Social Media Insights into Climate Change: Sentiment Analysis Using VADER and RoBERTa

Gerasimos Vonitsanos*, Andreas Kanavos† and Phivos Mylonas‡
*Computer Engineering and Informatics Department
University of Patras, Patras, Greece
mvonitsanos@ceid.upatras.gr
†Department of Informatics
Ionian University, Corfu, Greece
akanavos@ionio.gr
‡Department of Informatics and Computer Engineering
University of West Attica, Athens, Greece
mylonasf@uniwa.gr

***Abstract*—Sentiment analysis, a critical branch of Natural Language Processing (NLP), is pivotal for uncovering the emotional undertones within textual data, thereby revealing public sentiments on diverse topics. This study conducts a comparative analysis of two prominent sentiment analysis tools—VADER, a lexicon and rule-based approach, and RoBERTa, a transformer-based deep learning model. It focuses on their efficacy in analyzing tweets related to climate change, a topic of global significance that elicits a wide range of public opinions. The nuanced and dynamic nature of social media language poses unique challenges, particularly in contexts such as climate change discussions. We assess how effectively each model discerns positive, neutral, and negative sentiments across different categories of climate change-related tweets, delineated into Pro, Neutral, News, and Anti stances. Our findings indicate that RoBERTa generally outperforms VADER in capturing contextual nuances and providing detailed sentiment classifications. This detailed capability allows RoBERTa to better reflect the complex public opinions on climate change, offering this way invaluable insights to policymakers, researchers, and environmental advocates. This study not only aids in better understanding and engaging with public discourse on social media but also highlights the potential of advanced NLP tools in shaping environmental communication and policy formulation.**

***Index Terms*—Sentiment Analysis, Climate Change, Twitter, VADER, RoBERTa, Natural Language Processing, Social Media Analysis, Opinion Mining, Machine Learning, Text Mining, Computational Linguistics**

I. Introduction

Sentiment analysis, or opinion mining, is a critical field within Natural Language Processing (NLP) that focuses on identifying and categorizing opinions expressed in textual data. The primary goal of sentiment analysis is to ascertain the emotional tone behind words, which aids in understanding authors' attitudes, feelings, and opinions [22]. This field has experienced significant growth due to the proliferation of digital communication, meeting the needs of businesses, researchers, and policymakers to swiftly and accurately gauge public sentiment [36], [38].

979-8-3503-2771-7/23/$31.00 © 2023 IEEE

Methodologies in sentiment analysis are typically categorized into rule-based, machine-learning, and hybrid approaches. Rule-based systems utilize manually generated linguistic rules and predefined sentiment dictionaries. Although these systems are straightforward and interpretable, they often struggle to capture the nuanced expression of human emotions. Machine learning approaches, such as support vector machines (SVM), Naive Bayes, logistic regression, and advanced techniques involving deep learning models like Recurrent Neural Networks (RNN) and transformers such as Bidirectional Encoder Representations from Transformers (BERT), offer improvements by learning from data [40]. However, their effectiveness heavily depends on the quality and representativeness of the training datasets [2], [3], [39].

Twitter's ubiquity as a platform for real-time public expression makes it an ideal dataset for sentiment analysis, especially concerning global issues like climate change. The platform's rapid information dissemination capabilities and broad user engagement provide a rich dataset of public opinions, presenting unique opportunities and challenges for sentiment analysis [5], [8], [9], [23]. Twitter's structure allows for the capture of spontaneous and immediate reactions to global events, offering a live pulse on public sentiment that is unrivaled by other platforms.

Existing sentiment analysis models struggle with the complexity of social media language and the rapid evolution of online discourse [20], [21], [24]. There is a need for models that more accurately reflect the subtleties and dynamism of language used on social media, particularly in contexts like climate change, where public opinion is diverse and continually evolving.

This research contributes to the field of sentiment analysis in several key ways:

- We compare the effectiveness of two fundamentally different sentiment analysis models: VADER, a lexicon and rule-based approach, and RoBERTa, a transformer-based model, in analyzing sentiments expressed in tweets about climate change.

