Public Health 215 (2023) 83–90

Contents lists available at ScienceDirect

Public Health

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

Original Research

Evolution of social mood in Spain throughout the COVID-19 vaccination process: a machine learning approach to tweets analysis

A. Turón, A. Altuzarra, J.M. Moreno-Jiménez, J. Navarro*

Grupo Decisión Multicriterio Zaragoza (GDMZ), Dpt. Economía Aplicada, Facultad de Economía y Empresa, Universidad de Zaragoza, Gran Vía 2, 50003, Zaragoza, Spain

ARTICLE INFO

Article history:
Received 18 July 2022
Received in revised form
5 December 2022
Accepted 8 December 2022
Available online 14 December 2022

Keywords:
COVID-19 vaccination process
Sentiment analysis
Machine learning
Multivariate statistics
Tweets
Social mood

ABSTRACT

Objectives: This paper presents a new approach based on the combination of machine learning techniques, in particular, sentiment analysis using lexicons, and multivariate statistical methods to assess the evolution of social mood through the COVID-19 vaccination process in Spain.
Methods: Analysing 41,669 Spanish tweets posted between 27 February 2020 and 31 December 2021, different sentiments were assessed using a list of Spanish words and their associations with eight basic emotions (anger, fear, anticipation, trust, surprise, sadness, joy and disgust) and three valences (neutral, negative and positive). How the different subjective emotions were distributed across the tweets was determined using several descriptive statistics; a trajectory plot representing the emotional valence vs narrative time was also included.
Results: The results achieved are highly illustrative of the social mood of citizens, registering the different emerging opinion clusters, gauging public states of mind via the collective valence, and detecting the prevalence of different emotions in the successive phases of the vaccination process.
Conclusions: The present combination in formal models of objective and subjective information would therefore provide a more accurate vision of social reality, in this case regarding the COVID-19 vaccination process in Spain, which will enable a more effective resolution of problems.

Introduction

The COVID-19 outbreak has been declared a pandemic by the World Health Organization because of its high rate of spread, severity and its frequent outcomes of severe pneumonia, respiratory failure and death.[1] Vaccination has become the main available public resource against the pandemic. However, the prejudices or sentiments of the general public and political leaders, as reflected in social media, are having a significant impact on the progression towards achieving vaccination targets.[1,2]

Social media such as Twitter, Facebook, YouTube and LinkedIn, with billions of users worldwide,[3] represent the preferred sites for sharing, almost instantly and very easily, thoughts, feelings and opinions on all kinds of events.[4] Twitter[5] is one of the most active platforms with approximately 290.5 million monthly active users worldwide in 2020 and was projected to keep increasing up to over 340 million users by 2024.[6] Every second around 6000 tweets on average are tweeted, which corresponds to more than 350,000 tweets sent per minute, 500 million tweets per day and around 200 billion tweets per year.[7]

Tweets are real-time messages with a maximum length of 280 characters at a time. They can be analysed based on *hashtags*, which refer to the symbol (#) in Twitter (for instance: #COVID19), containing a combination of the word *hash* from 'hash mark' and the word *tag*, that marks something belonging to a specific category. Hashtags make it easy to quickly find messages about a topic of interest as well as to collect all the sentiments and opinions of people in one place or country.[8–11]

One of the most promising methods for content analysis in social media is sentiment analysis.[12,13] It can be understood as a set of approaches, techniques and tools that extracts people's opinions, feelings and thoughts from users' text data by means of natural language processing methods.[14] Sentiment analysis through social media is growing rapidly within the international scientific community as a useful tool to understand people's opinions and attitudes on any important situation or phenomenon that affects public opinion.[11,15] For instance, natural disasters,[11] the Syrian refugee crisis,[4] the UK-EU referendum,[16] the impact of Brexit,[17] presidential

* Corresponding author.
E-mail address: jnavarrol@unizar.es (J. Navarro).

https://doi.org/10.1016/j.puhe.2022.12.003

or general elections in the United States,[18,19] Indonesia[20] and India,[21] the world cup soccer tournament,[22] extremism in social media,[23] 2019 EVALI outbreak[24] and the COVID-19 outbreak.[25,26]

This article presents a new approach based on the combination of machine learning techniques, in particular, sentiment analysis using lexicons, and multivariate statistical methods to assess the evolution of social mood through the COVID-19 vaccination process in Spain via tweet messages. Sentiment analysis, or opinion mining, will allow us to carry out the quantitative scrutiny of those tweets by extracting subjective information from the detection of eight basic emotions (anger, fear, anticipation, trust, surprise, sadness, joy and disgust) and the assessment of polarity (valence), that is, the neutral, positive or negative connotation of the language used. Multivariate statistical methods, or data mining, will provide figures and graphics that can synthesise objective information and knowledge about the vaccination process; in particular, properties of social structures and the patterns of relationships among actors.

The proposed methodology has been applied to the analysis of 41,669 tweets from February 2020 to December 2021. It shows how the opinions expressed in social media can be analysed, so that the social mood of citizens can be detected, opinion groups and their leaders can be identified, and social support for government measures can be evaluated.[27–30] The present combination in formal models of objective and subjective information about the vaccination process provides a more accurate vision of reality, which will enable a more effective resolution of problems.

Vaccination process in Spain

The vaccination strategy in Spain was published on 2 December 2020, with 11 updates up to the end of the considered period for analysis.[31] Four phases were defined according to available doses (see Table 1). The population groups to be vaccinated were established in order of priority , following an assessment based on criteria that incorporated the risk of exposure and transmission, the existence of previous serious illness, and the socio-economic impact of the pandemic on each population group.[32]

Methods

The methodological approach was based on Social Web Mining complemented with natural language processing and social network analysis. Messages were collected from social networks, preprocessed, and then their features were extracted to perform an analysis of society's opinion and mood regarding that critical event, and the way people related to each other and exchanged information on that event on social networks. The chart in Fig. 1 shows the methodological procedure that consists of three steps and three stages for each step.

