

## **PRÁCTICA 2: Análisis del impacto de la pandemia en la salud mental de usuarios/as de Redes Sociales (parte II)**

Esta práctica es una continuación de la anterior, en la que trataremos de dar respuesta a otros retos adicionales. Para ello, utilizaremos herramientas de extracción de dimensiones psicológicas que nos permitirán estimar si existe algún tipo de relación entre los eventos acaecidos en estos dos últimos años y lo que las personas publican en sus redes sociales. No todo el mundo tiene un perfil de redes sociales, pero estudios previos han demostrado que estos contenidos suelen aportar una buena pincelada de la sociedad en general.

### **Pasos a seguir:**

1) para la extracción de dimensiones psicológicas latentes en el texto utilizaremos la librería *Empath* (<https://github.com/Ejhfast/empath-client>) desarrollada por la Universidad de Stanford.

Tal y como se especifica en la documentación, *Empath* proporciona una serie de categorías léxicas por defecto para las cuales podemos analizar nuestro texto. Sin embargo, estas categorías son demasiado amplias para nuestro caso de uso, analizar dimensiones psicológicas. Para ello, la librería nos da la opción de, a partir de una palabra semilla, obtener una serie de subcategorías léxicas más específicas de un dominio concreto. Podéis hacer pruebas de esto en el siguiente enlace: <http://empath.stanford.edu/>. **Pista:** tened cuidado también con el modelo que fijáis para el análisis, tened en cuenta que en este caso estamos trabajando con *tweets* y deberéis seleccionar uno que tenga unos tipos de textos similares.

Debéis pensar en qué conjuntos de palabras pueden estar asociadas a un determinado perfil psicológico y, a partir de eso, generar el correspondiente análisis en EMPATH. Por ejemplo, podéis optar por analizar perfiles relacionados con la ansiedad, el estrés, pensamientos suicidas, etc.

Continuaremos trabajando con el corpus de US data de la práctica anterior.

2) Hacer algún tipo de representación gráfica de esas dimensiones psicológicas, utilizando la granularidad temporal que mejor se ajuste para la extracción de conclusiones.

Si añadimos la curva de contagios a la representación de sentimiento del apartado anterior, nos permitirá extraer mejores conclusiones.

3) (apartado opcional, 0.5 puntos) Utilizar la herramienta *praw* (<https://github.com/praw-dev/praw>) o *alien* (<https://github.com/labteral/alien>) para realizar un análisis adicional con una nueva colección de datos extraídos de Reddit. Para obtener este 0.5 a mayores no sería necesario limitar los posts a una localización concreta como US, pero sí al periodo que nos interesa.

4) (apartado opcional, 0.5 puntos) Ampliar la colección con posts de Reddit, pero ahora **localizados en US**. Para esto podéis proponer diferentes estrategias que se os ocurran, pero una posible, sería filtrar por subreddits.

## Entregables:

- 1) Guión python (.py)
- 2) Python Notebook (.pynb)

Es fundamental que el Notebook sea autoexplicativo de todos los pasos (con celdas textuales acompañando a celdas con código y que contenga explícitamente los resultados -sin tener que ejecutar las celdas de nuevo-). Comprobad esto antes de enviar el Notebook. Cualquier proyecto de Analítica de Datos debe ser autodocumentado y sus experimentos fáciles de reproducir. Un aspecto clave en la evaluación de esta práctica reside en la calidad de las explicaciones y documentación que acompañéis al código dentro del Notebook.

- Valoración y Fecha de Entrega:
  - Esta práctica tiene una valoración de 3 puntos (sobre el total de 7 puntos de la parte práctica de la materia). 2 puntos se corresponden a la correcta realización de los apartados obligatorios -apartados de 1) a 2)- y el punto restante se corresponde con la correcta realización de los apartados optativos.

Fecha límite entrega: 10 de diciembre

Se permiten entregas retrasadas pero se reducirá la puntuación del siguiente modo:

- Cada día tarde reduce en un 10% de la máxima nota alcanzable (es decir, cada día tarde resta un 0.3 puntos de la nota que se os asigne al valorar la práctica).