

Los eventos electorales de diciembre 2020 y enero 2021 en Venezuela: un panorama en la red social Twitter

The electoral events of December 2020 and January 2021 in Venezuela: an overview on the social network Twitter

COLINA, Mairene ¹
COLINA, Katherine ²
FERNANDEZ, Alberto ³
MAIA, Manuel ⁴
TORREALBA, Daniela ⁵

Resumen

Se presenta un análisis de las opiniones de los usuarios de Twitter sobre acontecimientos de interés político-electoral ocurridos en Venezuela los días 6 diciembre 2020 y 5 enero 2021. Usando técnicas estadísticas y de inteligencia artificial se identificaron palabras y hashtag de mayor impacto, horas de mayor actividad, entre otros. Se detectó polarización en el sentir de los usuarios respecto a la Asamblea Nacional y la utilización de cuentas bots para influir en las tendencias de la red.

Palabras clave: minería texto, minería datos, twitter

Abstract

An analysis is presented on how Twitter users perceived the political-electoral related events of interest happening on December 6, 2020 and January 5, 2021 in Venezuela. Statistical and artificial intelligence methods were used to identify high impact words and hashtags, high activity time periods, among other. A notable polarization of the user's opinions about the Asamblea Nacional, and the use of bots accounts to skew tendencies were observed.

Key words: text mining, data mining, twitter

1. Introducción

Este estudio constituye una continuación del trabajo de revisión de los contenidos de la red social Twitter, en relación al tema electoral en Venezuela, promovido por la Organización No Gubernamental Red de Observación

¹ Profesora. Escuela de Matemática. Universidad Central de Venezuela. Venezuela. Email: mcolina@gmail.com

² Estudiante. Escuela de Computación. Universidad Central de Venezuela. Venezuela. Email: kathymcolina@gmail.com

³ Profesor. Escuela de Química. Universidad Central de Venezuela. Venezuela. Email: ferna.aj@gmail.com

⁴ Profesor. Escuela de Matemática. Universidad Central de Venezuela. Venezuela. Email: manuelpmaiaf@gmail.com

⁵ Profesora. Escuela de Matemática. Universidad Central de Venezuela. Venezuela. Email: danielatorrealbag@gmail.com

Electoral de la Asamblea de Educación (ROAE, www.redobservacion.org) y cuyos primeros resultados pueden verse en Colina y otros (2021).

Para la realización de este estudio, en primer lugar se descargaron los tweets que contenían el tema electoral venezolano los días 6 de diciembre de 2020, fecha de las elecciones parlamentarias (en lo sucesivo, 6D) y 5 de enero de 2021, fecha de instalación de la nueva Asamblea Nacional (en lo sucesivo, 5E). Para ello se hizo uso de una interfaz de programación de aplicaciones (API) de Twitter y códigos desarrollados en los lenguajes de programación Python y R. Posteriormente, se analizó el texto de estos tweets haciendo uso de técnicas estadísticas y de inteligencia artificial que facilitan el análisis de la información contenida en grandes volúmenes de datos y colecciones de documentos. Dichas técnicas se conocen como minería de datos y minería de texto. Para información más detallada sobre estas técnicas, se puede consultar Jo (2019).

Las herramientas de la minería de datos y de texto de las que los investigadores disponen hoy en día permiten con gran acierto identificar hechos y datos puntuales a partir de los contenidos de los tweets, agruparlos por contenido similar, determinar los temas que son tratados mediante la categorización automática de los textos, visualizar las relaciones entre los conceptos tratados, establecer correlaciones entre las palabras usadas en los textos de los tweets, detectar cuentas bots, entre otros aspectos.

Para este estudio se lograron descargar un total 84000 tweets, distribuidos de la siguiente manera:

- 17828 tweets emitidos el 6 de diciembre de 2020.
- 66172 tweets emitidos el 5 de enero de 2021.

Estos 84000 tweets constituyen una muestra representativa de las opiniones e informaciones vinculadas al tema electoral en Venezuela que circularon en la plataforma Twitter el 6D y el 5D. Lo primero a destacar es que el número de tweets emitidos el día 5E es casi cuatro veces mayor que los emitidos el día 6D.

Una vez pre procesados los tweets, se realizó en primer lugar un análisis exhaustivo de ambas fechas consideradas en conjunto, en el que se identificaron:

- a) las palabras más tuiteadas,
- b) los hashtags que recibieron más difusión,
- c) los diferentes temas abordados por los tuiteros y
- d) las palabras más correlacionadas con los vocablos *asamblea*, *elecciones* y *nueva*.

En segundo lugar se realizó un análisis más detallado donde se introducen otros elementos, pero esta vez considerando ambas fechas por separado, en el que se identificaron:

- a) las palabras más tuiteadas,
- b) el periodo del día que presentó mayor actividad,
- c) las cuentas bots que mostraron actividad.

2. Metodología

Para el manejo y procesamiento de la información se hizo uso de los lenguajes de programación Python y R, utilizando las facilidades de sus librerías y desarrollando códigos propios para fines específicos.

El esquema metodológico implementado sigue en primer término los siguientes procesos sucesivos, cuyos principios básicos están descritos en Colina y otros (2021):

- Extracción de la información a través de la API de Twitter.
- Pre procesamiento de la información, para disponer de data limpia y uniforme.
- Procesamiento de la información (minería de texto y datos):
 - Identificación de *hashtags* (etiquetas), para categorizar los tweets y determinar tendencias y mensajes divulgados.
 - Generación de bigramas, para identificar temas de conversación.

En segundo término, y para complementar el procesamiento de la información en este estudio, se realizó:

- Un estudio de la correlación entre las palabras más comunes utilizadas por los usuarios los días 6D y 5E.
- Un análisis sobre la línea de tiempo para identificar la hora en la cual hubo mayor actividad en la red los días 6D y 5E.
- Un análisis para determinar posibles cuentas bots que presentaron actividad los días 6D y 5E.

