



# Máster Smart Energy

Postgrado de DIGITAL ENERGY

Curso 2022/2023

## Sesión 1 – Introducción a Machine Learning

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

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



# Calendario

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

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

- Introducción a qué es la Inteligencia Artificial.
- Entender los conceptos básicos de Machine Learning.
- Explorar los diferentes tipos de aprendizaje en Machine Learning.
- Ejemplos de aplicaciones de big data y machine learning en el sector energético.



Pregunta!



[www.menti.com](http://www.menti.com)

(Introducir el código que se dirá el día de clase)



## Conceptos generales.

Conceptos básicos de Machine Learning.  
Crear un modelo de Aprendizaje Supervisado  
Aplicaciones de Big Data y Machine Learning  
en el sector energético



## Conceptos generales

Big Data

Inteligencia Artificial

Machine Learning

Deep Learning



## Big Data

- Datos más **variados**, que llegan en **volúmenes** cada vez mayores y con más **velocidad** (3 V's)
- Estos conjuntos de datos son tan voluminosos que el software de procesamiento de datos tradicional no puede gestionarlos.
- Todo dato que ocupe más de 65536 x 256 filas y columnas. ¿Por qué?
- Estos volúmenes masivos de datos pueden utilizarse para resolver problemas de negocio que antes no se podían abordar.



## Big Data





## Big Data

 **Oscar Wijsman**  
@oscarwijsman

#Bigdata analytics in 1939. Staff sorting 4M used London Underground tickets to analyse line use via @kerfors



9:47 PM · Aug 8, 2014 from Overbetuwe, Nederland · Twitter Web Client

72 Posted by u/700Tnechin 5 years ago

Computing Division at the Department of the Treasury, mid 1920s [936x585]  
[i.imgur.com/qZNHrH...](https://i.imgur.com/qZNHrH...)



7 Comments Share Save Hide Report 95% Upvoted



## Big Data

Algunos datos sobre datos...

Sensor data from a cross-country flight



Fuente: HP



270 millones de usuarios generan 100 GB por día



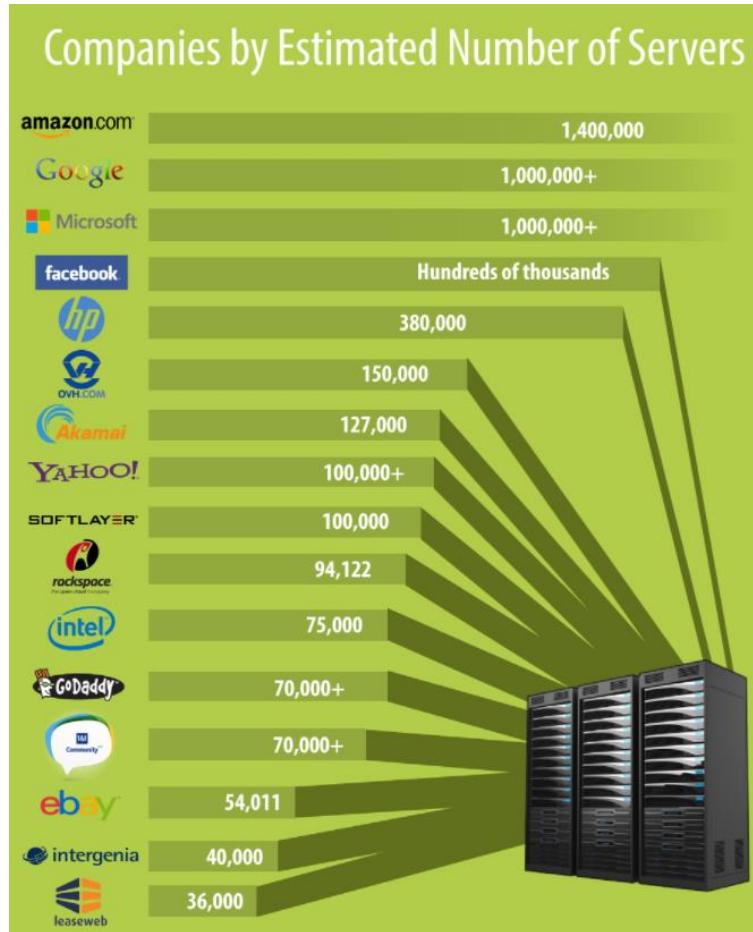
Un solo Phasor Measurement Unit (PMU) puede generar unos 300 GB de datos en un año.



Una turbina genera 588 GB por día



## Big Data



## Pero... ¿Dónde están estos datos?

- Cada día se crean 2,5 Zettabytes ( $10^{21}$  bytes).
- Ese ritmo no hace más que acelerarse.
- Solo en los **dos últimos años** se ha generado el **90%** de los datos del mundo.
- Estos datos alimentan los modelos de Machine Learning.

Fuente: Forbes - [How Much Data Do We Create Every Day?](#)



**Vídeo** muy interesante acerca de qué hacen con nuestros datos  
*“¿Por qué me vigilan, si no soy nadie?”*

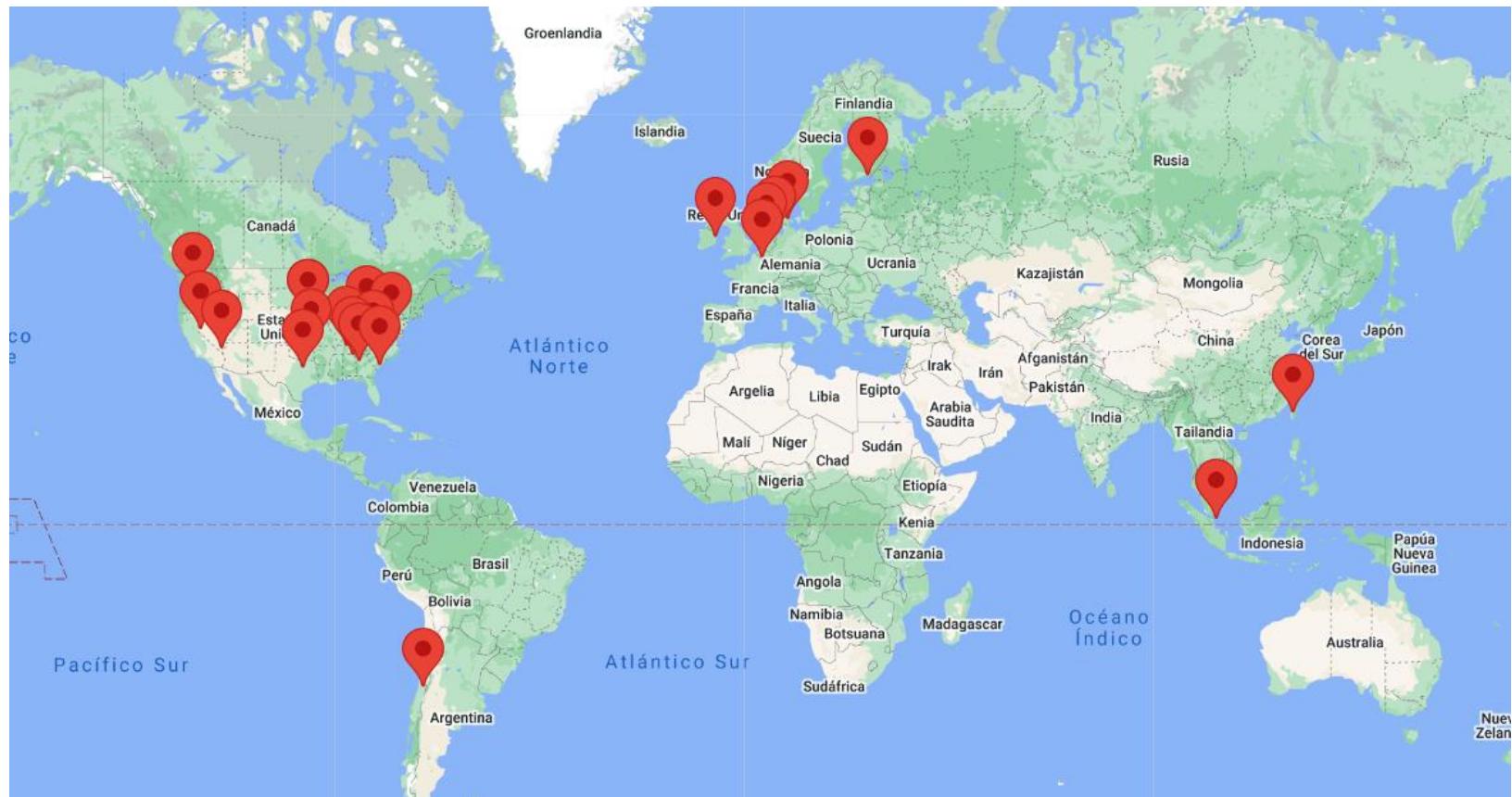


Big Data

Pero...¿Dónde están estos datos?

## Ubicación de los centros de datos de Google

<https://www.google.com/about/datacenters/locations/>





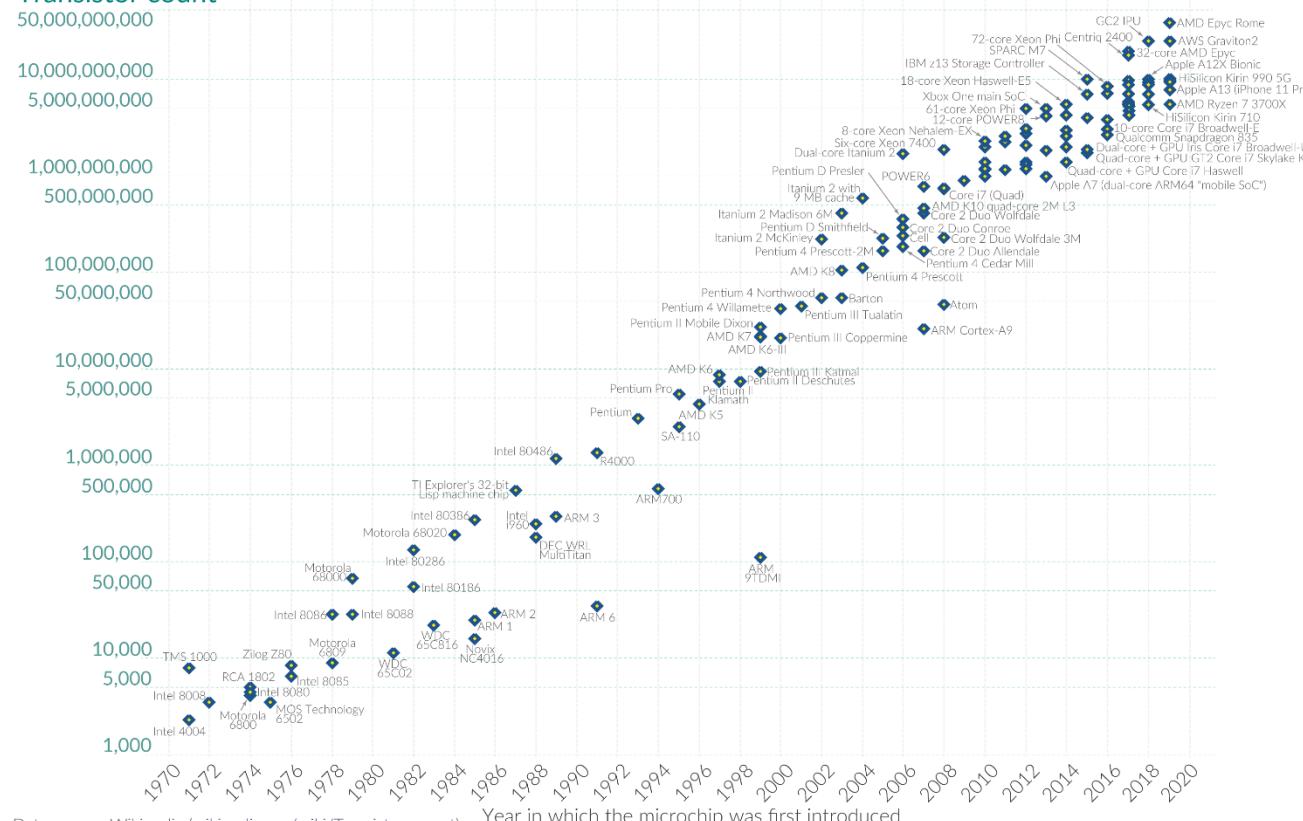
## Inteligencia Artificial El origen

Moore's Law: The number of transistors on microchips doubles every two years

Our World  
in Data

Moore's law describes the empirical regularity that the number of transistors on integrated circuits doubles approximately every two years. This advancement is important for other aspects of technological progress in computing – such as processing speed or the price of computers.

### Transistor count





## Inteligencia Artificial El futuro?





## Inteligencia Artificial      El futuro?

### Espías de EEUU avisan del potencial "devastador" de la computación cuántica para la seguridad digital

La carrera para lograr un ordenador cuántico viable preocupa a la NSA por su capacidad para romper sistemas cifrados, aunque Snowden reveló que también investiga formas de utilizar esta tecnología de forma ofensiva

Fuente: ABC

### *Computación cuántica: un salto tan grande como el que hubo entre el abaco y la informática actual*

Los próximos ordenadores solucionarán en segundos problemas que las máquinas convencionales más potentes tardarían miles de años en resolver

Fuente: [El País](#)



## Inteligencia Artificial

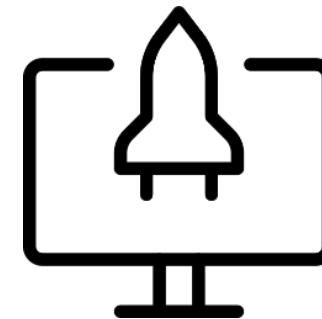
### Factores clave para su desarrollo

GRAN CANTIDAD DE  
DATOS GENERADOS  
(BIG DATA)



+

MEJORA  
COMPUTACIONAL





## Inteligencia Artificial

## Machine Learning

## Deep Learning

**IA:** Ciencia (parte de la informática) que desarrolla algoritmos inspirados en el comportamiento humano para tomar decisiones inteligentes

### Artificial Intelligence



Any technique that enables computers to mimic human intelligence. It includes *machine learning*

### Machine Learning



A subset of AI that includes techniques that enable machines to improve at tasks with experience. It includes *deep learning*

### Deep Learning



A subset of machine learning based on neural networks that permit a machine to train itself to perform a task.

