Journal of Cleaner Production 312 (2021) 127820

Contents lists available at ScienceDirect

Journal of Cleaner Production

journal homepage: www.elsevier.com/locate/jclepro

A novel sentiment analysis framework for monitoring the evolving public opinion in real-time: Case study on climate change

May El Barachi [a,*], Manar AlKhatib [b], Sujith Mathew [c], Farhad Oroumchian [a]

[a] Faculty of Engineering & Information Sciences, The University of Wollongong in Dubai, Knowledge Village, P.O. Box 20183, Dubai, United Arab Emirates
[b] Faculty of Engineering & IT, British University in Dubai, Dubai International Academic City, P.O. Box 345015, Dubai, United Arab Emirates
[c] College of Technological Innovation, Zayed University, Khalifa City B, P.O. Box 144534, Abu Dhabi, United Arab Emirates

ARTICLE INFO

Handling editor: Cecilia Maria Villas Bôas de Almeida

Keywords:
Smart cities
Sentiment analysis
Opinion leaders
Social media analytics
Climate change

ABSTRACT

Smart city analytics involves tracking, interpreting, and evaluating the sentiments and emotions that are shared via online social media channels. Sentiment analysis of social media posts has become increasingly prominent in recent years as a means of gaining insights into how members of the public perceive current affairs. The ongoing research in this domain has leveraged multiple types of sentiment analysis. However, although the existing approaches enable researchers to acquire retrospective insights into public opinion, they do not enable a real-time evaluation. In addition, they are not readily scalable and necessitate the analysis of a significant amount of posts (in the millions) to facilitate a more in-depth evaluation. The study outlined in this paper was designed to address these shortcomings by presenting a framework that facilitates a real-time evaluation of the evolution of opinions shared by prominent public features and their respective followers; that is, high-impact posts. The developed solution encompasses a sophisticated Bi-directional LSTM classifier that was implemented and tested using a dataset consisting of 278,000 tweets related to the topic of climate change. The outcomes reveal that the proposed classifier achieved the following accuracies: 88.41% for discrimination; 89.66% for anger; 87.01% for inspiration; and 87.52% for joy - with negative emotions being more accurately classified than positive emotions. Similarly, the sentiment classification performance yielded accuracies of 89.32% for support and 89.80% for strong support, as well as 88.14% for opposition and 87.52% for strong opposition. In addition, the findings of the study indicated that geographic location, chosen topic, cultural sensitivities, and posting frequency all play a critical role in public reactions to posts and the ensuing perspectives they adopt. The solution stands out from existing retrospective analysis methods because it does not rely on the analysis of vast quantities of data records; rather, it can perform real-time, high-impact content analysis in a resource-efficient and sustainable manner. This framework can be used to generate insights into how public opinion is developing on a real-time basis. As such, it can have meaningful application within social media analysis efforts.

1. Introduction

Smart cities are principally designed with the goals of improving quality of life for inhabitants while promoting economic growth with a focus on sustainable development and efficient delivery of services (Angelidou, 2015). In the smart city development context, intelligence-gathering consists of continuous assessment and tracking of events, residents, and infrastructural and other assets. Thereafter, gathered intelligence is utilized to continually enhance the efficiency of urban operations, such as predicting volume and type of resources required, promoting optimal municipal services, and informing citizen-friendly amendments to policies and regulations (Harrison et al., 2010). One specific challenge in this regard is how to build a scalable model which can efficiently gather contextual real-time data concerning residents, assets, or events within constraints on resources. This work presents an *innovative and sustainable technological solution for real-time monitoring of public opinion's evolution, in smart cities.*

Public opinion represents the views, desires, and wants of the majority of a population concerning a certain issue, whether political, commercial, social, or other (Glynn and Huge, 2008). Up-to-date information about people's views and preferences is important for any business seeking to maintain or enhance the position of its brands or products, as

* Corresponding author.
E-mail addresses: MaiElbarachi@uowdubai.ac.ae, Manaralkhatib09@gmail.com (M. El Barachi), Sujith.Mathew@zu.ac.ae (S. Mathew), farhadoroumchian@uowdubai.ac.ae (F. Oroumchian).

https://doi.org/10.1016/j.jclepro.2021.127820
Received 19 February 2021; Received in revised form 11 May 2021; Accepted 31 May 2021
Available online 6 June 2021

well as its overall reputation. It also enables businesses to identify and predict which trends will be profitable in various areas such as recreation, wellness, and lifestyle. When such data is astutely incorporated in product or service design, it enables businesses to become or remain competitive. Governments, too, frequently base policy decisions on trends and reactions in public opinion, and monitoring such opinion is also valuable in identifying extremism and predicting and reacting to potentially problematic trends and occurrences. In the worst cases, for example, civil disobedience or even rioting may result from a build-up of negative opinion about a particular issue, while expressions of extremism, if not checked, have been known to be predictors of violent incidents, such as the online statements published by gunmen who go on to carry out terrorist attacks (Kibble). If governments can track and assess public sentiment and opinion in real time, the resulting information base can inform decision-making and underpin the prediction, correction, and prevention of unwanted events, for example by indicating where pressure is building or where intervention is necessary. Monitoring public opinion is thus considered as a valuable tool in emergency preparedness, the preemption of violence, and the maintenance of public order in general.

There are multiple issues to be considered when attempting to capture and analyze public opinion. These include the diversity of the sources to be monitored, the challenging nature of both acquiring and analyzing data, and the sheer volume of the data to be assessed. Among the most common instruments used to assess public opinion are interviews and random sample surveys. While those techniques can be used to capture public opinion even in countries without Internet access, both tools, however, have shortcomings regarding limited sample size, potential for bias in questions, and collection methods which lack transparency. Moreover, collection and analysis of data through these methods are expensive, laborious, and time-consuming.

Ever-expanding social media platforms have yielded huge volumes of user-generated data which, if effectively harnessed and analyzed, can reveal popular perceptions of current affairs and be used as complement to interviews and surveys. How best to accomplish this remains an ongoing topic in the research community. One means is to identify social media posts made by opinion leaders and perform sentiment analysis on such content to obtain a rapid and accurate reflection of opinion among the wider public.

The two-step flow of communication model (Lazarsfeld et al., 1948) conceptualized opinion leaders as individuals possessing influence within the society and who share their perspectives via the media to shape public opinion and perception. Typically, an opinion leader not only has expertise in a specific area but also the strategic know-how to utilize social networks to their advantage, connecting with individuals both within and outside their social circles. In particular, opinion leaders are characterized by popularity and competency.

Sentiment analysis is part of the wider field of social media content analysis and refers to the practice of systematically identifying, extracting, assessing, and considering subjective data by the application of computational linguistics techniques (Pang and Lee, 2008). Sentiment analysis has gained significant attention due to the ability it provides into understanding the public sentiments and feelings. More recently, sentiment analysis was employed for opinion mining in order to extract intelligence about the public's opinion from social media content. For instance, the work presented in (Kermanidis and Maragoudakis, 2013) focused on extracting and analysing sentiments in tweets posted about the results of a Greek election. The research in (La-sheng and Al Baadani, 2018), meanwhile, sought to use machine learning lexical-based applications to model, classify, and analyze the sentiment of Arabic-language social media posts. The authors also proposed substituting a five-level scale (where 1 = highly positive and 5 = highly negative) for the standard three-level positive-negative-neutral sentiment one. The work presented in (Itani, 2018) assessed the performance of two classifiers for sentiment analysis (Lexicon and Naïve Bayes) and put forward a sentiment analysis training dataset covering spam, positive, and negative pattern groups (Plank et al., 2016). Furthermore, the authors of (Almuqren and Cristea, 2016) developed a framework to assess the satisfaction and churn rate of Saudi telecommunication companies' subscribers, through the analysis of Arabic tweets and the feelings they encapsulated. In relation to US elections, Karami and Elkouri gathered millions of tweets about Senator Sanders, applying a mixed-method analysis to explain his popularity (Karami and Elkouri, 2019). Lastly, the authors of (Karami et al., 2018) gathered millions of tweets in the runup to the 2012 US elections and conducted modelling of topics and analysis of sentiments to assess people's economic concerns about the elections.

Existing approaches present certain shortcomings. Firstly, they yield an aggregated, retrospective picture of affect. The work presented in (Karami et al., 2018), for example, which constitutes a study of economic concerns expressed at the time of the 2012 US elections, was actually conducted in 2018. Furthermore, existing solutions fail to track the development of sentiment and evolution of opinion in real time, although fine-grained real-time analysis is precisely what is required if strategies are to be changed in line with public reaction to ongoing events. Finally, scalability and efficiency are both lacking, as meaningful conclusions rely on the harnessing and analysis of millions of social media posts, which clearly undermines their usefulness as tools for real-time tracking.

