Heliyon 10 (2024) e32464

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

Exploring climate change discourse on social media and blogs using a topic modeling analysis

Tunahan Gokcimen, Bihter Das *

Department of Software Engineering, Technology Faculty, Firat University, 23119, Elazig, Turkey

ARTICLE INFO

Keywords:
Bibliometric Analysis
Latent Dirichlet Allocation(LDA)
BERTopic
Topic modeling
Climate change
Sentence similarity

ABSTRACT

Climate change is one of the most pressing global issues of our time, and understanding public perception and awareness of the topic is crucial for developing effective policies to mitigate its effects. While traditional survey methods have been used to gauge public opinion, advances in natural language processing (NLP) and data visualization techniques offer new opportunities to analyze user-generated content from social media and blog posts. In this study, a new dataset of climate change-related texts was collected from social media sources and various blogs. The dataset was analyzed using BERTopic and LDA to identify and visualize the most important topics related to climate change. The study also used sentence similarity to determine the similarities in the comments written and which topic categories they belonged to. The performance of different techniques for keyword extraction and text representation, including OpenAI, Maximal Marginal Relevance (MMR), and KeyBERT, was compared for topic modeling with BERTopic. It was seen that the best coherence score and topic diversity metric were obtained with OpenAI-based BERTopic. The results provide insights into the public's attitudes and perceptions towards climate change, which can inform policy development and contribute to efforts to reduce activities that cause climate change.

1. Introduction

Climate change is defined as the effects of global warming on the world climate system. Climate change has been considered as the most important global change problem for nearly 30 years. Human-induced (anthropogenic) climate change and its associated. As a result, changes in natural resources cause deterioration, severe weather, and climate events, and an increase in the severity of disasters. It also poses a significant threat to livelihoods and the sustainability of food security and has geopolitical consequences. According to Intergovernmental Climate Change the last evaluation report of the Panel (IPCC) [1], the main reason for the increase in temperature that causes climate change is human-induced factors. Climate change continues to make its impact felt all over the world. The increase in population, increasing housing needs, heavy traffic, and the decrease in the number of green areas are among the most important problems affecting air pollution [2,3]. The effects of climate change will increasingly reshape our security and peace in the coming years and will affect social geographies such as economy, population, energy, and industry and physical geographies such as atmosphere, air, water, ecology, vegetation, and soil. In short, it will be reflected in our lives in the form of serious social, economic, and political effects [4]. Governments have developed various restriction policies to reduce the factors that cause climate change. These restrictions are pollution tax, emissions trading policy and greenhouse gas emissions [5–7].

* Corresponding author.
E-mail addresses: tunahangokcimen@gmail.com (T. Gokcimen), bihterdas@firat.edu.tr (B. Das).

https://doi.org/10.1016/j.heliyon.2024.e32464
Received 26 January 2024; Received in revised form 30 May 2024; Accepted 4 June 2024
Available online 5 June 2024

It is very important to measure the public's knowledge, awareness, and reaction to climate change with comprehensive public opinion research. It is possible to learn how societies perceive climate change and how they interpret the developments they feel, from blog posts on the internet and social media platforms. Through these internet channels, people freely express their ideas and exchange information. It is undoubtedly very important to keep the pulse of society in this regard, but it also contributes to the adoption of fast and radical policies.

Climate change awareness surveys are studies to understand the public's perception of climate change [8]. It has often been used in the past to identify key factors, challenges, and barriers to environmental and climate-conscious behavior. In addition, it is aimed to shed light on the possible causes of air change and the factors affecting it. Since survey studies work on small amounts of data, information extraction is limited and is seen as a passive information disclosure process. In addition, due to the survey preparation process being long, it cannot measure the climate change sensitivity of the population in real-time, moreover, data collection requires a long time and high economic cost [9,10]. The amount of big data is increasing day by day with the social media shares and blog posts that people make every day. With the development of big data and natural language processing technologies, text analysis techniques have now begun to replace surveys [11].

Natural language processing (NLP), data visualization, and other techniques [12] have recently offered effective and efficient techniques to extract important and valuable information from social network data. Issues that concern society such as climate change are discussed in social media and blog posts, and people express their views on this issue. In addition, they share which topics they discuss with climate change, what their level of awareness is and what they do about the issue.

Climate change, stemming from the consequences of global warming on the Earth's climate system, stands as a foremost concern of our time, demanding concerted action from governments, communities, and individuals worldwide. Despite widespread recognition of the gravity of climate change, there exists a need for comprehensive assessments of public awareness, sentiments, and behaviors toward this multifaceted challenge. In this study, we endeavor to bridge this gap by delving into user-generated content from social media platforms and blog posts to discern prevailing attitudes and perceptions regarding climate change. By harnessing the power of natural language processing (NLP) and advanced data analytics techniques, we seek to extract valuable insights from vast troves of online discourse on climate-related topics. The specific objectives of our study include.

- To analyze user-generated content from social media platforms and blog posts to discern prevalent themes and sentiments related to climate change.
- To identify key research questions and hypotheses guiding our investigation into public perceptions and attitudes towards climate change.
- To employ advanced topic modeling techniques, such as BERTopic and Latent Dirichlet Allocation (LDA), to categorize and visualize topics within the collected dataset.
- To assess the semantic similarity of user comments and categorize them based on their relevance to identified climate change topics.
- To compare the performance of various keyword extraction and text representation techniques, including OpenAI, Maximal Marginal Relevance (MMR), and KeyBERT, in conjunction with BERTopic for topic modeling.

By elucidating these specific objectives, we aim to provide a clearer roadmap for our study, facilitating a more focused and coherent analysis of public perceptions of climate change. Through our research, we endeavor to contribute to the advancement of evidence-based policymaking and public discourse on this critical global issue. The following research questions (RQs) guided this study.

RQ1. How do individuals' perceptions and attitudes towards climate change vary across different topics and themes identified from user-generated content?

RQ2. What similarities exist among comments discussing similar topics, and how do these similarities contribute to understanding public discourse on climate change?

RQ3. How do different keyword extraction and text representation techniques, including OpenAI, Maximal Marginal Relevance (MMR), and KeyBERT, compare in their ability to identify and categorize climate change-related content from user-generated data?

The study investigates public discourse on climate change through analysis of user-generated content from social media platforms and blogs. The research questions guide the exploration of primary concerns, sentiments, and attitudes expressed by individuals regarding climate change, as well as the variations across different topics and themes. The hypotheses suggest that the content will demonstrate diverse sentiments and concerns influenced by individual characteristics and societal contexts, with variations across topics and themes. Furthermore, the study anticipates identifying similarities in language use and content structure within similar discussion contexts, providing insights into the collective perceptions and attitudes towards climate change in the public domain.

The study's contribution lies in its utilization of advanced topic modeling techniques, specifically BERTopic and LDA, to analyze public discourse on climate change. By employing these methods, the study provides a comprehensive exploration of the primary concerns, sentiments, and attitudes expressed by individuals across various topics and themes related to climate change. The use of BERTopic and LDA enables the identification and visualization of key topics from user-generated content on social media platforms and blogs, thus offering insights into the prevailing discourse surrounding climate change. This methodological approach enhances our understanding of public perceptions and attitudes towards climate change in a nuanced manner, thereby contributing to the broader discourse on environmental issues. Moreover, the comparison of BERTopic and LDA models adds to the methodological advancements in topic modeling research, showcasing their respective strengths and capabilities in analyzing user-generated data. The paper makes the following other contributions.

- By creating a new dataset, we identified and visualized the most important topics related to climate change using BERTopic and LDA from this dataset.
- We tried to catch the similarities in the comments written using sentence similarity and determined which topic category these comments belonged to.
- The performance of OpenAI, Maximal Marginal Relevance (MMR), and KeyBERT techniques are compared for keyword extraction and text representation for the topic modeling with BERTopic.

The rest of the paper is organized as follows. Section 2 presents the state-of-the-art related works on sentence similarity and topic modeling. Section 3 explains the dataset and methods for the experiment is explained in detail. Section 4 presents the experimental results and discussion. Section 5 concludes with the Conclusion.

2. Related works

This section provides a summary of studies state-of-art that have addressed the problem of sentence similarity and topic modeling.

2.1. *Studies on semantic similarity*

There are various approaches to finding word and sentence similarity in Natural Language Processing (NLP) applications [10,13]. These similarities can be found lexical or semantic. Recently, pre-trained word vectors created by deep learning have been used to find similarities.

Li et al., created word vectors for words by calculating the importance weight of the word with a corpus-based method to measure the similarity between words [14]. These word vectors were then used to measure the similarity between words. By representing words in a vector space based on their contextual usage within the corpus, they could effectively capture semantic relationships between words and compute their similarity. This approach allowed them to overcome the limitations of traditional methods and provide a more robust measure of word similarity.

Islam et al. propose a method focusing on short texts to measure semantic similarity between texts, utilizing both semantic word similarity and a normalized and modified version of the Longest Common Subsequence (LCS) string-matching algorithm. This study provides a significant tool that can be employed in various applications such as text mining and knowledge discovery. However, its limitation lies in the need for further information regarding the generalizability and applicability of the study to specific types of texts or languages [15].

Yoo et al. used a deep learning and lexical association analysis based on CNN and RNN methods to measure the similarity between two Korean sentences. In addition, the Bert pre-trained language model was used. Pearson and Spearman correlation coefficients were used to measure the performance of the system [16]. The limitation of this study lies in its focus on Korean sentences, which may restrict the generalizability of the proposed methodology to other languages or text types.