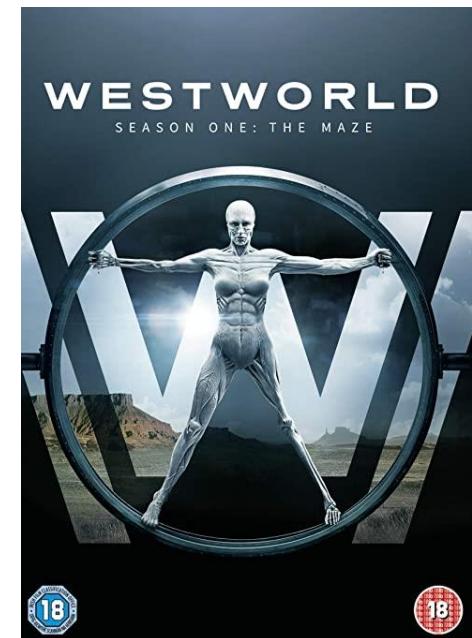
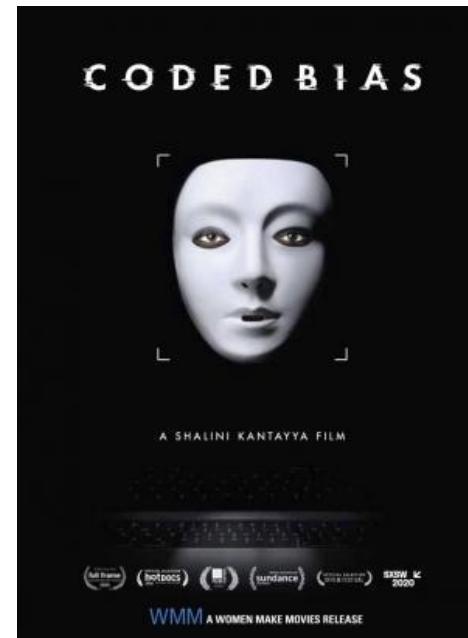


Digital Energy

Machine Learning



## AI SERIES!



TRAILER



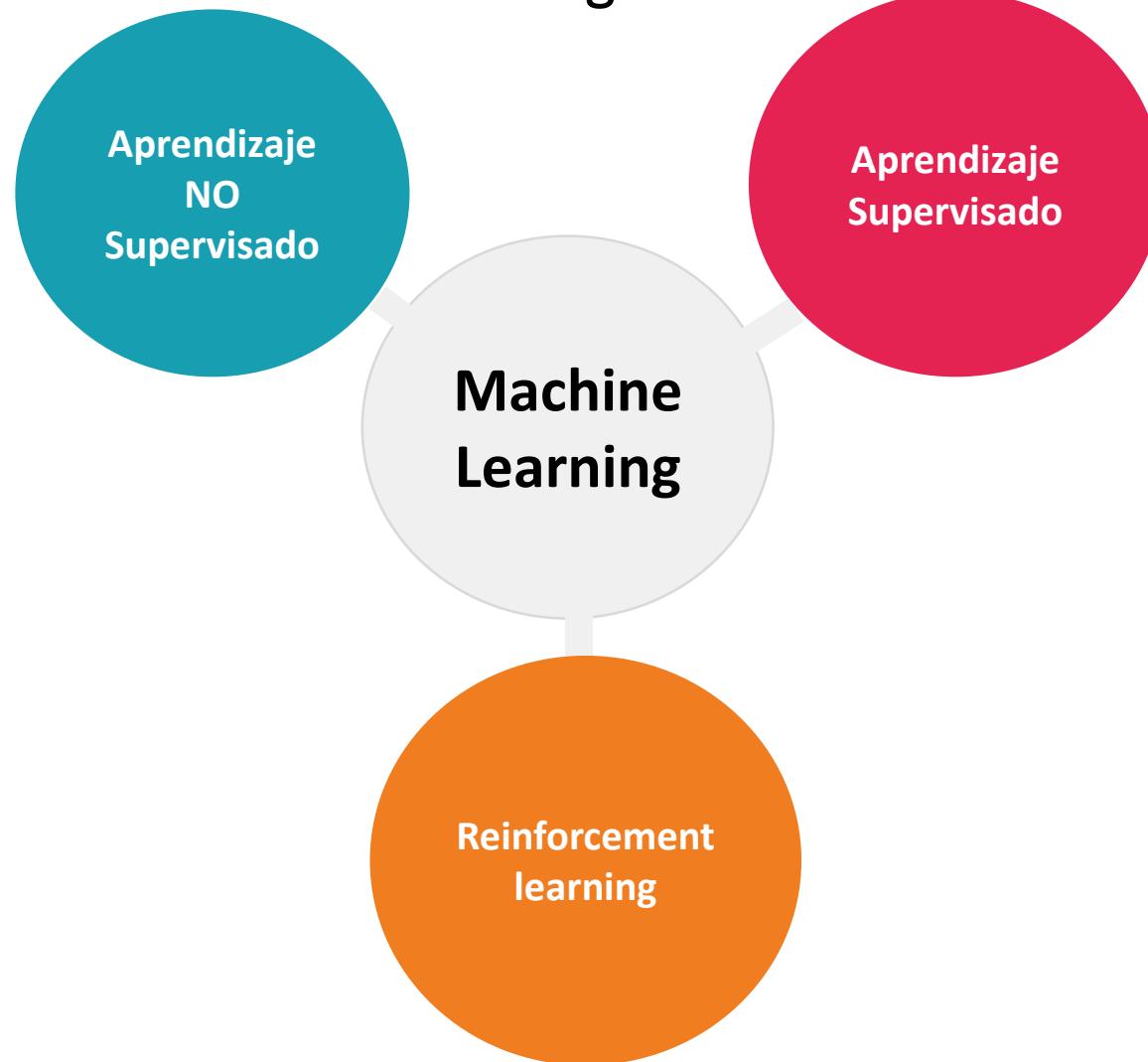
Conceptos generales.

## **Conceptos básicos de Machine Learning.**

Crear un modelo de Aprendizaje Supervisado  
Aplicaciones de Big Data y Machine Learning  
en el sector energético

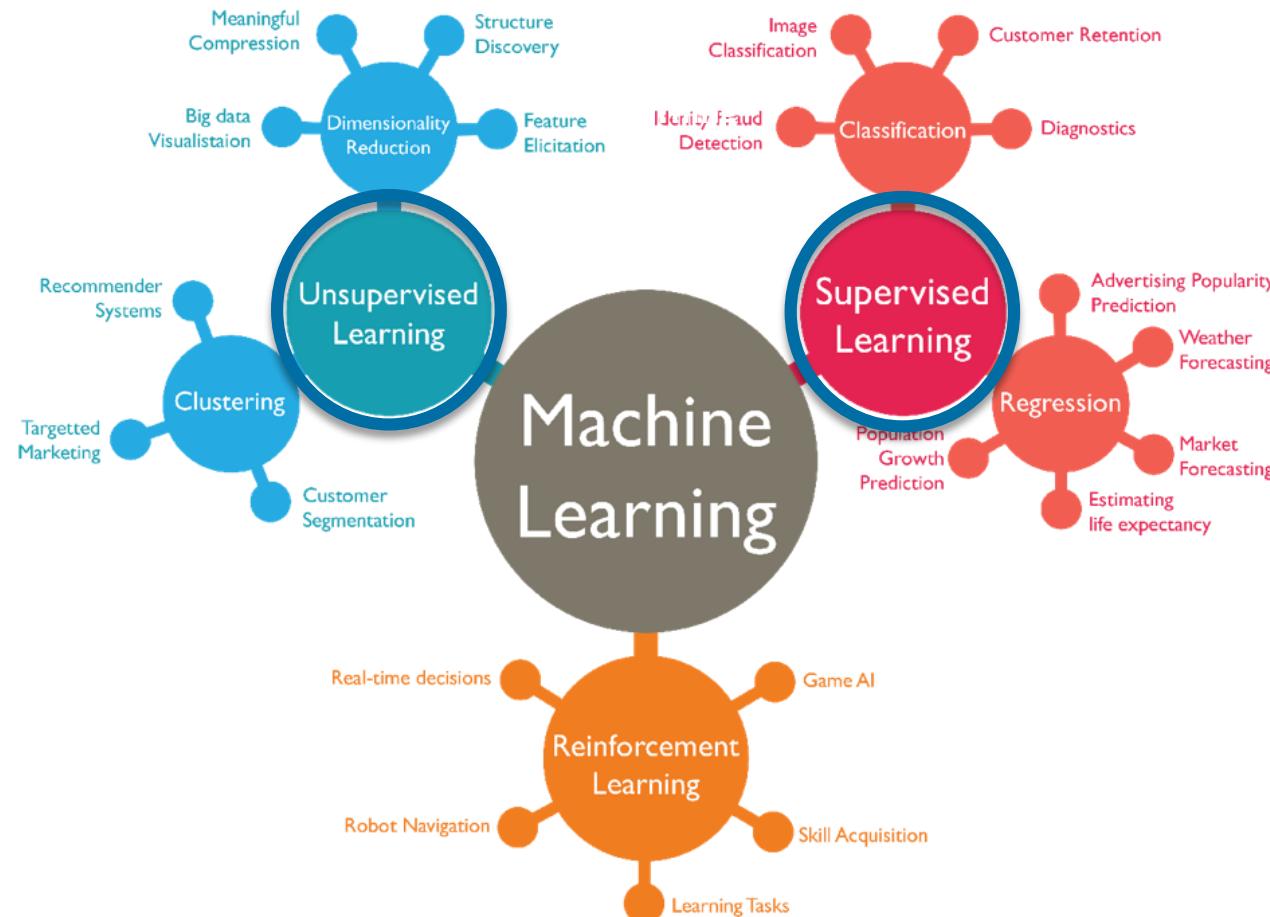


## Tipos de Machine Learning





## Tipos de Machine Learning



Source: Oracle AI and Data Science Blog.

<https://blogs.oracle.com/datascience/types-of-machine-learning-and-top-10-algorithms-everyone-should-know-v2>



## Aprendizaje Supervisado vs NO Supervisado

### Supervised Learning

X <sub>1</sub>	X <sub>2</sub>	X <sub>3</sub>	X <sub>p</sub>	Y

Target

### Un-Supervised Learning

X <sub>1</sub>	X <sub>2</sub>	X <sub>3</sub>	X <sub>p</sub>	Y

**NO target**

**Input variables:** features, predictors, or independent variables

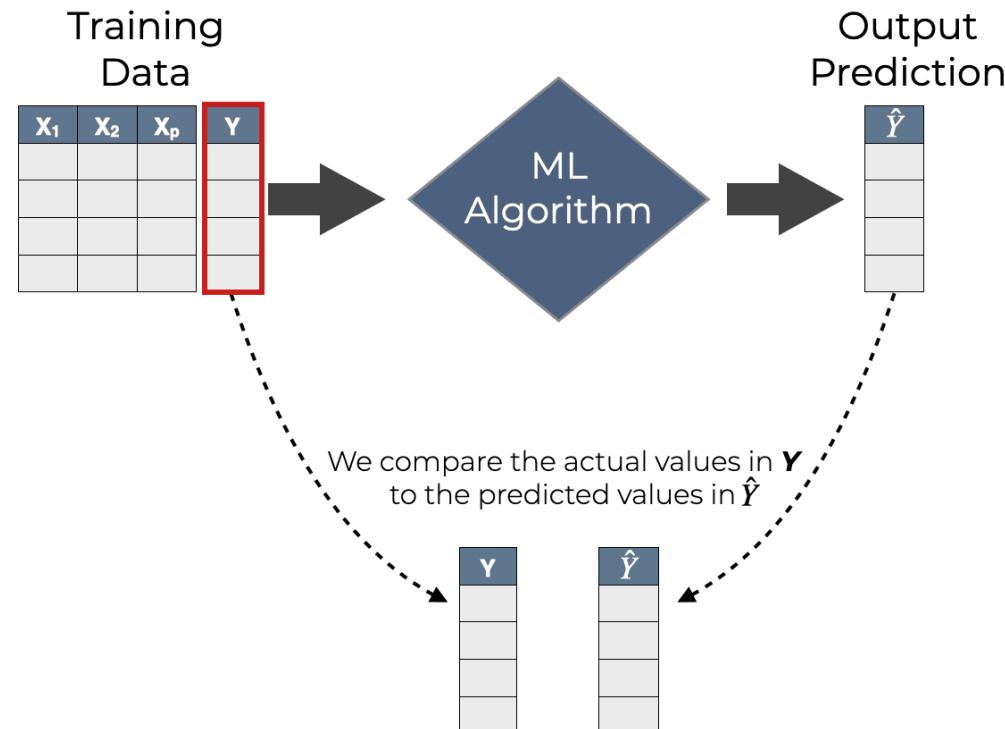
**Supervisado:** los datos tienen una variable de salida que estamos tratando de predecir.

