

2. Métodos de aprendizaje

DA05 - Introducción a la Inteligencia Artificial

DA - Data Advanced - Data Analytics Journey

DA05 - Introducción a la Inteligencia Artificial

2. Métodos de aprendizaje



- **Objetivos de aprendizaje**

- Conocer los diferentes métodos de aprendizaje que disponen las máquinas para poder replicar el comportamiento humano
- Conocer la familia de aprendizaje supervisado (predictivos) y sus aplicaciones
- Conocer la familia de aprendizaje no supervisado (descriptivos) y sus aplicaciones

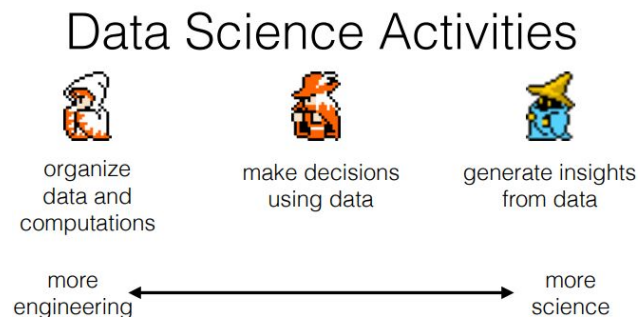


2. Métodos de aprendizaje

2. Métodos de aprendizaje

Ciencia de datos

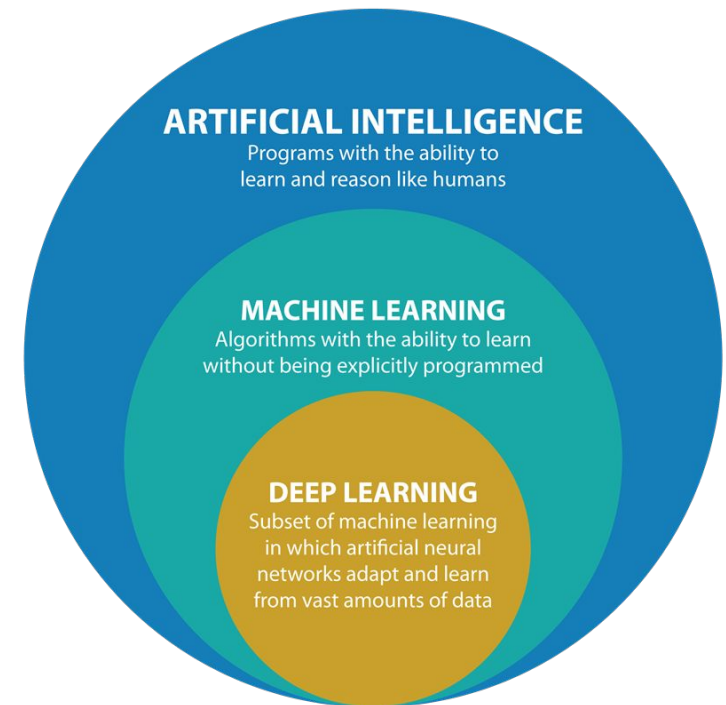
- Una palanca fundamental para la transformación de los modelos de negocio es la **ciencia de datos**
 - Se basa en el uso combinado de técnicas de aprendizaje automático (Machine Learning), inteligencia artificial, matemáticas, estadística, bases de datos y optimización
- El impulso de estas técnicas se debe a cuatro factores:
 - Incremento sin precedentes del volumen y tipología de datos disponibles
 - La conectividad y el acceso al dato
 - Mejora de los algoritmos utilizados
 - Aumento de la capacidad computacional de los sistemas



2. Métodos de aprendizaje

Machine Learning

- El Machine Learning es una rama de la Inteligencia Artificial enfocada al desarrollo de técnicas que permitan el aprendizaje de las máquinas
- Para saber cómo funciona y qué utilidad tiene esta disciplina, a continuación, ilustramos, a través de algunos ejemplos, la importancia del Machine Learning en la productividad de las empresas



2. Métodos de aprendizaje

Machine Learning (II)

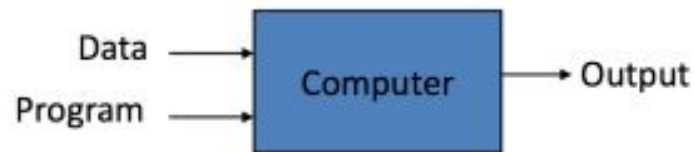
- Son un conjunto de métodos capaces de **detectar automáticamente patrones en los datos**
- El objetivo de esta detección de patrones es tener modelos para representar de manera simplificada esa realidad
- Un **modelo** es, en general, una **función o una estructura de datos**
 - Es una **representación específica hecha a partir de los datos**
- Un **algoritmo (o método)** es una secuencia de pasos con un fin.
 - El algoritmo se encarga de que el modelo aprenda de los datos (se ajuste a los datos)
 - En otras palabras, el algoritmo se encargará de crear esa función matemática para representar esa realidad a partir de los datos

2. Métodos de aprendizaje

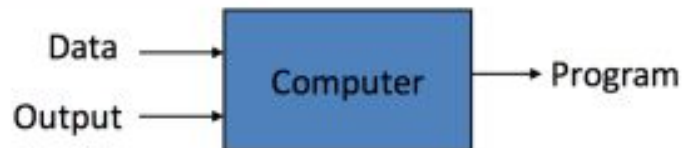
Machine Learning (III)

- **Modelo (entrenado):** resultado de la aplicación de un método sobre unos datos específicos
- El **método (de aprendizaje)** se encargará de dar valores a las variables que forman el modelo
- Por ello, la diferencia fundamental entre la programación tradicional y la de Machine Learning radica en los elementos de entrada y de salida que tiene un modelo y el otro:

Traditional Programming



Machine Learning



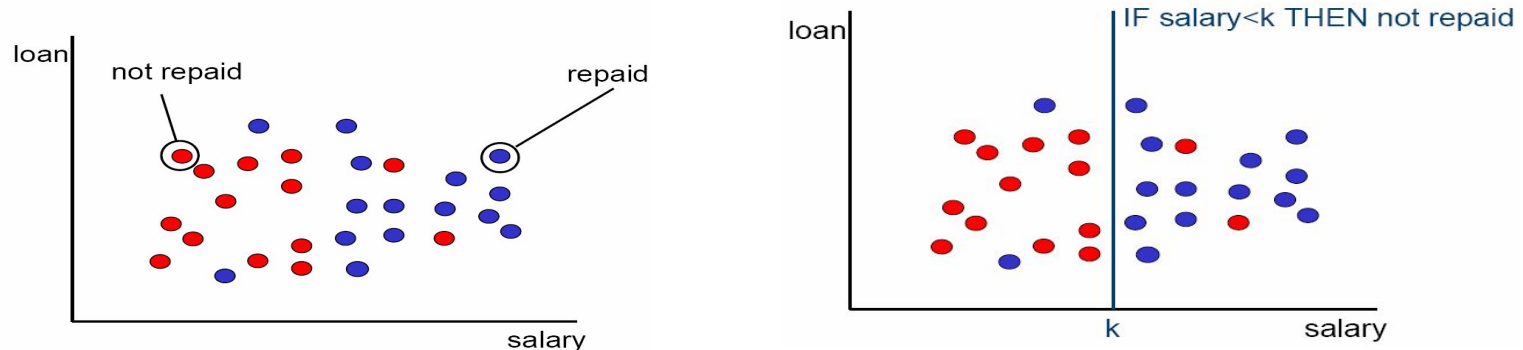
2. Métodos de aprendizaje

Machine Learning (IV)

- **Modelo = Método(Datos)**

- **Riesgo de un crédito**

- Modelo → Una regla: *“If salary < k THEN not repaid”*
- Método → *“Probar 1000 valores aleatorios de k y asignarlo al que mejor resultado dé”*
- Modelo (entrenado) → *“If salary < k THEN not repaid”*
- Siendo k el valor obtenido por el método



2. Métodos de aprendizaje

Machine Learning (V)

- Las técnicas de aprendizaje automático (o Machine Learning) están experimentando un auge sin precedentes
 - Si bien estas técnicas eran conocidas, diversos factores están provocando que su uso sea más intensivo cuando antes era minoritario
- Bajo la definición antes expuesta, el concepto lleva existiendo al menos desde los años 50
 - Es un periodo en el que se descubrieron y redefinieron diversos métodos estadísticos y se aplicaron al aprendizaje automático a través de algoritmos simples
 - Estuvieron circunscritos casi exclusivamente al ámbito académico

2. Métodos de aprendizaje

Machine Learning (VI)

INTERNATIONAL

Winner Of French Scrabble Title Does Not Speak French

July 21, 2015 · 1:00 PM ET

BILL CHAPPELL



3-Minute Listen

"Basically, what he does is, he looks at word lists and looks at dictionary pages... he can conjure up the image of what he has seen. He told me that if he actually hears a word, it doesn't stick in his brain. But if he sees it once, that's enough for him to recall the image of it. I don't know if that's a photographic memory; I just think it's something that his brain chemistry allows him to do."



2. Métodos de aprendizaje

Machine Learning - Componentes

- **Fuentes de información**

- Datos estructurados: bases de datos relacionales, sistemas de ficheros, etc.
- Datos no estructurados: transaccionales, mailing, CRM, voz, imágenes, etc.

- **Técnicas y algoritmos**

- Métodos de aprendizaje no supervisados
- Métodos de aprendizaje supervisados

- **Capacidad de autoaprendizaje**

- Reentrenamiento automático a partir de nueva información
- Combinación de modelos y reponderación/calibración

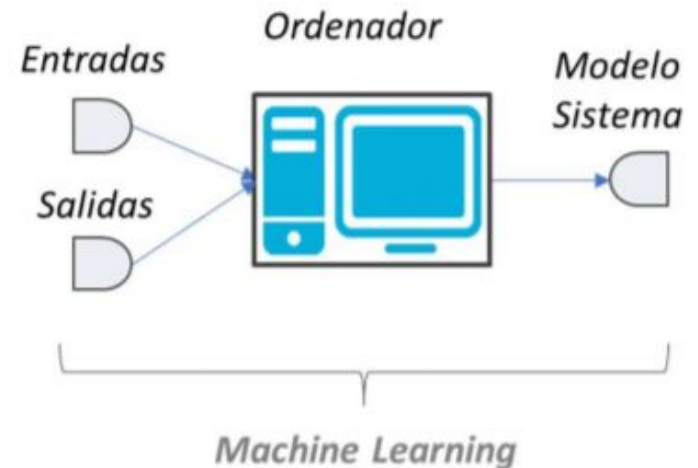
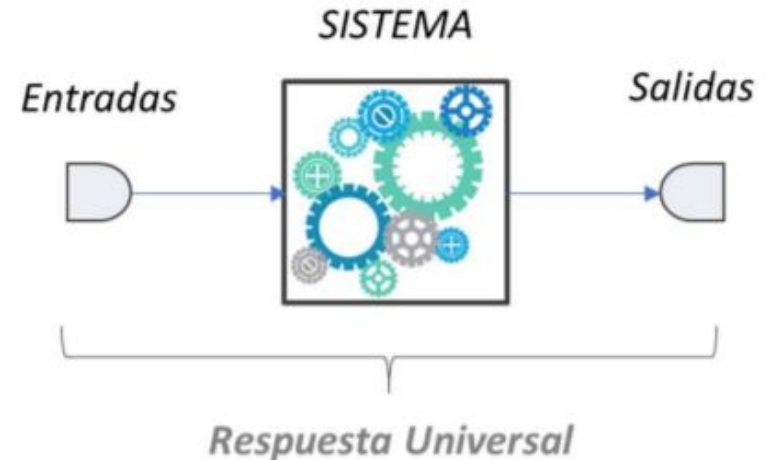
- **Sistemas y software para la visualización de las conclusiones**

- Visualización: Qlikview, Tableau, Power BI, etc.
- Programación: R, Python, Scala, Ruby, SAS, etc.

