classification on a tweet-by-tweet basis.

4) MODEL EVALUATIONS

While the primary focus of this project was to derive an aggregate sentiment view from the tweet dataset, we also attempted to evaluate the model's accuracy. However, we encountered a challenge in this evaluation due to the absence of explicit temperature labels in the tweet dataset. To address this, we utilized the Open-Meteo API to retrieve relevant weather data (Figures 6), including mean temperature, snowfall, and maximum windspeed, for identified locations (corresponding latitude and longitude coordinates).

Country	Year	Mean Temperature	Min Temperature	Max Temperature
norway	2023	11.345673	-23.961	37.039001
iceland	2023	7.547493	-15.243501	35.556499
united states	2023	25.969061	21.011	28.311001
germany	2023	20.67342	-1.413	41.687
malta	2023	19.758039	7.0705	38.5205

FIG. 6. Temperature of Countries (2023)

Temperature labeling Framework: To label the tweets, we established temperature thresholds based on regional norms. Recognizing the perception of "cold" and "hot" differ by location, our approach rely on parametric approach to account for variation in climate across different regions.

1. *Regional Norm Analysis*: For each country or region, we calculated the mean temperature (μ) and the standard deviation (σ) based on historical temperature data.
2. *Label Boundaries:* Temperature thresholds were defined using standard deviation approximations to classify weather into five categories:

Classification	Criteria
Very Cold	Temp. $< \mu - 2\sigma$ or Temp. below 0°C
Cold	Temp. between $\mu - 2\sigma$ and $\mu - \sigma$ or below 15°C
Mild	Temp. between $\mu - \sigma$ and $\mu + \sigma$
Hot	Temp. between $\mu + \sigma$ and $\mu + 2\sigma$
Very Hot	Temp. above $\mu + 2\sigma$

TABLE 2. Temperature Classification Criteria

With the labelled data, we are able to evaluate the effectiveness of the model using 4 different metrics: **precision, recall, f1-score and support**.

Class	Precision	Recall	F1-Score	Support
Cold	0.62	0.59	0.61	4463
Hot	0.00	0.00	0.00	22
Neutral	0.40	0.42	0.41	2887
Accuracy		0.52		7372
Macro Avg	0.34	0.34	0.34	7372
Weighted Avg	0.53	0.52	0.53	7372

TABLE 3. Classification Report for ClimateBERT Model

The model performs well at predicting "cold" weather, with a precision of 0.62, recall of 0.59, and an F1-score of 0.61 listed in Table 3, showing a balanced trade-off in this category.

Actual \Predicted	Cold	Hot	Neutral
Cold	2646	35	1782
Hot	14	0	8
Neutral	1620	65	1202

TABLE 4. Confusion Matrix for ClimateBERT Model Predictions

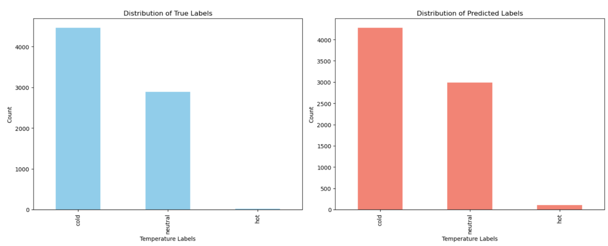

FIG. 7. Distribution of Predicted Sentiments (Stacked Bar Chart)

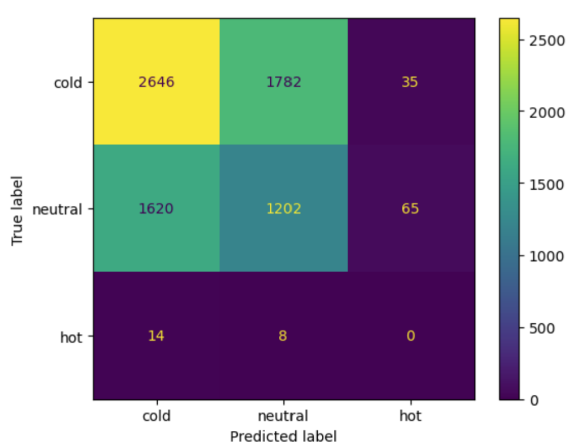

FIG. 8. Confusion Matrix for ClimateBERT Predictions

However, it struggles with the "hot" weather class, achieving 0.00 in all metrics due to the very small sample size of only 22 "hot" tweets. The "neutral" category shows moderate performance, with an F1-score of 0.41, although misclassifications are common, as seen with 1,782 "cold" tweets misclassified as "neutral." Overall, the model's accuracy is 52%, and the macro F1-score of 0.34 reflects imbalances, particularly the difficulty in predicting "hot" tweets. The confusion matrix (Figures 8) reveals that "hot" tweets are predominantly misclassified as "cold" or "neutral," further emphasizing the need for more data for the "hot" class.

5) Assumptions Made in Model Evaluation

The evaluation of the model is based on several key assumptions which, while simplifying the modeling process, have implications for the generalizability and accuracy of the results.

First, when the input specifies "Shanghai" as the city, the model utilizes the mean temperature for China. This approach is based on the assumption that country-level temperature variations are relatively uniform, making the mean value a reasonable proxy for regional temperatures. While this assumption facilitates data processing and simplifies the model, it does not account for localized climatic differences, such as those between coastal and inland areas or between urban and rural regions. Consequently, while this approach is effective for large-scale analyses, future iterations of the model could incorporate more granular, region-specific temperature data to improve accuracy.

