

Análisis de Sentimiento con Twitter API

Luis Henriquez, Felipe Quisel and Gianfranco Maiocchi

ARTICLE INFO

Keywords:

sentiment
tweet
subjectivity
API
python

Resumen

Conocer la opinión de un sector de la población puede ser tan importante para lanzar un producto en marketing, como conocer la opinión de los votantes, en política. En Ecuador, la red social Twitter se ha convertido en uno de los principales medios de interacción directa entre figuras políticas y población. Entonces, un estudio que refleja los sentimientos en español por modismos de cada región, nos brinda una gran oportunidad para estudiar 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 (SA) utilizando una herramienta de PNL adaptada a la variación del español utilizada en Ecuador, aprovechando que la mayor parte de la literatura se ha centrado en el idioma inglés, mientras que las adaptaciones para idiomas, al igual que el español, son mínimos y aún están en proceso debido a la complejidad inherente al idioma.

1. Introducción

El análisis de sentimientos (o minería de opiniones) es una técnica de **Procesamiento del Lenguaje Natural** (PNL), que es un campo de la inteligencia artificial (IA) combina el poder de la lingüística y la informática para estudiar las reglas y la estructura del lenguaje y crear sistemas inteligentes (ejecutados en aprendizaje automático y algoritmos de PNL) capaces de comprender, analizar y extraer significado del texto y el habla, por lo que aplicado al análisis de sentimientos se puede determinar si los datos de entrada en formato texto representan un sentimiento positivo, negativo o neutral.



Figura 1: Análisis de sentimientos.

Este se realiza a menudo en internet para ayudar a las empresas a monitorear el sentimiento de marca y producto en los comentarios de los clientes y comprender sus necesidades, también es importante mencionar que se este tipo de análisis se ha llevado a cabo también en la política. [1].

Tweepy en informática, es una biblioteca de Python, que trae métodos para extraer información de usuarios de Twitter, es de código abierto, y gratis, el cual se utilizará en esta investigación, para extraer información de un usuario de prueba y posteriormente, analizar el significado o sentimiento de esta.



Figura 2: Combinación de twitter con python: Tweepy.

2. Desarrollo

Utilizando el paquete **Tweepy** y **Textblob**, se crean las funciones `strip_undesired_chars` que limpiará el tweet en cuestión, seguido de `get_all_tweets`, que extrae los tweets del usuario y que posteriormente serán analizados. Luego, como se puede extraer información de cualquier usuario, se debe agregar una condición para el lenguaje del tweet, y traducirla al español. Finalmente se crea la función `main_function` que le pedirá al usuario ingresar el idioma del tweet, seguido del nombre del usuario a buscar, y finalmente imprimirá en pantalla la cantidad de tweets analizados.

A modo de demostración, se utilizó al usuario `sebastianpinera` para analizar los sentimientos de los tweets, como se verá en las figuras siguientes:

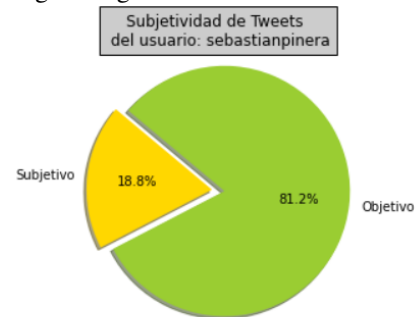


Figura 3: Subjetividad de los Tweets.

En la figura 3, se observa una subjetividad del 18.8% contra una objetividad del 81.2%. Esto quiere decir que los tweets realizados por `sebastianpinera` son realizados de forma relativa al tema en sí mismo.

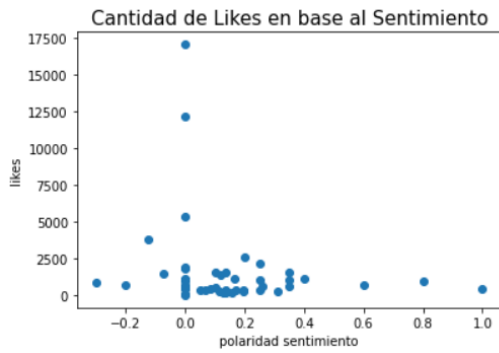


Figura 4: Sentimientos de los likes.

En este gráfico, se observa la cantidad de likes y su polaridad, ya sea negativo, positivo o neutro, el rango va entre -1 y 1. La tendencia es de sentimientos positivos, siendo que la mayoría de likes está entre 0.0 y 0.4.

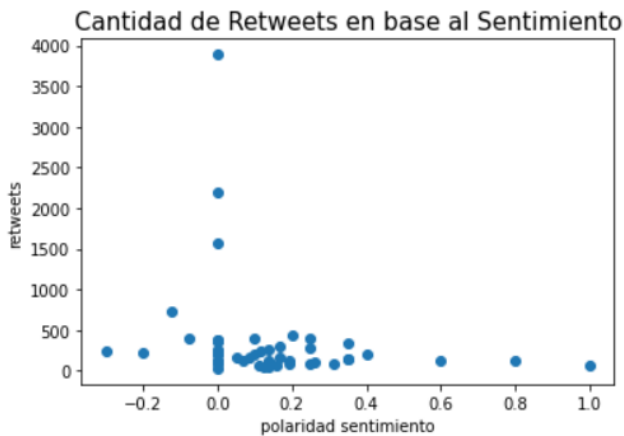


Figura 5: Sentimientos de retweets.

En este gráfico, se observa la cantidad de retweets y su polaridad, ya sea negativo, positivo o neutro, el rango va entre -1 y 1. La tendencia es de sentimientos positivos, siendo que la mayoría de likes está entre 0.0 y 0.4, similar al gráfico anterior.

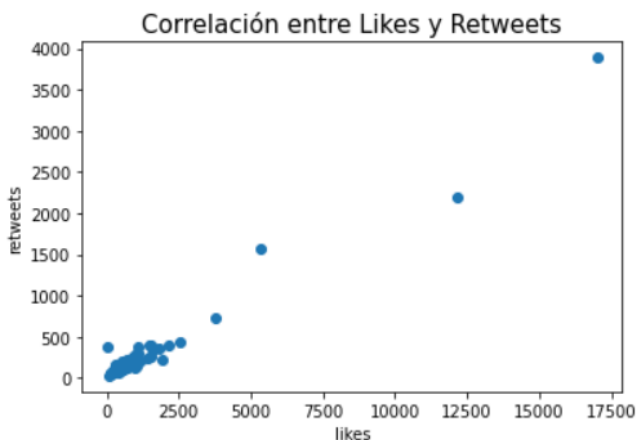


Figura 6: Correlación entre likes y retweets.

En este gráfico, se observa la relación entre la cantidad de retweets y likes por retweeteo. El rango mayor de likes en comparación con sus retweets está entre 0 y 2.500.



Figura 7: Wordcloud.

En la figura 7 se realizó el wordcloud de los tweets del usuario sebastianpinera.

3. Conclusión

El libre acceso a paquetes de tweepy y textblob, permite el análisis profundo de sentimientos de los tweets, y de la cantidad de likes que tiene el post, permitiendo recopilar datos de usuarios para realizar alguna estadística a elección o simplemente, monitorear alguna marca o producto que se esté vendiendo u ofreciendo, siendo de gran ayuda para empresas o comerciantes que buscan emprender en el área de negocios y necesitan conocer a profundidad el interés de usuarios en general.

Referencias

- [1] Computing, A., 2020. Sentiment analysis of facebook users reacting to political campaign posts. <https://dl.acm.org/doi/fullHtml/10.1145/3382735>.