

Monitoreo del Sentimiento de la Sociedad Boliviana en Redes Sociales: Aplicación Twitter

Freddy Richard Rojas Illatarco
Postgrado en Informática
Universidad Mayor de San Andrés
La Paz – Bolivia
frichardrojasi@gmail.com

Resumen—El Análisis de Sentimiento, también llamado Minería de Opinión, permite determinar si una opinión expresada en un texto es positiva o negativa. En este contexto, el Sentimiento de la Sociedad es equivalente a la Opinión Pública. El presente artículo contiene los resultados de la aplicación del modelo propuesto por el autor como alternativa complementaria a los sondeos de opinión empleando la información de Twitter. La aplicación del modelo inicia con el uso de las API de Twitter y algoritmos de Teoría de Grafos para encontrar a la población total de usuarios de Twitter en el país por departamento. El trabajo continúa con la recolección de Tweets y el uso de algoritmos de Aprendizaje Automático sobre los Tweets recolectados para determinar el Sentimiento de la Sociedad Boliviana.

Palabras clave—ciencia de datos, análisis de sentimiento, minería de opinión, aprendizaje automático, sondeos de opinión, twitter.

I. INTRODUCCIÓN

Los sondeos de opinión son mediciones estadísticas que, basadas en encuestas, se realizan para conocer la opinión pública general sobre ciertos temas. Los resultados de los sondeos permiten conocer a distintos niveles el sentimiento, a favor o en contra, de la sociedad sobre un tema específico.

Este tipo de estudios son llevados a cabo con métodos probados para asegurar su imparcialidad con márgenes de error aceptables. Sin embargo, sus resultados sólo reflejan la percepción de la sociedad en un determinado periodo de tiempo. Se tienen que repetir los estudios para determinar la evolución de la percepción. Además, estos estudios no miden el grado de honestidad de los encuestados, quienes pueden decir o no la verdad al encuestador.

Una alternativa complementaria a los sondeos de opinión, basados en encuestas, es la monitorización de las redes sociales a través del Análisis de Sentimiento de los textos publicados por las personas en Redes Sociales sobre diferentes temas [6].

Los sondeos de opinión y el Análisis de Sentimiento en las redes sociales se diferencian sustancialmente en que, en el primero el encuestador busca al encuestado y pide su opinión, y en el segundo el “encuestador” determina la opinión del “encuestado” a partir de las expresiones que éste vierte en Redes Sociales.

En el contexto nacional, algunos de los temas polémicos publicados en medios de prensa y redes sociales son: legalidad del aborto, pena de muerte, castración de violadores, alimentos transgénicos e ideología de género. Considerando que en

octubre del presente año (2019) se llevarán a cabo las elecciones generales para elegir al presidente y al vicepresidente del estado, otro de los temas polémicos del año será la reelección o no de los actuales mandatarios.

El autor del presente artículo propone un modelo, basado en estrategias de Ciencia de Datos, para determinar la Opinión Pública Boliviana a partir de la información de la red social Twitter. Los resultados presentados, son de la aplicación del modelo en una versión demostrativa para determinar la opinión pública respecto de la potencial reelección de los actuales mandatarios.

A continuación, se describe el estado de las tecnologías y fundamentos teóricos que fueron aplicados en la versión demostrativa.

A. Determinación de una Muestra Estadística

En el marco de los sondeos de opinión, resulta imposible estudiar una población en su totalidad, razón por la cual se hace necesario tomar muestras del total de la población. Una muestra es una porción representativa de individuos que permitirá generalizar los resultados de una investigación al total de la población [1] [2].

La fórmula para determinar el tamaño de una muestra es:

$$n = (Z^2 p (1-p)) / e^2 \quad (1)$$

Donde:

Z es el valor asociado al grado de confianza.
 p es la probabilidad de éxito en la muestra.
 e es el error muestral permitido.

El grado de confianza del 95% es el más usado en trabajos de investigación. El valor de Z asociado a un grado de confianza del 95% es 1.96. Por otra parte, si se desconoce la probabilidad de éxito p , entonces se puede emplear el valor convencional de 50% que asegura el tamaño de la muestra más grande posible.

