## Extra:

1. **Clasificación de enfermedades por sistema afectado (modelo supervisado)**

Objetivo: Dado un conjunto de características (por ejemplo, prevalencia, edad de aparición, tipo de herencia, síntomas), predecir a qué sistema del cuerpo afecta la enfermedad (neurológico, inmunológico, etc.).

Modelo: Árboles de decisión, Random Forest o XGBoost.

Datos necesarios: Variables categóricas/numéricas limpias, codificadas.

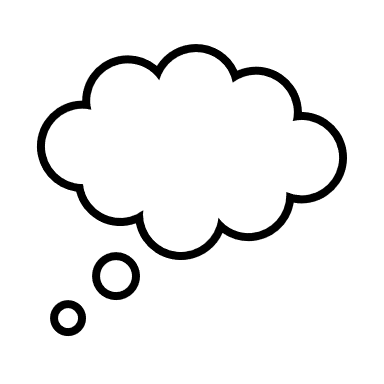
1. **Clustering (modelo no supervisado)**

Objetivo: Agrupar enfermedades raras con características similares (sin usar una variable objetivo).

Modelo: K-means, DBSCAN o PCA para visualización.

Identificar clusters de enfermedades poco estudiadas pero con rasgos comunes.

Explorar patrones ocultos que no se ven a simple vista.

ANÁLISIS FUTUROS

Analizar la interacción entre el fármaco cannabidiol (CBD) y la proteína (Ras/Rap GTPase-activating protein SynGAP) del gen SYNGAP1.

* ¿Hay datos que indiquen la existencia de una relación directa entre CBD y SYNGAP1?
* ¿Hay otras proteínas en el gen SYNGAP1 que puedan interaccionar con el CBD?
* ¿Qué efectos biológicos o rutas de señalización comparten el CBD y SYNGAP1?

¿Qué hacer?

1. Consultar bases de datos bioinformáticas para buscar interacciones existentes.
2. Analizar redes de proteínas para buscar unas similares a la de SYNGAP1.
3. Analizar la similitud química entre CBD y otros compuestos que interactúan con SYNGAP1.
4. Analizar enriquecimiento funcional para ver si están involucrados en rutas comunes.

Paso a paso:

1. STITCH (Search Tool for Interactions of Chemicals) y ChEMBL para verificar las interacciones CBD – SYNGAP1. [STITCH: chemical association networks](http://stitch.embl.de/)
2. STRING para encontrar proteínas similares a la de SYNGAP1.
3. Ver si hay otras proteínas similares a SYNGAP1. En ChEMBL o SwissTargetPrediction se puede buscar si existe interacción entre ellas y el CBD
4. Buscar fármacos que interactúen con SYNGAP1 y ver si son químicamente similares al CBD con RDKit.