iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 **Issue:** VIII **Month of publication:** August 2025

DOI: https://doi.org/10.22214/ijraset.2025.73573

Sentiment and Emotion Analysis of Climate Related Social Media Content

Priyanka Bakki[1], G. Praveen Babu[2]
[1]Post Graduate Student, M. Tech (Data Sciences), [2]Associate Professor, Department of Information Technology, UCESTH, Jawaharlal Nehru Technological University Hyderabad

Abstract: This study proposes a domain-specific sentiment and emotion analysis framework to address the limitations of general tools in capturing nuanced environmental discourse on social media. Using a hybrid model combining Pointwise Mutual Information (PMI) and Logistic Regression, along with NRCLex for emotion detection and BERTopic for topic modeling, the research analyses social media posts from 2014 to 2023. The findings show a predominance of negative sentiment, particularly on Twitter and Facebook, while Instagram displays more positivity. Common topics include climate change, plastic pollution, and air quality. The proposed model demonstrates improved accuracy over existing tools and is validated through expert annotations, offering a robust, multi-platform approach to understanding public perceptions of environmental issues—beneficial for policymakers, researchers, and communicators.
Keywords: Sentiment analysis, emotion analysis, social media, public perception, climate change, global warming, pointwise mutual information, Twitter, Instagram, YouTube.

I. INTRODUCTION

This paper focuses on analysing public sentiment and emotions related to environmental issues—particularly climate change—through social media content on Twitter, Instagram, and Facebook over a ten-year period (2014–2023) [1],[21]. By using a hybrid sentiment analysis model combining Pointwise Mutual Information (PMI) and Logistic Regression, the study classifies posts into positive, neutral, or negative sentiments and identifies emotions such as fear, trust, anticipation, anger, and sadness using NRCLex[17]. Topic modeling with BERTopic further reveals recurring themes like climate change, air quality, and plastic pollution. The findings show a predominance of negative sentiment, especially on Twitter and Facebook, while Instagram content is generally more positive [14],[19].

The paper's scope includes both sentiment and emotion analysis across platforms, with manual annotation by human raters—some of whom are domain experts—to validate the model. Compared to existing tools like VADER, spaCy, and Senti, the proposed model demonstrates improved contextual accuracy [3],[12]. This work highlights platform-specific differences in emotional expression and offers insights for researchers, policymakers, and environmental advocates seeking to design effective communication strategies and policy interventions [17].

The problem addressed is the lack of accurate, domain-specific sentiment analysis tools capable of handling informal, nuanced social media language. Existing models often fail to capture sarcasm, layered emotions, or platform-specific behavior [10],[22]. This study responds to that gap by offering a multi- platform analytical framework tailored to environmental discourse, helping to track shifts in public opinion and support more informed environmental action. Future work will include multimodal content such as images and videos to enhance analysis depth [22].

II. LITERATURE REVIEW

Research on sentiment and emotion analysis of environmental content in social media has increasingly focused on understanding public perception through platforms like Twitter, Reddit, and YouTube [3],[16],[17]. While early studies were based on consumer reviews, recent work shows that environmental discussions often carry negative sentiment, especially on topics like pollution and emissions, though some positivity is found in areas like sustainability [18],[19].

Social media sentiment effectively mirrors real-world environmental events and can even serve as an early warning system for disasters [17],[23]. Researchers have used lexicon-based tools (e.g., VADER, TextBlob)[12], machine learning models (e.g., Naïve Bayes, SVM, RNN)[4],[5], and deep learning methods alongside emotion detection tools to identify emotions like fear, trust, and anticipation. However, most studies focus narrowly on English-language Twitter data and use general-purpose models, which struggle with domain-specific language, sarcasm, and short-form content, reducing their accuracy and scope [11],[22].

To overcome these limitations, the current study introduces a comprehensive, domain-specific framework using a PMI-based sentiment classifier, which achieves higher accuracy (65%) than general tools like VADER [12]. It includes data collection from multiple platforms using targeted keywords, emotion detection through NRCLex, and topic modeling with BERTopic to link emotions with themes like emissions and plastic pollution [20]. The framework is validated through expert annotations and also explores sentiment-driven engagement patterns, revealing negativity bias on Twitter and Facebook and positivity bias on Instagram [14]. This cross-platform analysis over a ten-year period offers deeper insight into evolving public sentiment and emotional responses to environmental issues [17],[19].

