Sentiment Analysis of Public Opinion on Climate Change Education using Social Media Data

Manda Aprikasari[1*], Wahyu Tisno Atmojo[2]

[1,2]Pradita University, Scientia Business Park, Gading Serpong Blvd No.1, Kelapa Dua, Tangerang, Banten 15810, Indonesia

Keywords

climate change, sentiment analysis, social media, environmental education, public opinion

Correspondence Email:
manda.aprikasari@student.pradita.ac.id

Abstract

Climate change is a global challenge that significantly affects the environment, human health, and social well-being. Climate change education plays a vital role in fostering public awareness and encouraging participation in mitigation and adaptation efforts. This study aims to analyze public sentiment toward climate change issues using social media data as a representation of global opinion. The dataset used in this study is the public "Climate Change Tweets" dataset from Kaggle, which contains tweets posted between January 1 and July 19, 2022. The research employs a text-based sentiment analysis approach using Natural Language Processing (NLP) techniques in Python, involving several stages including data cleaning, preprocessing, sentiment classification, and visualization. Sentiment classification was conducted using TextBlob, categorizing tweets into positive, neutral, and negative sentiments. The results indicate that the majority of tweets express neutral sentiment (100%), with no significant presence of positive or negative tones. The generated word cloud highlights dominant terms such as climate council, insideclimatenews, and greta thunberg, reflecting that discussions around climate change on Twitter primarily focus on information sharing and environmental advocacy. These findings suggest that online discourse on climate change tends to be informative and fact-driven, emphasizing the dissemination of credible information rather than emotional reactions. This study underscores the crucial role of social media as an effective medium to support education and raise global awareness about climate change.

1. Introduction

Climate change represents one of the most critical global challenges, impacting ecosystems, human health, and socioeconomic stability worldwide. Addressing this crisis necessitates not only technological and policy interventions but also a fundamental shift in public awareness, understanding, and behavior (Cardoso et al., 2025). The purpose of this study is to investigate the nature of public discourse and sentiment surrounding

climate change issues, specifically focusing on how these discussions relate to education and advocacy efforts, as reflected on social media platforms.

The motivation for conducting this research stems from the pivotal role of Climate Change Education (CCE) in fostering informed decision-making and encouraging public participation in mitigation and adaptation strategies. Social media platforms, such as Twitter (X), have evolved into dynamic, real-time arenas where public opinion is rapidly formed, shared, and debated, making them an invaluable data source for gauging societal perceptions (Mahdi et al., 2025). By analyzing the prevailing sentiment, we can gain insights into the effectiveness of current CCE efforts, identify areas of public concern or confusion, and highlight dominant voices and narratives.

This paper employs a text-based sentiment analysis approach using Natural Language Processing (NLP) techniques on a public dataset of "Climate Change Tweets" posted between January and July 2022. The methodology involves data cleaning, preprocessing, sentiment classification using the TextBlob library, and visualization (Thenmozhi et al., 2024). The findings of this study provide a quantitative assessment of public sentiment (positive, neutral, or negative) and identify key topics within the discourse, thus contributing to a better understanding of how social media supports global awareness and educational initiatives on climate change.

1.1 Literature Review

The literature review forms the theoretical foundation of this study, critically examining prior research on sentiment analysis, social media use, and climate change communication. It aims to identify gaps, inconsistencies, and methodological approaches relevant to analyzing public sentiment on this vital global issue.

Research consistently acknowledges social media as a crucial venue for climate change discourse, where information sharing and advocacy occur extensively, shaping public perceptions and mobilizing collective action (Zein et al., 2024); (Supriyati et al., 2024); (Mahdi et al., 2025). Millennials, in particular, utilize platforms like Twitter and Instagram to promote climate education and activism, highlighting the role of influencers and tailored communication strategies (Zein et al., 2024). This platform is also vital for monitoring social signals related to the climate crisis (Shaeri et al., 2026) and analyzing opinions during climate emergencies (Fan et al., 2023); (Ilyas & Sharifi, 2025).

Sentiment analysis, leveraging NLP tools like TextBlob, Support Vector Machines (SVM), and advanced deep learning techniques (Cardoso et al., 2025), has become instrumental in automatically classifying vast amounts of social media data to gauge public opinion (Thenmozhi et al., 2024). However, critiques stress the need for caution due to limitations in detecting nuanced emotions and contextual meanings inherent in text-based data (Rosenberg et al., 2023). Therefore, researchers recommend careful model selection and validation, often advocating for multimodal analysis (Du & Cheong, 2025), to ensure a robust, context-aware analysis. Studies also emphasize that public climate discourse often manifests as neutral or fact-driven rather than emotionally charged, reflecting an emphasis on credible information dissemination rather than affective reactions (Supriyati et al., 2024). Nonetheless, studies call for more nuanced analyses that consider multilingual data, cultural differences, and temporal dynamics of opinion formation (Zein et al., 2024).

