



# Máster Smart Energy

Postgrado de DIGITAL ENERGY

Curso 2022/2023

## Sesión 4 – Modelos de Aprendizaje Supervisado (I): Clasificación

Sara Barja: [sara.barja@upc.edu](mailto:sara.barja@upc.edu)



# Calendario

	Lunes	Martes	Miércoles	Jueves
ABRIL	10	11	12 S1 – Introducción a Machine Learning	13
	17 S2 – Introducción a Python	18	19 S3 – Estadística descriptiva	20
	24 S4 – Modelos aprendizaje supervisado (I): Clasificación	25	26 S5 – Modelos de aprendizaje supervisado (II): Regresión	27
MAYO	1	2	3 S6 – Aplicación de AI en el sector eléctrico: Odit-e	4
	8 S7 – Modelos de aprendizaje no supervisado	9	10 S8 – Examen final	

Sara Barja : [sara.barja@upc.edu](mailto:sara.barja@upc.edu)

Marc Jené: [marc.jene@upc.edu](mailto:marc.jene@upc.edu)



## Objetivos de la sesión

- Conceptos generales.
- Tipos de clasificación.
- Aprendizaje supervisado clasificación: aplicaciones.
- Detectar y resolver el desequilibrio de clases.
- Principales modelos de clasificación.
- Aprender las métricas de evaluación de la clasificación.
- Presentación de un ejemplo práctico de un modelo de clasificación.



## **Conceptos generales.**

Tipos de Clasificación.

Aplicaciones de Clasificación.

Detectar y resolver el desequilibrio de clases.

Principales modelos de Clasificación.

Métricas de evaluación: Clasificación.

Ejemplo práctico de un modelo de Clasificación.



## Recapitulemos...

### MACHINE LEARNING

Aprendizaje  
Supervisado

Aprendizaje **NO**  
Supervisado

Reinforcement  
Learning

**Clasificación**

Clustering

Regresión

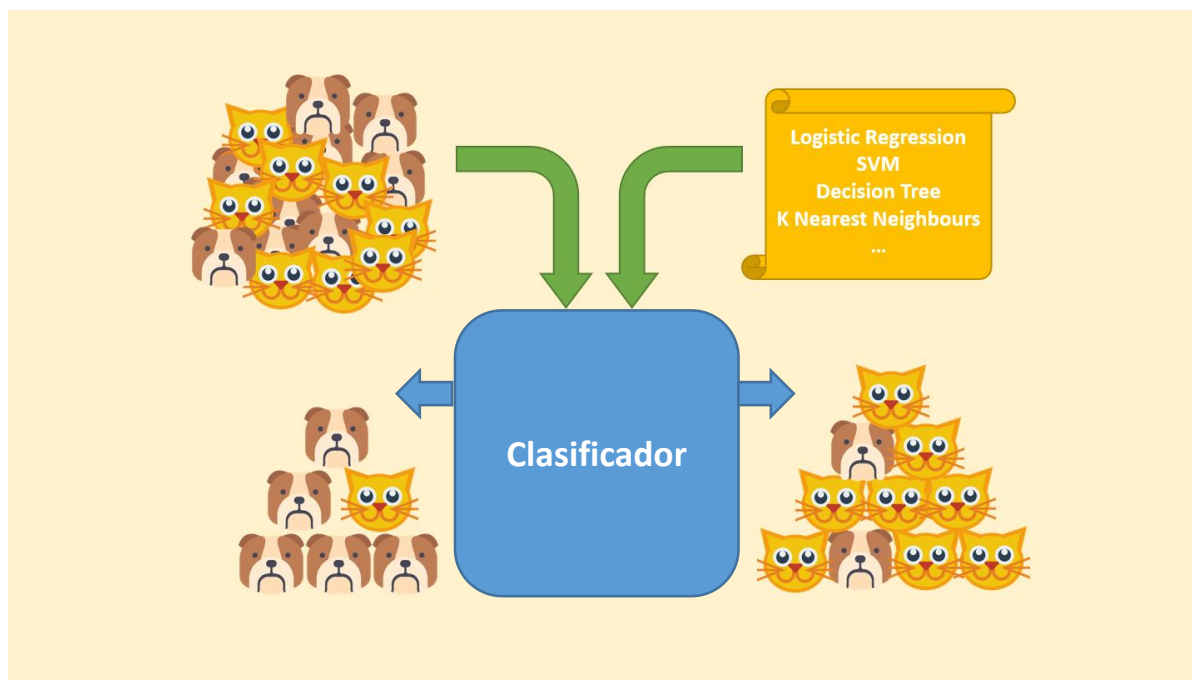
Reducir dimensiones

- Es necesario disponer de las variables de entrada y salida para entrenar
- Se utiliza para explorar grandes conjuntos de datos.
- Solo se necesitan datos de entrada



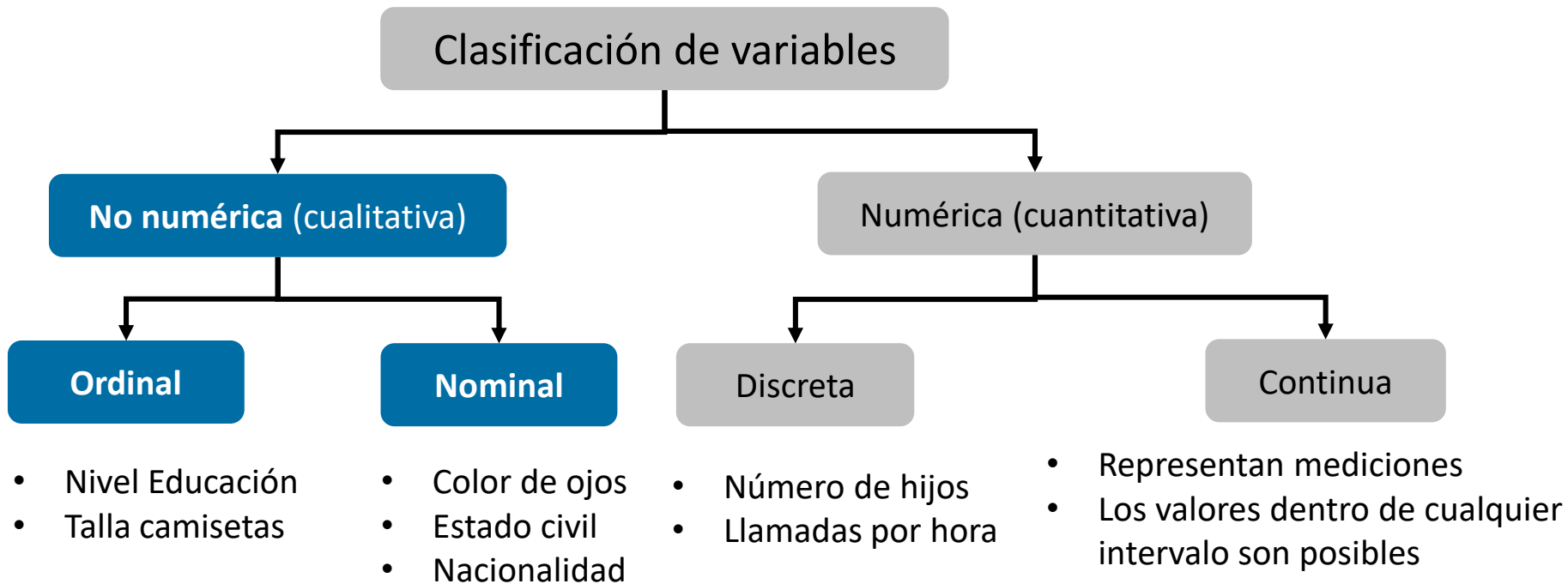
## ¿Qué es clasificación en Machine Learning?

La **clasificación** es una subcategoría del **aprendizaje supervisado** donde el objetivo es predecir una **clase categórica** basada en un conjunto de características/atributos y observaciones pasadas.





## Clasificación de variables



Es necesario codificar las variables no numéricas para presentarlas como números.



## ¿Cómo pasar de variables cualitativas a numéricas?

Los modelos de machine learning y deep learning requieren que todas las variables de entrada y salida sean **numéricas**.

### Datos de entrada

ID	Country	Population
1	Japan	127185332
2	U.S	326766748
3	India	1354051854
4	China	1415045928
5	U.S	326766748
6	India	1354051854

Datos no numéricos

[Más información](#)

### LABEL ENCODER

ID	Country	Population
1	0	127185332
2	1	326766748
3	2	1354051854
4	3	1415045928
5	1	326766748
6	2	1354051854

El problema aquí es que como hay diferentes números en la misma columna, el modelo malinterpretará los datos y supondrá que existe algún tipo de orden,  $0 < 1 < 2$ .

### ONE HOT ENCODING

ID	Country_Japan	Country_U.S	Country_India	Country_China	Population
1	1	0	0	0	127185332
2	0	1	0	0	326766748
3	0	0	1	0	1354051854
4	0	0	0	1	1415045928
5	0	1	0	0	326766748
6	0	0	1	0	1354051854





Conceptos generales.

## **Tipos de Clasificación.**

Aplicaciones de Clasificación.

Detectar y resolver el desequilibrio de clases.

Principales modelos de Clasificación.

Métricas de evaluación: Clasificación.

Ejemplo práctico de un modelo de Clasificación.



## Tipos de Clasificación

### CLASIFICACIÓN BINARIA

- Solo se pueden asignar dos clases diferentes (**0 o 1**).
- Cada observación sólo puede ser etiquetada como una clase.

### CLASIFICACIÓN MULTI-CLASE

- Se asignan múltiples categorías a las observaciones.
- Cada observación sólo puede ser etiquetada como una clase.

### CLASIFICACIÓN MULTI-ETIQUETA

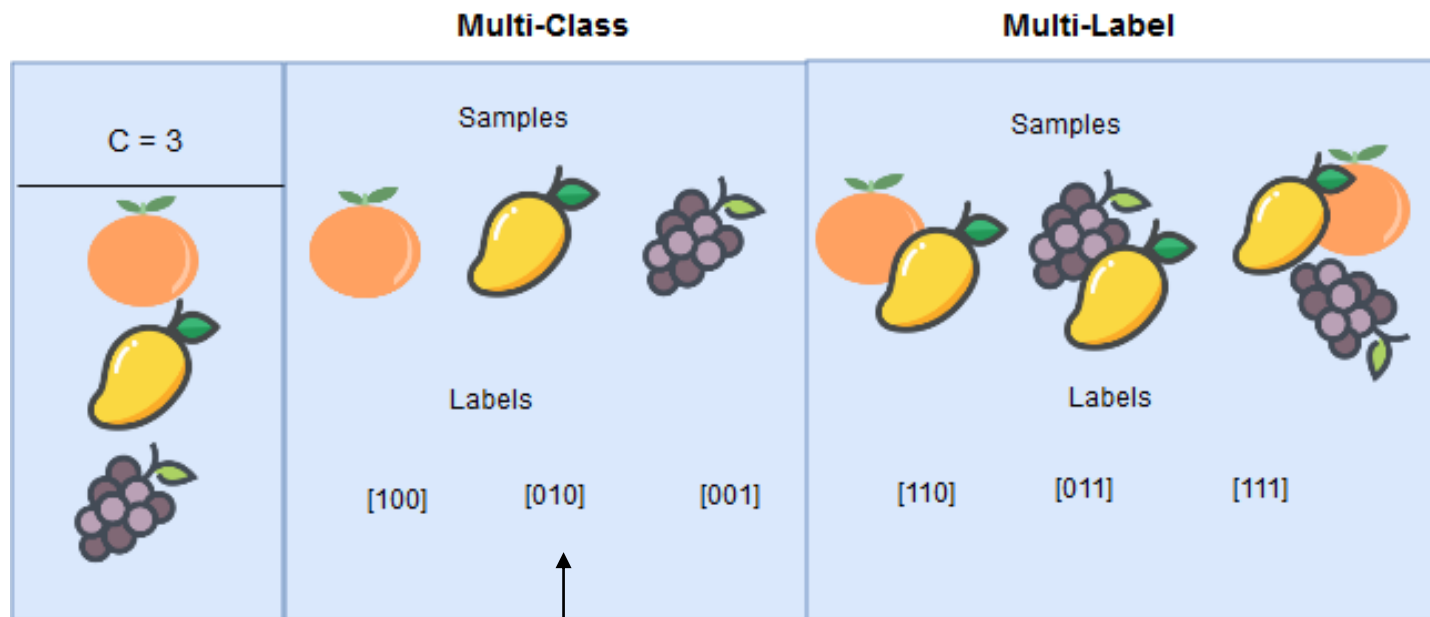
- Se pueden asignar múltiples categorías a las observaciones.
- Predice más de una etiqueta de clase para cada observación.

