

# Análisis computacional del uso de Twitter durante los debates presidenciales del 2022 en Costa Rica

Computational analysis of the use of Twitter during the 2022 presidential debates in Costa Rica

*Elías Chavarría-Mora*

*Universidad de Pittsburgh  
Pittsburgh, Estados Unidos  
elc117@pitt.edu*

**RESUMEN.** ¿Cómo se desarrolló la conversación alrededor de los candidatos presidenciales del 2022 en la red social *Twitter* durante los debates presidenciales de *RepreTel* y *Teletica*? En esta nota de investigación exploratoria, utilizo técnicas computacionales del tipo análisis de diccionario y aprendizaje no supervisado para modelación de temas para explorar esta pregunta, con el objetivo de aportar a la creciente cantidad de investigaciones cuantitativas sobre redes sociales en Costa Rica. Mientras los resultados son netamente exploratorios y no prueban hipótesis, estos no sugieren una gran presencia de cuentas automatizadas durante el debate. Rodrigo Chaves, como tema, claramente dominó la conversación en Twitter, en particular, en relación con las acusaciones de acoso sexual en su contra.

**PALABRAS CLAVE.** Ciencias sociales computacionales, elecciones, modelación de temas, redes sociales, debates presidenciales.

**ABSTRACT.** In which ways did the conversation evolve around presidential candidates in 2022 in the social network *Twitter* during the presidential debates of *RepreTel* and *Teletica*? In this exploratory research note, I use computational techniques such as dictionary analysis and unsupervised machine learning for topic modeling to explore this question, with the goal of adding to the growing number of quantitative research regarding social



networks in Costa Rica. While the results are merely exploratory and do not test any hypothesis, they do not suggest a great presence of automated accounts during the debates. Rodrigo Chaves, as a topic, clearly dominated the conversation in Twitter, in particular with regards to the accusations of sexual harassment against him.

KEYWORDS. Computational social sciences, elections, social networks, topic modeling, presidential debates.

## 1. INTRODUCCIÓN

¿Cómo se desarrolló la conversación alrededor de los candidatos presidenciales del 2022 en la red social *Twitter* durante los debates presidenciales de *Repretel* y *Teletica*? En esta nota de investigación, propongo utilizar una serie de técnicas de análisis computacional para acercarse de una forma exploratoria a los estilos de comunicación que fueron utilizados por los usuarios de *Twitter* durante los dos debates presidenciales que fueron transmitidos por los dos principales canales de noticias nacionales.

El tema del uso de redes sociales en campañas electorales en Costa Rica, tanto de parte de candidatos como de ciudadanos, ha venido cobrando importancia, puesto que estas redes cumplen un papel importante como forma en que la ciudadanía se informa sobre política (Centro de Investigación y Estudios Políticos - Universidad de Costa Rica 2018), y pueden resultar una herramienta útil y de bajo costo para realizar campañas para partidos que de otra forma no tendrían recursos para esto (Tremínio Sánchez 2022). Además de esto, candidatos a puestos públicos han demostrado tener interés en el uso de estas herramientas (Carazo Barrantes, Tristán Jiménez, y Siles González 2021). A su vez, se ha dado la voz de alarma sobre un incremento en el discurso agresivo e ideológicamente cargado en *Twitter* en Costa Rica (Bunse 2021).

Por todos los motivos previamente expuestos, es importante continuar el estudio de redes sociales y su efecto en política de una forma metodológicamente rigurosa, tal que se pueda aprovechar en su totalidad la enorme cantidad de datos producidas en estos espacios. Parte de esto implica también aprovecharse de las ventajas ofrecidas por métodos desarrollados en otras disciplinas, como ciencias computacionales o ciencia de datos, particularmente creados para utilizar el volumen y complejidad de la información producida en redes sociales.

El objetivo de este texto es aportar en esa dirección. Este puede dividirse en dos subobjetivos: uno sustantivo y otro metodológico. El objetivo sustantivo es realizar una descripción exploratoria del tipo de lenguaje y temas que se conversaron en Twitter durante los debates presidenciales del 2022 en Costa Rica. Además del análisis propiamente del uso de redes sociales que se dio en los debates presidenciales, se persigue como objetivo metodológico ser lo más claro y didáctico posible sobre las decisiones que se tomaron con respecto a los métodos implementados para poder ser un apoyo o contribuir al diálogo del análisis cuantitativo de textos en la ciencia política costarricense.

El texto procede de la siguiente forma. Primeramente, se presenta el desarrollo que se ha dado en el estudio de redes sociales en procesos políticos del país, así como un breve recuento contextual de las elecciones de 2022 en Costa Rica. Posteriormente, paso a describir el tipo de datos utilizados en este estudio, así como su proceso de recolección y preprocesamiento antes del análisis; además, se incluye una breve descripción de los datos. En la sección siguiente, explico el análisis utilizado: primeramente, un método basado en el uso de diccionarios LIWC basados en estudios psicométricos diseñados para identificar tipos de conversación; luego, un método de aprendizaje no supervisado para realizar modelación de temas (*topic modeling*), en donde específicamente utilizo Asignación Latente de Dirichlet (*Latent Dirichlet Allocation*, LDA) para conocer los principales temas de la conversación. En la sección siguiente, exploro la evolución de la conversación en torno a tres indicadores psicométricos, uno referente al nivel de autenticidad en la conversación, otro al nivel de complejidad de la conversación y el tercero referente al tono o sentimiento de la conversación, así como los resultados del LDA. En las secciones siguientes, discuto los resultados y concluyo.

## 2. REDES SOCIALES EN EL ANÁLISIS DE LA POLÍTICA COSTARRICENSE

Es importante reconocer que ya se han dado muy importantes pasos en Costa Rica en el análisis de sitios de redes sociales en la política, por ejemplo: el uso de redes sociales en procesos electorales como los nacionales de 2014 y municipales 2016 (Cruz Romero 2017a; 2017b), nacionales 2018 (Siles González, Carazo Barrantes, y Tristán

Jiménez 2019; Gómez Campos y Durán Monge 2020) y nacionales 2022 (Arce Chaves y Bonilla Cruz 2023), los diversos estudios en el libro *Democracia en digital* (Siles González 2020), así como el estudio de comentarios en YouTube relacionados con protestas (Alvarenga Forunier y Saénz Leandro 2022). Cabe mencionarse la multiplicidad de métodos adoptados, pero también, y como ya lo planteé previamente, que se han adquirido métodos novedosos dentro de la ciencia política para que se ajusten al objeto de estudio, como, por ejemplo, la teoría de grafos, el *eye-tracking*, la teoría de probabilidad, las técnicas de diccionario, el aprendizaje no supervisado y el *social listening*.

Al seguir esta línea de investigación, en este trabajo planteo explorar el tipo de lenguaje y los temas de la conversación de los usuarios de la red social Twitter durante los debates presidenciales del 2022 en los principales canales de televisión costarricenses. El tipo de lenguaje es analizado mediante el diccionario del proyecto LIWC, el cual es explicado más a fondo en el apartado de métodos. En otros contextos nacionales, LIWC se ha utilizado para estudiar el tipo de lenguaje utilizado en redes sociales con propósitos políticos, incluyendo tanto los mensajes generados por los votantes así como los candidatos en Twitter durante las elecciones presidenciales del 2016 en Estados Unidos (Jordan, Pennebaker, y Ehrig 2018), al analizar el uso de estilos de comunicación de parte de grupos de defensa de derechos en Facebook (Bail 2016; Bail, Brown, y Mann 2017). Además de redes sociales, LIWC ha sido también utilizado para analizar el tipo de lenguaje usado en discursos de líderes políticos en Estados Unidos, Reino Unido, Canadá, Australia y Europa (Jordan et al. 2019).

