

POPULARIDAD PRESIDENCIAL EN AMÉRICA LATINA ANÁLISIS DE SENTIMIENTO EN TWITTER*

Lic. Chirino Gutiérrez, Álvaro Limber

✉ *achirino@aru.org.bo*

RESUMEN

Este artículo presenta una propuesta para medir la popularidad de ocho presidentes de países de América Latina empleando la información de Twitter mediante el uso de scraping web en R. Las medidas están basadas en base a seguidores, favoritos, *retweets* y un análisis de sentimiento de los *tweets* de los usuarios hacia los presidentes. La información corresponde al mes de septiembre de 2019.

PALABRAS CLAVE

Twitter, presidentes América Latina, scraping, estadística, minería de texto, análisis de sentimiento.

ABSTRACT

This article presents a proposal to measure the opinions of eight presidents of Latin American countries using the information of Twitter through the use of web scraping in R. The measures are based on the base of followers, favorites, retweets and an analysis of feelings of the tweets of the users towards the presidents. The information corresponding to the month of September 2019.

KEYWORDS

Twitter, Latin American presidents, scraping, statistics, text mining, sentiment analysis.

1. MOTIVACIÓN

Actualmente las redes sociales se han convertido en una ventana para que las personas interactúen con una fluidez sin precedente, la llegada del internet, los teléfonos inteligentes han acelerado la comunicación. Una de las redes sociales más populares es el *Twitter*, esta plataforma permite a sus usuarios crear miniblogs o mensajes limitados a 140 caracteres y publicarlos de manera pública o privada, no existe una interacción horizontal entre los usuarios, la relación que existe es de carácter vertical dado que los usuarios deciden a que usuario seguir.

La manera que existe para interactuar es mediante los *tags* (etiquetas) por ejemplo el tag #Bolivia está orientado a etiquetar el mensaje con contenido relacionado al *tag*, otro ejemplo son los tags como @evoespueblo

que usan los usuarios para crear mensajes que incluyen al usuario del *tag*.

El *Twitter* es una de las herramientas empleadas por los presidentes de los distintos países, en América Latina todos los presidentes tienen cuentas activas, en este documento se exploran los datos de ocho presidentes; Mauricio Macri (Argentina), Evo Morales (Bolivia), Mario Abdo (Paraguay), Nicolás Maduro (Venezuela), Sebastián Piñera (Chile), Martin Vizcarra (Perú), Iván Duque (Colombia) y Lenin Moreno (Ecuador). La cuenta más antigua corresponde a Sebastián Piñera en febrero de 2008 y la más reciente cuenta corresponde a Evo Morales en abril de 2016.

Este documento presenta medidas de popularidad presidencial de ocho presidentes latinoamericanos, la información proviene de las cuentas de *Twitter* y fue obtenida

* Este documento está en el marco del proyecto de investigación: “Aplicación del Web Scraping en la Estadística”

mediante el uso del *scraping web* en R. En 2 se presentan los objetivos y alcances del trabajo, en 3 se detalla la metodología, en 5 presentan los resultados y finalmente en 6 se describen los hallazgos y recomendaciones del estudio.

2. OBJETIVOS Y ALCANCES

El objetivo central del documento es **desarrollar criterios para medir popularidad a nivel de los presidentes de ocho países latinoamericanos empleando información proveniente del Twitter.**

La información del *Twitter* es extraída mediante el uso de *Scraping web* usando el software estadístico R.

En cuanto a los alcances:

- La cobertura temporal de la información extraída del *Twitter* corresponde al mes de septiembre de 2019.
- Se define al español como el criterio de búsqueda para los presidentes.
- Se emplean el lexicón nrc¹ para el análisis de sentimiento.

3. METODOLOGÍA

El *scraping web* con R permite tener acceso a diferentes datos dentro del *Twitter*, a partir de esto se definen las siguientes medidas para medir la popularidad de los ocho presidentes.

- Basado en seguidores, favoritos y *retweets*
 1. Porcentaje de seguidores respecto la población total del país (*followers*).
 2. Número de favoritos por cada 100.000 seguidores basado en el promedio de los últimos 20 *tweets* (*fav20*).
 3. Número de favoritos por cada 100.000

seguidores basado en el promedio de los últimos 200 *tweets* (*fav200*).