- We investigate the impact of feature selection techniques on the performance of these models to identify which approaches best capture the nuances of social media language.
- By analyzing climate change-related tweets, we extract key themes and trends, providing insights into public sentiment that can inform policy decisions, guide communication strategies, and assist environmental organizations in their advocacy efforts.

These contributions aim to enhance understanding of public sentiment on climate change, offering tools and insights that can be applied by policymakers, businesses, and researchers to better interpret and respond to public opinion [11], [15]. By providing a more granified view of how different groups perceive climate issues, our findings can help tailor communication and policy strategies that are more effective and resonant with diverse audiences. Additionally, the insights derived from our analysis could guide businesses in making informed decisions about green practices and sustainability initiatives that align with public sentiments.

The remainder of this paper is organized as follows: Section II reviews related work, highlighting existing sentiment analysis techniques and the evolution of models applied to climate change discourse on social media platforms. Section III provides a detailed explanation of the methodology, outlining the specific architectures and configurations of the VADER and RoBERTa models, as well as the preprocessing techniques and feature selection methods employed in this study. Section IV presents the experimental results, comparing the performance of both models on climate change-related tweets and evaluating the impact of sentiment categories and feature selection on model accuracy. Finally, Section V concludes the paper by summarizing the key findings, discussing their implications for sentiment analysis in climate change discourse, and proposing future research directions to enhance model robustness and sentiment detection on social media. This structured approach ensures detailed research coverage and its contribution to Natural Language Processing and sentiment analysis.

II. Related Work

As climate change continues to be a critical global issue, understanding public sentiment and discourse surrounding it becomes increasingly important. Sentiment analysis, particularly through social media platforms such as Twitter and Facebook, serves as a key tool in capturing this public sentiment. It allows researchers to analyze and interpret the emotional tone and prevailing attitudes found in online discussions, providing insights into the collective public consciousness [41].

Research consistently indicates that discussions about climate change on social media are predominantly negative. Public expressions often reflect concern, anxiety, and fear about the planet's future, with sentiments intensifying during extreme weather events or following significant climate-related news [10]. These observations suggest that such events act as catalysts, amplifying public concern and engagement in environmental discourse.

Geographic analysis further reveals variations in sentiment by region, particularly in areas directly impacted by climate events like hurricanes or wildfires, where negative sentiment is markedly higher. This geographical disparity underscores the profound influence of direct environmental experiences on public sentiment [5]. The analysis extends to longitudinal studies, which highlight fluctuations in public mood in response to environmental reports or policy announcements. Peaks in sentiment during these events indicate the public's heightened concern at critical moments, potentially informing the strategic decisions of policymakers and environmental organizations [25].

Social media platforms not only serve as vital forums for disseminating scientific information and expressing personal opinions but also as arenas for mobilizing climate action. Emotionally charged content—whether about extreme events, personal narratives, or urgent calls to action—tends to engage users more significantly, leading to broader sharing and discussion. However, skepticism and denial about climate change also prominently feature in the discourse, often polarizing public opinion and presenting challenges in communication strategies [7].

Network analysis provides insights into how information about climate change proliferates across social media, identifying key influencers such as environmental NGOs, scientists, and journalists. These individuals and groups play crucial roles in shaping discourse and amplifying key messages, thus influencing public perceptions and actions concerning climate change [44].

Advanced techniques such as Latent Dirichlet Allocation (LDA) are employed to categorize tweets on climate change, revealing prevalent topics like ”renewable energy,” ”policy and regulation,” and ”climate science” [12]. These topics reflect the broad spectrum of issues that engage the public and influence conversations on social media, illustrating the diverse aspects of climate change that captivate and concern the global community [18].

