

# MINERÍA DE DATOS

Grado en Ingeniería Informática. Doble Grado INDAT  
UNIVERSIDAD DE VALLADOLID

## PRÁCTICA: Kmedias(init), Kmedoids y DBScan

Con el fichero de datos de prácticas anteriores correspondientes a “semillas de trigo”, aplicar mediante un sólo programa en python (en formato jupyter), los siguientes algoritmos:

- Kmedias con la elección de baricentros iniciales expuesto en clase (Selección inicial de centros – diapositiva 41), que combina el algoritmo de ward, con un número de baricentros al final de 12.
- Kmedoids, también con 12 clusters, usando la variante por defecto y con centros iniciales establecidos con k-medoid++
- DBScan jugando con el valor de eps y mínimo de muestras por clusters para que salgan entorno a los 12 clusters antes de eliminar los puntos aislados.

Empleando el método de la moda para etiquetar a los clusters, compárese con la etiqueta de cada muestra y calcúlese la tasa de aciertos. Exponga estos resultados y verifique qué método ofrece mejores resultados.

**Nota:** véase que no se pide reservar una parte de las muestras para aprendizaje, ni para test; todos los cálculos pedidos anteriormente, son llevados a cabo con el conjunto de datos al completo.