



UT.1 Introducción

IES Abastos

CE Inteligencia Artificial y Big Data

Sistemas de Aprendizaje Automático.

Bloques de la UT1:



1. Aprendizaje automático.
2. Aprendizaje automático supervisado.
3. Aprendizaje automático no supervisado.
4. Tipos de problemas aprendizaje automático.
5. Conjuntos de entrenamiento.
6. Fundamentos básicos estadísticos.

1. Aprendizaje automático.

¿ Qué pensáis
qué es el
aprendizaje
automático ?



1. Aprendizaje automático.



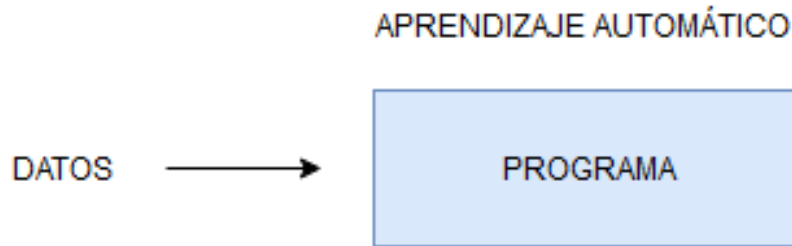
Se entiende por aprendizaje automático aquellos algoritmos que se ejecutan en los ordenadores para aprender automáticamente en **base a los datos** proporcionados.

Se trata de crear **programas** capaces de generalizar comportamientos a partir de los **datos** suministrados en forma de **ejemplos**.

El aprendizaje automático, también conocido como **machine learning**.

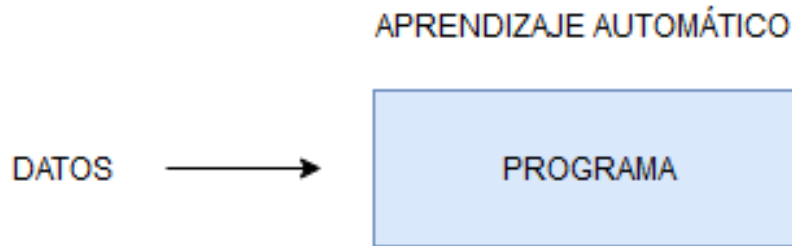
1. Aprendizaje automático.

En la siguiente figura podemos apreciar que vamos a disponer de unos datos y a través de un software vamos a realizar aprendizaje automático.

