

Fundamentos del Aprendizaje de Máquinas

Taller comercio y desarrollo regional

Módulo Aprendizaje de máquinas

Taller práctico para la elaboración informes

Educación continua | Universidad de los Andes
octubre 26 – noviembre 5

2021

Hoy

De qué vamos a hablar

1. Preliminares para este módulo

*Metodología de taller. Estructura del módulo.
Alcance y competencias.*

2. Calibrando motores:

JamBoard para alinear con nuestros proyectos.

3. Fundamentos del ML:

*¿Qué es el aprendizaje de máquinas? ¿Qué tipo
de retos enfrenta? ¿Cuáles son sus productos?*

4. Cómo se ve un proceso de análisis de ML

*Sus etapas, nociones de evaluación, ojo con el
sobreajuste – ¡Ejercicio En R!*



Photo by [The Coach Space](#) from [Pexels](#)

Preliminares

Arrancamos el módulo ML



Photo by [The Coach Space](#) from [Pexels](#)

¡Bienvenidos/as al módulo taller 3!

- Alfredo Eleazar Orozco Quesada ae.orozco10@uniandes.edu.co
- Comité de docencia, Anderer Sinn, educación, innovación mediante tecnología y diseño
- Pregunten mucho: yo **no me las se todas**, pero se moverme entre las herramientas y buscamos la respuesta entre todos (así es la vida real)
- Repositorio: <https://github.com/alorozco22/taller-comercio-2021>
- ¡Vamos a calentar la voz y abrir micrófonos! (no calentar en serio) pero sí vamos a conversar mucho, **preguntas activas**, etc.

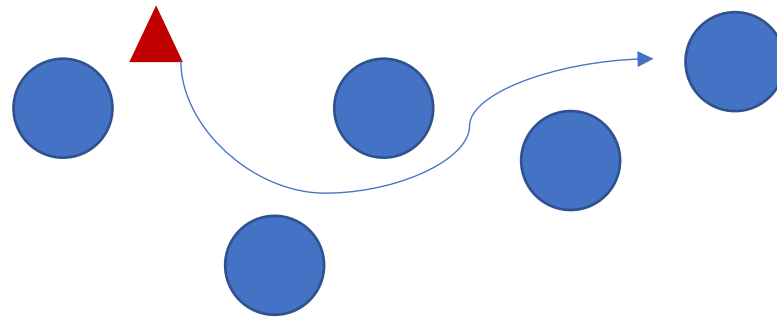
Alcance y competencias del módulo

- El objetivo de este módulo es potenciar sus proyectos
 - Desde las herramientas del curso
 - Desde el estado en el que se encuentren
- Lograr una competencia de familiarización con el aprendizaje de máquinas y sus técnicas
- Dejarles un esquema de pensamiento para que puedan profundizar autónomamente en las herramientas que más les sirvan

No nos vamos a volver expertos en todas las herramientas, pero vamos a entender para qué sirven y más o menos cómo se trabaja con ellas en R.

Curso Taller

- Curso **intensivo en práctica**: talleres acompañados
- Una parte de las clases tendremos exposición con preguntas activas
- Otra parte nos reunimos en grupos:
 - Avanzamos en los proyectos con una dinámica propuesta o bien libremente
 - El tutor va rotando por grupo para dar una retroalimentación personalizada



Curso Taller - Advertencias

- Vamos a ser muy concretos con la teoría, pero es importante comprenderla bien para sacarle el máximo a este taller.
- Vamos a pensar todo el tiempo **¿esto puede servirme para mi documento? ¿Para que otros retos podría servirme?**
- Aquí vamos a direccionar todas las explicaciones hacia la elaboración del documento, pero la idea es que se vayan con una noción de panorama.
 - Hoy es una exposición de fundamentos del aprendizaje de máquinas.
 - A partir de la otra clase, la teoría va a ser más reducida y habrá más dinámica de taller acompañado orientado hacia la elaboración del documento.

Estructura del módulo 3

- Vamos calibrando los tiempos según como sientan que vamos avanzando.

Clase	Estrategia
Clase 1 – Fundamentos del Aprendizaje de Máquinas	Exposición con partes de taller.
Clase 2 – Modelos de aprendizaje supervisado	Exposición – Avance en proyecto
Clase 3 – Modelos de aprendizaje no supervisado	Exposición – Avance en proyecto
Clase 4 – Técnicas de evaluación	Exposición – Avance en proyecto

Otros detalles

- Recuerden, nuestra clase es en la misma sala de siempre. Enlace en el repo.
- Vamos a trabajar en R, igual que en los cursos anteriores.
- En este curso, vamos a programar entre todos/as. Vamos a compartir pantalla, conversar errores de código entre todos/as, etc.
- **Regla de juego:** para preguntas activas primero intentamos todos/as responder, luego sí compartimos
- Vamos a tener un break de 15 minutos en medio de la clase
- ¿Dudas?

Pregunta:

En este punto... ¿cómo
estamos de datos?

1. No tenemos datos concretos, sólo del tipo de información que queremos usar.
2. Ya tenemos datos pero no los hemos movido mucho.
3. Ya tenemos datos y los hemos procesado limpiándolos lo suficiente para calcular un par de procedimientos.

Pregunta:

En este punto... ¿cómo van nuestros objetivos?

1. No hemos comenzado ningún tipo de análisis, todavía estamos formalizando la idea de lo que queremos hacer.
2. No hemos corrido análisis pero tenemos una estrategia más o menos clara de hacia dónde queremos ir.
3. Ya tenemos una intuición de cómo queremos abordar el proyecto y vamos en ruta.

Pregunta:

En este punto... ¿cuál es nuestro mayor reto?

1. Piensen desde las cuestiones **metodológicas** y técnicas, hasta tal vez el **tiempo** con el cuentan, la **planeación**, el **estado de avance** cómo están fluyendo en el **equipo**, etc.

Pregunta:

¿Cuál es la idea de nuestro informe?

1. ¿Qué objetivo general tenemos con este proyecto? ¿Qué queremos lograr? ¿Cuál es el enunciado?
2. Cuéntenme un poco de su tema, etc.

Jamboards

- 1: https://jamboard.google.com/d/1EydfEJWRsg4PNsqWnGefXe4b7sdCUXS_2oxSUOpDPUU/edit?usp=sharing
- 2: https://jamboard.google.com/d/1whER4YqSG718UtVidbO4yPBB9YSnUGmIS_6M8l6x4ul/edit?usp=sharing
- 3: https://jamboard.google.com/d/1tCwSXlssfoGQw_n7W6XWazidpT83qhQMXwzFaTR6y7o/edit?usp=sharing
- 4: <https://jamboard.google.com/d/1wCDbkMP1VpR3GHsopZ50B9569HFtYKGpYfM5k9Jgtq0/edit?usp=sharing>
- 5: https://jamboard.google.com/d/1RgJDHcn7XEQYm0y9VNRTubirCVSox9kcn_7AjzVizyE/edit?usp=sharing
- 6: <https://jamboard.google.com/d/1cBXrJug0PfQk4fEOoqmACbL8mXgl0Kw-gaAatXzml4c/edit?usp=sharing>
- 7: https://jamboard.google.com/d/1a9NEOPsd_3TxbVXqR9RS0XIOWUvwXekookqkuHM4uQ8/edit?usp=sharing

Fundamentos del ML

Las máquinas aprenden...