Existing sentiment and emotion analysis systems for environmental issues are limited by their reliance on generic tools not tailored to this domain. They mainly use lexicon-based approaches (e.g., VADER, TextBlob, Sent WordNet)[12],[23] and machine learning models (e.g., Naïve Bayes, SVM, RNN), which struggle with domain-specific language, sarcasm, and short-form content typical of social media[11]. Most prior research focuses narrowly on Twitter and English-language posts, limiting demographic and platform diversity. Additionally, these systems often analyse only basic sentiment polarity and overlook deeper emotional nuances. Tools like VADER show reduced accuracy (64%) on environmental datasets compared to domain-specific methods (PMI-based model at 65%) [12].

The study presents a comprehensive framework for sentiment and emotion analysis of environmental content across social media. It employs a PMI-based sentiment classifier that captures domain-specific nuances and outperforms general tools like VADER and spaCy. Data is collected from multiple platforms using targeted keywords to ensure diverse and representative coverage. Emotion analysis using NRCLex identifies key emotions such as fear, trust, and anticipation, particularly in negative posts. BERTopic is used for topic modeling, linking themes like emissions and plastic pollution with associated sentiments and emotions. The model is validated through expert-annotated data, achieving 65% accuracy. The study also examines how sentiment affects user engagement, revealing negativity bias on Twitter and Facebook and a more positive tone on Instagram. A decade-long, cross-platform comparison highlights evolving trends in public sentiment and emotional expression around environmental issues [17],[19].

III. METHODOLOGY

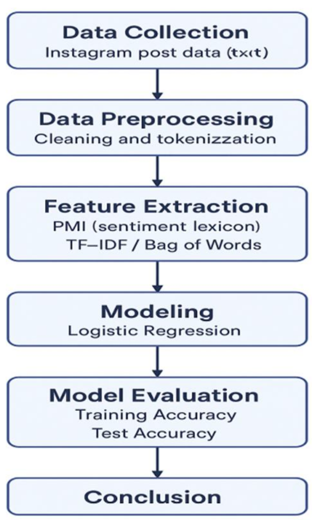


Figure 1: Methodology

A. Data Collection

The sentiment analysis framework uses two main types of datasets: Labelled Data – Pre-annotated social media texts (e.g., tweets) tagged with sentiment values (like 0 for negative, 4 for positive), used to train machine learning models to detect sentiment patterns. Unlabelled Data – Raw text from platforms like Instagram (captions, comments) without sentiment tags. Stored in formats like CSV and analysed using unsupervised methods such as Pointwise Mutual Information (PMI) to derive sentiment lexicons and trends. The quality, diversity, and size of both datasets are crucial, as they directly impact model accuracy, generalization ability, and overall sentiment analysis performance.

B. Training Dataset

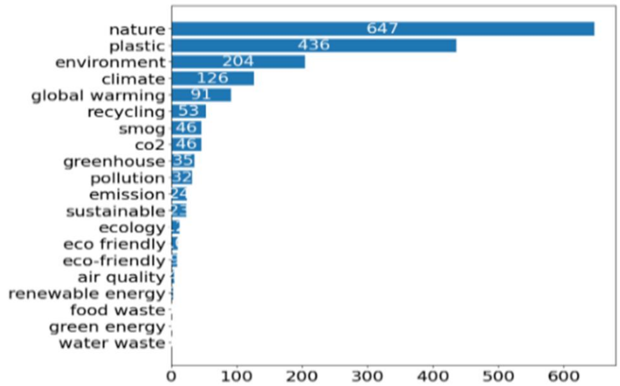


Figure 2: Environmental tweets by the keywords for training dataset

The sentiment analysis model was trained using the Sentiment140 dataset, which contains 1.6 million labelled tweets and is widely used in sentiment research. From an initial 800,000 tweets, a keyword-based filter targeting environmental topics (e.g., *climate*, *pollution*, *green energy*) was applied, resulting in 1,804 relevant tweets (946 positive, 858 negative). This filtered, balanced dataset served as the training data for the sentiment classifier. Notably, the keyword "nature" appeared most frequently (647 times), followed by other prominent terms like "plastic" and "environment", highlighting common themes in environmental discussions.

C. Testing Datasets

The testing dataset comprises pre-processed Instagram captions and comments labelled with true sentiments (positive or negative). It is kept separate from training data to ensure unbiased evaluation. Texts are transformed into numerical features (e.g., TF-IDF, BoW, PMI) and used to calculate metrics like accuracy.

