

Proyecto 2

Herramientas que veremos durante el curso:

- Automl, Watsonx, Bigml, Make.com, Notion, Knime, APIs de IA

Fase 1. Selección y Exploración del Dataset

- **Datasets sugeridos:**
 - Archivos de texto con reseñas de productos (por ejemplo, reseñas de Amazon, Yelp o TripAdvisor). Este tipo de dataset contiene datos no estructurados como texto libre que deberá analizarse.
 - Emails recibidos con novedades de idealista
 - Datasets de <https://dataverse.csuc.cat/>
- **Entregables:**
 - Análisis exploratorio del dataset: identificación de patrones iniciales, limpieza necesaria y desafíos.
 - Documentación de herramientas propuestas para cada etapa.

Fase 2. Limpieza y Estructuración de Datos

- **Tarea:**
 - Transformar el texto no estructurado en un formato estructurado, por ejemplo:
 - Extraer campos como puntuación, palabras clave, sentimiento, etc.
 - Normalizar texto (eliminar símbolos, convertir a minúsculas, etc.).
 - Transformar emails que van llegando en hojas de excel o tablas en **notion**
 - Automatizar este proceso utilizando:
 - **Make.com:** Para automatizar la extracción de texto desde diferentes fuentes (por ejemplo, APIs de reseñas, emails, archivos de texto...).
 - **KNIME:** Para procesar el texto y estructurarlo en un formato tabular.
- **Entregables:**
 - Un archivo CSV o base de datos estructurada.
 - Un flujo de trabajo documentado de las herramientas utilizadas.

Fase 3. Creación y Entrenamiento del Modelo Predictivo

- **Tarea:**
 - Entrenar un modelo de aprendizaje automático para predecir algún campo del dataset (o datasets) trabajados.
 - Utilizar herramientas como:
 - **BigML:** Para automatizar el entrenamiento y evaluación del modelo.
 - **AutoML o Watsonx:** Como alternativa para optimizar el rendimiento del modelo.
- **Entregables:**
 - Un modelo predictivo.
 - Evaluación del modelo con métricas adecuadas.