

Intensity quantification of public opinion and emotion analysis on climate change

Tishya Thukral¹, Ashwani Varshney² and Vibha Gaur^{1*}

Department of Computer Science, Acharya Narendra Dev College, University of Delhi, India¹

Department of Computer Science, University of Delhi, India²

Received: 13-July-2021; Revised: 09-October-2021; Accepted: 13-October-2021

©2021 Tishya Thukral et al. This is an open access article distributed under the Creative Commons Attribution (CC BY) License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Human-related activities are primarily accountable for climate change resulting in natural disasters. Therefore, it has become essential to analyze and evaluate public awareness of climate change globally. With the prevalence of social networks like Twitter, sentiment classification has been recognized as a powerful tool to determine public opinion and concern on such ecological issues. Therefore, this study proposes a framework to classify the tweets containing public opinion towards climate change using Bi-directional Long Short-Term Memory (Bi-LSTM) Networks. The proposed framework quantified the intensity of the public opinion classified by the Bi-LSTM model to measure the strength of the public concern towards climate change and validated it using three case studies: Earth Day, Delhi Air Pollution, and Australian Bushfires. The intensity values of the public sentiments concerning these events were obtained as 98.50%, 96.57%, and 98.33%, respectively. The proposed work was further augmented with a lexicon-based emotion analyzer to categorize the emotions associated with the tweets into positive, negative, neutral, and mixed to substantiate the results. This framework can be utilized before enforcing the policy decisions on the general public in any domain.

Keywords

Bi-LSTM networks, Sentiment analysis, Global warming, Social networks, Climate change.

1. Introduction

The large-scale shift in the weather patterns due to intensified global warming is becoming a severe issue globally with constant evolving outlooks. Rising temperatures, melting of glaciers, and changes in weather patterns are some of the indicators of climate change. Increased human emissions on the ozone layer and the environment itself are primary causes leading to changes in the climate [1]. Such anthropogenic undertakings trigger alarming climate hazards, affecting significant sectors such as human livelihood, ecosystem, and infrastructure [2]. However, there persists a gap between the climate perceptions of the populace and the scientific discoveries leading to a relatively limited understanding of the prevailing dangers [3]. News headlines may not expel such preconceptions regarding climate change with the prevailing research to derive meaningful conclusions for the public.

Therefore, developing appropriate learning techniques to analyze and evaluate public awareness of the critical issues has become an absolute necessity for a promising future. For example, there have been instances of mass protests during the implementation of acts such as the Citizenship Amendment Act of (CAA) 2019 [4] and Farm Bills 2020 [5] due to an absence of adequate understanding of the public on the issues. The implementation of these policies and strategies has resulted in a variety of societal responses due to a lack of an extensive comprehension of the public stance. Therefore, it has become essential to develop a framework that can be utilized by government and non-government organizations in assessing public opinion before enforcement of the policies and laws for the citizens. Hence, objective of this study is to develop a novel framework to assess and quantify public outlook that can assist the authorities in devising informed policies with thorough cognizance of public perceptions. The proposed framework is applied on the public opinion regarding climate change as human activities are attributed as its primary cause.

* Author for correspondence

With the growth of social media, environmental issues such as climate change have been debated extensively. Such deliberations lead to the formation of innumerable opportunities for researchers worldwide to draw inferences on the public stance by utilizing the data from social media platforms. Several machine learning and deep learning techniques have been designed and tested to validate public standpoints on the serious issues. Sentiment analysis is one of the robust and significant techniques employed to identify, mine, and categorize the public opinions from texts methodically [6]. It can assist in determining and understanding public sensitivity on various climate-related issues. Hence, the proposed study aims to address and analyze the public sentiments concerning climate change that can benefit the concerned authorities to formulate policies and make decisions for societal awareness. Additionally, it can also facilitate the citizens in improving their outlooks and contemplations about climate change.

Twitter is considered a powerful resource for providing data, with more than 192 million users sharing their opinions on numerous issues daily [7]. It forms a vital resource due to the availability of a massive amount of data that can be easily classified and segregated with hashtags.

This study proposes a novel framework to analyze and quantify the intensity of the public sentiments using Bi-directional Long Short-Term Memory (Bi-LSTM) multi-classification networks. The tweets concerning climate change were extracted and utilized to train the multi-classification model. As a result, the model determined whether a given tweet has a positive, negative, or a neutral outlook regarding climate change. The positively classified tweets were segregated. Their probability scores were utilized to compute the intensity value, indicating how strongly the public was concerned with climate change. The intensity value facilitates the concerned authorities and policymakers to formulate strategies concerning climate-related issues. The proposed framework was endorsed by computing the intensity values for three case studies, including Earth Day, Delhi Air Pollution, and Australian Bushfires. The study further presents an emotion analysis to supplement the sentiment classification results and gain further insights concerning the public opinion on climate change.

The organization of the rest of the paper is as follows: Section 2 briefly reports the related work

from the literature, Section 3 presents the proposed framework to estimate the intensity of the positive sentiments and classifies the associated emotions. Section 4 discusses the results and validates them using the undertaken case studies. Finally, Section 5 concludes the paper.

2.Literature review

Climate change is described as long-term alterations in regional and global weather patterns. It is a genuine ecological threat to humanity and wildlife. Climate change has become a significant area of research to examine the extent of shifts in weather patterns in the past, forecast the environmental conditions in the future, and understand the public outlook on it. These scientific undertakings can inform and guide environment-related decisions in various ways by facilitating an improved understanding of the risks posed by climate change and the necessity to act on them.

With the increase in the prevalence of social media, researchers have further directed their interest to evaluate and assess public opinions towards impending climate-related issues. Classifying and extracting valuable insights from the information accessible on social media is gaining research interest. Studies indicate that supervised machine learning and deep learning approaches have been utilized for addressing various environmental issues. Despite significant steps taken to understand several facets of climate change, quantification of the concerned public opinion has not been considered yet; however, human intervention is one of the primary causes of climate change.

Various machine learning techniques have been employed to analyze public opinion in several domains using sentiment analysis [8, 9].

Soumya and Pramod [10] developed a model utilizing various machine learning algorithms like a Support Vector Machine (SVM), Naïve Bayes (NB), and Random Forest (RF) to classify Malayalam texts into positive and negative sentiment. However, the size of the training corpora did not suffice, due to the lack of accessible datasets in the chosen language, highlighting the importance of data preparation for sound results. With the advancement in artificial intelligence techniques, deep learning models overtook machine learning models due to their capability to understand complex relationships between the parameters of the study.

Ansari et al. [11] developed a sentiment analysis model incorporating LSTM networks and feature extraction techniques, including Term Frequency-Inverse Document Frequency (TF-IDF), to evaluate people's opinions and their political orientation during the election process based on their activities on social media platforms. However, the dataset utilized in this research was not balanced, which signaled a necessary inclusion of better data sampling methods.

Heikal et al. [12] developed an ensemble model integrating Convolutional Neural Network (CNN) and LSTM models to predict and classify the sentiment of Arabic tweets into positive, negative, and neutral. However, the size of the utilized dataset could be improved by including more complex collaborative models. Researchers have also deployed a hybrid deep learning model using CNN and LSTM for sentiment classification referencing diverse domains on several issues such as the presidential elections, airline, and cinema reviews, with a scope of improvement in the size of the utilized corpora to yield better insights from the trained models [13].

Muhammad et al. [14] developed a self-evaluation mechanism based on a sentiment analysis using LSTM Networks and Word2vec embeds to classify customer reviews on the hotel services. However, the achieved accuracy can be further improved with the tuning of hyper-parameters.

Furthermore, several studies have also incorporated an analysis of emotions together with the sentiment using EmoLex or National Research Council of Canada (NRC) Lexicon, i.e., a custom lexicon with annotations referencing the primary eight emotions.

Bose et al. [15] incorporated an analysis of tweets on the 14th Gujarat Legislative Assembly Elections with the NRC lexicon to determine the overall emotional tone of the event.

Syed [16] presented a lexicon-based approach for incorporating sentiment and emotion analysis of tweets concerning technology brands, thereby improving their digital marketing approach.

