Identifying the Political Tendency of Social Bots in Twitter Using Sentiment Analysis: A Use Case of the 2021 Ecuadorian General Elections

Andres Quelal(✉) , Juan Brito , Mateo S. Lomas , Jean Camacho ,
Argenis Andrade , and Erick Cuenca

Yachay Tech University, Urcuquí, Ecuador
{andres.quelal,juan.brito,mateo.lomas,jean.camacho,
argenis.andrade,ecuenca}@yachaytech.edu.ec

Abstract. Sentiment analysis of social network data increasingly represents the real political scenario of many countries, which has turned bots into a powerful tool of influence, mainly due to their high efficiency. This work analyzes the messages on Twitter during the 2021 Ecuadorian presidential elections to determine sentiments and bots detection. We obtained a sample of 35,242 tweets corresponding to each candidate's first and second rounds. Our methodology consists of four phases: first, we perform data collection using the Twitter API; secondly, we preprocess the data; in the third phase, we perform sentiment analysis of the content of the tweets to understand their posture towards a candidate, and finally, we classify the users as bots or not. As a result, we discovered that bots and non-bots people on both sides had more positive feelings towards their respective candidates than unfavorable feelings against the other candidates.

Keywords: Sentiment analysis · Twitter bots · Political tendency · Social media

1 Introduction

Every year, there is a 9% growth in the number of social media users, and half of the internet traffic consists primarily of bots [17]. Part of the content of social media is composed of false or misleading news reports, hoaxes, conspiracy theories, click-bait headlines, junk science, and even satire [18]. In Ecuador, this is not the exception and more important, many of the relevant issues for the general population are received and discussed on social networks. For instance, the most followed users on Twitter in Ecuador respond to a localized and public profile, which means that the leading accounts in the country react to mainly national interests [3].

Although social media communication does not suppose any problem, there is the possibility of massive misinformation and conflicts generated by political

and economic interests. These conflicts and the spread of false news often do not only originate from a malicious person or group of people but also respond to a sophisticated set of technologies that include specialized bots that pose as ordinary users through fake accounts.

Twitter's popularity is extensive, giving facilities to do publications through bots, which has reached problems in the platform [4,7]. These social bots have an outsized role in disseminating articles from low-trust sites. The widespread dissemination of digital disinformation has been seen as a severe danger to democratic institutions [18].

Bots include programmed instructions to communicate in digital environments to accomplish tasks such as spam generation, blocking exchange points, launching denial of service attacks, deploying and replicating messages, publishing news, updating feeds, programming malware, phishing, and fraud clicks [16]. In the case of Twitter, many of them post directly through its Application Programming Interface (API). Still, frequently, their publications are disseminated through automation services or applications. It is essential to mention that sometimes the bot profiles lack the account's basic information, such as the username or profile photos [16]. Political bots, for example, are often used in conjunction with three types of political events: elections, scandals turn, and national security crises. Using bots during these situations aims to achieve simple goals such as filling the candidate's "followers" list or complex purposes such as harassing human rights activists or demobilizing citizens [16]. Due to its importance in citizen conversations, Twitter has become the preferred object of studies on the construction of public opinion in Ecuador [5].

If we look deeper at Twitter's role in Ecuador, it respond to mainly national and popular interests, ranging from politics to entertainment [3]. Also, the results of a study by [5] show a close relationship between the cyber-media agenda and the trending topics on Twitter in political and sports content. That wide is the scope of Twitter that during the second round of the presidential elections of Ecuador in 2017, automated accounts or Twitter bots played a central role in positioning campaign hashtags [16]. Taking all this into account, we can see that the utilization of Twitter bots in Ecuador is widespread.

This paper aims to analyze the political trend of tweets in Ecuador during the period of the 2021 presidential elections. This analysis is intended to use sentiment analysis and bots detection techniques. The results are analyzed using various visualizations to represent the political trend in this period. The details of the implementation can be found in the following Google Colaboratoy[1].