[Más info](#)

TIPOS DE CLASIFICACIÓN	NÚMERO DE ETIQUETAS	NÚMERO DE CLASES DIFERENTES
Clasif. binaria	1	2
Clasif. multi-clase	1	>2
Clasif. multi-etiqueta	>1	>2



## Multi-Class vs Multi-Etiqueta



One hot encoding



Conceptos generales.

Tipos de Clasificación.

## **Aplicaciones de Clasificación.**

Detectar y resolver el desequilibrio de clases.

Principales modelos de Clasificación.

Métricas de evaluación: Clasificación.

Ejemplo práctico de un modelo de Clasificación.

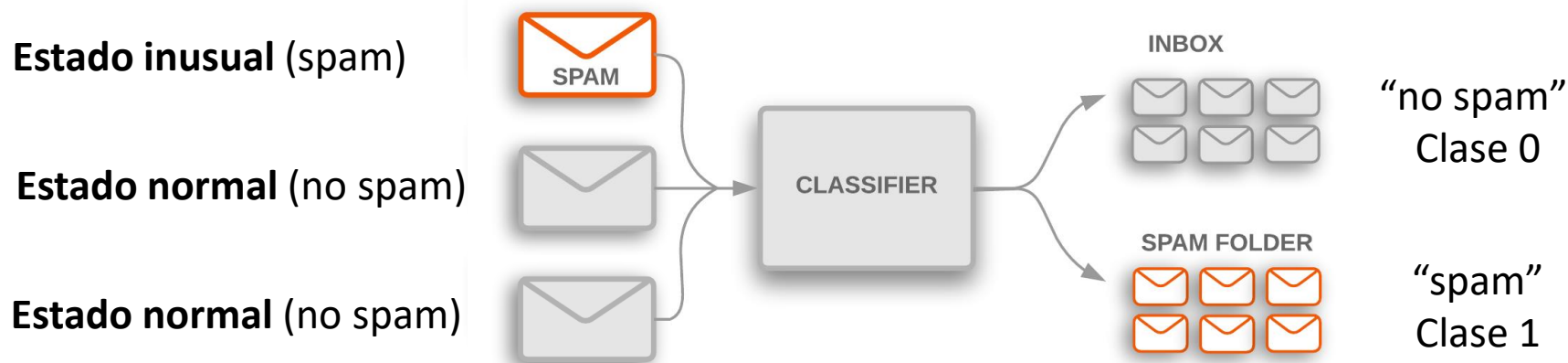


## Aplicaciones clasificación

### APLICACIONES CLASIFICACIÓN BINARIA

- Detección de spam (“spam” o “no spam”)

Las tareas de clasificación binaria implican una clase que es el **estado normal** (etiqueta de clase 0) y otra clase que es el **estado inusual** (etiqueta de clase 1).





## Aplicaciones clasificación

### APLICACIONES CLASIFICACIÓN BINARIA

- Covid detection (“COVID-19” or “non-COVID-19”)

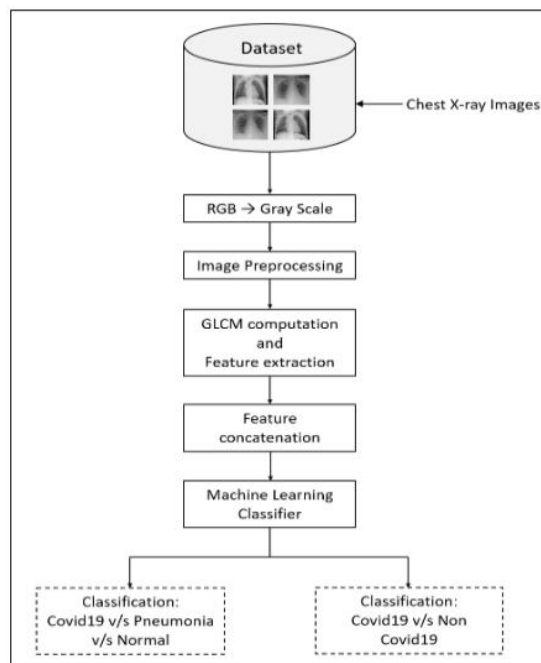


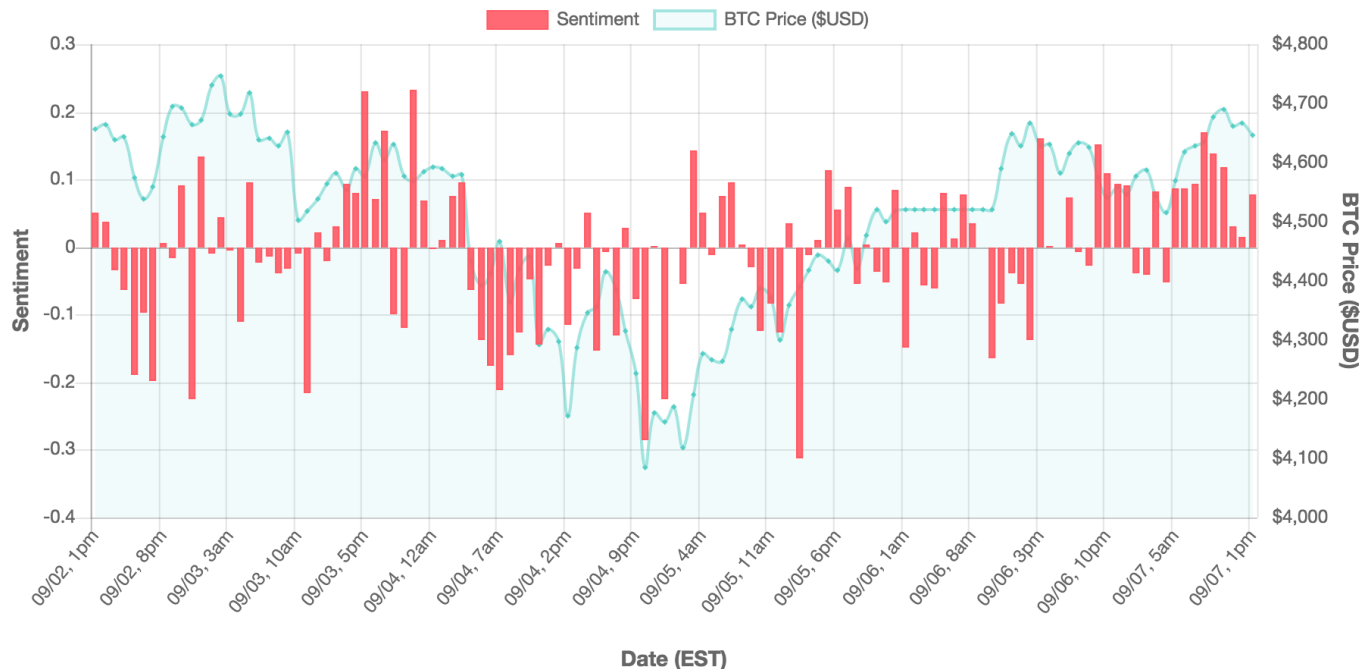
Figure 1: Block Diagram Describing Procedure of Identification of Covid19 using GLCM Feature Extraction Method



## Aplicaciones clasificación

### APLICACIONES CLASIFICACIÓN BINARIA

- **Detección de clientes descontentos** (“contento” o “no contento”)
  - Conocido también como como **Sentiment Analysis**.
  - Se utiliza el método **Natural Language Processing (NLP)**





## Aplicaciones clasificación

### APLICACIONES CLASIFICACIÓN MULTI-CLASE

- Reconocimiento de imágenes: identificar si se trata de un perro, gato, elefante o serpiente.
- Clasificación de caras (Image recognition).
- **Clasificación de las especies de plantas**
- Reconocimiento óptico de caracteres

Más de **10.000** especies de plantas diferentes



**PICTURE THIS**

<https://www.picturethisai.com/>



**Diagnóstico automático de problemas**

Haz una foto de las partes enfermas de una planta para conseguir las causas del problemas y sugerencias de tratamiento.



**Identifica plantas con una foto**

Solo tienes que hacer o subir una foto de una planta para conseguir resultados de identificación de la planta precisos e instantáneos gracias a nuestra revolucionaria tecnología de inteligencia artificial.

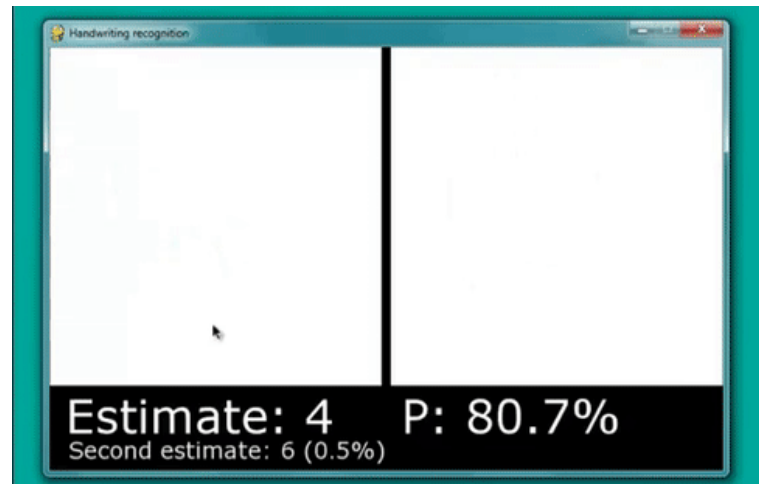
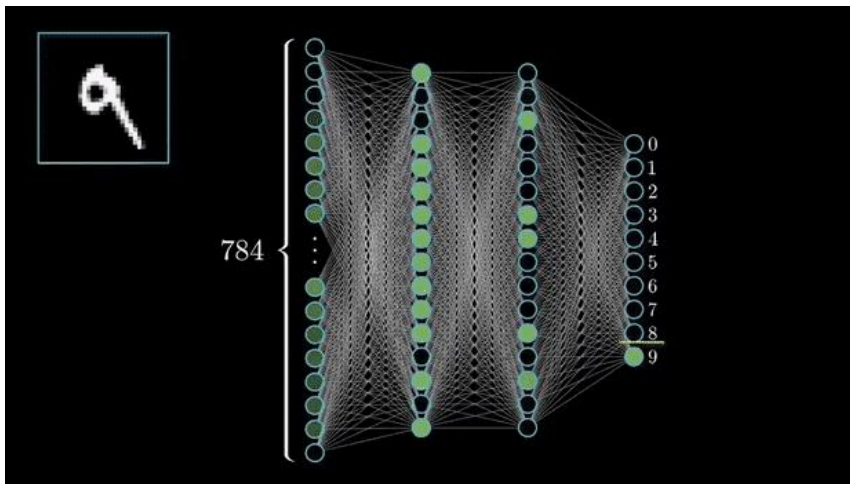