In the study conducted by Pawar et al., in 2018, the similarity between the two sentences was measured with the technique called WordNet. Sentence vector is created by using word similarity in sentences [17]. The reliance on WordNet for measuring sentence similarity is limited by the coverage and accuracy of WordNet's lexical database, potentially impacting the robustness of the similarity assessments.

Mamdouh's study measured similarity between sentences using Word Embedding and WordNet. In addition, word order similarity was used to measure the similarity of two sentences with different meanings [18]. Mamdouh's approach, despite its effectiveness in combining pre-trained word vectors, WordNet, and word order similarity, may encounter limitations related to potential gaps in WordNet coverage, variability in the quality of pre-trained word vectors, and the adequacy of word order similarity in capturing semantic nuances across diverse text types and contexts.

Lee et al. developed a corpus-based algorithm to measure grammatical and semantic similarity between two sentences [19]. While showing significant performance improvements in comparing sentences with arbitrary syntax and structure, Their grammar-based semantic similarity algorithm may face limitations related to the generalizability of its grammar rules to diverse linguistic patterns and the scalability of its approach to longer sentences or those with complex grammatical structures.

Tayal et al. measured the meaning between two sentences with an approach based on the WordNet structure [20]. Their method for determining semantic sentence similarity through WordNet-based word senses, while effective in many cases, may underestimate similarity scores in certain instances, leading to inaccurate results. This limitation arises from the reliance on WordNet's processing of available vocabulary, which may not capture all nuances of meaning accurately.

Ahmad et al. presented a hybrid algorithm that cares about sentence similarity and word order using a word database and corpus. The performance of the proposed algorithm was measured on two data sets. A high Pearson correlation result of 0.89 was obtained for the algorithm word and sentence similarity [21]. Ahmad and Faisal's hybrid methodology for computing semantic similarity between sentences may encounter limitations in accurately capturing nuances of meaning, particularly in specialized domains, despite its incorporation of lexical databases and word order information.

Apart from these, semantic search and sentence similarity have been used in many natural language processing applications such as sentiment analysis, natural language understanding, machine translation, answering questions, and developing chatbots [22].

Ferreira et al. examined the similarity between two sentences with a similarity algorithm that they proposed as lexical, syntactic, and semantic in their comprehensive study. They measured the performance of the proposed algorithm on different datasets such as the dataset by Li and collaborators and SemEval 2012. They used WordNet to evaluate the meaning of words [23].

2.2. Studies on topic modeling

Kirelli et al. analyzed this data by collecting social media tweets about global warming. Using machine learning methods, they categorize tweets as positive, negative, and neutral. The Support vector Machine (SVM) method obtained the highest performance with 74.63 % [24].

Barachi et al. used the Bi-directional Long Short-term Memory (BiLSTM) method to determine the mood of the public regarding global warming and climate change. They achieved the highest accuracy performance of 89.80 % [25].

Tong et al. conducted a study to model the most discussed topics in Tweets and Wikipedia. The latent Dirichlet Allocation model was used in the study [26].

Bergstedt et al. analyzed the comments on climate change in the Arctic and Alaska by collecting them from relevant Tweets. Two-class sentiment analysis as positive and negative, four-class cluster analysis, and daily distribution of tweets were made [27].

Samson et al. used an LDA and Bert-based model to analyze Twitter data on climate change. They categorized the data into support, neutral, and anti-groups. Uncased-Bert showed 93.50 % accuracy performance as working performance [28].

Chen et al. analyzed 491,279 tweets from the Twitter resource to measure the public's attitude towards climate change, so that the role of social robots in critical issues was determined [29]. In another study by Loureiro et al. analyzed tweets made in countries such as the United Kingdom, America and Spain and analyzed public opinion on climate change [30].

Wu et al. collected data from 169,592 blog posts from China's largest social media, Weibo, and analyzed the public's sensory orientation regarding climate change. In practice, a spatio-temporal difference analysis was made using Bidirectional Long Short-Term Memory (BiLSTM) and Latent Dirichlet Allocation (LDA), and the public's sensitivity to climate change was discussed. As a result of the analysis, it is understood that attention was drawn to three issues such as "green development/energy transformation", "the adverse effects of climate change on human life" and impact of climate change [31].

Ebeling et al. investigated the effects of political polarization with BERTopic and LDA using Twitter posts of two political groups in Brazil on social isolation [32].

3. Materials and methods

In this section, the data set collected from social media and blog posts about climate change, the text preprocessing steps applied to the data set, sentence similarity, and topic modeling are explained in detail. The study aims to analyze the attitudes and sensitivity of the public on climate change by using social media and blog posts. In addition, it is aimed to determine the most critical and current issues about climate change by making subject modeling with BERTopic and LDA algorithms. In this study, BERTopic and LDA modeling were used to analyze individuals' perceptions and attitudes towards climate change. The reasons for choosing these models include the following: BERTopic and LDA were chosen for topic modeling in this study due to their proven effectiveness in handling large-scale text data and their ability to generate interpretable topics [28,32]. BERTopic leverages contextualized word embeddings

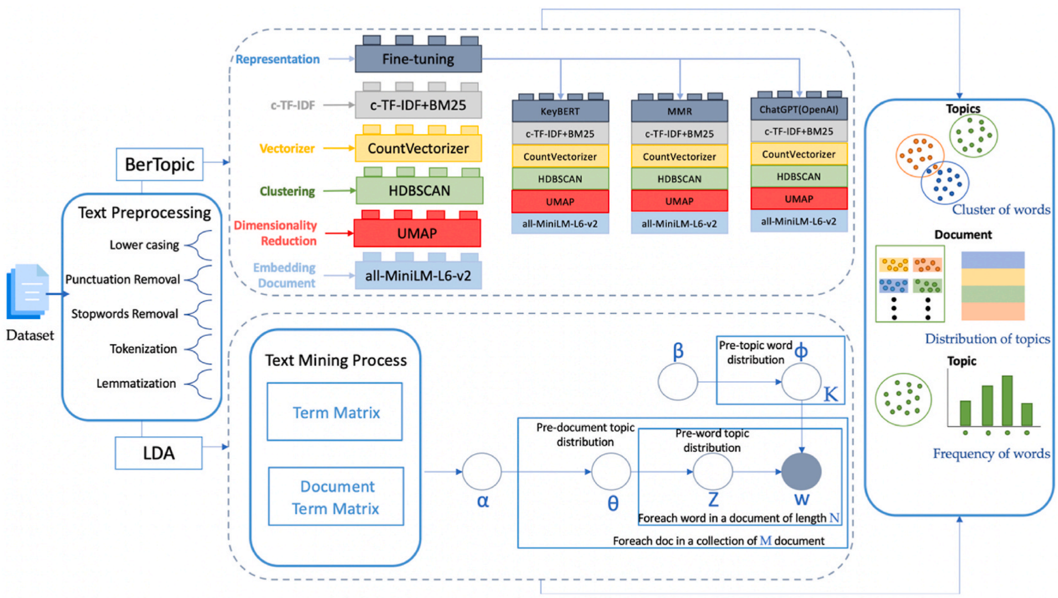


Fig. 1. The flowchart of the algorithm for topic modeling.

provided by BERT, which allows for a more nuanced understanding of semantic relationships within the text. On the other hand, LDA is a well-established probabilistic model that provides a clear probabilistic interpretation of topics based on word co-occurrence patterns. Another technique used in topic modeling in the literature is Non-negative Matrix Factorization (NMF). It is generally used in more limited areas. NMF matrix factors document and identify thematic parts under certain conditions, but are not as capable of creating distinct themes as BERTopic and LDA. BERTopic, with its contextualized embeddings, excels in capturing subtle semantic relationships and nuances in language usage, which is particularly relevant in the context of analyzing public discourse on climate change where the language can be complex and varied. LDA, on the other hand, offers a more traditional approach with a solid probabilistic foundation, making it suitable for comparison and complementarity with BERTopic. Both BERTopic and LDA generally perform well on large data sets and facilitate knowledge discovery with their visualization capabilities. For these reasons, BERTopic and LDA were determined as the most appropriate and effective options for this study. The semantic similarity of the words and sentences written by the users was also used with BERTopic in the classification of critical topics. In the BERTopic algorithm, OpenAI, Maximal Marginal Relevance (MMR), and KeyBERT techniques are used for keyword extraction and text representation. Their performance was compared in the representation part. Fig. 1 shows the flowchart of the topic modeling study.

3.1. Data collection

We collected data that contains 10.000 documents from social media, blogs, and forums about climate change written in English. The new dataset created by collecting the data can be accessed from this link: https://github.com/Gokcimen/Climate_Data. Fig. 2 shows the distribution of the dataset. In the data collection process, we sourced documents from various social media platforms, including Twitter, Facebook, and Instagram, as well as popular blogging platforms such as Medium, WordPress, and Tumblr. Forum posts were collected from platforms like Reddit and Quora. The selection criteria included relevance to climate change topics, the English language, and a balanced representation across different platforms. The keywords used as selection criteria are from 11 categories such as climate and consumption systems, global warming, energy transition, climate change, lowering energy and material intensities, strategies for managing resource supply and demand, valorizing waste, sustainable production and consumption, sustainability and productivity of resource use, resource productivity improvement, and recycling.