2   Related Work

Many works have already studied if social bots on Twitter or other social media have a particular influence over public opinion on politics, science, or different polemic topics. For instance, Pastor-Galindo et al. [15] analyze the impact of bots on Spain's elections during the 2019 campaign period and emphasize specific dates where activity was higher. An important aspect of this work is the

---

[1] https://shorturl.at/dftz6, last access: August 2022.

methodology the authors implement to spot the bots on Twitter and realize if they influenced the elections. Figure 1 shows the methodology adopted by the authors. It shows a pipeline divided into three main processes: data collection, data analysis, and knowledge extraction.

The data collection first sets the query parameters to obtain the tweets from those events related to those topics with a crawler and harvester. Then, the data analysis tests this processed data and the feature discovered over multiple options. This leads to an augmented data set with the individual evaluation of the sentiment analysis. In the final step, they do the knowledge extraction by using this augmented data set on a supervised learning technique to classify their political inclination, whether they are humans or not. Using an unsupervised learning approach, they analyze the friendship graphs, the whole pre-processed data, and the augmented data set they got. All of it lets them identify the possible presence of bots.
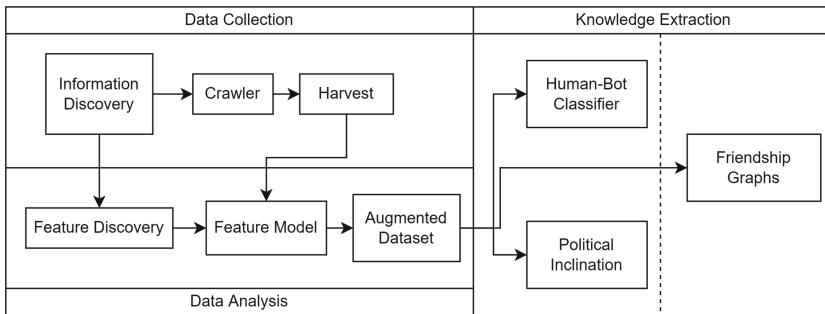

Fig. 1. Research methodology adopted on [15] refereed to elections in Spain.

2.1   Sentiment Analysis

A way to understand the content of users' tweets is text analysis through sentiment analysis. It involves studying tweets' opinions, sentiments, attitudes, and emotions to understand the behavior on social networks of a relevant or trending topic. In Computer Science, there is an area concerned with providing computers with the ability to understand the text and the context of words, called Natural Language Processing (NLP). This area aims to process human language, either speech or text. Sentimental Analysis is part of NLP to understand the writer's purpose, feelings, or emotion from a text.

Many works and papers are dedicated to analyzing sentiment from a tweet's text. For instance, Ibrahim et al. [13] presented a work centered mainly on the sentimental analysis to predict presidential elections. In this work, the authors highlight the importance of cleaning those tweets that computer bots, paid users, and fanatic users could generate. All these kinds of tweets are considered noise and difficult to predict. They use a technique to divide the tweets into sub-tweets

using limiters, such as commas, points, question marks, etc. They associate the sub-tweets to the respective politics using their words or names. This score represents the sentiment evaluation; the sub-tweets can be classified as positive or negative to the politician with an associated tweet. Also, using the positive sub-tweets only tends to get more accurate results in predicting any behavior, in this case, who will win elections. This work's value leads principally to how the authors process the data, where phrases get associated with an emotion and a politician. It is mentioned that bots usually talk well only of one of the politicians and bad about the rest.

2.2   Bots Detection

There are some ways to classify/detect if a user is a bot. One technique is using the universal score distribution. On a range [0,1], this score evaluates how likely an account is to be a human or a bot, where 1 is more likely to be a bot and 0 a human. So it is possible to set a threshold to decide in what range we classify them as humans and in what range we classify them as bots. A good range for humans could be: $[0 \leq U_{score} \leq 0.85]$, where the range for bots will be $[0.85 < U_{score} \leq 1]$. This score is calculated based on polarity and subjectivity. Polarity gives us if the sentiment is positive or negative and a value.