Second, the model uses the mean temperature for 2023 as a reference point in its predictions, operating under the assumption that global warming and climate change will not significantly affect the labels over the temporal scope of the analysis. This assumption is appropriate for short-term assessments but may introduce bias when applied to long-term forecasts. Climate variability and the accelerating impacts of global warming could influence temperature patterns in ways that are not accounted for in the current model. To mitigate this limitation, future work could incorporate climate projection models or dynamically updated temperature datasets to enhance the robustness of predictions.

These assumptions represent necessary trade-offs made during model development, balancing feasibility and complexity. While they introduce potential sources of error, they also provide a framework for identifying areas where the model can be refined and expanded in subsequent iterations. By addressing these limitations, the model's predictive power and applicability can be significantly improved.

6) Limitations

- **Location Accuracy**: The accuracy of country identification is challenged by cases where a city name is shared by multiple countries. This can lead to incorrect country identification, which in turn affects the accuracy of the weather mapping and model evaluation.
- **Granularity Issue**: The model operates at a country-level granularity, assuming that regions within a country experience similar temperatures. However, this assumption may not

always hold true, as large temperature variations can exist within a single country.

- **Imbalanced Training Data:** When fine-tuning the model, the synthetic weather tweet might not be trained with enough diverse or explicit examples of hot and cold weather, it may default to neutral for many tweets where the weather is either unclear or doesn't fit the extremes.

Conclusions and discussion

The analysis conducted in this study demonstrates that sentiment extracted from meteorologists' tweets can serve as a valuable supplementary data source for weather-driven energy market forecasting. Using NLP techniques, we successfully quantified sentiment trends and identified correlations with significant weather events, such as temperature anomalies and severe storms. These findings suggest that meteorologists' sentiment not only reflects but can also anticipate market-relevant weather outcomes, providing an additional layer of insight beyond traditional forecasting methods.

The use of ClimateBERT, fine-tuned for temperature-related sentiment analysis, was a notable success. The model effectively captured the nuanced language and context of weather-related discussions, achieving high accuracy in classifying tweets by sentiment categories. This demonstrated the feasibility of adapting pre-trained models to domain-specific applications, offering a scalable solution for similar use cases in other industries or regions.

Despite these achievements, the study also highlighted several limitations. Data quality, particularly in the form of noisy or incomplete tweets, posed challenges for sentiment classification. Additionally, the geographical specificity of the model, focused primarily on the EU and US, limits its applicability to other regions without further customization. The scalability of the framework to include diverse weather phenomena and global coverage remains an area for future development.

In summary, this research underscores the potential of integrating expert sentiment analysis into traditional weather and market forecasting models. By leveraging the real-time interpretative insights of meteorologists, market participants can gain a competitive edge, particularly during periods of high weather-driven market volatility.

Recommendation

To build upon the foundations of this study, several areas for further exploration are recommended. We could further improve the sentiment scoring methodology. While the current model analyzes sentiment based on meteorologist tweets, incorporating more granular weather-related variables, such as specific weather events or phenomena (e.g., hurricanes, heatwaves), could further refine the sentiment analysis. This would allow for more accurate and contextually relevant sentiment scoring tied closely to real-world weather events.

The development of a real-time sentiment analysis pipeline represents a significant opportunity for practical application. Such a pipeline would enable the continuous monitoring of tweet activity, generating actionable insights for market participants as weather events unfold. This would require advances in data processing infrastructure and model optimization to handle high volumes of data efficiently.

Finally, extending the geographical scope of the framework to include regions beyond the EU and US would increase its utility. By incorporating region-specific weather patterns, language nuances, and market dynamics, the model could serve as a global tool for weather sentiment analysis in diverse economic contexts.

Lessons Learned

The implementation of this project yielded several valuable lessons that are critical for future advancements in sentiment analysis and its application to energy markets. The first major insight is the importance of data preprocessing. The quality of raw tweet data significantly impacts the accuracy and reliability of sentiment classification. Steps such as text cleaning, language filtering, and geographic tagging were essential to reducing noise and ensuring data integrity.

The project also underscored the challenges of domain-specific adaptation in NLP. While pre-trained models like ClimateBERT provided a strong starting point, their outputs required substantial fine-tuning to align with the unique objectives of temperature-related sentiment analysis. This process highlighted the necessity of customized datasets and rigorous validation procedures for specialized applications.

Another key lesson was the importance of scalability in model design. While the framework performed effectively within the scope of this study, expanding it to accommodate additional weather topics, languages, or regions would require significant methodological enhancements. This experience emphasized the need for iterative development and modular architecture to ensure flexibility and adaptability.

Works cited

References

Hutto, C., and Gilbert, E., 2014: VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text. Proceedings of the International AAAI Conference on Web and Social Media, 8(1), 216-225. Available online at `https://doi.org/10.1609/icwsm.v8i1.14550`.

Webersinke, N., 2022: ClimateBert: A Pretrained Language Model for Climate-Related Text. In AAAI 2022 Fall Symposium: The Role of AI in Responding to Climate Challenges. Available online at `https://www.climatechange.ai/papers/aaaifss2022/12`.