A partir de (1) se obtiene la fórmula para determinar el error muestral dado el tamaño de una muestra:

$$e = Z \sqrt{(p (1-p) / n)} \quad (2)$$

Esta fórmula fue empleada para determinar el error muestral de la muestra obtenida por la versión demostrativa.



Para referenciar este artículo (IEEE):

[N] F. Rojas, «Monitoreo del Sentimiento de la Sociedad Boliviana en Redes Sociales: Aplicación Twitter», *Revista PGI. Investigación, Ciencia y Tecnología en Informática*, n° 8, pp. 71-76, 2020.

B. Algoritmos de Detección de Comunidades

Un grafo puede ser descrito como un conjunto de nodos y las relaciones que los conectan. Los algoritmos de detección de comunidades evalúan cómo se agrupan o dividen grupo de nodos en un grafo, así como su tendencia a fortalecerse o separarse [3] [4].

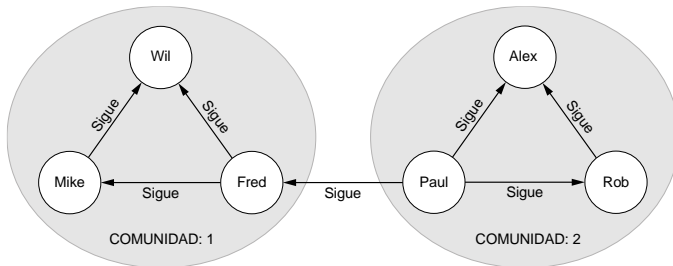


Fig. 1. Ejemplo de la aplicación de un algoritmo de detección de comunidades
Fuente: Adaptación de la documentación de Neo4j

Neo4j [11] es un gestor de base de datos de grafos que contiene un motor de persistencia embebido e implementado en Java. Neo4j cuenta con algoritmos para detección de comunidades implementados (Fig. 1), entre los que está el algoritmo de Louvain y el algoritmo de Propagación de Etiquetas.

Estos algoritmos fueron empleados en la versión demostrativa para determinar la localización, a nivel de departamento, de los usuarios de Twitter que no hubieran declarado explícitamente su localización.

C. Clasificadores Bayes Ingenuo

Los clasificadores Bayes Ingenuo son algoritmos de clasificación de Aprendizaje Automático que tienen base en el Teorema de Bayes el cual expresa la probabilidad condicional de un evento aleatorio A dado otro evento aleatorio B [5] [7].

De acuerdo a este teorema, dadas dos características, la probabilidad posterior de que se dé una característica A dado que se dio una característica B es:

$$P(A|B) = (P(B|A) P(A)) / P(B) \quad (3)$$

Donde:
 P(B|A) es la probabilidad de que se dé B dado A.
 P(A) es la probabilidad de A.
 P(B) es la probabilidad de B.

Este clasificador permite predecir clases de una manera fácil y rápida cuando se trata de problemas de clasificación binarios y multiclase.

TextBlob [12] es una librería Python destinada a procesar datos de textos basado en la plataforma NLTK [13]. TextBlob proporciona una API simple para tareas comunes de Procesamiento del Lenguaje Natural como el etiquetado gramatical, extracción de frases nominales, análisis de sentimientos, clasificación y traducción.

Los clasificadores Bayes Ingenuo están implementados en NLTK y son accesibles desde TextBlob.

Estos clasificadores fueron empleados en la versión demostrativa para determinar el sentimiento, a favor o en contra, de las opiniones expresadas en los Tweets.

D. Análisis de Negocios

El Análisis de Negocios es el uso de datos, tecnología de la información, análisis estadístico, métodos cuantitativos, modelos matemáticos y modelos basados en computadora, orientado a la toma de decisiones. Comienza con la recopilación, organización y manipulación de datos y está soportado por tres componentes principales: Análisis Descriptivo, Análisis Predictivo y Análisis Prescriptivo [8].

Las funciones matemáticas utilizadas en los modelos analíticos predictivos incluyen: función lineal, logarítmica, polinómica, de potencia y exponencial.