Step 1: Corpus Determination

Stage 1.1. Data collection

We used a data set of 300,286 tweets in Spanish, posted between 27 February 2020 and 31 December 2021, that is, from the beginning of the pandemic until the end of the main stage of the vaccination process in Spain. The tweets were extracted from Twitter using the *twitterR* package, written in R programming language, accessing Twitter API 2.0. and searching in the full historical Twitter database. The search key was built from the following hashtags: #covid; #covid19; #*Yomevacuno* (I'm getting vaccinated); #*Yonomevacuno* (I'm not getting vaccinated); #*Negacionista* (denialist). The key string used to query the database was (*covid* OR *covid19*) AND (*Yomevacuno* OR *Yonomevacuno* OR *negacionista*).

It was referring to COVID and vaccination and to the pro- and anti-vaccine positions. The search terms were written in Spanish, and the condition that the messages be written in Spanish was added.

The attributes extracted from each tweet and its author were stored in two separate tables in the database according to the scheme shown in Table 2.

Other R packages such as *httr*, *RCurl* or *jsonlite* were used to extract the information from the Twitter API, in addition to *RMySQL* to manage the data through a MySQL database.

Stage 1.2. Data preprocessing

The tweets were preprocessed to eliminate all elements of the data that are susceptible to inconsistency or ambiguity, or, for reasons of efficiency, unnecessary in the subsequent analysis (punctuation marks, symbols or numbers, and words that do not provide meaning). This means that from a total of 7,377,533 words, 5,813,263 were preserved after the depuration; in other words, 21.20% of the words were suppressed. The preprocessing was carried out using the *stringr* R package.

Table 1
Spanish vaccination phases according to available doses.

Phase/description	Duration	Population group
Phase 0/Development, authorisation and evaluation	From February 27 till 18 December 2020 (1st update)	
Phase 1/First available doses	From 19 December 2020 till 26 February 2021 (4th update)	• Residents and staff in nursing homes and care centres for the elderly and the highly dependent • Front-line health and social personnel • Other health and social care staff • Non-institutionalised major dependents
Phase 2/More available doses	From 27 February 2021 till 11 May 2021 (7th update)	• Over 80 years • People between 70 and 79 and people with very high-risk conditions • People between 60 and 65 • People between 66 and 69 • Other health and social care workers • Workers with an essential social function • People between 50 and 59
Phase 3/Widely available vaccine	From 12th May 2021 till 31st December 2021	• People between 40 and 49 • People between 30 and 39 • People between 20 and 29 • People between 12 and 19 • People between 5 and 11 • Booster doses

Table 2
Structure of the database.

Tweet		Author	
Tweet ID	Text	Author ID	Registration date
Author ID	Hashtags	Author name	Location
Creation date	Is retweeted	Username	Description

Table 3
Filters for corpus determination.

Filter	Number of tweets
Tweets collected	300,286
Tweets containing location	188,392
Authors geolocated in Spain	28,285
Authors geolocated in Spain with indication of region	24,394
Tweets posted by authors geolocated in Spain	41,669

Stage 1.3. Geolocation of the authors

To select the tweets written by Spanish authors, the geographical location of the authors was identified, when possible, from the information contained in the location field. This was done by calling the *Nominatim* geocoding service, an Open Data project/of *OpenStreetMap*.[33] A total of 188,392 tweets were posted by authors that contained information in this field, of which Nominatim obtained a location determined by its latitude, longitude and country. It was shown that 28,285 authors were from Spain and writing in Spanish, of which 24,394 had indication of the region.

The study considered the tweets sent by these 28,285 Spanish authors. In total, there were 41,669 tweets that constituted the corpus of the study, being some of them retweets of other authors (Table 3).

Step 2: Social mood evolution

Stage 2.1: Social network analysis

The most relevant network interaction was considered to be the retweet because the number of retweets was very abundant in the corpus and the action of sharing or retweeting a text implied personal interest from the person who retweeted. Given the list of 28,285 Spanish users, all their messages that were retweets were selected, and the authors of the original message were extracted