## Aplicaciones clasificación

### APLICACIONES CLASIFICACIÓN MULTI-CLASE

- Reconocimiento de imágenes: identificar si se trata de un perro, gato, elefante o serpiente.
- Clasificación de caras (Image recognition).
- Clasificación de las especies de plantas
- **Reconocimiento óptico de caracteres**



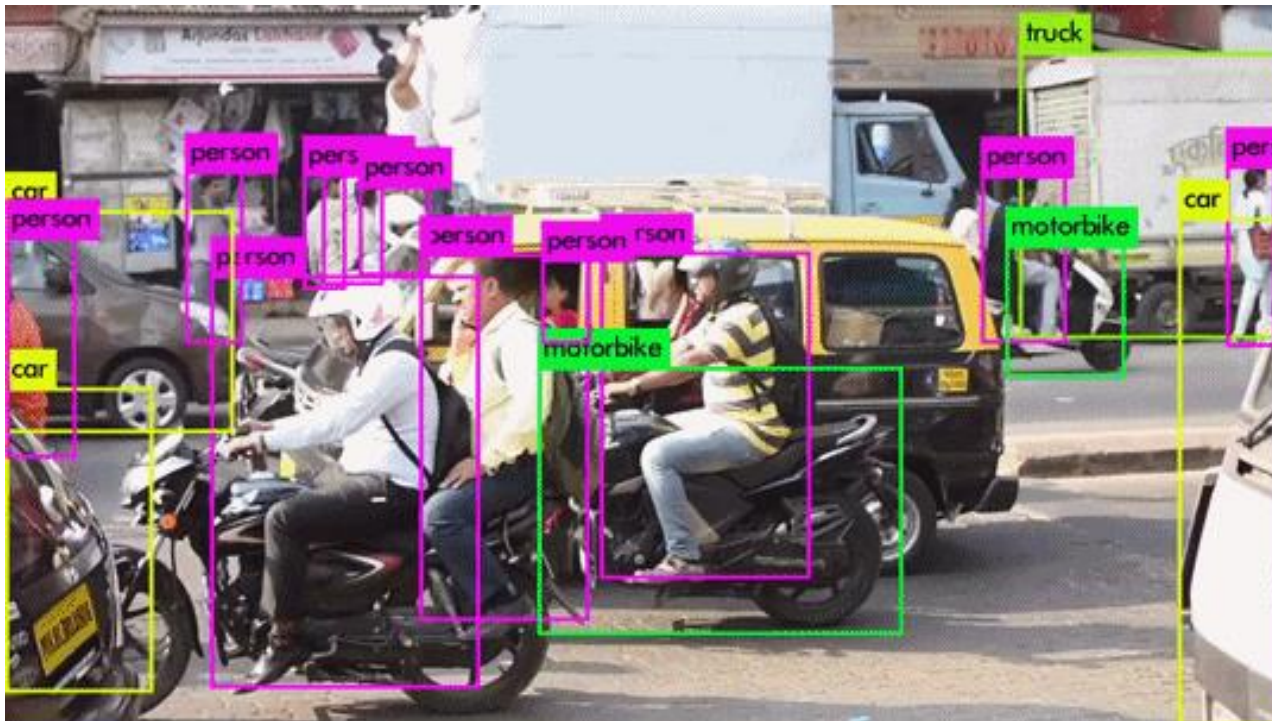
Reconocimiento óptico de caracteres



## Aplicaciones clasificación

### APLICACIONES CLASIFICACIÓN MULTI-ETIQUETA

- **Reconocimiento** de múltiples objetos conocidos en una fotografía (bicicleta, manzana, persona...).





## Aplicaciones clasificación

### APLICACIONES CLASIFICACIÓN MULTI-ETIQUETA

- **Clasificación de noticias:** un artículo puede ser sobre **deportes**, una **persona** y un **lugar** al mismo tiempo.

**Table 1.** Class label of Indonesian new

Class Label	Explanation
1	Politic
2	Law
3	Economy
4	Social
5	Culture
6	Technology
7	Life Style
8	Sport
9	Entertainment
10	Education
11	Defense
12	Health
13	Others

**Table 2.** Dataset of Indonesian news and label.

No	News Article	Class Label
1	The dynamics of dismantling the 2019 presidential candidate pair continues. Whoever is a floating figure can be juxtaposed. Moreover, a number of continually emerging are considered alternative candidates.	1
2	Today (30/11) Butet Kartaredjasa will present 138 visual works at the exhibition at the National Gallery Jakarta. The visual work takes the painted media. This way of art is deliberately done by Butet to conduct social criticism of the community and the government.	5,9
3	The Vice President of the Republic of Indonesia, Jusuf Kalla delivered a written warning to the Minister of Youth and Sports Imam Nahrawi to allocate more funds to prepare for a number of sports ahead of the 2018 Asian Games.	1,3,8



## Aplicaciones clasificación

### LA VANGUARDIA (22/04/2021)

EL AVANCE DE LAS NUEVAS TECNOLOGÍAS

## Europa no quiere un Gran Hermano: límites estrictos a la vigilancia con inteligencia artificial

- El despliegue normativo presentado por la Comisión Europea es restrictivo, con prohibición de la vigilancia masiva, salvo contadas excepciones



Una pantalla muestra los datos que proporciona un sistema de reconocimiento de peatones y vehículos en Pekín (Giles Sabine / Bloomberg)





## Aplicaciones clasificación

LA VANGUARDIA (08/04/2023)

 LA VANGUARDIA 

### Barcelona instalará 17 cámaras inteligentes en el paseo de Gràcia

- El dispositivo redoblará la presión sobre los delincuentes multirreincidentes



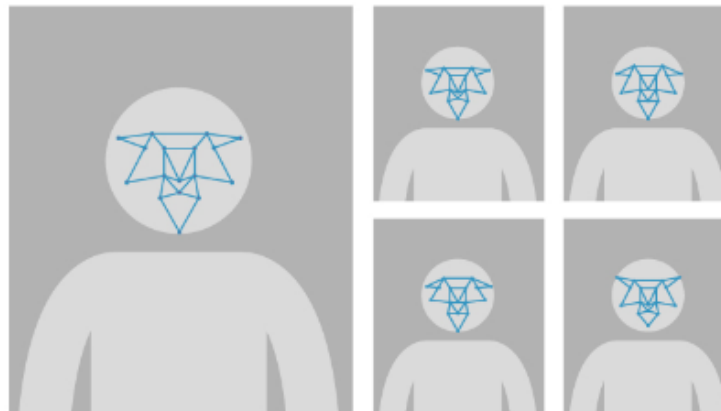
Estos aparatos permitirán clasificar a los viandantes en función de su género, edad, ropa, gafas...



## Aplicaciones clasificación

### Así funciona la técnica biométrica de verificación de identidad

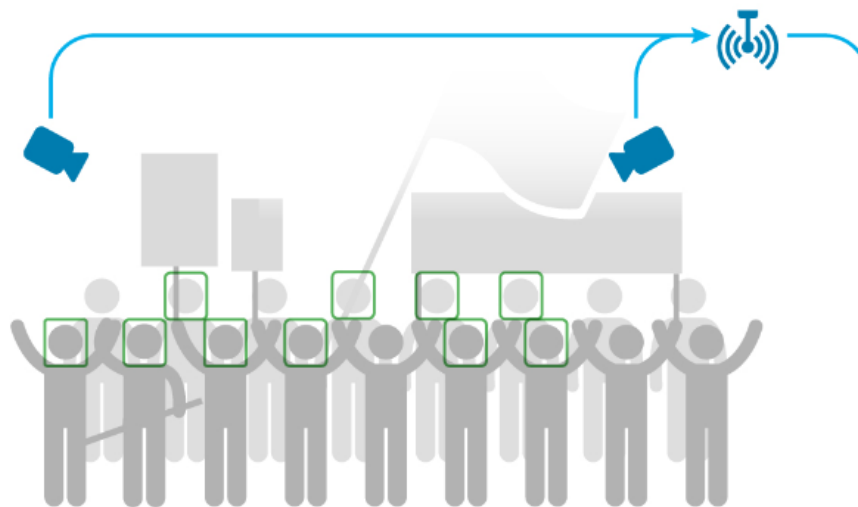
- 1 Previamente, un software **escanea los rostros de las fotografías** de la base de datos policial y extrae los rasgos geométricos de cada cara



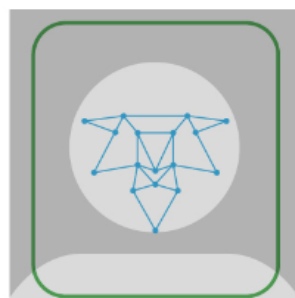


## Aplicaciones clasificación

- ② Unas **cámaras** instaladas en zonas estratégicas **escanean a los asistentes** a un evento masivo o en zonas de paso muy concurridas



- ③ Las imágenes de los rostros **se comparan en tiempo real con la base de datos** policial...

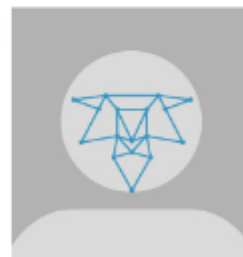
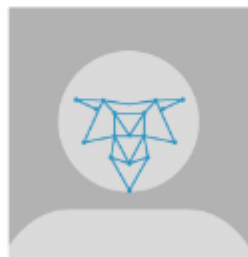
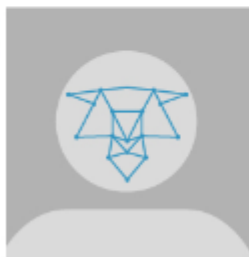
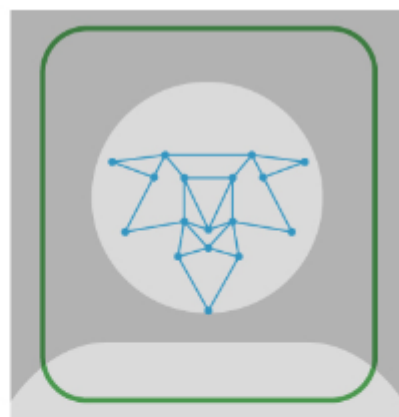




## Aplicaciones clasificación

LA VANGUARDIA (08/04/2023)

- ③ Las imágenes de los rostros se comparan en tiempo real con la base de datos policial...