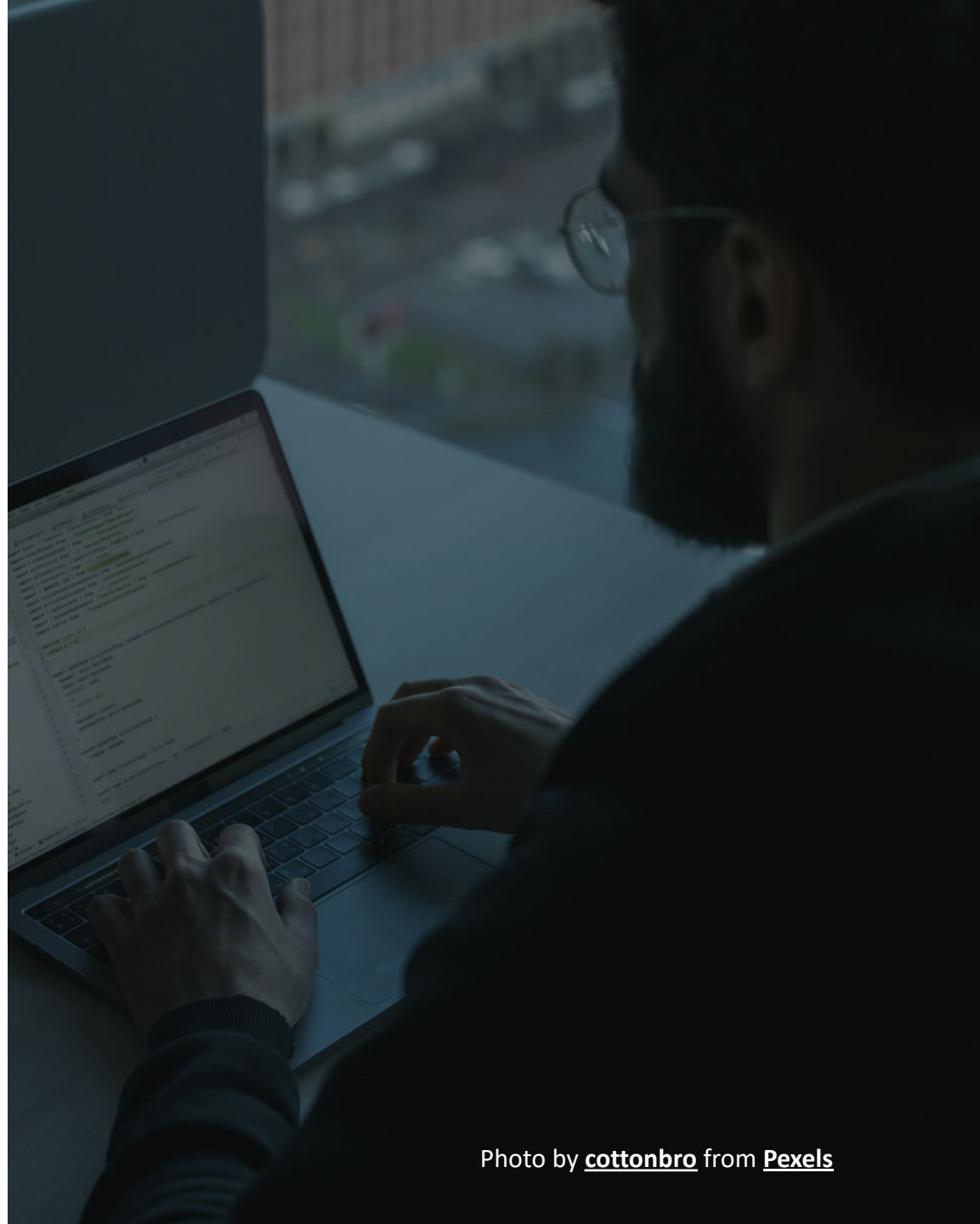


Photo by [cottonbro](#) from [Pexels](#)

Qué podemos hacer con ML

- **Marketing:** llega un nuevo cliente, y la máquina lo clasifica en un segmento u otro.
- **Finanzas:** Se presentan datos de transacciones, y la máquina predice con algo de probabilidad si hay fraude o no.
- **Agricultura:** Se registran nuevos datos de condiciones climáticas y la máquina predice cuándo van a estar lista la cosecha.

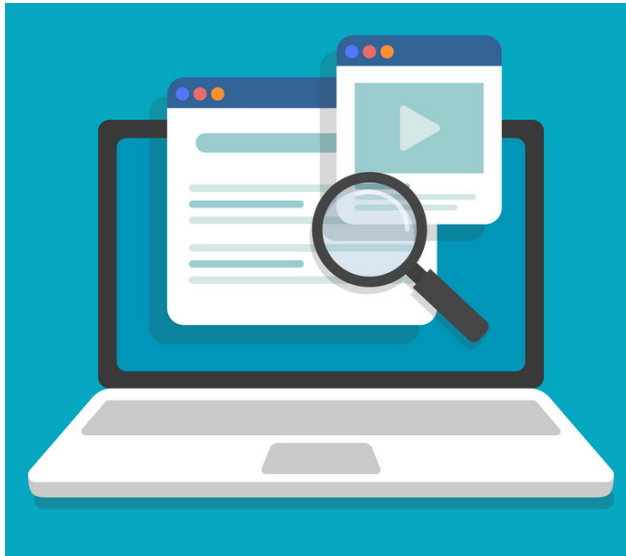


Analítica predictiva

- Hasta ahora hemos **visto hacia el pasado**: ¿qué cantidades están relacionadas entre sí?
- ¿Qué podemos decir sobre los datos y eventos registrados?
- ¿Qué podemos aprender nosotros sobre la información capturada?
- En aprendizaje de máquinas queremos **entrenar un programa** con datos del pasado, para que **cuando tengamos datos nuevos** la máquina pueda reaccionar a ellos