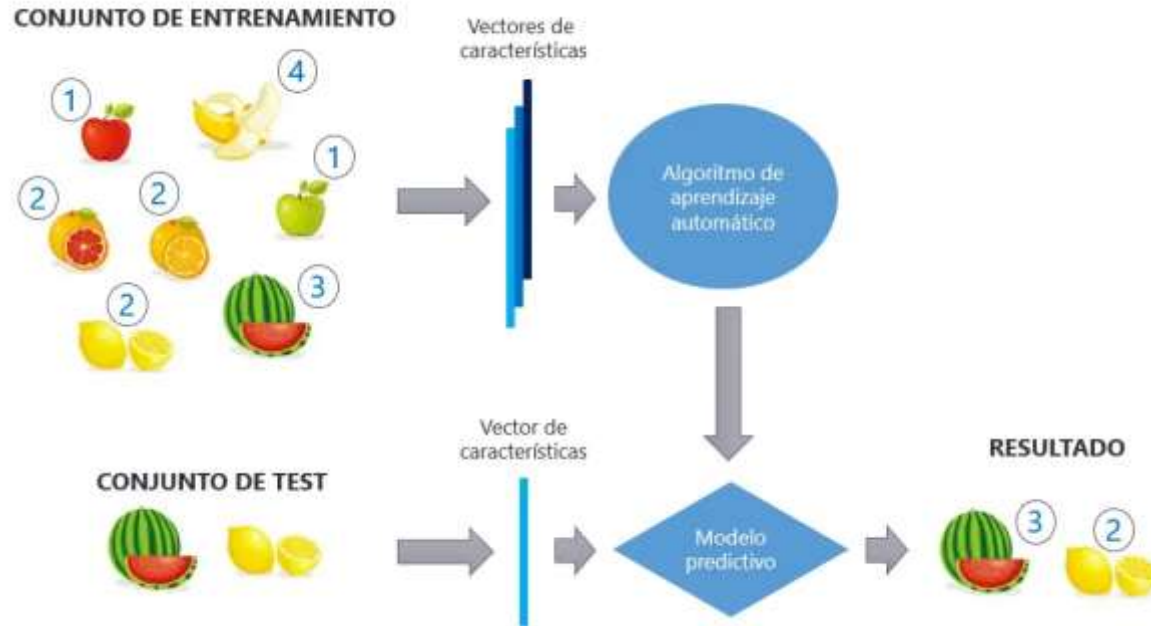


1. Aprendizaje automático.

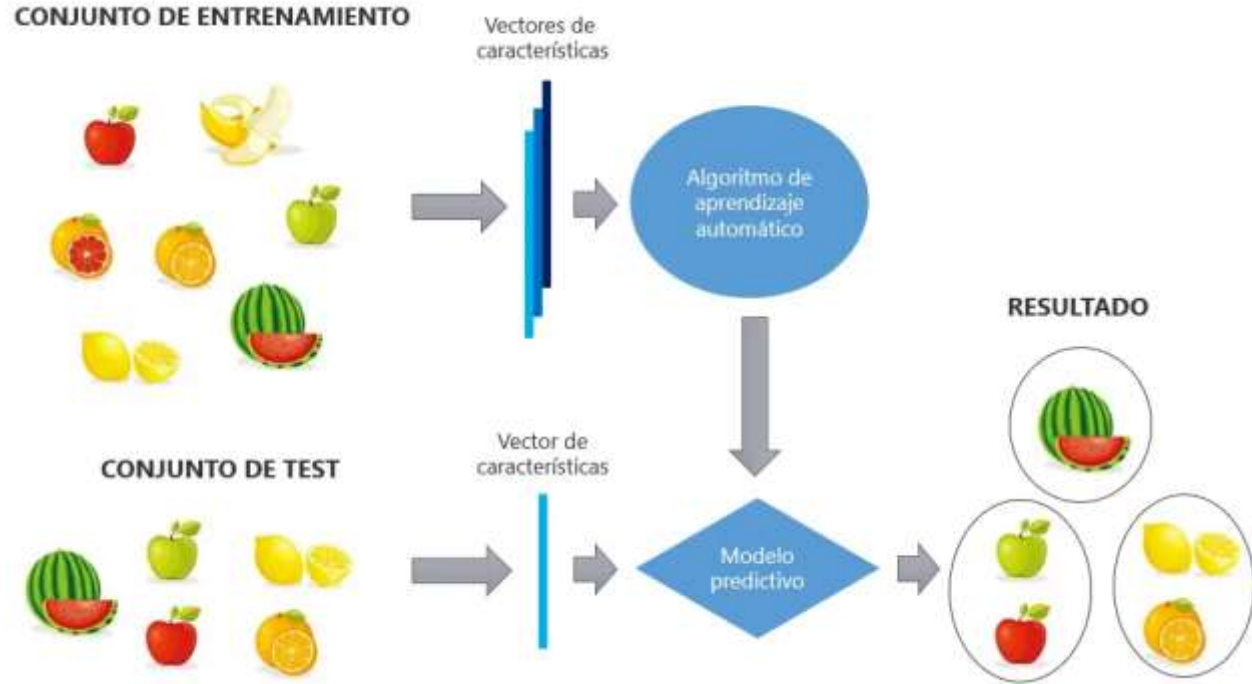
Nuestro software tiene que aprender. ¿ Cómo pensáis que lo hace?



1. Aprendizaje automático.

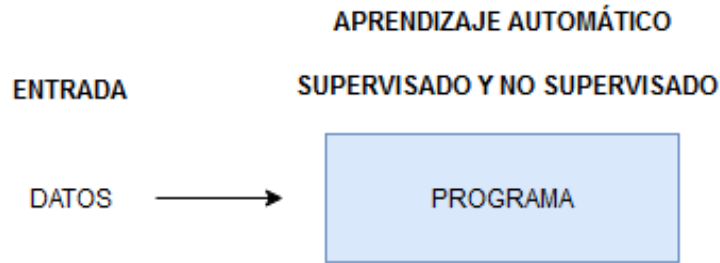


1. Aprendizaje automático.



1. Aprendizaje automático.

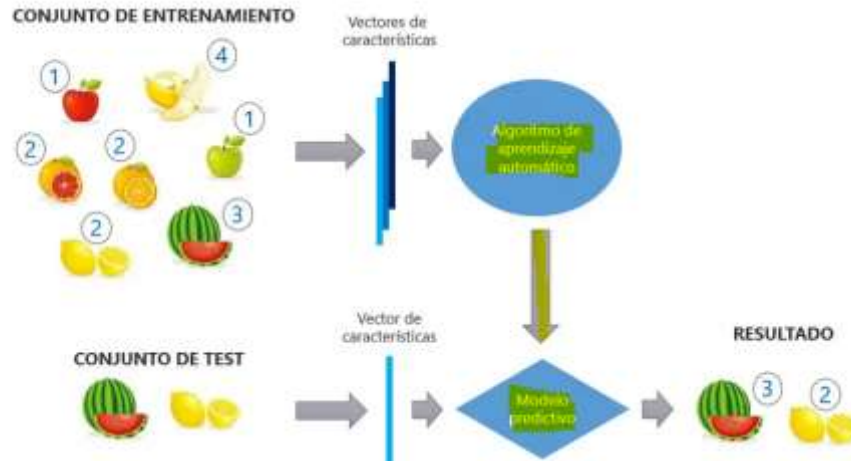
Existen dos tipos de aprendizaje: Supervisado y no supervisado.