In summary, while sentiment analysis via social media offers a valuable lens for observing the global narrative on climate change in real time, it also presents several challenges. These include the complexity of language, the rapid evolution of online discourse, and the need for sophisticated analytical techniques to parse and understand vast amounts of data accurately [4]. This paper underscores the role of advanced analytical techniques in identifying and interpreting public sentiment, offering essential insights that can assist stakeholders in crafting more responsive and informed environmental policies and strategies.

III. Methodology Foundations

In this study, we employ the RoBERTa model to analyze tweets related to climate change, refining its capabilities on a dataset of sentiment-labeled tweets. This strategic choice leverages RoBERTa's advanced capacities for high-precision sentiment classification, particularly adept at unraveling the complexities of social media text. By leveraging this model, we aim to provide a nuanced and profound insight into public sentiment towards climate change, facilitating a deeper understanding of how public opinion forms and evolves in response to various climatic events and news.

A. VADER: Lexicon and Rule-Based Sentiment Analysis

VADER (Valence Aware Dictionary for sEntiment Reasoning) is a lexicon and rule-based tool designed specifically for the sentiment analysis of social media texts, such as tweets. It combines a sentiment lexicon with a set of grammatical rules to effectively evaluate the sentiment of short texts. Renowned for its simplicity and efficiency, VADER is particularly adept at capturing textual nuances such as capitalization, modifiers, and conjunctions, making it ideal for real-time sentiment analysis. This model has demonstrated robustness and reliability across various contexts, including microblogging platforms and social media, and is widely favored for preliminary sentiment assessments. VADER's unique ability to quickly process and analyze large volumes of data in real-time makes it an invaluable tool for the initial screening and analysis of sentiments in tweets [17].

B. RoBERTa: Enhanced Transformer Model for Deep Contextual Understanding

RoBERTa, an enhanced version of BERT (Bidirectional Encoder Representations from Transformers), is optimized for complex language understanding tasks. It improves upon BERT by training on a larger corpus and adjusting key hyperparameters, such as batch size and learning rates. This model excels at capturing deep contextual relationships within text, making it particularly effective for detailed sentiment analysis. In this research, RoBERTa was fine-tuned on a specific dataset to boost its efficacy in classifying sentiments expressed in social media text, often involving nuanced and context-dependent emotions. The enhancements in RoBERTa allow for a more detailed understanding of sentiment, enabling the model to capture subtle expressions of emotion that are crucial for accurately gauging public opinion on sensitive issues like climate change [27].

C. Preprocessing Techniques

The preprocessing of text data is a critical step in preparing inputs for sentiment analysis, involving multiple stages aimed at cleaning and standardizing the data for optimal analytical outcomes [13]. The detailed steps include:

1) **Data Cleansing**: This step involves the removal of punctuation, special characters, and numbers from the text. Such elements are typically non-contributory towards sentiment analysis, and their removal ensures the text is free from extraneous noise, thereby readying it for further processing [14].
2) **Tokenization**: This process segments the text into smaller units, or tokens, using both word-level and sentence-level tokenization. Tokenization is essential for capturing the granularity of sentiment and preparing the text for more nuanced analysis [19], [29].
3) **Stop Words Removal**: Commonly occurring words that generally do not carry significant meaning, such as ”the”, ”is”, and ”at”, are removed. This reduces the data's dimensionality and focuses the analysis on more impactful content, crucial for accurate sentiment assessment [16], [29].
4) **Lemmatization and Stemming**: Both techniques are employed to reduce words to their base or root form. Lemmatization considers the context and converts words to their meaningful base forms, while stemming cuts off the ends of words. We specifically utilize the Porter Stemmer, a well-established method known for its efficiency and effectiveness in reducing words to their stems [34], [35]. Although stemming, particularly with the Porter algorithm, is more aggressive than lemmatization, it is particularly useful in handling large datasets as it significantly reduces the computational complexity involved in processing the text. Employing these methods enhances the consistency and accuracy of the data analysis, allowing for more effective processing of textual data for sentiment analysis [6], [19], [28].