El contexto electoral en general, y los debates políticos en particular, son un espacio valioso para explorar dinámicas de comportamiento en redes sociales. No puede obviarse la situación de desalineamiento o realineamiento (Perelló y Navia 2021) del sistema de partidos que ha llevado cada vez más a la fragmentación del sistema y a la alta volatilidad en la intención de voto (Programa Estado de la Nación en Desarrollo Humano Sostenible 2018). Es posible que los debates representen un espacio apropiado para que los indecisos pasen a apoyar a un candidato con base en la calidad de su participación en el debate. Esto puesto que, además, son un espacio que permite dar a conocer propuestas políticas de partidos poco conocidos (Ovares Sánchez 2022).

Las elecciones presidenciales y legislativas de Costa Rica en el 2022 han sido caracterizadas por diversas fuentes como unas que continúan con tendencias de elecciones anteriores, como la des/realineación del sistema de partidos, el incremento en el número de partidos, pero con una dificultad en su diferenciación y el decrecimiento en lealtad partidaria y en participación electoral. A su vez, estas elecciones giraron en torno a temas económicos de austeridad, a temas de corrupción, temas relacionados con derechos de las mujeres, particularmente en relación con las acusaciones de acoso sexual en contra del eventual ganador de la contienda, Rodrigo Chaves. (Ovares Sánchez 2022; Treminio Sánchez 2022; Programa Estado de la Nación en Desarrollo Humano Sostenible 2022).

### 3. DATOS

Utilicé *academictwitteR*, un paquete del lenguaje de programación para análisis estadístico R, para acceder al API académico de Twitter y obtener todos los tweets relacionados con los debates electorales de los dos principales canales de televisión de Costa Rica durante la primera ronda de la elección del 2022. Los *tweets* fueron obtenidos durante ventanas de tiempo correspondientes a cada debate, a saber, para el caso del debate de Repretel, que tuvo lugar el jueves 3 de febrero, iniciando a las 8 p.m. hora de Costa Rica, y para el caso de Teletica, que tuvo lugar el viernes 4 de febrero, a partir de las 7 p.m. hora de Costa Rica. Ambos debates duraron alrededor de tres horas, sin embargo, considerando que la discusión alrededor de los debates se extiende por horas luego de su finalización, extendí la recolección de *tweets* hasta las 2 a.m.

Para capturar los *tweets* relevantes, utilicé una lista de palabras que incluyen el nombre completo del candidato presidencial participando en el debate (por ejemplo, “José María Villalta” o “Eliécer Feinzaig”), el mismo nombre sin tildes (por ejemplo, “Jose Maria Villalta”), variaciones comunes del nombre de algunos candidatos (por ejemplo, “Eli Feinzaig”), el nombre de los partidos, y varios *hashtags* relevantes, incluyendo el respectivo para cada debate. En total, descargué 476 *tweets* relacionados con el debate de Repretel y 4016 tweets relacionados con el debate de Teletica, para un total de 4495 tweets, provenientes de 2489 cuentas diferente de Twitter. Los *queries* o términos de búsqueda utilizados específicamente fueron los siguientes, con la única diferencia siendo el último término,

respectivamente, *#DebateRepretel* o *#DebateTeletica*, para cada caso correspondiente en vez del genérico *#debateCANAL*:

'Eliecer Feinzaig OR Eliécer Feinzaig OR Eli Feinzaig OR Fabricio Alvarado OR Lineth Saborío OR Lineth Saborio OR Welmer Ramos OR Rodrigo Chaves OR José María Figueres OR Jose Maria Figueres OR José María Villalta OR Jose Maria Villalta OR *#decision2022* OR *#elecciones2022* OR *#DebateCANAL*'

Para descargar la información de Twitter, utilicé un acceso a la interfaz de programación de aplicaciones (*application programming interface*, API) 2.0 de Twitter, que da acceso al archivo completo de tweets para investigadores académicos, en vez de dar solo acceso a *tweets* recientes. El paquete *academictwitteR* crea una base de datos en formato de R *dataframe* que incluye treinta y dos variables por cada *tweet* individual, entre ellas, el texto completo de cada *tweet*, nombre de usuario y número de identificación del autor y fecha de creación.

Una vez unificadas las bases de *tweets* respectivas a los debates de Repretel y Teletica, realicé un preprocesamiento del texto utilizando el paquete *Tidytext*, que usa la lógica *Tidyverse* de organización de bases de datos. En este proceso, convertí todas las letras mayúsculas en minúsculas, eliminé trozos de código de contenido comunes en *tweets*, pero irrelevantes para lenguaje natural humano tales como "https", "rt", "t.co" y "amp", así como saltos de línea (" $\backslash r \backslash n$ ") y salvé el texto con codificación UTF-8, esto último es importante para que se conserven vocales con tilde en los textos de los *tweets*.

Posteriormente, usé expresiones regulares para identificar si se mencionaba al candidato en un *tweet*. Las expresiones regulares son utilizadas para identificar patrones en textos, por ejemplo, empleé la expresión regular "(eli | eli[eé]cer) | (feinzaig) | (eli | eli[eé]cer feinzaig)", dado que esta abarca a todas las siguientes variaciones: "eliecer feinzaig", "eliécer feinzaig", "eli feinzaig", "feinzaig", "eli", "eliecer" y "eliécer". Específicamente, la expresión regular indica lo siguiente: la barra | es un operador lógico "o", (es decir, que aparezca "eli" o "eli[eé]cer"), mientras que las letras dentro de un paréntesis cuadrado las trata como si tuvieran un operador lógico "y" (es decir, eli[eé]cer indica que se capturen tanto "eliecer" como "eliécer"). Nótese que las mayúsculas son irrelevantes, porque en el paso anterior de preprocesamiento se convirtieron todas

las letras en minúscula.

Las expresiones regulares utilizadas capturan si se menciona en el *tweet* el nombre y apellido del candidato, pero también solo el nombre o solo el apellido. Esto puede causar que existan falsos positivos, donde, por ejemplo, se hable de otra persona de nombre “rodrigo” o de apellido “chaves” que no fuera el candidato, sin embargo, considerando como se creó la base de datos, es decir, el query que se envió al API de Twitter, no deberían existir muchos casos así. También, nótese que la excepción a esto son los casos de José María Figueres y José María Villalta para quienes, por tener el mismo primer nombre, utilicé una variación en las expresiones regulares para que no se incluyeran los casos donde solo se menciona el primer nombre “José María”. Las expresiones regulares empleadas se incluyen a continuación:

(eli | eli[eé]cer) | (feinzaig) | (eli | eli[eé]cer feinzaig)

(fabricio) | (alvarado) | (fabricio alvarado)

(lineth) | (sabor[ií]o) | (lineth sabor[ií]o)

(welmer) | (ramos) | (welmer ramos)

(rodrigo) | (chaves) | (rodrigo chaves)

(jos[ée] mar[fi]a figueres) | (figueres)

(jos[ée] mar[fi]a villalta) | (villalta)

Una vez identificado esto, uso una serie de bucles *if-else* para clasificar todos los tweets en las siguientes categorías: menciona solo a Feinzaig, menciona solo a Alvarado, menciona solo a Saborío, menciona solo a Ramos, menciona solo a Chaves, menciona solo a Figueres, menciona solo a Villalta, y una categoría de residuo: o no menciona a ninguno o menciona a más de uno. La idea detrás de esto es mantener el análisis enfocado en el tipo de lenguaje utilizado cuando es posible identificar al candidato por el cual se utiliza el lenguaje. Es decir, si se menciona a dos candidatos, no es posible

saber si el candidato A o B es el que influye en el tipo de lenguaje que se usa. Como último paso previo al análisis, utilicé el paquete *Lubridate* para que R identifique correctamente los datos de fecha y pueda manipularlos de forma correcta.

El cuadro 1 muestra la cantidad de *tweets* dentro de la base de datos que menciona exclusivamente a cada uno de los candidatos. Lo primero que salta a la vista es la enorme diferencia en cantidad de menciones: Rodrigo Chaves, el eventual vencedor de la contienda electoral, es mencionado más que cualquier otro candidato, el doble después del segundo lugar, Fabricio Alvarado. Esto puede ser debido a la naturaleza controversial de muchas de las declaraciones de Chaves, o tal vez puede explicarse por las acusaciones que han salido a la luz mucho después de la campaña sobre el supuesto uso de *troles* pagados por parte del partido de Chaves en la campaña. Ambas explicaciones no son mutuamente excluyentes, es más, tendría sentido el uso de *troles* para *astro-turfing* (es decir, inflar falsamente la relevancia) de un candidato y que esto, a su vez, genere más *engagement* de parte de usuarios reales de la red, pero una hipótesis de este tipo se escapa de los objetivos de esta investigación.

