Heliyon 10 (2024) e32572

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

Analyzing digital societal interactions and sentiment classification in Twitter (X) during critical events in Chile

Pablo A. Henríquez [a,*], Francisco Alessandri [b]

[a] *Facultad de Administración y Economía, Universidad Diego Portales, Santiago, Chile*
[b] *London School of Economics and Political Science, United Kingdom*

ARTICLE INFO

Dataset link: https://github.com/pabhenriquez/Indicadores

Keywords:
Sentiment classification
Twitter
Engagement
Inclusivity

ABSTRACT

This study explores the influence of social media content on societal attitudes and actions during critical events, with a special focus on occurrences in Chile, such as the COVID-19 pandemic, the 2019 protests, and the wildfires in 2017 and 2023. By leveraging a novel tweet dataset, this study introduces new metrics for assessing sentiment, inclusivity, engagement, and impact, thereby providing a comprehensive framework for analyzing social media dynamics. The methodology employed enhances sentiment classification through the use of a Deep Random Vector Functional Link (D-RVFL) neural network, which demonstrates superior performance over traditional models such as Support Vector Machines (SVM), naive Bayes, and back propagation (BP) neural networks, achieving an overall average accuracy of 78.30% (0.17). This advancement is attributed to deep learning techniques with direct input–output connections that facilitate faster and more precise sentiment classification. This analysis differentiates the roles of influencers, press radio, and television handlers during crises, revealing how various social media actors affect information dissemination and audience engagement. By dissecting online behaviors and classifying sentiments using the RVFL network, this study sheds light on the effects of the digital landscape on societal attitudes and actions during emergencies. These findings underscore the importance of understanding the nuances of social media engagement to develop more effective crisis communication strategies.

1. Introduction

Over the past few years, there has been a significant surge in the creation of text-based content on social media platforms. For example, as of July 2018, Twitter (X) boasted 326 million active users who collectively sent over 500 million tweets daily.[1] Social media fosters virtual connections among users, facilitating the expression of views and the formation of ties via posts, comments, messages, and likes. It offers individuals a platform to instantly and effortlessly convey emotions, ideas, and viewpoints [1–3].

Online social platforms such as Facebook, Instagram, and Twitter (X) provide millions of individuals with unlimited access to information and connectivity [4]. The content created on such platforms has been proven to have a significant impact on society as a whole. From social and political discussions [5,6] and emergency and disaster responses [7,8], social media conversation affects the offline, physical world in tangible ways. This tendency, combined with the rapidly spreading nature of virtual content, has transformed online opinions into valuable assets.

* Corresponding author.
E-mail address: pablo.henriquez@udp.cl (P.A. Henríquez).

[1] https://www.statista.com/statistics/282087/number-of-monthly-active-twitter-users/.

https://doi.org/10.1016/j.heliyon.2024.e32572
Received 21 February 2024; Received in revised form 4 June 2024; Accepted 5 June 2024

Available online 11 June 2024

The prevalence of social crises in modern times presents a significant and challenging issue that cannot be promptly resolved. These crises, which pose threats to societal well-being, have garnered the attention of both government authorities and community members alike [9]. Recognizing and understanding situations that endanger social life are crucial when considering a social crisis. Early identification allows for the mitigation of adverse effects and resolution of this issue. In times of crisis, access to reliable information is critical for effective problem-solving. While many organizations leverage the capabilities of social media, only a few specifically focus on aiding individuals during social crises [10]. Platforms such as Twitter have become essential for users, not only as a means for academics to gather data through tweets but also as a tool for monitoring societal crises [11]. Twitter serves as a rich source of data for analyzing topics and detecting crises. The initial act of a societal event is often observed first on social media platforms [12]. People may refrain from being direct about things in real life; however, extensive discussions, blog posts, and short messages proliferate on the Internet via social media. Nonetheless, it is important to note that such online content can range from genuine to fraudulent, potentially fueling rumors on social media platforms [13].

Building on this understanding of social media's role in crisis communication, it is apparent that influencers play a pivotal role in shaping narratives during such times. At the peak of a crisis, researchers have found that, through qualitative analysis of Twitter communication, valuable insights can be gleaned regarding people's perceptions, the nature of the critical event, and its level of visibility [14]. With their vast reach and trusted authority, influencers have been found to significantly expedited the propagation process and augmented the scale of information dissemination. Their ability to connect with a large audience and the trust they command owing to their authenticity and position [15] further highlight social media's transformative power. By leveraging their influence, they can either amplify the dissemination of critical and reliable information during crises or, conversely, contribute to the spread of rumors and misinformation.

Sentiment Analysis (SA) is a growing task [16], aims to classify opinions and sentiments expressed in text, particularly within the realm of social media analysis. Known also as opinion mining, this application seeks to determine the orientation of opinions—positive, negative, or neutral—expressed in text streams. Building a classification model that can accurately perform sentiment classification requires an analysis of tweets with both positive and negative sentiments. It is noteworthy that the training time for such a sentiment classification model increases with an increasing volume of training tweets [17].

Artificial Neural Networks (ANN) are popular in Machine Learning (ML) for tasks such as classification and regression. However, the detailed process of adjusting their settings can slow their learning [18]. The popular way to train an ANN is through backpropagation (BP). To solve this problem [19], a single hidden layer neural network with random weights (RWSLFN) has been proposed, with random weight assignment between the input and hidden layers and least-squares estimation of the output weights as the training method. A functional link neural network known as the Random Vector Functional Link (RVFL) network (RVFLN) has been proposed [20]. Although it is similar to the RWSLFN reported in [19], it has direct input–output connections between the input and output neurons. Extreme Learning Machine (ELM) [21] is an RVFL variant that omits direct links and adopts the primitive structure of single hidden layer feedforward neural networks (SLFN). ELM is essentially a variation of the RVFL network [22,23]. Various studies, including those by [24–26,22,23,27], have highlighted the efficacy of RVFL networks, particularly emphasizing the crucial role of the direct input—output link in enhancing network performance.

Recently, a Deep RVFL model [28,27], RVFL learning with privileged information [29], and variance-embedded RVFL [30] have been proposed to enhance the generalization of the RVFL based models. The variations, improvements, and applications of the RVFL models are discussed in [31]. The work presented in [27] is one of the most groundbreaking attempts to propose multilayer RVFL networks. The authors introduced a novel deep neural network using RVFL networks called Deep RVFL (D-RVFL). They evaluated the performance of the D-RVFL on human sentiments using Twitter. In Table 1 we present a review of the most commonly used applications using RVFL models. From this review, we observe that most studies using RVFL networks do not focus on sentiment classification.

Understanding which information (content and type) resonates most significantly with the audience can serve as a pivotal strategy for enhancing engagement [32]. The integration of an RVFL network into SA underscores the broader implications of understanding the specific content types with which the audience resonates, thereby enhancing engagement and fostering sustained dialogue with the public. This strategy is crucial for calming people, establishing trust, and navigating crises. Consequently, this study examines the role of content creators and social media platforms in shaping societal attitudes and actions during natural disasters and social uprisings by assessing their engagement and impact in such situations. This comprehensive examination highlights the critical role of SA in crisis communication and management. The major highlights of this study are as follows:

- We obtained a set of tweets concerning several significant events in Chile, including the COVID-19 pandemic in 2020, the social uprising in October 2019, and the wildfires that transpired in January 2017 and February 2023.
- We introduced new indicators to measure sentiment, inclusivity, engagement, and societal impact of Spanish social media content, offering a comprehensive toolkit to analyze this content.
- This study outlines how influencers, the press, radio stations, and television uniquely affect information sharing during emergencies. This demonstrates the distinct ways in which each group influences the flow of information.
- We used the Deep Random Vector Functional Link (D-RVFL) to improve the accuracy of tweet sentiment classification.
- We perform a comparative analysis of different classifiers to establish the superiority of the RVFL and deep random vector functional link (D-RVFL) networks over traditional models in social media sentiment classification.

Table 1
Application of RVFL Models.

Year	Literature	Application
2023	[33]	Direction-of-arrival (DOA)
2023	[34]	ORL database
2023	[35]	UCI datasets
2023	[36]	Solar exposure forecasting
2023	[37]	Diagnosis of enlarged lymph nodes
2023	[38]	Leaf disease detection
2023	[39]	Wind power forecasting
2022	[40]	Alzheimer's disease diagnosis
2022	[41]	UCI datasets
2021	[42]	Nonlinear system identification
2021	[43]	Stock price prediction
2021	[44]	Assessing dry weight of hemodialysis patients
2021	[45]	forecasting of oil production in China
2021	[46]	Time series forecasting
2021	[47]	Short term solar power forecasting
2020	[48]	COVID-19 cases forecasting
2020	[49]	Nonlinear system identification
2020	[50]	Brain MRI image classification problem
2020	[51]	Respiratory motion prediction
2020	[52]	Forecasting seawater power consumption and productivity
2019	[53]	Water quality analysis
2019	[54]	Detecting Brain abnormalities
2019	[55]	UCI datasets
2019	[56]	Forecasting crude oil price
2018	[27]	Twitter sentiment classification
2018	[57]	Indian summer monsoon rainfall prediction

The rest of the article is structured as follows. The Related Work section presents a literary review of related work. Methodology section discusses the research methodology. Results section provides the experimental results, and the Discussion and Conclusion sections conclude the article.

2. Related works

Social media has emerged as an important medium for governments and citizens to understand and explain crisis situations, make public decisions, and act accordingly [9]. Government agencies in various countries use special media to communicate and manage crises [58]. Studies have shown that public health organizations effectively engage the public on social media during past crises, with platforms such as Instagram showing higher levels of interaction, indicating their potential for strategic health risk communication [59]. How economic crises affect Generation Z's engagement with a city's social media in the past has been investigated [60]. This has been related to the city's image, brand personality, and residents' satisfaction, revealing that crises altered past engagement patterns. In [61], it was found that during the COVID-19 crisis, Chinese government agencies' use of social media for citizen engagement was complex; media richness tended to reduce engagement, whereas interactive communication increased it. Crisis news content and government actions enhanced engagement, with the impact varying depending on the emotional tone of the posts. In 2020, radiology residency programs expanded the use of social media to engage students and promote diversity, equity, and inclusion during the NRMP application cycle. They utilized platforms such as Twitter, Instagram, and Facebook with a consistent approach to expand outreach [62].

Building on these insights, a recent study analyzed Tweets from leaders and healthcare organizations in countries with high COVID-19 resilience. This analysis revealed that the UAE Prime Minister and the Canada's Public Health Agency had the most significant societal associations online [63]. These leaders showed an acute awareness of individual factors on social media, aligning with user preferences for these platforms during health crises. Furthermore, [58] delved into the coproduction of disaster risk communications between the government and citizens during Hurricane Sandy, analyzing over 132,922 #sandy tweets. Networked citizen interactions on social media significantly amplified the government's communication and agility. This case study underlines the critical role of social media policy governance networks in enabling effective public service coproduction and emphasizes the importance of social media in crisis communication and management.

A significant amount of research has been conducted on the analysis of tweets about natural disasters. For example, [64] provided a network analysis of official Twitter accounts activated during the Charleston, West Virginia water contamination crisis in 2014. Another study [8] analyzed the activity of the 2010 earthquake in Chile. Furthermore, the authors conducted a preliminary study of certain social phenomena, such as the circulation of false rumors and confirmed news. A pivotal study [65] proposed significant advancements in crisis-related information detection on Twitter, with an absolute improvement of 16.55 points in identifying crisis events and an enhancement of 21.71 points in classifying different types of crisis information. In [66], a comprehensive literature review was conducted to identify social media users during a disaster. This framework can be used to facilitate the development of disaster social media tools, formulation of disaster social media implementation processes, and scientific study of disaster social media effects. Another study examined tsunami warnings in Padang, Indonesia and the reactions of Twitter users [67]. In [68], the shifting perceptions of international students during the COVID-19 crisis were explored by analyzing 6,501 Twitter posts from January to April 2020. The study revealed a transition from initial stereotypes and discrimination to empathy and support after universities closed, including being labeled as disease carriers. Another study [69], investigated misinformation on Facebook and Twitter during the initial months of the Russia–Ukraine conflict, using large datasets to track the spread of Russian propaganda and unreliable content. They revealed the pivotal role of superspreaders and reported that only 8–15% of such posts were moderated. Additionally, the study found a right-leaning bias among sharers of this content, emphasizing platform vulnerabilities and the critical need for improved moderation to safeguard online discourse integrity.