2. Métodos de aprendizaje

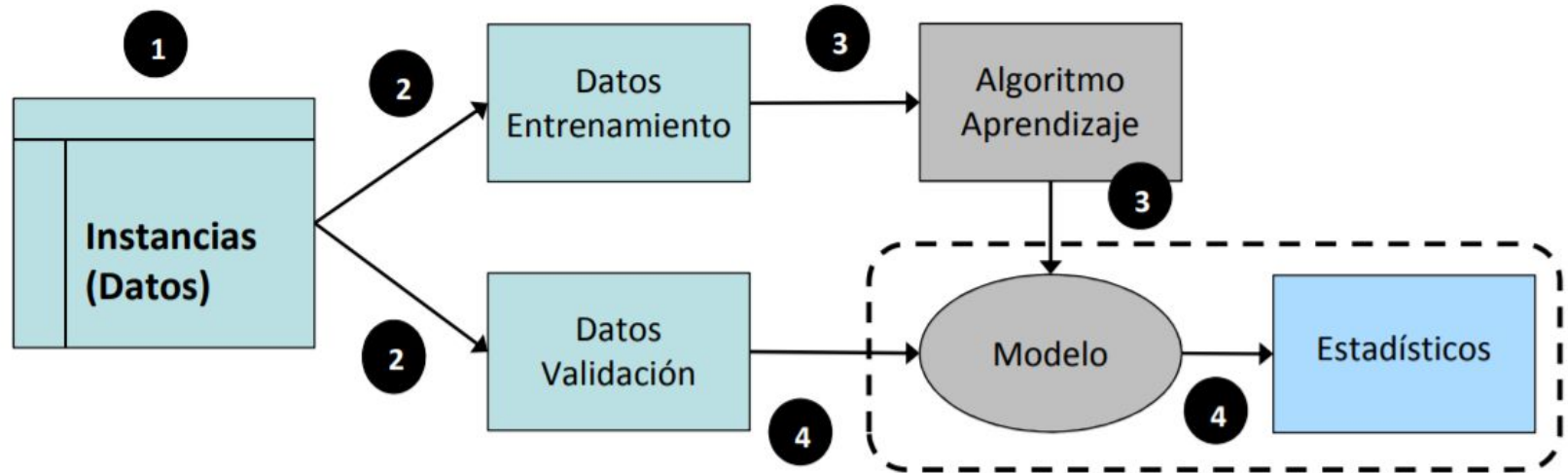
El aprendizaje

- ¿Qué significa **aprender**?
 - No hay una única definición
 - Es un proceso de inducción del conocimiento → A partir de una observación o varias observaciones, se puede reconocer un patrón. Aprender está relacionado con generalizar, con saber reaccionar a casos nuevos.
 - El patrón viene dado en forma de modelo matemático
- ¿Por qué **automático**?
 - Problemas demasiados complicados para resolverlos a mano
 - Grandes volúmenes de datos
 - Alta dimensionalidad → Elevado número de variables implicadas



2. Métodos de aprendizaje

Diagrama de un sistema de aprendizaje

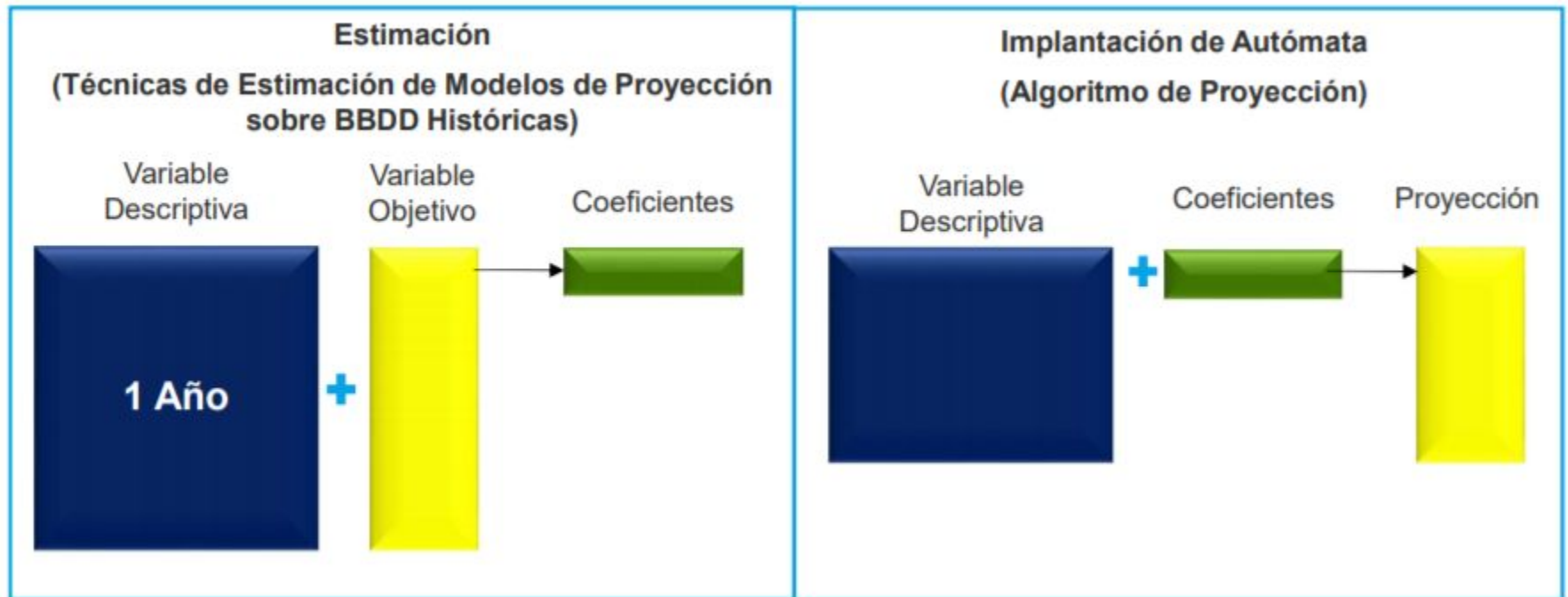


- (1) Los datos iniciales, tras ser preprocesados, se presentan al sistema de aprendizaje, en forma de instancias y atributos.
- (2) El conjunto de instancias se dividen en dos grupos. Los datos de entrenamiento servirán para generar el modelo de conocimiento. Los datos de validación servirán para valorar la calidad/fidelidad del modelo.
- (3) El algoritmo de aprendizaje, genera el patrón que mejor se adecua a las instancias presentadas
- (4) Se valora la capacidad de generalización del modelo, frente a nuevas instancias (datos validación). Se utilizan diferentes estadísticos/métricas.

2. Métodos de aprendizaje

Diagrama de un sistema de aprendizaje (II)

- En general, los trabajos de Machine Learning se basan en detectar patrones de una base de datos histórica a efectos de programar un proceso repetitivo optimizado en base a dicha historia



2. Métodos de aprendizaje

Ejemplo

- **Empresa retail japonesa**

- La temporada de verano en Japón se ha adelantado unos 10 días. El director de Ventas de una empresa retail del país sabe que, durante las primeras semanas del periodo estival, el volumen de venta de las camisetas de manga corta supone el 40% del total. Conoce el impacto que tiene el clima, también el del precio y el de los productos de la competencia. Por eso, decide enviar más mercancía a las tiendas
- El año anterior ocurrió precisamente lo contrario: la temporada de verano se inició con temperaturas bajas, por lo que las ventas de un producto similar a las camisetas de manga corta se atrasaron. El director de Ventas decidió potenciar la venta de camisetas de manga larga con una promoción, de modo que consiguió mejorar sus resultados un 20% a la vez que se incrementó la actividad de los clientes en redes sociales con relación a estas prendas.

2. Métodos de aprendizaje

Ejemplo (II)

- **Planta de fabricación de maquinaria agrícola**

- El ingeniero de una planta de fabricación de maquinaria agrícola reconoce que los nuevos componentes incorporados en el último modelo de uno de sus productos no se comportan de acuerdo con las expectativas de la compañía.
- Cuenta con datos históricos que indican el nivel de fallo de nuevos componentes durante las primeras semanas por tiempo de uso (horas/kilómetros), por piezas de repuesto, clima y geografía. A su vez, conoce las tendencias en términos de costes por máquina y fallos por máquina durante el período de garantía. Desde su experiencia, el nivel de fallos supera el volumen esperado según el tiempo de uso, lo que le lleva a pensar en que está frente a un fallo epidémico emergente.

2. Métodos de aprendizaje

Ejemplo (III)

- En ambos casos, tanto el director de Ventas como el ingeniero, nuestros expertos, contaban con conocimientos generados en base a su experiencia
- Ambos han desarrollado lo que se conoce como “*comportamiento automático*”: utilizar la información histórica almacenada internamente o procedente de fuentes externas y aprender con el fin de tomar decisiones inteligentes
- La característica fundamental de los algoritmos de Machine Learning es su extraordinaria capacidad para extraer patrones que se escapan a la mayoría de métodos aplicados históricamente

2. Métodos de aprendizaje

Metodología: CRISP-DM

1. Conocimiento del Negocio

- Objetivos / Entorno
- Supuestos, restricciones, riesgos y contingencias
- Definir análisis a realizar

2. Conocimiento de los datos

- Recolección y entendimiento de los datos.
- Exploración, evaluación de la calidad

6. Implementación

- Desarrollo de plan de implementación
- Auditoría de implementación
- Mantenimiento

5. Evaluación

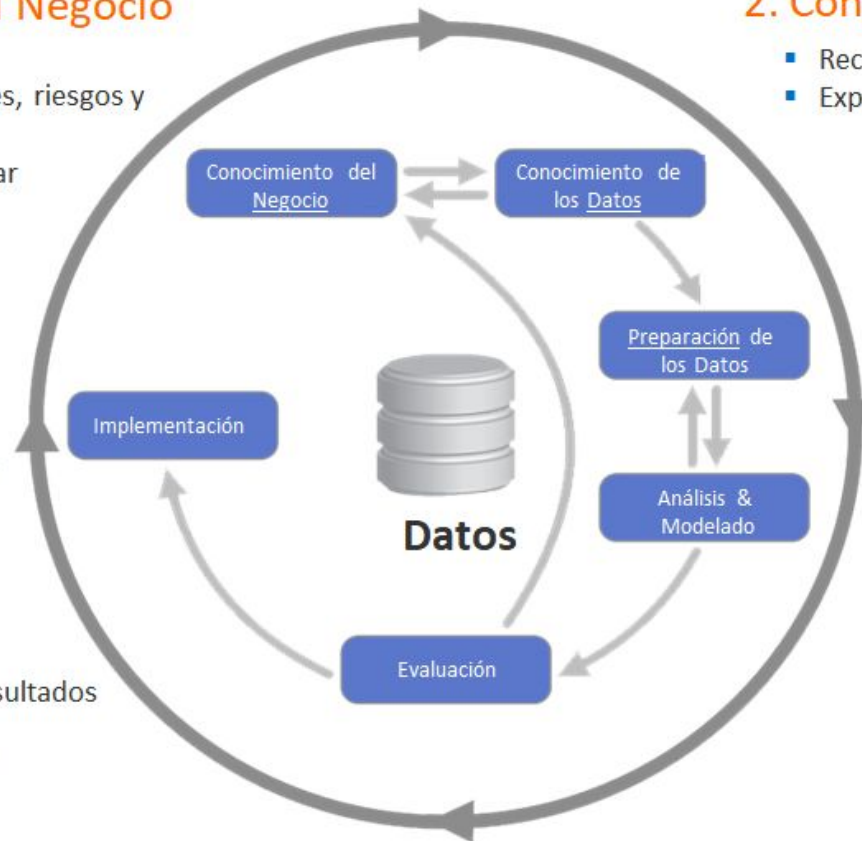
- Evaluar e interpretar resultados
- Auditar el proceso
- Definir próximos pasos