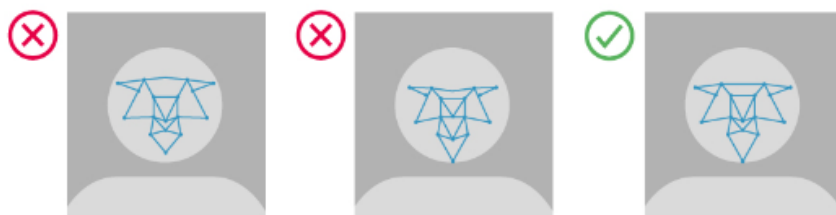


...detectando posibles coincidencias



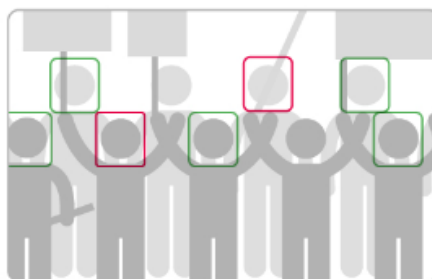


## Aplicaciones clasificación



...detectando **posibles coincidencias**

- ④ De esta manera la policía y otros cuerpos de seguridad puede **visualizar al momento si hay coincidencias**





Conceptos generales.

Tipos de Clasificación.

Aplicaciones de Clasificación.

**Detectar y resolver el desequilibrio de clases.**

Principales modelos de Clasificación.

Métricas de evaluación: Clasificación.

Ejemplo práctico de un modelo de Clasificación.



## Desequilibrio de clases

**DESEQUILIBRIO ENTRE CLASES:** sucede cuando las clases (etiquetas) no se encuentran balanceadas. El modelo predecirá peor la clase con menos observaciones al tener menos instancias con las que entrenar.

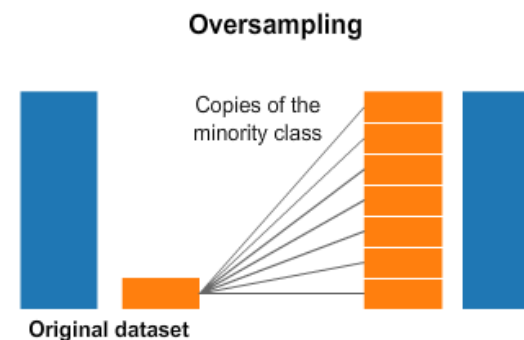
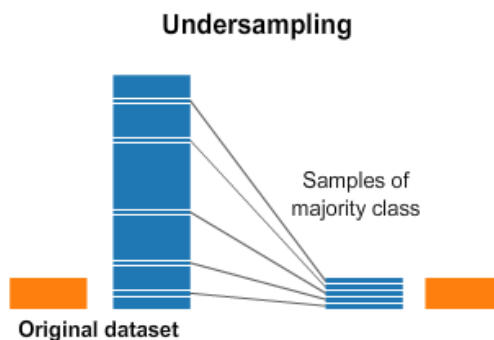
### SOLUCIONES

- **Conseguir más datos** (si es posible).
- **Sobre-muestreo (over-sampling):** duplicación de las observaciones de la clase minoritaria.
- **Sub-muestreo (under-sampling):** eliminación de datos de clase mayoritaria.
- Cambiar la **métrica de evaluación**.
- Generar **datos sintéticos** de la clase minoritaria.
- **StratifiedShuffleSplit()**

### EJEMPLOS

- Detección de fraude energético.

Referencia: [8 Tactics to Combat Imbalanced Classes in Your Machine Learning Dataset](#)





Conceptos generales.

Tipos de Clasificación.

Aplicaciones de Clasificación.

Detectar y resolver el desequilibrio de clases.

## **Principales modelos de Clasificación.**

Métricas de evaluación: Clasificación.

Ejemplo práctico de un modelo de Clasificación.



## Modelos supervisados de Clasificación

Se estudian los siguientes modelos:

- Logistic Regression
- Support Vector Machines (SVM)
- K-nearest neighbors (k-NN)
- Árboles de decision
- Ensemble Random Forest
- Neural Network (MLP)



## Paramétricos vs No paramétricos

**Paramétricos:** el modelo tiene un conjunto limitado de parámetros

- Logistic Regression
  - Naïve Bayes
  - Redes neuronales
- 
- Eficientes: sencillos de entrenar
  - Menos complejos

**No paramétricos:** la complejidad aumenta con el número de muestras

- Vecinos más próximos K-NN
  - Kernel SVM
  - Árboles de decisión
- 
- Más flexibles
  - Computacionalmente costosos



## Logistic Regression

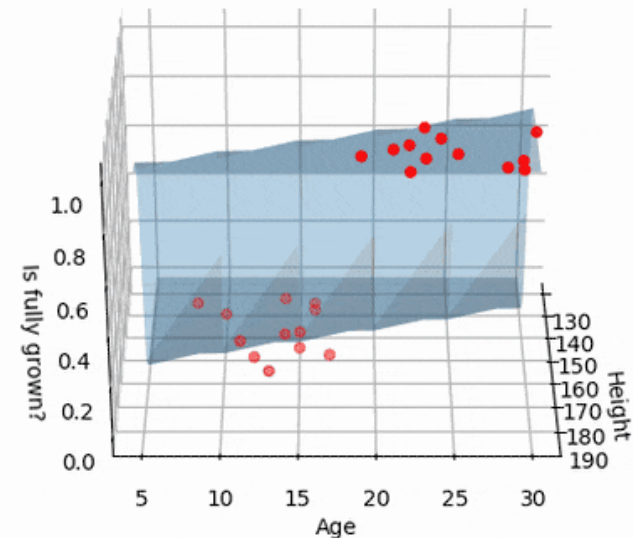
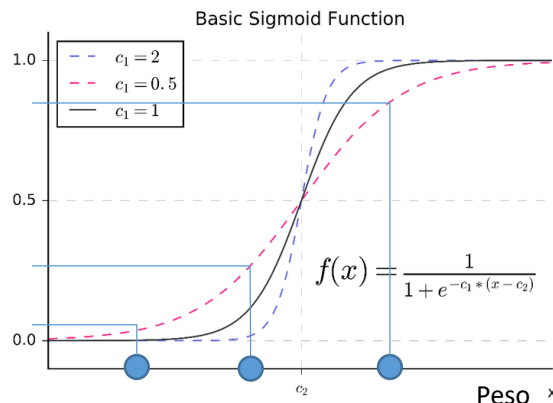
**DEFINICIÓN:** las probabilidades que predicen el resultado de una variable categórica se modelan utilizando una función logística “sigmoid function”.

**VENTAJAS:** uno de los más utilizados para clasificación binaria. Funciona muy bien en clases linealmente separables.

**DESVENTAJAS:** Funciona sólo para clasificación binaria.

Tiene  
sobrepeso

NO tiene  
sobrepeso



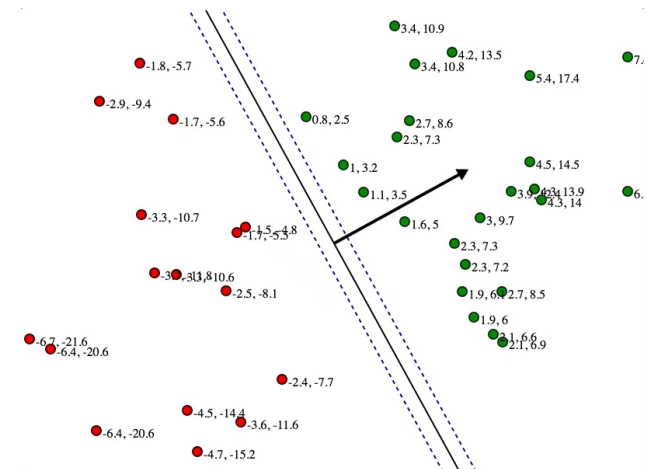
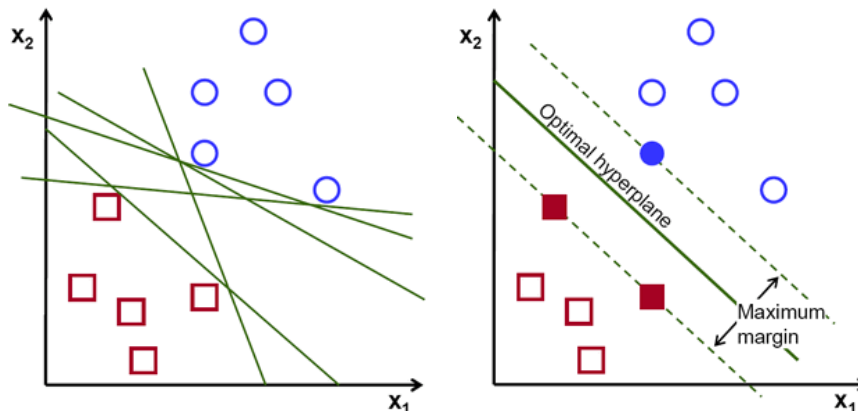
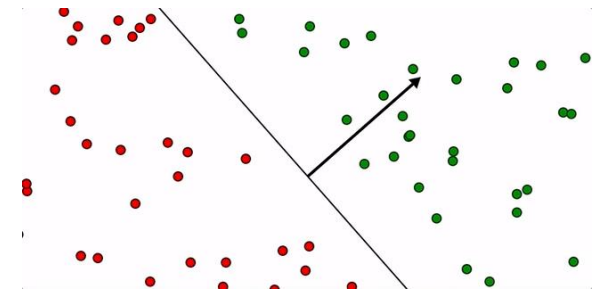


## Support Vector Machines

**DEFINICIÓN:** el objetivo de SVM es encontrar un hiperplano óptimo de decisión que clasifique los puntos de datos, maximizando el margen entre esta línea y los puntos de muestra cercanos a este hiperplano.

**VENTAJAS:** Es eficaz en espacios de grandes dimensiones y es eficiente en cuanto a la memoria.

**DESVENTAJAS:** el algoritmo no da buenos resultados si *características > muestras*.





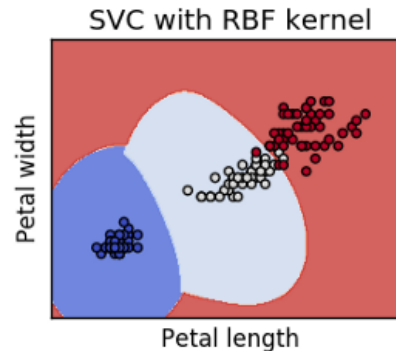
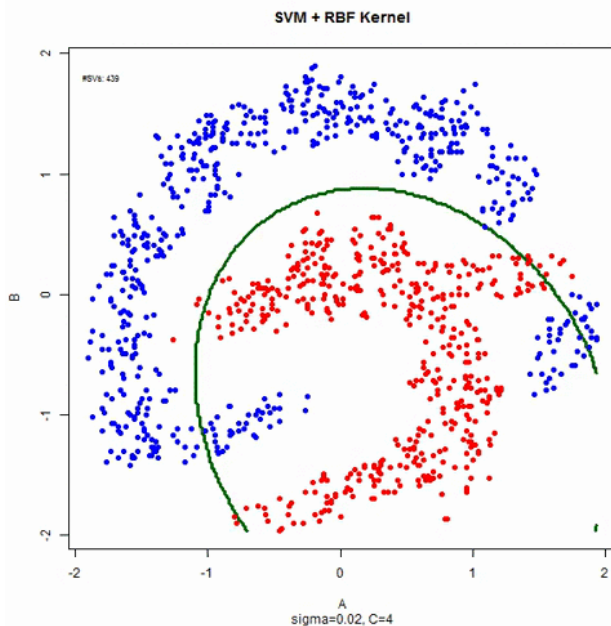


## Support Vector Machines

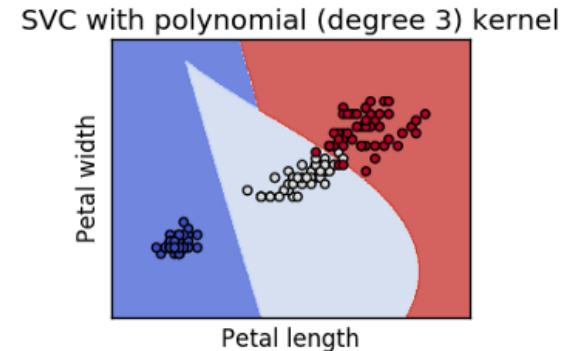
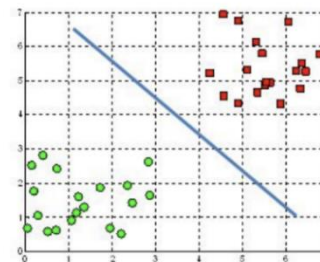
¿Pero qué pasa cuando los datos no son separables linealmente?