Selenium Web driver was used to collect data from websites. Table 1 presents detailed information about the dataset.

3.2. Data preprocessing

In this section, text normalization was performed with the preprocessing steps of the collected data. The steps in this section are as follows:

Text converting to lowercase: In this step, all uppercase letters in comments are converted to lowercase.

Removing numbers/punctuation: Some numbers and punctuation not relevant to the research have been removed.

Stop words removal: Stop words have been removed for topic modeling, where the text will be divided into different groups so that more attention can be paid to the words that determine the meaning of the text.

Lemmatization: In this step, we converted each of the tokens in the sentence into dictionary form to perform morphological analysis of a word called lemma. We converted plural words into singular and tense verbs into infinitives.

Tokenization: In this step, the text is divided into smaller parts called tokens, such as words. These are the smallest components called tokens. Fig. 3 shows the preprocessing operations for some example sentences of data set-

3.3. Background information on topic modeling

Topic modeling (TM) is a type of text mining in which unsupervised and supervised statistical data are used to investigate abstract topics within texts [33]. Unstructured texts are made understandable by TM algorithms and the information in the texts is made available by editing. Thus, time savings will be achieved. With TM, clusters called topics in the text are categorized. The frequency of these topics in the text and the fact that words sharing a similar theme are in a group are taken into account. With TM algorithms,

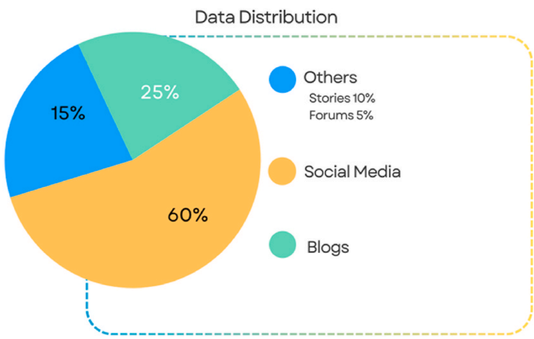


Fig. 2. The distribution of the dataset.

Table 1
The detailed information on the data.

Data source		Data size (%)	Description	Number of topics	Date	Language
Social media	Twitter	24 %	Tweets and comments from some accounts	10	From March 2022 to March 2023	English
	Facebook	10 %				
	Instagram	26 %				
Blogs	Medium	8.6 %	Blog posts	8	From March 2022 to March 2023	English
	WordPress	9.4 %				
	Tumblr	7 %				
Other	Reddit	7 %	Real-time stories	4	From March 2022 to March 2023	English
	Quora	4 %				
	Wattpad	4 %				

	label	label_without_stopwords	label_without_punctuation	tokenization_utt	label_lemmatized
0	valorizing waste is the process of turning was...	valorizing waste process turning waste resource.	valorizing waste process turning waste resource	[valorizing, waste, process, turning, waste, r...	valorizing waste process turning waste resource
1	It is a sustainable approach to waste management.	sustainable approach waste management.	sustainable approach waste management	[sustainable, approach, waste, management, .]	sustainable approach waste management
2	The aim is to reduce waste, conserve resources...	aim reduce waste, conserve resources, reduce p...	aim reduce waste conserve resources reduce pol...	[aim, reduce, waste, ,,conserve, resources,...	aim reduce waste conserve resource reduce poll...
3	Valorization can take many forms, including re...	valorization take many forms, including recycle...	valorization take many forms including recycle...	[valorization, take, many, forms, ,,including...	valorization take many form including recycling...
4	It is important to prioritize waste reduction...	Important prioritize waste reduction reuse con...	Important prioritize waste reduction reuse con...	[important, prioritize, waste, reduction, reus...	Important prioritize waste reduction reuse con...

Fig. 3. The preprocessing for dataset.

similar or synonymous words can be determined and documents containing those words can be listed [34]. In this study, the performances of the two most used and popular algorithms in the literature such as LDA and BERTopic for subject modeling will be examined comparatively, and the sentence similarity technique will be used together for BERTopic. The semantic similarities of words and sentences in subject modeling will also be examined.

Latent Dirichlet Allocation (LDA) Algorithms: LDA is a probability-based topic modeling method and identifies topics based on word weight in a text. At the core of LDA, topics have a probability distribution over words, and text documents have a probability distribution over topics. Each topic has a distribution over the word string [35]. LDA is an unsupervised learning algorithm and does not need predefined words. The number of topics is determined and labels are assigned to the topics according to the classes. As a result of these assignments, the model produces various local and global statistics. Local statistics determine how many words are assigned to the topics in each document. Global statistics show how many times each word is assigned to each topic for the whole document [36]. Table 2 shows the used parameters of the LDA model in the experiment.

BERTopic Algorithms: BERTopic, developed by Maarten Grootendorst in 2020 [37], is a topic modeling algorithm that combines transformative embeds and clustering model algorithms. Although similar in structure to Top2Vec, this technique is based on the procedure for generating subject representation with c-TF-IDF. BERTopic allows subject modeling by calculating the subject representation at each time step without the need to run the model several times. Thus, with BERTopic, topics can be easily interpreted and visualizations are possible. BERTopic topic modeling takes place in three stages: document placement, document clustering, and topic representation. In document placement, document placements are extracted, in document editing, Uniform Manifold Approximation and Projection (UMAP) is used for placement size reduction [38], and HDBSCAN is used for semantically clustering documents. Clustering technique is very important for the process. The accuracy of subject representations parallels clustering performance. HDBSAN is used because BERTopic captures structures of different densities well. In creating topic representation, topics are extracted and reduced using c-TF-IDF + BM25 [39]. In BERTopic, the TF-IDF document is set to cluster/categorical/topic to get an accurate representation of topics from the word bag matrix. c-TF-IDF is the corrected representation of TF-IDF and addresses the features that distinguish one set of documents from others. We used all-MiniLM-L6-v2 sentence transformers as the embedding model in the BERTopic, as it can capture the semantic similarity between documents quite well. Table 3 shows the used parameters of HDBSCAN, UMAP and BERTopic models in the experiment.

Table 2
The parameters of the LDA model.

Model	Parameters
LDA Model	corpus = doc_term_matrix, num_topics = 20, id2word = dictionary, passes = 15, random_state = 100, chunksize = 2000

Table 3
The parameters of the models.

Models	Parameters
HDBSCAN Model	min_cluster_size = 15, metric = 'euclidean', gen_min_span_tree = True, cluster_selection_method = 'eom', prediction_data = True
UMAP Model	n_neighbors = 15, n_components = 5, min_dist = 0.0, metric = 'cosine'
BERTopic Model	language = "english" calculate_probabilities = True ngram_range=(3,3)

3.4. Topic representation

Bag-of-Words and c-TF-IDF are key components of BERTopic. This method, which is quite fast, quickly generates several keywords regardless of the clustering task [40]. Topics are easily and quickly updated after training without the need to retrain the model. Although these provide good topic representations, there are a number of representation models that allow further fine-tuning of topic representations. In this study, OpenAI, KeyBERT, and MMR models were used and their performances were also compared.

Maximal Marginal Relevance (MMR): MMR stands for Maximal Marginal Relevance. It's a technique used in information retrieval to select a subset of relevant documents from a larger set of search results [41]. MMR prioritizes the difference and diversity of the document group. It uses two properties to represent a documentation group. One of these features is the importance of a document in itself (relevance), and the other is how different the document is from other selected documents (diversity). In BERTopic, MMR is used to rank the topics based on their similarity to the entire document corpus and to ensure that the most representative and diverse topics are selected for analysis.

KeyBERT: It is a Python library that uses the BERT language model to extract keywords or keyphrases from a given text. It's a keyphrase extraction algorithm that is based on fine-tuning a BERT model on the specific task of keyphrase extraction [42]. In BERTopic, keyBERT is used to extract the most relevant and informative words or phrases from the documents in the corpus and to represent the documents in a high-dimensional vector space. In short, keyBERT is a powerful tool that works with large volumes of text data that needs to extract meaningful keywords quickly and accurately.

OpenAI: OpenAI is the organization that developed and trained the GPT (Generative Pre-trained Transformer) models, including GPT-3, which is the largest and most powerful language model released by OpenAI [43]. ChatGPT, the large language model that was trained by OpenAI using the GPT-3 architecture [44]. In BERTopic, the OpenAI model is used to generate embeddings for each document in a corpus, which are then clustered to identify the most relevant topics. The quality of these embeddings is a key factor in the accuracy and interpretability of the resulting topics. In BERTopic, the OpenAI model is used to generate embeddings for each document in a corpus, which are then clustered to identify the most relevant topics. The quality of these embeddings is a key factor in the accuracy and interpretability of the resulting topics. Table 4 shows the selection of keywords of the KeyBERT, MMR, OpenAI and LDA models on a sample sentence.

3.5. Background information on sentence similarity

Sentence similarity is the determination of how similar two texts are to each other. Semantic similarity is used in many natural

Table 4
Keyword selection by KeyBERT, MMR, OpenAI, and LDA.

