**Introducción**

Por años se han utilizado metodologías como las encuestas por los investigadores para medir los pensamientos e intenciones de una población de individuos. Estas incluyen un número de desventajas en las que se incluyen el esfuerzo humano y su coste económico y en relación a su tiempo de aplicación. Con la llegada de los medios sociales, hay una abundante información online donde las personas expresan sus sentimientos con respecto a una gran variedad de temas (Bermingham, A. y Smeaton, 2011).

Twitter es un servicio de "microblogging" que comenzó a dar servicio en 2006 y posee más de 20 millones de visitas mensuales. En Twitter, cada usuario puede publicar breves mensajes de una longitud de 140 caracteres también llamados "tweets", que son visibles por los demás usuarios (Tumasjan et al., 2010).

El flujo público de las publicaciones de todos los usuarios de todo el mundo constituye una enorme fuente de información a tiempo real (Tumasjan et al., 2010).

La idea que subyace a estas publicaciones es dar a conocer el estado personal de cada usuario (Tumasjan et al., 2010); sin embargo, eventos de especial relevancia para la vida de las personas como son las elecciones, o eventos políticos de gran importancia, influyen en el contenido de las publicaciones de cada usuario. Los mensajes de Twitter son usados frecuentemente para compartir sentimientos sobre el mundo que nos rodea, y la disponibilidad de este contenido social, generado en estos sitios web, crea nuevas oportunidades para el estudio de la opinión pública (Nakov et al., 2013).

Sin embargo, este punto resulta controvertido, pues mientras algunos analistas políticos tienen la vista en Twitter como un indicador de las diversas opiniones políticas (Skemp, 2009; citado en Tumasjan et al., 2010); otros autores han afirmado que la mayoría de los mensajes constituyen un "parloteo sin sentido" (pearanalytics, 2009; citado en Tumasjan et al., 2010).

Parloteo, o no, se ha de considerar que las opiniones expresadas en las redes sociales juegan un papel crucial a la hora de influenciar el comportamiento de la opinión pública en áreas como la compra de un producto (Bai, 2011; Eirinaki, Pisal y Singh, 2012; citado en Mostafa, 2013).

**Método**

Obtención de los datos

Se han examinado 164.424 tweets extraídos diariamente entre las fechas 24-04-2016 y 10-05-2016. Para su obtención se ha utilizado los softwares libres: R y Rstudio (R Core Team, 2015). Además de la librería “twitteR” que permite extraer desde la API de twitter mensajes los diferentes usuarios en función de una serie de parámetros de búsqueda.

Los criterios de búsqueda utilizados hacen referencia a aquellos partidos políticos que obtuvieron un mayor número de votos en las pasadas elecciones españolas del 20-12-2015. A saber: Partido Popular (“Partido+Popular”; “PP”); Partido Socialista Obrero Español (“Partido+Socialista”, “PSOE”); Podemos (“Podemos”); Ciudadanos (“Ciudadanos”, “Cs”, “C´s). Además, para evitar tweets en idiomas diferentes al español se han filtrado todos aquellos tweets que no estuviesen en este idioma.

Compilación de los datos y eliminación de “retweets”

Las búsquedas realizadas proporcionan además del mensaje de texto (“tweet”) información sobre el usuario, fecha y hora de la entrada y si el mensaje ha sido tomado por un usuario y publicado en su perfil (“retweet”).

Puesto que, salvo el mensaje del usuario, el resto de información es poco útil en la investigación que se ha llevado a cabo, se ha extraído el mensaje aislado, salvándolo como un archivo “.txt” y eliminando los “retweet”, puesto que al tomar el mensaje de otro usuario se almacenarían mensajes idénticos que podría perjudicar el corpus de datos final.

Filtrado de los datos

Posteriormente a la búsqueda, se han recopilado las búsquedas en un corpus usando para ello el paquete “tm” (Text Mining) del softwate R. Posteriormente se han eliminado los signos de puntuación, los acentos, se han convertido todos los caracteres a minúscula y se han suprimido las stopwords (palabras como determinantes y preposiciones que no aportan significado al mensaje y, que al ser las palabras más frecuentes en el lenguaje, alteraría el análisis de aquellas palabras que aportarían significado al mensaje). Es importante destacar que todo ello son herramientas que ofrece el programa “tm” y que se ha considerado el castellano como idioma a la hora de seleccionar y suprimir las stopwords.

**Referencias**

Nakov, P., Rosenthal, S., Kozareva, Z., Stoyanov, V., Ritter, A. y Wilson, T. (2013). SemEval-2013 Task 2: Sentiment Analysis in Twitter. Second Joint Conference on Lexical and Computational Semantics, 2, 312-320.

Tumasjan, A., Sprenger, T. O., Sander, P. G., & Welpe, I. M. (2010). Predicting elections with Twitter: What 140 characters reveal about political sentiment. Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media. Menlo Park, CA: The AAAI Press, 178-185.

Bermingham, A., & Smeaton, A. (2011). On using twitter to monitor political sentiment and predict election results. In Proceeding of IJCNLP conference, Chiang Mai, Thailand.

R Core Team. (2015). R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria. URL <https://www.R-project.org/>.

Mostafa, M.M. (2013). More than words: Social networks´ text mining for consumer brand sentiments. *Expert System with Application, 40,* 4241-4245.