The final publication is available at

https://doi.org/10.1109/INTERCON52678.2021.9532665

Additional Information

Politics and Social Media: Reaction to the presidential debates during the Peruvian electoral campaign according to Twitter

Rosmery Ramos-Sandoval
Centre for Interdisciplinary Science and Society Studies
Universidad de Ciencias y Humanidades
Lima, Peru
rramos@uch.edu.pe

Amparo Blazquez-Soriano
Department of Economics and Social Sciences
Universitat Politècnica de València
Valencia, Spain
mdeblzso@upvnet.upv.es

Abstract— **This research explores sentiments of live-tweeting during the first 2021 Peruvian Presidential debate. We focus on citizens' social media engagement during political debates, because this may have the potential to affect opinion formation about candidates, as well as to encourage (or discourage) election participation. To reach this goal, we explored the contents of the tweets posted during the Peruvian Presidential debate with sentiment analysis techniques and then categorized them into as reflecting positive or negative feelings. Results show a different pattern in each of the three days of the Peruvian Presidential debate. Findings suggest that by means of analyzing social media data, there is a chance to better understand citizen's reactions and opinions in Peru regarding a political highlight, such as a presidential election.**

Keywords— Twitter, social media, presidential debates, sentiment analysis, Naïve Bayes.

I. INTRODUCTION

The opportunities and challenges of the Internet for citizens to access and participate in political discourses are major strands of discussion within the academic debate on the nature of contemporary democracy [1, Ch. 24]. In this regard, Social Networking Sites (SNS) are not only changing the supply side of politics but also the demand side by reshaping citizens political behaviour [2]. For instance, it has been pointed out that social media engagement during political debates has the potential to affect opinion formation about candidates and political issues, as well as to encourage election participation [3]–[6]. Therefore, there non-traditional media have an enormous potential to engage citizens into politics, to make their opinions more visible and to influence on such opinions. Indeed, SNS do not only fulfil the role of improving the access to information, but they may also represent a tool for inclusivity in the political dynamics of a country for an important part of the population.

Regarding the opportunities that social media and the Internet bring in the political field, previous research works from Ceron [7], [8] pointed to the increased relevance of social media in everyday politics, particularly from microblogging platforms like Twitter. In addition, Tusmajan et al. [9] described multiple advantages of extracting public opinions from microblogging content, such as obtaining user's unedited expressions under a greater variety of topics. Political opinions extracted from social media platforms have also been used to help in forecasting social and economic indexes [10]. Hence, the contents and feelings shared on Twitter could be an accurate reflect of the offline political and economic landscape. Furthermore, Twitter allows users to create their frames regarding political events and not simply to rely on the frames created by media organizations [11]. Therefore, the social media platform is a clear example of the opportunities that citizens have nowadays to be listened, not only by

politicians but also by their community. Thus, we could say that voters today could use better tools to avoid giving uninformed votes or support, while politicians can have first-hand information about the needs and requests of their electorate.

The important expansion of digital media favour candidates to be more open to confronting proposals during debates, since now voters are more susceptible not only to direct, but also to indirect political information. Delli Carpini et al. [12] pointed out the value of political deliberation for democracy, including its role for societal consensus-building, reducing conflict, conferring greater tolerance for opposing views, increasing political efficacy and encouraging engagement in public life. Benoit et al. [13] showed that the influence from political debates on voters' decisions may be direct and indirect, whereas, for candidates, the face-to-face confrontation of debates provides them with an opportunity to answer to opponents and get political yields. Furthermore, in that sense, the expansion of SNS has boosted users to share their real-time reactions, and thus, these reactions would not respond to a traditional post-debate political analysis of discourses by experts, but rather to nearly genuine feedback of the impressions of citizens.

A. Presidential Debates on Twitter

Since the first televised presidential debate in modern history, 1960 Kennedy-Nixon, research has revealed that televised debates serve for at least two purposes apart from providing information about the political proposals: a) helping voters to be more informed about the candidates, and b) helping political parties to get swing voters by emphasizing their candidate's strengths and their opponent's weaknesses. Meanwhile, Camaj [6] highlights the relevance of televised political debates as places to provide high deliberative quality, and those spaces have been enhanced with the SNS, that are opening the way to a new echo platform and different activism, as the "clicktivism".

