RESEARCH ARTICLE

Understanding Environmental Posts: Sentiment and Emotion Analysis of Social Media Data

DANIYAR AMANGELDI[1], **AIDA USMANOVA**[2], **AND PAKIZAR SHAMOI**[1]

[1]School of Information Technology and Engineering, Kazakh-British Technical University, Almaty 050000, Kazakhstan
[2]Institute of Information Systems, Leuphana University Lüneburg, 21335 Lüneburg, Germany

Corresponding author: Pakizar Shamoi (p.shamoi@kbtu.kz)

ABSTRACT Social media is now the predominant source of information due to the availability of immediate public response. As a result, social media data has become a valuable resource for comprehending public sentiments. Studies have shown that it can amplify ideas and influence public sentiments. This study analyzes the public perception of climate change and the environment over a decade from 2014 to 2023. Using the Pointwise Mutual Information (PMI) algorithm, we identify sentiment and explore prevailing emotions expressed within environmental tweets across various social media platforms, namely Twitter, Reddit, and YouTube. Accuracy on a human-annotated dataset was 0.65, higher than Vader's score but lower than that of an expert rater (0.90). Our findings suggest that negative environmental tweets are far more common than positive or neutral ones. Climate change, air quality, emissions, plastic, and recycling are the most discussed topics on all social media platforms, highlighting its huge global concern. The most common emotions in environmental tweets are fear, trust, and anticipation, demonstrating the wide and complex nature of public reactions. By identifying patterns and trends in opinions related to the environment, we hope to provide insights that can help raise awareness regarding environmental issues, inform the development of interventions, and adapt further actions to meet environmental challenges.

INDEX TERMS Sentiment analysis, emotion analysis, social media, public perception, climate change, global warming, pointwise mutual information, Twitter, Reddit, YouTube.

I. INTRODUCTION

Environmental issues are among the most pressing challenges facing society today. In 2018, the United Nations IPCC issued a report warning of a climate change catastrophe within 12 years [1]. The crucial need for environmental sustainability in the current era of climate change is stated by recent reports [1] and research on sustainability [2].

Social media platforms are crucial for environmental advocacy, as they enable individuals and organizations to share information, raise awareness, and mobilize support for common goals. Analyzing social media data involves exploring thoughts and opinions on various domains [3].

The associate editor coordinating the review of this manuscript and approving it for publication was Seifedine Kadry.

Analyzing social media data is useful to understand people's opinions on various topics. Understanding the sentiment and emotion expressed in environment-related posts is important for several reasons. On social media, sentiment analysis can identify key issues and concerns and reveal patterns of public opinion and attitudes towards environmental issues. Analyzing sentiment and emotion can help identify factors that encourage participation in environmental discussions on social media. Therefore, this aims to address the gap in knowledge regarding the sentiment and emotion expressed in the comments related to the environment.

This study aims to reveal the public perception of environmental problems. The objective will be achieved by collecting data from popular social media platforms over the last decade, including Twitter, Reddit, and YouTube. The textual data will

be analyzed to understand the prevailing emotions related to environmental issues over the past ten years and the factors influencing public attitudes toward ecological awareness. Additionally, the study explores whether using specific social media networks impacts the emotional background of users.

We aim to answer the following questions using textual data from Twitter, Reddit, and YouTube in the period from 2013 to 2023:

- How has the public perception of environmental problems changed over the decade of data?
- What were the prevailing emotions and topics associated with this change?
- Does the utilized social media network influence emotional background and promote distinct behavior?

The main contributions of the study may be summarised as follows:

- *Use of Multiple Social Media Platforms.* We provide a comprehensive analysis of user opinions and discussions on environment-related topics by utilizing data from multiple social media platforms like Reddit, YouTube, and Twitter. The multi-platform approach enables one to account for each platform's diverse user demographics and communication styles.
- *Analysis of Posts Over a Decade.* Analyzing posts over a decade allows us to examine trends and changes in public opinion regarding the environment, ecology, and global warming. We captured the evolution of discussions, the impact of significant events or policy changes, and the shifting attitudes and awareness among the online community.
- *Emotion Analysis besides Sentiments.* The emotion analysis complements the sentiment analysis to provide a deeper understanding of the emotional experiences and affective reactions associated with environmental discussions.
- *Holistic Understanding of Public Perception.* Analysis was done using **popular** posts from these social networks.

The paper has been structured in the following way. This Introduction is Section I. Section II contains an overview of the literature on sentiment analysis and opinion-mining research. Section III is concerned with the methodology used for this study. Data collection and description are covered in Section IV. Section V presents the findings of the research. Next, the Discussion is presented in Section VI. Finally, Section VII provides concluding remarks and recommendations for future enhancements to the methodology.

II. RELATED WORK
A. OPINION MINING IN SOCIAL MEDIA
1) OVERVIEW
Social media has become a significant platform for public discussions and opinions on various topics, including ecology and the environment. It is capable of shaping public opinions and amplifying ideas.

Sentiment Analysis, also known as Opinion mining, is one of the essential methods for understanding public views and getting insights into current trends. Such analysis has proven useful for further decision-making in various domains. Studies like [4] and [5] classified public opinions into negative and positive, based on Amazon reviews and Twitter comments, respectively. Sentiment analysis conducted by [6] and [7] studied restaurant and hotel reviews to generate personalized review recommendations. Reference [8] analyzed product reviews to acquire valuable information for marketing analysis. Another related work uses a similar approach for monitoring YouTube movie reviews [9]. Reference [10] developed a framework that supports airlines in addressing customer complaints and improving services during global events like the COVID-19 pandemic through social media sentiment analysis focusing on sarcasm detection. Reference [11] analyzed 2 billion posts and comments from Reddit to identify toxic comments. The authors hope to bring more awareness to the online harassment problem experienced by many people nowadays and potentially prevent toxic behavior on social networks. The other recent study analyzed public sentiment regarding the vegan diet using Twitter data [12]. It finds that public perception has a growing positive trend despite persistent fears associated with veganism. These insights have important implications for health programs, government initiatives, and efforts to reduce veganism-related negative emotions.

Studies like [13] and [14] analyzed public reactions to COVID-19, providing insights into sentiment patterns during the pandemic. A paper conducted by [15] highlights the significance of monitoring public sentiment in social media for decision-making processes and emphasizes the importance of understanding the causes of sentiment spikes, with a focus on extracting relevant topics using the Latent Dirichlet allocation (LDA) method.

2) ENVIRONMENTAL TWEETS
The literature on sentiment analysis of social media data related to environmental issues has shown that it can provide insights into the public's attitudes and emotions toward various environmental topics. Sentiment analysis has been found to align with attitudes towards renewable energy, sustainability, and pollution [16].

Specific terms have been indicated to carry a high percentage of negative sentiment in social media discussions, such as topics of CO_2 and fracking, and expressions like "climate impact," "gas emission," and "charcoal," which consistently emerge with high levels of negative sentiment [17]. In contrast, research by [18] unveiled a positive trend in discussions about sustainable consumption and a higher number of positive sentiments surrounding climate change [19]. Yet, when there is significant climate change discussion on Twitter, the discourse tends to be notably negative [20]. Users tend to express negative sentiments about

the term "global warming" rather than "climate change" [21], [22], [23].

Social media information, namely Twitter comments, has been used successfully to analyze public sentiment toward air pollution [24], [25], [26]. Over 50% of comments express a sentiment of dissatisfaction in the environmental tweets [24]. The studies also prove that sentiments are consistently aligned with the actual air quality, hence public perception in social media has the potential for air quality forecasting [25]. Such a conclusion suggests that social media can function as an early warning system for natural disasters, with users providing real-time information and expressing concerns about the impact on their daily lives.

A study conducted by [27] and [28] performed a sentiment analysis on social media posts to track people's emotional responses towards river pollution. The research demonstrates the potential of social media data for tracking emotional responses to environmental issues. The findings indicate that people tend to express more negative emotions than positive ones when discussing river pollution.

Further studies demonstrated the effectiveness of social media information during environmental challenges, disasters, and public resilience toward environmental actions [29], [30], [31], [32]. Reference [31] performed sentiment analysis of tweets to derive the status of floods. Reference [32] established a correlation between wildfires in Spain and Portugal and the well-being of individuals expressed through sentiments.

Recent studies like [30] and [33] indicate that the analysis of sentiments in Twitter data can effectively reveal public attitudes towards environmental actions. This is demonstrated by concerns surrounding field experiments, including potential adverse environmental impacts, the risk of progressing towards extensive implementation, and apprehension surrounding poor governance or even the illegality of such projects.

Various methods have been employed to analyze environmental tweets. Several studies show that document embeddings combined with neural networks are effective for environment-related tweets. For example, a recent study by [34] presents a data processing pipeline combining traditional methods and deep learning, such as VADER, TextBlob, and LSTMs, specifically for environmental tweets. A similar pipeline approach involving Naïve Bayes and RNN was employed in a recent study on Sentiment Analysis on global warming-related tweets [35].

Another study focused on the sentiments and emotions of 34 opinion leaders about global warming, using text mining and classification [36]. Their findings suggest that examining the posts of the opinion leaders is effective considering their influence. In another research, Latent Dirichlet Allocation (LDA) and SentiWordNet, a broad sentiment vocabulary, were employed to examine the sentiments to see how different publications approach stories about climate change differently [37]. Their findings reveal diverse perspectives on the same problem offered in different news sources.

Several works in environmental tweets analysis used emotion detection besides sentiments [36], [38], [39], [40].

Previous studies performed sentiment analysis of environmental issues on generic, domain-independent textual data. Such an approach lacks domain and context-specific training, which could limit capturing all public sentiments towards environmental issues. An obstacle in analyzing sentiments on social media is the challenge of identifying the root causes of particular emotions due to the limited length of posts. Though sentiment analysis can provide useful information on the overall emotional reactions of the public toward environmental problems, it may not be adequate to comprehend the underlying reasons behind these emotions completely. Further analysis, such as topic modeling or emotion analysis, can help address this challenge.

Although sentiment analysis in the natural environment context is a significant field, most research within this domain relies on English-language data and Twitter datasets [41]. Our study uses only English comments from social media. However, adding Reddit and YouTube comments can expand user demographics and reveal new trends.

Further research is needed to develop domain-specific sentiment analysis models that better capture public sentiment towards environmental issues. Improved understanding of public perception can inform environmental decision-making. Therefore, this study aims to investigate people's emotional responses to environmental issues by analyzing social media data, specifically Twitter, Reddit, and YouTube.

B. SENTIMENT ANALYSIS METHODS

Sentiment analysis methods fall into two categories: Lexicon-based and Machine Learning (ML)-based approaches [42], [43], [44]. Lexicon-based methods can further be split into dictionary-based techniques and corpus-based methods. These methods rely on predefined dictionaries to assess sentiment based on positive, negative, or neutral words and/or employ statistical models based on large text datasets to understand the context. ML-based methods can be divided into Supervised and Unsupervised Learning. Supervised Learning uses labeled data to train classification models like SVM, KNN, DTC, and LR. Unsupervised Learning, on the other hand, uncovers patterns in data using clustering, topic modeling, and mapping algorithms [45], [46], [47].