4. Número de *retweets* por cada 100.000 seguidores basado en el promedio de los últimos 20 *tweets* (*retweets20*).
 5. Número de *retweets* por cada 100.000 seguidores basado en el promedio de los últimos 200 *tweets* (*retweets200*).
- Basado en los mensajes de los presidentes y de los usuarios hacia los presidentes
 1. *Wordclouds* de los tweets presidenciales
 2. Análisis de sentimiento en base a los *tweets* de los usuarios hacia los presidentes.

Donde:

$$followers_i = \frac{S_i}{P_{2019,i}} \quad (1)$$

$$fav\bar{20}_i = \frac{fav\bar{20}_i}{S_i} * 100.000 \quad (2)$$

$$fav\bar{200}_i = \frac{fav\bar{200}_i}{S_i} * 100.000 \quad (3)$$

$$retweets20_i = \frac{ret\bar{20}_i}{S_i} * 100.000 \quad (4)$$

$$retweets200_i = \frac{ret\bar{200}_i}{S_i} * 100.000 \quad (5)$$

Con S_i la cantidad de seguidores registrados a septiembre de 2019 del presidente de i , $P_{2019,i}$ es la población proyectada del país del presidente i para el 2019. $fav\bar{20}_i$ y $fav\bar{200}_i$ el promedio de favoritos de los últimos 20 y 200 *tweets* presidenciales respectivamente. $ret\bar{20}_i$ y $ret\bar{200}_i$ el promedio de *retweets* de los últimos 20 y 200 *tweets* presidenciales respectivamente.

3.1. WORDCLOUDS

Los *wordclouds* presidenciales son nubes de palabras que se construyen a partir de

¹ El NRC Emotion Lexicon es una lista de palabras en inglés y sus asociaciones con ocho emociones básicas <https://saifmohammad.com/WebPages/NRC-Emotion-Lexicon.htm>

Popularidad presidencial en América Latina análisis de sentimiento en Twitter

las frecuencias de ocurrencias de palabras provenientes de los *tweets* que publican los presidentes.

3.2. ANÁLISIS DE SENTIMIENTO

Para el análisis de sentimientos de los *tweets* de los usuarios hacia los presidentes se siguen los siguientes pasos:

1. Se extraen los *tweets* de los usuarios hacia los presidentes, no se toma en cuenta los *retweets*
2. Se eliminan caracteres, números y stopword de cada *tweet*. Quedando únicamente las palabras de interés
3. En base al *lexicón nrc* se identifican ocho emociones para cada palabra, los sentimientos son: enojo, expectación, disgusto, miedo, alegría, tristeza, sorpresa y confianza.
4. Se agregan las ocho emociones para todos los *tweets* y se genera la proporción de emociones global.

4. DATOS

Los datos provienen de las siguientes fuentes:

1. Las cuentas de *Twitter* de ocho presidentes, la información corresponde a los *tweets* y actualizaciones hasta el mes de septiembre de 2019. El total de *tweets*

explorados es de 20.699, no incluye los *retweets* de los presidentes.

2. Los *tweets* provenientes de los usuarios en donde se emplea un *tag* para alguno de los ocho presidentes. El total de *tweets* explorados es de 275.536, no incluye los *retweets* que realizan los usuarios y corresponde al mes de septiembre.
3. Las proyecciones de población provienen de los datos del Banco Mundial mediante la librería *wbstats* de R.

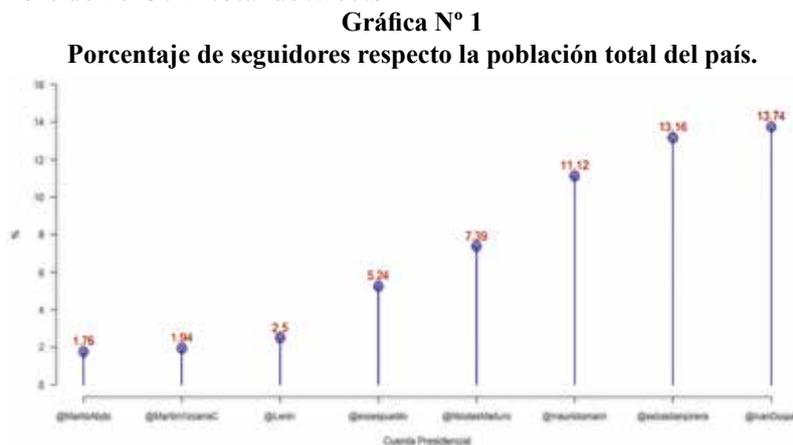
La información fue recolectada empleando R con las librerías *twitterR*, *wordcloud*, *rvest*, *tidyverse*, *syuzhet* y *wbstats*.