1. Aprendizaje automático.

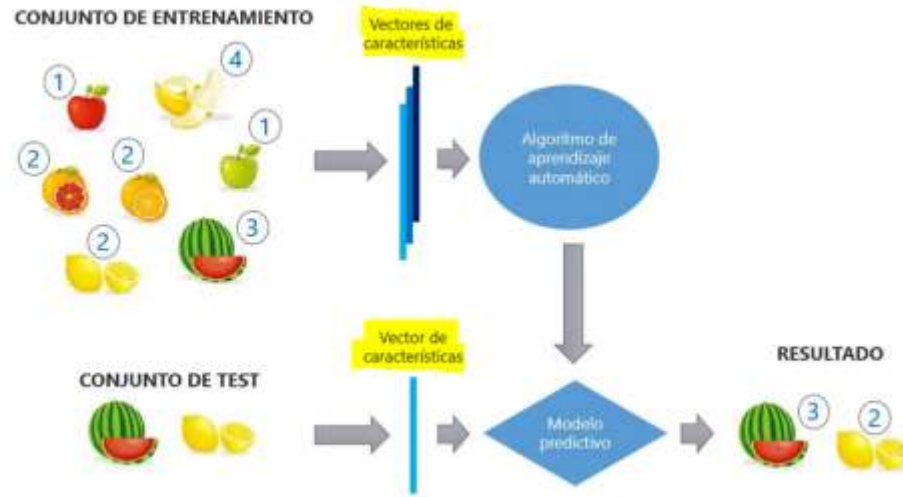
Por otro lado del ejemplo anterior también podemos extraer:

- a) Nuestro software va a tener una parte de predicción y otra de aprendizaje.



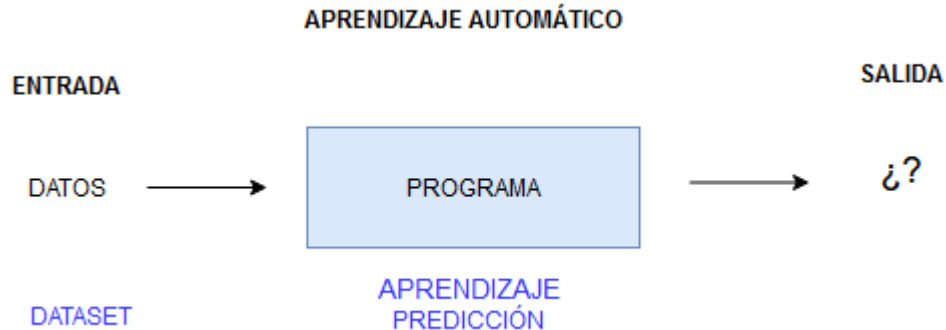
1. Aprendizaje automático.

b) Los datos de entrada tienen un formato de vectores de características. (Dataset)



1. Aprendizaje automático.

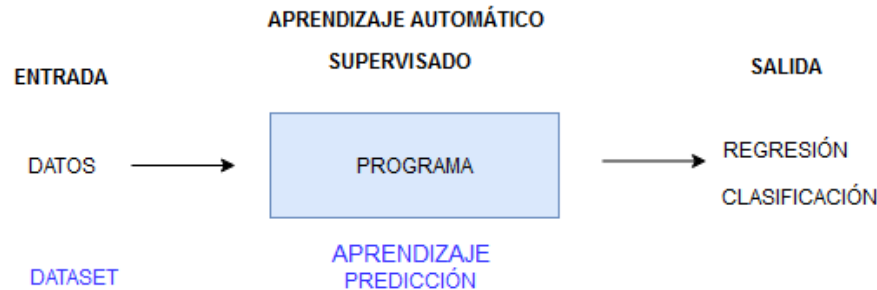
Al ser un programa, existirán datos de salida ¿ Qué tipo de datos podrían ser ?



2. Aprendizaje automático supervisado.

Predecir la temperatura que va a hacer mañana a una determinada hora sería una **regresión**.

Detectar qué ciudadanos al entrar al metro tienen puesta o no la mascarilla sería una **clasificación**.



2. Aprendizaje automático supervisado.

5 ejemplos de regresión y otros de clasificación



2. Aprendizaje automático supervisado.



<https://www.youtube.com/watch?v=vTNDXO2jiwM>

2. Aprendizaje automático supervisado.



<https://www.youtube.com/watch?v=VOC3huqHrss>

2. Aprendizaje automático supervisado.



Resumen aprendizaje supervisado.

- El aprendizaje supervisado tiene como objetivo generalizar las respuestas sobre datos no observados, utilizando para ello **ejemplos observados previamente**.
- En el caso de los problemas de regresión la variable respuesta y es una **variable numérica continua**.
- En el caso de los problemas de clasificación la variable respuesta y es una variable con **2 o más categorías o clases**.