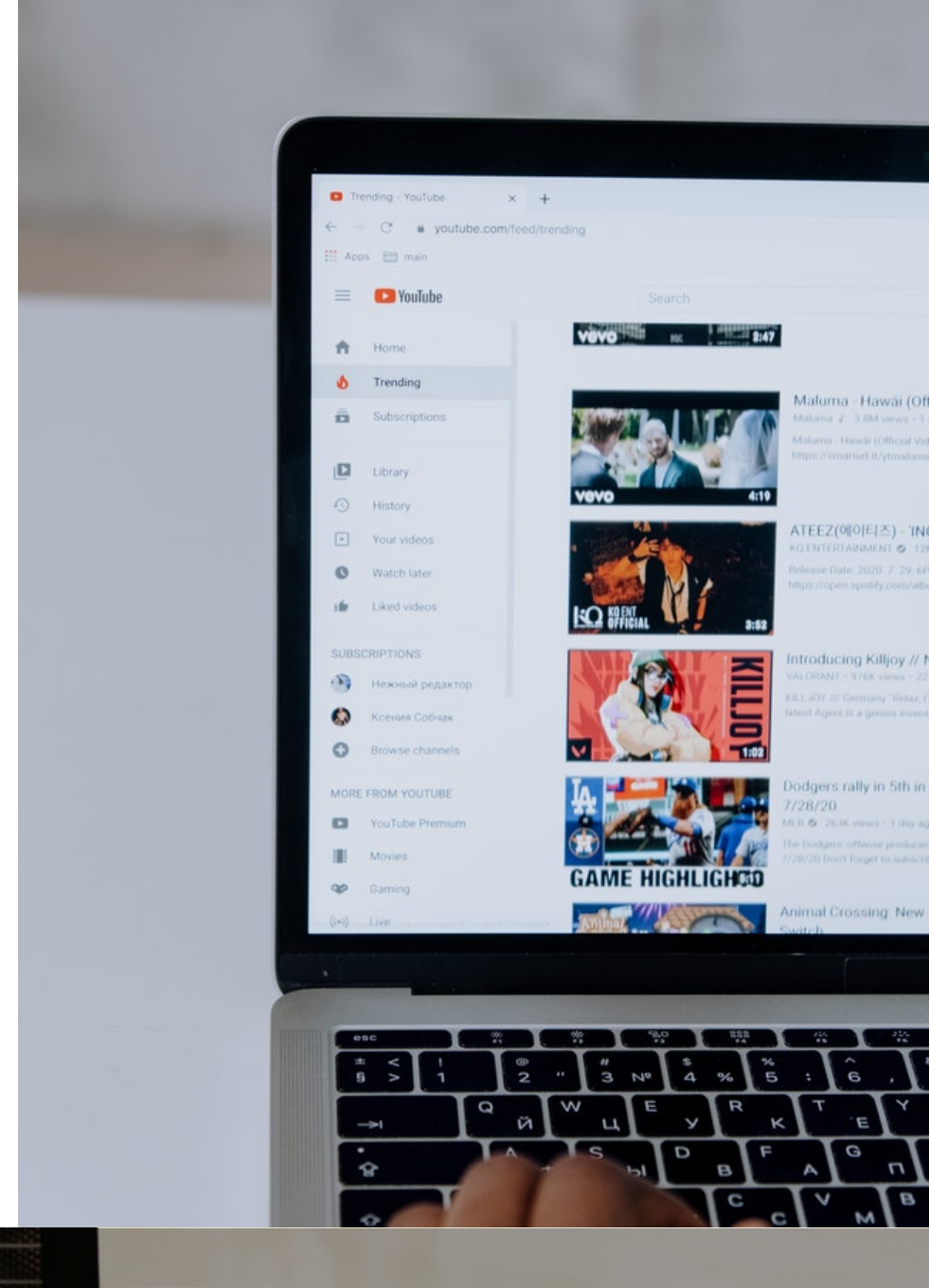
Analítica predictiva



- Por lo general un ejercicio descriptivo: **informes, análisis, documentos, estimaciones, etc.**
- Por lo general, un ejercicio predictivo: **herramienta que se puede usar en el futuro**

Analítica predictiva

- A veces utilizamos ML para ejercicios descriptivos:
 - **Cuando queremos procesar mucha información que luego podemos analizar:** tenemos tweets y queremos ver cómo se relacionan sus emociones con otros datos.
 - **Cuando queremos encontrar relaciones de agrupamiento entre los datos:** tenemos canastas de consumo, y nos damos cuenta de que pañales y cerveza son productos muy cercanos.



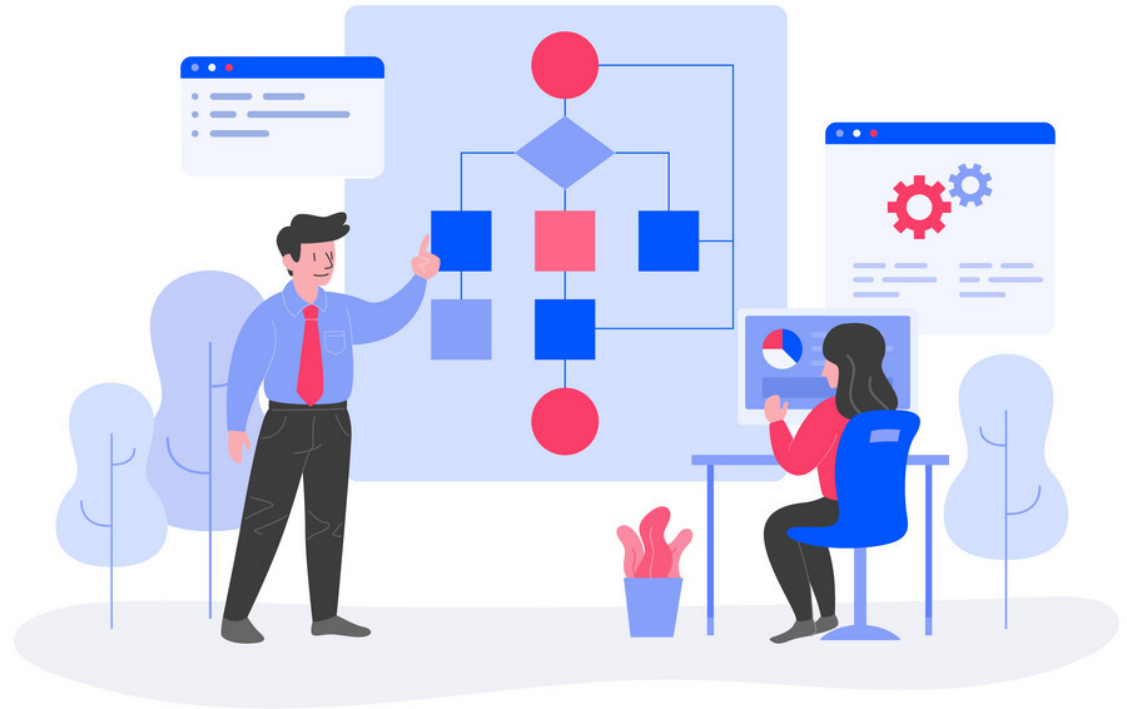
Pregunta:

¿Sus informes son descriptivos o predictivos?

En un principio... programación explícita

Computer Algorithm

- En computación trabajamos con algoritmos:
Instrucciones para que un computador solucione una tarea concreta
- Programarlo explícitamente suele tomar tiempo y conocimiento detallado de los problemas: preparar chocolate



Pero había retos difíciles de conquistar

- Visión por computador: ¿es una naranja o es una manzana?