Data Sources:

1) *Twitter*: The study collected 284,440 environment-related tweets from 2013 to 2023 using snscrape, applying consistent environmental keywords to match the training data. To ensure balance and diversity, the dataset was limited to 100 tweets per keyword per month, creating a representative sample of long-term environmental discussions on Twitter.
2) *Instagram*: Instagram is a valuable platform for sentiment analysis due to its high user engagement and abundance of user-generated content such as captions, comments, and hashtags, which effectively capture public sentiment.
3) *Facebook*: Facebook data was gathered from official news outlet pages (e.g., BBC, CNN, Euronews), focusing on environment-related videos from 2014 to 2023. Using custom scripts, the study extracted 5,468 user comments from 1,998 videos, offering valuable insights into public sentiment on environmental topics.

D. Data Analysis

To extract meaningful insights, an initial exploratory analysis was performed, followed by data pre-processing to clean and standardize the content by removing noise and stop words. After refining the data, Sentiment Analysis, Emotion Detection, and Topic Modelling were applied to identify underlying patterns and key themes.

E. Pre-Processing

The data pre-processing workflow involved several transformation steps to prepare raw tweet data for analysis:

1) *Text Cleaning*: Removed links, mentions, hashtags, emojis, and special characters.
2) *Lowercasing*: Converted all text to lowercase for consistency.
3) *Tokenization*: Split sentences into individual words (tokens).
4) *Slang & Abbreviation Handling:* Replaced informal terms with standard language.
5) *Stop Word Removal*: Filtered out common, low-value words like "and" or "the."

F. Sentiment Analysis

Pointwise Mutual Information (PMI) measures the strength of association between words by comparing their joint occurrence against independence. Semantic Orientation (SO) of a word is calculated by comparing PMI with positive and negative word sets, normalized by word frequency.

Words with higher SO scores relate to positive sentiment, lower scores to negative. Overall comment sentiment is computed by aggregating individual word SO scores.

G. Emotion Analysis

Goes beyond polarity to identify specific emotions like joy, anger, fear, sadness, etc. Focused on negative sentiment comments using NRCLex, detecting 8 emotions with an intensity threshold (>0.25) to identify dominant emotions.

H. Topic Modelling

Used BERTopic, a transformer-based method combining embeddings with class-based TF-IDF, to uncover themes in comments across platforms. Separate topic modelling was also performed on comments linked to specific emotions (fear, trust, anticipation) to explore nuanced discussions.

I. Dataset Annotation

Manually annotated 100 environmental tweets on an 11-point scale (-5 to +5) for fine-grained sentiment. Annotation by six articipants, including two domain experts with double weighting. Final scores were weighted averages; outlier annotations were identified but none removed. Sentiment distribution: 44% positive, 56% negative, 0% neutral. Inter-Annotator Agreement (Cohen's Kappa) averaged 0.525, indicating moderate agreement, with experts showing higher consistency.

J. Accuracy Evaluation

Compared the PMI-based model against VADER, spaCy, and Senti tools on the annotated dataset. PMI model achieved 0.65 accuracy, outperforming VADER (0.64), Senti (0.57), and spaCy (0.44). Expert annotator (Ph.D. in ecology) scored 0.90 accuracy, setting an upper performance bound. Highlighted a polarizing tweet about "zero waste" with high annotator disagreement, illustrating the subjective and contextual complexity of environmental sentiment annotation.

IV. EXPERIMENTAL RESULTS

A. Sentiment Detection Results

Twitter: Predominantly negative sentiment consistently over the years, reflecting critical or concerned discourse on environmental issues. Instagram: Increasing positive sentiment over time, indicating more optimistic or supportive conversations. Facebook: Moderate negative sentiment, less intense than Twitter. This shows platform-specific communication styles: Instagram leans positive, while Twitter and Facebook focus more on problems and criticism.

B. Emotion Detection Results

Analysed only negative sentiment comments using NRCLex. Most frequent emotions: fear, trust, anticipation. Twitter showed rising fear, trust, and anticipation, with spikes around 2020 (possibly linked to pandemic-related trust). Facebook saw increased fear, sadness, trust, and anticipation from 2020 onward, linked to media coverage growth. Emotional responses vary with platform engagement and global events.

