

Influencia de redes sociales en el análisis de sentimiento aplicado a la situación política en Ecuador

(Influence of social networks on the analysis of sentiment applied to the political situation in Ecuador)

Estevan Gómez-Torres¹, Roger Jaimes², Orlando Hidalgo², Sergio Luján-Mora³

Resumen:

Conocer la opinión de un sector de la población puede ser tan importante para lanzar un producto, como para saber la opinión de los electores. En Ecuador, la red social Twitter se ha convertido en uno de los principales medios de interacción directa entre figuras políticas y la población. Por lo cual un estudio que refleje los sentimientos en español por modismos de cada región, nos da una gran oportunidad de analizar la relación entre el nivel de aceptación en Twitter de un candidato y los resultados electorales. El aporte de este artículo es el análisis de sentimientos (AS) mediante una herramienta para PNL adaptada a la variación del español en el Ecuador, aprovechando que la mayoría de la literatura se ha centrado en el idioma inglés, mientras que adaptaciones para lenguajes, como el español, son mínimas y siguen en proceso por la complejidad inherente al lenguaje.

Palabras clave: análisis de sentimiento; Twitter; minería de texto; Stanford NLP, Rafael Correa.

Abstract:

Knowing the opinion of a sector of the population can be as important to launch a product in marketing, as to know the opinion of voters, in politics. In Ecuador, the social network Twitter has become one of the main means of direct interaction between political figures and the population. So a study that reflects feelings in Spanish by idioms of each region, gives us a great opportunity to study the relationship between the level of acceptance on Twitter of a candidate and the election results. The contribution of this article is the analysis of feelings (SA) using a tool for NLP adapted to the variation of Spanish used in Ecuador, taking advantage of the fact that most of the literature has focused on the English language, while adaptations for languages, like Spanish, they are minimal and are still in process due to the complexity inherent in the language.

Keywords: sentiment analysis; Twitter; text mining; Stanford NLP, Rafael Correa.

1. Introducción

La cantidad de datos producidos en la actualidad a nivel global es bastante alta. Empresas, gobiernos, universidades y, en general, todas las organizaciones los producen a gran escala, relacionados con sus actividades. Dichos datos se recopilan en grandes repositorios, principalmente en bases relacionales que permiten el almacenamiento estructurado de la información. Agregado a estos registros, se genera más información diariamente de la fuente más importante de todas: Internet. Los produce por millones debido al uso masivo de redes sociales, servicios de mensajería, blogs, wikis y comercio electrónico, entre otros.

¹ Universidad Tecnológica Equinoccial, Quito – Ecuador (estevan.gomez@ute.edu.ec).

² Universidad de las Fuerzas Armadas, Quito – Ecuador (rogerjaimes1@espe.edu.ec, ojhidalgo@gmail.com).

³ Universidad de Alicante, Alicante– España (sergio.lujan@ua.es).

La comunidad académica tiene mucho que hacer a partir de la gran cantidad de datos en la Web. En ello participan activamente, tanto el ciudadano común, como las grandes empresas y los actores políticos de una nación. Y lo hacen mediante diferentes herramientas en la nube al dejar sus comentarios, opiniones e incluso reseñas sobre todo tipo de temas, en su lengua materna. Sin lugar a dudas que en la actualidad, las redes sociales sirven para informar, expresar opiniones y sentimientos diversos sobre diferentes aspectos de la sociedad, productos, servicios, aficiones y demás. Por eso, empresas y colectivos en general han mostrado su interés en las opiniones y sentimientos que sus clientes y potenciales usuarios tienen sobre sus actividades. Es así que la red social Twitter se ha convertido en una de las herramientas más comúnmente usadas para determinar las opiniones de los usuarios acerca de una gran variedad de temas. Si tomamos el entorno de la comunicación política, el análisis de contenido y cualquier estudio cuantitativo de los mensajes publicados en Twitter permiten la identificación de patrones de comportamiento entre sus usuarios, así como puntos de inflexión en las corrientes de opinión (Bravo-Márquez, Mendoza, & Poblete, 2014).

Debido al amplio rango de estratos sociales que tienen acceso a la red social Twitter, la naturaleza de los *tuits* contiene en su mayoría emoticones, abreviaturas, terminología específica, jergas, entre otros modismos, lo que dificulta su procesamiento. Las técnicas usuales de procesamiento del lenguaje natural (PLN) deben adaptarse a estas características del lenguaje y proponer nuevas aproximaciones para poder afrontar con éxito el problema. En tal sentido los miles de mensajes existentes en la red de microblogging Twitter se deben evaluar, implementando técnicas para analizar los textos de forma automática y con porcentaje de precisión aceptable, a fin de establecer las condiciones que conlleven a calificar los contenidos extraídos de la interacción entre los usuarios. El análisis de sentimiento en Twitter es formulado como respuesta a esta necesidad del mercado. Además de esto, las computadoras ya están comenzando a adquirir la capacidad de expresar y reconocer el afecto, y pronto tendrán la capacidad de "tener emociones" (Baldasari, 2017). Esto se ha estado construyendo desde la aparición de la computación afectiva, que busca que las computadoras interpreten el estado emocional de los humanos y se adapten a su comportamiento, proporcionándoles una respuesta adecuada a estas emociones. Por lo tanto el análisis de sentimiento funciona de la siguiente manera: "una vez que se registran las diferentes entradas de la información del usuario hay que clasificarlas y, dado que normalmente se suelen producir varias entradas en un mismo instante de tiempo, será necesario determinar cuáles son más importantes a la hora de procesarlas mediante técnicas de fusión multimodal y de gestionar la respuesta adecuada por parte del sistema".