**SVM kernels:** método de análisis de patrones.

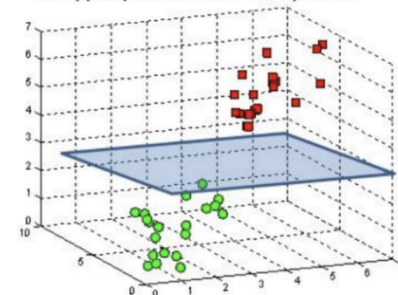
- Curvas no lineales de separación
- Casos donde los conjuntos de datos no pueden ser completamente separados.



A hyperplane in  $\mathbb{R}^2$  is a line



A hyperplane in  $\mathbb{R}^3$  is a plane



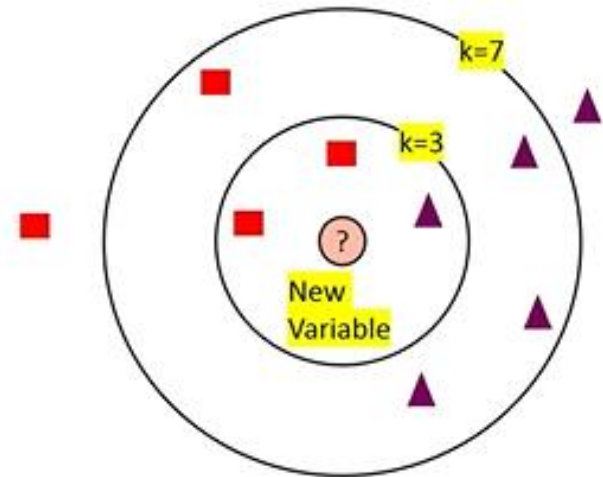


### K-NN

**DEFINICIÓN** : La clasificación se calcula a partir de la mayoría simple de los votos de los  $k$  vecinos más cercanos de cada punto, es decir, se basa en como están clasificados sus vecinos.

**VENTAJAS** : Este algoritmo es simple de implementar, robusto a los datos de entrenamiento ruidosos, y efectivo si los datos de entrenamiento son pequeños y sin muchas dimensiones.

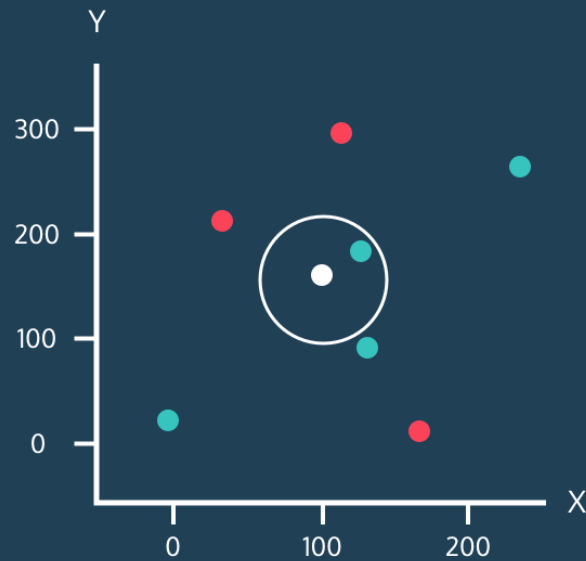
**DESVENTAJAS** : La necesidad de determinar el valor de  $K$  y el costo del cálculo es alto ya que necesita calcular la distancia euclídea de cada instancia a todas las muestras de entrenamiento.



$K \downarrow$ , sesgo tendrá mucho ruido.  
 $K \uparrow$ , tiempo computación muy alto.



## K-NN



● # Green = 1

K = 1

● # Red = 0

Prediction: Green



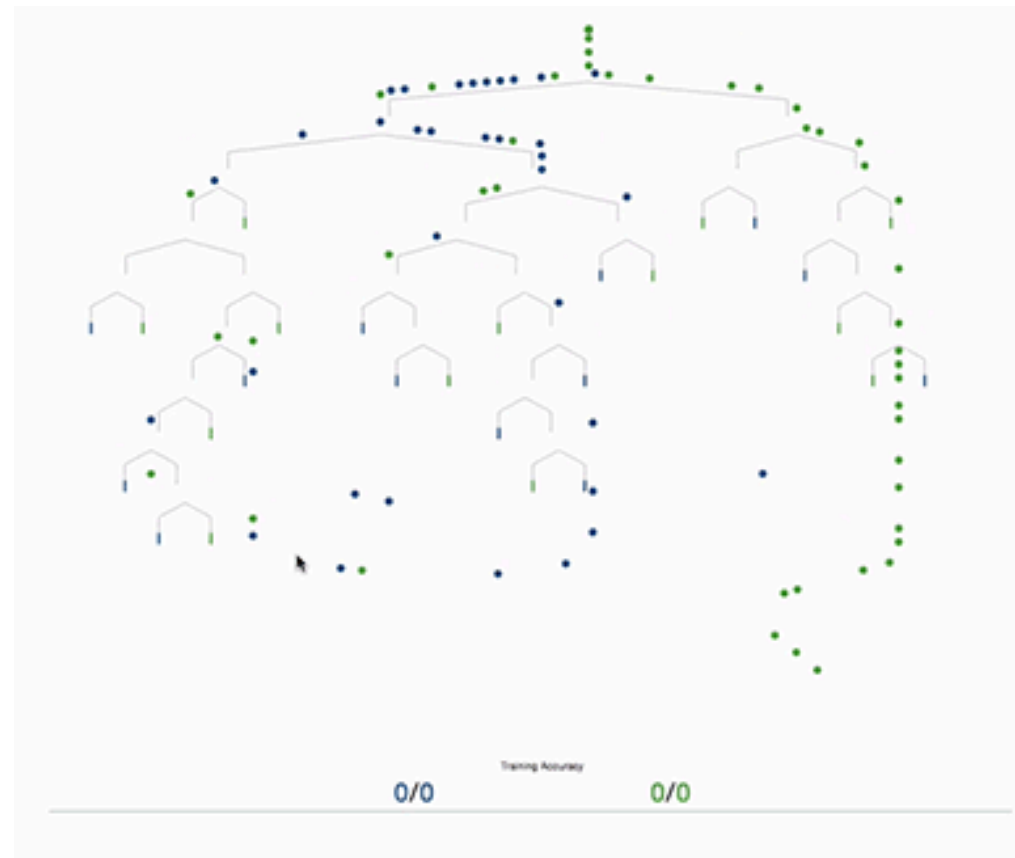
## Árboles de decisión

**DEFINICIÓN** : Dado un dato de atributos junto con sus clases, un árbol de decisión produce una secuencia de reglas que pueden utilizarse para clasificar los datos.

**VENTAJAS** : es sencillo de entender y visualizar, requiere poca preparación de datos (no es necesario normalizar) y es muy rápido de entrenar y evaluar.

### DESVENTAJAS

- Cuando hay muchas variables, riesgo de overfitting: control de la complejidad.
- Inestabilidad: un pequeño cambio en los datos puede modificar ampliamente la estructura del árbol.





## Ensemble: Random Forest

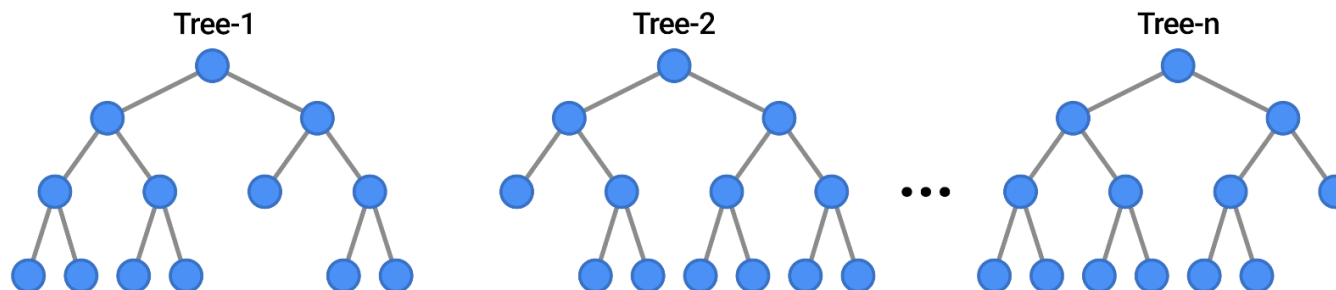
Es un **algoritmo tipo ensambladores**. Estos están formados por un **grupo de modelos predictivos** que permiten alcanzar una mejor precisión y estabilidad del modelo.

### VENTAJAS :

- Reducción del overfitting.
- Es más preciso que los árboles de decisión en la mayoría de los casos.
- Puede manejar miles de variables de entrada e identificar las más significativas (**feature selection**).

**DESVENTAJAS** : Predicción lenta en tiempo real.

## EXAMPLES





## Neural Network (MLP)

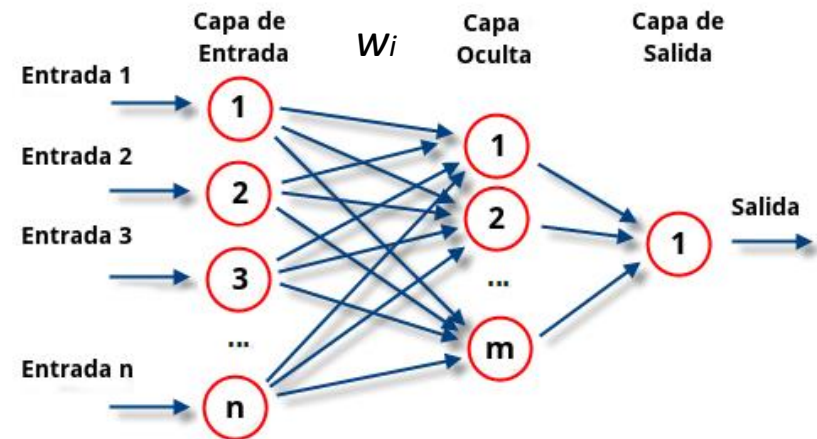
**DEFINICIÓN:** es una red neuronal artificial de tipo “feedforward” totalmente conectada.

