

MACHINE LEARNING

Manuel Luján Vilchez

PROYECTO FINAL



MANUEL LUJÁN VILCHEZ

RESPONSABLE DEL DESARROLLO

**Fecha:** 01/09/2023

**DESARROLLO**

1. Seleccionar un *dataset* de OpenML según las indicaciones previas.
2. Cargarlo con la función auxiliar de Scikit-learn: [from sklearn.datasets import fetch\_openml](https://scikit-learn.org/stable/datasets/loading_other_datasets.html#downloading-datasets-from-the-openml-org-repository).
3. Plantear un notebook o “cuaderno del científico” de ML donde iremos tomando anotaciones, hipótesis/experimentos y sus resultados, etc., documentando el proceso como memoria del proyecto.
4. Analizar mínimamente los datos: ¿están limpios y completos?, ¿en qué formato?, ¿hay muchos *outliers* a limpiar?, ¿nos permitirán construir las características de forma sencilla?
5. Construir un *dataset* inicial con los datos listos para utilizar. Para ello, podemos guardar el *dataset* en un archivo CSV exportado y comprobar si existe en dicho entorno.

* Si ya existe el archivo, el código que limpia y prepara los datos no se ejecuta, sino que lo importa.
* Si no existe, el código limpia y prepara los datos y exporta el *dataset* a CSV.

1. Plantear al menos 3-4 hipótesis del modelo a entrenar:

* conjunto de características a utilizar (y cuáles no)
* relaciones entre las mismas
* variable objetivo
* tipo de modelo
* …

1. Para cada una de las hipótesis, entrenamos un modelo simple y contrastamos la hipótesis evaluándolo.

Para dicho entrenamiento, usamos un porcentaje de los ejemplos disponibles y sólo un *dataset* de entrenamiento y test (sin validación), ya que lo que buscamos es obtener una idea rápida sobre la idoneidad de las mismas.

1. Tras evaluar las hipótesis, nos decantamos por un tipo de modelo y comenzamos a entrenarlo en profundidad, siguiendo las explicaciones vistas en el curso:

* Escoger un conjunto de características.
* Preprocesar los datos, procesos y *subsets* de validación, y test final.
* Comprobación de hiperparámetros.
* Evaluación del modelo con métricas en detalle.
* …

1. En dicho proceso de entrenamiento final, buscamos *iterar, iterar e iterar*, planteando múltiples posibilidades o hipótesis, comprobándolas y anotando sus resultados.
2. En cada iteración, evaluamos el modelo con detenimiento:

* Conjunto de métricas de evaluación
* Representación de errores o residuos
* Validación
* Hiper-parámetros
* Desviación vs sobreajuste
* …

**CONCLUSIONES**

Debes explicar claramente tu proyecto.

Y deberás redactar unas conclusiones donde se indique si se ha llegado a obtener algún resultado que satisfaga alguna de las preguntas iniciales.