

Machine Learning aplicado a los negocios

Sergio Hernán Valenzuela Cámara

sergio@tecnopolis.ai



Tecnópolis . Ai

Febrero, 2021

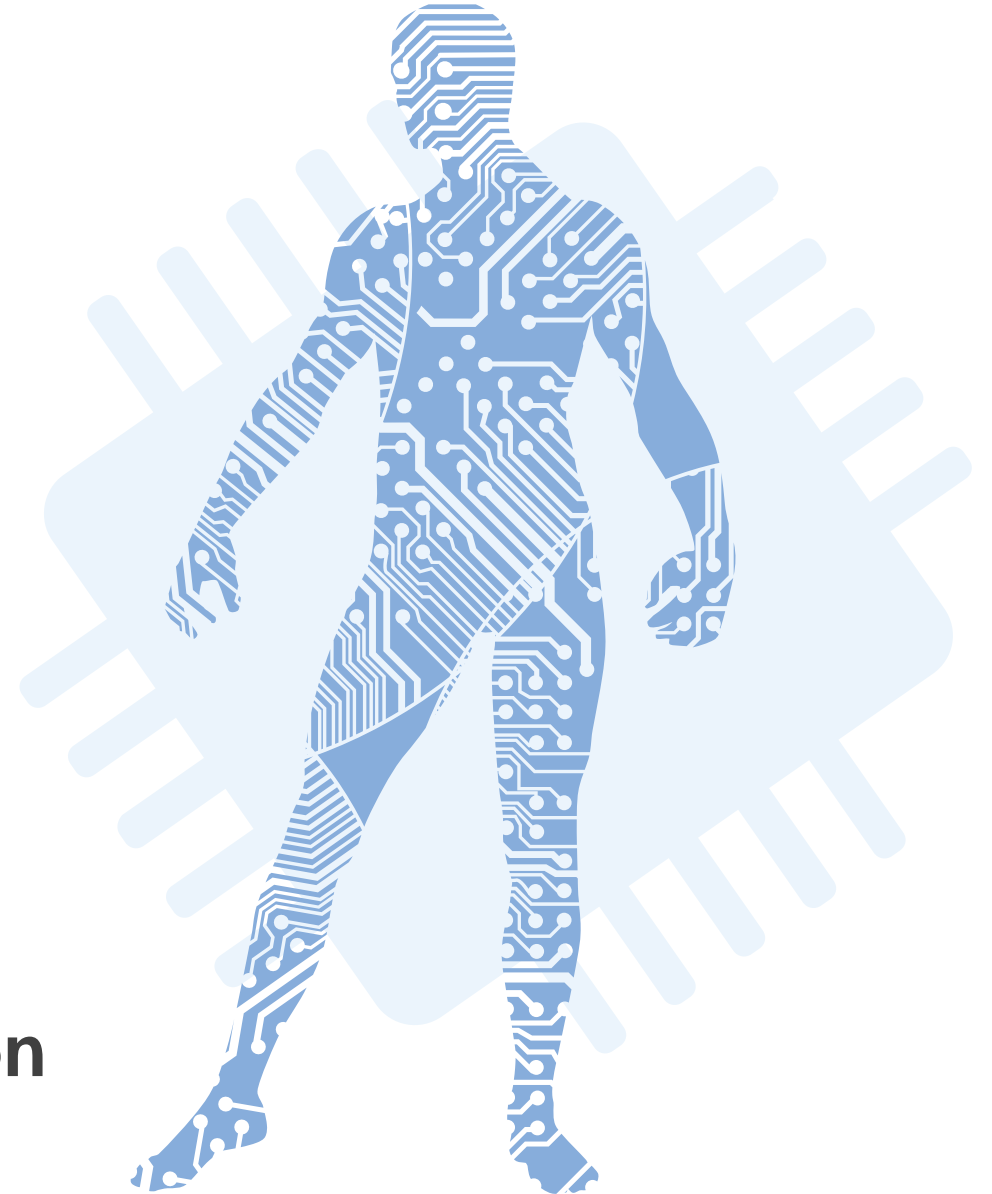
Temario

01 Entendiendo la IA

02 Construyendo IA

03 Creando valor con IA

04 Sistema de Recomendación



Entendiendo la INTELIGENCIA ARTIFICIAL

Capítulo 01

¿Cómo definimos la inteligencia?



Aprendizaje

Tener la capacidad de obtener y procesar nueva información.



Razonamiento

Ser capaz de manipular la información de diversas formas.



Comprensión

Considerar el resultado de la manipulación de la información.



Validar información

Determinar la validez de la información manipulada.



Ver relaciones

Adivinar cómo los datos validados interactúan con otros datos.



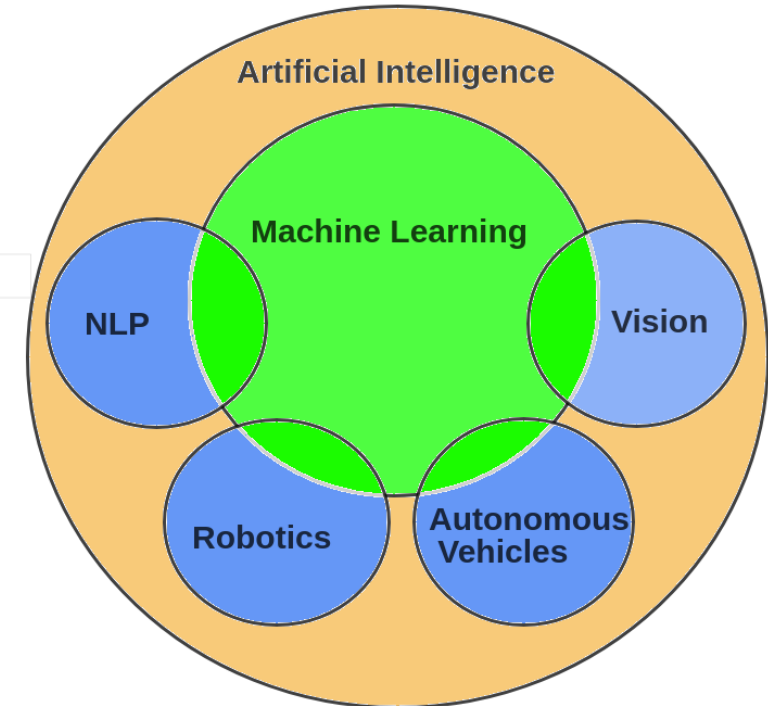
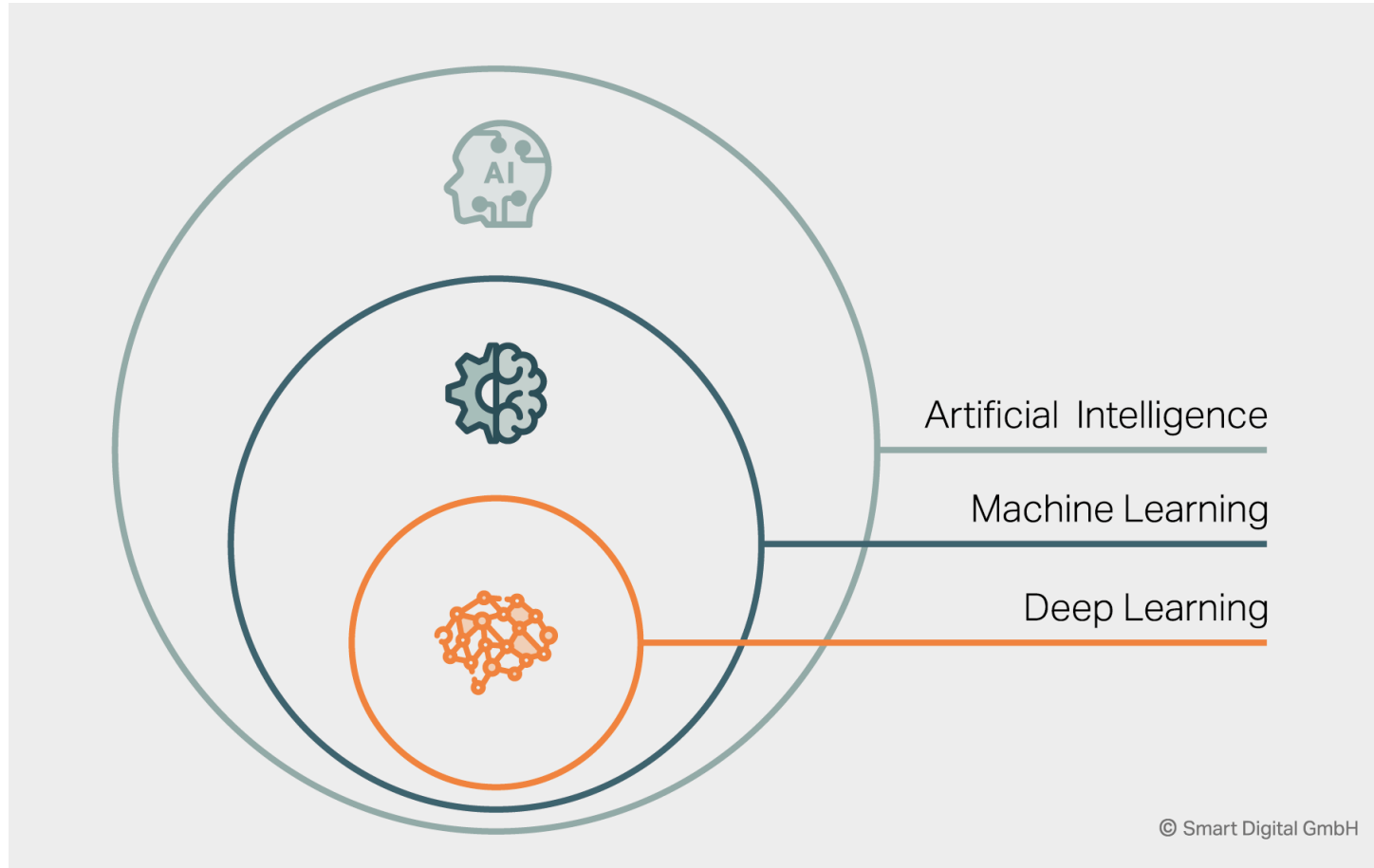
Considerar significados

Aplicar verdades a situaciones particulares de una manera consistente con su relación



¿ Qué es la
Inteligencia
Artificial ?

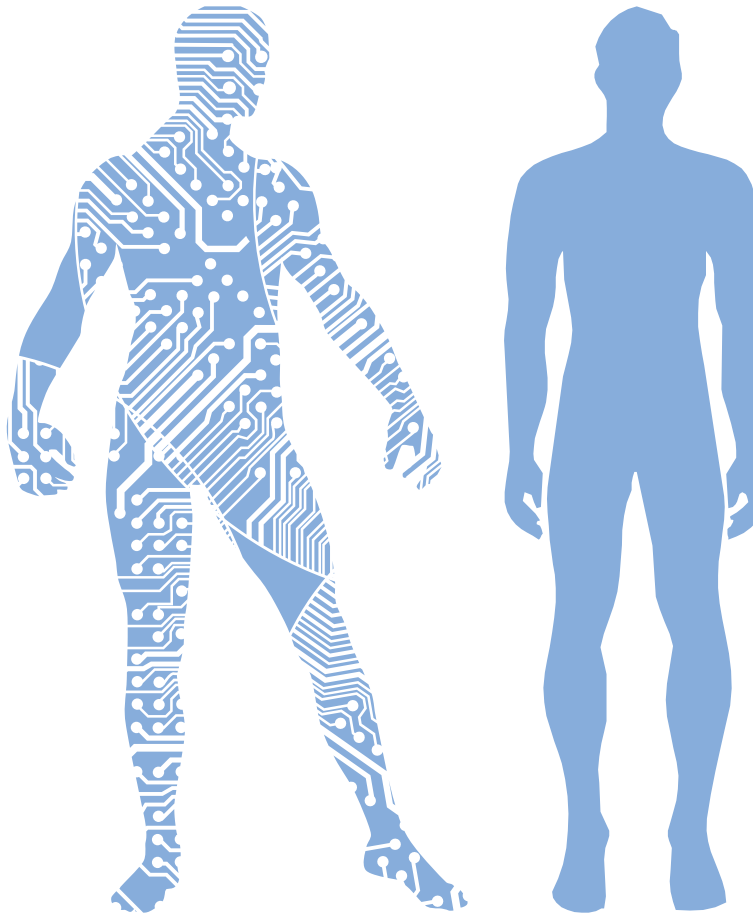
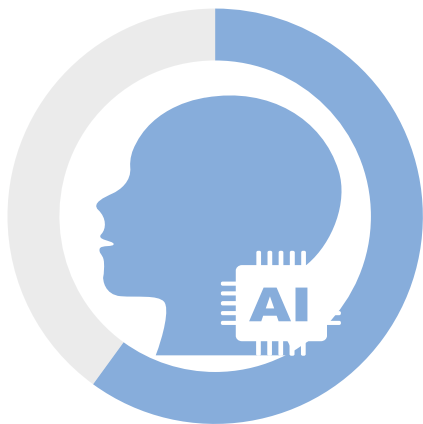
Inteligencia Artificial, Machine Learning y Deep Learning



AI Inteligencia Artificial

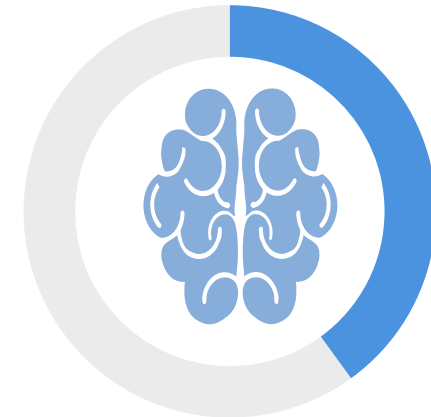
Definición concisa

El esfuerzo por automatizar las tareas intelectuales que normalmente realizan los humanos.



Definición mas simple

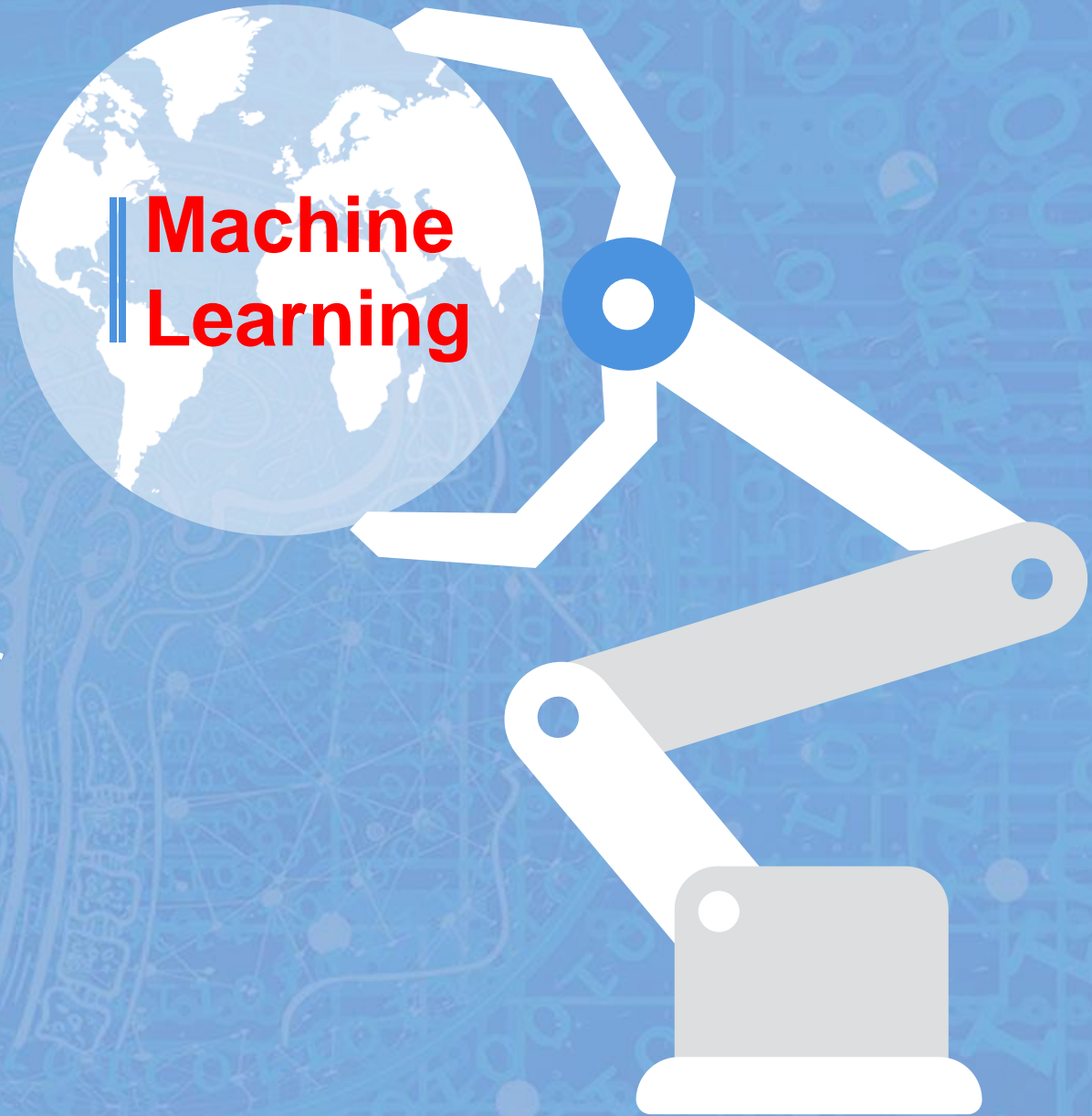
Resolver problemas como lo hacen los seres humanos.