Este estudio realiza una adaptación de la herramienta Stanford NLP (Group S. N., 2018) al lenguaje regional ecuatoriano, a través de la revisión de expresiones regionales, modismos, mensajes con sentidos discordantes, abreviaciones, entre otras características propias del lenguaje español.

Podemos mencionar que en la evolución de técnicas para el estudio de sentimiento en inglés existen varios trabajos de investigaciones satisfactorias, tales como: 1) Léxico para polaridad y enfoques del estilo en bloques de texto para subjetividad (Medhat, Hassan, & Korashy, 2014), 2) Clasificadores formados por componentes diversificados (datos textuales, emoticones y léxicos) (Jungherr, 2015), 3) Clasificación de sentimientos mediante BayesNaïve y Support Vector Machines en aprendizaje automático (machine learning) (Da Silva, Hruschka, & Hruschka, 2014), 4) Predicciones de sentimientos basados en la Teoría de la disposición (Yu & Wang, 2015).

La aplicación de técnicas de procesamiento de lenguaje natural (PLN) a eventos políticos en Twitter permitieron evaluar la influencia de actores políticos en comunidades (Caton, Hall, & Weinhardt, 2015), comportamientos, interacciones con seguidores (Dang-Xuan, Stieglitz, Wladarsch, & Neuberger, 2013) e influencia de los medios tradicionales (Guo & Vargo, 2015), (Vargo, Guo, McCombs, & Shaw, 2014).

Es notable que las herramientas para el PLN, como diccionarios, lexicones y corpus, se están traduciendo a idiomas distintos al inglés (Prata, Soares, Silva, Trevisan, & Letouze, 2016), pero al español, aún están en fase de maduración por la complejidad de sintaxis y semántica. Los trabajos de análisis de sentimiento sobre Twitter fueron implementados después de la creación de la red social Twitter, luego de convertirse en un medio masivo de comunicación e interacción social y su contenido llega a ser utilizado como mapa perceptual de sus usuarios.

Por su parte Stanford Natural Language Processing (Manning et al., 2014), es una de las herramientas más populares por sus características, ya que proporciona un conjunto de utilidades de tecnología de procesamiento de lenguaje natural tales como: reconocimiento de las formas básicas de las palabras, análisis morfológico, análisis sintáctico, identificación de nombres propios, normalización de fechas, tiempos y cantidades numéricas, planificación de frases, generación de frases, identificación de citas textuales, correlación entre menciones y entidades analizadas.

Entre otras, estas características, junto con las ecuaciones de ponderación para el score de sentimiento sobre una frase y un diccionario de palabras en español bastante robusto (Vargo, Guo, McCombs, & Shaw, 2014), nos han permitido realizar el presente análisis de sentimiento.

2. Trabajos relacionados

El Análisis de Sentimiento tiene varios estudios y mejoras realizadas con anterioridad por varios investigadores. El problema se ha tratado usando técnicas de aprendizaje automático, en el trabajo de (Bo, Lee, & Vaithyanathan, 2002) o desde aproximaciones basadas en el conocimiento o no supervisadas. Estudios que analizan y proponen métodos para determinar la polaridad, básicamente, sobre el dominio de opiniones de las frases a analizar. (Turney, 2002), trabaja sobre comparativa entre la codificación automática y manual en el análisis de sentimientos sobre campañas de marketing en redes de microblogging, (Lipka, 2018) con su trabajo de Modelado de tareas de clasificación de texto no estándar, (Bernard, Zhang, Sobel, & Chowdury, 2009), representación abstracta de los mensajes contenidos en un *tuit* en vez de compararlos con palabras estáticas, para el análisis de la metadata contenida en los mismos (O'Connor, Krieger, & Ahn, 2010). Los resultados y guía de estos trabajos nos permitieron escoger la herramienta Stanford NLP como principal recurso para el procesamiento de lenguaje natural, por el conjunto de características relevantes que esta posee. En (Barbosa & Feng, 2010) se recopilan las técnicas y métodos que innovaron el mundo del AS, al exponer la dependencia de los clasificadores a dominios y temas, al notar que se pueden unir varias técnicas para suplir falencias individuales. En (Vinodhini & Chandrasekaran, 2012) se detallan las tendencias de investigaciones en Twitter para el AS. Allí nombran varios problemas como la escasez de datos luego de la limpieza de la base inicial, así como el trabajo que se debe realizar para pulir herramientas que no sean enfocadas al idioma inglés. Es digno notar el esfuerzo llevado a cabo en la competición SemEval-2013, donde se dedicaron tareas específicas para el AS en Twitter para el idioma español (Martínez-Cámara, Martín-Valdivia, Ureña-López, & Montejó-Ráez, 2012). Los resultados de todos estos trabajos han sido muy útiles al cimentar el camino para mejorar el nivel de exactitud en el AS, fruto del cual Stanford NLP se convirtió en nuestra principal fuente de diccionarios y lexicones para el presente estudio de sentimientos.