Default representation
Recycling|can|help|reduce|the|environmental|impact|of|the|fashion|industry|by|promoting|sustainable|fashion|production|practices|and|reducing|fashion|waste
KeyBERT
sustainable|waste|production|reduce|promoting|impact|recycling|industry|environmental|practices|reducing|production|fashion
MMR
Recycling|reduce|environmental|fashion|industry|promoting|sustainable|production|practices|waste|valorization|
LDA
Recycling|reduce|environmental|impact|industry|promoting|sustainable|fashion|production|practices|reducing|waste
OpenAI
recycling and sustainability:reduce the environmental impact of the fashion industry

language processing applications, such as sentiment analysis, translation, Question&Answer (Q&A), developing chatbots, and developing search engines to retrieve information [21]. In sentence similarity, the semantic information of the texts is extracted and how similar they are to each other is calculated with vectors [45]. The models used for similarity take the list of the source sentence and the sentence to be searched for similarity and calculate a similarity score. The semantic similarity between the two vectors is calculated by comparing the word insertions. For the similarity between the two vectors, cosine similarity was used in this study.

One of the most popular techniques, Cosine similarity, is used to measure similarity between vectors by calculating the cosine angle between two vectors for a multidimensional space in Fig. 4. It uses the dot product of two vectors to find the cosine angle. The cosine is equal to 1 when the angle is 0, and less than 1 when the angle is greater than 0.

4. Experimental results and discussion

Experimental results are shown separately for LDA and BERTopic models.

4.1. Research results for LDA model

Experimental results obtained using the LDA model will be represented in terms of "Intertopic distance map", "Coherence score", "Most talked combined topics in reviews", and "Distribution of topics".

In the LDA model, with the help of hashtags on social media and blogs, the twenty most popular topics that users talk about climate change are determined. These topics are global warming, climate change, sustainability and productivity of resource use, Energy transition, Sustainable production, Resource productivity improvement, Recycling, Including strategies for managing resource supply and demand, Valorizing waste, Lowering energy and Material intensities, energy consumption, energy management, water pollution, resource management, sustainable resource, ecosystem, innovation, lifecycle, global effects, water pollution. Fig. 5 shows climate change topics and their consistency.

Topic 13 overlaps 15 with topic 2 with 18 and 6, and topic 12 with 9 and 17. Topic 15 draws attention to actions such as focus, supply, management, requires, and operation related to climate change, while topic 13 focuses on general effects such as global effects, economy, and sector. words seem to be emphasized. In Topic 18, it is seen that the public focuses on the negative effects of climate change on areas such as fishing, air, agriculture, farming, and oceans. Topic 6, on the other hand, draws attention to industrial technologies and renewable systems. Topics 9 and 12 draw attention to the public's concern about the negative impact of environmental pollution on biodiversity, ocean, and water resources. Again, in topic 1, it is seen that the public is most concerned about the decrease in water resources and believes that the whole world will experience water shortages in the future. Studies conducted in recent years draw attention to the public's shares about climate change, and it is desired to measure the public's knowledge and interest in this field. The current study shows parallelism with other studies [46–48], and it is seen that the public is concerned about climate change. Fig. 6 shows the most talked about topics in reviews of both social media and blogs.

As can be seen in Fig. 6, the most talked about topic by the public is managing resource and demand. It is seen that the public wants to continue the natural life and stated that the environment and resources should be protected. Consuming of natural resources and climate change such matters cause air, water and soil pollution to such an extent that they threaten all living things. It is seen that the public is conscious of this issue and therefore it deals with the evaluation and recycling of wastes. So, the second most talked about topics are valorizing waste and recycling. As a matter of fact, various sustainable development goals all over the world emphasize the importance of having holistic environmental pollution [49,50]. Fig. 7(a) gives the distribution of the most talked about topics in topic 0. Fig. 7(b) shows topic 1, Fig. 7(c) shows topic 2, and finally, Fig. 7(d) shows topic 3. Fig. 8 shows the word clouds for topics 1–4.

The most spoken and most mentioned words in Fig. 8(a) are climate, change, food and impact. In Topic 0, it is seen that the public is concerned that climate change will have a negative impact on food such as agriculture and fruit and vegetable production. The topic that the public talks the most by using hashtags is climate change. While topics such as sustainability, production, responsibility, and consumption are discussed in Fig. 8(b), it is seen that the topic of sustainability, which is focused in Fig. 8(c), comes to the fore. In addition, the words promotion and importance are the most common words in conversations. In Fig. 8(d), it is predicted that the public mostly talks about pollution in social media and blogs, and comments are made about the effect of this pollution on water and other renewable resources.

In the analyzes carried out with the LDA model, the twenty topics that users talked about most about climate change were determined with the help of hashtags on social media and blogs. These topics include widely discussed themes such as global warming,

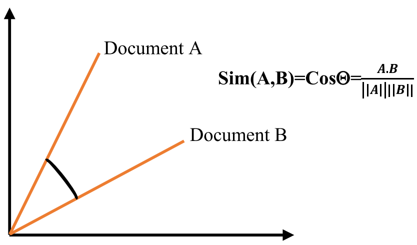


Fig. 4. The Cosine similarity.

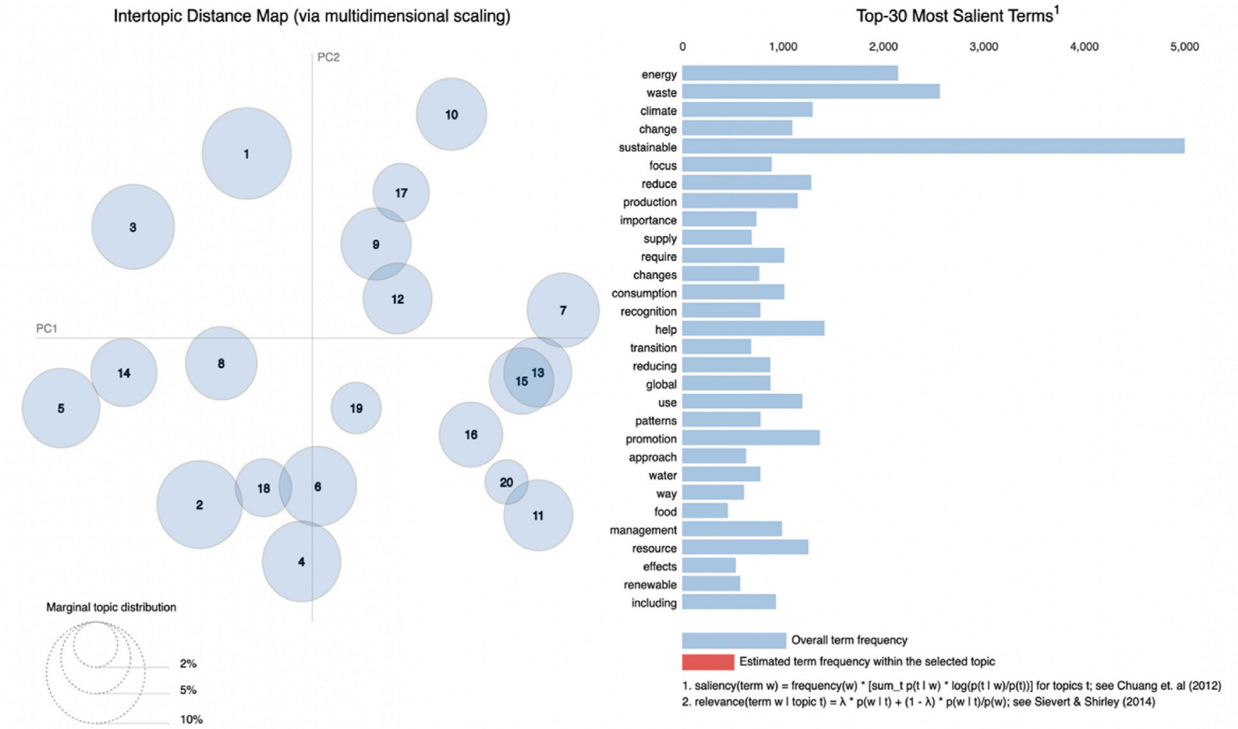


Fig. 5. The intertopic distance map.

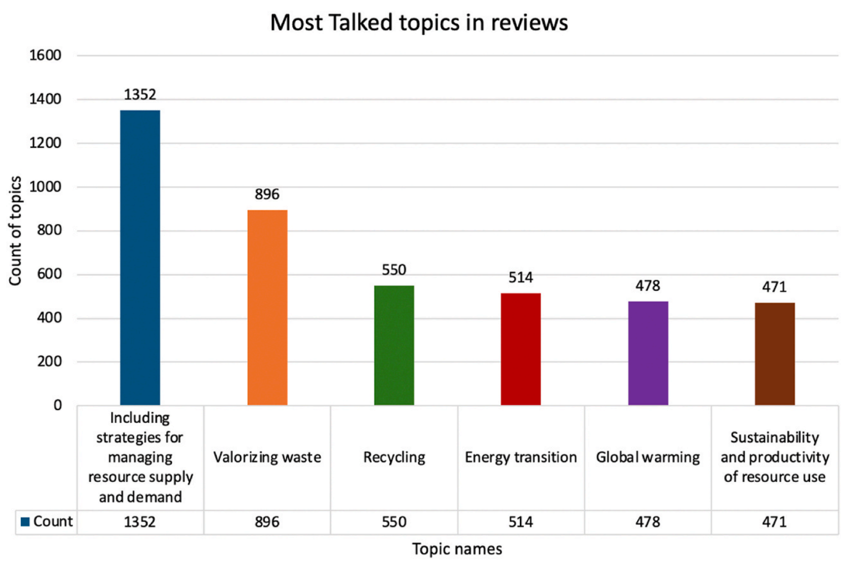


Fig. 6. The most talked about topics in reviews.

sustainability, energy transition and recycling. The results obtained are consistent with the findings in the literature [29,31]. In particular, it highlights the central role of issues such as sustainability and resource management in the public's perception of climate change.