**No Supervisado:** la variable objetivo está ausente. Sigue habiendo variables de "entrada", pero no hay objetivo.



## Aprendizaje Supervisado vs NO Supervisado

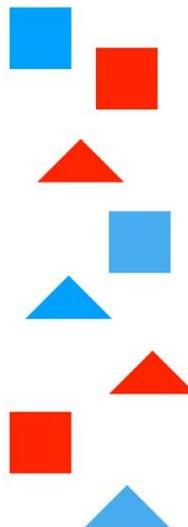
**IN SUPERVISED LEARNING, THE TARGET VARIABLE  $Y$  SUPERVISES THE MODELING PROCESS**





## Aprendizaje Supervisado vs NO Supervisado

Conjunto de datos



👁 Aprendizaje Supervisado

Conjunto de datos

Etiquetas

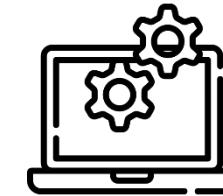


Clasificación de formas

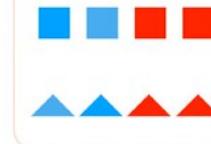


👁 Aprendizaje NO Supervisado

Conjunto de datos



Clustering





## Aprendizaje Supervisado vs NO Supervisado

### Supervised Learning (Classification Algorithm)



Cat



Cat



Not Cat



Not Cat



Predictive Model

→ Cat

### Unsupervised Learning (Clustering Algorithm)

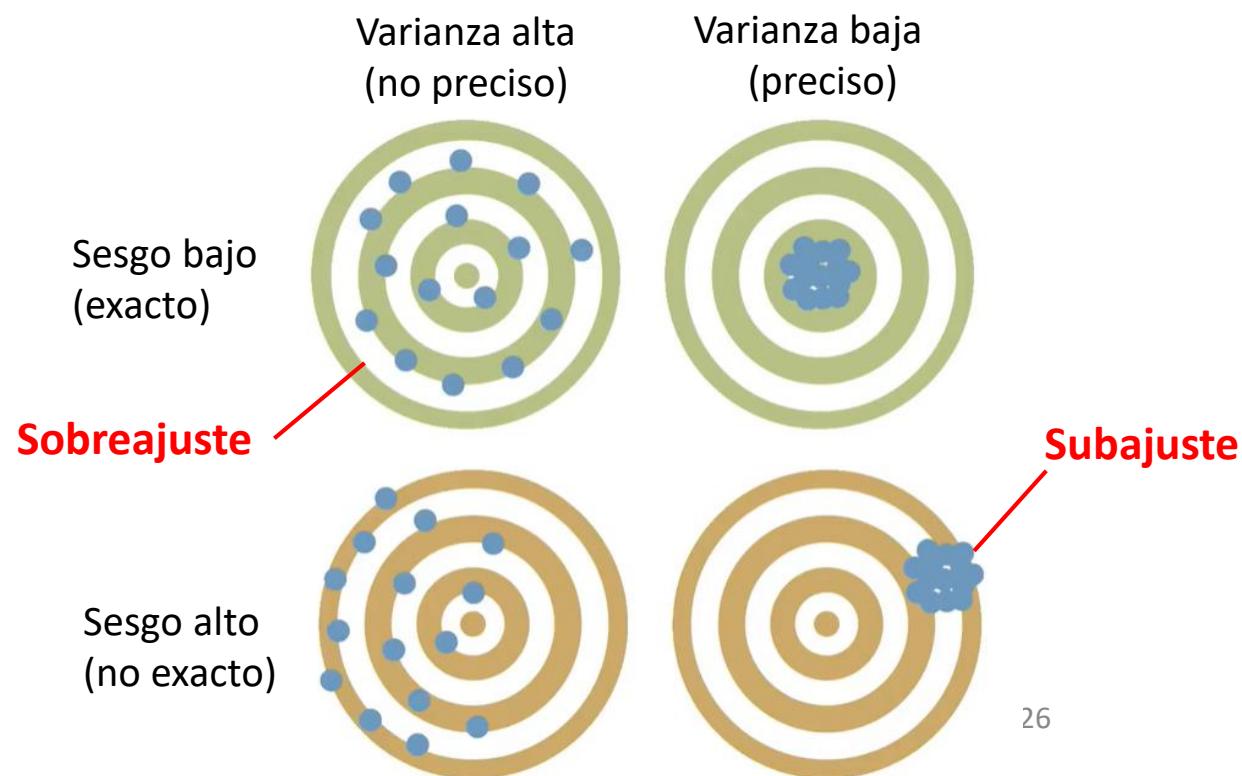




## Sesgo y varianza (Bias and Variance)

- **Sesgo (bias):** mide lo lejos que se encuentra el valor estimado respecto al valor real.
- **Varianza (variance):** diferencia entre los valores estimados.

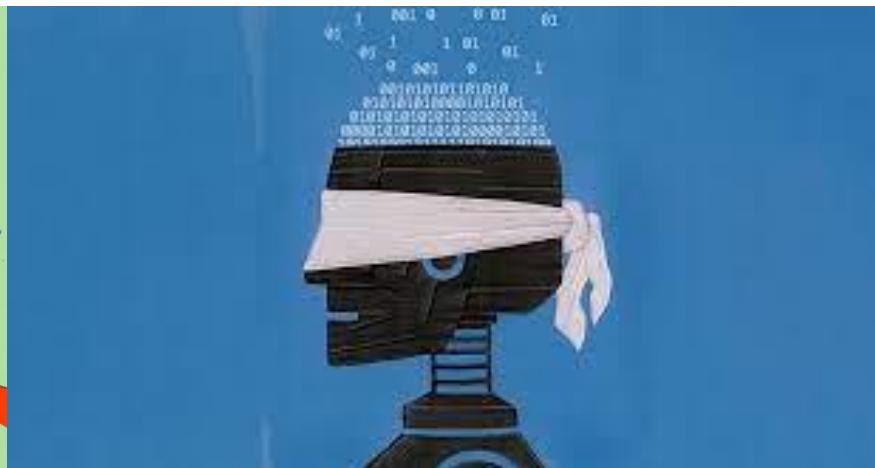
*Supongamos que el valor verdadero se sitúa en el centro de la diana y los puntos azules son las predicciones.*





## AI Bias (AI sesgo)

¿Qué os inspiran estas imágenes?





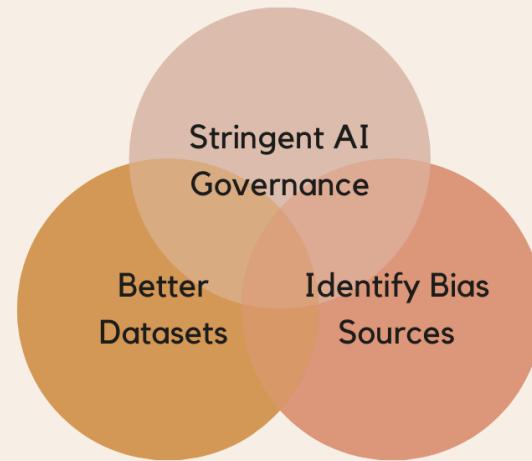
## AI Bias (AI sesgo)

- Se trata de un fenómeno que surge cuando un algoritmo ofrece resultados sistemáticamente sesgados como consecuencia de supuestos erróneos del proceso de aprendizaje automático.
- Normalmente afecta a cierto grupos o minorías.
- Al puede reforzar y aumentar los sesgos humanos ya existentes.
- Puede que la **IA ética** no esté en el centro de los debates, pero debería estarlo.



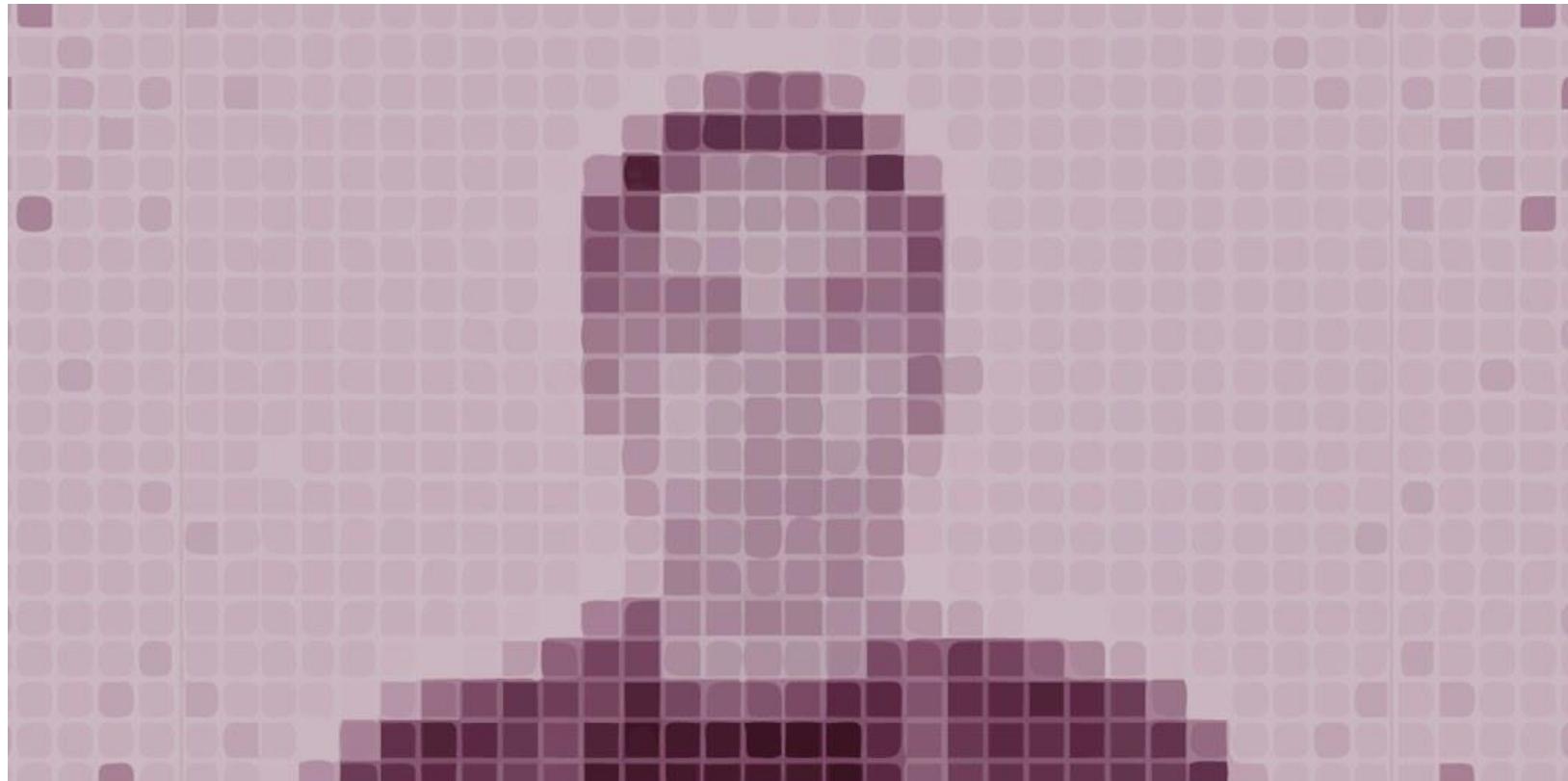
## AI Bias. ¿Cómo mitigarla?

### WAYS TO ELIMINATE OR MITIGATE BIAS IN AI





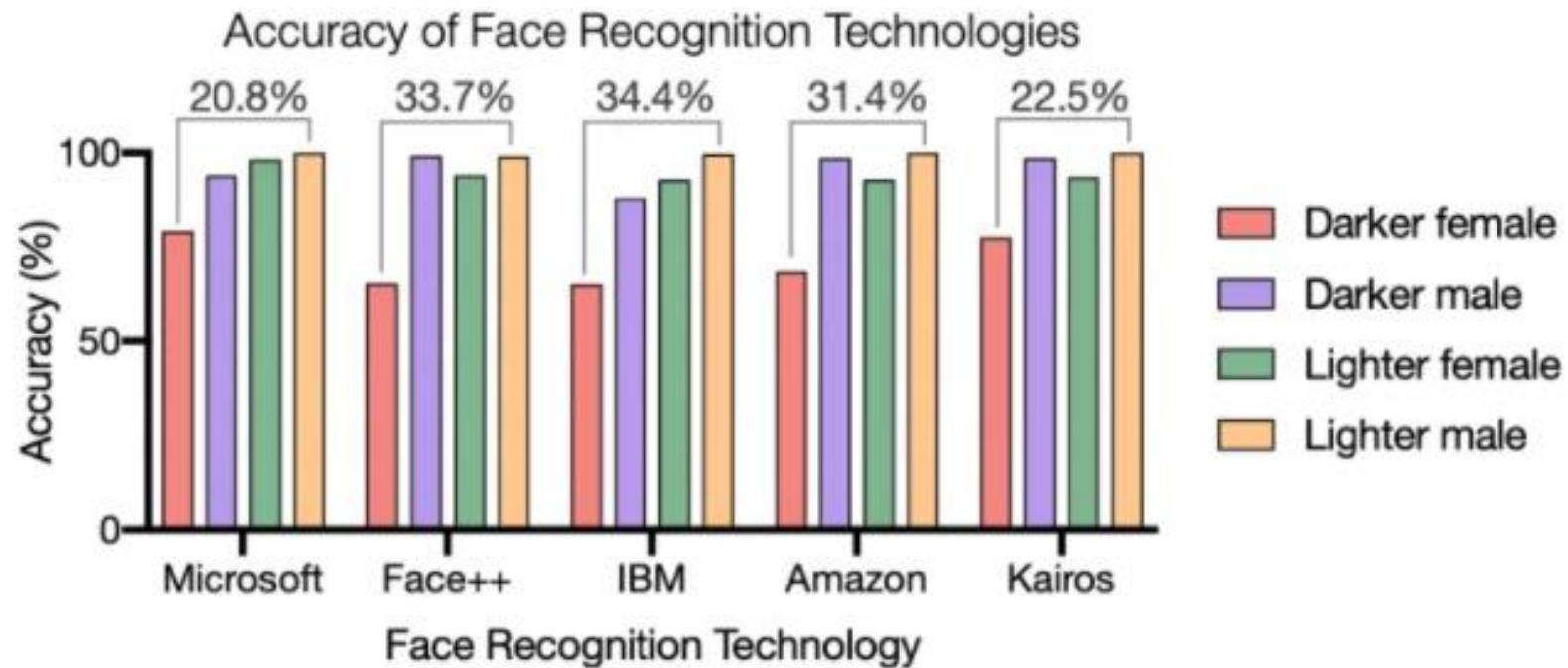
## Ejemplo de AI sesgada: The Gender Project



Fuente: Hardvard University <https://sitn.hms.harvard.edu/flash/2020/racial-discrimination-in-face-recognition-technology/>



## Ejemplo de sesgo: The Gender Project





## Ejemplo de sesgo: The Gender Project

RETAIL OCTOBER 11, 2018 / 1:04 AM / UPDATED 5 YEARS AGO

# Amazon scraps secret AI recruiting tool that showed bias against women

Reuters. Article [link](#)

# Machine Bias

There's software used across the country to predict future criminals. And it's biased against blacks.

by Julia Angwin, Jeff Larson, Surya Mattu and Lauren Kirchner, ProPublica

May 23, 2016

Propublica . <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>



## Subajuste y sobreajuste

**Sobreajuste (overfitting):** el modelo coincide demasiado con los datos de entrenamiento que no puede realizar predicciones correctas con datos nuevos.

**Subajuste (underfitting):** el modelo es demasiado simple y no se ajusta a los datos de entrenamiento.

El objetivo es obtener un modelo intermedio que sea capaz de obtener buenos resultados con datos nuevos (**generalización**).

