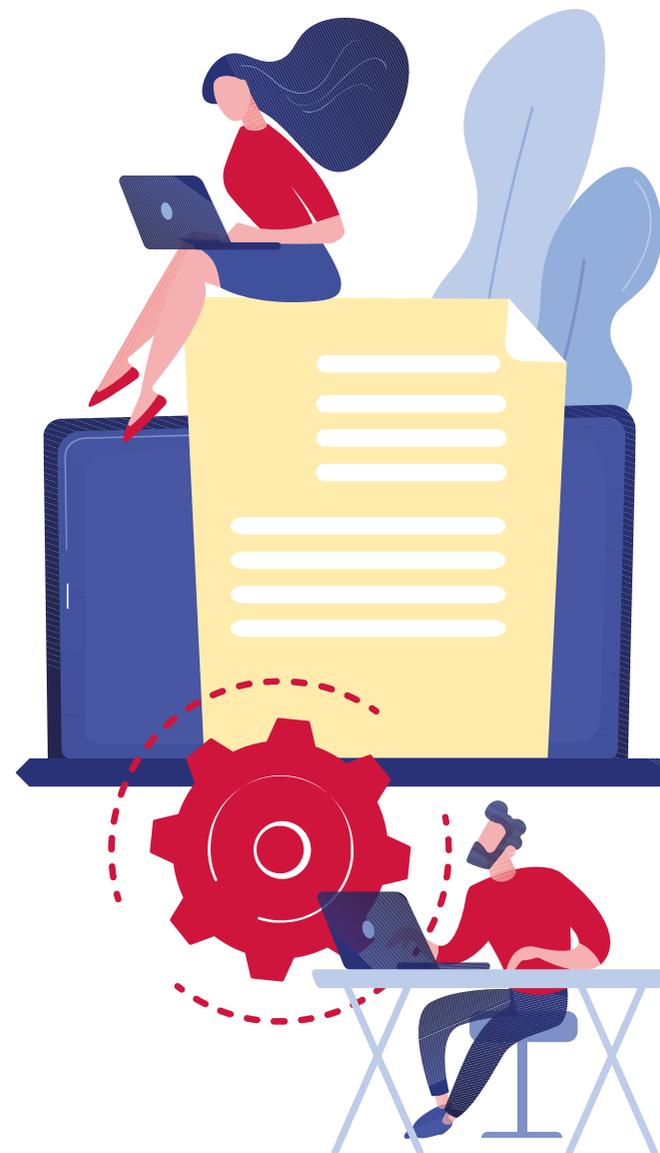


OFICINA
Acelera
pyme

Aplicación de la Ciencia de Datos a la Industria 4.0

1 diciembre 2022



VICEPRESIDENCIA
SEGUNDA DEL GOBIERNO
MINISTERIO
DE ASUNTOS ECONÓMICOS
Y TRANSFORMACIÓN DIGITAL

SECRETARÍA DE ESTADO
DE DIGITALIZACIÓN
E INTELIGENCIA ARTIFICIAL

red.es



UNIÓN EUROPEA

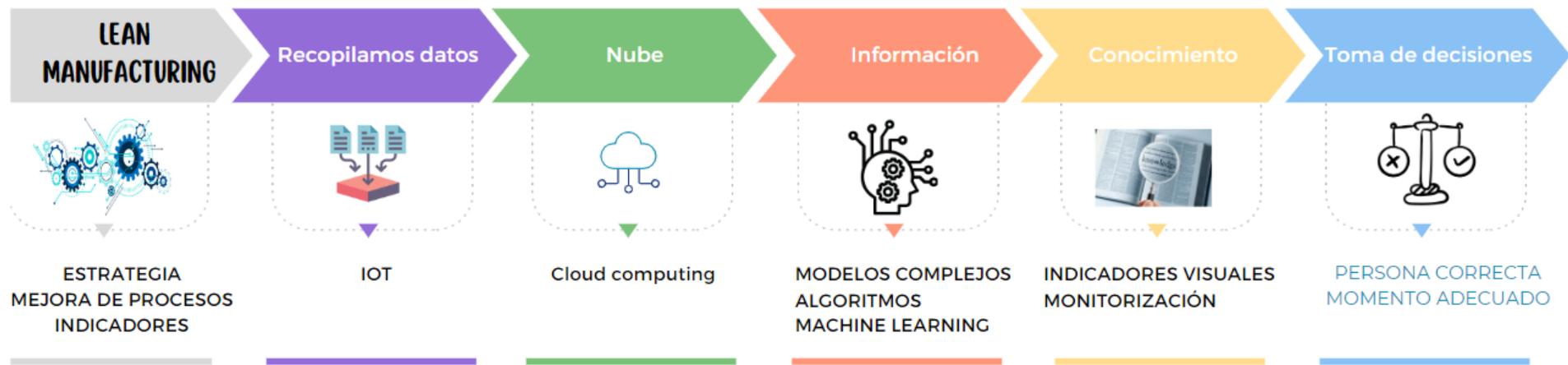
Fondo Europeo de Desarrollo Regional
“Una manera de hacer Europa”



Rosa Martínez Álvarez - Castellanos

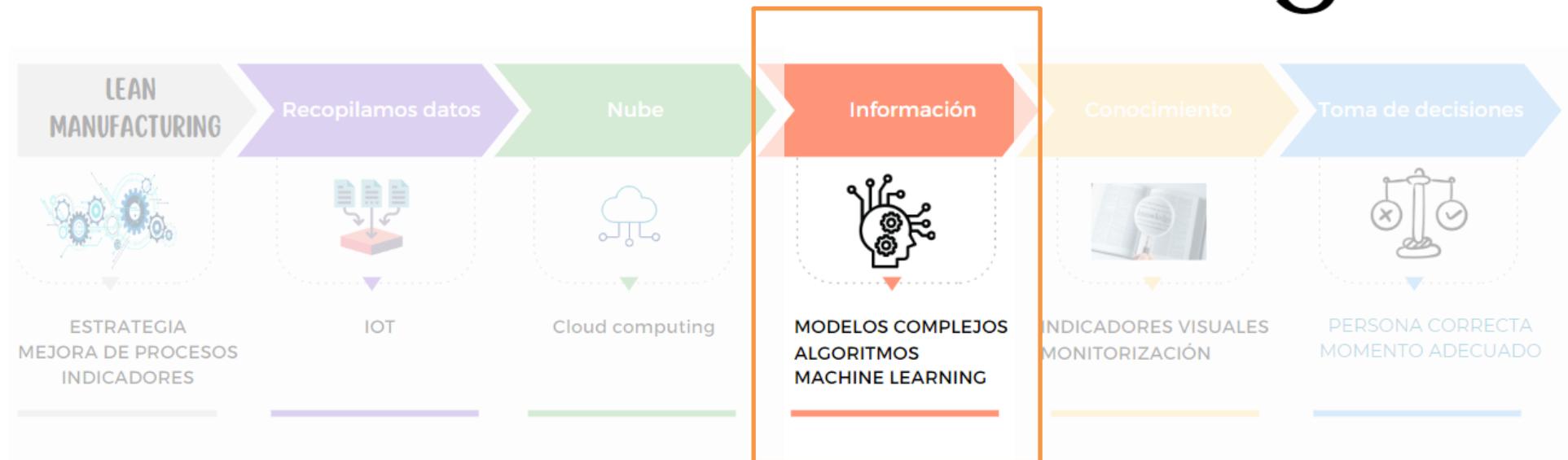
- Máster en Ingeniería de Telecomunicación y graduada en Ing. De Sistemas de Telecomunicación
- Responsable de Ciencia de Datos en CTN.
- Dirige el desarrollo técnico de proyectos de ciencia de datos aplicados a diferentes sectores (medio ambiente, naval, acuicultura, energía, etc.)
- Participación en diversos proyectos europeos y nacionales: DigiSafeCage, Demo-Bluesmartfeed, SmartLEAN, NeXOS, entre otros.

METODOLOGIA PARA LA transformación digital

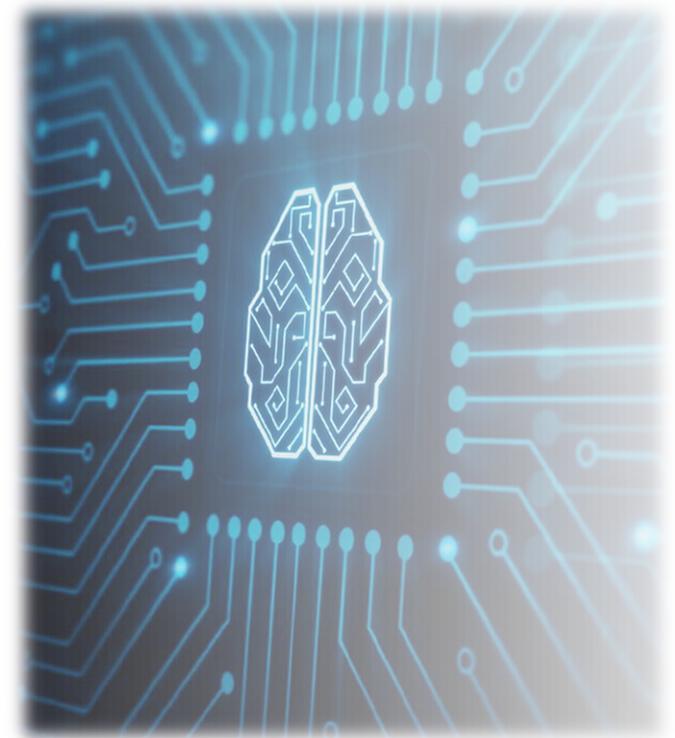


METODOLOGIA PARA LA

transformación digital



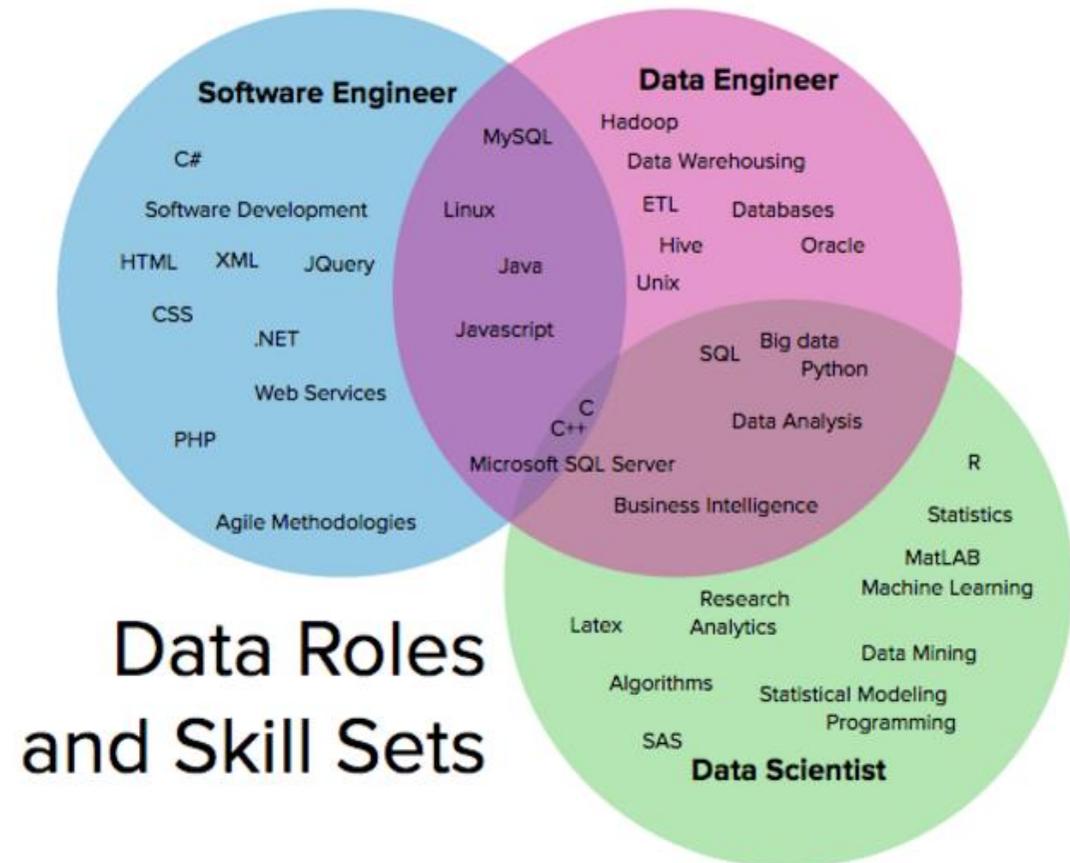
- 1. Contexto y evolución de la Ciencia de datos**
- 2. Metodologías para descubrir conocimiento**
- 3. Conceptos de ciencia de datos: métodos, técnicas y métricas**
- 4. Aplicaciones de la Ciencia de datos a la Industria 4.0**



Introducción

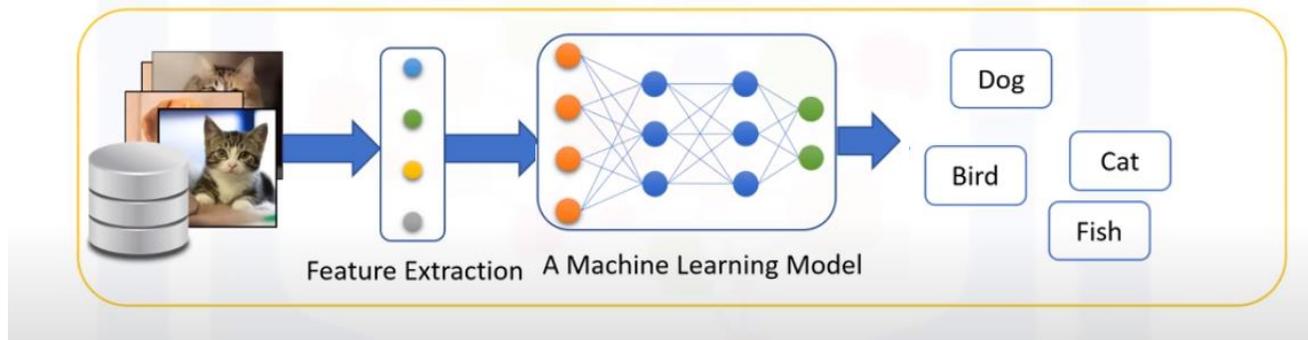
Contexto y evolución

La **Ciencia de Datos** aparece como una “aptitud” en la elaboración de sistemas para la toma de decisiones en distintos ámbitos de negocio



Machine Learning

“Machine Learning es un campo de estudio que dota a los ordenadores con la capacidad de aprender sin estar programado explícitamente”
(Arthur Samuel, 1959)



Machine Learning

“Machine Learning es un campo de estudio que dota a los ordenadores con la capacidad de aprender sin estar programado explícitamente”
(Arthur Samuel, 1959)