In line with that, Hosch-Dayican et al. [2], who monitored Twitter users' activities during the 2012 Dutch Parliamentary election campaign, showed that Twitter appears to be particularly common for conducting electoral campaigning, particularly of two different types: Persuasive campaigning & negative campaigning. However, for Freelon and Karpf [14] the defining moments of the 2012 US Presidential debates offered a window into how media events and political humour mixed together within the hybrid media system. A gathering from these approaches was a recent study conducted by Robertson et al. [15] during the 2016 US Presidential debates, which showed that the majority of Tweets posted during the debate provided humorous and derisive commentaries, in addition to substantive criticism to the candidates and the debate process and mostly void of insulting remarks. These

works show, thus, the existing possibilities to real-time monitor the reactions to political debates on social networks.

Additionally, it is possible that the active and open participation of all users, whether or not they are actively immersed in the electoral contest, could be manifested and perceived through the emotionality of the texts posted on Twitter. This emotionality can been assessed through sentiment analysis (SA) classification techniques, which have the potential to create knowledge that could be employed by decision support systems and decision-makers [8], [16].

In this study, we explore the sentiments expressed on Twitter in relation to a series of political debates. For this purpose, a case study was designed to explore the predominant sentiments of Twitter users, in real-time, regarding the 2021 Peruvian Presidential Debates, particularly during the first round of elections.

The paper is organized as follows. Section 2 describes the case study. Section 3 discusses the method for data collection and data processing. Section 4 presents the exploratory results on citizen's reactions during the 2021 Peruvian Presidential Debate. Subsequently, Section 5 provides some conclusions.

II. 2021 PERUVIAN ELECTION CONTEXT

After an extremely conflictive five-year political period, with a total of four different presidents in Peru between 2017-2020, General Elections were scheduled to be held in April 2021 in compliance with the *Decreto Supremo* Number 122-2020-PCM [17]. In the race for the presidency in 2021, 18 candidates competed for the most important Peruvian political office (Table I).

Under the Peruvian legislation applicable to the electoral process, the presidential elections were going to be held under a two-round voting system. The first round took place on April 11, 2021, and if no candidate received an absolute voting majority, the second round would take place on June 6, 2021.

TABLE I.   Presidential candidates of 2021 Peruvian elections

Party	Party Candidate
1. Partido Nacionalista Peruano	Ollanta Humala
2. Frente Amplio por Justicia, Vida y Libertad	Marco Arana
3. Partido Morado	Julio Guzmán
4. Perú Patria Segura	Rafael Santos
5. Victoria Nacional	George Forsyth
6. Acción Popular	Jhony Lescano
7. Avanza País	Hernando De Soto
8. Podemos Perú	Daniel Urresti
9. Juntos por el Perú	Verónica Mendoza
10. Partido Popular Cristiano	Alberto Beingolea
11. Fuerza Popular	Keiko Fujimori
12. Unión por el Perú	José Vega
13. Renovación Popular	Rafael López Aliaga
14. Renacimiento Unido Nacional	Ciro Gálvez
15. Somos Perú	Daniel Salaverry
16. Perú Libre	Pedro Castillo
17. Democracia Directa	Andrés Alcantara
18. Alianza para el Progreso	Cesar Acuña

Note: Adapted from the Official Ballot Paper used in the first round of scheduled presidential elections.

III. METHODS

Given the exploratory nature of the research, this study designed and implemented a sentiment analysis model of Tweets that were posted in Spanish. The main goal was to measure public opinions regarding the Peruvian 2021 Presidential debates during the first round of presidential elections, using machine learning techniques.

A. Data Collection

Data were collected using the Streaming Twitter API tool, which is designed to collect data in real-time. In order to collect relevant tweets, we employed a sampling technique to gather relevant hashtags, following the logic of the snowball sampling method [2]. Therefore, the data gathering process proposed in this paper was done by specifying the query search from mentions and multiple hashtags, which were combined with the date-time specification, extracting tweets posted between March 29-31, 2021. A total of 106,931 tweets related to a list of 19 tags were extracted (Table II).

Regarding Table II, it also shows 7 hashtags for which tweets were collected each day, considering the official tag for the presidential debate *#DebatePresidencialJNE*, and the candidates' participation according to the official program proposed by the *Jurado Nacional de Elecciones* (JNE, by its Spanish acronym), a constitutional autonomous agency from the Peruvian State, throughout the consultation period.