**Cuadro 1. Cantidad de tweets que mencionan a cada candidato**

Candidato	Cantidad de <i>tweets</i> que lo menciona
Fabricio Alvarado	850
Rodrigo Cháves	1.754
Eliécer Feinzaig	445
José María Figueres	197
Welmer Ramos	260
Lineth Saborío	352
José María Villalta	122
Categoría residual	515

Fuente: Elaboración propia con datos obtenidos de Twitter.

#### 4. MÉTODOS

Para analizar el corpus (colección de documentos textuales, en este caso, el texto de cada tweet), utilicé primeramente una técnica de

diccionario. En este tipo de análisis, el diccionario se refiere a una base de datos donde se clasifican palabras, por ejemplo, esta técnica es muy comúnmente usada para análisis de sentimiento. Para este caso, utilizo el diccionario y software de *Linguistic Inquiry and Word Count* (LIWC), particularmente la versión en español del diccionario (Ramírez-Esparza *et al.* 2007) y la versión del *software* LIWC-22.

LIWC clasifica cientos de miles de palabras en docenas de categorías relevantes para análisis psicométrico (Boyd *et al.* 2022), bajo la lógica general de que la proporción del tipo de palabras usadas (categorías gramaticales, funciones específicas o sentimientos) puede darnos indicaciones ya sea sobre la persona que escribe o sobre el tipo de lenguaje usado. En ese sentido, se acerca a una lógica de “bag-of-words” o “bolsa de palabras” utilizado en *machine learning* supervisado, es decir, la estructura de las oraciones, el orden de las palabras en sí mismo, no es tomado en cuenta.

Para el análisis, reviso tres índices calculados a partir de estudios previos relacionados con el proyecto LIWC: *Authenticity*, *Analytic* y *Tone*. La versión del programa LIWC-22 no calcula automáticamente estos índices al usar un diccionario en español, pero es trivial calcularlos con base en el porcentaje de cada texto que corresponde a ciertos tipos de palabras. *Authenticity* se refiere al grado en el cual las personas escriben sin autocensurarse, la fórmula para su cálculo fue identificada mediante una regresión logística con selección hacia adelante (Newman *et al.* 2003). Se reproduce la fórmula a continuación, cada variable corresponde a la proporción de ese tipo de palabras en un texto:

*Authenticity* = Pronombre de primera persona + palabras que indican intuición o percepción + palabras que indican diferenciación + palabras que indican relatividad – palabras que indican discrepancia – pronombres de tercera persona.

*Analytic* se refiere al nivel en que la persona utiliza pensamiento lógico y complejo; la fórmula para su cálculo fue identificada mediante un análisis de componentes principales (Pennebaker *et al.* 2014) y se reproduce a continuación, bajo la misma lógica de que cada variable corresponde a la proporción de ese tipo de palabras en un texto:

*Analytic* = 30 + artículos + preposiciones – pronombre personales

– pronombres impersonales – verbos auxiliares – conjunciones – adverbios – palabras de negación.

Cada uno de estos índices tiene un valor entre -100 y 100. *Tone* es el índice final y el más sencillo de todos y corresponde a un simple análisis de sentimiento: palabras con sentimiento negativo restadas a palabras con sentimiento positivo, su valor es un número entre 0 y 100.

Finalmente, como un ejercicio exploratorio, llevo a cabo un modelado de temas (*topic modeling*) en el corpus de *tweets* mediante una técnica de machine learning o aprendizaje automático no supervisado, específicamente uso la Asignación Latente de Dirichlet (*Latent Dirichlet Allocation*, LDA). Como otras formas de *topic modeling*, LDA es una técnica de reducción de dimensionalidad que busca encontrar temas latentes en los textos basados en la distribución de palabras usadas. Específicamente, LDA es un modelo generativo probabilístico bayesiano, donde la intuición básica es: En un corpus o colección de textos, cada texto es una mezcla al azar de una serie de temas latentes, y cada palabra tiene una distribución probabilística pertenencia a cada tema (Blei, Ng, y Jordan 2003).

LDA básicamente ayuda a encontrar temas latentes en los textos al encontrar las palabras más relevantes, siendo similar a un análisis de factores o de componentes principales, pero aplicado a textos. Para el análisis de LDA, utilizo el lenguaje de programación *Python*, utilizando los paquetes *Pandas* para estructura de datos, *Sklearn* para aprendizaje automático, *NLTK* para procesamiento de lenguaje natural y *Numpy* para cálculos matemático.

Como primer paso, extraje el texto de los *tweets* obtenidos durante los debates y los salvé como una lista en *Python*, tal que cada texto corresponde a un elemento de la lista y es de tipo 'string'. Como paso siguiente, realicé una 'tokenización', es decir, tomé cada texto y convertí cada palabra (no el termino) en una unidad de análisis o *token*. Por ejemplo, dado el texto "una manzana es una manzana", en esta frase existen tres términos: "una", "manzana" y "es". Sin embargo, existen cinco *tokens*: "una", "manzana", "es", "una" y "manzana". Para este proceso, utilicé el método *word\_tokenize* de *NLTK*.

Luego de esto, creé una matriz documento-término (*document-term matrix*) en la cual cada fila corresponde a cada texto único proveniente de un *tweet*, y cada columna corresponde a cada ter-

mino. Para este caso, es una matriz de 4495 x 324, indicando que el corpus es de 4495 textos (mismo número que el total de *tweets*, lo cual de paso indica que no se ha dado ningún error en la vectorización), y el vocabulario tiene 324 términos únicos. Para cada celda, calculé la medida conocida como *tf-idf* (*term frequency - inverse document frequency*, o frecuencia de término - frecuencia inversa de documento). La frecuencia de término se refiere al número de tokens que corresponden a cada termino en la columna, para el documento en la fila. Al volver al ejemplo de “una manzana es una manzana”, “manzana” tendría un *tf* de dos. La frecuencia inversa de documento es el número de documentos en el corpus donde aparece el término, dividido entre el número total de documentos en el corpus, bajo la idea de que, entonces, el *tf-idf* ayuda a restarle peso a palabras que aparecen en muchos de los documentos y no son, por tanto, relevantes para discriminar entre textos (por ejemplo la palabra “una”). La fórmula para *tf-idf*, donde *i* indica término, *j* documento y *n* el total de documentos (Aggarwal 2018; Bengfort, Bilbro, y Ojeda 2018), es:

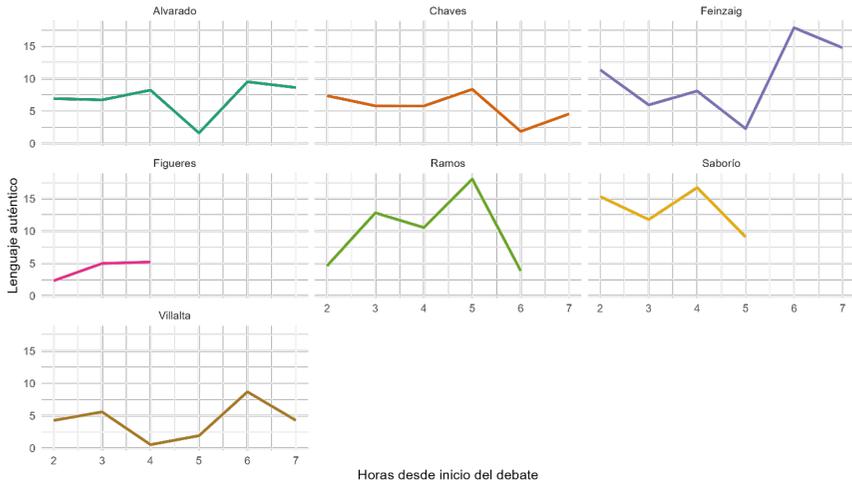
$$tf - idf = tf_{ji} * \log\left(\frac{n}{n_i}\right)$$

La matriz documento-término fue creada utilizando el método *TfidfVectorizer* de NLTK, de acuerdo con los siguientes parámetros: no se incluyeron en la matriz palabras que aparecen en menos de veinticinco documentos o palabras que aparecen en más del 90 % de los documentos; también se excluyeron a palabras en la lista de *stopwords* de NLTK para el idioma español (o sea, palabras demasiado comunes para tener un significado relevante, tales como “la”).

