

# Métricas de Evaluación en Aprendizaje No Supervisado

Clustering y Segmentación de Clientes

# Contenido

1. Introducción al Aprendizaje No Supervisado
2. Métricas Internas
3. Métricas Externas
4. Ejemplos de Evaluación
5. Conclusión

# 1. Aprendizaje No Supervisado

- No se tienen etiquetas o valores “verdaderos” para entrenar.
- Objetivos comunes:
  - **Clustering:** Agrupar clientes con comportamientos similares.
  - **Reducción de dimensionalidad:** Visualizar datos complejos.
  - **Detección de anomalías:** Identificar comportamientos atípicos.

💡 Ejemplo:

Segmentación de clientes según frecuencia de compra y gasto promedio.

## 2. Métricas Internas (sin etiquetas)

Evaluamos la calidad de los clusters usando solo los datos:

### Silhouette Score

- Mide qué tan cercano está un punto a su cluster comparado con otros.
- Rango:  $[-1, 1]$  (1 = excelente separación)

💡 Ejemplo:

Clientes con alto gasto y frecuencia alta → cluster bien separado.

## Davies–Bouldin Index

- Evalúa compactación y separación de clusters.
- Entre más pequeño, mejor.

## Calinski-Harabasz Index

- Relación dispersión entre clusters vs dispersión dentro de clusters.
- Entre más grande, mejor.

### 3. Métricas Externas (requieren etiquetas)

Se usan cuando tenemos etiquetas reales para evaluar la coincidencia:

- **Adjusted Rand Index (ARI)**

Mide similitud entre clusters y etiquetas reales.

Rango:  $[-1, 1]$ , donde 1 = coincidencia perfecta.

- **Normalized Mutual Information (NMI)**

Mide información compartida entre clusters y etiquetas.

Rango:  $[0, 1]$

- **Fowlkes-Mallows Index (FMI)**

Combina precisión y recall sobre pares de puntos.

Rango:  $[0, 1]$

💡 Ejemplo:

Comparar clusters de clientes con categorías de suscripciones conocidas.

## 4. Ejemplo Práctico: Segmentación de Clientes

Supongamos que tenemos un dataset con:

- Frecuencia de compra
- Gasto promedio
- Tipo de producto comprado

### Visualización de clusters

Cluster 1: Clientes frecuentes y alto gasto  
Cluster 2: Clientes ocasionales y bajo gasto  
Cluster 3: Clientes medianos en ambos