### SUBAJUSTE

Sesgo alto

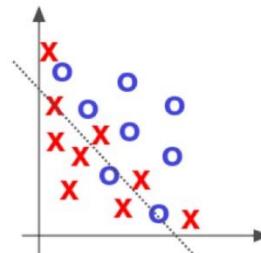
Varianza baja

### SOBREAJUSTE

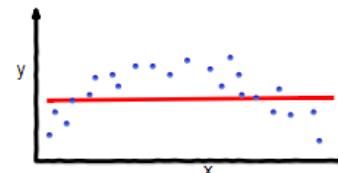
Sesgo bajo

Varianza alta

### Clasificación



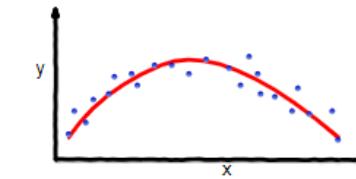
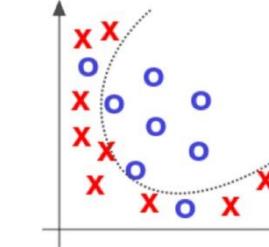
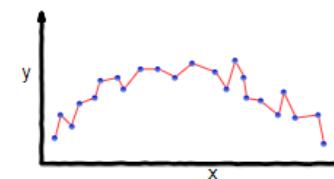
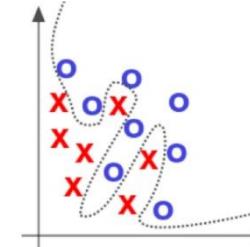
### Regresión



### BUEN AJUSTE

Sesgo bajo

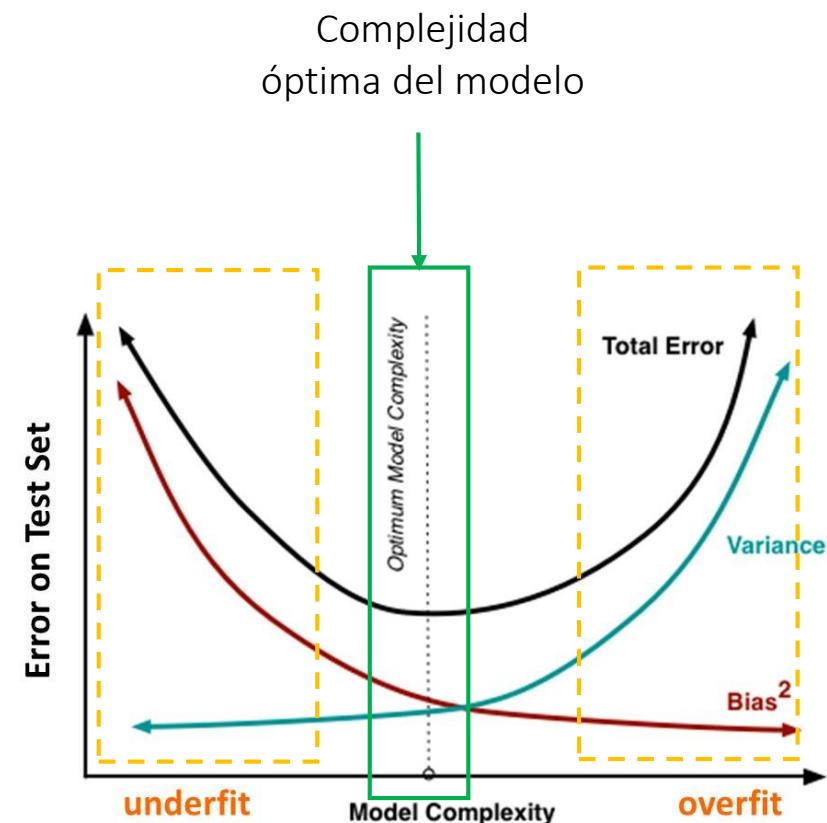
Varianza baja





## Errores de predicción

- El **error de predicción** para cualquier algoritmo de aprendizaje automático se puede dividir en tres partes:
  - Error de sesgo
  - Error de varianza
  - Error irreducible
- La complejidad óptima del modelo viene dada por un  $\downarrow$  **sesgo** y  $\downarrow$  **varianza**.
- A medida que aumenta la complejidad del modelo, aumentan las posibilidades de sobreajuste, es decir, la varianza aumenta.

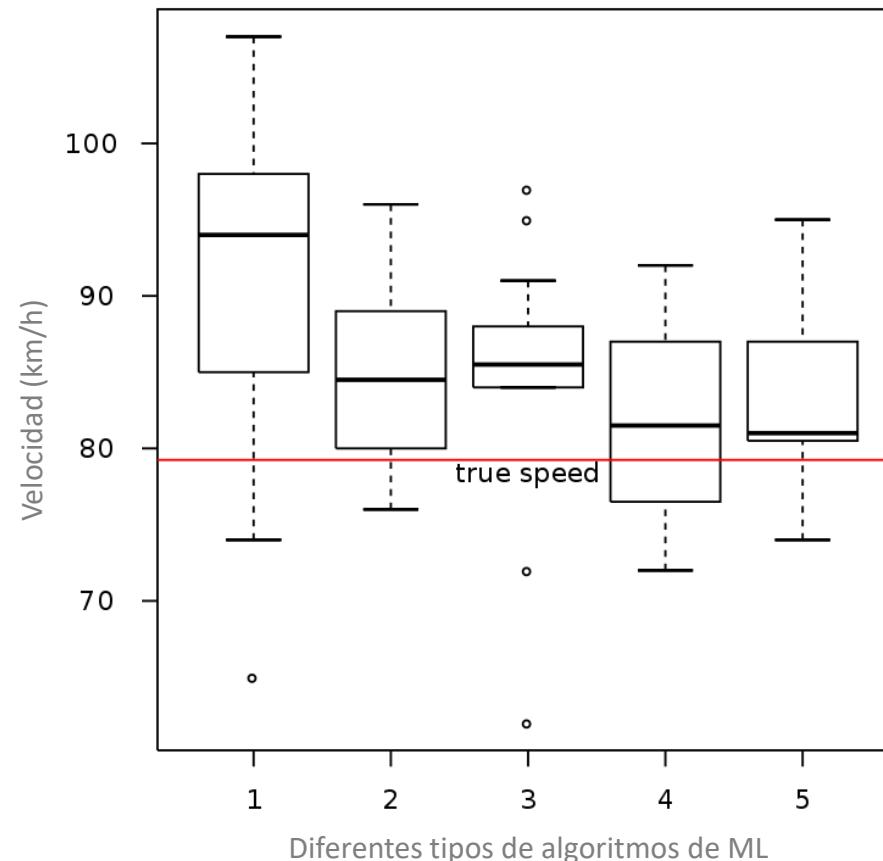




## Errores de predicción

Ejemplo: resultados de diferentes modelos para predecir la velocidad

¿Cuál tiene un menor error de sesgo? ¿Y cuál más varianza? ¿Con cual os quedaríais?





Conceptos generales.

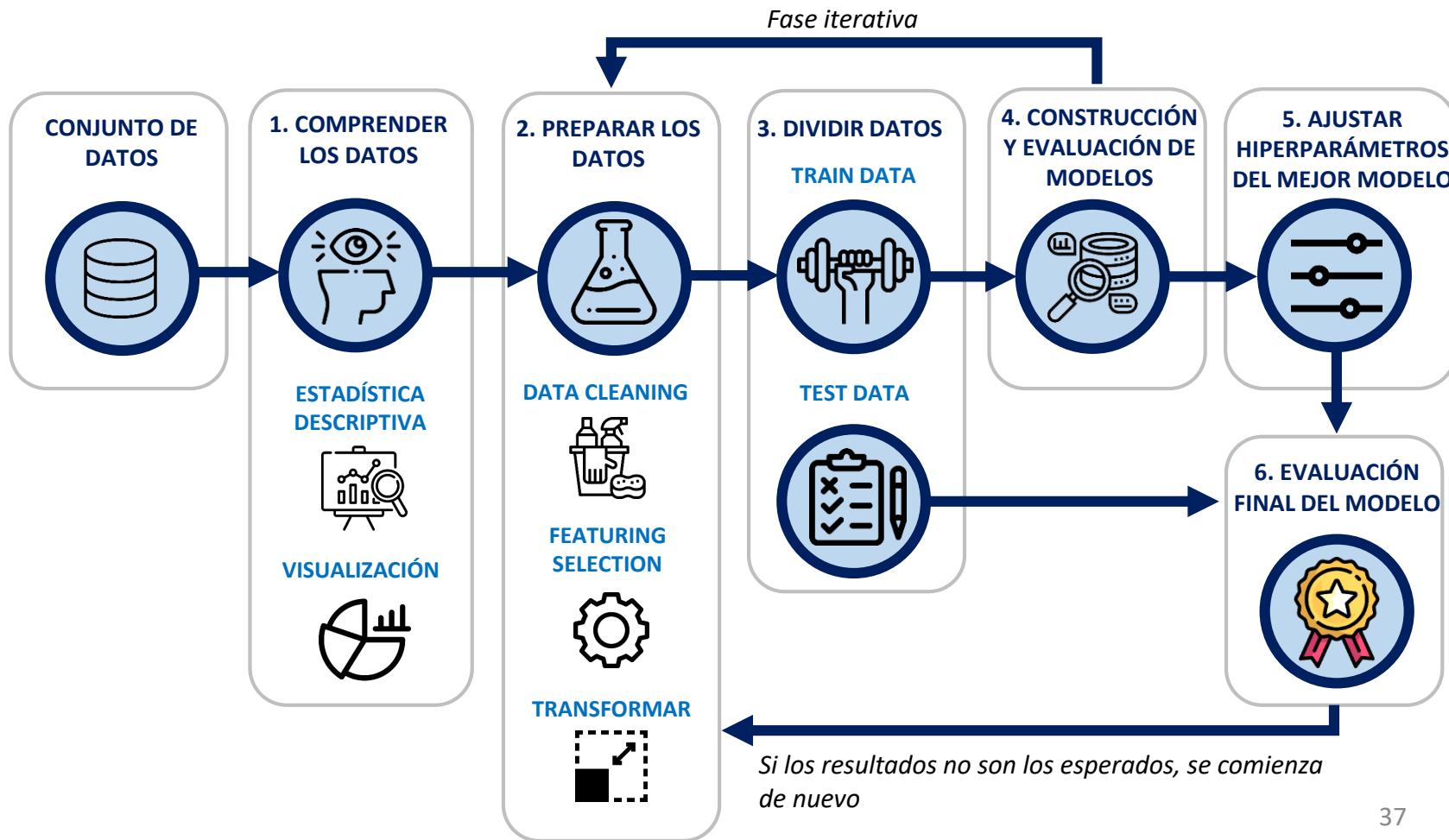
Conceptos básicos de Machine Learning.

## **Crear un modelo de Aprendizaje Supervisado**

Aplicaciones de Big Data y Machine Learning en el sector energético



## Crear un modelo de Machine Learning

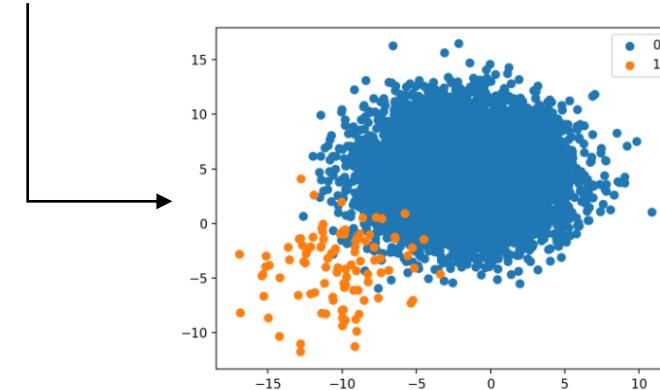




## 1. Comprender los datos

Primero debemos entender como son los datos, y para ello utilizamos herramientas estadísticas descriptivas y visuales.

- Observación de los datos.
- Dimensión (nº filas x nº columnas).
- Tipo de datos (string, float, etc.)
- Resumen de estadística descriptiva: media, mediana, desviación estándar, valor mín./máx...
- **Distribución de la etiqueta/clase** (solo clasificación): detectar si la etiqueta está balanceado o no.
- Correlación entre atributos.



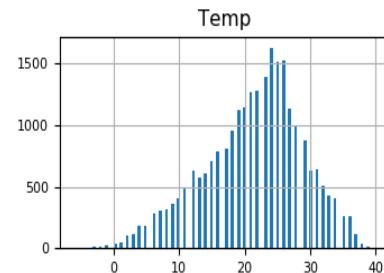
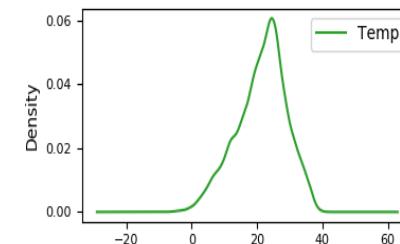
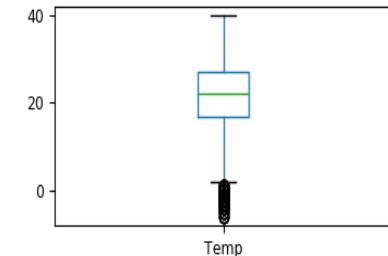


## 1. Comprender los datos

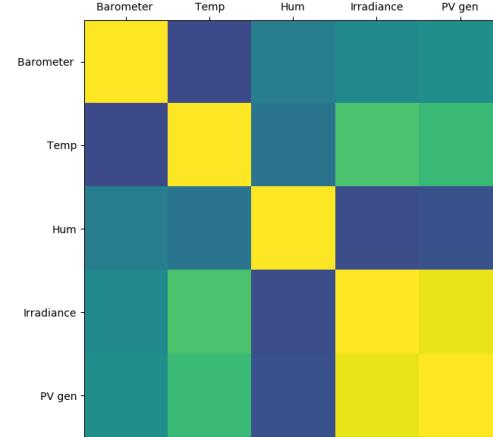
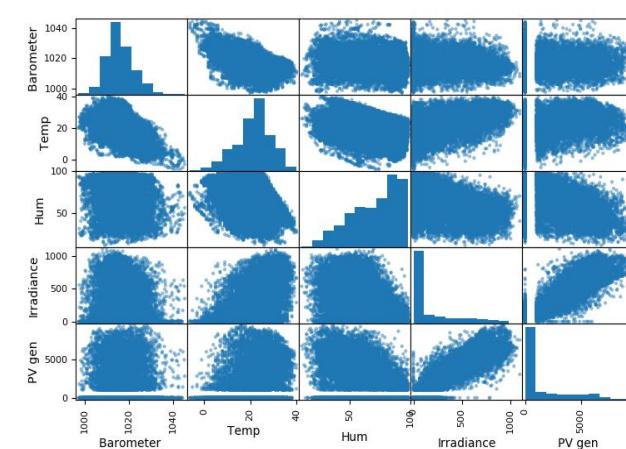
Primero debemos entender como son los datos, y para ello utilizamos herramientas estadísticas descriptivas y visuales.