Surgimiento ...

Alan Turing en su histórico artículo de 1950 “Maquinaria de cómputo e inteligencia” que presentó la prueba de Turing, así como los conceptos clave que darían forma a la IA. Turing citaba a Ada Lovelace mientras reflexionaba si las computadoras de uso general podrían ser capaz de aprender y originalmente llegó a la conclusión de que sí podían.

A. M. Turing, “Computing Machinery and Intelligence,” *Mind* 59, no. 236 (1950): 433-460.





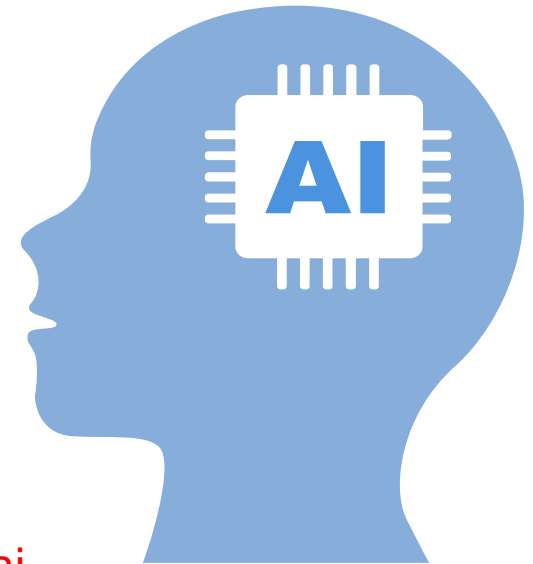
¿ Podría un
computador
ir más allá de
“lo que
nosotros
sabemos
como
ordenar que
se realice” y
aprender por
sí mismo
cómo realizar
una tarea
específica ?

El Machine Learning (Aprendizaje Automático) surge de la siguiente pregunta ...

¿puede un computador sorprendernos?

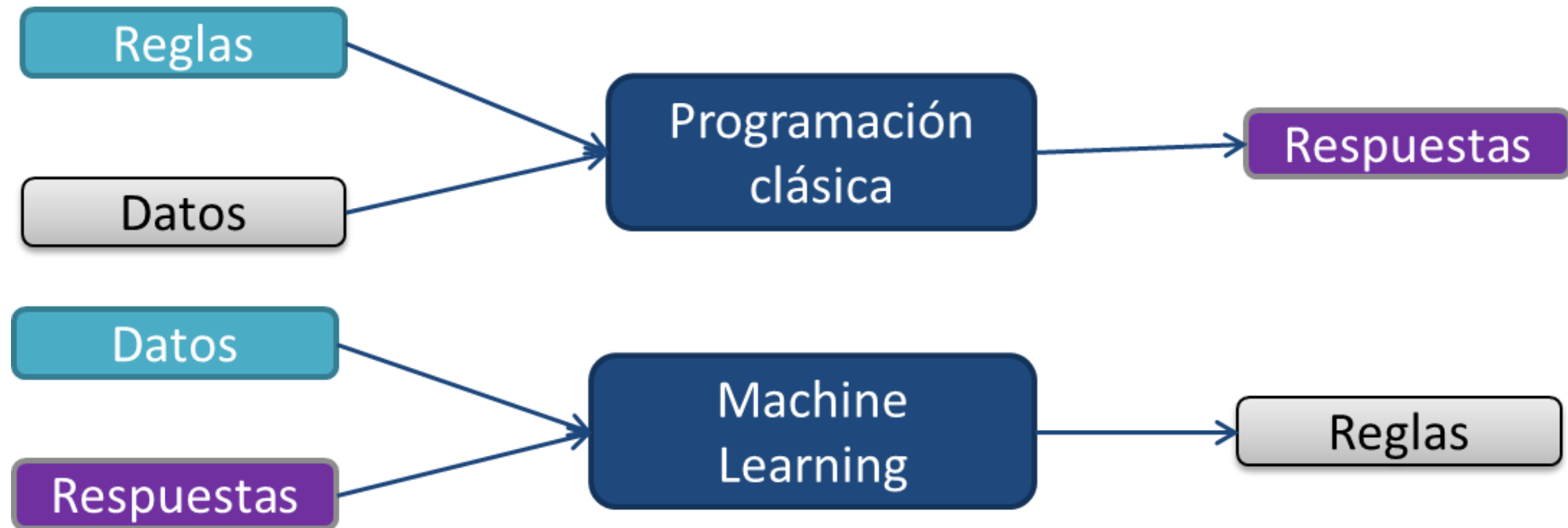
En vez de que los programadores
elaboren el procesamiento de
datos reglas a mano ...

¿podría un computador aprender estas reglas automáticamente al observar los datos?



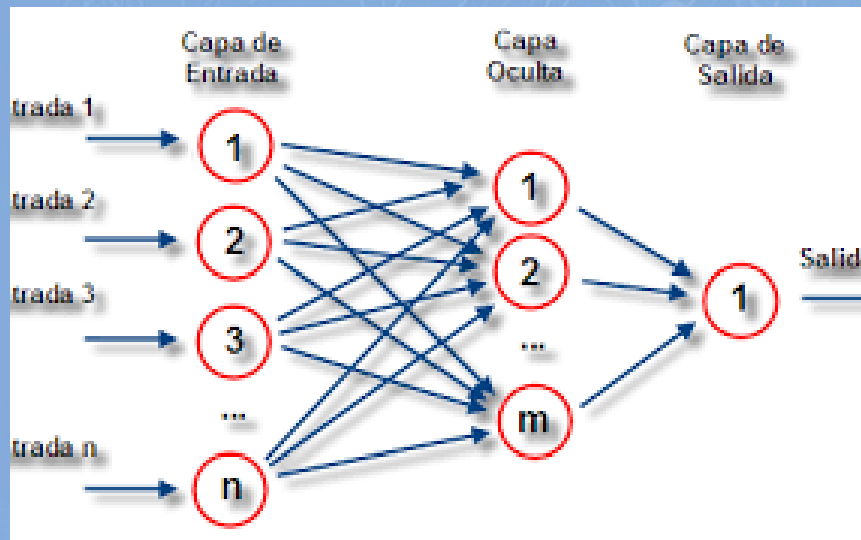
Machine Learning

El nuevo paradigma en la programación

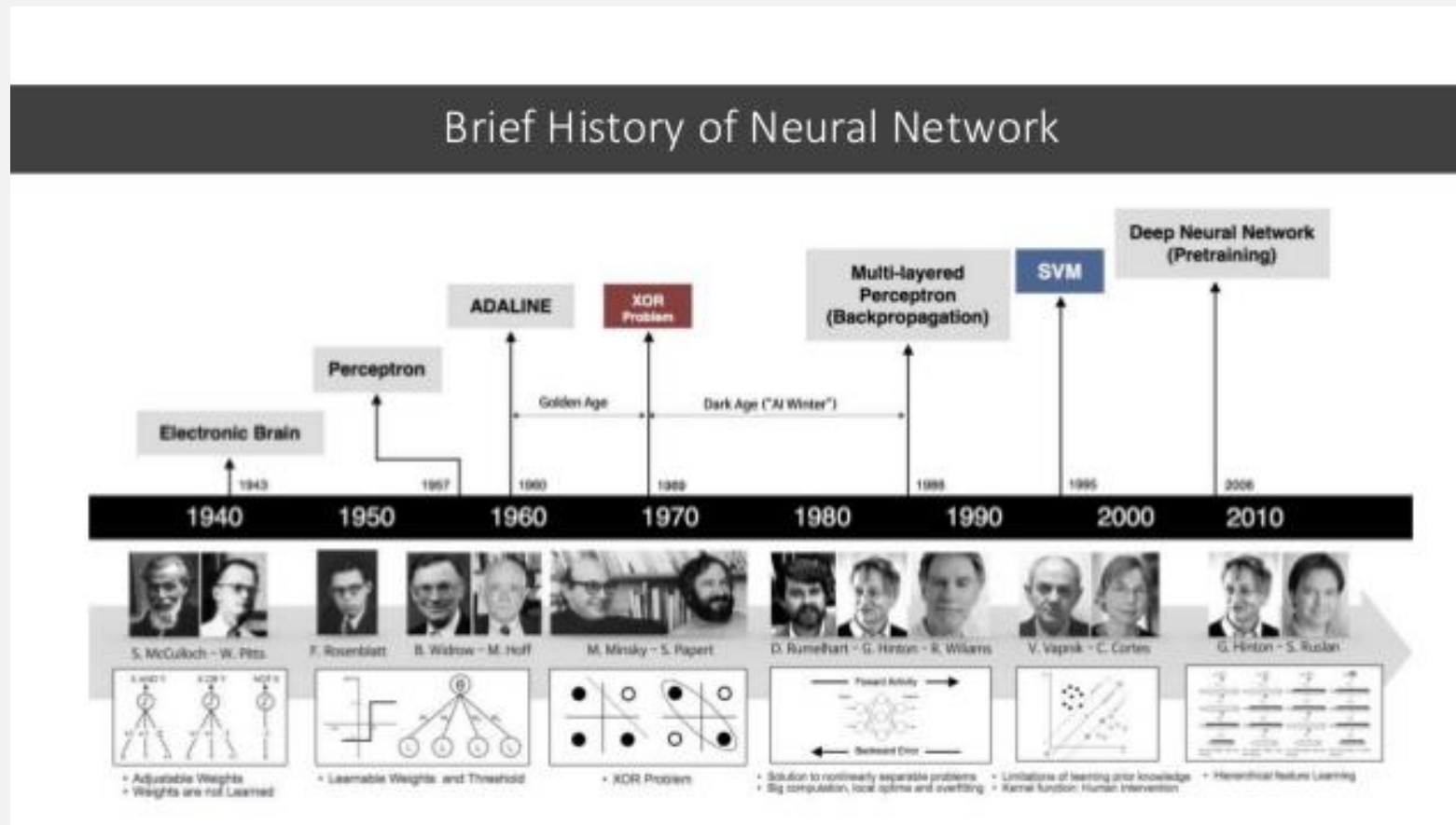


Deep Learning (Aprendizaje Profundo)

Es un sub campo de del Machine Learning que utiliza redes neuronales para el proceso de aprendizaje. Los algoritmos se encuentran en diferentes capas (entrada, oculta y salida) neuronales compuestas por pesos (números).



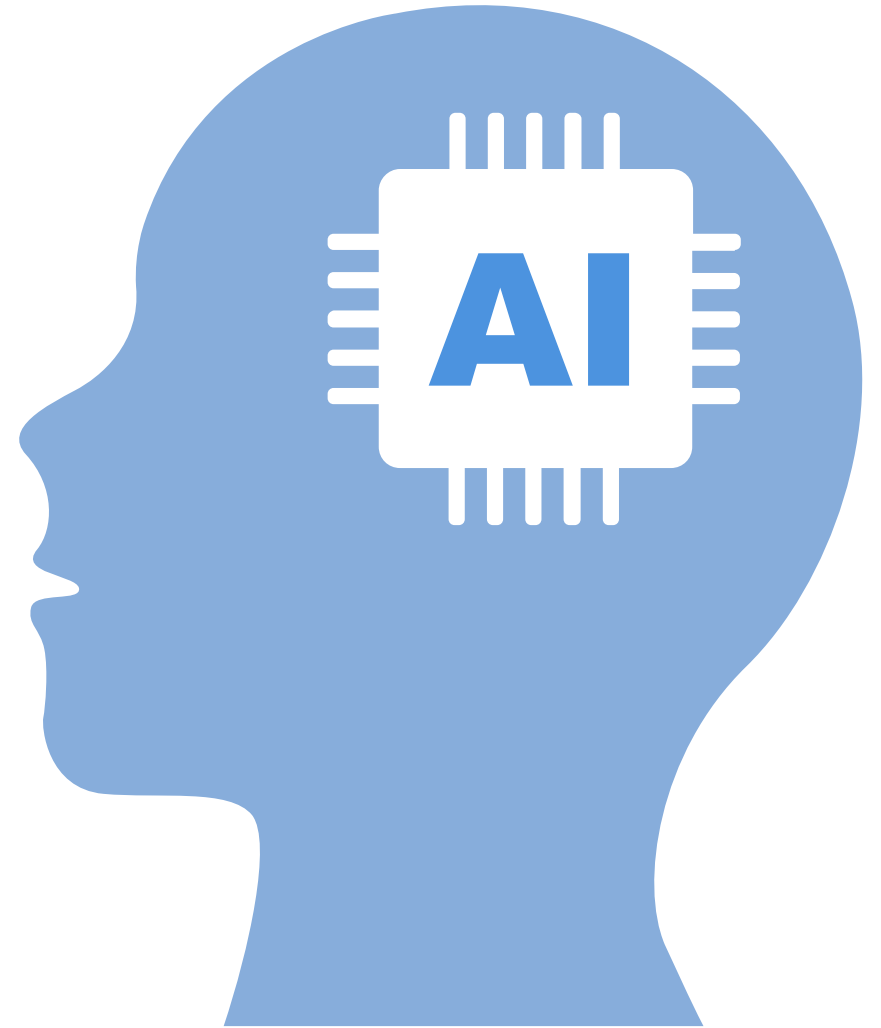
HISTORIA Y EVOLUCIÓN DE LAS REDES NEURONALES



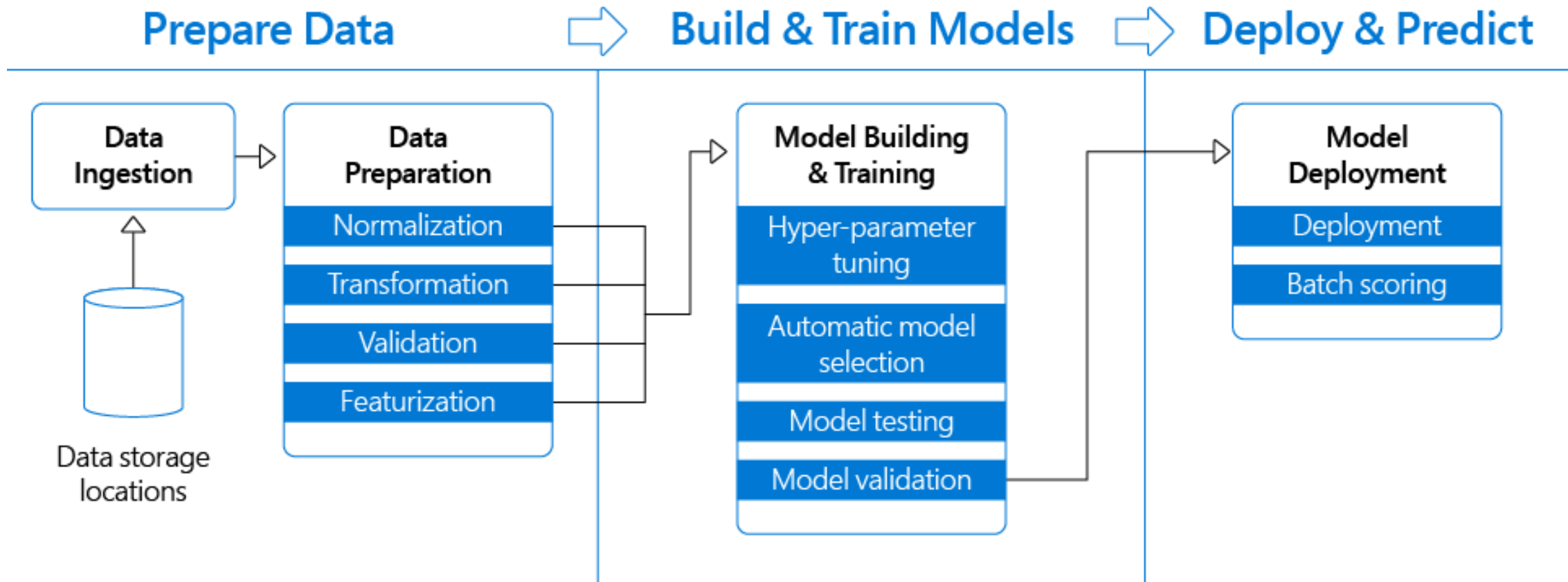
Constuyendo

INTELIGENCIA ARTIFICIAL

Capítulo 02



Pipeline (tubería) Machine/Deep Learning



Un Pipeline de Machine Learning es una secuencia de operaciones sobre el dataset que va de un estado inicial hasta el modelo