La librería Python de Aprendizaje Automático Scikit-Learn [14] contiene un conjunto amplio de herramientas simples y eficientes para Minería de Datos, Análisis de Datos y Análisis de Negocios. Complementariamente, la librería gráfica Plotly [15] cuenta con una amplia variedad de herramientas para la construcción de Cuadros de Mando Integral.

Estas librerías fueron empleadas en la versión demostrativa para determinar tendencias y pronósticos a partir de los resultados del Análisis de Sentimiento aplicado sobre los Tweets.

II. MÉTODOS Y MATERIALES

El modelo propuesto por el autor plantea la identificación de la mayor cantidad posible de los usuarios de Twitter en el país y el uso de este conjunto como muestra de la población boliviana. Las opiniones expresadas por este conjunto permitirán determinar el sentimiento de la muestra y, a través de él, la opinión pública de la población.

A continuación, se describen brevemente las etapas principales de la aplicación del modelo propuesto.

A. Cuentas Semilla

El punto de partida del modelo es la identificación de Cuentas Semilla. Una cuenta semilla es una cuenta de Twitter con una gran cantidad de seguidores.

Para la identificación de las cuentas semilla se utilizaron los reportes de estadísticas de uso de Twitter hechos por *SocialBakers* al 10 de junio de 2019 [10]. Este portal ofrece reportes de cuentas con más seguidores de marcas comerciales, celebridades, comunidad, entretenimiento, medios de comunicación, lugares, sociedad y deportes.

Se encontraron 63 cuentas de Twitter con más seguidores en Bolivia. De cada cuenta semilla, sólo se tienen los nombres de pantalla elegidos por los usuarios, con lo cual el paso final de la identificación de Cuentas Semilla es obtener los metadatos de estas cuentas.

B. Recolección de Seguidores y Amigos

Esta etapa consiste en la recolección de los seguidores y de los amigos de las cuentas dadas en una lista. Las cuentas nuevas encontradas conforman una nueva lista. Cada proceso de recolección es denominado ciclo. Se pueden ejecutar N ciclos.

En la Fig. 1, las relaciones respecto del usuario Fred son: el usuario Paul es seguidor y los usuarios Mike y Wil son amigos.

Las API estándar de Twitter permiten una cantidad máxima de solicitudes de información por ventana. Cada ventana es equivalente a 15 minutos. Estas cantidades están resumidas en la siguiente tabla.

TABLA I. TASAS DE TRANSFERENCIA DE LAS API DE TWITTER

Método de API de Twitter	Método de Tweepy	Tasa/Ventana
GET users/show	API.get_user	900
GET followers/list	API.followers	15
GET friends/list	API.friends	15
GET statuses/user_timeline	API.user_timeline	900

El primer método es empleado para obtener los metadatos de las cuentas semilla. El segundo y tercer método son empleados para obtener la lista de metadatos de seguidores y amigos de una cuenta específica respectivamente. El último método es empleado para obtener los Tweets de una determinada cuenta.

Debido a las bajas tasas de transferencia del segundo y tercer método, el proceso de recolección de seguidores y amigos fue el más largo de la aplicación del modelo.

La recolección se ejecutó en un periodo de 60 días, del 24 de junio al 22 de agosto de 2019, y obtuvo un total de 33,598,048 cuentas, sin embargo, sólo 4,031,077 son cuentas únicas.

C. Determinación de Localizaciones

En el marco del modelo, una localización es equivalente a unos de los departamentos del territorio boliviano: Beni, Chuquisaca, Cochabamba, La Paz, Oruro, Pando, Potosí, Santa Cruz o Tarija.

Entre los metadatos de las cuentas recolectadas está la localización declarada por los usuarios. A partir de este dato se pudo determinar que 2.08% de los usuarios indican la localidad donde se encuentra en el territorio boliviano, el 3.35% indica que se encuentra en Bolivia únicamente sin señalar una localidad en particular y el 52.08% no indican ninguna localidad.