Even though there are other techniques for building sentiment analysis models, we picked the point-wise mutual information (PMI) approach (it is Lexicon-based) proposed by [48], [49], [50], [51] for its interpretability and robustness to statistical bias in small sample sizes [12]. In NLP applications, the PMI or MI evaluates the chance of two-word co-occurrence relative to the random probability, adding greater meaning to the semantic proximity of the terms. By calculating PMI, sentiment analysis algorithms can better understand the contextual and semantic relationships between words and sentiments, thus improving the accuracy of sentiment classification and providing more nuanced
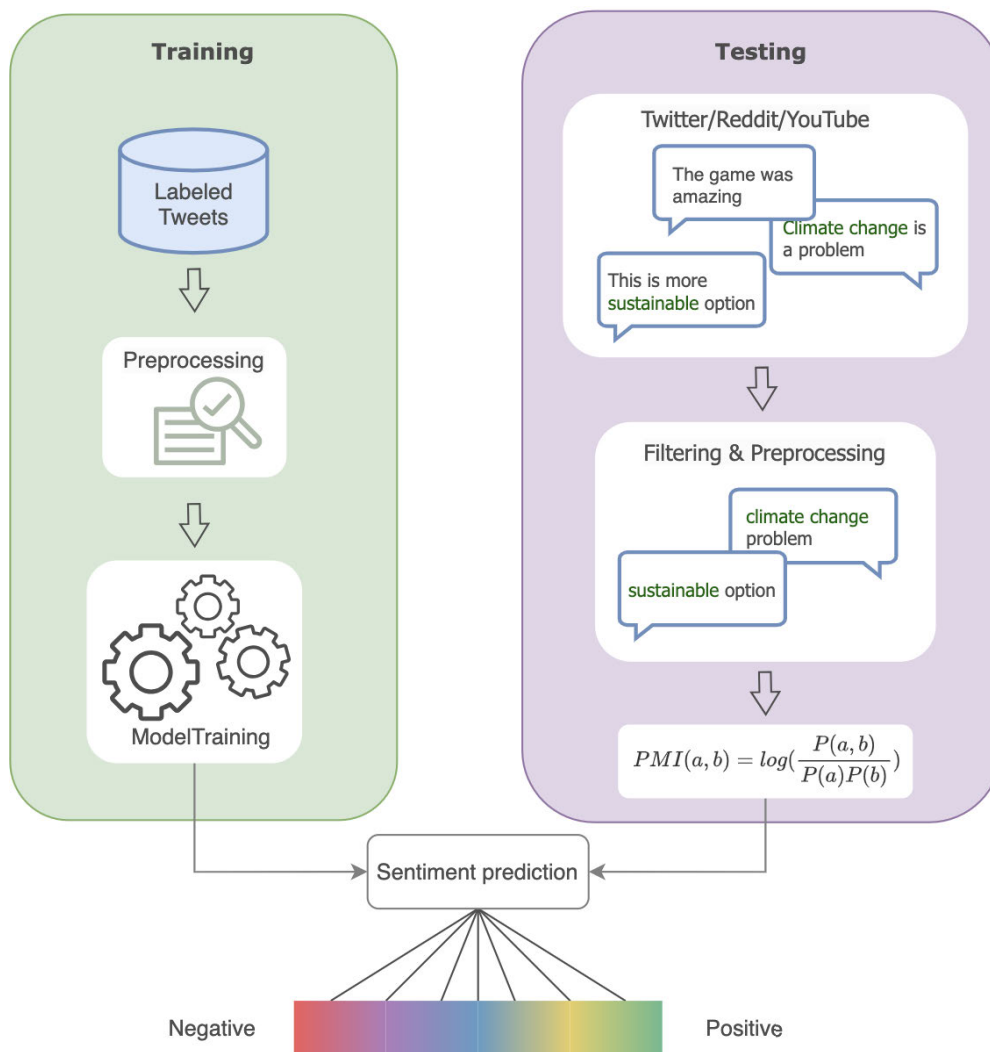
FIGURE 1. Study workflow: The work consists of two parts. Training is done on the labeled tweets dataset. Testing is performed on web-scraped data from Twitter, Reddit, and YouTube. The trained model is then applied to test data to generate sentiment prediction scores for each comment.

insights into text data [12]. Another reason for employing the aforementioned approach is that it will be used for feature selection in this study. The principles of this approach were initially referred to as mutual information (MI) [52], [53].

Our goal in employing the strategy is to evaluate how well the words intended to be connected with particular sentiment classes (PMI measures) can function as features while building the sentiment classifier model. The PMI-based method of sentiment analysis is employed in a lot of research [54], [55], [56], [57], [58], [59], [60], [61], [62].

III. METHODS

The schematic representation of the methodology in Figure 1 shows the steps involved in the sentiment analysis process. The methodology includes data collection, pre-processing, training a model, and applying the algorithm for sentiment analysis.

A. DATA COLLECTION

Two datasets are required to perform sentiment analysis on environmental posts: the training and testing datasets. The labeled dataset contains pre-annotated tweets with sentiment labels (0 for negative and 4 for positive), which are used to train the sentiment analysis model. The unlabelled dataset is scraped from social media networks and is used for further analysis. The aim is to use the knowledge gained from the labeled training dataset to create a model that can accurately predict the sentiment of new, unlabelled tweets in the testing dataset.

1) TRAINING DATASET

To train our sentiment analysis model, Sentiment140 dataset[1] was utilized. The dataset is extensively utilized in research

---

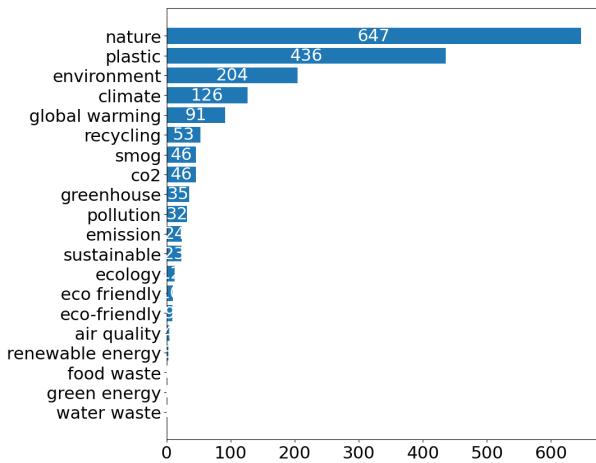[1] https://www.kaggle.com/datasets/kazanova/sentiment140

FIGURE 2. Environmental tweets by the keywords for training dataset.

on sentiment analysis due to its vast size and diverse content. It comprises a vast collection of textual information gathered from Twitter, covering various topics and sentiments. The dataset comprises around 1.6 million tweets, each labeled with a positive or negative sentiment polarity. This makes it possible to conduct supervised training of the model.

The dataset was filtered based on the following keywords: *"climate"*, *"global warming"*, *"environment"*, *"nature"*, *"pollution"*, *"plastic"*, *"green energy"*, *"food waste"*, *"water waste"*, *"greenhouse"*, *"recycling"*, *"air quality"*, *"eco-friendly"*, *"emission"*, *"renewable energy"*, *"sustainable"*, *"zero waste"*, *"carbon dioxide"*, *"ecology"*, *"smog"*, *"biodiversity"*. We collected 1804 environmental tweets from a dataset of 800,000 instances, equally divided between positive and negative sentiments. The subset of filtered tweets consists of an almost equal number of tweets: 946 positive and 858 negative. Based on the Figure 2, among the keywords, *nature* emerges as the most prominent, dominating the conversation with 647 tweets for training, followed by *"plastic"* and *"environment"*.

2) TESTING DATASETS

To perform comprehensive sentiment analysis, we scraped textual information from 3 popular social media platforms: Twitter, Reddit, and YouTube. To ensure the data collection process is consistent across all platforms, we used the same search keywords(mentioned in Section III-A1) to gather relevant posts from each platform.

a: TWITTER

Twitter is a popular micro-blogging platform with 1.3 billion users who send out 500 million tweets daily [63]. Scraping environmental tweets from 2013 to 2023 was performed with snscrape[2] Python library. Tweets were filtered based on the keywords applied in filtering training data. To ensure a balanced representation over time and manage the size of the

dataset, we limited the collection to 100 tweets per keyword each month. This approach provides a diverse and relevant set of environmental tweets that can be further analyzed and evaluated. In total, 284,440 environmental tweets for analysis and evaluation were retrieved.

b: REDDIT

To obtain a testing dataset from Reddit, the PRAW[3] (Python Reddit API Wrapper) library was utilized to scrape environmental posts from the period 2013 to 2023. The same set of keywords was utilized for subreddit searches. To ensure temporal balance, the collection was limited to a maximum of 100 posts per keyword monthly, which resulted in 38,251 environmental Reddit posts.

c: YOUTUBE

To mitigate bias and ensure more heterogeneous public feedback, we scraped comments from popular news channels on YouTube, namely Euronews, CNN, Sky News, BBC, NBC, CBC, and ABC. We collected data from YouTube by sending requests to the website and obtaining links to the relevant videos. To get the videos relevant to the topic, we manually selected Playlists from the channels related to the environment, e.g., climate change, global warming, etc. Once we got the links, we sent requests to the respective websites and scraped 100 comments containing relevant keywords in the content. Selected videos and scraped comments were published within the 2014 and 2023 time frames. Overall, we retrieved 1998 relevant videos with 5468 relevant comments.

B. DESCRIPTION OF ENVIRONMENTAL POSTS AND COMMENTS

Table 1 illustrates the evolution of engagement trends over the years across Twitter, Reddit, and YouTube. It becomes apparent that each platform exhibits distinct patterns in the proportion of popular posts. For instance, while Reddit consistently maintains a high percentage of popular posts (over 95%), Twitter and YouTube steadily increased, suggesting a higher user engagement.

Defining what constitutes a popular post across different social media platforms is a crucial aspect of our analysis because each platform has its own metrics for engagement, and setting criteria for popularity allows for consistent evaluation. Here is how we define a *popular post* for:

- Twitter: at least one like
- Reddit: at least one upvote
- YouTube: at least one like

We will use popular posts from our scraped data from different social network systems for further analysis.

Next, Figure 3 presents word clouds of text from posts scrapped from Twitter, Reddit, and YouTube. It could be noted that the absence of visually dominant words in YouTube comments within the word cloud indicates a
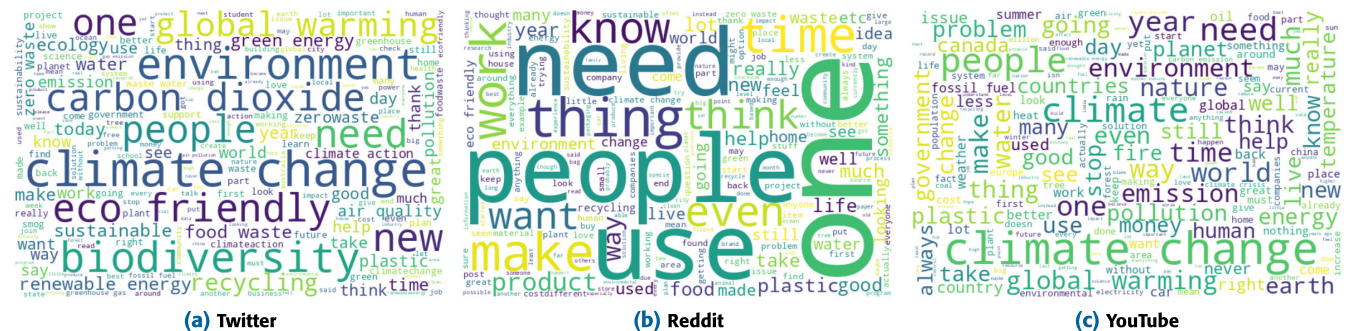
---

[2]https://github.com/JustAnotherArchivist/snscrape

[3]https://praw.readthedocs.io/en/stable/

FIGURE 3. Word clouds for *popular* posts in social media.