There are multiple attempts to detect social bots using machine learning techniques. Some authors use "Blacklists" [21] to extract features of tweets generated by bots and then pass these features to a Decorate classifier [12]. Others prefer comparing the results obtained with more traditional techniques, such as Decision Trees, Random Forest Algorithm, k-Nearest Neighbor Algorithm, Support Vector Machine (SVM), Logistic Regression, Neural Networks, and Naive Bayes Classifier [1,2,6,9,14,17,20]. Moreover, other studies combine some of these previous techniques in the denominated Ensemble Learning, obtaining better results than using only one of them [10,19]. For instance, Lingan et al. [11] proposed using Deep Q Learning for detecting social bots and influential users in online social networks providing a 5–9% improvement of precision over other existing algorithms. Furthermore, different approaches compare probabilistic techniques (Approximate Entropy, Sample Entropy) along with machine learning for detecting automated behavior on Twitter [8]. Most of the results of these works may also be used to analyze the role of social bots in the context of presidential elections.

3   Methodology

3.1   Data Collection

Data is available from the Twitter platform to request objects or fields such as tweets, users, spaces, lists, media, polls, and locations through its API[2]. Considering the user's information, we can obtain various attributes, such as id, a

---

[2] https://developer.twitter.com/en/docs/twitter-api, last access: August 2022.

screen name (used to communicate online), description, URL, verified (if the
user is authenticated) location, list of followers, list of following, list of favorite
(used for liked tweets).

The dataset considers the topic's selection, description of the data, and acqui-
sition time. Ecuador Elections 2021 is the input request topic, where the pres-
idential candidates Guillermo Lasso (CREO political party) and Andrés Arauz
(UNES political party) are the prominent mentions. We also collect tweets for
the vice-presidential candidates' Alfredo Borrero and Carlos Rabascall for CREO
and UNES political parties, respectively. The first and second round of the pres-
idential elections from November 30, 2020 to February 2, 2021 is the acquisition
period of the dataset. Table 1 shows the query parameters used to collect the
dataset.

Table 1. Parameters used in the querys to obtain the dataset of tweets.

Candidate	Guillermo Lasso	Andrés Arauz
Keywords for querys	GuillermoLasso	AndrésArauz
	LassoGuillermo	ArauzAndrés
	"Guillermo Lasso"	"Andrés Arauz"
	Lasso	Arauz
	"Alfredo Borrero"	"Carlos Rabascall"
	Borrero	Rabascall
	AlfredoBorrero	CarlosRabascall
	BorreroAlfredo	RabascallCarlos
Start time 1st round	2020-12-15 T17:00:00Z	2020-12-15 T17:00:00Z
End time 1st round	2021-02-07 T23:30:00Z	2021-02-07 T23:30:00Z
Start time 2nd round	2021-02-07 T17:00:00Z	2021-02-07 T17:00:00Z
End time 2nd round	2021-04-12 T23:30:00Z	2021-04-12 T23:30:00Z
Tweet fields	id	id
	text	text
	created_at	created_at
	author_id	author_id
	public_metrics	public_metrics

The number of tweets generated in one day with the theme Elections of
Ecuador in 2021 was enormous, so obtaining all the data for its respective
analysis became unrealistic considering the available computational limitations.
The solution to this problem was obtaining a certain number of daily tweets.
Although it considerably biases the results, it does not remove the possibility of
analyzing and drawing accurate conclusions. The decision was made to obtain
around 400 tweets per day. These tweets will correspond to each candidate's first
and second rounds. A total of 35,242 tweets were collected. The results where
stored in a CSV file.