A study has examined Twitter use during and after Typhoon Haiyan devastated the Philippines [70]. The authors explored the usage time, geographic location, type of stakeholders (e.g., ordinary citizens and journalists), and social media engagement to forecast these uses. They demonstrated that different stakeholders used social media mostly for the dissemination of second-hand information, to help coordinate relief efforts, and to pay tribute to those affected.

A study did a Twitter investigation of the real-time interaction of events, such as earthquakes, and proposed an algorithm to monitor tweets and detect a target event [7]. The researchers devised a tweet classifier based on features such as keywords in a tweet, the number of words, and the context thereof. In [71], a SA approach applied to a set of tweets related to a natural disaster in Italy (a study on the 2014 Genoa Flash Floods) was presented. They identified tweets that may provide useful information from a disaster management perspective. In [72], was utilized text mining and SA of Twitter data to identify recent social crises by comparing the findings with those of reputable newspapers. Employing a hybrid method, they achieved high identification rates for the top crises between February 27 and March 11, 2020. This study demonstrates the potential of social media analytics in detecting and understanding social issues, offering a valuable tool for effectively addressing them.

2.1. A BERT framework to SA of tweets

Recent developments in Sentiment Analysis (SA) have witnessed a substantial move towards more advanced techniques. This shift goes beyond the traditional approaches that relied on lexicons and machine learning (ML). Instead, it embraces the methodologies of deep learning. A seminal study [73] compared various SA techniques, including BERT, and found that it outperformed the others, especially when analyzing IMDB reviews and social media content. This study highlighted the adaptability of SA across various domains like marketing, politics, economics, and healthcare. However, it also identified a lack of application in areas like emergency response.

At the same time, another study [74] reinforced the effectiveness of machine learning (ML) and ensemble learning techniques in sentiment analysis (SA). This study brought to light popular and successful mechanisms within the field. While it offered valuable contributions to our understanding of current practices, a more in-depth comparison between the methodologies was not provided.

The growing appreciation for BERT, a state-of-the-art deep learning framework, has become increasingly evident in SA of tweets. A prior study [75] suggested a groundbreaking improvement to BERT by incorporating emotion-cognitive reasoning, with the goal of enhancing sentiment classification in tweets related to emergencies. While this approach shows promise, its reliance on basic emotional rules and lexicons necessitates further development.

An initial investigation conducted by [76] explored the influence of preprocessing techniques on BERT's effectiveness for analyzing tweets in both English and Italian. This study illuminates the ideal preparation methods for informal Twitter text to optimize BERT's performance. In a similar vein, another study by [77] demonstrated the harmonious interaction between BERT's contextual embeddings and deep learning classifiers, such as CNN and Bi-LSTM, for sentiment analysis across various tweet datasets. These studies reinforce BERT's ability to handle the informal and contextual nature of tweets, suggesting that exploring other transformer models could be a promising direction for future advancements.

A study by [78] underscored the difficulty posed by the scarcity of annotated datasets for training powerful deep learning models. This study utilized transfer learning techniques, including BERT, to analyze public sentiment on Twitter regarding HPV vaccines. The results showcased BERT's superiority in performance. Further emphasizing the significance of domain-specific knowledge, a study by [79] demonstrated that integrating external sentiment knowledge from the domain into BERT can mitigate the limitations of training data, especially for aspect-based sentiment analysis.

Research conducted on Arabic sentiment analysis [80] has highlighted the need for advancements in language-specific methodologies and datasets. This research emphasizes the importance of detecting sarcasm within Arabic natural language processing (NLP). Similarly, a study on Vietnamese SA [81] advocates for the use of PhoBERT and acknowledges challenges like limitations on input length. Finally, a study by [82] demonstrated that BERT and GPT-3 outperform traditional SA methods on standard datasets. Their findings revealed GPT-3's superior performance when analyzing COP9 tweets, reinforcing the effectiveness of pre-trained models for sentiment analysis with limited data annotations.

3. Methodology

3.1. Data collection

We collected datasets using the academic API[2] provided by Twitter to collect information on non-protected users. In the data collection stage, we employed a rigorous process of curating keywords. These keywords were drawn from official documentation,

[2] https://developer.twitter.com/en/products/twitterapi/academic-research. We systematically searched for and extracted relevant tweets from Twitter using the academictwitteR package, which is developed in the R programming language [83]. However, it should be noted that this academic API is no longer accessible.

Table 2
Duration of the event.

Event	Duration
Wildfire 2017	1 month
Wildfire 2023	2 months
COVID-19	9 months
Social Uprising	14.5 months

media narratives, and preliminary analyses of social media activity. This thorough approach ensured the dataset's relevance and richness. It allowed us to align the keywords with the subtle contextual variations of each chosen event, ultimately capturing the most representative data for our analysis.

Four different critical events in Chile were chosen for this study:

- Chilean forest fires of 2017: The dataset contains a collection of 195,044 tweets. The keywords such as #bomberosdechile, #Chileenllamas, #EstadoDeCatastrofe were systematically selected to capture the broad societal impact and public response over the event duration from 12/01/2017 to 01/02/2017.
- Chilean forest fires of 2023: A total of 415,791 tweets were captured from 01/01/2023 to 02/28/2023. To ensure focused and relevant data extraction, the exclusive keyword #IncendiosForestales was used.
- Chilean Social Protests of 2019: This dataset includes 15,892,881 tweets collected from 10/10/2019 to 01/31/2020. To capture the complex and varied facets of the protests, multiple keywords were used including #EstallidoSocial, #PlazaDeLaDignidad, #PrimeraLinea, and #Carabineros, among others. These keywords were selected to accurately reflect the diverse opinions, significant moments, and key actors involved in the unfolding social dynamics during the protests.
- COVID-19 pandemic: The dataset comprises 2,118,000 tweets, gathered through the keywords #COVID2019chile and #CoronaVirusEnChile, reflecting the public discourse from the onset of the pandemic restrictions in Chile to their gradual easing.

Our selection of these diverse events, including the forest fires of 2017 and 2023, the 2019 protests, and the COVID-19 pandemic, was driven by two main goals. First, we aimed to investigate the far-reaching effects of various crisis types on Chilean society. Second, we sought to capitalize on the substantial amount of social media engagement surrounding these events to conduct robust sentiment analysis. Given the profound societal impact of these events, they offer a rich and conducive environment for analyzing online social interactions across diverse crisis scenarios.

The timeframe for each dataset was meticulously chosen to coincide precisely with the duration of the event itself. This ensured we captured the societal sentiment as it unfolded, from the initial stages of each crisis to its eventual resolution. This strategic approach to timing guaranteed our analysis comprehensively covered the trajectory of public opinion and societal responses throughout the events.

Following the meticulous selection of keywords and timeframes, the tweets underwent preprocessing to eliminate noise and extraneous information. This preprocessing involved removing elements like URLs, punctuation marks, symbols, numbers, emojis, non-Spanish and non-ASCII characters, hashtags, and user mentions.

3.2. Societal interactions

Building on the collected data, this study analyzes four significant events that profoundly impacted Chilean society, encompassing both natural disasters and social uprisings. Specifically, we focus on the COVID-19 pandemic in 2020, the social uprising in October 2019, and the wildfires that occurred in January 2017 and February 2023. These events were of national importance, dominating discussions in both the media and social networks over varying periods, as illustrated in Table 2.

After preprocessing the tweets, our focus shifted towards identifying key communicators for each event. We compared the accounts of prominent radio stations, television channels, press agencies, and online news outlets with those of a select group of influencers. The influencers for each event were identified based on a specific strategy that revolved around their engagement and reach during the event. This strategy entailed the following steps:

- Throughout the event duration, we specifically selected hash tags that achieved a Trending Topic status in Chile.
- The selected hashtags were subsequently refined to include only those tags directly associated with the event. Subsequently, all tweets containing these specific hashtags within the designated timeframe were downloaded.
- Tweets were grouped by account and non-institutional accounts with the highest total retweets were identified as influencers.

Further details regarding the tweets authored by each account, including the influencers, can be found in Table 3. We analyzed the results of these accounts using three parameters: account engagement and impact, tweet sentiment, and tweet diversity and inclusivity. Fig. 1 depicts the overall research framework. Finally, the average of the normalized indicators was used to observe the accounts' societal impact, both before and after the events.

The engagement and impact metrics (as defined in Eq. (1)) were derived from two key factors: the influence of the social network account and the level of engagement achieved by each individual tweet posted during the event. This dual approach allows us to factor in both account outreach and the viral nature of specific tweets [63]:

Table 3
Distribution of tweets made from selected accounts.

Account type	Account name	Wild fire 2017 total tweets	Wildfire 2023 total tweets	COVID-19 total tweets	Social Uprising total tweets
press and radio	adnradiochile	4044	7125	33560	48974
	biobio	4929	9444	35715	55940
	Cooperativa	10919	16625	84270	139520
	eldesconcierto	1511	3471	18771	28543
	elmostrador	2103	3916	21352	32095
	InterferenciaCL		263	1143	1849
	latercera	9804	10307	55291	85348
	PublimetroChile	6722	3153	25971	45249
	Tele13_Radio	4087	6780	54290	75779
	theclinicccl	1749	8898	28332	45592
television	24HorasTVN	7090	10429	51034	78740
	CHVNoticias	7	6210	22672	32343
	meganoticiascl	2438	7681	35816	57856
	T13	5443	7800	36244	58494
influencers	ablanch4	3307			
	alegriagonzaa	4848			
	bomberoschillan	400			
	CEBioBio	1378			
	InfoNuble	1232			
	INFORMADORCHILE	1712			
	PiensaPrensa	1834			19812
	reddeemergencia	4317			
	S_Schwartzmann	2418			
	ayala_rodolfo		667		
	Cachoescalona1		511		
	camilaemiliasv		3021		
	CapuchaCreativa		823		
	Chileno17039890		1417		
	JackoProu		1518		
	MrRangerR1		924		
	rsumen		1958		
	SpreenDMC		240		
	csantander23			245	427
	El_Ciudadano			19684	30413
	FelipeParadaM			2334	5629
	FrancoBassoSotz			2382	
	Izkia			3820	
	Pa__tty			9435	
	PolarBearby			9749	
	RockandRolec			6943	
	UPLaRadio			18997	
	Vitalicio7020			24924	
	andres20ad				3806
	Chileokulto				11699
	GAMBA_CL				5693
	hernan_sr				24078
	JoviNomas				32625
	vagoilustrado				6174
	Total	82,292	113,181	602,974	926,678

$$impact = \frac{tweets \times \log(\sqrt{\frac{followers}{following+1}} + 1)}{antiquity^2} \quad (1)$$

Thus, the impact is increased by the number of tweets and followers of the users. However, the impact is diminished by the account's age, which is measured by the number of days between the first of January of the ongoing year and the date on which the account was created, because older accounts are expected to have more tweets and followers and the number of accounts following.

