# Machine Learning No Supervisado

Este documento detalla el contenido, la estructura visual y las metáforas clave para explicar los conceptos fundamentales del aprendizaje no supervisado.

## 1. Clustering (Agrupación)

Título: CLUSTERING: Agrupando lo Similar

Metáfora Visual: Clasificación de canicas de diferentes colores en cuencos separados.

| **Sección** | **Contenido (Español)** | **Diseño Clave** |
| --- | --- | --- |
| **Definición** | Algoritmos que unen puntos de datos comparables basándose en sus características inherentes o cercanía. | Diagrama de dispersión 2D/3D con puntos de datos de diferentes colores, claramente separados en 3-4 grupos (clusters). |
| **Objetivo** | Descubrir estructuras y patrones ocultos en datos sin etiquetar. | Flecha que apunta de "Datos sin Estructura" a "Datos Estructurados (Agrupados)". |
| **Algoritmos Comunes** | - K-means: Basado en la distancia al centroide.  - Jerárquico: Construye una jerarquía de clusters (dendrograma).  - DBSCAN: Agrupa puntos que están densamente conectados (útil para formas complejas).  - GMMs: Modelos de Mezcla Gaussiana. | Iconos representativos para cada algoritmo. |

## 2. Dimensionality Reduction (Reducción de Dimensionalidad)

Título: REDUCCIÓN DE DIMENSIONALIDAD: De lo Complejo a lo Esencial

Metáfora Visual: Comprimir una imagen 3D a una representación 2D, conservando la información crítica.

| **Sección** | **Contenido (Español)** | **Diseño Clave** |
| --- | --- | --- |
| **Definición** | Técnicas para reducir el número de variables o características, manteniendo la información más importante del conjunto de datos. | Imagen de dos cajas: "Alta Dimensionalidad (e.g., 100 variables)" "Baja Dimensionalidad (e.g., 2 variables)". |
| **Objetivo** | Visualizar datos complejos y reducir el ruido para mejorar la eficiencia del modelo. | Icono de lupa sobre datos, con texto: "Mejora la interpretabilidad y reduce el costo computacional." |
| **Algoritmos Comunes** | - PCA (Análisis de Componentes Principales): Encuentra las direcciones de máxima varianza.  - t-SNE: Visualización no lineal (ideal para mapas de características).  - Autoencoders: Redes neuronales que aprenden una codificación eficiente. | Visualización de un eje de datos que pasa de un espacio 3D a un plano 2D. |

## 3. Anomaly Detection (Detección de Anomalías)

Título: DETECCIÓN DE ANOMALÍAS: Encontrando al "Punto Raro"

Metáfora Visual: Un campo de esferas grises con una sola esfera roja brillante y solitaria (el outlier).

| **Sección** | **Contenido (Español)** | **Diseño Clave** |
| --- | --- | --- |
| **Definición** | Algoritmos para identificar instancias que se desvían drásticamente de la norma o son poco comunes en un conjunto de datos. | Un gran grupo denso de puntos normales con un círculo o punto de datos alejado y marcado con una "X". |
| **Aplicaciones** | Detección de fraude, monitoreo de fallos en equipos, ciberseguridad. | Iconos de una tarjeta de crédito (fraude) y un servidor (fallo). |
| **Algoritmos Comunes** | - Z-score / Distancia de Mahalanobis: Métodos estadísticos.  - Isolation Forest: Aísla anomalías de forma eficiente.  - One-Class SVM: Define un límite alrededor de los datos normales. | Diagrama de un límite o frontera dibujado alrededor del grupo de datos normales, dejando a la anomalía fuera. |

## 4. Association Rule Learning (Aprendizaje de Reglas de Asociación)

Título: REGLAS DE ASOCIACIÓN: Descubriendo Relaciones Ocultas

Metáfora Visual: Análisis de la cesta de la compra (Market Basket Analysis).

| **Sección** | **Contenido (Español)** | **Diseño Clave** |
| --- | --- | --- |
| **Definición** | Algoritmos que buscan correlaciones o conexiones interesantes entre diferentes elementos de un conjunto de datos. | Un diagrama de flujo que muestra la relación: **SI** (compra Mantequilla y Pan) **ENTONCES** (es probable que compre Leche). |
| **Métricas Clave** | - Soporte: Frecuencia con la que aparecen los ítems.  - Confianza: Probabilidad de que B sea comprado dado que A fue comprado (A → B).  - Lift (Elevación): Cuánto más probable es que A y B se compren juntos. | Iconos de porcentajes y flechas para indicar las métricas. |
| **Algoritmos Comunes** | - Apriori: Encuentra conjuntos de ítems frecuentes.  - FP-growth: Una alternativa más rápida que usa un árbol de patrones. | Diagrama con un carrito de compras y flechas conectando ítems. |

## 5. Generative Models (Modelos Generativos)

Título: MODELOS GENERATIVOS: Creando Datos Nuevos y Realistas

Metáfora Visual: Un artista que aprende el estilo de la pintura (la distribución de probabilidad) y crea una obra completamente nueva.

| **Sección** | **Contenido (Español)** | **Diseño Clave** |
| --- | --- | --- |
| **Definición** | Modelos que aprenden la distribución de probabilidad subyacente de los datos para crear nuevas muestras similares a los datos de entrenamiento. | Un proceso de tres pasos: **1. Datos Reales** **2. Modelo Generativo** **3. Datos Sintéticos (Nuevos)**. |
| **Aplicaciones** | Generación de imágenes, creación de texto, aumento de datos (Data Augmentation). | Imágenes de ejemplo: un rostro generado por IA (nuevo) junto a rostros reales. |
| **Algoritmos Comunes** | - GANs (Redes Adversarias Generativas): Compuestas por un Generador y un Discriminador que compiten.  - VAEs (Autoencoders Variacionales): Aprenden un espacio latente para muestrear nuevos datos. | Diagrama simple de un GAN (dos cajas en conflicto). |

## 6. Density Estimation (Estimación de Densidad)

Título: ESTIMACIÓN DE DENSIDAD: Mapeando la Distribución de Probabilidad

Metáfora Visual: Un mapa topográfico de una cordillera, donde los picos indican mayor densidad de datos.

| **Sección** | **Contenido (Español)** | **Diseño Clave** |
| --- | --- | --- |
| **Definición** | Algoritmos que calculan la función de densidad de probabilidad (PDF) subyacente de los datos para comprender su distribución y organización. | Un gráfico de dispersión 2D con un mapa de contorno (isobaras de color) superpuesto, mostrando dónde están los "picos" de datos. |
| **Objetivo** | Entender la estructura interna del conjunto de datos e identificar regiones de alta o baja probabilidad. | Lupa sobre las regiones más densas del gráfico. |
| **Algoritmos Comunes** | - KDE (Estimación de Densidad del Kernel): Utiliza funciones de núcleo para estimar la PDF (genera una superficie suave).  - GMMs (Modelos de Mezcla Gaussiana): Modela la distribución como una combinación de múltiples distribuciones Gaussianas (campanas). | Iconos de un "suavizador" (KDE) y dos o tres curvas de campana superpuestas (GMMs). |