3.2   Data Pre-procesing

The preprocessing and data cleaning process provides a balanced data set. Object attributes such as text were processed using NLP techniques. Tweets' attributes were converted into a usable format for sentiment analysis and bot recognition. For this purpose, data processing methods such as:

– **Punctuation's marks removal:** Twitter messages often contain symbols, numbers, and punctuation such as: $'!"\#\$\ \&\ \backslash'()^{*+,-.1;;}\Leftrightarrow\Rightarrow?@[11]^\wedge-\{|\}\sim1$. These preprocessed entities reduce ambiguous and unnecessary expressions for our dataset. All of these punctuation marks were removed using an NLP library. Also, HTML references, mentions, and hashtags were cleaned from our dataset.
– **Tokenization:** The tokenization task aims at splitting a text stream into smaller units called tokens. Tokens are composed of words, phrases, or other meaningful elements that can show a trend of the most common words found in our dataset. For example, the text: *"Durante las elecciones de este 7 de febrero, recuerda cumplir con los protocolos de bioseguridad establecidos."* will become as:

   ['Durante', 'elecciones', 'febrero', 'cumplir', 'protocolos', 'bioseguridad']

– **Stopwords removal:** Some tweet words do not have a significant influence on the sentence. Stopword removal removes common and frequent irrelevant words in our dataset using the NLTK python library.

3.3   Sentiment Analysis

We used Python libraries such as NLTK, specifically TextBlob, to compute the sentiment score. TextBlob is a library that allows complex analysis and operations on textual data.

3.4   Bots Detection

For bot detection, it was used the Botometer platform[3]. However, the API has limitations on the request per day on its free version; nevertheless, the way to detect if an account is a bot or human was the same with other libraries.

4   Results and Discussion

4.1   Statistical Information

Figures 2a and 2b show that the number of accounts that get less than 20 interactions is more than the 70%. In the first and second rounds, we can appreciate the users' interactions do not have a uniform distribution, even though most get

---

[3] https://botometer.osome.iu.edu, last access: August 2022.

20 or fewer actions (tweet, RT, like). Also, it could be expected that get more interactions on Fig. 2b than on Fig. 2a because, on the second round, tension could be even higher than in the first round. Still, accounts from both political sides got similar behaviors.

If we take the average of the sum of all the different interactions (retweet, reply, like, quote) of the bots per game, as reflected, convincing results are not appreciated. The results obtained are generally biased by obtaining a small data set. Many of the possible interactions that bots and people, in general, could have will not be reflected. It is estimated that, on average, there are 2,000 tweets every 10 min; our dataset does not even represent 1% of the entire data set. Another limitation was the fact that the Botometer has restrictions on the number of requests that we can obtain. In this case, it is limited to 500 requests per day; in general, resource limitations prevent us from getting reliable results.
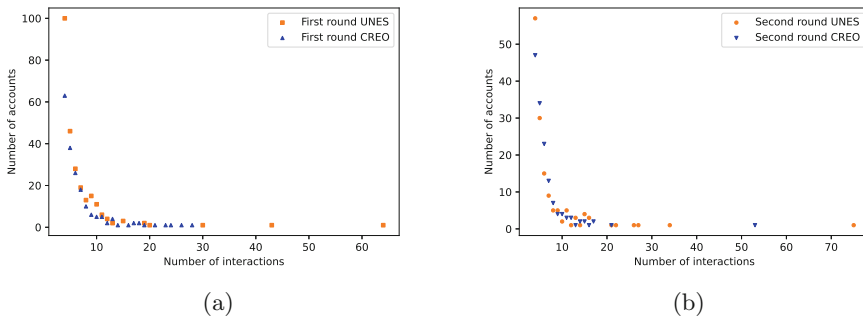

Fig. 2. Distribution of accounts with number of interactions in (a) first round and (b) second round.

4.2   Word Cloud Analysis

A practical way to explore the dataset's content is using a Word Cloud visualization. It is a visual representation object for word processing, which shows the frequency of words. For example, our dataset contains reference tweets of two presidential candidates. In Fig. 3a, the Word Cloud representation gives us a better approximation of user opinions in general. Word Cloud helps us understand the users' behavior, where the most used word was "Lasso". In the sentimental analysis, we checked this trend for each candidate. In Fig. 3b, the Arauz word cloud gives us that the most common word was "Andrés Arauz". Some word in this word cloud shows us words controversial events that happened to the candidate.