5. RESULTADOS

Siguiendo la metodología descrita y en base a los datos recolectados, en esta sección se presentan los resultados.

5.1. BASADO EN SEGUIDORES, FAVORITOS Y RETWEETS

La Gráfica N° 1 presenta el indicador *follower* por presidente, los resultados se presentan de forma ascendente, se aprecia que los presidentes con más seguidores respecto la población del país son Iván Duque (Colombia) y Sebastián Piñera (Chile) mientras los con menos seguidores



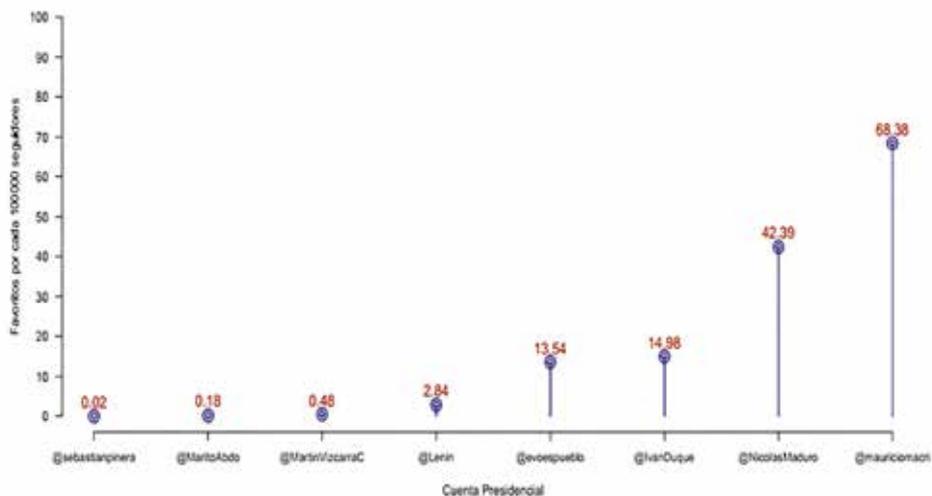
Fuente: Elaboración del Autor

son Mario Abdo (Paraguay) y Martin Vizcarra (Perú).

Las Gráficas N° 2 y N° 3 presentan los indicadores basados en la cantidad de favoritos, en ambas figuras se aprecia que los presidentes con los valores más altos son Nicolás Maduro (Venezuela) y Mauricio Macri (Argentina) mientras los más bajos son Sebastián Piñera (Chile) Martín Viscarra (Perú) y Mario Abdo (Paraguay). Notar que la cantidad de favoritos

Gráfica N° 2

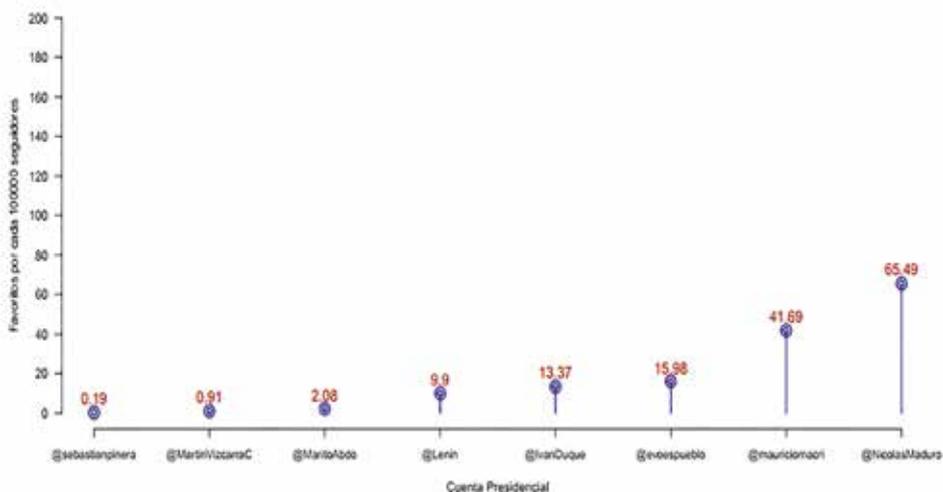
Promedio del Número de Favoritos por cada 100.000 seguidores basado en los últimos 20 tweets presidenciales.