### VENTAJAS:

- Puede aplicarse a problemas complejos no lineales.
- Funciona bien con grandes datos de entrada.
- Predicciones rápidas tras el entrenamiento.

### DESVENTAJAS:

Los cálculos son difíciles y requieren mucho tiempo.



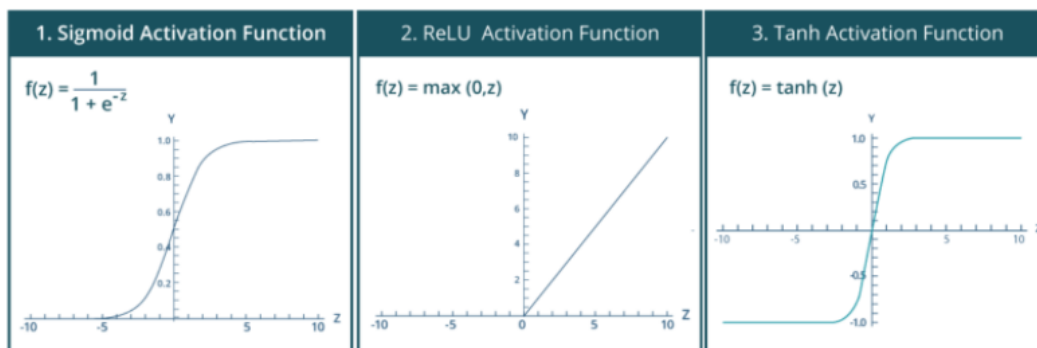


## Neural Network: Deep Learning

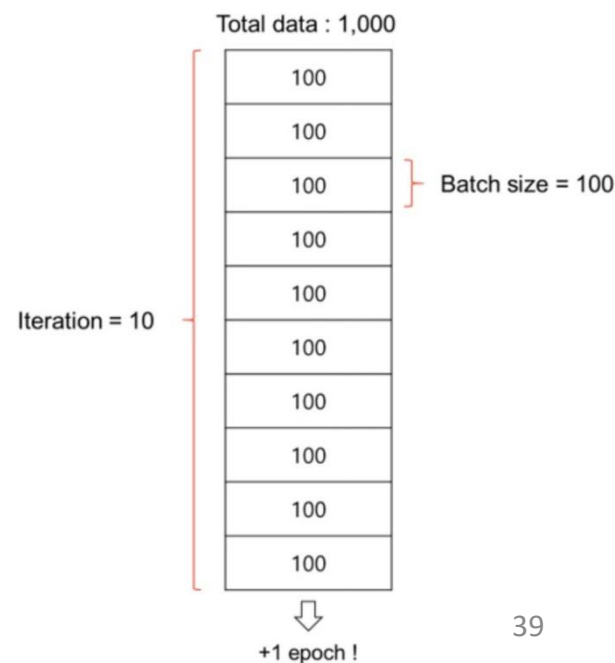
Veamos como funcionan las redes neuronales con ejemplos (deep learning)

<http://playground.tensorflow.org/>

### Funciones de activación



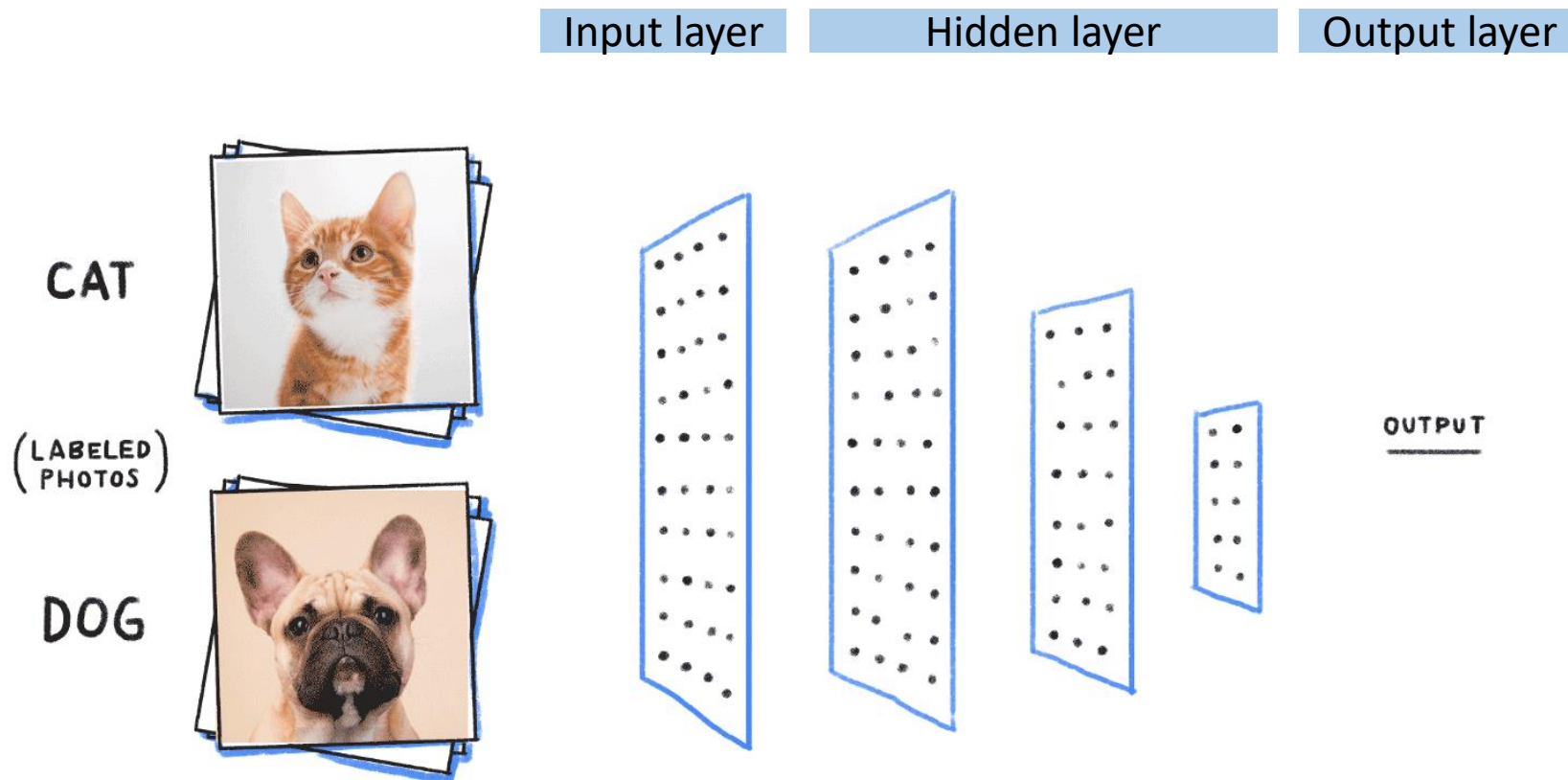
### Batch, epoch e iteraciones







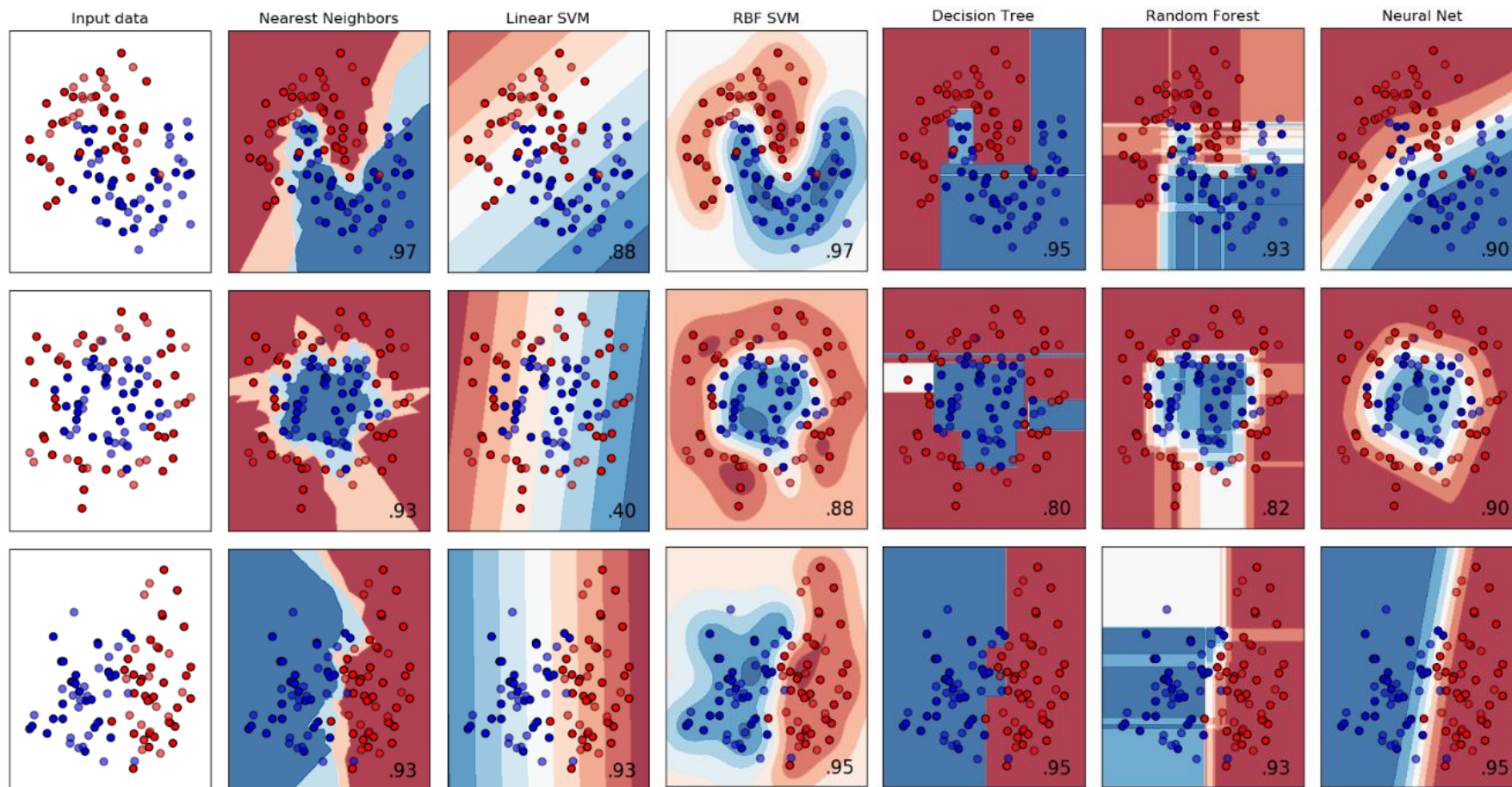
## Neural Network: Deep Learning







## Algoritmos Supervisados de Clasificación

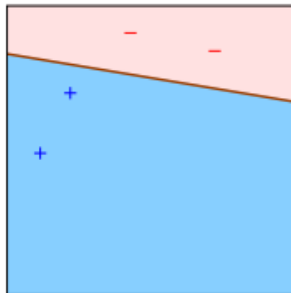




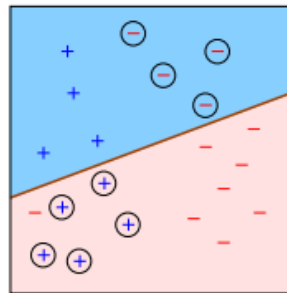
## Construir un clasificador preciso

Para un buen rendimiento del **test**, es necesario:

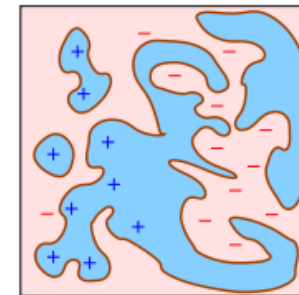
- Suficientes datos de **entreno** (así se evita el *underfitting*).
- Obtener un buen resultado de los datos de **entrenamiento**.
- El clasificador no debe ser demasiado "complejo" (así se evita el *overfitting*).