(a)   (b)

Fig. 3. Word Cloud representations of opinions about (a) Guillermo Lasso and (b) Andrés Arauz.

4.3   Sentiment and Polarity Score

Figure 4 shows the volume of tweets per sentiment for every political party. We can see that both parties have a significant volume of positively related sentiments. But the "Neutral" sentiment is as prevalent as positive sentiments, we can see negative sentiments towards parties, but they are not significantly larger than the others.

Table 2 shows in percentages how positive and negative emotions are present in both parties and rounds. They are above 40%, which is an excellent parameter for determining tendencies and intent to vote for that candidate. Both do not differ much, but we must analyze more data to distinguish between parties comprehensively.

Table 2. Sentiment analysis for both candidates in both rounds

Candidate	First round		Second round	
	Positive sentiment	Negative sentiment	Positive sentiment	Negative sentiment
Andrés Arauz	41.99%	23.22%	41.31%	21.35%
Gillermo Lasso	42.44%	22.79%	40.24%	23.60%

In Fig. 5, the polarity score shows a better understanding of user behavior in all presidential elections. Based on the polarity categorization, the scores were classified such that if the score is less than zero, the sentiment is negative, if the score is equal to zero, the sentiment is neutral, and if the score is greater than zero, the sentiment is positive. In relevant events, the decrease in polarity score shows us that users have a negative tendency at this stage. The positive polarity score varies for each stage. The overall trend varies a lot for each date, but it gives us a better understanding of how public opinion was.
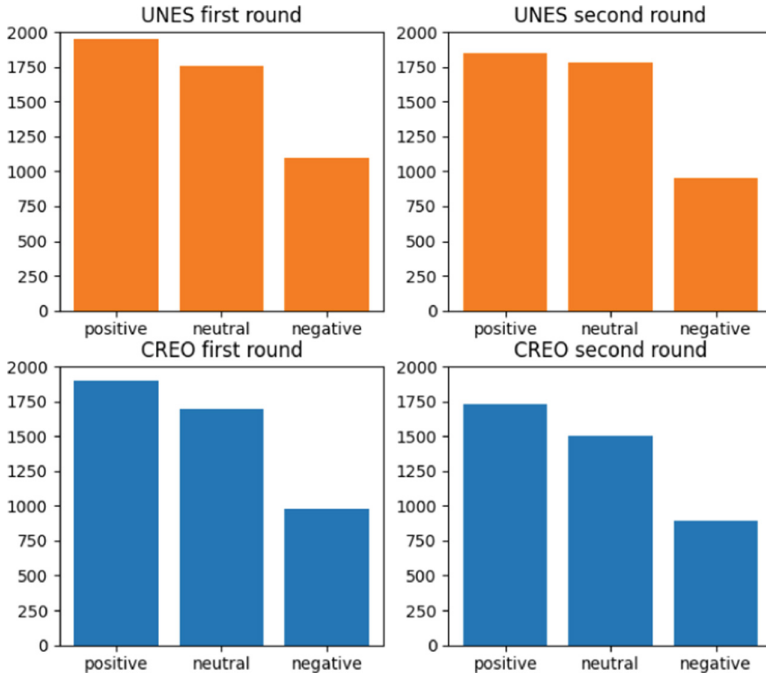
Fig. 4. Volume of tweets per sentiment

4.4   Bots Detection Results

We decided to use Botometer, which is an API that is specialized in the detection of bots. A limitation was the number of daily requests. We split the data set to get a sample to reach some results. There, we got the number of interactions, the politic they are with, and based on the number of interactions, it is viable to infer if they had any relevant participation.

The number of interactions for Andrés Arauz was 8,493 and for Lasso 8,029. In Figs. 6a and 6b, we can appreciate that in different tweets with a certain periodicity, there are some publications with many more interactions. This can represent the publications that turned viral, and as much as Lasso and Arauz, we got a similar number of tweets with more than 4,000 interactions.