La metodología usada en esta segunda parte del procesamiento de la información, se describe a continuación.

2.1. Correlación entre palabras

La correlación entre dos palabras presentes en un conjunto de datos que contiene texto indica la frecuencia con la que aparecen juntas, en relación a la frecuencia con la que aparecen separadas, y mide de esta manera el grado de asociación que presentan. Indica en este caso que, si la correlación es grande entre dos palabras, es más probable que aparezcan juntas en un tweet o que no aparezca ninguna de las dos, a que aparezca una sola de ellas en un tweet.

Una medida comúnmente usada de este tipo de correlaciones binarias es el coeficiente φ . Si las palabras son X e Y, este coeficiente será

$$\varphi = \frac{n_{11}n_{00} - n_{10}n_{01}}{\sqrt{n_{1\cdot}n_{0\cdot} - n_{\cdot 0}n_{\cdot 1}}}$$

donde los índices n_{**} son el número de tweets que contienen o no las palabras X ó Y, según se especifica en la tabla siguiente

Tabla 1
Valores usados para calcular el coeficiente φ .

	Contiene Y	No contiene Y	Total
Contiene X	n_{11}	n_{10}	$n_{1\cdot}$
No Contiene X	n_{01}	n_{00}	$n_{0\cdot}$
Total	$n_{\cdot 1}$	$n_{\cdot 0}$	

Fuente: Silge, J. y Robinson, D. (2017).

Una elección adecuada de los pares de palabras y el estudio de la correlación entre ellas permite obtener conclusiones acerca del discurso subyacente a un texto. Se eligieron pares de palabras, con alta frecuencia de aparición, en consonancia con los eventos del 6D y 5E, y para mostrar el grado de correlación entre ellas se hizo uso de recursos de la teoría de grafos que facilitan su visualización.

2.2. Línea de tiempo

Cada mensaje contenido en la data extraída en los dos días analizados, cuenta con la hora en que fue emitido. Esto permite contabilizar la cantidad de mensajes por hora e identificar, para cada uno de los dos días en estudio, la hora del día que presentó mayor actividad.

2.3. Identificación de cuentas bots

Con el objetivo de identificar cuentas bots que mostraron actividad estos dos días donde se produjeron eventos relevantes en cuanto al tema electoral, se establecieron estrategias para detectar dos tipos bien diferenciados de cuentas. A saber:

1. Cuentas cuyo contenido de los tweets no muestra información de utilidad y solo hacen uso de tendencias y mencionan múltiples cuentas. Este tipo de cuentas no suelen permanecer activas mucho tiempo, pues el algoritmo de la plataforma de Twitter las suele suspender por incumplir las políticas de uso.
2. Cuentas cuyos tweets tienen información de utilidad, pero suelen hacer retweet de forma masiva saturando la red y con ello posicionando etiquetas.

Para la detección de estos dos tipos de cuentas se aplicó una técnica de minería de datos vinculada al aprendizaje estadístico, que permite con una precisión del 95% clasificar las cuentas como bots o no. Esta técnica se conoce como XGBoost (derivado del inglés Extreme Gradient Boosting) y se trata de un algoritmo predictivo supervisado, implementado en R, que utiliza el Principio de Remuestreo o Boosting.

La idea detrás del Remuestreo es la de generar múltiples árboles de clasificación de forma secuencial, haciendo que cada uno de éstos tome los resultados de la clasificación anterior para generar un árbol o modelo más “fuerte”, con mejor poder predictivo y mayor estabilidad en sus resultados. Para conseguir un modelo más fuerte se emplea un algoritmo de optimización. Durante el entrenamiento, los parámetros de cada modelo son ajustados de forma iterativa tratando de encontrar el mínimo de la raíz del error cuadrático medio. Cada modelo es comparado con el anterior y si un nuevo modelo tiene mejores resultados, entonces se toma este como base para realizar nuevas modificaciones. Si, por el contrario, tiene peores resultados, se regresa al mejor modelo anterior y se modifica ese de una manera diferente. Este proceso se repite hasta llegar a un punto en el que la diferencia entre modelos consecutivos es insignificante, lo cual nos indica que hemos encontrado el mejor modelo posible, es decir la mejor clasificación, y adicionalmente se obtiene un valor de probabilidad, el cual nos indica que tan probable o certera es dicha clasificación. Para mayor información sobre esta técnica del Remuestreo o Boosting y su implementación, ver Friedman (2001) y Friedman (2002).

Como es de esperarse, si el valor de probabilidad asignado por el algoritmo a una cuenta en particular es cercano a 1, entonces es altamente probable que la cuenta en cuestión sea una cuenta bot. En este caso, y como criterio para determinar si una cuenta particular es bot o no, convenimos que si la probabilidad arrojada por el algoritmo para dicha cuenta es mayor a 0.8, entonces la consideraremos como cuenta bot.