3. Preparación de los datos

- Selección, limpieza, personalización y agregación de los datos

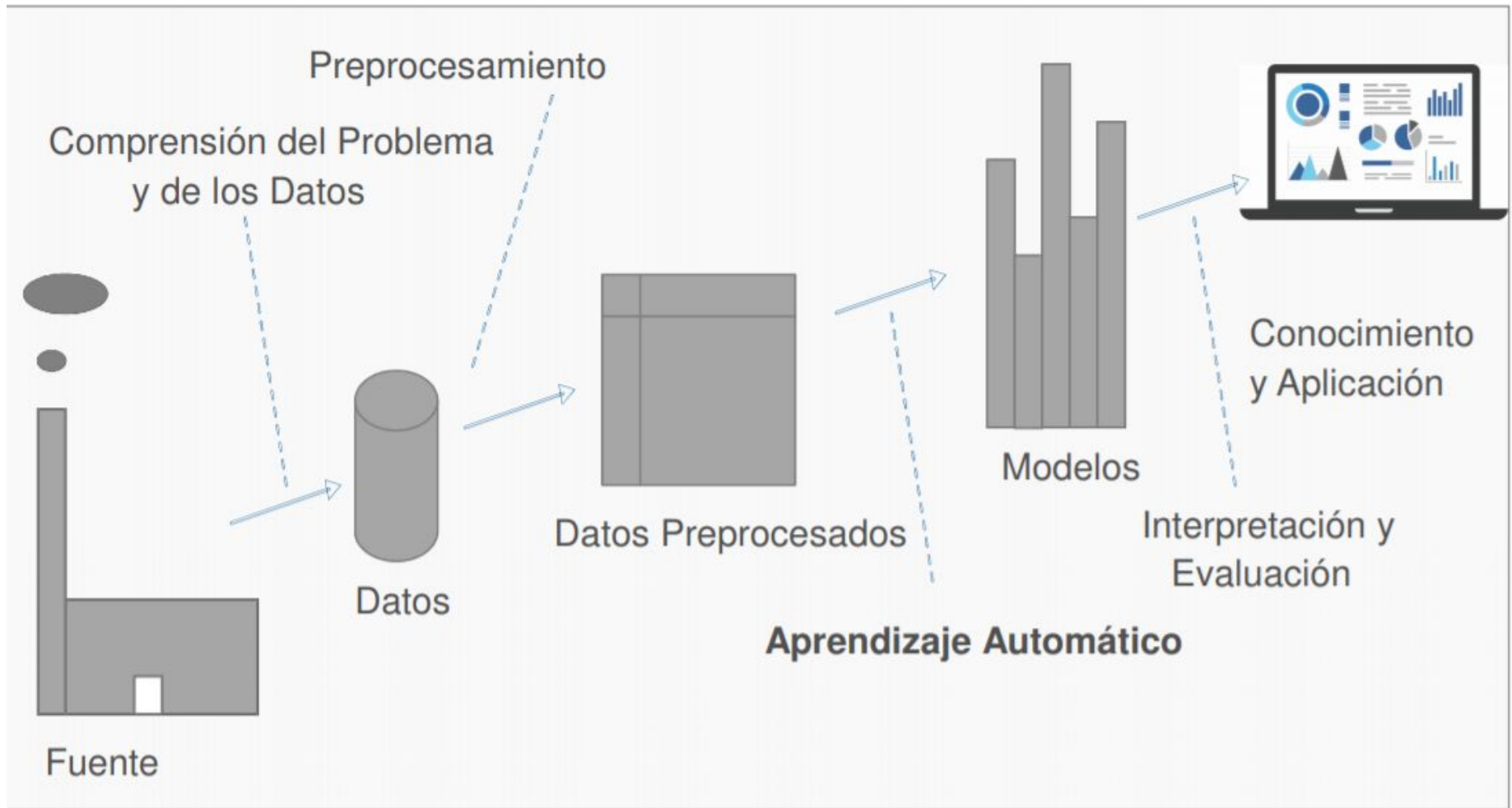
4. Análisis & Modelado

- Definir la técnica más adecuada de Data Mining
- Desarrollo del modelo / solución



2. Métodos de aprendizaje

Metodología CRISP-DM: ¿dónde se sitúa?



2. Métodos de aprendizaje

Metodología CRISP-DM: (3) Preparación de datos

- Las técnicas de Machine Learning (generalmente)
 - Trabajan con una única tabla de datos
 - Trabajan únicamente con los datos que hay en esa tabla
 - No consideran información (por evidente que pueda ser) más allá de los valores almacenados en la tabla
 - Si tenemos en un campo fecha “01/01/2016”
 - No saben que ese día fue viernes, y que el viernes es fin de semana (o no)
 - No saben que ese día es Año Nuevo, y que es festivo
 - No saben que el segundo “01” significa “enero”
- Por eso es importante la preparación de datos para su procesamiento por estas técnicas
 - Por eso hablaremos de **datasets**, en lugar de tablas, bases de datos...
 - Por eso hablaremos de **atributos** en lugar de columnas, campos...
 - Estos atributos pueden extraerse de la base de datos
 - En (muchas) ocasiones deben de calcularse a partir de la información almacenada

2. Métodos de aprendizaje

Metodología CRISP-DM: (3) Preparación de datos (II)

- Son dependientes del problema a abordar
- Deben ser relevantes para la pregunta:
 - Quiero estudiar las ventas por mes
 - “01/01/2016” → Atributo MES = Enero
 - Quiero ver si recibo más compras a primeros de mes
 - “01/01/2016” → Atributo DIA_DEL_MES = 1
 - Quiero ver si hay más ventas días festivos
 - “01/01/2016” → Atributo FESTIVO = TRUE
 - Quiero ver si las ventas se elevan en fines de semana
 - “01/01/2016” → Atributo FIN_DE_SEMANA = FALSE
 - Si tengo datos de ventas, y quiero estudiar clientes, para cada cliente, puedo extraer (entre otros)
 - Número de compras, Antigüedad,...
 - Frecuencia de compra (compras por mes)
 - Compra promedio (gasto promedio)
 - Categoría de productos adquiridos más frecuentemente
 - Preferencias de envío (ordinario o urgente / nacional o internacional)
 - Tipo de cliente (Normal o premium)

2. Métodos de aprendizaje

Metodología CRISP-DM: (3) Preparación de datos (III)

- Deben ser relevantes para la pregunta:
 - Si tengo documentos, puedo extraer (entre otros)
 - Número de palabras, Idioma, Contiene Imágenes?, Temática
 - Cuántas veces aparece el término “Big Data” (o cualquier otro). Ojo:
 - “Big Data” ≠ “big data” ≠ “big-data” ≠ “Big data”
 - Requerirá de una preparación especial (dejar caracteres, minúscula, sin acentos...)
 - “Candidato” ≠ “Candidata”
 - Se puede extraer la raíz “Candidat” → “Candidato” = “Candidata” = “Candidatura”
 - Procedencia del autor
 - Fecha de creación/modificación → antigüedad
 - Si tengo enlaces, puedo extraer (entre otros)
 - Dominio, Activo o Caído, Sitio web, Contiene enlaces...
 - [Todo lo comentado para documentos, sobre el contenido]
 - Si tengo imágenes, puedo extraer (entre otros)
 - Procesando con software especializado → Qué objetos aparecen
 - Sin software especializado
 - Color promedio, mediano, modal... Color promedio, mediano, modal... en determinada zona
 - Porcentaje de rojo, verde, azul, etc.

2. Métodos de aprendizaje

Metodología CRISP-DM: (3) Preparación de datos (IV)

- Famosa frase de Richard Bellman “*The curse of dimensionality*”
 - Durante décadas la metodología para construir modelos predictivos fue evitar la “*maldición de la dimensión*”
 - Si había muchas variables, la receta era buscar unas pocas variables que tuvieran la mayor información y utilizar estas

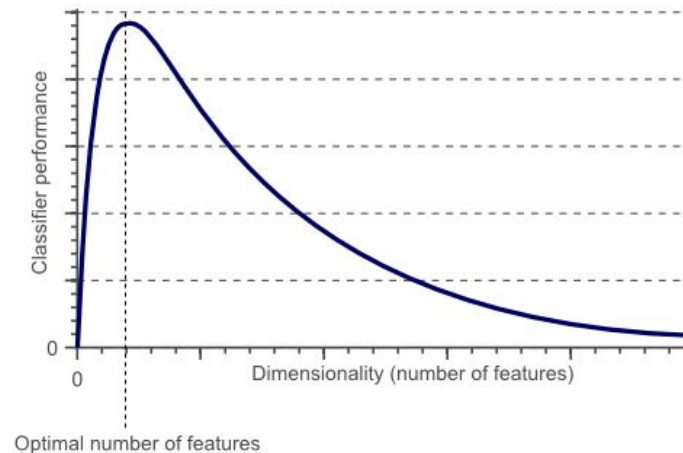


Figure 1. As the dimensionality increases, the classifier's performance increases until the optimal number of features is reached. Further increasing the dimensionality without increasing the number of training samples results in a decrease in classifier performance.

2. Métodos de aprendizaje

Metodología CRISP-DM: (3) Preparación de datos (V)

- La tradición siempre ha sido que "*high dimensionality*" es peligrosa, pero en algunos casos puede ser al contrario
 - Reducir la dimensión también reduce la información disponible para predicción
 - En lugar de reducir el número de variables, también se pueden incrementar estas para solucionar problemas complejos
 - Ejemplo: SVM e hiperplanos.
- ¿Entonces?
 - **Práctica común en el modelado de datos:** Seleccionar variables (por un experto o data-driven) y reducir la dimensión de los datos (PCA, SVD).
 - **Práctica común en modelado algorítmico:** Generar nuevas variables para mejorar la capacidad predictiva y utilizar algoritmos robustos al crecimiento del número de variables. (SVM)

2. Métodos de aprendizaje

Metodología CRISP-DM: (3) Preparación de datos (VI)

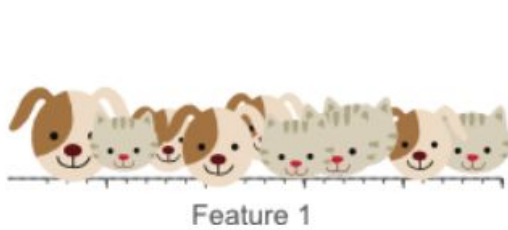


Figure 2. A single feature does not result in a perfect separation of our training data.

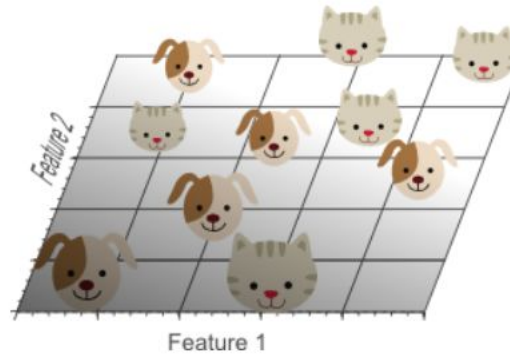


Figure 3. Adding a second feature still does not result in a linearly separable classification problem: No single line can separate all cats from all dogs in this example.

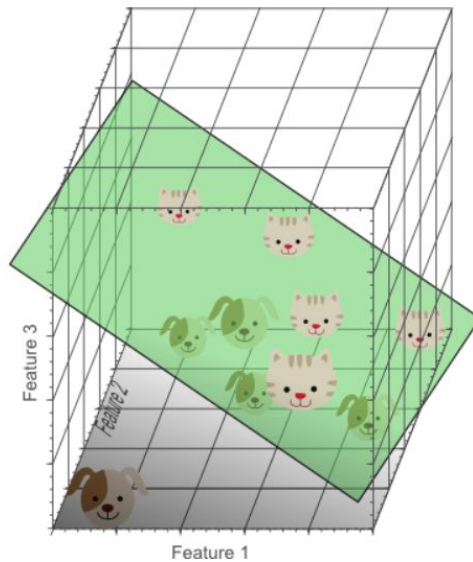


Figure 5. The more features we use, the higher the likelihood that we can successfully separate the classes perfectly.