Based on that, considering the original data set was of 35.242 tweets, only those users with more than three interactions and a threshold of 0.85, where those users with a score bigger than that were considered candidates to be bots. So we got 17 possible bots: 3 tend to support Guillermo Lasso, and 14 support Andrés Arauz (See Fig. 7). We have to consider that these detected bots are not from the total users of the whole dataset used but instead from a reduced sample. Eventually, this does not say anything about who candidate got more bots, but with these bots spotted, it is possible to look for how many times they interacted.
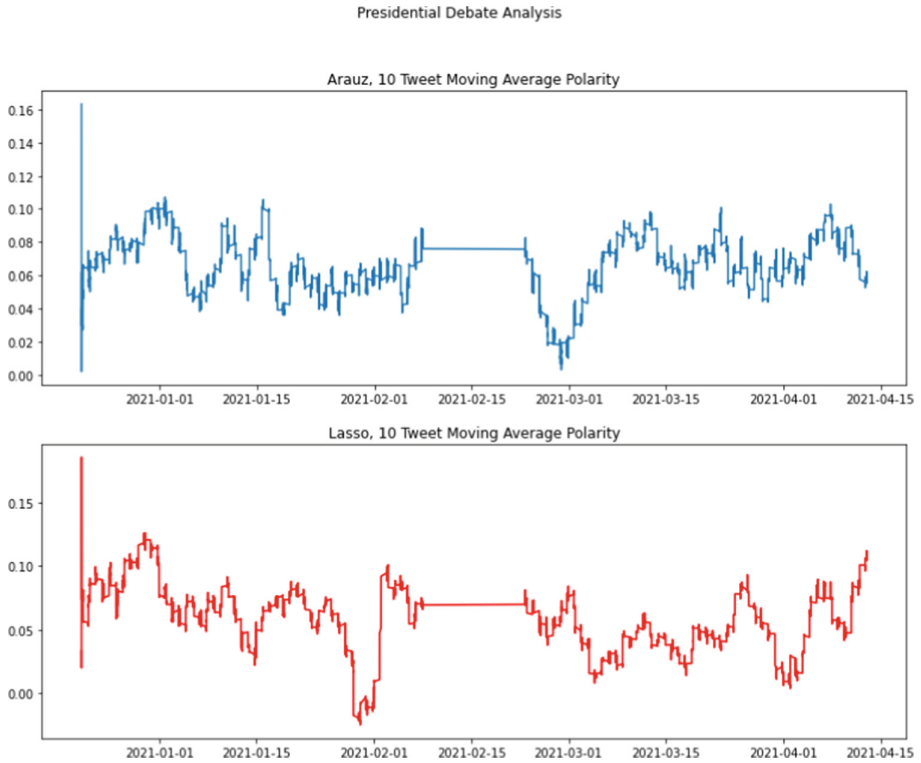
Presidential Debate Analysis

Arauz, 10 Tweet Moving Average Polarity

Lasso, 10 Tweet Moving Average Polarity

Fig. 5. Presidential data analysis/polarity scores.
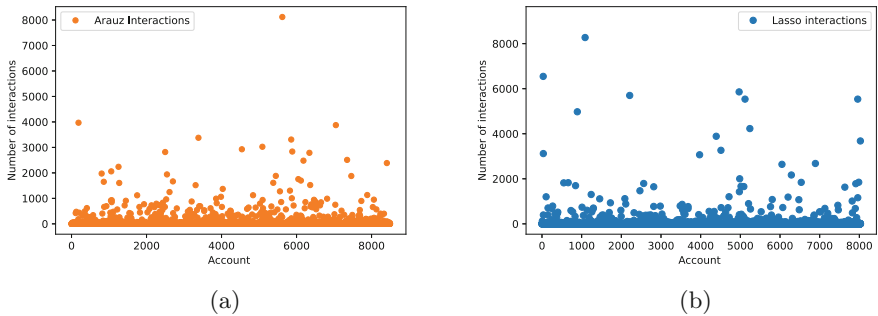
Fig. 6. Interactions in a certain tweet by (a) Guillermo Lasso and (b) Andrés Arauz.

