

Modelo De Intención Del Voto En El Proceso De Elecciones Presidenciales Ecuatorianas 2017

Jefferson Beltrán¹, Luis Torres²
Michelle Guerra³, Yasmina Atarihuana⁴

**(Facultad de Ingenieria Ciencias Físicas y Matemáticas, Universidad Central del Ecuador*

***(Facultad de Ingenieria Ciencias Físicas y Matematicas, Universidad Central del Ecuador*

****(Facultad de Ingenieria Ciencias Físicas y Matemáticas, Universidad Central del Ecuador*

*****(Facultad de Ingenieria Ciencias Físicas y Matematicas, Universidad Central del Ecuador*

RESUMEN: Con la finalidad de aprovechar los datos generados en redes sociales y transformarlos en información, el presente trabajo tiene el propósito de crear un modelo para medir la popularidad de un candidato en las elecciones presidenciales de Ecuador del año 2017, utilizando Twitter como la red social para el análisis. En el desarrollo de este se realizó la elección de personalidades políticas más sonadas del país en el mes de mayo del 2016, de cada uno de ellos se extraen los tweets donde se los mencionan, consecutivamente someterlos a la limpieza de su texto para poder emplear minería de opinión mediante un método supervisado, para determinar la polaridad de cada tweet, y así generar el modelo de popularidad a través de la clasificación de opiniones generadas por un tweetero. Con los datos obtenidos del modelo se muestran la aceptación que tiene cada precandidato.

Palabras Clave: Redes Sociales; Twitter; Minería de Opinión; Score.Sentiment; Análisis de Sentimiento.

I. INTRODUCCIÓN

El hombre es un ser naturalmente sociable, nos encontramos en una era donde compartir en redes sociales ocupa gran importancia en nuestras vidas, muchos buscan compartir sus experiencias y ser partícipes de las de otros, ya sean sus amigos o personas ajenas a su círculo más próximo.

En la actualidad las redes sociales forman una interacción social permanente que involucra a personas, grupos y organizaciones; conectados por diferentes tipos de relaciones tales como gustos, aficiones, política, etc., que proporcionan información para la toma de decisiones. Políticos están tomando en serio las publicaciones en diferentes redes sociales para enterarse de su popularidad y así realizar adecuadamente sus acciones de campaña. La falta de un Analizador de Sentimientos que permita catalogar opiniones mediante técnicas de minería de datos en redes sociales que se considera costoso y trabajoso, hace que necesiten herramientas de minería de texto que permita alcanzar resultados sobre la opinión de una persona hacia un político.

Profundizar en Twitter por la brevedad de sus mensajes, facilidad de publicación y además de la simplicidad de extracción de información, nos hace inclinarnos por tratar de conocer el contexto de cada tweet por personalidad política.

El objetivo principal es crear un modelo que permita medir la popularidad de un político utilizando Twitter como la red social para el análisis. El interés de hacer un modelo es comprender de qué manera los tweeteros escriben hacia una personalidad política y conocer su popularidad a través del valor de cada tweet que puede ser positivo, negativo o neutral.

Al analizar los tweets se emplea minería de opinión para clasificar su polaridad con aproximaciones supervisadas al necesitar a priori un diccionario de palabras positivas y negativas.

Las limitaciones al momento de extraer la información de un tuitero están sujetas a Twitter. Como una limitación al analizar el léxico de los tweets se encuentra las reglas gramaticales que tiene el idioma español.

II. REVISIÓN BIBLIOGRÁFICA

La habilidad de entender rápidamente actitudes del consumidor y reaccionar adecuadamente es algo que Expedia Canadá aprovechó cuando se dieron cuenta de que había un crecimiento continuo de feedback negativo hacia la música de uno de sus anuncios de televisión. El peor anuncio, lo repiten tanto que llega a ser

molesto, y probablemente no vuelva a usar Expedia, solo porque su anuncio es tan irritante.” El análisis de sentimiento que la marca llevó a cabo reveló que la música de uno de sus anuncios llegaba a ser muy irritante tras varias escuchas, y los consumidores se estaban dirigiendo a las redes sociales para desahogarse. Un par de semanas después de que el anuncio se emitiera por primera vez, más de la mitad de los comentarios de la conversación online eran negativos. En vez de anotarse un fallo con el anuncio, Expedia supo dirigir el sentimiento negativo de una manera divertida y consciente al emitir una nueva versión del anuncio que mostraba como rompían el violín que tanto molestaba.