## 5. ANÁLISIS

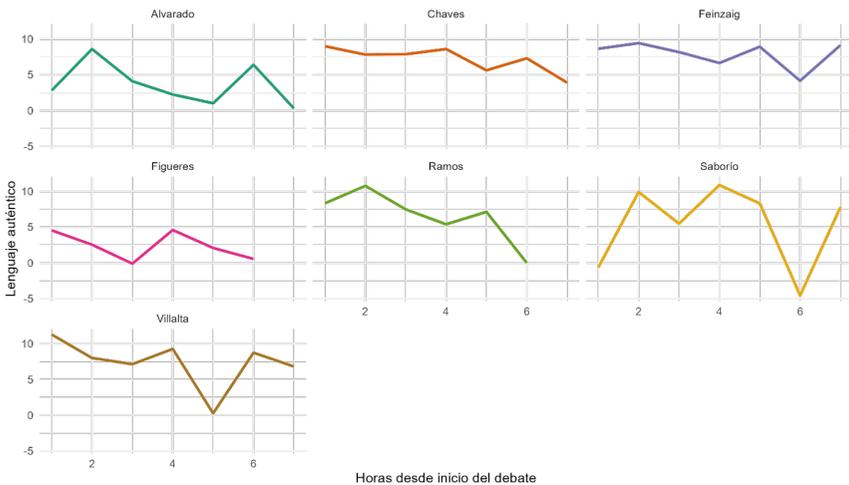
Para esta sección, se analizan en orden el uso de lenguaje auténtico (*Authenticity*), lenguaje analítico (*Analytic*) y el sentimiento (*Tone*) de los mensajes en Twitter sobre cada candidato; para cada tipo de medición, se estudian primero los mensajes que tuvieron lugar durante el debate de Repretel, seguido por los mensajes relativos al debate transmitido por Teletica. Reportaré, además, en cada caso, el promedio ( $\mu$ ) y la desviación estándar por candidato ( $\sigma$ ). Posterior a esto se presentan los resultados de la LDA.

**Figura 1. Nivel de lenguaje auténtico utilizado en el debate de Repretel**



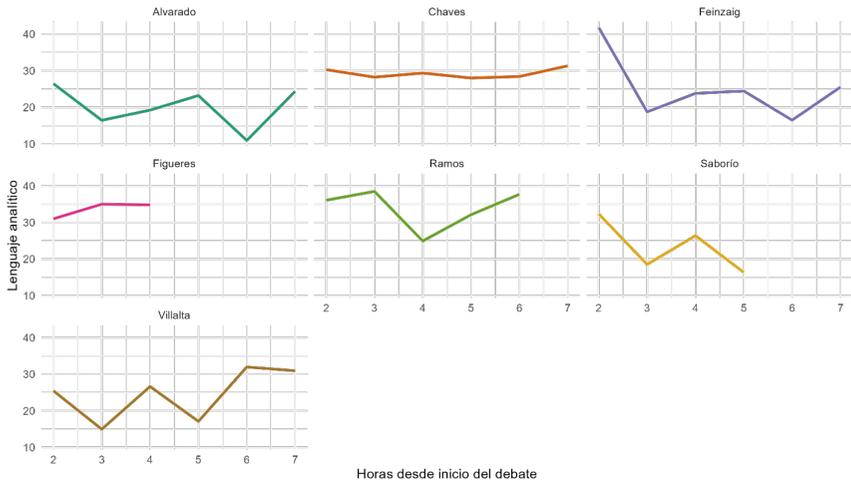
Fuente: Elaboración propia con datos obtenidos de Twitter.

**Figura 2. Nivel de lenguaje auténtico utilizado en el debate de Teletica**



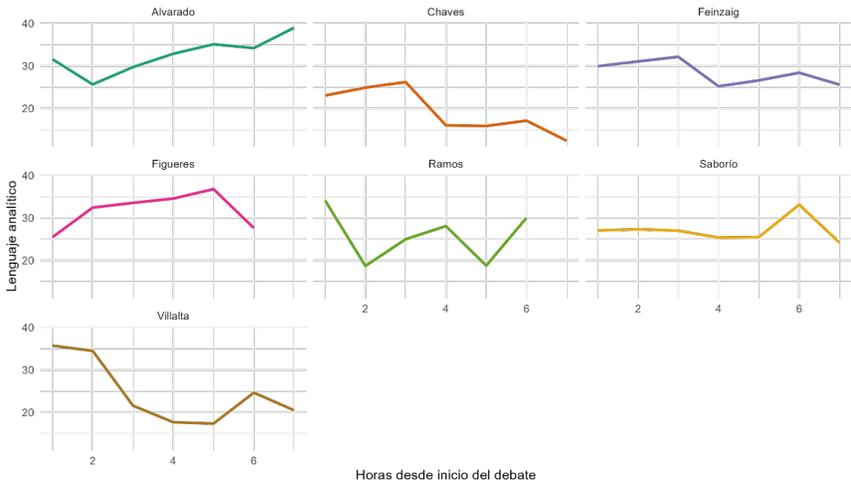
Fuente: Elaboración propia con datos obtenidos de Twitter.

**Figura 3. Nivel de lenguaje analítico utilizado en el debate de Repretel**



Fuente: Elaboración propia con datos obtenidos de Twitter.

**Figura 4. Nivel de lenguaje analítico utilizado en el debate de Teletica**



Fuente: Elaboración propia con datos obtenidos de Twitter.

## 5.1. Lenguaje auténtico

Como se puede ver en la figura 1 (página anterior), la variación en autenticidad en el tipo de lenguaje utilizado para hablar sobre los candidatos durante el debate de Repretel es baja, moviéndose entre 0 y 15 en la escala. Al estar sobre 0, esto indica que el lenguaje está por encima del punto medio y aparece como más sincero y poco autocensurado en vez de aparecer como premeditado. Debe de mencionarse que los mensajes mencionando únicamente a Figueres desaparecen temprano en la noche, y un poco después sucede lo mismo con Saborío, aunque los mensajes sobre ella muestran un nivel alto en lenguaje auténtico ( $c = 14,1$ ,  $\sigma = 10,1$ ). Otros dos casos en los cuales la conversación sobre candidatos presentó un nivel alto de lenguaje auténtico fue para Eliécer Feinzaig ( $c = 8,65$ ,  $\sigma = 11$ ) y Welmer Ramos ( $c = 10,5$ ,  $\sigma = 9,18$ ). Chaves ( $c = 6,22$ ,  $\sigma = 11,1$ ) y Alvarado ( $c = 6,6$ ,  $\sigma = 9,94$ ) presentan patrones similares, mientras que la conversación sobre Figueres ( $c = 3,98$ ,  $\sigma = 2,75$ ) y Villalta ( $c = 3,56$ ,  $\sigma = 6,23$ ) presenta los niveles más bajos de lenguaje auténtico en el debate de Repretel.