In the same way, as in Figs. 2a and 2b, we got the total interactions only of the bot accounts, comparing the amount of interactions. Eventually, a bot tends to get a superior number of interactions in contrast to the people's average interactions; this can be interpreted as a way of influence. Seventeen bots are too few, but those can create a ton of movement on the network and have a

direct influence over viral publications; because of that amount of iterations, we can say they get some relevant influence.
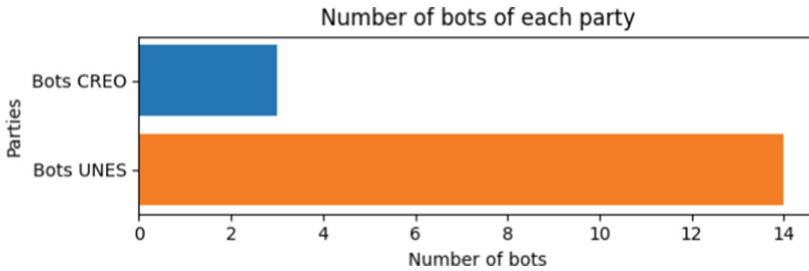

Fig. 7. Number of bots detected of each Party

5   Conclusion and Future Work

This paper presents a sentiment analysis of Twitter users during the 2021 Ecuadorian Presidential Race. It contains an intriguing examination of user sentiments, the potential that these users are bots, and how these sentiments relate to the official votes received by presidential contenders.

We obtained positive sentiments toward both candidates Guillermo Lasso (CREO political party) and Andrés Arauz (UNES political party) that were more significant in both rounds. We can say that the bots used from each side focused more on speaking good things about their supported candidate than speaking against the opposite candidate. Also, the influence of bots can vary where most bots have a certain amount of interactions, not that far from the number of interactions humans do. Still, a few bots have several interactions way more significant than the average. Based on the number of interactions, we can infer that those bots could be responsible for the vitality of certain publications.

For future work, we plan to try this methodology in a more extensive dataset. We could also apply this to a new electoral process before the final results are revealed to try to predict it.

References

1. Alothali, E., Hayawi, K., Alashwal, H.: Hybrid feature selection approach to identify optimal features of profile metadata to detect social bots in twitter. Soc. Netw. Anal. Min. **11**(1), 1–15 (2021). https://doi.org/10.1007/s13278-021-00786-4
2. de Andrade, N., Rainatto, G., Lima, F., Silva Neto, G., Paschoal, D.: Machine learning and bots detection on twitter. Int. J. Sci. Res. (IJSR) **8**, 001–011 (2019)
3. Barredo Ibáñez, D., Arcila Calderón, C., Barbosa Caro, E.: El perfil de los usuarios de Twitter más influyentes en Ecuador y la influencia del mensaje en la captación de seguidores. Observatorio **10**, 219–230 (2016). https://doi.org/10.15847/obsOBS10420161004

4. Chu, Z., Gianvecchio, S., Wang, H., Jajodia, S.: Detecting automation of twitter accounts: are you a human, bot, or cyborg? IEEE Trans. Dependable Secure Comput. **9**(6), 811–824 (2012). https://doi.org/10.1109/TDSC.2012.75

5. Coronel, P., García, J., Vera, M.: Twitter y la opinión pública en Ecuador: discursos, emisores y agendas. In: La Innovación de la Innovación: Del Medio al Contenido Predictivo. Actas del III Simposio Internacional sobre Gestión de la Comunicación (XESCOM 2018), pp. 697–713 (2018)

6. Deekshith, G.: Twitter bots detection using machine learning techniques. Int. J. Res. Appl. Sci. Eng. Technol. **9**, 1536–1541 (2021). https://doi.org/10.22214/ijraset.2021.36637