La mayor parte de los sistemas de Análisis de Opiniones se centran en el tratamiento de textos en inglés. En el caso de textos escritos en español, probablemente el sistema más relevante sea *The Spanish SO Calculator* [1], desarrollado en la Universidad Simon Fraser de Canadá. Este sistema, además de resolver las opiniones almacenadas a nivel individual en adjetivos, sustantivos, verbos y adverbios; trata modificadores de la polaridad como son la negación o los intensificadores (*\muy*, *\poco*, *\bastante*, ...). También detecta y descarta el sentimiento relegado en el contenido no fáctico del texto, representado, por ejemplo, mediante expresiones condicionales o subjuntivas.

La manera más habitual de tratar todas estas construcciones lingüísticas es a nivel léxico y en este aspecto *The Spanish SO Calculator* no es una excepción. En lo que respecta al tratamiento de la negación, Taboada [7] utiliza información morfológica para identificar el alcance de la negación, mientras que Yang [9] considera dicho alcance como los términos situados a la derecha de la negación y Fernández [2] emplea una heurística que asume que los tres elementos a continuación de una negación son los que deben cambiar su polaridad. Para la intensificación, Fernández [2] considera de nuevo que los tres términos a la derecha son los que deben variar su polaridad. Taboada propone [7] que además de los intensificadores propiamente dichos, se trate como aspectos del discurso, la conjunción *\pero* o las mayúsculas.

III. METODOLOGÍA

Los datos fueron extraídos desde la API de Twitter. Se eligió los precandidatos presidenciales más sonados en el ambiente político de Ecuador. Se obtuvo 12 precandidatos presidenciales, de los cuales se extrajo todos los tweets donde se mencionen aquellos personajes.

Para cada precandidato se tuvo un máximo de 100000 tuits.

Los tuits extraídos fueron de dos fechas diferentes para realizar la comparación de actividad de cada precandidato en la red social.

Fecha 1: 2016-03-26 / 2016-04-04

Fecha 2: 2016-06-06 / 2016-06-15

Como dato importante se podría decir que el punto de equilibrio para tomar las muestras de los tweets sería el Terremoto en Ecuador el 16 de abril de 2016.

Tabla 1. Cuentas de Twitter de Precandidatos

Precandidato	Cuenta de Twitter
Alberto Acosta	@AlbertoAcostaE
Cynthia Viteri	@CynthiaViteri6
Dalo Bucaram	@dalo10
Gabriela Rivadeneira	@GabrielaEsPais
Jaime Nebot	@CynthiaViteri6
Jorge Glas	@JorgeGlas
Guillermo Lasso	@LassoGuillermo
Lenin Hurtado	@lenhurtado
Lenin Moreno	@LeninMorenoPAIS
Rafael Correa	@MashiRafael
Mauricio Rodas	@MauricioRodasEC
Ramiro González	@RamiroGonzalezJ

La limpieza de los datos es una de las etapas más importantes de este proyecto para obtener mejores resultados al ejecutar el modelo.

Al momento de la extracción de los tweets se encontró varios problemas para analizarlos posteriormente, por tal razón se ve la necesidad de realizar la siguiente limpieza en un tweet:

- Remover el RT de los retweet.
- Convertir todo el texto a minúsculas.
- Eliminar el @ y lo que tiene a continuación.

- Remover signos de puntuación.
- Eliminar números.
- Remover los links.
- Eliminar las Tabulaciones.
- Quitar los espacios en blanco del inicio del texto.
- Quitar los espacios en blanco del final del texto.
- Remover textos iguales.
- Eliminar tildes.

A continuación, se presenta un tweet y el resultado de la limpieza aplicada en este.