The current research seeks to address the shortcomings identified above by developing a framework which will enable real-time, high-impact social media content analysis in a resource-efficient and sustainable manner. This is achieved by analysing the sentiments contained in social media content posted by opinion leaders and the response from their followers in real time, thus revealing developments in opinion over time. The proposed solution was realized and tested using Twitter data collected over a 15 months span. Climate change was used as case study in our research, and a deep learning-based LSTM model was employed to classify the sentiments and emotions conveyed by 278,264 tweets posted by the environmental activist Greta Thunberg and her followers between 20/07/2019 and 30/10/2020. The obtained results yielded sophisticated reporting that provide insights about the evolution of Greta Thunberg's emotions over time as expressed by her tweets, the reaction of her followers to her posts, and a location-based sentiment analysis highlighting the street's opinion in different regions of the world. Those results complement existing studies on the impact of Greta's tweets on climate change activism (Sabherwalet al., 2021) (Roederer, 2020). The work in (Sabherwalet al., 2021) relied on a survey of 1303 US adults to investigate whether exposure to Greta's speech predicts collective action on climate change. The findings suggest that young activists such as Greta may motivate collective action, with a stronger effect on those who share the same political ideology. In (Roederer, 2020), the author investigates whether a Greta Thunberg effect that impacts corporate sentiments and tweeting behaviour about climate change exists. The findings indicate that the Greta effect does exist and that companies tend to react mainly to positive tweets from Greta.

Unlike existing approaches, the proposed framework offers a scalable technological solution to efficiently harnessing real-time development of public opinion. Moreover, as it offers situational awareness, it can be of use to government bodies seeking to monitor, predict, and react to real-time development in public opinion in relation to events underway. The contributions of this work are as follows:

1. The proposed solution represents an advanced framework focusing on the real-time tracking and sophisticated reporting of public opinion evolution.
2. Contrary to existing approaches focusing on the retrospective data analysis after events' occurrence, our approach is envisioned as a technological tool for capturing the public's opinion about events as they unfold, in an efficient, scalable, and timely manner. The

intelligence offered by such tool could allow for timely intervention and adjustment of policies as needed.

3. Our solution offers sentiment and emotion analysis capabilities as means to identify general feelings about current events as well as emotions underpinning those feelings and triggering actions. Such fine-grained classification can help in the inference of problematic behaviors and tendencies. Eight emotions and sentiments were addressed in this work.
4. Our framework combines several techniques to accurately classify the sentiments and emotions expressed through public opinion. Those techniques include the automatic extraction of social media content, the identification of opinion leaders, the classification of emotions and sentiments, and location-based analysis.

In the coming section, we discuss our deep learning approach for monitoring public opinion. The validation and extensive testing results are presented in Section 3. Conclusions are highlighted in Section 4.

2. Sentiment analysis framework for monitoring public opinion evolution

This work aims at the creation of a decision support tool that can be used by government entities for sentiment analysis and real time information gathering about the evolution of public opinion over time. Fig. 1 depicts the main components of our framework, which consists of: 1) a data acquisition module responsible of automated social media data extraction using scripts; 2) a data pre-processing module for cleaning and preparation of the data for the next stages; 3) an opinion leaders identification module to identify the most influential social media users in a certain domain; 4) a targeted data acquisition module focusing on the collection of the posts of the identified prominent public figures and their followings; and 5) a deep learning based analysis and reporting module yielding advanced sentiment intelligence reports.

2.1. Dataset

As case study, we focused on Twitter data and we selected a topic that generated numerous discussions, debates and changing opinions on social media, namely: ***Climate Change***. Beside the importance of this topic and its potential impact for the future of the earth, the subject of climate change has been heavily discussed by opinion leaders and prominent figures on social media and has led to important reactions from followings. This justifies its choice as case study for our research on public opinion monitoring and sentiment analysis. It should be noted that our framework in not limited to Twitter data, and can be used for the analysis of any social media textual content that can be automatically extracted using scripts (e.g. Facebook data).

Greta Thunberg is an activist and prominent figure on social media, with a large following on Twitter of more than 4.9 million users. We collected Greta's tweets and the replies of those who follow her, as means to build our dataset. The hashtags ***#fridaysforfuture, #climatestrike, #climatestrikeonline, #schoolstrike4climate, #facetheclimateemergency, #climateaction, #climateemergency, #climatecrisis, #digitalstrike, and #climatejustice*** were used, to collect 6240 tweets posted by Greta between 20/07/2019 and 30/10/2020, for emotion analysis. Furthermore, 282,000 replies generated in reaction to the most popular 40 tweets posted by Greta were collected to gain insight into followers' sentiments and reactions. Our dataset characteristics are summarized in Table 1.

Over the studied period of 16 months, Greta generated a large number of tweets (i.e. 6240), with an average of 13–14 tweets posted per days. Fig. 2 consists of a word cloud highlighting the topics found in Greta's tweets. As shown, the most prominent topics discussed were: climate week, future, crisis, global change, strikes, activists, and leaders. Furthermore, Greta's posts tend to generate important reactions, namely: 13 million replies and 89 million retweets for the posts she made in the studied period. Fig. 3 depicts the high level of engagement with Greta's posts achieved over the period of July 2019 to August 2020.

2.2. Cleaning and pre-processing of dataset

Cleaning and pre-processing constitute an essential step needed for the removal of irrelevant information and the preparation of social media data for classification and analysis.

Fig. 4 shows the steps we employed to achieve that task, namely:

- Eliminating irrelevant information such as dates, times, numbers, URLs, as well as re-tweet and mention symbols.
- Replacement of emoticons with their emojis equivalent, to capture the emotions behind those elements.

Table 1
Dataset characteristics.

Total Number of tweets	6240
Total number of followers	157,924,181
Total Number of retweets	89,793,779
Total number of replies	13,368,665
Start date	20/07/2019
End Date	30/10/2020

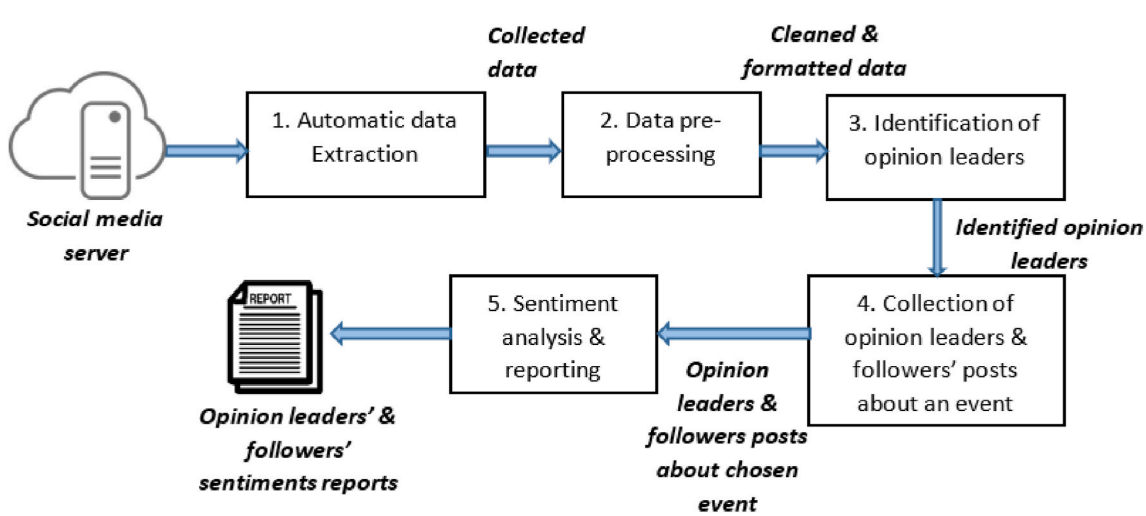


Fig. 1. Overview of sentiment analysis framework for real-time monitoring of public opinion.