In our analysis, interactions represent measurable engagement metrics exhibited by Twitter users towards tweets related to the studied events. Specifically, an "interaction" refers to any of these user actions on a tweet: like, retweet, quote tweet, or reply. These actions signify the level of user engagement and information dissemination across the platform. To calculate the daily engagement indicator ($Daily\_EI$), we first summed the number of interactions for each tweet. This sum was then multiplied by the account's impact score. This impact score considers factors like follower count, follower-to-following ratio, and account age, providing a holistic perspective of its potential reach and influence. Through this approach, we were able to quantify the engagement and influence of tweets and accounts during the critical events under study. This analysis offers valuable insights into the dynamics of information

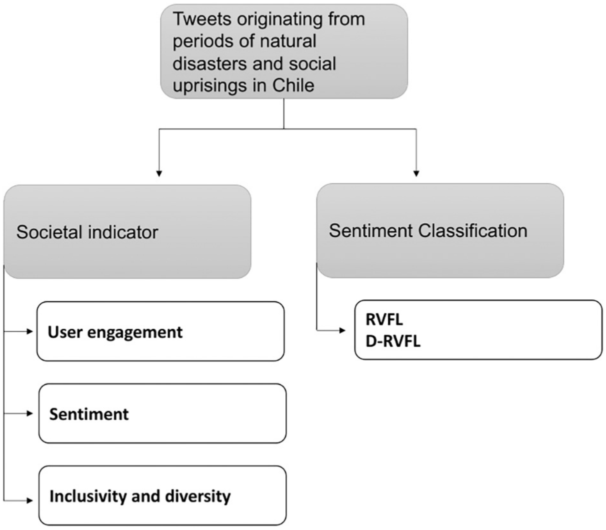

Fig. 1. Overall research framework.

diffusion and societal response on Twitter. The daily indicator, $Daily\_EI$ is obtained by multiplying the impact and daily average engagement of the tweets, and the indicator EI is the average of the $Daily\_EI$, as shown in Eq. (2) and Eq. (3) [84]:

$$Daily\_EI = impact \times \frac{\sum_{tweets} interactions}{4 \times \#of\_tweets} \tag{2}$$

$$EI = \frac{\sum_{days} Daily\_EI}{\#of\_days} \tag{3}$$

To calculate the sentiment indicator $Sent$, we analyzed each tweet using the Robertuito [85] model to obtain its positive, negative, and neutral emotions. As a mode, Robertuito is regarded as the highest benchmark of language model in Spanish. Robertuito is a pretrained language model for user-generated content in Spanish, trained following the RoBERTa guidelines on 500 million tweets. After calculating the sentiment and labeling each tweet as neutral, positive, or negative, the $Sent$ indicator is calculated for each user. To do so, we calculate the proportion of each sentiment for the user, and assign the values 10^{-6} if the user has mostly neutral tweets, the proportion of positive tweets if the user has mostly positive ones, and minus the proportion of negative tweets if the user mostly tweeted negatively (Eq. (4)) [63,14].

$$Sent = \begin{cases} 10^{-6} & if \quad max = neutral \\ \frac{\#Positive\_tweets}{\#Total\_tweets} & if \quad max = positive \\ -\frac{\#Negative\_tweets}{\#Total\_tweets} & if \quad max = negative \end{cases} \tag{4}$$

The third indicator Div analyzes the diversity and inclusivity of the language used by the user. This indicator uses a dictionary with terms in five dimensions, considering diversity in terms of gender, age, ethnicity, structure, and sectors of the economy, as shown in Table 4. This indicator was calculated as the number of dictionary terms, divided by the total number of tweets (Eq. (5)).

$$Div = \frac{\#words\_in\_dictionary}{\#Total\_tweets} \tag{5}$$

Finally, the compound Societal Impact indicator is created with the average of the normalized engagement indicator, sentiment indicator, and diversity indicator, as shown in Eq. (6) [63]:

$$SocietalImpact = \frac{EI + |Sent| + Div}{3} \tag{6}$$

3.3. Sentiment classification

3.3.1. TF-IDF

After preprocessing, the tweets were converted into a matrix of numeric vectors. We used a method called term frequency–inverse TF-IDF. This technique consists of term frequency (TF) and inverse document frequency (IDF). The TF focuses on the raw word count in a document, whereas the IDF focuses on how the frequency of a word is measured. Eq. (7) shows the TF formulations:

Table 4
Dictionary.

Dimension	Terms
Gender	hombre, mujer, trans, lgbt, mujeres, hombres, padres, madres, queer, homosexual transexual, pansexual, gay, lesbiana
Age	niños, niño, niña, jóvenes, adultos, adultos mayores abuelos, abuelas, familia, familias, abuela, abuelo
Cultural inferences	extranjero, extranjeros, peruano, venezolano, colombiano, haitiano
Ethnicity	mestizo, mapuches, mapuche, originarios, aymara, selknam, diaguita, indígenas
Economic sectors	salud, educación, construcción, minería, agricultura, energía, transporte, retail

Table 5
Social Uprising (October 2019).

Account type	Data	Positive	Negative	Total
Influencers	Train	10919	79039	89958
	Test	4707	33847	38554
Press & Radio	Train	37644	212043	249687
	Test	16083	90926	107009
Television	Train	15516	78884	94400
	Test	6814	33644	40458

Table 6
COVID-19.

Account type	Data	Positive	Negative	Total
Influencers	Train	7468	35783	43251
	Test	3055	15482	18537
Press & Radio	Train	17354	101672	119026
	Test	7594	43418	51012
Television	Train	6975	41744	48719
	Test	2949	17931	20880

Table 7
Wildfires (January 2017).

Account type	Data	Positive	Negative	Total
Influencers	Train	2115	6459	8574
	Test	938	2737	3675
Press & Radio	Train	6491	21208	27699
	Test	2923	8948	11871
Television	Train	2069	6400	8469
	Test	930	2736	3666

$$TF(t,d) = \frac{\text{Frequency of term (t) in the document (d)}}{\text{Total word in the document (d)}} \tag{7}$$

IDF's purpose is to calculate the informativeness of the word in a document. We need IDF because it helps minimize the weight of frequent terms and lends the infrequent terms a high impact. IDF can be computed using Eq. (8).

$$IDF(t) = log_2(\frac{\text{Total documents (N)}}{\text{Total documents with term (df(t))}}) \tag{8}$$

The TF-IDF expression in Eq. (9) can be obtained by combining Eq. (7) and Eq. (8) [86]:

$$TF - IDF = tf.idf(t,d,N) = tf(t,f).idf(t,N) \tag{9}$$

3.3.2. Data labeling

SA can be defined as a process that automates the mining of attitudes, opinions, views and emotions from text, speech, tweets and database sources through NLP. SA involves classifying text opinions into categories, such as *positive* or *negative* or *neutral*.

The labels in Tables 5, 6, 7 and 8, representing datasets from various events, such as the Chilean protests in October 2019, COVID-19, and wildfires (January 2017 and February 2023), are evidently imbalanced in terms of positive and negative classifications across

Table 8
Wildfires (February 2023).

Account type	Data	Positive	Negative	Total
Influencers	Train	1497	5995	7492
	Test	648	2564	3212
Press & Radio	Train	6145	24825	30970
	Test	2666	10608	13274
Television	Train	2430	12897	15327
	Test	5496	1074	6567

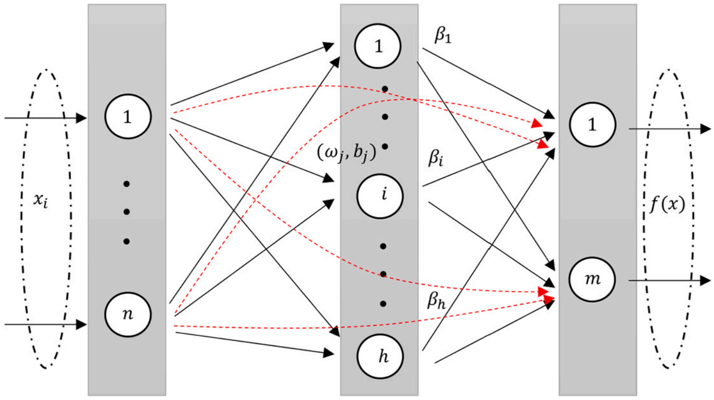

Fig. 2. Diagram illustrating an RVFL network, highlighting the direct links between input and output neurons (dashed red arrows).

different account types and data splits (training and testing). This imbalance can compromise the efficacy of ML models. Hence, there is an urgent need to implement algorithms or techniques that balance these labels to ensure reliable and unbiased analytical outcomes. To address the class imbalance within our data, we employed the Synthetic Minority Oversampling Technique (SMOTE) [87]. This technique generates new, synthetic data points based on existing ones to achieve a more balanced distribution. Here's how it works: for each data sample x_i belonging to the minority class $S_{\min}$, the algorithm identifies its (k) nearest neighbors. In our experiments, we set the value of k to 5.

3.4. Random vector functional link neural networks (RVFLN)

The Random Vector Functional Link Neural Network (RVFLN) [88] is a special type of neural network. Different from most, it connects the input directly to the output layer, bypassing the multiple layers typically found in these networks. This straightforward design, highlighted in our study and other studies [31,89], makes it faster and potentially more efficient for certain tasks, as illustrated in Fig. 2.

RVFL uses a novel approach for weight initialization. All weights and biases between the input and hidden layers are randomly initialized. Once set, these parameters remain unchanged during training. On the other hand, the output weights, shown as black solid lines in the diagram, are computed either via the Moore–Penrose pseudo-inverse or through ridge regression.

An intrinsic feature of RVFL is the direct connection (red line) between the input and output layers. This not only simplifies the architecture, but also acts as a regularization method, curbing the potential for overfitting.

One of RVFL's main merits is its efficiency. Given that there is no iterative weight tuning in the training phase, it boasts quicker convergence, computational ease, and often a diminished training error compared with other neural network methodologies.

In the RVFL network, enhancement nodes are used to map data from the input layer to the hidden layer, represented as $g(w_j x_i + b_j)$ where $g(\cdot)$ is the activation function, w_j is the weight of the jth enhancement node, b_j is the threshold and x_i the input vector. Input nodes are essentially a linear combination of inputs, expressed as $\sum_{j=h+1}^{h+n} \beta_j x_i$, with β_j being the weight terms. The RVFL network structure is represented concisely in a given format [18]:

$$\hat{y}_i = \sum_{j=1}^{h} \beta_j g(w_j x_i + b_j) + \sum_{j=h+1}^{h+n} \beta_j x_i. \tag{10}$$

Eq. (10) can be written in a compact form as follows

$$\mathbf{H}\boldsymbol{\beta} = \hat{\mathbf{Y}}, \tag{11}$$

where

$$\mathbf{H} = \begin{bmatrix} g(w \cdot x_1 + b_1) & \cdots & g(w_h \cdot x_1 + b_h) & x_1^T \\ \vdots & \ddots & \vdots & \\ g(w_1 \cdot x_N + b_1) & \cdots & g(w_h \cdot x_N + b_h) & x_N^T \end{bmatrix}$$

The output matrix $\boldsymbol{\beta}$ can be determined analytically with the least-squares solutions, under the constraint of minimum least-square $\min_{\beta}||\boldsymbol{\beta}||$ and $\min_{\beta}||\mathbf{H}\boldsymbol{\beta} - \mathbf{Y}||$ can be calculated as

$$\hat{\boldsymbol{\beta}} = \mathbf{H}^{\dagger}\mathbf{Y}, \tag{12}$$

where $\mathbf{H}^{\dagger}$ is the Moore-Penrose generalized inverse of matrix $\mathbf{H}$ [90] (Eq. (12)). The groundtruth $\mathbf{Y}$ is a matrix of size $N \times m$.

3.4.1. Deep RVFL network

The Deep Random Vector Functional Link (D-RVFL) network builds upon the traditional RVFL network by introducing concepts from representation learning or deep learning. In contrast to standard RVFL networks with a single hidden layer, D-RVFL utilizes multiple hidden layers. Each layer is fully interconnected and incorporates direct input-output connections that bypass the intermediate hidden layers. These direct connections play a crucial role in mitigating the vanishing gradient problem, a common challenge in deep neural networks.

Unlike its single-layered predecessor, the Deep Random Vector Functional Link (D-RVFL) network incorporates multiple hidden layers. Each layer is densely interconnected and includes direct input-output connections that bypass the intervening hidden layers. These direct connections play a critical role in alleviating the vanishing gradient problem, a frequent obstacle in deep networks.

By eliminating the requirement for backpropagation through the use of pre-defined, random weights and direct training of the output layer, D-RVFL networks achieve a substantial reduction in training time compared to traditional deep learning models [91]. D-RVFL networks also boast simplified configuration and implementation due to the absence of hidden layer weight tuning. Only the output weights require calculation, leading to diminished complexity and a lower risk of overfitting. The D-RVFL network architecture presents a compelling alternative to conventional neural network models, especially for applications where training speed and model parsimony are paramount. Its resilience to common neural network issues like overfitting, coupled with its ability to effectively manage complex patterns, makes it a valuable tool for expeditious yet precise data analysis [92,31].

While the D-RVFL network's stacked, hierarchical hidden layer structure offers general adaptability in terms of network dimensions (both width and depth), we will focus on a simpler case for this analysis. Here, we consider a stack of L hidden layers, with each layer containing an identical number of h hidden nodes [28].

The D-RVFL architecture, as referenced in [28,27], comprises multiple layered stacks. Within this structure, all the parameters of the hidden layers are initialized randomly and remain unchanged throughout the training period. Only the parameters of the output layer are computed analytically.

The output of the first hidden layer is then defined as follows:

$$\mathbf{H^{(1)}} = \mathbf{g}(\mathbf{XW^{(1)}}) \tag{13}$$

Every layer $l > 1$ (Eq. (13)) it is defined as [28,27,92,31]:

$$\mathbf{H^{(l)}} = \mathbf{g}(\mathbf{H^{l-1}W^{(l)}}) \tag{14}$$

The weights and biases of the hidden neurons are randomly generated within a suitable range and kept fixed during the training. $g()$ denotes the non-linear activation function (Eq. (14)). The input to the output layer is then defined as ([28,27]):

$$\mathbf{D} = [\mathbf{H^{(1)}H^{(2)}}...\mathbf{H^{(l-1)}H^{(l)}X}] \tag{15}$$

This design structure (Eq. (15)) is very similar to that of a standard RVFL network, wherein the input to the output layer consists of non-linear features from the stacked hidden layers along with the original features (as shown in Fig. 3).