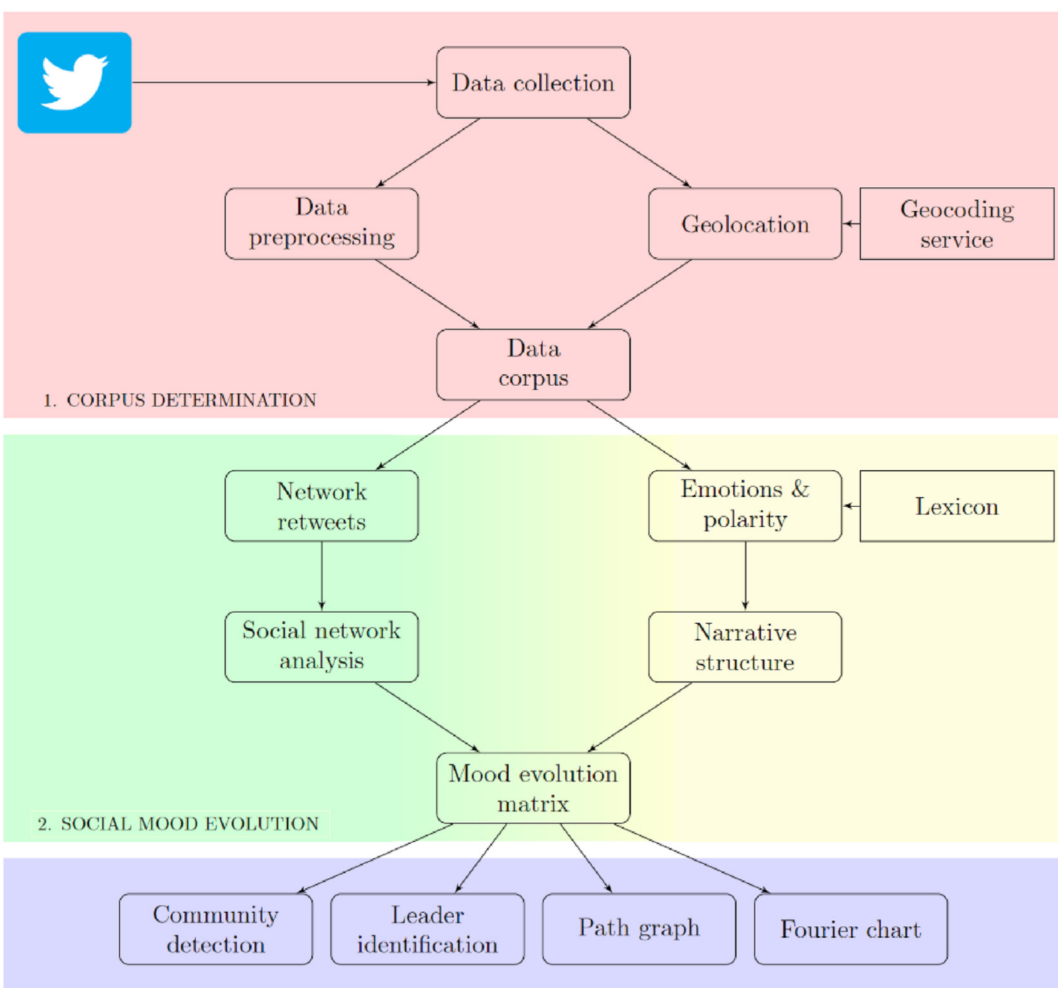

Fig. 1. Methodology flow diagram for the study of social mood evolution.

(although these may not be geolocated in Spain). A network was created based on the following methodological considerations:

- The network was a directed graph, the origin of each arc was the node corresponding to the author who retweeted a message and destination was the node that represented the author of the original tweet.
- The nodes were the users who had published tweets and retweets.
- The size of the nodes was proportional to the in-degree, representing the volume of retweets that has been made of their tweets.
- The colours of the nodes indicate communities. These communities have been calculated with the *Gephi software*,[34] which uses the algorithm described in.[35]
- The colour of the edges is the same as in the origin node, whereas their size is proportional to the number of messages from the destination node that the origin node has retweeted.
- The position of each node in the graph has been calculated using the *Force Atlas 2* algorithm,[36] an energy model for network spatialisation so that the more retweets a node has, the more focused it will be with respect to the nodes connected to it.

The resulting network contained 10,021 nodes and 17,340 edges, which represents a very low density, practically zero. Also, the average degree of the network is 1.73. This means that few retweets were made, and usually, the same authors were retweeted.

The analysis reveals the most influential users because of the size of their node (number of times a message of theirs has been retweeted) and their position within the cluster to which they belong (the more focused, the larger this size is). And the more compact a community is, the more relationships appear between its members. On the other hand, the different communities are closer to each other depending on how many nodes of each one are related to the other. The more relationships there are between two communities, the closer they would be.

Stage 2.2: Sentiment analysis

The 41,669 tweets were analysed, applying text mining by means of the *Syuzhet* 1.0.6 package[37] and *RStudio* 1.1.419, according to the general procedure already shown in Fig. 1.

As a first step, the sentiment was evaluated with *NRC Word-Emotion Association Lexicon Version 0.92*.[38–40] This lexicon is a list of English words and their associations with eight basic emotions (anger, fear, anticipation, trust, surprise, sadness, joy and disgust) and two sentiments (negative and positive). For each tweet, the valence was also obtained, that is, the difference between the number of positive and negative words, as well as the number of words associated with each of the above emotions and sentiments. We then examined how emotions were distributed throughout the text. To do this, several descriptive statistics were obtained (minimum, maximum, Q1, Q3, mean, and median) with which an overall assessment of each tweet could be achieved.

Stage 2.3: Mood evolution Matrix

After performing the social network and sentiment analysis (Stage 2.2 and Stage 2.3), the result is a matrix where the rows are the different tweets (41,669) and the columns (40) are grouped into the following information blocks:

- Tweet variables (8 columns): id, author_id, date, text, clean text, hashtag, retweeted (yes or no), retweeted_id.
- User variables (14 columns): name, username, created_at, location, description, type, lat, lng, country, city, region, postal code, cod_region, id_region.
- Emotions (8 columns): eight basic emotions (anger, fear, anticipation, trust, surprise, sadness, joy and disgust).
- Sentiments (4 columns): polarity (negative or positive), valence and number of sentiment words.
- Statistics (6 columns): six descriptive statistics (min, max, Q1, Q3, mean and median).

Results

This section presents the results corresponding to Step 3 of the methodology (Graphic Visualization). It includes illustrations of community detection, leader identification and path and Fourier graphs.

Community detection

Fig. 2 analyses the evolution of the retweet network during the phases of the process.

The most striking result is that two differentiated nuclei emerged, with very few interconnections between them, are distinguished in each phase: on the left, groups linked to the official sources of the Government and the health administrations of Spain, journalists and media (provaccine messages); on the right, accounts disseminating denialist and antivaccine messages. In Phase 0, there were 3818 users (2746 pro- and 1072 anti-vaccination); in Phase 1, 7758 users (5726 pro- and 2032 anti-vaccination); in Phase 2, 3510 users (2883 pro- and 627 anti-vaccination); and in Phase 3, 5637 users (2698 pro- and 2939 anti-vaccination). The composition and size of both pro- and anti-vaccine groups are clearly related to the variations produced in the social mood that will appear later in Fig. 3.

Leader identification

As can be seen in Table 4, there were several leaders involved in the different communities.