3. Resultados y discusión

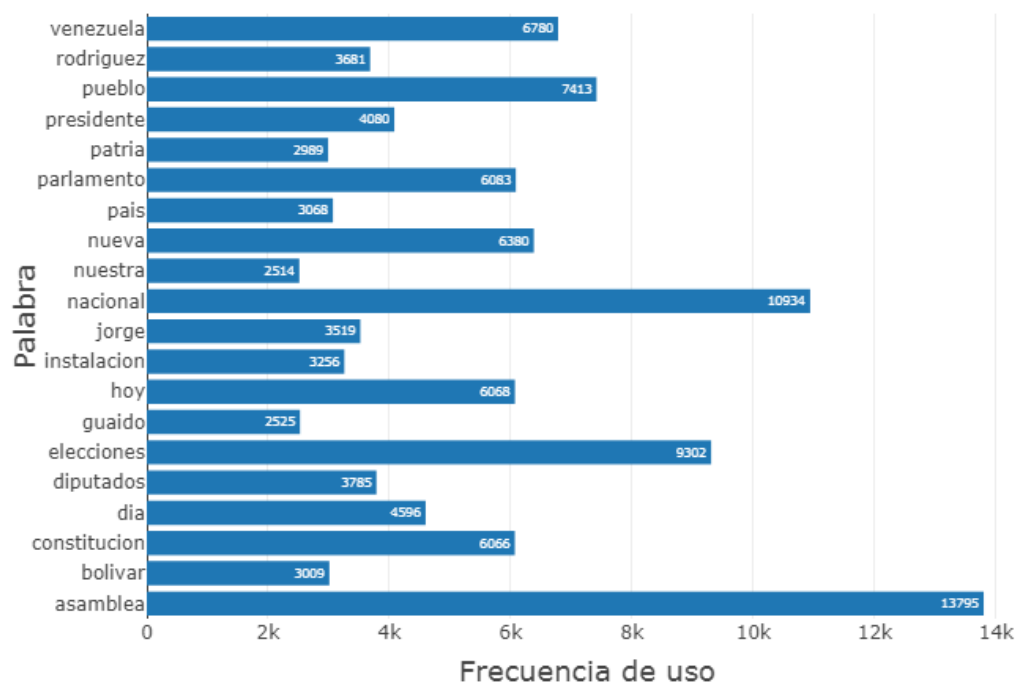
En este capítulo se presentan los resultados obtenidos al analizar el conjunto de tweets recopilados. En primer lugar considerando ambas fechas, 6D y 5E, conjuntamente, es decir se analizó la totalidad de los tweets emitidos los dos días. Tal como se hizo en el artículo Colina y otros (2021), se realizó un análisis de tendencias, se determinaron las palabras más utilizadas y se aplicó el modelo de bigramas para determinar los principales temas de conversación presentes en los textos de los tweets. Seguidamente, haciendo uso de técnicas estadísticas y recursos de la teoría de grafos, se determinaron las palabras presentes en los tweets que mostraron mayor correlación con los vocablos *asamblea*, *elecciones* y *nueva*. Para finalizar, se consideraron ambas fechas en

estudio de manera independiente, con el fin de hacer un análisis más exhaustivo. En este caso, se identificaron las palabras más utilizadas por los usuarios al emitir opiniones respecto al tema electoral, se determinó el periodo de cada día analizado qué mostró mayor actividad por parte de los tuiteros y, para finalizar, se identificaron probables cuentas bots que mostraron actividad en ambas fechas.

3.1. Análisis de datos agrupados.

Al ser eventos que no son independientes, en primer lugar se determinaron las palabras usadas con más frecuencia por los usuarios en este conjunto de tweets de ambas fechas, luego se procedió a graficar el número de veces que se menciona en función de la palabra. El gráfico siguiente muestra el resultado obtenido.

Gráfico 1
Palabras más tuiteadas

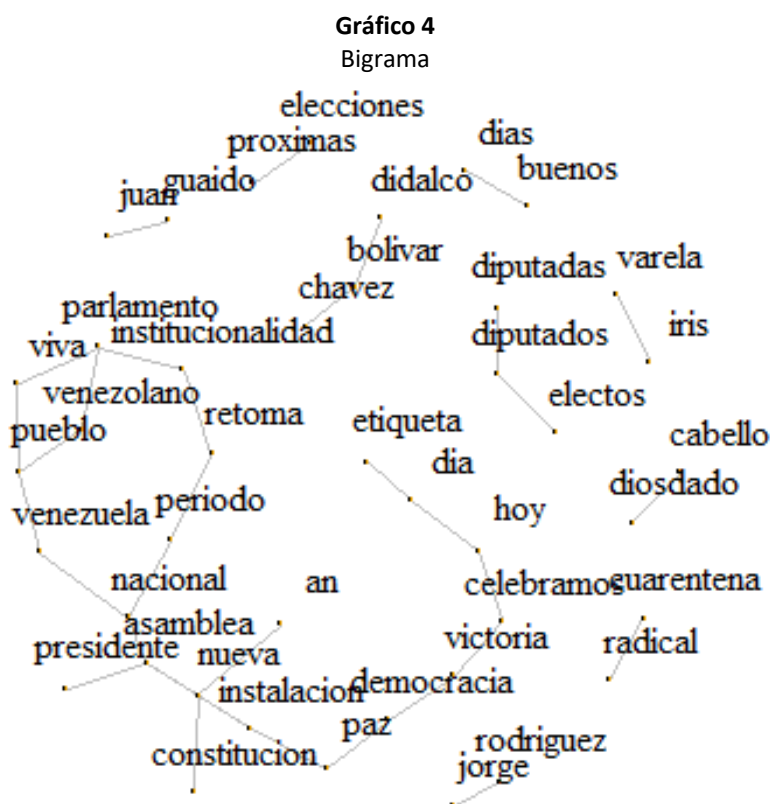


Fuente: Los autores

Las dos palabras más usadas o tuiteadas fueron, como puede verse en el gráfico 1, *asamblea* y *nacional*. Entre el resto de los vocablos usados, destacan también dos nombres de personajes públicos, a saber, Jorge Rodríguez y Juan Guaidó. El primero (mucho más nombrado por los tuiteros), el presidente de la nueva Asamblea Nacional y el segundo, presidente de la Asamblea Nacional saliente.

Otra forma de visualizar lo anterior es mediante una nube de palabras. Para generar esta nube se consideraron términos que aparecieran en el conjunto de tweets con una frecuencia de al menos 1000 repeticiones. El tamaño de las letras de las palabras en la nube está determinado por el número de veces que se menciona (más menciones = mayor tamaño) y el color vincula aquellas palabras de frecuencia de uso equivalente. La nube obtenida se muestra en el gráfico a continuación.