The sigmoidal activation function $g(x) = 1/(1+\exp(-x))$ was used in our simulations. The maximum number of hidden layers was set to ten and the best performance was achieved. All experimental runs were performed 100 times and the averages and standard deviations were recorded. We divided all datasets into 70% for training and 30% for testing, and the division was performed randomly. All simulations were performed using the open-source R software environment for statistical computing running on an Apple M2 Max computer with 32 GB RAM.

4. Results

The results of each account for each event are shown in Tables 9, 10, 11 and 12. All values are between -1 and 1, in keeping with the methodology set forth under the previous section.

When obtaining the sentiment indicator for all four sets, all TV accounts obtained a value of 0.001, meaning that these accounts publish mostly neutral tweets with no difference in either natural or social events. However, the influencers showed more neutral results in the 2017 wildfire. They are more negative during the COVID-19 pandemic. In both the 2023 wildfire and social uprising, most influencers had negative indicators. Finally, the press and radio have a few accounts with mostly negative results, whereas the rest are mostly neutral but more open to presenting negative emotions about any event. In three of the four events, @ayala_rodolfo in the 2023 wildfire (-0.783), @GAMBA_CL in the 2019 social uprising (-0.946), and @Vitalicio7020 in the COVID pandemic (-0.603)-

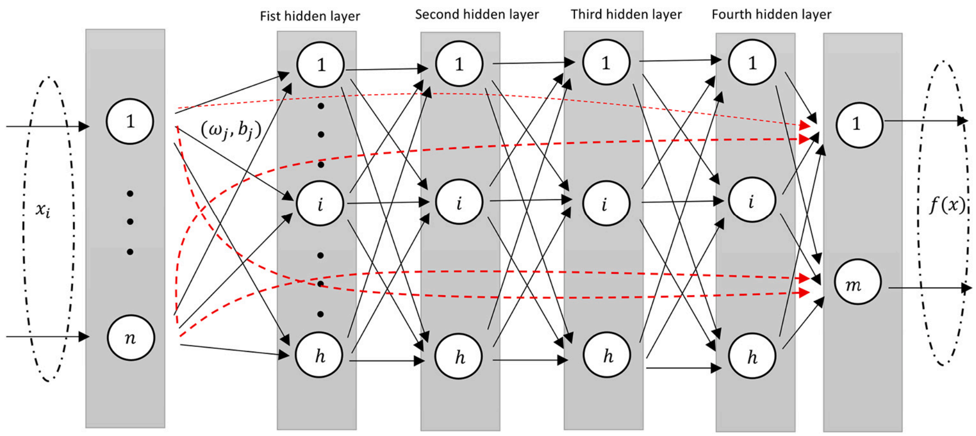

Fig. 3. Diagram illustrating an Deep-RVFL network, highlighting the direct links between input and output neurons (dashed red arrows).

Table 9
Account indicator results during the 2017 wildfires.

Account type	Name	Sentiment	Inclusion	Engagement	Societal indicator
Influencers	ablanch4	0.001	0.04	0.04	0.03
	alegriagonzaa	0.001	0.04	0.08	0.04
	bomberoschillan	0.001	0.05	0.01	0.02
	CEBioBio	0.001	0.04	0.02	0.02
	InfoNuble	0.001	0.06	0.00	0.02
	INFORMADORCHILE	-0.523	0.09	0.27	0.30
	PiensaPrensa	-0.531	0.07	0.08	0.23
	reddeemergencia	0.001	0.05	0.26	0.10
	S_Schwartzmann	0.001	0.10	0.08	0.06
Press & Radio	adnradiochile	0.001	0.06	0.39	0.15
	biobio	0.001	0.12	0.98	**0.37**
	Cooperativa	0.001	0.05	0.19	0.08
	eldesconcierto	**-0.613**	**0.16**	0.06	0.28
	elmostrador	-0.482	0.10	0.24	0.27
	latercera	0.001	0.08	0.29	0.12
	PublimetroChile	0.001	0.09	0.20	0.10
	theclinicc l	-0.576	0.13	0.16	0.29
Television	24HorasTVN	0.001	0.06	**1.00**	0.36
	CHVNoticias	0.001	0.00	0.00	0.00
	meganoticiascl	0.001	0.07	0.55	0.20
	T13	0.001	0.06	0.26	0.11

influencers obtained the most negative value, with the only exception being the online press portal @eldesconcierto, which obtained a score of -0.613 during the 2017 wildfires.

The inclusivity and diversity indicators showed similar levels between influences, press & radio accounts, and television accounts during all four events. However, it can be observed that natural disasters, like wildfires, show less inclusivity and diversity than events concerning social issues, such as the Chilean protests, the COVID-19 pandemic, and its measures. The highest values for the 2017 wildfire are obtained by the online news portal @eldesconcierto (0.16), for the 2023's wildfires the influencer @Cachoescalona1 (0.18) obtained the highest values, for the 2019 social uprising, it was the online news portal @InterferenciaCL (0.22), and for the COVID-19 pandemic, the influencer @Pa_tty (0.29) recorded the highest value.

Regarding engagement, influencers show worse results in the instructor than in the press & radio, and television accounts. This may be explained by the greater number of followers of the latter, which indicates that institutional accounts have broader public outreach. Simultaneously, in Figs. 4, 5 and 6, we observe that engagement peaked on the day the social uprising began for every account. However, as the other events did not have a day that directly marked the beginning of the event but instead grew consistently over time, there is no such peak for wildfires or COVID-19. When observing the total engagement indicator, we observe that, in three of the four events, higher values are obtained by accounts of established media agencies - @24HorasTVN, the state-owned television channel during the 2017 wildfires; @PublimetroChile, the most read newspaper of the country during the 2023's wildfires, and @biobio, a 60-year-old radio with a national outreach, during the social uprising- the COVID pandemic being the only exception, as the influencer @FelipeParadaM obtained the highest score.

Finally, the societal indicator for each account was calculated across all events. The influencers, press & radio, and television accounts achieved similar results for each event. This is normal because accounts with higher engagement are mostly neutral. Both wildfires show that the top scores in the indicator mostly belong to press & radio accounts, followed by influencers and top TV

Table 10
Account indicator results during the 2023 wildfires.

Account type	Name	Sentiment	Inclusion	Engagement	Societal indicator
Influencers	ayala_rodolfo	**-0.783**	0.06	0.00	0.28
	Cachoescalona1	-0.589	**0.18**	0.23	0.33
	camilaemiliasv	-0.497	0.09	0.74	**0.44**
	CapuchaCreativa	0.001	0.17	0.06	0.08
	Chileno17039890	-0.567	0.06	0.05	0.23
	MrRangerR1	0.001	0.07	0.00	0.02
	rsumen	-0.61	0.15	0.02	0.26
Press & Radio	adnradiochile	0.001	0.09	0.29	0.13
	biobio	-0.454	0.11	0.73	0.43
	Cooperativa	0.001	0.06	0.14	0.07
	eldesconcierto	-0.519	0.08	0.05	0.22
	elmostrador	0.001	0.14	0.11	0.09
	InterferenciaCL	-0.487	0.17	0.01	0.22
	latercera	0.001	0.09	0.24	0.11
	PublimetroChile	0.001	0.11	**1.00**	0.37
	thecliniccl	0.001	0.17	0.05	0.07
Television	24HorasTVN	0.001	0.09	0.18	0.09
	CHVNoticias	0.001	0.11	0.13	0.08
	meganoticiascl	0.001	0.13	0.44	0.19
	T13	0.001	0.09	0.18	0.09

Table 11
Account indicator results during the 2019 social uprising.

Account type	Name	Sentiment	Inclusion	Engagement	Societal indicator
Influencers	andres20ad	-0.77	0.21	0.01	0.33
	Chileokulto	-0.552	0.11	0.03	0.23
	csantander23	-0.534	0.07	0.00	0.20
	El_Ciudadano	0.001	0.16	0.03	0.06
	FelipeParadaM	-0.489	0.08	0.01	0.20
	GAMBA_CL	**-0.946**	0.14	0.02	0.37
	hernan_sr	-0.562	0.07	0.00	0.21
	JoviNomas	-0.551	0.09	0.01	0.22
	PiensaPrensa	-0.538	0.14	0.09	0.26
	vagoilustrado	-0.563	0.15	0.00	0.24
Press & radio	adnradiochile	0.001	0.09	0.30	0.13
	biobio	0.001	0.13	**1.00**	**0.38**
	Cooperativa	0.001	0.07	0.22	0.10
	eldesconcierto	0.001	0.13	0.04	0.06
	elmostrador	0.001	0.13	0.22	0.12
	InterferenciaCL	-0.509	**0.22**	0.04	0.25
	latercera	0.001	0.09	0.13	0.07
	PublimetroChile	-0.48	0.11	0.31	0.30
	thecliniccl	0.001	0.13	0.22	0.11
Television	24HorasTVN	0.001	0.11	0.82	0.31
	CHVNoticias	0.001	0.12	0.06	0.06
	meganoticiascl	0.001	0.10	0.15	0.08
	T13	0.001	0.10	0.32	0.14

accounts, while in COVID-19 and the Chilean protests the top scores were held mostly by influencers. The highest value in the indicator was held by @biobio during the 2017 wildfires (0.37) and in the 2019 social uprising (0.38), being the only account that obtained the highest result in any indicator across two different events, while influencers obtained the highest value in the other two events @camilaemiliasv (0.44) of 2023 wildfires, and @FelipeParadaM (0.53) in the COVID pandemic.

4.1. Classification

This section compares different ML methods, highlighting how the RVFL and its advanced version, the Deep RVFL, perform across various scenarios and types of social media accounts.

In Table 15, our findings demonstrate that the RVFL network exhibits relatively stable accuracy across varying neuron counts. This stability suggests robustness in the model's performance regardless of the changes in the dimensionality of the hidden layer. Specifically, we observed that during the COVID-19 event, influencers achieved the highest accuracy with 40 neurons at 75.53% (0.09). By contrast, the press & radio, and television categories reached their highest accuracies with 100 neurons, recording 77.52% (0.07) and 77.75% (0.12), respectively. The Chilean protests showed all account types performing optimally at the 40-neuron mark. For both the Wildfire events, there was a noticeable trend in which the accuracy increased with the number of neurons, with the

Table 12
Indicators results for the accounts during the COVID-19 pandemic.

