

# Guía preparación Prueba 1

1. Defina qué es Minería de Datos.
2. Explique cuál es el valor de la Minería de Datos.
3. Explique las 4 metodologías de Minería de Datos vistas en clases.
4. Explique por qué es necesario escalar/normalizar variables.
5. ¿Cuál es la diferencia entre Normalización MinMax y Estandarización?
6. Explique la diferencia entre variables categóricas y numéricas.
7. Para los puntos que están en la tabla, ejecute *a mano* K-Means con  $k=2$ . Explique con claridad todos sus pasos y cálculos. Use como centroides iniciales los puntos (1,1) y (5,6).

x	y
0	1
0	0
1	0
4	4
5	4
6	6

- a. ¿Cuáles son los centroides finales?
  - b. Indique matemáticamente a qué cluster pertenece cada punto.
8. Repita el ejercicio pero ahora utilizando DBSCAN con los puntos de la tabla anterior.
    - a. Utilizando  $\text{MinPts}=2$  y  $\text{eps}=2$ , indique con claridad si un punto pertenece a un cluster o es ruido.
    - b. Utilizando  $\text{MinPts}=1$  y  $\text{eps}=1$ , indique con claridad si un punto pertenece a un cluster o es ruido.
  9. Realice un clustering jerárquico con los puntos de la Tabla 1.
    1. Utilice single linkage
    2. Utilice complete linkage
    3. Utilice average linkage
  10. Calcule las métricas de SSE, SSB y Silueta para los centroides finales del ejercicio 7
  11. Para las siguientes transacciones.

Transacción ID	Item
1	Chocolate, galleta
2	Dulce, chocolate, galleta
3	Helado, Chocolate
4	Galleta, dulce
5	Dulce, helado

1. Cuáles son los ítem sets con soporte = 30%

2. Cuáles reglas cumplen con confidence = 60%