In a study presented by Dubey [17], a country-wise analysis of tweets during the COVID-19 outbreak has been carried out, focusing on analyzing the emotions felt by the citizens during a specified period of the outbreak with the aid of the NRC emotion lexicon.

However, the scope of the research was limited to a specific number of countries which could have been extended further. Furthermore, researchers have also assessed the sentiments and emotions towards the United States and India leadership during the COVID-19 outbreak, with a scope of improvement in the size of the utilized dataset for better results [18].

Although the prevalence of sentiment and emotion analysis is visible in numerous fields, its utilization in the field of climate change is still a novel opportunity for researchers. Sentiment analysis may be helpful in the area of climate change as understanding public beliefs on climate change directly correlates to the response of the concerned authorities towards it. Furthermore, people who do not realize the impact of climate change may be counseled to mitigate the climate risks.

Kirelli and Arslankaya [19] performed sentiment analysis of Turkish texts to understand society's receptivity towards climate change, integrating various machine learning techniques such as NB, SVM, K-Nearest Neighbor (KNN) for improvement in the accuracy and performance measures. However, this study only included Turkish texts as part of the conducted analysis.

Another study was undertaken using sentiment and emotion analysis by Loureiro and Alló [20] to extract renewable energy preferences and policies formulated and assessed for the public residing in the UK and Spain. EmoLex was utilized to address a multitude of aspects concerning public sentiment towards the environment. However, it was not robust enough to assess a context-specific language.

Reyes-Menendez et al. [21] presented a topic-based sentiment approach to understanding social, economic, environmental, and cultural factors related to the sustainable environment goals and public health that concerned most Twitter users for policymaking initiatives. However, the studies mentioned above can be improved by employing advanced deep learning models like Bi-LSTM networks since the deep learning algorithms such as Recurrent Neural Networks (RNNs) suffer from the 'vanishing gradient problem,' thereby hindering the comprehension of long sequences of textual information [22, 23]. Since sentiment classification requires an adequate understanding of the contextual information, algorithms such as LSTM and CNN may be utilized to provide adequately accurate results due to their ability to identify long-term text

dependencies [24]. However, the efficacy of these algorithms does not facilitate a thorough understanding of the underlying context due to the prevalence of a sizeable number of features for classification. Bi-LSTM consisting of two LSTMs exhibits greater efficacy in terms of performance measures for an application using text processing such as sentiment classification. It increases the information flow, thus improving the model's learning capacity [23]. Bi-LSTM incorporates forward and backward LSTM layers to facilitate the extraction of contextual information in both directions, ensuring the detection of richer semantic information for thorough text classification [24].

Researchers have also utilized Bi-LSTM networks to address various domains, apart from sentiment classification.

Sunny et al. [25] utilized RNN techniques involving LSTM and Bi-LSTM for a stock market prediction, providing improved accuracy compared to other regression methods. However, the hyper-parameter tuning of both models could be further enhanced to improve the performance of the model.

Li et al. [26] utilized Bi-LSTM in conjunction with CNN to address the classification of large-scale news texts, highlighting a necessary inclusion of methods to eliminate words interfering with the subject context to improve the accuracy. Several studies have utilized Bi-LSTM for issues relating to the agriculture addressing mapping of crop production [27] and natural resources [28], forecasting urban water demand by establishing the correlation between climate change and the demand.

Yin et al. [29] developed a hybrid Bi-LSTM model to forecast short-term Evapotranspiration (ET_0) for crop water requirements. However, the presented model could not be directly related to the other climatic regions in the world for predictions.

Nevertheless, the utilization of Bi-LSTM for classifying climate perceptions has been minimal; hence this work utilizes Bi-LSTM networks for predicting and quantifying the intensity of the public opinion regarding climate change and is further augmented with a lexicon-based emotion classification. Therefore, the proposed study presents a novel approach to facilitate extensive public contemplations on pressing issues such as climate change. The study may enable the government and non-government organizations to measure the extent

of public concern towards the prevailing environmental issues, aiding them to implement strategies with active societal involvement and support. Furthermore, the policymaking initiatives may be enhanced with a deeper understanding of the public stance facilitated by the incorporated emotion classification, thereby ensuring an improvement in the ecological decisions for the future.

3.Methods

Climate change is one of the primary causes of the frequent occurrence of natural calamities. Increasing temperatures, a deviation in weather patterns, and the melting of glaciers are prominent indicators of change in the climate. To mitigate detriments caused by the change, public awareness is the need of the hour, which can be facilitated with social media platforms due to their exponential capacity to disseminate information. Various studies have been undertaken to analyze public perceptions of environment-related issues using artificial intelligence techniques, including RF [30], decision tree methods [30], Artificial Neural Networks (ANN) [31], NB [32], and SVM [32].

However, quantifying the strength of public opinion concerning climate change has not been considered yet. Nevertheless, quantification may help determine the intensity of public belief, signaling an improvement for assessing and strengthening the ideologies of the public on climate change in the right direction. Along with the intensity computation, a comprehension of the exhibited emotional response may provide an improved understanding of global anxieties on climate change, imparting help of government and non-government organizations in devising policies and tactics to address the environmental issues with an anticipated consensus.

Hence, a framework is proposed to address the above-mentioned necessity that classified the climate-related tweets and quantified the intensity value expressing public concern regarding climate change using Bi-LSTM networks. The intensity value signifies the degree to which Twitter users are conscious of climate change, which government and non-government organizations may utilize in policymaking on environmental issues. Moreover, the framework incorporated an emotional analysis to understand the emotions associated with tweets concerning climate change. The proposed framework was implemented using Python 3.8 programming environment and mainly consisted of four major phases:

- (i) **Preparation of the data:** It involved the gathering of data from online repositories. The collection of data was followed by pre-processing to remove any undesirable components. Finally, the pre-processed data was tokenized to convert it into a suitable form for training the model.
- (ii) **Training the multi-classification model:** The Bi-LSTM networks were utilized in the multi-classification of tweets. These networks can effectively utilize the available information in both directions, thus improving the understanding of language context. Once the model was trained and tested, it classified tweets into the positive, negative, and neutral class labels.
- (iii) **Quantification of intensity value:** The predicted positive class labels were segregated to obtain the probability scores of the tweets returned by the

- trained model. The tweets were assigned identifiers based on the probability scores utilized to assess and quantify the strength of the public sentiment.
- (iv) **Emotion Classification:** The predicted positive class labels were segregated to conduct a lexicon-based emotion analysis. The tweets were assigned identifiers based on their classification into four emotion categories: positive, negative, neutral, and mixed emotions, to obtain emotional distribution graphs.

The proposed framework for assessing the intensity of public sentiments and subsequent emotion classification is depicted in *Figure 1* and explained in the following subsections.

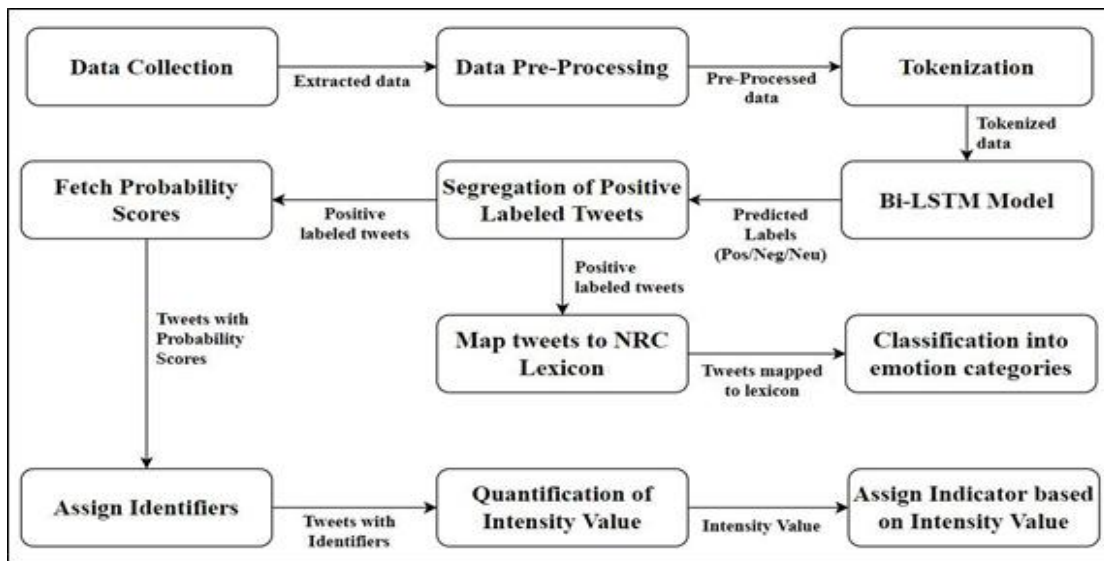


Figure 1 Proposed framework

3.1 Data collection

The collection of appropriate corpora is the most crucial step for ensuring the quality of the obtained results. Hence, it necessitates significant effort to obtain the most relevant data for the study. This

research utilized a corpus gathered from Kaggle, an open-source platform [33]. The gathered corpus included tweets concerning climate change. *Table 1* describes the significance of class labels.

