

Práctica 2: Análisis de tuits

Análisis de Redes Sociales 2024/25 (1C)

Fecha	Descripción de la versión
2024-10-17	Enunciado de la práctica 2

Objetivo

El objetivo de esta práctica es realizar una investigación en conjuntos de datos extraídos de Twitter, para descubrir relaciones entre distintos tipos de datos del contenido.

Se podrá usar uno o varios de los conjuntos de datos propuestos por el profesor, disponibles en el campus virtual.

Se deberán utilizar técnicas computacionales para extraer y justificar las conclusiones obtenidas, implementando el código en el repositorio de prácticas.

Requisitos

- Utilizar `python`, gestionado con `uv`, y formateado con `black`.
- Utilizar `pandas` para gestionar los datos en el código.
- Utilizar una o varias librerías de procesamiento del lenguaje natural, como `NLTK`, `spaCy` o `TextBlob`, para procesar el contenido de los tuits. La instalación y configuración de estas librerías no es trivial, así que se recuerda que está disponible el foro del CV para preguntas y para ayudarnos entre nosotros.
- Utilizar `click` para gestionar la interfaz de mi aplicación, que deberá recibir un conjunto de datos en formato `csv` y realizar los distintos cálculos, procesos de extracción de información, etc., y la generación de los gráficos y/o tablas necesarios para el informe final.
- Escribir un **informe** en `markdown`. Las fuentes `markdown` deberán formar parte del repositorio `GitHub`. Se usará `pandoc` para convertir las fuentes en un documento `PDF`, que se entregará también.
- El informe deberá estar formateado y redactado adecuadamente.

- Las conclusiones del informe deberán estar justificadas con evidencia computacional reproducible, usando el código entregado y los conjuntos de datos elegidos.
- **No** se permitirá el uso de notebooks de jupyter como entrega, ni para el código ni para el informe, aunque se podrán usar como herramienta de exploración si algún estudiante los encuentra cómodos.

Contenido del Informe

El informe debe contener descubrimientos, análisis, conclusiones... (“insights”) sobre los datos investigados. Ejemplos de insights pueden ser:

- Los temas más habituales, temas puntuales, temas sorprendentes...
- La evolución de un tema (término, conjunto de términos, tema implícito...) a lo largo del tiempo.
- La relación o falta de ella de temas entre sí, o de temas con otras variables numéricas primitivas (retuits, likes...) o derivadas (sentimiento, polaridad...), o entre distintas variables numéricas.

Se valorará la originalidad en las cuestiones exploradas, así como la variedad de ellas. No se requerirán unas conclusiones sorprendentes, sino una demostración del dominio de distintas técnicas de análisis de contenido, así como la reflexión e interpretación sobre las cuestiones planteadas.

Se valorará el uso de gráficos informativos y de una narrativa sintética pero explicativa sobre las conclusiones obtenidas.

Información adicional

Se recuerda a los estudiantes que el proyecto final a realizar para la asignatura consistirá en repetir el trabajo de las prácticas con las redes sociales asignadas a cada grupo, por lo que no es mala idea ir haciendo esta práctica con un ojo puesto en las redes asignadas.

En el campus virtual hay un ejemplo básico de markdown, junto con el PDF generado, que incluyen además algunos ejemplos de los tipos de gráficas que se podrían hacer.

Entrega

1. El informe se entregará, en formato PDF, en el campus virtual, en la tarea habilitada para tal efecto.
2. El código se entregará a través del mismo repositorio git utilizado para la práctica 1. Se deberá crear una etiqueta entrega_p2 en el repositorio, que será la versión del código que el profesor corregirá. El commit deberá estar subido al repositorio remoto a tiempo para la fecha límite de entrega.
3. El repositorio debe incluir todas las instrucciones necesarias para ejecutar el programa en el archivo README, así como la documentación adicional que se considere oportuna. El repositorio debe incluir también cualquier archivo de configuración que sea necesario y las fuentes markdown del informe. El repositorio **no deberá incluir** los conjuntos de datos.
4. La evaluación de la tarea se realizará en el campus virtual. Se ha incluido una rúbrica donde se puede ver los criterios de corrección que se usarán, que se recomienda consultar antes de empezar a realizar la práctica.
5. La fecha de entrega límite es: **2024-11-05 23:55**.