TABLE 1. Social media popularity.

	Twitter			Reddit			YouTube		
Year	Total	Popular	Popular, %	Total	Popular	Popular,%	Total	Popular	Popular,%
2013	27600	2553	9.25	624	595	95.35	0.0	0.0	0.00
2014	27600	4535	16.43	705	671	95.18	1.0	1.0	100.00
2015	27600	5535	20.05	650	638	98.15	7.0	5.0	71.43
2016	27559	7648	27.75	916	899	98.14	40.0	26.0	65.00
2017	27600	9016	32.67	1122	1111	99.02	72.0	54.0	75.00
2018	27526	11019	40.03	2069	2030	98.12	75.0	38.0	50.67
2019	27075	11628	42.95	3785	3673	97.04	420.0	267.0	63.57
2020	26632	12399	46.56	4521	4384	96.97	205.0	143.0	69.76
2021	26310	13439	51.08	5630	5469	97.14	476.0	356.0	74.79
2022	27450	13159	47.94	7634	7425	97.26	2060.0	1640.0	79.61
2023	11488	5514	48.00	8895	8674	97.52	2112.0	1613.0	76.37

diverse range of discussions. Conversely, in the context of environmental posts on Reddit, dominant words like *"people"* in the word cloud suggest that discussions on environmental issues often intertwine with human-related aspects. This could imply a focus on how environmental problems impact people directly or indirectly, such as through policies, lifestyle changes, activism, or societal impacts. In the word cloud generated from Twitter posts about the environment, there are dominant words like *"environment"*, *"carbon dioxide"*, *"climate change"* and *"eco-friendly"*. These terms represent key focal points in discussions on Twitter regarding environmental issues.

a: TWITTER

After analyzing the language distribution of the collected tweets, we discovered that 89% of the tweets were in English. This indicates that the English language is predominantly used in environmental discussions. Japanese accounted for approximately 3% of the tweets, followed by French and Spanish with 2% each. The remaining 5% consisted of tweets in different languages, including those with fewer than 3,000 instances. This language breakdown provides valuable

insights into the linguistic composition of the environmental discourse captured in the testing dataset.

As per the data presented in Table 2, it can be observed that around 60% of the tweets in the dataset did not receive any likes, replies, retweets, or quotes. This indicates that more than half of the tweets in the dataset had limited visibility or did not resonate well with the audience, resulting in minimal engagement. However, by analyzing the tweets that received at least one like, we can gain insights into the engagement and interaction patterns of tweets that have gathered some level of attention from users.

Figure 4 shows how often certain words were used in a collection of popular tweets about the environment. The analysis reveals that the three most commonly used words were "biodiversity", "climate action", and "ecology". These words are important because they represent key themes in environmental and sustainability discussions. They are popular on Twitter because they align with current global environmental concerns and sustainability efforts. The fact that these words are frequently used shows that people are recognizing the need to protect biodiversity and take action to maintain ecology. This makes them highly relevant in environmental discussions.

TABLE 2. Engagement metrics of environmental tweets.

	replyCount	retweetCount	likeCount	quoteCount
count	284440.00	284440.00	284440.00	284440.00
mean	0.39	1.41	4.10	0.10
std	8.22	54.48	235.69	3.71
min	0.00	0.00	0.00	0.00
50%	0.00	0.00	0.00	0.00
70%	0.00	0.00	1.00	0.00
90%	1.00	1.00	3.00	0.00
max	2241.00	17840.00	111318.00	1560.00
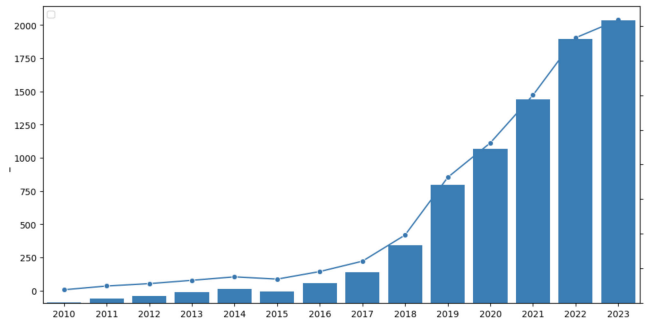
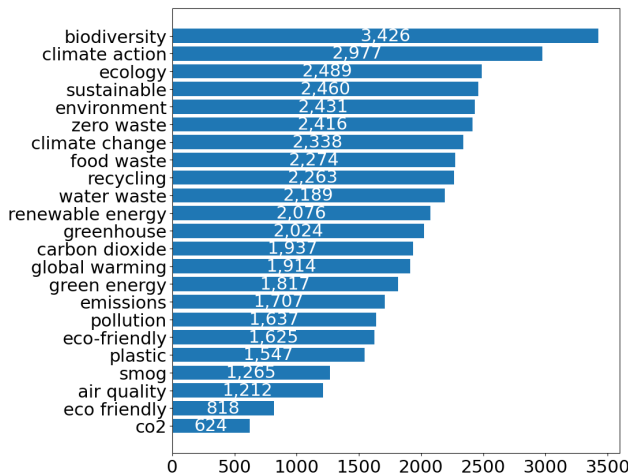**FIGURE 6.** Number of popular environmental Reddit posts over time.

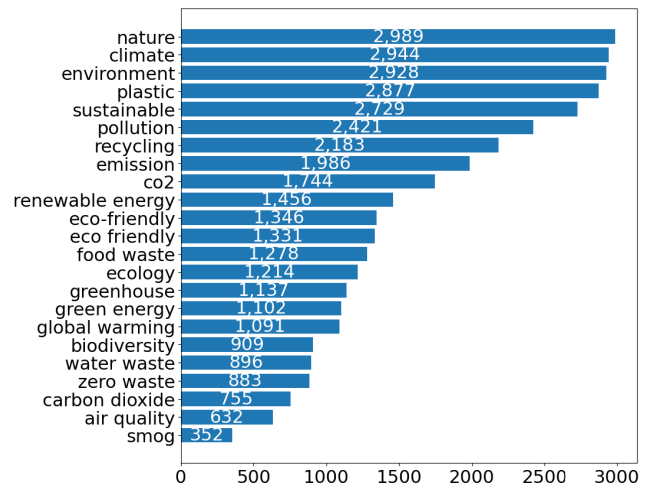**FIGURE 4.** Popular environmental tweets by the keywords.

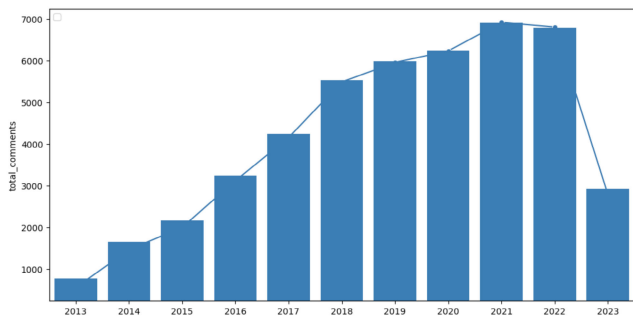**FIGURE 7.** Popular Reddit posts by the keywords.

FIGURE 5. Number of popular environmental tweets over time.

The number of popular environmental tweets has been increasing over the years. According to Figure 5, the count has risen from 556 in 2013 to 6,916 in 2021. This upward trend can be attributed to the growing accessibility and prevalence of social media platforms. As more people join these platforms and engage in online conversations, the opportunity to share and discuss environmental topics becomes more widespread. This, in turn, contributes to the overall increase in the number of environmental tweets.

b: REDDIT

According to Figure 6, the evolution of Reddit posts over time has a noticeably upward trend, showing a significant increase in posts over the years. From a modest start in 2008 with 85 posts, the numbers gradually rose, with a more rapid increase observed after 2017. The consistent rise in the number of posts reflects a growing interest and engagement in the topic of environment. Despite Reddit scraping for only half of 2023, the popularity of posts in this period still surpasses that of previous years, indicating a sustained trend of high engagement and interaction on the platform throughout 2023.

With 38,251 posts analyzed, the average number of upvotes per post stands at approximately 157. More than half of the posts received at least 15 upvotes, showing the community's interest in environmental topics. The presence of posts with exceptionally high upvotes, reaching up to 48,700, indicates the existence of standout content that captures widespread attention and engagement of society.

Figure 7 demonstrates the frequency of popular Reddit posts by keywords. The high frequency of keywords like *"climate"*, *"environment"* and *"nature"* suggests a significant Reddit focus on broad environmental topics. Additionally, terms like *"sustainable"* and *"renewable energy"* indicate
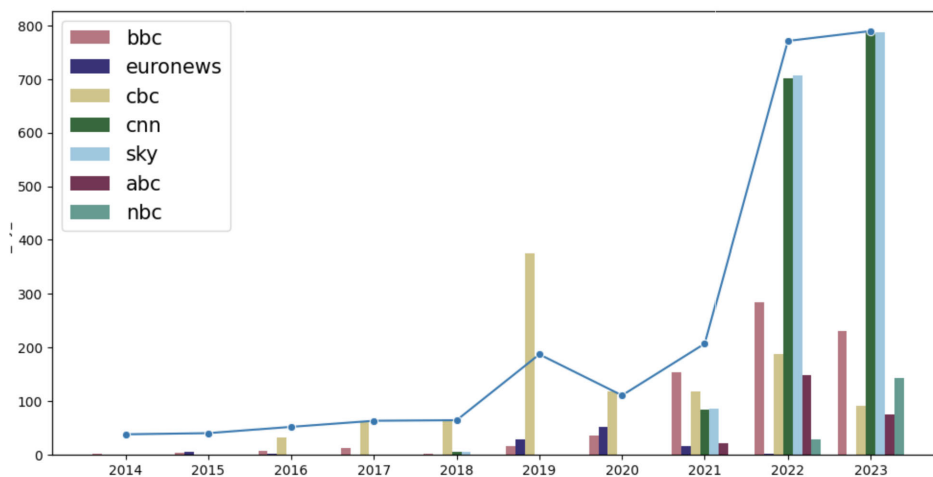
FIGURE 8. Number of comments on climate change and environment under YouTube videos from 2014 to 2023.

a growing interest in ecologically clean practices within the community.

c: YOUTUBE

The scraped comments from YouTube were not distributed evenly. Most comments were gathered from Sky News, CNN, and CBC channels, with 1584, 1577, and 1045 comments, respectively. The remaining four channels correspond to BBC 745, ABC 244, NBC 172, and Euronews 101 comments Out of 5468 gathered comments, 4151 received at least one vote, which was considered in the further analysis.

As seen from Figure 8 the number of discussions on environmental topics has constantly risen since 2014. There is a sudden increase in the number of comments from the CBC channel, likely triggered by Australian wildfires in 2019. The overall trend was then followed by an observable decline in 2020, which could be associated with the public focus shifting toward the COVID-19 pandemic. In the past two years, the numbers have drastically increased, with the environment being one of the hot topics of discussion.

C. DATA ANALYSIS

We perform the following data analysis to understand the nature of our data better and remove any non-relevant information. The pre-processing step includes noise reduction, standardization, stop-words removal, etc. Once we get the clean data, we perform Sentiment and Emotion Analysis with Topic Modeling.

1) PRE-PROCESSING

Data pre-processing involves several steps to transform raw textual data into a format suitable for analysis (see Figure 9):

- *Cleaning*: When cleaning tweets, irrelevant elements such as URLs, special characters, hashtags, and mentions are removed to ensure only relevant content remains.