Al recopilar información sobre análisis político en Twitter, se pueden citar los trabajos de (Barberá & Rivero, 2012), (Bustos & Capilla, 2013) y (Congosto, 2015) en Europa; en Sudamérica existen varias iniciativas de investigación universitaria en pregrado, pero pocas investigaciones formales. Se puede destacar un análisis del balotaje en Argentina (Robins, Frati, Alvarez, Texier, & Loto, 2012), Venezuela (Artigas, Muñoz, Luengo, Chourio, & Fernández, 2012) o Chile (Guevara, Pino, Mendoza, Pacheco, & Olivares, 2013); sin embargo, Ecuador carece de estudios formales.

En inicio, puede decirse que analizar el sentimiento en Twitter supone asignar a cada mensaje publicado un valor relacionado con la carga emocional que transmite. En relación con esta carga emocional se distinguen algunos tipos de variables diferentes según explica Bravo-Márquez et al. (Medhat, Hassan, & Korashy, 2014):

- Polaridad: nos permite medir cuál es el resultado del sentimiento (positivo o negativo). En el caso de Stanford NLP, se introduce un valor neutro para clasificación de oraciones.

- Intensidad: permite asignar un número para relacionarlo con cuán intenso es un sentimiento. Es posible distinguir los valores de intensidad y clasificarla entre positiva y negativa.

- Emoción: clasifica el texto a base del listado de emociones. Estas pueden ser de alegría, tristeza, enfado, agresión o ira por dar unos ejemplos. Analizar grandes volúmenes de datos trae una dificultad añadida, que es evaluar el sentimiento de la polaridad, ya que este puede variar según el codificador; es por eso que se debe tomar muy en cuenta la evaluación de fiabilidad en estas herramientas.

En la mayoría de análisis de sentimientos, el acercamiento más popular es el de polaridad. En un estudio de (Da Silva, Hruschka, & Hruschka, 2014) se presenta una categorización entre los niveles de análisis. El sentimiento puede ser evaluado en distintos niveles, ya sean como entidades individuales, oraciones simples y complejas, o como un documento en su totalidad. El nivel inicial es el análisis de sentimiento para una determinada entidad, la misma que puede ser cualquier cosa, asociación, o personaje. Cuando se evalúa un texto en cambio, este se limita a analizar cómo está construida la oración. Finalmente, al evaluar el sentimiento en un documento completo, es regla tomar el concepto completo de la información expuesta. En el contexto de Twitter, no se distingue entre estos niveles, en consecuencia, se utiliza al *tuit* como un documento independiente. Tomado de (Da Silva, Hruschka, & Hruschka, 2014), las técnicas de análisis de sentimiento más usadas, se dividen en dos grupos: técnicas basadas en el aprendizaje automático (machine learning approach) y técnicas que son basadas en diccionarios de palabras (lexicon-based approach).

Cada *tuit* debe ser clasificado en una escala de polaridad que puede tener varios estados, el modelo más utilizado tiene seis niveles: sin sentimiento, negativo fuerte, negativo, neutral, positivo y positivo fuerte.

En general, el resultado de la evaluación del nivel de exactitud de estos procedimientos debe superar 85 % para considerarse exitosa.

Iniciar el análisis de sentimientos de una base de opiniones, requiere categorizar los mismos (tuits) a causa del lenguaje que expresan los usuarios en Twitter, como terminologías específicas, modismos, errores ortográficos, mal uso de símbolos gramaticales y de puntuación, los cuales impiden analizar las muestras sin limpiarlas. Se añade a la complejidad el hecho que las herramientas están diseñadas específicamente para el inglés, lo cual conlleva mucho esfuerzo en estos trabajos a fin de lograr una categorización adecuada para obtener resultados aceptables en los estudios.

Es de consenso general el uso del Tuitmotif para categorizar *tuits*, explicado en O'Connor et al. (Bernard, Zhang, Sobel, & Chowdury, 2009), además la herramienta Freeling (Padró & Stanilovsky, 2012); el uso de ellas facilitó la agrupación de varios grupos de interés (modismos y expresiones regionales).

Varios estudios difieren en los métodos de correlación estadística entre resultados y muestreo inicial, pero las conclusiones coinciden en que los diccionarios necesitan resultados confiables, tampoco existe una relación efectiva de los resultados del análisis de sentimiento y los electorales.