C. Topic Modelling

Using BERTopic, five major themes emerged across platforms: Climate Change: Dominant topic; mixed views on urgency and policy. Air Quality and Emissions: Focus on transportation and electric vehicles (notably on Facebook). Recycling and Plastic: Debates on plastic usage and individual fvs corporate responsibility. Biodiversity: Mainly on Twitter, concerns over species loss. Energy and Environmental Responsibility: Facebook users criticized political leaders on environmental actions. Themes reflect shared concerns but show nuanced platform differences.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue VIII Aug 2025- Available at www.ijraset.com

D. Positivity Bias and Engagement

Defined “viral” posts by platform-specific engagement thresholds. Instagram: Positive posts receive higher engagement, supporting a positivity bias. Twitter and Facebook: Negative posts attract more attention, amplifying critical or controversial content. Example averages: Twitter negative tweets had more retweets; Instagram positive posts had more upvotes.

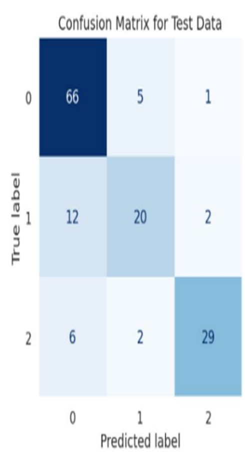
E. Semantic Orientation of Words

Positive associations: “alternative,” “electric” (approval of green tech). Negative associations: “fire,” “oil,” “fossil,” “animal” (ecological harm concerns). Mildly negative: “gas,”

“Pollution,” “bag.” Moderately positive: “clean,” “planet.” This analysis helps understand emotional perceptions of environmental terms.

F. Output

```
          Classification report for training set
-------------------------------------------------------
              precision    recall  f1-score   support

           0       1.00      1.00      1.00       288
           1       1.00      1.00      1.00       135
           2       1.00      1.00      1.00       145

    accuracy                           1.00       568
   macro avg       1.00      1.00      1.00       568
weighted avg       1.00      1.00      1.00       568

           Classification report for test set
-------------------------------------------------------
              precision    recall  f1-score   support

           0       0.79      0.92      0.85        72
           1       0.74      0.59      0.66        34
           2       0.91      0.78      0.84        37

    accuracy                           0.80       143
   macro avg       0.81      0.76      0.78       143
weighted avg       0.81      0.80      0.80       143
```

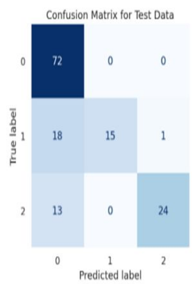


Logistic Regression

Accuracy	80.42%
Precision	81.09%
Recall	76.29%
F1-score	78.08%

Figure3: logistic Regression: Using Bag of words

```
              Classification report for training set
-------------------------------------------------------------
              precision    recall  f1-score   support

           0       0.92      1.00      0.96       288
           1       1.00      0.84      0.91       135
           2       0.99      0.96      0.97       145

    accuracy                           0.95       568
   macro avg       0.97      0.93      0.95       568
weighted avg       0.95      0.95      0.95       568

                Classification report for test set
-------------------------------------------------------------
              precision    recall  f1-score   support

           0       0.70      1.00      0.82        72
           1       1.00      0.44      0.61        34
           2       0.96      0.65      0.77        37

    accuracy                           0.78       143
   macro avg       0.89      0.70      0.74       143
weighted avg       0.84      0.78      0.76       143
```



Logistic Regression

Accuracy	77.62%
Precision	88.63%
Recall	69.66%
F1-score	73.64%

Figure 4: Logistic Regression: Using TF-IDF

```
=== Bag of Words Model ===

📊 Results for Bag of Words
✅ Train Accuracy: 100.00%
✅ Test Accuracy: 90.21%
Classification Report:
              precision    recall  f1-score   support

           0       1.00      0.22      0.36        18
           1       0.90      1.00      0.95       125

    accuracy                           0.90       143
   macro avg       0.95      0.61      0.66       143
weighted avg       0.91      0.90      0.87       143

=== TF-IDF Model ===

📊 Results for TF-IDF
✅ Train Accuracy: 100.00%
✅ Test Accuracy: 88.81%
Classification Report:
              precision    recall  f1-score   support

           0       1.00      0.11      0.20        18
           1       0.89      1.00      0.94       125

    accuracy                           0.89       143
   macro avg       0.94      0.56      0.57       143
weighted avg       0.90      0.89      0.85       143
```

