

MODELO DE PREDICCIÓN LINEAL DEL NÚMERO DE TWEETS ENVIADOS POR LOS USUARIOS DE TWITTER

Manuel Suárez Gutiérrez

RESUMEN

La predicción del total de mensajes enviados diarios a través de la red social online de Twitter, permite establecer una tendencia sobre su uso, asimismo identificar si tiene una preferencia por los usuarios o si por el contrario se encuentra en un proceso de decaída. Para este análisis se tomó una muestra por cinco semanas comprendidas durante el mes de octubre y noviembre de 2016, y a través de estos datos seleccionar un modelo de predicción para formular un pronóstico sobre el comportamiento de los Tweets para la sexta semana.

ABSTRACT

The prediction of the total of messages sent daily through the online social network of Twitter, allows to establish a tendency on its use, and also to identify if it has a preference for the users or if on the contrary it is in a process of decay. For this analysis a sample was taken for five weeks comprised during the month of October and November of 2016, and through these data select a prediction model to formulate a forecast on the behavior of the Tweets for the sixth week.

INTRODUCCIÓN

El presente ensayo tiene como propósito, hacer una predicción sobre el número de mensajes que son enviados a través de la red social online de Twitter, estos mensajes son denominados como Tweets y cuentan con la peculiaridad de tener una limitante de 140 caracteres, lo cual hace que sean sumamente concretos. Además, estos mensajes cuentan con diversas características, una de ellas es la capacidad de agregar Hash Tags (#), las cuales son etiquetas muy concretas y unificadas, permitiendo a los analistas de redes sociales identificar los Trending Topics, (los temas más populares), en las redes sociales online. Para la formulación de un pronóstico será necesario seguir el procedimiento planteado en el apartado de metodología, el cual establece que se requiere tomar una muestra por un

periodo de cinco semanas, con la finalidad de obtener una estimación para la sexta semana. Posteriormente en el apartado de resultados, se detallará el modelo seleccionado para hacer la predicción y se justificará el mismo. Por último, en las conclusiones se valida si el procedimiento propuesto es el más apto para medir el comportamiento de las redes sociales online.

METODOLOGÍA

Para la realización de la predicción de los Tweets, fue necesario como primer aspecto analizar diversas herramientas disponibles en Internet para medir el uso de las redes sociales online, pero sobre todo que permitiera capturar el total de Tweets enviados en un día. Por lo que se seleccionó la herramienta de

“internet live stats”¹, la cual busca difundir estadísticas en tiempo real sobre el uso del Internet y redes sociales, sin embargo, su principal inconveniente, es que no almacena un historial sobre los contadores mostrados para estadísticas diarias, estos se reinician a la media noche. Por lo que, para tener una información precisa sobre el número de Tweets enviados en un día, se debe tomar la medida segundos antes de que el contador se reinicie. Derivado de lo anterior, a través de la herramienta seleccionada se tomaron 34 lecturas consecutivas diarias, durante el periodo comprendido entre el 2 de octubre y el 4 de noviembre de 2016, en el horario de las 23:59 Horas. Teniendo como resultado los datos mostrados en la Tabla 1.

A partir de los datos observados y

Tabla 1
Mediciones de Tweets por día

Día	Domingo	Lunes	Martes	Miércoles	Jueves	Viernes	Sábado
Semana 1	635,920,485	636,099,859	636,252,126	636,357,190	636,439,682	636,591,825	636,696,047
Semana 2	636,215,215	636,338,319	636,923,194	637,023,079	636,975,222	637,042,362	636,905,571
Semana 3	636,860,097	636,711,573	636,786,835	637,689,517	637,116,835	637,439,939	637,670,043
Semana 4	637,593,147	637,326,251	637,436,355	637,676,459	637,826,563	638,510,889	638,440,993
Semana 5	638,284,097	637,977,201	638,127,305	638,567,409	638,802,211	639,020,769	N/D

Fuente. Elaboración propia, con datos de <http://www.internetlivestats.com/> (2016)

¹. <http://www.internetlivestats.com/>

medidos se elabora una serie histórica, a la cual se le seleccionará el método de predicción más adecuado con la finalidad de obtener estimaciones o pronósticos de valores futuros para la cuarta semana.

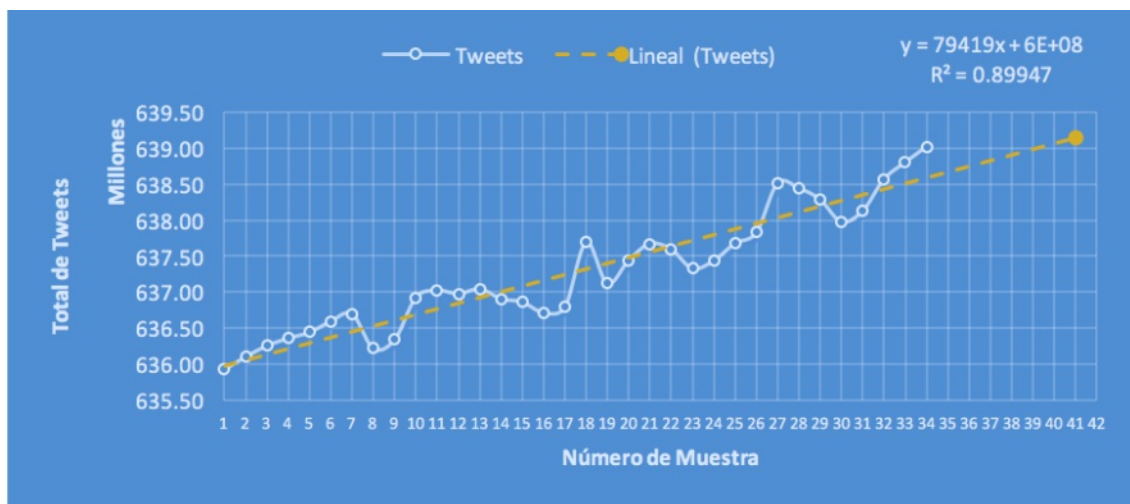
RESULTADOS

Al analizar los resultados obtenidos a partir del número de Tweets enviados en cada uno de los días comprendidos en el periodo de estudio, se tiene un comportamiento lineal, sin embargo, por el tipo de análisis se debe tomar en cuenta que existen agentes externos que influyen en el resultado, ejerciendo

un efecto para que los usuarios de las redes sociales online tengan una participación activa tanto por encima del valor medio como por debajo de este valor. Además, se logra identificar días de la semana que son más activos en las redes sociales online (viernes y sábados) y días menos activos (domingos y lunes).

Como criterio para seleccionar el modelo de predicción del comportamiento de los mensajes enviados diarios en Twitter, se tomó el valor del coeficiente de correlación R^2 , obteniendo para la ecuación lineal un valor de 0.89947, el cual, al ser valor cercano a 1 permite dar un valor aproxima-

Gráfica 1
Pronóstico de Tweets con una tendencia lineal.



Fuente. Elaboración propia, 2016

do de la conducta de los Tweets para 7 días, en donde se espera que para el séptimo día se tenga un valor cercano a los 639.5 Millones de Tweets por día.

CONCLUSIONES

La muestra obtenida permite establecer con mayor probidad el comportamiento de Twitter para los próximos 7 días, aunque existen fenómenos externos que pueden llegar a inferir en el comportamiento de los usuarios, sin embargo, estos afectan con menor medida al tener una mayor cantidad de información. Se propone como próxima investigación ampliar la muestra para poder hacer un análisis por día de la semana e identificar concretamente los días de mayor y menor actividad en Twitter.