

# Memoria Práctica 7

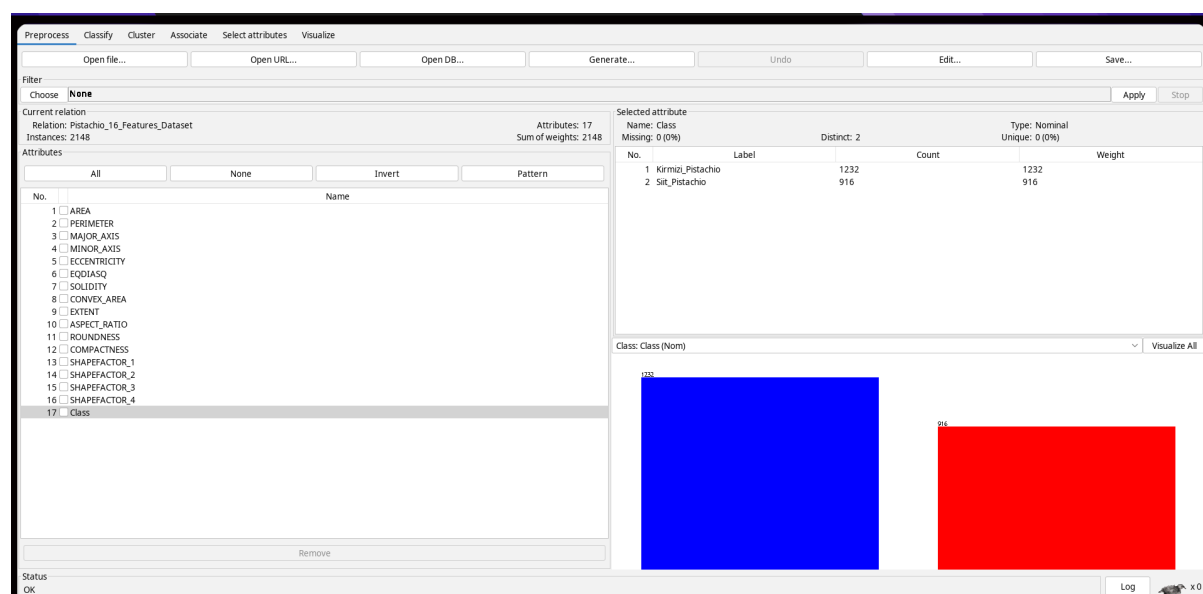
## 1. Elección y descripción del dataset.

Para llevar a cabo esta práctica, hemos escogido un conjunto de datos de la página web <https://www.kaggle.com/> y, concretamente el siguiente:

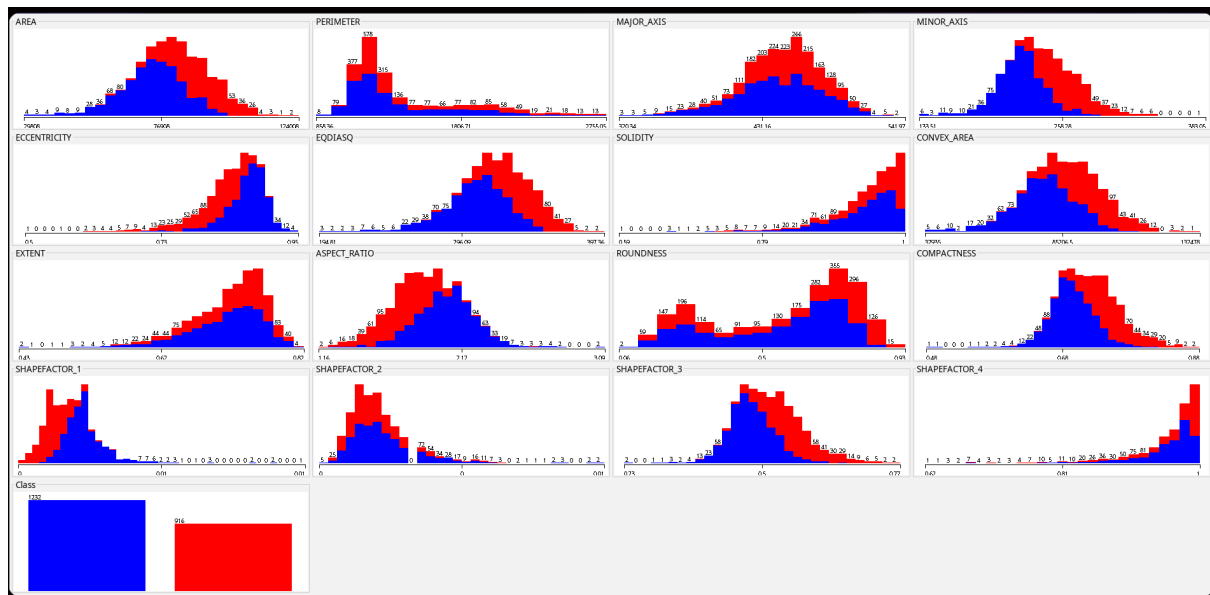
<https://www.kaggle.com/datasets/muratkokludataset/pistachio-dataset>

El dataset está en formato arff y, por lo tanto, ya está preparado para analizarse con Weka. De dicho dataset hemos hecho una partición, donde el 20% de los datos se destina al conjunto de prueba y el 80% restante se destina al conjunto de entrenamiento.