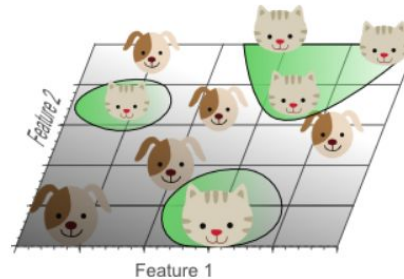


Figure 6. Using too many features results in overfitting. The classifier starts learning exceptions that are specific to the training data and do not generalize well when new data is encountered.

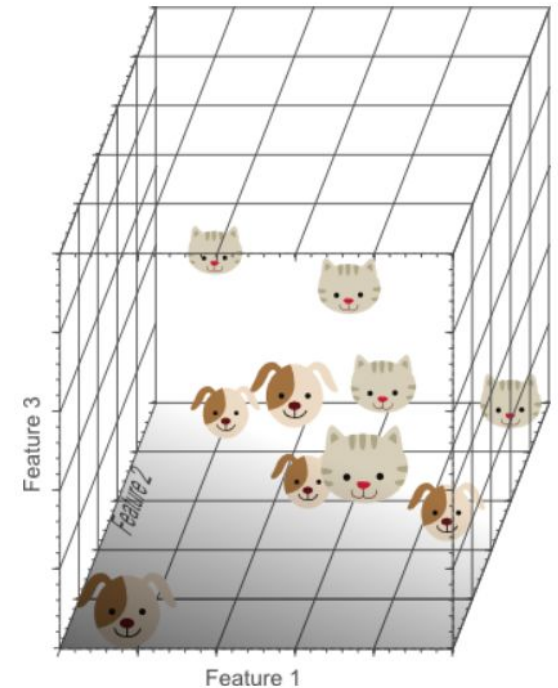


Figure 4. Adding a third feature results in a linearly separable classification problem in our example. A plane exists that perfectly separates dogs from cats.

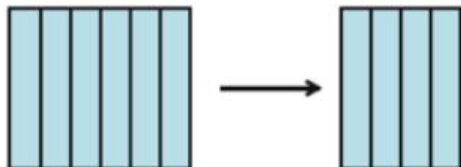
Lectura: <https://www.visiondumy.com/2014/04/curse-dimensionality-affect-classification/>

2. Métodos de aprendizaje

Metodología CRISP-DM: (3) Preparación de datos (VII)

Preprocesado de Datos:

Selección/Compresión de Atributos



Selección de Instancias



Discretización/Numerización de Atributos

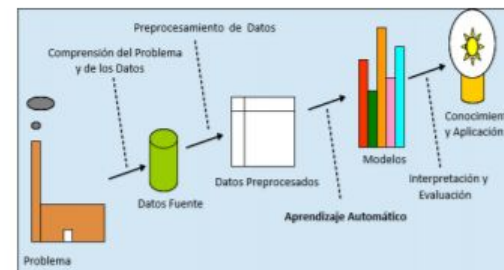
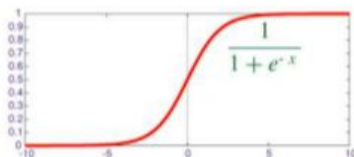


Transformación de Atributos

$\log x$

Normalizar

Peso/Capacidad



Fecha	Temp_1	Temp_2	Temp_3	Altura	Diametro	Esp_Fon	Media_Mecheros
3/4/2016 6:00:00 PM	549.5	547	531	265	107.3	6.458016	0.01625
3/4/2016 7:00:00 PM	549.5	549	533	265	107.3	6.458016	0
3/4/2016 8:00:00 PM	549.5	546	530	265	107.3	6.458016	0.0375
3/4/2016 9:00:00 PM	549.5	545	529	265	107.3	6.458016	0.05
3/4/2016 10:00:00 PM	549.5	544	530	265	107.3	6.458016	0.07
3/4/2016 11:00:00 PM	549.5	547	531	265	107.3	6.458016	0.0025
3/5/2016 12:00:00 AM	549.5	547	532	265	107.3	6.458016	0.01875
3/5/2016 1:00:00 AM	549.5	545	529	265	107.3	6.458016	0.04125
3/5/2016 2:00:00 AM	549.5	545	529	265	107.3	6.458016	0.03375
3/5/2016 3:00:00 AM	549.5	545	529	265	107.3	6.458016	0.0475
3/5/2016 4:00:00 AM	549.5	546	531	265	107.3	6.458016	0.04375
3/5/2016 5:00:00 AM	549.5	547	532	265	107.3	6.458016	0.02125
3/5/2016 6:00:00 AM	549.5	544	530	265	107.3	6.458016	0.0525
3/5/2016 7:00:00 AM	549.5	545	528	265	107.3	6.458016	0.04625
3/5/2016 8:00:00 AM	549.5	544	529	265	107.3	6.458016	0.04375
3/5/2016 9:00:00 AM	549	545	529	265	107.3	6.458016	0.07
3/5/2016 10:00:00 AM	548.5	546	531	265	107.3	6.458016	0.01875
3/5/2016 11:00:00 AM	549.5	547	532	265	107.3	6.458016	0.015

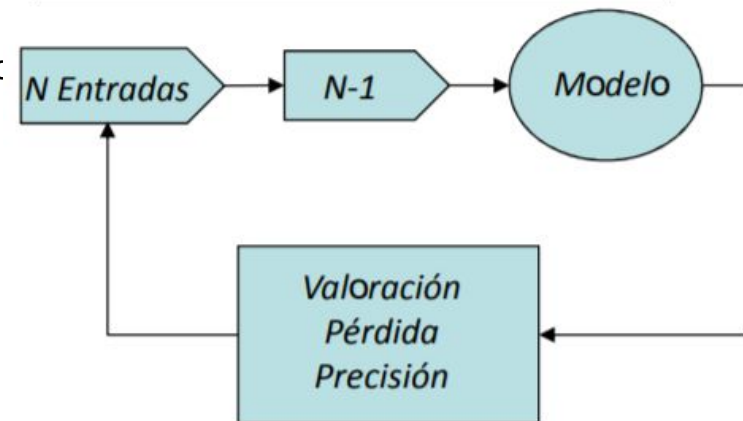
2. Métodos de aprendizaje

Metodología CRISP-DM: (3) Preparación de datos (VIII)

● Preprocesamiento: Selección de Atributos

- Un menor número de atributos implica:
 - Un modelo más simple
 - Facilita el proceso de estimación de parámetros (entrenamiento)
 - Facilita la comprensión del modelo/fenómeno
- Por ello, es muy útil seleccionar el menor subconjunto de variables de entrada que dé una respuesta aceptable
- Un método muy empleado es la selección hacia atrás (reducción iterativa): se parte de un conjunto formado por N variables de entrada
 - Se elimina temporalmente una de las variables
 - Se entrena el modelo y se valida el resultado
 - Si la pérdida de precisión se quita la variable y se vuelve a iterar

Selección entradas: reducción iterativa



2. Métodos de aprendizaje

Metodología CRISP-DM: (3) Preparación de datos (IX)

● Preprocesamiento: Comprensión de Atributos

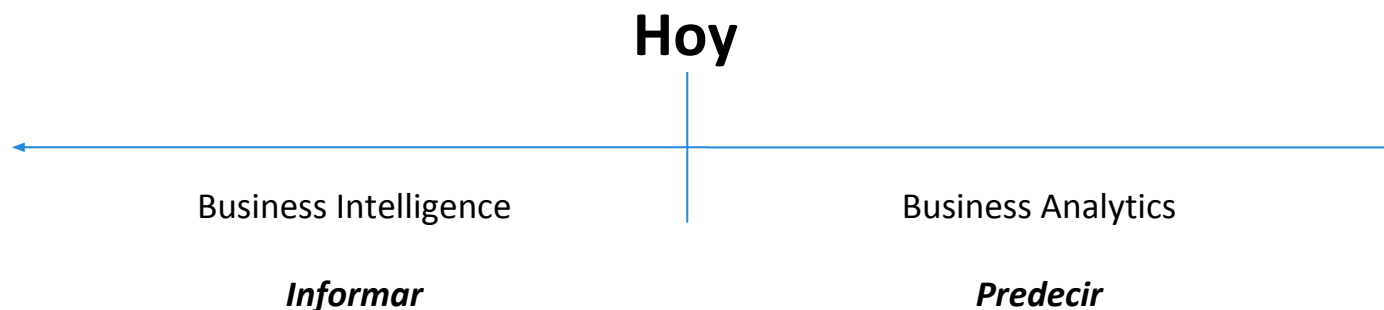
- Los atributos de partida siguen contribuyendo al modelo
- Lo que se hace es comprimir su expresión; es decir, buscar un sistema de referencia de los mismos diferentes, que ocupe menos espacio, que resulte menos “pesado”, y por lo tanto agilice el resultado
- La técnica más empleada es el Análisis de Componentes Principales
 - Ejemplo: Dos variables (entradas al modelo) de partida: altura y diámetro de botella. Se calculan las componentes principales: (PC1 y PC2) para que se cumpla:
 - Mayor varianza en PC1
 - Después mayor varianza en PC2
 - De acuerdo con la figura, se podría representar cada punto (altura, diámetro) sólo con el valor de PC1, sin perder demasiada información se comprime el número de variables iniciales



2. Métodos de aprendizaje

Metodología CRISP-DM: (4) Modelo: las dos culturas

- Hay dos objetivos por los que analizar datos:
 - **Predecir → ¿Qué podría ocurrir?**
 - Ser capaz de predecir qué valores de salida se darán para unos valores de entrada determinados
 - **Informar → ¿Qué ha ocurrido?**
 - Extraer información acerca de cómo el modelo, para unas variables de entrada, produce unas variables de salida
- La forma de conseguir estos dos objetivos varía en función de la cultura empleada
 - (1) La cultura estadística
 - (2) La cultura algorítmica



2. Métodos de aprendizaje

Metodología CRISP-DM: (4) Modelo: las dos culturas (II)

- Es interesante realizar una comparación entre las dos culturas de modelado analítico de datos
 - **(1) Cultura estadística**
 - Asume que hay un proceso estocástico que produce los datos
 - Cultura asociada a la estadística tradicional
 - **(2) Cultura algorítmica**
 - Optimización de una función a través de un algoritmo (software implementado en un conjunto de pasos que siguen unas determinadas reglas)
 - Cultura asociada al **Machine Learning**
 - La comunidad estadística clásica ha estado centrada principalmente en el uso exclusivo de modelos heurísticos (1).
 - El modelado algorítmico en teoría y práctica se ha desarrollado rápidamente en áreas ajenas a la estadística clásica (IA), en lo que se han llamado modelos empíricos


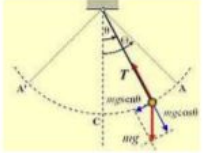






2. Métodos de aprendizaje

Metodología CRISP-DM: (4) Modelo: las dos culturas (III)

- Ambas culturas son complementarias y se apoyan una a otra




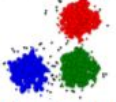
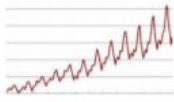


Modelos Heurísticos

Basados en los principios de la Física

	Cinemática	 $\sum F = m \cdot a$ $\frac{dS}{dt} \geq 0$ $P + \frac{1}{2}\rho v^2 + \rho gh = \text{constante}$
	Termodinámica	
	Electromagnetismo	
	Mecánica de fluidos	
	Óptica	
	Acústica	
	física moderna	