These preprocessing steps are crucial in refining the text data, enhancing the models' accuracy and efficiency for subsequent sentiment analysis. Each step is carefully designed to ensure that the data fed into our sentiment analysis models is as clean and structured as possible, maximizing the potential for accurate and insightful results.

IV. Experimental Evaluation

Understanding how different sentiment analysis tools perform when applied to social media discourse on climate change is crucial for extracting meaningful insights from vast amounts of textual data. This section details the experiments conducted to compare the effectiveness of two prominent sentiment analysis models, VADER and RoBERTa, in classifying sentiments expressed in tweets. Through a rigorous evaluation process, we aim to discern which model better captures the complexities of public sentiment on climate change, thereby providing deeper analytical insights that could influence environmental discourse and policy-making.

A. Dataset

The dataset utilized in this study comprises a carefully curated collection of 43,943 tweets spanning from April 27, 2015, to February 21, 2018. These tweets are centered around various facets of climate change discussions on Twitter, providing a rich corpus for analysis [1]. Rigorous annotation was undertaken by three independent reviewers, focusing on the high reliability of sentiment classification, with only those tweets receiving unanimous agreement on their sentiment categorization being retained for further analysis. This stringent selection criterion ensures the precision and accuracy of our dataset, which is vital for the validity of sentiment analysis. The tweets were categorized into four distinct sentiment groups, allowing for nuanced analysis of public opinion:

- **News (2)**: Tweets providing factual updates about climate change, typically linking to news articles or reports.
- **Pro (1)**: Tweets expressing affirmative support for the scientific consensus on anthropogenic climate change.
- **Neutral (0)**: Tweets that discuss climate change in a neutral or unbiased manner, without any apparent support or denial.
- **Anti (-1)**: Tweets that express skepticism or outright denial of anthropogenic climate change.

B. Results

The analysis revealed distinct patterns in sentiment across different categories of tweets, with RoBERTa generally outperforming VADER in detecting nuanced emotional expressions. Table I presents the sentiment scores assigned by VADER and RoBERTa across different tweet categories. These results are visualized in Figure 1, which illustrates the distribution of sentiments as analyzed by each model.

Both models reveal that the lowest positive sentiment is associated with Anti tweets, indicating a consistent negative tone in this category. VADER shows the highest positive sentiment on neutral tweets, whereas RoBERTa peaks for Pro tweets, illustrating a variance in how each model interprets positive sentiment within these contexts. Moreover, RoBERTa consistently produces higher positive sentiment scores across all categories compared to VADER, suggesting its heightened sensitivity to positive sentiment nuances.

VADER displays uniformly high neutral sentiment scores across all categories, consistently ranging between 0.8 and 0.85. Conversely, RoBERTa shows more significant variation, with news tweets receiving the highest neutral scores at 0.6 and reaction tweets the lowest at 0.4. This indicates that VADER treats tweets more homogeneously neutral, applying similar neutrality scores regardless of content type. In contrast, RoBERTa registers more nuanced differences in neutrality, reflecting its ability to capture subtler variations in tone and context across different tweet categories.

Regarding negative sentiment, both models agree that Anti tweets exhibit the highest negative sentiment, emphasizing a shared detection of negativity in this category. However, RoBERTa assigns higher negative sentiment scores across all categories than VADER, with a notably higher score for Anti tweets (0.5 compared to VADER's 0.12). This suggests that RoBERTa has a more robust capability for detecting negative sentiment, particularly in Anti tweets, reflecting its greater sensitivity to the intensity of negative emotions.

These results underline the differences between rule-based and transformer-based models in detecting and quantifying sentiment. The evaluation reveals a distinct performance gap between RoBERTa and VADER, with RoBERTa demonstrating a more favorable interpretation of supportive content and a more nuanced recognition of objective or factual information in news tweets. RoBERTa's performance suggests that transformer-based models may be more adept at capturing subtleties in emotional tone more effectively than rule-based models, providing a more refined evaluation of sentiment across diverse tweet types.