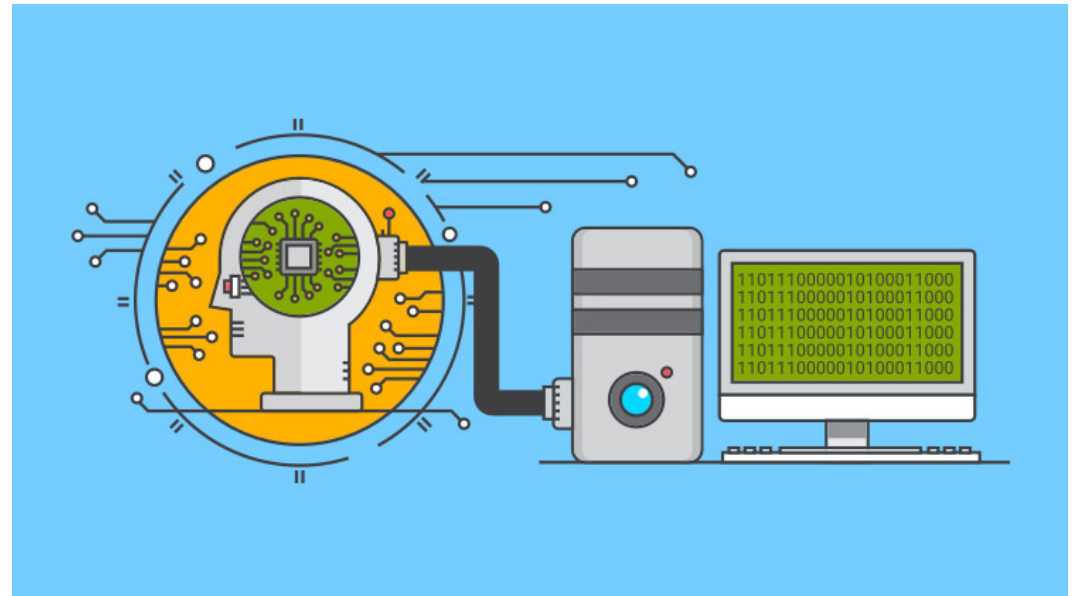


Pero había retos difíciles de conquistar



Las máquinas aprenden a partir de la experiencia

- En *Machine Learning*, no le damos instrucciones directamente a la máquina.
- Damos datos y les decimos a la máquina cómo emplear matemáticas para ajustar sus algoritmos.



Las máquinas aprenden a partir de la experiencia

- Formulamos instrucciones generales que llamamos modelos.
- Esos modelos relacionan la información que ingresamos con potenciales predicciones, decisiones o cursos de acción.
- Pero en lugar de escribir el modelo completo, dejamos ***parámetros*** abiertos, sin definir.
- Esas cantidades se ajustan con los datos que les proporcionamos a las máquinas.

Al final, las máquinas adivinan lo mejor que pueden



Podemos predecir cosas ante nuevos clientes, por ejemplo.

- Una vez un modelo ha sido ***entrenado*** con datos, la máquina puede hacer predicciones en situaciones nuevas.
- ¡Note! Estos modelos ya le permiten trabajar a las máquinas con probabilidades y lógica difusa!

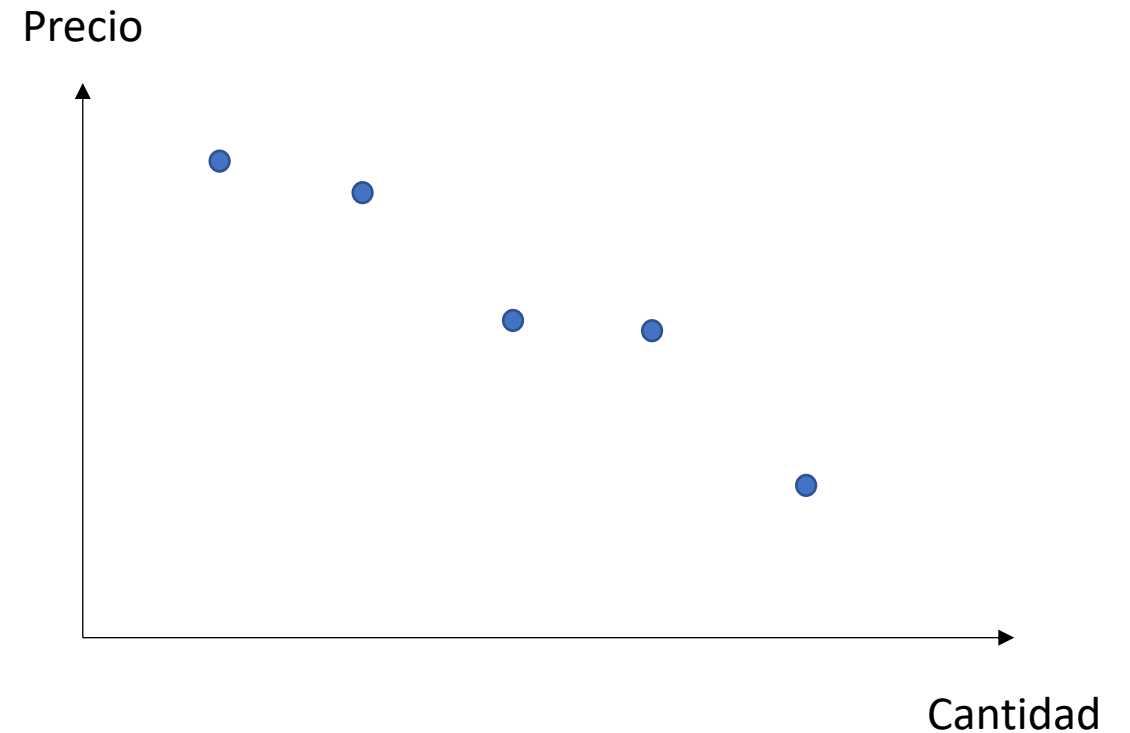
Conozcan a Clara



- Ella trabaja en un departamento de mercadeo para una empresa de productos
- Su empresa cuenta una encuesta que hizo acerca de cuánto comprarían a diferentes precios
- Ella quiere aprender algo de esos datos históricos

Clara decide ir por una aproximación gráfica

- Grafica para cada precio cuántos productos se venderían.
- Ella observa que mientras más costosos los productos, menor cantidad se vende.