Tweet antes de la limpieza: “RT @RamiroGonzalez: Sin estar de acuerdo, seguimos apoyando nuestro compromiso es con los ecuatorianos. # GabrielDiaz # EcuadorRenace http...”

Tweet después de la limpieza: “sin estar de acuerdo seguimos apoyando nuestro compromiso es con los ecuatorianos gabrieldiaz ecuadorrenace”

Para lograr clasificar un tweet como positivo o negativo se emplea el uso de diccionarios de palabras Positivas y Negativas, de esta manera proceder a comparar las palabras que conforman el Tweet con las palabras que se encuentran en el diccionario, con la finalidad de obtener el total de palabras positivas y negativas que conformen el tweet. El diccionario de palabras fueron tomadas del diccionario The General Inquirer, [6] el cual fue traducido a idioma español, contando así con los datos registrados en la Tabla 2 presentada a continuación:

Tabla 2. Diccionario de Palabras Positivas y Negativas.

<i>Palabras</i>	<i>Cantidad</i>
Positivas	2006
Negativas	4783

La técnica de minería de datos se aplicó al análisis de sentimientos en la polaridad de sus publicaciones que se lo realiza mediante una clasificación Score el cual está definido como $Score = \#palabras\ positivas - \#palabras\ negativas$, que permite realizar una comparación de los tweets con el diccionario de datos.

Tabla 3. Parámetros de score.sentiment

Score	Publicación
Score > 1	Positivo
Score < 0	Negativo
Score =0	Neutral

Al tener la polaridad de cada texto se procedió a dar un valor a los tweet, positivos (1), negativos (0) y neutros (2). Un tuitero puede tener varios tweet, cada uno de ellos depende de su estado de ánimo al realizar las diferentes publicaciones en la red social. Se le asignó al tuitero el tipo de comentario que tenga más peso, si tiene los tres tipos de comentarios se le asigna como comentario neutro.

Tabla 4. Modelo de popularidad

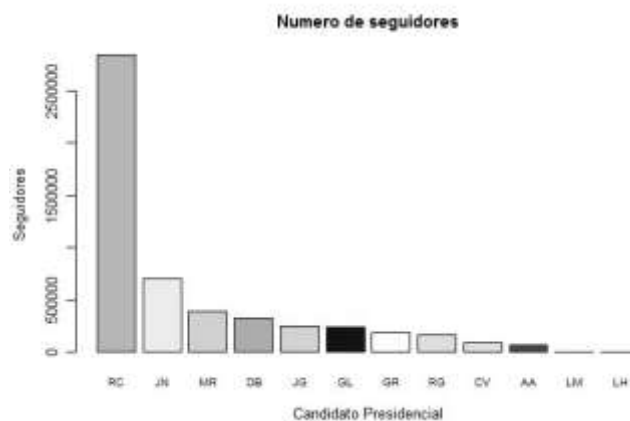
Candidato	Tweetero	Texto/Tweet	Polaridad	Intención
X	Tweetero 1	T1	+	N
		T2	-	
		Tn	N	
	Tweetero 2	T1	+	-
		T2	-	
		Tn	N	
	Tweetero n	T1	+	+
		T2	-	
		Tn	N	
	Tweetero 1	T1	+	N
		T2	-	

Y	Tweetero 2	Tn	N	+
		T1	+	
		T2	-	
	Tweetero n	Tn	N	-
		T1	+	
		T2	-	

Al obtener la afinidad de cada tuitero pudiendo ser positivo, negativo o neutral hacia un precandidato presidencial se puede obtener el nivel de popularidad por la cantidad de tweeteros que comentan de manera positiva, con estos datos se realizó varios análisis para profundizar la actividad que tiene cada político en Twitter.

IV. RESULTADOS

Los principales resultados que se obtuvieron del Análisis de Opiniones en Twitter por precandidato son: número de seguidores que tiene un precandidato, datos obtenidos del Twitter del mes de agosto del 2016.