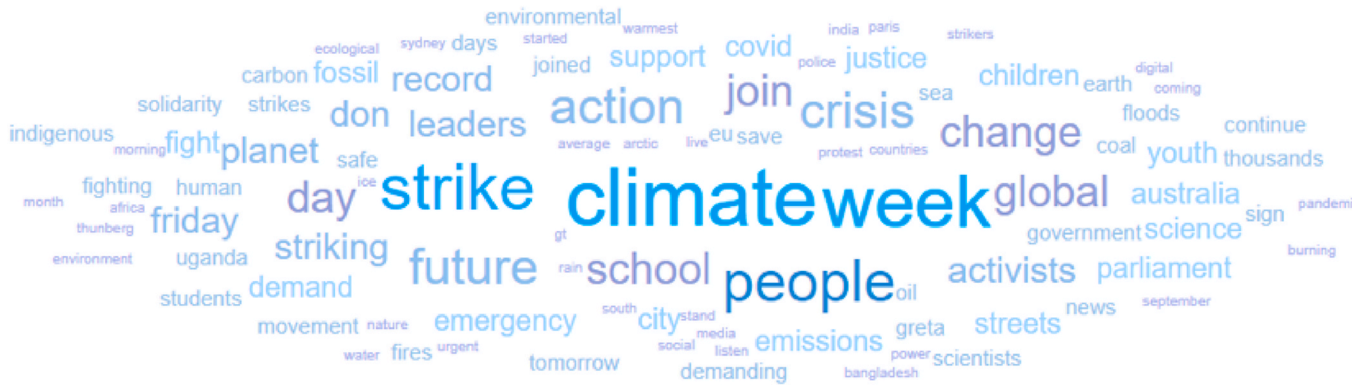


Fig. 2. Word cloud presentation of the topics found in Greta Thunberg's tweets.

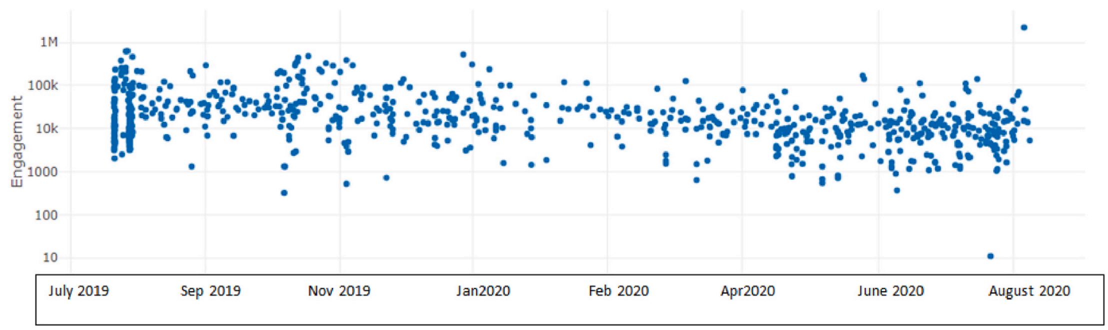


Fig. 3. Level of engagement over Twitter, achieved by Greta Thunberg.

Fig. 4. Methodology for data preprocessing.

- Tokenization of emojis by placement of spaces between successive elements.
- Words' normalization and removal of unwanted punctuation marks and symbols.
- Replacement of elongations with single letter occurrence (e.g. replacing 'soooooooo greaaaaaaaaaat' with 'so great').
- In natural language processing, it is important to analyze the relationship between words. This can be conducted using on-hot encoding or word embedding. Adopting the Word2Vec word embedding approach (Church, 2017) in our model, we prepare data for training by substituting each tweet with a 2D vector representation. Word embedding is a word representation technique in which words with similar meanings have a similar representation. Word2-Vec is a popular word embedding method that enables the efficient learning of words embedding (or meaning) from a text corpus. In this work, we relied on the pre-trained Word2Vec model (Church, 2017) to represent tweet words as real-valued vectors in a pre-defined vector space. This vector representation is used to train an LSTM model, in order to capture the words' long-term dependencies.
- Finally, we process unknown words using an approach inspired by FastText (Wu and Manber, 1992), in which n-grams of the unknown words are used to attempt and find a suitable embedding in the model. More specifically, the N-gram technique predicts surrounding words, given a current word.

2.3. Opinion leaders' identification methodology

We have previously proposed an algorithm for opinion leaders' identification based on competency and popularity metric, which we leveraged in this work, namely: The OLFinder (Aleahmad et al., 2016). The algorithm computes the popularity score as function of the number of followers, while relying on the Latent Dirichlet Allocation (LDA) and the Divergence from Randomness (DFR) to compute a person's competency in a certain domain. Finally, an opinion leadership score is calculated win a field (d) using equation (1) below:

Fig. 5 shows an overview of the OLFinder algorithm steps. For more details, reference (Aleahmad et al., 2016) can be consulted.

2.4. Sentiment and emotion analysis approach

In this work, we conduct sentiment and emotion analysis for advanced insights about public opinion. Despite their interchangeable usage, the words 'emotion' and 'sentiment' have different meanings. While sentiment represents a general feeling or thought, emotion signifies a state that influence our behavior and is triggered by a stimulus (Sailunaz and Alhajj, 2019). While sentiment analysis addresses the classification of general sentiment into positive or negative feelings, emotion analysis provides insights about the specific psychological states (i.e. emotions) behind those feelings. In this work, we employ sentiment analysis to gain insights about the reactions of Greta's followers about her posts, while we conduct emotion analysis to

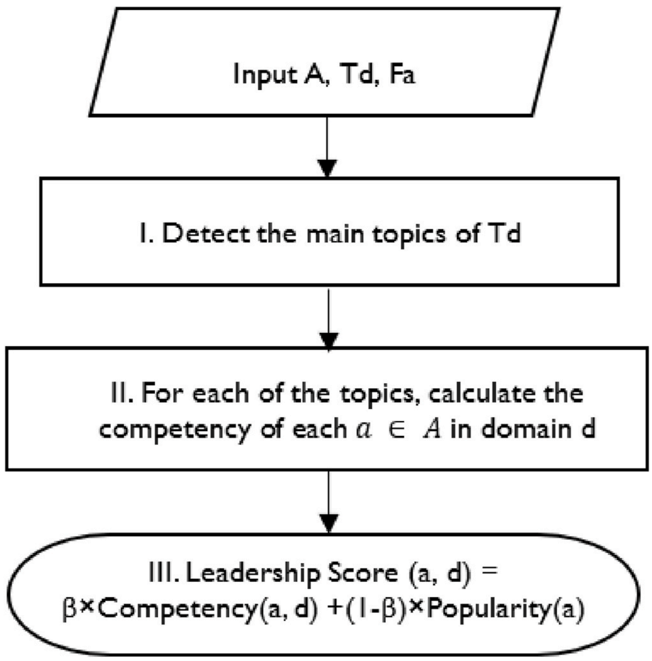


Fig. 5. Steps of the opinion leaders' identification algorithm.

understand the emotions that underpin Greta's posts and her actions. The emotions and sentiments considered in this work are presented in Table 2.

As shown, we chose four categories of sentiments that are suitable for representing followers' opinions about climate change related debates, namely: 1) Support; 2) strong support; 3) opposition; and 4) strong opposition. As for the emotions considered when analyzing Greta's tweets, four main categories were considered: 1) Joy; 2) Inspiration; 3) Anger; and 4) Discrimination. Those four emotions were chosen as they were the most prominent emotions observed when we examined Greta's speech.

To conduct emotion and sentiment analysis, we employed a 6-stage approach, as depicted in Fig. 6. Following the data pre-processing and word embedding steps, we rely on a deep learning LSTM model, followed by a ReLu activation, Dropout and SoftMax layers to achieve accurate text classification.

2.4.1. Using LSTM model for sentiment and emotion analysis

Considered as a sub-field of Artificial Intelligence (AI), machine learning enables computer systems to automatically learn from data patterns rather than being programmed following particular rules. While conventional supervised machine learning and probabilistic models, such as Naïve Bayes and Bag of words, have been used for various natural language processing tasks, such techniques require carefully chosen features. Given the nature of social media content which is both diverse and short in size, it is very challenging to obtain useful classification features.

Deep learning models are machine learning models that consist of neural networks with multiple hidden layers. Due to their complex architectures and their ability to continuously improve their learning, deep neural networks are able to process complex data and uncover hidden patterns. Convolutional neural networks (CNNs) (Kim, 2014) and Recurrent neural networks (RNNs) (Elman, 1990) are two popular types of deep learning models. While CNNs can learn local features from words or sentences, RNNs have the ability to learn dependencies between words in sequential data, thus making such model suitable for sentiment analysis related tasks. Long Short-term Memory models (LSTMs) are a type of RNNs designed to consider dependencies in longer data sequences and address the vanishing gradient problem faced by traditional RNNs. This ability is enabled by memory capabilities that maintain the operations' states on previous vectors in the sequence (Sundermeyer et al., 2012), thus capturing not only the meaning but the context in which a word is used. Due to this capability, LSTMs were found to be efficient models for the sentiment classification of short text such as social media tweets (Zhang et al., 2018). Fig. 7 depicts the LSTM model we used for both sentiment and emotion analysis.