C. Discussion

The evaluation highlights significant differences in how VADER and RoBERTa interpret and classify sentiments within tweets about climate change. RoBERTa's superior capability to detect nuanced sentiments, particularly in the Pro and News categories, suggests its potential for providing deeper insights into public discourse, which is critical for shaping informed environmental policies.

1) Influence of Psychological and Societal Factors on Sentiment: The analysis also sheds light on the complex interplay of psychological and societal factors that influence public sentiment towards climate change. For instance, the positive sentiment among individuals indifferent to climate change often stems from a lack of concern or disbelief in the severity of the issue. Factors such as complacency, optimism bias, and exposure to misinformation play significant roles in shaping these attitudes [7], [26], [30], [42].

Optimism bias and complacency, where individuals believe that climate change will not personally affect them, can lead to a dismissive attitude towards proactive measures [42]. Additionally, the media's portrayal of climate change as a distant issue may reinforce such indifference [7]. On the other hand, coping mechanisms that involve maintaining a positive outlook despite the looming crisis highlight how psychological strategies influence public perception and discussion around climate change [37].

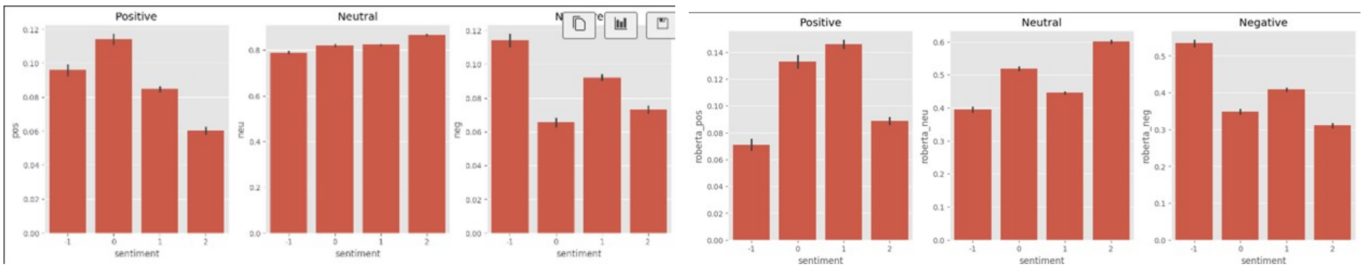
2) Polarization Between Supporters and Sceptics: The negative sentiment expressed by both supporters of climate change action and sceptics is emblematic of the polarized nature of this debate. Supporters often feel frustration and despair due to the slow pace of action and the continuous opposition, which can intensify negative emotions and advocacy efforts [33], [43]. Conversely, sceptics' distrust towards the scientific consensus and perceived threats to economic interests or personal freedoms fuel their negative sentiments [26], [31].

This polarization is crucial for understanding the broader discourse dynamics, as it not only affects individual sentiments but also influences collective action and policy-making. Recognizing these underlying sentiments can aid policymakers and environmental organizations in crafting strategies that address both the concerns of supporters and the skepticism of detractors.

3) Implications for Sentiment Analysis Tools:: The findings from this study underscore the importance of selecting appropriate sentiment analysis tools that can accurately interpret complex and context-dependent public sentiments. Transformer-based models like RoBERTa, with their advanced capabilities to handle nuanced language, appear more suited for analyzing sentiments on contentious topics like climate change, where expressions are often influenced by deep-seated beliefs and external factors.

TABLE I
COMPARATIVE SENTIMENT ANALYSIS USING VADER AND ROBERTA FOR TWEETS ON CLIMATE CHANGE

Tweet Type	VADER Scores			RoBERTa Scores		
	Positive	Neutral	Negative	Positive	Neutral	Negative
Anti	0.7	0.8	0.12	0.08	0.4	0.5
Neutral	0.11	0.8	0.04	0.13	0.5	0.3
Pro	0.09	0.8	0.05	0.15	0.45	0.35
News	0.06	0.84	0.06	0.10	0.6	0.3