Table 1 Tweet class distribution

Class label	Description
Positive (1)	The tweet supports the belief of human-made climate change
Neutral (0)	The tweet neither supports nor refutes the belief in human-made climate change
Negative (-1)	The tweet does not believe in human-made climate change

The snapshot of the extracted tweets is given in *Figure 2*.

Data collection was followed by data pre-processing to clean the tweets and thus making them appropriate for further processing.

	sentiment	message
0	-1	@tiniebeany climate change is an interesting h...
1	1	RT @NatGeoChannel: Watch #BeforeTheFlood right...
2	1	Fabulous! Leonardo #DiCaprio's film on #climat...
3	1	RT @Mick_Fanning: Just watched this amazing do...
5	0	Unamshow awache kujinga na iko global warming ...
...
43936	1	#AikBaatSunIThi \nGlobal warming has negative ...
43938	1	Dear @realDonaldTrump,\nYeah right. Human Medi...
43939	1	What will your respective parties do to preven...
43941	0	RT @taehbeingextra: i still can\$qt believe th...
43942	1	@Likeabat77 @zachhaller \n\nThe wealthy + foss...

Figure 2 Raw training data

3.2 Data pre-processing

Data pre-processing is a significant stage for ensuring that data is ready to use for classification. The quality of data is crucial to model training and accurate classification.

Data pre-processing included removing additional characters, including user-handles, retweets, Uniform Resource Locators (URLs), punctuations, hashtags, emoticons, special placeholder characters, other Hypertext Markup Language (HTML) references, and numeric values with the help of the RegEx module [34]. Contractions were replaced with their corresponding full-length words.

Furthermore, case-folding was performed to convert all the words in lower case. *Figure 3* presents the tweets after pre-processing.

	sentiment	text
0	-1	climate change interesting hustle global warml...
1	1	watch beforetheflood right travels world tackl...
2	1	fabulous leonardo dicaprios film climate chang...
3	1	watched amazing documentary leonardodicaprio C...
4	0	unamshow awache kujinga na iko global warming
...
34662	1	aikbaatsunithi global warming negative effect ...
34663	1	dear yeah right human mediated climate change ...
34664	1	respective parties prevent climate change glob...
34665	0	still canqt believe gif taehyung saved human r...
34666	1	wealthy fossil fuel industry know climate chan...

Figure 3 Training tweets after pre-processing

Data sampling was employed to ensure better accuracy for classification, which balanced the tweets for positive and negative class labels. There are primarily two methods for sampling; undersampling involves taking a random subset of the majority class small enough to match the number of observations in the minority class. In contrast, oversampling involves taking repeated random samples from the minority class until as many observations as the majority class [35]. Both sampling methods were incorporated to balance the tweets. Hence, the positive class was undersampled using the module sklearn [36]. In contrast, the negative and neutral classes were oversampled to create a balanced dataset, with all three classes attaining 20,000 tweets each, as shown in *Figure 4*.

```
-1 20000
1 20000
0 20000
Name: sentiment, dtype: int64
```

	sentiment	text
0	1	dont know ton paris agreement dangerous trump ...
1	0	writing none deny climate change describing hu...
2	1	global warming huh
3	1	climate denier posts link debunking climate ch...
4	0	foolish world leaders stay fixated paris costl...
...
59995	-1	republicans dont ignore climate change simply ...
59996	-1	dear global warming couldnt real signed confus...
59997	-1	foxnews socialistsen believes climate change l...
59998	1	unchecked climate change going stupendously ex...
59999	0	httpcoojzp vysgv science reveals pluto trust ...

Figure 4 Training data after resampling

Pre-processing was followed by tokenization to convert the tweets into a compatible format for model training.

3.3 Tokenization

Tokenization involves breaking a piece of text into smaller units called tokens [37]. The tweets need to undergo tokenization to diversify the understanding of the words, thereby facilitating the model training. The tweets after the pre-processing were transformed into tokenized words forming separate entities using the Keras module [38]. Once the tokenization was carried out, 26,064 unique tokens were obtained. The generated input sequence was padded to obtain the tweets of the same length. Finally, the resulting tweets were used for training the classification model.

3.4 Training the model

Bi-LSTM networks have been utilized for the multi-classification of tweets. A Bi-directional LSTM or Bi-LSTM is a sequence processing model that consists of two LSTM networks: one taking the input

in a forward direction and the other in a backward direction [39]. The LSTM architecture involving a series of recurrently connected subnets facilitates the processing with layers comprising a cell, an input gate, an output gate, and a forget gate [40]. The cell remembers values over arbitrary time intervals, and the three gates regulate the flow of information into and out of the cell. Although a single LSTM network is very effective in capturing sequence information that can aid in classification, a Bi-LSTM network is more appropriate in capturing the meaning of the texts under different contexts owing to its capability of understanding in forward and backward directions [39]. Bi-LSTM networks can effectively increase the amount of information available to the network, improving the context and understanding.

Therefore, Bi-LSTM networks were utilized for the desired multi-classification of tweets. The tokenized tweets were split into training and testing sets with a ratio of 90:10, respectively, using the sklearn module. After the division of datasets, the hyper-parameters were specified to tune the model for training. The model utilized an embedding layer and a spatial dropout layer, thereby facilitating a reduction in the network parameter count, increasing computation speed, and regulating overfitting. A Bi-LSTM layer with a dropout rate of 0.2 and a recurrent dropout rate of 0.2 and categorical cross-entropy as the loss function with softmax activation function was utilized for multiclass classification. Adam optimizer was used to handle sparse gradients on noisy

problems [41]. *Table 2* summarizes the assimilated hyper-parameters.

Table 2 Training hyperparameters

Hyper-parameters	Value
Batch Size	32
Epochs	10
Activation	Softmax
Loss Strategy	Categorical Cross-entropy
Optimizer	Adam
Spatial Dropout	0.2
Recurrent Dropout (Bi-LSTM Layer)	0.2
Dropout (Bi-LSTM Layer)	0.2
Early Stopping Patience	3 Epochs
Early Stopping min_delta	0.0001

The performance of the trained model was evaluated using the Classification Accuracy [42] metric, which is defined as the ratio of correct predictions to the total input samples. The best-case accuracy of 94.98% was achieved for the trained model, implying that out of every 100 tweets, the model classified almost 95 tweets correctly. *Figure 5 a)* and *Figure 5 b)* display the model's efficiency using accuracy and loss graphs, respectively.

The trained and tested Bi-LSTM model was then utilized for making predictions of tweets into three class labels: positive, negative, and neutral, out of which only positive labeled tweets were utilized for evaluation of the intensity value.