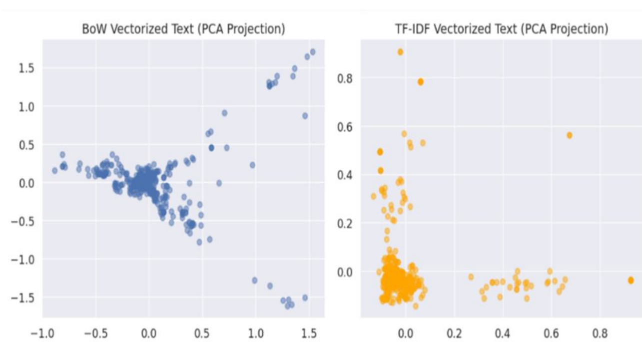


Figure 5: Pointwise Mutual Information: Using Bad of Words and TF-IDF

V. CONCLUSION

This research demonstrates the value of sentiment analysis in uncovering public opinion on environmental issues over a ten-year span using data from multiple social media platforms. The study revealed a dominance of negative sentiment in environmental discussions, with climate change emerging as the most frequently mentioned topic. Other recurring themes included air quality, emissions, plastic waste, and recycling, indicating their continued public relevance. Emotion analysis showed that fear, trust, and anticipation were the most common emotional responses, reflecting the public's complex emotional engagement with these issues. These findings offer valuable insights for policymakers and environmental advocates to shape more effective and resonant communication strategies. However, the study faced challenges in sentiment classification accuracy due to the prevalence of negative tone, and the use of sarcasm and irony in social media posts, which complicated accurate interpretation.

VI. FUTURE SCOPE

Future research could be enhanced through multimodal analysis, combining image and text analysis for a more comprehensive understanding of public sentiment. Additionally, comparing social media sentiments with those from traditional media (e.g., newspapers, TV, radio) could reveal differences or alignments in public opinion, offering a more holistic view of environmental communication across media formats.

REFERENCES

[1] Masson-Delmotte, P. Zhai, H. O. Pörtner, D. Roberts, J. Skea, P. R. Shukla, A. Pirani, W. Moufouma-Okia, C. Péan, R. Pidcock, S. Connors, J. B. R. Matthews, Y. Chen, X. Zhou, M. I. Gomis, E. Lonnoy, T. Maycock, M. Tignor, and T. Waterfield, ''IPCC, 2018: Global Warming of 1.5 ◦C. An IPCC Special Report on the impacts of global warming of 1.5 ◦C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change, sustainable development, and efforts to eradicate poverty,'' Intergovernmental Panel Climate Change (IPCC), Cambridge Univ. Press, Cambridge, U.K., New York, NY, USA, Geneva, Switzerland, p. 616, 2018, doi: 10.1017/9781009157940.

[2] H. Dagevos and J. Voordouw, ''Sustainability and meat consumption: Is reduction realistic?'' Sustainability, Sci., Pract. Policy, vol. 9, no. 2, pp. 60–69, Oct. 2013.

[3] O. Y. Adwan, M. Al-Tawil, A. Huneiti, R. Shahin, A. Abu Zayed, and R. Al-Dibsi, ''Twitter sentiment analysis approaches: A survey,'' Int. J. Emerg. Technol. Learn., vol. 15, no. 15, p. 79, Aug. 2020.

[4] K. L. S. Kumar, J. Desai, and J. Majumdar, ''Opinion mining and sentiment analysis on online customer review,'' in Proc. IEEE Int. Conf. Comput. Intell. Comput. Res. (ICCIC), Dec. 2016, pp. 1–4.

[5] Jumadi, D. S. Maylawati, B. Subaeki, and T. Ridwan, ''Opinion mining on Twitter microblogging using support vector machine: Public opinion about state Islamic University of Bandung,'' in Proc. 4th Int. Conf. Cyber IT Service Manage., Apr. 2016, pp. 1–6.

[6] I. K. C. U. Perera and H. A. Caldera, ''Aspect based opinion mining on restaurant reviews,'' in Proc. 2nd IEEE Int. Conf. Comput. Intell. Appl. (ICCIA), Sep. 2017, pp. 542–546.

[7] V. B. Raut and D. D. Londhe, ''Opinion mining and summarization of hotel reviews,'' in Proc. Int. Conf. Comput. Intell. Commun. Netw., Nov. 2014, pp. 556–559.

[8] A. Jeyapriya and C. S. K. Selvi, ''Extracting aspects and mining opinions in product reviews using supervised learning algorithm,'' in Proc. 2nd Int. Conf. Electron. Commun. Syst. (ICECS), Feb. 2015, pp. 548–552.

[9] M. Wöllmer, F. Weninger, T. Knaup, B. Schuller, C. Sun, K. Sagae, and L.-P. Morency, ''YouTube movie reviews: Sentiment analysis in an audio- visual context,'' IEEE Intell. Syst., vol. 28, no. 3, pp. 46–53, May 2013.