4.2. Research results for BERTopic model

Experimental results obtained using the BERTopic model will be represented in terms of "Topic word score", "hierarchical clustering", "intertopic distance map", "term score decline per topic", "similarity matrix", and "word cloud".

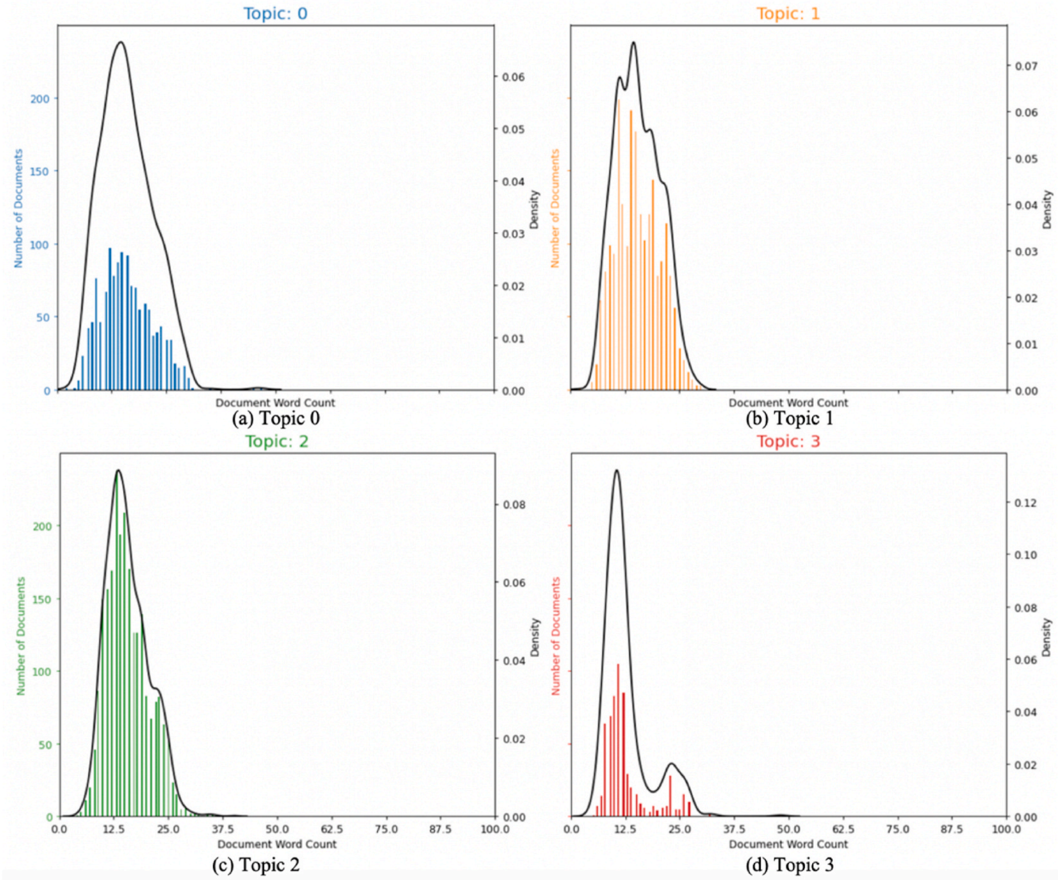


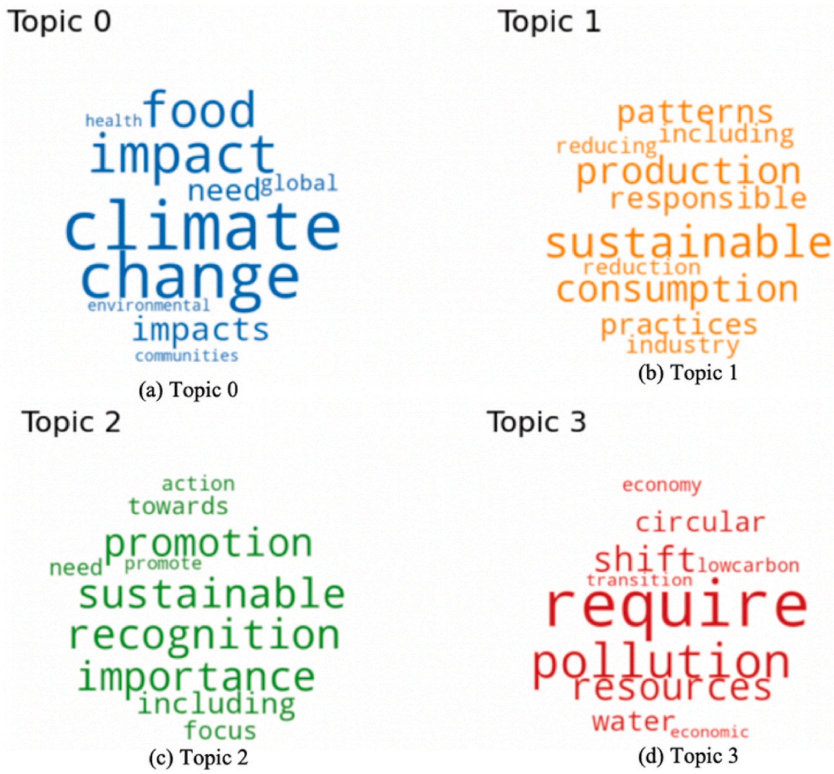
Fig. 7. Distribution of document word counts by dominant topic.

In BERTopic model, the performance of MMR, keyBERT and OpenAI techniques was compared for meaningful keywords extraction. Since the best coherence score performance is provided by the OpenAI technique, the figures obtained with OpenAI have been added to the paper. Fig. 9 shows the hierarchical clustering of the topics.

As seen in Fig. 9, the topics obtained with the BERTopic model formed a cluster according to their similarities. While these clusters are 7 with 17, 6 with 14, 16 with 2, 10 with 15, 0,5 and 13 represent a similar cluster. 12,4,1,8 and 11 constitute another subject cluster. When we examine the topics in more detail in the clusters, it is seen that the people argue that the recycling of food will reduce agriculture and waste. In the comments, it is seen that there are related conversations between effective supply and demand and supply chain organizations. In another cluster, it is seen that there is a relationship between metal density and the global effects of climate change and environmental protection topics. Fig. 10 shows the hierarchical documents and topics according to topics.

Fig. 10 shows the distribution of topics discussed by the public according to social media(D1) and blogs + forums(D2) documents. It is seen that people are discussing the effects of climate change on the ecosystem, global warming, the Arctic ice sea, and tourism on social media. While the number of topics discussed in social media is high, it is seen that fewer topics are mentioned in blogs and forums. In these documents, scientific subjects such as circular model business, permafrost melting release, and waste reduction were mentioned. At the same time, climate change is one of the most talked about topics in both blogs and social media. Fig. 11 shows the similarity matrix of topics.

As seen in Fig. 11, there are similarities between the topics discussed by the public. According to the matrix, the "severe effect world" topic is very similar to the 4th and 7th topics. Although the comments the public writes differ, they are discussing similar hot topics. Although some topics are very similar, it is understood that the topic of "permafrost melting release" bears little resemblance to the others or even no similarity to the tourism industry. Experimental results further elucidate and validate the effectiveness of the approach of detecting sentence similarity and associating comments with topic categories. Sentence similarity analysis identified similarities in user comments across different social media platforms and blogs, helping to uncover common themes and emotional trends. In particular, sentence similarity analysis allowed us to identify similar expressions and emotional tones associated with specific topic categories. This approach allowed us to better understand how certain topics and emotional trends are common and how they may vary across different user groups. Additionally, association with topic categories helped us determine the distribution and importance of specific topics and keywords in comments. In this way, it was possible to understand which topics users focused on more and which topics received more attention. Therefore, the approach of identifying sentence similarity and correlating comments with



(a) Topic 0 (b) Topic 1 (c) Topic 2 (d) Topic 3

Fig. 8. The word cloud for topics 1-4.

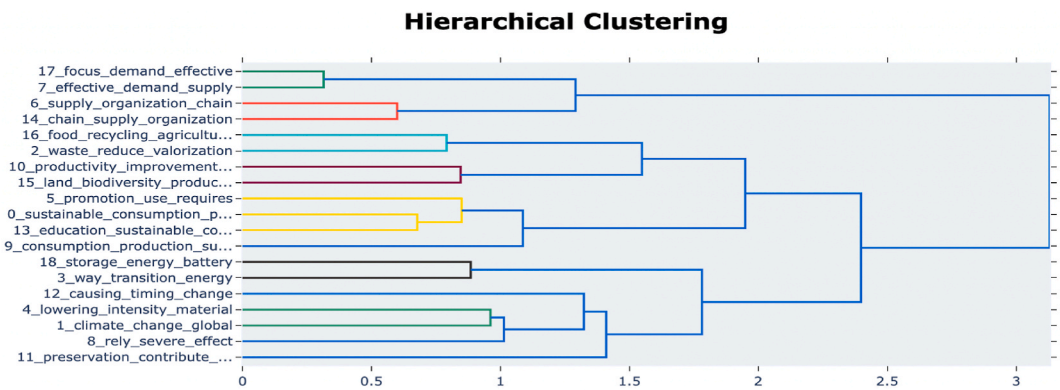


Fig. 9. The hierarchical clustering for topics.

topic categories provided a comprehensive analysis, helping us to understand more deeply the contribution of users to climate change discussions on online platforms. The cosine similarity technique played a crucial role in the approach of detecting sentence similarity and associating comments with topic categories. By utilizing cosine similarity, we were able to quantitatively measure the similarity between pairs of sentences or comments based on their semantic content. This allowed us to identify similarities in user comments across different social media platforms and blogs, even if the wording differed slightly. For example, in the context of climate change discourse, cosine similarity enabled us to identify common themes and emotional trends expressed by users, regardless of the specific wording or language used in their comments.