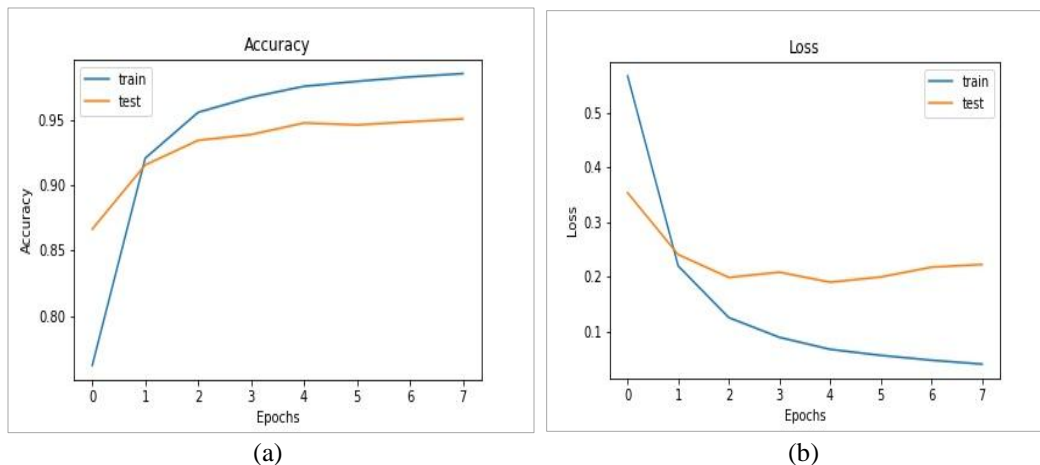


Figure 5 a) Training and testing accuracy b) Training and testing loss

3.5 Quantification of intensity value

This phase of the framework incorporated calculating the intensity value for the positive-labeled class to measure the degree of public belief towards the prevalence of climate change. The predictions made by the trained model were segregated to ensure the provision of tweets belonging to the positive-labeled class only. Furthermore, the probability scores of the extracted positive labeled tweets were fetched in a range of 0-1, as shown in *Table 3*. For simplicity, the tweets were assigned identifiers as per their probability scores and divided into four equal quartiles without the loss of generality. These identifiers were utilized for computing the aggregated score using Algorithm 1.

Table 3 Range of positive probability scores and corresponding values assigned to calculate intensity value

Probability score	Value assigned (Identifier)
(0.0 - 0.25]	1
(0.25 - 0.50]	2
(0.50 - 0.75]	3
(0.75 - 1.0]	4

Algorithm 1. Algorithm for Computation of Intensity Value

tweetIntensity(score):

Value = 0

IF (Score is between 0 and 0.25) THEN:

ADD 1 to Value

IF (Score is between 0.25 and 0.50) THEN:

ADD 2 to Value

IF (Score is between 0.50 and 0.75) THEN:

ADD 3 to Value

ELSE:

ADD 4 to Value

RETURN (Value/4)

The final intensity value was calculated by dividing the normalized aggregated score by the number of positive labeled tweets. The intensity value obtained was multiplied by 100 to get the percentage signifying how intensely people attribute man-made alterations to climate change. *Table 4* describes the intensity indicator based on the obtained intensity value. The obtained intensity indicator determines the strength of public belief relating to climate change. A low-intensity value may indicate that the public does not consider human alterations as the primary cause of climate change. Government organizations may initiate public awareness programs to strengthen their beliefs in the prevalent issue. However, a high-

intensity value may indicate a firm public belief for the same. The study may also help the government and policy makers in proper decision-making, considering the public opinion before enforcing the rules and regulations regarding climate change.

Table 4 Intensity Indicator corresponding to the Intensity Value

Intensity value (%)	Intensity indicator
Less than 25	Weak
Between 25 and 50	Moderate
Between 50 and 75	Strong
More than 75	Very Strong

3.6 Emotion analysis

The final step in this study incorporated a lexicon-based approach that classified the positively predicted tweets, i.e., the tweets that supported the prevalence of human-initiated climate change, into different emotions. The NRC Word-Emotion Association Lexicon (EmoLex) [43, 44], defined as a list of English words and their associations with eight basic emotions demonstrated by Plutchik [45]: anger, fear, anticipation, trust, surprise, sadness, joy, and disgust. Emotion analysis of the segregated tweets was applied and the positively labeled tweets were examined for the frequency of words reflecting different emotions as described in the lexicon. Then, the tweets were further divided into four categories of emotions, a dominant prevalence of positive emotions, i.e., tweets exhibiting emotions of anticipation, trust, surprise, or joy; a dominant prevalence of negative emotions, i.e., tweets exhibiting emotions of anger, fear, sadness, or disgust; mixed emotion, i.e., tweets demonstrating an equivalent prevalence of both positive and negative emotions and finally neutral emotion, i.e., tweets lacking both positive and negative emotions. *Table 5* lists the numeric identifiers that were assigned to the aforementioned emotion categories for facilitating graph projections.

Table 5 Numeric identifiers corresponding to emotion category

Numeric identifier	Emotion category
1	Positive (anticipation, trust, surprise, or joy)
-1	Negative (anger, fear, sadness, or disgust)
2	Mixed (both positive and negative)
0	Neutral (lacking both positive and negative emotions)

A substantial prevalence of positive emotion signifies that the public outlook is increasingly optimistic about driving decisive, cohesive climate action. However, a substantial prevalence of negative emotion signifies that the populace is distressed about the prevailing and forthcoming climate adversities, signaling a desolation at the current state of environmental affairs.

The following section evaluated the performance of the proposed framework and validated the results on three real-world scenarios.

4. Results

The proposed framework was utilized for measuring the strength of public opinion concerning climate change. As mentioned below, the framework was applied to three case studies; *Earth Day*, *Delhi Air Pollution*, and *Australian Bushfires* explained below briefly. The corpora for the intensity calculation were crawled from Twitter.

1) **Earth Day:** It is celebrated on April 22 to establish support for environmental protection. This day aims to raise awareness among the masses about climate change and global warming every year, motivating them to promote conservation at the earliest to secure the foundation of a sustainable future [46]. The event fills the landscape with meaningful global conversations, performances, and webinars addressing major environmental issues while pushing forward suggestions such as embracing opportunities for green jobs and sustainable stimulus. Marking the 50th year of the celebration, *Earth Day 2020* inspired the populace to take urgent steps to address climate-related variations, highlighting the need to reduce unchecked environmental damage [46]. Studying the public outlook on an event widely acknowledged as a demonstration of solidarity for environmental protection is considered a necessity to identify the fundamental public convictions regarding climate change.

2) **Delhi Air Pollution:** Delhi has been classified as the most polluted capital globally in several instances. The citizens of Delhi and its neighboring areas struggle to breathe as the air pollution intensifies with each passing year [47]. Air pollution in Delhi may be attributed to crop burning in the Northern Indian States, vehicle transmissions, or heavy industry [47]. Moreover, air pollution significantly relates to greenhouse gas emissions, thus affecting factors relating to climate

change [48]. With an observable interrelation with climate change and human health [49], it becomes essential to study the public opinion regarding a region adversely affected by air pollution.

3) **Australian Bushfires:** These fires heavily impacted various regions of Australia in 2019 and 2020, including the state of New South Wales, such as the North Coast, the Hunter Region, and the Wollondilly in Sydney's far west. Thousands of fires burning across the states led to a severe loss of human life and damage to wildlife [50]. Exacerbated by a conjunction of several causes, including the underlying climate change, the bushfires triggered a catastrophic impact on all forms of life during a series of cumulative events, generating a string of responses from the populace facing unprecedented challenges. Therefore, studying the public outlook is considered vital to gain insights regarding their 1) beliefs concerning the role of climate change in the incidence of bushfires and 2) emotional response concerning the state of the region.

The public opinion on the three events was analyzed using the proposed framework. The tweets containing the public opinion were classified, and intensity values of the sentiments were quantified. Furthermore, the framework was augmented with a classification of emotions for all three scenarios. The obtained results of the studies are discussed in the following subsections.

4.1 Case study 1 (Earth Day 2020)

The dataset for this case study consisted of 10,222 tweets scraped for the year 2020 with the following hashtags: '#EarthDay,' '#EarthDay2020,' and '#CarbonCapture.' This dataset was extracted using the Twitter search API, which is an open-source Python package that provides users with a very convenient way to access tweets from Twitter profiles of the general public [51].