[10] A.-M. Iddrisu, S. Mensah, F. Boafo, G. R. Yeluripati, and P. Kudjo, ''A sentiment analysis framework to classify instances of sarcastic sentiments within the aviation sector,'' Int. J. Inf. Manage. Data Insights, vol. 3, no. 2, Nov. 2023, Art. no. 100180.

[11] H. Almerekhi, H. Kwak, and B. J. Jansen, ''Investigating toxicity changes of cross-community redditors from 2 billion posts and comments,'' PeerJ Comput. Sci., vol. 8, p. e1059, Aug. 2022.

[12] E. Shamoi, A. Turdybay, P. Shamoi, I. Akhmetov, A. Jaxylykova, and A. Pak, ''Sentiment analysis of vegan related tweets using mutual information for feature selection,'' PeerJ Comput. Sci., vol. 8, p. e1149, Dec. 2022.

[13] A. R. Pratama and F. M. Firmansyah, ''COVID-19 mass media coverage in English and public reactions: A west-east comparison via Facebook posts,'' PeerJ Comput. Sci., vol. 8, p. e1111, Sep. 2022.

[14] M. O. Faruk, P. Devnath, S. Kar, E. A. Eshaa, and H. Naziat, ''Perception and determinants of social networking sites (SNS) on spreading awareness and panic during the COVID-19 pandemic in Bangladesh,'' Health Policy Open, vol. 3, Dec. 2022, Art. no. 100075.

[15] A. Giachanou, I. Mele, and F. Crestani, ''Explaining sentiment spikes in Twitter,'' in Proc. 25th ACM Int. Conf. Inf. Knowl. Manage., Oct. 2016, pp. 2263–2268.

[16] B. Sluban, J. Smailovic, M. Juric, I. Mozetic, and S. Battiston, ''Com-munity sentiment on environmental topics in social networks,'' in Proc. 10th Int. Conf. Signal-Image Technol. Internet-Based Syst., Nov. 2014, pp. 376–382.

[17] E. Rosenberg, C. Tarazona, F. Mallor, H. Eivazi, D. Pastor-Escuredo, F. Fuso-Nerini, and R. Vinuesa, ''Sentiment analysis on Twitter data towards climate action,'' Results Eng., vol. 19, Sep. 2023, Art. no. 101287.

[18] C. Cubukcu-Cerasi, Y. S. Balcioglu, A. Kilic, and F. Huseynov, ''Embracing green choices: Sentiment analysis of sustainable consumption,''Eurasia Proc. Sci. Technol. Eng. Math., vol. 23, pp. 253–261, Oct. 2023.

[19] F. Jost, A. Dale, and S. Schwebel, ''How positive is 'change' in climate change? A sentiment analysis,'' Environ. Sci. Policy, vol. 96, pp. 27–36, Jun. 2019.

[20] B. Dahal, S. A. P. Kumar, and Z. Li, ''Topic modeling and sentiment analysis of global climate change tweets,'' Social Netw. Anal. Mining, vol. 9, no. 1, pp. 1–20, Dec. 2019.

[21] M. Lineman, Y. Do, J. Y. Kim, and G.-J. Joo, ''Talking about climate change and global warming,'' PLoS One, vol. 10, no. 9, Sep. 2015, Art. no. e0138996.

[22] W. Shi, H. Fu, P. Wang, C. Chen, and J. Xiong, ''Climatechange vs. globalwarming: Characterizing two competing climate discourses on Twitter with semantic network and temporal analyses,'' Int. J. Environ. Res. Public Health, vol. 17, no. 3, p. 1062, Feb. 2020.

[23] T. E. Taufek, N. F. M. Nor, A. Jaludin, S. Tiun, and L. K. Choy, ''Public perceptions on climate change: A sentiment analysis approach,'' GEMA Online J. Lang. Stud., vol. 21, no. 4, pp. 209–233, Nov. 2021.

[24] L. Rocca, D. Giacomini, and P. Zola, ''Environmental disclosure and sentiment analysis: State of the art and opportunities for public-sector organisations,'' Meditari Accountancy Res., vol. 29, no. 3, pp. 617–646, Jun. 2021.

[25] Y. Tao, F. Zhang, C. Shi, and Y. Chen, ''Social media data-based sentiment analysis of tourists' air quality perceptions,'' Sustainability, vol. 11, no. 18,p. 5070, Sep. 2019.