(a) VADER

(b) RoBERTa

Fig. 1. Distribution of Positive, Neutral, and Negative Sentiment Scores across Different Tweet Categories by VADER and RoBERTa

V. CONCLUSIONS AND FUTURE WORK

Climate change elicits a broad spectrum of emotions and opinions, which are vividly expressed on social media platforms like Twitter. Through advanced sentiment analysis tools such as RoBERTa and VADER, this study has analyzed tweets related to climate change, revealing distinct emotional responses that vary significantly among different groups. The predominant sentiment among those indifferent to climate change is surprisingly positive, often driven by complacency, misinformation, and psychological coping mechanisms that frame the issue as distant and non-threatening [42]. Conversely, both supporters of climate action and sceptics primarily express negative sentiments, albeit for different reasons: supporters feel frustration and despair due to the slow pace of action and opposition, while sceptics harbor distrust towards the scientific consensus and the institutions advocating for policy changes [31].

This study underscores the complexity of public sentiments towards climate change and highlights the role of socio-psychological factors in shaping these attitudes. Understanding these dynamics is crucial for policymakers, educators, and communicators who aim to foster a more informed and constructive dialogue about climate change. Effective communication strategies must address these diverse viewpoints and psychological barriers to increase public engagement and support for climate action.

Future research should explore longitudinal sentiment analysis to monitor shifts in public opinion over time, particularly in response to significant environmental events or policy changes. This approach can provide insights into the dynamic nature of public sentiment and help assess the impact of communication strategies. Additionally, understanding how contextual factors influence sentiment expression could refine strategies for engaging different demographic and cultural groups.

Expanding sentiment analysis to multilingual and intercultural studies would offer a more comprehensive global perspective on climate change opinions. Such research could inform the creation of culturally sensitive communication strategies that resonate across diverse populations. Moreover, advanced emotion detection models could delve deeper into the specific emotions driving public attitudes, such as fear, anger, or hope, enhancing the effectiveness of targeted messaging.

Addressing the impact of misinformation is crucial, as it significantly shapes public sentiment. Research focused on the sources and spread of misinformation could lead to more effective strategies to counter its influence and promote scientific literacy [32]. Understanding the mechanisms through which misinformation affects climate change perceptions and decision-making processes can enable the design of targeted educational and communication campaigns that more effectively change minds and behaviors. Additionally, studying the correlation between expressed sentiments and actual behavioral changes could provide valuable insights into the efficacy of interventions aimed at promoting proactive responses to climate change.

As we advance our understanding of how sentiments towards climate change are formed and evolve, it is vital to integrate these insights into practical strategies that motivate and sustain public engagement and action. Recognizing the complex interplay between sentiment and behavior allows us to design interventions that not only raise awareness but also drive collective and individual action. By bridging the gap between emotional responses and informed action, we can more effectively tackle the multifaceted challenges of climate change. Strategies that address both emotional triggers and factual understanding can enhance resilience and adaptability in our societies, ensuring that responses to climate change are both swift and effective.

REFERENCES

[1] Twitter climate change sentiment dataset Online; accessed on 09 August 2024

[2] Acheampong, F., Henry, N.M., Chen, W.: Comparative analyses of bert, roberta, distilbert, and xlnet for text-based emotion recognition. In: 2020 17th International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP). pp. 117–121. IEEE (2020)

[3] Alexopoulos, A., Drakopoulos, G., Kanavos, A., Mylonas, P., Vonitsanos, G.: Two-step classification with SVD preprocessing of distributed massive datasets in apache spark. Algorithms **13**(3), 71 (2020)

[4] Alexopoulos, A., Drakopoulos, G., Kanavos, A., Sioutas, S., Vonitsanos, G.: Parametric evaluation of collaborative filtering over apache spark. In: 5th South-East Europe Design Automation, Computer Engineering, Computer Networks and Social Media Conference (SEEDA-CECNSM). pp. 1–8. IEEE (2020)

[5] An, X., Ganguly, A.R., Fang, Y., Scyphers, S.B., Hunter, A.M., Dy, J.G.: Tracking climate change opinions from twitter data pp. 1–6 (2014)