To better identify the leaders of the different communities, @sanidadgob corresponds to the official account of the Spanish Ministry of Health; @We_T_Resistance is an account positioned against the vaccination process; @salvadorilla (at the time Minister of Health of Spain); @rimbaudarth is an account positioned with the thesis of @We_T_Resistance; @publico_es is a media positioned in favour of the process; @Javier_CB is a very heterogeneous community with media presence but with very low activity on the network; and @daandina is a facultative working in public health. Clearly, the two most prominent leaders are the Government (1581 retweets) and the deniers (992 retweets).

Path and fourier graphs

The protocol described in Sections 3 and 4 (and Fig. 1) was applied to the 41,669 tweets. Fig. 3 shows the Fourier plot trajectory that represents emotional valence vs percentage of tweets (tweets date). From this analysis of tweets, we can see how the mental state or social mood of Spanish people has been changing through the different phases of the vaccination process (in different colours).

As shown in Fig. 3, the highest value of valence is found at Phase 1 (orange), between 4 and 6 January 2021, corresponding with the start of vaccination in Spain with Pfizer-BioNTech COVID-19 vaccine and the approval of Moderna COVID-19 (MD) vaccine by the European Medicines Agency. While the lowest value of valence is found at Phase 3 (green), between 4 and 6 August 2021, corresponding with the announcement of the need for booster doses and the debate on compulsory vaccination. On the other hand, we should note that the biggest fluctuations were produced in Phase 2

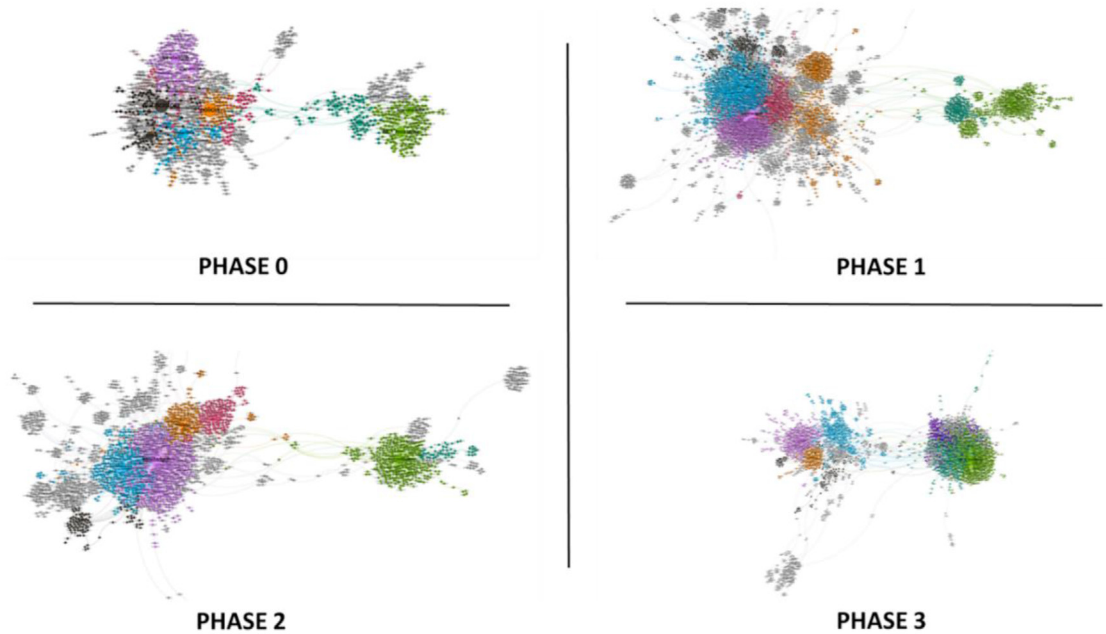

Fig. 2. Retweets network of the vaccination phases. The nodes are the users, and the arcs point goes from the retweeter to the author of the original tweet. The most retweeted authors are highlighted, and seven relatively clear clusters can be distinguished (each of them is formed by more than 2.5% of the total nodes and coloured in different colours). Within each cluster, those with highest number of retweets have been distinguished, appearing as the largest nodes in the graph.

(yellow) and Phase 3 (green) because of discordant health decisions on the Astra Zeneca vaccine.

Fig. 4 shows the percentage of words for each emotion according to each of the phases. It shows that the highest values for the main two emotions of the population during COVID-19 (fear and sadness)[41] were found at Phases 0 and 3. However, the highest value of joy and trust (more positive emotions) were shown in Phases 1 and 2, coinciding with the results obtained in Fig. 3 where the positive valences were in Phases 1 and 2.

The same pattern can be observed in Fig. 5 where we analyse the percentage of words for each phase according to each of the emotions. It is worth noting that the highest percentages of words