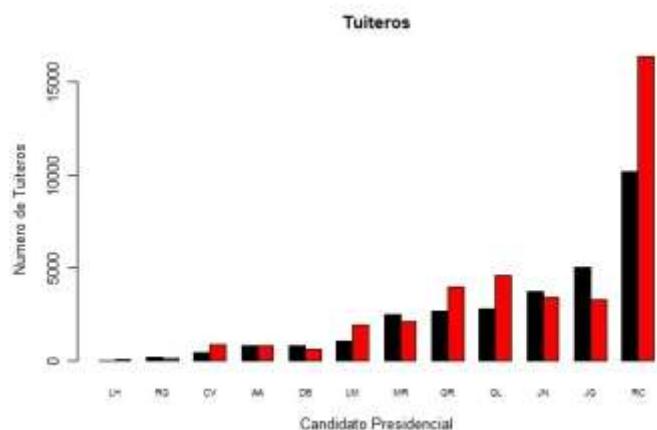


Como se observa en el gráfico el que tiene mayor número de seguidores es el precandidato Rafael Correa y lleva una gran diferencia al segundo precandidato que es Jaime Nebot. Se concluye que el Ec. Rafael Correa es el más popular de todos los precandidatos en la red social Twitter.

Tabla 5. Número de seguidores por precandidatos

Precandidato	Seguidores
Alberto Acosta	68.096
Cynthia Viteri	90.291
Dalo Bucaram	324.058
Gabriela Rivadeneira	190.961
Jaime Nebot	698.594
Jorge Glas	246.530
Guillermo Lasso	238.814
Lenin Hurtado	2.744
Lenin Moreno	9.303
Rafael Correa	2.844.155
Mauricio Rodas	386.937
Ramiro González	168.319

Esta grafica representa el número de tuiteros que han comentado sobre el precandidato presidencial en las dos fechas señaladas anteriormente para conocer su actividad en la red social Twitter.



Para un jefe de campaña lo importante es que su candidato se encuentre presente en los comentarios en Twitter y que siempre este en constante crecimiento. Al observar el gráfico, no todos los precandidatos crecieron en publicaciones, como es el caso de Ramiro González, Alberto Acosta, Dalo Bucaram, Mauricio Rodas, Jaime Nebot y Jorge Glas.

Tabla 6. Número de tuiteros que han comentado por precandidato en diferentes fechas

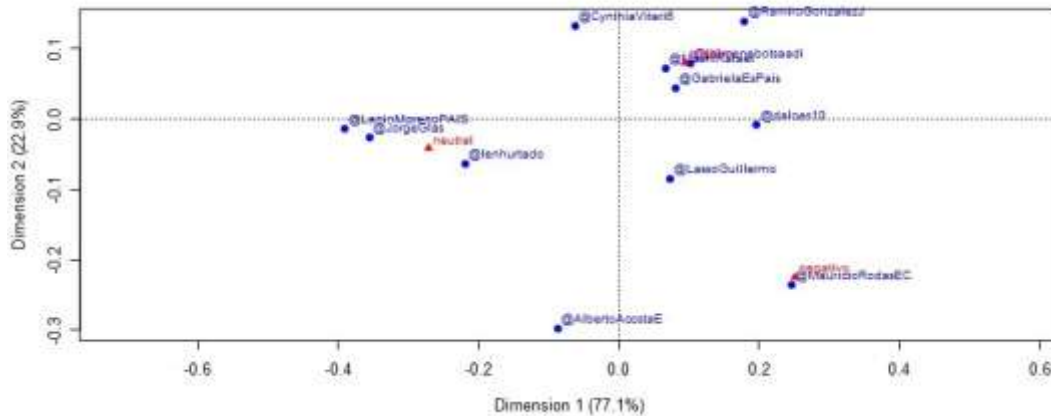
Precandidato	Fecha1	Fecha2
Alberto Acosta	829	850
Cynthia Viteri	454	919
Dalo Bucaram	526	672
Gabriela Rivadeneira	2689	3996
Jaime Nebot	3734	3393
Jorge Glas	5014	3312
Guillermo Lasso	2777	4578
Lenin Hurtado	52	101
Lenin Moreno	1069	1946
Rafael Correa	10143	16360
Mauricio Rodas	2507	2120
Ramiro González'	205	168

Al comparar los precandidatos se tenía el problema de que cada uno de ellos tenía diferentes muestras por lo que se realizó un análisis factorial de correspondencias (AFC) que asigna un peso proporcional a su importancia en el total. De esta forma se evita que al trabajar con perfiles se privilegie a las clases de efectivos pequeños.