The tweets collected in this corpus underwent pre-processing, which included removing numeric values, emoticons, HTML references, retweets, URLs, punctuations, user handles, hashtags, and case-folding. The Bi-LSTM model was trained to categorize tweets into three class labels: Positive, Negative, and Neutral. Furthermore, it was observed that most of the predictions made by the model belonged to the positive class label, signifying the solid public belief in the prevalence of human-made climate change. The classification of tweets mainly in

the positive class resulted from many deliberations on this occasion involving the general public and the dedicated researchers for initiating the actions required to mitigate the damage caused due to human alterations in the environment [52].

The probability scores of the segregated positive-labeled tweets were fetched and utilized. The distribution of the obtained probability scores is shown in *Figure 6*, where X-axis shows the 9,139 positive-labeled tweets and the Y-axis represents the corresponding probability scores ranging from 0 to 1. The fetched probability scores were utilized to compute the intensity value. The intensity value of the positive class was obtained using the formula described in Sec. 3.5 and was found to be 98.50%. The obtained intensity value confirmed that the tweets on Earth Day favor the prevalence of increasing climate threats often initiated by human

alterations. Moreover, the fetched positive-labeled tweets were utilized to classify emotions felt by the populace regarding climate change. The positive-labeled tweets underwent a lexicon-based analysis utilizing EmoLex. The emotion scores for the tweets were aggregated to form two primary categories, i.e., positive and negative emotions with subsequent categorization into four types as described in Section 3.6. The tweets were then assigned the respective identifiers for demonstrating the distribution of emotions presented in *Figure 7*.

The graph suggested a high prevalence of positive emotion with the majority of the tweets belonging to identifier 1, signifying that the citizens are optimistic about bringing a substantiated change to achieve sustainability, thereby confirming a relation with the theme of Earth Day 2020, i.e., climate action [52].

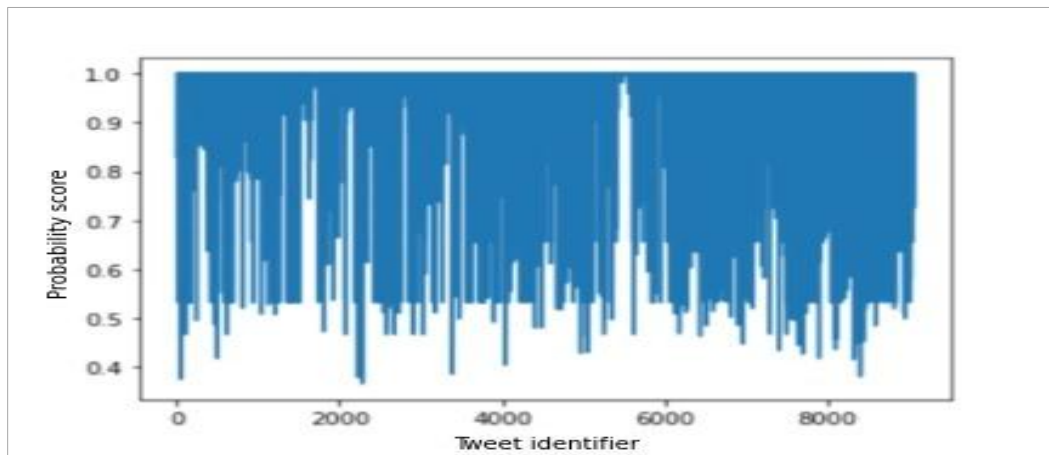


Figure 6 Probability scores for positive labeled tweets for corpus 1 (Earth Day 2020)

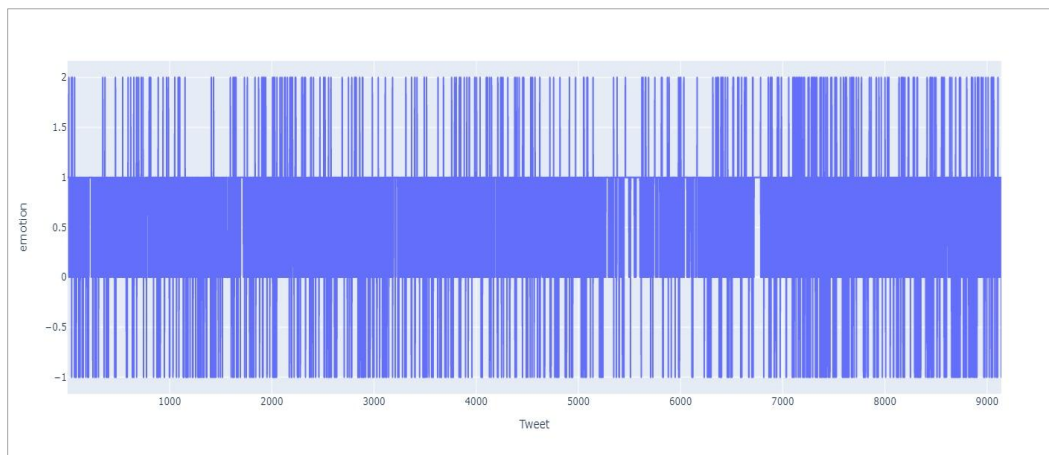


Figure 7 Distribution of emotions for corpus 1 (Earth Day 2020)

4.2 Case study 2 (Delhi air pollution 2019-2020)

The proposed framework was also applied to another dataset, Delhi Air Pollution 2019 and 2020. The dataset consisted of 10,000 tweets scraped for the years 2019 and 2020 with the following hashtags: '#DelhiPollution,' '#DelhiAQI,' and '#DelhiAirQuality.' This case study was undertaken as air pollution in conjunction with climate change has detrimental effects on human health and the environment [53, 54]. Hence, it is essential to study the strength of public belief in the correlation between the prevailing pollution and the arising climate adversities. Finally, the graph in *Figure 8* shows the attained probability scores where the X-axis represents the 5,462 positive-labeled tweets, and the Y-axis shows the corresponding probability scores ranging from 0 to 1. These probability scores were used for computing the intensity value of positive tweets, and the value was obtained as 96.57%. The calculated intensity value confirmed the public belief of human alterations in damaging the environmental equilibrium. The results emphasized that more people realize the exigency to improve the

state of their contaminated surroundings, thereby addressing the link between air pollution and climate-related variations.

However, a sizeable portion of the predictions relating to negative and neutral labels demonstrated a need for the formulation of awareness campaigns to improve the public outlook.

The emotion distribution of the positive labeled tweets is demonstrated in *Figure 9*. The majority of the tweets corresponded to positive and negative emotions with identifier values of 1 and -1, respectively, thereby providing insight into the public response towards the overall situation of the polluted environment. A prevalence of positive emotions signified strong optimism towards improved air quality during the COVID-19 lockdown [55], also signifying support for eco-friendly practices. In contrast, a high prevalence of negative emotions suggested distress at the prolonged toxic state of the environment.

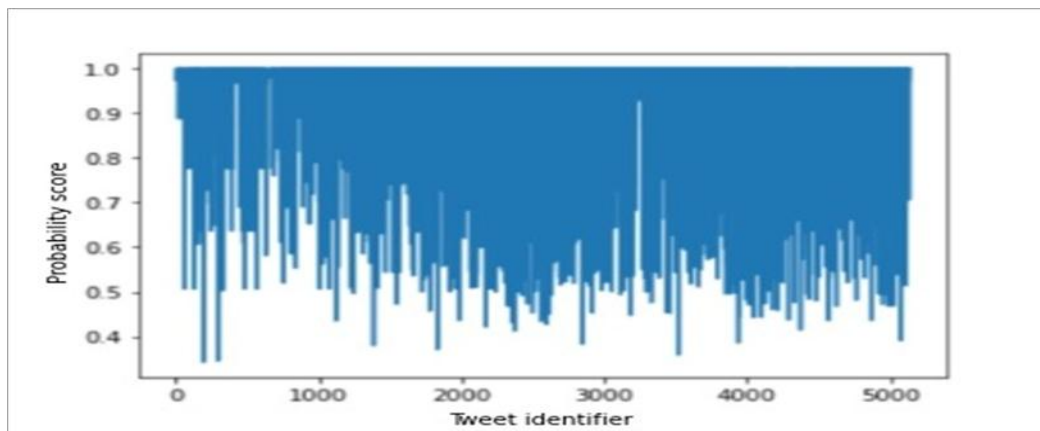


Figure 8 Probability scores for positive labeled tweets for corpus 2 (Delhi air pollution 2019-2020)