TABLA II. TIPOS DE LOCALIDAD DE CUENTAS RECOLECTADAS

Tipo	Cuentas	(%)
Con Localización	83,918	2.08%
Localizados en Bolivia	134,909	3.35%
Sin Localización	2,099,487	52.08%
Localizados en el Exterior	1,712,763	42.49%
Total	4,031,077	100.00%

La determinación de localizaciones tiene por objetivo aplicar los algoritmos de detección de comunidades en un grafo compuesto por las cuentas recolectadas excluyendo las localizadas en el exterior del país. Las localizaciones conocidas fueron empleadas como puntos de referencia para determinar a qué departamento pertenece una comunidad detectada.

El grafo requerido fue construido en Neo4j con 2,020,593 nodos y 14,111,514 relaciones. Los nodos son las cuentas de los usuarios de Twitter y las relaciones son las conexiones entre las cuentas.

Se ejecutaron los siguientes algoritmos: Louvain sin comunidades predefinidas, Louvain con comunidades predefinidas y Propagación de Etiquetas. El algoritmo de Louvain sin comunidades predefinidas detecta la menor cantidad de comunidades: 85 comunidades en total. A partir de este resultado, se obtuvo la siguiente distribución de cuentas por departamento.

TABLA III. DISTRIBUCIÓN DE CUENTAS A PARTIR DE ALGORITMOS DE DETECCIÓN DE COMUNIDADES

Departamento	Con Localización	Con (%)	Con/Sin Localización y localizados en Bolivia (%)	
			Con/Sin Localización y localizados en Bolivia	(%)
Beni	1,498	(1.79%)	1,592	(0.08%)
Chuquisaca	4,479	(5.34%)	20,511	(1.02%)
Cochabamba	12,177	(14.51%)	89,179	(4.41%)
La Paz	29,168	(34.76%)	1,421,204	(70.34%)
Oruro	2,687	(3.20%)	2,687	(0.13%)
Pando	531	(0.63%)	531	(0.03%)
Potosí	1,587	(1.89%)	1,587	(0.08%)
Santa Cruz	27,729	(33.04%)	469,838	(23.25%)
Tarija	4,062	(4.84%)	13,492	(0.67%)
Total	83,918	(100.00%)	2,020,621	(100.00%)

En síntesis, el proceso de determinación de localidades estableció que 2,020,621 cuentas corresponden a usuarios localizados en el país.

D. Recolección de Tweets

El proceso de recolección de Tweets se ejecutó en un periodo de 12 días, del 6 al 17 de septiembre, periodo en el cual se recolectaron 105,125,489 Tweets de las cuentas de usuario localizados en el país.

Considerando las limitaciones de tiempo y las tasas de transferencia de los métodos de las API de Twitter, se determinó obtener hasta los 200 primeros Tweets y Retweets de cada cuenta de usuario.

Se descartaron direcciones URL, signos de puntuación y de control de los contenidos de los Tweets.

E. Palabras Clave

Las palabras clave relativas al tema de análisis fueron extraídas de reportes de tendencias en Twitter publicadas en el portal TrendsMap.com [16] a nivel de Hashtags al 10 de junio de 2019.

TABLA IV. PALABRAS CLAVE DEL TEMA DE ANÁLISIS

Palabras Clave		
evoespueblo	21f	somosmas
evo	boliviadijono	fuerzaevo
ema	noesno	boliviadigna
linera	boliviadiceno	boliviadicesi
agl	evonuncamas	procesodecambio
	elecciones2019	2020-2025

La versión demostrativa prevé el análisis de Tweets aplicado exclusivamente sobre palabras compuestas por letras y números, descartando signos de puntuación y de control, razón por la cual las palabras clave no llevan el signo numeral que precede a los Hashtags.

F. Determinación del Sentimiento de los Tweets

Este proceso se hizo empleando un Clasificador Bayes Ingenuo implementado en la librería Python TextBlob. El clasificador requiere una lista de entrenamiento y una lista de evaluación para determinar el grado de precisión del clasificador.