- *Case Folding*: Text converted to lowercase to standardize text and avoid word duplication.
- *Tokenization*: Breaking down sentences into individual words or tokens facilitates further analysis and processing by separating each word. This step also helps remove punctuation and split hashtags or compound words into meaningful units.
- *Slang Lookup*: Social media texts often contain slang words and abbreviations. These slang words and abbreviations are replaced with their corresponding full forms or standard equivalents to make the text easier to understand. This step helps to improve the readability and comprehensibility of the text.
- *Stopwords Removal*: Stopwords are common words in a language that don't carry significant meaning. They are removed in textual data pre-processing to reduce noise and focus on meaningful content.

After completing the steps of textual data pre-processing, the raw comments and posts are transformed into a clean and standardized format ready for further analysis and interpretation. [64].

2) SENTIMENT ANALYSIS

This study used Pointwise Mutual Information (PMI) to measure the association between words and their sentiment orientations. PMI calculates the statistical dependence between two words by comparing their co-occurrence in a given corpus with their individual occurrences. Specifically, PMI measures the logarithm of the ratio between the observed co-occurrence probability of two words and the expected probability if they were independent. This helps us understand how closely related two words are in terms of their sentiment orientation [12] (as seen from Equation 1).

$$\text{PMI}(w_1, w_2) = \log\left(\frac{\text{P}(w_1, w_2)}{\text{P}(w_1) \cdot \text{P}(w_2)}\right) \quad (1)$$
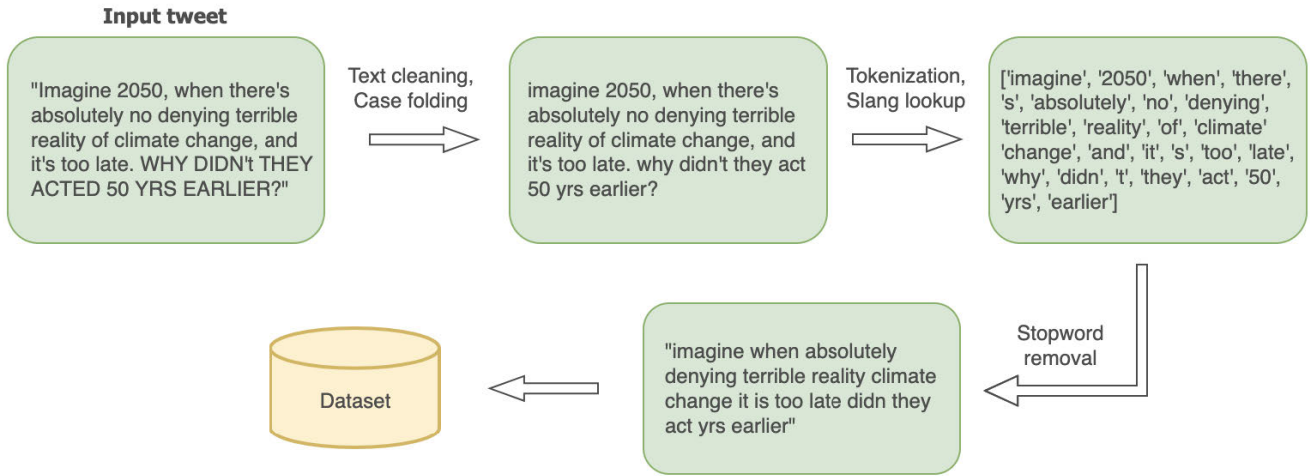
Input tweet

FIGURE 9. Comment pre-processing steps. Firstly, the texts are brought to lowercase, and various special characters are removed. This is followed by tokenization and transforming slang into full forms. As a last step, stopwords are removed from comments, creating a clean dataset.

where $PMI(w_1, w_2)$ represents the observed co-occurrence probability of $word_1$ and $word_2$, and $P(w_1)$ and $P(w_2)$ represent their individual occurrence probabilities.

To determine the semantic orientation of a word, we use the PMI scores between the target word and a set of positive (p) and negative (n) sentiment words. The semantic orientation (SO) is calculated by subtracting the accumulated PMI scores with negative sentiment words from the accumulated PMI scores with positive sentiment words and then dividing the result by the frequency of the target word. This formula is shown in Equation 2.

$$SO(w) = \frac{\sum_{p \in P} PMI(w, p) - \sum_{n \in N} PMI(w, n)}{\text{word freq.}(w)} \quad (2)$$

where P and N represent positive and negative sentiment words, respectively, and word freq.(word) represents the frequency of the target word in the dataset.

This approach allows us to capture the sentiment associations of individual words based on their co-occurrence patterns with positive and negative sentiment words, providing insights into the semantic orientation of the words in our sentiment analysis.

Finally, the sum of individual sentiment scores results in a sentiment score of a comment (as seen from Equation 3), which is a numerical measure where a positive score indicates a positive sentiment, a negative score suggests a negative sentiment and a score of zero shows a lack of strong emotional tone.

$$\text{CommentSentiment(C)} = \sum_{c \in C} SO(c) \quad (3)$$

3) EMOTION ANALYSIS

While sentiment analysis focuses on the polarity of opinions (positive, negative, neutral), emotion analysis helps to dive into the specific emotional states expressed in the posts (e.g., happiness, sadness, anger, fear, surprise). Analyzing

emotional experiences and affective reactions associated with environmental discussions enables us to capture a more nuanced understanding of public opinion. NRCLex [65] is utilized to identify the emotional effect of comments.

In this study, we decided to focus on comments classified as negative in the Sentiment Analysis step. NRCLex is applied to observe the emotion distribution among filtered comments. NRCLex contains 11 emotions, out of which we used only 8, removing *positive*, *negative*, and *anticip*. The emotion intensity range is between 0 and 1. Each comment consists of a combination of various emotions with one prevailing emotion. We considered an emotion prevailing if the intensity score was over 0.25.

We use the following emotions from NRCLex: *fear, anger, anticipation, trust, surprise, sadness, disgust, joy*. We excluded positive and negative emotions as we used our sentiment classifier.

4) TOPIC MODELING

Topic Modeling clusters information into bigger groups and helps to identify present topics in textual datasets. BERTopic [66] is a common tool to perform Topic Modeling. The tool helps to divide textual information into meaningful clusters based on their semantic meaning.

In our study, we applied BERTopic on all comments from each platform and comments labeled as prevailing emotions from the previous step. We separately performed Topic Modeling on comments related to *fear*, *trust*, and *anticipation* emotions.

IV. DATASET ANNOTATION

We annotated the dataset because there were no human-labeled or classifier-trained tweets in this environment context. We randomly sampled 100 tweets from our previously mentioned dataset, focusing on environment-related content.

A sentiment analysis dataset annotated on an 11-point scale implies that each instance in the dataset is assigned a sentiment label on a scale ranging from -5 to 5.

Six human subjects performed the annotation process. Among them was one expert in ecology and one Ph.D. in Sustainable Engineering. They were given a double weight since they better understood the context and sentiment conveyed by specific terms within discussions about climate change.

All six participants in our annotation process have formally passed through the informed consent procedure, demonstrating their understanding and willingness to participate in the study. Figure 10 shows a screenshot of the Google Sheets form. The 11-point annotation approach proposed in [67] was used to annotate the test dataset. Each annotator was needed to assign a score to a tweet's opinion based on a perceived value ranging from -5 (showing significant discontent) to +5 (for exceptionally positive tweets).

The total sentiment score for each tweet was derived as a weighted average of all six annotators (A), with experts (Expert Annotators, EA)receiving a twofold weighting, as mentioned above.

$$\text{Total Sentiment Score}_{\text{tweet}} = \frac{\sum_{i=1}^{n} w_i \times \text{Annotator}_i}{\sum_{i=1}^{n} w_i} \quad (4)$$

where w_i is the weight assigned to annotator i, and Annotator_i represents the sentiment score assigned by annotator i, n is the number of annotators.

For expert annotators (EA) receiving a twofold weighting, non-expert annotators (A) receiving weight 1:

$$w_{\text{EA}} = 2, w_{\text{A}} = 1$$

We applied the strategies described in [67]: if 60% or more annotator labels are considered outliers, the annotator judgments are removed from the job. We utilize the formula Equation 5 to determine whether a judgment A_i, j is an outlier [67]:

$$|A_{i,j} - avg(A_{i',j})| > std_t(t_j), \quad (5)$$

where $std_t(t_j)$ is the standard deviation of all scores given for a tweet t_j.

As a result, no outliers were revealed since we got the following proportion of outlier labels: $EA_1 = 8\%$, $EA_2 = 25\%$, $A_4 = 29\%$, $A_5 = 35\%$, $A_6 = 36\%$.

Each tweet in a trial dataset eventually received a positive, neutral, or negative score based on weighted average scoring (see Table 3). With the classification threshold set at -0.1 and $+0.1$, the annotation process produced the following sentiment distribution: positive (44%), neutral (0%), and negative (56%).

Inter-Annotator Agreement (IAA) [68] measures the level of agreement between multiple annotators in their assessments of environmental tweets. Specifically, we use Cohen's Kappa (κ) as a measure to quantify the level of agreement among the annotators.

The formula for Cohen's Kappa is given by:

$$\kappa = \frac{P_o - P_e}{1 - P_e} \quad (6)$$

where:

$$P_o = \frac{\text{Number of agreements}}{\text{Total number of annotations}}$$

$$P_e = \sum_i \left(\frac{\text{Total annotations by annotator } i}{\text{Total number of annotations}} \right)^2$$

So, P_o represents the observed agreement - the proportion of times the annotators agree, while P_e denotes the expected agreement, the hypothetical probability of chance agreement.

First, we converted the annotations of each subject from an 11-point scale to positive (1), neutral (0), or negative(-1) to calculate agreement. Then, we calculate Pairwise Cohen's Kappa. Finally, we calculate the average Cohen's Kappa across all pairs of annotators to get an overall measure of the agreement.

The heatmap in Figure 11 visualizes the Cohen's Kappa scores for each pair of annotators, providing insight into their level of agreement. The values range from -1 to 1, where 1 indicates perfect agreement, 0 indicates no agreement, and -1 indicates perfect disagreement.

The value of κ can range from -1 (complete disagreement) to 1 (complete agreement). A value of 0 indicates that the agreement is no better than chance. The average Cohen's Kappa score across all pairs of annotators is 0.525 in our case, which suggests a *moderate* level of agreement among the annotators. Experts' agreement is very high.

V. EXPERIMENTAL RESULTS
The experiments involved three stages (as seen in Figure 12): Sentiment prediction, Emotion prediction, and Topic modeling. The following sections describe each stage of the analysis process.

A. SENTIMENT DETECTION RESULTS
Figure 14 and Table 4 demonstrate the sentiment distribution over the years on three different social network systems: Twitter, Reddit, and YouTube. The comparison indicates a distinct sentiment pattern among these platforms. Contrary to the prevalent negative sentiments found consistently on Twitter throughout the years, Reddit exhibits a rising trend in positive expressions. In contrast, YouTube portrays a slightly higher frequency of negative sentiments, indicating a different sentiment landscape within the platform. Twitter and YouTube predominantly showcase negative sentiments, suggesting a prevalence of critical or adverse expressions among their user bases. However, the noticeable increase in positive sentiments on Reddit reveals a contrasting sentiment trend, showcasing an evolving and comparatively more optimistic user engagement over time.

Figure 13 presents some examples of tweets classified as positive and negative.

