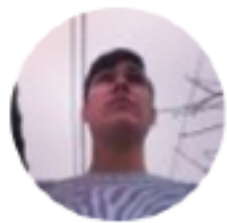


Carga y obtención

Twitter ofrece una gran potencia para problemas de minería de opinión. Si analizamos la **anatomía de un tuit**, podremos trazar un claro paralelismo entre este y la quintupla de opinión definida por Liu [1] en el que dice que una opinión estará formada por un quintupla de compuesta por: **Entidad, emisor, aspecto, orientación y momento temporal**.



Jose Angel Diaz @joseangeldiazg · 26 jul.

¿Quieres aprender #DataScience? ¿No has visto nunca #R ni #Python?
Entonces #DataCamp es tu mejor amigo...



1



Los datos han sido obtenidos de **Twitter** usando **Scrapy** para evitar las restricciones de la API de Twitter parseando la web de Twitter y almacenando los tuits en **MongoDB**.

El data set final está compuesto por 1.7M de tuits obtenidos en EEUU durante distintos meses de la primera mitad del año 2016. Para cada tuit guardamos: **id, datetime, has_media, is_reply, is_retweet, nbr_retweet, nbr_favorurite, nbr_reply, text, url, userid, username**.

[1] Liu B., Zhang L. (2012) A Survey of Opinion Mining and Sentiment Analysis. In: Aggarwal C., Zhai C. (eds) Mining Text Data. Springer, Boston, MA

Metodología

OBTENCIÓN DE DATOS