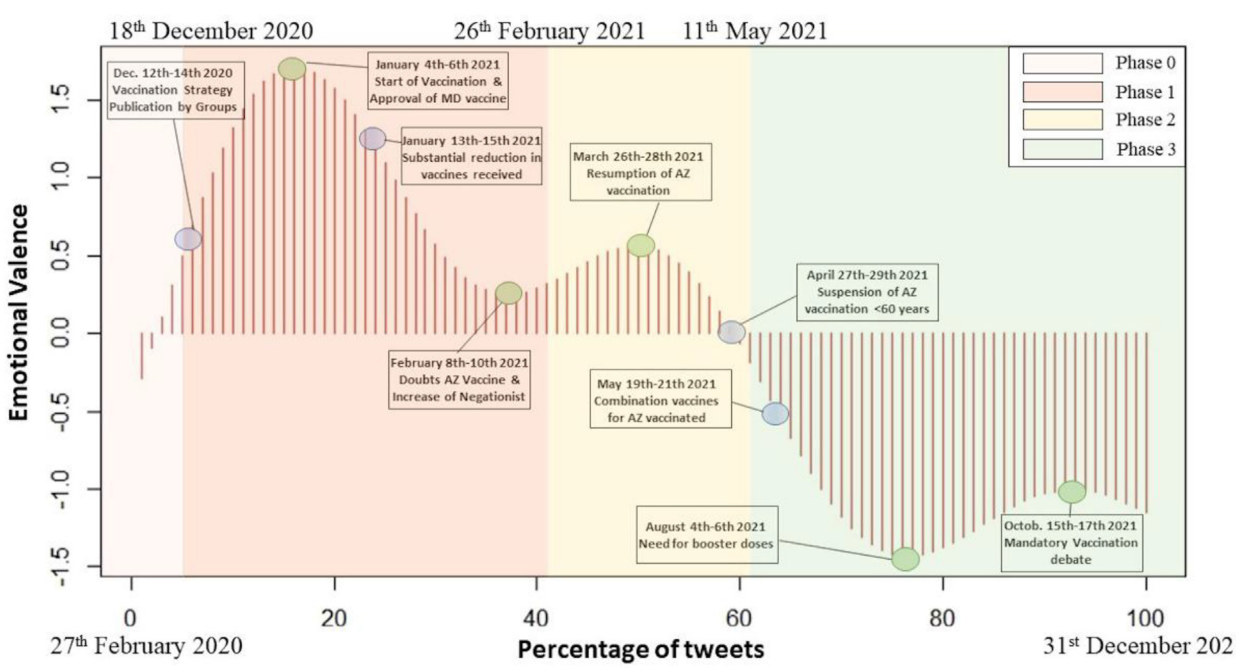

Fig. 3. Fourier plot trajectory of the tweets with the four phases (differently coloured). It represents emotional valence vs percentage of tweets (tweets date). In the upper side, the positive sentiments, and in the lower side, the negative ones. Local hotspots (green circles) and areas of trend change (purple circles) were marked by analysing the content of these tweets and relating them to relevant news and political decisions.

Table 4
Most retweeted authors in each community.

Community	Number of members (%)	Username	Number of retweets	Number of retweets (community)
Pink	1508 (15.05%)	@sanidadgob	1581	45.76%
Orange	446 (4.45%)	@daandina	42	6.87%
Black	774 (7.725)	@salvadorilla	347	17.08%
Fuchsia	426 (4.25%)	@Javier__CB	85	13.78%
Blue	1162 (11.60%)	@publico_es	158	7.52%
Green	1811 (18.07%)	@We_T_Resistance	922	23.28%
Emerald	292 (2.91%)	@rimbaudarth	171	34.76%

expressing the most negative emotions (anger, disgust, fear and sadness) are found in Phase 3, where the vaccines were widely available, but nevertheless, many doubts arose about the vaccination process with the news of the need for new doses or even compulsory vaccination. On the other hand, the most positive emotions (trust and joy) were in Phase 1, coinciding with the first available doses and the start of the vaccination process in Spain.

Discussion

This study has obtained a series of congruent results regarding the social networks involved, the evolution of social mood coupled with the dynamics of these networks, and the sentiment analysis represented in the plot trajectory. This overall congruence between the different kinds of obtained results may be interpreted as a very promising aspect of the approach.

Let us first point out that, regarding the evolution of social networks depicted in Fig. 2, the clustering dynamics during the four phases distinguished is surprisingly accurate, capturing the evolution of public opinion during the vaccination process. The analysis of the network of retweets not only shows the interconnections and clustering of the community of tweeters around interest groups but also shows how the structure of these groups varies throughout the process. It can be seen how public health decisions and other environmental circumstances that cause the changes in mood are translated not only into how tweeters are grouped but also who their referents are when it comes to sharing information. In addition, we can see in the network dynamics that clustering around two compact groups, of pro-vaccines and anti-vaccines, polarises the position of individuals in two communities with extremely few interconnections. These 'radical' divisions occur because of, and are exacerbated by, increasing conflict in communications about contentious topics such as lockdowns and compulsory vaccination.

Table 4 indicates the importance of public health communication from official sources (@sanidadgob and @salvadorilla) because their retweets from other users can reach far more people that are not following the official accounts. This means a cost-effective communication strategy for public health promotion.[42] In this regard, we may realise that most international political leaders are progressively turning to social networks to broadcast information about the pandemics, response plans, public health measures and connection with citizens.[43] This implies a series of strategic choices to use a more positive frame to influence opinion and action and to encourage compliance with public health norms and standards. The choice of positive frames may guide the national conversation away from seeking 'blame' for the pandemic towards a supportive mood necessary to implement the public health strategies required.[44] Finally, identifying and monitoring those social leaders whose opinions most closely reflect the needs or demands of society will contribute to make more realistic and effective public health decisions.

The prevalence of the different emotions during each of the phases shown in Figs. 4 and 5 would correlate well with the above. The high levels of anger, disgust, fear and sadness in phase 3 would document, as already said, the news about the new doses needed and the compulsory vaccination. The mental fatigue after the prolonged lockdowns and the stress for such long periods of uncertainty and pandemic fears are indeed reflected in the emotional arousal seen in these final phases.

