

UNIVERSIDAD TECNOLÓGICA DE CHIHUAHUA

Tecnologías de la Información: Desarrollo y Gestión de Software



Diagnostico Unidad 4: Análisis no supervisado

IDGS91N - Kevin Iván Aguirre Silva

Extracción de Conocimiento en Bases de Datos - Ing.

Luis Enrique Mascote Cano

Chihuahua, Chih., 29 de noviembre de 2025

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

Define en una frase qué es un cluster:

Es un grupo de puntos de datos que son similares entre sí y disímiles a los puntos de otros grupos.

Menciona un escenario donde usarías reducción de dimensionalidad:

Visualización de datos de muy alta dimensión (más de 3) para facilitar el análisis exploratorio.

¿Para qué sirve detectar outliers antes del clustering?

Para evitar que distorsionen la formación de los centros de los clusters, especialmente en algoritmos sensibles como K-Means.