Account type	Name	Sentiment	Inclusion	Engagement	Societal indicator
Influencers	csantander23	-0.592	0.09	0.02	0.23
	El_Ciudadano	0.001	0.18	0.07	0.08
	FelipeParadaM	-0.510	0.09	**1.00**	**0.53**
	FrancoBassoSotz	0.001	0.15	0.00	0.05
	Izkia	0.001	0.20	0.20	0.13
	Pa_tty	0.001	**0.29**	0.19	0.16
	PolarBearby	0.001	0.12	0.19	0.10
	RockandRolec	0.001	0.03	0.01	0.01
	UPLaRadio	0.001	0.11	0.02	0.04
	Vitalicio7020	**-0.603**	0.11	0.12	0.28
Press & radio	adnradiochile	0.001	0.10	0.64	0.25
	biobio	0.001	0.13	0.54	0.23
	Cooperativa	0.001	0.08	0.31	0.13
	eldesconcierto	0.001	0.13	0.25	0.13
	elmostrador	0.001	0.14	0.32	0.15
	InterferenciaCL	-0.495	0.21	0.04	0.25
	latercera	0.001	0.10	0.17	0.09
	PublimetroChile	0.001	0.12	0.77	0.30
	theclinicc l	0.001	0.15	0.39	0.18
Television	24HorasTVN	0.001	0.12	0.56	0.23
	CHVNoticias	0.001	0.13	0.16	0.09
	meganoticiascl	0.001	0.11	0.36	0.16
	T13	0.001	0.11	0.30	0.14

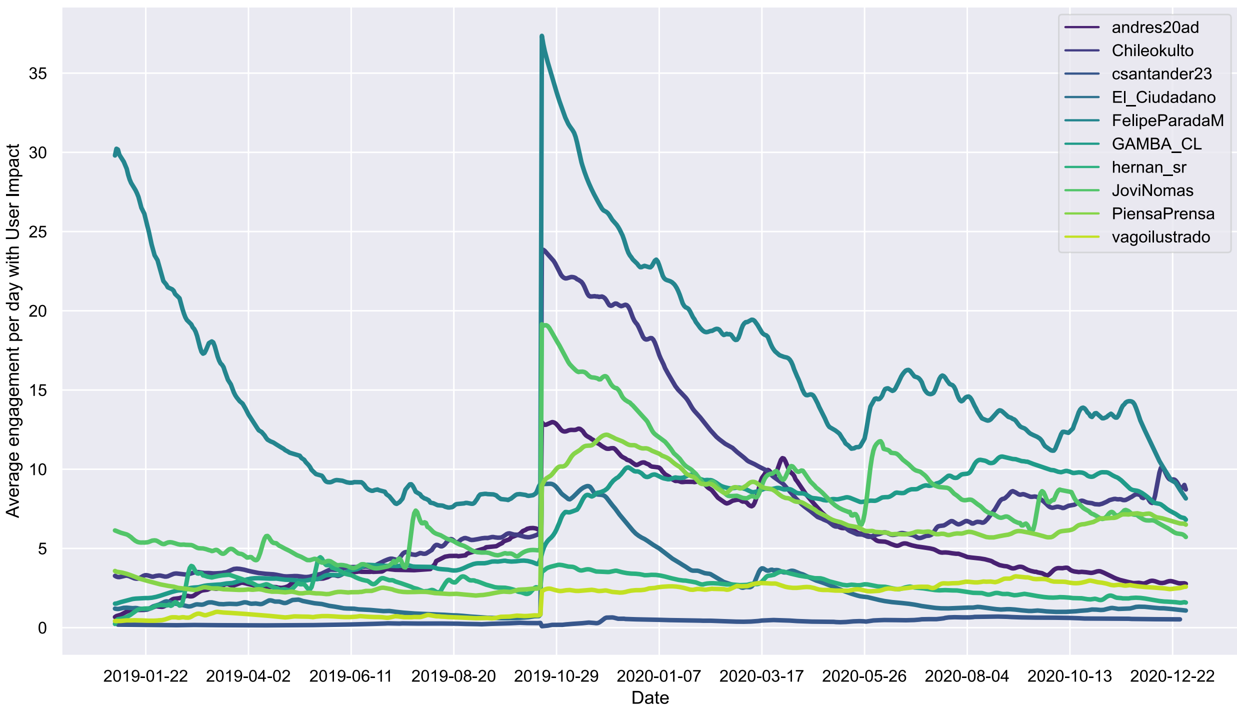

Fig. 4. Daily results for influencer engagement indicator during the Chilean protests and nine months prior.

influencer category reaching a peak accuracy of 100 neurons during the Wildfires in January 2017 and February 2023, with accuracies of 79.41% (0.19) and 80.57% (0.28), respectively. Our overall average accuracy across all events and neuron configurations was 77.21% (0.13), indicating a consistent model performance.

As shown in Table 16, the D-RVFL network exhibited a slight increase in accuracy as the neuron count increased. For instance, during the COVID-19 pandemic, influencers achieved the highest accuracy with 120 neurons at 76.46% (0.11). Press & radio and television categories delivered their best performance with 100 and 120 neurons, achieving accuracies of 78.23% (0.13) and 78.67% (0.16) respectively. Most notably, during the Wildfires (February 2023) event, the television account type achieved an accuracy

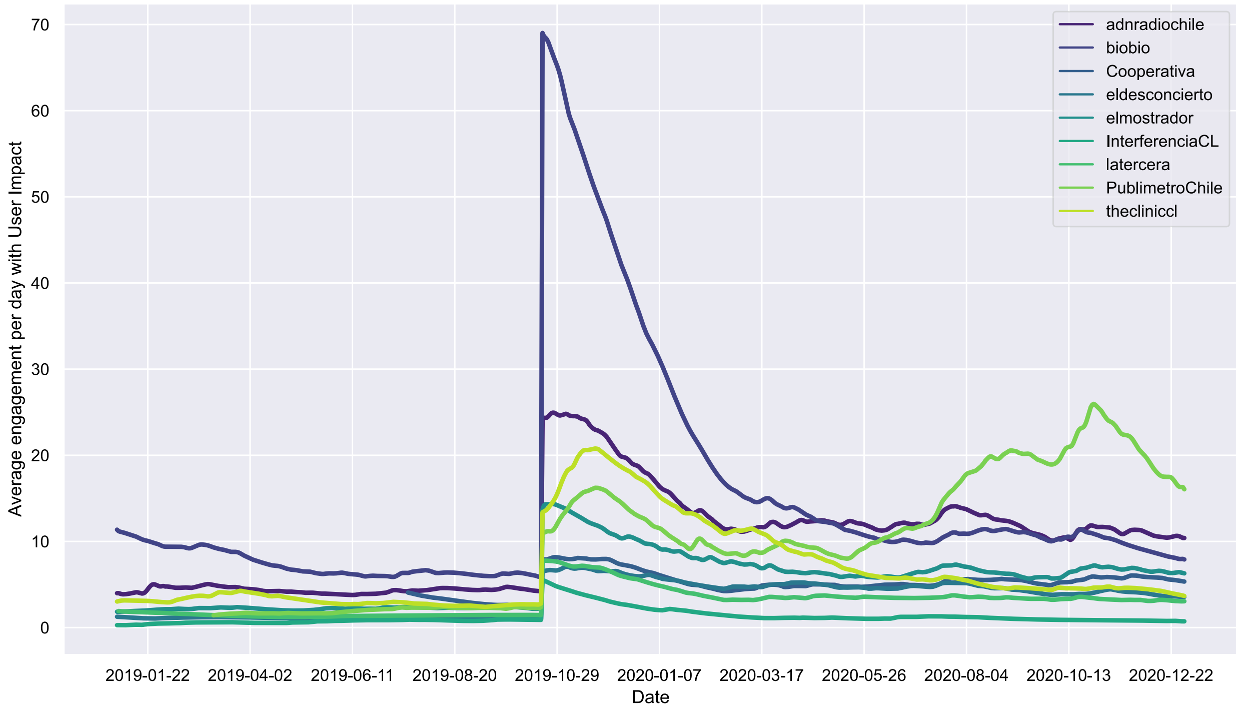

Fig. 5. Daily engagement indicator results of press and radio accounts during the social uprising and nine months prior.

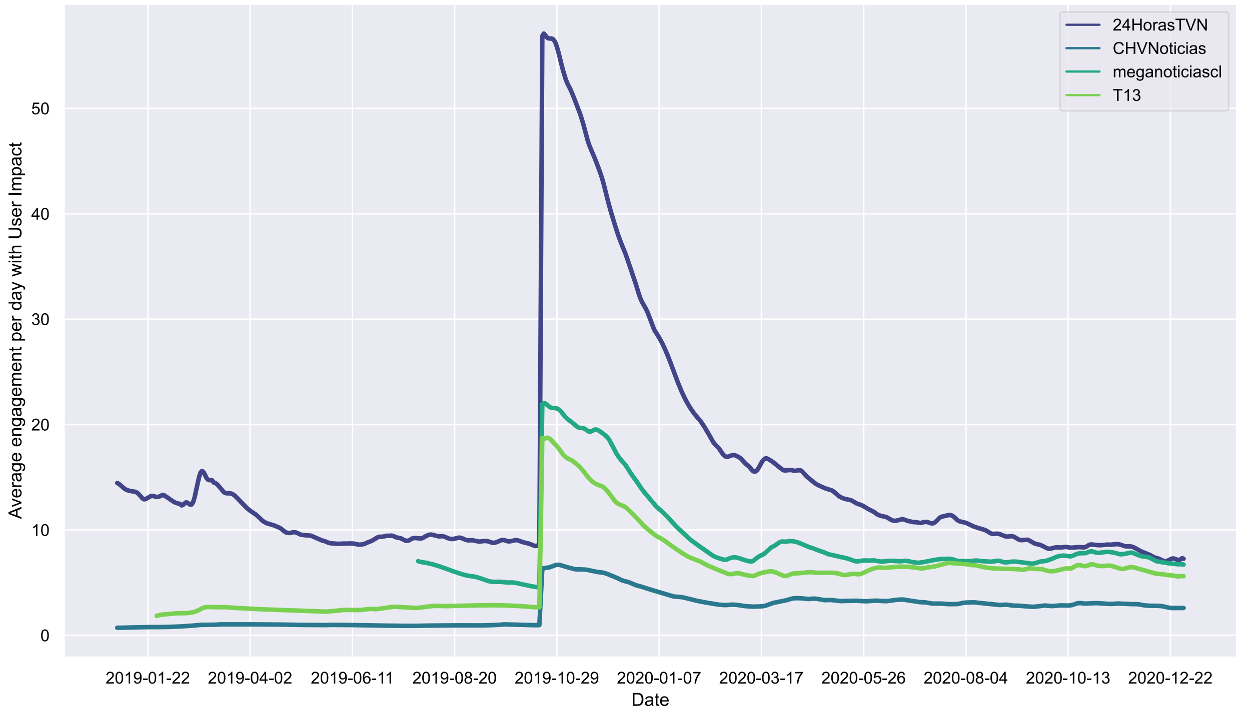

Fig. 6. Daily engagement indicator results of television accounts during the social uprising and nine months prior.

of 81.41% (0.29) with 100 neurons. The overall average accuracy for the D-RVFL network was 78.30% (0.17), outperforming the standard RVFL.

Table 17 provides a performance comparison of the various classifiers, including RVFL and D-RVFL. Our analysis indicates that traditional classifiers such as SVM, naive Bayes, and backpropagation (BP) neural networks were consistently outperformed by the

Table 13
Accuracy Results for COVID-19 Twitter Dataset.

Algorithm	Accuracy (%)
SVMBFTAN	82.8
BFTAN	72.5
TAN	69.8
Naive Bayes	72.13
SVM	69.62
RF	70.50
RVFL	**72.77**
D-RVFL	**74.88**

Table 14
Accuracy Results for Expo-2020 Twitter Dataset.

Algorithm	Accuracy (%)
SVMBFTAN	90.84
BFTAN	88.50
TAN	87.62
Naive Bayes	87.30
SVM	90.45
RF	91.50
RVFL	**90.77**
D-RVFL	**92.81**

RVFL and D-RVFL models. This is particularly evident in the Wildfires (Feb 2023) event, where the D-RVFL network achieved an accuracy of 80.87% (0.25) for influencers, which was significantly higher than that of the traditional models. The average accuracy across all events and account types for D-RVFL is 78.30% (0.17), highlighting its superior performance.

Our findings highlight the benefits of using RVFL and D-RVFL networks to classify information from various datasets. We found that D-RVFL networks with more neurons achieved higher accuracy than traditional methods. This demonstrates their strong capability to classify tweet sentiments accurately.

To evaluate the performance of various classification algorithms, we conducted a comparative analysis involving six algorithms on two separate events. Our analysis leveraged the sentiment analysis (SA) approach and data size established in [93], but applied it to these two new events. The algorithms included in this comparison were: Support Vector Machine (SVM), Bayes Factor Tree Augmented Naive Bayes (SVMBFTAN), Bayes Actor Tree Augmented Naive Bayes (BFTAN), Tree Augmented Naive Bayes (TAN), Naive Bayes, Random Forest (RF), Random Vector Functional Link (RVFL), and Deep Random Vector Functional Link (D-RVFL).

Our comparative analysis revealed that our approach yielded significantly higher accuracy on both datasets compared to the benchmarks established in [93]. As shown in Table 13, the D-RVFL network demonstrably surpassed all the conventional models evaluated using the COVID-19 Twitter data, solidifying its superior analytical capacity. For the Expo-2020 Twitter dataset (as detailed in Table 14), the D-RVFL network not only exceeded the peak accuracy achieved by the best performing RF model in the benchmark study [93], but it also underscores the substantial advantages of incorporating deep learning advancements into network architectures for intricate sentiment analysis tasks. These findings conclusively demonstrate that the D-RVFL network outperforms the existing methodologies outlined in [93], bringing about significant advancements in social media analytics.

4.1.1. Rationale for selecting RVFL and D-RVFL networks

When choosing methodologies for our sentiment analysis tasks, we prioritized the unique strengths and recent advancements of RVFL and D-RVFL networks in the field of Natural Language Processing (NLP). Our decision to favor RVFL and D-RVFL over other machine learning methods stems from several key factors.