2. Metodología

Para el desarrollo de este estudio en minería de datos, se tomó como fuente de análisis el contenido de la red de microblogging Twitter. Se utilizó Stanford NLP, que está

disponible para descargar, bajo la Licencia Pública General de GNU. Es una herramienta de aprendizaje automático que tomará elementos de datos y los ubicará en clases determinadas. Se adaptó a Stanford NLP para que sea un clasificador de entropía máximo, también conocido como clasificador softmax, equivalente a modelos de regresión logística multiclase. Modelos como estos tienen base en tecnologías de aprendizaje en la nube. (Manning, y otros, 2014)

Para la resolución de este tipo de casos de minería de datos existen algunas metodologías que permiten organizar el proyecto de mejor manera. A continuación se detalla el proceso metodológico para ejecutarlo:

1) Recopilación de información: inicia con la captación de información, se realiza la lectura de los *tuits* por medio de la ejecución de una aplicación que utiliza el API de Twitter con la herramienta “LINQ to Twitter” en .NET que permite contar con una conexión mediante un servicio con la red social, la búsqueda se ejecuta con la palabra “Rafael Correa”, para su posterior registro en base y captura o *snapshot* como respaldo tal como se presenta en la *Figura 1*.

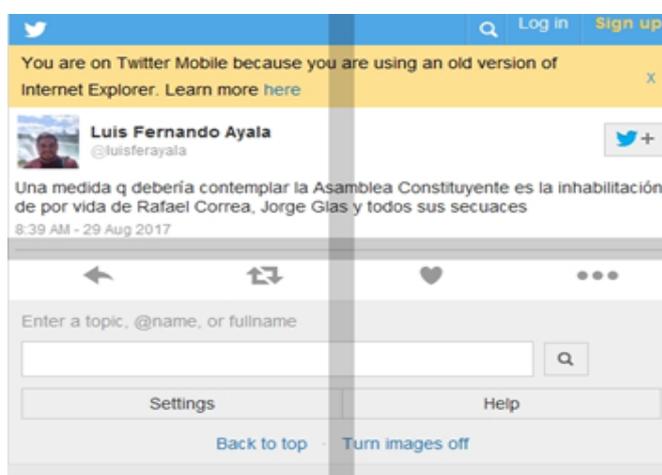


Figura 1. Ejemplo de captura realizada por el sistema posterior a la búsqueda de palabras claves.

Rango de fechas para la recolección de datos: las fechas claves de la investigación son los siguientes rangos de la campaña electoral.

- Precampaña: 28 octubre 2016 - 18 febrero 2017.
- Campaña: 19 febrero 2017 - 2 abril 2017.
- Poscampaña: 3 abril 2017 - 3 mayo 2017.

Información recopilada: por cada *tuit*, se obtuvieron los siguientes campos para el análisis.

- Nombre del usuario.
- Demografía de ubicación.
- Texto publicado.

Depuración: para el análisis de los *tuits* se utilizó el usuario y el texto para una posterior depuración.

Tamaño de muestra: el cálculo del tamaño correcto de la muestra, se basó en la distribución normal, definido de la siguiente forma:

- Tamaño de la población electoral o padrón electoral: 10708969 votantes habilitados.
- Margen de error: 5 %.
- Nivel de Confianza: 95 %.
- Viabilidad Conocida: 50 %.

A partir de los valores anotados arriba, el tamaño n de su muestra y el margen de error E se calculan así:

$$x = Z \left(\frac{c}{100} \right)^2 r(100 - r)$$

$$n = N \frac{x}{((N - 1)E^2 + x)}$$

$$E = \sqrt{\frac{(N - n)x}{n(N - 1)}}$$

Con el resultado de esta ecuación, se pudo determinar que el tamaño de la muestra óptima es de 385 casos.

2) Pasos de la limpieza de información: una vez realizada la búsqueda del personaje a investigar, la evidencia existente se limpia como se detalla a continuación: a) Eliminar el @ o remover el direccionamiento a personajes, b) Eliminar números, c) Eliminar los links, d) Eliminar las tabulaciones, e) Remover el RT del *retuit*, f) Eliminar caracteres especiales, g) Eliminar espacios que estén en blanco al inicio y final de la publicación, h) Identificar mediante ubicación geográfica longitud y latitud de electorado.

La asignación del valor positivo y negativo se evidencia en la *Figura 2*. El proceso de limpieza se lo puede evidenciar en la *Tabla 1*.

3) Análisis de sentimiento: tras clasificar los *tuits*, se evalúan los sentimientos conociendo el valor para clasificar los comentarios negativos, neutrales y positivos. Se necesita un diccionario de palabras positivas y negativas con sus respectivas calificaciones, según lo destaca el StanfordNLP con el contenido que se muestra en la *fig. 3*. Se realiza un diccionario de palabras; y según la palabra, se le asigna un valor positivo o valor negativo.

