



Módulo VI - Aprendizaje NO supervisado. Clase 19:

Tipos de aprendizaje NO supervisado





¿Ponemos a grabar el
taller?

OBJETIVOS de hoy



- Aprendizaje No Supervisado
- Tipo de algoritmos en Aprendizaje No Supervisado
 - Clustering
 - Reducción de la Dimensionalidad



REPASEMOS

Machine Learning

El objetivo del **aprendizaje automático** o machine learning es crear un **modelo** que permita resolver una **tarea dada**.



Machine Learning

Este **modelo** se entrena usando **datos**.
Mediante este proceso, aprende a
encontrar **características escondidas**
en los datos que le permiten hacer
predicciones.



Eligiendo algoritmo

Tarea

Definir de forma clara el objetivo

1

Información

Con qué datos se cuenta para lograr el objetivo

2

APRENDIZAJE
SUPERVISADO

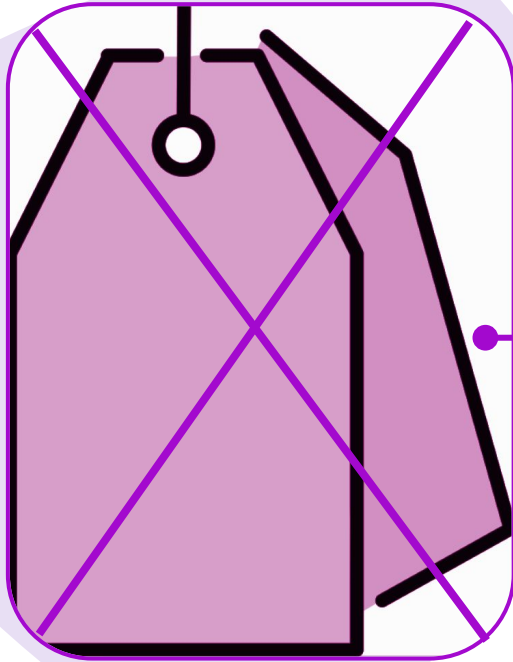
APRENDIZAJE
**NO
SUPERVISADO**





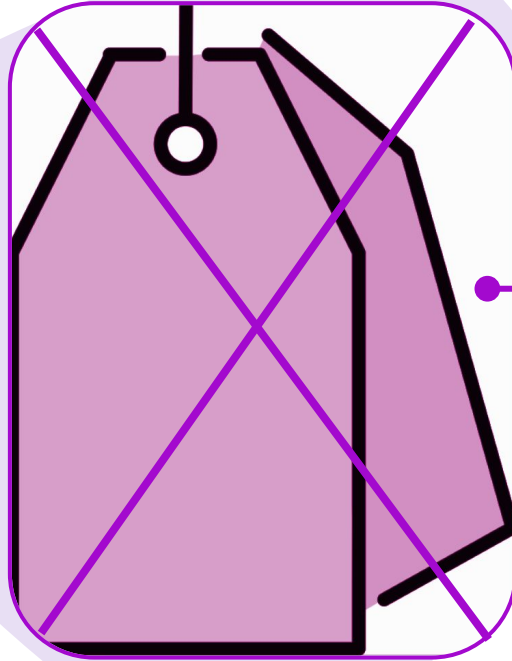
Aprendizaje NO supervisado

Aprendizaje No Supervisado



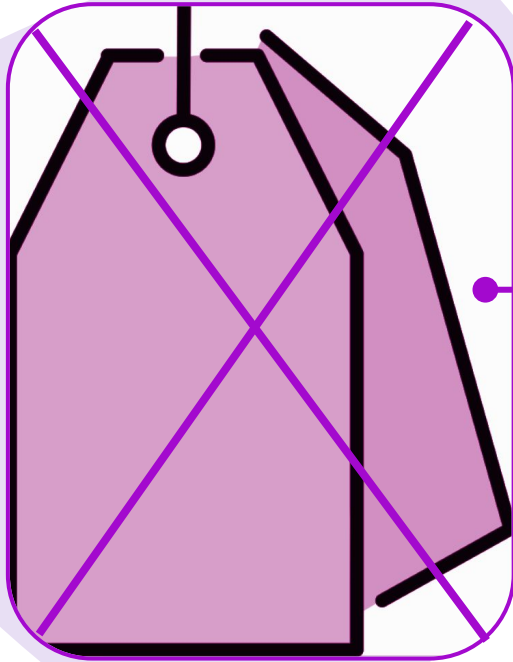
Se llama **Aprendizaje No Supervisado** a la colección de métodos o técnicas que trabaja con datos que no tienen asociados una etiqueta (una clase o un valor), o sea, que **NO** están **etiquetados**.