Fuente: Elaboración del Autor

Gráfica N° 3

Promedio del Número de Favoritos por cada 100.000 seguidores basado en los últimos 200 tweets presidenciales.



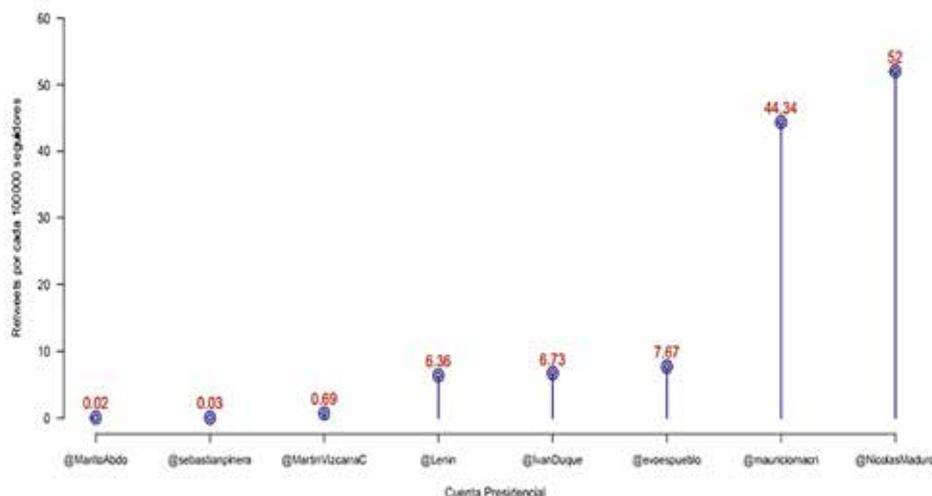
Fuente: Elaboración del Autor

por cada 100.000 seguidores es muy bajo en todos los casos, logrando en el mejor de los casos 69 favoritos..

Las Gráficas N° 4 y N° 5 presentan los indicadores basados en la cantidad de retweets, el

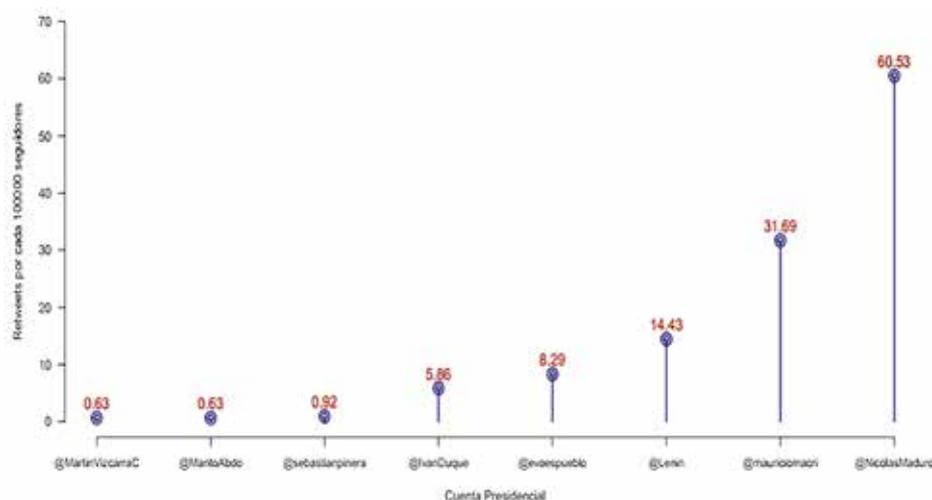
Popularidad presidencial en América Latina análisis de sentimiento en Twitter

Gráfica N° 4
Numero de Retweets por cada 100.000 seguidores basado en el promedio de los últimos 20 tweets presidenciales



Fuente: Elaboración del Autor

Gráfica N° 5
Numero de Retweets por cada 100.000 seguidores basado en el promedio de los últimos 200 tweets presidenciales



Fuente: Elaboración del Autor

comportamiento es similar a los indicadores de favoritos tanto en los presidentes con valores más altos y bajos.

5.2. BASADO EN LOS MENSAJES DE LOS PRESIDENTES Y DE LOS USUARIOS HACIA LOS PRESIDENTES

En la Gráfica N° 6 se presenta las nubes de palabras basadas en los *tweets* de los ocho presidentes, mientras que en la Gráfica N° 7 se presenta la nube de palabras de los

usuarios.