Posteriormente se construye y se analiza el bigrama obtenido a partir de los textos contenidos en este conjunto de tweets. En este caso se determinaron los pares de palabras más frecuentes en los diferentes textos, lo cual genera eslabones que se pueden graficar y que permiten ver de forma resumida la información. De esta manera se pueden determinar los principales temas de conversación considerando estos dos días en conjunto. El bigrama que se obtenido se muestra a continuación.



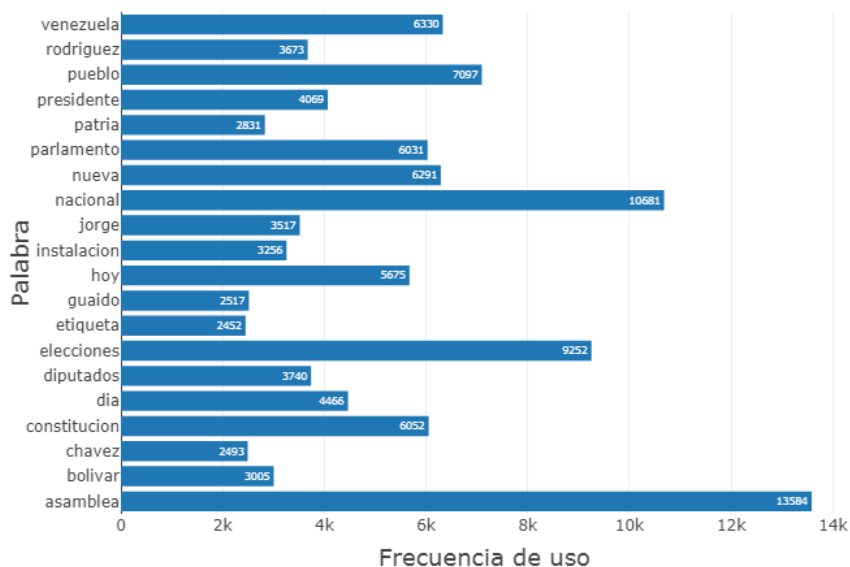
Fuente: Los autores

Siguiendo las flechas en el gráfico (ver grafico 4) y la agrupación de las palabras, se puede ver que los usuarios de la red hablaron estos dos días, 6D y 5E, principalmente de:

- La Asamblea Nacional. La instalación de este Parlamento para un nuevo período constitucional y cómo se percibe por un sector (gobierno) como una victoria y una vuelta a la institucionalidad y a la paz. Vinculada a esta misma conversación se menciona una celebración de esto por el mismo sector.
- La cuarentena motivada por la Covid 19.
- Unas próximas elecciones.
- Los diputados electos Jorge Rodríguez, Iris Valera y Didalco Bolívar, miembros de la directiva de la nueva Asamblea Nacional.
- Juan Guaidó, el presidente del Parlamento saliente.
- El diputado electo Diosdado Cabello, quien se mencionó como posible miembro de la directiva del nuevo Parlamento.
- Los diputados electos, en general.

El siguiente gráfico muestra las palabras más usadas por los tuiteros el día 5 de enero de 2021 (5E), fecha de instalación de la Asamblea Nacional elegida el 6D.

Gráfico 8
Palabras más tuiteadas el 5E



Fuente: Los autores

En el gráfico (ver gráfico 8) se puede ver que las palabras *asamblea*, *nacional* y *elecciones* son las más usadas por los tuiteros en sus conversaciones. Aparecen también los nombres de Jorge Rodríguez, Simón Bolívar, Juan Guaidó y, en menor medida, Hugo Chávez.

En el gráfico a continuación se muestra la nube de palabras obtenida para el 5E. Para generar esta nube se consideraron palabras que aparecieran con una frecuencia de al menos 1000 repeticiones.

Gráfico 9
Nube de palabras más usadas el 5E

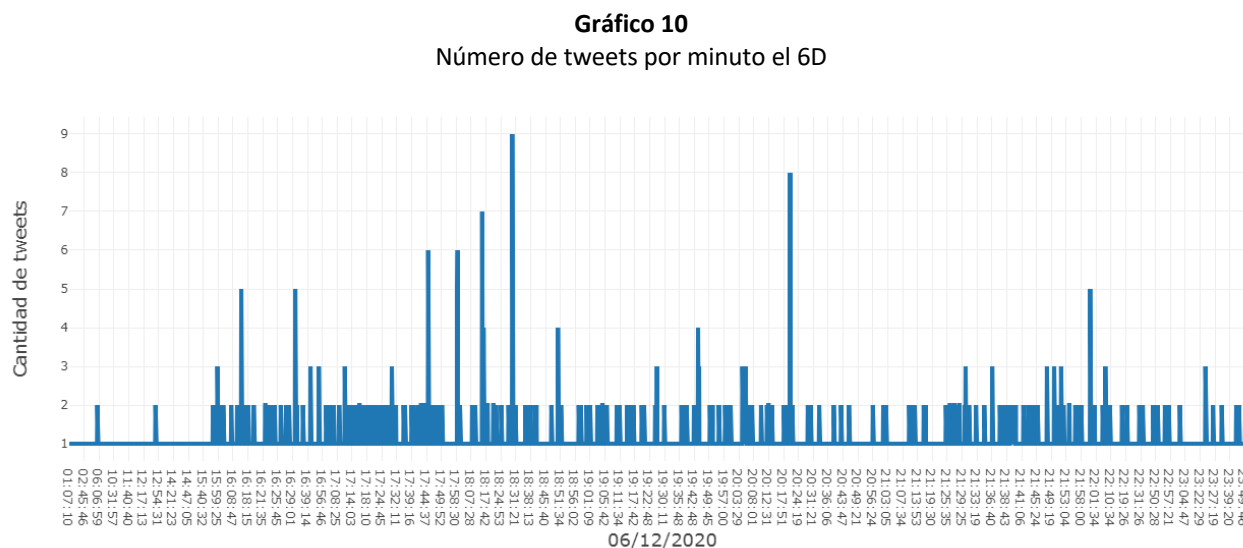


Fuente: Los autores

Se observa que las palabras *asamblea* y *nacional* son las más destacadas en la nube de palabras para el 5E (ver gráfico 9).