Modelos Empíricos

Basados en observaciones directas del fenómeno estudiado

 Regresión	 Clasificación	 Probabilidad	 Clustering
 Series Temporales	 Reglas	 Tasa Fallos	

2. Métodos de aprendizaje

Metodología CRISP-DM: (4) Modelo: las dos culturas (IV)

(Breiman) Statistical Modeling--The Two Cultures.pdf

1 / 33



Statistical Science
2001, Vol. 16, No. 3, 199–231

Statistical Modeling: The Two Cultures

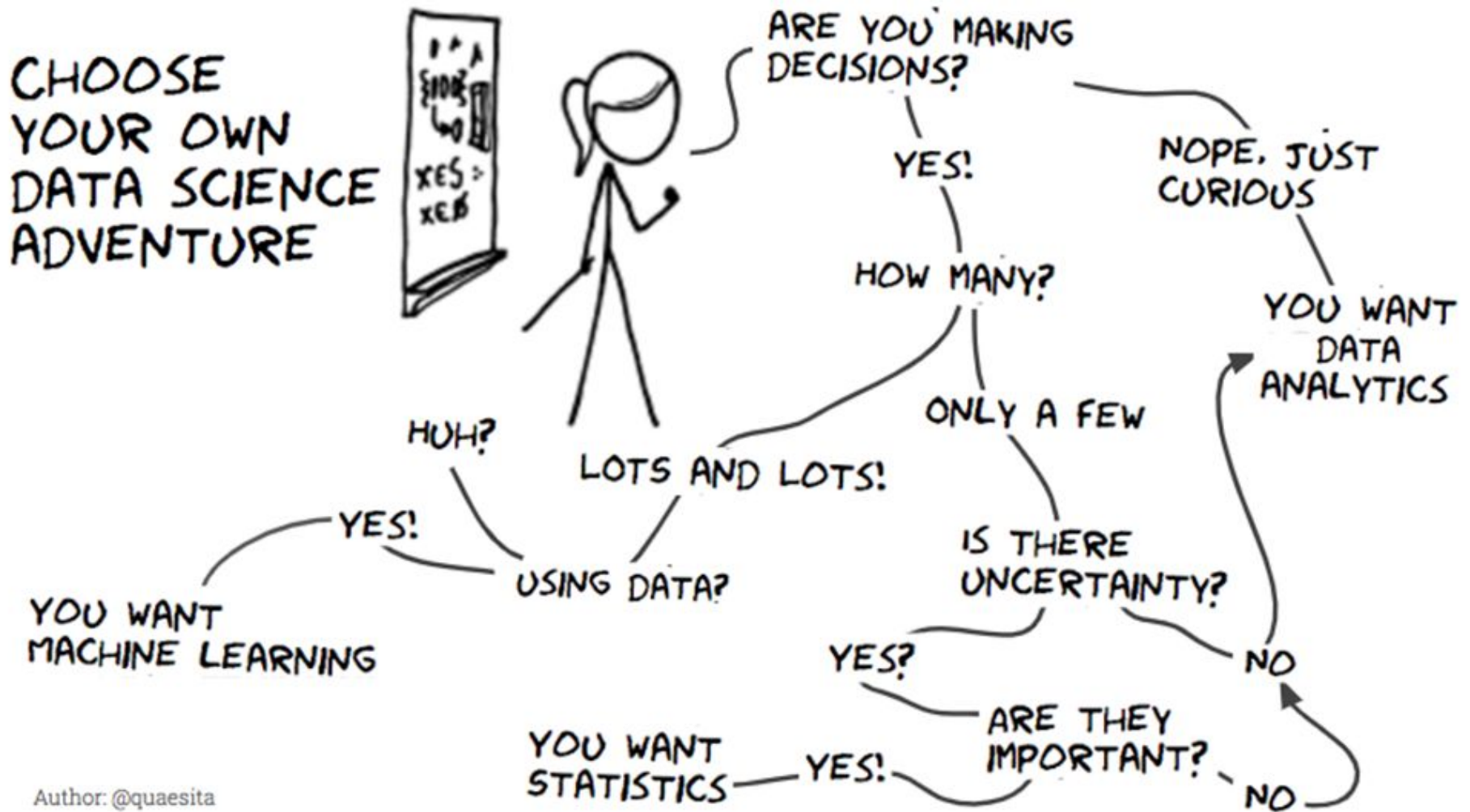
Leo Breiman

Abstract. There are two cultures in the use of statistical modeling to reach conclusions from data. One assumes that the data are generated by a given stochastic data model. The other uses algorithmic models and treats the data mechanism as unknown. The statistical community has been committed to the almost exclusive use of data models. This commitment has led to irrelevant theory, questionable conclusions, and has kept statisticians from working on a large range of interesting current problems. Algorithmic modeling, both in theory and practice, has developed rapidly in fields outside statistics. It can be used both on large complex data sets and as a more accurate and informative alternative to data modeling on smaller data sets. If our goal as a field is to use data to solve problems, then we need to move away from exclusive dependence on data models and adopt a more diverse set of tools.



2. Métodos de aprendizaje

Metodología CRISP-DM: (4) Modelo: las dos culturas (V)



Author: @quaesita

2. Métodos de aprendizaje

Metodología CRISP-DM: (4) Modelo: las dos culturas (VI)

- Cuanto mayor número de variables (dimensiones) tengamos en nuestro problema, más difícil será la separación entre el ruido y la señal
- ¿Qué hacer con el **enfoque estadístico**?
 - **Selección de variables** (los que conocen el dominio o basado en el análisis de datos) o **reducción de la dimensionalidad** (ACP)
- ¿Qué hacer en el **enfoque algorítmico**?
 - Los algoritmos son bastante robustos con muchas variables
 - Es más, a mayor número de variables, más precisos serán → en muchas ocasiones incluso se busca aumentar su número

2. Métodos de aprendizaje

Metodología CRISP-DM: (4) Modelo: las dos culturas (VII)

- Pensemos en una tabla, como Excel, donde hay columnas, filas y celdas...
 - **Columna** → una columna incluye los datos de un determinado tipo. Todos los datos en ella deben de tener la misma escala y tener un significado relativo.
 - Calidad de datos
 - **Fila** → cada fila representa una entidad u observación
 - Las columnas describen las propiedades de esta
 - **Celda** → valor en una determinada fila y columna
 - Booleano, categoría, entero, real, texto, etc.

◇	A	B	C	D
1		Column 1	Column 2	Column 3
2	Row 1	2.2	2.3	1
3	Row 2	2.3	2.6	0
4	Row 3	2.1	2	1

2. Métodos de aprendizaje

Metodología CRISP-DM: (4) Modelo: las dos culturas (VIII)

- Perspectiva estadística

- Variables dependientes
- Variables independientes

◇	A	B	C
1	X1	X2	Y
2	2.2	2.3	1
3	2.3	2.6	0
4	2.1	2	1
5			

- Perspectiva algorítmica

- Fila → Entidad, instancia, ejemplo, etc.
- Columna → Atributo, característica, etc.
- Podemos hablar de atributos de entrada y salida

◇	A	B	C	D
1		Attribute 1	Attribute 2	Output Attribute
2	Instance 1	2.2	2.3	1
3	Instance 2	2.3	2.6	0
4	Instance 3	2.1	2	1
5				

2. Métodos de aprendizaje

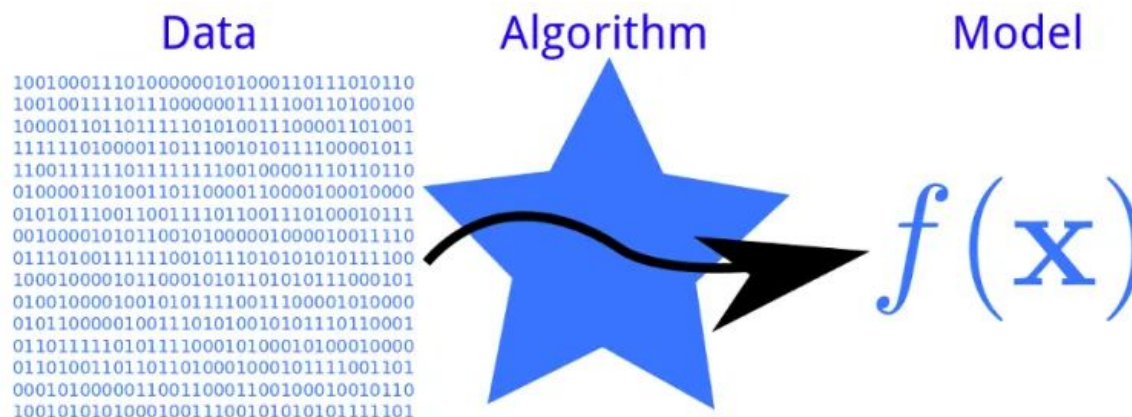
Metodología CRISP-DM: (4) Modelo: las dos culturas (IX)

- **Perspectiva Estadística**

- Variable de Salida = $f(\text{Variables de Entrada})$
- Variable de Salida = $f(\text{Vector de Entrada})$
- Variable dependiente = $f(\text{Variables Independientes})$
- $Y=f(X)$

- **Perspectiva Algorítmica**

- Salida = programa(atributos de entrada)
- Predicción = programa(instancia)

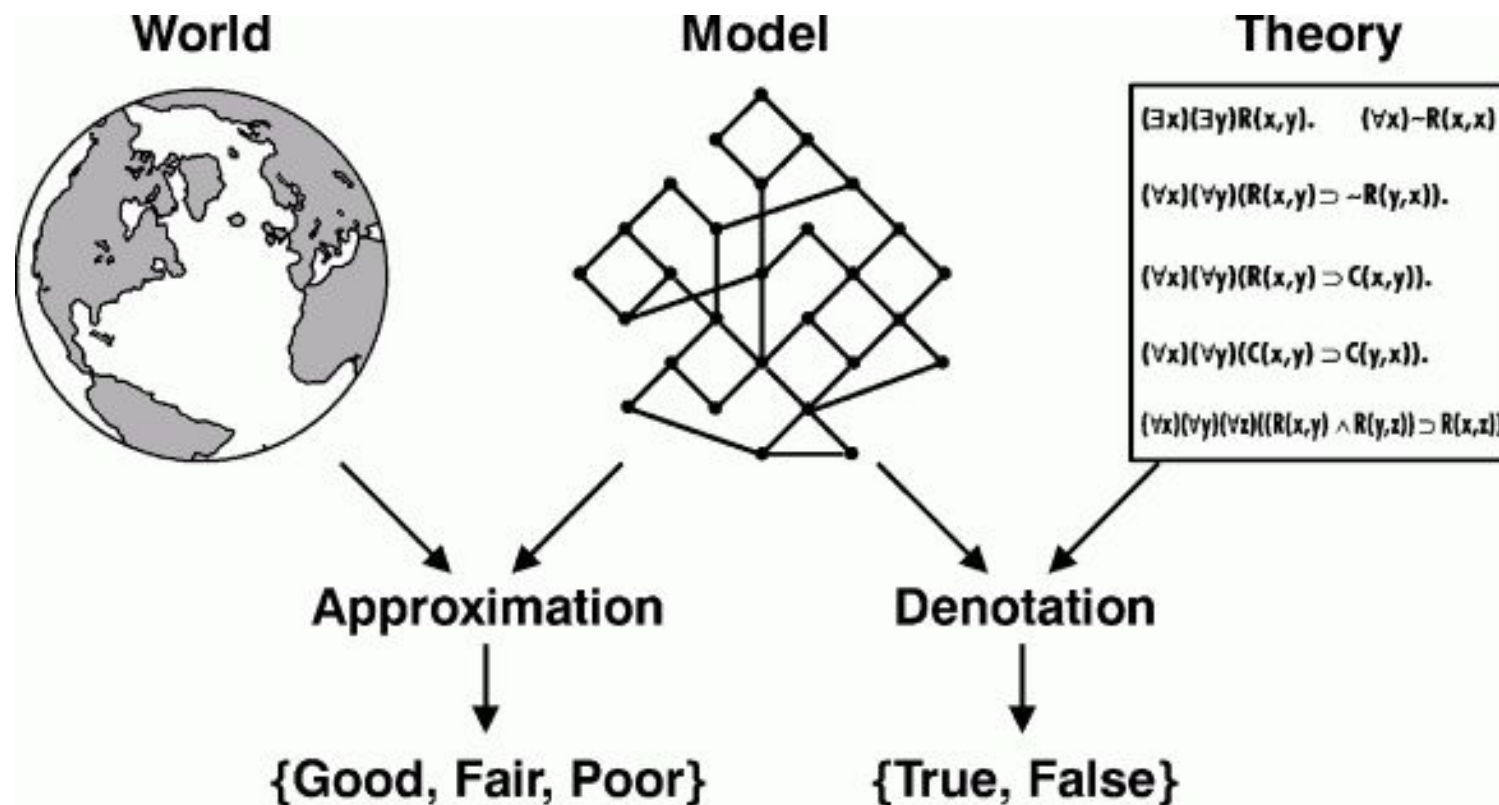


2. Métodos de aprendizaje

Metodología CRISP-DM: (4) Modelo: las dos culturas (X)