[6] Bird, S., Klein, E., Loper, E.: Natural Language Processing with Python: Analyzing Text with the Natural Language Toolkit. O'Reilly Media, Inc. (2009)

[7] Boykoff, M.T.: Who Speaks for the Climate?: Making Sense of Media Reporting on Climate Change. Cambridge University Press (2011)

[8] Camarillo, M.G., Ferguson, E., Ljevar, V., Spence, A.: Big changes start with small talk: Twitter and climate change in times of coronavirus pandemic. Frontiers in Psychology **12**, 661395 (2021)

[9] Clark, A., Fox, C., Lappin, S.: The handbook of computational linguistics and natural language processing **118** (2012)

[10] Cody, E.M., Reagan, A.J., Mitchell, L., Dodds, P.S., Danforth, C.M.: Climate change sentiment on twitter: An unsolicited public opinion poll. PloS One **10**(8), e0136092 (2015)

[11] Devika, M.D., Sunitha, C., Ganesh, A.: Sentiment analysis: A comparative study on different approaches. Procedia Computer Science **87**, 44–49 (2016)

[12] Dritsas, E., Trigka, M., Vonitsanos, G., Kanavos, A., Mylonas, P.: Aspect-based community detection of cultural heritage streaming data. In: 12th International Conference on Information, Intelligence, Systems & Applications (IISA). pp. 1–4. IEEE (2021)

[13] Dritsas, E., Vonitsanos, G., Livieris, I.E., Kanavos, A., Ilias, A., Makris, C., Tsakalidis, A.K.: Pre-processing framework for twitter sentiment classification. In: Artificial Intelligence Applications and Innovations (AIAI). vol. 560, pp. 138–149. Springer (2019)

[14] Feldman, R., Sanger, J.: The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data. Cambridge University Press (2007)

[15] Ford, J.D., Tilleard, S.E., Berrang-Ford, L., Araos, M., Biesbroek, R., Lesnikowski, A.C., MacDonald, G.K., Hsu, A., Chen, C., Bizikova, L.: Big data has big potential for applications to climate change adaptation. Proceedings of the National Academy of Sciences **113**(39), 10729–10732 (2016)

[16] Gupta, V., Lehal, G.S.: A survey of common stemming techniques and existing stemmers for indian languages. Journal of Emerging Technologies in Web Intelligence **5**(2), 157–161 (2013)

[17] Hutto, C., Gilbert, E.: Vader: A parsimonious rule-based model for sentiment analysis of social media text. In: International AAAI Conference on Web and Social Media. vol. 8, pp. 216–225 (2014)

[18] Jang, S.M., Hart, P.S.: Polarized frames on "climate change" and "global warming" across countries and states: Evidence from twitter big data. Global Environmental Change **32**, 11–17 (2015)

[19] Jurafsky, D.: Speech and language processing (2000)

[20] Kafeza, E., Kanavos, A., Makris, C., Pispirigos, G., Vikatos, P.: T-PCCE: twitter personality based communicative communities extraction system for big data. IEEE Transactions on Knowledge and Data Engineering **32**(8), 1625–1638 (2020)

[21] Kafeza, E., Kanavos, A., Makris, C., Vikatos, P.: T-PICE: twitter personality based influential communities extraction system. In: International Congress on Big Data. pp. 212–219. IEEE Computer Society (2014)

[22] Kanavos, A., Perikos, I., Hatzilygeroudis, I., Tsakalidis, A.K.: Emotional community detection in social networks. Computers & Electrical Engineering **65**, 449–460 (2018)

[23] Kanavos, A., Vonitsanos, G., Mohasseb, A., Mylonas, P.: An entropy-based evaluation for sentiment analysis of stock market prices using twitter data. In: 15th International Workshop on Semantic and Social Media Adaptation and Personalization (SMAP). pp. 1–7. IEEE (2020)