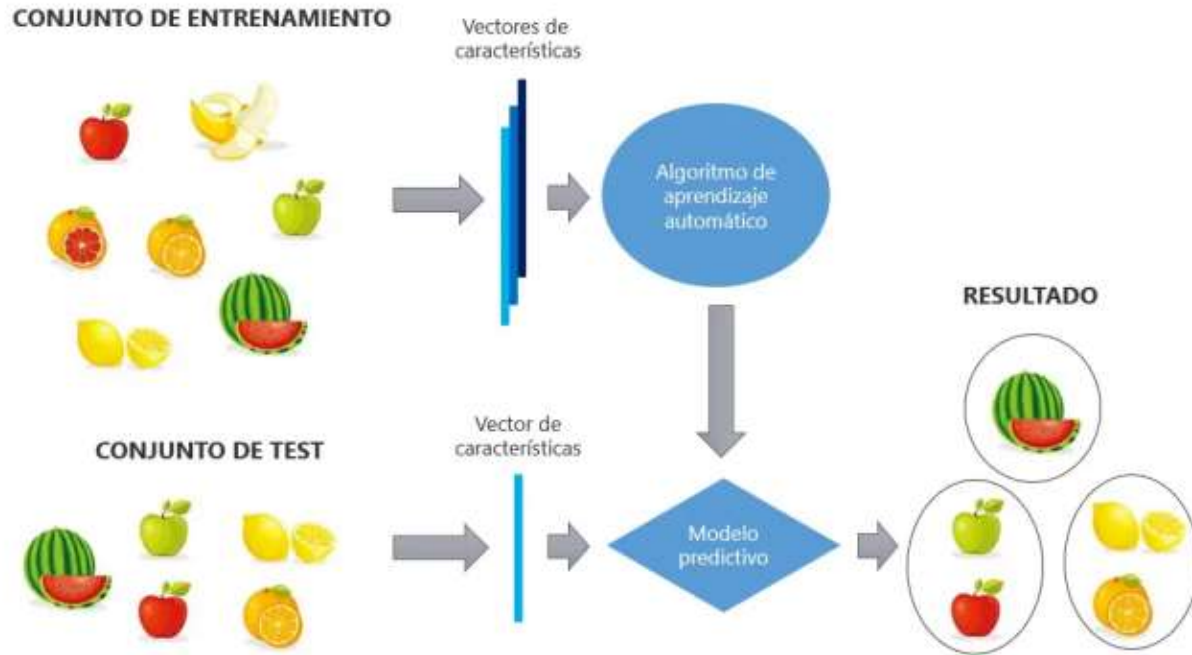
3. Aprendizaje automático no supervisado.



Aprendizaje no supervisado el caracteriza por lo siguiente:

- No disponemos de ejemplos con un objetivo o etiqueta conocido.
- En este tipo de problemas no se tiene conocimiento sobre las categorías o valores de los datos observados y se desea buscar la estructura oculta en los datos.
- El proceso de modelado se lleva a cabo sobre un conjunto de ejemplos que solo contiene las entradas del sistema, y el algoritmo tiene que ser capaz de reconocer patrones para diferenciar entre los datos existentes.

3. Aprendizaje automático no supervisado.



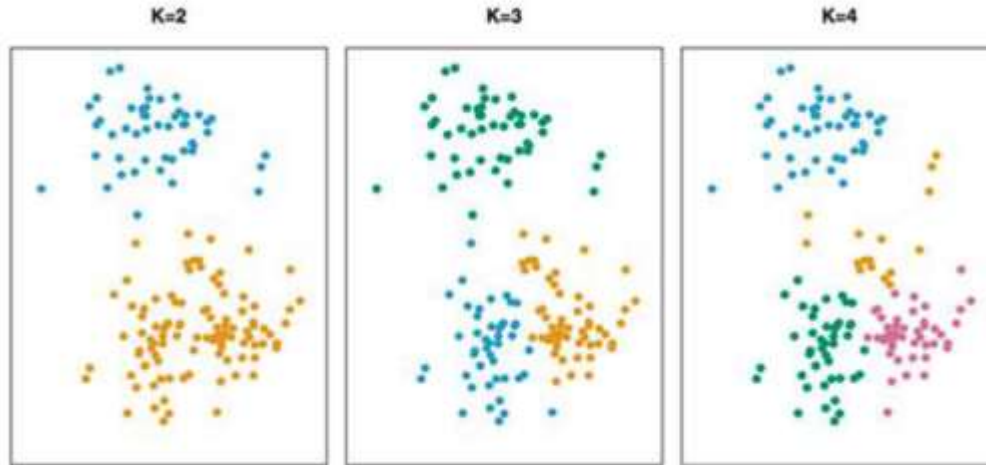
3. Aprendizaje automático no supervisado.



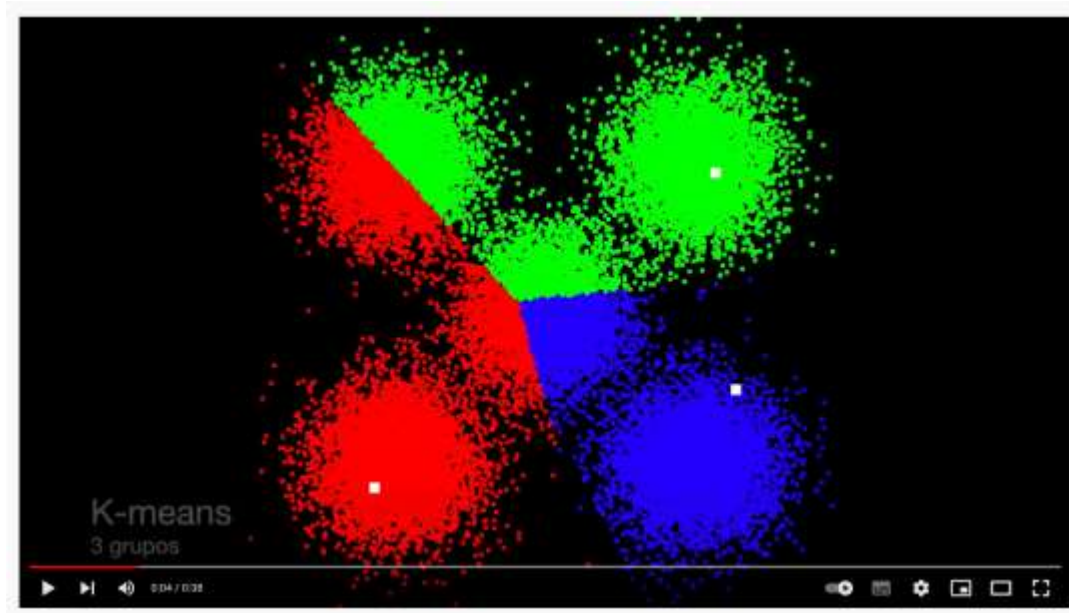
La evaluación de este tipo de técnicas no es sencilla porque no se conoce el número de errores producidos. En concreto, cualquier método de evaluación sobre este tipo de técnicas no pueden calcularse con las etiquetas asignadas a las instancias sino con la separación que el algoritmo hace utilizando para ello métricas de similitud entre los miembros de cada una de las clases.