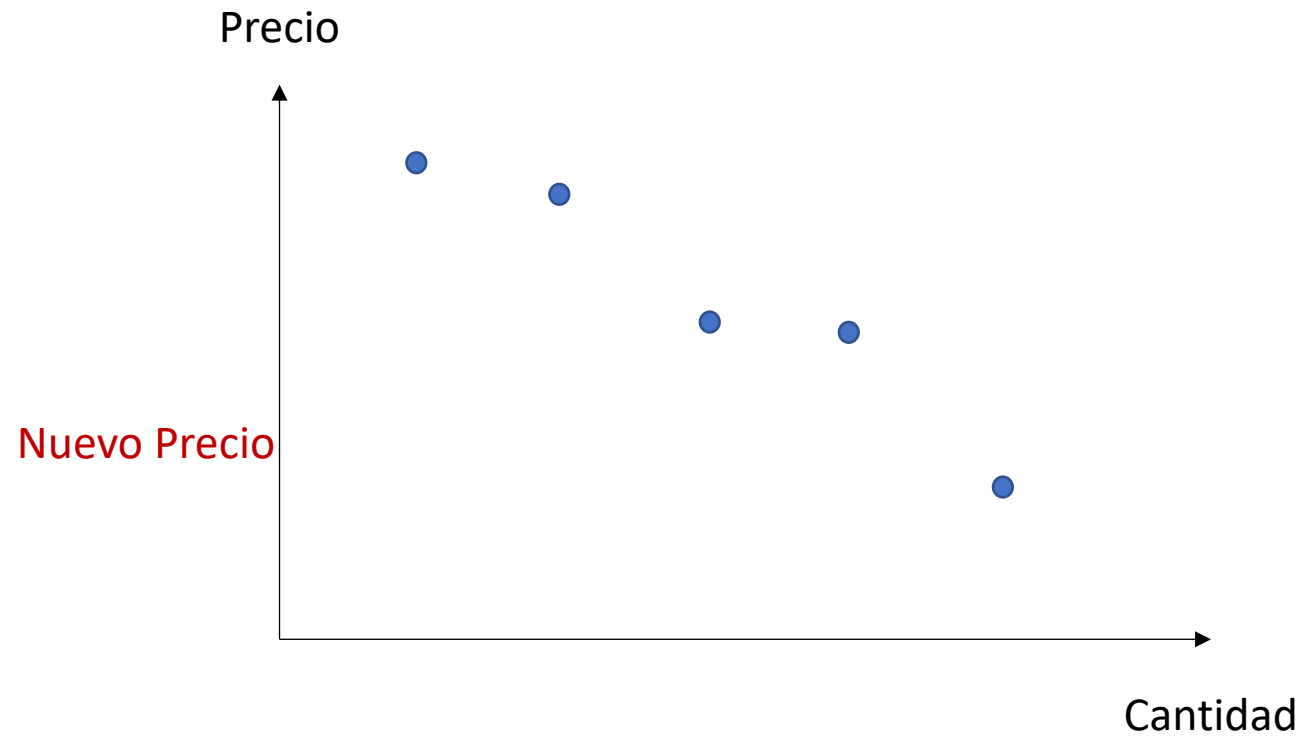
Clara quiere ver el futuro



- Clara quiere saber cuántos productos se venden si ponen un **precio nuevo**, para el que no se preguntó

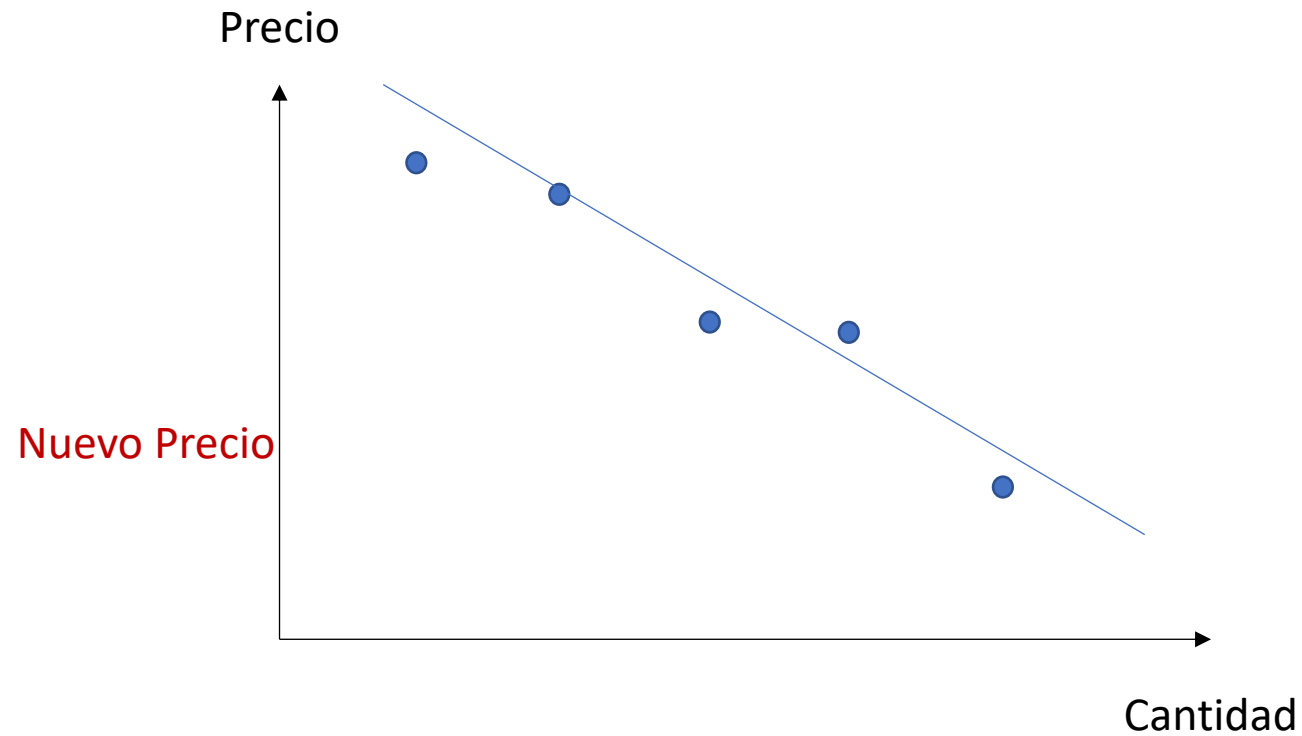
Clara encuentra un modelo

- Vuelve de nuevo a los datos y los observa detenidamente...