3.2.2. Línea temporal

Acercando la lupa aún más sobre los mensajes de los tuiteros minuto a minuto durante todo el día 6 de diciembre de 2020, se obtuvo un gráfico (ver gráfico 10) donde puede observarse que el periodo de mayor actividad se encuentra entre las 4:00 p.m. y las 12 p.m., alcanzando un pico notable a las 6:30 p.m., momento en el que estaban cerrando la gran mayoría de las mesas de votación.



[6] "#6Dic | Javier Bertucci: \"Nuestro pueblo ya no aguanta más sueños, no aguanta más esta situación difícil que enfrenta todos los días. No caigamos en fantasías, lo único real que tenemos es el voto y es lo único vinculante para generar El Cambio en el país\". #ElCambioVa <https://t.co/IZrt9Xw7ir>"

[7] "He cumplido responsablemente con mi derecho y deber ciudadano. La vía es electoral, pacífica y constitucional, el voto es el único instrumento de cambio verdadero para el conflicto que vive nuestro país. #ElCambioVA <https://t.co/5eWIBuz89M>"

[8] "#6Dic | Javier Bertucci: \"Nuestro pueblo ya no aguanta más sueños, no aguanta más esta situación difícil que enfrenta todos los días. No caigamos en fantasías, lo único real que tenemos es el voto y es lo único vinculante para generar El Cambio en el país\". #ElCambioVa <https://t.co/IZrt9Xw7ir>"

[9] "#6Dic | Javier Bertucci: \"Nuestro pueblo ya no aguanta más sueños, no aguanta más esta situación difícil que enfrenta todos los días. No caigamos en fantasías, lo único real que tenemos es el voto y es lo único vinculante para generar El Cambio en el país\". #ElCambioVa <https://t.co/IZrt9Xw7ir>"

[10] "Los que deciden una elección en #venezuela son los más pobre y la mayoría no tiene redes sociales GRCIAS A DIOS.\n#6D #ElCambioVa"

[11] "#6Dic | Nuestra lucha la damos hoy desde los centros electorales, eligiendo a quienes deseamos que nos representen en el parlamento nacional. Hoy más que nunca, los venezolanos sabemos qué queremos para nuestro país y lo demostraremos con nuestro voto. #ElCambioVa <https://t.co/X95Jg6dfTR>"

[12] "#ElCambioVa <https://t.co/6M24FsgjVu>"

[13] "Ya ejercí mi derecho. #ElCambioVa"

[14] "He cumplido responsablemente con mi derecho y deber ciudadano. La vía es electoral, pacífica y constitucional, el voto es el único instrumento de cambio verdadero para el conflicto que vive nuestro país. #ElCambioVA <https://t.co/5eWIBuz89M>"

[15] "#6Dic | Javier Bertucci: \"Generaremos un gran acuerdo político desde la Asamblea Nacional, para activar el referéndum revocatorio y presentarle al país un mecanismo realizable y una vía pacífica para salir de conflicto\". \n\n#ElCambioVa <https://t.co/XU6VPfnyQK>"

[16] "¡El día ha llegado!\n\nHoy comienza #ElCambio en nuestro país. Porque para influenciar los poderes de Venezuela, debemos tomar espacios como la #AN. ¡Así que vota este #6Dic abajo y a la derecha!\n\n#ElCambioVA <https://t.co/WoILmW8m1I>"

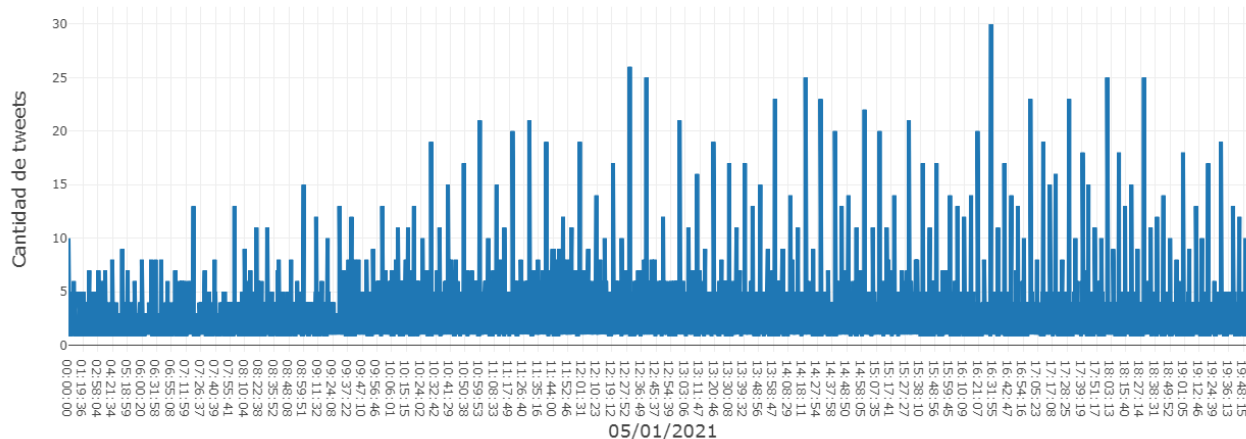
[17] "¡Nuestra democracia es gobernable! Con ello se alude a dos niveles fundamentales de la política: los procesos democráticos para la conformación de gobiernos legítimos y el ejercicio gubernamental eficiente con vocación de servicio ciudadano. #ElCambioVa <https://t.co/H5FXPF68Oc>"

[18] "¡Debemos restablecer la democracia en Venezuela!\n\nDurante muchos años nos hemos preguntado cuándo comenzó el declive de la democracia en #Venezuela. Sin entender, que comenzó el día que decidimos abstenernos a votar y dejar la gobernabilidad en manos del oficialismo.\n\n#ElCambioVa"

Entre los tweets mostrados se nota claramente la presencia de textos que se repiten textualmente, a pesar de que provienen de cuentas diferentes, como por ejemplo los etiquetados como [1], [2], [3], [7] y [14]. Esto indicaría el posible uso de cuentas bots para reproducir mensajes en la plataforma con el propósito de crear de forma artificial tendencias e influir en la percepción de la gente sobre la motivación de la participación en el evento electoral de 6D. Este día, de los 17828 tweets analizados, se identificaron 14099 tweets idénticos.