Modelos empíricos: algoritmos

Modelos teóricos: estadística



Fuente: <https://www.w3.org/wiki/HttpUrisAreExpensive>

2. Métodos de aprendizaje

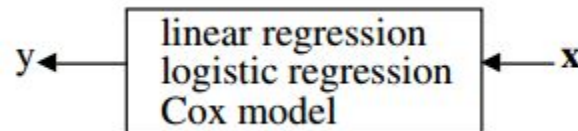
Metodología CRISP-DM: (4) Modelo: Estadística

- Para cualquier persona vinculada con la estadística puede hablarse de dos tipos de estadística, una que podemos denominar
 - **Estadística descriptiva**
 - Primero son los datos y es a partir de estos que se busca manifestar la información relevante para los problemas planteados.
 - **Estadística inferencial**
 - Paradigma central de la estadística → decidir entre varias hipótesis a partir de las consecuencias observadas
 - Incorporar la aleatoriedad dentro de la decisión



2. Métodos de aprendizaje

Metodología CRISP-DM: (4) Modelo: Estadística (II)



variables de salida = f (variables de entrada, ruido aleatorio, parámetros)

- Los valores de los parámetros son estimados a partir de los datos y el modelo empleado para predecir o informar
 - La evaluación del modelo se lleva a cabo utilizando tests de ajuste estadísticos y examinando los residuos.
 - Cultura de utilización: 98% de los estadísticos

2. Métodos de aprendizaje

Metodología CRISP-DM: (4) Modelo: Estadística (III)

● ¿Problemas?

- Las conclusiones las sacamos a partir de los modelos
- Las suposiciones iniciales a veces no se cumplen
- No se produce una evaluación del modelo
 - Los modelos predictivos, por ello, suelen tener problemas de precisión
- Los modelos de datos no se manejan bien en áreas como el reconocimiento de voz o imagen
- Uno nunca sabe hasta qué punto los patrones documentados en los datos son espurios, producto quizás del azar, del prolongado esfuerzo del *data scientist* de corroborar sus preconcepciones
- Con bases de datos amplias, incluso el *data scientist* más perspicaz puede perderse y no descubrir patrones empíricos interesantes
 - Con la llegada del *Big Data* este problema de mero manejo de los datos se hace cada vez más agudo, en especial en aplicaciones industriales

2. Métodos de aprendizaje

Metodología CRISP-DM: (4) Modelo: Estadística (IV)

- Casi todos los artículos empiezan con la siguiente frase:
"Assume that the data are generated by the following model..."
- Seguido de las matemáticas que explican la inferencia, contraste de hipótesis y suposiciones.
 - Estadísticos en investigación aplicada consideran el modelado de datos como la plantilla para el análisis estadístico: cada vez que tienen que solucionar un problema piensan en el modelo que genera los datos.
- Este enfoque tiene su origen en la suposición de que un buen estadístico puede encontrar un modelo que represente un mecanismo complejo de la naturaleza
 - Después se estiman los parámetros de este modelo y se obtienen conclusiones
 - Pero...cuando un modelo se ajusta a los datos para obtener conclusiones cuantitativas, las conclusiones son acerca del mecanismo del modelo, no sobre el mecanismo de la naturaleza.
 - Por tanto... si el modelo es una simulación pobre de la naturaleza, las conclusiones pueden ser erróneas

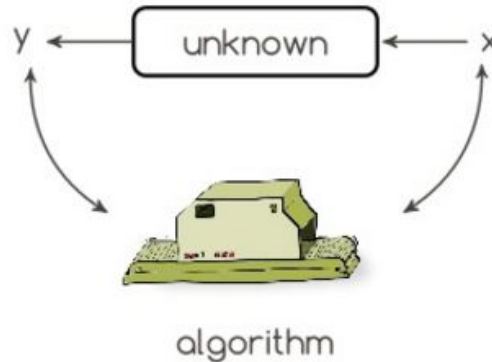
2. Métodos de aprendizaje

Metodología CRISP-DM: (4) Modelo: Estadística (V)

- Los modelos de estadística clásicos se suelen limitar a regresiones logísticas en clasificación y regresiones lineales múltiples en regresión
 - Cuando los datos son generados por observaciones no controladas de sistemas complejos la suposición a-priori de que el proceso genera los datos por un modelo paramétrico puede dar lugar a conclusiones erróneas.
 - Los modelos de datos no deberían ser las únicas herramientas utilizadas para analizar datos.
 - Los modelos de datos fallan en áreas como procesamiento de imágenes y speech recognition.
 - Su objetivo no es la predicción.

2. Métodos de aprendizaje

Metodología CRISP-DM: (4) Modelo: Algoritmos o Machine Learning



- Encontrar una función $f(X)$ que minimice la pérdida $\rightarrow L(Y, f(X))$
- Las mecánicas, la lógica que transforma entradas en salidas, se desconoce \rightarrow y así se asume
 - Por ello, el objetivo se convierte en encontrar un algoritmo que imite lo mejor posible (de ahí lo de minimizar pérdida) esas mecánicas de funcionamiento

2. Métodos de aprendizaje

Metodología CRISP-DM: (4) Modelo: Algoritmos o Machine Learning (II)

- El aprendizaje automático (*Machine Learning*) es un conjunto de algoritmos diseñados para permitir que un ordenador aprenda sobre patrones en los datos
- En vez de especificar una larga lista de atributos de los objetos a estudiar (como hacían los sistemas expertos antiguos), comienzan con:
 - Un simple modelo del mundo para clasificar observaciones
 - Una serie de reglas para modificar tal modelo según el éxito del mismo

2. Métodos de aprendizaje

Metodología CRISP-DM: (4) Modelo: Algoritmos o Machine Learning (III)

- El modelado, así, se resume en un problema de optimización de una función
 - $X \rightarrow$ variable de entrada
 - $Y \rightarrow$ variable de salida
 - $f(X) \rightarrow$ función que minimiza la pérdida para la predicción de la salida
- Por todo ello, se podría decir que la principal diferencia entre el **enfoque estadístico** y el **algorítmico** es que el primero trata de encontrar la verdadera mecánica y el segundo imitarla de la mejor manera posible

2. Métodos de aprendizaje

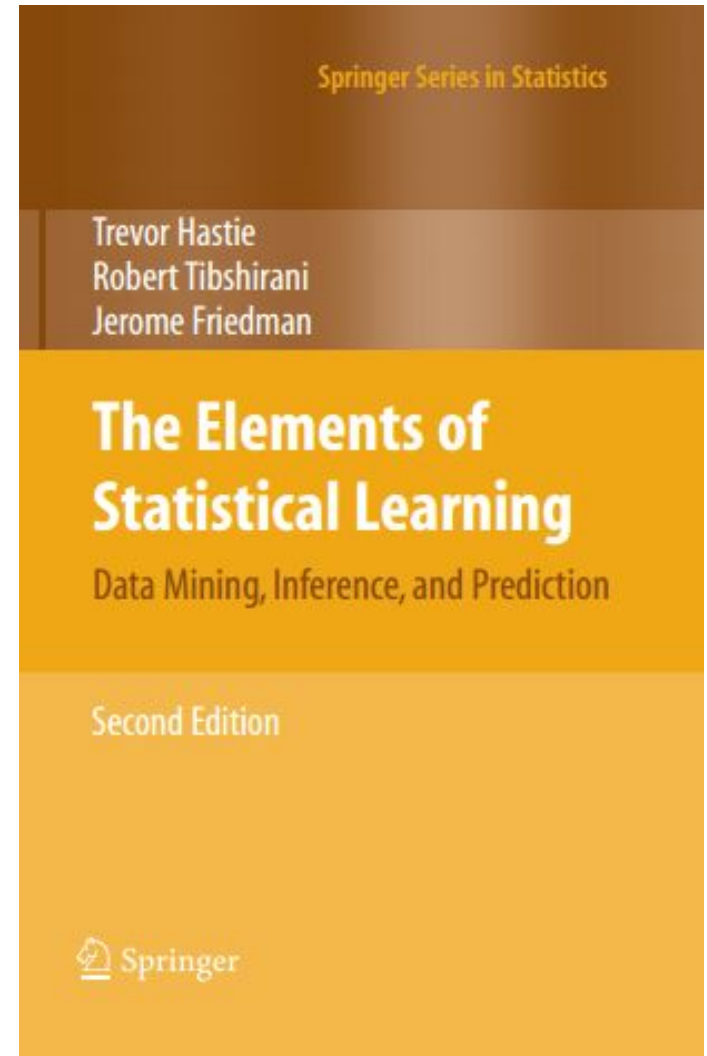
Metodología CRISP-DM: (4) Modelo: Comparación

Estadística	Algoritmos / Machine Learning
Modelo	Red
Observaciones, individuos	Ejemplos (patrones)
Variables	Características, inputs, outputs
Variables explicativas	Inputs
Variables de respuesta	Outputs, targets
Residuos	Errores
Estimación	Entrenamiento, Aprendizaje
Criterio de ajuste	Función de error, Coste
Parámetros	Pesos, coeficientes sinápticos
Regresión, discriminación	Aprendizaje supervisado
Clasificación	Aprendizaje no supervisado

2. Métodos de aprendizaje

Metodología CRISP-DM: (4) Modelo: ¿qué cultura elijo?

- Resumiendo mucho, podemos decir que
 - La **Inteligencia Artificial** ha estado más preocupada en ofrecer soluciones algorítmicas con un coste computacional aceptable
 - La **Estadística** se ha preocupado más del poder de generalización de los resultados obtenidos, esto es, poder inferir los resultados a situaciones más generales que la estudiada



2. Métodos de aprendizaje

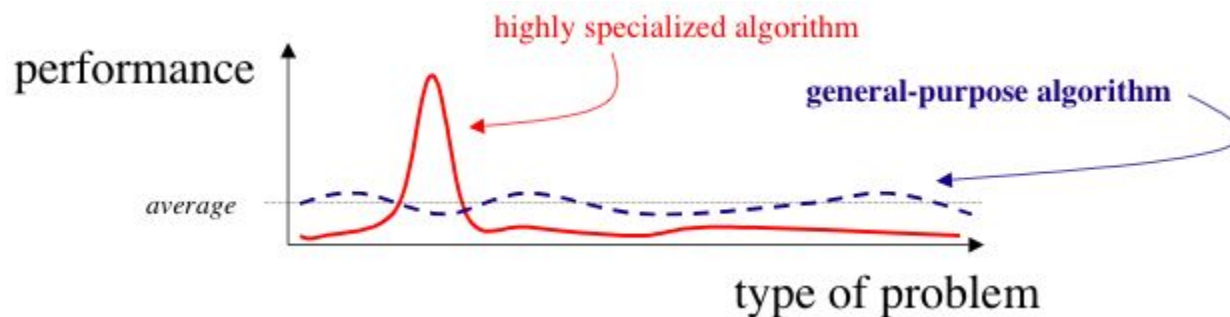
Metodología CRISP-DM: (4) Modelo: ¿qué cultura elijo? (II)

- La mayoría de los usuarios quieren validar sus hipótesis con modelos.
- Quieren modelos que sean fáciles de interpretar en función de sus preguntas.
- Para ellos, es más importante tener modelos que sean fáciles de interpretar que modelos que tengan una buena capacidad predictiva.
- Por ello, muchas veces vamos a necesitar combinar los dos enfoques.
- Uno de los principales problemas del modelado de datos son los cambios frecuentes en la naturaleza de los problemas estadísticos.

