

**UNIVERSIDAD TECNOLÓGICA DE
CHIHUAHUA
DESARROLLO Y GESTIÓN DE SOFTWARE**



**Extracción de Conocimiento en Bases de
Datos**

**Diagnóstico Unidad 4: Análisis no
supervisado**

Docente:

Enrique Mascote

Presentan:

Ian Carlos Chávez Rojo

Grupo:

IDGS91N

Fecha: 28/11/2025

Diagnóstico Unidad 4 Análisis no supervisado

Formato: 7 preguntas (4 opción múltiple, 3 respuesta corta)

1. ¿Cuál no es algoritmo de clustering?

- a) K-Means
- b) DBSCAN
- c) **PCA**
- d) Jerárquico

2. PCA sirve para:

- a) Clasificar imágenes
- b) Reducir dimensiones**
- c) Detectar outliers
- d) Ajustar hiperparámetros

3. Silhouette score mide:

- a) Coherencia interna de clusters**
- b) Precisión del modelo
- c) MSE
- d) Número de clusters

4. ¿Qué tipo de datos requiere t-SNE?

- a) Tablas relacionales
- b) Datos de alta dimensión**
- c) Series de tiempo
- d) Texto puro

Respuesta Corta

5. Define en una frase qué es un cluster.

No he investigado bien, pero creo que es un grupo de datos o elementos, que pueden ser iguales o diferentes entre sí. Es decir, que tienen cosas o características en común

6. Menciona un escenario donde usarías reducción de dimensionalidad.

La verdad no lo he utilizado o visto, pero según he aprendido en esta materia, puede ser que sea algo para visualizar datos de una manera que sea por ejemplo en alta dimensión

7. ¿Para qué sirve detectar outliers antes del clustering?

Pues no me se los términos pero según lo que he visto en esta materia, funciona para mejorar la calidad de los clusters, debido a que elimina datos no muy comunes.