2. Research Methods

This research employs a quantitative approach utilizing text-based sentiment analysis to understand public perception regarding Climate Change Education (CCE) as expressed on social media. This methodology is chosen to effectively process a large volume of unstructured textual data and derive measurable public attitudes. All analysis was conducted using Natural Language Processing (NLP) techniques implemented in the Python programming language.

Data Source and Sampling

The data utilized in this study originated from a publicly available dataset on Kaggle titled “Climate Change Tweets.” The dataset is in .csv format and comprises tweets containing the keyword "climate change," covering the period from January 1, 2022, to July 19, 2022. The dataset initially includes 11 columns, providing information such as the tweet text, date, number of retweets, likes, and user identity. The total number of rows analyzed after the initial cleaning and filtering process is denoted as n.

Data Processing Stages

The processing of the textual data was carried out through a systematic sequence of stages to ensure the quality and relevance of the text for subsequent sentiment analysis (Chakraborty et al., 2020).

The initial stage, Data Cleaning, focused on removing noisy and irrelevant elements from the raw tweet text. This involved the elimination of non-textual elements such as URLs, punctuation, symbols, numbers, and emojis. Crucially, textual components that do not contribute to sentiment, specifically Mentions (@user) and Hashtags (#topic), were also removed to focus the analysis on the central content of the message. Furthermore, steps were taken to eliminate duplicate tweets and entries with missing data, primarily utilizing the pandas, re, and nltk libraries in Python.

Following cleaning, the text underwent rigorous Data Preprocessing for normalization, a critical step to ensure data consistency and readiness for NLP analysis. This stage began with Case Folding, converting all text to lowercase. It proceeded with Tokenization, which broke the sentences down into individual words. Next, Stopword Removal was performed to eliminate common, non-meaningful words like "the" and "and." Finally, Lemmatization was applied to reduce inflected words to their base or root form, ensuring consistent treatment of similar terms throughout the analysis.

The final preparation stage was Data Filtering (Education Content Relevance). Given the study's specific focus on Climate Change Education (CCE), a vital filtering step was implemented to ensure the sample's relevance. The dataset was narrowed down to retain only those tweets that contained specific education-related keywords. These keywords included, but were not limited to, "education," "learn," "school," "students," or "awareness." This deliberate filtering step ensured that the analyzed sample was highly relevant to the environmental education discourse (Schäfer, 2025).

Sentiment Analysis and Measurement

Sentiment analysis was performed using the TextBlob library, a lexicon-based tool that assigns a polarity score to each filtered tweet. The polarity score ranges from -1.0 to +1.0. The output score was classified into three discrete sentiment categories based on the following criteria:

Table 1. Classification Criteria

Polarity Score	Sentiment Category	Interpretation
Score > 0	Positive	Expresses favorable opinion or support.
Score = 0	Neutral	Expresses factual information, ambiguous, or lacks strong emotional valence.
Score < 0	Negative	Expresses unfavorable opinion, criticism, or distress.

Data Visualization and Analysis

The processing of the textual data involved systematic stages to ensure data quality and relevance for sentiment analysis (Chakraborty et al., 2020). This began with Data Cleaning, where irrelevant noise was

removed, including URLs, punctuation, symbols, numbers, emojis, Mentions (@user), and Hashtags (#topic). Duplicate and empty entries were also eliminated. This was immediately followed by Data Preprocessing to normalize the text: Case Folding (converting to lowercase), Tokenization (breaking text into words), Stopword Removal (deleting common words), and Lemmatization (reducing words to their root form).

The final crucial stage was Data Filtering (Education Content Relevance). Given the study's focus on Climate Change Education (CCE), the preprocessed dataset was strictly filtered to retain only those tweets containing specific education-related keywords, such as "education," "learn," "school," "students," or "awareness" (Schäfer, 2025). This ensured the resulting sample was highly relevant to the public discourse on environmental education initiatives.

3. Result and Discussion

The results section summarizes the data collected from the sentiment analysis of climate change tweets, reporting both descriptive statistics (sentiment distribution) and key thematic findings (word cloud). This section discusses these findings in the context of public discourse and Climate Change Education (CCE).