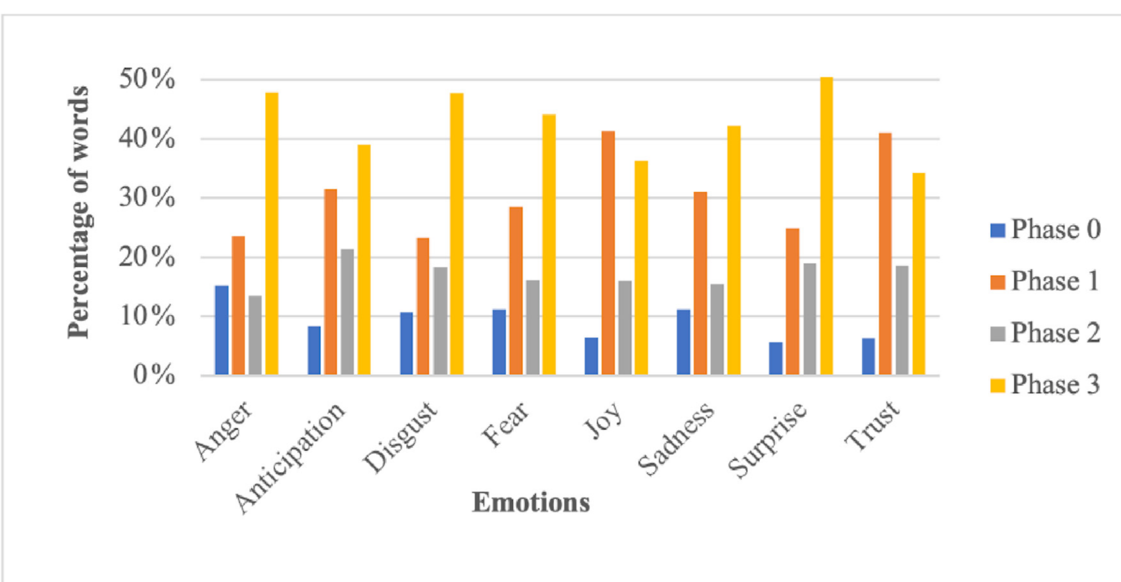

Fig. 4. Percentage of words per emotion according to each of the phases.

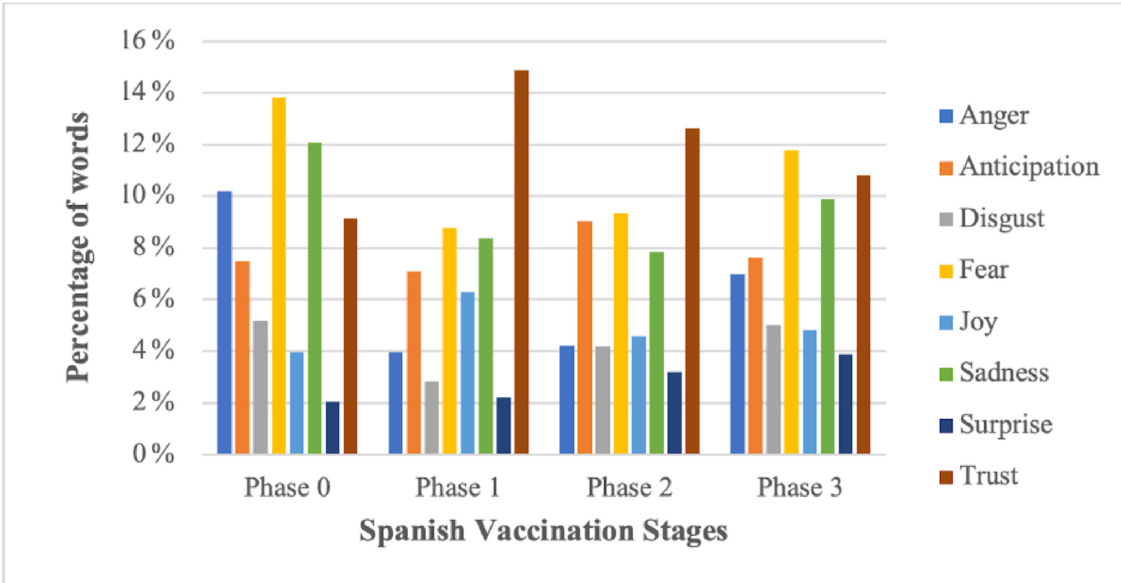

Fig. 5. Percentage of words per phase according to each of the emotions.

The specific results of sentiment analysis in the Fourier plot also show a remarkable congruence with the development of the four phases and the most notable events during the vaccination process. Although the way to obtain the valence of each tweet may look rather coarse, there is a considerable degree of theoretical sophistication in this evaluation of emotional valence. Some of the most accepted theories of emotions rely on two-dimensional spaces where valence becomes one of the fundamental dimensions.[45–49] The six basic emotions due to Paul Eckman[50] are generally maintained, although it is also generally accepted the need to enlarge these basic emotions.[51,52]

Sentiment analysis indeed offers an exciting panorama of emerging tools and paradigms to explain the emergence of social moods and emotional contagion phenomena that are so important in our societies, including the current 'epidemic of loneliness'.[53,54]

Looking at the limitations of the present approach, we have to consider the existing complementarity between the sentiment analysis technic using lexicons, as herein developed, and the machine learning and deep learning models (supervised and unsupervised).[55] Lexicon-based models are to be preferred where the data sets are small and the available computational resources limited under the condition of slightly lower performance.[56] The supervised models perform fine for the specific domain they have been trained. But this specific training becomes an important limitation for addressing different domains or brand-new topics such as the present COVID-19 pandemic. The unsupervised learning approaches do not hinge on the domain or topic of the training data, overcoming the difficulty of labelled training data collection and creation, although they need an extensive learning process and the subsequent computational resources. The hybrid technique is the combination of both lexicon and deep learning approaches. This combination improves the performance of classification, makes the detection and measurement of sentiment at the concept level and provides high accuracy results.[57]

Conclusions

The new approach developed combines machine learning techniques (sentiment analysis and data mining) with multivariate analysis methods (SNA and text mining). Free software, that is very easy to access and use, has been used to do this. We are currently working on a research project aiming at integrating all these software tools into a Decision Support System, easier to use and interpret the results.

The sentiment analysis approach has proven its validity to evaluate the social mood of citizens in different time scales, registering the different clusters that emerged, gauging public states of mind via the collective valence and detecting the prevalence of the different emotions in the successive phases of the pandemic.

The approach has also shown, albeit rather indirectly, social support for public policies. Overcoming the conceptual limitations around the study of emotions may considerably enrich the perspectives and applications of sentiment analysis and similar kinds of studies, particularly thinking in the emerging mental pathologies—and not only in viral pandemics—around the 'information society'.