Un análisis similar se aplicó el 5E. Puede observarse (ver gráfico 11) que la actividad comienza muy temprano (poco después de media noche) y aumenta paulatinamente durante todo el día, alcanzando su momento de mayor actividad a las 4:30 p.m., momento culminante de la instalación de la Asamblea Nacional.

Gráfico 11
Número de tweets por minuto el 5E



Fuente: Los autores

Algunos tweets emitidos después de esa hora pico (4:30 p.m.), como puede verse en la siguiente lista, en general reflejan comentarios acerca del significado de la Asamblea Nacional, tanto para el gobierno, como para la oposición, y la incidencia de ésta en el devenir de la Nación y en la lucha política.

[1] "Venezuela es un paciente agonizante, que está en cuidados intensivos y en un hospital que no tiene insumos ni electricidad. Por eso nosotros haremos todo lo contrario a lo que quiere la dictadura #ANLegitimaSigueDePie <https://t.co/6pCk0Z1vrD>"

[2] "Nosotros estamos dando la cara y trabajando siempre. Unos perseguidos, otros en la cárcel, otros torturados, otros en el exilio y otros hasta dieron su vida. Nos ha costado sangre, sudor y lágrimas, pero solo así se ha construido la libertad en todo el mundo #ANLegitimaSigueDePie <https://t.co/Xns0KPlu3q>"

[3] "Mientras ellos intentan matar a Venezuela, nosotros tratamos de revivirla; mientras ellos se empeñan en asfixiarla, a nosotros nos toca darle oxígeno; mientras ellos se empeñan en destruirla, nosotros nos empeñamos en reconstruirla #ANLegitimaSigueDePie <https://t.co/KZFcLicsBK>"

[4] "Hoy nos enfrentamos a un nuevo desafío que nos mantiene en el constante esfuerzo de lograr la liberación de nuestra nación, secuestrada por un grupo de delincuentes que han llevado al país a la peor de las miserias y de la destrucción #ANLegitimaSigueDePie <https://t.co/opdkqk78b5>"

[5] "Chavismo retoma control del Parlamento con promesa de 'exorcizar' era Guaidó. #NuevaAsambleaNacional - #LVBPxTLT - #ANLegitimaSigueDePie"

[6] "En este punto de la historia, los venezolanos tenemos que hacer un acto de contrición. Todos sin excepción. Porque durante estos años hemos cometido errores por acción o por omisión. Todos debemos rectificar, todos debemos actuar #ANLegitimaSigueDePie <https://t.co/GS6xsevpRc>"

[7] #NuevaAsambleaNacional @PetraYsabelina @Alex2021Jsr @Jose71951453 @mariann12876209 @escobarj291 @Marcos16308564 @GiannaP75346516 @Daniela10828048 @Thais64101031 <https://t.co/vaxSpmM9Z5>

[8] "A todos los venezolanos debe unirnos el objetivo de liberar a nuestra nación. Mientras la dictadura intenta matar y destruir nuestra Venezuela, nosotros tratamos de revivirla y luchamos para reconstruirla #ANLegitimaSigueDePie <https://t.co/saeaoIjAfJ>"

[9] "Los factores democráticos y legítimos, asumieron la Continuidad Constitucional de la AN #ANLegitimaSigueDePie , firmes en su compromiso con Venezuela y preservar el único poder legítimo que tienen los venezolanos. #FalsaANDeMaduroFracasó <https://t.co/tyaFUEPIfv>"

[10] "Cuántas Asambleas hay? 1 2 3 El pato maché #ANLegitimaSigueDePie #NuevaAsambleaNacional #LassoNuncaSeraPresidente"

[11] "El País sigue en su lucha entre la barbarie y la civilización, el atraso y el desarrollo. Venceremos la sombra de la miseria con la luz del progreso. Juntos, AN legítima y pueblo de Venezuela vamos a recuperar la democracia, la libertad y la paz. #ANLegitimaSigueDePie"

[12] "La #FalsaANDeMaduroFracasó. Los legítimos diputados estamos aquí, de pie. Porque nos corresponde representar a millones de venezolanos, asumimos la continuidad Constitucional. #ANLegitimaSigueDePie <https://t.co/USPrY3jK0c>"

[13] "Yo me pregunto como es que el narcogobierno instala con elecciones fraudulentas la disquera asamblea y toma Posesión del palacio federal? es que hacen lo que les dala gana, no hay justicia, donde está guaido? #ANLegitimaSigueDePie"

[14] "La Libertad de Venezuela es como un secreto a voces, todos lo saben pero tienen recelo de decirlo #ANLegitimaSigueDePie"

[15] "#ANLegitimaSigueDePie <https://t.co/jhIW7DizCO>"

[16] @Mippcivzla #5Ene |  ¡ETIQUETA DEL DÍA!  #NuevaAsambleaNacional

[17] "Éstos ESCUECAS ladrones insisten en su estupidez #ANLegitimaSigueDePie...bueno...en la estupidez de quienes les creen...si son bien mmmgv..."