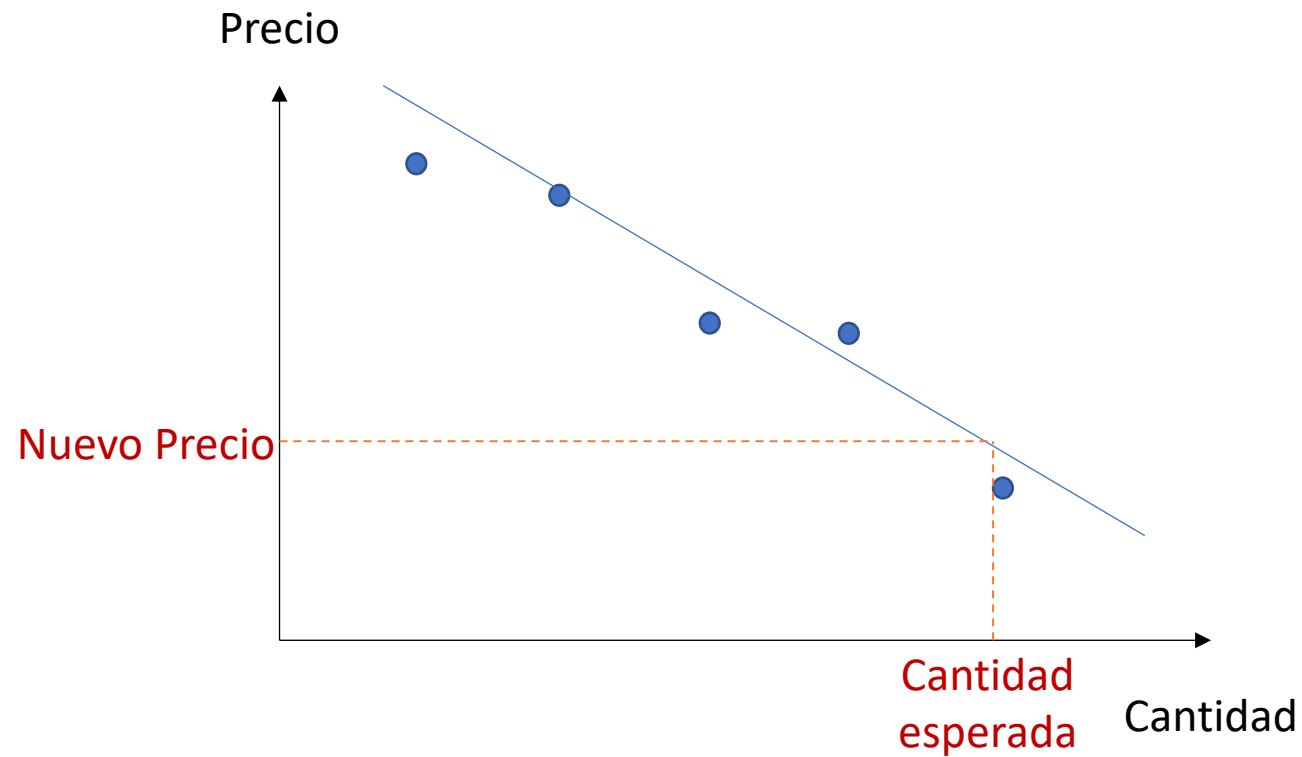
Clara encuentra un modelo

- Se da cuenta de que tal vez una línea describe muy bien lo que pasa

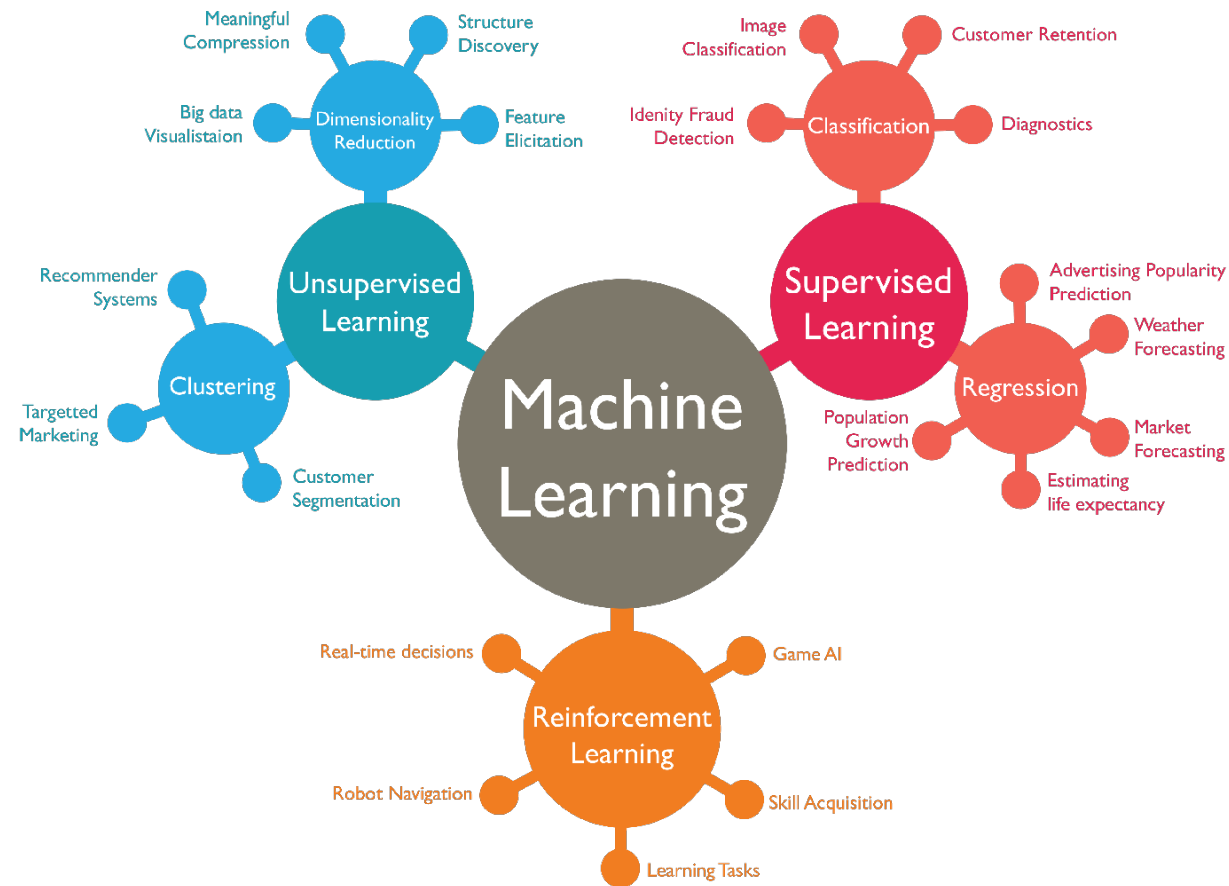


Clara encuentra un modelo

- Ahora puede predecir a partir de su modelo para el nuevo precio



Tipos de retos



Aprendizaje supervisado

Identificador	Dato útil 1	Dato útil 2	ETIQUETA
Alfredo	12	7432	100.000
Juliana	16	9375	150.000

- Contamos con datos “históricos” con una respuesta correcta registrada: esa etiqueta puede construirla un ser humano o registrarse.
- Clasificación: Los datos útiles conducen a una categoría.
- Regresión: Los datos útiles conducen a predecir un valor continuo.

Aprendizaje no supervisado

Identificador	Dato útil 1	Dato útil 2
Compra de las 2 pm	Pañal	7432
Compra de las 2:01 pm	Leche	9375

- Tenemos bases de datos que no tienen etiquetas, queremos encontrar estructuras y relaciones detrás de los datos mismos
- Reducción de dimensionalidad: quitamos columnas que nos dan información de sobra.
- Clustering: Identificamos grupos de filas que tienen similitud entre sí.

El proceso del aprendizaje de máquinas

Cómo planeamos un análisis de *Machine Learning* – Caso con modelo de clasificación lineal