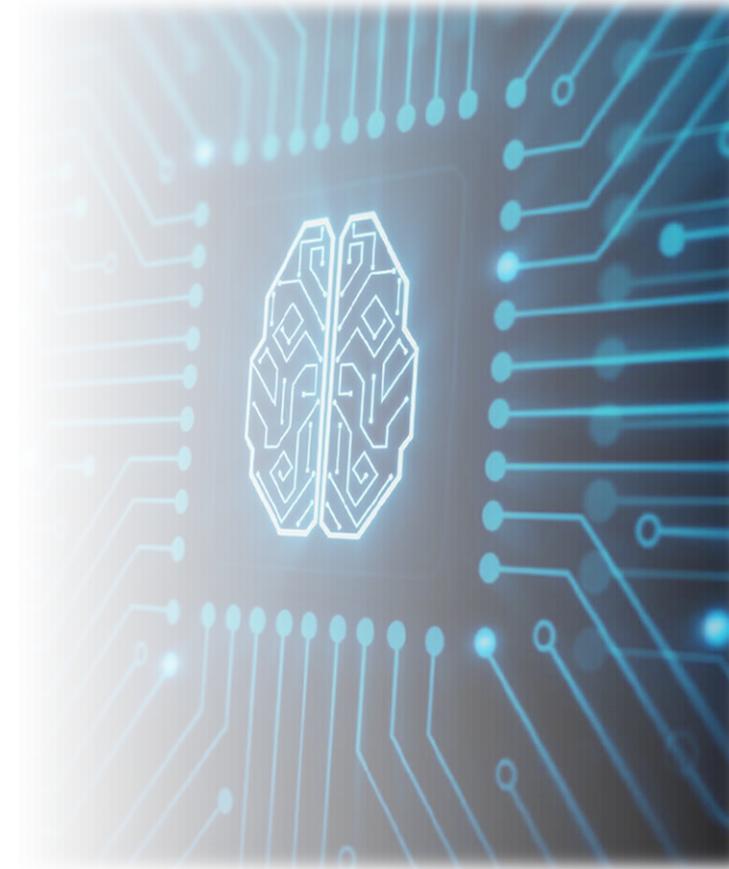


Metodologías para descubrir conocimiento

Descubrir conocimiento

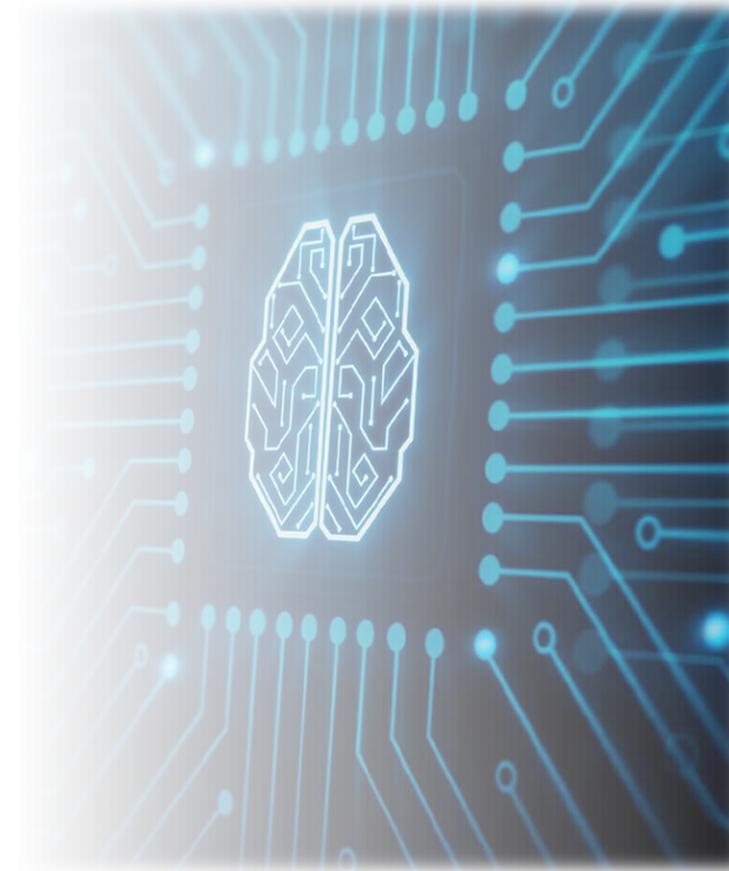
PARADIGMA

- Mis datos son valiosos para mi (in → in).
 - Datos internos útiles para la organización.
 - Inteligencia empresarial clásica... Muchas oportunidades todavía.
- Esos datos son valiosos para mi (out → in).
 - Datos externos útiles para la organización.
 - Medios sociales, Internet, datos abiertos, ... Muchas oportunidades nuevas.
- Mis datos son valiosos para otros (in → out).
 - Datos internos útiles para otras organizaciones.
 - Mis datos tienen utilidad para otros, ... Muchas oportunidades nuevas.
- Esos datos son valiosos para otros (out → out).
 - Datos externos útiles para otras organizaciones.
 - Estos datos tienen utilidad para otros, ... ¡Científico de datos freelancer!
- Creando datos (Φ → out).
 - Coleccionar datos que pueden tener valor. ¡Emprendedor de datos!



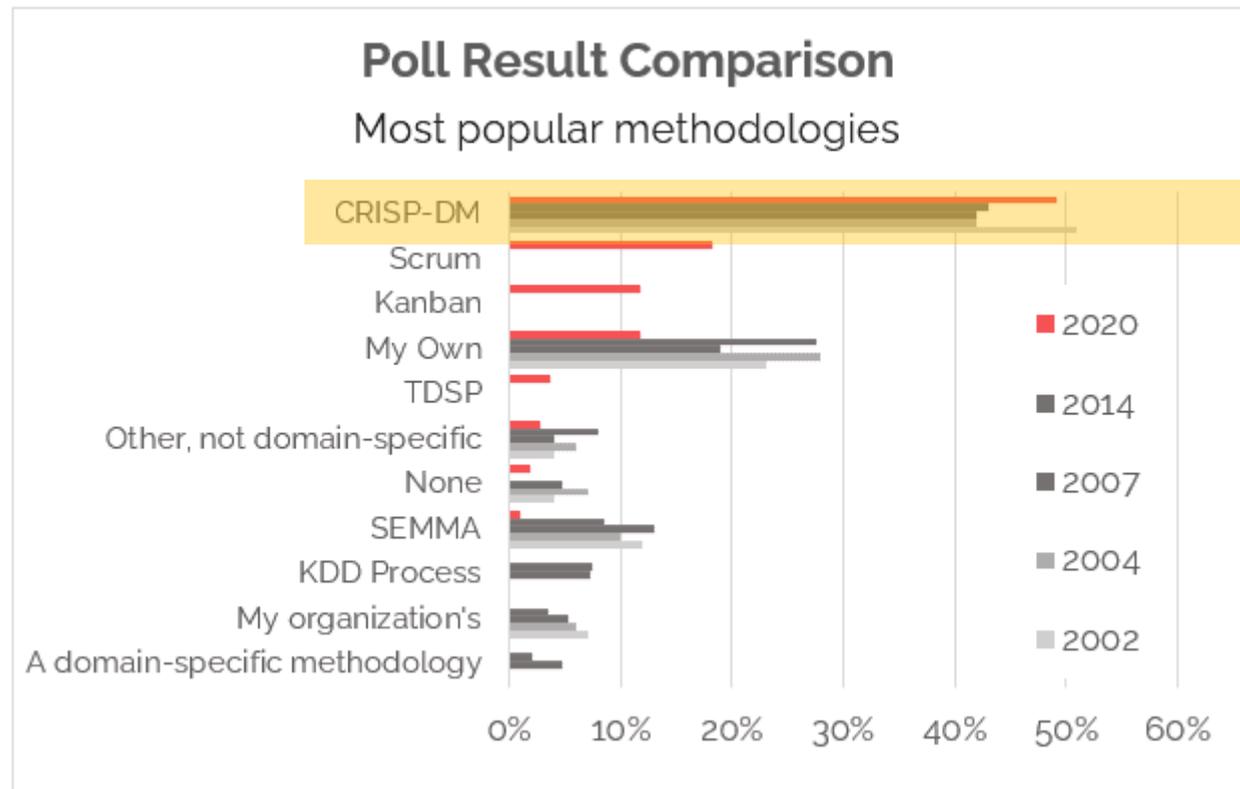
Descubrir conocimiento

Flujo general



Descubrir conocimiento

Metodologías

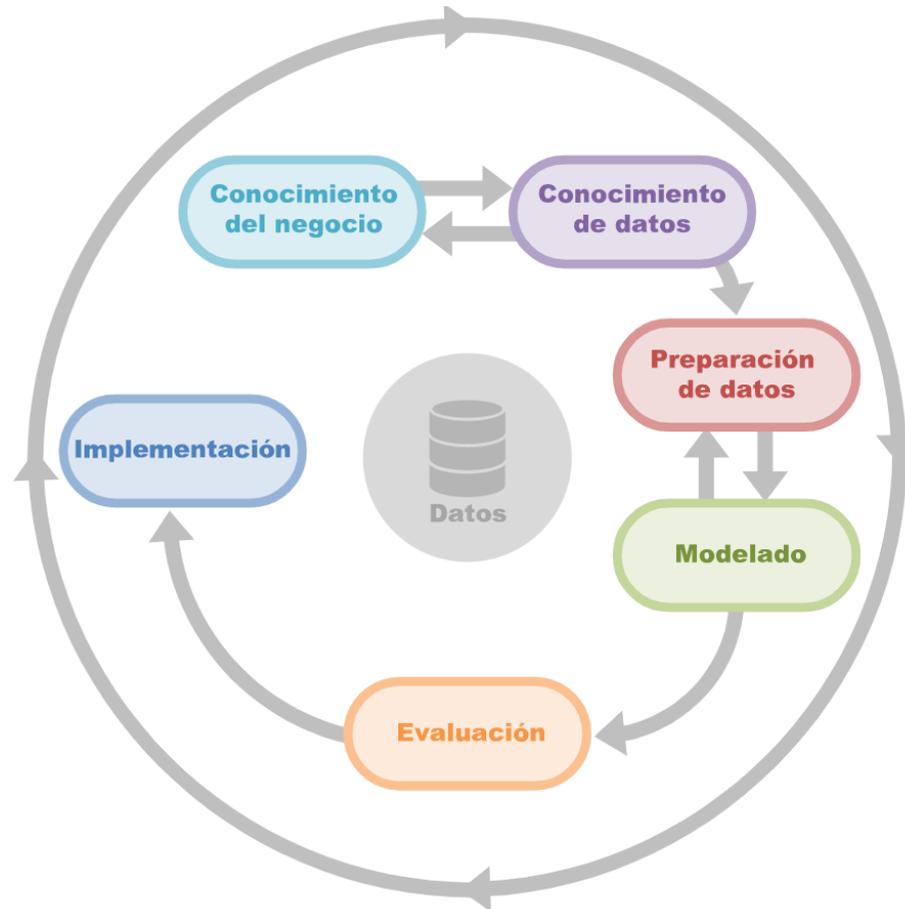


Fuente: [CRISP-DM is Still the Most Popular Framework](#)



Descubrir conocimiento

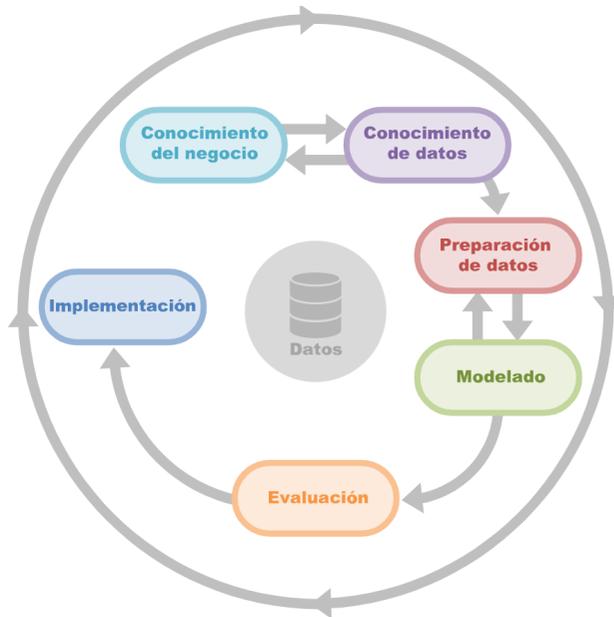
CRISP DM



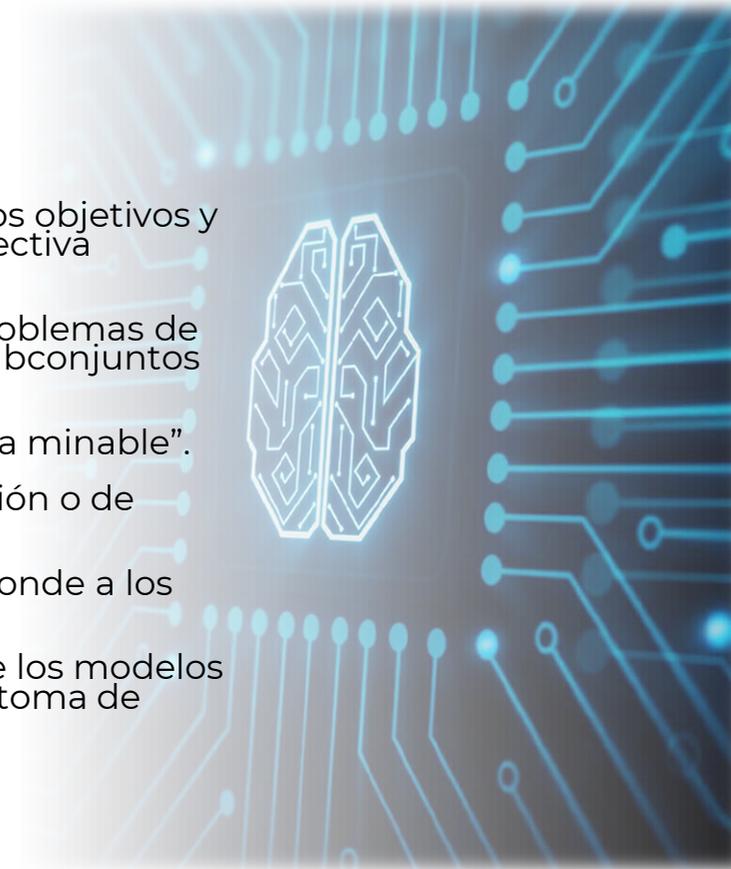
Wirth, R., & Hipp, J. (2000, April). CRISP-DM: Towards a standard process model for data mining. In Proceedings of the 4th international conference on the practical applications of knowledge discovery and data mining (Vol. 1, pp. 29-40).