Table 2
Considered emotions and sentiments.

Sentiments considered	Emotions considered
Support	Joy
Strong support	Inspiration
Opposition	Anger
Strong opposition	Discrimination

Our architecture leverages a bi-directional LSTM model, in which a backward LSTM layer maintains previous words' context and a forward LSTM maintains next words' context, thus capturing the meaning and the contextual relationship between words – a relationship that is critical for analyzing sentiments and emotions. As shown in the figure, tokenized tweets forming word vectors (from the word embedding step) are used as input to the forward and backward LSTM layers (each of hidden size h). The output of the two LSTM layers are concatenated to form a vector of size 2 h. The ReLu fully connected layer is activated after this operation, followed by a dropout layer to reduce overfitting, and ending with the evaluation of the sentiment using the SoftMax layer, based on the following rules:

Table 3 shows the model's optimal configuration parameters obtained through experimentation.

3. Validation and results' analysis

The topic of climate change has resulted in numerous debates and polarized views on social media. Defined as the large-scale change in weather patterns and the global warming resulting from greenhouse gazes' emissions, climate change has far reaching and serious consequences affecting forests, coastal areas, agriculture, water resources, species and natural areas, as well as health and infectious diseases. We chose global warming as our case study, due to the importance of this topic and its effect on the sustainability of life on earth. Moreover, different stakeholders, such as companies and the private sector, government entities, and citizens have often shared opposing views on this topic. As shown in table four, 278,264 tweets were collected about the case study and used for the validation of our framework. 80% of the dataset was used for training and 20% of the dataset for testing. Furthermore, three types of analysis were conducted: 1) Emotion analysis of all the tweets posted by Greta Thunberg between July 2019 and October 2020 to gain an insights about the emotions that underpin her posts and actions; 2) sentiment analysis of Greta's followers' replies in relation to her top 40 tweets (tweets that generated a large number of reactions); and 3) Sentiment analysis of Greta's followers analyzed by location, to gain an understanding of how culture and geographic location can impact perception. The results of the three categories of analysis conducted are presented in the coming sections.

3.1. Sentiment and emotion analysis results

3.1.1. Evolution of Greta Thunberg's emotions over time

Fig. 8 depicts the emotional frequency analysis of 6240 tweets posted by Greta Thunberg between July 20, 2019 and October 30, 2020, while Table 8 and Fig. 9 highlight followers' reactions to those tweets. As shown in Fig. 8, the predominant emotions exhibited by Greta's posts are related to the anger emotion (with 43.7% of tweets associated with anger), followed by joy (with 38.1% of tweets). The inspiration emotion comes in third place with 17.4% of tweets associated with it. The discrimination emotion is negligible in Greta's speech, with 0.8% of

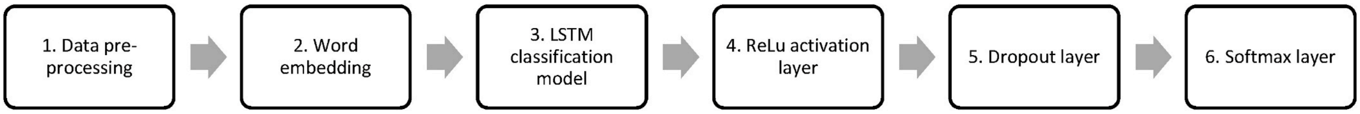


Fig. 6. Overview of sentiment and emotion analysis approach.

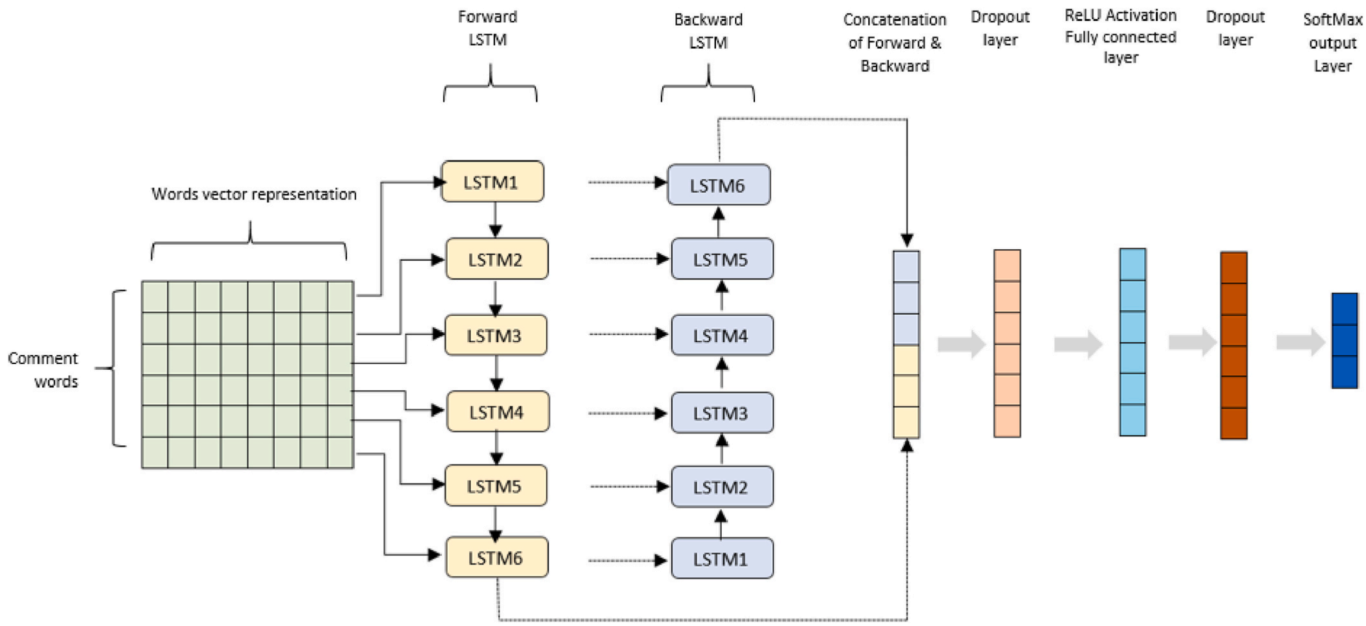


Fig. 7. Bi-directional LSTM model for sentiment and emotion classification.

Table 3
LSTM model configuration parameters.

Hyper-parameters	Value
LSTM hidden state dimension	200
Number of units in fully connected layer	30
Dropout rate	0.5
Learning rate	0.001
Number of epochs	10
Batch size	50

Table 4
Characteristics of dataset analyzed in relation to the climate change case study.

Analysis type	Data type	Number of tweets	Date range
Emotion Analysis	Tweets posted by Greta	6240	20/07/2019 to 30/10/2020
Sentiment Analysis	Greta's followers' reactions to top 40 tweets	213,982	20/07/2019 to 31/08/2020
Location-based Sentiment Analysis	Greta's followers' reactions to top 40 tweets – analyzed by geographic location	58,042	20/07/2019 to 31/08/2020
Total number of tweets analyzed	**278,264 tweets**		

tweets associated with it. A more in-depth analysis shows significant activity and associated number of emotions in the month of September 2019, while the least amount of emotions/activities was observed in July 2019. In September 2019, Greta was very active on Twitter with 1311 tweets and re-tweets posted. During that month, she shared information about the all schools strike week related activities, posting photos about events happening in Portugal, Stockholm, Rome, London, Mumbai, Lausanne, Freiburg, and Germany. During that month, she also traveled to give an important speech to the United Nations at the New York Climate Action Summit, and shared many tweets about her speech and participation. Throughout this month, Greta showed a mix of emotions with 495 tweets associated with the Anger emotion about climate change and how it is impacting earth, 492 tweets associated with Joy emotion and 317 tweets related to inspiration about the strikes taking place and the environmental activism the world is witnessing, followed by 7 tweets related to discrimination about her age and mental health issues. October 2019 was also an active month, as Greta continued her visits to other countries to join the climate strike in Denver, Colorado, Wyoming, Idaho, Montana and Alberta. She exhibited 212 tweets associated with Joy, 165 tweets associated with anger, 109 tweets associated with inspiration, and 3 tweets associated with discrimination.