Firstly, RVFL networks are renowned for their superior computational efficiency and faster training times. Unlike traditional deep learning models that rely heavily on backpropagation for parameter tuning, RVFL networks utilize pre-defined, random weights within their hidden layers. This streamlined approach significantly reduces training time and computational requirements, which are crucial for processing large datasets in sentiment analysis. Furthermore, D-RVFL networks capitalize on these strengths of RVFLs by incorporating deep learning architectures, ultimately enhancing the network's capacity to identify intricate data patterns.

Table 18 showcases the computational efficiency of various classifiers, including RVFL and D-RVFL, across diverse event types and account categories. As expected, the RVFL and D-RVFL networks exhibit significantly lower computational times compared to conventional machine learning classifiers like SVM, Naive Bayes, and Backpropagation (BP), highlighting their suitability for real-time sentiment analysis applications.

Table 15
RVFL performance for different neuron numbers in terms of accuracy and standard deviation.

Event	Account type	40 Neurons	60 Neurons	80 Neurons	100 Neurons	120 Neurons
COVID-19	Influencers	75.53 (0.09)	75.20 (0.10)	75.02 (0.08)	74.12 (0.07)	75.37 (0.09)
	Press & Radio	76.84 (0.07)	76.12 (0.09)	76.45 (0.08)	77.52 (0.07)	77.42 (0.08)
	Television	77.47 (0.11)	77.12 (0.10)	77.15 (0.12)	77.75 (0.12)	77.47 (0.13)
Social Uprising	Influencers	77.11 (0.07)	76.88 (0.07)	77.04 (0.06)	76.85 (0.08)	77.13 (0.08)
	Press & Radio	74.88 (0.05)	74.02 (0.07)	75.15 (0.05)	75.21 (0.07)	75.19 (0.06)
	Television	75.44 (0.05)	74.64 (0.06)	75.84 (0.04)	75.88 (0.05)	75.61 (0.07)
Wildfires (Jan 2017)	Influencers	78.69 (0.17)	78.72 (0.18)	79.01 (0.21)	79.41 (0.19)	79.22 (0.20)
	Press & Radio	77.25 (0.11)	77.41 (0.14)	76.98 (0.12)	76.57 (0.13)	76.49 (0.14)
	Television	76.12 (0.17)	75.76 (0.14)	77.12 (0.18)	77.21 (0.21)	76.94 (0.19)
Wildfires (Feb 2023)	Influencers	79.52 (0.26)	79.61 (0.31)	79.66 (0.29)	80.57 (0.28)	80.29 (0.27)
	Press & Radio	75.40 (0.15)	75.45 (0.17)	75.34 (0.21)	75.91 (0.12)	75.05 (0.14)
	Television	80.26 (0.19)	79.78 (0.19)	79.12 (0.19)	79.50 (0.21)	81.33 (0.22)
Average		77.04 (0.12)	76.72 (0.14)	76.99 (0.14)	77.21 (0.13)	77.29 (0.14)

Table 16
Deep RVFL performance for different neuron numbers in terms of accuracy and standard deviation.

Event	Account type	40 Neurons	60 Neurons	80 Neurons	100 Neurons	120 Neurons
COVID-19	Influencers	76.24 (0.11)	75.78 (0.12)	75.94 (0.09)	75.37 (0.09)	76.46 (0.11)
	Press & Radio	77.79 (0.09)	76.79 (0.11)	77.56 (0.09)	78.23 (0.13)	77.92 (0.11)
	Television	77.88 (0.13)	77.73 (0.14)	77.82 (0.15)	78.67 (0.16)	77.86 (0.16)
Social Uprising	Influencers	78.99 (0.11)	77.78 (0.11)	77.89 (0.12)	78.56 (0.07)	78.88 (0.09)
	Press & Radio	75.21 (0.06)	74.88 (0.08)	75.78 (0.06)	77.12 (0.09)	76.96 (0.07)
	Television	75.89 (0.09)	75.35 (0.08)	76.34 (0.06)	77.67 (0.07)	77.31 (0.08)
Wildfires (Jan 2017)	Influencers	79.67 (0.19)	79.45 (0.19)	79.88 (0.19)	80.17 (0.25)	80.21 (0.22)
	Press & Radio	77.56 (0.12)	77.78 (0.15)	77.88 (0.13)	77.12 (0.14)	77.23 (0.16)
	Television	76.42 (0.18)	76.72 (0.16)	77.42 (0.20)	77.56 (0.25)	77.22 (0.22)
Wildfires (Feb 2023)	Influencers	80.12 (0.24)	80.22 (0.30)	80.46 (0.27)	80.87 (0.25)	80.44 (0.27)
	Press & Radio	75.52 (0.18)	75.77 (0.21)	76.12 (0.22)	76.88 (0.19)	75.31 (0.18)
	Television	80.32 (0.21)	80.22 (0.19)	80.19 (0.20)	81.41 (0.29)	81.39 (0.27)
Average		77.55 (0.14)	77.37 (0.15)	77.77 (0.15)	78.30 (0.17)	78.10 (0.16)

Table 17
Performance comparison of various classifiers with RVFL and D-RVFL in terms of accuracy.

Event	Account type	SVM	Naive Bayes	BP	RVFL	D-RVFL
COVID-19	Influencers	61.07 (1.21)	62.82 (1.11)	64.19 (1.09)	74.12 (0.07)	75.37 (0.09)
	Press & Radio	62.56 (1.09)	63.65 (1.02)	65.23 (1.20)	77.52 (0.07)	78.23 (0.13)
	Television	62.78 (1.12)	63.89 (1.11)	65.79 (1.21)	77.75 (0.12)	78.67 (0.16)
Social Uprising	Influencers	63.45 (0.98)	64.12 (1.14)	66.78 (0.98)	76.85 (0.08)	78.56 (0.07)
	Press & Radio	64.12 (1.07)	65.31 (1.08)	66.91 (1.04)	75.21 (0.07)	77.12 (0.09)
	Television	65.13 (1.11)	65.89 (1.01)	66.65 (1.19)	75.88 (0.05)	77.67 (0.07)
Wildfires (Jan 2017)	Influencers	66.12 (1.78)	65.87 (1.14)	67.12 (1.17)	79.41 (0.19)	80.21 (0.22)
	Press & Radio	65.88 (1.16)	65.91 (1.05)	66.21 (1.11)	76.57 (0.13)	77.23 (0.16)
	Television	66.08 (1.15)	66.12 (1.01)	66.24 (1.25)	77.21 (0.21)	77.22 (0.22)
Wildfires (Feb 2023)	Influencers	66.67 (1.16)	66.92 (1.13)	67.24 (1.02)	80.57 (0.28)	80.87 (0.25)
	Press & Radio	66.56 (1.12)	66.78 (1.18)	67.34 (1.26)	75.91 (0.12)	76.88 (0.19)
	Television	66.48 (1.17)	66.89 (1.18)	67.39 (1.17)	79.50 (0.21)	81.41 (0.29)
Average		64.74 (1.18)	65.35 (1.09)	66.42 (1.14)	77.21 (0.13)	78.30 (0.17)

5. Discussion and future directions

This section explores some intriguing insights gleaned from this study regarding the strengths and limitations of the proposed approach. These observations warrant further discussion.

Our approach offers a broader and more comprehensive perspective on critical events compared to traditional social science methods. This includes in-depth interviews, focus groups, questionnaires, and conventional big data analysis of social media plat-

Table 18
Computational time (in seconds) of various classifiers.

Event	Account type	SVM	Naive Bayes	BP	RVFL	D-RVFL
COVID-19	Influencers	80.12	50.56	107.17	30.67	32.19
	Press & Radio	100.12	80.12	134.78	45.14	48.09
	Television	60.10	51.14	111.45	32.11	33.22
Social Uprising	Influencers	142.11	101.15	223.87	78.45	81.49
	Press & Radio	214.41	188.45	312.89	121.12	123.56
	Television	143.21	109.23	225.46	79.15	82.56
Wildfires (Jan 2017)	Influencers	30.14	21.56	38.12	14.78	16.21
	Press & Radio	52.12	41.16	78.19	22.78	24.91
	Television	30.88	21.14	39.55	15.09	16.76
Wildfires (Feb 2023)	Influencers	31.11	22.12	39.44	15.01	16.33
	Press & Radio	54.17	43.19	81.46	24.46	27.11
	Television	45.11	38.58	66.89	21.83	24.52

forms. However, we acknowledge the persistent challenges associated with coding and classifying Twitter data. This can hinder the identification of relevant communication patterns, especially when the volume of tweets and retweets surges during critical events. Furthermore, discrepancies between Twitter data and ground truth information obtained from physical sensors can lead to authorities making misinformed decisions. To mitigate these issues, we propose developing methods to minimize data collection errors and introducing new time-saving techniques to improve geolocation accuracy.

Data retrieved through Twitter's APIs (including the Filter API and Sample API) may not always be fully indicative of the platform's overall communication, making it challenging to evaluate the data's representativeness. To address these potential biases inherent in using these APIs, we employed deep neural networks for sentiment analysis. This approach capitalized on the sentiment conveyed by hashtags as a significant classification feature. By incorporating this strategy, we were able to mitigate the biases and noise potentially present within the dataset, potentially strengthening the reliability of our results.

One challenge, as identified by [94], is the duplication of hashtags, both individually and collectively. Over time, this repetition can indicate confirmation bias and the development of echo chambers. To address potential bias in hashtag trends, [95] proposed a method that involves extracting representative samples from the Sample API (which offers an unbiased view of tweets) and calculating confidence intervals for hashtag activity. However, the possibility that the Sample API itself may be biased, as suggested by [96], necessitates closer examination and the development of more refined data collection strategies.

To address these data collection obstacles, [97] proposed a strategy that bolsters the trustworthiness of Twitter data. Their method involves sampling user accounts through the REST API, which yielded a substantial dataset encompassing 1.6 million accounts and 500 million tweets. This dataset is noteworthy for containing primarily Japanese users, recognized for their high level of activity on Twitter. The data is updated monthly using cost-efficient cloud services. This approach demonstrably enhances the correlation between keyword trends and other patterns when compared to alternative API-based methods. This is corroborated by the remarkably high correlation score of 0.97 achieved with the Crimson Hexagon dataset.

This study also highlights a potential bias in the emotional makeup of Twitter samples due to the platform's inherent sampling methods. While there's a common perception that negative tweets might be more prevalent during critical events, the true distribution of sentiment classes within the data may not precisely reflect this. To address this class imbalance issue, we implemented weighting strategies or oversampling techniques like SMOTE. These techniques help improve the classifiers' ability to accurately predict sentiment across datasets with uneven class distributions. This is a significant area for further exploration, and although we only examined the use of SMOTE in this study, other approaches to mitigating class imbalance problems warrant investigation within this context, as suggested by [98].

Moving forward, the convergence of hashtag-level and tweet-level sentiment analysis has the potential to strengthen the validity of sentiment assessments. However, training deep RVFL network classifiers for specific events presents challenges due to the transient and event-dependent nature of the data. This characteristic may limit the effectiveness of trained models when applied to future, similar events. To address these limitations, we propose exploring online learning techniques or unsupervised learning algorithms for deep RVFL networks. These approaches can continuously train models using large-scale datasets, improving their adaptability. This capability, combined with the network's capacity to handle big data, positions deep RVFL networks as promising instruments for sentiment analysis in evolving and critical event scenarios.

We acknowledge the vital importance of considering a wider range of contextual variables that influence social media behavior during emergencies. Future research should delve into how cultural backgrounds, socioeconomic conditions, and political landscapes impact information sharing and processing across diverse social media platforms and geographic locations. To accomplish this, we propose broadening our analysis beyond Twitter to encompass platforms like Facebook and Instagram, facilitating a cross-platform investigation. This expansion has the potential to yield a more nuanced understanding of the distinct dynamics and user interactions characteristic of each platform.

Furthermore, to glean more profound understandings of social media users' motivations and behaviors during critical events, we will incorporate qualitative methodologies such as user interviews and content analysis. This approach will equip us to investigate the types of content that generate user interest and how these preferences evolve across various crisis stages and contexts.

6. Conclusion

The indicators showed that the higher inclusivity scores were mostly obtained by influencers and independent online news portals, and not by established mass media. This can be related to not having to follow organizational guidelines but instead taking the personal decision of having a more inclusive language or focusing on topics where inclusivity is a major theme.

We also observed that the influencers' characteristics changed depending on the event. Because the social uprising was an event with higher political and social components, most confrontational people became influencers, which was reflected in the higher negative scores on the sentiment indicator. This explains why stronger sentiments increase an account's engagement and reach but only when the event is favorable to such comments.