Para el análisis de sentimiento se empleó el gráfico de telaraña que permite visualizar las ocho emociones, esto se presenta en la Gráfica N° 8. El patrón recurrente es que la emoción con mayor frecuencia es la confianza seguida de la tristeza y el miedo, la presencia marcada de la confianza puede deberse en parte a los *tweets* de las distintas entidades públicas de los distintos países,

Gráfica N° 8
Emociones hacia los presidentes de parte de los usuarios



(a) Mauricio Macri



(b) Evo Morales



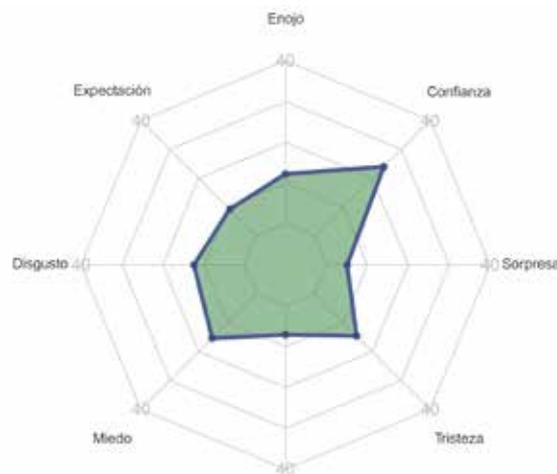
(c) Mario Abdo



(d) Nicolás Maduro



(e) Sebastian Piñera

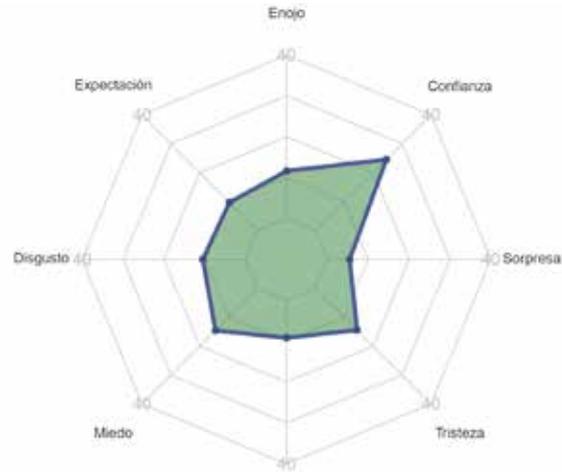


(f) Martín Viscarra

Popularidad presidencial en América Latina análisis de sentimiento en Twitter



(g) Iván Duque



(h) Lenin Moreno

Fuente: Elaboración del Autor

para siguientes aproximaciones se podría explorar la relación entre la emoción y los horarios en los que se publican.

6. HALLAZGOS Y RECOMENDACIONES

Los hallazgos del documentos pueden resumirse en (1) El *Twitter* es un espacio de generación de información masiva que brinda la oportunidad de explorar diversos aspectos de la población vinculada a ella, (2)

el R es una de las herramientas estadísticas actuales con mayor versatilidad para explorar las nuevas tendencias de las ciencia de datos y (3) el análisis de sentimiento es una técnica dentro de la minería de texto que aún necesita adaptaciones para el contexto boliviano. En cuanto a las recomendaciones para futuros estudios (1) incorporar el análisis de sentimiento diferenciando las horas y días para evaluar si las emociones están condicionadas a ciertos momentos del tiempo y (2) a partir de los indicadores es posible realizar un monitoreo diario.

BIBLIOGRAFÍA

Chen, L.-P. (2019). Text mining in practice with R.

Feldman, R., y Sanger, J. (2006). The Text Mining Handbook.

Iacus, S. M. (2015). Automated Data Collection with R - A Practical Guide to Web Scraping and Text Mining (Vol. 68) (n.o Book Review 3).

Naldi, M. (2019). A review of sentiment computation methods with R packages., 1–11. Descargado de <http://arxiv.org/abs/1901.08319>

Singh, S., y Choudhary, S. S. (2017). Social Media Data Analysis: Twitter Sentimental Analysis Using R Language. International Journal of Advances in Electronics and Computer Science, 4 (11), 13–17. Descargado de [http:// iraj.in](http://iraj.in)

Zhao, Y., y Cen, Y. (2013). Data Mining Applications with R.