# Enunciado práctica NLP + recomendadores + generativos

**NLP**

1. A partir del dataset FinancialPhraseBank (<https://huggingface.co/datasets/financial_phrasebank>) utilizando únicamente las frases financieras con un 100% de coincidencia en la etiqueta realizar las siguientes representaciones tokenizando de forma adecuada:

\* Bag of words

\* TF-IDF

\* Word-to-vect

Visualizar las representaciones de las frases obtenidas mediante alguna técnica de reducción de dimensionalidad si fuera necesario.

2. Mediante el dataset Financial PhraseBank (para las frases 100% agree con el dataset de clase) construir un clasificador de sentimientos con las siguientes técnicas.

Comparar sobre test los resultados de vuestro modelo con algún modelo de análisis de sentimiento ya implementado como nltk y etc:

Se evaluarán los resultados en función de la distancia al benchmark 0.98 acc (allagree), la complejidad de la solución (embeddings utilizados y entrenados, etc) y la explicación del proceso.

\* Redes recurrentes

\* Redes convolucionales

\* Transformers

3. Hacer fine tuning de un modelo BERT pre entrenado aplicado al mismo dataset.

4. Usar un checkpoint de un modelo t5-small pre entrenado para summarization en el dataset *cnn\_dailymail*(<https://huggingface.co/datasets/cnn_dailymail>) y hacer fine tuning para obtener un modelo que resuma noticias.

**Sistemas de recomendación**

1. Completar el modelo NeuMF sobre el dataset movilens visto en clase y analizar los resultados del modelo. Aplicar si es necesario algún tipo de preprocesado sobre los datos de entrada para obtener una mejor performance.

2. Extraer los embeddings del modelo entrenado en el ejercicio anterior y realizar una visualización de los usuarios e items. Seleccionar usuarios cercanos y comprobar si las recomendaciones realizadas mediante usuarios cercanos tienen un valor alto en la predicción del modelo.

**Modelos generativos**

1. Queréis hacer un sistema de búsqueda por similitud. Tenéis un dataset (FashionMNIST train) sobre el que podéis entrenar o calcular cosas. Ahora tenéis un dato arbitrario de input (imagen *"query"*, cogida de FashionMNIST test) del cual queréis encontrar otro existente dentro de una base de datos (FashionMNIST test, sin la imagen *"query"*) lo más similar posible al dato de input. Desarrolla el sistema y evalúa cualitativamente con ejemplos de su funcionamiento.

2. Diseña un VAE para generar series de precios de las empresas constituyentes del S&P500.

3. Imaginemos que del dataset de train de FashionMNIST sólo tenemos 100 datos etiquetados disponibles y hemos perdido las etiquetas del resto; no obstante, conservamos las imágenes. Queremos entrenar un clasificador lo más preciso posible. Para ello:

- Entrena un modelo supervisado usando únicamente los datos etiquetados de los que disponemos. Elige parámetros y arquitectura, argumentando la elección de forma razonada.

- Ahora usa de alguna forma el espacio latente de un modelo no supervisado pre-entrenado con los datos no etiquetados. Prueba diferentes modelos y comenta, detalladamente, los resultados. ¿Cómo varían los resultados con el número de datos etiquetados?

Para EVALUACIÓN en cada uno de los casos, tenemos disponible los datos de test con sus etiquetas.

4. Haz un modelo generativo condicional (GAN) sobre FashionMNIST usando sus etiquetas como condicionante. Explica las decisiones razonadamente.