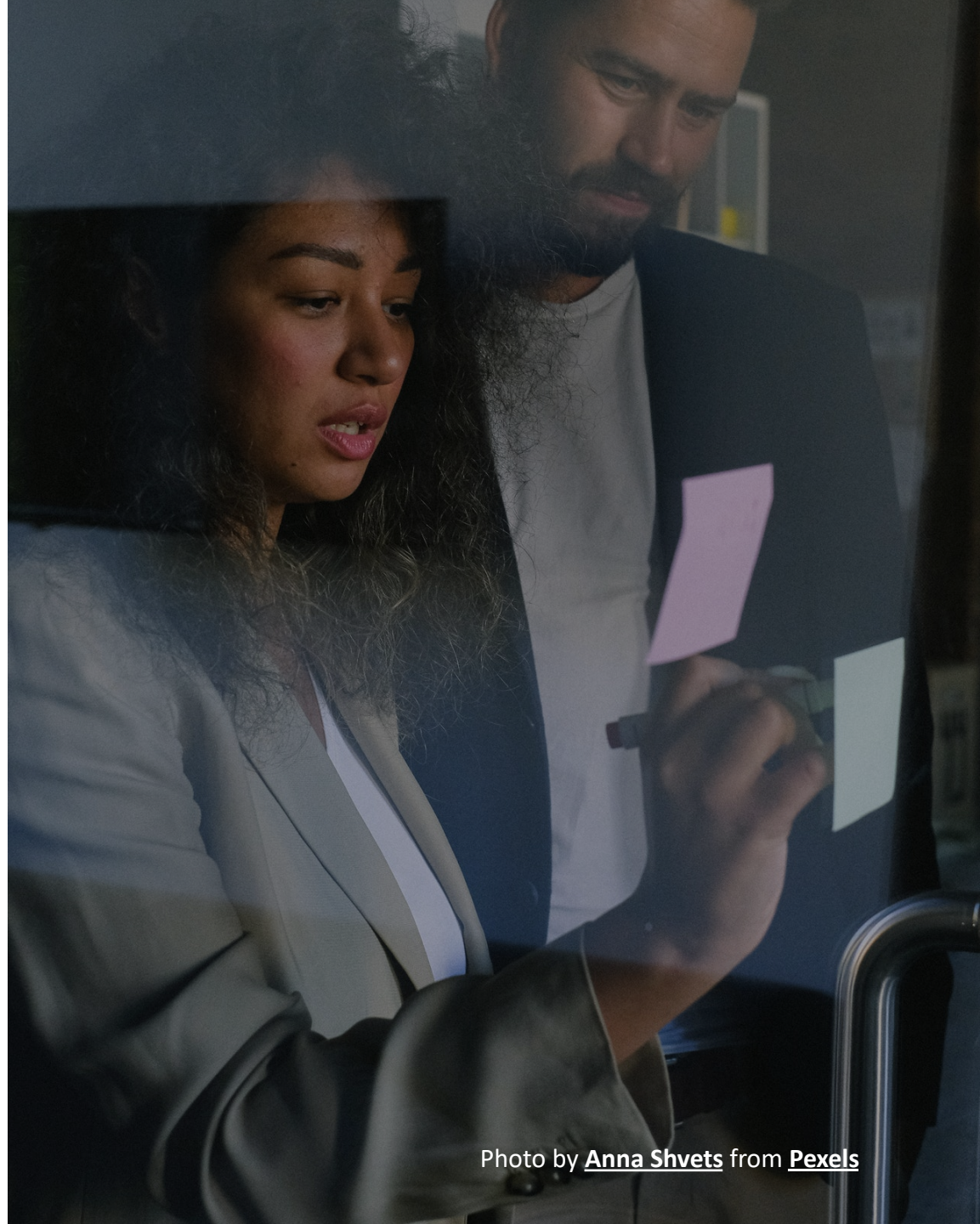


Photo by [Anna Shvets](#) from [Pexels](#)

Etapas de todo proceso de ML



1. Limpieza
2. Exploración descriptiva
3. Agrupación: datos de entrenamiento y datos de prueba*
4. Entrenamiento
5. Evaluación*

* Veremos en más detalle en un momento o al final del curso

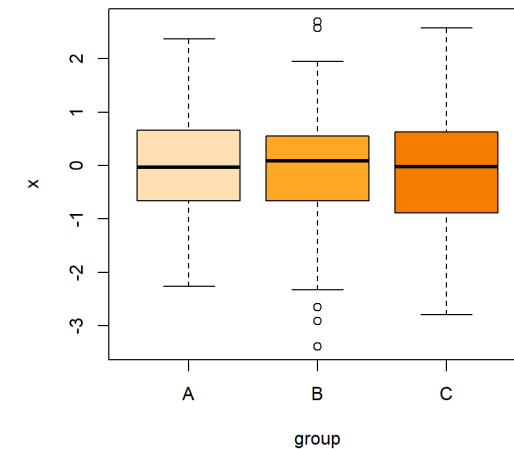
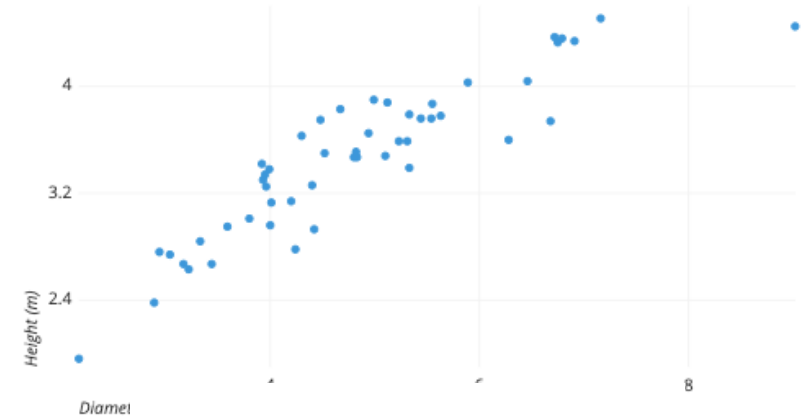
Limpieza

- Limpieza puede llegar a ser el 60% del tiempo que tome un proyecto.
- Buscamos que los datos queden listos para ser utilizados:
 - Revisamos observaciones atípicas
 - Eliminamos errores en los datos
 - Convertimos los formatos adecuadamente
 - Hacemos preprocesamiento si el modelo lo requiere...



Descripción: exploración de relaciones

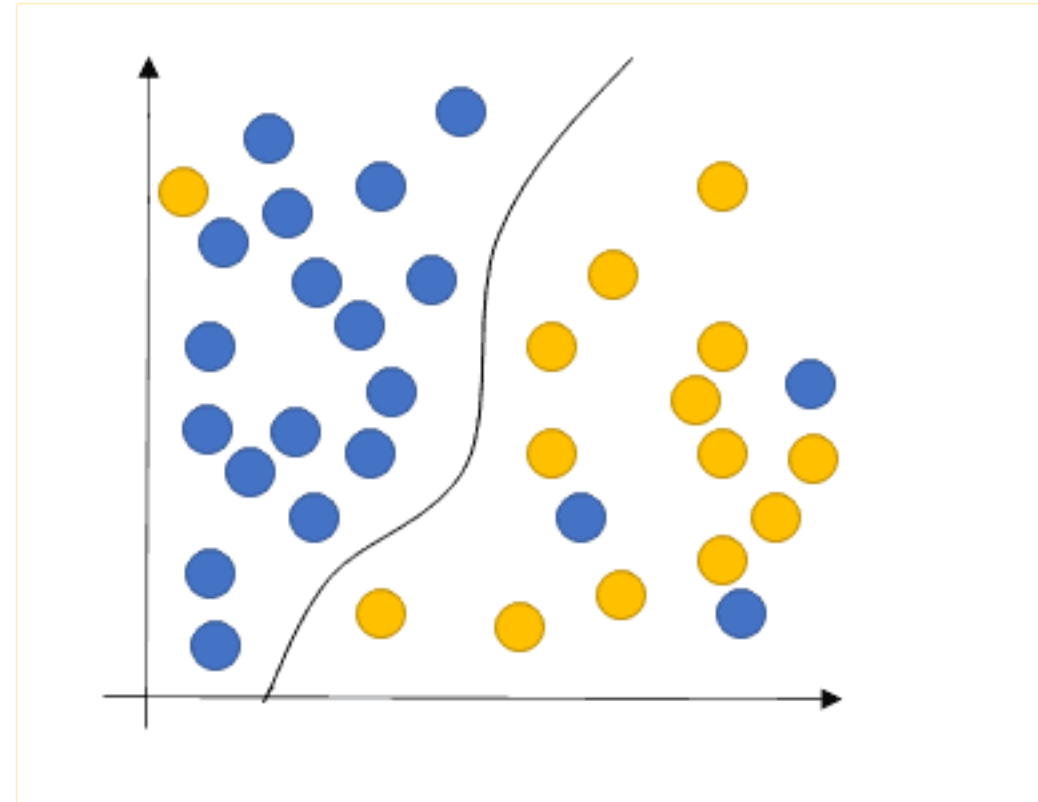
- Esta etapa exploratoria nos permite identificar datos más útiles que otros para el objetivo de predicción.
- Identificamos variables que contienen información sobre otras variables.
- Podemos irnos por aproximaciones como correlaciones o regresiones
- También podemos graficar en diagramas de dispersión o barras por grupos para notar patrones.