Finally, July 2019 was observed as the month with the least engagement from Greta on twitter. During that month, she was still working on convincing followers that she does not support any political movement and is against all forms of fascism and violence. During that month, followers were commenting that she is still a child who does not understand climate change and that she should focus on going to school and getting an education. The top emotion observed during that month is anger, and the least observed is inspiration.

It is interesting to note that followers' reactions track a similar trend as tweets' posting behavior. Indeed, months with a high number of posts trigger more reactions than months with less frequent posting activities. As shown in Table 5 and Fig. 9, September 2019, October 2019, December 2019, January 2020, February 2020, and March 2020 are months that saw important reactions from Greta's followers and those months were associated with increased posting activities from Greta. In general, likes are the preferred mode of reaction, followed by retweets, then replies (see Table 6).

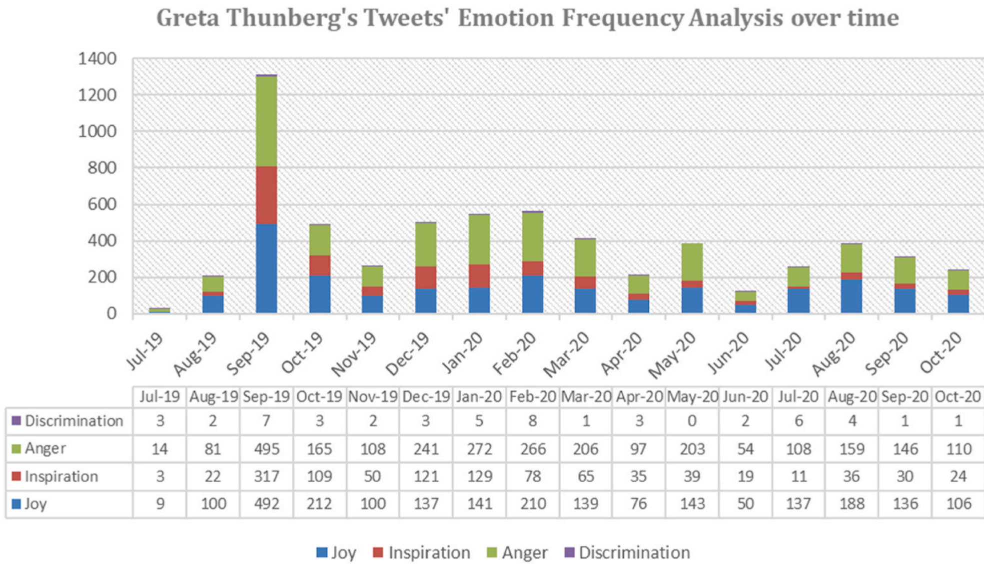


Fig. 8. Greta Thunberg –Tweets' Emotion frequency analysis over the months.

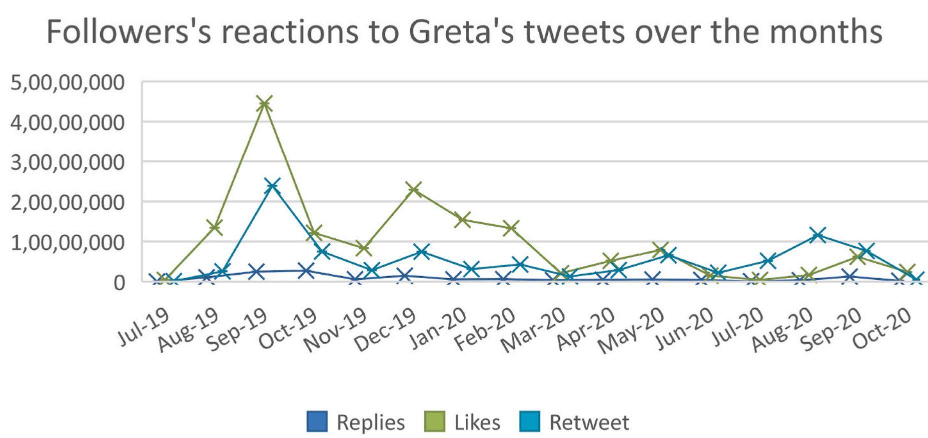


Fig. 9. Greta Thunberg – Followers' reactions to tweets over the months.

Table 5
Statistics about followers' reactions about Greta Thunberg's posts over the months.

Reactions/Months	Replies	Likes	Retweets
July 2019	43,609	438,182	119,939
August 2019	1,069,830	13,474,705	2,579,891
September 2019	2,480,770	44,505,077	23,904,141
October 2019	2,717,715	12,143,689	7,514,051
November 2019	581,214	8,321,286	2,848,776
December 2019	1,464,741	22,934,717	7,473,449
January 2020	590,619	15,404,431	3,133,014
February 2020	620,092	13,306,684	4,285,868
March 2020	374,072	2,095,344	1,310,795
April 2020	472,949	518,5344	289,4235
May 2020	564,197	7,917,836	6,576,558
June 2020	412,468	1,487,624	2,209,156
July 2020	119,567	429,1179	5,182,705
August 2020	343,911	16,55,868	11,612,256
September 2020	1,279,280	6,191,011	7,660,937
Ocober 2020	233,631	2,433,204	488,008
Total	**13,368,665**	**157,924,181**	**89,793,779**

3.1.2. *Analysis of followers' reactions*

In order to gain deeper insights of Greta's emotions conveyed through her tweets and the sentiments those tweets yielded with her followers, we selected 40 of her tweets that generated the most reactions and conducted emotional analysis for those individual tweets as well as sentiment analysis of the followers replies to those tweets. Fig. 10 shows a fluctuation of Greta's emotions with ongoing events and indicate that more important reactions are associated with tweets conveying joy and discrimination emotions. On the other hand, tweets conveying anger and inspiration emotions seem to generate less reactions.

Table 6
Performance evaluation of emotion classification – Greta's 6240 tweets.

Category	Accuracy	Precision	Recall	F-Measure
Discrimination	88.41%	89.62%	88.7%	89.16%
Anger	89.66%	90.71%	90.41%	90.56%
Inspiration	87.01%	87.15%	88.52%	87.83%
Joy	87.52%	87.87%	89.95%	88.90%

Fig. 11 depicts followers' reactions sentiment analysis, in relation to Greta's most popular 40 tweets. The January 3rd' 2020 tweet related to her birthday celebration (classified under the joy emotion) garnered the highest support (95%). The lowest support (19%) was obtained for the October 3rd' 2019 tweet (classified under inspiration), in which Greta was talking about the CO2 budgets and the science behind the impact of climate change. In general, we observed that when Greta posts inspirational tweets about change and activism, she receives positive and

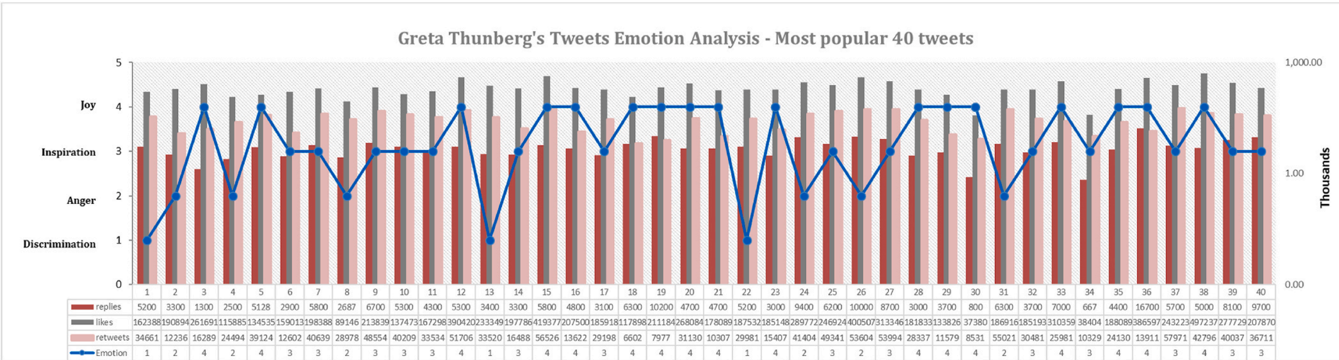


Fig. 10. Emotion Analysis of Greta Thunberg's most popular 40 tweets.