### GRÁFICOS UNIVARIABLE

**Histogramas****Gráficos de densidad****Boxplots**

### GRÁFICOS MULTIVARIABLE

**Matriz de correlación****Matriz de dispersión**



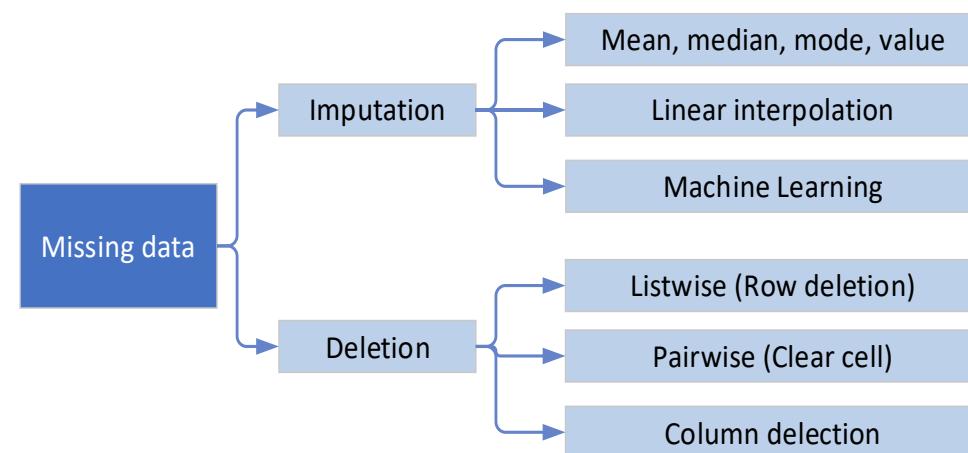
## 2. Preparar los datos

**DATA CLEANING:** proceso de detección y corrección (o eliminación) de registros corruptos o inexactos de un conjunto de registros, tabla o base de datos



### MISSING DATA

Típicamente representado con los indicadores "NaN" o "Null". La mayoría de los algoritmos no pueden manejar esos valores perdidos, así que tenemos que ocuparnos de ellos a través de diferentes técnicas antes de alimentar con datos a nuestros modelos.





## 2. Preparar los datos

**DATA CLEANING:** proceso de detección y corrección (o eliminación) de registros corruptos o inexactos de un conjunto de registros, tabla o base de datos

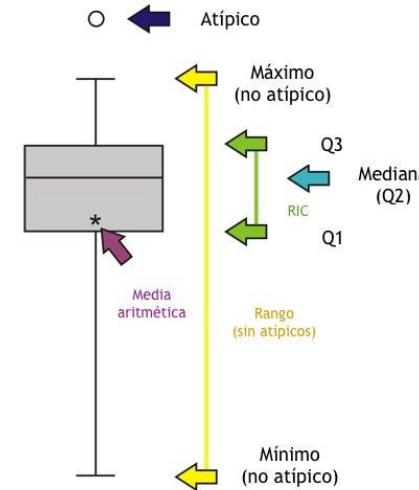
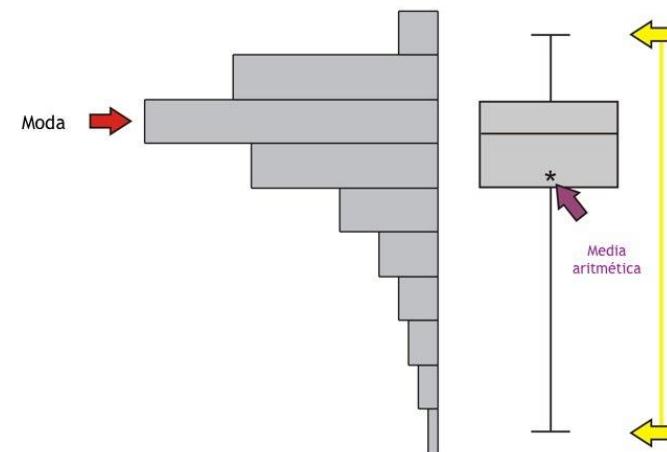
**OUTLIERS (VALORES ATÍPICOS):** observación que es numéricamente distante del resto de los datos. **OJO!** A veces aportan información importante y no deben ser eliminados.

Para visualizar los valores atípicos, se utiliza el diagrama Box Plot:



### Métodos:

- Z-score
- IQR

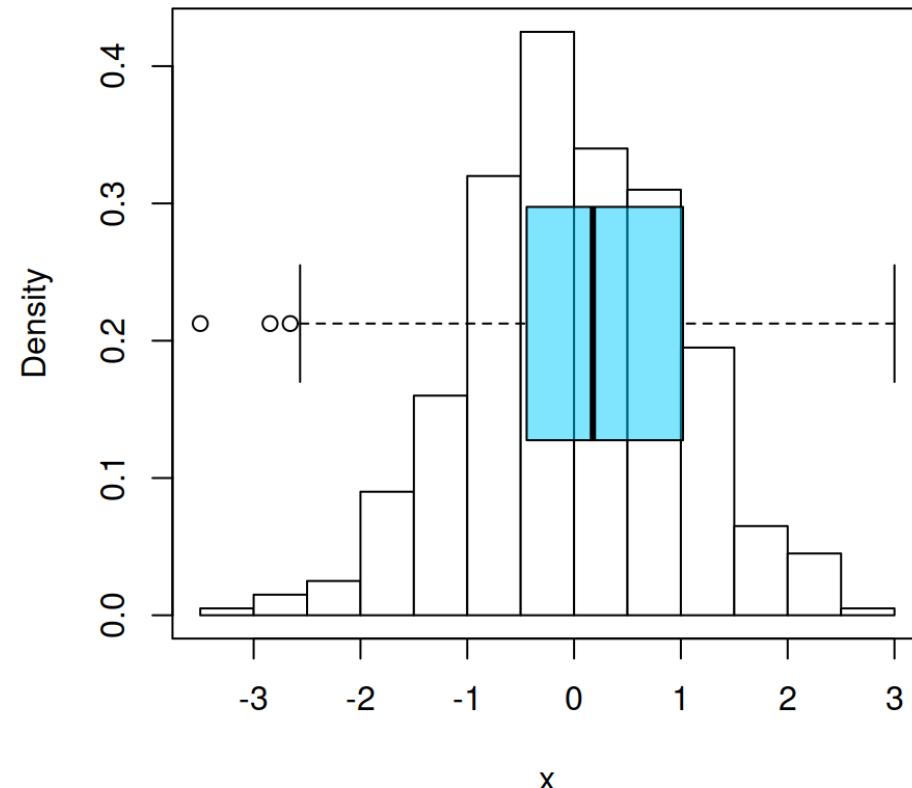


Explicación gráfica del diagrama Box Plot



## 2. Preparar los datos

Histograma y boxplot



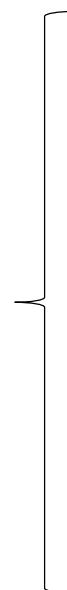


## 2. Preparar los datos

**FEATURING SELECTION:** Proceso de selección de un subconjunto de características relevantes para su uso en la construcción de modelos. Las características irrelevantes o parcialmente relevantes pueden afectar negativamente al rendimiento del modelo.



FEATURING  
SELECTION

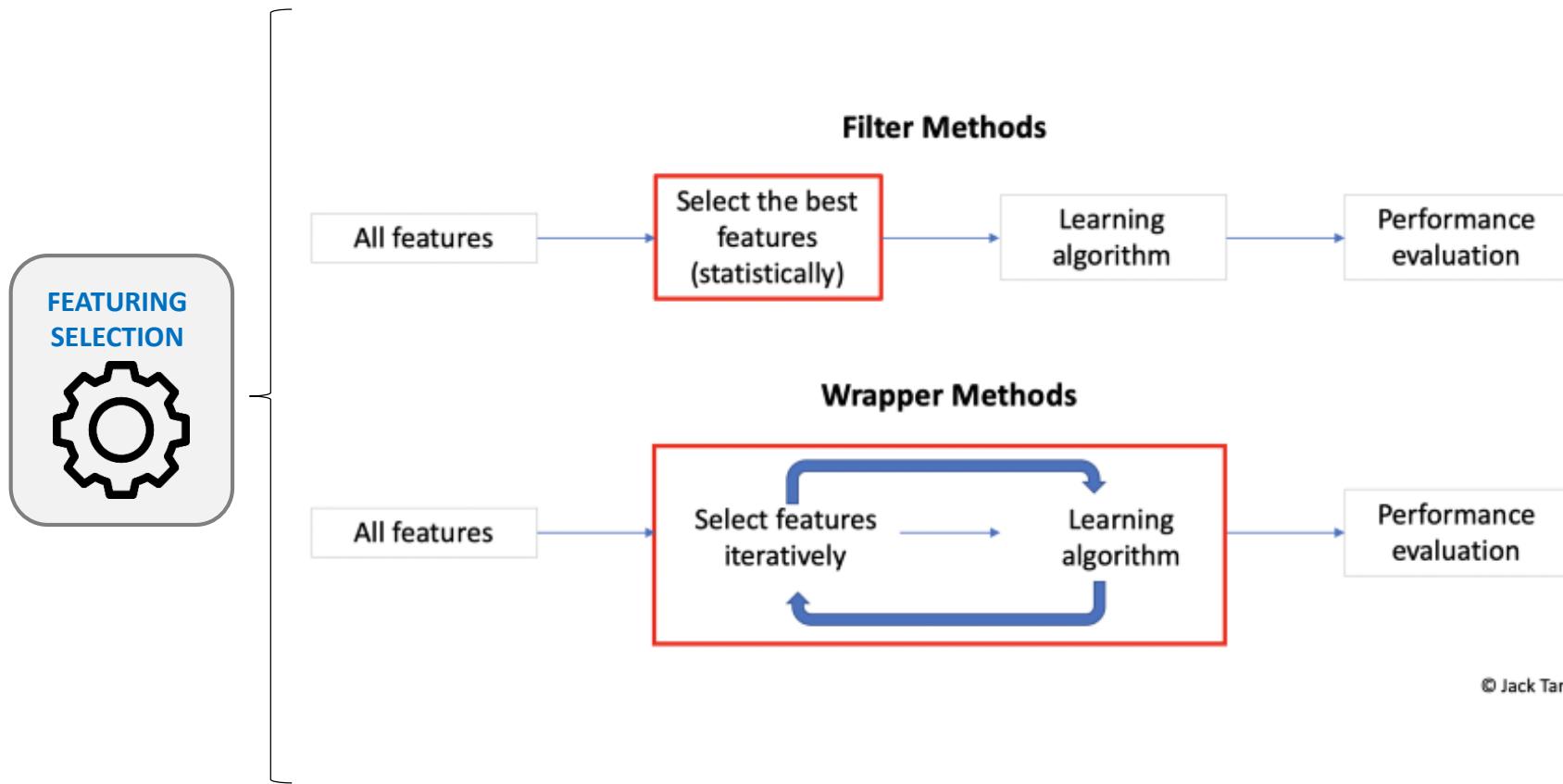


¿Cuáles son los **beneficios** de realizar la selección de atributos antes de modelar sus datos?

- Mejora la precisión.
- Reduce el tiempo de entrenamiento.



## 2. Preparar los datos



© Jack Tan



## 2. Preparar los datos

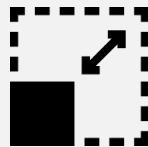
**TRANSFORMACIÓN DE LOS DATOS:** los datos pueden no estar en el formato adecuado o pueden requerir transformaciones para hacerlos más útiles.

**ESCALADO** Paso importante en la fase de pre-procesamiento, ya que la mayoría de los algoritmos de aprendizaje funcionan mucho mejor cuando los atributos están en la misma escala (rango).

**NORMALIZACIÓN:** La normalización es una buena técnica que se utiliza cuando no se conoce la distribución de los datos o cuando se sabe que la distribución no es gaussiana. Consiste en reescalar el rango de características dentro del rango [0, 1].

**ESTANDARIZACIÓN:** consiste en centrar las columnas de las características en la media 0 con una desviación estándar de 1 (distribución normal con media cero y desviación unitaria). Asume que los atributos tienen una distribución gaussiana. Esto no tiene porque ser estrictamente cierto, pero la técnica es más eficaz si su distribución es normal. Mantiene información útil sobre los valores atípicos y hace que los algoritmos sean menos sensibles a ellos.