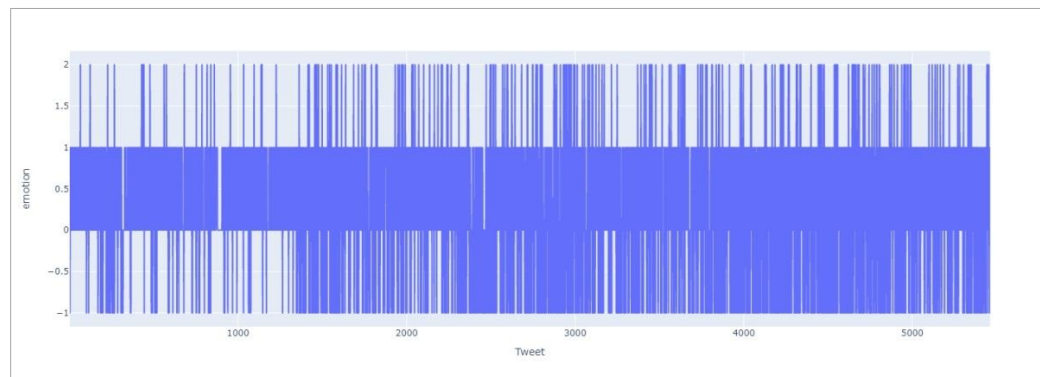


Figure 9 Distribution of emotions for corpus 2 (Delhi air pollution 2019-2020)

4.3 Case study 3 (Australian bushfires 2019-2020)

Further endorsing the scope of the study globally, the proposed framework was investigated using another dataset of tweets on Australian Bushfires 2019-2020 consisting of 10,056 tweets scraped with the hashtag: '#Australian Bushfires.'

The study on Australian Bushfires was taken up as these bushfires often cause significant health and economic impact on the population. Hence, comprehending a vast multitude of opinions held by individuals and communities worldwide to measure the extent of accountability subjected to climate change as the underlying cause of exacerbated bushfires was the aim of the study.

Once the concerned tweets were classified into the various classes, the probability scores were obtained

for the positively classified tweets. The probability distribution is depicted in *Figure 10*, where the x-axis shows the 8,326 positive-labeled tweets, and the y-axis represents the corresponding probability scores ranging from 0 to 1. For this dataset, the computed intensity value of the positive class was obtained as 98.33%. The result indicated the strength of public belief in human-made climate change as a significant contributor to the severity of bushfires. Furthermore, the emotion distribution is demonstrated in *Figure 11*. A high prevalence of positive emotion, i.e., tweets corresponding to identifier 1, signified the trust in collective action to overcome the crisis [56], whereas an overwhelmingly high prevalence of negative emotion, i.e., tweets corresponding to identifier -1, highlighted agony of the public.

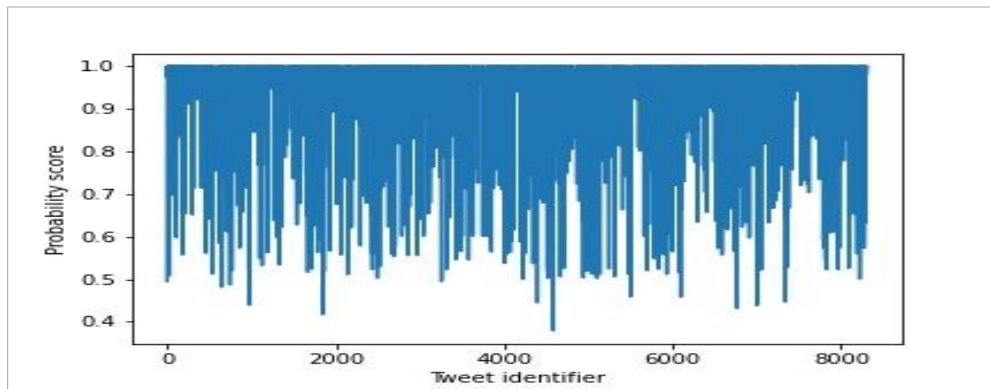


Figure 10 Probability scores for positive labeled tweets for corpus 3 (Australian bushfires 2019-2020)

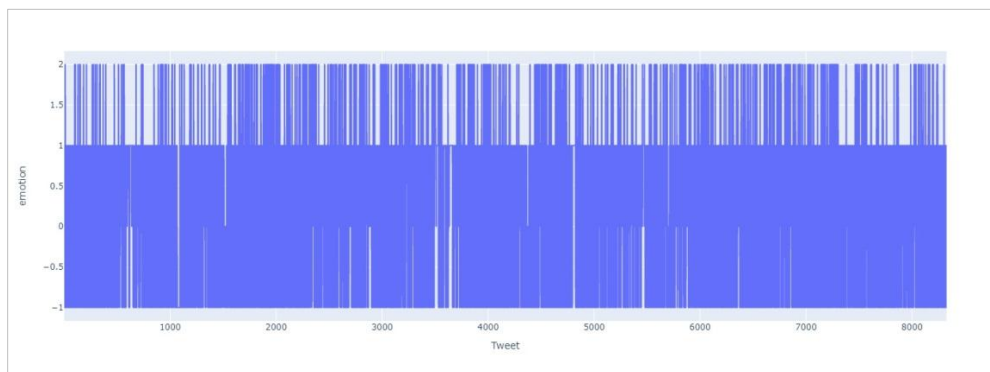


Figure 11 Distribution of emotions for corpus 3 (Australian Bushfires 2019-2020)

The quantified intensity values with their corresponding intensity indicators for each case study are presented in *Table 6*. As can be seen from *Table 6*, the intensity indicator obtained for each case study was Very Strong, implying that most Twitter users strongly believed that man-made alterations are a leading cause of climate change.

Along with validating the public belief in the prevalence of human-initiated climate change, the emotion analysis in the proposed framework provided further insights, demonstrating the various emotions felt by the general populace concerning climate action. The study results may be utilized by various government and non-government

organizations for policy decisions considering public opinion before enforcing any rules or regulations

regarding climate change.

Table 6 Quantified intensity values for the case studies

Case study	Intensity value	Intensity indicator
Earth Day 2020	98.50%	Very Strong
Delhi Air Pollution 2019-2020	96.57%	Very Strong
Australian Bushfires 2019-2020	98.33%	Very Strong

5. Discussion

The proposed work presents a novel framework that government organizations could utilize before enforcing policy decisions to the general public. In the past, it has been observed that implementing government policies without cognizance of the public stance has resulted in various societal responses, sometimes inciting protests. Hence, it is essential to develop a solution that addresses such issues by facilitating comprehension of public opinion before executing these strategies. Public opinion can be classified as positive, negative, and neutral, utilizing a viable deep learning framework like Bi-LSTM networks. The neutral public opinion may be transformed to a positive outlook by disseminating relevant knowledge among the populace. However, the negative opinion may require substantial effort to bring about a difference in the viewpoints, with awareness and counseling programs. The quantification of the public opinion may assist the government in achieving a clear understanding of the public consent before implementing the decisions to the general public.

Therefore, the proposed study proposes a distinct approach for assessing public opinion on climate change, with an opportunity for its application in any other domain. It facilitates an improved comprehension of the public believes regarding human-made climate change with the quantification of the extracted opinion and is further augmented with an emotion analysis to provide meaningful insights. The incorporated intensity quantification has been utilized to estimate the strength of belief in the prevalence of human-initiated climate change among the demographic of Twitter users and validated the results on three case studies: Earth Day, Delhi Air Pollution, and Australian Bushfires. The findings from the proposed framework may help the organizations formulate strategies for environmental problems with an expected consensus from the populace.

