**Descripción de la Práctica a realizar**

El objetivo de esta actividad será la creación de un dataset a partir de los datos contenidos al web. Tenéis que indicar las siguientes características del dataset general:

1. Título del dataset. Poned un título que sea descriptivo.

Titulo del dataset: TweetsPuraVida

1. Subtítulo del dataset. Agregad una descripción ágil de vuestro conjunto de  datos por vuestro subtítulo.

100 primeros Tweets al hashtag #PuraVida, desde el dia 9/11/2017 hasta el 1/10/2017, #PuraVida (Concepto de Costa Rica refiriéndose a Naturaleza y Positivismo).

1. Imagen. Agregad una imagen que identifique vuestro dataset visualmente



1. Contexto. ¿Cuál es la materia del conjunto de datos?

La materia recogida son los tweets generados en twitter ligados al hashtag #PuraVida en la que se recogen comentarios sobre el concepto PuraVida y la repercusión que este tiene. Se observa que aunque sea un concepto nacido en Costa Rica este se expande por todo el mundo. Es un concepto que contempla un estado personal muy agradable y recomendable y nos ayuda a buscar la parte positiva de las cosas además del cuidado de la naturaleza.

1. Contenido. ¿Qué campos incluye? ¿Cuál es el periodo de tiempo de los datos y  cómo se ha recogido?

Se han recogido 100 tweets y se para ello se ha utilizado la Api de Twitter.

Primero se ha abierto la conexión con Twitter pasándole los parámetros de conexión, había sido necesario antes pedírselos a Twitter ya que dispone de estos para cada usuario de Twitter y para cada aplicación que se vaya a desarrollar.

Luego se ha pedido a Twitter que entregue 100 Tweets sobre el hadstag #PuraVida y que no fueran mas antiguos que el 01/10/2017. Como resultado se han conseguido los 100 Tweets pero al ser un tema al que la gente aporta mucho y les gusta solo hemos recibido Tweets desde el dia 07/11/2017 hasta el dia 9/11/2017 ya que se ejecutó el dia 9 y el dia 7/10/2017 ya se habían recogido 100.

Todo ello utilizando desde R, con R Studio y la Api de Twitter.

Los campos para cada uno de los tweets recibidos son:

Text, Favorited, favoriteCount, replyToSN, created, truncated, replyToSID, id, replyToUID, statusSource, screenname, retweetCount, isRetweet, retweeted, longitude, latitude

Todos los campos son importantes, el Text (la descripción de lo que se ha puesto y si se ha agregado alguna foto o dibujo), created(Fecha de creación), statusSource(desde donde se ha creado, incluido dispositivo), screenname(quien ha generado el tweet), longitude y latitude(desde donde se genera), aunque estos dos últimos campos pocas veces están informados. Luego, información también importante es si el Tweet se ha retuiteado y cuantas veces, si se tiene como favorito. La verdad es que hay mucha información que recoger y que viene muy bien si quieres utilizar este canal para conseguir que se hable de algo y ver de que forma se esta reaccionando.

1. Agradecimientos. ¿Quién es propietario del conjunto de datos? Incluid citas de  investigación o análisis anteriores.

Agradezco en primer lugar a los autores de las URL utilizadas para poder descubrir de que forma podía trabajar con R para llevar a cabo la práctica.

-Para poder recoger los tweets y pasarlos a un fichero con format csv

Autor: Justin Meyer

<http://rprogramming.net/write-csv-in-r/>

-Para ver como utilizar la App de Twitter desde R

Autor: Bogdan Rau, MPH

<http://bogdanrau.com/blog/collecting-tweets-using-r-and-the-twitter-search-api/>

-A Twitter y sus desarrolladores por publicar la URL donde se pueden conseguir los códigos necesarios para así poder acceder desde R a la red de Twitter

<https://dev.twitter.com/apps>

Por último a Laia Subirats, por la paciencia y dedicación que le ha supuesto el ayudarme a conseguir la programación en R y así poder obtener este resultado.

Como propietario del conjunto de datos recogidos pienso son cada uno de los screenname o autores de los diferentes Tweets recogidos aunque son de acceso público.

1. Inspiración. ¿Por qué es interesante este conjunto de datos? ¿Qué preguntas  le gustaría responder la comunidad?

Me parece interesante poder ver que una idea del positivismo de #PuraVida puede llegar tan lejos y conseguir que se enganche y lo tenga en cuenta tanta gente, con el resultado de esta práctica se puede ver eso. Además de que su potencia de propagación es enorme.

1. Licencia. Seleccionad una de estas licencias y decid porqué la habéis  seleccionado:
   * ○  Released Under CC0: Public Domain License
   * ○  Released Under CC BY-NC-SA 4.0 License
   * ○  Released Under CC BY-SA 4.0 License
   * ○  Database released under Open Database License, individual contents  under Database Contents License
   * ○  Other (specified above)
     + - 1. ○  Unknown License

Para este caso, selecciono una licencia **CC BY-ND** de forma que permite que otros copien, distribuyan, muestren y realicen solo copias originales de mi trabajo. Si quieren modificar mi trabajo, primero deben obtener mi permiso.

1. Código: Hay que adjuntar el código con el que habéis generado el dataset, preferiblemente con R o Python, que os ha ayudado a generar el dataset

[Codigo en R](Pura%20Vida%20sin%20claves.R)

1. Dataset: Dataset en formato CSV

[Dataset en CSV](tweets.csv)

[Dataset en Excel](tweets.xlsx)