Insuficientes  
datos de entreno



Error de entreno  
demasiado alto



Clasificador  
demasiado complejo.



Conceptos generales.

Tipos de Clasificación.

Aplicaciones de Clasificación.

Detectar y resolver el desequilibrio de clases.

Principales modelos de Clasificación.

**Métricas de evaluación: Clasificación.**

Ejemplo práctico de un modelo de Clasificación.



## Evaluación de Algoritmos de Clasificación

Existen muchas formas de evaluar “cómo de bien” funciona el algoritmo.

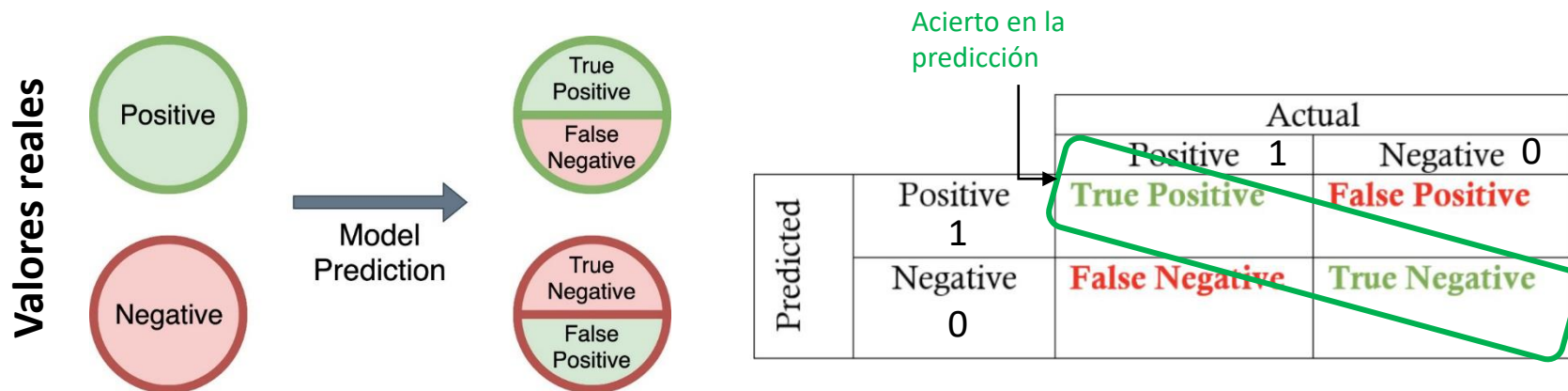
### **Métricas de Clasificación**

- Matriz de confusión
- Acierto
- Recall
- Precision
- F-1 Score
- MCC (Matthews correlation coefficient)
- AUC-ROC



## Matriz de confusión

Una matriz de confusión es una tabla que resume el rendimiento de un algoritmo de clasificación de tipo supervisado.



- **Verdaderos positivos (TP):** Cuando predcimos un positivo y el resultado verdadero es positivo.
- **Falsos positivos (FP):** Cuando predcimos un positivo y el resultado verdadero es negativo.
- **Falsos negativos (FN):** Cuando predcimos un negativo y el resultado verdadero es positivo.
- **Verdaderos Negativos (TN):** Cuando predcimos un negativo y el resultado real es negativo.

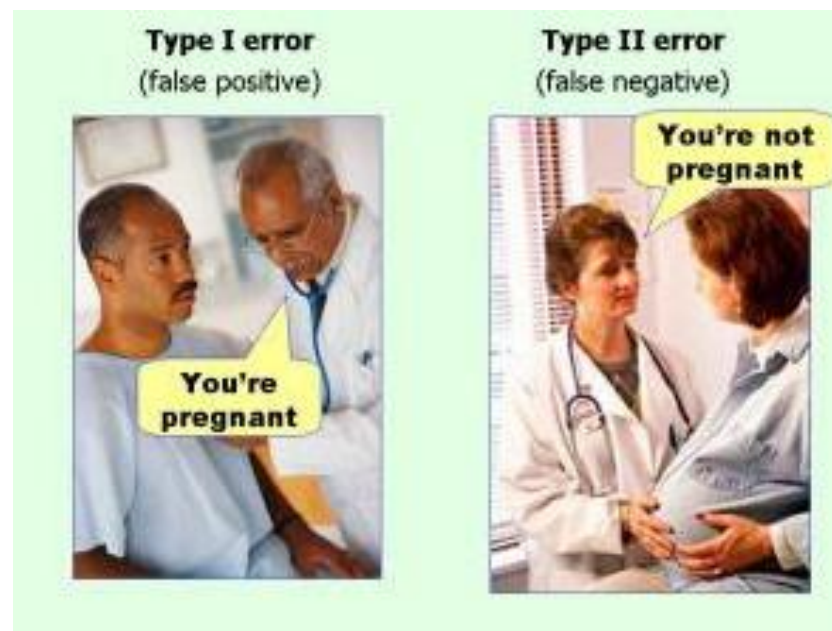




## Matriz de confusión

Una matriz de confusión es una tabla que resume el rendimiento de un algoritmo de clasificación de tipo supervisado.

		Actual	
		Positive 1	Negative 0
Predicted	Positive 1	True Positive	False Positive
	Negative 0	False Negative	True Negative



**Objetivo:** evitar falsos positivos y falsos negativos

- **Falsos positivos (FP):** Cuando predecimos un positivo y el resultado verdadero es negativo.
- **Falsos negativos (FN):** Cuando predecimos un negativo y el resultado verdadero es positivo.



## Accuracy (exactitud)

Fracción de predicciones que se realizaron correctamente en un modelo de clasificación.

$$Accuracy = \frac{Total\ de\ aciertos\ (TP + TN)}{Total\ de\ muestras\ (N)} = \frac{270 + 9}{314} = 0,88$$

Predicciones	Valor real		Total
	Actual - Cancer 1	Actual - NOT Cancer 0	
	Predicted - Cancer 1	TP=9 FP=5	14
	Predicted - NOT Cancer 0	FN = 30 TN = 270	300
Total	39	275	N=314

Matriz de confusion para el modelo de Hawkins

Pero... cuidado!

(Accuracy paradox)

Es necesario utilizar otras métricas de evaluación (recall, f1, etc).





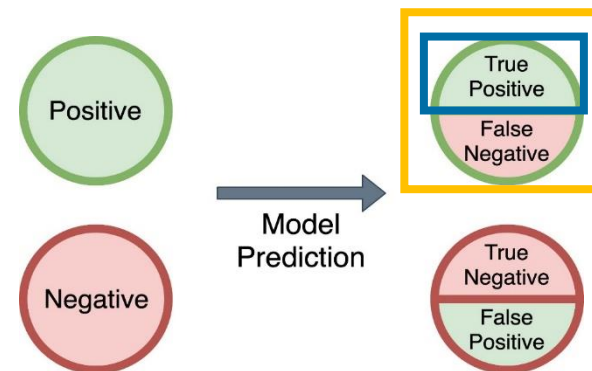
## Recall (Sensibilidad)

Número de predicciones de una clase que fueron correctamente identificadas entre el total de observaciones de esa clase.

Responde a la siguiente pregunta: **de todas las etiquetas positivas posibles, ¿cuántas identificó correctamente el modelo?**

Valor real

Predicciones	Valor real		Total
	Actual - Cancer <b>1</b>	Actual - NOT Cancer <b>0</b>	
	Predicted - Cancer <b>1</b>	TP=9 FP=5	14
	Predicted - NOT Cancer <b>0</b>	FN = 30 TN = 270	300
Total	39	275	N=314



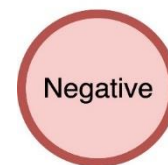
$$Recall = \frac{TP}{TP + FN} = \frac{9}{9 + 30} = 0,23$$



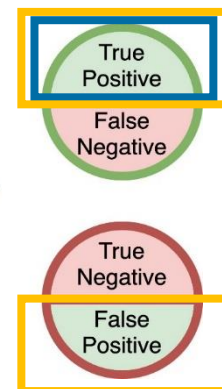
## Precisión

Identifica la frecuencia con la que un modelo predijo correctamente la clase positiva.  
 Responde a la siguiente pregunta: **de todas las predicciones que he etiquetado como cáncer, ¿cuántas eran realmente cáncer?**

		Valor real		
		Actual - Cancer <b>1</b>	Actual - NOT Cancer <b>0</b>	Total
Predicciones	Predicted - Cancer <b>1</b>	TP=9	FP=5	14
	Predicted - NOT Cancer <b>0</b>	FN = 30	TN = 270	300
	Total	39	275	N=314



Model  
Prediction



$$Precision\ Clase\ 1 = \frac{TP}{TP + FP} = \frac{9}{9 + 5} = 0,64$$



## F1-Score

- Combina las medidas de **Precision** y **Recall** en un solo valor.
- Se calcula haciendo la media armónica entre la Precisión y el Recall.
- Hace más fácil el comparar el rendimiento combinado de la Precisión y el Recall entre varias soluciones.

$$F1 - score = 2x \frac{Precision \cdot Recall}{Precision + Recall} = 2 \cdot \frac{0,64 \cdot 0,23}{0,64 + 0,23} = 0,33$$



## MCC (Matthews correlation coefficient)

Solo para **clasificación binaria**

MCC = 1 → Predicción perfecta  
MCC = 0 → Predicción aleatoria  
MCC = -1 → Predicción totalmente errónea

Valor real

Predicciones

	Actual - Cancer <b>1</b>	Actual - NOT Cancer <b>0</b>	Total
Predicted - Cancer <b>1</b>	TP=9	FP=5	14
Predicted - NOT Cancer <b>0</b>	FN = 30	TN = 270	300
Total	39	275	N=314

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} = 0,34$$



**My Boss when finds out my  
Model is 90% accurate!**

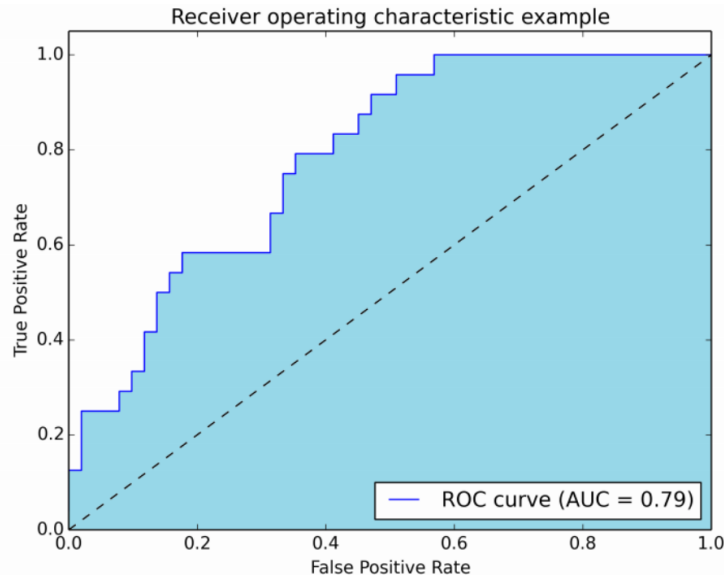


**Me: Trained my Model  
on an imbalanced Dataset  
with 90% of one class!**





## Curva ROC / AUC



La **curva ROC** (receiver operating curve) es un gráfico que muestra el rendimiento de un modelo de clasificación en todos los umbrales de clasificación. Esta curva representa dos parámetros:

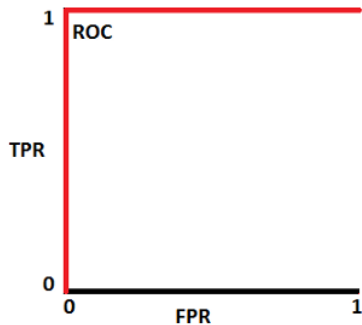
- Tasa de verdaderos positivos (TPR).
- Tasa de falsos positivos (FPR)

**AUC** significa "área bajo la curva ROC". AUC mide toda el área bidimensional por debajo de la curva ROC completa (cálculo integral).