2. Métodos de aprendizaje

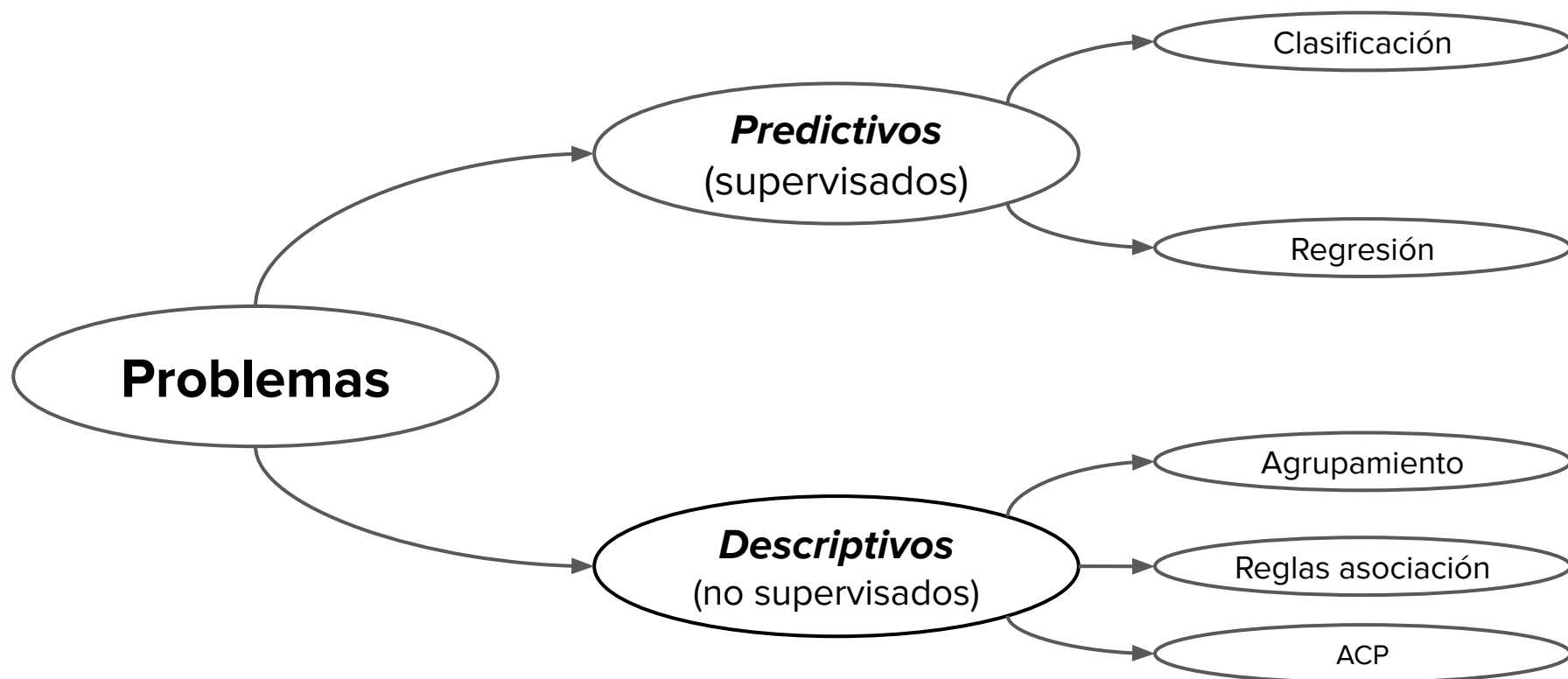
Metodología CRISP-DM: (4) Modelo: Teorema del Free Lunch (III)

- *“No learning algorithm has an inherent superiority over other learning algorithms for all the problems”*
 - *“Any two algorithms are equivalent when their performance is averaged across all possible problems”*
 - El teorema No-Free-Lunch (NFL) nos dice que no existe un algoritmo de aprendizaje universal.
 - Una implicación práctica es que los algoritmos de machine learning no van a funcionar si no existe una estructura previa, unas hipótesis de partida en el espacio de posibles algoritmos, que refleje un conocimiento previo sobre la tareas específica a resolver.

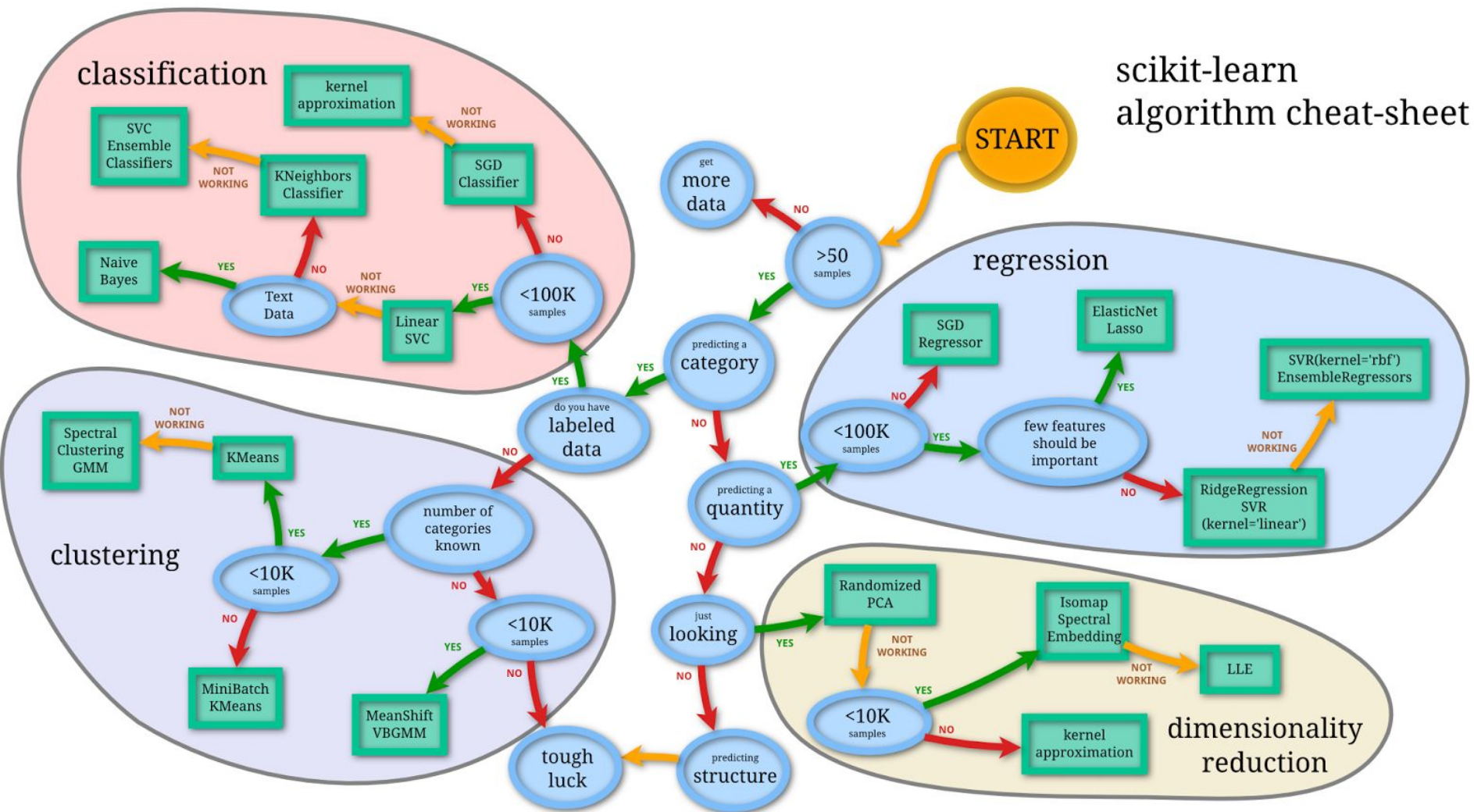


2. Métodos de aprendizaje

Metodología CRISP-DM: (4) Modelo: ¿qué problemas?



Metodología CRISP-DM: (4) Modelo: ¿qué problemas? (II)



2. Métodos de aprendizaje

Metodología CRISP-DM: (4) Modelo: ¿qué problemas? (III)

- **Supervisados (Predictivos)**

- Cuando disponemos del valor que quisiéramos que nuestro modelo diera ante una determinada entrada
- Tenemos datos “etiquetados”
 - Con la clase deseada o Con el valor esperado
- Realizan predicciones del valor de salida a partir de datos

Deuda	Salario	Moroso
100.000	10.000	SI
110.000	30.000	NO
80.000	50.000	NO
90.000	45.000	NO

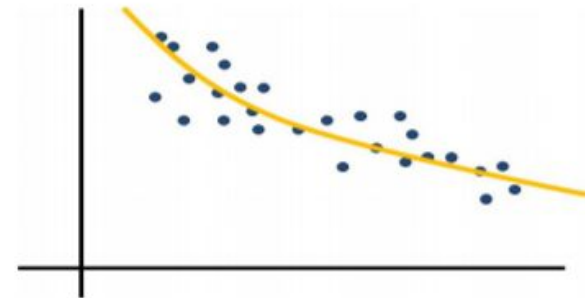
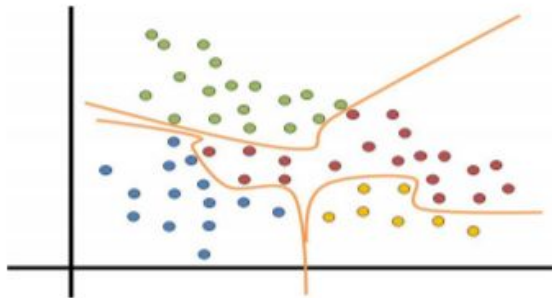
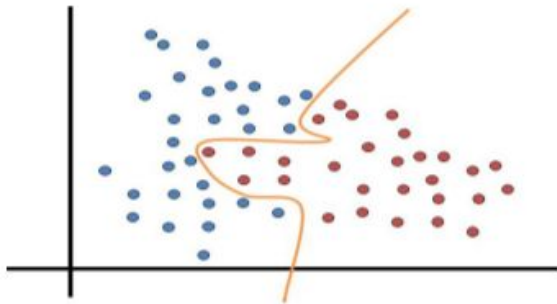
Salario	Edad	Préstamo
10.000	25	100.000
30.000	50	110.000
50.000	45	20.000
45.000	27	250.000

2. Métodos de aprendizaje

Metodología CRISP-DM: (4) Modelo: ¿qué problemas? (IV)

● Supervisados (Predictivos)

- Clasificación: Cuando la variable a predecir es una categoría.
 - Binaria: {Sí, No}, {Azul, Rojo}, {Fuga, No Fuga}...
 - Múltiple: {Comprará Producto1, Producto2...}...
 - Ordenada: {Riesgo Bajo, Medio, Alto}...
- Regresión: Cuando la variable a predecir es una cantidad
 - Precio, cantidad, tiempo, etc.