TABLE II.   Hashtag tweets collected by day, March 29-31, 2021

Debate's day	Hashtag Archives	Number of Tweets archived by day	Time Frame
Day 1	#DebatePresdencialJNE #VeronikaMendoza #AlbertoBeingolea #KeikoFujimori #MarcoArana #CesarAcuña #GeorgeForsyth	10,001	03-29-21 19:22-20:00 (UTC-5)
Day 2	#DebatePresdencialJNE #PedroCastillo #DanielUrresti #AndresAlcantara #DeSoto #HDSoto #OllantaHumala #JoseVega	46,892	03-30-21 18:00-20:00 (UTC-5)
Day 3	#DebatePresdencialJNE #LopezAliaga #YohnyLescano #JulioGuzman #Salverry #RafaelSantos #CiroGalvez	50,038	03-31-21 18:00-20:00 (UTC-5)

B. Data Cleaning

The Twitter raw data extracted needed extensive cleaning. The first step was the pre-processing, in order to convert words to vectors. This step has the aim to reduce noise in the text and to improve the performance and accuracy of the sentiment classification. The Python's library Natural Language Toolkit (NLTK) was used, applying the following techniques to explore the text content [18]:

- Tokenization, which is used to break the text down into words and symbols.
- Stop words removal, to delete common words in the given language that do not have an important meaning.
- Stemming, which is a task used to transform words into their root words.

The Natural Language Processing (NLP) analyses used in the tokenization process, allow us to exclude the presence of special Twitter tokens like user names, hashtags, URLs, emoticons (e.g., @JNE_Peru; #DebatePresdencial; ☺; https://) and punctuations (e.g., "¿"; "…"). Whereas stop words and stemming were employed as algorithms to remove

common words (e.g., "ese", "esta") and morphological affixes from words, leaving only the stems of the words.

C. Lexical normalization

Given that tweets extracted were posted in Spanish, the pre-processing task was very relevant due to its multilingual characteristics [19]. After the conversion of the texts of the hashtags' datasets obtained during the 2021 Peruvian Presidential Debates each token is passed through a set of basic modules imported from a Corpus previously trained. The language model was estimated from the TASS Corpus generated by Villena-Román et al. [20], which contained over 70,000 tweets written in Spanish between November 2011 and March 2012, with topics related to politics, economics, communication, mass media, and culture.

D. Polarity Classification

Assessing the content of tweets becomes essential for understanding the patterns of electoral campaigning [2]. This research performs subjectivity classification on individual sentences, obtained from people's opinions posted on Twitter, during the 2021 Peruvian Presidential Debate. In line with this, we employed a machine learning method [21], applying a binary classification task of labelling an opinionated text that classifies tweets as positive or negative.

The development of the classification scheme was conducted under the Naïve Bayes technique, which is a method of classification that uses the probability theory to find the most likely of the possible classifications. Naïve Bayes classifiers, gives us a way of combining the prior probability and conditional probabilities in a single formula, which we can use to calculate the probability of each of the possible classification in turn [22]. Therefore, under the independent assumptions of the Bayes Theorem (1), it can find the probability of occurrence of c, since t has occurred:

$$P(c|t) = \frac{P(t|c)P(c)}{P(t)} \qquad (1)$$

Under the Naïve Bayes technique, this research designs a model to find the probability of the assignment of a sentiment class c, which might be positive (☺) or negative (☹), given the occurrence of a tweet t, assuming that the predictors and features are independent and that the presence of one particular feature does not affect the other. Then, this learning method takes the training set m as input and returns the learned classification function (Fig. 1).
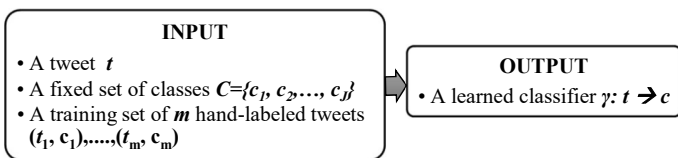

Fig. 1. The Naïve Bayes Classification modelled.

IV. RESULTS

The exploratory approach to finding trends from Twitter user' real-time dominant sentiments during the 2021 Peruvian Presidential Debate, allowed us to address some frequency of live users commentaries, which may contribute to drawing a first impression about how each candidate performed during the debates. In line with this, Table III shows erratic hashtag frequencies, identified during each debate's day.

Although the official label proposed by the official organisation *"#DebatePresidencialJNE"* is predominant, among the hashtags proposed in Table II, we can identify that the predominant labels are suggested based on the candidate's performance and spontaneously proposed by the users.

TABLE III. Frequency of hashtags tweeted by citizens during debates, March 29-31, 2021

Debate's day	Top 10 hashtags	Number of hashtags
Day 1	#DebatePresidencialJNE	7,377
	#TodosConAcu	961
	#Acu	361
	#DebateBicentenario	199
	#De	103
	#DebatePresiden	83
	#Elecciones	74
	#Debate	74
	#ElCambioEsForzay	66
	#MejorBeingolea	66
Day 2	#DebatePresidencialJNE	25,426
	#DebatePresidencial	13,086
	#Deb	590
	#Elecciones	498
	#perudecide	480
	#HernandoDeSoto	476
	#OllantaPresidente	316
	#Debate	313
	#DebatePresidencialJN	202
	#DeSoto	193
Day 3	#DebatePresidencialJNE	30,510
	#DebatePresidencial	11,066
	#LopezAliagaDeudor	1,428
	#VotaMorado	1,053
	#DebateElectoralJNE	704
	#LopezAliagaHablaHuevadas	539
	#Porky	522
	#Elecciones	492
	#JulioPresidente	476
	#Debat	450

NOTE: Day 1=March 29; Day 2=March 30; Day 3=March 31.

