# TEMA 3 – HOJA DE EJERCICIOS III

Se plantean una serie de ejercicios relacionados con la representación de textos mediante modelado latente de temas.

**Ejercicio 1. Probabilistic Topic Model. LDA.**

Dado un fichero en formato JSON con noticias se quiere obtener los diferentes topics utilizando LDA. En este primer ejercicio se proporciona el notebook completo hacerlo, los pasos son los siguientes:

1. Lectura del fichero y carga de datos.
2. Preprocesamiento básico de los textos.
3. Entrenamiento del algoritmo Latent Dirichlet Allocation de sklearn. Se recomienda ver la documentación de la librería para ver los diferentes parámetros del algoritmo.
4. Visualización de resultados.
5. Evaluación de los resultados.

Se pide ejecutar el código proporcionado y realizar ajustes sobre él para ver cómo cambian los resultados:

1. Cambiar el número de topics, ahora está a 2, pero revisando el fichero con las noticias, en realidad serían más.
2. Cambiar el preprocesamiento de los textos, eliminando stopwords al menos y lo que se considere para ver si mejoran los resultados.

¿Cómo impacta en los resultados los diferentes cambios de los apartados anteriores?

**Ejercicio 2. Probabilistic Topic Model. LDA. Inferencia.**

Sobre la base del ejercicio anterior se quiere añadir una nueva noticia y realizar inferencia para ver los topics en ella.

El texto de la noticia a inferir es el siguiente:

> "Trump ordena suspender toda la ayuda militar de Estados Unidos a Ucrania tras su bronca a Zelenski."

> Se tiene que mostrar en qué % se ajusta la noticia a los diferentes topics existentes.