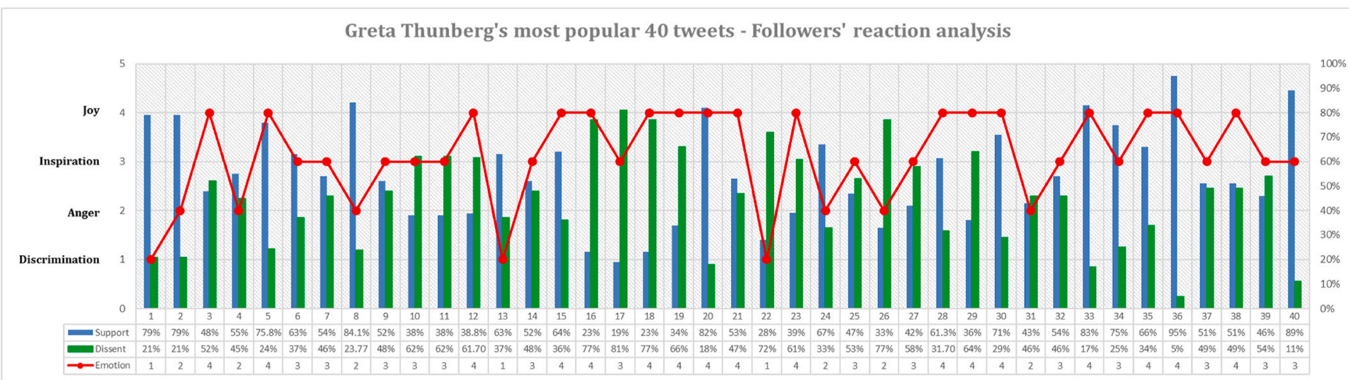


Fig. 11. Greta Thunberg's most Popular 40 Tweets - Followers' reactions analysis.

supportive replies. However, when she tweets about science, budget and political news related to climate change, she receives negative reactions often accusing her that those posts are not her tweets, but rather an adult's tweets using her for their agenda.

In terms of emotions' classification accuracy, the following results were obtained: 88.41% for discrimination; 89.66% for anger; 87.01% for inspiration; and 87.52% for joy. This implies that negative emotions are more accurately classified than positive emotions. Similarly, the sentiment classification performance yielded accuracies of 89.32% for support and 89.80% for strong support, as well as 88.14% for opposition and 87.52% for strong opposition. In that case, positive sentiment was classified with slightly higher accuracy than negative sentiment (See Table 7).

3.2. Geographic location based analysis

To understand the street's opinion in different regions, on the topic of climate change, we conducted a location-based sentiment analysis of 50,055 replies from Greta's followers. As shown in Table 8 and Fig. 12, the support sentiment is led by Sweden (Greta's home country) with 11% of total support replied, followed by the UK with 9% of support replies. The least support is witnessed by the US followers (6% of support replies) followed by German followers (3% of supportive replies).

Table 7
Performance evaluation of sentiment classification – Greta's followers 213,982 replies.

Category	Accuracy	Precision	Recall	F-Measure
Support	89.32%	90.34%	90.54%	90.44%
Strong support	89.80%	89.51%	88.65%	89.08%
Opposition	88.14%	88.47%	89.11%	88.79%
Strong opposition	87.52%	87.37%	87.22%	87.29%

Table 8
Location based sentiment analysis statistics of Greta's followers replies.

Country	Support	Strong support	Opposition	Strong opposition
USA	2914	1555	5224	3104
UK	4570	4354	4634	4502
Sweden	5564	3209	2499	1652
Germany	1551	1602	1602	1519
Grand total	**50,055 replies**			

Similarly, strongly supportive replies are shown for British followers (9%), then Swedish followers (6%), then American and German followers (3% each). As for the opposition and strong opposition, it is witnessed from American and UK followers, then Swedish and German followers. Based on those results, we can observe that Swedish followers are mainly supportive to their homegirl activism. British followers are split between supporters and dissenters. American followers fall mainly in the dissenters' camp, while German followers are evenly distributed between support and dissent. This shows clearly that culture can play an important role in public opinion (see Table 9).

To delve deeper in this analysis, we selected 2 popular tweets, and further examined their sentiment analysis distribution across the world's regions. As shown in the table, the first tweet (Dec. 13, 2019) resulted in the highest number of replies with sentiments fluctuating between strong opposition and opposition. The second tweet dated November 7th' 2019 resulted mainly in support and strong support replies. Such support can be explained by the events that occurred during that period, such as the rainforest wild fires and the global climate strikes week (in September 2019), followed by forest fires in Bolivia (in October 2019), then the nomination of Greta Thunberg as youth power by the Time magazine (in December 2019).

Fig. 13 illustrates followers' sentiments associated in reaction to

Table 9
Location based sentiment analysis of top 2 tweets and related followers' replies.

Date	Likes	Retweets	Replies	Strong support	Support	Opposition	Strong opposition
12/13/2019[a]	390420	51706	5300	2.80%	36%	45.9%	15.8%
11/7/2019[b]	89073	28944	2687	7.50%	76.62%	4.54%	19.23%

[a] https://twitter.com/GretaThunberg/status/1205220013034094592.

[b] https://twitter.com/GretaThunberg/status/1192469138167541760.

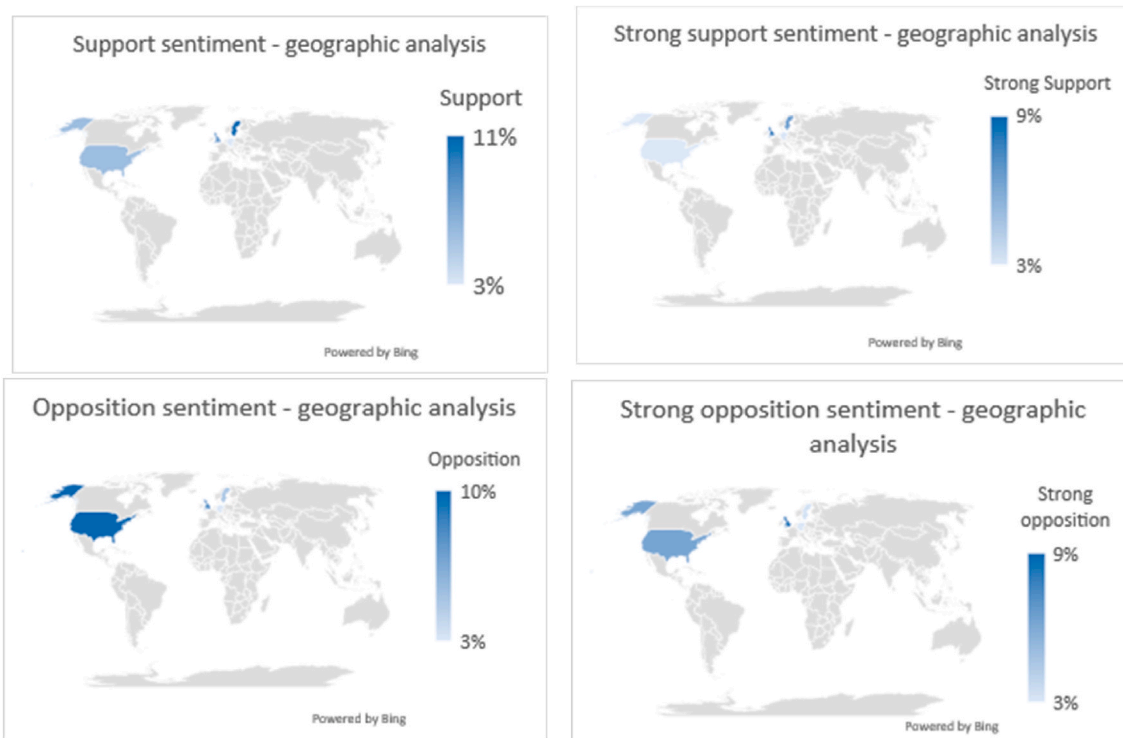


Fig. 12. Location based sentiment analysis results for 50,055 followers' replies.

Greta Thunberg's tweet posted on December 13, 2019 that is categorized as angry emotion. In their replies, many followers accused her of being political and not caring about the actual cause. To conducted location-based sentiment analysis, the replies sentiments were analyzed and categorized based on the followers' countries. Five main category locations were noted: 1) the USA, 2) the UK, 3) Sweden (Greta's home country), 4) Germany, and 5) others including the others countries like Canada, Australia, Spain, Middle East, France, …etc. Swedish followers displayed the biggest support (5.75%) in their replies, while the lowest support (0.15%) was obtained from the American followers. Furthermore, the percentage of positive sentiment replies was 38% (with 6% from Sweden, 2% from US, UK, and Germany respectively, and 26% from other countries). The percentage of negative sentiment replies was 62%, thus showcasing a largely negative sentiment about this tweet's topic (i.e. the political nature of Greta's actions).