Como ve el computador una imagen

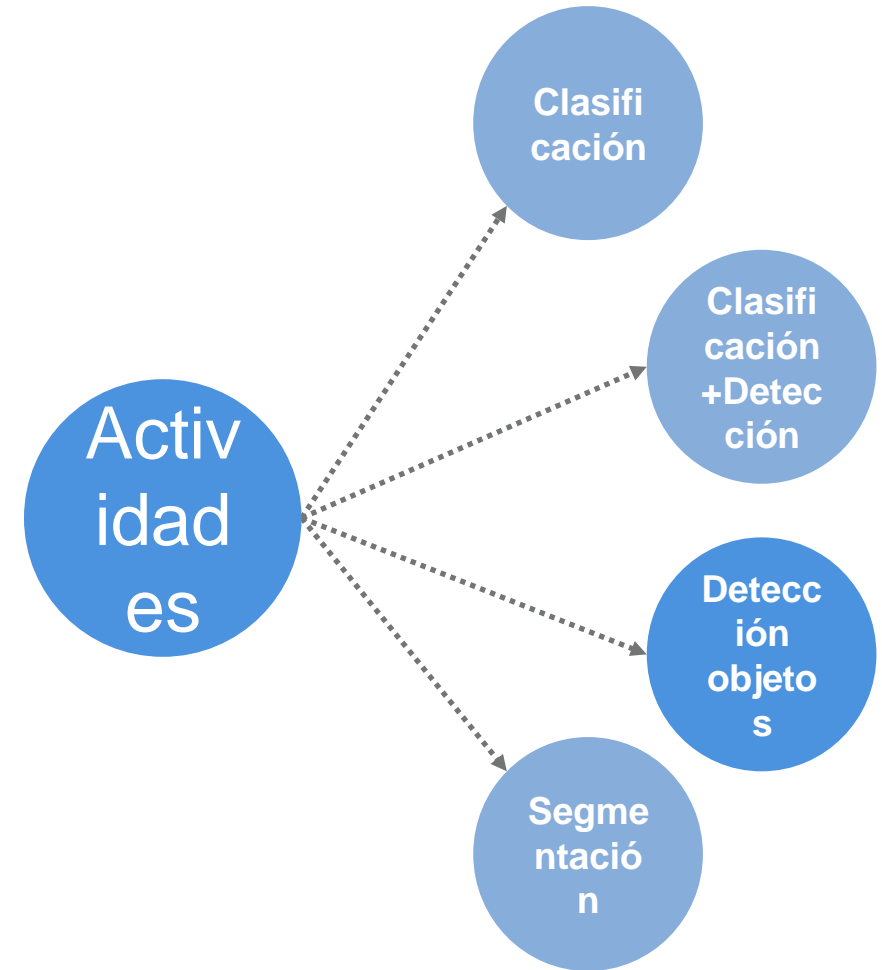
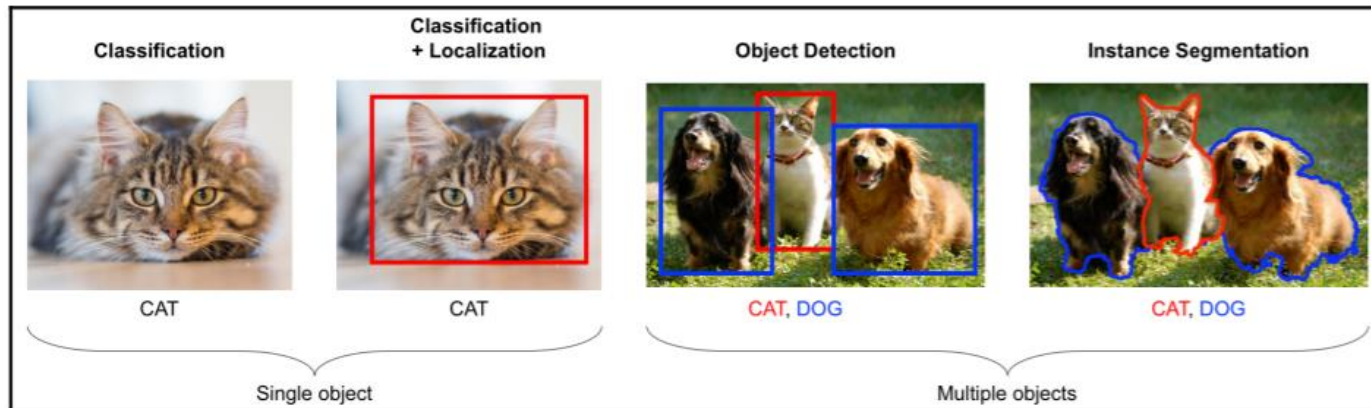


```
207 202 136 89 119 103 117 179 216 213 210 212 216 215 215 220 217 219 124 218 215 222 215 202 211 224 216 219 219 221 215 219 219 215 211 206 198 195 193 189
202 206 130 81 105 104 130 212 215 198 202 214 213 216 221 215 216 216 214 220 228 218 184 155 196 220 217 217 216 222 216 215 207 199 192 190 190 191 187 178
204 203 147 70 105 94 174 216 211 208 204 209 219 217 214 221 221 223 214 219 215 187 172 168 176 215 220 219 216 221 214 206 194 193 193 191 189 191 192 188
201 202 154 74 86 114 202 180 172 202 213 209 214 218 215 214 211 217 216 224 203 169 176 186 165 213 222 222 216 215 206 197 195 198 202 201 196 197 197 194
200 215 155 68 77 178 202 170 160 144 176 219 218 211 217 216 222 222 220 215 188 171 179 168 166 213 217 218 210 204 198 195 200 198 197 198 199 199 190 176
205 208 174 25 128 207 201 187 194 146 138 177 208 217 217 212 200 208 214 187 155 167 181 161 170 213 209 208 203 197 196 198 197 196 197 197 197 197 188 173
203 196 171 31 179 195 203 192 197 201 160 125 154 189 181 162 128 145 170 152 137 163 167 149 169 213 206 206 203 200 202 203 198 200 202 200 198 198 191 179
204 213 143 111 188 203 205 191 177 194 185 149 117 99 201 119 113 105 121 119 136 167 148 125 164 212 207 209 209 209 210 206 211 205 199 197 200 202 190 170
203 204 175 171 193 197 196 202 189 176 162 155 152 98 64 126 125 130 78 137 119 141 147 113 145 207 208 208 209 212 208 208 205 205 208 201 203 198 188 156
200 205 182 177 195 197 194 201 194 158 161 150 130 133 110 115 113 106 81 127 130 96 122 91 102 205 219 203 209 213 202 207 203 203 205 196 196 192 184 151
199 205 188 180 192 195 195 201 205 134 129 136 143 143 114 91 67 102 114 102 119 111 135 127 121 193 216 214 213 216 212 212 211 207 204 194 195 194 186 145
199 202 189 177 184 192 196 202 194 145 147 165 172 152 134 100 99 93 133 136 122 114 88 120 135 177 206 213 210 212 213 210 214 208 203 196 199 196 179 122
198 199 190 176 179 191 197 199 206 200 185 177 154 137 161 136 113 117 165 128 107 146 102 43 101 180 224 214 215 215 205 207 205 203 202 198 198 190 186 95
195 196 193 178 176 190 196 196 200 188 124 129 130 108 129 155 129 130 137 104 127 120 124 92 52 132 207 207 205 210 210 209 203 203 202 196 193 187 189 95
195 196 196 178 168 183 192 193 213 180 118 151 149 124 109 144 158 145 97 94 135 94 104 121 94 129 202 213 203 204 209 202 206 204 200 196 194 189 168 85
196 196 197 174 159 176 189 193 198 168 159 178 111 130 135 130 174 124 62 122 146 128 89 84 122 150 200 202 210 207 202 207 201 199 196 189 199 187 150 48
193 192 191 179 153 170 183 193 197 207 179 158 128 160 165 134 121 85 73 110 150 170 148 124 109 142 170 163 180 192 196 200 197 199 190 202 185 176 134 43
197 188 195 183 144 158 180 193 191 203 166 160 171 189 162 151 100 96 147 116 137 188 148 138 139 143 77 81 42 82 130 160 189 203 203 184 190 178 131 46
198 181 181 182 148 154 179 186 197 206 187 171 170 200 172 135 146 114 138 128 174 127 116 122 110 41 100 118 30 6 39 50 104 140 193 193 191 171 116 30
189 193 182 184 160 148 176 177 189 196 205 179 136 164 180 169 172 168 163 167 143 112 96 83 81 105 105 110 119 25 43 118 73 60 102 170 182 164 116 52
182 188 176 188 151 112 159 170 174 183 196 202 183 166 155 172 196 190 167 133 63 92 83 75 113 113 87 122 138 80 4 81 136 68 3 93 150 155 103 73
186 183 183 193 143 91 139 166 184 198 183 178 193 198 191 190 181 199 174 125 48 51 101 93 52 0 111 129 106 104 14 16 120 111 0 17 104 150 118 116
183 174 179 174 160 122 119 140 171 217 210 182 174 189 206 187 206 178 192 123 50 70 67 9 6 83 134 87 130 104 64 20 71 135 71 0 54 110 126 130
154 150 149 129 171 156 88 90 159 201 187 178 179 168 190 199 190 205 184 108 39 50 67 96 109 129 71 59 122 149 86 22 38 105 116 22 34 62 104 96
113 107 84 61 109 133 83 76 128 182 135 82 105 188 135 152 178 188 138 57 68 108 105 109 108 105 80 68 99 109 90 79 0 30 134 77 4 27 46 81
144 141 144 128 132 145 130 100 161 202 178 119 97 57 34 29 110 130 98 103 112 77 74 86 120 81 51 96 112 86 92 79 47 3 99 123 54 24 24 38
194 187 195 192 179 182 186 184 184 203 209 178 156 132 111 99 78 54 120 119 77 83 87 89 59 65 96 104 85 80 68 58 61 25 37 133 109 10 24 27
200 198 191 194 196 193 193 190 188 191 198 173 133 124 131 120 106 107 89 102 97 71 38 4 9 44 81 57 37 34 5 36 41 46 2 77 150 56 11 21
198 206 197 197 202 198 199 199 195 201 202 178 118 98 102 85 76 44 43 94 49 0 16 64 103 128 101 107 117 75 20 35 61 88 45 4 113 100 0 43
197 200 201 200 197 198 203 198 195 193 188 190 162 155 128 52 51 128 135 111 139 164 131 169 172 148 90 95 121 101 38 34 8 59 66 3 82 103 0 25
202 194 199 199 195 200 199 193 198 191 186 190 186 206 177 113 151 197 170 165 135 147 156 99 111 64 35 68 96 86 26 42 28 49 52 27 79 99 41 12
195 189 194 191 191 192 184 190 186 193 198 168 133 156 253 143 186 173 155 117 49 80 85 31 84 62 8 52 79 68 41 28 30 21 45 40 78 100 55 26
184 184 185 186 185 184 182 181 185 187 185 142 84 70 86 65 154 156 106 109 73 16 91 83 104 65 3 50 79 73 48 26 37 30 26 11 70 81 0 3
207 207 208 208 208 207 206 205 203 210 176 122 73 53 49 25 101 121 92 91 69 94 104 92 99 95 48 51 50 52 28 1 0 1 3 8 55 81 57 73
208 209 209 210 209 208 208 207 218 201 124 43 3 0 2 0 0 80 65 58 17 2 48 55 47 36 0 1 0 2 15 46 74 90 99 114 112 142 147 148
205 205 206 206 206 205 204 204 200 205 174 141 128 119 114 110 117 126 129 136 131 125 127 112 108 110 111 121 121 131 131 148 157 164 162 168 165 163 173 164
208 208 209 209 208 208 208 209 213 208 205 205 203 204 203 203 203 208 208 204 201 202 206 196 201 199 188 185 198 190 189 185 182 177 175 174 175 177 174
207 207 207 207 207 207 207 209 204 204 209 210 210 211 208 212 208 207 202 207 214 208 207 210 202 202 209 208 202 195 202 197 196 196 194 194 193 184 185
207 207 207 207 207 207 207 208 206 205 206 206 207 207 203 199 205 210 211 205 207 209 203 203 207 207 202 202 202 202 208 207 208 206 203 201 200 199 199 201
207 207 207 207 207 207 208 208 204 213 213 205 207 210 207 207 207 208 206 206 210 207 203 209 205 206 203 208 209 200 206 202 202 204 200 207 207 202 206 200
```

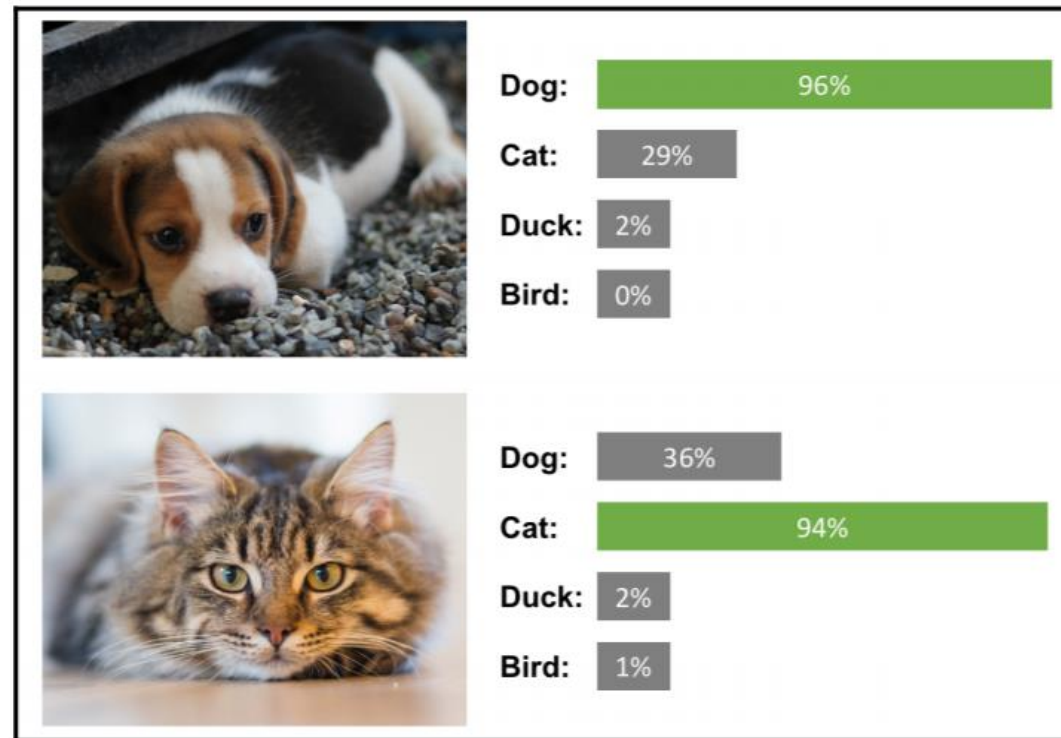


Procesamiento de Imágenes

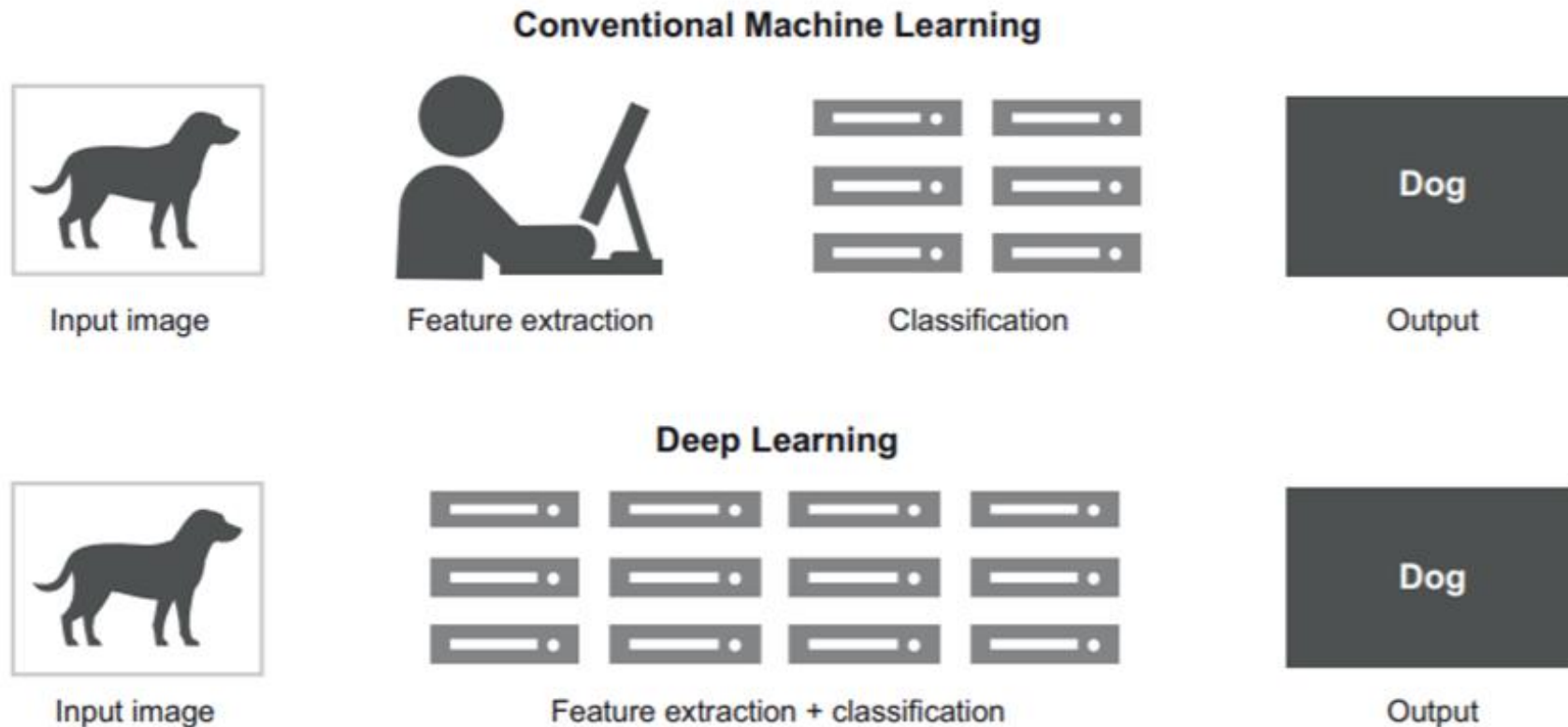
La detección o localización es una tarea que encuentra un objeto en una imagen y localiza el objeto con un cuadro delimitador.