Para el caso del debate de Teletica (figura 2), se puede observar que el nivel de autenticidad en el lenguaje utilizado en Twitter para hablar de los candidatos está aún más acotado, por debajo de 10, pero aún por sobre 0 con una excepción, y en los casos se observa cierta tendencia hacia la baja. De esta manera, no hay mucha diferencia en los estadísticos descriptivos de cada caso, aunque sí es de importancia mencionar cómo Saborío ahora cae por debajo de 0 en un punto, y sus estadísticos son ( $c = 8,57$ ,  $\sigma = 11,1$ ). Como ya se dijo, todos los tweets sobre los candidatos presentan una tendencia hacia lenguaje cada vez menos auténtico, y mientras los tweets sobre Saborío, Ramos ( $c = 8,29$ ,  $\sigma = 12,4$ ), Feinzaig ( $c = 7,83$ ,  $\sigma = 9,62$ ) y Chaves ( $c = 7,68$ ,  $\sigma = 12,2$ ) mantienen estadísticos similares, los valores son mucho menores para Alvarado ( $c = 3,25$ ,  $\sigma = 8,3$ ) y Figueres ( $c = 2,36$ ,  $\sigma = 10,2$ ).

## 5.2. Lenguaje analítico

Con respecto al uso de lenguaje más analítico o complejo, como se observa en la figura 3, aquí sí es posible ver una mayor variación entre el lenguaje usado para hablar sobre los candidatos. Los *tweets* sobre Figueres son los que presentan lenguaje más analítico ( $c =$

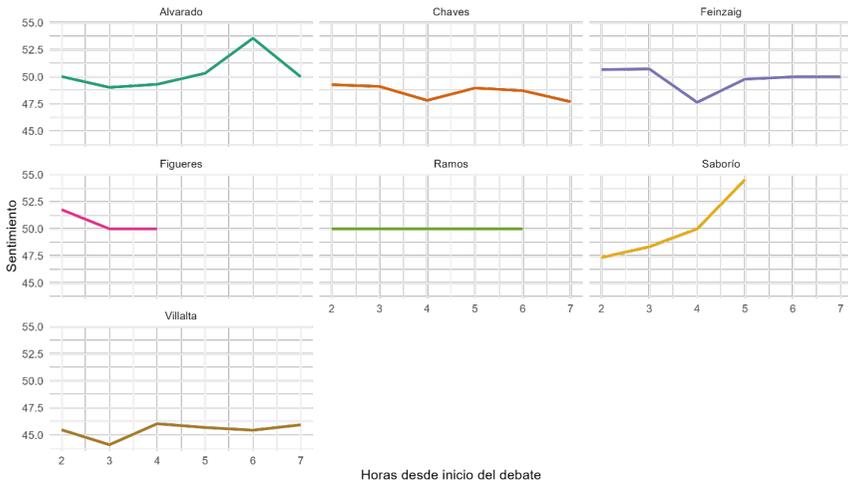
33,3,  $\sigma = 4,41$ ), aunque, como se mencionó antes, las personas dejaron de “twittear” mensajes que solo lo mencionan a él temprano en la noche, lo cual no nos permite saber si no hubiera presentado una evolución en alguna dirección a lo largo de la noche. Esto es de hecho lo interesante de la naturaleza de los datos como serie de tiempo, mientras que los estadísticos descriptivos no varían mucho (básicamente entre 20 y 30), sí podemos ver evoluciones interesantes, particularmente cómo el lenguaje utilizado para hablar sobre Feinzaig ( $c = 24,7$ ,  $\sigma = 13,4$ ), Saborío ( $c = 25,8$ ,  $\sigma = 19$ ) y Alvarado ( $c = 21,2$ ,  $\sigma = 16,3$ ) se vuelve menos analítico a lo largo de la noche, mientras que el lenguaje se vuelve cada vez más analítico para hablar de Ramos ( $c = 32,1$ ,  $\sigma = 14,3$ ), pero en especial para hablar sobre Villalta ( $c = 23,1$ ,  $\sigma = 13,6$ ). Por supuesto, también es importante notar los altos valores de las desviaciones estándar, que indican los saltos en las series de tiempo.

Para el caso del debate de Teletica (figura 4), tenemos totalmente un *reverse* en los casos de Alvarado y Villalta. Es decir, el tipo de lenguaje para hablar de Alvarado ( $c = 31,8$ ,  $\sigma = 16,2$ ) se vuelve cada vez más analítico, mientras que los mensajes sobre Villalta ( $c = 25,5$ ,  $\sigma = 17$ ) presentan una caída constante en su nivel de complejidad. También se presenta una caída constante en el nivel de complejidad en los *tweets* sobre Chaves ( $c = 21,5$ ,  $\sigma = 19,9$ ), llegando a los niveles más bajos que se registraron. Figueres ( $c = 33,6$ ,  $\sigma = 13,3$ ), por su parte, parece tener una leve mejora, Saborío ( $c = 26,5$ ,  $\sigma = 16,3$ ) no parece presentar mucha variabilidad, mientras que Ramos ( $c = 23,2$ ,  $\sigma = 18,6$ ) y Feinzaig ( $c = 28,3$ ,  $\sigma = 15,2$ ) no tiene patrones muy claros.

### 5.3. Análisis de sentimiento

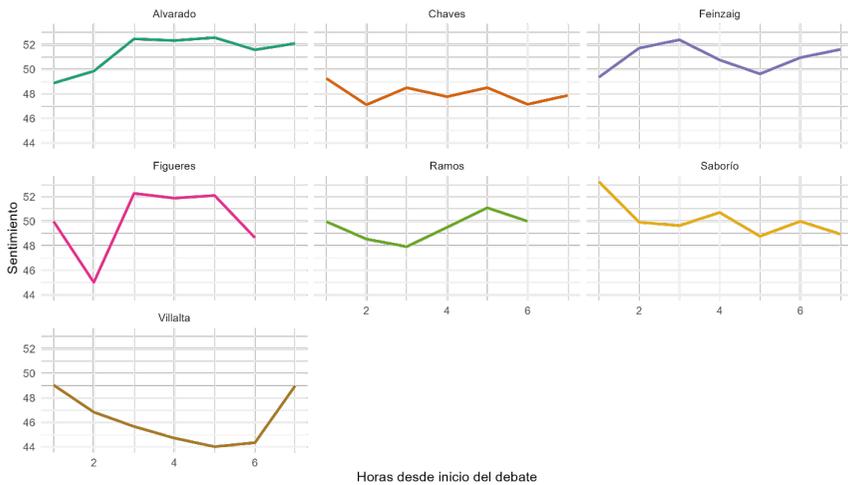
Con respecto al análisis de sentimiento o la tonalidad de los mensajes, las diferencias son minúsculas y se encuentran casi todos los mensajes alrededor de 50 en la escala, lo cual corresponde a un tono sumamente neutro. La figura 5 sí muestra un incremento en el sentido positivo con respecto a los *tweets* que mencionan a Alvarado ( $c = 49,8$ ,  $\sigma = 3,64$ ) y Saborío ( $c = 48,5$ ,  $\sigma = 3,75$ ), una más clara caída hacia sentimientos negativos en los *tweets* que mencionan a Chaves ( $c = 48,8$ ,  $\sigma = 3,81$ ), mientras que una vez más Figueres ( $c = 50,7$ ,  $\sigma = 1,79$ ) tiene pocas observaciones para el debate de Repretel, y Feinzaig ( $c = 49,8$ ,  $\sigma = 3,64$ ) y Ramos ( $c = 50$ ,  $\sigma = 0$ ) no presen-

**Figura 5. Nivel de análisis de sentimiento en el lenguaje utilizado en el debate de Repretel**



Fuente: Elaboración propia con datos obtenidos de Twitter.

**Figura 6: Nivel de análisis de sentimiento en el lenguaje utilizado en el debate de Teletica**



Fuente: Elaboración propia con datos obtenidos de Twitter.