3. Aprendizaje automático no supervisado.

En el siguiente ejemplo se muestra la separación entre clases que llevaría a cabo un algoritmo de agrupamiento para dos, tres y cuatro clases.



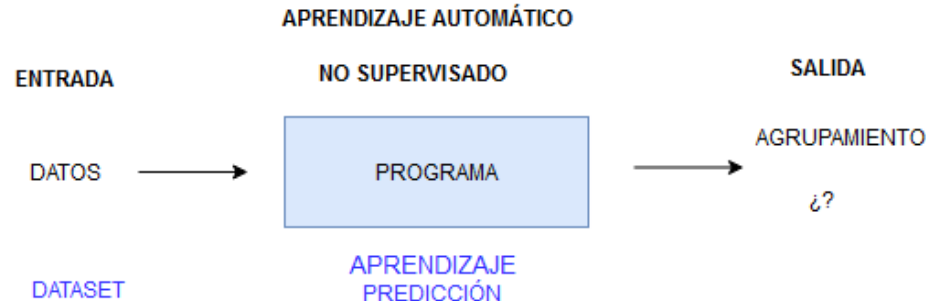
3. Aprendizaje automático no supervisado.



<https://www.youtube.com/watch?v=4WJv49WYVcl>

3. Aprendizaje automático no supervisado.

Las técnicas de **agrupamiento** son un ejemplo de aprendizaje no supervisado donde no se conocen las clases o valores de cada una de las instancias. El objetivo de estas técnicas es buscar la estructura oculta en los datos.



3. Aprendizaje automático no supervisado.



Una anomalía es una observación que es significativamente diferente del resto de observaciones.

La detección de anomalías implica que esa observación es sospechosa de haber sido generada por un mecanismo diferente del resto de observaciones.

Los algoritmos de detección de anomalías también son conocidos con el nombre de detección de outliers.

3. Aprendizaje automático no supervisado.

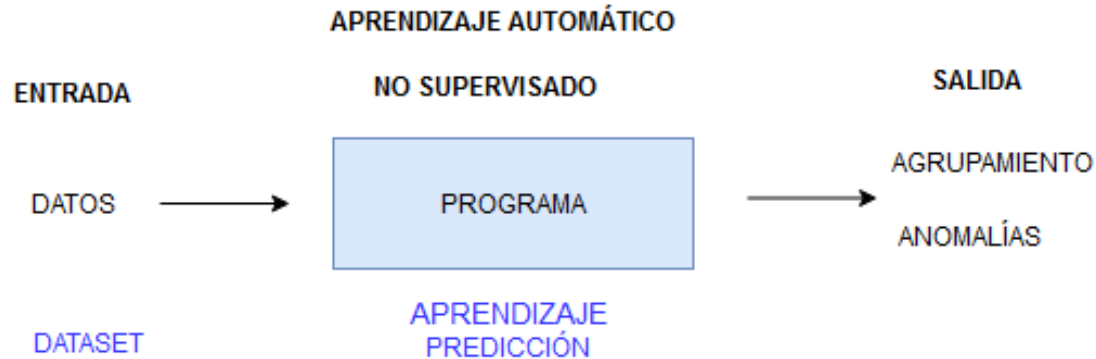


Estos problemas tienen gran aplicación en entornos como el fraude, la detección de intrusos, la detección de defectos estructurales, problemas médicos o errores en texto.