Social media and blog posts over the 12-month period were collected and used to analyze the content of the word cloud and public concern about climate change. As can be seen in Fig. 12, the words that are most discussed and worried by the public are promotion, climate change, sustainable resource, help reduce.

Table 5 shows the results of performance metrics of the LDA model, BERTopic model with MMR, OpenAI, and keyBERT models for topic modeling. In the study, evaluation metrics commonly used in the literature were selected for topic modeling [51,52]. These

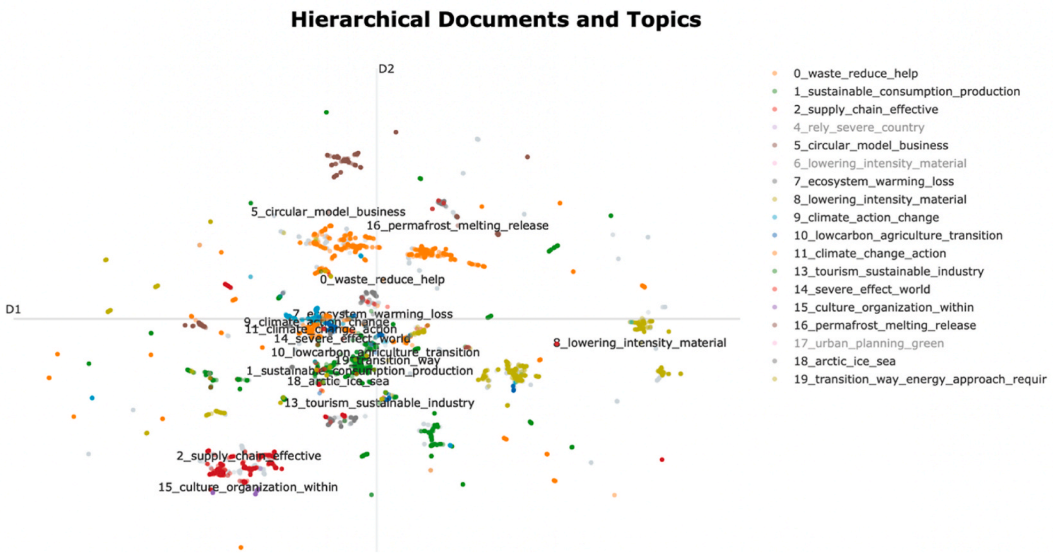


Fig. 10. The hierarchical documents and topics for document types.

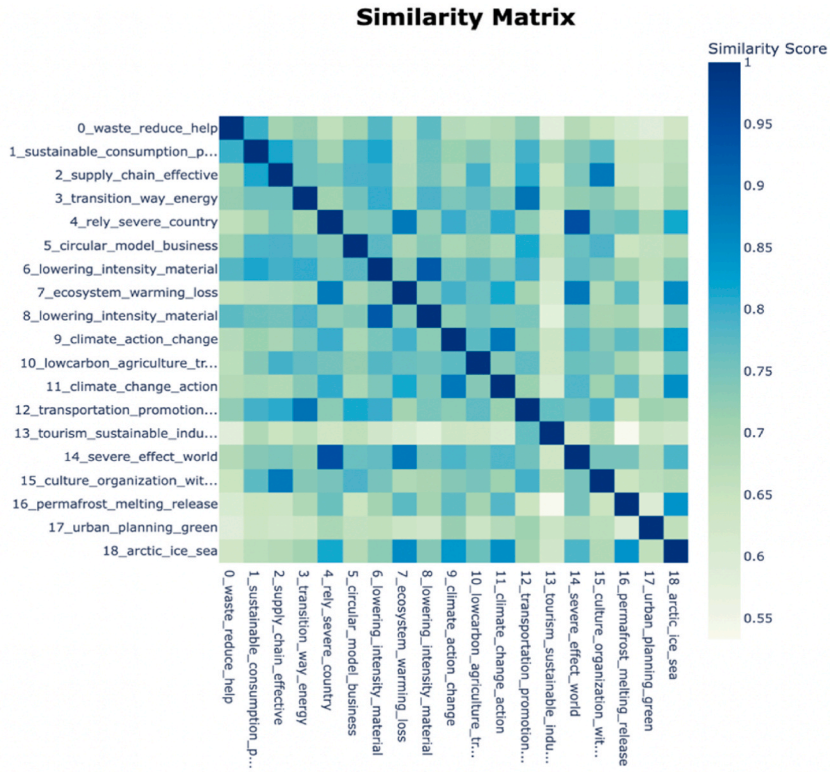


Fig. 11. The similarity matrix of topics.

metrics: Topic Coherence, Normalized Pointwise Mutual Information (NPMI), (C_v) and Topic Diversity. Topic Coherence measures the coherence of the topics a model produces and evaluates the relationship between the words contained in those topics. NPMI measures the likelihood of certain words being used together, thereby determining the coherence and meaningfulness of a topic. While the C_v measure evaluates the similarities and consistencies between words contained in a particular topic, Topic Diversity measures the diversity of topics produced by a model. These metrics are used to evaluate the performance of topic modeling methods, helping to determine the quality and consistency of the topics produced by each model. In particular, a model with a high topic consistency and

Fig. 12. The climate change word cloud.

Table 5
The results of performance metrics for models.

Models	Topic Coherence NPMI	Topic Coherence (C_v)	Topic Diversity
NMF	0.048	0.5164	0.75
LDA	0.051	0.5220	0.81
BERTopic with keyBERT	0.1063	0.6521	0.83
BERTopic with MMR	0.124	0.6690	0.89
BERTopic with Open AI	**0.0978**	**0.7348**	**0.91**

diversity score is considered to perform better. Therefore, these criteria play an important role in the field of topic modeling and are used to compare different model options. According to the given results, BERTopic with MMR and BERTopic with Open AI models are the best-performing methods as they have the highest topic consistency and diversity scores.

Table 5 shows that the BERTopic model performs much better than the LDA and NMF models. It is also seen that the NMF model has the lowest topic consistency and diversity compared to other models. NMF operates under the constraint that all values in the input data are greater than zero and non-negative. This means that NMF may be limited in fully representing complex topics in certain texts. They also have limitations in accurately understanding contrasting topics or contradictory themes in texts. For these reasons, it is thought to have poor performance. On the other side, when we compare the performance of BERTopic and LDA, we can see that the performance of LDA is lower in the experiment. This is because BERTopic and LDA models use different algorithms for topic modeling. BERTopic extracts topics using a machine learning model called BERT, while LDA identifies topics in text based on the probabilistic model. Unlike LDA, BERTopic is not dependent on vocabulary and word order. Thus, it can more accurately capture the relationships between different variations of word order and word usage. Also, unlike BERTopic LDA, it does not consider each word individually but instead focuses on groups of words. It allows it to more accurately capture the broader meaning of a text. Moreover, unlike BERTopic LDA, it offers a more flexible approach to determining hyperparameters such as several subjects. BERTopic can automatically determine the number of topics in a text, users do not need to manually select the number of topics. In the analyses made with the BERTopic model, the performance of MMR, KeyBERT, and OpenAI techniques was compared. The results show that the OpenAI-based BERTopic model provides a higher consistency score than other techniques. These results emphasize the contribution of the study to the literature, especially in terms of information extraction and keyword extraction. The reason why OpenAI outperforms MMR and KeyBERT is because OpenAI is one of the most advanced machine learning models recently released and is based on GPT-3, which provides better natural language processing capabilities. When choosing keywords, OpenAI takes into account not only word frequency and position but also word meanings, which helps in choosing more meaningful and consistent keywords. Analyzes with LDA and BERTopic models identified which main topics users were discussing in content about climate change on social media and blogs. For example, themes such as sustainability, resource management, and environmental impact were among the topics commonly discussed among users. These results and analysis of user-generated content suggest that social media and blogs play an important role in the perception of climate change and that certain issues reflect public concern.