Nuevo	CODIGO	PALABRA	V. POSITIVO	V. NEGATIVO	EXCLUIR
Editar	1	lay	1	9	<input type="checkbox"/>
Editar	2	icaramba	9	1	<input type="checkbox"/>
Editar	3	Denuncian	10	35	<input type="checkbox"/>
Editar	4	ino	1	9	<input type="checkbox"/>
Editar	5	Salud	6	1	<input type="checkbox"/>
Editar	6	Viva	12	1	<input type="checkbox"/>
Editar	7	desprotegiendo	5	35	<input type="checkbox"/>
Editar	8	empezarán	50	1	<input type="checkbox"/>
Editar	9	renovable	35	1	<input type="checkbox"/>
Editar	10	pendejadas	1	123	<input type="checkbox"/>

Figura 2. Diccionario de palabras con el valor a base de ponderación positiva y negativa

Se realiza el desarrollo de la limpieza de datos tomando en cuenta que siempre en toda información que se recibe, tanto de sensores como de redes sociales, existe la posibilidad de que se introduzcan datos inválidos entre la información recibida, como se muestra en la *Figura 3*.

```

Noticia: 1789167
Resumen:SECTORES POLITICOS SE SUMAN A INICIATIVA DE CONSULTA POPULAR
Positive Score for 39.4254581489484
Negative Score for 0.574541851051628
POSITIVO

Noticia: 1789166
Resumen:UN HOMBRE RECIBIO IMPACTO DE BALA POR NO DEJARSE ROBAR EN SOCIO VIVIENDA DOS
Positive Score for 1.42570675713534
Negative Score for 98.5242132428646
NEGATIVO

Noticia: 1789165
Resumen:FISCALIA PIDE INCREMENTO DE PENA EN CASO PETROEQUADOR
Positive Score for 02.4582849862743
Negative Score for 17.5397130937257
POSITIVO

Noticia: 1789164
Resumen:GLAS OPINA QUE ES VIABLE QUE EL PRESIDENTE HAGA CONSULTA POPULAR
Positive Score for 34.1248913632887
Negative Score for 45.8751086317193
NEUTRAL
    
```

Figura 3. Limpieza, análisis y calificativo del *tuit* registrado

Se le atribuyó al usuario el tipo de comentario que tenga más peso, si tiene los tres tipos se le califica como comentario neutral, como se muestra en la *Tabla 1*.

Tabla 1. Asignación de sentimiento por mensaje

	Tuitero	Texto/Tuit	Polaridad	Resultado	Latitud	Longitud	
Personaje		Txt 1	+	N	La @1	Lo @1	
	"@Usuario 1"	Txt 2	-				
		Txt n	N				
			Txt 1	+	(+)	La @2	Lo @2
	"@Usuario 2"	Txt 2	-				
		Txt n	+				
			Txt 1	+	(-)	La @3	Lo @3
	"@Usuario n"	Txt 2	-				
		Txt n	-				

La técnica de minería de datos, nos entrega resultados a base del proceso realizado con StanfordNLP, cuya funcionalidad compara el texto del *tuit* con el diccionario de palabras positivas y negativas cargado al inicio del análisis, a fin de proporcionar una calificación acorde con la realidad. Si el aplicativo arroja una respuesta mayor a 60 % se conserva el valor del resultado, caso contrario se le calificará como neutro; así se refleja en la *Tabla 1*.

3. Resultados

Efectuado el análisis de sentimiento de los *tuits* publicados en el rango de fecha estudiada, los resultados obtenidos de las opiniones de los usuarios del sitio de microblogging Twitter se muestran en la *Tabla 2*, donde puede verse la polarización de resultados.

- Tamaño de la muestra de la Precampaña Electoral: 17818 *tuits* (28 octubre 2016 - 18 febrero 2017).

Tabla 2. Polarización de resultados de la Precampaña Electoral

	POSITIVO	NEUTRAL	NEGATIVO
TUITS	4264	4727	8827

En la *Figura 4* los resultados de esta investigación muestran que entre los *tuits* analizados correspondientes a la precampaña electoral Ecuador 2017 se enviaron 17818 *tuits* a la red social de Twitter, de ellos, 8827 mensajes corresponden al calificativo "negativo", otorgándole la mayor representatividad a este sentimiento. Se evidenció en dichos comentarios que usuarios descalificaron al expresidente Rafael Correa y su gestión, asimismo otros criticaron irregularidades en servicios públicos, inseguridad, entre otros. La *Figura 4* muestra los resultados de los sentimientos para la precampaña electoral. Mientras que 4727 *tuits* fueron de carácter "neutral" ocupando el segundo lugar, en menor proporción se ubicó el sentimiento "positivo" con 4264 mensajes, donde los usuarios manifestaron: amor, gratitud y apoyo al ex primer mandatario ecuatoriano. Asimismo, cabe destacar que, agrupados los *tuits* de carácter positivo y neutral, superan a los mensajes con sentimiento negativo.

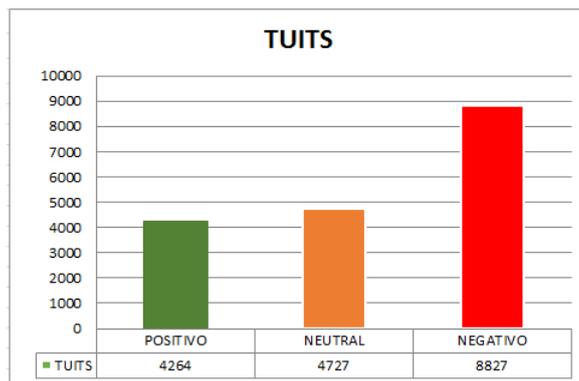


Figura 4. Resultados de sentimientos para la precampaña electoral

- Tamaño de la muestra de la Campaña Electoral: 5766 tuits. La *Tabla 3* muestra la polarización de resultados de la campaña electoral (19 febrero 2017 - 2 abril 2017).