7. Edwards, C., Edwards, A., Spence, P., Shelton, A.: Is that a bot running the social media feed? Testing the differences in perceptions of communication quality for a human agent and a bot agent on twitter. Comput. Hum. Behav. **33**, 372–376 (2014). https://doi.org/10.1016/j.chb.2013.08.013

8. Gilmary, R., Venkatesan, A., Vaiyapuri, G.: Detection of automated behavior on twitter through approximate entropy and sample entropy. Pers. Ubiquit. Comput. (2021). https://doi.org/10.1007/s00779-021-01647-9

9. Khanday, A.M.U.D., Khan, Q.R., Rabani, S.T.: Identifying propaganda from online social networks during COVID-19 using machine learning techniques. Int. J. Inf. Technol. **13**(1), 115–122 (2020). https://doi.org/10.1007/s41870-020-00550-5

10. Kirn, S.L., Hinders, M.K.: Bayesian identification of bots using temporal analysis of tweet storms. Soc. Netw. Anal. Min. **11**(1), 1–17 (2021). https://doi.org/10.1007/s13278-021-00783-7

11. Lingam, G., Rout, R.R., Somayajulu, D.V.L.N.: Adaptive deep Q-learning model for detecting social bots and influential users in online social networks. Appl. Intell. **49**(11), 3947–3964 (2019). https://doi.org/10.1007/s10489-019-01488-3

12. Melville, P., Mooney, R.J.: Constructing diverse classifier ensembles using artificial training examples. In: Eighteenth International Joint Conference on Artificial Intelligence, pp. 505–510 (2003)

13. Mochamad, I., Omar, A., Alfan, W.F., Mirna, A.: Buzzer detection and sentiment analysis for predicting presidential election results in a twitter nation. In: 2015 IEEE International Conference on Data Mining Workshop (ICDMW), pp. 1348–1353 (2015). https://doi.org/10.1109/ICDMW.2015.113

14. Narayan, N.: Twitter bot detection using machine learning algorithms. In: 2021 Fourth International Conference on Electrical, Computer and Communication Technologies (ICECCT), pp. 1–4 (2021). https://doi.org/10.1109/ICECCT52121.2021.9616841

15. Pastor-Galindo, J., et al.: Spotting political social bots in twitter: a use case of the 2019 Spanish general election. IEEE Trans. Netw. Serv. Manage. **17**(4), 2156–2170 (2020). https://doi.org/10.1109/TNSM.2020.3031573

16. Puyosa, I.: Political bots on twitter in #Ecuador2017 presidential campaigns. Contratexto (27), 39–60 (2017). https://doi.org/10.26439/contratexto.2017.027.002

17. Ramalingaiah, A., Hussaini, S., Chaudhari, S.: Twitter bot detection using supervised machine learning. J. Phys. Conf. Ser. **1950**, 012006 (2021). https://doi.org/10.1088/1742-6596/1950/1/012006

18. Shao, C., Ciampaglia, G.L., Varol, O., Yang, K.C., Flammini, A., Menczer, F.: The spread of low-credibility content by social bots. Nat. Commun. **9**(1) (2018). https://doi.org/10.1038/s41467-018-06930-7

19. Shukla, H., Jagtap, N., Patil, B.: Enhanced twitter bot detection using ensemble machine learning. In: 2021 6th International Conference on Inventive Computation

Technologies (ICICT), pp. 930–936 (2021). https://doi.org/10.1109/ICICT50816.2021.9358734

20. Souza, S., Rezende, T., Nascimento, J., Chaves, L., Soto, D., Salavati, S.: Tuning machine learning models to detect bots on twitter. In: 2020 Workshop on Communication Networks and Power Systems (WCNPS), pp. 1–6 (2020). https://doi.org/10.1109/WCNPS50723.2020.9263756

21. Swe, M.M., Nyein Myo, N.: Fake accounts detection on twitter using blacklist. In: 2018 IEEE/ACIS 17th International Conference on Computer and Information Science (ICIS), pp. 562–566 (2018). https://doi.org/10.1109/ICIS.2018.8466499