[24] Kanavos, A., Vonitsanos, G., Mylonas, P.: Clustering high-dimensional social media datasets utilizing graph mining. In: International Conference on Big Data. pp. 3871–3880. IEEE (2022)

[25] Kirilenko, A.P., Molodtsova, T., Stepchenkova, S.O.: People as sensors: Mass media and local temperature influence climate change discussion on twitter. Global Environmental Change **30**, 92–100 (2015)

[26] Lewandowsky, S., Ecker, U.K., Seifert, C.M., Schwarz, N., Cook, J.: Misinformation and its correction: Continued influence and successful debiasing. Psychological Science in the Public Interest **13**(3), 106–131 (2012)

[27] Liu, Y.: Roberta: A robustly optimized bert pretraining approach. arXiv Preprint arXiv:1907.11692 (2019)

[28] Lovins, J.B.: Development of a stemming algorithm. Mech. Transl. Comput. Linguistics **11**(1-2), 22–31 (1968)

[29] Manning, C.D.: Introduction to information retrieval (2008)

[30] Marx, S.M., Weber, E.U., Orlove, B.S., Leiserowitz, A., Krantz, D.H., Roncoli, C., Phillips, J.: Communication and mental processes: Experiential and analytic processing of uncertain climate information. Global Environmental Change **17**(1), 47–58 (2007)

[31] McCright, A.M., Dunlap, R.E.: The politicization of climate change and polarization in the american public's views of global warming, 2001–2010. The Sociological Quarterly **52**(2), 155–194 (2011)

[32] Mohasseb, A., Kanavos, A.: Grammar-based question classification using ensemble learning algorithms. In: 18th International Conference on Web Information Systems and Technologies (WEBIST). Lecture Notes in Business Information Processing, vol. 494, pp. 84–97. Springer (2022)

[33] Norgaard, K.M.: Living in Denial: Climate Change, Emotions, and Everyday Life. MIT Press (2011)

[34] Paice, C.D.: Another stemmer. In: ACM Sigir Forum. vol. 24, pp. 56–61 (1990)

[35] Porter, M.F.: An algorithm for suffix stripping. Program **14**(3), 130–137 (1980)

[36] Sahoo, C., Wankhade, M., Singh, B.K.: Sentiment analysis using deep learning techniques: A comprehensive review. International Journal of Multimedia Information Retrieval **12**(2), 41 (2023)

[37] Stoll-Kleemann, S., O'Riordan, T., Jaeger, C.C.: The psychology of denial concerning climate mitigation measures: Evidence from swiss focus groups. Global Environmental Change **11**(2), 107–117 (2001)

[38] Taboada, M.: Sentiment analysis: An overview from linguistics. Annual Review of Linguistics **2**(1), 325–347 (2016)

[39] Tan, K.L., Lee, C.P., Anbananthen, K.S.M., Lim, K.M.: Roberta-lstm: A hybrid model for sentiment analysis with transformer and recurrent neural network. IEEE Access **10**, 21517–21525 (2022)

[40] Vernikou, S., Lyras, A., Kanavos, A.: Multiclass sentiment analysis on covid-19-related tweets using deep learning models. Neural Computing and Applications **34**(22), 19615–19627 (2022)

[41] Vonitsanos, G., Kanavos, A., Mylonas, P.: Decoding gender on social networks: An in-depth analysis of language in online discussions using natural language processing and machine learning pp. 4618–4625 (2023)

[42] Weber, E.U.: Experience-based and description-based perceptions of long-term risk: Why global warming does not scare us (yet). Climatic Change **77**(1), 103–120 (2006)

[43] Whitmarsh, L., Seyfang, G., O'Neill, S.: Public engagement with carbon and climate change: To what extent is the public 'carbon capable'? Global Environmental Change **21**(1), 56–65 (2011)

[44] Williams, H.T., McMurray, J.R., Kurz, T., Lambert, F.H.: Network analysis reveals open forums and echo chambers in social media discussions of climate change. Global Environmental Change **32**, 126–138 (2015)