Integrating the performance metrics of ML classifiers with the sentiment, inclusivity, engagement, and societal indicators from social media account analyses provides a comprehensive overview of information dissemination during critical events.

The consistent performance of RVFL and D-RVFL across different neuron configurations, as highlighted by ML classifier evaluations, underscores the potential of these networks in predicting complex data patterns. This robustness is particularly relevant when analyzing social media content, which is characterized by its dynamic and multifaceted nature.

A SA across the four events revealed that television accounts maintained a neutral stance, potentially reflecting a systematic approach to broadcasting without emotional bias. By contrast, influencers showed more variability in sentiment, with a proclivity for negativity during the COVID-19 and 2023 wildfires, indicating a more personal or individualized expression of reactions to events. The press and radio accounts were generally neutral but had a greater propensity to express negative sentiments, possibly reflecting the critical role of journalism in scrutinizing events.

The inclusivity and diversity indicators suggest that social issues may elicit more diverse viewpoints and discussions across the analyzed account types than natural disasters. This could be due to the social ramifications of such events, which might engage a broader spectrum of voices and opinions.

Engagement levels point to institutional accounts having a wider reach and impact, possibly because of their established follower bases. However, during the social uprising, a spike in engagement was observed across all account types, implying that certain events can dramatically heighten public attention and interaction regardless of the usual engagement patterns.

In conclusion, the interplay between ML classifier performance and social media analysis provides insights into how different account types react to and are perceived during various events. While ML models, such as RVFL and D-RVFL, can efficiently classify and predict data patterns, sentiment and engagement indicators offer a nuanced understanding of human responses to the data, especially in the context of influencers and institutional accounts during times of crisis. The higher accuracy of D-RVFL in classification tasks suggests that deeper models may be better at capturing the complexities of social media discourse, which is reflected in the diverse sentiments and engagement levels observed across events.

Author contribution statement

All authors listed have significantly contributed to the development and the writing of this article.

CRediT authorship contribution statement

Pablo A. Henríquez: Writing – review & editing, Writing – original draft, Visualization, Validation, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Francisco Alessandri:** Formal analysis, Data curation, Conceptualization.

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.

Data availability

Data associated with this study can be found online at https://github.com/pabhenriquez/Indicadores.

Funding statement

This work was supported by ANID/PIA/ANILLOS ACT210096; ANID FONDECYT Iniciación 2023 grant No. 11230396.

References

[1] N. Öztürk, S. Ayvaz, Sentiment analysis on Twitter: a text mining approach to the syrian refugee crisis, Telemat. Inform. 35 (1) (2018) 136–147, https://doi.org/10.1016/j.tele.2017.10.006.

[2] A. Agarwal, B. Xie, I. Vovsha, O. Rambow, R.J. Passonneau, Sentiment analysis of Twitter data, in: Proceedings of the Workshop on Language in Social Media (LSM 2011), 2011, pp. 30–38.

[3] H. Saif, Y. He, H. Alani, Semantic sentiment analysis of Twitter, in: The Semantic Web–ISWC 2012: 11th International Semantic Web Conference, Boston, MA, USA, November 11-15, 2012, Proceedings, Part I, vol. 11, Springer, 2012, pp. 508–524.
[4] H. Kwak, C. Lee, H. Park, S. Moon, What is Twitter, a social network or a news media?, in: Proceedings of the 19th International Conference on World Wide Web, WWW'10, ACM, New York, NY, USA, 2010, pp. 591–600.
[5] P.T. Metaxas, E. Mustafaraj, Soc. Med. Elections Sci. 338 (6106) (2012) 472–473.
[6] M.D. Conover, E. Ferrara, F. Menczer, A. Flammini, The digital evolution of occupy Wall Street, PLoS ONE 8 (5) (2013) 1–5.
[7] T. Sakaki, M. Okazaki, Y. Matsuo, Earthquake shakes Twitter users: real-time event detection by social sensors, in: Proceedings of the 19th International Conference on World Wide Web, WWW'10, ACM, New York, NY, USA, 2010, pp. 851–860.
[8] M. Mendoza, B. Poblete, C. Castillo, Twitter under crisis: can we trust what we RT?, in: Proceedings of the First Workshop on Social Media Analytics, SOMA'10, ACM, New York, NY, USA, 2010, pp. 71–79.
[9] P. Panagiotopoulos, J. Barnett, A.Z. Bigdeli, S. Sams, Social media in emergency management: Twitter as a tool for communicating risks to the public, Technol. Forecast. Soc. Change 111 (2016) 86–96, https://doi.org/10.1016/j.techfore.2016.06.010.
[10] Y. Jin, L.L. Austin, Social Media and Crisis Communication, Taylor & Francis, 2017.
[11] O.D. Apuke, E.A. Tunca, Social media and crisis management: a review and analysis of existing studies, LAÜ Sosyal Bilim. Derg. 9 (2) (2018) 199–215.
[12] N.A.B. Ibrahim, R.B. Musa, R.B. Abdul Wahab, The effect of social media depends on social media intelligence among graduates, in: Regional Conference on Science, Technology and Social Sciences (RCSTSS 2014) Business and Social Sciences, Springer, 2016, pp. 835–843.
[13] A. Bodaghi, J. Oliveira, The theater of fake news spreading, who plays which role? A study on real graphs of spreading on Twitter, Expert Syst. Appl. 189 (2022) 116110.
[14] G.A. Ruz, P.A. Henríquez, A. Mascareño, Sentiment analysis of Twitter data during critical events through Bayesian networks classifiers, Future Gener. Comput. Syst. 106 (2020) 92–104, https://doi.org/10.1016/j.future.2020.01.005.
[15] C. Lou, S. Yuan, Influencer marketing: how message value and credibility affect consumer trust of branded content on social media, J. Interact. Advert. 19 (1) (2019) 58–73, https://doi.org/10.1080/15252019.2018.1533501.
[16] B. Liu, Sentiment Analysis: Mining Opinions, Sentiments, and Emotions, Cambridge University Press, 2015.
[17] R. Pandarachalil, S. Sendhilkumar, G.S. Mahalakshmi, Twitter sentiment analysis for large-scale data: an unsupervised approach, Cogn. Comput. 7 (2) (2015) 254–262.
[18] L. Zhang, P. Suganthan, A survey of randomized algorithms for training neural networks, Inf. Sci. 364 (Supplement C) (2016) 146–155.
[19] W.F. Schmidt, M.A. Kraaijveld, R.P.W. Duin, Feedforward neural networks with random weights, in: Proceedings, 11th IAPR International Conference on Pattern Recognition, vol. II. Conference B: Pattern Recognition Methodology and Systems, 1992, pp. 1–4.
[20] Y.-H. Pao, G.-H. Park, D.J. Sobajic, Learning and generalization characteristics of the random vector functional-link net, Neurocomputing 6 (2) (1994) 163–180, backpropagation, Part IV.
[21] G.-B. Huang, Q.-Y. Zhu, C.-K. Siew, Extreme learning machine: theory and applications, Neurocomputing 70 (1) (2006) 489–501.
[22] D.P. Mesquita, J.P.P. Gomes, L.R. Rodrigues, S.A. Oliveira, R.K. Galvão, Building selective ensembles of randomization based neural networks with the successive projections algorithm, Appl. Soft Comput. (2017).
[23] N. Vuković, M. Petrović, Z. Miljković, A comprehensive experimental evaluation of orthogonal polynomial expanded random vector functional link neural networks for regression, Appl. Soft Comput. (2017).
[24] L. Tang, Y. Wu, L. Yu, A non-iterative decomposition-ensemble learning paradigm using rvfl network for crude oil price forecasting, Appl. Soft Comput. (2017).
[25] L. Zhang, P. Suganthan, A comprehensive evaluation of random vector functional link networks, Inf. Sci. 367 (Supplement C) (2016) 1094–1105.
[26] Y. Ren, P. Suganthan, N. Srikanth, G. Amaratunga, Random vector functional link network for short-term electricity load demand forecasting, Inf. Sci. 367 (Supplement C) (2016) 1078–1093.
[27] P.A. Henríquez, G.A. Ruz, Twitter sentiment classification based on deep random vector functional link, in: 2018 International Joint Conference on Neural Networks (IJCNN), 2018, pp. 1–6.
[28] Q. Shi, R. Katuwal, P. Suganthan, M. Tanveer, Random vector functional link neural network based ensemble deep learning, Pattern Recognit. 117 (2021) 107978, https://doi.org/10.1016/j.patcog.2021.107978.
[29] P.-B. Zhang, Z.-X. Yang, A new learning paradigm for random vector functional-link network: Rvfl+, Neural Netw. 122 (2020) 94–105, https://doi.org/10.1016/j.neunet.2019.09.039.
[30] M. Ganaie, M. Tanveer, A. Malik, P. Suganthan, Minimum variance embedded random vector functional link network with privileged information, in: 2022 International Joint Conference on Neural Networks (IJCNN), 2022, pp. 1–8.
[31] A. Malik, R. Gao, M. Ganaie, M. Tanveer, P.N. Suganthan, Random vector functional link network: recent developments, applications, and future directions, Appl. Soft Comput. 143 (2023) 110377, https://doi.org/10.1016/j.asoc.2023.110377.
[32] E. Bonsón, S. Royo, M. Ratkai, Citizens' engagement on local governments' Facebook sites. An empirical analysis: the impact of different media and content types in western Europe, Gov. Inf. Q. 32 (1) (2015) 52–62, https://doi.org/10.1016/j.giq.2014.11.001.
[33] J. He, X. Li, P. Liu, L. Wang, H. Zhou, J. Wang, R. Tang, Ensemble deep random vector functional link for self-supervised direction-of-arrival estimation, Eng. Appl. Artif. Intell. 120 (2023) 105831.
[34] S. Zhang, L. Xie, Deep incremental random vector functional-link network: a non-iterative constructive sketch via greedy feature learning, Appl. Soft Comput. 143 (2023) 110410.
[35] S. Shiva, M. Hu, P.N. Suganthan, Online learning using deep random vector functional link network, Eng. Appl. Artif. Intell. 125 (2023) 106676.
[36] M. Jamei, M. Karbasi, M. Ali, A. Malik, X. Chu, Z.M. Yaseen, A novel global solar exposure forecasting model based on air temperature: designing a new multi-processing ensemble deep learning paradigm, Expert Syst. Appl. 222 (2023) 119811.
[37] W. Jiao, S. Song, H. Han, W. Wang, Q. Zhang, Artificially intelligent differential diagnosis of enlarged lymph nodes with random vector functional link network plus, Med. Eng. Phys. 111 (2023) 103939.
[38] C. Sarkar, D. Gupta, B.B. Hazarika, 1-norm twin random vector functional link networks based on universum data for leaf disease detection, Appl. Soft Comput. 148 (2023) 110850.
[39] M.A. Al-qaness, A.A. Ewees, H. Fan, L. Abualigah, A.H. Elsheikh, M. Abd Elaziz, Wind power prediction using random vector functional link network with capuchin search algorithm, Ain Shams Eng. J. 14 (9) (2023) 102095.
[40] A.K. Malik, M. Ganaie, M. Tanveer, P. Suganthan, A.D.N.I. Initiative, et al., Alzheimer's disease diagnosis via intuitionistic fuzzy random vector functional link network, IEEE Trans. Comput. Soc. Syst. (2022).
[41] M. Ganaie, M. Tanveer, A.K. Malik, P. Suganthan, Minimum variance embedded random vector functional link network with privileged information, in: 2022 International Joint Conference on Neural Networks (IJCNN), IEEE, 2022, pp. 1–8.
[42] D. Samal, P.K. Dash, R. Bisoi, Modified added activation function based exponential robust random vector functional link network with expanded version for nonlinear system identification, Appl. Intell. 52 (5) (2022) 5657–5683.
[43] J. Chen, S. Yang, D. Zhang, Y.A. Nanehkaran, A turning point prediction method of stock price based on rvfl-gmdh and chaotic time series analysis, Knowl. Inf. Syst. 63 (10) (2021) 2693–2718.
[44] X. Guo, W. Zhou, Q. Lu, A. Du, Y. Cai, Y. Ding, et al., Assessing dry weight of hemodialysis patients via sparse Laplacian regularized rvfl neural network with l 2, 1-norm, BioMed Res. Int. (2021) 2021.