The study's findings provide valuable insights into climate change discourse on social media platforms and blogs. Through topic modeling and sentence similarity analysis, we uncovered prominent themes and sentiments expressed by the public regarding climate change. In the context of topic modeling, our analysis revealed several important themes in climate change discourse. For example, issues related to sustainability, resource management, and environmental impact are closely aligned with themes commonly discussed in the literature [24–26]. Our study, relative to others, contributes to a detailed understanding of societal perceptions of climate change by explaining the pervasiveness and interconnectedness of these issues in public discourse. Furthermore, the examination of sentence similarity sheds light on the semantic nuances underlying public discussions on climate change. According to previous research on semantic similarity and sentiment analysis [19,21,23], we can distinguish patterns in language use and emotion expression across different social media platforms and blogs. For example, the similarity in language usage between discussions on climate change impacts and sustainable practices reflects a growing awareness of environmental issues among the public. Moreover, our study's integration of topic modeling and sentence similarity analysis offers a comprehensive perspective on climate change discourse, capturing both thematic trends and linguistic patterns. This interdisciplinary approach is in line with recent advancements in computational social science and natural language processing [28–32]. By synthesizing findings from these analyses with insights from the literature, it could be discerned underlying factors shaping public perceptions of climate change and identify potential areas for future research. Our study differs from the studies in the literature in a few points:

First, we used big data and text mining methods instead of traditional methods (survey or interview) [8–10]. In this way, we have obtained more objective and real-time results. Secondly, we compared the performance of two important topic modeling methods such as LDA and BERTopic on the same dataset. In the BERTopic method, we also compared the MMR, KeyBERT, and OpenAI word extraction techniques and determined that the OpenAI-based BERTopic performed best. Thus, we obtained a different approach and a better result than the methods used in other studies in the literature [25,31,53,54].Third, in our study, we also examined the differences in attention and emotion orientations to climate change-related issues by using the comments of people in different regions and user types on social media, blogs, and forum sites. This provided us with a different approach and perspective than other studies in the literature.

The identified topics in our experimental results align with current climate change discourse in both policy and media to a significant extent. Themes such as sustainability, energy transition, resource management, and environmental impact are commonly discussed in both scientific literature and public discourse [24–26]. This alignment suggests that the topics identified through our analysis accurately reflect the ongoing conversations surrounding climate change in various societal spheres. Understanding these topics and their prominence in public discourse could be crucial for policymakers and media professionals seeking to engage the public on climate change issues effectively. By recognizing the key themes and sentiments expressed by the public, policymakers can tailor their strategies and policies to resonate with public concerns and priorities. Likewise, media outlets can leverage these insights to shape their coverage of climate change in a way that resonates with their audience, fostering greater awareness and engagement. Overall, our findings underscore the importance of aligning communication efforts with the topics and sentiments prevalent in public discourse to maximize their impact on addressing the challenges posed by climate change.

The results contribute to understanding the public's perception of climate change and inferring information on this subject. However, this study has some limitations. The dataset used in this study was collected over a limited time period and focused on English-language content, potentially limiting its representativeness and generalizability to other regions and demographics. Future studies can address these limitations by collecting data over longer periods of time, expanding the dataset to include content from different linguistic and cultural backgrounds, and using datasets from different geographical regions. Additionally, investigating temporal changes in public perceptions and comparing perceptions across different regions or demographic groups can provide a more comprehensive understanding of climate change discourse.

5. Conclusion and future studies

The importance of understanding public perceptions and attitudes towards climate change cannot be overstated in the face of its significant global ramifications. In order to contribute to this understanding, a content analysis study was conducted using natural language processing techniques and data visualization techniques in order to learn about the concerns of the public about climate change from various social media and blogs and to measure the level of consciousness and awareness. In the study, the performance of two important topic modeling methods, LDA and BERTopic, were compared on the same data set and it was seen that the BERTopic with Open AI model showed the best performance overall because it had the highest topic consistency and topic diversity scores. Our study differs from previous studies in the literature by using big data and text mining methods, comparing different topic modeling and keyword extraction techniques, and examining differences in attention and emotion orientations to climate change-related issues among people and user types in different regions. social media, blogs and forum sites. These findings provide a unique perspective on public perceptions and attitudes towards climate change and can inform policy development and efforts to reduce climate change-causing activities.

In future studies, it is planned to develop a more comprehensive data collection and analysis method for the climate change. For this purpose, it is aimed to collect more data, including other data on social media platforms, especially visual data (photos and videos). Also, an international perspective can be captured to more thoroughly examine the differences between different cultures and countries. Attitudes of different cultures and countries regarding climate change and the shaping of the social, economic and political structure of these cultures can be examined. In addition, it is planned to provide suggestions to make the use of social media platforms more effective in order to increase awareness about climate change with future studies. Social media platforms, blogs, and forums are powerful tools for the dissemination of information, awareness and events about climate change. However, the results of this study revealed that some social media, blogs users spread misleading or false information and others were uninterested in climate change. Therefore, social media and blogs platforms can take advantage of the results of such studies to develop more effective methods to curb the spread of such misleading and false information. In addition, the findings from our study can be used by institutions and organizations fighting climate change. Policy makers around climate change can benefit from this type of research to understand the issues and concerns that come to public attention.

CRediT authorship contribution statement

Tunahan Gokcimen: Visualization, Validation, Software. **Bihter Das:** Writing – review & editing, Writing – original draft, Visualization, Supervision, Methodology, Conceptualization.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:Bihter Das reports financial support was provided by Arcelik A.S. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

This work is supported by the Republic of Turkey, Ministry of Science, Technology and Industry project named "AI Based Smart Digital Assistant Customer Dialog Bot project" and project code AR-22-087-0001. It is funded by R&D project within the scope of law 5746 by the Arçelik Digital Transformation, Big Data and Artificial Intelligence R&D Center.

References

[1] IPCC, Summary for policymakers, in: V. Masson-Delmotte, et al. (Eds.), Climate Change 2021: the Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change, Cambridge University Press. In Press, 2021.
[2] N.N.M. Hlapisi, Enhancing hybrid spectrum access in CR-IoT networks: reducing sensing time in low SNR environments, Mesopotamian Journal of Computer Science 2023 (Mar. 2023) 47–52.
[3] M. Turkes, M.T. Turp, N. An, T. Ozturk, M.L. Kurnaz, Impacts of climate change on precipitation climatology and variability in Turkey, in: N.B. Harmancioglu, D. Altinbilek (Eds.), Water Resources of Turkey, World Water Resources. Cham: Springer International Publishing, 2020, pp. 467–491, https://doi.org/10.1007/978-3-030-11729-0_14.
[4] N. An, M.T. Turp, M. Türkeş, M.L. Kurnaz, Climate change effects on agricultural production: a short review, Curr. Inves. Agri. Curr. Res. 8 (3) (2020) 288–292, https://doi.org/10.32474/CIACR.2020.08.000288.
[5] Y. Li, Y. Wei, Estimating the potential CO2 emission reduction in 97 contracting countries of the Paris agreement, Clim. Chang. Econ. 12 (1) (2021) 2150004, https://doi.org/10.1142/S2010007821500044.
[6] M. Stuhlmacher, S. Patnaik, D. Streletskiy, K. Taylor, Cap-and-Trade and emissions clustering: a spatial-temporal analysis of the European union emissions trading scheme, J. Environ. Manag. 249 (2019) 109352, https://doi.org/10.1016/j.jenvman.2019.109352.
[7] Y. Wei, P. Gong, J. Zhang, L. Wang, Exploring public opinions on climate change policy in 'big data era'—a case study of the European union emission trading system (EU-ETS) based on twitter, Energy Pol. 158 (2021) 112559, https://doi.org/10.1016/j.enpol.2021.112559.
[8] F. Jost, A. Dale, S. Schwebel, How positive is 'change' in climate change? A sentiment analysis, Environ. Sci. Policy 96 (2019) 27–36, https://doi.org/10.1016/j.envsci.2019.01.013.
[9] A.P. Kirilenko, S.O. Stepchenkova, Public microblogging on climate change: one year of Twitter worldwide, Global Environ. Change 26 (2014) 171–182.
[10] A.P. Kirilenko, T. Molodtsova, S.O. Stepchenkova, People as sensors: mass media and local temperature influence climate change discussion on Twitter, Global Environ. Change 30 (2015) 92–100.
[11] Mehmet Pekkaya, Zafer Uysal, Aytaç Altan, Seçkin Karasu, Artificial intelligence-based evaluation of the factors affecting the sales of an iron and steel company, Turk. J. Electr. Eng. Comput. Sci. 32 (1) (2024), https://doi.org/10.55730/1300-0632.4055. Article 4.
[12] C. Yanarateş, S. Okur, A. Altan, Performance analysis of digitally controlled nonlinear systems considering time delay issues, Heliyon 9 (10) (Oct. 2023) e20994, https://doi.org/10.1016/j.heliyon.2023.e20994.
[13] T. Kenter, M. de Rijke, Short text similarity with word embeddings, in: Proceedings of the 24th ACM International on Conference on Information and Knowledge Management, 2015, pp. 18–23. Melbourne, Australia, Oct.
[14] Y. Li, D. McLean, Z. Bandar, J. O'Shea, K.A. Crockett, Sentence similarity based on semantic nets and corpus statistics, IEEE Trans. Knowl. Data Eng. 18 (8) (2006) 1138–1150.
[15] A. Islam, D. Inkpen, Semantic text similarity using corpus-based word similarity and string similarity, ACM Trans. Knowl. Discov. Data 2 (2) (Jul. 2008) 1–10.
[16] Y. Yoo, T.-S. Heo, Y. Park, K. Kim, A novel hybrid methodology of measuring sentence similarity, Symmetry 13 (2021) 1442, https://doi.org/10.3390/sym13081442.
[17] A. Pawar, V. Mago, Calculating the similarity between words and sentences using a lexical database and corpus statistics, arXiv, doi: 10.48550/arXiv.1802.05667 (Feb.2018).
[18] M. Farouk, Sentence semantic similarity based on word embedding and WordNet, in: Proceedings of the 13th International Conference on Computer Engineering and Systems (ICCES), 2019. Cairo, Egypt.
[19] M.C. Lee, J.W. Chang, T.C. Hsieh, A grammar-based semantic similarity algorithm for natural language sentences, Sci. World J. 2014 (2014).
[20] M.A. Tayal, M.M. Raghuwanshi, L. Malik, Word net based method for determining semantic sentence similarity through various word senses, in: Proceedings of the 11th International Conference on Natural Language Processing, 2014. Goa, India.
[21] F. Ahmad, M. Faisal, A novel hybrid methodology for computing semantic similarity between sentences through various word senses, Int. J. Cogn. Comput. Eng. 3 (2022) 58–77, https://doi.org/10.1016/j.ijcce.2022.02.001.
[22] Baeldung, Semantic similarity of two phrases [Online]. Available: https://www.baeldung.com/cs/semantic-similarity-of-two-phrases. (Accessed 28 April 2023).
[23] R. Ferreira, R.D. Lins, S.J. Simske, F. Freitas, M. Riss, Assessing sentence similarity through lexical, syntactic and semantic analysis, Comput. Speech Lang 39 (2016) 1–28, https://doi.org/10.1016/j.csl.2016.01.003.
[24] Y. Kirelli, S. Arslankaya, Sentiment analysis of shared tweets on global warming on twitter with data mining methods: a case study on Turkish language, in: Computational Intelligence and Neuroscience, vol. 2020, Hindawi Limited, 2020, pp. 1–9, https://doi.org/10.1155/2020/1904172.
[25] M. El Barachi, M. AlKhatib, S. Mathew, F. Oroumchian, A novel sentiment analysis framework for monitoring the evolving public opinion in real-time: case study on climate change, J. Cleaner Prod. 312 (2021) 127820, https://doi.org/10.1016/j.jclepro.2021.127820.
[26] Z. Tong, H. Zhang, A text mining research based on LDA topic modelling, in: Computer Science & Information Technology (CS & IT). The Sixth International Conference on Computer Science, Engineering and Information Technology, Academy & Industry Research Collaboration Center (AIRCC), 2016, https://doi.org/10.5121/csit.2016.60616.