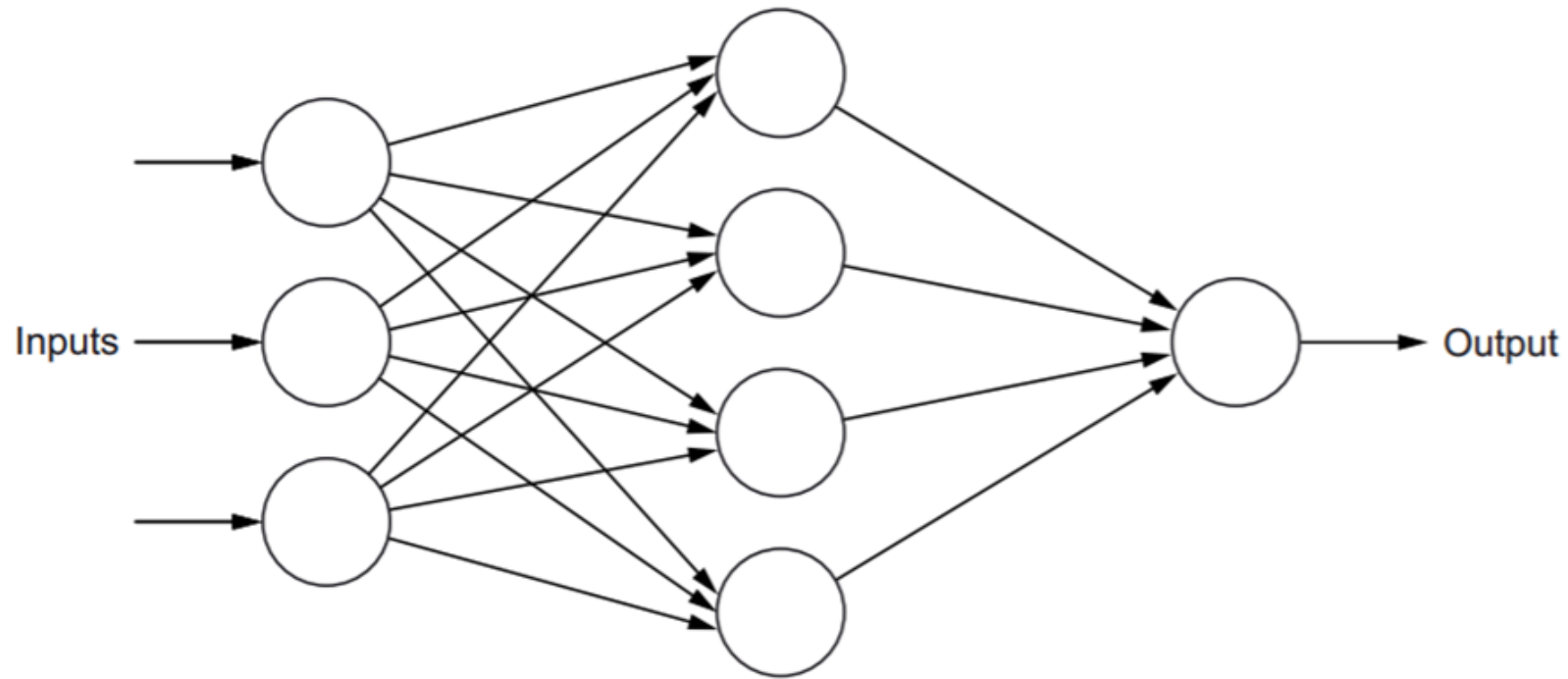
Clasificación Imágenes



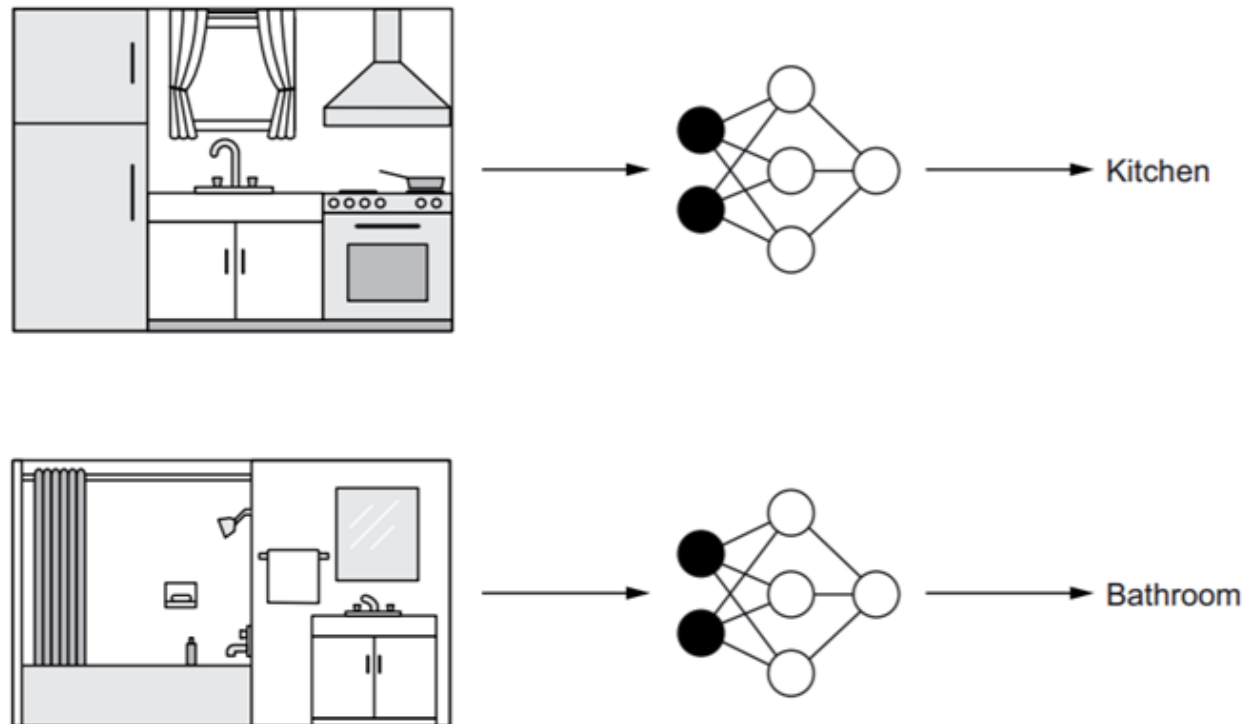
Machine Learning los ingenieros deben desarrollar algoritmos para extraer características que se pueden alimentar al modelo. Un modelo DL no necesita este complejo paso preliminar, es automático

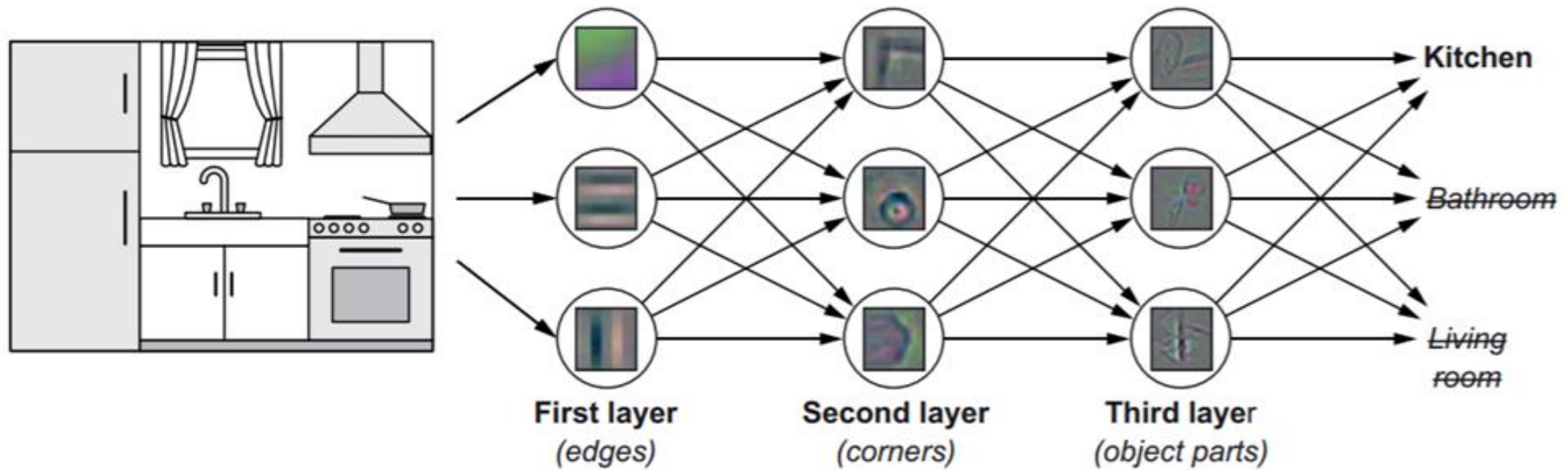


Red Neuronal



En la clasificación de imágenes lo que queremos hacer es crear una red que recibe los píxeles de la imagen, realizar algunos cálculos internos y genera la clase (con suerte) correcta para la imagen de entrada.

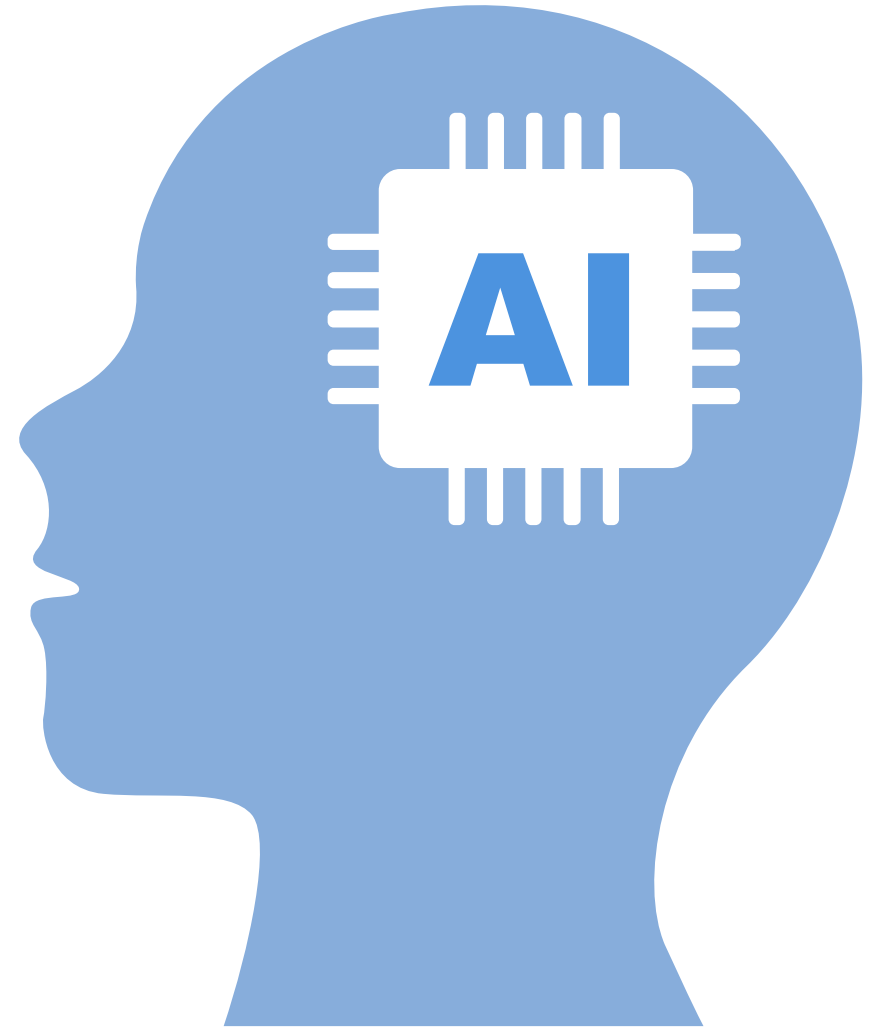




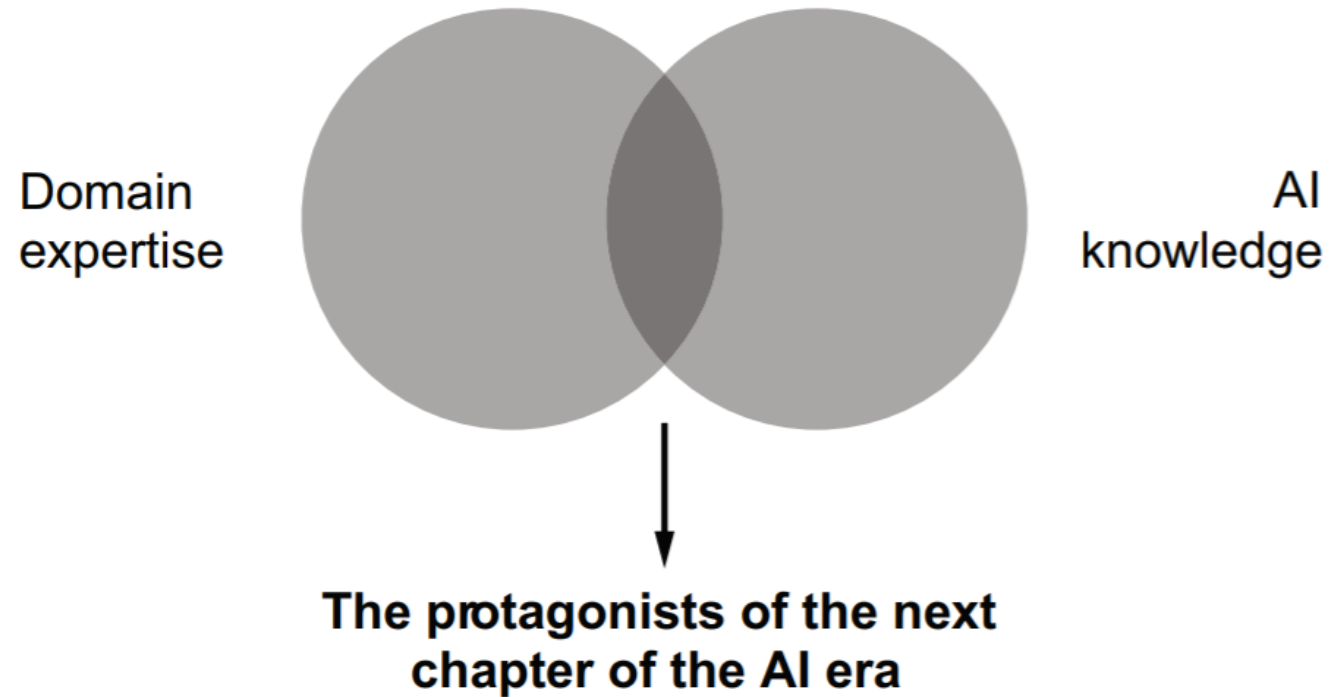
Creando valor con

INTELLIGENCIA ARTIFICIAL

Capítulo 03




























El líder de IA es una persona que combina un dominio específico de la experiencia y conocimiento de IA para resolver desafíos específicos de la industria.