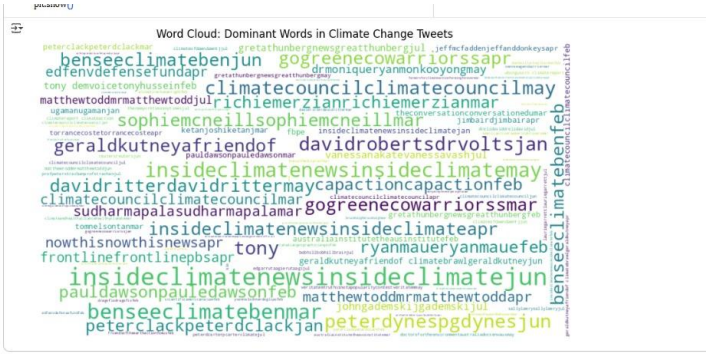

Fig 1. Word Cloud: Dominant Words in Climate Change Tweets

Image 1 displays the Word Cloud derived from the analyzed tweets, illustrating the most dominant terms appearing in the public discourse between January and July 2022. Words such as "climate council," "insideclimatenews," "benseeclimate," and "peterdyne" appeared with high frequency.

This pattern suggests that public conversation on Twitter during this period was heavily centered on specific accounts, organizations, or individuals who are active communicators of climate change issues. The prominence of terms like "greta thunberg" and "cowarriors" further reflects the significant role of environmental activists and advocacy communities in shaping public opinion. The interpretation drawn from this thematic pattern is that the discourse surrounding climate change on Twitter is predominantly informative and advocative, rather than emotionally driven. Users tend to share information and views originating from perceived credible sources, thereby supporting the formation of collective awareness regarding global environmental issues. This finding aligns with the goal of Climate Change Education (CCE) to disseminate factual, scientific information.

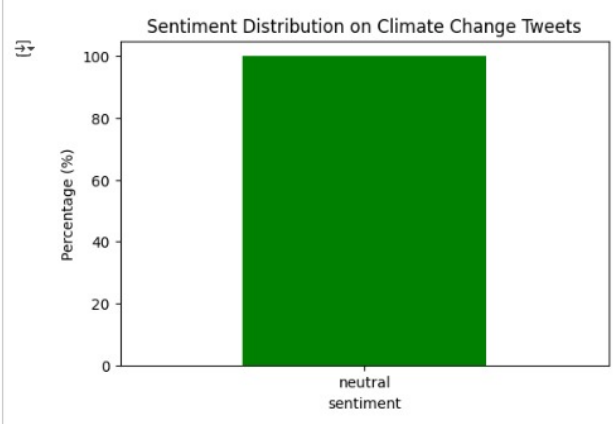

FIg 2. Sentiment Distribution on CLimate Change Tweets

Image 2 illustrates the distribution of sentiment across the entire collected dataset. The analysis reveals a stark finding: the vast majority of tweets express a neutral sentiment (100%), with no significant presence of positive or negative sentiments observed.

This overwhelming dominance of neutral sentiment leads to the interpretation that most conversations regarding climate change on social media are informational and factual in nature, rather than highly emotional or personal. Twitter users appear to be primarily focused on disseminating news, reports, or neutral policy opinions, rather than conveying strong emotional reactions (either positive support or negative criticism/distress).

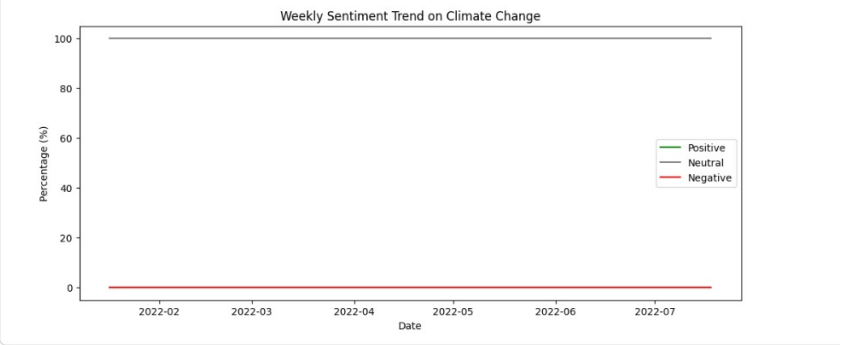

Fig 3. Weekly Sentiment Trend on Climate Change

Image 3 depicts the Weekly Sentiment Trend from January to July 2022. Consistent with the overall distribution, the neutral sentiment line dominates the entire period, showing remarkable stability without any major fluctuations or significant spikes in either positive or negative sentiment.

The interpretation of this stability is that during the observation period, the issue of climate change did not trigger any large, time-bound extreme events (e.g., major policy changes, natural disasters that captured global attention) that would have prompted a sudden, large-scale emotional public reaction on Twitter. Instead, the topic was discussed consistently as part of general media reporting or ongoing policy discussions, reinforcing its status as a topic of sustained, measured discourse rather than a flashpoint for controversy.