2. Métodos de aprendizaje

Metodología CRISP-DM: (4) Modelo: ¿qué problemas? (V)

● No Supervisados (Descriptivos)

- Cuando no disponemos del valor que quisiéramos que nuestro modelo diera ante una determinada entrada
- Su objetivo es modelar y describir la estructura o distribución interna de los datos
- Muchas aplicaciones reales hacen uso de estos datos
 - Hay más datos, y son baratos
 - Etiquetarlos puede ser costoso
 - Fáciles de obtener
- **Agrupamiento - Clustering:** Buscan encontrar grupos dentro de los datos de elementos similares → permite así crear perfiles
 - Clientes con hábitos de compra similares
 - Productos vendidos en fechas similares
- **Asociación:** buscan reglas que describen la mayor parte posible de los datos de los que se disponen → permite crear reglas de co-ocurrencia de valores de variables
 - Productos que se compran juntos

2. Métodos de aprendizaje

Algoritmos

- La mejora de los algoritmos ha permitido tanto optimizar el tratamiento de grandes volúmenes de datos (a través de técnicas de escalado, resampling, etc.) como obtener métodos más eficientes y robustos y tratar valores faltantes, variables no numéricas, valores atípicos, etc.
- A pesar de que la mayor parte de los algoritmos fueron desarrollados antes del año 2000, es ahora cuando las empresas están invirtiendo mayores esfuerzos en su implementación, obteniendo mejores resultados que los alcanzados por los humanos

2. Métodos de aprendizaje

Algoritmos (II)

Three Eras of Automation

If this wave of automation seems scarier than previous ones, it's for good reason. As machines encroach on decision making, it's hard to see the higher ground to which humans might move.

ERA ONE 19TH CENTURY

Machines take away the **dirty and dangerous**—industrial equipment, from looms to the cotton gin, relieves humans of onerous manual labor.

ERA TWO 20TH CENTURY

Machines take away the **dull**—automated interfaces, from airline kiosks to call centers, relieve humans of routine service transactions and clerical chores.

ERA THREE 21ST CENTURY

Machines take away **decisions**—intelligent systems, from airfare pricing to IBM's Watson, make better choices than humans, reliably and fast.

SOURCE THOMAS H. DAVENPORT AND JULIA KIRBY
FROM "BEYOND AUTOMATION," JUNE 2015

© HBR.ORG

2. Métodos de aprendizaje

Algoritmos (III)

- La Inteligencia Artificial implica emular la mecánica de aprendizaje → de este modo, es posible utilizar algoritmos de Machine Learning a fin de buscar el algoritmo de aprendizaje que mejor actualice el algoritmo de proyecto
- Son muchos los algoritmos que permiten esto

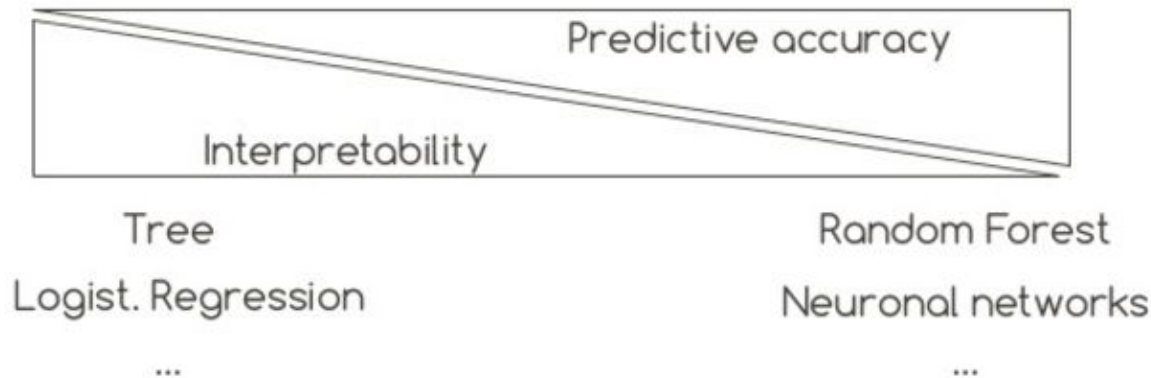
- | | | |
|------------------------------|----------------------|--------------------------|
| • Logistic Regression | • Euclidean Distance | • Bayes |
| • GLM (General Linear Model) | • Kalman Filtering | • Beta Regression |
| • Decision Trees | • Factorial Analysis | • Support Vector Machine |
| • Random Forests | • Neural Networks | • Gradient Boosting |
| • Clustering | • Hopfeld | • Unsupervised Learning |
| • K-Means | • Perceptron | • Steepest Descent |
| | • Genetic Algorithms | • Discriminant Analysis |

2. Métodos de aprendizaje

Algoritmos (IV)

- Algoritmos que nos hacen tener un dilema o trade-off

- ¿Queremos precisión o interpretación?
- ¿O un equilibrio?



- *Ejemplo*: un árbol de decisión vs. un “Random Forest”
 - Un árbol se lee y entiende rápidamente; pero ofrece muchas veces conclusiones débiles
 - Un *Random Forest*, que agrega muchos árboles, tiene una capacidad predictiva muy precisa, pero es difícil de interpretar muchas veces

2. Métodos de aprendizaje

Algoritmos (V)

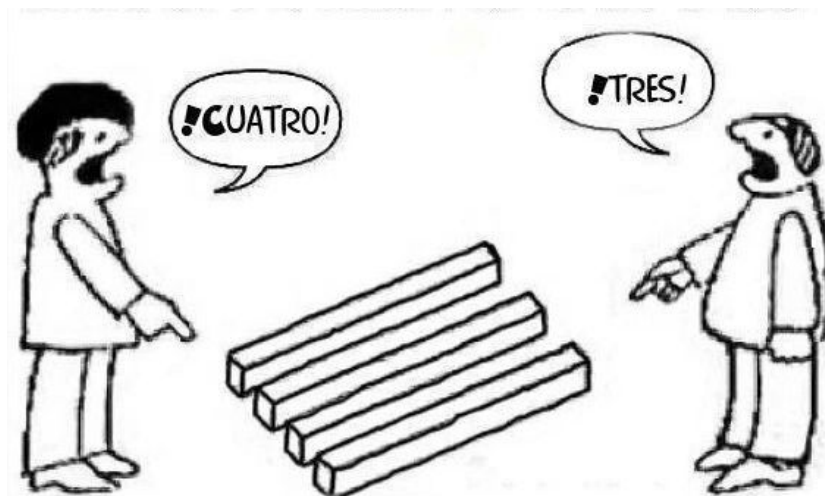
- **Trade off: Precisión vs. Interpretación (simplicidad)**

- La navaja de Occam se suele interpretar como: *menos es mejor*
- Desafortunadamente en predicción, simplicidad (interpretabilidad) y precisión están en conflicto. *Ejemplo: regresión lineal vs redes neuronales.*
- Los árboles de decisión por ejemplo tienen muy buena interpretación pero la precisión es menor que la de los random forests.
- Dilema de Occam: *"La precisión generalmente requiere modelos de predicción más complejos. Funciones simples e interpretables no suelen hacer los modelos predictivos más precisos".*

2. Métodos de aprendizaje

Elección de un algoritmo u otro

- Rashomon es un film japonés en el que 4 testigos cuentan diferentes versiones de un accidente ajustadas a los hechos pero contradictorias entre sí.
 - Existen siempre una multitud de ecuaciones distintas para un problema determinado que proporcionan ratios de errores mínimos similares.
 - Suele haber múltiples modelos diferentes que describen una situación con la misma precisión
- De ahí toma el nombre “*Efecto Rashomon*”

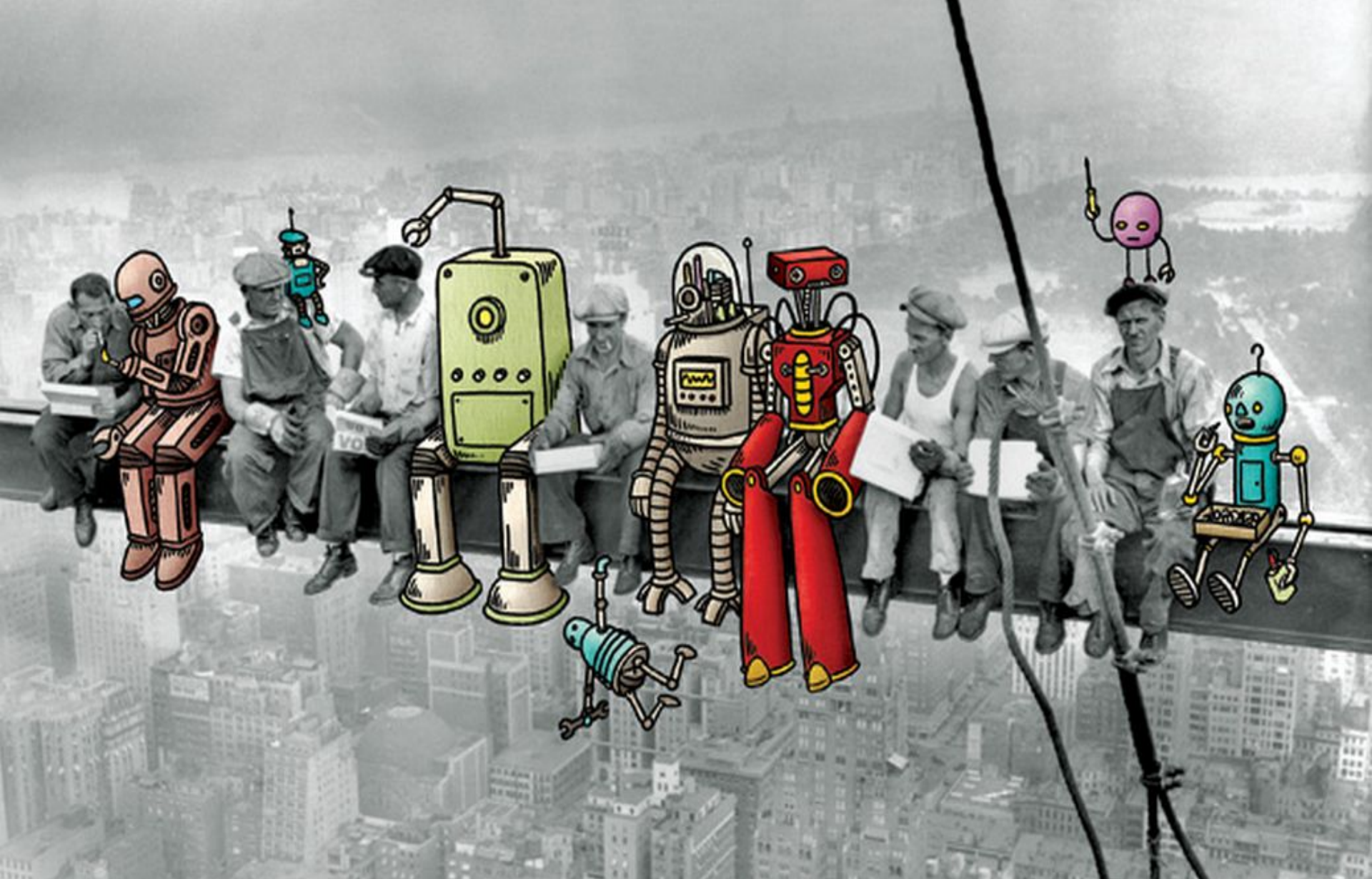




I believe that at the end of the century
the use of words and general educated
opinion will have altered so much that
one will be able to speak of machines
thinking without expecting to be
contradicted.

— *Alan Turing* —

AZ QUOTES



2. Métodos de aprendizaje

DA05 - Introducción a la Inteligencia Artificial

DA - Data Advanced - Data Analytics Journey