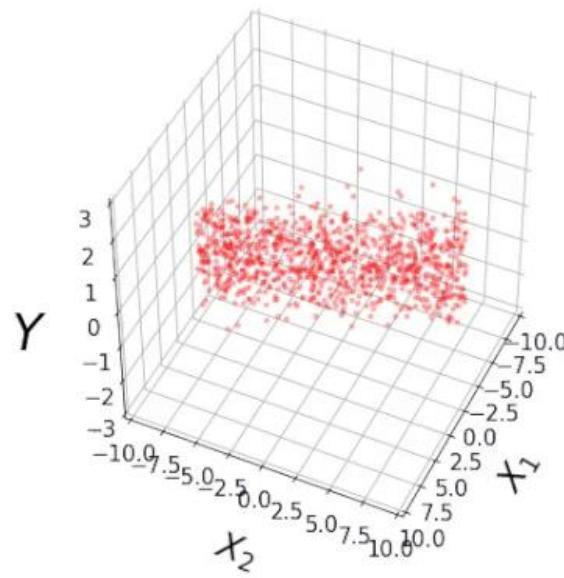
TRANSFORMACIÓN  
DE LOS DATOS



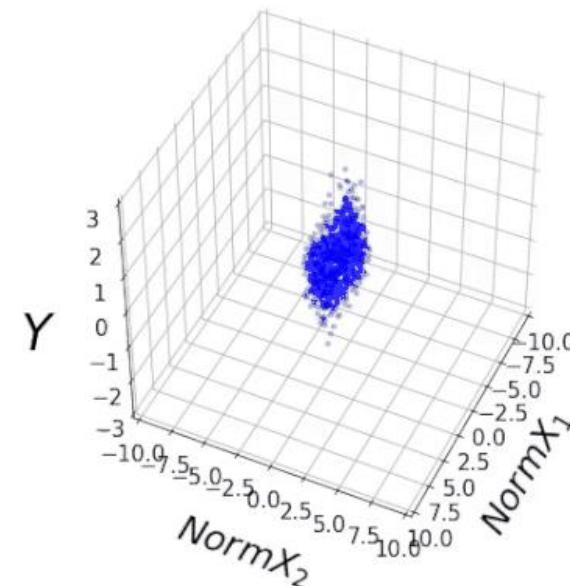


## 2. Preparar los datos

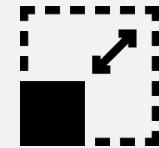
Original



Normalized



TRANSFORMACIÓN  
DE LOS DATOS



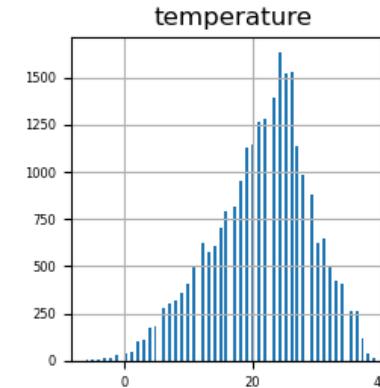
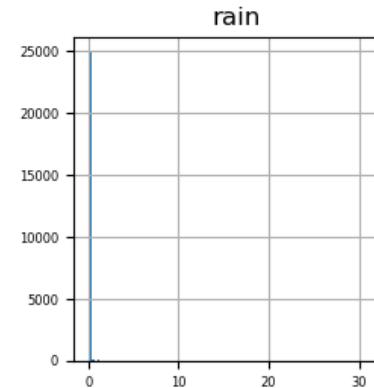
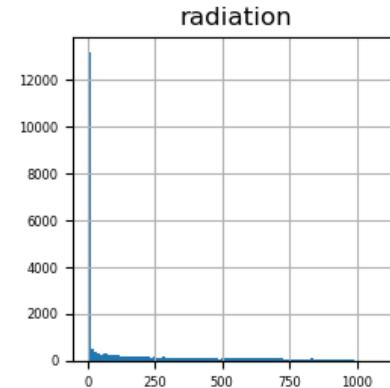
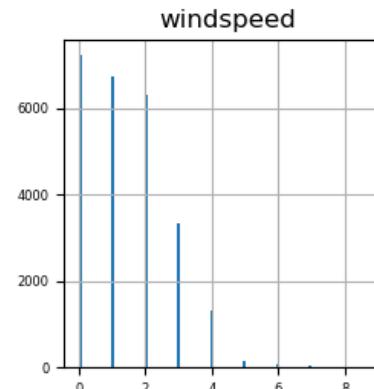
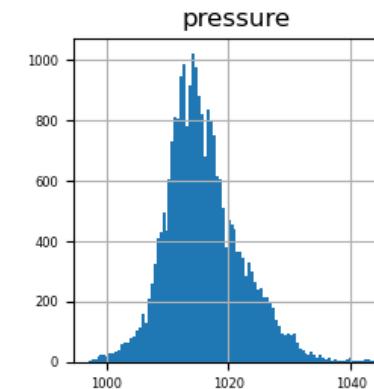
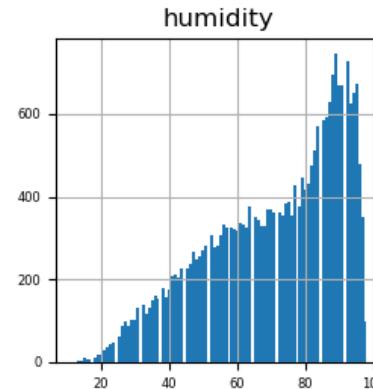
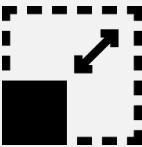


## 2. Preparar los datos

**TRANSFORMACIÓN DE LOS DATOS:** los datos pueden no estar en el formato adecuado o pueden requerir transformaciones para hacerlos más útiles.

¿Hace falta escalar estos datos?

TRANSFORMACIÓN  
DE LOS DATOS





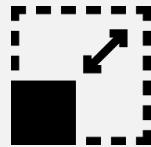
## 2. Preparar los datos

**TRANSFORMACIÓN DE LOS DATOS:** los datos pueden no estar en el formato adecuado o pueden requerir transformaciones para hacerlos más útiles.

**DATOS CATEGÓRICOS:** Los algoritmos de Machine Learning requieren que las variables de entrada (atributos/características) y salida (variable objetivo) sean números, por lo que los datos categóricos deben ser codificados en números.

- **Integer encoding:** cada etiqueta única es mapeada a un entero.
- **One hot encoding:** cada etiqueta es mapeada a un vector binario.

TRANSFORMACIÓN  
DE LOS DATOS



INTEGER ENCODING

Talla
S
S
M
L
M
S



Talla
1
1
2
3
2
1

Animal

Perro
Perro
Gato
Vaca
Gato
Perro

ONE HOT ENCODING

Perro	Gato	Vaca
1	0	0
1	0	0
0	1	0
0	0	1
0	1	0
1	0	0



### 3. Dividir los datos

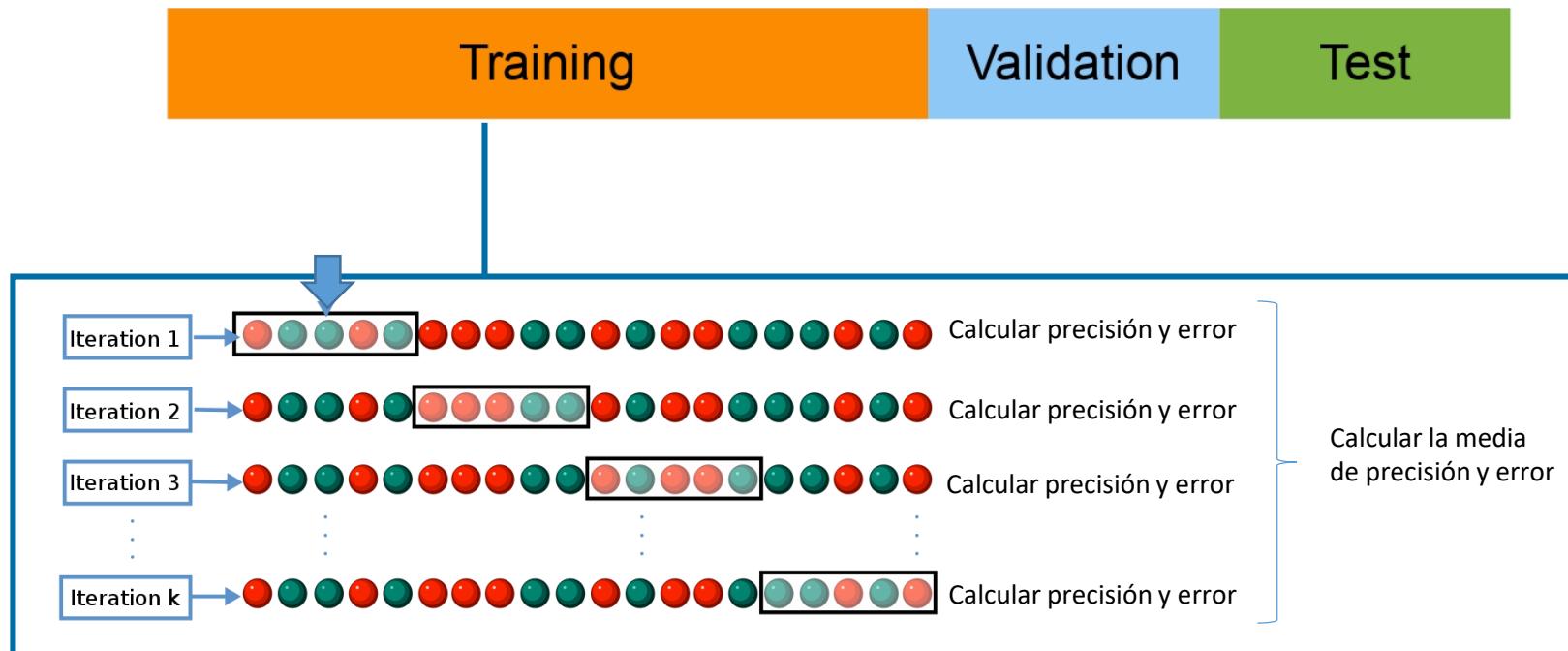
- **Datos de entrenamiento (train):** subconjunto de los datos utilizados para entrenar al modelo.
- **Datos de validación (validation):** subconjunto de datos que se usa para ajustar hiperparámetros.
- **Datos de prueba (test):** Subconjunto que se usa para testear un modelo.





### 3. Dividir los datos

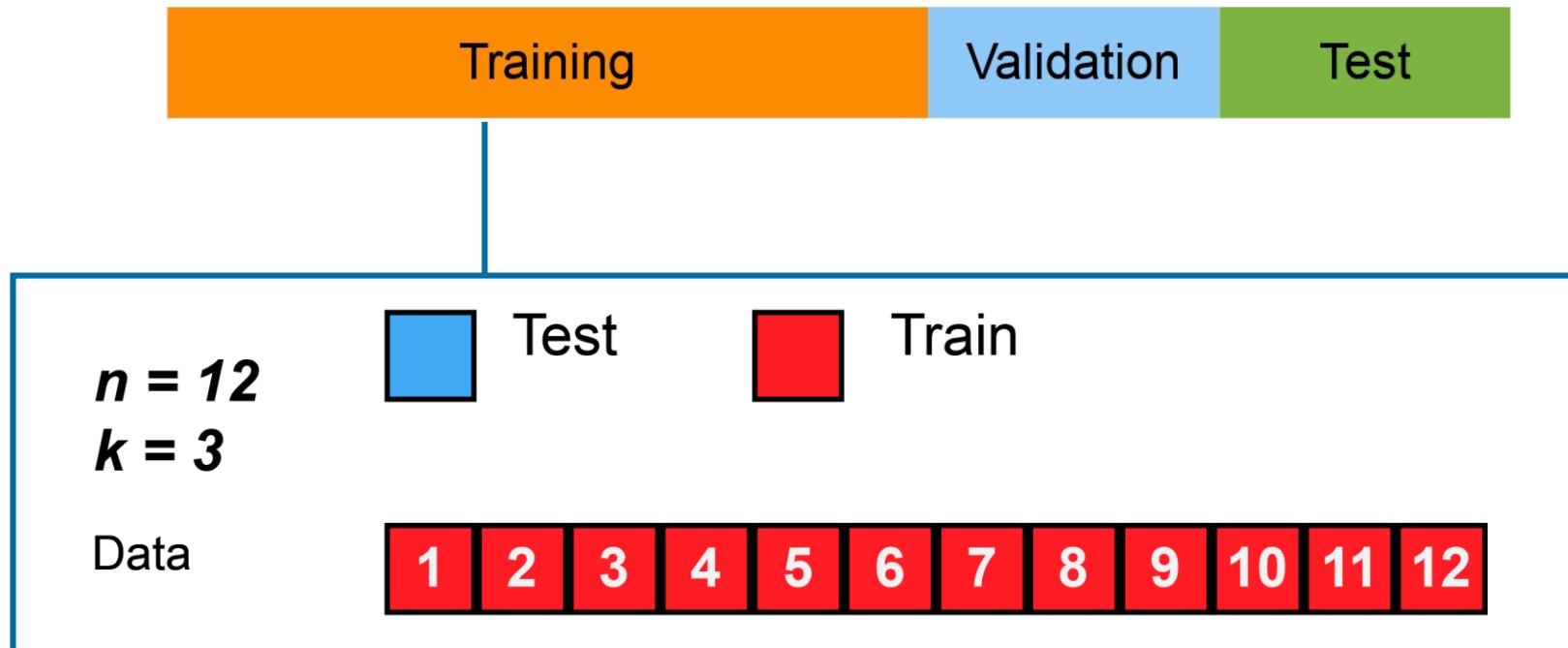
**VALIDACIÓN CRUZADA:** Técnica para evaluar modelos que coge los datos originales y crear a partir de ellos dos conjuntos separados: un primer conjunto de entrenamiento, y un segundo conjunto de *validación* (diferentes a los datos de validación en azul).





### 3. Dividir los datos

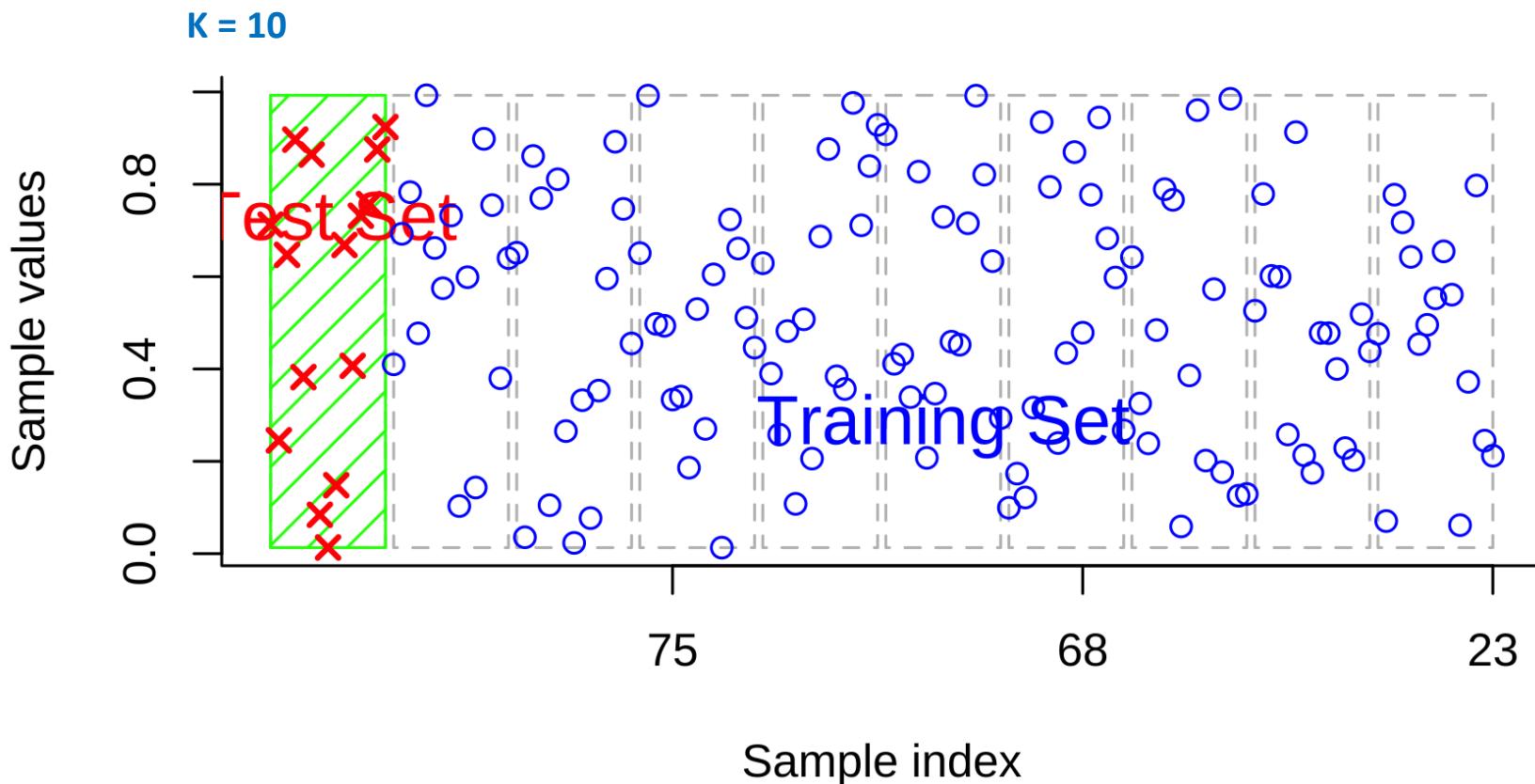
**VALIDACIÓN CRUZADA:** Técnica para evaluar modelos que coge los datos originales y crear a partir de ellos dos conjuntos separados: un primer conjunto de entrenamiento, y un segundo conjunto de *validación* (diferentes a los datos de validación en azul).