TABLE 3. Sentiment analysis for selected tweets based on manual annotation. Expert annotators were marked with asterisks, and they were given twice the weight.

#	Tweet	EA1*	EA2*	A3	A4	A5	A6	Weighted AVG	Annotation	Sentiment
1	Nestlé helps farm...	3	2	3	3	1	4	2.63	1	Positive
2	Geology research...	-2	0	0	-2	-1	-2	-1.13	-1	Negative
3	New paper demo...	0	0	1	-1	0	2	0.25	1	Positive
4	If climate change...	4	5	3	2	3	3	3.63	1	Positive
5	Climate change th...	-4	-1	-2	3	-2	3	-1	-1	Negative
...	...	...	...	...	...	...	...	...	...	
100	Supporting the tra...	5	4	5	4	3	4	4.25	1	Positive

FIGURE 10. An example form (Google Sheets) featuring comments or posts from Twitter/Reddit/Youtube given to each annotator individually for annotation. The following study's score descriptions and data annotation techniques were taken from [67]. The page design was taken from [12]. The form contains 100 random posts from our unlabeled environment-related dataset.

B. EMOTION DETECTION RESULTS

Once we get only comments with negative sentiment scores, we input them into the NRClex tool to get emotion distribution in each comment. Comments with scores less than -0.1 were considered negative. Figure 15 shows each social media platform's mean emotion intensity score over the years. We can clearly see that *fear*, *trust*, and *anticipation* were the most prevailing emotions over the years throughout all social media platforms.

In the case of Twitter, all three emotions were growing simultaneously over time. *Trust* and *anticipation* peaked in 2020, which could be attributed to COVID-19 and tons of information being spread through social media during that period. Many users tended to agree and listen to posts from

medical professionals, thus increasing trust and anticipation. Emotion *fear* also grows steadily with fewer fluctuations.

Unfortunately, YouTube has a major lack of information between 2014 and 2019. This happened because news channel accounts on YouTube only recently started actively publishing videos related solely to the environment and climate change, thus creating an information gap. However, we can observe an increasing number of *fear* emotions with its peak in 2022. Such a trend indicates growing user anxiety towards environmental challenges and acknowledging existing problems. *Trust*, *anticipation* and *sadness* show a visible growth in the last three years. Both *trust* and *anticipation* could be attributed to the nature of the data, as it was scraped from the YouTube accounts of popular news

TABLE 4. Sentiment analysis statistical information per year.

Platform	Year	Popular	Positive	Positive, %	Negative	Negative, %	Neutral	Neutral, %	Emotion
Twitter	2014	1530	494	32.29	783	51.18	253	16.54	Anticipation
	2015	2101	671	31.94	1086	51.69	344	16.37	Surprise
	2016	3279	1096	33.42	1598	48.73	585	17.84	Anticipation
	2017	4364	1366	31.30	2296	52.61	702	16.09	Surprise
	2018	5930	1876	31.64	3079	51.92	975	16.44	Fear
	2019	6419	2008	31.28	3430	53.44	981	15.28	Fear
	2020	6725	2039	30.32	3655	54.35	1031	15.33	Anticipation
	2021	7513	2226	29.63	4108	54.68	1179	15.69	Surprise
	2022	7405	2211	29.86	4045	54.63	1149	15.52	Anticipation
	2023	3037	886	29.177	1687	55.55	464	15.28	Fear
	Total	**48887**	**15060**	**30.81%**	**26079**	**53.35%**	**7748**	**15.85%**	**Anticipation**
Reddit	2014	108	49	45.37	34	31.48	25	23.15	Trust
	2015	88	35	39.77	32	36.36	21	23.86	Fear
	2016	150	63	42.00	58	38.67	29	19.33	Anticipation
	2017	230	111	48.26	65	28.26	54	23.48	Fear
	2018	432	195	45.14	144	33.33	93	21.53	Anger
	2019	881	395	44.84	290	32.92	196	22.25	Fear
	2020	1144	559	48.86	381	33.30	204	17.83	Surprise
	2021	1494	708	47.39	496	33.20	290	19.41	Surprise
	2022	1947	843	43.30	682	35.03	422	21.67	Surprise
	2023	2079	946	45.50	729	35.06	404	19.43	Fear
	Total	**8732**	**3977**	**45.55%**	**2972**	**34.04%**	**1783**	**20.42%**	**Fear**
YouTube	2014	1	0	0	1	100.00	0	0	-
	2015	5	3	60.00	2	40.00	0	0	Fear
	2016	26	11	42.31	10	38.46	5	19.23	Anticipation
	2017	38	15	39.47	17	44.74	6	15.79	Trust
	2018	36	15	41.67	17	47.22	4	11.11	Fear
	2019	201	70	34.83	102	50.75	29	14.43	Anger
	2020	129	50	38.76	57	44.19	22	17.05	Trust
	2021	269	120	44.61	115	42.75	34	12.64	Fear
	2022	1036	411	39.67	480	46.33	145	14.00	Fear
	2023	1015	281	37.54	486	47.88	148	14.58	Fear
	Total	**2756**	**1076**	**39.04%**	**1287**	**46.70%**	**393**	**14.26%**	**Fear**

channels, indicating users' trust towards information released on the official news channel accounts.

C. TOPIC MODELING

As mentioned, Topic Modeling was performed by BERTopic [66]. To get an insight into all scraped comments, we fed the cleaned sentences into BERTopic. The results correspond to the "Overall" column in Table 5. The comments from each platform were clustered into a maximum of five topics. The topics reflect the most common themes for the discussion on each social media platform. Clearly, *Climate change* represents the biggest topic cluster across all platforms. *Air*

quality, *Emissions*, *Plastic*, and *Recycling* also seem to be common discussions in all three datasets.

Within *Climate change* topic, users raise their concerns about global warming and criticize governmental ignorance. It was common for users to demonstrate skepticism towards climate change's real threat; some users justified it as the natural process for the planet and named the opposite opinion a "Climate hysteria."

Air quality and *Emissions* seems to be mostly addressing vehicles and transport. Many users discussed the emissions released into the atmosphere during flights. On YouTube, users are mostly encouraged to switch
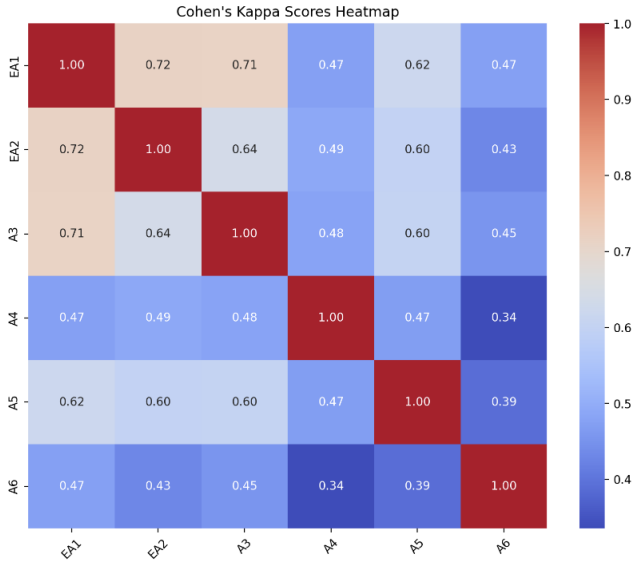
FIGURE 11. The heatmap showing the correlation between the ratings of different annotators - two expert annotators and four general annotators. Cohen's Kappa for each pair of annotators.
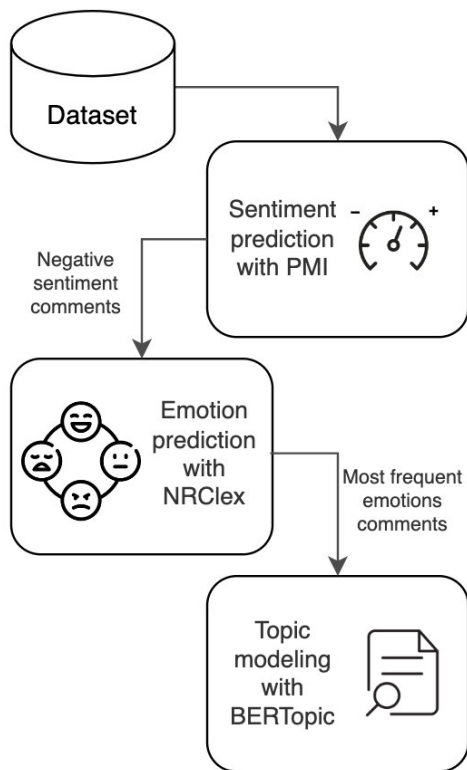

FIGURE 12. Analysis flow.

from gas and petrol-fueled cars to electric and hybrid alternatives.

Recycling topic includes discussions about the ban on single-use plastic and companies shifting to recycled plastic or paper. Some users raised their concerns about whether recycling plastic individually will have any effect or whether
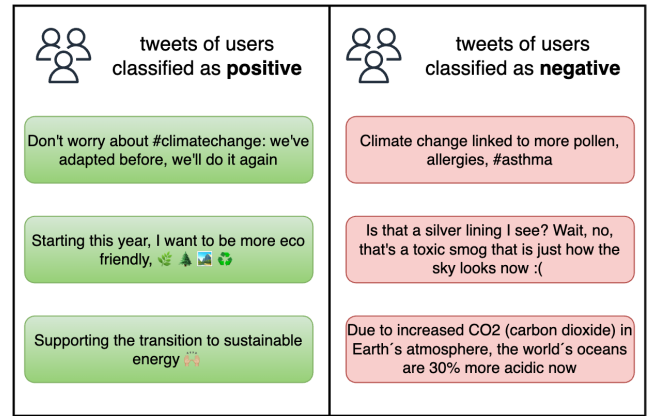

FIGURE 13. Examples of tweets classified as positive and negative.

the big companies should take responsibility and incorporate green practices.

On YouTube, several users expressed the irony of world leaders flying on private jets to the summit to discuss climate change. Such discussions were likely sparked by news reports about outcomes of the "Climate Summit" or similar events.

Within Twitter, discussions about biodiversity were also popular. Users worry about the receding planet's biodiversity and the role of human activity in that problem.

Overall, we could see similar topics and patterns across all three social media platforms, indicating that the platform does not have a notable effect on user comments and discussions. It was also observed that there is still a good portion of users who consider climate change propaganda and its effects as a natural process for the planet that has been occurring before. Table 6 provides comment examples per social media platform and the associated topic and emotion.

We examined data from three social networks and observed common trends with slight differences, likely stemming from the distinct nature of each platform, different user demographics, communication styles, etc. For instance, YouTube's inclusion of visual data in the form of videos may impact the associated comments.

D. POSITIVITY BIAS TEST

Whether more positive or more negative posts are more popular on each platform? The impact of sentiment on the virality of content in social media has been a subject of considerable interest. While some studies indicate a tendency for negative content to be shared more frequently [69], [70], contrasting findings suggest that positive information on social media is often more likely to garner likes and retweets [71]. Using our collected dataset, we aim to investigate and analyze the relationship between sentiment and content sharing.

For the analysis, we filtered out only viral, extremely popular tweets, Reddit posts, and YouTube comments, with at least 30 retweets (RT $\geq$ 30) for Twitter, at least 200 upvotes for Reddit, and 100 likes for YouTube comments.

TABLE 5. Topic clusters per social media platform and prevailing emotion.