Tabla 3. Polarización de resultados de la Campaña Electoral

	POSITIVO	NEUTRAL	NEGATIVO
TUI TS	1335	1641	2790

En la *Figura 5*, durante la campaña electoral, al evaluar 5766 tuits, se contabilizaron 2790 comentarios de rechazo, 1641 fueron neutrales y 1335 mensajes positivos. Así mismo los tuits positivos y neutrales superaron a los negativos.

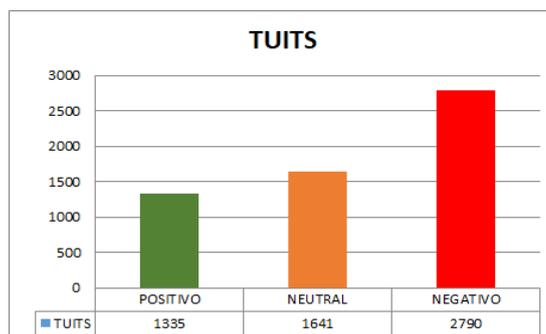


Figura 5. Resultados de sentimientos para la campaña electoral

- Tamaño de la muestra de la Poscampaña Electoral: 4641. La *Tabla 4* muestra la polarización de los resultados de la Postcampaña Electoral (3 abril 2017 - 3 mayo 2017).

Tabla 4. Polarización de resultados de la Poscampaña Electoral.

	POSITIVO	NEUTRAL	NEGATIVO
TUI TS	1096	1130	2415

Culminada la campaña, en la *Figura 6*, se emitieron 4641 *tuits*, reiterando la tendencia de reducción de mensajes de usuarios hacia Rafael Correa, 2415 fueron calificados negativos por contener palabras de repudio.

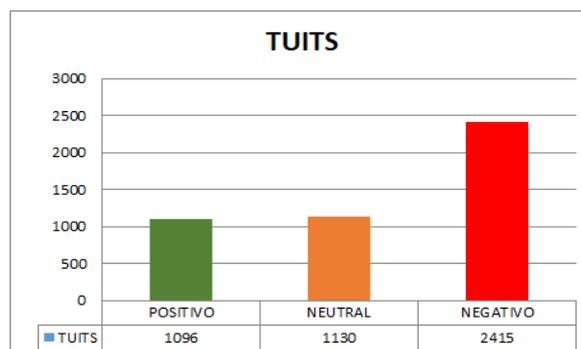


Figura 6. Resultados de la postcampaña electoral

Los mensajes neutrales ocuparon el segundo lugar similar a la precampaña y campaña, en menor proporción 1096 tuits fueron de respaldo calificadas como positivas. Es preciso resaltar que terminado el proceso electoral desapareció la prevalencia de los sentimientos positivos y neutrales sumados sobre los negativos como se destacó en las *figuras 5 y 6*, es decir que en la postcampaña los tuits negativos fueron absolutos.

Muestra consolidada para la precampaña, campaña y postcampaña: 28225. La Tabla 5 indica la polarización de resultados totales de la campaña electoral.

Tabla 5. Polarización de resultados totales de la Campaña Electoral

	POSITIVO	NEUTRAL	NEGATIVO
TUI TS	6695	7498	14032

La *Figura 7* nos enseña los resultados de la campaña electoral con un resultado positivo de 6695 *tuits*, Neutrales se muestran 7498 y Negativos 14032. Evidenciándose que hay un rechazo a la imagen política del expresidente Rafael Correa, lo cual se traduce en un nivel de aceptación del nuevo mandatario Lenin Moreno.

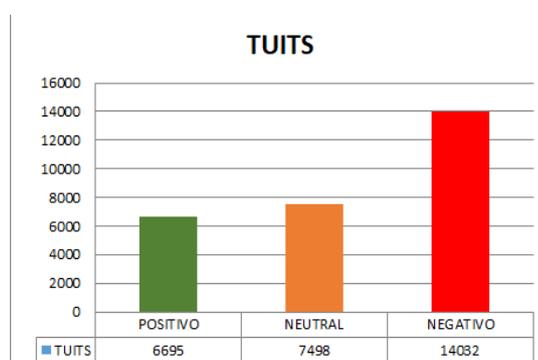


Figura 7. Resultados consolidados de la campaña electoral

A continuación se analizan los resultados generales obtenidos en las elecciones presidenciales de Ecuador 2017.

Resultados Generales Presidenciales Ecuador 2017. La Figura 8 detalla los porcentajes de votos obtenidos por los candidatos presidenciales Guillermo Lasso correspondiente al Movimiento Alianza Creo-Suma y Lenin Moreno Garcés correspondiente al Movimiento Alianza País.