Finally, Fig. 14 depicts the location-based sentiment analysis of Greta's tweet posted on November 7th, 2019. In this case, the followers' replies were mainly positive with 84% positive sentiment and 16% negative sentiment, with key support received from Sweden, the US and Germany.

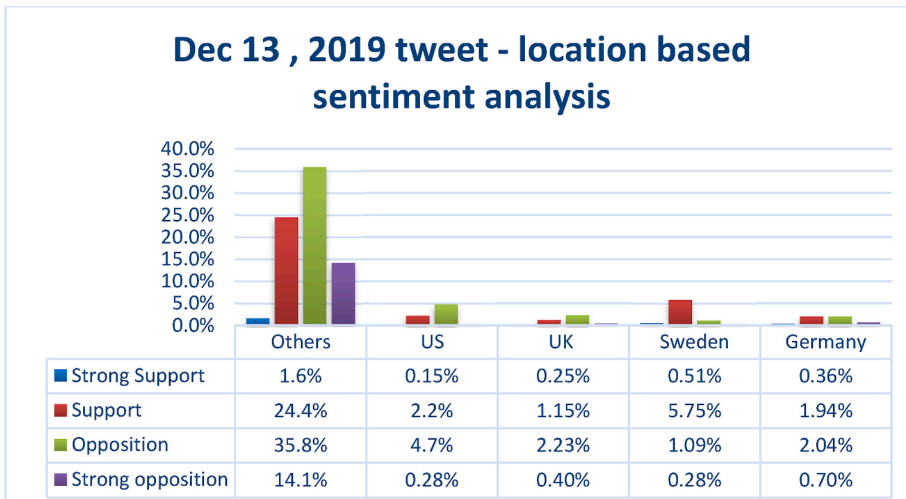
4. Conclusions

Smart cities are envisioned as the cities of the future that combine technological advancement with sustainability, efficient delivery of services, and citizens' welfare. Today's world is both fast-moving and hyper-connected; hence, occurrences of all types are met with large-scale and immediate public reaction which, if harnessed, constitutes very important predictive data for governments. Hence, the efficient and sustainable extraction and analysis of such opinion is vital if governments are to devise and implement the necessary corrective measures, prevent violence, and ensure law and order.

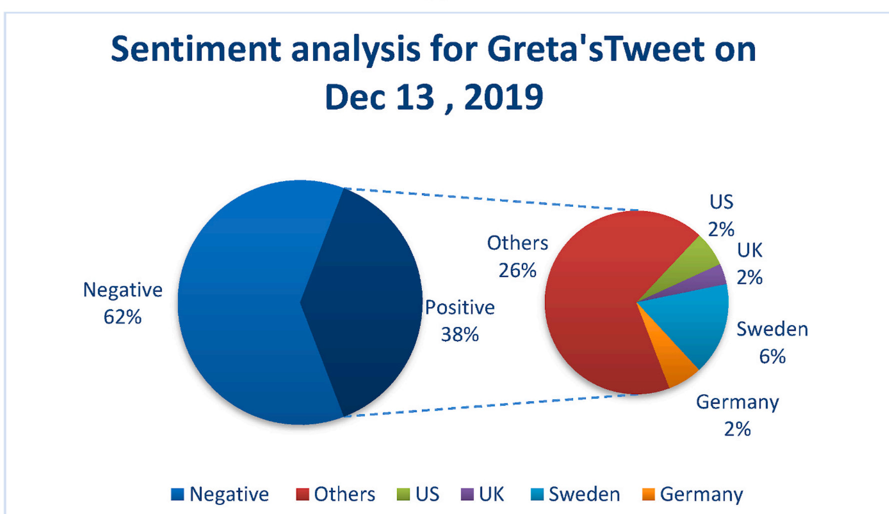
The solution put forward by the current study offers smart cities a means of monitoring public opinion's time evolution which is innovative, efficient, scalable, and sustainable. Social media posts generated by prominent public figures and their followings are analyzed to shed light on real-time developments in public reaction to matters ranging from natural disasters to political developments to sports events. Hence, analysis will enable quantification and prediction of public opinion to unfolding events by government actors, who will thus be able to decide how to act in light of the same.

The framework developed in this research leverages multiple techniques simultaneously to generate real-time reports on the development of public sentiments and emotions. In the framework, four classes of sentiment (strong agreement, agreement, disagreement, and strong disagreement) along with four categories of emotion (Joy, inspiration, anger, and discrimination) were detected by means a bi-directional LSTM deep learning model. The model was tested using of 278,264 tweets posted by the environmental activist Greta Thunberg and her followers between 20/07/2019 and 30/10/2020, in relation to climate

(a)



(b)



(c)

Fig. 13. Location based sentiment analysis results for December 13th, 2019 tweet: a) December 13th tweet; b) location-based sentiment analysis statistics; c) geographic distribution of sentiments detected.

change. Three types of results were presented, including Greta's emotions evolution over time, followers' reactions, and location-based sentiment analysis highlighting the street's opinion in different regions of the world.

The obtained results showed that the predominant emotions exhibited by Greta's posts are related to the anger emotion (with 43.7% of tweets associated with anger), followed by joy (with 38.1% of tweets). The inspiration emotion comes in third place with 17.4% of tweets associated with it. The discrimination emotion is negligible in Greta's speech, with 0.8% of tweets associated with it. Furthermore, the fluctuation between those emotions is closely linked to events and developments in relation to climate change. For instance, climate related strikes were mainly associated with inspiration as an emotion.

On the other hand, followers' reactions seemed to mimic Greta's social media posting behavior, with months with a high number of posts leading to more reactions than months with less frequent posting activities. This result is expected, since tweets from a prominent figure can act as trigger to reactions from followers. As for the preferred mode to

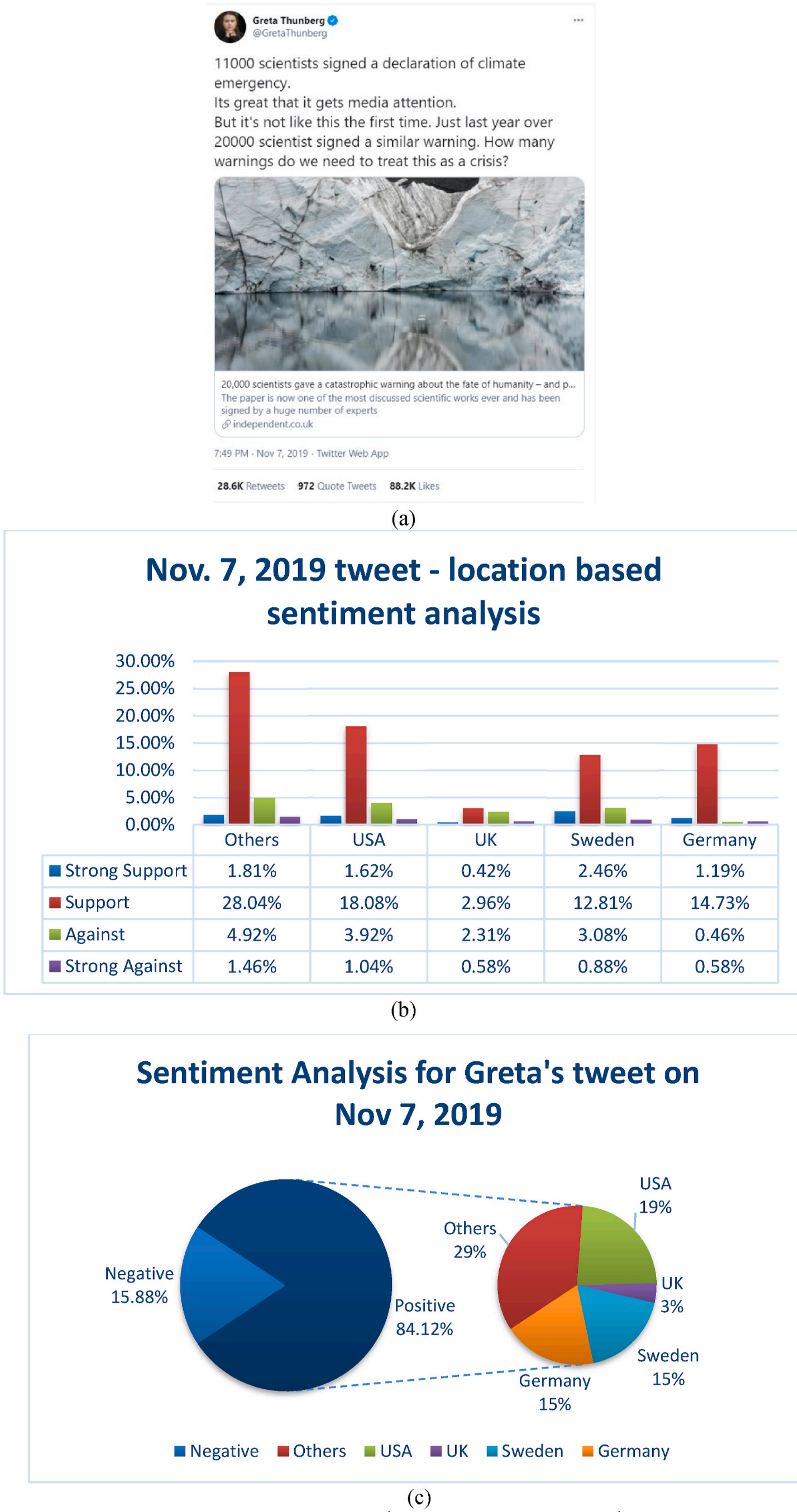


Fig. 14. Location based sentiment analysis results for November 7th, 2019 tweet: a) November 7th tweet; b) location-based sentiment analysis statistics; c) geographic distribution of sentiments detected.