	Overall	Fear	Trust	Anticipation
Twitter	Climate change Wastewater Plastic recycling Biodiversity ecology CO_2 and oxygen	Climate change Biodiversity Plastic recycling Air quality Foodwaste and vegan	Climate and energy Biodiversity and food Green energy CO_2 and air	Emissions Energy and carbon Ecology conservation Eco-friendly Air quality
Reddit	Plastic Air quality level	Air and cars Single plastic use Climate change	Plastic recycling Air quality monitor	Store use plastic Eco friendly
YouTube	Climate and people Plastic bags Private jets	Climate and people Private jets Hot summer	Carbon emissions Change and warming Energy and emissions	Oil and plastic Climate change and ice Time to start

TABLE 6. Sample comments with their corresponding topic, emotion, and the social media platform.

Comment	Emotion	Topic	Platform
Investigation done by @CTVToronto reveals poor air quality in #TDSB classrooms could contribute to headaches and sleepiness	Anticipation	Air quality and emissions	Twitter
A Series of NASA Satellite Images Showcasing the Improvement in United States Air Quality From 2005 to 2011	Trust	Air quality and emissions	Twitter
Some were concerned the city's air quality, made worse by commercial expansion, would worsen allergies and asthma	Fear	Air quality and emissions	Twitter
So I got an air quality monitor showing PM2.5 levels of 150 µg/m3. That can't be good right? My cat has asthma attacks constantly and I'm wondering if its linked	Trust	Air quality and emissions	Reddit
Smoking flares do not count when reducing emissions	Trust	Air quality and emissions	YouTube
How in the fuck can people look at the world around us and think climate change isn't real	Trust	Climate change	Twitter
When they ban jet flights to the Maldives and make everyone sail or swim there we'll their Global Warming hysteria seriously	Fear	Climate change	Twitter
Duh climate has always changed since 1990's it's nothing new it goes from cold , warm , dry , tornadoes, thunderstorms, rain , earthquakes	Trust	Climate change	YouTube
It seems so terrible. Increasingly the real harm of climate change is getting bigger, year by year	Fear	Climate change	YouTube
Ironic that everybody is worrying about climate change now when we knew decades ago this was coming	Anticipation	Climate change	YouTube
The newest climate report from the Intergovernmental Panel on Climate Change (IPCC) says that climate change is "unequivocally" caused by humans and warns that global temperatures are expected to reach a significant warming milestone in the next 20 years	Fear	Climate change	Reddit
The Climate Summit burned a hole in the Ozone with all the private jets that flew there	Fear	Private jets	YouTube
As much as 40% of the food produced in America ends up at the dump, off-gassing methane and contributing to climate change	Fear	Foodwaste	Twitter
Ppl are more comfortable w/ Recycling b/c Reducing and Reusing challenge capitalism too much	Anticipation	Recycling and eco-friendly	Twitter
Sustainability and environment expert, Dr. Jody Tishmack, provides an insightful explanation of our current water crisis, the impact it will have on society, and the solutions needed to manage it	Trust	Recycling and eco-friendly	Reddit

As each social media platform is unique, the filtering parameters also differ. Figure 16 shows viral media posts' retweets/upvotes/likes as a function of their sentiment score. The data was fitted into a polynomial function. The graph demonstrates the positivity bias on Reddit and the negativity bias on Twitter and YouTube.

Our findings partly back up previous research on the influence of sentiment on information spread [71]. Based on their research, positive messages are more likely to be shared and liked due to a phenomenon known as positivity bias. In our case, only Reddit confirms the positivity bias, probably because the context of the environment is mostly negative.
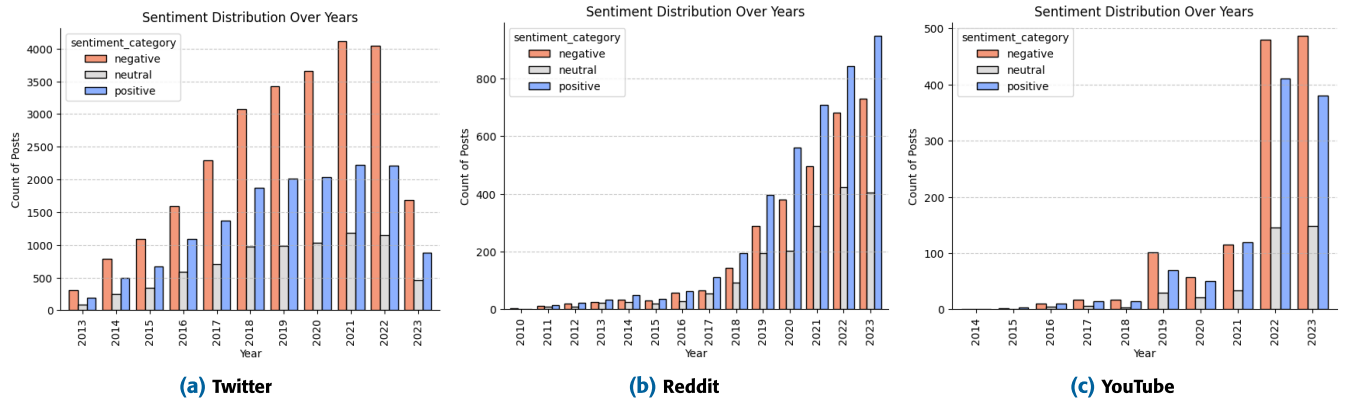
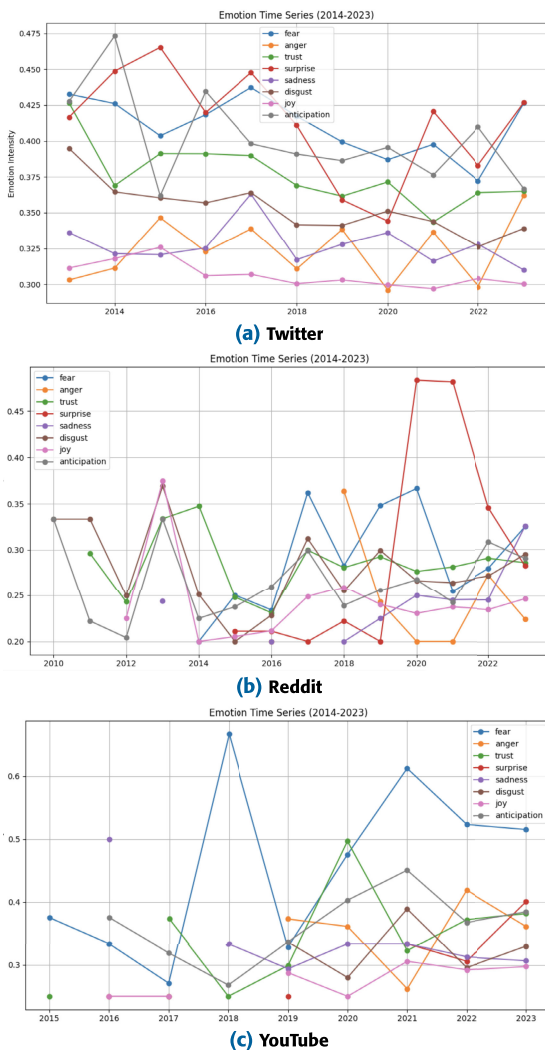**FIGURE 14.** Sentiment scores distribution over the years.

FIGURE 15. Mean emotion intensity scores over the years. For some emotions, the line is breaking due to the mean value of that emotion being lower than 0.25 in that year.

Next, the average number of retweets for negative tweets is larger (7.37 and 5.41 for negative and positive tweets,

respectively) for Twitter. At the same time, the average number of likes/upvotes for negative comments is smaller for YouTube (23.45 and 30.95 for negative and positive comments, respectively) and Reddit (56.11 and 60.22 for negative and positive posts/comments, respectively).

E. CONTEXT-SPECIFIC SEMANTIC ORIENTATION OF WORDS

The analysis of semantic orientation (SO) scores for some context-specific keywords yielded interesting insights into the sentiment associations of these words. The SO scores indicate the sentiment associations within the context of the sentiment analysis. Based on Figure 17, among these keywords, *"alternative"* and *"electric"* stand out with notably positive scores, suggesting strong positive associations, potentially indicative of favorable perceptions towards alternative solutions and electric-related aspects. Conversely, words like *"fire"*, *"animal"*, *"humans"*, *"oil"*, *"clothes"* and *"fossil"* exhibit extremely negative sentiment scores, hinting at severe negative associations within the context, perhaps highlighting concerns about environmental degradation, social issues, or adverse impacts.

Keywords such as *"pollution"*, *"bag"*, *"gas"* demonstrate a small degree of negativity, indicating concerns or associations that lean towards negative aspects within the context.

"Clean" and *"planet"* show moderately positive scores, hinting at positive associations potentially linked to cleanliness or considerations for the well-being of the planet.

These semantic scores collectively reflect a nuanced landscape of sentiments and associations surrounding these keywords within the specified domain, showcasing a wide range from deeply negative to strongly positive perceptions and concerns.

Analyzing the sentiment analysis scores of specific keywords in the dataset helps us better understand how they are perceived and associated with sentiment. This provides valuable insights into subtle sentiment patterns and serves as a foundation for further discussions and interpretations in sentiment analysis in environmental contexts.
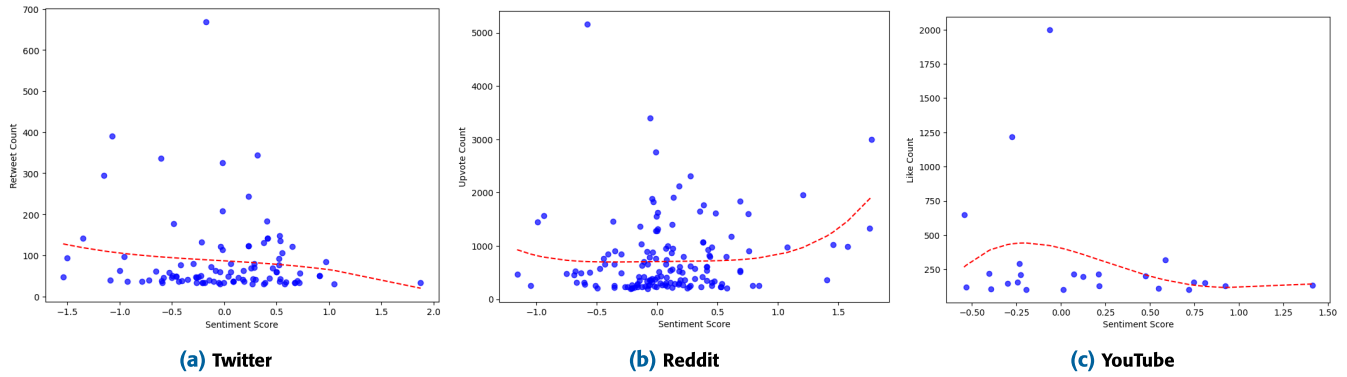
FIGURE 16. The number of retweets/upvotes/likes received by extremely popular posts as a function of the sentiment score represented in them. We can see the positivity bias on Reddit and the negativity bias on Twitter and YouTube.
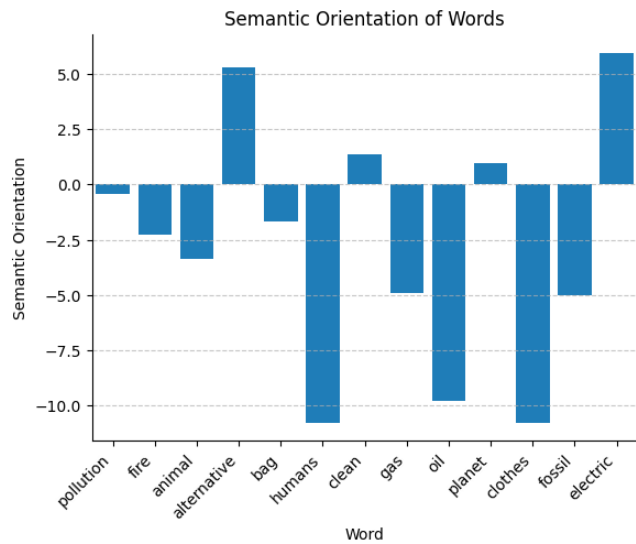

FIGURE 17. Words semantic orientation scores.

