



# Problema Sesión 03

Introducción a la IA: Machine Learning

## PREPARACIÓN DE DATOS PARA UN MODELO DE ML

Hemos explorado los diferentes aspectos que involucran la preparación de datos, así como todas sus posibilidades, por lo que ha llegado el momento de que preparemos un dataset y que, una vez listo, podamos entrenar un modelo de machine learning.

La exploración y preparación de los datos requiere de atención y paciencia. Pues bien, si sigues el fastbook recomendado en este lab mientras realizas este ejercicio, ¡seguro que conseguirás dejar tus datos a punto!

### Objetivos de este ejercicio

La idea principal de esta actividad es que puedas cargar, explorar, analizar y aplicar los distintos procedimientos involucrados en la preparación de datos (limpieza, transformación, *splitting*...) a un conjunto de datos real, ¿con qué intención?

Con la que nos compite a nosotros, los expertos en IA: que puedan servirnos para entrenar un modelo de machine learning.

Si aplicas en esta práctica todas las pautas vistas en el fastbook y en la clase en directo, podrás superar el próximo test y, todavía más importante, embargarte en futuros proyectos sin miedo a esta tarea.

## Descripción de la actividad

En primer lugar, deberás elegir un conjunto de datos sobre el que aplicar el preprocesado: puedes [clickar en este enlace hacia la web Kaggle](#) y descubrir una gran cantidad de ellos. Importante: necesitas registrarte para poder descargar estos datasets.

¿Y qué es (o por qué) Kaggle?

*Kaggle* es una web muy conocida que cuenta con un amplio repositorio de enseñanzas, foros, datasets..., así como los notebooks (en la sección Code, cuando clicamos en un dataset) realizados por su comunidad. Estos notebooks nos enseñan desde la exploración y preparación de datasets hasta la aplicación de los modelos y técnicas de machine learning más novedosos, por lo que es una fuente de conocimiento y consulta muy utilizada.

Cuando uses su buscador, descubrirás diferentes temáticas, a través de palabras clave. Quizá si estás interesado en alguna en concreto, te motives aún más a la hora de realizar esta tarea. Si además filtras los resultados por los más votados o activos, descubrirás que muchos usuarios los han tratado, ventaja que te da, porque puedes aprender de ellos.

En definitiva: te invitamos a que dediques un poco de tiempo a esta exploración.

Pero volvamos al tema que nos concierne: nuestra actividad. Intenta que el dataset esté explicado en su descripción para que lo puedas entender (sección *Data Card*), que tenga sentido, sea variado, con campos numéricos y categóricos, etc. para que puedas aplicar todo lo aprendido, y que no sea excesivamente grande en cuanto a campos y variables para que no sea inabarcable.

¡Y ojo!, también te tropezarás con datasets de imágenes y otros tipos de datos, pero en nuestro caso **buscamos solo datos estructurados en tablas**, ¡recuérdalo!

Si esta tarea de búsqueda te resulta compleja o no tienes tiempo, unas líneas más abajo, en este documento, te compartimos unos datasets como ejemplos. Puedes elegir cualquiera de ellos y utilizarlos; o tomarlos de referencia para buscar uno similar y que te guste más.

- <https://www.kaggle.com/datasets/neuromusic/avocado-prices>
- <https://www.kaggle.com/datasets/jsphyg/weather-dataset-rattle-package>

- <https://www.kaggle.com/datasets/ahsan81/hotel-reservations-classification-dataset>

## Formato de entrega

Realiza el preprocesado de datos en un **Jupyter notebook** indicando tu nombre y apellido. Preferible si eliges el formato *.ipynb* (el propio de los Jupyter notebooks) para que sean interactivos o, si no, en formato PDF.

Y ya sabes: comparte el archivo en el campus, en el apartado 'Entrega', y podrás acceder al test final.

## Criterios de corrección

Para autoevaluar el ejercicio en la sesión con el profesor o durante la tutoría, comprueba que has seguido un criterio lógico a la hora de aplicar distintos procesos de la preparación de datos: ten en cuenta que no existe una guía exacta que se aplique siempre de la misma manera, pero el fastbook te ayudará a entender los pasos que debes seguir. No obstante, cualquier duda que te haya surgido al realizar este problema traslázalas a tu instructor.



[Qualentum.com](https://Qualentum.com)