Aprendizaje No Supervisado



A diferencia del Aprendizaje Supervisado, el objetivo ya no es predecir una etiqueta, sino **encontrar patrones** en el conjunto de datos.

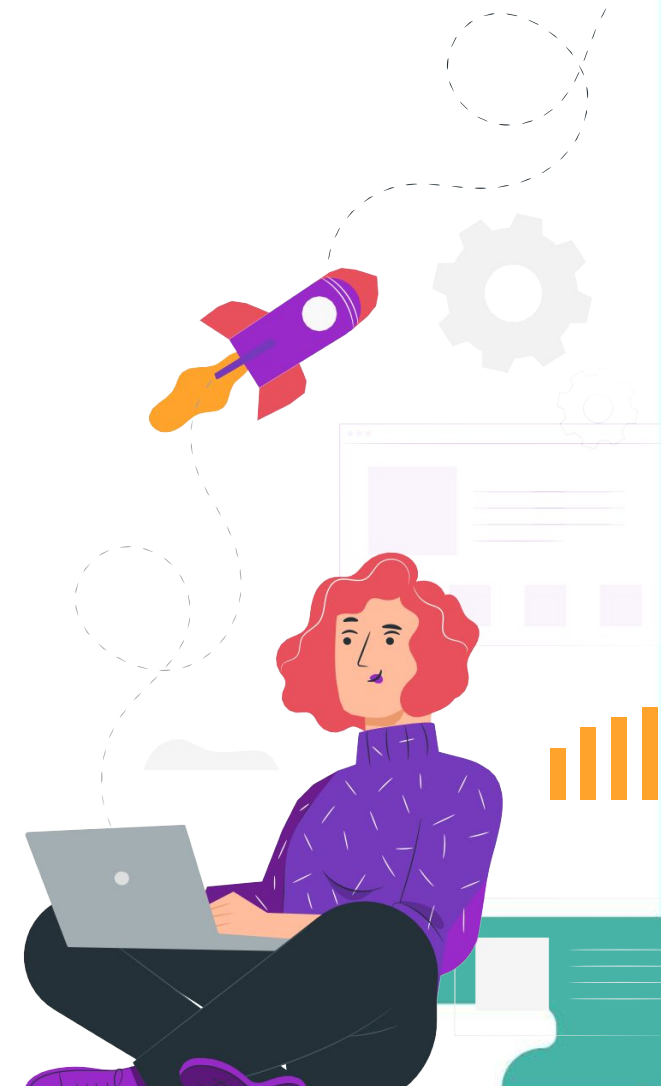
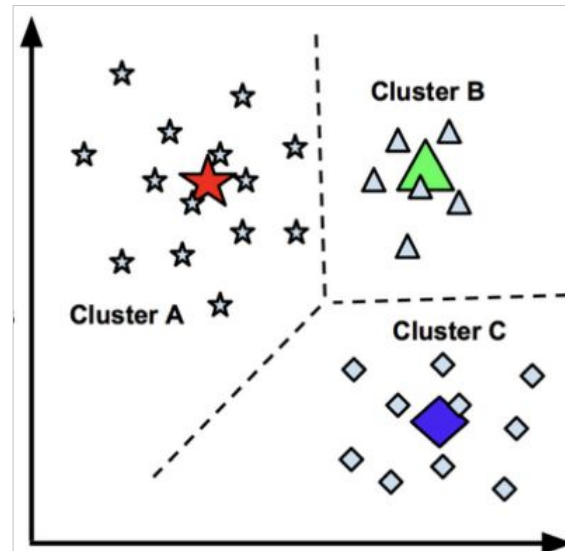
Aprendizaje No Supervisado



- **Clustering**
- **Reducción de la Dimensionalidad**

Clustering

El **objetivo** del clustering o agrupamiento es encontrar grupos (**clusters**) en los cuales las instancias pertenecientes sean parecidas.