F. ACCURACY EVALUATION

To assess the effectiveness of our method, we use the human-annotated dataset previously discussed and the well-known sentiment analysis models VADER [72] and spaCy [73], and Senti [74], [75].

We examine the classification algorithms using metrics like F_1 score, Accuracy, Precision, and Recall.

Precision represents the ratio of the number of true positive predictions to the total number of positive predictions.

$$Precision = \frac{TP}{TP + FP}$$

Recall shows accurate positive predictions when compared to the total number of actual positives [76]:

$$Recall = \frac{TP}{TP + FN}$$

F_1 score is [76]:

$$F_1 = 2 * \frac{Precision * Recall}{Precision + Recall}$$

Accuracy is a measure that represents the ratio of correct predictions to the total number of predictions made [76]:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Table 7 shows evaluation results for each method. Sentiment prediction accuracy, precision, recall, and F1 scores were calculated for PMI-based, VADER, spaCy, and Senti classifiers. Additionally, we provided evaluation results for annotator EA_1, who holds a Ph.D. in ecology and demonstrated the lowest outliers rate among all annotators (12%) for comparison. Based on the results presented in Table 7, it can be observed that our classifier (0.65) outperforms VADER (0.64), Senti (0.57), and spaCy (0.44) for the selected context. However, it still falls short of the accuracy achieved by individual human raters with specialized context knowledge (0.90). Although VADER is known to be sensitive to social media lexicon, its accuracy decreases when used in domain-specific environmental tweets. The dataset used for training has limited variability and context-specific vocabulary, which is why a PMI-based classifier performs better.

Table 8 compares the Mutual-Information scoring function to the VADER, spaCy, Senti, and expert annotator scores, along with the example tweets from the labeled dataset.

The last row in Table 8 represents the most controversial tweet with the highest standard deviation(std = 4.22) in our manual annotation experiment. Despite "zero waste" typically being associated with positivity, the tweet explores the challenges of adhering to this concept in the modern world. The heightened standard deviation reflects the varying interpretations among annotators, underscoring the nuanced sentiment and real-world complexities embedded in the discussion of "zero waste."

TABLE 7. Sentiment prediction accuracy, precision, recall, F1 for the proposed PMI-based, VADER, spaCy, Senti classifiers. We also provided the evaluation for annotator EA_1, who has a Ph.D. in ecology, just for comparison.

	Accuracy	Precision	Recall	F1
Proposed PMI classifier	0.65	0.65	0.65	0.64
VADER classifier	0.64	0.80	0.64	0.71
spaCy classifier	0.44	0.44	1.00	0.61
Senti classifier	0.57	0.72	0.57	0.64
EA_1, PhD in Ecology	0.90	0.94	0.9	0.92

TABLE 8. Sentiment analysis results for 100 tweets: VADER, spaCy, Senti, Expert Annotator, and Pmi-based methods.

Tweet	VADER	spaCy	Senti	EA1	PMI score	PMI sent.	Annotated AVG	Annotated label
Imagine 2050, when there's absolutely no denying terrible reality of climate change, and it's too late, WHY DIDN'T THEY ACT 50 YRS EARLIER?	Positive	Positive	Negative	-5 (Negative)	-0.19	Negative	-4.63	Negative
Imagine Mexico City with all electric cars and no chocking smog	Negative	Positive	Positive	3 (Positive)	-2.31	Negative	3.25	Positive
Lets take action people... Save energy, cut down red meat and practice zero waste	Positive	Positive	Neutral	5 (Positive)	1.7	Positive	3.88	Positive
Make a difference! Join the fight to defend the teaching of evolution and climate change	Negative	Positive	Negative	4 (Positive)	-0.33	Negative	2.75	Positive
I want to go zero waste but there's no such thing as zero waste art supplies, everything comes in aluminum tubes and I need to use lots of paper towels and ...aren't good for the planet either:(	Negative	Positive	Positive	-5 (Negative)	-0.19	Negative	-2.75	Negative

VI. DISCUSSION

Let's explore how our research findings align with previous studies. Contrary to the findings of recent research [17], [18], which emphasize a dominant trend of positive sentiments on Twitter, our study unveils a significantly different scenario. For example, in our analysis, a considerable 54.7% of tweets from the year 2021 exhibited negative sentiments, sharply contrasting with their reported figure of 19.7% negative sentiments. Interestingly, this higher prevalence of negative sentiments in our findings is consistent with the results of other studies [20], [22].

Reference [77] conducted a study involving clustering Twitter data into various topics, aligning with our own topic modeling efforts. Notably, themes such as *"climate change"* and *"carbon emissions"* surfaced in both studies, underscoring their mutual support.

Our findings align with the research by [38], which concluded that anticipation (fear) is the most evoked emotion in environmental posts. However, this contrasts the analysis presented by [78], where anger was identified as the predominant emotion in tweet-based emotion detection. While their results highlight anger, our study reveals anticipation and fear as the more dominant emotions in our analyzed data.

It's worth highlighting that similar studies like [17], [79], [80], and others have revealed a consistent pattern of relatively low accuracy in sentiment classification for environmental tweets, although these studies used different datasets. The outcome of our study supports this claim, as our classifier achieved an accuracy of 65%. Despite their widespread use, VADER, Senti, and spaCy yielded relatively modest accuracies in this specific context, recording 0.64, 0.57, and 0.44, respectively, on the human-annotated dataset we utilized in this study. However, in sentiment analysis and NLP, the accuracy above 80% is considered a good performance. This underscores the ongoing necessity for enhanced support in refining sentiment classification models to identify sentiments within environmental data effectively.

VII. CONCLUSION

This study demonstrated the value of sentiment analysis in understanding the public perception of environmental topics over a decade. We explored multiple social networks for a broader perspective, included emotion detection to capture a wide range of emotional responses, and employed topic modeling techniques to identify specific environmental topics.

Our findings show that negative environmental tweets are much more common than positive or neutral ones (X% vs Y% and Z% on average). Climate change is the primary topic across all social media platforms, emphasizing its widespread

global concern. Additionally, discussions on *air quality, emissions, plastic,* and *recycling* consistently appear in all datasets, highlighting their ongoing relevance. The prevailing emotions expressed in environmental tweets are *fear, trust* and *anticipation*, indicating public reactions' diverse and complex nature.

These findings contribute to understanding how social media platforms perceive and discuss environmental topics. The results can be used to inform policymakers, organizations, and governments about the public's priorities, concerns, and suggestions related to environmental issues. This can lead to the formulation of more effective environmental policies.

Our study faced a key limitation: low sentiment analysis accuracy in environmental tweets. This challenge arises from the predominantly negative tone surrounding climate change discussions, making it challenging to identify nuanced positive sentiments. Additionally, we have observed that such tweets often contain irony and sarcasm, which can also hinder accuracy.

As for future works, we plan to incorporate multimodal analysis and use images besides text. We also aim to compare and contrast social media sentiments with sentiments expressed in traditional media, like news articles, TV, and radio.

REFERENCES

[1] V. 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, "Community 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.

[26] B. Wang, N. Wang, and Z. Chen, "Research on air quality forecast based on web text sentiment analysis," *Ecolog. Inform.*, vol. 64, Sep. 2021, Art. no. 101354.

[27] S. Shan, J. Peng, and Y. Wei, "Environmental sustainability assessment 2.0: The value of social media data for determining the emotional responses of people to river pollution—A case study of Weibo (Chinese Twitter)," *Socio-Econ. Planning Sci.*, vol. 75, Jun. 2021, Art. no. 100868.

[28] A. Reyes-Menendez, J. Saura, and C. Alvarez-Alonso, "Understanding WorldEnvironmentDay user opinions in Twitter: A topic-based sentiment analysis approach," *Int. J. Environ. Res. Public Health*, vol. 15, no. 11, p. 2537, Nov. 2018.

[29] E. M. Cody, A. J. Reagan, L. Mitchell, P. S. Dodds, and C. M. Danforth, "Climate change sentiment on Twitter: An unsolicited public opinion poll," *PLoS One*, vol. 10, no. 8, Aug. 2015, Art. no. e0136092.

[30] F. Müller-Hansen, T. Repke, C. M. Baum, E. Brutschin, M. W. Callaghan, R. Debnath, W. F. Lamb, S. Low, S. Lück, C. Roberts, B. K. Sovacool, and J. C. Minx, "Attention, sentiments and emotions towards emerging climate technologies on Twitter," *Global Environ. Change*, vol. 83, Dec. 2023, Art. no. 102765.

[31] L. Bryan-Smith, J. Godsall, F. George, K. Egode, N. Dethlefs, and D. Parsons, "Real-time social media sentiment analysis for rapid impact assessment of floods," *Comput. Geosci.*, vol. 178, Sep. 2023, Art. no. 105405.

[32] M. L. Loureiro, M. Alló, and P. Coello, "Hot in Twitter: Assessing the emotional impacts of wildfires with sentiment analysis," *Ecolog. Econ.*, vol. 200, Oct. 2022, Art. no. 107502.

[33] C. Li, L. Li, J. Zheng, J. Wang, Y. Yuan, Z. Lv, Y. Wei, Q. Han, J. Gao, and W. Liu, "China's public firms' attitudes towards environmental protection based on sentiment analysis and random forest models," *Sustainability*, vol. 14, no. 9, p. 5046, Apr. 2022.

[34] D. Nikolova, G. Mircheva, and E. Zdravevski, "Application of traditional and deep learning algorithms in sentiment analysis of global warming tweets," in *Smart Objects and Technologies for Social Good*, P. J. Coelho, I. M. Pires, and N. V. Lopes, Eds. Cham, Switzerland: Springer, 2024, pp. 49–61.

[35] D. T. Joy, V. Thada, and U. Srivastava, "Sentiment analysis on global warming tweets using Naïve Bayes and RNN," in *Data Engineering for Smart Systems*, P. Nanda, V. K. Verma, S. Srivastava, R. K. Gupta, and A. P. Mazumdar, Eds. Singapore: Springer, 2022, pp. 225–234.

[36] M. Alkhatib, M. El Barachi, S. S. Mathew, and F. Oroumchian, "Using artificial intelligence to monitor the evolution of opinion leaders' sentiments: Case study on global warming," in *Proc. 5th Int. Conf. Smart Sustain. Technol. (SpliTech)*, Sep. 2020, pp. 1–6.

[37] Y. Jiang, X. Song, J. Harrison, S. Quegan, and D. Maynard, "Comparing attitudes to climate change in the media using sentiment analysis based on latent Dirichlet allocation," in *Proc. EMNLP Workshop, Natural Lang. Process. Meets Journalism*, Copenhagen, Denmark, O. Popescu and C. Strapparava, Eds. Copenhagen, Denmark: Association for Computational Linguistics., 2017, pp. 25–30.