Finally, the combination in formal models of objective and subjective information, in this case about the COVID-19 vaccination process in Spain, will provide a more accurate vision of social reality, which will enable a more effective resolution of problems.

Author statements

Ethical approval

This work did not need to be approved by an ethics committee, as we used public information and messages from the social network Twitter.

Funding

This research is partially funded by the 'Grupo Decisión Multicriterio Zaragoza' research group (S35-20R) and the project 'Participación Ciudadana Cognitiva y Decisiones Públicas. Aplicaciones Sociosanitarias' (Ref. LMP35_21), both supported by grants from the Regional Government of Aragon and FEDER funds. The funders had no role in the design of the study; in the writing of the article, or in the decision to publish the results.

Competing interests

None declared.

Authors' contributions

A.T. contributed to conceptualisation, methodology, software, data curation, formal analysis, and writing, reviewing and editing the article. A.A. contributed to formal analysis and reviewing and editing the article. J.M.M.-J. contributed to conceptualisation, methodology, reviewing and editing, and funding acquisition. J.N. contributed to conceptualisation, methodology, software, data curation, formal analysis, and writing, reviewing and editing the article.

References

1. Sattar NS, Arifuzzaman S. COVID-19 vaccination awareness and aftermath: public sentiment analysis on twitter data and vaccinated population prediction in the USA. *Appl Sci* 2021 Jun 30;**11**(13):6128.
2. Yousefinaghani S, Dara R, Mubareka S, Papadopoulos A, Sharif S. An analysis of COVID-19 vaccine sentiments and opinions on Twitter. *Int J Infect Dis* 2021 Jul;**108**:256–62.
3. Kapoor KK, Tamilmani K, Rana NP, Patil P, Dwivedi YK, Nerur S. Advances in social media research: past, present and future. *Inf Syst Front* 2018 Jun 1;**20**(3):531–58.
4. Öztürk N, Ayvaz S. Sentiment analysis on Twitter: a text mining approach to the Syrian refugee crisis. *Telematics Inf* 2018 Apr;**35**(1):136–47.
5. https://twitter.com/; 2022.
6. https://www.statista.com/statistics/303681/twitter-users-worldwide/; 2022.
7. https://www.internetlivestats.com/twitter-statistics/; 2022.
8. Chandra Pandey A, Singh Rajpoot D, Saraswat M. Twitter sentiment analysis using hybrid cuckoo search method. *Inf Process Manag* 2017 Jul;**53**(4):764–79.
9. Zimbra D, Abbasi A, Zeng D, Chen H. The state-of-the-art in twitter sentiment analysis. *ACM Trans Manag Inf Syst* 2018 Sep. 5;**9**(2):1–29.
10. Naseem U, Razzak I, Musial K, Imran M. Transformer based deep intelligent contextual embedding for twitter sentiment analysis. *Future Generat Comput Syst* 2020 Dec;**113**:58–69.
11. Mendon S, Dutta P, Behl A, Lessmann S. A hybrid approach of machine learning and lexicons to sentiment analysis: enhanced insights from twitter data of natural disasters. *Inf Syst Front* 2021 Sep. 14;**23**(5):1145–68.
12. Giachanou A, Crestani F. Like it or not. *ACM Comput Surv* 2016 Nov 11;**49**(2):1–41.
13. Antonakaki D, Fragopoulou P, Ioannidis S. A survey of Twitter research: data model, graph structure, sentiment analysis and attacks. *Expert Syst Appl* 2021 Feb;**164**:114006.
14. Hasan MdR, Maliha M, Arifuzzaman M. Sentiment analysis with NLP on twitter data. In: *2019 international conference on computer, communication, chemical, materials and electronic engineering (IC4ME2)*. IEEE; 2019. p. 1–4.
15. Zahra K, Imran M, Ostermann FO. Automatic identification of eyewitness messages on twitter during disasters. *Inf Process Manag* 2020 Jan;**57**(1):102107.
16. Agarwal A, Singh R, Toshniwal D. Geospatial sentiment analysis using twitter data for UK-EU referendum. *J Inf Optim Sci* 2018 Jan 2;**39**(1):303–17.
17. Ilyas SHW, Soomro ZT, Anwar A, Shahzad H, Yaqub U. Analyzing Brexit's impact using sentiment analysis and topic modeling on Twitter discussion. In: *The 21st annual international conference on digital government research*. New York, NY, USA: ACM; 2020. p. 1–6.
18. Somula R, Dinesh Kumar K, Aravindharamanan S, Govinda K. *Twitter sentiment analysis based on US presidential election 2016*. 2020. p. 363–73.
19. Chaudhry HN, Javed Y, Kulsoom F, Mehmood Z, Khan ZI, Shoaib U, et al. Sentiment analysis of before and after elections: twitter data of U.S. Election 2020. *Electronics (Basel)* 2021 Aug 27;**10**(17):2082.
20. Budiharto W, Meiliana M. Prediction and analysis of Indonesia Presidential election from Twitter using sentiment analysis. *J Big Data* 2018;**5**(1):51. C.
21. Sharma A, Ghose U. Sentimental analysis of twitter data with respect to general elections in India. *Proc Comput Sci* 2020;**173**:325–34.
22. Patel R, Passi K. Sentiment analysis on twitter data of world cup soccer tournament using machine learning. *IoT* 2020 Oct 10;**1**(2):218–39.
23. Asif M, Ishtiaq A, Ahmad H, Aljuaid H, Shah J. Sentiment analysis of extremism in social media from textual information. *Telematics Inf* 2020 May;**48**:101345.