El dataset elegido tiene 2148 instancias con 17 atributos y donde el objetivo es clasificar en dos tipos de pistachos: Kirmizi y Siirt, en este caso, hay 1232 de Kirmizi y 916 de Siit.



Información principal del dataset y visualización de los dos tipos de clase.



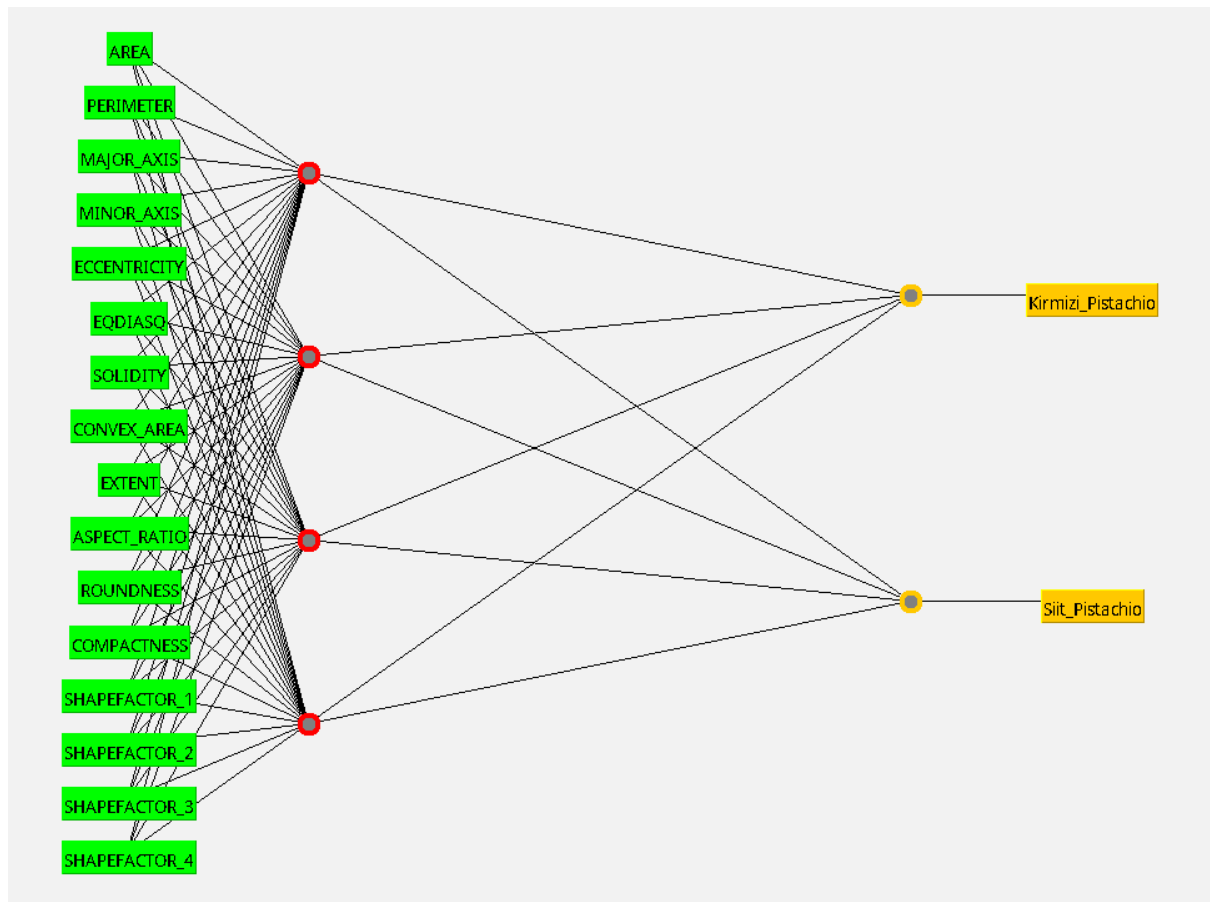
Visualización de las gráficas para cada atributo.

## 2. Análisis del Dataset

La mejor red neuronal con cross validation con configuración:

- **Validation threshold** → 100
- **ValidationSetSize** → 10
- **Training Time** → 2000

<b>Root Mean Squared Error</b>	0.3074
<b>Root Relative Squared Error</b>	29.5076 %



- Root Mean Squared Error y Root Relative Squared Error

MinNumObject / Confidence Factor	1	2	3
0.05	0.328 66.3208 %	0.3263 65.9845 %	0.3253 65.7712 %
0.12	0.3281 66.3357 %	0.3257 65.8654 %	0.3245 65.6151 %
0.2	0.332 67.1237 %	0.3268 66.0812 %	0.3261 65.9322 %
0.25	0.3301 66.7481 %	0.326 65.9228 %	0.3282 66.3674 %
0.3	0.3317 67.064 %	0.3263 65.9804 %	0.3266 66.2355 %
0.5	0.3362 67.3772 %	0.3312 66.9773 %	0.3335 67.4297 %



Árbol que obtiene el mejor resultado.

Si comparamos el resultado, podemos observar que el mejor árbol de decisión se obtiene con un confidence factor de 0.12 y con número mínimo de objetos igual a 3. Este árbol obtiene un error mayor al perceptrón multicapa ya que éste lleva a cabo un entrenamiento mucho más exhaustivo del dataset.