reaction, likes were preferred, followed by retweets, then replies. More in-depth analysis of followers' reactions showed that more important reactions are associated with tweets conveying joy and discrimination emotions, while limited reaction was observed in relation to tweets conveying anger and inspiration emotions. It seems that Joy and inspiration act as stronger triggers to followers, leading to more important reaction to such tweets.

In terms of support or opposition, we observed that when Greta posts inspirational tweets about change and activism, she receives positive and supportive replies. However, when she tweets about science, budget and political news related to climate change, she receives negative reactions often accusing her that those posts are motivated by a political agenda.

In terms of classification accuracy, our model yielded 88.41% for discrimination; 89.66% for anger; 87.01% for inspiration; and 87.52% for joy - with negative emotions are more accurately classified than positive emotions. Similarly, the sentiment classification performance yielded accuracies of 89.32% for support and 89.80% for strong support, as well as 88.14% for opposition and 87.52% for strong opposition - with positive sentiment classified with slightly higher accuracy than negative sentiment.

Interesting results were obtained when conducting location-based sentiment analysis of followers replies, that is analyzed based on geographic location. The results show that culture can play an important role in shaping public opinion. In this case, we observed that Swedish followers were mainly supportive of Greta's activism - an expected result, since Greta is Swedish, not to mention that Sweden is considered as one of the most sustainable countries in the world. British followers on the other hand were split between supporters and dissenters, depending on their political ideologies, while American followers fell mainly in the dissenters' camp due to the skepticism of the Trump administration on the issue of climate change.

To sum up, several important findings were made as part of this work. Firstly, using both sentiment and emotion analysis can yield meaningful insights into both popular opinion and the more specific emotions which underlie it, enhancing its value as means to predict how people will behave. Secondly, tracking the evolution of public opinion and identifying the factors influencing it is critical for making effective decisions. Lastly, the results show that gathering and analyzing high-impact content posted on social media constitutes a sustainable and efficient means of assessing public opinion.

As future work, we plan to focus on the identification of extreme views that can act as precursors to violent actions. More specifically, there is an interest in the real-time monitoring of extreme views and the prediction of radicalization on social media platforms. Furthermore, the ability to analyze images and conduct facial sentiment analysis is envisaged to improve the capabilities of our solution. Finally, addressing the issue of fake discourse/account on social media is of interest to us.

Funding

Zayed University RIF grant number R19099 was used to fund this work.

CRediT authorship contribution statement

May El Barachi: Conceptualization, Methodology, Project administration, Writing – original draft, & editing. **Manar AlKhatib:** Data collection, Formal analysis, Software, implementation and testing, Writing – original draft. **Sujith Mathew:** Conceptualization, and paper review. **Farhad Oroumchian:** Design and modelling, Project administration.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Aleahmad, A., Karisani, P., Rahgozar, M., Oroumchian, F., 2016. OLFinder: finding opinion leaders in online social networks. J. Inf. Sci. 42 (5), 659–674. https://doi.org/10.1177/0165551515605217.

Almuqren, L., Cristea, A.I., 2016. Framework for sentiment analysis of Arabic text. In: Proceedings of the 27th ACM Conference on Hypertext and Social Media, pp. 315–317. Halifax, Nova Scotia, Canada.

Angelidou, M., September 2015. "Smart Cities: A Conjuncture of Four Forces," Cities Journal, vol. 47. Elsevier, pp. 95–106. https://doi.org/10.1016/j.cities.2015.05.004.

Church, K., 2017. Word2Vec. Nat. Lang. Eng. 23 (1), 155–162. https://doi.org/10.1017/S1351324916000334.

Elman, J.L., 1990. Finding structure in time. Cognit. Sci. 14 (2), 179–211.

Glynn, C.J., Huge, M.E., 2008. Public opinion. In: The International Encyclopedia of Communication. Wiley Online Library. https://doi.org/10.1002/9781405186407.wbiecp124.

Harrison, C.B., Eckman, R., Hamilton, P., Hartswick, J., Kalagnanam, J., Paraszczak, J., Williams, P., 2010. Foundations for smarter cities. IBM J. Res. Dev. 54 (4), 1–16. https://doi.org/10.1147/JRD.2010.2048257.

Itani, M., 2018. Sentiment Analysis and Resources for Informal Arabic Text on Social Media. Ph.D. Dissertation. Sheffield Hallam University.

Karami, A., Elkouri, A., 2019. Political popularity analysis in social media. In: Proceedings of the International Conference on Information. https://doi.org/10.1007/978-3-030-15742-5_44.

Karami, A., Bennett, L.S., He, X., 2018. Mining public opinion about economic issues: twitter and the U.S. Presidential election. Int. J. Strat. Decis. Sci. 9 (1), 18–28. https://doi.org/10.4018/IJSDS.2018010102.

Kermanidis, K.L., Maragoudakis, M., 2013. Political sentiment analysis of tweets before and after the Greek elections of May 2012. Int. J. Soc. Netw. Min. 1 (3–4), 298–317.

H. Kibble, "New Zealand Mosque Shooter Manifesto," Available Online: https://www.haydenkibble.com/new-zealand-mosque-shooter-manifesto-read-online/.

Kim, Y., 2014. Convolutional neural networks for sentence classification. In: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP 2014), pp. 1746–1751.

La-sheng, Y., Al Baadani, S., 2018. A Sentiment Analysis Approach Based on Arabic Social Media Platforms. *DEStech Transactions on Engineering and Technology Research*. https://doi.org/10.12783/dtetr/icmeit2018/23467.

Lazarsfeld, P.F., Berelson, B., Gaudet, H., 1948. The People's Choice: How the Voter Makes up His Mind in a Presidential Campaign. *Columbia University Press*.

Pang, B., Lee, L., 2008. Opinion mining and sentiment analysis. Found. Trends Inf. Retr. 2 (1–2), 1–135.

Plank, B., Søgaard, A., Goldberg, Y., 2016. Multilingual part-of-speech tagging with bidirectional long short-term memory models and auxiliary loss. In: Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, pp. 412–418. https://doi.org/10.18653/v1/P16-2067. Berlin, Germany.

Roederer, F., 2020. The Greta Thunberg Effect: an Analysis of Corporate Tweets on Climate Change. Master dissertation. Nova School of Business and Economics (NSBE), Lisbon, Portugal. Retrieved from. https://run.unl.pt/handle/10362/108599.

Sabherwal, A., et al., 2021. The Greta Thunberg effect: familiarity with Greta Thunberg predicts intentions to engage in climate activism in the United States. J. Appl. Soc. Psychol. 51 (4), 321–333. https://doi.org/10.1111/jasp.12737.

Sailunaz, K., Alhajj, R., September 2019. Emotion and sentiment analysis from twitter text. J. Comput. Sci. 36, 101003. https://doi.org/10.1016/j.jocs.2019.05.009.

Sundermeyer, M., Schlüter, R., Ney, H., 2012. LSTM neural networks for language modeling. In: Proceedings of Interspeech, pp. 194–197.

Wu, S., Manber, U., October 1992. Fast text searching: allowing errors. Commun. ACM 35 (10), 83–91. https://doi.org/10.1145/135239.135244.

Zhang, L., Wang, S., Liu, B., 2018. Deep learning for sentiment analysis: a survey. In: Wires data mining and knowledge discovery, 8. https://doi.org/10.1002/widm.1253, 4.