Descubrir conocimiento

CRISP DM



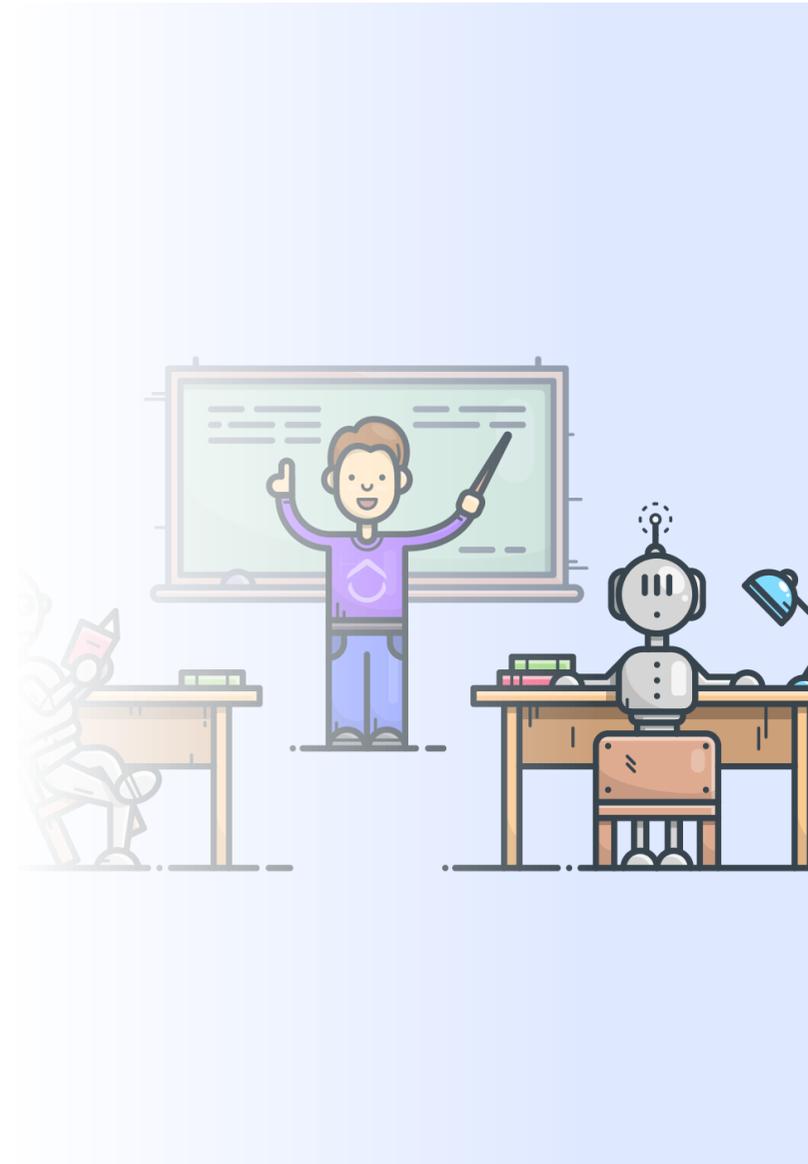
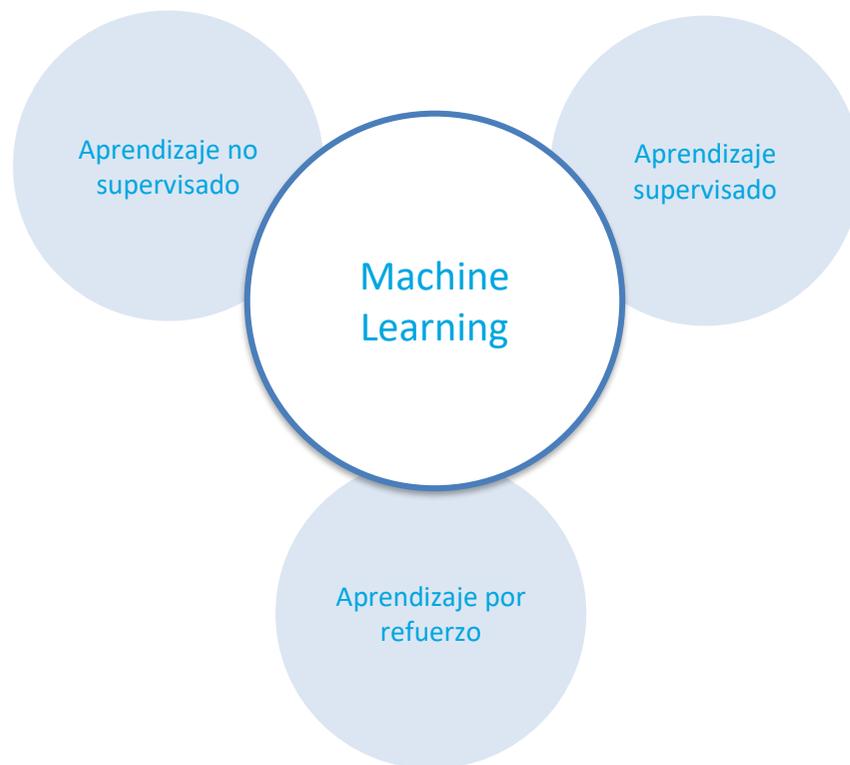
- 1) **Comprensión de negocio:** comprender los objetivos y requisitos del proyecto desde una perspectiva empresarial
- 2) **Conocimiento de datos:** identificar los problemas de calidad y discernir datos potenciales o subconjuntos que pueden ser interesantes de analizar
- 3) **Preparación de los datos:** obtener la “vista minable”.
- 4) **Modelado:** aplicar técnicas de modelización o de minería de datos
- 5) **Evaluación:** comprobar si el modelo responde a los requisitos establecidos
- 6) **Implementación:** explotar el potencial de los modelos extraídos, integrarlos en los procesos de toma de decisiones.



Conceptos de ciencia de datos

Métodos y técnicas

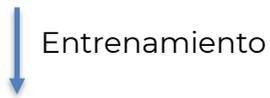
Métodos



Métodos

APRENDIZAJE SUPERVISADO

Guiado por la inteligencia humana, la observación y unos resultados conocidos

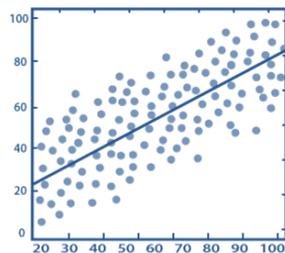


Dotamos al modelo de conocimiento para predecir valores futuros



Clasificación

Regresión



APRENDIZAJE NO SUPERVISADO

No hay guía ni instrucciones sobre la interpretación de los datos no conocidos.

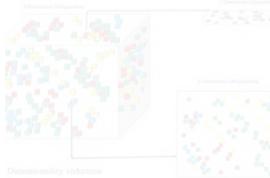


El modelo trabaja por sí solo para descubrir información sobre los datos



Reducción de la dimensionalidad

Clustering (agrupamiento)

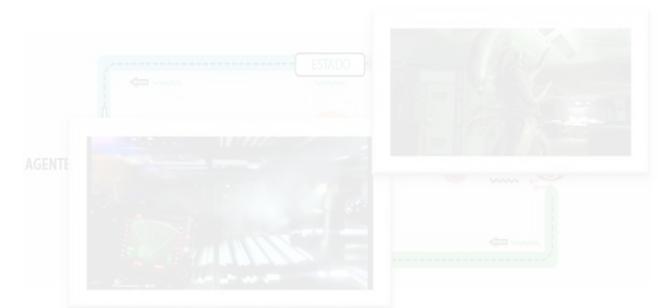


APRENDIZAJE POR REFUERZO

Enseña a un agente cómo elegir una acción para maximizar recompensas



El modelo trabaja por sí solo para descubrir información sobre los datos



Métodos

APRENDIZAJE SUPERVISADO

Guiado por la inteligencia humana, la observación y unos resultados conocidos

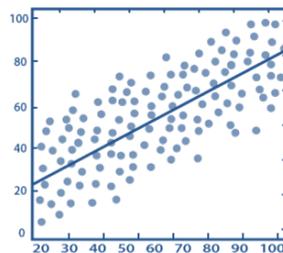
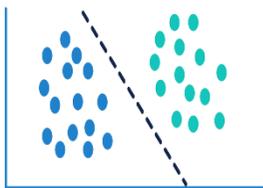


Dotamos al modelo de conocimiento para predecir valores futuros



Clasificación

Regresión



APRENDIZAJE NO SUPERVISADO

No hay guía ni instrucciones sobre la interpretación de los datos no conocidos.

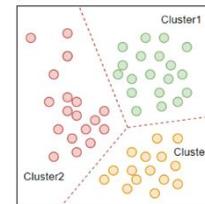
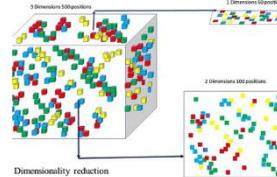


El modelo trabaja por sí solo para descubrir información sobre los datos



Reducción de la dimensionalidad

Clustering (agrupamiento)



APRENDIZAJE POR REFUERZO

Enseña a un agente cómo elegir una acción para maximizar recompensas



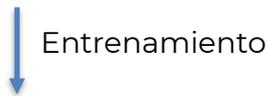
El modelo trabaja por sí solo para descubrir información sobre los datos



Métodos

APRENDIZAJE SUPERVISADO

Guiado por la inteligencia humana, la observación y unos resultados conocidos



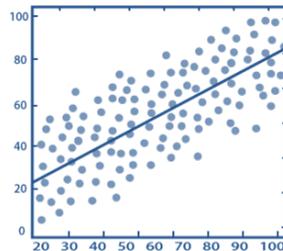
Dotamos al modelo de conocimiento para predecir valores futuros



Clasificación



Regresión



APRENDIZAJE NO SUPERVISADO

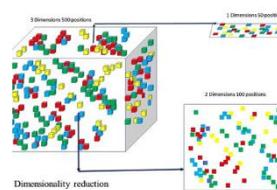
No hay guía ni instrucciones sobre la interpretación de los datos no conocidos.



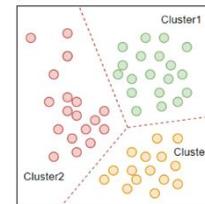
El modelo trabaja por sí solo para descubrir información sobre los datos



Reducción de la dimensionalidad



Clustering (agrupamiento)

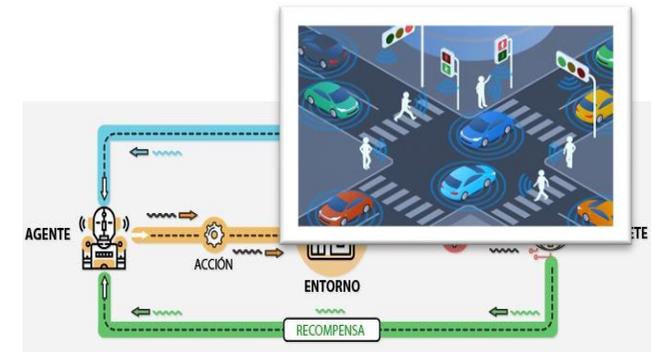


APRENDIZAJE POR REFUERZO

Enseña a un agente cómo elegir una acción para maximizar recompensas



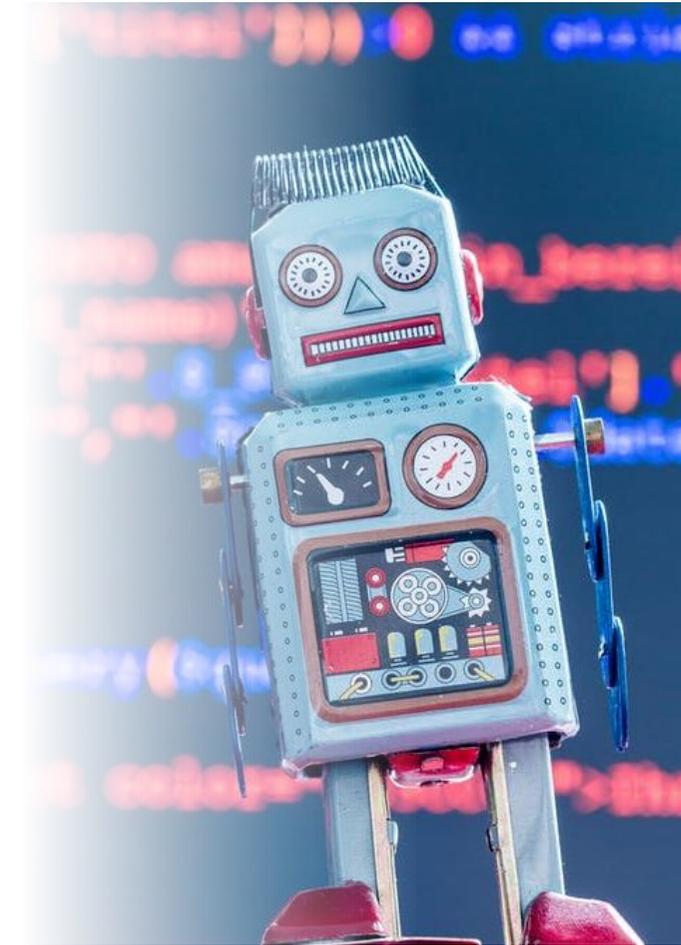
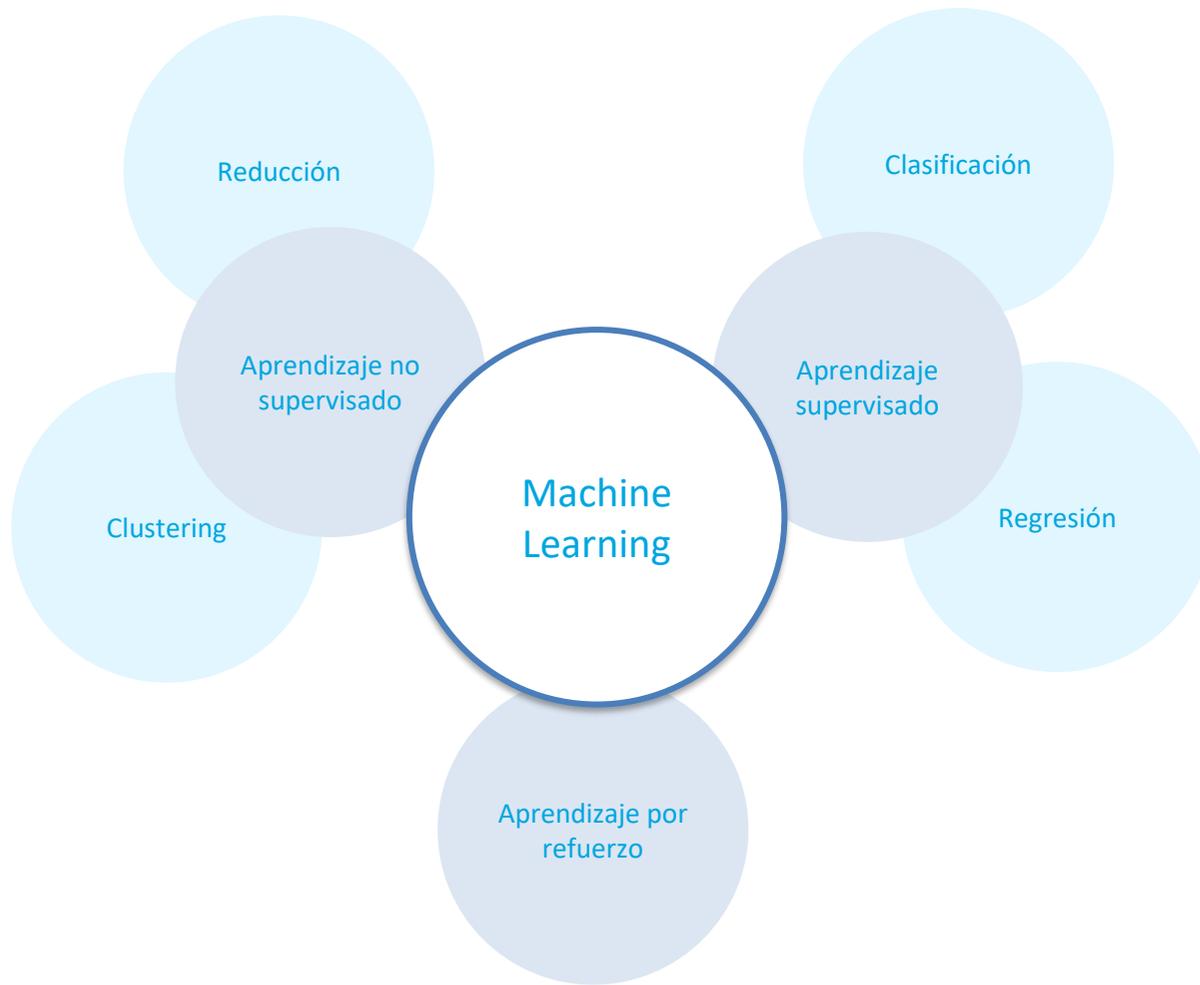
El modelo trabaja por sí solo para descubrir información sobre los datos