Sistema de Recomendación

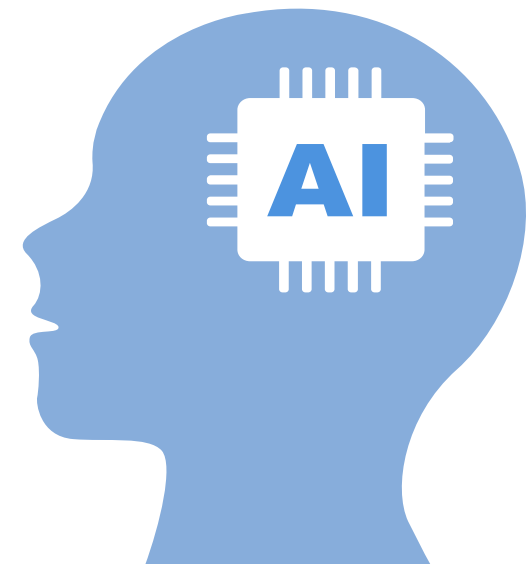
Capítulo 04

¿Qué es un sistema de recomendación?

**Sistema de recomendación
utilizando Machine Learning**

Un sistema de recomendación no es solo un algoritmo elegante. También se trata de comprender los datos y sus usuarios . Los súper algoritmos requieren súper hardware y mucho. Más datos crea otros desafíos, como cómo acceder a ellos lo suficientemente rápido.



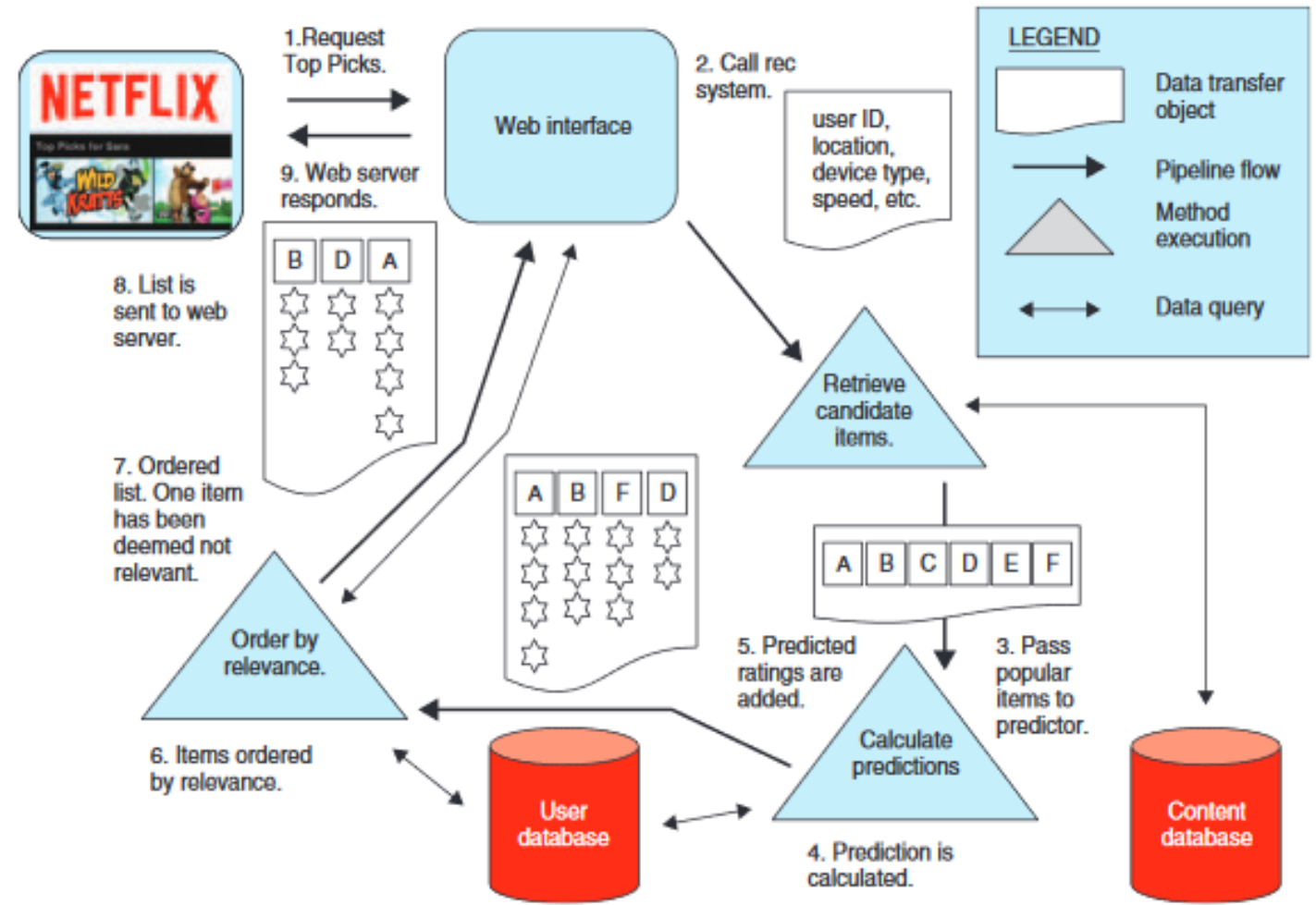


Sistema de Recomendación

Un sistema de recomendación calcula y proporciona contenido relevante al usuario basado en conocimiento del usuario, del contenido y de las interacciones entre ellos.

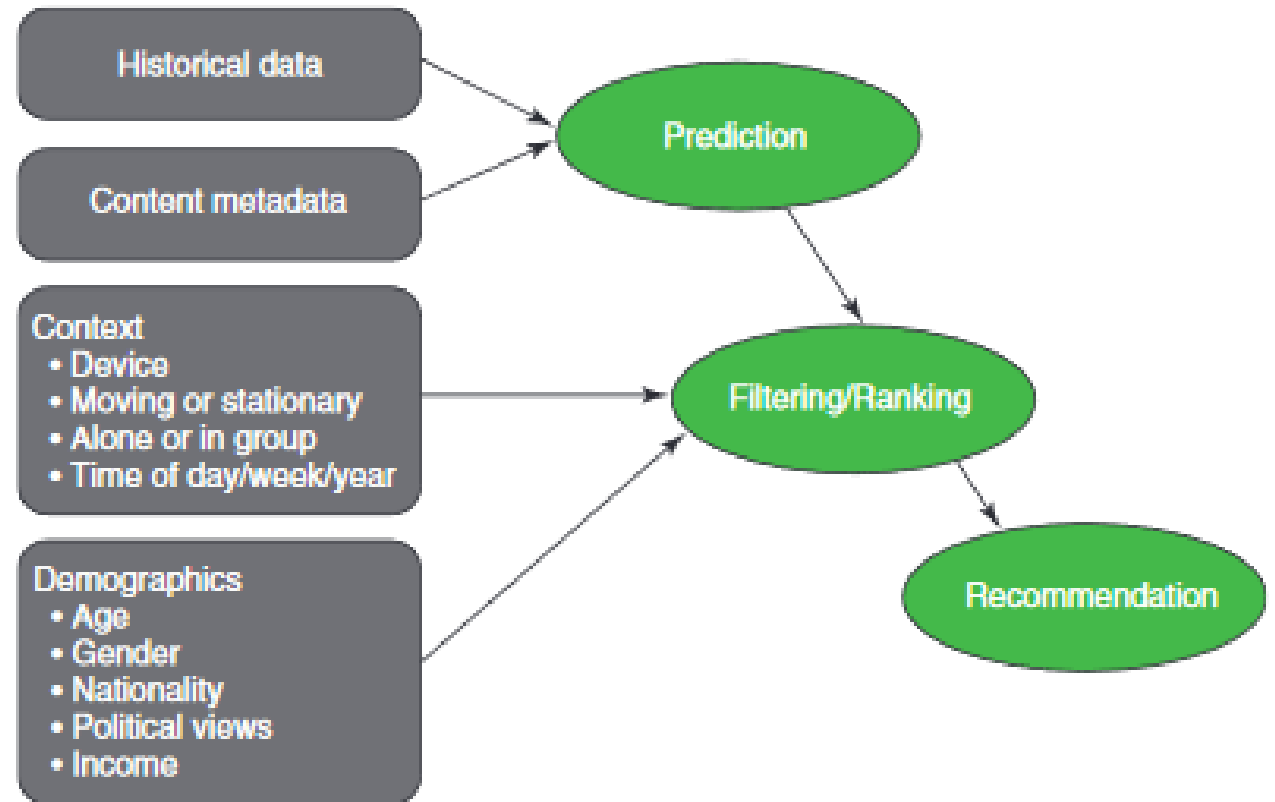
¿Cómo recomienda Netflix ?

1. Se recibe una solicitud para el top de recomendaciones.
2. El servidor llama al sistema de recomendaciones, el cual consiste en una secuencia de métodos y retorna del catálogo las películas/series que mas se asemejen al gusto del usuario.
3. El top 5(normalmente 100 o mas) de los elementos pasan a la siguiente etapa.
4. Se calculan predicciones de ratings en los elementos, usando las preferencias del usuario conseguidas del database del usuario. Si algún elemento posee un rating predicho muy bajo estos elementos se retiran de la lista.
5. Los elementos ahora con la predicciones de ratings pasan al proceso de ordenarlos por importancia.
6. Los elementos se ordenan de acuerdo al gusto del usuario, al contexto y demografía. Inclusive el proceso puede intentar hacer que la lista de elementos sea lo mas diversa posible
7. Los elementos ahora están ordenados por relevancia, e incluso algunos fueron eliminados puesto que se estimo que no eran relevante para el usuario.
8. El Sistema de recomendaciones retorna la lista al servidor.
9. Y el servidor le entrega la lista al usuario.



¿Qué información utiliza Netflix?

- Para la predicción:
 - Emplea el historial del usuario.
 - Y los metadatos del contenido en los catálogos.
- Para el filtro/ranking:
 - Usa el contexto en el que se encuentra el usuario.
 - Y la demografía del usuario.



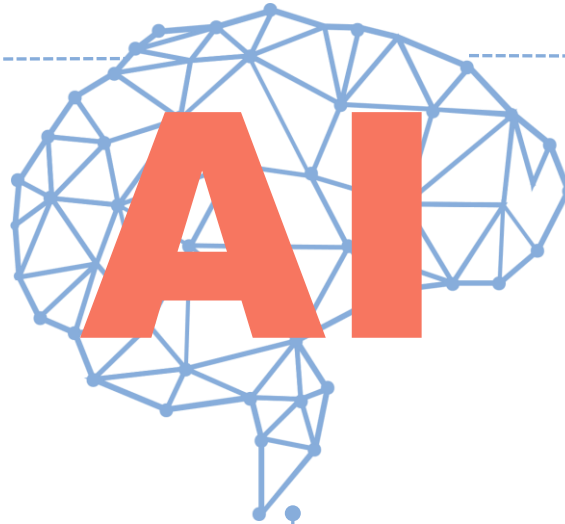
A Taxonomía de los Sistemas de Recomendación



¿ Qué algoritmos se utilizan ?

COLABORATIVE FILTERING

Usan los datos de uso para generar recomendaciones. Por ejemplo, si un usuario ve un conjunto de películas parecido a las vistas por otro usuario, se recomiendan las películas vistas por este otro usuario al usuario original.



CONTENT BASED FILTERING

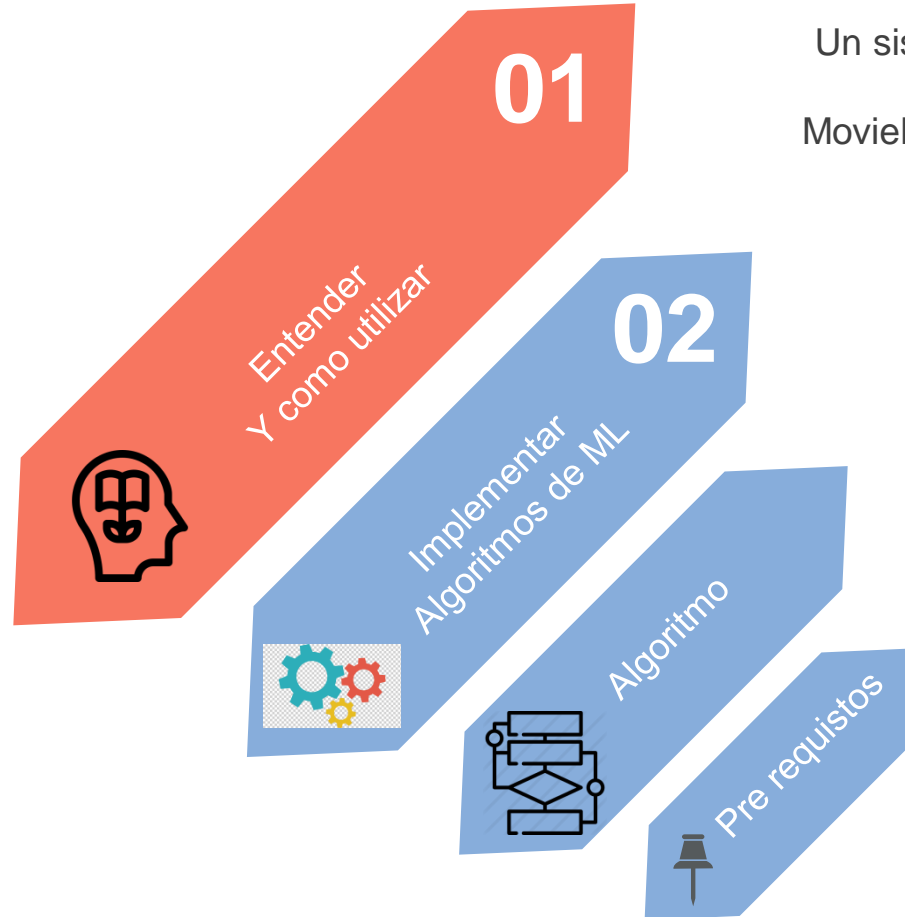
Usan los metadatos del contenido disponible y los perfiles de los usuarios, para generar recomendaciones. Por ejemplo, a un usuario que le gusta películas de ficción se les recomienda películas de ficción.

HIBRID RECOMMENDERS

Son una combinación de ambos
COLABORATIVE FILTERING y CONTENT
BASED FILTERING.

..

Objetivos del curso



01.- Entender y como utilizar

Un sistema de recomendación para generar recomendaciones en un conjunto de datos de películas, en este caso las evaluaciones de películas del sitio MovieLens, que usaremos para generar un sistema de recomendaciones para los usuarios.

02.- Implementar algoritmos de Machine Learning

Entendiendo que podemos implementar algoritmos de aprendizaje automático.

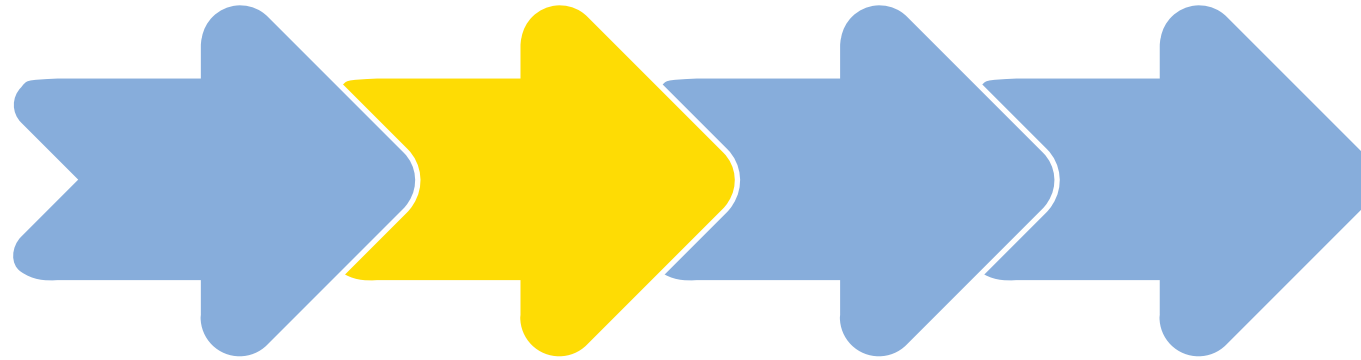
Algoritmo

Vamos implementar una variación de k-NN, lo que significa "los k vecinos más cercanos" de un usuario que ha visto películas. Implementando este algoritmo desde cero, en un cuaderno de Python mientras trabajamos con los conjuntos de datos y generamos recomendaciones basadas en ellos.

Pre requisitos

Python, Pandas y Numpy.

A Ambiente de trabajo



01

Ambiente de Análisis

Utilizaremos una herramienta de Google llamada **Colaboratory** (es necesario tener una cuenta de Gmail), que es un Jupyter notebook que será guardado en **Drive** y no requiere configuración.

02

Sobre **Colaboratory**

El código de su cuaderno Python se ejecuta en una máquina virtual dedicada a su cuenta. Las máquinas virtuales se reciclan durante un cierto tiempo de inactividad o si la ventana está cerrada. Para restaurar su computadora portátil, es posible que deba volver a cargar su archivo csv y ejecutar las opciones Runtime y Restart and run all ...

03

¿Se puede utilizar otro ambiente Jupyter notebook para este curso?