Para la construcción de las listas de Tweets de entrenamiento y evaluación, se obtuvo una muestra aleatoria de 10,000 Tweets que contienen cualquiera de las palabras clave. Se eligieron los primeros 500 Tweets con opiniones positivas y los primeros 500 Tweets con opiniones negativas. La lista de entrenamiento fue compuesta con 400 Tweets con opiniones positivas y 400 Tweets con opiniones negativas. La lista de evaluación fue compuesta con los restantes Tweets con opiniones positivas y opiniones negativas.

Luego de entrenar al clasificador con la lista de entrenamiento, se aplicó la clasificación sobre la lista de evaluación. El 95% de los Tweets de evaluación recibieron la misma clasificación, este valor representa el grado de precisión del clasificador.

Para determinar la opinión de una cuenta de usuario, en un determinado periodo de tiempo, se calculó el promedio de las opiniones emitidas a través de la cuenta, donde una opinión positiva fue representada por el valor 1 y una opinión negativa fue representada por el valor -1.

G. Cuadro de Mando Integral

El Cuadro de Mando Integral presenta los resultados de los siguientes indicadores clave:

1. Tasa de cuentas de usuario que emiten opinión respecto del tema de análisis.
2. Tasa de opiniones a favor, en contra y neutras de cuentas de usuario que emiten opinión respecto del tema de análisis total.
3. Tasas de opiniones a favor, en contra y neutras de cuentas de usuario que emiten opinión respecto del tema de análisis total por departamento.
4. Tasas de opiniones a favor y en contra de cuentas de usuario que emiten opinión respecto del tema de análisis de forma diaria comparada con resultados de otras fuentes.

Para el cuarto indicador clave, se recopilaron los resultados de los 10 sondeos de opinión publicados por la prensa nacional entre el 25 de abril y el 11 de septiembre de 2019, donde:

- a) La intención de voto a favor de los actuales mandatarios fue considerada como una opinión a favor de su reelección.
- b) La intención de voto a favor de otros candidatos fue considerada como una opinión en contra.
- c) La intención de voto por ningún candidato, blanco o nulo, y “no sabe/no responde” fue considerada como una opinión neutra.

III. RESULTADOS

A. Indicadores del Cuadro de Mando Integral

1. La tasa de cuentas de usuario que emiten opinión respecto del tema de análisis es del 5.04%. En otras palabras, de 2,020,621 cuentas, 101,933 cuentas emitieron opinión respecto del tema de análisis hasta el momento en el que se hizo la recolección de Tweets (Fig. 2).

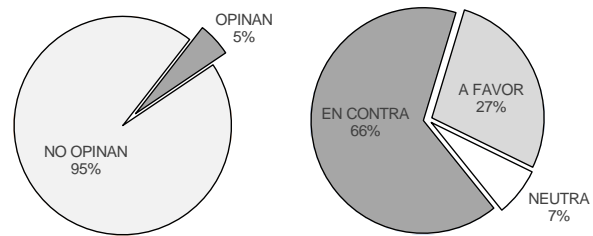


Fig. 2. Tasas de cuentas de cuentas de usuario que emiten opinión y el Tipo de opinión que emiten.

2. De la tasa de opiniones a favor, en contra y neutras de cuentas de usuario que emiten opinión respecto del tema de análisis, las opiniones en contra constituyen el 65.52%. En otras palabras, de 101,933 cuentas que emitieron opinión, 66,791 cuentas emitieron opiniones en contra (Fig. 2).
3. De las tasas de opiniones a favor, en contra y neutras de cuentas de usuario que emiten opinión respecto del tema de análisis total por departamento, las opiniones en contra son mayoría en 8 de los 9 departamentos del país: 66.82% en el Beni, 76.80% en Cochabamba, 63.64% en La Paz, 67.81% en Oruro, 74.03% en Pando, 63.95% en Potosí, 83.48% en Santa Cruz y 80.26% en Tarija. En contraste, el 56.19% de las opiniones en Chuquisaca son a favor de la reelección de los actuales mandatarios.
4. De las tasas de opiniones a favor y en contra de cuentas de usuario que emiten opinión respecto del tema de análisis de forma diaria comparada con resultados de sondeos de opinión, la tendencia de las opiniones a favor es descendente con una pendiente igual a -0.0005 (Fig. 3) y la tendencia de las opiniones en contra es ascendente con una pendiente igual a 0.0004 (Fig. 4).