Métodos

Aprendizaje supervisado	Aprendizaje no supervisado	Aprendizaje por refuerzo
<ul style="list-style-type: none">• Datos conocidos• Predicción valores futuros• Entorno controlado• Diferentes métodos de evaluación	<ul style="list-style-type: none">• Datos no conocidos• Descubrimiento de información• Entorno no controlado• Pocos métodos de evaluación	<ul style="list-style-type: none">• Proceso de decisión• Sistema de reconocimiento• Aprendizaje de series de acciones

Técnicas



Técnicas: Aprendizaje supervisado

CLASIFICACIÓN

Proceso de predecir categorías o etiquetas de clases discretas o nominales:

Características del conjunto de datos

ID	GRUPO	RADIO	TEXTURA	FORMA	TAMAÑO	SIMETRIA	CLASE
1000025	5	1	5	2	3	1	benigno
1003248	5	4	6	7	6	1	benigno
1003788	3	1	4	2	3	1	maligno
1012365	6	8	8	3	3	2	benigno
1005789	4	1	2	2	2	7	maligno
1002573	8	10	10	2	1	0	maligno
1028830	1	1	2	7	9	1	benigno
1033078	2	1	3	2	2	5	benigno
1033786	2	2	1	2	6	2	maligno
1018578	4	2	2	2	2	3	?

Clases o etiquetas



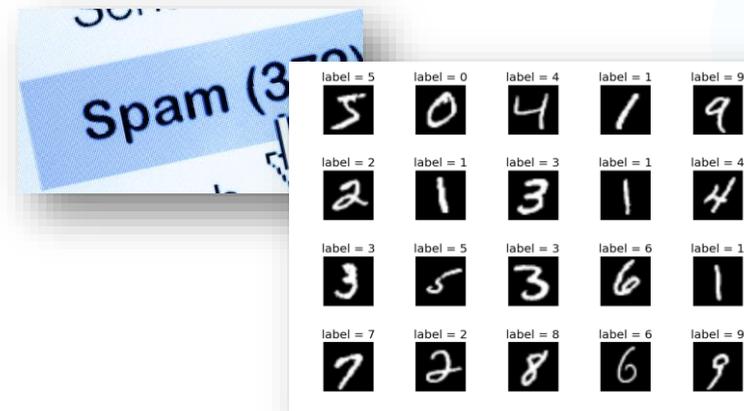
Técnicas: Aprendizaje supervisado

CLASIFICACIÓN

Aplicaciones

Muchos problemas se pueden expresar como asociaciones entre variables de características y variables objetivo. Especialmente cuando se tienen datos etiquetados. Ejemplos:

- Filtrado de emails
- Reconocimiento de voz
- Clasificación de documentos



Reducción

Aprendizaje no supervisado

Clustering

Machine Learning

Aprendizaje supervisado

Regresión

Clasificación

Técnicas: Aprendizaje supervisado

CLASIFICACIÓN

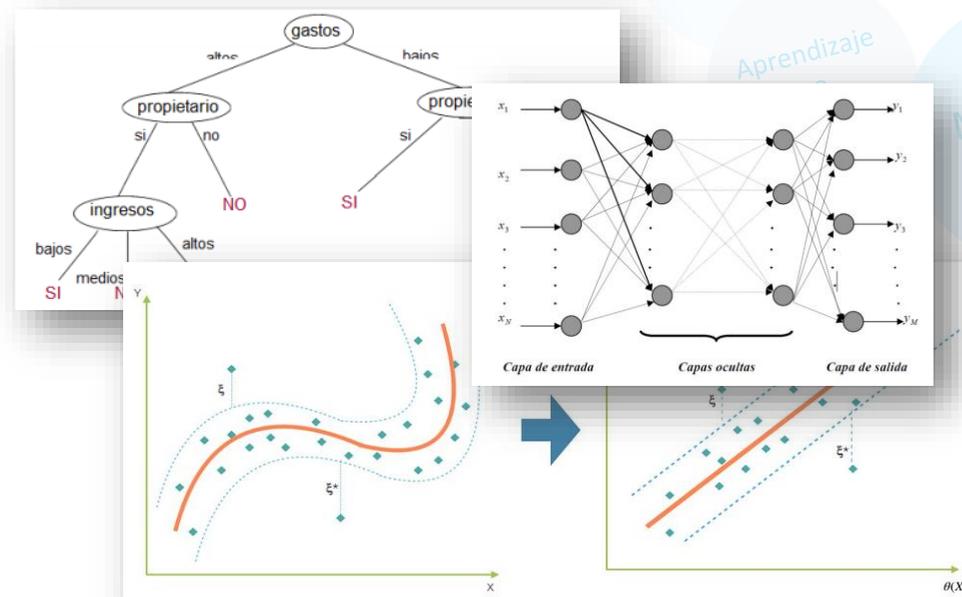
Aplicaciones

Muchos problemas se pueden expresar como asociaciones entre variables de características y variables objetivo. Especialmente cuando se tienen datos etiquetados. Ejemplos:

- Filtrado de emails
- Reconocimiento de voz
- Clasificación de documentos

Algoritmos

- Árboles de decisión
- Máquinas de vector soporte (SVM)
- Clasificador de Naive Bayes
- Redes neuronales
- Redes de función de base radial



Reducción

Aprendizaje

Clasificación

Aprendizaje
supervisado

Regresión

Machine
Learning

Técnicas: Aprendizaje supervisado

REGRESIÓN

Proceso de predicción de valores (atributo objetivo) continuos

ID MOTOR	TAMAÑO	CILINDRADAS	CONSUMO DE COMBUSTIBLE	EMISIONES CO2
0	2	4	8.5	196
1	2.4	4	9.6	221
2	2.5	4	5.9	136
3	3.5	6	11.2	255
4	3.5	6	10.6	244
5	3.5	6	10.0	230
6	3.5	4	10.1	232
7	3.7	6	11.1	255
8	3.7	6	11.4	267
9	2.4	4	9.2	?

Valores continuos

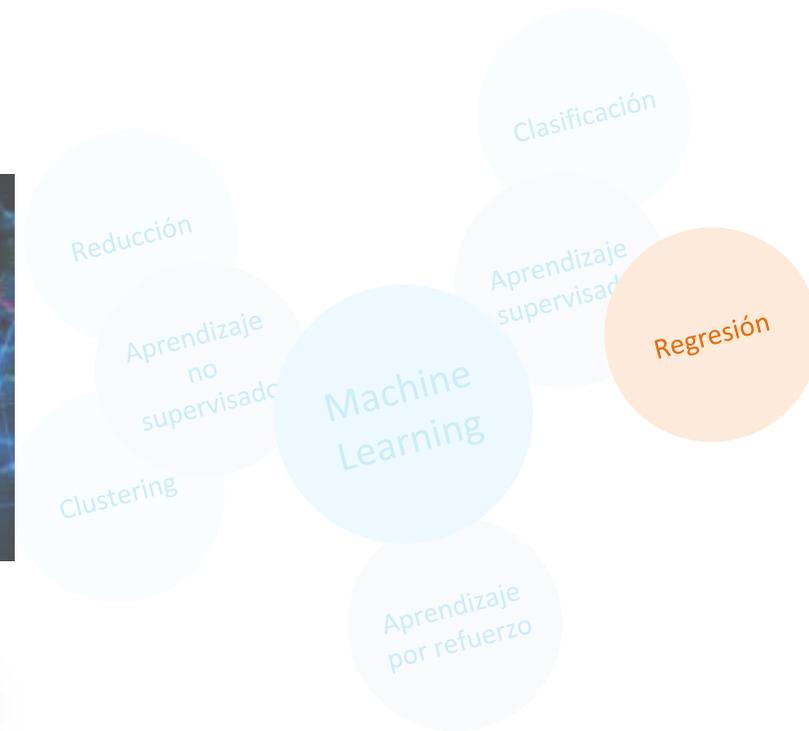
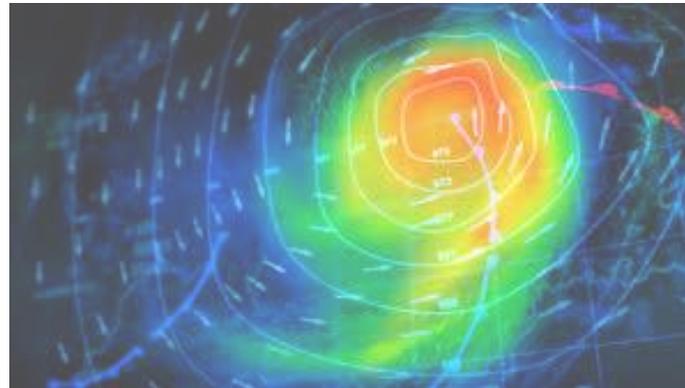


Técnicas: Aprendizaje supervisado

REGRESIÓN

Aplicaciones

- Predicción de ventas
- Análisis de satisfacción
- Ingresos de empleados
- Estimación de precios
- Tiempo meteorológico
- N° de pacientes en urgencias
- Largo etc.



Técnicas: Aprendizaje supervisado

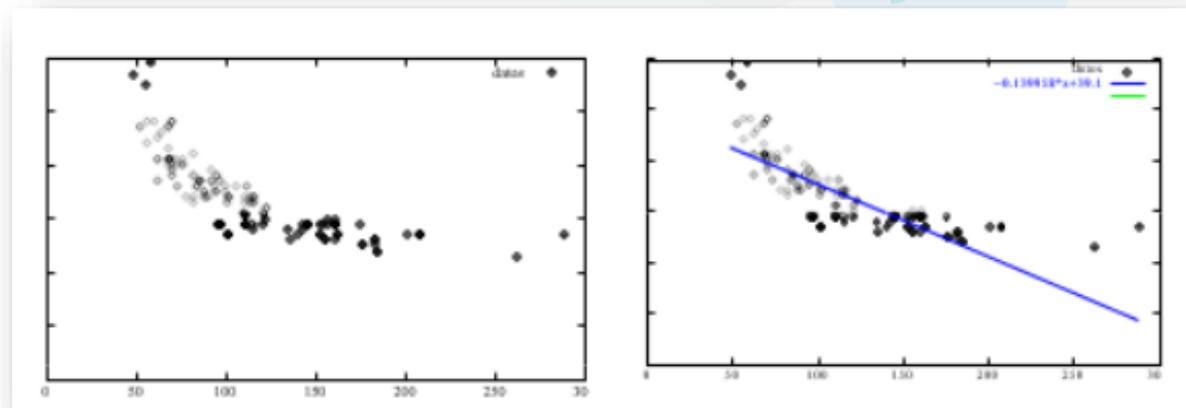
REGRESIÓN

Aplicaciones

- Predicción de ventas
- Análisis de satisfacción
- Ingresos de empleados
- Estimación de precios
- Tiempo meteorológico
- Nº de pacientes en urgencias
- Largo etc.

Algoritmos

- Regresión lineal
- Regresión polinómica
- Regresión lineal bayesiana
- Redes neuronales regresivas
- Bosques de decisión regresivos
- Maquinas de vector soporte regresivas