[38] M. L. Loureiro and M. Alló, "Sensing climate change and energy issues: Sentiment and emotion analysis with social media in the U.K. and Spain," *Energy Policy*, vol. 143, Aug. 2020, Art. no. 111490.

[39] F. Qiao and J. Williams, "Topic modelling and sentiment analysis of global warming tweets," *J. Organizational End User Comput.*, vol. 34, no. 3, pp. 1–18, Dec. 2021.

[40] M. El Barachi, M. AlKhatib, S. Mathew, and F. Oroumchian, "A novel sentiment analysis framework for monitoring the evolving public opinion in real-time: Case study on climate change," *J. Cleaner Prod.*, vol. 312, Aug. 2021, Art. no. 127820.

[41] M. O. Ibrohim, C. Bosco, and V. Basile, "Sentiment analysis for the natural environment: A systematic review," *ACM Comput. Surv.*, vol. 56, no. 4, pp. 1–37, Apr. 2024.

[42] M. V. Mäntylä, D. Graziotin, and M. Kuutila, "The evolution of sentiment analysis—A review of research topics, venues, and top cited papers," *Comput. Sci. Rev.*, vol. 27, pp. 16–32, Feb. 2018.

[43] M. Rodríguez-Ibánez, A. Casánez-Ventura, F. Castejón-Mateos, and P.-M. Cuenca-Jiménez, "A review on sentiment analysis from social media platforms," *Expert Syst. Appl.*, vol. 223, Aug. 2023, Art. no. 119862.

[44] G. Gautam and D. Yadav, "Sentiment analysis of Twitter data using machine learning approaches and semantic analysis," in *Proc. 7th Int. Conf. Contemp. Comput. (IC3)*, Aug. 2014, pp. 437–442.

[45] M. S. Neethu and R. Rajasree, "Sentiment analysis in Twitter using machine learning techniques," in *Proc. 4th Int. Conf. Comput., Commun. Netw. Technol. (ICCCNT)*, Jul. 2013, pp. 1–5.

[46] F. Hemmatian and M. K. Sohrabi, "A survey on classification techniques for opinion mining and sentiment analysis," *Artif. Intell. Rev.*, vol. 52, no. 3, pp. 1495–1545, Oct. 2019.

[47] A. Ligthart, C. Catal, and B. Tekinerdogan, "Systematic reviews in sentiment analysis: A tertiary study," *Artif. Intell. Rev.*, vol. 54, no. 7, pp. 4997–5053, Oct. 2021.

[48] K. Church and P. Hanks, "Word association norms, mutual information, and lexicography," *Comput. Linguistics*, vol. 16, pp. 22–29, Jul. 2002.

[49] P. D. Turney and M. L. Littman, "Measuring praise and criticism: Inference of semantic orientation from association," *ACM Trans. Inf. Syst.*, vol. 21, no. 4, pp. 315–346, Oct. 2003.

[50] P. D. Turney and M. L. Littman, "Unsupervised learning of semantic orientation from a hundred-billion-word corpus," 2003, *arXiv: cs/0212012*.

[51] P. D. Turney, "Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews," 2002, *arXiv: cs/0212032*.

[52] R. M. Fano, *Transmission of Information: A Statistical Theory of Communication*. Cambridge, MA, USA: MIT Press, 1961.

[53] D. Jurafsky, J. H. Martin, P. Norvig, and S. Russell, *Speech and Language Processing*. London, U.K.: Pearson Education, 2014.

[54] N. Bindal and N. Chatterjee, "A two-step method for sentiment analysis of tweets," in *Proc. Int. Conf. Inf. Technol. (ICIT)*, Los Alamitos, CA, USA, Dec. 2016, pp. 218–224.

[55] A. D. Vo and C. Y. Ock, "Sentiment classification: A combination of PMI, sentiWordNet and fuzzy function," in *Proc. Int. Conf. Comput. Collective Intell.* New York, NY, USA: Springer, 2012, pp. 373–382.

[56] R. Feldman, "Techniques and applications for sentiment analysis," *Commun. ACM*, vol. 56, no. 4, pp. 82–89, Apr. 2013.

[57] H. Hamdan, P. Bellot, and F. Bechet, "Sentiment lexicon-based features for sentiment analysis in short text," *Res. Comput. Sci.*, vol. 90, no. 1, pp. 217–226, Dec. 2015.

[58] H. Utama, "Sentiment analysis in airline tweets using mutual information for feature selection," in *Proc. 4th Int. Conf. Inf. Technol., Inf. Syst. Electr. Eng. (ICITISEE)*, Nov. 2019, pp. 295–300.

[59] M. Bonzanini, *Mastering Social Media Mining With Python: Acquire and Analyze Data From All Corners of the Social Web With Python* (Community Experience Distilled). Birmingham, U.K.: Packt Publishing, 2016.

[60] P. R. Kanna and P. Pandiaraja, "An efficient sentiment analysis approach for product review using turney algorithm," *Proc. Comput. Sci.*, vol. 165, pp. 356–362, Dec. 2019.

[61] A. Pesaranghader, S. Muthaiyah, and A. Pesaranghader, "Improving gloss vector semantic relatedness measure by integrating pointwise mutual information: Optimizing second-order co-occurrence vectors computed from biomedical corpus and UMLS," in *Proc. Int. Conf. Informat. Creative Multimedia*, Sep. 2013, pp. 196–201.

[62] L. Du, X. Li, and D. Lin, "Chinese term extraction from web pages based on expected point-wise mutual information," in *Proc. 12th Int. Conf. Natural Comput., Fuzzy Syst. Knowl. Discovery (ICNC-FSKD)*, Aug. 2016, pp. 1647–1651.

[63] A. Matt, "Twitter statistics 2022: Stats, user demography and facts," WebsiteRating.com, QL, Australia, Tech. Rep. 1, Apr. 2022.

[64] P. Pandey, "Basic tweet preprocessing in Python," Towards Data Sci., Toronto, ON, Canada, Tech. Rep. 1, 2020.

[65] S. M. Mohammad and P. D. Turney, "NRC emotion lexicon," *Nat. Res. Council, Canada*, vol. 2, p. 234, Jun. 2013.

[66] M. Grootendorst, "BERTopic: Neural topic modeling with a class-based TF-IDF procedure," 2022, *arXiv:2203.05794*.

[67] A. Ghosh, G. Li, T. Veale, P. Rosso, E. Shutova, J. Barnden, and A. Reyes, "SemEval-2015 task 11: Sentiment analysis of figurative language in Twitter," in *Proc. 9th Int. Workshop Semantic Eval. (SemEval)*, Jun. 2015, pp. 470–478.

[68] A. J. Viera and J. M. Garrett, "Understanding interobserver agreement: The Kappa statistic," *Fam. Med.*, vol. 37, no. 5, pp. 360–363, 2005.

[69] J. P. Schöne, D. Garcia, B. Parkinson, and A. Goldenberg, "Negative expressions are shared more on Twitter for public figures than for ordinary users," *PNAS Nexus*, vol. 2, no. 7, pp. 1–11, Jul. 2023.

[70] J. P. Schöne, B. Parkinson, and A. Goldenberg, "Negativity spreads more than positivity on Twitter after both positive and negative political situations," *Affect. Sci.*, vol. 2, no. 4, pp. 379–390, Oct. 2021.

[71] E. Ferrara and Z. Yang, "Quantifying the effect of sentiment on information diffusion in social media," *PeerJ Comput. Sci.*, vol. 1, p. e26, Sep. 2015.

[72] C. J. Hutto and E. Gilbert, "VADER: A parsimonious rule-based model for sentiment analysis of social media text," in *Proc. Int. AAAI Conf. Web Social Media*, 2014, pp. 216–225.

[73] M. Honnibal and I. Montani, "spaCy 2: Natural language understanding with Bloom embeddings, convolutional neural networks and incremental parsing," *To Appear*, vol. 7, no. 1, pp. 411–420, 2017.

[74] F. Sebastiani and A. Esuli, "SentiWordNet: A publicly available lexical resource for opinion mining," in *Proc. 5th Int. Conf. Lang. Resour. Eval.*, Genoa, Italy, N. Calzolari, K. Choukri, A. Gangemi, B. Maegaard, J. Mariani, J. Odijk, and D. Tapias, Eds., May 2006, pp. 1–6.

[75] B. Ohana and B. Tierney, "Sentiment classification of reviews using SentiWordNet," in *Proc. 9th Int. Conf. Inf. Technol. Telecommun.*, 2009, pp. 1–9.

[76] D. Powers and Ailab, "Evaluation: From precision, recall and F-measure to ROC, informedness, markedness and correlation," *J. Mach. Learn. Technol.*, vol. 2, pp. 2229–3981, Jan. 2011.

[77] D. Effrosynidis, G. Sylaios, and A. Arampatzis, "Exploring climate change on Twitter using seven aspects: Stance, sentiment, aggressiveness, temperature, gender, topics, and disasters," *PLoS One*, vol. 17, no. 9, Sep. 2022, Art. no. e0274213.

[78] G. A. Veltri and D. Atanasova, "Climate change on Twitter: Content, media ecology and information sharing behaviour," *Public Understand. Sci.*, vol. 26, no. 6, pp. 721–737, Aug. 2017.

[79] N. Mucha, "Sentiment analysis of global warming using Twitter data," Ph.D. dissertation, Dept. Comput. Sci., North Dakota State Univ., Fargo, North Dakota, 2018.

[80] Y. Kirelli and S. Arslankaya, "Sentiment analysis of shared tweets on global warming on Twitter with data mining methods: A case study on Turkish language," *Comput. Intell. Neurosci.*, vol. 2020, pp. 1–9, Sep. 2020.

AIDA USMANOVA received the B.S. degree in information systems from Kazakh-British Technical University, Almaty, Kazakhstan, in 2020, and the M.S. degree in artificial intelligence from the University of Hamburg, Hamburg, Germany, in 2023. She is currently pursuing the Ph.D. degree in natural sciences with Leuphana University Lüneburg, Lüneburg, Germany. She presented her work at the Bonn Sustainable AI Conference, in 2023. Her research interests include natural language processing, knowledge extraction, and cognitive computing.

PAKIZAR SHAMOI received the B.S. and M.S. degrees in information systems from Kazakh-British Technical University, Almaty, Kazakhstan, in 2011 and 2013, respectively, and the Ph.D. degree in engineering from Mie University, Tsu, Japan, in 2019. In her academic journey, she has held various teaching and research positions at Kazakh-British Technical University, where she has been a Professor with the School of Information Technology and Engineering, since August 2020. She is the author of one book and more than 28 scientific publications. Her research interests include artificial intelligence and machine learning in general, with a focus on fuzzy sets and logic, soft computing, representing and processing colors in computer systems, natural language processing, computational aesthetics, and human-friendly computing and systems. She received awards for the best paper at conferences five times. She took part in the organization and worked in the organization committee (as the Head of the Session and responsible for special sessions) of several international conferences, such as IFSA-SCIS 2017, Otsu, Japan; SCIS-ISIS 2022, Mie, Japan; and EUSPN 2023, Almaty. She served as a Reviewer for several international conferences, including IEEE: SIST 2023, SMC 2022, SCIS-ISIS 2022, SMC 2020, ICIEV-IVPR 2019, and ICIEV-IVPR 2018.

• • •

DANIYAR AMANGELDI received the B.S. degree in computer systems and software from Kazakh-British Technical University, Almaty, Kazakhstan, in 2022, where he is currently pursuing the M.S. degree in software engineering. He is a Software Developer in a leading e-commerce company in Kazakhstan. His research interests include natural language processing and machine learning.