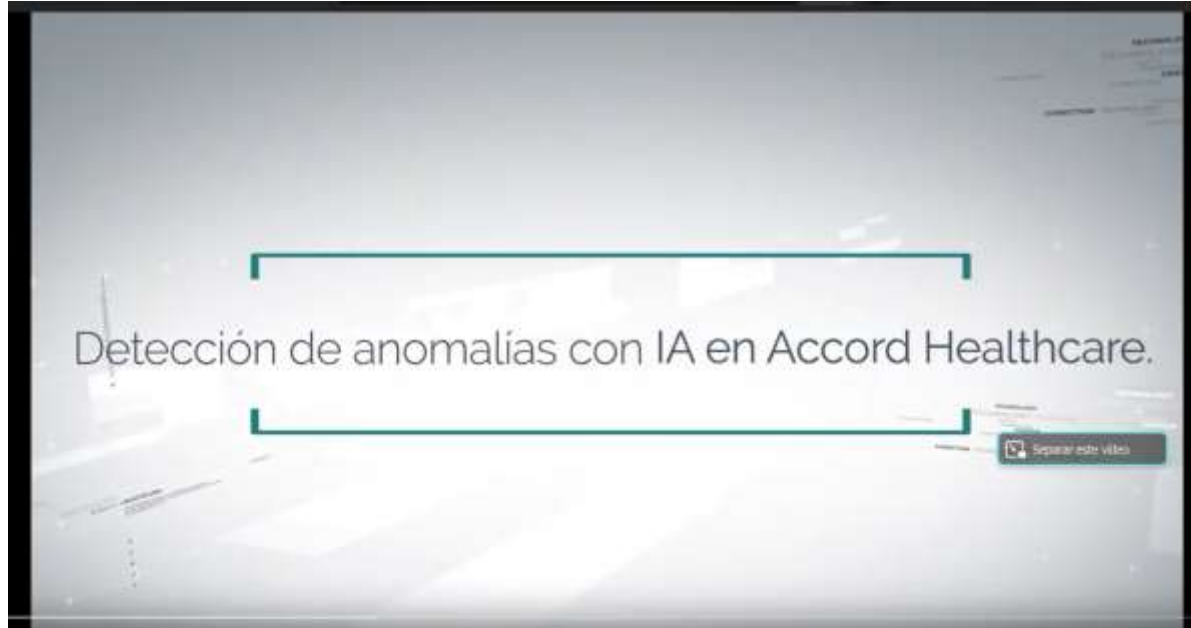
La detección de anomalías de forma no supervisada busca anomalías en conjuntos de datos sin etiquetar con la hipótesis de que la mayoría de las instancias en el conjunto de datos son normales y observando aquellas que menos se parecen a la mayoría.

3. Aprendizaje automático no supervisado.

Nuestro esquema de flujos de datos quedaría de la siguiente forma:



3. Aprendizaje automático no supervisado.



<https://www.youtube.com/watch?v=qk4tbnxy0-I>

4. Tipos de problemas aprendizaje automático:



Aprendizaje supervisado.

- Problemas de regresión.
- Problemas de clasificación

Aprendizaje no supervisado.

- Problemas de agrupamiento.
- Problemas de detección de anomalías.

4. Tipos de problemas aprendizaje automático:

Indicar un problema que se puede solucionar con varios tipos de aprendizaje automático.



4. Tipos de problemas aprendizaje automático:



A lo largo del **curso** y la **asignatura** veremos que un mismo **problema** se puede **solucionar de distintas formas**.

Por ejemplo, la detección de anomalías se puede solucionar utilizando aprendizaje no supervisado, si no conocemos ejemplos previos de anomalías. O bien utilizando aprendizaje supervisado para entrenar un clasificador que distinga entre ejemplos normales o anormales.

5. Conjuntos de datos de entrenamiento

Si nos proporcionan
un conjunto de datos
de entrenamiento.

¿ Qué tipo de
aprendizaje es ?



5. Conjuntos de datos de entrenamiento



La clave es su capacidad de generalizar situaciones del futuro en función de los datos históricos observados, por ejemplo:

- Predecir la temperatura que va a hacer mañana a una determinada hora sería una **regresión**.
- Detectar que ciudadanos al entrar al metro tienen puesta o no la mascarilla sería una **clasificación**.

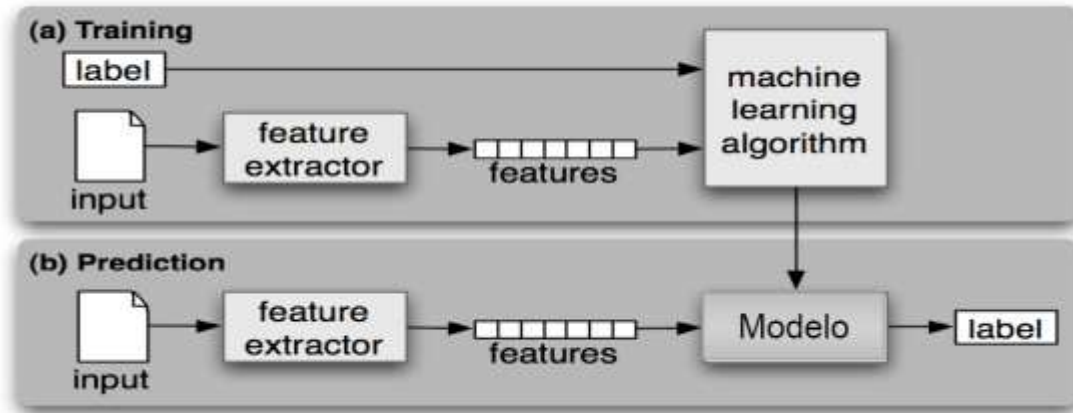
5. Conjuntos de datos de entrenamiento

Ahora vamos a ver como funciona el software o programa a muy alto nivel.



5. Conjuntos de datos de entrenamiento

Esquema funcionamiento aprendizaje supervisado.



5. Conjuntos de datos de entrenamiento



En la fase de entrenamiento (a) se extraen las características o variables relevantes de los datos de entrada para construir un modelo por medio de algoritmos de aprendizaje automático.