Reducción

Aprendizaje
no
supervisadoMachine
LearningAprendizaje
supervisado

Clasificación

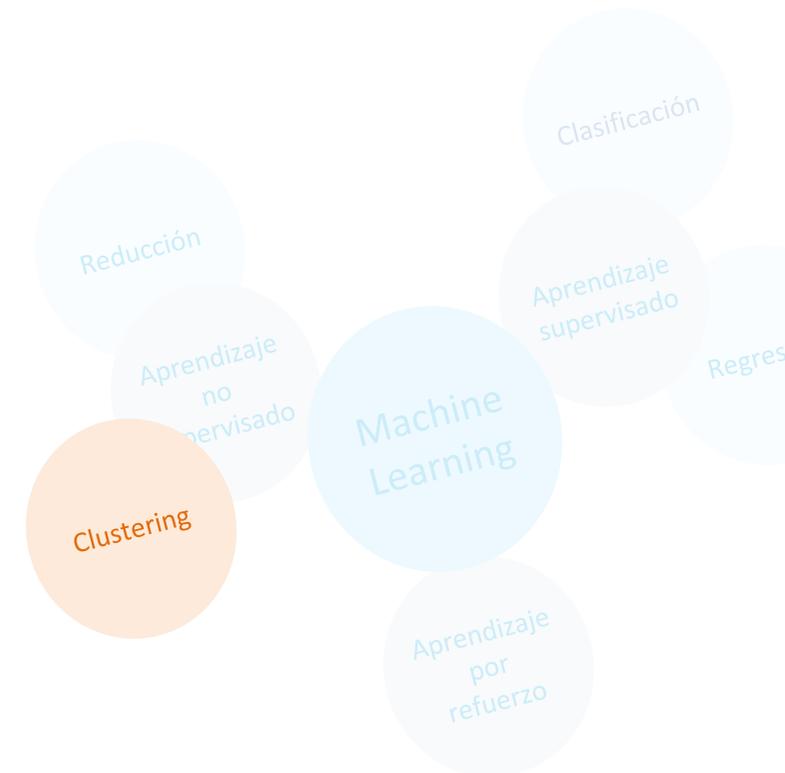
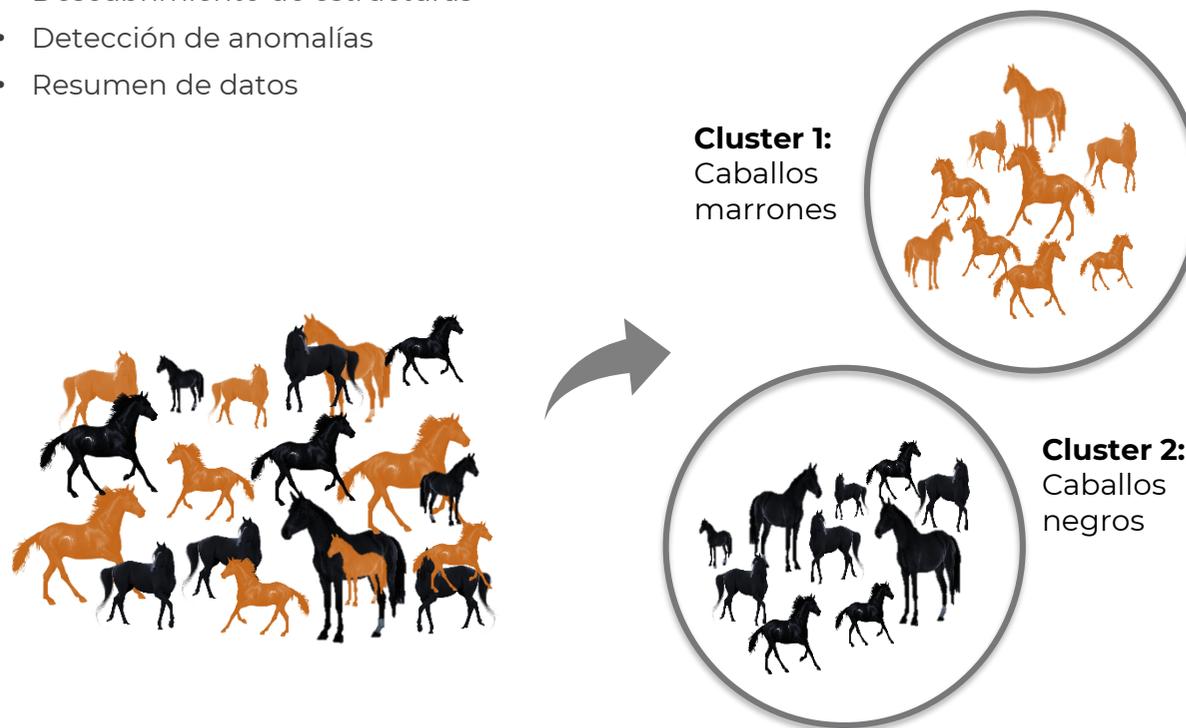
Regresión

Técnicas: Aprendizaje no supervisado

CLUSTERING

Agrupar objetos o conjuntos de datos similares entre sí que sean distintos a los objetos de otros agrupamientos mediante:

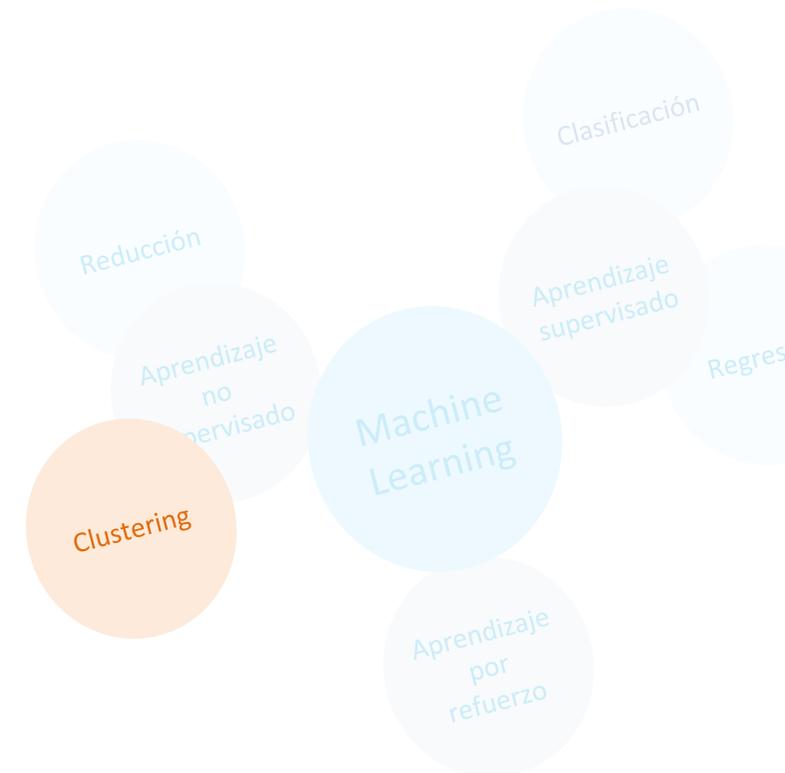
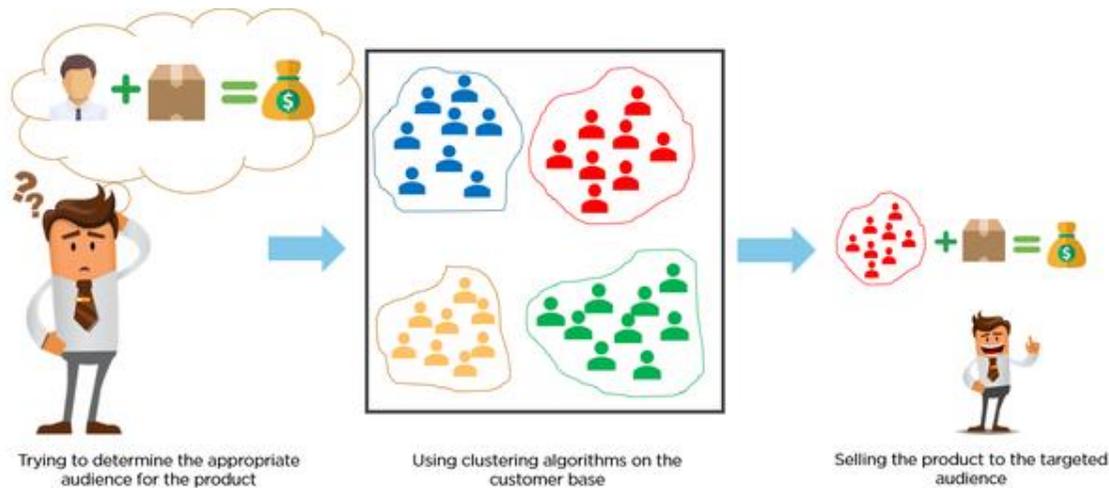
- Descubrimiento de estructuras
- Detección de anomalías
- Resumen de datos



Técnicas: Aprendizaje no supervisado

CLUSTERING

- Un clúster es un grupo o conjunto de objetos similares a cualquier otro incluido en el mismo cluster y distintos a los objetos incluidos en otros grupos.
- Clustering es la técnica de segmentar una población heterogénea en un número de subgrupos homogéneos o clusters.
- El clustering es subjetivo.



Técnicas: Aprendizaje no supervisado

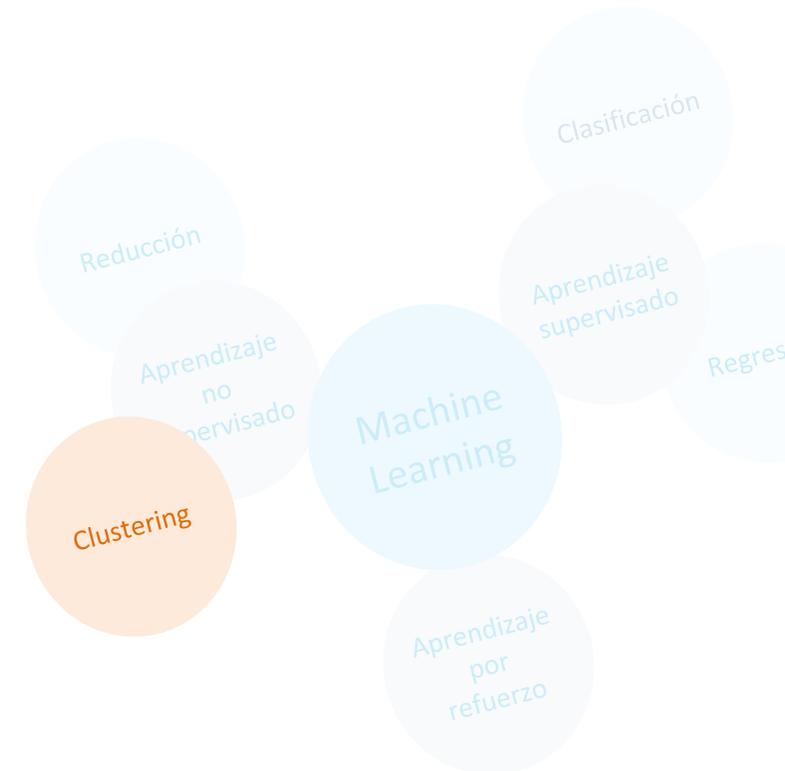
CLUSTERING

Aplicaciones

- Marketing:
 - Identificación de patrones de compra en los clientes.
 - Recomendación de artículos a nuevos clientes.
- Banca:
 - Detección de fraudes.
 - Identificación de clientes de alto riesgo.
- Seguridad:
 - Análisis exploratorio de los datos.
 - Detección de outliers o ruido.
 - Encontrar duplicados en un conjunto de datos.
 - Etapa de pre-procesado.
- Medicina:
 - Caracterización de patrones de comportamiento del paciente.
- Biología:
 - Agrupación de marcadores genéticos para identificar los lazos familiares.

¿Para qué utilizarlo?

- Análisis exploratorio de los datos.
- Detección de outliers o ruido.
- Encontrar duplicados en un conjunto de datos.
- Etapa de pre-procesado.

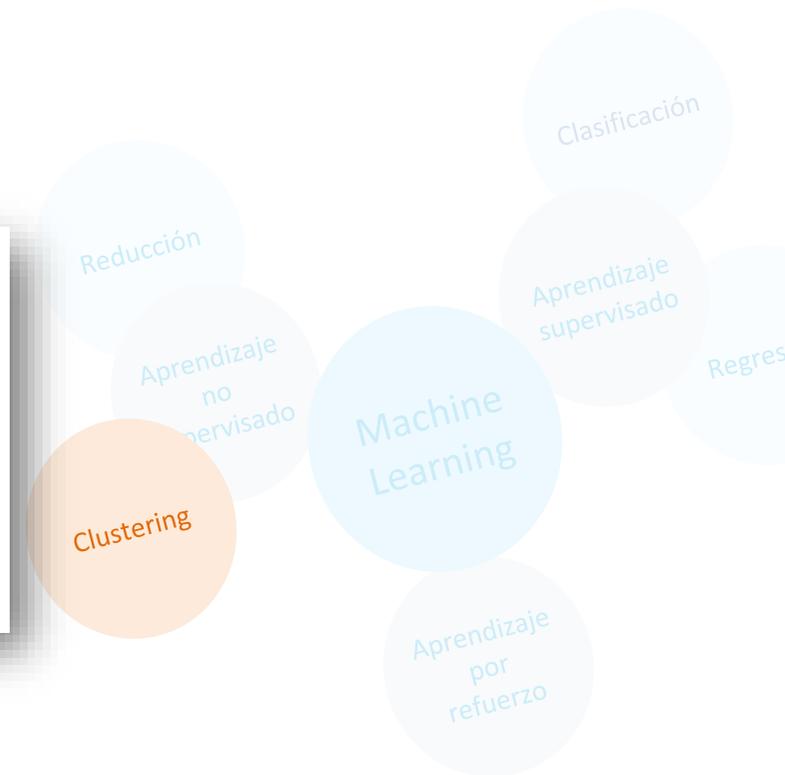
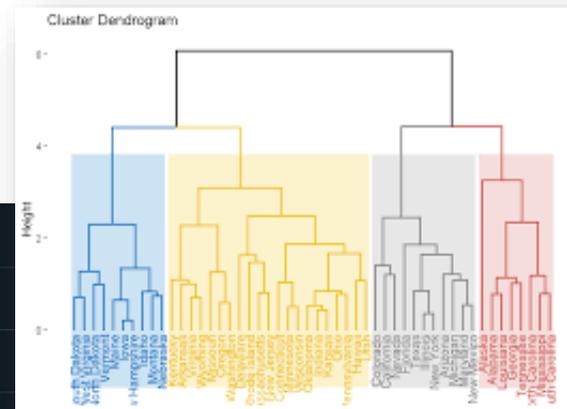
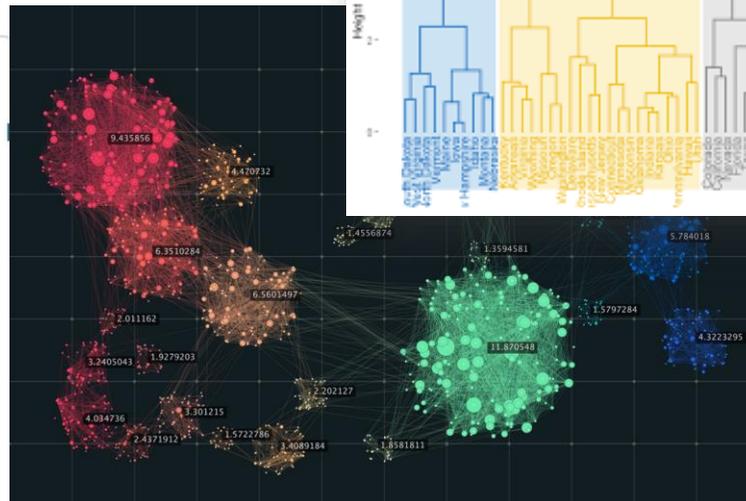
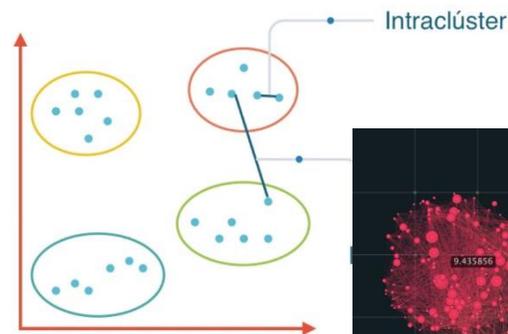


Técnicas: Aprendizaje no supervisado

CLUSTERING

Algoritmos

- Clustering basado en segmentación o partición (K-Means)
- Clustering jerárquico (Algoritmos aglomerativos)
- Clustering basados en densidad (DBSCAN)