- El AUC oscila entre 0 y 1.
- Un modelo cuyas predicciones son un 100% incorrectas tiene un AUC de 0; otro cuyas predicciones son un 100% correctas tiene un AUC de 1.0

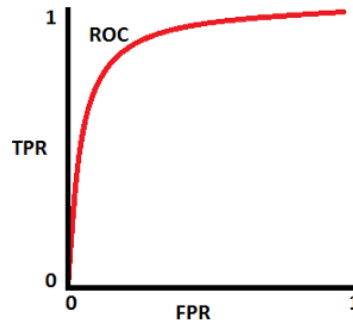


## Curva ROC / AUC



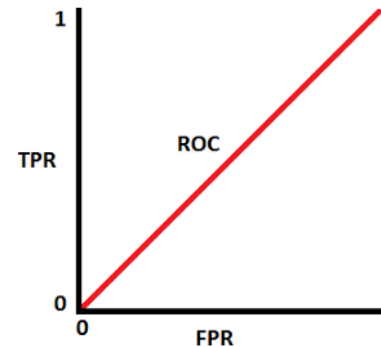
### Predicción perfecta.

- $AUC = 1$ .
- $FPR=0$ ,  $TPR=1$ .



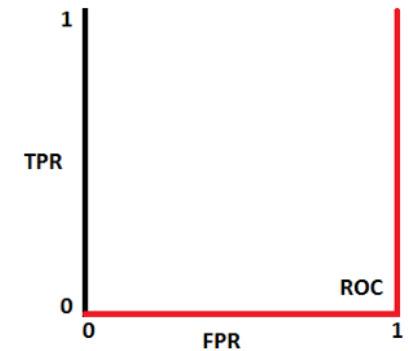
### Predicción buena/mediocre.

- $AUC = (1-0,5)$
- Si AUC es 0.85, hay un 85% de probabilidades de que el modelo distinga correctamente las clases.



### Predicción aleatoria.

- $AUC = 0,5$ .
- $FPR=TPR$ .
- Es tan exacto como voltear una moneda.
- ¡Este es el peor de los casos!



### Predicción inversa.

- $AUC = 0$ .
- $FPR=1$ ,  $TPR=0$ .
- El algoritmo predice perfectamente el inverso de las clases
- Situación altamente improbable.





Conceptos generales.

Tipos de Clasificación.

Aplicaciones de Clasificación.

Detectar y resolver el desequilibrio de clases.

Principales modelos de Clasificación.

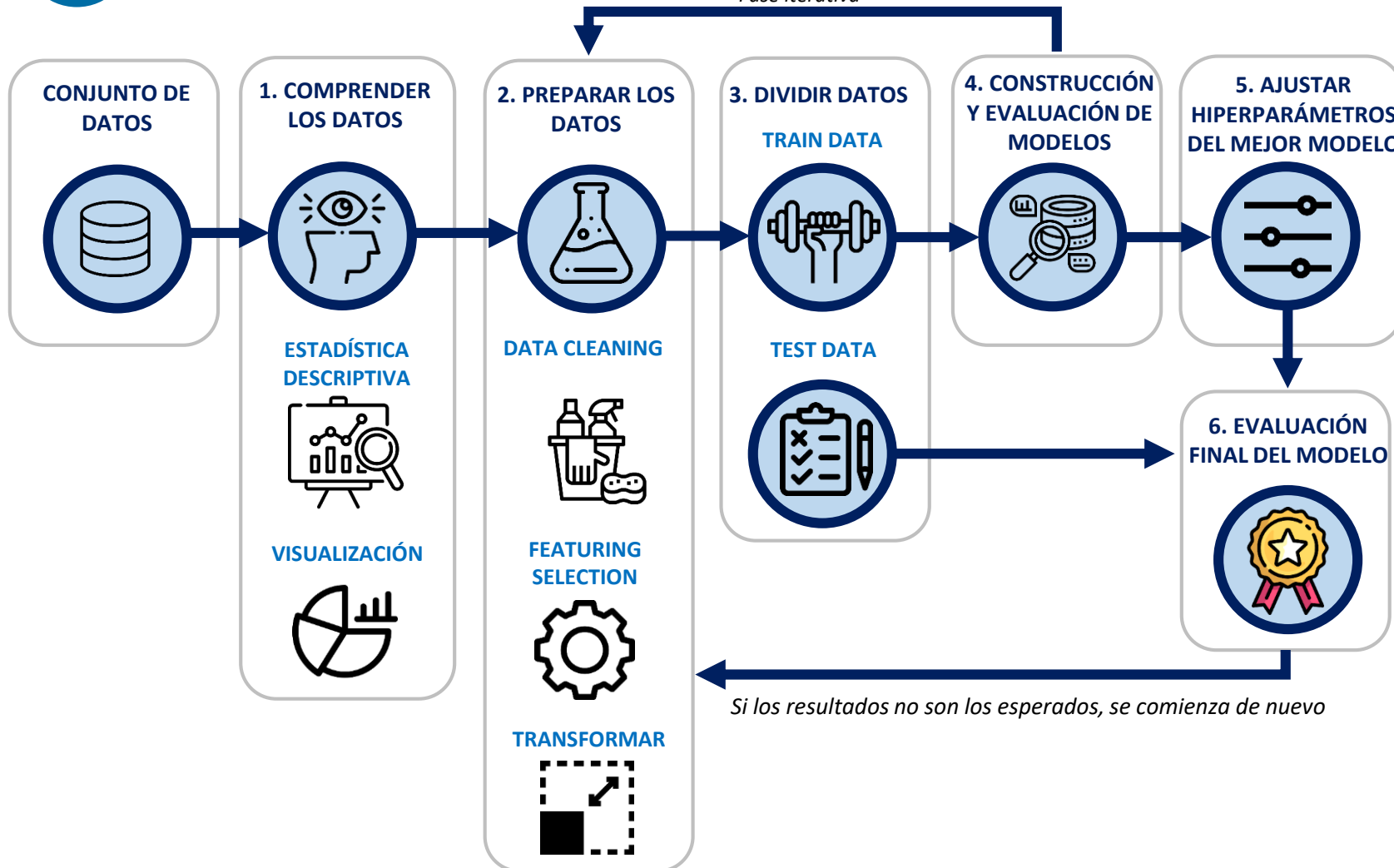
Métricas de evaluación: Clasificación.

**Ejemplo práctico de un modelo de Clasificación.**



## Recordatorio antes de la práctica

*Fase iterativa*





## Machine Learning y Deep Learning con Python

- Librerías para machine learning



- Fácil de utilizar (intuitivo)
- Contiene una amplia variedad de modelos de clasificación, regression, clustering y dimensional reduction.

- Librerías para deep learning



Keras





## Documentación de Scikit-learn



<https://scikit-learn.org/stable/>

**Guía de usuario** [http://scikit-learn.org/stable/user\\_guide.html](http://scikit-learn.org/stable/user_guide.html)

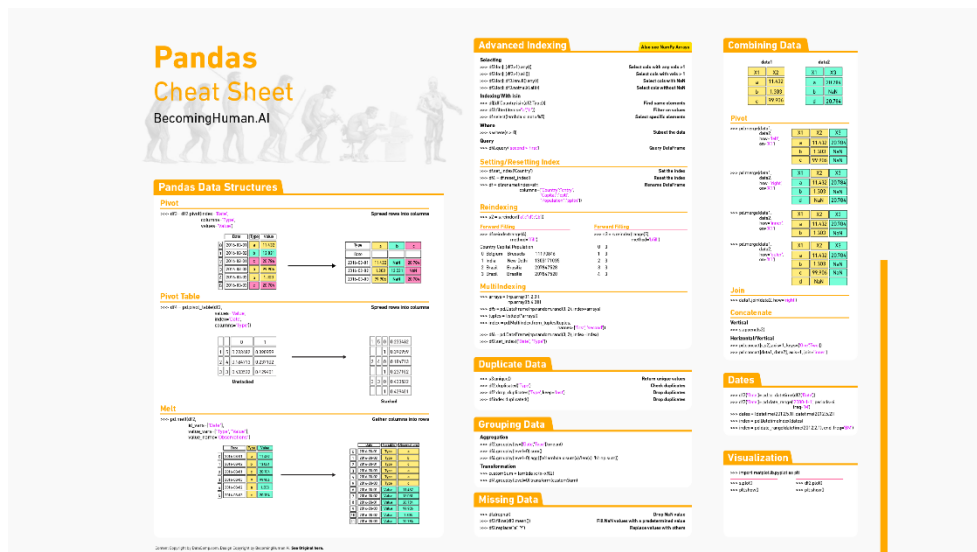
**Tutorial básico** <https://scikit-learn.org/stable/tutorial/basic/tutorial.html>

**Ejemplos** [http://scikit-learn.org/stable/auto\\_examples/index.html](http://scikit-learn.org/stable/auto_examples/index.html)



## Más información

- [Cheat-sheets de Scikit-learn Machine Learning, Pandas, Matplotlib...](#)
- [Aprendizaje automático \(Coursera\)](#)



The collage features several cheat sheets for data science tools:

- Pandas Cheat Sheet:** Titled "Becoming Human AI", it covers Pandas Data Structures, Pivot, Pivot Table, and Merge.
- Advanced Indexing:** Includes sections on Selecting, Reindexing, and Multi-Indexing.
- Combining Data:** Covers Concatenation, Joins, and Stacking.
- Duplicate Data:** Focuses on Identifying and Removing duplicates.
- Grouping Data:** Details Grouping, Aggregating, and Transforming data.
- Missing Data:** Provides methods for Identifying, Dropping, and Filling missing values.
- Dates:** Lists various date-related operations and functions.
- Visualization:** Offers quick references for plotting with Matplotlib and Seaborn.



## Ejercicio de Clasificación





## Bibliografía

- **“Advanced Data Analytics for Power Systems”** - Ali Tajer, Samir M. Perlaza, H. Vincent Poor.
- **“Hands-on Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems”** Aurelien Geron
- **“Python Data Science Handbook: Tools and Techniques for Developers: Essential Tools for working with Data”**. Jake VanderPlas