[18] #NuevaAsambleaNacional @Yohander2911 @Ender4059  @will6942  @ch_cdg  @BaenaDegas @caciquetiunatuy  @SotoTuitero  @JesusYormin  @danielarujano  @mpsa25  @mercedeslouzaod  @kerny70  @1_yanny  @luiscarrillo66  @ritik_b  @amelia74698445  @kermys4 @jooz_paraguana2 @Maricar12737912 @ArelYR09038179

Entre los tweets de la lista anterior se aprecian textos cuya única intención es promover una etiqueta, como por ejemplo los textos hemos etiquetado como [7], [16] y [18]. Esto indicaría el posible uso de cuentas bots para posicionar un hashtag en la plataforma con el propósito de crear tendencias de forma artificial y de esta manera influir en cómo percibió la gente la instalación de la Asamblea Nacional elegida el 6D.

3.2.3. Cuentas bots

La presencia, en los datos recopilados estos dos días, de tweets idénticos emitidos por cuentas distintas y tweets cuya única finalidad era promover etiquetas y mencionar otras cuentas, indicaría, como ya se mencionó, la posible actividad de cuentas bots. Para finalizar, se muestran los resultados obtenidos, mediante la técnica

XGBoost, en la búsqueda e identificación de posibles cuentas bots que hayan mostrado actividad en los dos días analizados. Esto incluye el conteo de la totalidad de tweets emitidos por estas cuentas y también la valoración de los contenidos que divulgaron.

Para el día 6 de diciembre de 2020, en los datos recopilados se contabilizaron 9581 tweets que fueron generados por 547 cuentas diferentes cuya probabilidad de ser cuenta bot es igual o mayor a 0.8, con un nivel de confianza del 95%, y que según el criterio establecido son cuentas bots. Estos 9581 tweets representan un relevante 53.74% del total de tweets analizados, lo que indica una importante actividad de cuentas bots para ese día. La siguiente tabla muestra, en relación a los datos recopilados el 6D, el número total de cuentas diferentes identificadas, el número de cuentas bots detectadas en este total por el algoritmo y la cantidad de tweets emitidos por todas éstas.

Tabla 2
Número de cuentas, cuentas bot y tweets el 6D

	Total	Tweets emitidos
Cuentas	2799	17828
Cuentas bots	547	9581

Fuente: Los autores.

En relación al contenido de estos 9581 tweets emitidos por cuentas bots (ver tabla 2), se determinó que en la mayoría de estos se utilizaron las etiquetas #SoyComunicadorClap, #Parlamentarias2020 y #YoSiVoto, acompañadas de menciones a distintas cuentas de usuario de Twitter, en algunos casos dirigentes políticos relacionados al gobierno nacional. Un ejemplo de esta clase de tweet se muestra en la siguiente figura, el cual fue emitido por el usuario lrrc2020, cuya cuenta fue identificada como bot con una probabilidad de 0.9856982 por el algoritmo de detección.

Figura 1
Tweet emitido por cuenta bot el 6D

<p>Tod@s a Votar LaPatriaLoReclama @NicolasMaduro @FreddyBernal @CClaPalavecino1 #SoyComunicadorClap #Parlamentarias2020 https://t.co/qa8fjbtEvv</p>

Fuente: Twitter.

A continuación, en la tabla 3, se muestran algunas cuentas identificadas en los datos recopilados el día 6 de diciembre de 2020 y cuya probabilidad de ser bot es mayor a 0.97. En dicha tabla La columna **Probabilidad** indica la probabilidad de ser cuenta bot asignada por el algoritmo de detección.

Tabla 3

Algunas cuentas identificadas como bots el 6D.

Usuario	Probabilidad	Usuario	Probabilidad
lrrc2020	0.9856982	Shuail	0.9754518
1_Mejor_Consejo	0.9754518	tuiteroactivo11	0.9754518
Alejandro06611430	0.9754518	JoseLuisReina19	0.9754518
ArturoD03807596	0.9754518	LUCILA81431780	0.9754518
MeiberTrejo	0.9754518	Ramonci46644876	0.9754518
Sanjuan35940848	0.9754518	Delfin77146846	0.9754518

Fuente: Los autores.

Por otra parte, para el día 5 de enero de 2021 se contabilizaron 2086 tweets que fueron generados por 174 cuentas diferentes cuya probabilidad de ser cuenta bot es igual o mayor a 0.8, con un nivel de confianza del 95%, y que según el criterio establecido son cuentas bots. Estos 2086 tweets representan apenas un 3.15% del total de tweets analizados. La siguiente tabla muestra, en relación a los datos recopilados el 5E, el número total de cuentas diferentes identificadas, el número de cuentas bots detectadas en este total por el algoritmo y la cantidad de tweets emitidos por todas éstas.

Tabla 4

Número de cuentas, cuentas bot y tweets el 5E.

	Total	Tweets emitidos
Cuentas	26134	66172
Cuentas bots	174	2086

Fuente: Los autores.

En los 2086 tweets emitidos por cuentas bots el 5E (ver tabla 4) se utilizó la etiqueta #NuevaAsambleaNacional acompañada de la frase “Etiqueta del día”, además de mencionar distintas cuentas de usuario de Twitter. Un ejemplo de esta clase de tweet se muestra en la siguiente figura, el cual fue emitido por el usuario william15666773, cuya cuenta fue identificada como bot con una probabilidad de 0.9754518 por el algoritmo de detección.

Figura 2

Tweet emitido por cuenta bot el 5E.

🍷🍷🍷🍷 #NuevaAsambleaNacional @Ender4059👉@will6942👉@ch_cdg👉@BaenaDegas
 @caciquetianatuy👉@SotoTuitero👉@ItaloPsv👉@danielarujano👉@mpsa25👉@mercedeslouzaod
 👉@kerny70👉@1_yanny👉@luiscarrillo66👉@ritik_b👉@amelia74698445👉@kermys4
 @jooz_paraguana2 @Maricar12737912 @Patriotas_20k @RYorm

Fuente: Twitter.

En la siguiente tabla se muestran algunas cuentas identificadas como bots por el algoritmo de detección con una probabilidad superior a 0.97, y que mostraron actividad el día 5 de enero de 2021. La columna **Probabilidad** indica la probabilidad de ser cuenta bot asignada por el algoritmo de detección.