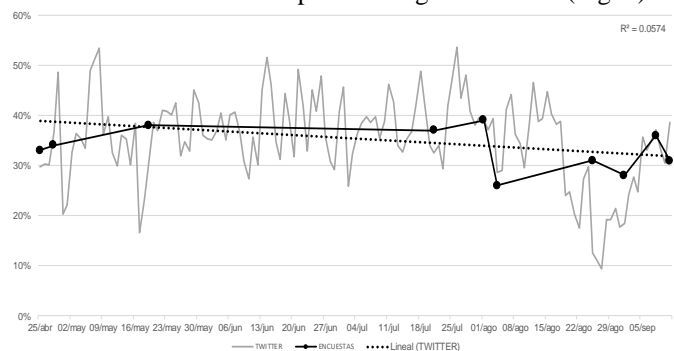


Fig. 3. Tendencia de las opiniones a favor de la reelección de los actuales mandatarios respecto de sondeos de opinión.

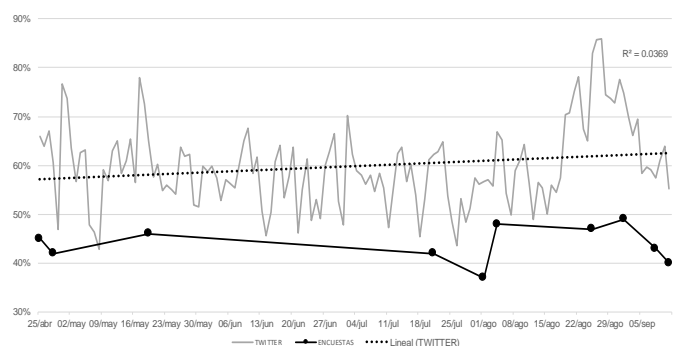


Fig. 4. Tendencia de las opiniones en contra de la reelección de los actuales mandatarios respecto de sondeos de opinión.

De acuerdo a las tendencias de las opiniones a favor y en contra, al 20 de octubre de 2019:

- 29.87% estará a favor de la reelección de los actuales mandatarios.
- 64.17% estará en contra de la reelección de los actuales mandatarios.

Producto del análisis de la desviación estándar de las opiniones a favor y en contra, se determinó que el 49.12% de las cuentas no son constantes en cuanto a sus opiniones. Es decir, cerca de la mitad de los usuarios emiten opiniones positivas y negativas (Fig. 5).

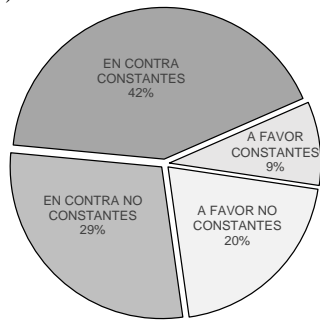


Fig. 5. Clasificación de cuentas que emiten opinión según constancia del tipo de opiniones que emiten.

B. Error muestral

Siendo que la versión demostrativa obtuvo una muestra de 101,933 cuentas únicas de usuario de Twitter que emiten opinión respecto del tema de análisis, con un grado de confianza del 95% y probabilidad de éxito del 50%, el error muestral calculado a partir de la ecuación (2) es 0.31%.

C. Resultados Adicionales

De 2,020,621 cuentas localizadas en el país, el pico en cuanto al año de creación de dichas cuentas, es el 2016 con el 14.40%. Ese año muchas autoridades de estado crearon sus cuentas en Twitter, entre ellas el primer mandatario (15 de abril de 2016).

Se confirma el carácter público de la información disponible en Twitter, el 97.53% de las cuentas recolectadas son públicas, con lo cual las opiniones de estas cuentas son públicas. Sólo el 2.41% de las cuentas recolectadas son privadas.

De 105,125,489 Tweets recolectados, el 35.64% fueron creados el año 2019. Esto sugiere que, si bien existe un importante número de usuarios, la mayoría no publica Tweets.