The first case study on Earth Day 2020 suggests an optimistic and action-oriented public outlook that conforms to the event's theme, i.e., climate action

necessitating strategies for preventing environmental detriments. The obtained intensity value of 98.50% indicates a strong belief in the attribution of man-made alterations as a leading cause of climate change. The results present a positive emotional response comprising anticipation, trust, surprise, or joy, as shown in *Figure 7*. It implies that people comprehend the existing challenges posed by climate change and intend on discovering solutions to prevent such unchecked environmental damage. The obtained results may encourage the policymaking organizations to devise strategies and initiatives involving the active participation of the populace in solving the environmental crisis.

Delhi Air Pollution represents an interrelation between human-generated pollution and the prevailing climate change, highlighting a need to address the public belief. The intensity of the public concern towards man-made alterations affecting climate change was obtained as 96.57%. The results obtained from the emotion classification suggests a dominant prevalence of both positive and negative emotions, as shown in *Figure 9*. The prevalence of positive emotion specifies public joy, attributing to a few months of clean skies during the COVID-19 lockdown. On the other hand, the prevalence of negative emotion, which comprised anger, fear, sadness, and disgust, indicates public discontent, highlighting the need for more robust and well-formulated strategies to address the prevalent issue.

The third case study on Australian Bushfires was utilized to quantify public opinion in the wake of the climate adversity that transpired during the years of 2019 and 2020. The obtained intensity value of 98.33% demonstrates a strong belief in human-made climate change contributing to the bushfires. *Figure 11* demonstrates projections of the obtained emotion categories and indicates a pervasive presence of positive emotion involving anticipation, trust, surprise, or joy and overwhelming negative emotion involving anger, fear, sadness, and disgust. The positive emotion can be ascribed to trust in the community effort carried out during the disastrous incidence of the fires. However, the negative

emotional signals, public anguish and distrust in handling the calamitous situation, confirmed by the occurrence of mass protests [57]. The obtained results confirm that the public is deeply concerned with the arising climate adversities, indicating the need for meaningful action in environmental affairs.

Hence, the three case studies demonstrate a serious concern of the public regarding the rampant climate change often initiated by man-made alterations. Along with the quantification of intensity values of the positive-labeled tweets, the provision of negative and neutral-labeled tweets implying disbelief in human-made climate change can be significantly valuable for non-profit organizations, aiding them to formulate awareness campaigns and strategies based on the insights from the obtained results.

However, the proposed methodology incorporates quantification and classification of public opinion only in the English language, signaling a possible diversification with the inclusion of multiple languages in the future. The conducted research addresses an environmental issue of significance, i.e., climate change. Though the framework can be easily applied to the necessitating application in other societal domains such as healthcare and education. The proposed framework facilitates a comprehensive understanding of the public opinion and sentiments that may be utilized by government and non-government organizations in vital decision-making. Complete list of abbreviations is shown in *Appendix I*.

6. Conclusion and future work

Human intervention is the primary cause of climate change resulting in changing weather patterns, rising temperatures, melting glaciers, and finally, a warmer Earth planet. Hence, it has become essential to study the public outlook on such grave issues. Social media platforms like Twitter offer an exceptional opportunity for analyzing public beliefs and sentiments. This work proposes a novel framework to assess the public belief on climate change using Sentiment and Emotion Analysis. The research utilizes Bi-LSTM networks to classify climate-related tweets into positive, negative, and neutral categories with a mean accuracy of 94.9%. The positively classified tweets were segregated, and their probability scores were obtained to quantify the intensity value signifying the strength of public belief. Furthermore, a lexicon-based emotion classification was augmented to understand better the public outlook associated with the tweets. Three case

studies concerned with climate change: Earth Day, Delhi Air Pollution, and Australian Bushfires were undertaken to verify the proposed framework results. The intensity values for the studies mentioned above were obtained as 98.50%, 96.57%, and 98.33%, respectively, which signified a strong foundation of public support to manage the prevalence of man-made climate adversities. The concerned authorities may utilize the study results in formulating policy decisions and taking better mitigation measures considering the public opinion beforehand.

The proposed work may be extended to support languages other than English. The addition of more languages may facilitate expanding the scope of the study. The framework was utilized for gaining insights regarding the public stance on climate change, though it can be applied to other domains such as healthcare, education, and the economy. Therefore, policymakers may utilize the study for improved decision-making with consideration of public opinion before devising strategies and regulations.

Acknowledgment

None.

Conflicts of interest

The authors have no conflicts of interest to declare.

References

- [1] Thompson LG. Climate change: the evidence and our options. *The Behavior Analyst*. 2010; 33(2):153-70.
- [2] Fawzy S, Osman AI, Doran J, Rooney DW. Strategies for mitigation of climate change: a review. *Environmental Chemistry Letters*. 2020; 18:2069-94.
- [3] Bergquist P, Goldberg MH, Gustafson A, Leiserowitz A, Rosenthal SA, Marlon J. Information about the human causes of climate change influences climate causal attribution, risk perceptions, and policy support. *Thinking & Reasoning*. 2021:1-13.
- [4] Bhushan B. Citizens, infiltrators, and other the nature of protests against the citizenship amendment act. *South Atlantic Quarterly*. 2021; 120(1):201-8.
- [5] Narayanan S. Understanding farmer protests in India. *Academics Stand Against Poverty*. 2021; 1(1):137-44.
- [6] Ain QT, Ali M, Riaz A, Noureen A, Kamran M, Hayat B, et al. Sentiment analysis using deep learning techniques: a review. *International Journal of Advanced Computer Science and Applications*. 2017; 8(6):424-33.
- [7] <https://www.oberlo.in/blog/twitter-statistics>. Accessed 2 July 2021.
- [8] Nahar KM, Jaradat A, Atoum MS, Ibrahim F. Sentiment analysis and classification of arab jordanian facebook comments for jordanian telecom companies using lexicon-based approach and machine learning.

