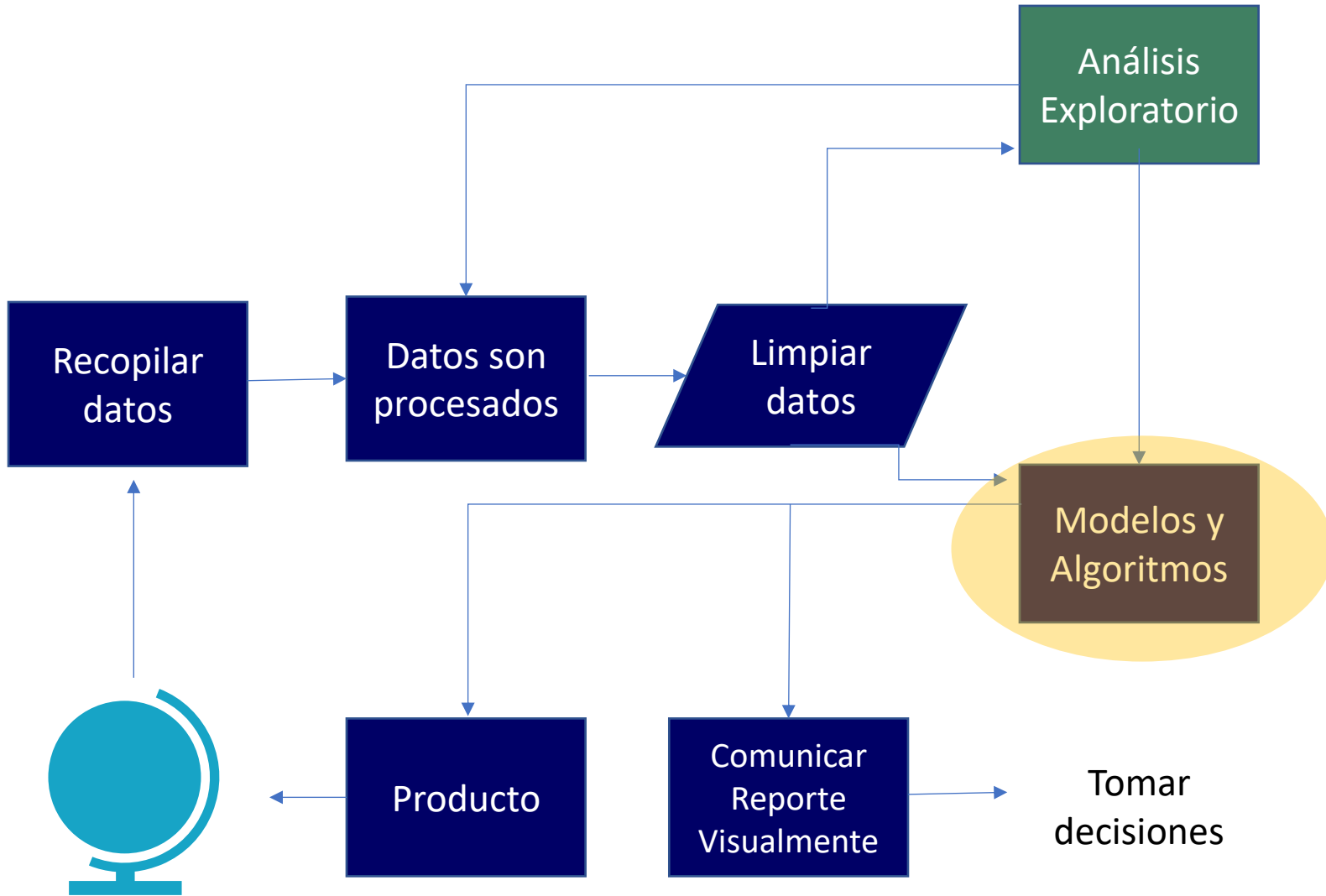




Modelos y Algoritmos

Proceso de Ciencia de Datos

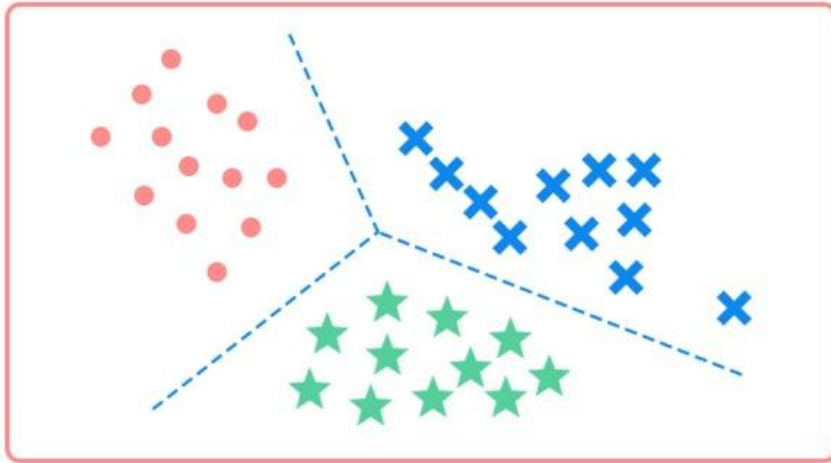


MODELADO

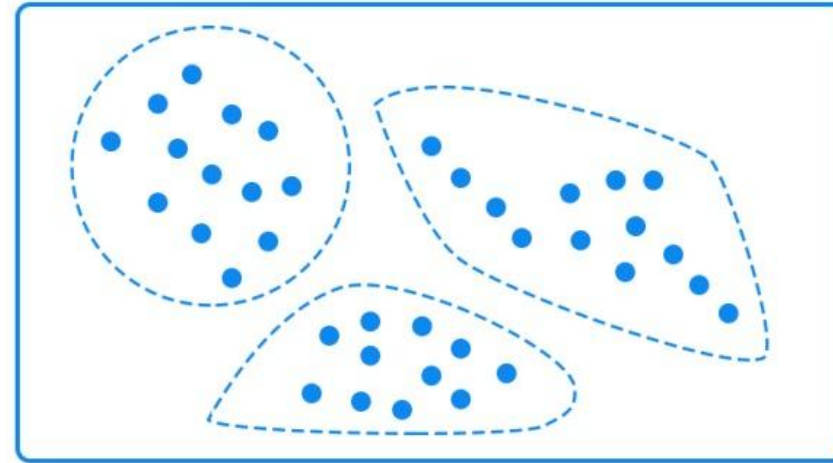


Cómo funciona el machine learning?

Tipos de aprendizaje automático



Supervised learning



Unsupervised learning

Actividad: a qué aprendizaje pertenece?

Predecir el precio de una casa

Diagnosticar una enfermedad
basado en síntomas

Encontrar grupos de usuarios
de Instagram similares

Predecir el número de clicks
en un anuncio

Segmentar clientes
basándonos en sus patrones de
compras

Clasificar a un correo en spam
o no spam

Encontrar reglas como
"Comprar pan → recomendar
mantequilla"

Supervisado

No Supervisado

Tipos de modelos de datos

REGRESIÓN - S

- ☐ Predice valores continuos
- ☐ Se quiere entender la relación entre Y y Xi
- ☐ Ejemplos:
 - Predecir ingresos de una persona a partir de su edad
 - Precio de un producto en función de costos, transporte



CLASIFICACIÓN - S

- ☐ Predice valores discretos o clases
- ☐ No busca una formula matemática, sino un corte empírico que funcione mejor
- ☐ Ejemplos:
 - correo spam/no spam
 - diabetes/no diabetes
 - Ganar/perder



CLUSTERING - U

- ☐ Para particionar los datos en grupos (clusters)
- ☐ El objetivo es dividir los datos de forma que los valores dentro del mismo cluster sean muy parecidos y puntos en diferentes clusters sean diferentes.
- ☐ No hay una "Y", sólo se tienen Xi y se quiere entender qué son (datos sin etiqueta)
- ☐ Ejemplo: Segmentación de clientes basada en el historial de compras para marketing



SERIES DE TIEMPO - S

- ☐ Datos ordenados en tiempo
- ☐ Busca entender la dinámica pasada del tiempo
- ☐ Ejemplos:
 - Precio de una acción
 - Ventas de clientes a través del tiempo

Tarea 2: elige el problema de aprendizaje supervisado (preparación del proyecto final)

Objetivo

Comenzar a pensar en un problema del mundo real que te gustaría explorar utilizando el aprendizaje automático **supervisado**. El objetivo es encontrar **dos conjuntos de datos** relacionados con un tema que te interese:

1. Un conjunto de datos para un problema de clasificación
2. Un conjunto de datos para un problema de regresión

La tarea es INDIVIDUAL. Pero este es un paso de preparación para el proyecto final, que se completará en equipos asignados al azar.

Tarea 2: elige el problema de aprendizaje supervisado (preparación del proyecto final)

 Qué necesitas hacer?

1. Elige uno o dos temas que realmente te interesen (Ejemplos: salud, deportes, educación, marketing, finanzas, medio ambiente, psicología, juegos, música, etc.)
2. Describe brevemente tu tema (3 a 5 oraciones):
3. ¿Por qué te interesa?
4. ¿Qué tipo de preguntas podrían responderse con datos?
5. Encuentra dos conjuntos de datos relacionados con los temas:
 - Un conjunto de datos adecuado para un problema de clasificación
 - Un conjunto de datos adecuado para un problema de regresión
6. Para cada conjunto de datos, incluye lo siguiente:
 - Nombre del conjunto de datos
 - Enlace al conjunto de datos (o cargue el archivo CSV)
 - Variable objetivo (lo que intentas predecir)
 - Breve explicación de por qué es clasificación o regresión

Tarea 2: elige el problema de aprendizaje supervisado (preparación del proyecto final)

Entrega

- Documento breve (.docx, .pdf o .txt) con sus descripciones.
- Incluir enlaces o archivos para ambos conjuntos de datos.
- Sube todo a la plataforma de canva

Notas importantes

- No es necesario limpiar ni analizar los datos todavía.
- El proyecto final se realizará en equipos, pero esta tarea es individual.
- Asignaremos equipos al azar y los proyectos seleccionados pueden basarse en los conjuntos de datos que propongas.

Tarea 2: elige el problema de aprendizaje supervisado (preparación del proyecto final)



Criterio de Evaluación

Criterio	Puntos
Tema descrito claramente	2
Dataset de clasificación	3
Dataset de regression	3
Los datasets son accesibles (link/archivo)?	2
Total	10

Aprendizaje Supervisado

Cómo se evalúan los modelos?