24. Kasson E, Singh AK, Huang M, Wu D, Cavazos-Rehg P. Using a mixed methods approach to identify public perception of vaping risks and overall health outcomes on Twitter during the 2019 EVALI outbreak. *Int J Med Inform* 2021 Nov;**155**:104574.
25. Garcia K, Berton L. Topic detection and sentiment analysis in Twitter content related to COVID-19 from Brazil and the USA. *Appl Soft Comput* 2021 Mar;**101**:107057.
26. Manguri K H, Ramadhan R N, Mohammed Amin R, Twitter P. Sentiment analysis on worldwide COVID-19 outbreaks. *Kurd J Appl Res* 2020 May 19:54–65.
27. Turón A, Aguaron J, Escobar MT, Moreno-Jiménez JM. Decision analysis in e-cognocracy using dynamic social networks. In: *ICDSST2020 conference*. Zaragoza; 2020.
28. Scott J. Social network analysis. *Sociology* 1988 Feb 2;**22**(1):109–27.
29. West R, Paskov HS, Leskovec J, Potts C. Exploiting social network structure for person-to-person sentiment analysis. *Trans Assoc Comput Linguist* 2014 Dec;**2**:297–310.
30. https://www.sanidad.gob.es/profesionales/saludPublica/prevPromocion/vacunaciones/covid19/docs/COVID-19_EstrategiaVacunacion.pdf; 2022.
31. https://www.sanidad.gob.es/profesionales/saludPublica/ccayes/alertasActual/nCov/vacunaCovid19.htm; 2022.
32. https://www.vacunacovid.gob.es/; 2022.
33. https://www.openstreetmap.org; 2022.
34. http://gephi.org; 2022.
35. Blondel VD, Guillaume JL, Lambiotte R, Lefebvre E. Fast unfolding of communities in large networks. *J Statist Mech Theory Exp [Internet]* 2008;**2008 Oct**(10):P10008. https://doi.org/10.1088/1742-5468/2008/10/p10008.
36. Jacomy M, Venturini T, Heymann S, Bastian M. ForceAtlas2, a continuous graph layout algorithm for handy network visualization designed for the Gephi software. *PLoS One* 2014 Jun 10;**9**(6):e98679.
37. Jockers M. *Syuzhet: extracts sentiment and sentiment derived plot arcs from text, R package version 1.0-6*. https://cran.r-project.org/web/packages/syuzhet/syuzhet.pdf; 2020.
38. Bravo-Marquez F, Frank E, Mohammad SM, Pfahringer B. Determining word-emotion associations from tweets by multi-label classification. In: *2016 IEEE/WIC/ACM international conference on Web intelligence (WI)*. IEEE; 2016. p. 536–9.
39. Mohammad SM, Turney PD. Crowdsourcing a word-emotion association lexicon. *Comput Intell* 2013 Aug;**29**(3):436–65.
40. Mohammad SM. Sentiment analysis. In: *Emotion measurement*. Elsevier; 2021. p. 323–79.
41. de las Heras-Pedrosa C, Sánchez-Núñez P, Peláez JI. Sentiment analysis and emotion understanding during the COVID-19 pandemic in Spain and its impact on digital ecosystems. *Int J Environ Res Publ Health* 2020 Jul 31;**17**(15):5542.
42. Liang H, Fung ICH, Tse ZTH, Yin J, Chan CH, Pechta LE, et al. How did Ebola information spread on twitter: broadcasting or viral spreading? *BMC Publ Health* 2019 Dec 25;**19**(1):438.
43. Rufai SR, Bunce C. World leaders' usage of Twitter in response to the COVID-19 pandemic: a content analysis. *J Public Health* 2020 Aug 18;**42**(3):510–6.
44. Wang Y, Croucher SM, Pearson E. National leaders' usage of twitter in response to COVID-19: a sentiment analysis. *Front Commun (Lausanne)* 2021 Sep. 8:6.
45. Handel S. *Classification of emotions*. http://www.theemotionmachine.com/classification-of-emotions/; 2012.
46. SeshathriAathithyan S, Sriram Mv, Prasanna S, Venkatesan R. Affective — hierarchical classification of text — an approach using NLP toolkit. In: *2016 international conference on circuit, power and computing technologies (ICCPCT)*. IEEE; 2016. p. 1–6.
47. Russell JA, Barrett LF. Core affect, prototypical emotional episodes, and other things called emotion: dissecting the elephant. *J Pers Soc Psychol* 1999;**76**(5):805–19.
48. Russell JA. A circumplex model of affect. *J Pers Soc Psychol* 1980;**39**(6):1161–78.
49. Barrett LF. Solving the emotion paradox: categorization and the experience of emotion. *Pers Soc Psychol Rev* 2006 Feb 21;**10**(1):20–46.
50. Ekman P. An argument for basic emotions. *Cognit Emot* 1992 May 7;**6**(3–4):169–200.
51. Ekman P. Basic emotions. In: Dalgleish T, Power M, editors. *Handbook of cognition and emotion*. Sussex (UK): John Wiley & Sons; 1999. p. 45–60.
52. Navarro J, Marijuán PC. The natural, artificial, and social domains of intelligence: a triune approach. *Proc West Mark Ed Assoc Conf* 2022. In press.
53. Bernal AI. Audiencias y temas de noticias en medios online y Twitter. El caso de elpais.com. In: *V congreso internacional de Ciberperiodismo y Web 20*. Audiencias Activas y Periodismo.; 2013.
54. Navarro J, Turón A, Altuzarra A, Lahoz-Beltra R. Comparative sentiment analysis of COVID-19: a machine learning approach. In: *7th ICDSST 2021 (international conference on decision support System technology)*. Loughborough (Reino Unido); 2021.
55. Verma B, Thakur RS. *Sentiment analysis using lexicon and machine learning-based approaches: a survey*. 2018. p. 441–7.
56. Catelli R, Pelosi S, Esposito M. Lexicon-based vs. Bert-based sentiment analysis: a comparative study in Italian. *Electronics (Basel)* 2022 Jan 26;**11**(3):374.
57. D'Andrea A, Ferri F, Grifoni P, Guzzo T. Approaches, tools and applications for sentiment analysis implementation. *Int J Comput Appl* 2015;**125**(3).