A. 2021 Peruvian Presidential Debate's sentiments on Twitter

The purpose of this research was to focus on the contents of tweets posted during the 2021 Presidential Debates, to provide a more detailed appreciation of the citizens' response about the Peruvian Presidential Candidates performance and proposals. For this analysis, we proposed to assess the citizens' emotional reactions through the extraction and analysis of tweets of each debate's day between March, 29-31.

The polarity examination model of individual debate's day tweets had an accuracy rate of 80.47%, showing an optimal performance in terms of classification under the Naïve Bayes approach. Regarding sentiment classification, polarity classification scores of Twitter conversations shows variations in trends between each debate. Fig. 2 shows rating scores obtained from tweets posted during the Debate's Day 1 (Positive=36.17%; Negative=66.83%), with a visible polarisation of sentiments, where negative sentiments are prevalent over positive ones on this day. Fig. 3 shows a balanced polarisation classification, indeed visibly neutral in the case of Day 2 (Positive=49.85%; Negative=50.15%). Finally, tweets posted on Day 3 (Positive=52.71%; Negative=47.29%) were positioned slightly towards a positive position (Fig. 4).

The most interesting findings concerning the tweets posted during the three Presidential Debates, are related to the potential to identify tweet sentiments associated with citizens likely to vote to candidates of political parties on the debate.
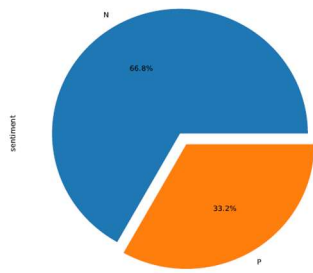
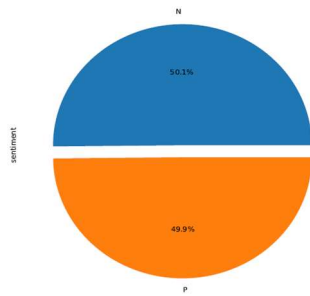Fig. 2. Distribution of tweet sentiments classification during Day 1.

Fig. 3. Distribution of tweet sentiments classification during Day 2.
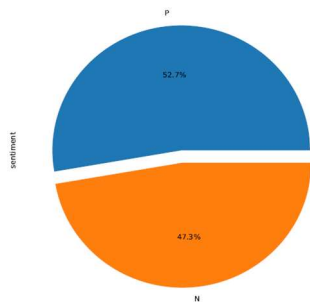

Fig.4. Distribution of tweet sentiments classification during Day 3.

V. CONCLUSIONS

The present study has focused on exploring the sentiments related to the Peruvian Presidential Debates by means of extracting and analysing the comments that were posted on Twitter in real-time. The results from this analysis show a different pattern in each of the three days of the Peruvian Presidential Debate. When examining the tags of each debate's day, based on terms of frequency we may identify low levels of parties mentions by day. Therefore, a relationship between the tags frequencies and sentiments identified in the Twitter posts during the Peruvian Presidential debate, might not be clearly defined.

To explore the sentiments of Twitter users' discussions, would be helpful as an information driver of general public opinion about Candidate's political agendas exposed during the televised debate. From our empirical sentiment analysis model, we identified more negative associations between the candidates and Twitter users on the first day of the debate, in contrast with the reactions posted during the third day of the debate. This may point out to the individual polemic associated with the candidates who participated on the first and third day. In contrast, the observed moderation on the second debate day might be associated with the lack of visibility in opinion polls of some of the candidates who participated in that session.

In future studies, we plan to apply a predictive approach to provide more insights and test the predictability of the electoral results in Peru given the data extracted from SNS.

REFERENCES

[1] K. Weller, A. Bruns, J. Burgess, M. Merja, and P. Cornelius, Twitter and Society, vol. 52, no. 02. Bern, Switzerland: Peter Lang, 2014.

[2] B. Hosch-Dayican, C. Amrit, K. Aarts, and A. Dassen, "How Do Online Citizens Persuade Fellow Voters? Using Twitter During the 2012 Dutch Parliamentary Election Campaign," Soc. Sci. Comput. Rev., vol. 34, no. 2, pp. 135–152, 2016.