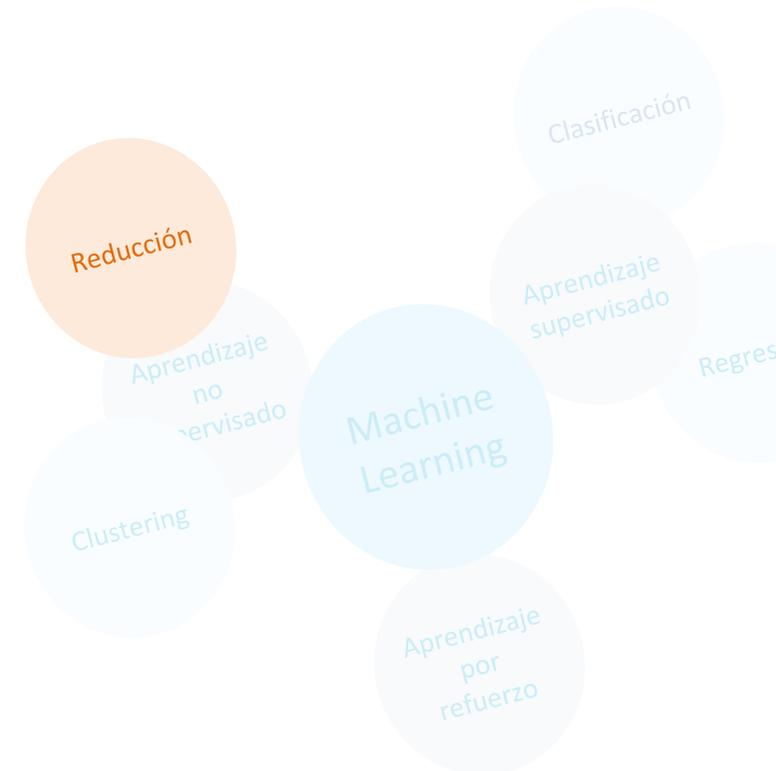
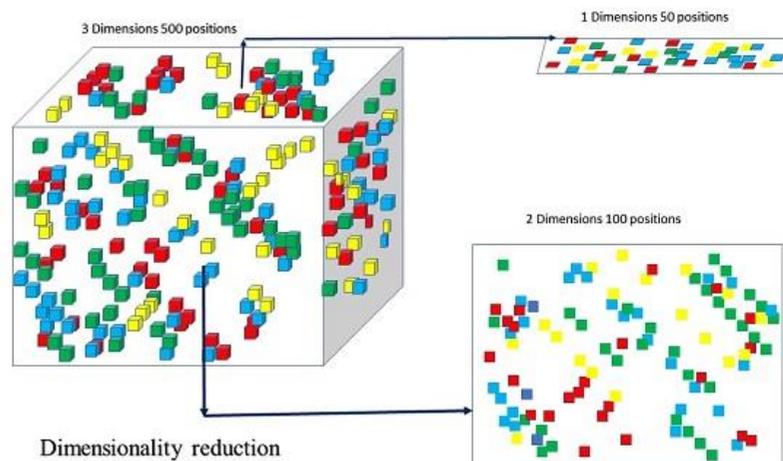


Técnicas: Aprendizaje no supervisado

REDUCCIÓN DE LA DIMENSIONALIDAD

Proceso de reducir el número de variables aleatorias del conjunto de datos mediante la obtención de un conjunto de variables principales.

- Identificar y eliminar las variables irrelevantes
- No siempre el mejor modelo es el que más variables contiene
- Mejora del rendimiento computacional
- Reducción de la complejidad



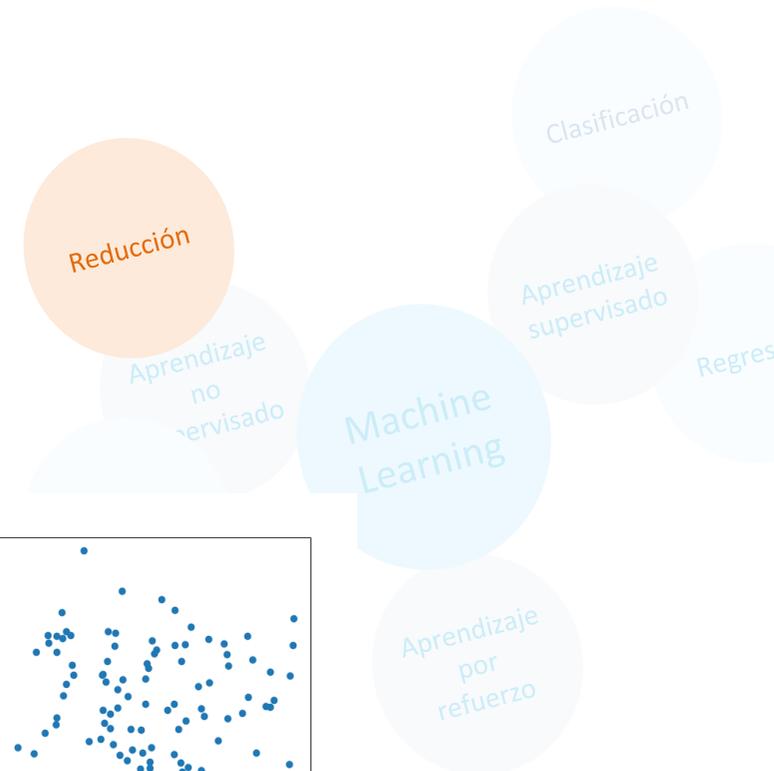
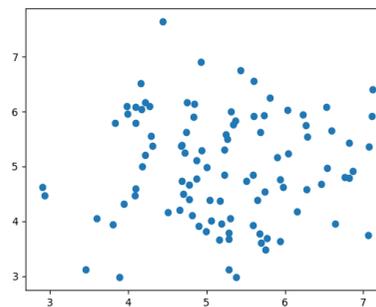
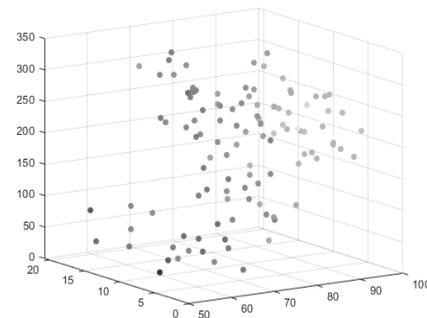
Técnicas: Aprendizaje no supervisado

REDUCCIÓN DE LA DIMENSIONALIDAD

Proceso de reducir el número de variables aleatorias del conjunto de datos mediante la obtención de un conjunto de variables principales.

- Identificar y eliminar las variables irrelevantes
- No siempre el mejor modelo es el que más variables contiene
- Mejora del rendimiento computacional
- Reducción de la complejidad

	total_bill	tip	sex	smoker	day	time	size
0	16.99	1.01	Female	No	Sun	Dinner	2
1	10.34	1.66	Male	No	Sun	Dinner	3
2	21.01	3.50	Male	No	Sun	Dinner	3
3	23.68	3.31	Male	No	Sun	Dinner	2
4	24.59	3.61	Female	No	Sun	Dinner	4



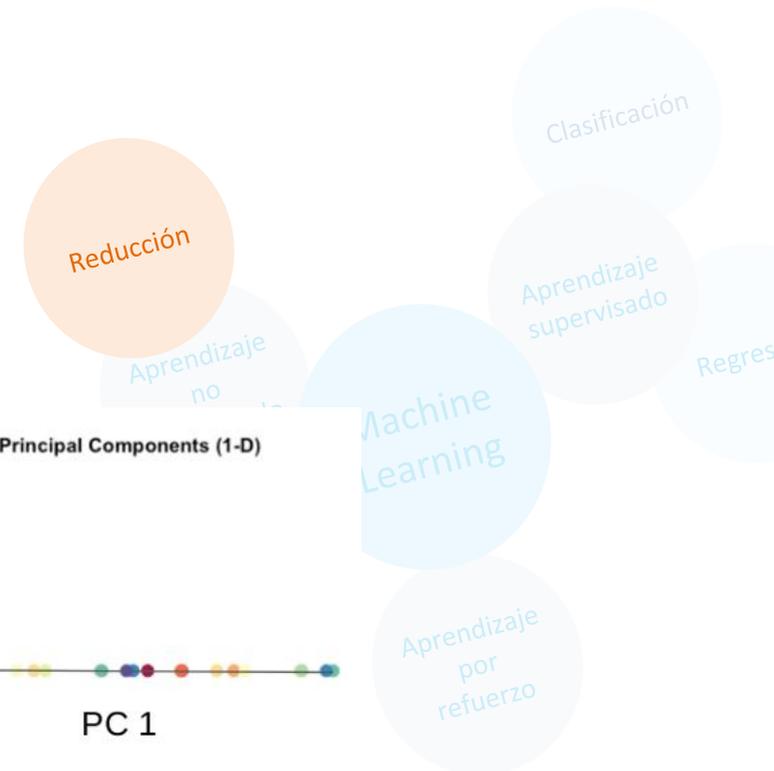
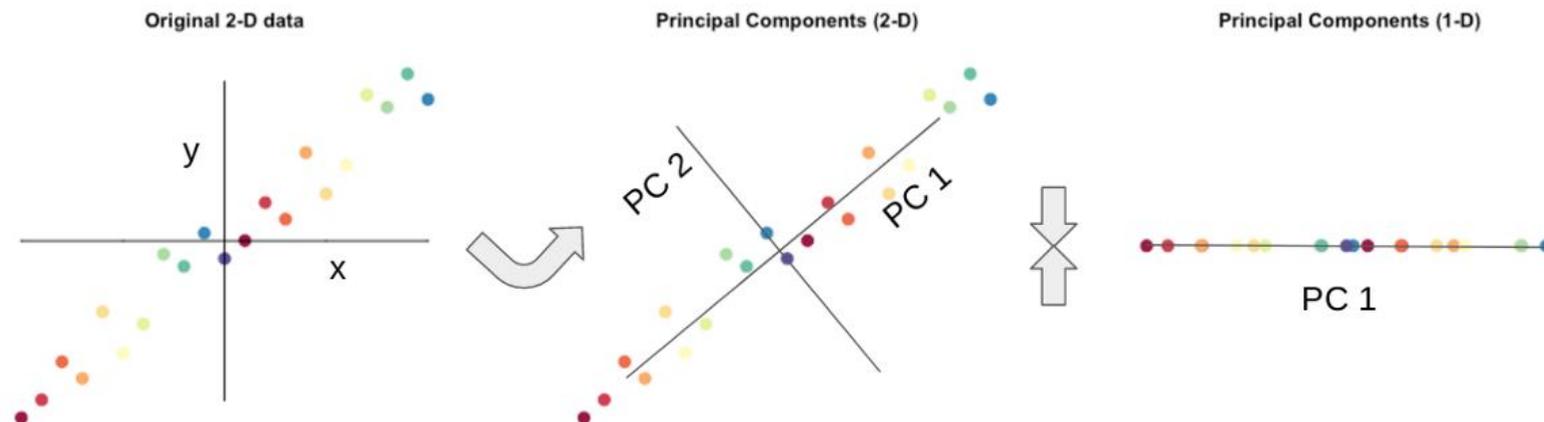
Técnicas: Aprendizaje no supervisado

REDUCCIÓN DE LA DIMENSIONALIDAD

Análisis de Componentes Principales (PCA)

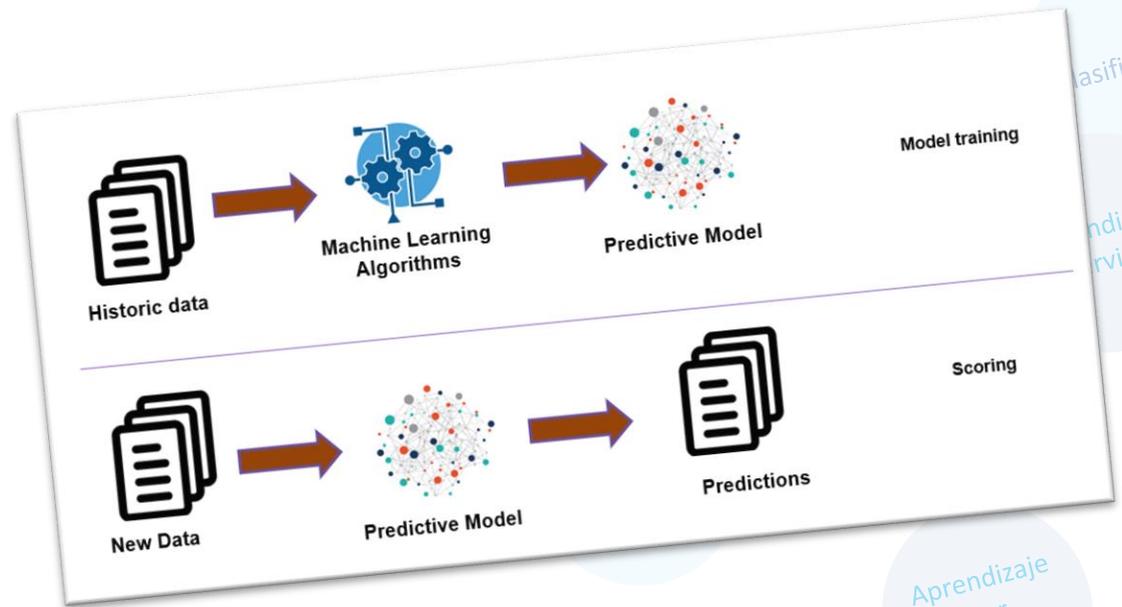
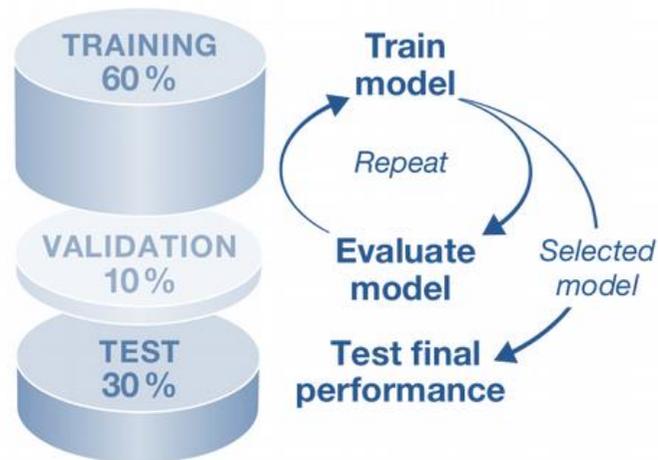
Objetivo: Encontrar aquellas direcciones de máxima varianza en datos de gran dimensionalidad y proyectarlos en un nuevo espacio de igual o menor dimensión al original

- Proyección ortogonal de los datos en un espacio lineal de menor dimensión, conocido como *componente principal*, con la varianza de los datos proyectados máxima.