[27] H. Bergstedt, A. Ristea, B. Resch, A. Bartsch, Public perception of climate change in Alaska: a case study of opinion-mining using twitter, in: GI_Forum, vol. 1, Osterreichische Akademie der Wissenschaften, 2018, pp. 47–64, https://doi.org/10.1553/giscience2018_01_s4.
[28] S.E. Uthirapathy, D. Sandanam, Topic modelling and opinion analysis on climate change twitter data using LDA and BERT model, Procedia Comput. Sci. 218 (Jan. 2023) 908–917, https://doi.org/10.1016/j.procs.2023.01.071.
[29] C.F. Chen, W. Shi, J. Yang, H.H. Fu, Social bots' role in climate change discussion on Twitter: measuring standpoints, topics, and interaction strategies, Adv. Clim. Change Res. 12 (6) (2021) 913–923.
[30] M.L. Loureiro, M. Alló, Sensing climate change and energy issues: sentiment and emotion analysis with social media in the UK and Spain, Energy Pol. 143 (2020) 111490.
[31] M. Wu, R. Long, F. Chen, H. Chen, Y. Bai, K. Cheng, H. Huang, Spatio-temporal difference analysis in climate change topics and sentiment orientation: based on LDA and BiLSTM model, Resour. Conserv. Recycl. 188 (2023) 106697.
[32] R. Ebeling, C.A. Córdova Sáenz, J. Nobre, K. Becker, The effect of political polarization on social distance stances in the Brazilian COVID-19 scenario, J. Inf. Data Manag. 12 (1) (2021) 1889.
[33] S. Prabha, N. Sardana, Question tags or text for topic modeling: which is better, Procedia Comput. Sci. 218 (2023) 2172–2180.Regenerate response
[34] A. Abdelrazek, Y. Eid, E. Gawish, W. Medhat, A. Hassan, Topic modeling algorithms and applications: a survey, Inf. Syst. 112 (2023) 102131, https://doi.org/10.1016/j.is.2022.102131.
[35] S.E. Uthirapathy, D. Sandanam, Topic modelling and opinion analysis on climate change twitter data using LDA and BERT model, Procedia Comput. Sci. 218 (2023) 908–917, https://doi.org/10.1016/j.procs.2023.01.071.
[36] Q. Xie, X. Zhang, Y. Ding, M. Song, Monolingual and multilingual topic analysis using LDA and BERT embeddings, Journal of Informetrics 14 (3) (2020) 101055, https://doi.org/10.1016/j.joi.2020.101055.
[37] M. Grootendorst, BERTopic: neural topic modeling with a class-based TF-IDF procedure, arXiv (2022), https://doi.org/10.48550/ARXIV.2203.05794.
[38] L. McInnes, J. Healy, J. Melville, UMAP: Uniform manifold approximation and projection for dimension reduction (Version 3), arXiv (2018), https://doi.org/10.48550/ARXIV.1802.03426 [Online]. Available:.
[39] K. Koruyan, BERTopic konu modelleme tekniği kullanılarak müşteri şikayetlerinin sınıflandırılması, İzmir Sosyal Bilimler Dergisi 4 (2) (2022) 66–79, https://doi.org/10.47899/ijss.1167719.
[40] V.R. Hananto, Implementation of dynamic topic modeling to discover topic evolution on customer reviews, Jurnal Online Informatika 8 (2) (Dec. 2023) 2, https://doi.org/10.15575/join.v8i2.963.
[41] L. Yang, D. Ji, M. Leong, Document reranking by term distribution and maximal marginal relevance for Chinese information retrieval, Inf. Process. Manag. 43 (2) (Mar. 2007) 315–326, https://doi.org/10.1016/j.ipm.2006.07.011.
[42] L. Yang, D. Ji, M. Leong, Document reranking by term distribution and maximal marginal relevance for Chinese information retrieval, Inf. Process. Manag. 43 (2) (Mar. 2007) 315–326, https://doi.org/10.1016/j.ipm.2006.07.011.
[43] D. Menon, K. Shilpa, Chatting with ChatGPT': analyzing the factors influencing users' intention to Use the Open AI's ChatGPT using the UTAUT model, Heliyon 9 (11) (Nov. 2023) e20962, https://doi.org/10.1016/j.heliyon.2023.e20962.
[44] A. Koubaa, B. Qureshi, A. Ammar, Z. Khan, W. Boulila, L. Ghouti, Humans are still better than ChatGPT: case of the IEEEXtreme competition, Heliyon 9 (11) (Nov. 2023) e21624, https://doi.org/10.1016/j.heliyon.2023.e21624.
[45] Intellica, AI, comparison of different word embeddings on text similarity - a use case in NLP [Online]. Available: https://intellica-ai.medium.com/comparison-of-different-word-embeddings-on-text-similarity-a-use-case-in-nlp-e83e08469c1c, 2019.
[46] H.T.P. Williams, J.R. McMurray, T. Kurz, F. Hugo Lambert, Network analysis reveals open forums and echo chambers in social media discussions of climate change, Global Environ. Change 32 (May 2015) 126–138, https://doi.org/10.1016/j.gloenvcha.2015.03.006.
[47] M. Karimiziarani, W. Shao, M. Mirzaei, H. Moradkhani, Toward reduction of detrimental effects of hurricanes using a social media data analytic Approach: how climate change is perceived? Climate Risk Management 39 (Jan. 2023) 100480 https://doi.org/10.1016/j.crm.2023.100480.
[48] L. Tuitjer, P. Dirksmeier, Social media and perceived climate change efficacy: a European comparison, Digital Geography and Society 2 (Jan. 2021) 100018, https://doi.org/10.1016/j.diggeo.2021.100018.
[49] C.Y. Park, et al., Impact of climate and socioeconomic changes on fire carbon emissions in the future: sustainable economic development might decrease future emissions, Global Environ. Change 80 (May 2023) 102667, https://doi.org/10.1016/j.gloenvcha.2023.102667.
[50] C. Srisunont, T. Srisunont, S. Babel, Development of models for sustainable green mussel cultivation under climate change events, Ecol. Model. 473 (Nov. 2022) 110141, https://doi.org/10.1016/j.ecolmodel.2022.110141.
[51] S. Sawant, J. Yu, K. Pandya, C.-K. Ngan, R. Bardeli, An enhanced BERTopic framework and algorithm for improving topic coherence and diversity. 2022 IEEE 24th Int Conf on High Performance Computing & Communications; 8th Int Conf on Data Science & Systems; 20th Int Conf on Smart City; 8th Int Conf on Dependability in Sensor, Cloud & Big Data Systems & Application, HPCC/DSS/SmartCity/DependSys), Hainan, China, 2022, pp. 2251–2257, https://doi.org/10.1109/HPCC-DSS-SmartCity-DependSys57074.2022.00332.
[52] Z. Cao, S. Li, Y. Liu, W. Li, H. Ji, A novel neural topic model and its supervised extension, Proc. AAAI Conf. Artif. Intell. 29 (1) (2015), https://doi.org/10.1609/aaai.v29i1.9499.
[53] G. Xu, Y. Meng, X. Qiu, Z. Yu, X. Wu, Sentiment analysis of comment texts based on BiLSTM, IEEE Access 7 (2019) 51522–51532.
[54] J. Wang, B. Meng, T. Pei, Y. Du, J. Zhang, S. Chen, G. Zhi, Mapping the exposure and sensitivity to heat wave events in China's megacities, Sci. Total Environ. 755 (2021) 142734.