Por otro lado, la aplicación más utilizada para publicar Tweets es "Twitter para Android" y la mayoría de ellos tienen datos de ubicación geográfica (Fig. 6).

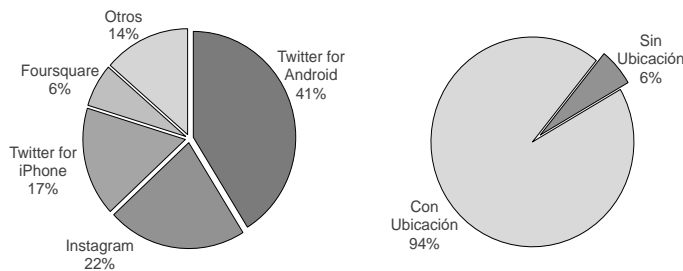


Fig. 6. Tweets por tipo de dispositivo usado y Disponibilidad de datos de ubicación geográfica en Tweets.

IV. DISCUSIÓN

Una encuesta hecha el año 2016 permitió a la AGETIC establecer que aproximadamente 1.1 millones de internautas usan Twitter [9]. De acuerdo al informe del Estado de las TIC de la AGETIC, los internautas son personas de 14 o más años de edad, de ambos sexos y pertenecientes a todos los niveles socioeconómicos.

El presente trabajo encontró más de 2 millones de cuentas de usuarios de Twitter, lo cual concuerda con la estimación hecha por la AGETIC tomando en cuenta el crecimiento de la red social en el país. Si bien los resultados de la aplicación de algoritmos de detección de comunidades son discutibles y merecen mayor análisis, este proceso permitió identificar a los usuarios localizados en el país empleando Teoría de Grafos.

Los resultados publicados el 11 de septiembre de 2019 del sondeo de opinión de la iniciativa "Tu Voto Cuenta" de la UMSA, la Fundación Jubileo y varios medios de comunicación, establecen que el 31% de la población votará por los actuales mandatarios. La tendencia de las opiniones a favor de la versión demostrativa del modelo establece que, a la misma fecha, el 31.85% de las cuentas de usuarios de Twitter están a favor de la reelección de los actuales mandatarios.

El modelo propuesto por el autor de ninguna manera pretende sustituir a los sondeos de opinión. Está claro que estos estudios son llevados a cabo con métodos ampliamente probados que aseguran imparcialidad y márgenes de error aceptables. El modelo propuesto es planteado como una alternativa para obtener información complementaria a los sondeos de opinión.

Una implementación del modelo en producción podría soportar N temas de análisis los cuales podrían ser evaluados de forma paralela, con resultados disponibles en periodos cortos de tiempo.

Además, el análisis de la desviación estándar de las opiniones permitiría determinar el grado de constancia o seguridad de los usuarios de Twitter. Permitiría también identificar los aspectos que provocan opiniones en contra y a favor respecto de un determinado tema. Por ejemplo, si el tema de análisis fuera la Legalización del Aborto, una persona podría estar en contra de ella, pero a favor de ella en determinados escenarios.

Finalmente, los Tweets recolectados por la versión demostrativa permitieron determinar que el 94% de ellos contienen datos de ubicación geográfica. Esto abre la puerta a una investigación complementaria de análisis de datos georeferenciados.

REFERENCIAS

- [1] Blanco, C. Encuesta y Estadística, Métodos de Investigación Cuantitativa en Ciencias Sociales y Comunicación. Editorial Brujas. Argentina. 2011.
- [2] Marradi, A., Archenti, N., Piovani, J.I. Metodología de las Ciencias Sociales. Editorial Emecé. Argentina. 2007.
- [3] Robinson, I., Webber, J. y Eifrem, E. Graph Databases, New Opportunities for Connected Data. (Bases de Datos de Grafos, Nuevas Oportunidades para Datos Conectados). Editorial O'Reilly. EE.UU. 2015.
- [4] Missaoui, R. Y Idrissa, S. Social Network Analysis – Community Detection and Evolution. (Análisis de Redes Sociales – Detección y Evolución de Comunidades). Editorial Springer. Canada. 2014.
- [5] Bird, S., Klein, E. y Loper, E. Natural Language Processing with Python. (Procesamiento de Lenguaje Natural con Python). Editorial O'Reilly. EE.UU. 2009.