### 3. Dividir los datos

VALIDACIÓN CRUZADA:



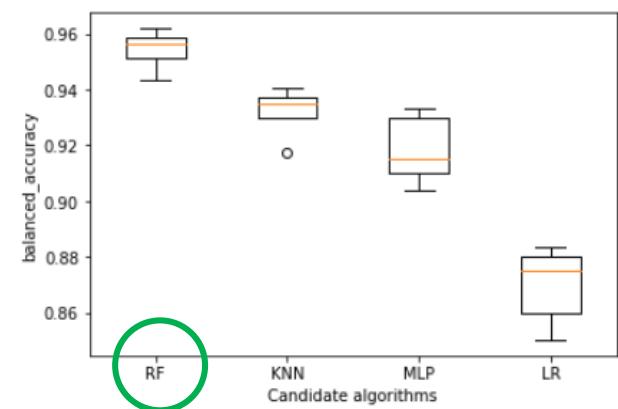


## 4. Construcción y evaluación de modelos

### Pasos principales para la construcción y evaluación de modelos

- Se seleccionan diferentes modelos de Machine Learning que se ajusten a lo que se quiere predecir.
- Se construye, entrena y valida cada modelo.
- Se comparan resultados y se escoge aquel modelo que demuestre mejor resultado, es decir, el que sea más exacto y preciso (bajo sesgo y baja varianza).
- Una vez escogido el modelo, se ajustan sus **hiperparámetros**.

Comparación de algoritmos con métrica de evaluación: balanced\_accuracy



Para estimar la habilidad de los modelos de aprendizaje, se utiliza un método llamado **cross-validation** (validación cruzada).

En este método se indica previamente:

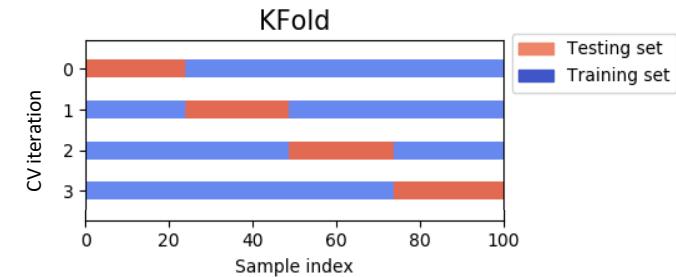
- Un número de pliegues en los que dividiremos nuestros datos (normalmente k=10).
- Una métrica de evaluación que variará dependiendo de la naturaleza del problema (regresión, clasificación...)



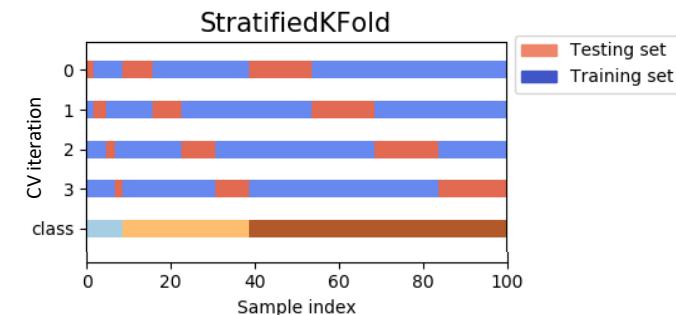
## 4. Construcción y evaluación de modelos

### TIPOS DE CROSS-VALIDATION:

**K-FOLD:** los datos de muestra se dividen en K subconjuntos. Uno de los subconjuntos se utiliza como datos de prueba o validación y el resto ( $k-1$ ) como datos de entrenamiento. Finalmente se realiza la media aritmética de los resultados de cada iteración para obtener un único resultado.



**STRATIFIED K-FOLD:** es una variación del K-fold que devuelve pliegues estratificados: cada conjunto contiene aproximadamente el mismo porcentaje de muestras de cada clase. Sólo para modelos de **clasificación** con el **target no balanceado** (más muestras de una clase que de otra).

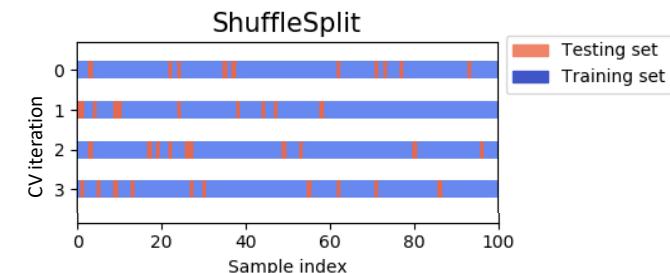




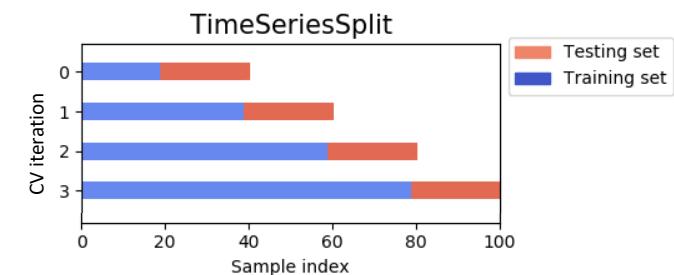
## 4. Construcción y evaluación de modelos

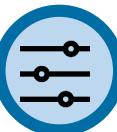
### TIPOS DE CROSS-VALIDATION:

**SHUFFLE:** ([Variante para clasificación: \*StratifiedShuffleSplit\*](#)) Este método consiste en dividir aleatoriamente el conjunto de datos de entrenamiento y prueba o validación. La ventaja es que la división de datos entrenamiento-prueba no depende del número de iteraciones. En cambio, con este método hay algunas muestras que quedan sin evaluar y otras que se evalúan más de una vez, es decir, los subconjuntos de prueba y entrenamiento se pueden solapar.



**TIME SERIES:** se utiliza para entrenar datos que son series temporales dependientes. Los conjuntos de entrenamiento sucesivos son superconjuntos de la iteración anterior.

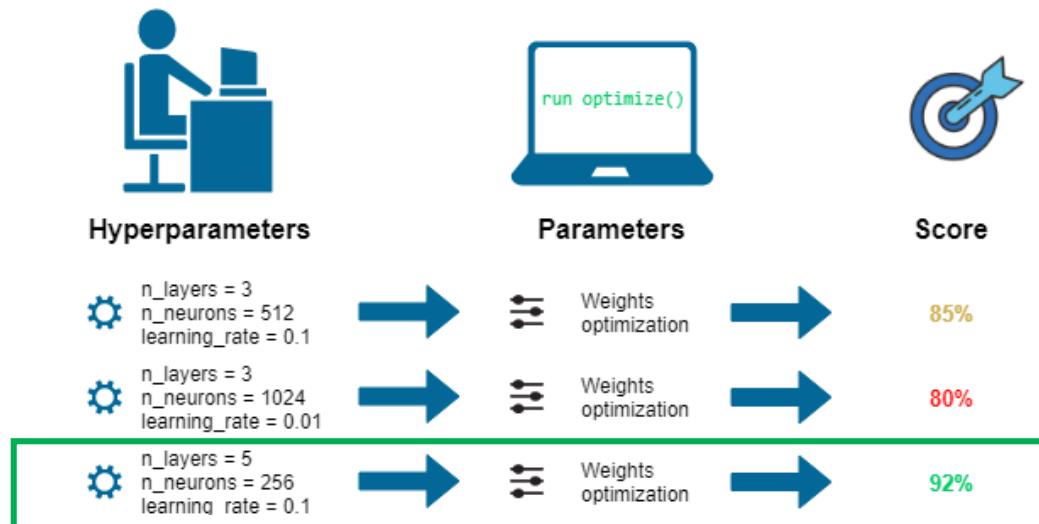


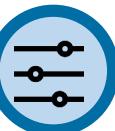


## 5. Ajustar hiperparámetros

### OPTIMIZACIÓN DE HYPERPARÁMETROS

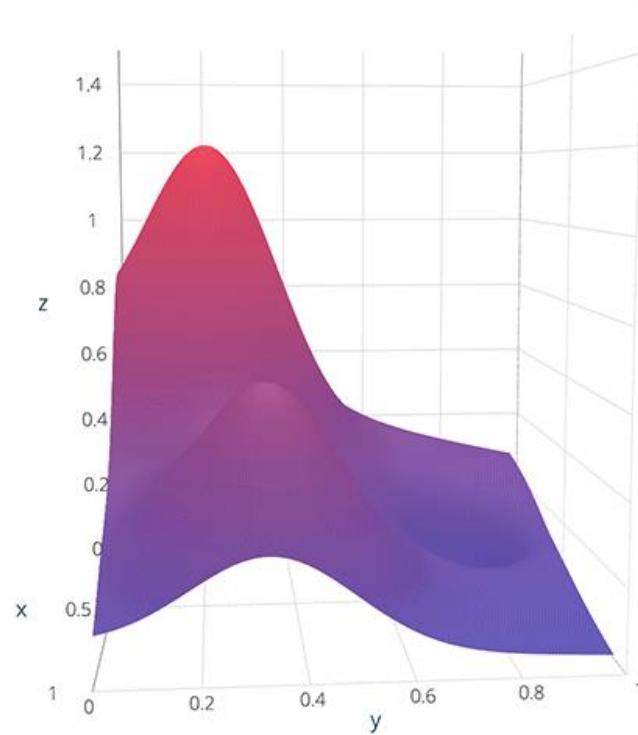
- Cada modelo de Machine Learning tiene un amplio número de parámetros que se utilizan para controlar el proceso de aprendizaje. Se llaman hiperparámetros a los parámetros del modelo que pueden ser modificados y, dependiendo del conjunto de datos, pueden dar lugar a un aumento del rendimiento del modelo.
- El proceso de encontrar el mejor conjunto parámetros para un modelo y un conjunto de datos se conoce como **optimización de hiperparámetros**.
- Se suele utilizar el subconjunto de datos de validación para ajustar los hiperparámetros.
- Un método común para la optimización de hiperparámetros se conoce como **GridSearch**.

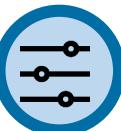




## 5. Ajustar hiperparámetros

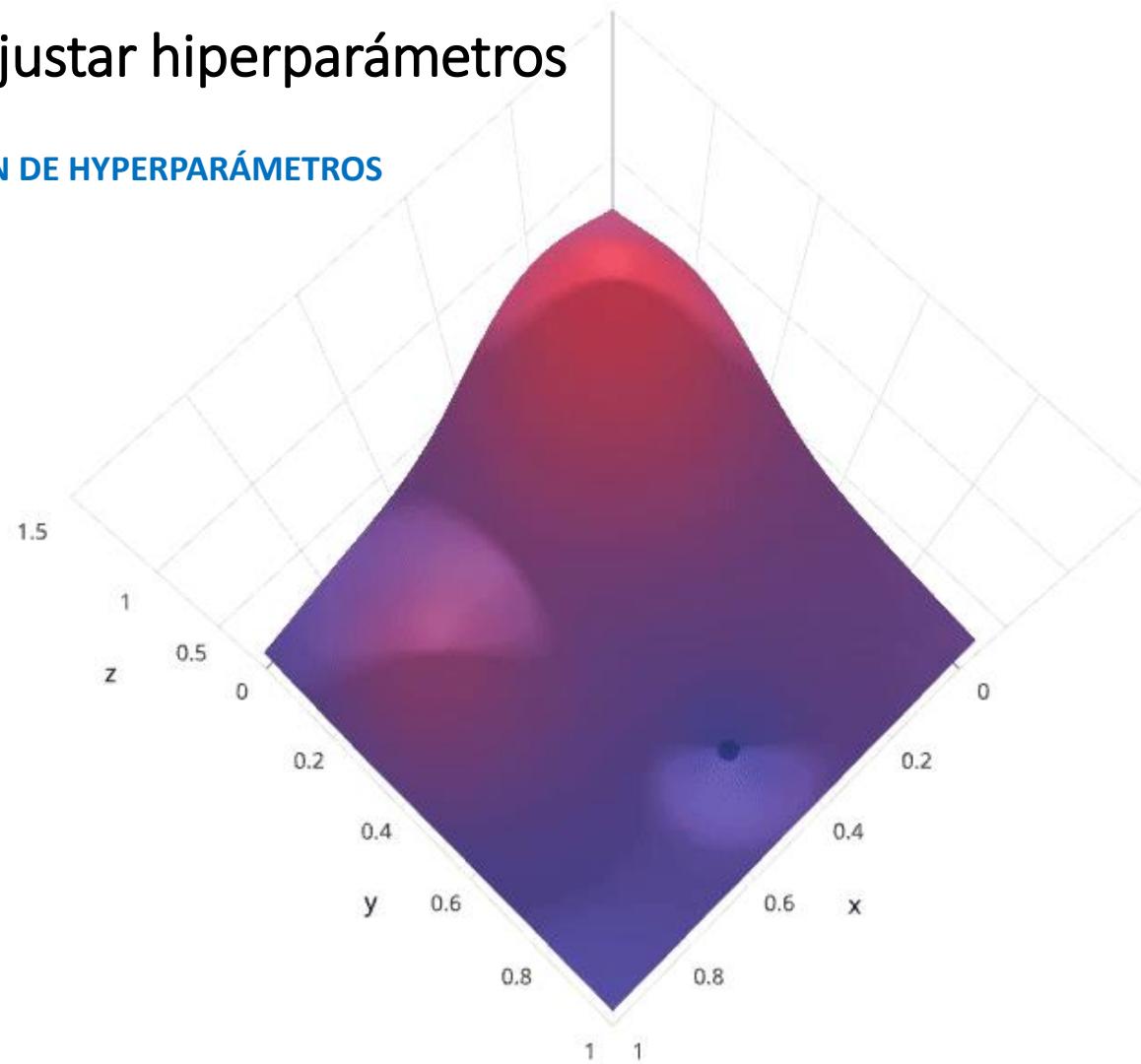
### OPTIMIZACIÓN DE HYPERPARÁMETROS

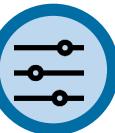




## 5. Ajustar hiperparámetros

### OPTIMIZACIÓN DE HYPERPARÁMETROS





## 5. Ajustar hiperparámetros

### EJEMPLOS de HIPERPARÁMETROS DE ALGORITMOS

#### sklearn.ensemble.RandomForestClassifier()

```
class sklearn.ensemble.RandomForestClassifier(n_estimators=100, *, criterion='gini', max_depth=None, min_samples_split=2,
min_samples_leaf=1, min_weight_fraction_leaf=0.0, max_features='auto', max_leaf_nodes=None, min_impurity_decrease=0.0,
min_impurity_split=None, bootstrap=True, oob_score=False, n_jobs=None, random_state=None, verbose=0, warm_start=False,
class_weight=None, ccp_alpha=0.0, max_samples=None)
```

[\[source\]](#)

##### Parameters:

###### **n\_estimators : int, default=100**

The number of trees in the forest.