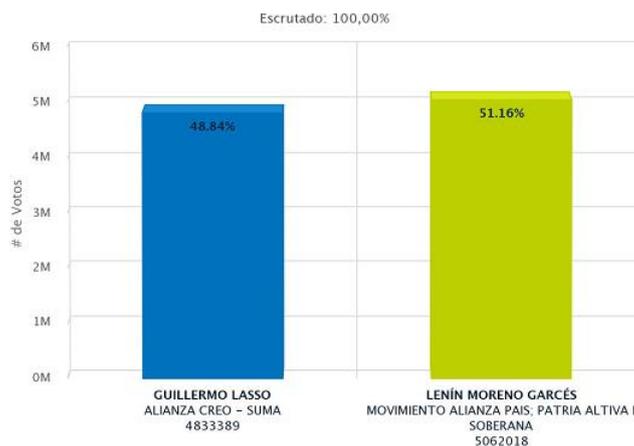


Figura 8. Resultados de las elecciones en Ecuador 2017
(Fuente: CNE Ecuador)

En todo el proceso electoral desde la precampaña, campaña y postcampaña desarrollado entre el 28 octubre de 2016 hasta el 3 mayo de 2017 se emitieron 28225 tuits, en dicho período se mantuvo el liderazgo de los mensajes con sentimientos negativos registrando 14032 tuits, seguidos por 7498 comentarios de tipo neutral con palabras intermedias, es decir carentes de afecto u odio, mayormente informativas, finalmente 6695 mensajes fueron calificados positivos en las fechas estudiadas. En la investigación se demostró que no existe relación entre la alta emisión de mensajes con sentimientos negativos y los resultados electorales, puesto que Rafael Correa recibió en la red social Twitter 14032 tuits en este tiempo, sin embargo, su candidato presidencial Lenin Moreno y su partido Alianza País fueron los ganadores de los comicios presidenciales y legislativos. Moreno obtuvo 51.16 % de votos sobre su más cercano contrincante Guillermo Lasso con 48.84 %.

4. Conclusiones y recomendaciones

Los resultados en las elecciones Ecuador 2017 no coincidieron con los resultados de Twitter, en consecuencia se determinó que la relación entre tendencias en redes de microblogging y los resultados electorales incluye más variables que las consideradas en el presente trabajo.

Además, la diferencia entre la tendencia hacia el expresidente Rafael Correa en la poscampaña y los resultados del actual presidente Lenin Moreno marcan un distanciamiento en la opinión general de los ecuatorianos. Podemos decir que Lenin Moreno se ha consolidado como personaje político con propio peso y es ajeno a influencias externas en su agrupación política, en especial del exmandatario Rafael Correa.

Al momento de efectuar un posterior análisis de un muestro aleatorio de los *tuits* analizados se pudo identificar que los usuarios después de un tiempo eliminan ciertas publicaciones, se bloqueó su cuenta y/o la convirtieron en cuenta privada.

Las redes sociales en la actualidad son una herramienta influyente para captar seguidores, facilitando además una comunicación a gran escala con las diversas masas sociales.

Los algoritmos de aprendizaje prometen soluciones en el área del procesamiento del lenguaje natural, así como la necesidad de generar nuevas técnicas para análisis de texto (sintáctico, análisis léxico, resolución de duplicidad, etc.).

En los medios sociales se genera un amplio y constante intercambio de información, completamente escaso de regulación, donde los usuarios se expresan con libertad y en oportunidades evaden su responsabilidad por lo publicado.

Al contar con el calificativo de cada texto se valoraron los *tuits*, positivos, neutros y negativos. Un usuario puede tener varias publicaciones y en cada una de ellas diversas inclinaciones ya sea política, social o de ánimo para realizar diferentes mensajes.

Para investigaciones futuras se recomienda depurar el diccionario de palabras, dado que la investigación actual dio como resultado que las ponderaciones a usarse, quedan muchos modismos y abreviaturas que no son cubiertas por el material actual.

El presente trabajo marca la pauta para futuras investigaciones en el campo social y político, para seguir la evolución de la influencia de las redes sociales en procesos electorales en Ecuador.