Posteriormente en la fase predicción (b) se realiza una extracción de variables similares sobre las que se aplica el modelo previamente entrenado para obtener el resultado estimado.

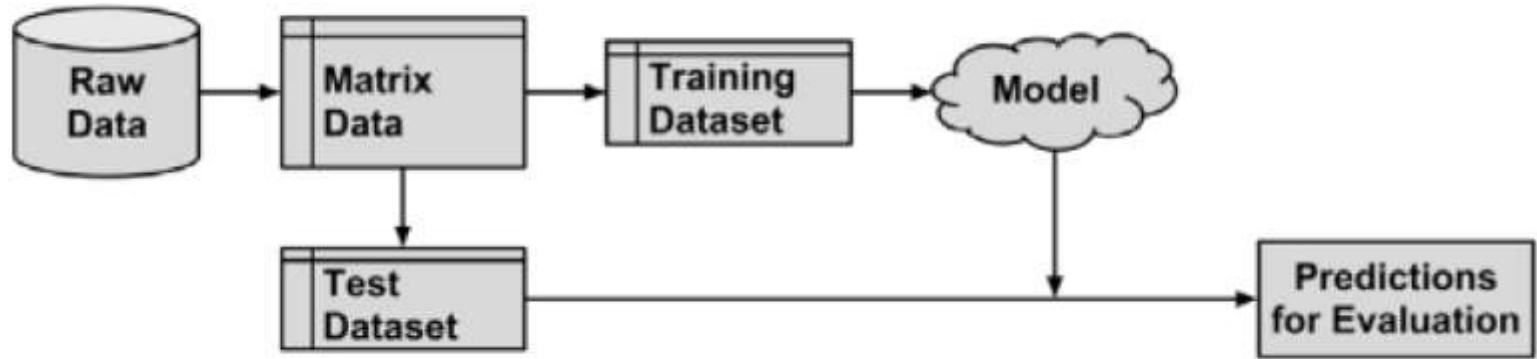
5. Conjuntos de datos de entrenamiento



Un modelo puede ser capaz de tener un error mínimo con datos históricos (conjunto de entrenamiento) y no ser capaz de predecir bien los valores futuros. Por esta razón el error en el conjunto de entrenamiento suele ser un valor muy optimista y debe de ser interpretado con cautela.

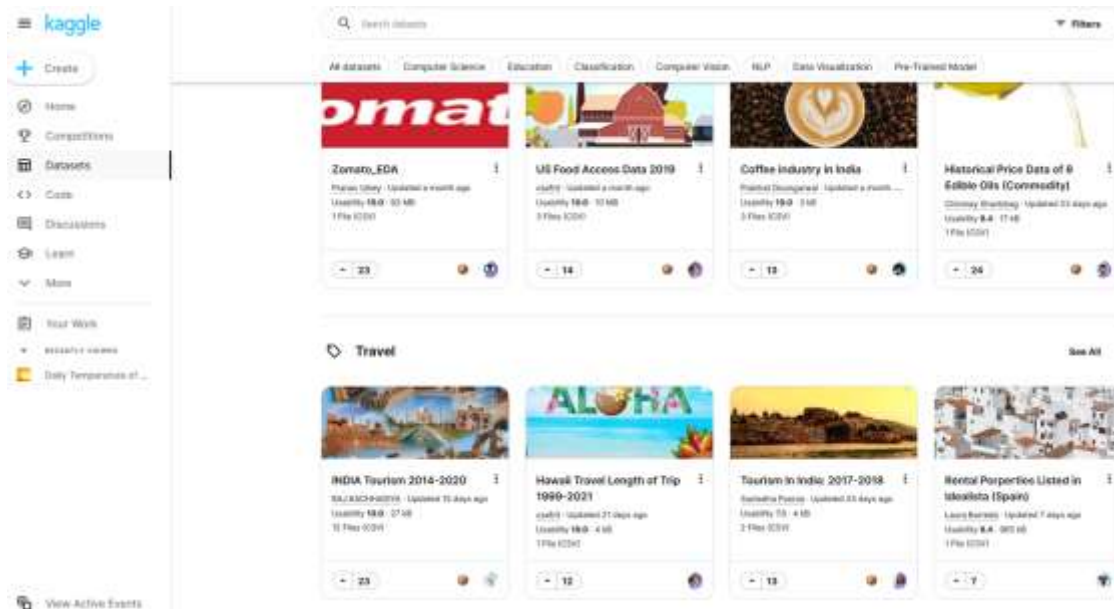
Para solucionar el problema anterior, una buena práctica es utilizar un conjunto de datos de test. Este conjunto de datos de test puede estar formado con un subconjunto de los datos de entrenamiento.

