

ESQUEMA (PASOS A SEGUIR EN LA PRÁCTICA)

- 1) Carga de datos y exploración de estos (función cargado de datos del dataset Airbnb)**
- 2) Normalización del valor a predecir**

En el caso de datos numéricos y categóricos:

- 3) Selección de las variables que son de interés**
- 4) Procesamos las variables numéricas y categóricas (normalización de datos)**
- 5) Búsqueda y asignación de valores perdidos**
- 6) Split de los datos**
- 7) Generación del modelo (optimización del número de capas, el número de neuronas (recordar que construiremos un encoder y por tanto, el número de neuronas se va reduciendo, ver que funciones de activación se utilizan, etc.)**
- 8) Compilación del modelo (seleccionar funciones de pérdidas y optimizador junto con su learning_rate)**
- 9) Entrenar y optimizar hiperparámetros**
- 10) Rescalar el precio a los valores reales**

En el caso de las imágenes:

- 1) Normalizamos sus valores**
- 2) Mismo split que en el caso de los datos numéricos y categóricos**
- 3) Desarrollamos el modelo convolucional (tenemos varias opciones: fine-tuning, transfer learning o from scratch).**
- 4) Compilamos y entrenamos el modelo optimizando hiperparámetros. Evitamos overfitting con las técnicas que hemos visto (Dropout, max-pooling, regularización) etc.**

Podemos realizar el modelo híbrido de distintas formas:

- Early fusion: sacar características de imágenes (generarnos un vector de características por imagen), concatenarlo con los datos numéricos y categóricos e introducir el vector concatenado a un modelo.**
- Late fusion: crear un modelo con dos ramas, cada rama procesará cada tipo de datos (una rama para los datos numéricos y categóricos y otra para las imágenes), unimos las dos ramas y diseñamos el clasificador final.**