Fases y criterios de evaluación

FASES



Aprendizaje por refuerzo

clasificación

rendizaje supervisado

Regresión

Fases y criterios de evaluación

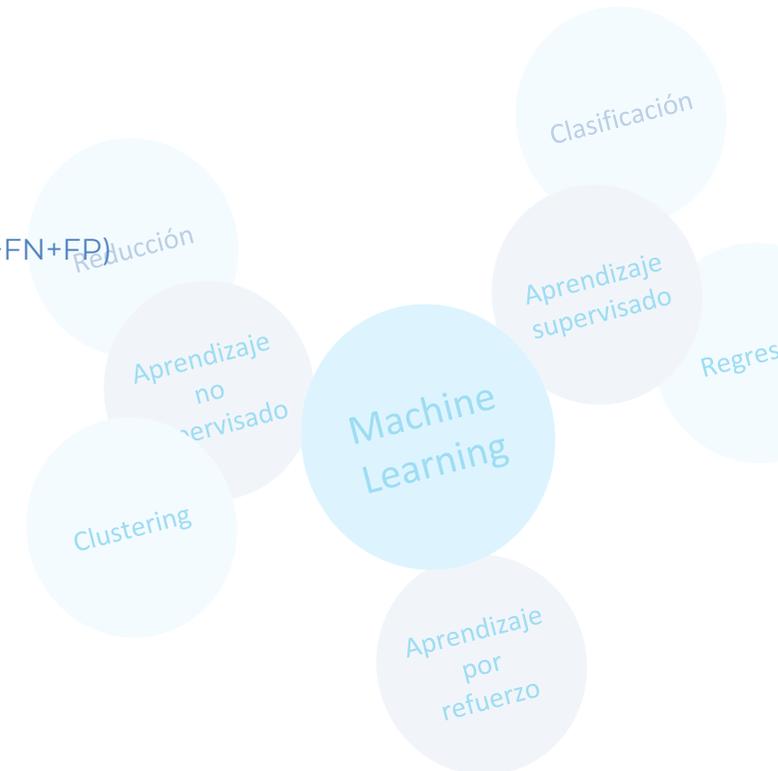
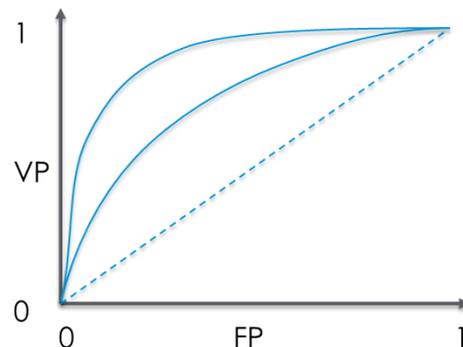
CRITERIOS DE EVALUACIÓN

- Precisión: métrica que permite evaluar el porcentaje de acierto que tiene el algoritmo → Matriz de confusión

		Predicción	
		Positivos	Negativos
Observación	Positivos	Verdaderos Positivos (VP)	Falsos Negativos (FN)
	Negativos	Falsos Positivos (FP)	Verdaderos Negativos (VN)

- Exactitud: $(VP+VN)/(VP+VN+FN+FP)$
- Precisión: $VP/(VP+FP)$
- Especificidad
- Sensibilidad
- ...

- Curva ROC: permite comparar visualmente distintos modelos de clasificación:



REALIDAD VIRTUAL & REALIDAD AUMENTADA

REALIDAD VIRTUAL

Representación de escenas e imágenes de objetos producidos por un sistema informático, que da la sensación de su existencia real.

- Ubicación del usuario en un entorno completamente virtual cubriendo su vista



<https://revistaderobots.com/rv/definicion-de-que-es-la-realidad-virtual/>

REALIDAD AUMENTADA

Integración de contenidos gráficos sobre una vista del mundo real

- Información extra virtual sobre elementos y entornos reales



<https://smartechgroup.es/realidad-aumentada-en-puntos-de-venta/>

Aplicaciones de la Ciencia de Datos a la Industria 4.0

Mantenimiento Predictivo

¿Qué es?

Es una técnica que se enfoca principalmente en **prevenir fallos o errores** en los equipos. De esta manera, todas las actividades empresariales se vuelven eficientes y fiables.



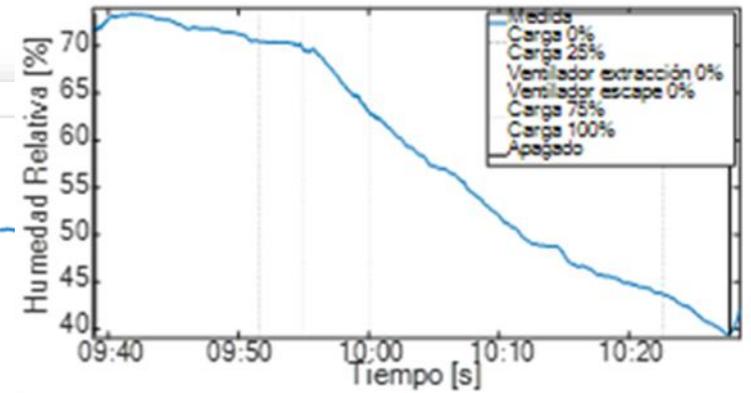
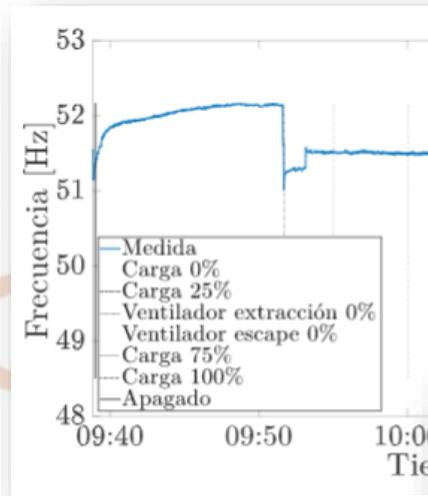
- Disminuir el riesgo de accidentes.
- Descender notablemente las reparaciones costosas por fallos que deben arreglarse rápidamente.
- Mejorar la confianza de los empleados.
- Reducir los errores en tareas diarias.
- Reducir el núm. de inspecciones innecesarias.
- Disminuir paradas imprevistas.



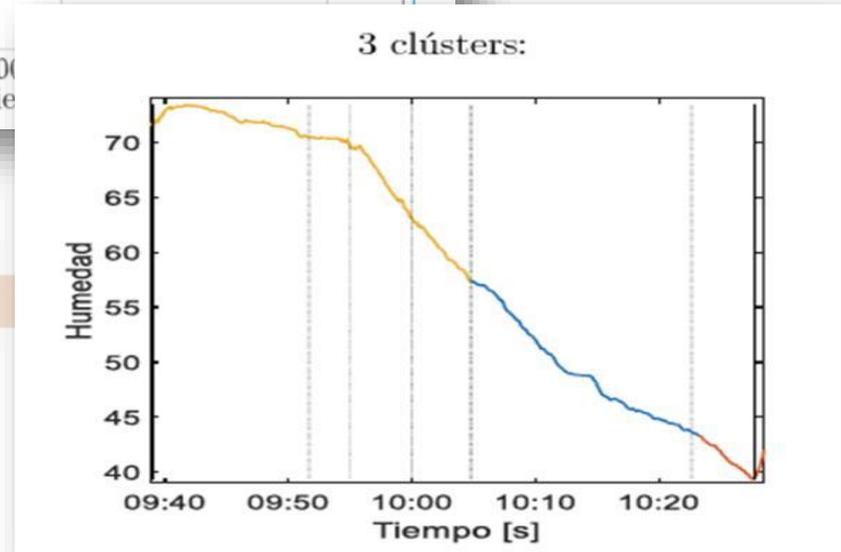
- Contratación de personal especializado.
- Complejidad al detectar el desgaste de las piezas de la maquinaria que su propia rotura



Mantenimiento Predictivo



UNIDAD DE CONTROL
(monitorización del
barco)



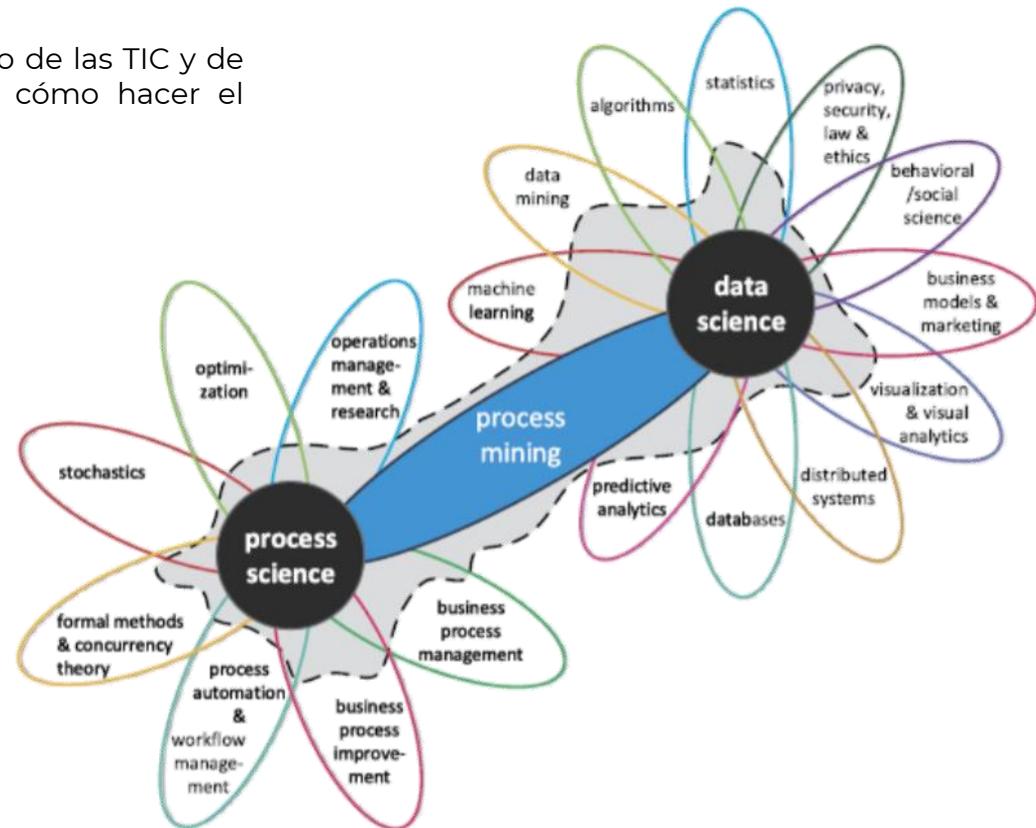
Predicción de stock

MINERÍA DE PROCESOS

El “descubrimiento de procesos” (combina conocimiento de las TIC y de las ciencias de gestión) centrada en el proceso y en cómo hacer el modelado, más que en aprender de los eventos.

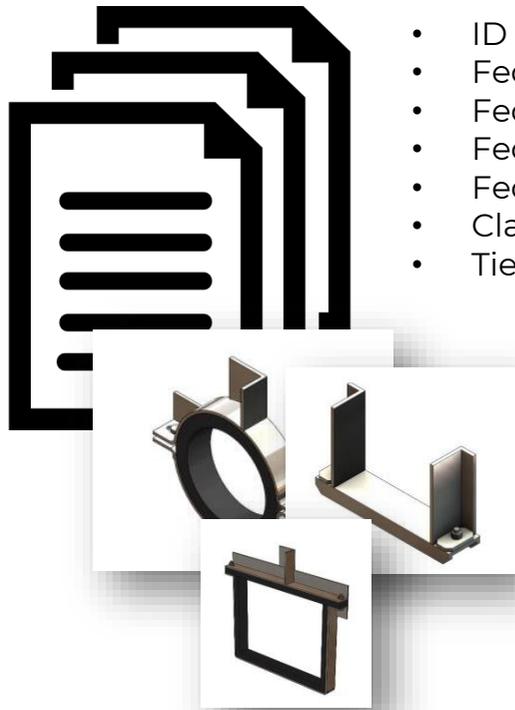
Aplicación:

- Descubrimiento de procesos
- Análisis de conformidad
- Mejora del proceso

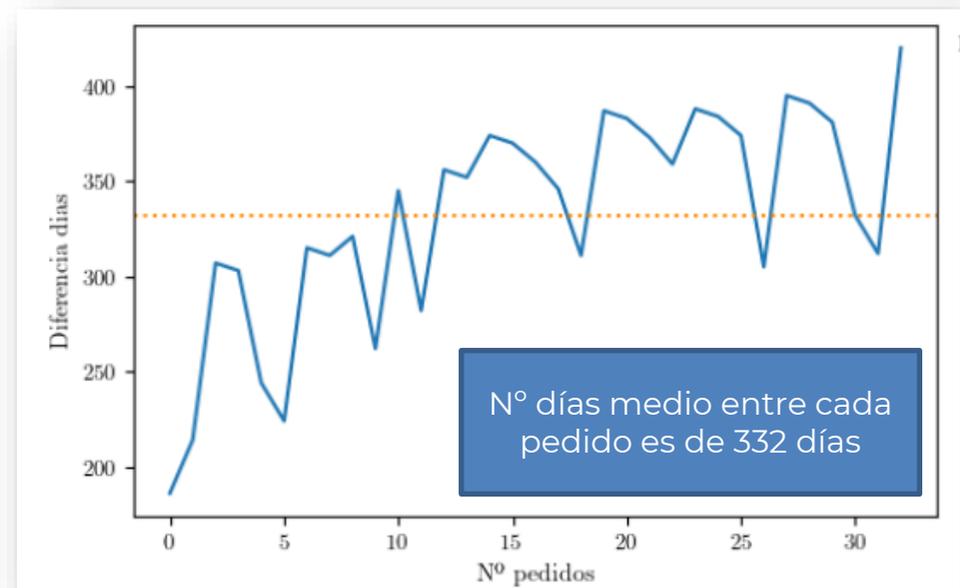


PREDICCIÓN DE STOCK

MINERÍA DE PROCESOS + CIENCIA DE DATOS



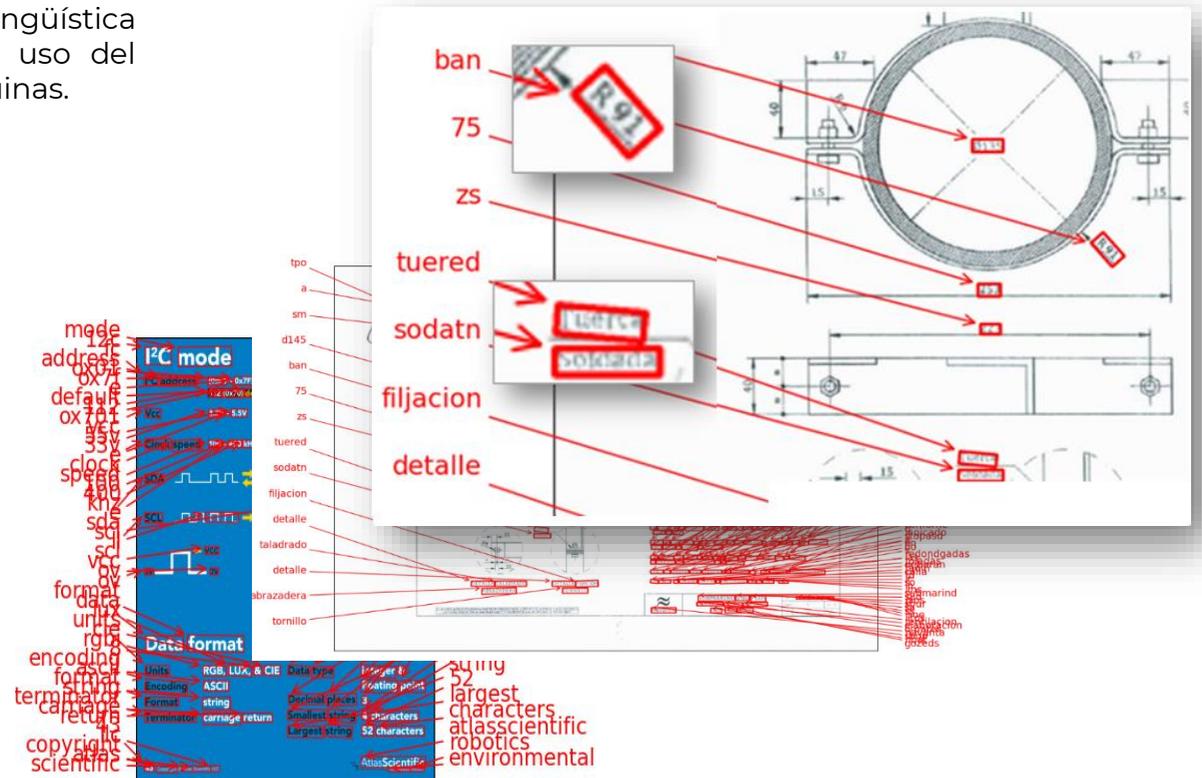
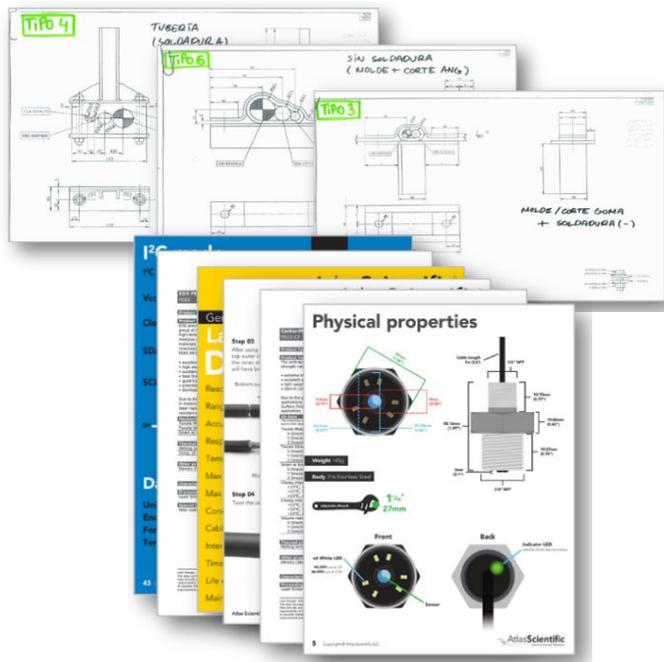
- ID pieza por pedido realizado y en curso
- Fecha de lanzamiento de pedido.
- Fecha de inicio de fabricación
- Fecha estimada de finalización de la fabricación
- Fecha real de finalización de la fabricación
- Clasificación de soportes por tipología
- Tiempos estimados de fabricación por tipología de soportes.



Digitalización de documentación técnica

Procesado Natural del lenguaje (NPL)

El procesamiento del lenguaje natural (PLN o NLP) es un campo dentro de la inteligencia artificial y la lingüística aplicada que estudia las interacciones mediante uso del lenguaje natural entre los seres humanos y las máquinas.



Gemelo Digital

¿Qué es?

Es una **representación digital** de un objeto, proceso o servicio físico: desde un motor a reacción o parques eólicos a edificios o ciudades enteras. Estas réplicas virtuales son utilizadas para hacer simulaciones antes de que se creen e implementen cambios en los objetos reales, con el fin de recopilar datos para predecir cómo funcionarán.

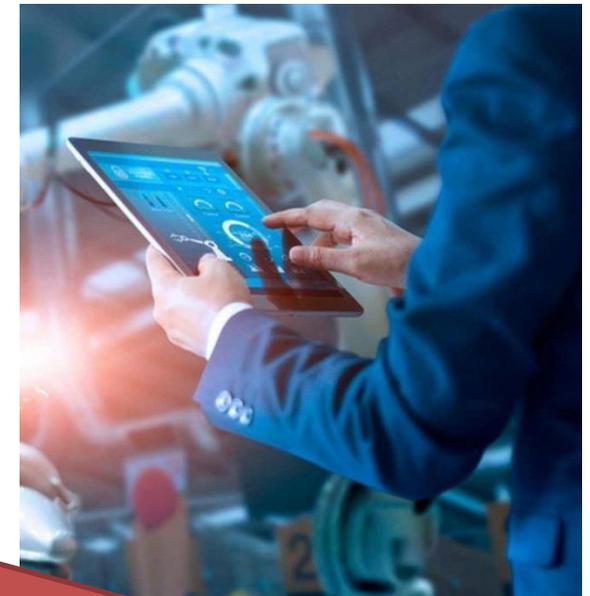
- Sensores IoT se recopilan los datos sobre el estado de los objetos/procesos/máquinas en tiempo real.
- Se analizan y procesan para crear un modelo digital.



Gemelo Digital

APLICACIONES

TOLVA ECOLÓGICA

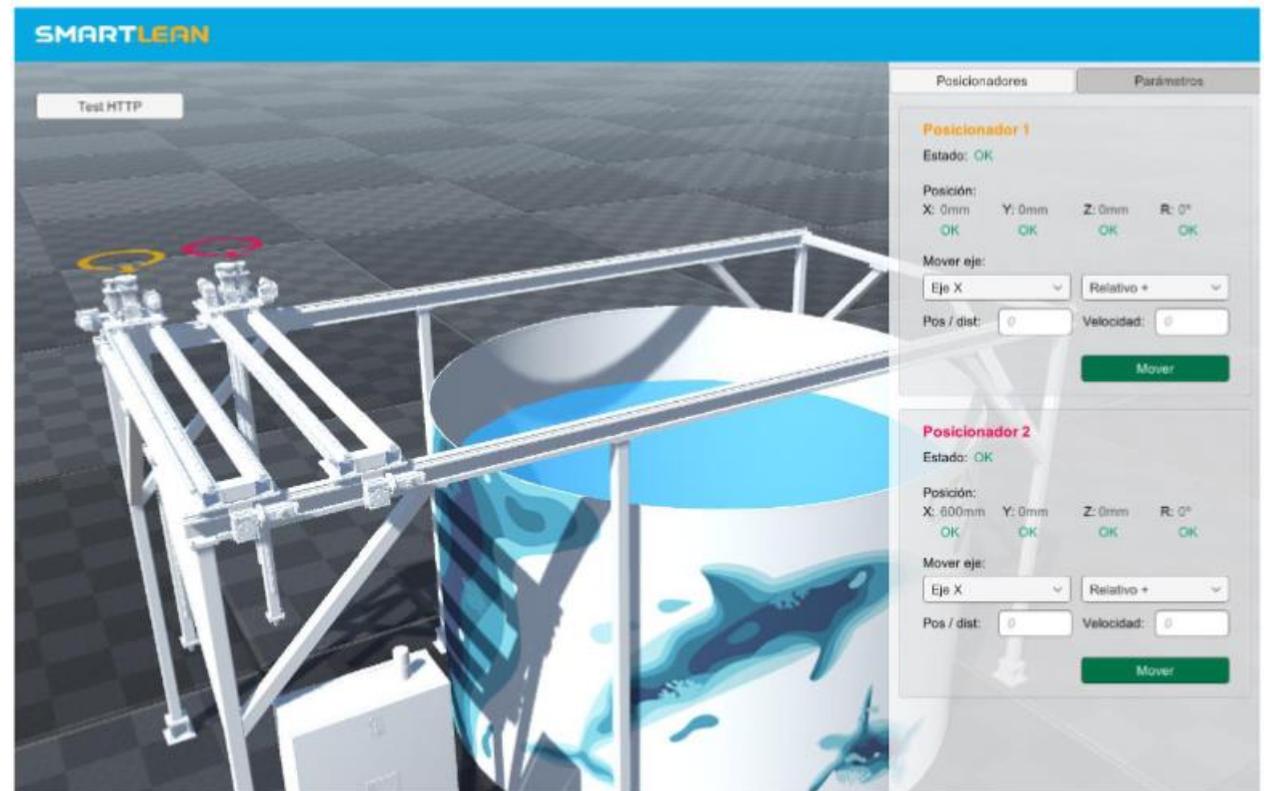
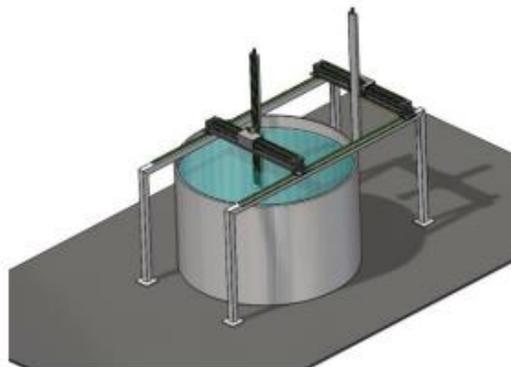
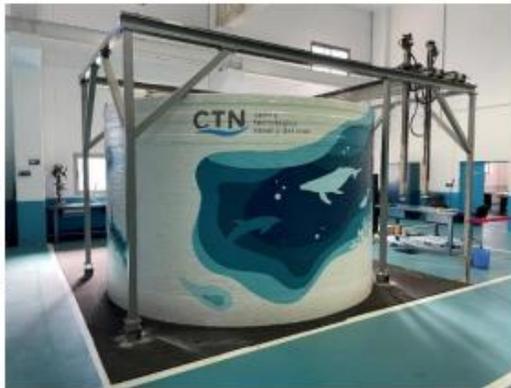


Reducción del impacto medioambiental mediante mantenimiento preventivo

Gemelo Digital

APLICACIONES

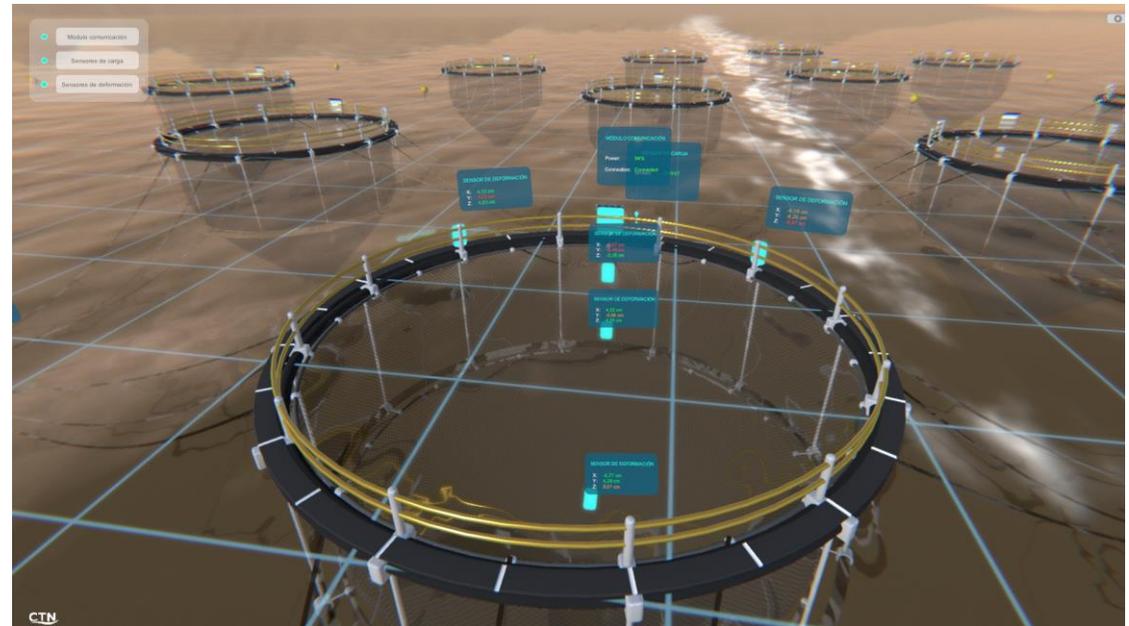
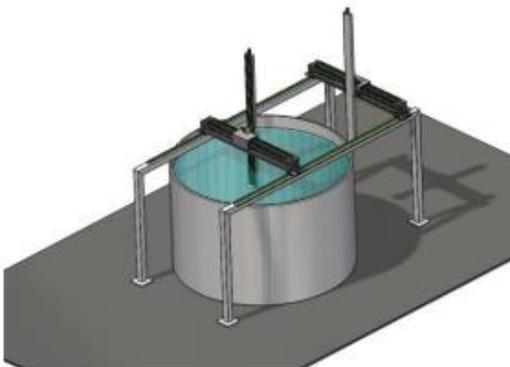
SISTEMAS DE POSICIONAMIENTO GEORREFERENCIADOS



Gemelo Digital

APLICACIONES

CASO AISLADO: SENSORIZACIÓN DE VIVEROS ACUÍCOLAS



REALIDAD VIRTUAL EN I4.0

FABRICACIÓN

- Monitorización de procesos y seguimiento de desarrollo
- Elaboración de pruebas con elementos virtuales antes de la aplicación real
- Nuevas posibilidades de diseño y personalización
- Visión pormenorizada del producto diferenciando entre diferentes piezas, capas o fases



<https://thinkmobiles.com/blog/augmented-reality-manufacturing/>

MANTENIMIENTO

- Detección de puntos calientes y elementos susceptibles de provocar fallos
- Inspecciones visuales remotas
- Guiado de procesos tanto presenciales como a distancia



https://www.prolink.com.au/technology/augmented-reality/shutterstock_ar-maintenance_low-res/

LOGÍSTICA

- Organización de cargas y distribución de mercancías
- Programación de tareas y gestión eficaz del trabajo de los operarios
- Accesibilidad a la información de pedidos
- Manipulación de objetos en remoto



<https://pierrejeanquenardel.wordpress.com/>

Muchas gracias



Nombre: **Rosa Martínez Álvarez - Castellanos**

Puesto: **Responsable de Ciencia de Datos**

Correo electrónico: **rosamartinez@ctnaval.com**

LinkedIn: **rosamartinezalc**



SECRETARÍA DE ESTADO
DE DIGITALIZACIÓN
E INTELIGENCIA ARTIFICIAL

red.es



UNIÓN EUROPEA

Fondo Europeo de Desarrollo Regional

“Una manera de hacer Europa”