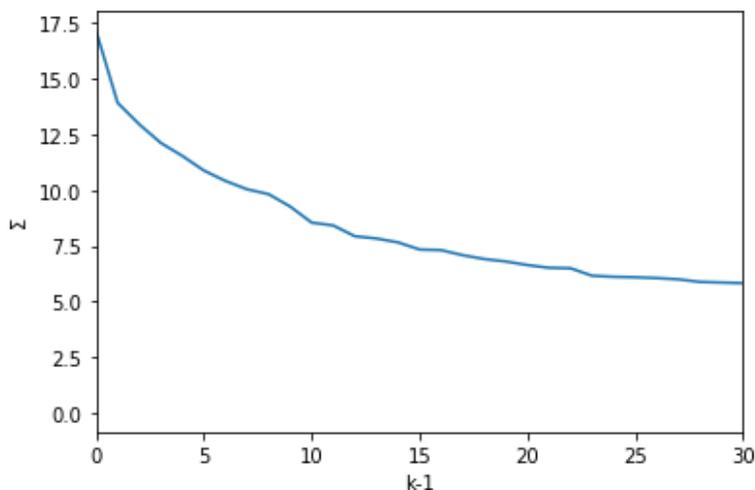
tan mucha variación interesante.

Para el caso del debate de Teletica (figura 6), se ve algo muy similar al caso del debate de Repretel: no hay mucha variación en los sentimientos de los *tweets* relevantes a los candidatos, y todos tienen un tono relativamente neutro. Claramente se observa un incremento en los sentimientos positivos sobre Alvarado ( $c = 52,1$ ,  $\sigma = 4,98$ ), Feinzaig ( $c = 51,1$ ,  $\sigma = 4,3$ ), Ramos ( $c = 48,6$ ,  $\sigma = 5,1$ ) y Figueres ( $c = 50,9$ ,  $\sigma = 5,44$ ), una disminución para Villalta ( $c = 45,8$ ,  $\sigma = 5,52$ ), Chaves ( $c = 47,8$ ,  $\sigma = 5,3$ ) y Saborío ( $c = 50,1$ ,  $\sigma = 4,78$ ), pero las diferencias propiamente son pequeñas.

#### 5.4. LDA

A la hora de realizar LDA, al igual que con otras formas de *topic modeling*, el número de componentes o temas (designados por  $k$ ) no puede obtenerse de forma automática por el algoritmo, sino que debe de ser elegido por el investigador. Dada la naturaleza exploratoria de este estudio, opto por realizar una descomposición en valores singulares (*single value decomposition*, SVD) en la matriz documento-término para averiguar, mediante un análisis de *scree plot*, cuantos temas utilizar.

**Figura 7: Scree plot de la descomposición en valores singulares de la matriz documento-término**



Fuente: Elaboración propia con datos obtenidos de Twitter.

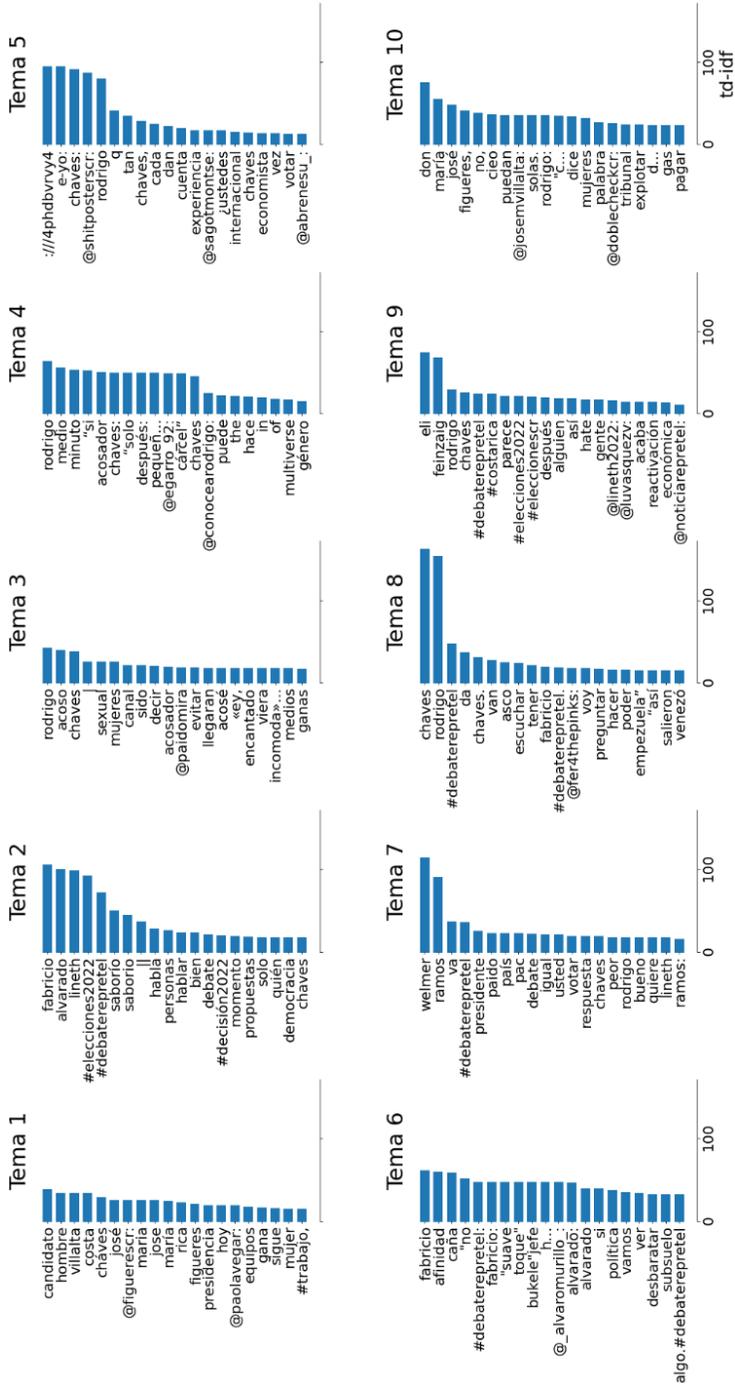
SVD es una técnica de descomposición matricial similar al análisis de factores y al análisis de componentes principales, ha sido utilizada para modelación de temas y, en este caso, la utilizo de forma exploratoria para obtener la figura 7 (página anterior), un *scree plot* que indique una cantidad de temas apropiados para utilizar como parámetros en la LDA. El análisis del *scree plot* es completamente visual, el eje x indica la cantidad de temas, o sea, de valores propios de la matriz, ordenados desde el que explica la mayor cantidad de variación hacia abajo, y se escoge la cantidad de temas correspondientes a la última caída relevante en la magnitud de valor propio, indicada en el eje y. Basado en esto, realizo la LDA con diez temas.

La figura 8 (página siguiente) muestra las veinte palabras más relevantes según el tf-idf para cada tema obtenido mediante LDA. El primer tema parece unir a Figueres y a Villalta, probablemente por compartir primer nombre y, además, hace referencia en varias palabras a la elección y el país, dado la alta magnitud del valor propio, puede ser una especie de tema “atrapa todo”. El segundo tema también tiene varias referencias a la elección, pero parece enfocarse mucho tanto en Alvarado como en Saborío, potencialmente porque se estaba hablando en conjunto o referenciando a ambos candidatos en los mismos *tweets*, o en *tweets* similares.

El tercer tema es sumamente interesante. Es claramente el tema del actual presidente, Rodrigo Chaves, e incluye varias referencias sobre acoso sexual, entre las palabras más utilizadas están “acoso”, “sexual”, “mujeres”, “acosador” y “acosé”. También aparece “incomoda”, la cual puede acercarse al tema de acoso sexual, pero también al discurso de Chaves de cómo su candidatura incomoda a grupos de poder económico en el país, por ejemplo, también aparece la palabra “medios”. El tema cuatro presenta muchas similitudes con el tres, al estar también enfocado en Chaves y tener algunos términos sin relevancia política en inglés (*multiverse*), lo cual sugiere que podría ser preferible optar por un k menor en el LDA, también nótese que alrededor de este punto hay una caída fuerte en la magnitud de los valores propios.