Agrupación: el problema del sobre-ajuste

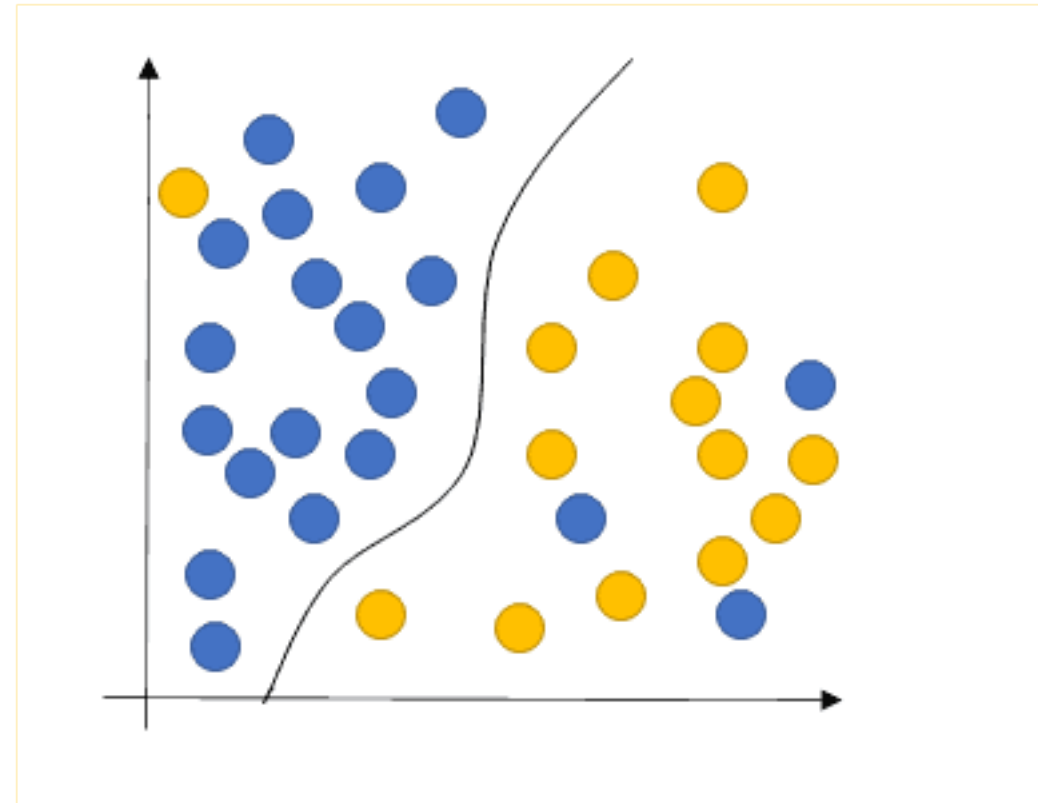
Noten esta clasificación:

- Tenemos observaciones azules, y amarillas dispersas para valores de x , y .
- La máquina puede equivocarse al predecir algunas de estas observaciones.
- Podríamos crear curvas más detalladas.



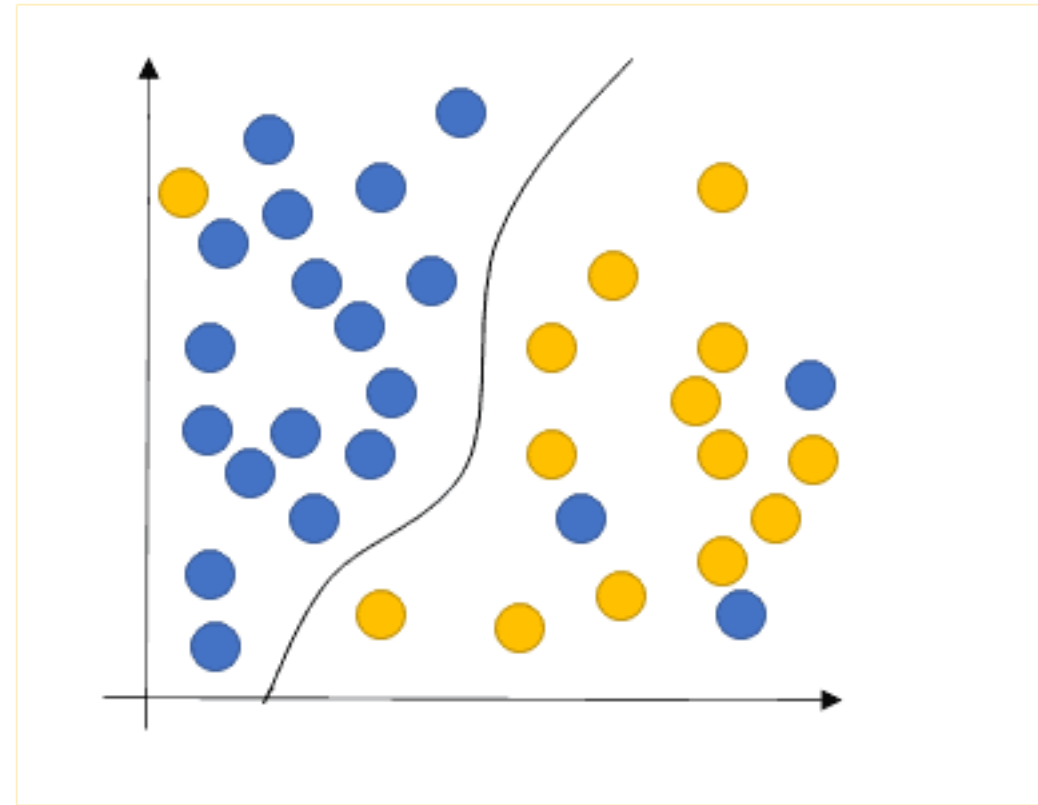
Agrupación: el problema del sobre-ajuste

- Cuando ajustamos un modelo en exceso, su error de predicción en datos conocidos es muy bajo.
- Pero cuando se ponen datos nuevos, falla mucho. ¡Y recuerden que entrenamos estas cosas para usarlas en datos nuevos!



Agrupación: el problema del sobre-ajuste

- Siempre hay que distribuir nuestros datos en grupos: (ej.) *train*, *test*
- Entrenamos nuestros modelos con los datos de entrenamiento (*train*)
- Luego revisamos que los errores de predicción no incrementen mucho en datos nuevos (*test*).



Entrenamiento

- En la etapa de entrenamiento, nuestro modelo prueba diferentes parámetros a partir de los datos que le damos.

¡Recuerden usar sólo los datos *train*!

$$y = 5 + 6x$$

$$y = 5 + 6.1x$$

$$y = 4 + 6.6x$$

...

Evaluación

- Al evaluar queremos ver qué tanto nuestro modelo se equivoca al predecir.
- Con datos conocidos...
- Con datos nuevos...
- Qué tipo de errores comete más...

Veremos una clase sobre esto.

		True Class	
		Positive	Negative
Predicted Class	Positive	TP	FP
	Negative	FN	TN

Manos al código

Veamos un código de ML de principio a fin



Photo by [William Fortunato](#) from [Pexels](#)