- [6] Pozzi, F., Fersini, E., Messina, E. y Liu, B. *Sentiment Analysis in Social Networks. (Análisis de Sentimiento en Redes Sociales)*. Editorial Elsevier. EE.UU. 2017.
- [7] Sarkar, D. *Text Analytics with Python: A Practical Real-World Approach to Gaining Actionable Insight from Your Data. (Análisis de Texto con Python: Un Enfoque Práctico del Mundo Real para Obtener Información Procesable de sus Datos)*. Editorial Apress. India. 2016.
- [8] Evans, J. *Business Analytics. (Análisis de Negocios)*. Editorial Pearson. EE.UU. 2017.
- [9] AGETIC, Agencia de Gobierno Electrónico y Tecnologías de Información y Comunicación. *Estado de las Tecnologías de Información y Comunicación en el Estado Plurinacional de Bolivia*. 2018. Obtenido de: <https://agetec.gob.bo/pdf/estadotic/AGETIC-Estado-TIC.pdf> Visitado en: 15/04/2019.
- [10] Socialbakers.com. *Twitter statistics - Society in Bolivia. (Estadísticas de Twitter – Sociedad en Bolivia)*. 2019. Obtenido de: <https://www.socialbakers.com/statistics/twitter/profiles/bolivia/society/> Visitado en: 15/04/2019.
- [11] Neo4j.com. *Neo4j Graph Platform – The Leader in Graph Databases. (Plataforma de Grafos Neo4j – El Líder en Bases de Datos de Grafos)*. 2019. Obtenido de: <https://neo4j.com> Visitado en: 15/04/2019.
- [12] Readthedocs.io. *TextBlob Simplified Text Processing Documentation. (Documentación de Procesamiento de Texto Simplificado con TextBlob)*. 2019. Obtenido de: <https://textblob.readthedocs.io/en/dev/index.html> Visitado en: 15/04/2019.
- [13] Nltk.org. *Natural Language Toolkit Documentation. (Documentación del Kit de Herramientas de Lenguaje Natural)*. 2019. Obtenido de: <https://www.nltk.org> Visitado en: 15/04/2019.
- [14] Scikit-Learn.org. *Machine Learning in Python – Scikit-Learn Documentation. (Aprendizaje Automático en Python – Documentación de Scikit-Learn)*. 2019. Obtenido de: <https://scikit-learn.org/stable/index.html> Visitado en: 15/04/2019.
- [15] Plot.ly. *Plotly Python Graphing Library. (Biblioteca de Gráficos Python Plotly)*. 2019. Obtenido de: <https://plot.ly/python> Visitado en: 15/04/2019.
- [16] Trendsmap.com. *Twitter Trends in Bolivia. (Tendencias de Twitter en Bolivia)*. 2019. Obtenido de: <https://www.trendsmap.com/local/bolivia> Visitado en: 10/06/2019.

Breve CV del autor

Freddy Richard Rojas Illatarco es Licenciado en Ingeniería de Sistemas por la Universidad Tecnológica Boliviana (2000). Actualmente realiza la maestría en Gerencia Estratégica de Tecnologías y Sistemas de Información en el Postgrado en Informática de la Universidad Mayor de San Andrés.

Cuenta con más de 20 años de experiencia profesional, especialista en estadísticas, análisis de tráfico y desarrollo de sistemas complejos a bajo nivel; y en el desarrollo de soluciones de alta disponibilidad y alto rendimiento, procesamiento masivo de datos, administración de bases de datos, administración de redes y sistemas Unix/Linux.

Fue responsable del área de estadísticas de la empresa ENTEL SA. Fue miembro del equipo de desarrollo de las plataformas de señalización: OCS, SMSC, USSD, EIR y GWSMS de la empresa ENTEL SA. Actualmente forma parte de la empresa NOLOGIN SRL como Project Manager. Email: frichardrojasi@gmail.com.