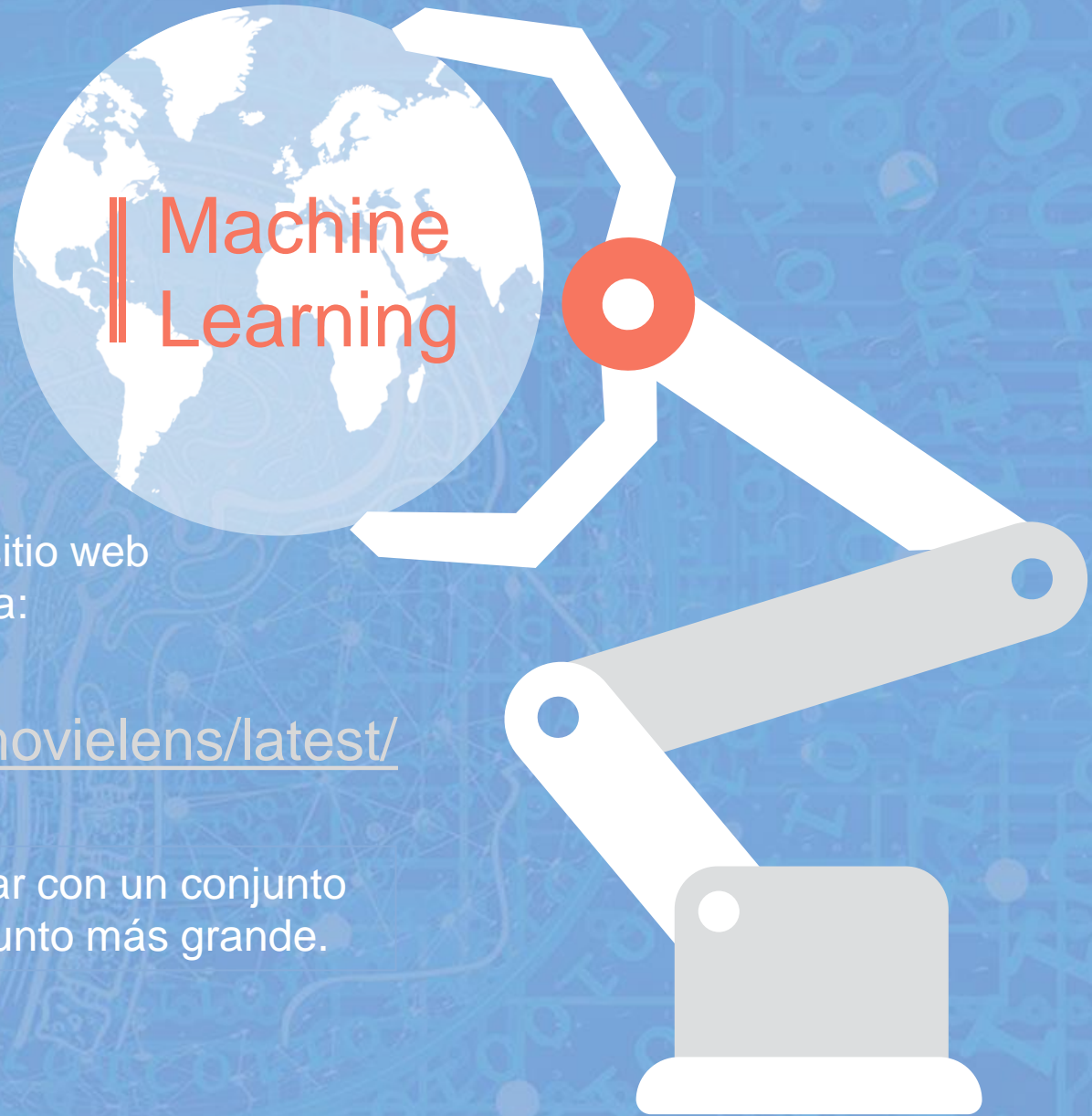
Por supuesto que sí, se puede emplear **Anaconda**.

04

Caso de utilizar **Anaconda**

Se tiene que instalar o si ya instalada en la plataforma correspondiente del sistema operativo que se esté usando.

Descargando los datos ...



Puede descargar el conjunto de datos del sitio web de grouplens, en la versión más actualizada:

<https://grouplens.org/datasets/movielens/latest/>

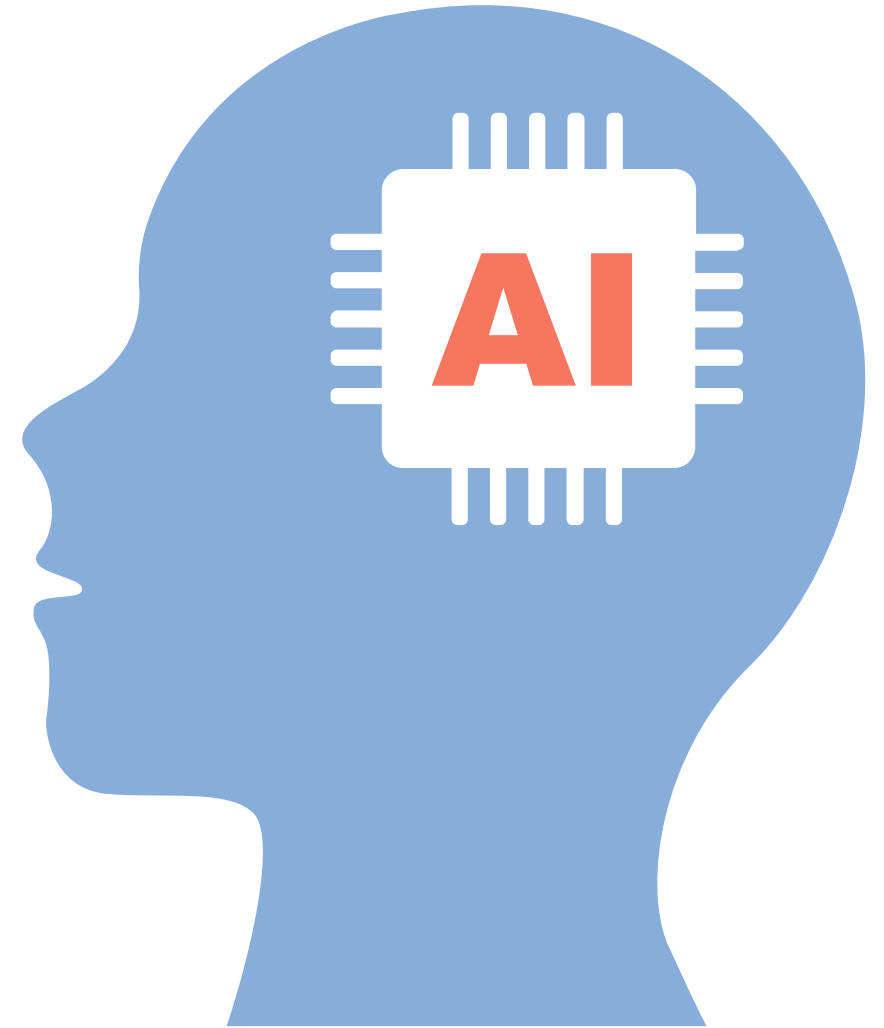
Elija la versión **ml-latest-small** para trabajar con un conjunto de datos reducido o **ml-latest** para el conjunto más grande.



¿Qué es una heurística?

Heurística de Recomendación,

entendiendo lo que es colaboración



Explicar ...

Abrir cuaderno en Colab

Explicación de pasos a seguir ...

.- Cargar y leer archivos con Pandas.
.- Renombrar columnas.

.- Usar **describe** para información numérica.
Usar **value_counts** para contar la frecuencia.

.- Insertar una columna.
.- Ordenar por cantidad (decreciente) con la función **sort_values()**
Parámetro **ascending=False**.

.- Agrupar datos con **groupby()** y hacer consulta.

Paso 1

Paso 2

Paso 3

Paso 4

Total votos, nota promedio y posibles dificultades de heurísticas simples

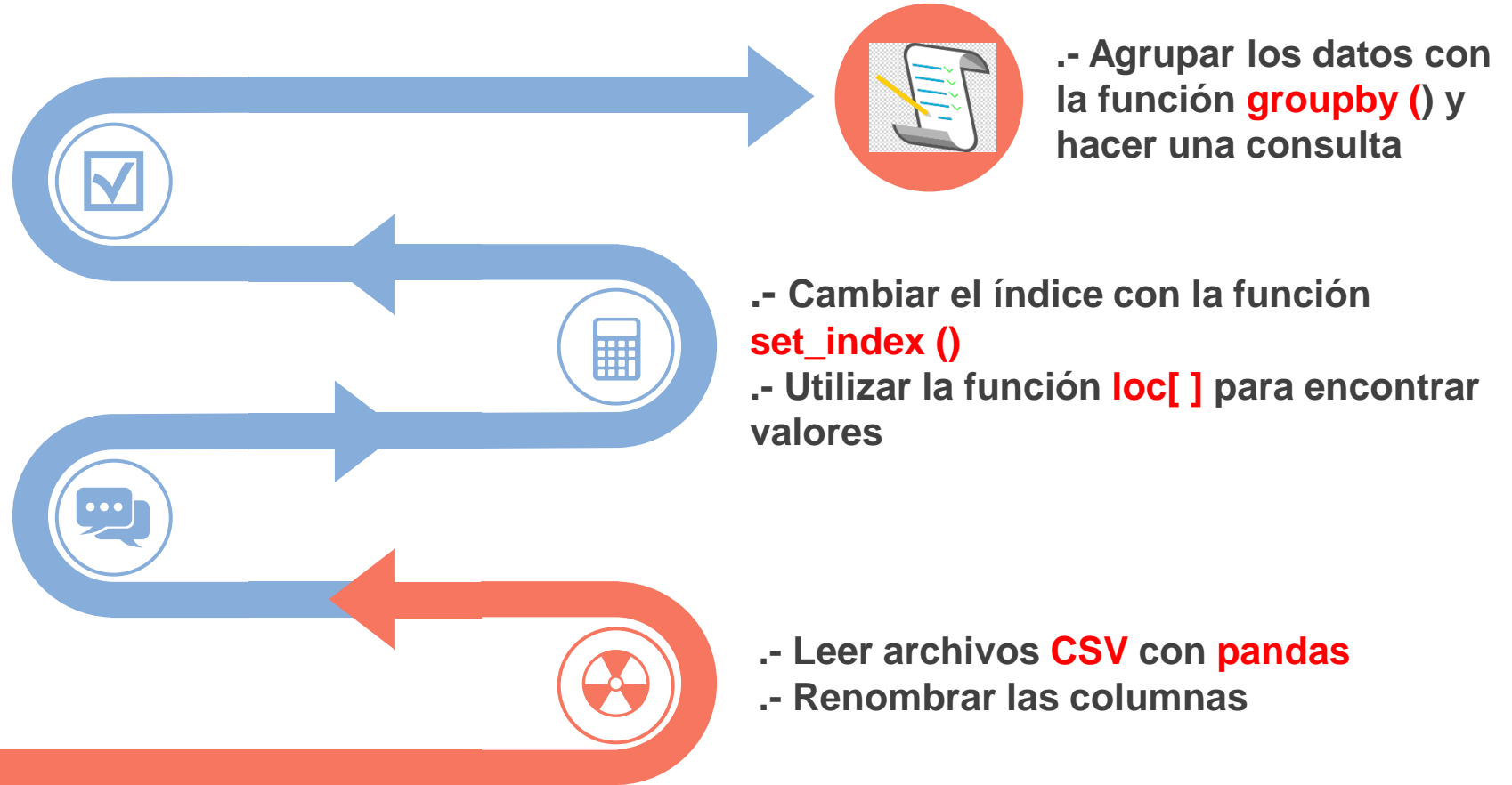


Colab ...

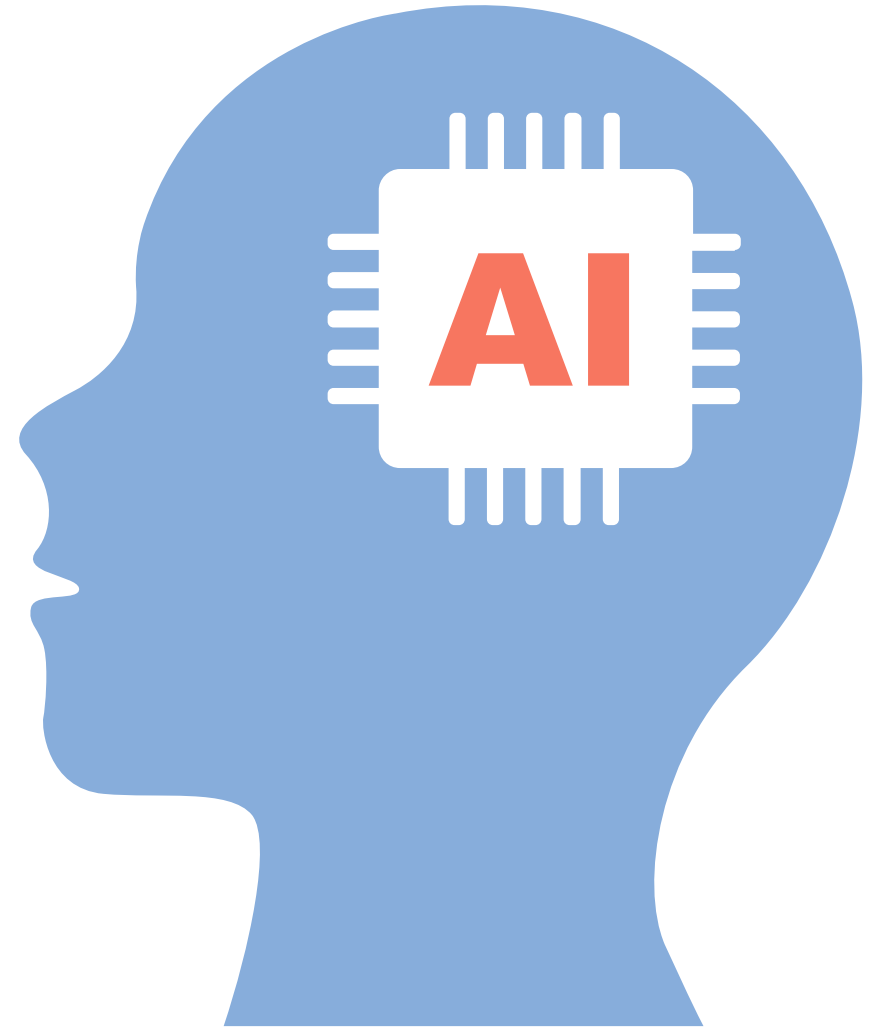
A ¿Qué hemos aprendido hasta aquí?

- .- Hacer una inclusión de columna
- .- Ordenar por cantidad con la función **sort_values()**, ordenar de forma decreciente con el parámetro **ascending = False**

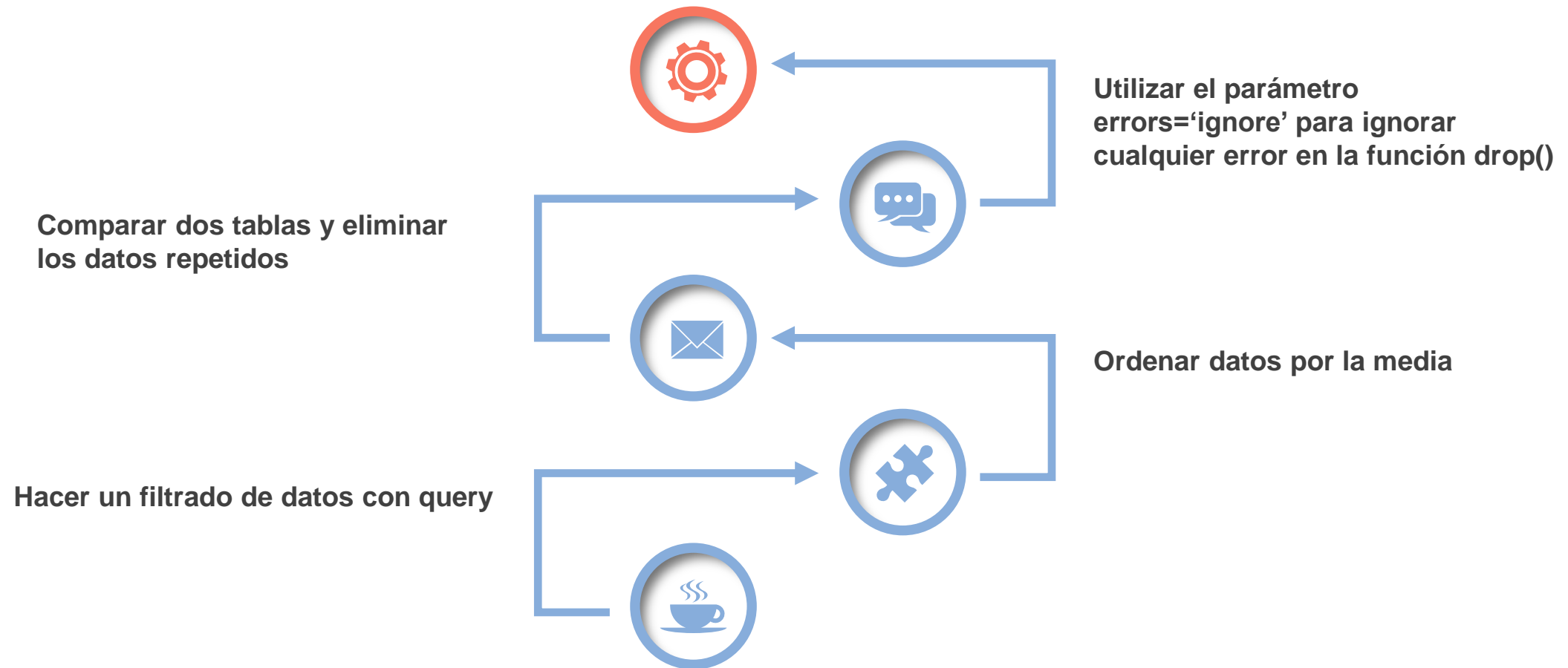
- .- Utilizar la función **describe** para ver información numérica
- .- Usar **value_counts** para contar la frecuencia



REPASANDO ANTERIOR AVANCE ...