El quinto tema es también poco interesante; sin embargo, el sexto tema parece enfocarse en Alvarado y la mención de Bukele y la palabra “afinidad” pueden deberse a que se hablaba de la afinidad que existe entre el candidato de Nueva República y el presidente de El Salvador. Los temas restantes parecen capturar a candidatos específicos, Ramos en el séptimo y Figueres en el décimo, el tema 8

Figura 8: Temas en el modelo LDA



Fuente: Elaboración propia con datos obtenidos de Twitter.

parece ser negativo tanto hacia Chaves como Alvarado, al incluir palabras como “asco”, “venezó” y “empezuela”. El tema nueve menciona a Feinzaig, Chaves y temas económicos, tal vez por el esperado *expertise* o *valence* de ambos candidatos en dicho tema.

## 6. DISCUSIÓN

Lo primero a mencionar con respecto al análisis de LIWC es que la variación entre la forma de hablar sobre los candidatos es baja, no importa si el indicador se refiere a la autenticidad, nivel de análisis o el sentimiento presentado en el texto de los *tweets*. El nivel de lenguaje auténtico es un proxy para indicarnos que no hay mucho planeamiento previo, es decir, que la gente se está comunicando de una forma honesta. Bajo esta lógica, no existe inferencia causal, pero niveles bajos en *Authenticity* podría indicar *bots* o personas pagadas para seguir guiones en su forma de postear.

El hecho de que el nivel de lenguaje auténtico se mantenga por arriba de 0 es interesante, pues sugiere que al menos el lenguaje tienda a ser más auténtico que inauténtico, lo cual propone que tal vez no hay intervención de *bots* o agentes pagos en la conversación, o al menos no de agentes poco sofisticados, sino agentes que pueden emular a un usuario estándar. Es curiosa la tendencia hacia la baja durante el debate de Teletica, particularmente por el mayor tráfico durante este, pero el cambio no es tan grande para inferir de forma exploratoria alguna hipótesis.

El nivel de lenguaje analítico indica qué tan compleja y racional fue la conversación alrededor de cada candidato. El nivel y variación es más alta en este caso comparado con el lenguaje auténtico. En el caso de Repretel, la conversación alrededor de Feinzaig, Saborrío y Alvarado se volvió cada vez menos analítica, mientras que lo inverso sucedió con Villalta. Sin embargo, para el debate transmitido por Teletica, la situación se invierte para Alvarado y Villalta, y muy importantemente, hay una constante caída para Chaves.

Lo anterior es interesante, ya que una conversación que se caracteriza por ser seria, respetuosa y enfocada en política pública se distingue por altos niveles de lenguaje analítico, mientras que conversaciones más emotivas o combativas tendrían niveles bajos en esta escala. Particularmente, Feinzaig, Alvarado y Chaves serían los candidatos naturales a ser atacados por grupos progresistas en redes sociales, al ser los candidatos libertario, neopentecostal y con

acusaciones de acoso sexual, respectivamente. Por su parte, Villalta sería el candidato natural por atacar para usuarios conservadores, al ser el candidato de un partido socialista. Esto podría explicar los cambios en esta escala.

El análisis de sentimiento presenta variaciones sumamente bajas entre candidatos, el lenguaje no es ni muy positivo ni muy negativo. Al menos en este sentido, no considero que se pueda ver un grado fuerte de apoyo o desagrado hacia algunos de los candidatos. Si bien se mantiene un poco la lógica mencionada con respecto al lenguaje analítico, por ejemplo, en el creciente sentimiento positivo hacia Alvarado y el decreciente hacia Chaves, los cambios son bajos.

El análisis LIWC se enfoca únicamente en *tweets* que tratan sobre un solo candidato. La LDA puede darnos pistas sobre la conversación relevante a más de uno, ya que, inclusive, vemos temas que engloban a varios candidatos. Sin embargo, el hecho de que varios temas traten solo sobre un candidato, por ejemplo, el sexto, el séptimo y el décimo, sugieren que al menos los temas parecen realmente separarse entre los candidatos, es decir, que se hablaba de uno a la vez, pero no de varios.

Además, una vez más se nota el total dominio de Chaves como tema de conversación, al ocupar varios temas. Esto no significa que la conversación fuera positiva, como ya se menciona, es fuerte su ligamen con las acusaciones de acoso sexual en la conversación en Twitter durante los debates. Es interesante considerar la contradicción presentada aquí: El análisis de sentimientos de LIWC de los *tweets* que solo mencionan a Chaves no sugiere que existe un fuerte sentimiento en contra de él, sin embargo, la LDA de todos los tweets claramente lo ligan con un tema que tendría una polaridad negativa: acoso sexual.

Además, es importante notar cómo existe congruencia en al menos uno de los temas identificados por la LDA y los que han sido identificados como preponderantes durante el proceso de elecciones por la literatura especializada, por ejemplo, en los artículos previamente citados de Treminio Sánchez (2022) y de Ovares Sánchez (2022). Particularmente, se ve una congruencia en la importancia que tuvo la discusión alrededor de las acusaciones de acoso sexual en contra de Rodrigo Chaves. Además, no debe dejar de mencionarse que el tráfico de conversación alrededor de Chaves es mucho mayor que para el caso de cualquier otro candidato. No es posible

explorar en este trabajo los motivos de esto, pero es relevante saber que la conversación estaba concentrada en la persona que eventualmente ganó la contienda electoral, independientemente de si la conversación giró alrededor de las acusaciones

## 7. CONCLUSIONES

Esta nota de investigación está planteada con dos subobjetivos. El primero de estos es sustantivo, y consiste en explorar cómo se desarrolló la conversación alrededor de los candidatos presidenciales del 2022 en la red social Twitter durante los debates presidenciales de Repretel y Teletica. El segundo subobjetivo es metodológico, y se centra en ser descriptivo y didáctico en el uso de dos técnicas computacionales de análisis de texto, el análisis de diccionario utilizando LIWC y la Asignación Latente de Dirichlet. Esto plantea a este estudio dentro de la creciente literatura en la ciencia política costarricense que busca aprovechar nuevos métodos tomados de ciencias computacionales y de datos, así como del acceso a mayores cantidades de datos, que han sido producto tanto de innovaciones tecnológicas así como del uso creciente de redes sociales por parte de la ciudadanía.

En el texto, describo el uso de un paquete del lenguaje de programación R para acceder al API de Twitter y descargar los datos de texto y metadatos de mensajes enviados por esa plataforma durante la duración de los debates, que mencionaban al menos a uno de los candidatos presidenciales que participaron en los debates. Describo también los diversos pasos para la preparación de los datos para su análisis, seguido de una explicación de los indicadores psicométricos calculados mediante el diccionario de LIWC, los cuales indican el grado en que una persona escribe sin autocensurarse (es decir, de forma espontánea), la complejidad del lenguaje y, finalmente, el sentimiento de un texto. Asimismo, también se incluye una explicación del LDA como una técnica de aprendizaje automático no supervisado para identificar temas en un cuerpo de textos, lo cual se logra mediante la representación del cuerpo de textos como una matriz documento-término y el cálculo del tf-idf, o frecuencia de término - frecuencia inversa de documento.

El análisis de LIWC indica que la variación con respecto a cómo se habla sobre cada candidato es baja. No es posible considerar a los resultados como una prueba de lo siguiente, pero estos pare-

cen al menos sugerir que, entre los mensajes de Twitter analizados, estos no fueron mayoritariamente creados por cuentas automatizadas. Es decir, los análisis exploratorios son más consistentes con la idea de que la mayoría de los usuarios que tuitearon durante los debates eran, efectivamente, humanos, y no cuentas de *bots*. En nivel de lenguaje analítico, cae alrededor de candidatos que podrían considerarse "controversiales" en tanto a pertenecer a tendencias ideológicas minoritarias (libertaria, socialista o neopentecostal) o por controversias relacionadas con el candidato, como con el caso de las acusaciones de acoso sexual contra Rodrigo Chaves. Sobre este último punto, las acusaciones contra Chaves son identificadas por la LDA como uno de los principales temas de conversaciones, lo cual es congruente con la literatura sobre las elecciones, y el dominio de Chaves como tema de conversación, lo cual queda claro tanto con ver la LDA como con el número de tweets que lo mencionan.