[3] A. Chadwick, B. O'Loughlin, and C. Vaccari, "Why People Dual Screen Political Debates and Why It Matters for Democratic Engagement," J. Broadcast. Electron. Media, vol. 61, no. 2, pp. 220–239, 2017.

[4] J. B. Houston, J. Hawthorne, M. L. Spialek, M. Greenwood, and M. S. McKinney, "Tweeting During Presidential Debates: Effect on Candidate Evaluations and Debate Attitudes," Argumentation and Advocacy, vol. 49, no. 4, pp. 301–311, 2013.

[5] J. A. Gottfried, B. W. Hardy, R. L. Holbert, K. M. Winneg, and K. H. Jamieson, "The changing nature of political debate consumption: Social media, multitasking, and knowledge acquisition," Polit. Commun., vol. 34, no. 2, pp. 172–199, 2017.

[6] L. Camaj, "Real time political deliberation on social media: can televised debates lead to rational and civil discussions on broadcasters' Facebook pages?," Inf. Commun. Soc., vol. online, pp. 1–18, 2020.

[7] A. Ceron, Social Media and Political Accountability: Briding the gap between citizens and politicians. 2017.

[8] A. Ceron, L. Curini, S. M. Iacus, and G. Porro, "Every tweet counts? How sentiment analysis of social media can improve our knowledge of citizens' political preferences with an application to Italy and France," New Media Soc., vol. 16, no. 2, pp. 340–358, 2014.

[9] A. Tumasjan, T. O. Sprenger, P. G. Sandner, and I. M. Welpe, "Election forecasts with Twitter: How 140 characters reflect the political landscape," Soc. Sci. Comput. Rev., vol. 29, no. 4, pp. 402–418, 2011.

[10] C. Colonescu, "The effects of Donald Trump's Tweets on US Financial and Foreign Exchange Markets," Athens J. Bus. Econ., vol. 4, no. 4, pp. 375–388, 2018.

[11] J. Hawthorne, J. B. Houston, and M. S. McKinney, "Live-Tweeting a Presidential Primary Debate: Exploring New Political Conversations," Soc. Sci. Comput. Rev., vol. 31, no. 5, pp. 552–562, 2013.

[12] M. X. Delli Carpini, F. Lomax Cook, and L. R. Jacobs, "Public deliberation, discursive participation, and citizen engagement: A review of the empirical literature," Annu. Rev. Polit. Sci., vol. 7, pp. 315–344, 2004.

[13] W. L. Benoit, G. J. Hansen, and R. M. Verser, "A Meta-analysis of the Effects of Viewing U.S. Presidential Debates," Commun. Monogr., vol. 70, no. 4, pp. 335–350, 2003.

[14] D. Freelon and D. Karpf, "Of big birds and bayonets: hybrid Twitter interactivity in the 2012 Presidential debates," Inf. Commun. Soc., vol. 18, no. 4, pp. 390–406, 2015.

[15] C. Robertson, W.H. Dutton, R. Ackland, and T-Q. Peng, "The democratic role of social media in political debates: The use of Twitter in the first televised US presidential debate of 2016", J. Inf. Technol. Politics, vol. 16, no. 2, pp. 105-118.

[16] F. A. Pozzi, E. Fersini, E. Messina, and B. Liu, Challenges of Sentiment Analysis in Social Networks: An Overview, vol. 1. Elsevier Inc., 2017.

[17] Diario Oficial El Peruano, DECRETO SUPREMO No 122-2020-PCM, no. 2. 2020, pp. 7–8.

[18] S. Bird, E. Klein, and E. Loper, Natural language processing with Python: analyzing text with the natural language toolkit. O'Reilly Media, Inc., 2009.

[19] K. Dashtipour, S. Poria, A. Hussain, and E. Cambria, "Multilingual Sentiment Analysis : State of the Art and Independent Comparison of Techniques," Cognit. Comput., vol. 8, no. 4, pp. 757–771, 2016.

[20] J. Villena-román, E. Martínez-cámara, S. Lana-serrano, and J. C. González-cristóbal, "TASS - Workshop on Sentiment Analysis at SEPLN TASS - Taller de Análisis de Sentimientos en la SEPLN," Proces. del Leng. Nat., vol. 50, pp. 37–44, 2013.

[21] B. Pang and L. Lee, "Opinion mining and sentiment analysis," Found. Trends Inf. Retr., vol. 2, no. 1–2, pp. 1–135, 2008.

[22] M. Bramer, Principles of data mining, vol. 180. Springer, 20