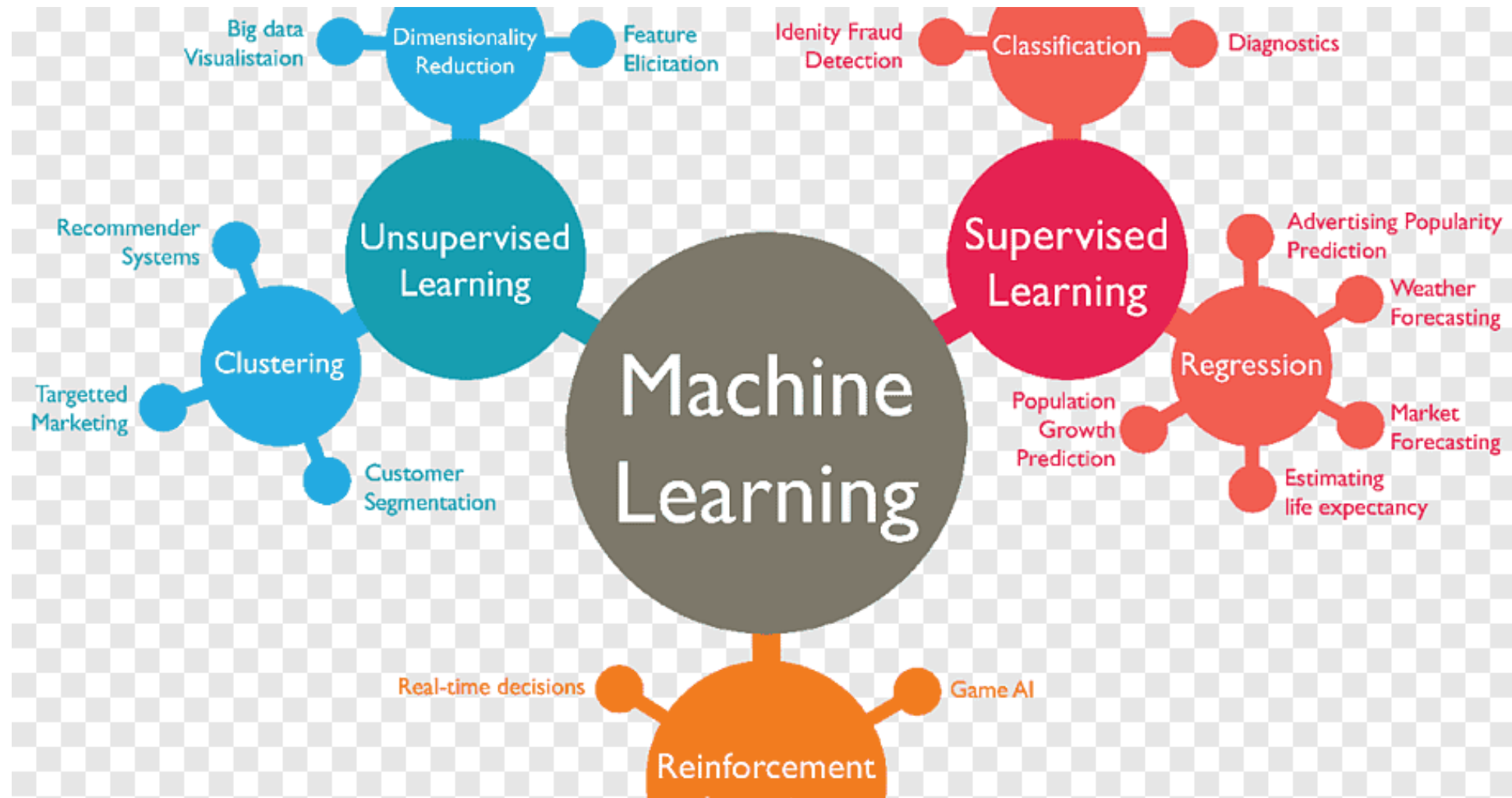
¿Qué hemos aprendido hasta aquí?





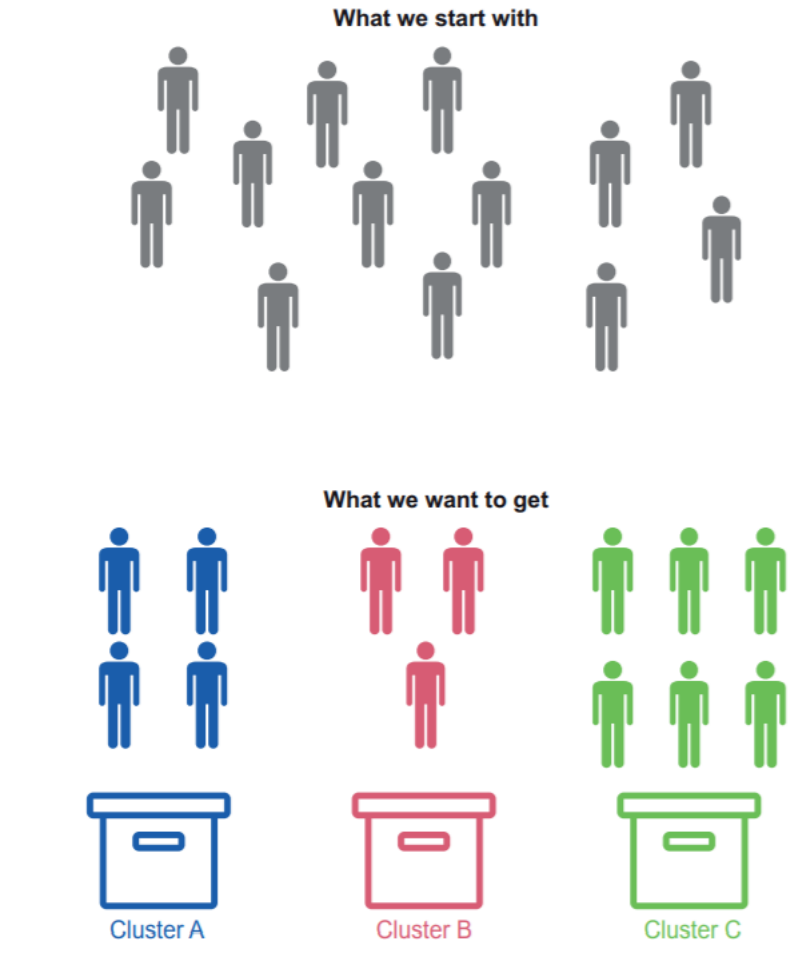
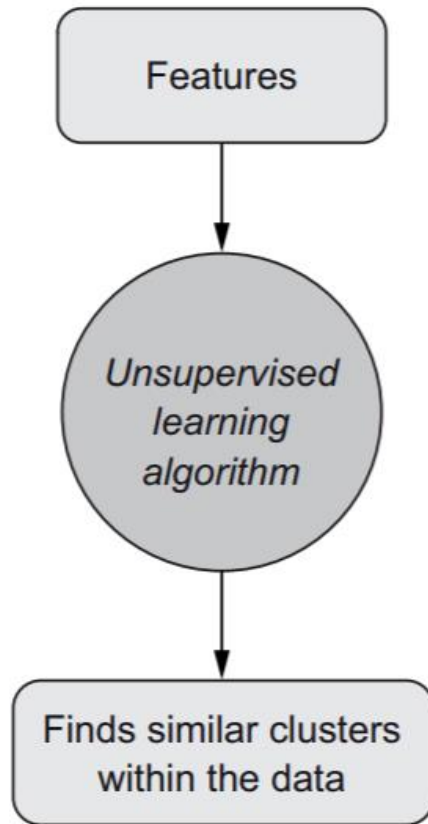
¿ Tienen
preguntas?

A Categorías del Machine Learning



AI Aprendizaje No Supervisado

Unsupervised Learning





Quiero ir al cine, hay dos películas nuevas que se están estrenando:

- Martín Rivas
- Padre Gallo

Recomendación basada en filtro colaborativo

... ¿Cuál de esas dos películas voy a ver?

... ¿Qué película recomiendan para mí?



Recomendación basada en filtro colaborativo

... El problema de recomendar, sea una película, música, libro, un producto comprado, una marca o cualquiera que sea la recomendación, puede ser visto de distintas maneras ...

Recomendación basada en filtro colaborativo

... Y una forma puede ser, buscar personas que tengan un gusto similar al mío.

...Esto es, quien ya vió películas que yo también ví y que además vieron películas que yo no he visto aún.

Recomendación basada en filtro colaborativo

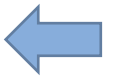
	Martín Rivas	Padre Gallo	El Padrino	Pretty Woman
Sergio	?	?	8	10
Favio	8	7	3	8
Naomi	5	9	7	9



Viendo este pequeño conjunto de tres usuarios de un sistema ... ¿Qué película desea recomendar para Sergio?

Recomendación basada en filtro colaborativo

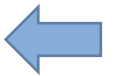
	Martín Rivas	Padre Gallo	El Padrino	Pretty Woman
Sergio	?	?	8	10
Favio	8	7	3	8
Naomi	5	9	7	9



Para recomendar veremos ... ¿Quien es el usuario(a) más parecido, más próximo a Sergio? ...(ver notas de Favio y Naomi)

Recomendación basada en filtro colaborativo

	Martín Rivas	Padre Gallo	El Padrino	Pretty Woman
Sergio	?	?	8	10
Favio	8	7	3	8
Naomi	5	9	7	9



Viendo simplemente la tabla, pareciera que tiene más sentido Naomi como usuaria más parecida ...



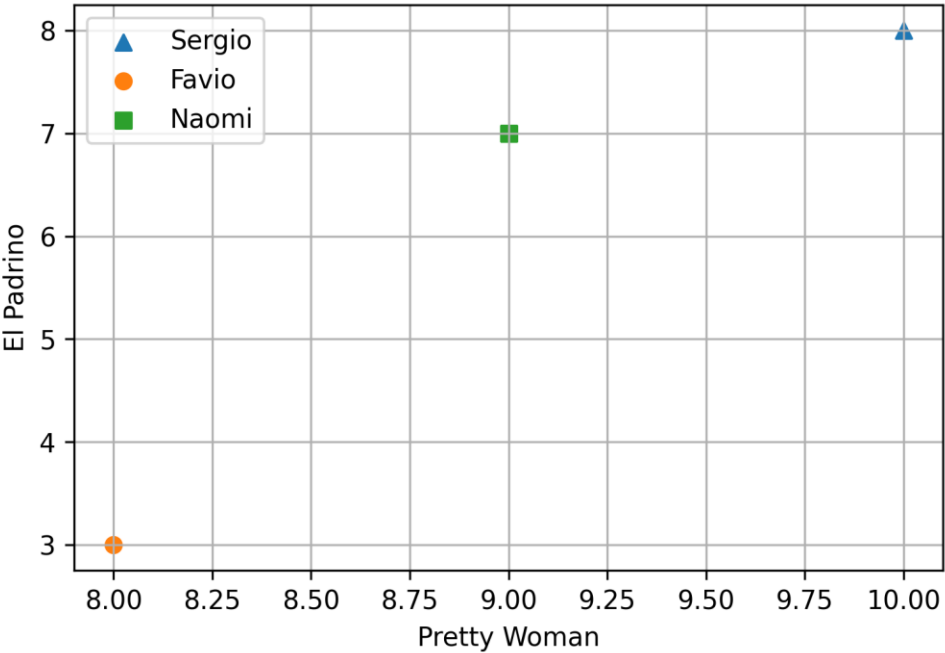
**También lo
podemos ver
de otra
manera ...**

AI Recomendación basada en filtro colaborativo

	Martín Rivas	Padre Gallo	El Padrino	Pretty Woman
Sergio	?	?	8	10
Favio	8	7	3	8
Naomi	5	9	7	9



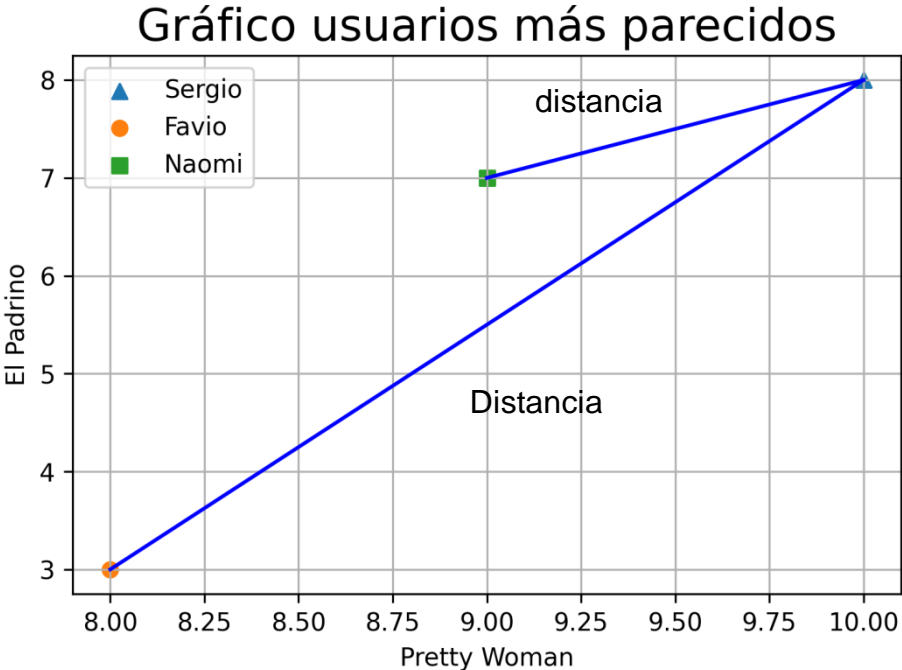
Gráfico usuarios más parecidos



... Visualmente en el plano, Naomi está más próxima de Sergio.

AI Recomendación basada en filtro colaborativo

	Martín Rivas	Padre Gallo	El Padrino	Pretty Woman
Sergio	?	?	8	10
Favio	8	7	3	8
Naomi	5	9	7	9



... Veamos distancia de Naomi a Sergio y de Favio a Sergio.

Recomendación basada en filtro colaborativo

	Martín Rivas	Padre Gallo	El Padrino	Pretty Woman
Sergio	?	?	8	10
Favio	8	7	3	8
Naomi	5	9	7	9

¿Cómo calculamos la distancia?
Triángulo rectángulo... **Teorema de Pitágoras**.