- Jordanian Journal of Computers and Information Technology. 2020; 6(3):247-62.
- [9] Al-agma I, Abu-dahrooj O. Multi-level analysis of political sentiments using twitter data: a case study of the Palestinian-Israeli conflict. *Jordanian Journal of Computers and Information Technology*. 2019; 5(3):195-215.
- [10] Soumya S, Pramod KV. Sentiment analysis of malayalam tweets using machine learning techniques. *ICT Express*. 2020; 6(4):300-5.
- [11] Ansari MZ, Aziz MB, Siddiqui MO, Mehra H, Singh KP. Analysis of political sentiment orientations on twitter. *Procedia Computer Science*. 2020; 167:1821-8.
- [12] Heikal M, Torki M, El-makky N. Sentiment analysis of Arabic tweets using deep learning. *Procedia Computer Science*. 2018; 142:114-22.
- [13] Behera RK, Jena M, Rath SK, Misra S. Co-LSTM: convolutional LSTM model for sentiment analysis in social big data. *Information Processing & Management*. 2021; 58(1).
- [14] Muhammad PF, Kusumaningrum R, Wibowo A. Sentiment analysis using word2vec and long short-term memory (LSTM) for Indonesian hotel reviews. *Procedia Computer Science*. 2021; 179:728-35.
- [15] Bose R, Dey RK, Roy S, Sarddar D. Analyzing political sentiment using twitter data. In *information and communication technology for intelligent systems 2019* (pp. 427-36). Springer, Singapore.
- [16] Syed AZ. Applying sentiment and emotion analysis on brand tweets for digital marketing. In *Jordan conference on applied electrical engineering and computing technologies 2015* (pp. 1-6). IEEE.
- [17] Dubey AD. Twitter sentiment analysis during COVID-19 outbreak. Available at SSRN: <https://ssrn.com/abstract=3572023>
- [18] Dubey AD. Decoding the twitter sentiments towards the leadership in the times of COVID-19: a case of USA and India. Available at SSRN: <https://ssrn.com/abstract=3588623>.
- [19] Kirelli Y, Arslankaya S. Sentiment analysis of shared tweets on global warming on twitter with data mining methods: a case study on Turkish language. *Computational Intelligence and Neuroscience*. 2020.
- [20] Loureiro ML, Alló M. Sensing climate change and energy issues: sentiment and emotion analysis with social media in the UK and Spain. *Energy Policy*. 2020.
- [21] Reyes-menendez A, Saura JR, Alvarez-alonso C. Understanding# world environmentday user opinions in twitter: a topic-based sentiment analysis approach. *International Journal of Environmental Research and Public Health*. 2018; 15(11):1-18.
- [22] Hochreiter S. The vanishing gradient problem during learning recurrent neural nets and problem solutions. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*. 1998; 6(2):107-16.
- [23] Elfaik H. Deep bidirectional LSTM network learning-based sentiment analysis for arabic text. *Journal of Intelligent Systems*. 2021; 30(1):395-412.
- [24] Jang B, Kim M, Harerimana G, Kang SU, Kim JW. Bi-LSTM model to increase accuracy in text classification: combining word2vec CNN and attention mechanism. *Applied Sciences*. 2020; 10(17):1-14.
- [25] Sunny MA, Maswood MM, Alharbi AG. Deep learning-based stock price prediction using LSTM and Bi-directional LSTM model. In *novel intelligent and leading emerging sciences conference 2020* (pp. 87-92). IEEE.
- [26] Li C, Zhan G, Li Z. News text classification based on improved Bi-LSTM-CNN. In *international conference on information technology in medicine and education 2018* (pp. 890-3). IEEE.
- [27] Crisóstomo DCFH, Abílio DCJO, Ferreira DCOL, Pozzobon DBP, Dos SDRM, Olino DAA, et al. Rice crop detection using LSTM, Bi-LSTM, and machine learning models from sentinel-1 time series. *Remote Sensing*. 2020; 12(16):1-25.
- [28] Hu P, Tong J, Wang J, Yang Y, De OTL. A hybrid model based on CNN and Bi-LSTM for urban water demand prediction. In *congress on evolutionary computation 2019* (pp. 1088-94). IEEE.
- [29] Yin J, Deng Z, Ines AV, Wu J, Rasu E. Forecast of short-term daily reference evapotranspiration under limited meteorological variables using a hybrid bi-directional long short-term memory model (Bi-LSTM). *Agricultural Water Management*. 2020.
- [30] Xavier LC, Silva SM, Carvalho TM, Pontes FJD, Souza FFD. Use of machine learning in evaluation of drought perception in irrigated agriculture: the case of an irrigated perimeter in Brazil. *Water*. 2020; 12(6):1-20.
- [31] Biesbroek R, Badloe S, Athanasiadis IN. Machine learning for research on climate change adaptation policy integration: an exploratory UK case study. *Regional Environmental Change*. 2020; 20(3):1-13.
- [32] An X, Ganguly AR, Fang Y, Scyphers SB, Hunter AM, Dy JG. Tracking climate change opinions from twitter data. In *workshop on data science for social good 2014* (pp. 1-6).
- [33] <https://www.kaggle.com/edqian/twitter-climate-change-sentiment-dataset>. Accessed 23 March 2021.
- [34] <https://pypi.org/project/regex/>. Accessed 23 March 2021.
- [35] Johnson JM, Khoshgoftaar TM. Survey on deep learning with class imbalance. *Journal of Big Data*. 2019; 6(1):1-54.
- [36] <https://scikit-learn.org/stable/>. Accessed 23 March 2021.
- [37] Rai A, Borah S. Study of various methods for tokenization. In *applications of internet of things 2021* (pp. 193-200). Springer, Singapore.
- [38] <https://keras.io/>. Accessed 25 March 2021.
- [39] Huang Y, Jiang Y, Hasan T, Jiang Q, Li C. A topic BiLSTM model for sentiment classification. In *proceedings of the 2nd international conference on innovation in artificial intelligence 2018* (pp. 143-7).
- [40] Hochreiter S, Schmidhuber J. Long short-term memory. *Neural Computation*. 1997; 9(8):1735-80.

[41] Kingma DP, Ba J. Adam: a method for stochastic optimization. arXiv preprint arXiv:1412.6980. 2014.

[42] Chen P, Ye J, Chen G, Zhao J, Heng PA. Robustness of accuracy metric and its inspirations in learning with noisy labels. arXiv preprint arXiv:2012.04193. 2020.

[43] Mohammad SM, Turney PD. Crowdsourcing a word-emotion association lexicon. *Computational Intelligence*. 2013; 29(3):436-65.

[44] Mohammad S, Turney P. Emotions evoked by common words and phrases: using mechanical turk to create an emotion lexicon. In proceedings of the NAACL HLT 2010 workshop on computational approaches to analysis and generation of emotion in text 2010 (pp. 26-34).

[45] Plutchik R. The nature of emotions: human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice. *American Scientist*. 2001; 89(4):344-50.

[46] Dietz T. Earth day: 50 years of continuity and change in environmentalism. *One Earth*. 2020; 2(4):306-8.

[47] Sharma S, Mathur S. Analyzing the patterns of Delhi's air pollution. In *advances in data sciences, security and applications 2020* (pp. 33-44). Springer, Singapore.

[48] <https://www.unep.org/news-and-stories/story/air-pollution-and-climate-change-two-sides-same-coin>. Accessed 2 July 2021.

[49] Manisalidis I, Stavropoulou E, Stavropoulos A, Bezirtzoglou E. Environmental and health impacts of air pollution: a review. *Frontiers in Public Health*. 2020; 8:1-13.

[50] Vardoulakis S, Marks G, Abramson MJ. Lessons learned from the Australian bushfires: climate change, air pollution, and public health. *JAMA Internal Medicine*. 2020; 180(5):635-6.

[51] <https://github.com/twintproject/twint>. Accessed 20 June 2021.

[52] <https://www.unep.org/news-and-stories/story/why-earth-day-more-important-ever>. Accessed 2 July 2021.

[53] Dholakia HH, Garg A. Climate change, air pollution and human health in Delhi, India. In *climate change and air pollution 2018* (pp. 273-88). Springer, Cham.

[54] Kinney PL. Interactions of climate change, air pollution, and human health. *Current Environmental Health Reports*. 2018; 5(1):179-86.

[55] https://www.washingtonpost.com/world/asia_pacific/india-coronavirus-delhi-clean-air-pollution/2020/04/10/ac23dd1e-783e-11ea-a311-adb1344719a9_story.html. Accessed 9 July 2021.

[56] <https://www.bbc.com/news/world-australia-50887982>. Accessed 9 July 2021.

[57] <https://www.bbc.com/news/world-australia-50748419>. Accessed 24 September 2021.



Tishya Thukral is an undergraduate student at the Department of Computer Science, Acharya Narendra Dev College, University of Delhi, India, pursuing her Bachelor's degree in Computer Science. Her research interests include Machine Learning, Deep Learning, and Natural Language

Processing.
Email: tishya.thukral.cs.andc.du@gmail.com



Ashwani Varshney is a graduate student at the Department of Computer Science, University of Delhi, India, pursuing his Master's degree in Computer Science. His research interests include Machine Learning, Deep Learning, Natural Language

Analytics.
Email: ashwani.mcs19.du@gmail.com



Vibha Gaur received Ph.D. Degree in Computer Science from Department of Computer Science, University of Delhi, Delhi, India. She is a professor in the Department of Computer Science, Acharya Narendra Dev College. She has published more than 40 papers in international journals and conferences.

Her research interests include Requirement Engineering, Software Quality, Fuzzy Logic, etc.
Email: vibhagaur@andc.du.ac.in

Appendix I

S. No.	Abbreviation	Description
1	ANN	Artificial Neural Network
2	Bi-LSTM	Bi-directional Long Short-Term Memory
3	CAA	Citizenship (Amendment) Act
4	CNN	Convolutional Neural Network
5	DL	Deep Learning
6	EmoLex	NRC Word-Emotion Association Lexicon
7	ET0	Evapotranspiration
8	HTML	Hypertext Markup Language
9	KNN	K- Nearest Neighbor
10	LSTM	Long Short-Term Memory
11	NB	Naïve Bayes
12	NRC	National Research Council of Canada
13	RF	Random Forest
14	RNN	Recurrent Neural Network
15	SVM	Support Vector Machine
16	TF-IDF	Term Frequency- Inverse Document Frequency
17	URL	Uniform Resource Locators