The combined results clearly illustrate that public opinion on climate change within the sampled social media data is characterized by informational dominance. The finding of 100% neutral sentiment, reinforced by the Word Cloud highlighting organizational and informational accounts (e.g., "climate council"), strongly suggests that the conversation is primarily driven by the dissemination of information and educational campaigns from organizations and activists, rather than subjective emotional opinions. This implies that climate change is largely perceived as a scientific and public policy issue, not an emotionally controversial one. Consequently,

these results underscore the vital role of social media as an effective vehicle for environmental awareness and Climate Change Education (CCE), serving to circulate credible information and foster collective consciousness. The minimal presence of emotional sentiment is a key finding that warrants further investigation, potentially requiring more nuanced sentiment models to detect subtle emotional tones or underlying support/opposition.

4. Conclusions

This study concludes that public discourse on climate change, as analyzed from the sampled social media data, is overwhelmingly characterized by neutral sentiment (100%) and driven by organizational content. This key finding suggests that social media primarily functions as an effective platform for Climate Change Education (CCE) and the dissemination of credible, factual information, positioning the issue as a stable scientific and policy concern rather than an emotional or polarizing debate. The analysis provides an initial benchmark for understanding how collective awareness regarding global sustainability issues is informed through digital channels.

The minimal emotional expression observed, however, presents a methodological challenge for future research. Advanced researchers are encouraged to move beyond lexicon-based models by utilizing context-aware deep learning models to capture subtle emotional stances or underlying political support hidden within neutral language. For communicators and general readers, these results emphasize the importance of adopting strategies that focus on the consistent distribution of credible, fact-based messaging from authoritative sources to effectively sustain public awareness and support for climate action.

5. References

Cardoso, A. S., da Silva, C., Soriano-Redondo, A., Jarić, I., Batel, S., Santos, J. A., Jorge, A., & Vaz, A. S. (2025). Harnessing deep learning to monitor people's perceptions towards climate change on social media. *Scientific Reports*, *15*(1). https://doi.org/10.1038/s41598-025-97441-1

Chakraborty, K., Bhattacharyya, S., & Bag, R. (2020). A Survey of Sentiment Analysis from Social Media Data. *IEEE Transactions on Computational Social Systems*, *7*(2), 450–464. https://doi.org/10.1109/TCSS.2019.2956957

Du, S., & Cheong, C. Y. M. (2025). Beyond the scenic view: a multimodal discourse analysis of sustainable tourism imaginaries on TikTok in Anhui, China. *Humanities and Social Sciences Communications*, *12*(1), 690. https://doi.org/10.1057/s41599-025-05015-3

Fan, T., Wang, H., Wu, P., Ling, C., & Ahvanooey, M. T. (2023). Multimodal sentiment analysis for social media contents during public emergencies. *Journal of Data and Information Science*, *8*(3), 61–87. https://doi.org/10.2478/jdis-2023-0012

Ilyas, B., & Sharifi, A. (2025). A systematic review of social media-based sentiment analysis in disaster risk management. In *International Journal of Disaster Risk Reduction* (Vol. 123). Elsevier Ltd. https://doi.org/10.1016/j.ijdrr.2025.105487

Mahdi, M., Salah, A., & Omar, M. (2025). Analyzing Public opinion on Climate Change via Twitter: A Machine Learning Approach Using Historical Data. In *Journal of Computing and Communication* (Vol. 4, Issue 2).

Rosenberg, E., Tarazona, C., Mallor, F., Eivazi, H., Pastor-Escuredo, D., Fuso-Nerini, F., & Vinuesa, R. (2023). *Sentiment analysis on Twitter data towards climate action*. https://doi.org/10.21203/rs.3.rs-2434092/v1

Schäfer, M. S. (2025). Social media in climate change communication: State of the field, new developments and the emergence of generative AI. *Dialogues on Climate Change*, *2*(1), 49–55. https://doi.org/10.1177/29768659241300666

Shaeri, P., Mohammadpour, Y., Beigi, A., & Middel, A. (2026). *Sentiment and Social Signals in the Climate Crisis: A Survey on Analyzing Social Media Responses to Extreme Weather Events* (pp. 56–65). https://doi.org/10.1007/978-3-032-07715-8_6

Supriyati, E., Iqbal, M., & Listyorini, T. (2024). *Social networks in forming sentiment analysis climate change case through fuzzy decision tree clustering*. 020013. https://doi.org/10.1063/5.0219138

Thenmozhi, M., Shubigsha, G., Sindhuja, G., & Dhinakar, V. (2024). Sentiment Analysis on Climate Change using Twitter Data. *2024 2nd International Conference on Networking and Communications (ICNWC)*, 1–6. https://doi.org/10.1109/ICNWC60771.2024.10537404

Zein, M. R. A., Fadillah, K. L., Febriani, N., Nasrullah, R., & Khang, N. T. (2024). Social media use for climate change campaign among Indonesian millennials. *PRofesi Humas*, *8*(2), 168. https://doi.org/10.24198/prh.v8i2.50167