Tabla 5
Algunas cuentas identificadas como bots el 5E.

Usuario	Probabilidad	Usuario	Probabilidad	Usuario	Probabilidad
eliztorres	0.9754518	tonogonzalez18	0.9848609	nunezvolcanes	0.9754518
jesika2323	0.9754518	joseluis0715	0.9821625	wendyvillla2020	0.9754518
jdmendezr71	0.9754518	william15666773	0.9754518	luislop21413453	0.9754518
mara38100484	0.9754518	blancas79919200	0.9754518	jesus_doctor1	0.9754518
mariaplanells4	0.9754518	soraima98929733	0.9754518	lis14445779	0.9754518
ortiz_nurys	0.9754518	deniscuevas6	0.9754518	kerny70	0.9754518
italopsuv	0.9754518	caciquetiunatuy	0.9754518	hendrick_558	0.9754518
kmfereira	0.9754518	arenadecafe	0.9754518	yorlin_v	0.9754518

Fuente: Los autores.

En resumen, y considerando ambos días, se contabilizó un total de 11667 tweets que fueron generados por 686 cuentas diferentes cuya probabilidad de ser bot es igual o mayor a 0.8, con un nivel de confianza del 95%, y que según el criterio establecido son cuentas bots. De este total de 686 cuentas bots se identificaron 35 de ellas que emitieron tweets en ambas fechas analizadas. Es importante señalar que la mayoría de cuentas bots por lo general son suspendidas por Twitter en muy poco tiempo, pues incumplen las políticas de uso relativo al spam y a la manipulación de la plataforma.

4. Conclusiones

Lo primero a destacar en el análisis realizado es que el número de tweets emitidos el día 5E es casi cuatro veces mayor que los emitidos el día 6D. Así mismo, el número de cuentas diferentes identificadas el 5E es casi diez veces más que las identificadas el 6E. Todo esto pudiera interpretarse como que la instalación de la Asamblea Nacional generó mayor impacto en el contenido de Twitter que la propia elección de sus integrantes.

La Asamblea Nacional, como era de esperarse, fue el tema principal de opinión. En este sentido, el análisis textual evidenció que lo que más destacado fueron, aparte de la instalación de ésta el día 5 de enero de 2021, las discusiones y opiniones acerca de los posibles integrantes que conformarían la directiva. Por otra parte, el análisis correlacional entre palabras permitió concluir que la percepción acerca de la figura de la Asamblea se mostró marcadamente dividida. Una parte de los tuiteros la calificaron como fraudulenta, la otra como legítima.

El análisis de la actividad de los tuiteros mediante líneas temporales evidenció una actividad esporádica el 6D, con su punto de mayor actividad a la hora del cierre de los centros electorales, y en cuanto al 5E, una actividad constante a lo largo del día, con el punto de mayor a la hora de la instalación de la directiva de la Asamblea Nacional.

En cuanto a los personajes que obtuvieron mayor mención en la plataforma por parte de los usuarios de Twitter al opinar sobre el tema electoral en Venezuela y los eventos del 6D y 5E, estos fueron los integrantes de la directiva de la Asamblea Nacional instalada el 5E, Jorge Rodríguez, Iris Valera y Didalco Bolívar, el presidente de la Asamblea Nacional saliente, Juan Guaidó y el diputado electo el 6D, Diosdado Cabello.

Una conclusión importante a la que arribó este estudio de los mensajes en la red social Twitter fue que se comprobó la utilización de cuentas bots para crear, de forma artificial, tendencias e influir en la intención y percepción de la gente, en ambas fechas, a través del envío masivo de mensajes. En el conjunto total de datos recopilados se logró identificar, con un nivel de confianza del 95%, un total de 686 cuentas bots diferentes que emitieron tweets en al menos uno de los dos días analizados. La mayor actividad de cuentas bots se detectó el 6D, donde los tweets emitidos por este tipo de cuentas representan un significativo 53.74% del total de tweets recopilados ese día. Se comprobó que el 6 de diciembre de 2020 se utilizaron las etiquetas #SoyComunicadorClap, #Parlamentarias2020 y #YoSiVoto en 9581 tweets generados por 547 cuentas bots, lo que representa el 19.54% del total de cuentas diferentes identificadas ese día. De igual manera, en los datos recopilados se identificaron 174 cuentas bots que promovieron masivamente la etiqueta #NuevaAsambleaNacional a través de 2086 tweets, el día 5 de enero de 2021. En este caso, estos 2086 tweets representan un escaso 3.15% del total de tweets emitidos ese día y el número de cuentas bots detectadas, apenas el 0.66%, evidenciando que, aunque hubo actividad de cuentas bots el día de la instalación de la Asamblea Nacional, esta ocurrió en mucha menor medida que el día de las elecciones parlamentarias, donde la actividad de cuentas bots, en cuanto al número y actividad de éstas, fue muy evidente y significativa.

Referencias bibliográficas

- Colina, M., Colina, K., Fernández, A., Maia, M. y Torrealba, D. (2021). Un recorrido electoral en Twitter Venezuela durante el periodo enero de 2019 a diciembre de 2020. *Espacios*, 42(12), 94-114. Recuperado de: <http://www.revistaespacios.com/a21v42n12/21421208.html> o DOI: 10.48082/espacios-a21v42n12p08
- Friedman, J. (2001). Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*, 29(5), 1189-1232.
- Friedman, J. (2002). Stochastic Gradient Boosting. *Computational Statistics & Data Analysis*, 38(4), 367-378.
- Jo T. (2019). *Text mining: Concepts, Implementation, and Big Data Challenge*. New York: Springer.
- Silge, J. y Robinson, D. (2017). *Text Mining with R: A Tidy Approach*. Sebastopol, U.S.A.: O'Reilly Media.



Esta obra está bajo una [Licencia Creative Commons](https://creativecommons.org/licenses/by-nc/4.0/)
Atribución-NoComercial 4.0 Internacional