En la figura se evidencia que los precandidatos que tienen tuiteros que comentan de mejor manera son: Rafael Correa, Jaime Nebot, Gabriela Rivadeneira y junto a ellos se proyectan Dalo Bucaram, Cynthia Viteri y Ramiro González.

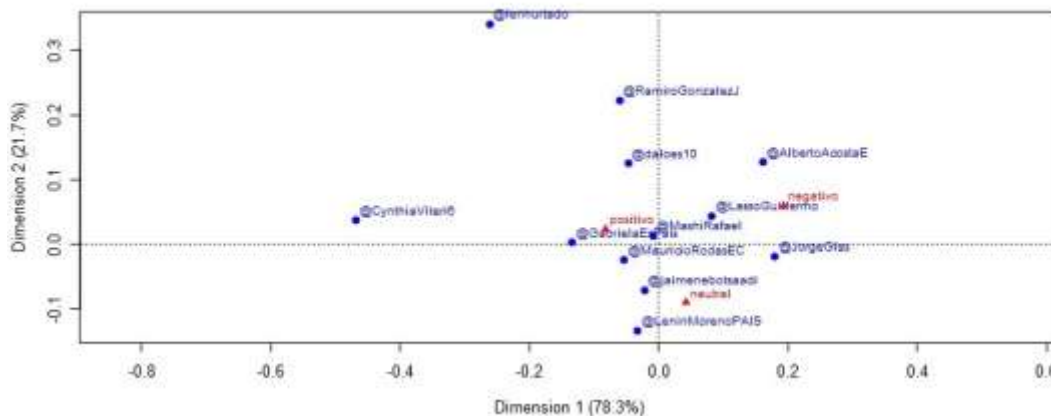
En el caso de tuiteros que comentan de manera neutral están los precandidatos: Lenin Moreno, Jorge Glas, Lenin Hurtado y Alberto Acosta.

Por el lado de los tuiteros negativos se encuentra Mauricio Rodas y Guillermo Lasso.



En la fecha 2 se encontró que los precandidatos que tienen tweeteros que comentan de mejor manera son: Rafael Correa, Gabriela Rivadeneira, Mauricio Rodas y junto a ellos se proyectan Dalo Bucaram, Cynthia Viteri, Ramiro González y Lenin Hurtado.

Por los tweeteros neutrales están los precandidatos: Jaime Nebot, Lenin Moreno y Jorge Glas. En los tweeteros negativos se encuentra Guillermo Lasso y Alberto Acosta.



Al comparar las dos fechas se observó que algunos candidatos se mantienen constantes en el apego de sus tweeteros como: Rafael Correa y Gabriela Rivadeneira.

Por otro lado, un precandidato que mejoro fue Mauricio Rodas que en la fecha 1 estaba con más tweeteros negativos y para la fecha 2 apareció con mayor cantidad de tweeteros positivos, al contrario de él está el precandidato Jaime Nebot que en la Fecha 1 estuvo con tweeteros positivos y en la fecha 2 se encontró que tweeteaban más personas de manera negativa.

En las dos fechas se encuentran constantes los candidatos: Lenin Moreno y Jorge Glas, con mayor cantidad de tweeteros que comentan de manera neutral.

Al momento de proyectarse hacia tweeteros positivos se encontraron los precandidatos: Cynthia Viteri, Dalo Bucaram, Ramiro González y Lenin Hurtado.

Para el precandidato Guillermo Lasso en las dos fechas se situó en el cuadrante de los tweeteros que comentaban negativamente.

Al revisar los comentarios de todos los precandidatos se puede observar que la mayoría de ellos tiene mayor cantidad de tweeteros que hablan positivamente y como segundo lugar comentarios neutrales dejando por debajo a los tweeteros que comentan negativamente.

Un caso diferente es el del precandidato Jorge Glas en donde la fecha 1 tiene una gran cantidad de tweeteros que comentan neutralmente y para la fecha 2 incrementan la cantidad de tweeteros con comentarios positivos.