[45] A. Alalimi, L. Pan, M.A. Al-Qaness, A.A. Ewees, X. Wang, M. Abd Elaziz, Optimized random vector functional link network to predict oil production from tahe oil field in China, Oil Gas Sci. Technol. 76 (2021) 3.
[46] R. Gao, L. Du, K.F. Yuen, P.N. Suganthan, Walk-forward empirical wavelet random vector functional link for time series forecasting, Appl. Soft Comput. 108 (2021) 107450.
[47] D.R. Dash, P. Dash, R. Bisoi, Short term solar power forecasting using hybrid minimum variance expanded rvfln and sine-cosine Levy flight pso algorithm, Renew. Energy 174 (2021) 513–537.
[48] B.B. Hazarika, D. Gupta, Modelling and forecasting of Covid-19 spread using wavelet-coupled random vector functional link networks, Appl. Soft Comput. 96 (2020) 106626.
[49] T. Chakravorti, P. Satyanarayana, Non linear system identification using kernel based exponentially extended random vector functional link network, Appl. Soft Comput. 89 (2020) 106117.
[50] S.-Y. Lu, S.-H. Wang, Y.-D. Zhang, A classification method for brain mri via mobilenet and feedforward network with random weights, Pattern Recognit. Lett. 140 (2020) 252–260.
[51] A. Rasheed, A. Adebisi, K.C. Veluvolu, Respiratory motion prediction with random vector functional link (rvfl) based neural networks, J. Phys. Conf. Ser. 1626 (2020) 012022.
[52] F. Essa, M. Abd Elaziz, A.H. Elsheikh, Prediction of power consumption and water productivity of seawater greenhouse system using random vector functional link network integrated with artificial ecosystem-based optimization, Process Saf. Environ. Prot. 144 (2020) 322–329.
[53] A.M. Hussein, M. Abd Elaziz, M.S.A. Wahed, M. Sillanpää, A new approach to predict the missing values of algae during water quality monitoring programs based on a hybrid moth search algorithm and the random vector functional link network, J. Hydrol. 575 (2019) 852–863.
[54] D.R. Nayak, R. Dash, B. Majhi, U.R. Acharya, Application of fast curvelet Tsallis entropy and kernel random vector functional link network for automated detection of multiclass brain abnormalities, Comput. Med. Imaging Graph. 77 (2019) 101656.
[55] G.A. Ruz, P.A. Henríquez, Random vector functional link with naive Bayes for classification problems of mixed data, in: 2019 IEEE 31st International Conference on Tools with Artificial Intelligence (ICTAI), IEEE, 2019, pp. 1749–1752.
[56] R. Bisoi, P. Dash, S. Mishra, Modes decomposition method in fusion with robust random vector functional link network for crude oil price forecasting, Appl. Soft Comput. 80 (2019) 475–493.
[57] Y. Dash, S.K. Mishra, S. Sahany, B.K. Panigrahi, Indian summer monsoon rainfall prediction: a comparison of iterative and non-iterative approaches, Appl. Soft Comput. 70 (2018) 1122–1134.
[58] A.T. Chatfield, C.G. Reddick, All hands on deck to tweet sandy: networked governance of citizen coproduction in turbulent times, in: Agile Government and Adaptive Governance in the Public Sector, Gov. Inf. Q. 35 (2) (2018) 259–272, https://doi.org/10.1016/j.giq.2017.09.004.
[59] J.P. Guidry, Y. Jin, C.A. Orr, M. Messner, S. Meganck, Ebola on Instagram and Twitter: how health organizations address the health crisis in their social media engagement, Public Relat. Rev. 43 (3) (2017) 477–486, https://doi.org/10.1016/j.pubrev.2017.04.009.
[60] C.-V. Priporas, N. Stylos, I.E. Kamenidou, City image, city brand personality and generation z residents' life satisfaction under economic crisis: predictors of city-related social media engagement, J. Bus. Res. 119 (2020) 453–463, https://doi.org/10.1016/j.jbusres.2019.05.019.
[61] Q. Chen, C. Min, W. Zhang, G. Wang, X. Ma, R. Evans, Unpacking the black box: how to promote citizen engagement through government social media during the Covid-19 crisis, Comput. Hum. Behav. 110 (2020) 106380, https://doi.org/10.1016/j.chb.2020.106380.
[62] J.L. Johnson, N. Bhatia, D.L. West, N.M. Safdar, Leveraging social media and web presence to discuss and promote diversity, equity, and inclusion in radiology, J. Am. Coll. Radiol. 19 (1, Part B) (2022) 207–212, https://doi.org/10.1016/j.jacr.2021.09.020.
[63] M.K. Baxi, J. Philip, V. Mago, Resilience of political leaders and healthcare organizations during Covid-19, PeerJ Comput. Sci. 8 (2022) e1121, https://doi.org/10.7717/peerj-cs.1121.
[64] M.C. Getchell, T.L. Sellnow, A network analysis of official Twitter accounts during the West Virginia water crisis, Comput. Hum. Behav. 54 (Supplement C) (2016) 597–606.
[65] N.R. Paul, D. Sahoo, R.C. Balabantaray, Classification of crisis-related data on Twitter using a deep learning-based framework, Multimed. Tools Appl. 82 (6) (2023) 8921–8941.
[66] J.B. Houston, J. Hawthorne, M.F. Perreault, E.H. Park, M. Goldstein Hode, M.R. Halliwell, S.E. Turner McGowen, R. Davis, S. Vaid, J.A. McElderry, S.A. Griffith, Social media and disasters: a functional framework for social media use in disaster planning, response, and research, Disasters 39 (1) (2015) 1–22.
[67] K.M. Carley, M. Malik, P.M. Landwehr, J. Pfeffer, M. Kowalchuck, Crowd sourcing disaster management: the complex nature of Twitter usage in padang Indonesia, Saf. Sci. 90 (Supplement C) (2016) 48–61.
[68] J. Mittelmeier, H. Cockayne, Global representations of international students in a time of crisis: a qualitative analysis of Twitter data during Covid-19, Int. Stud. Sociol. Educ. 32 (2) (2023) 487–510.
[69] F. Pierri, L. Luceri, N. Jindal, E. Ferrara, Propaganda and misinformation on Facebook and Twitter during the Russian invasion of Ukraine, in: Proceedings of the 15th ACM Web Science Conference 2023, 2023, pp. 65–74.
[70] B. Takahashi, E.C. Tandoc, C. Carmichael, Communicating on Twitter during a disaster: an analysis of tweets during Typhoon Haiyan in the Philippines, Comput. Hum. Behav. 50 (Supplement C) (2015) 392–398.
[71] D. Buscaldi, I. Hernandez-Farias, Sentiment analysis on microblogs for natural disasters management: a study on the 2014 Genoa floodings, in: Proceedings of the 24th International Conference on World Wide Web, WWW'15 Companion, ACM, New York, NY, USA, 2015, pp. 1185–1188.
[72] S. Rahman, N. Jahan, F. Sadia, I. Mahmud, Social crisis detection using Twitter based text mining-a machine learning approach, Bull. Electr. Eng. Inform. 12 (2) (2023) 1069–1077.
[73] M. Rodríguez-Ibánez, A. Casánez-Ventura, F. Castejón-Mateos, P.-M. Cuenca-Jiménez, A review on sentiment analysis from social media platforms, Expert Syst. Appl. (2023) 119862.
[74] D. Tiwari, B. Nagpal, B.S. Bhati, A. Mishra, M. Kumar, A systematic review of social network sentiment analysis with comparative study of ensemble-based techniques, Artif. Intell. Rev. 56 (11) (2023) 13407–13461.
[75] B. Wan, P. Wu, C.K. Yeo, G. Li, Emotion-cognitive reasoning integrated bert for sentiment analysis of online public opinions on emergencies, Inf. Process. Manag. 61 (2) (2024) 103609.
[76] M. Pota, M. Ventura, H. Fujita, M. Esposito, Multilingual evaluation of pre-processing for bert-based sentiment analysis of tweets, Expert Syst. Appl. 181 (2021) 115119.
[77] A. Bello, S.-C. Ng, M.-F. Leung, A bert framework to sentiment analysis of tweets, Sensors 23 (1) (2023) 506.
[78] L. Zhang, H. Fan, C. Peng, G. Rao, Q. Cong, Sentiment analysis methods for hpv vaccines related tweets based on transfer learning, in: Healthcare, vol. 8, MDPI, 2020, p. 307.
[79] A. Zhao, Y. Yu, Knowledge-enabled bert for aspect-based sentiment analysis, Knowl.-Based Syst. 227 (2021) 107220.
[80] I.A. Farha, W. Magdy, A comparative study of effective approaches for Arabic sentiment analysis, Inf. Process. Manag. 58 (2) (2021) 102438.
[81] D. Van Thin, D.N. Hao, N.L.-T. Nguyen, Vietnamese sentiment analysis: an overview and comparative study of fine-tuning pretrained language models, ACM Trans. Asian Low-Resour. Lang. Inf. Process. 22 (6) (2023) 1–27.
[82] S. Elmitwalli, J. Mehegan, Sentiment analysis of cop9-related tweets: a comparative study of pre-trained models and traditional techniques, Front. Big Data 7 (2024) 1357926.
[83] C. Barrie, J.C. ting Ho, academictwitter: an r package to access the Twitter academic research product track v2 api endpoint, J. Open Sour. Softw. 6 (62) (2021) 3272, https://doi.org/10.21105/joss.03272.

[84] M. Ángeles Oviedo-García, M. Muñoz-Expósito, M. Castellanos-Verdugo, M. Sancho-Mejías, Metric proposal for customer engagement in Facebook, J. Res. Interact. Mark. 8 (4) (2014) 327–344.
[85] J.M. Pérez, D.A. Furman, L. Alonso Alemany, F.M. Luque, RoBERTuito: a pre-trained language model for social media text in Spanish, in: Proceedings of the Thirteenth Language Resources and Evaluation Conference, European Language Resources Association, Marseille, France, 2022, pp. 7235–7243.
[86] H.C. Wu, R.W.P. Luk, K.F. Wong, K.L. Kwok, Interpreting tf-idf term weights as making relevance decisions, ACM Trans. Inf. Syst. 26 (3) (Jun. 2008), https://doi.org/10.1145/1361684.1361686.
[87] N.V. Chawla, K.W. Bowyer, L.O. Hall, W.P. Kegelmeyer, Smote: synthetic minority over-sampling technique, J. Artif. Intell. Res. 16 (1) (2002) 321–357.
[88] Y.-H. Pao, G.-H. Park, D.J. Sobajic, Learning and generalization characteristics of the random vector functional-link net, Neurocomputing 6 (2) (1994) 163–180.
[89] P.A. Henríquez, G.A. Ruz, A non-iterative method for pruning hidden neurons in neural networks with random weights, Appl. Soft Comput. (2018), https://doi.org/10.1016/j.asoc.2018.03.013.
[90] P.A. Henríquez, G.A. Ruz, Extreme learning machine with a deterministic assignment of hidden weights in two parallel layers, Neurocomputing 226 (Supplement C) (2017) 109–116.
[91] P.A. Henríquez, G.A. Ruz, Noise reduction for near-infrared spectroscopy data using extreme learning machines, Eng. Appl. Artif. Intell. 79 (2019) 13–22.
[92] Q. Shi, M. Hu, P.N. Suganthan, R. Katuwal, Weighting and pruning based ensemble deep random vector functional link network for tabular data classification, Pattern Recognit. 132 (2022) 108879, https://doi.org/10.1016/j.patcog.2022.108879.
[93] S.M. Alhashmi, A.M. Khedr, I. Arif, M. El Bannany, Using a hybrid-classification method to analyze twitter data during critical events, IEEE Access 9 (2021) 141023–141035.
[94] D. Kowald, E. Lex, Studying confirmation bias in hashtag usage on Twitter, arXiv preprint arXiv:1809.03203, 2018.
[95] F. Morstatter, H. Liu, Discovering, assessing, and mitigating data bias in social media, Online Soc. Netw. Media 1 (2017) 1–13.
[96] J. Pfeffer, K. Mayer, F. Morstatter, Tampering with Twitter's sample api, EPJ Data Sci. 7 (1) (2018) 50.
[97] A. Hino, R.A. Fahey, Representing the twittersphere: archiving a representative sample of Twitter data under resource constraints, Int. J. Inf. Manag. 48 (2019) 175–184.
[98] J. Ah-Pine, E.-P. Soriano-Morales, A study of synthetic oversampling for Twitter imbalanced sentiment analysis, in: Workshop on Interactions Between Data Mining and Natural Language Processing (DMNLP 2016), 2016.