Bibliografía

- Artigas, D., Muñoz, A., Luengo, F., Chourio, X., & Fernández, A. (2012). Caracterizando las elecciones venezolanas a través de Twitter. Caso: #26s. *Anuario electrónico de estudios en Comunicación Social "Disertaciones"*, 5(1), 57-76.
- Babiera, T. (2016). Técnicas para el análisis del sentimiento en Twitter: Aprendizaje Automático Supervisado y SentiStrength. *Revista DÍGITOS*, 33-50.
- Baldasarri, S. (2017). Computación Afectiva: tecnología y emociones para mejorar la experiencia de usuario. 14,15.
- Barberá, P., & Rivero, G. (2012). ¿Un tweet, un voto? Desigualdad en la discusión política en Twitter. *I Congreso Internacional en Comunicación Política y Estrategias de Campaña*.
- Barbosa, L., & Feng, J. (2010). Robust sentiment detection on twitter from biased and noisy data. *En Proceedings of the 23rd International Conference on Computational Linguistics*, 36–44.
- Bernard, J., Zhang, M., Sobel, K., & Chowdury, A. (2009). Twitter power: Tweets as electronic word of mouth. *Journal of the American society for information science and technology*, 60(11), 2169–2188.
- Bo, P., Lee, L., & Vaithyanathan, S. (2002). Thumbs up? Sentiment classification using machine learning techniques. *PROCEEDINGS OF EMNLP*, 79–86.
- Bravo-Marquez, F., Mendoza, M., & Poblete, B. (2014). Meta-level sentiment models for big social data analysis. *Knowledge-Based Systems*, 69(1), 86–99.
- Bustos, J., & Capilla, L. (2013). Twitter y la polarización del debate político: análisis del caso #objetivodeguindos y #aznara3. *Revista Historia y Comunicación Social*, 18, 499-509.
- Caton, S., Hall, M., & Weinhardt, C. (2015). How do politicians use Facebook? An applied Social Observatory. *Big Data & Society*, 2(2).
- Congosto, M. (2015). Elecciones Europeas 2014: Viralidad de los mensajes en TwitterRedes. *Revista Hispana para el Análisis de Redes Sociales*.
- Da Silva, N., Hruschka, E., & Hruschka, E. (2014). Tweet sentiment analysis with classifier ensembles. *Decision Support Systems*, 66, 170–179.
- Dang-Xuan, L., Stieglitz, S., Wladarsch, J., & Neuberger, C. (2013). An Investigation of Influentials and the Role of Sentiment in Political Communication on Twitter During Election Periods. *Information, Communication & Society*, 16(5), 795–825.
- Group, S. N. (2018). *Stanford NLP Group*. Obtenido de Stanford NLP Group: <https://nlp.stanford.edu/software/>
- Group, T. S. (2018). *The Stanford NLP Group*. Obtenido de The Stanford NLP Group: <https://nlp.stanford.edu/>
- Guevara, M., Pino, D., Mendoza, M., Pacheco, C., & Olivares, M. (2013). Chile y el Ecosistema de las Elecciones Políticas en Twitter. *IV Congreso Internacional de Informática del Norte de Chile, Coquimbo-Chile*.

- Guo, L., & Vargo, C. (2015). The Power of Message Networks: A Big-Data Analysis of the Network Agenda Setting Model and Issue Ownership. *Mass Communication and Society*, 18(5), 557–576.
- Jungherr, A. (2015). Analyzing Political Communication with Digital Trace Data: The Role of Twitter Messages in Social Science Research. *Cham: Springer International Publishing Switzerland*.
- Lipka, N. (2018). Modeling Non-Standard Text Classification Tasks. Bauhaus-Universität Weimar, Germany.
- Manning, C., Surdeanu, M., Bauer, J., Finkel, J., Bethard, S., & McClosky, D. (2014). The Stanford CoreNLP Natural Language Processing Toolkit. *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, 55-60.
- Martínez-Cámara, E., Martín-Valdivia, M., Ureña-López, L., & Montejó-Ráez, A. (2012). Sentiment analysis in twitter. *Natural Language Engineering*, 1-28.
- Medhat, W., Hassan, A., & Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey. *Ain Shams Engineering Journal*, 5(4), 1093-1113.
- O'Connor, B., Krieger, M., & Ahn, D. (2010). Tweetmotif: Exploratory search and topic summarization for twitter. *William W. Cohen y Samuel Gosling, editores, Proceedings of the Fourth International Conference on Weblogs and Social Media*.
- Padró, L., & Stanilovsky, E. (2012). Freeling 3.0: Towards wider multilinguality. *Proceedings of the Language Resources and Evaluation Conference (LREC), Istanbul, Turkey*.
- Prata, D., Soares, K., Silva, M., Trevisan, D., & Letouze, P. (2016). Social Data Analysis of Brazilian's Mood from Twitter. *International Journal of Social Science and Humanity*, 6(3), 179–183.
- Robins, D., Frati, F., Alvarez, J., Texier, J., & Loto, L. (2012). Balotaje Argentina 2015 a partir de un análisis de sentimiento de tweets. *Zenodo*.
- Turney, P. (2002). Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews. *ACL*, 417–424.
- Vargo, C., Guo, L., McCombs, M., & Shaw, D. (2014). Network Issue Agendas on Twitter During the 2012 U.S. Presidential Election. *Journal of Communication*, 64, 296–316.
- Vinodhini, G., & Chandrasekaran, R. (2012). Sentiment analysis and opinion mining: A survey. *International Journal*, 2(6).
- Wilson, T., Kozareva, Z., Nakov, P., Rosenthal, S., Stoyanov, V., & Ritter, A. (2013). Sentiment analysis in twitter. *Proceedings of the International Workshop on Semantic Evaluation. SemEval*.
- Yu, Y., & Wang, X. (2015). World Cup 2014 in the Twitter World: A big data analysis of sentiments in U.S. sports fans' tweets. *En Computers in Human Behavior*, 48, 392–400.