<i>Candidato</i>	<i>Fecha1</i>			<i>Fecha2</i>		
	<i>Positivo</i>	<i>Negativo</i>	<i>Neutral</i>	<i>Positivo</i>	<i>Negativo</i>	<i>Neutral</i>
<i>Alberto Acosta</i>	331	172	326	440	219	191
<i>Cynthia Viteri</i>	266	39	149	559	173	187
<i>Dalo Bucaram</i>	511	148	197	519	21	132
<i>Gabriela Rivadeneira</i>	1580	369	740	2203	634	1159
<i>Jaime Nebot</i>	2270	488	976	1808	730	855
<i>Jorge Glas</i>	2208	402	2404	2047	449	816
<i>Guillermo Lasso</i>	1476	482	819	2177	1078	1323
<i>Lenin Hurtado</i>	24	6	22	76	16	9
<i>Lenin Moreno</i>	466	75	528	1054	278	614
<i>Rafael Correa</i>	6040	1283	2820	9248	2932	4180
<i>Mauricio Rodas</i>	1292	623	592	1222	333	565
<i>Ramiro González</i>	134	26	45	107	34	27

V. CONCLUSIONES Y TRABAJO FUTURO

Este proyecto no se trata de predicción electoral se torna muy complejo decir por quien va a votar una persona y más aún si son en las redes sociales por la complejidad que existe en el idioma español, por el sesgo que se encuentra al momento de realizar un análisis de este tipo, estimar el grado y la ausencia de participación es difícil y además hay que considerar que la red social Twitter no es la más usada por los ecuatorianos.

Al comparar los precandidatos se tuvo el problema del tamaño de muestras por cada uno de ellos por lo que se decidió realizar un análisis factorial de correspondencias que nos permitió asignar un peso proporcional a su importancia en el total y así evitar que se privilegie a los precandidatos de muestras pequeñas.

REFERENCIAS

- [1]. Brooke, J., M. Toloski, y M. Taboada. 2009. Cross-Linguistic Sentiment Analysis: From English to Spanish. En Proceedings of the International Conference RANLP-2009, p_ aginas 50{54, Borovets, Bulgaria. ACL
- [2]. Fernández Anta, A., P. Morere, L. Nuñez Chiroque, y A. Santos. 2012. Techniques for Sentiment Analysis and Topic Detection of Spanish Tweets: Preliminary Report. En TASS 2012 Working Notes, Castell_on, Spain..
- [3]. Girvan M, Newman M. E. J (2002), "Community structure in social and biological networks", Recuperado el 04/05/2016 de: <http://www.pnas.org/content/99/12/7821>.
- [4]. Larrosa J, (2014/05/02), "Detección de Comunidades- Algoritmo de Girvan Newman" [Mensaje en un Blog], Recuperado de <http://ars-uns.blogspot.com/2014/05/deteccion-de-comunidades-algoritmo.html>
- [5]. Pang, B. , L. Lee, y S. Vaithyanathan. 2002. Thumbs up? Sentiment classification using machine learning techniques. En Proceedings of EMNLP pag 79-86.
- [6]. Stone P, (2002/09/02) "Welcome to the General Inquirer home Page" [Sitio Web] Obtenido de <http://soybibliotecario.blogspot.mx/2016/09/normas-apa-2016-aprende-citar-posts-de.html?m=1>
- [7]. Taboada, M., J. Brooke, M. Toloski, K. Voll, y M. Stede. 2011. Lexicon-based methods for sentiment analysis. Computational Linguistics, 37(2):267-307.
- [8]. UNESCO (1996): La educación encierra un tesoro. Informe a la UNESCO de la Comisión internacional sobre la educación para el s. XXI. Madrid, Santillana- UNESCO.
- [9]. Yang, K. 2008. WIDIT in TREC 2008 blog track: Leveraging multiple sources of opinion evidence. En E.M. Voorhees y Lori P. Buckland, editores NIST Special Publication 500-277: The Seventeenth Text Retrieval Conference Proceedings (TREC 2008)