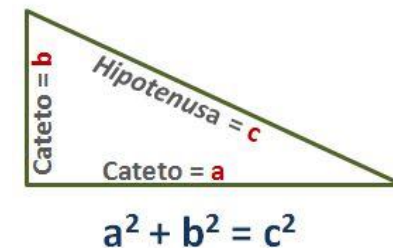
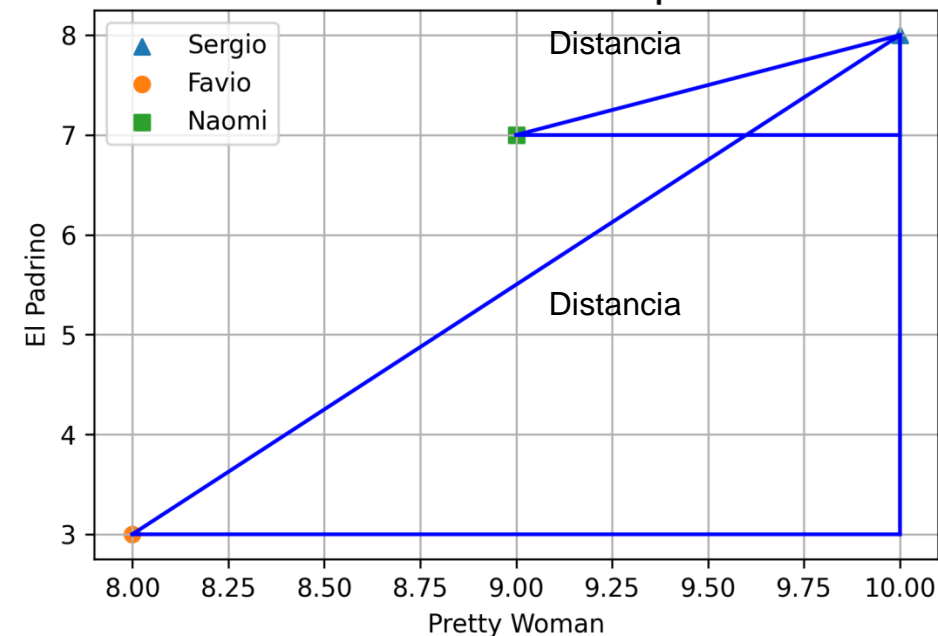


Gráfico usuarios más parecidos

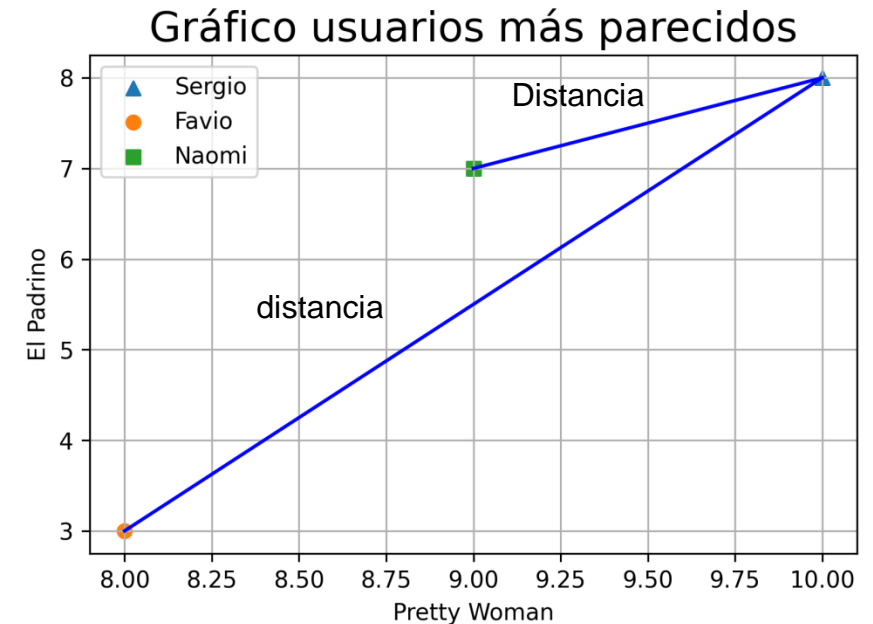




Recomendación basada en filtro colaborativo

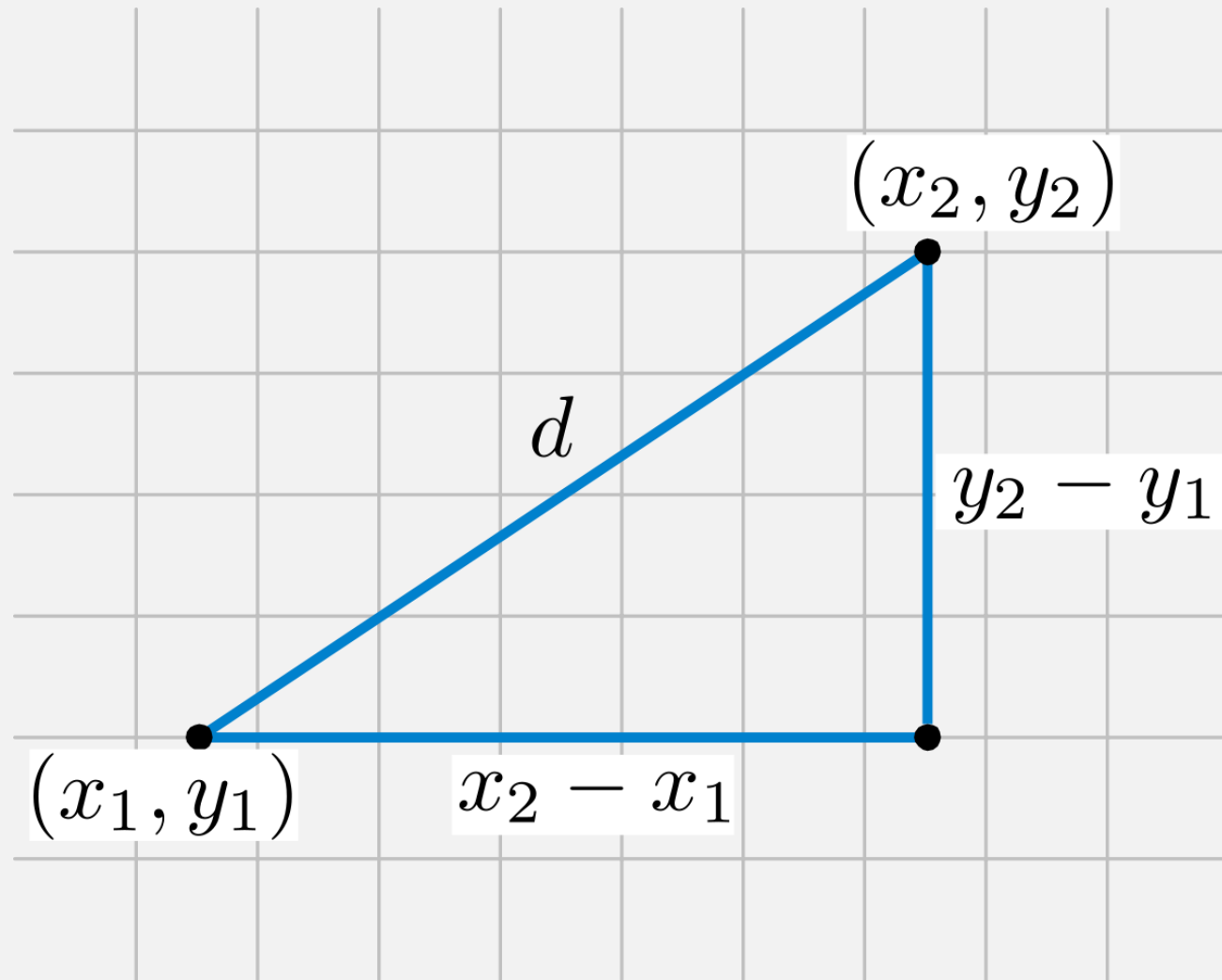
	Martín Rivas	Padre Gallo	El Padrino	Pretty Woman
Sergio	?	?	8	10
Favio	8	7	3	8
Naomi	5	9	7	9

Siendo Naomi la usuaria más próxima a Sergio, podemos ASUMIR que las películas que Sergio NO vió van a tener una nota próxima a la de Naomi.



Podemos crear una lista de recomendación (en este orden):

1. Padre Gallo
2. Martín Rivas



Definiendo
**Distancia
Euclidiana**
...en Colab.



Pensando en nuestro conjunto de datos ...

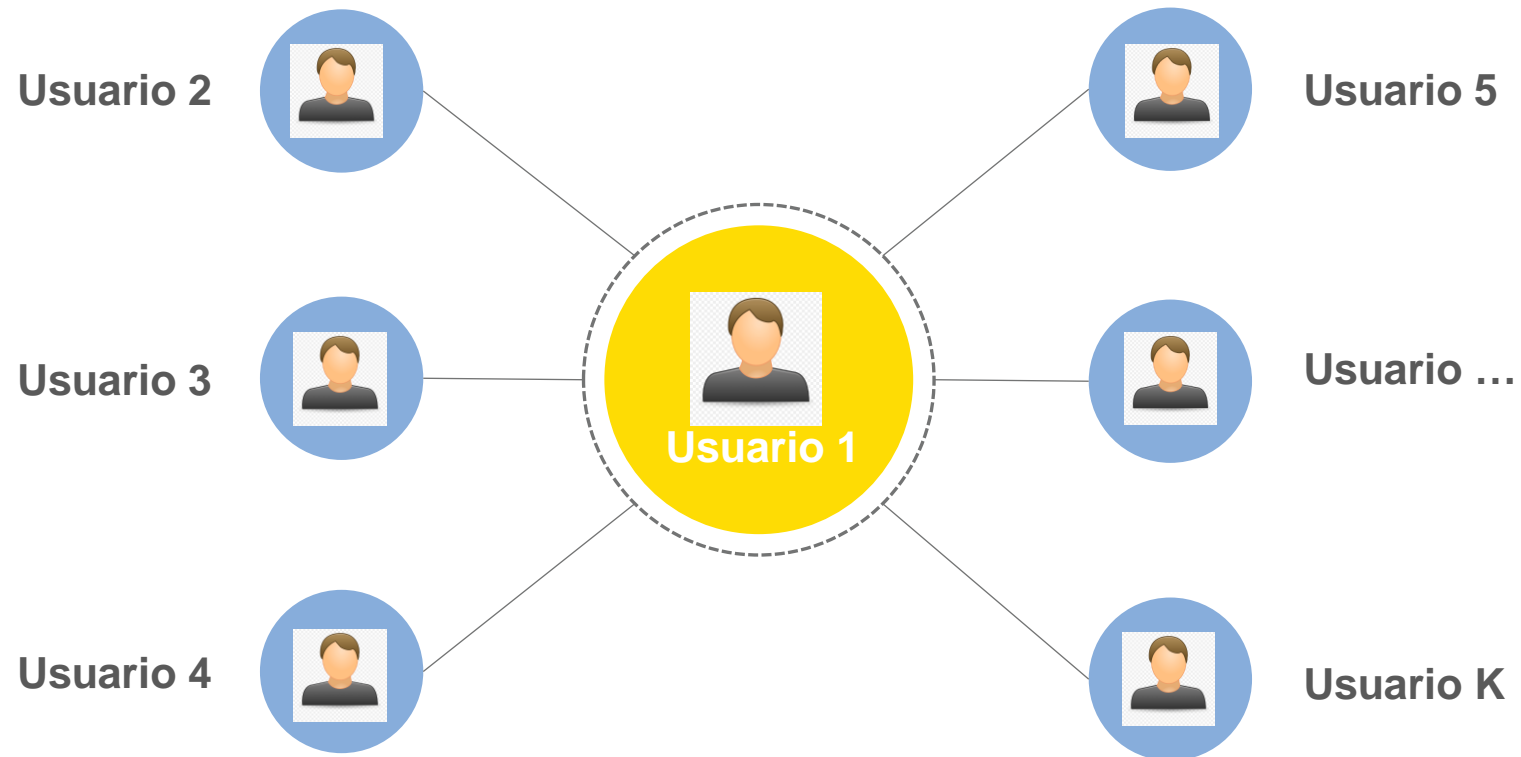
¿Si queremos calcular la distancia entre el usuario 1 y el usuario 4?

... ¿Cómo lo hacemos?



...Primero tenemos que extraer las notas del usuario 1 y luego las notas del usuario 4.

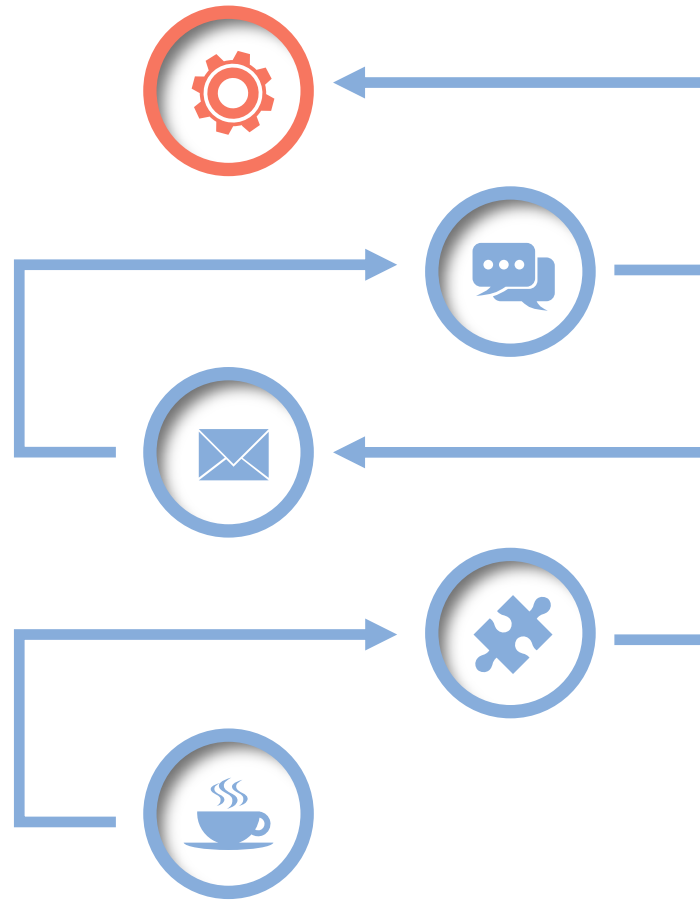
AI Implementando la distancia entre usuarios ... en Colab



¿Qué hemos aprendido hasta aquí?

- + Crear un array con **numpy** para hacer la substracción de las listas.
- + Utilizar la función `np.linalg.norm` para calcular el valor de las distancias.

- + Buscar datos similares en relación a las características que poseen.
- + Visualizar gráficamente la distancia entre los datos.



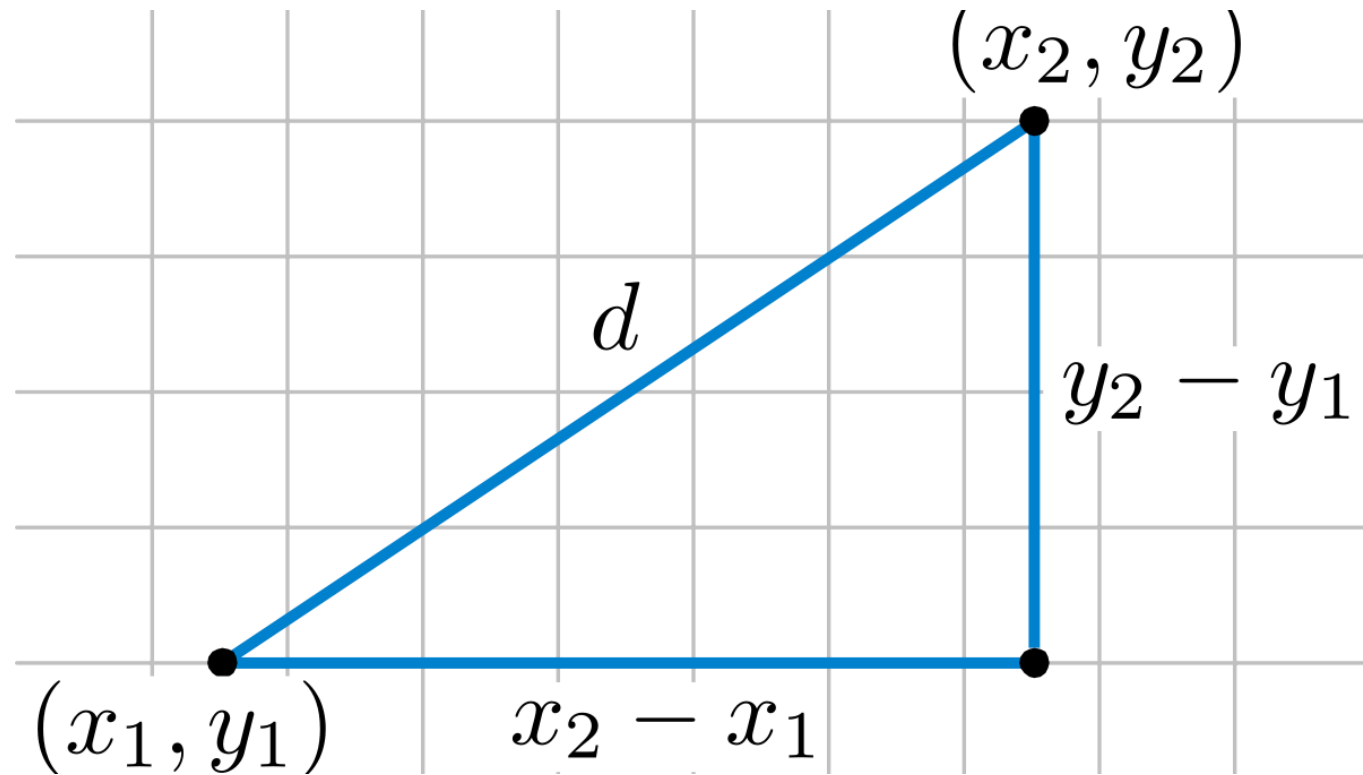
- + Utilizar la función **join** para juntar los datos.
- + Utilizar **lsuffix** y **rsuffix** para juntar los datos adyacentes.
- + Utilizar a função **dropna** para remover los valores nulos.

- + Añadir leyenda y crear una línea en el gráfico.
- + Utilizar la función **sqrt** para calcular la raíz cuadrada.



¿ Tienen
preguntas?

Calculando la distancia entre un usuario y todos los demás ... Colab



Sergio H. Valenzuela Cámara - TecnoPolis.Ai – sergio@tecnopolis.ai

Ordenando usuarios por DISTANCIA y lidiando con casos extremos



Colab ...



Gracias
Por su atención

...Continuará en el siguiente episodio.
Ahí nos vemos !!!



¿ Tienen preguntas?