## REFERENCIAS

- Aggarwal, Charu C. 2018. *Machine Learning for Text*. Cham: Springer International Publishing. <https://doi.org/10.1007/978-3-319-73531-3>.
- Alvarenga Forunier, Ximena, y Ronald Saéñz Leandro. 2022. «Lecturas Metodológicas Desde El NLP : Estudiando YouTube Como Espacio de Recepción de La Protesta Social En Costa Rica En Tiempos de COVID-19». *Anuario del Centro de Investigación Estudios Políticos*, n.º 13: 233–79. <https://doi.org/10.15517/aciep.v0i13.47564>.
- Arce Chaves, Laura, y Cristian Bonilla Cruz. 2023. «Observatorio de Comunicación Digital: Influencia de Las Redes Sociales En La Primera Ronda Electoral Del 2022 En Costa Rica». *Revista de Derecho Electoral*, n.º enero: 95–119. [https://doi.org/10.35242/RDE\\_2023\\_35\\_6](https://doi.org/10.35242/RDE_2023_35_6).
- Bail, Christopher A. 2016. «Emotional Feedback and the Viral Spread of Social Media Messages about Autism Spectrum Disorders». *American Journal of Public Health* 106, n.º 7: 1173–80. <https://doi.org/10.2105/AJPH.2016.303181>.
- Bail, Christopher A., Taylor W. Brown y Marcus Mann. 2017. «Channeling Hearts and Minds: Advocacy Organizations, Cognitive-Emotional Currents, and Public Conversation». *American Sociological Review* 82, n.º 6: 1188–1213. <https://doi.org/10.1177/0003122417733673>.
- Bengfort, Benjamin, Rebecca Bilbro y Tony Ojeda. 2018. *Applied Text Analysis with Python: Enabling Language-Aware Data Products with Machine*

- Learning*. O'Reilly Media, Inc.
- Blei, David M., Andrew Y. Ng, y Michael I. Jordan. 2003. «Latent Dirichlet Allocation». *Journal of Machine Learning Research*, n.º 3: 993–1022.
- Boyd, Ryan L., Ashwini Ashokkumar, Sarah Seraj y James W. Pennebaker. 2022. *The Development and Psychometric Properties of LIWC-22*. Austin, TX: University of Texas at Austin. <https://www.liwc.app>.
- Bunse, Simone. 2021. *Global Cyber Troops Country Profile: Costa Rica*. UK: Oxford University.. <https://comprop.oii.ox.ac.uk/wp-content/uploads/sites/127/2021/01/Cyber-Troop-Costa-Rica-2020.pdf>.
- Carazo Barrantes, Carolina, Larissa Tristán Jiménez y Ignacio Siles González. 2021. «En Palabras de Los Candidatos : La Comunicación En Redes Sociales Durante La Campaña Electoral Costarricense de 2018». *Anuario del Centro de Investigación Estudios Políticos*, n.º 12: 1–33. <https://doi.org/10.15517/ACIEP.V0I12.43489>.
- Centro de Investigación y Estudios Políticos - Universidad de Costa Rica. 2018. *Informe de Resultados de La Encuesta de Opinión Sociopolítica, Enero 2018*. San José, Costa Rica: Universidad de Costa Rica. <https://ciep.ucr.ac.cr/index.php/proyectos/encuestas-de-opinion>.
- Cruz Romero, Roberto. 2017a. «Política Digital: El Uso de Facebook En Política Electoral En Costa Rica (I)». *Revista de Derecho Electoral*, n.º 23: 133–51.
- . 2017b. «Política Digital: El Uso de Facebook En Política Electoral En Costa Rica (II)». *Revista de Derecho Electoral* 2017, n.º 24: 83–106.
- Gómez Campos, Steffan, y Esteban Durán Monge. 2020. «Construcción de La Imagen Pública En Redes Sociales (Campaña Electoral 2017-2018)». En *Elecciones 2018 en Costa Rica : Retrato de una Democracia Amenazada*, editado por Ronald Alfaro-Redondo y Felipe Alpizar Rodríguez, 109–49. San José, Costa Rica: CONARE - PEN.
- Jordan, Kayla N., James W. Pennebaker y Chase Ehrig. 2018. «The 2016 U.S. Presidential Candidates and How People Tweeted About Them». *SAGE Open* 8, n.º3. <https://doi.org/10.1177/2158244018791218>.
- Jordan, Kayla N., Joanna Sterling, James W. Pennebaker y Ryan L. Boyd. 2019. «Examining Long-Term Trends in Politics and Culture through Language of Political Leaders and Cultural Institutions». *Proceedings of the National Academy of Sciences of the United States of America* 116, n.º 9: 3476–81. <https://doi.org/10.1073/pnas.1811987116>.
- Newman, Matthew L., James W. Pennebaker, Diane S. Berry y Jane M. Richards. 2003. «Lying Words: Predicting Deception from Linguistic Styles». *Personality and Social Psychology Bulletin* 29, n.º 5: 665–75. ht-

[tps://doi.org/10.1177/0146167203029005010](https://doi.org/10.1177/0146167203029005010).

- Ovares Sánchez, Carolina. 2022. «Elecciones Nacionales Costa Rica 2022: Entre Fuerzas Políticas Tradicionales y Fuerzas Emergentes». *Revista Elecciones* 21, n.º 23: 253–66. <https://doi.org/10.53557/elecciones.2022.v21n23.09>.
- Pennebaker, James W., Cindy K. Chung, Joey Frazee, Gary M. Lavergne y David I. Beaver. 2014. «When Small Words Foretell Academic Success: The Case of College Admissions Essays». *PLoS ONE* 9, n.º 12: 1–10. <https://doi.org/10.1371/journal.pone.0115844>.
- Perelló, Lucas y Patricio Navia. 2021. «Abrupt and Gradual Realignments: The Case of Costa Rica, 1958–2018». *Journal of Politics in Latin America* 13, n.º 1: 86–113. <https://doi.org/10.1177/1866802x20967733>.
- Programa Estado de la Nación en Desarrollo Humano Sostenible. 2018. «Informe Estado de La Nación 2018». San José, Costa Rica.
- . 2022. «Informe Estado de La Nación 2022». San José, Costa Rica.
- Ramírez-Esparza, Nairán, James W Pennebaker, Florencia Andrea García y Raquel Suriá. 2007. «La Psicología Del Uso de Las Palabras: Un Programa de Computadora Que Analiza Textos En Español». *Revista Mexicana De Psicología* 24, n.º 1: 85–99.
- Siles Gonzáles, Ignacio, ed. 2020. *Democracia en Digital: Facebook, Comunicación y Política en Costa Rica*. San José, Costa Rica: Universidad de Costa Rica, Centro de Investigación en Comunicación.
- Siles González, Ignacio, Carolina Carazo Barrantes, y Larissa Tristán Jiménez. 2019. «Comunicación y Política En Clave Digital: Las Redes Sociales y El Proceso Electoral 2017-2018». En *Tiempos de Travesía. Análisis de las Elecciones del 2018 en Costa Rica*, editado por Manuel Rojas Bolaños and Ilka Treminio Sánchez, 175–96. San José, Costa Rica: FLACSO.
- Treminio Sánchez, Ilka. 2022. «Elecciones Presidenciales y Legislativas Costa Rica 2022». *Análisis Carolina*, n.º 5: 1. [https://doi.org/10.33960/AC\\_05.2022](https://doi.org/10.33960/AC_05.2022).

**ELÍAS CHAVARRÍA-MORA** es costarricense. *Master of Arts* en Ciencia Política por la Universidad de Pittsburgh, y licenciatura y bachillerato en ciencias políticas por la Universidad de Costa Rica. Actualmente es candidato a doctor en ciencia política por la Universidad de Pittsburgh, anteriormente ha sido asistente de investigación en el Programa Estado de la Nación del Consejo Nacional

ELÍAS CHAVARRÍA-MORA

de Rectores (CONARE) y en el Centro de Investigación y Estudios Políticos de la Universidad de Costa Rica.

**ORCID:** <https://orcid.org/0000-0001-6424-3915>.