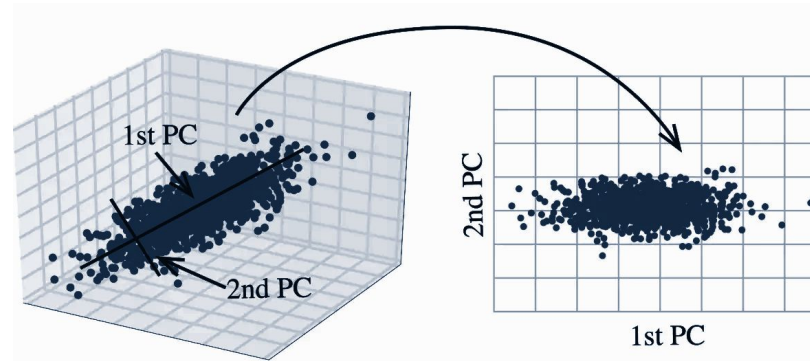
Clustering

El objetivo del clustering o agrupamiento es encontrar grupos (**clusters**) en los cuales las instancias pertenecientes sean parecidas.

- **Aplicaciones**
 - Investigación de mercado
 - Sistemas de recomendación
 - Medicina y Biología
- **Algoritmos**
 - K-means
 - DBSCAN
 - Hierarchical Clustering
 - Fuzzy C-Means
 - Gaussian Mixture Models

Reducción de la dimensionalidad

Consiste en **reducir** la cantidad de **features** de un dataset, pero **reteniendo** la mayor cantidad de **información** posible.



Reducción de la dimensionalidad

Consiste en **reducir** la cantidad de **features** de un dataset, pero **reteniendo** la mayor cantidad de **información** posible.



- **Aplicaciones**

- Reducir la complejidad del input en un modelo de regresión o clasificación
- Visualización
- Detectar features relevantes en datasets

- **Algoritmos**

- Principal Component Analysis
- Multidimensional Scaling
- t-SNE: t-distributed Stochastic Neighbor Embedding
- LDA: Linear Discriminant Analysis



Descanso

Nos vemos en 10 minutos



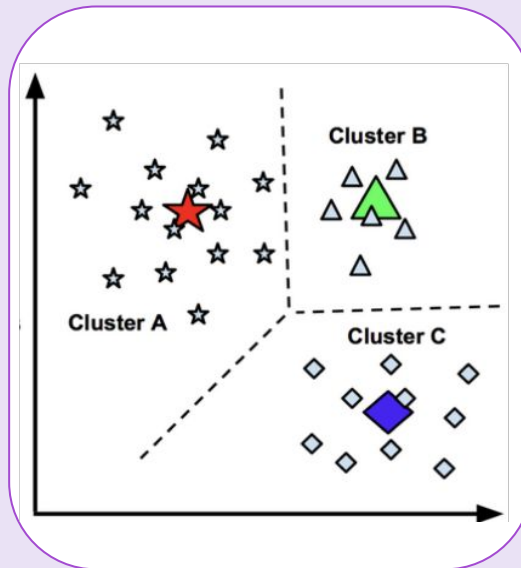
KMEANS

K-Means

K-Means tiene como objetivo separar los datos en un número **k** dado de clusters, ubicando a las instancias que estén dentro de una región cercana dentro de un mismo cluster.

K-Means

Centroide



Encuentra un número **k** de centroides, uno por cada cluster, tal que la distancia entre los centroides y los datos más cercanos sea la **mínima posible**.

A continuación, cada instancia se identifica en el grupo del centroide más cercano

K-Means

Se utiliza un algoritmo **iterativo** hasta llegar al resultado.

- A** Se inicializan los k Centroides (de manera aleatoria o siguiendo algún criterio)
- B** Se encuentra el centroide más cercano de cada instancia
- C** Se Actualizan los centroides (Nuevo centroide: Promedio de las posiciones de las instancias en un mismo cluster)
- D** Se repiten los pasos 2 y 3 hasta que la posición del centroide ya no varíe.



Repasamos en Kahoot

Tercera pre-entrega

Hoy se realiza la tercera pre-entrega.
La misma incluirá los desafíos vistos:

- Trabajar sobre el modelo de clasificación: desde el análisis exploratorio hasta la evaluación y selección
- Para ello, les pediremos ajustar cualquier modelo supervisado, teniendo en cuenta lo visto en las clases:
 - ajustar los modelos de regresión y clasificación vistos
 - evaluar los modelos y evitar el overfitting optimizando los hiperparámetros

Presentarán lo trabajado entregando el link a su Github en el foro del aula virtual





¿DUDAS?

FUNDACIÓN
YPF

¡Muchas gracias!