5. Conjuntos de datos de entrenamiento



5. Conjuntos de datos de entrenamiento

Ejemplo: <https://www.kaggle.com/datasets>



5. Conjuntos de datos de entrenamiento

A	B	C	D	E	F	G	H	I
	provincia	comunidad autor	título	precio	habitaciones	metros	total inmuebles/comunidad	
0	La Coruña	Galicia					1751	
1	La Coruña	Galicia	Chalet pareado en A Zap	1100	4	243	1751	
2	La Coruña	Galicia	Chalet pareado en Elmña	1100	4	243	1751	
3	La Coruña	Galicia	Ático en avenida Pasaxe	450		Planta 4ª exterior con asce	1751	
4	La Coruña	Galicia	Piso en calle de Fernanc	1000	4	100	1751	
5	La Coruña	Galicia	Piso en calle Manuel Jes	450	2	76	1751	
6	La Coruña	Galicia	Piso en Os Mallos, A Co	800	4	100	1751	
7	La Coruña	Galicia	Piso en ronda de Outeiro	650	3	122	1751	
8	La Coruña	Galicia	Piso en travesía de Meic	400	4	107	1751	
9	La Coruña	Galicia	Piso en calle Concepción	800	2	75	1751	
10	La Coruña	Galicia	Piso en calle da Pía, Ca	450	3	99	1751	
11	La Coruña	Galicia	Piso en ronda de nelle, 1	600	3	87	1751	
12	La Coruña	Galicia	Piso en Someso - Matog	825	3	80	1751	
13	La Coruña	Galicia	Piso en Juan Flórez-San	700	2	80	1751	
14	La Coruña	Galicia	Ático en Editor Francisco	245	1	46	1751	
15	La Coruña	Galicia	Chalet pareado en calle	550	4	90	1751	
16	La Coruña	Galicia	Piso en avenida de Villag	750	3	95	1751	
17	La Coruña	Galicia	Dúplex en calle Garrucha	650	2	90	1751	
18	La Coruña	Galicia	Piso en calle Rosalía de	360	2	65	1751	
19	La Coruña	Galicia	Piso en plaza de La Blaz	850	3	110	1751	

5. Conjuntos de datos de entrenamiento

	A	B	C	D	E	F	G	H	I
1725	1723	La Coruña	Galicia	Dúplex en plaza de Salu	2000	2	Planta 3ª exterior con asce	1751	
1726	1724	La Coruña	Galicia	Piso en calle Arquitecto	800	5		100	1751
1727	1725	La Coruña	Galicia	Piso en Salgueiros de	418	1		50	1751
1728	1726	La Coruña	Galicia	Piso en calle Panasquei	370	2		60	1751
1729	1727	La Coruña	Galicia	Piso en Escorial, 2. Ribi	1200	4		140	1751
1730	1728	La Coruña	Galicia	Piso en calle Antonio VII	615	3		95	1751
1731	1729	La Coruña	Galicia	Piso en ronda de Nella,	470	4		115	1751
1732	1730	La Coruña	Galicia	Piso en calle Agri do Mi	485	2		80	1751
1733	1731	La Coruña	Galicia	Piso en camoteira Baños	850	3		100	1751
1734	1732	La Coruña	Galicia	Piso en calle Frouxeiras,	500	2		64	1751
1735	1733	La Coruña	Galicia	Piso en avenida cristina,	350	2		70	1751
1736	1734	La Coruña	Galicia	Piso en Lugar Vixoi, Bar	330	2		60	1751
1737	1735	La Coruña	Galicia	Piso en calle das Oblata	550	2		56	1751
1738	1736	La Coruña	Galicia	Piso en calle Alameda, 4	630	3		72	1751
1739	1737	La Coruña	Galicia	Piso en Xosé Ramón Ba	475	3		105	1751
1740	1738	La Coruña	Galicia	Piso en travesía Praia, 1	1150	1		47	1751
1741	1739	La Coruña	Galicia	Piso en Ciudad Vieja - C	390	1		35	1751
1742	1740	La Coruña	Galicia	Piso en capitán Juan var	700	3		80	1751
1743	1741	La Coruña	Galicia	Piso en calle San Pedro,	700	3		100	1751
1744	1742	La Coruña	Galicia	Piso en calle Vázquez d	550	2		78	1751
1745	1743	La Coruña	Galicia	Piso en Rúaªa Fanny Gar	650	2		59	1751
1746	1744	La Coruña	Galicia	Dúplex en calle Concepç	680	4		154	1751
1747	1745	La Coruña	Galicia	Piso en calle de San Pe	1100	4		114	1751
1748	1746	La Coruña	Galicia	Piso en Os Matos, A Ce	400	2		80	1751
1749	1747	La Coruña	Galicia	Piso en puerta del Camir	1000	4		200	1751
1750	1748	La Coruña	Galicia	Piso en calle Estanque,	590	2		75	1751
1751	1749	La Coruña	Galicia	Piso en Monte Alto - Zali	850	3		110	1751
1752	1750	La Coruña	Galicia	Piso en Doctor Teixeiro,	1200	2		107	1751
1753	1751	La Coruña	Galicia	Piso en O Milladoiro, An	420	2		75	1751
1754	1752	Alava	País Vasco	Piso en calle Sol, 7. Moi	580	1		65	138
1755	1753	Alava	País Vasco	Dúplex en plaza de Salu	2000	2	Planta 3ª exterior con asce	1751	

5. Conjuntos de datos de entrenamiento



Si queremos predecir el precio del alquiler La Coruña (Regresión).

Tenemos 1751 datos de entrenamiento ya que sabemos el precio (es la etiqueta)

Lo que hacemos es dividir el datos en dos conjuntos:

- 80 % (aprox) datos para el entrenamiento.
- 20 % (aprox) datos de test. Con los datos de test vamos a saber el error real de nuestro modelo.