*Changed in version 0.22: The default value of n\_estimators changed from 10 to 100 in 0.22.*

###### **criterion : {"gini", "entropy"}, default="gini"**

The function to measure the quality of a split. Supported criteria are "gini" for the Gini impurity and "entropy" for the information gain. Note: this parameter is tree-specific.

###### **max\_depth : int, default=None**

The maximum depth of the tree. If None, then nodes are expanded until all leaves are pure or until all leaves contain less than min\_samples\_split samples.

###### **min\_samples\_split : int or float, default=2**

The minimum number of samples required to split an internal node:

- If int, then consider min\_samples\_split as the minimum number.
- If float, then min\_samples\_split is a fraction and ceil(min\_samples\_split \* n\_samples) are the minimum number of samples for each split.



## 6. Evaluación final del modelo

- Evaluar al modelo final con métricas adecuadas:
  - Métricas de clasificación: Precision-Recall, ROC-AUC, acierto ...
  - Métricas de regresión: RMSE, MAE,  $R^2$  ...
  - Métricas de clustering: Rand Index, información mútua...

***"Si no puedes medirlo no puedes mejorarlo"***

- Finalmente, usamos el mejor modelo hiperparametrizado para predecir la **variable objetivo** utilizando el conjunto de **datos de prueba (test)**.
- Se obtiene un informe para evaluar su rendimiento en detalle.
- Repetir el proceso en caso de no obtener unos resultados satisfactorios. **Es un proceso iterativo!!!**



Conceptos generales.

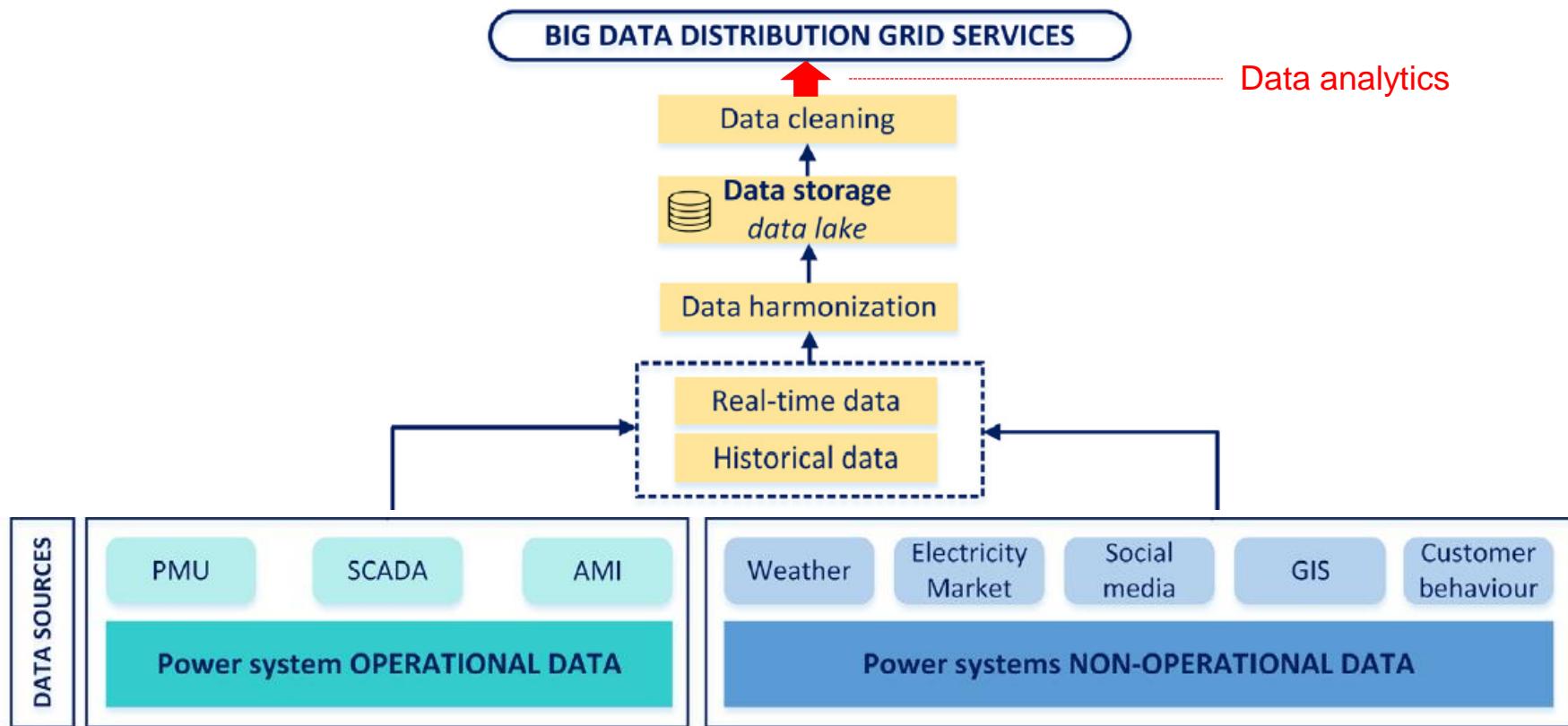
Conceptos básicos de Machine Learning.

Crear un modelo de Aprendizaje Supervisado

## **Aplicaciones de Big Data y Machine Learning en el sector energético**



## De los datos a las aplicaciones





## Aplicaciones Machine Learning en el Sector Energético

**Operación y monitorización de la red:** topología, observación y estabilidad de la red.

**Mantenimiento predictivo** de transformadores, cableado, etc.

**Detección de faltas** en la red de baja tensión.

**Detección de fraude.**

**Predicción** de demanda, generación, precio de la electricidad en el mercado diario.

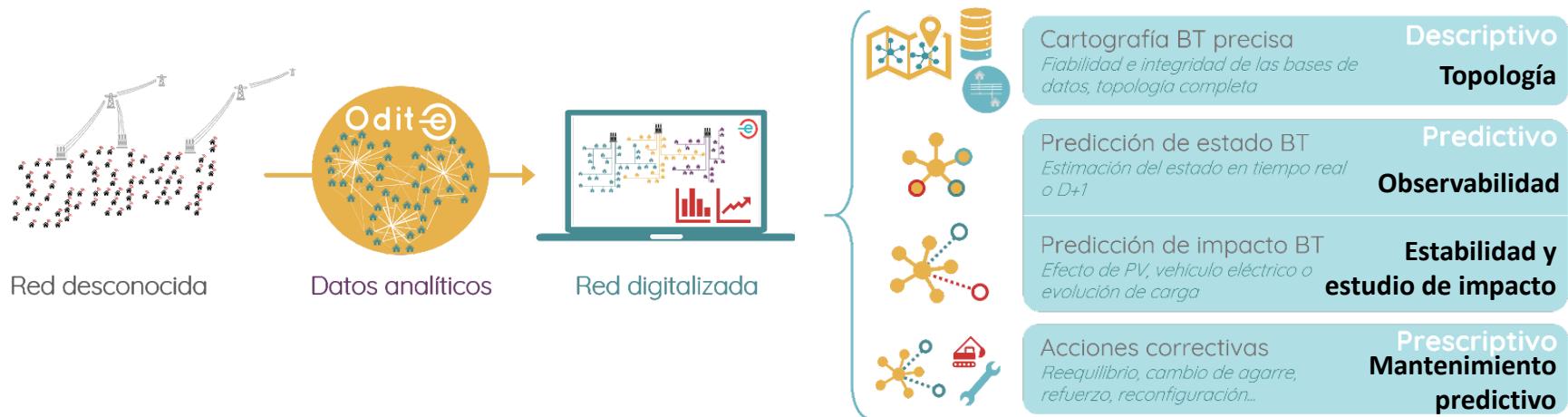
**Sistema de gestión energético.**

**P2P Trading and cyber security.**

**Plan de inversión y de equipos en la red.**



## Operación y monitorización de la red: topología y observación



Servicios que ofrece la empresa Odit-e



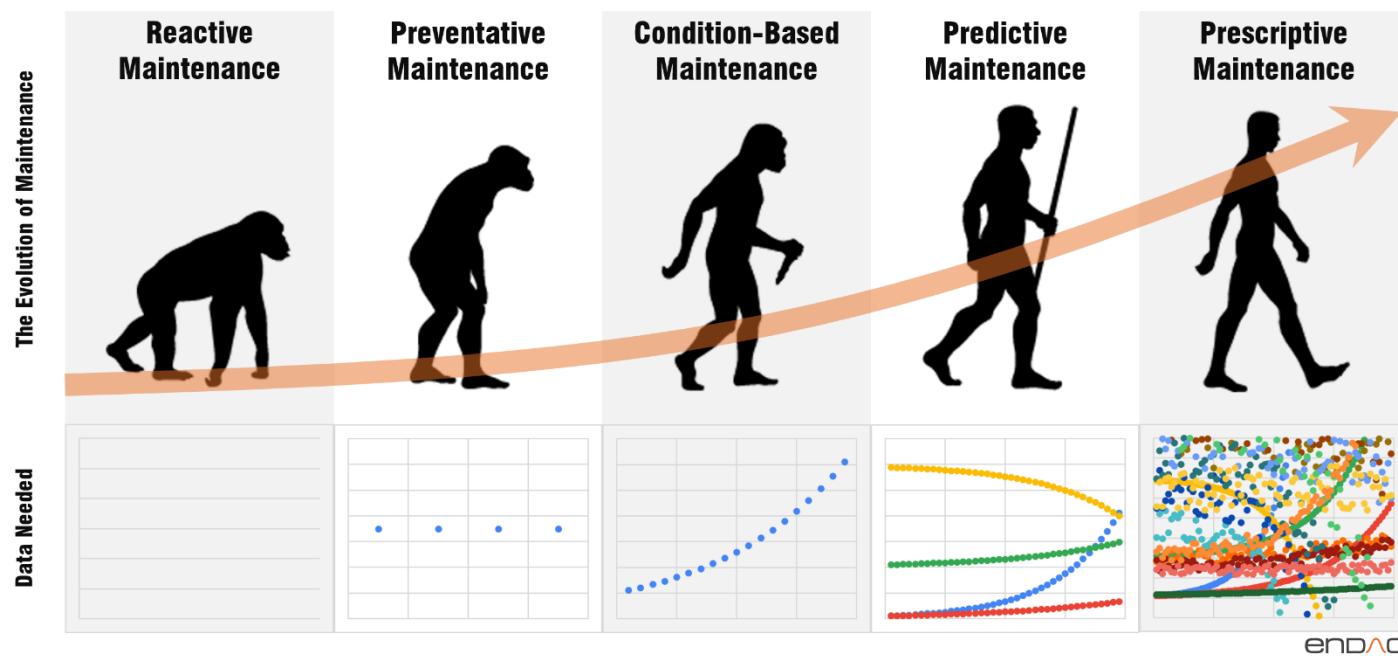
## Operación y monitorización de la red: topología y observación

Data sources for distribution network topology estimation.

Data source	Input measurement
Electricity market	Electricity prices
PMU	Voltage phasor
SCADA	Power injections Voltage magnitude Current Switch status
Smart meters	Power injections Voltage magnitude

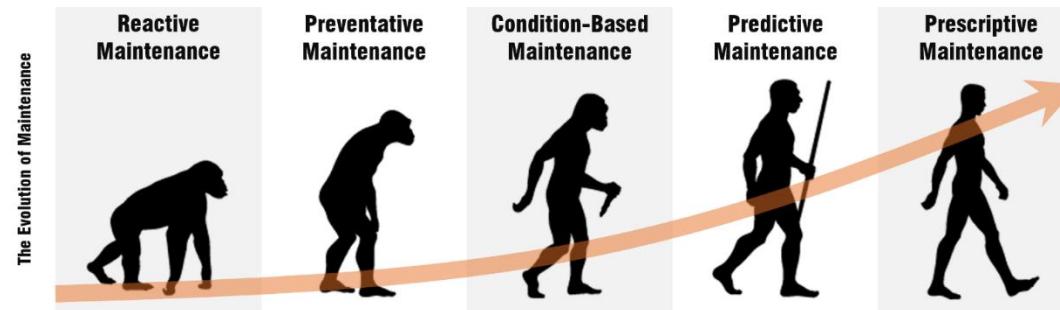


**Mantenimiento predictivo:** transformadores, cableado, aerogeneradores, etc.





## Mantenimiento predictivo: transformadores, cableado, aerogeneradores, etc.

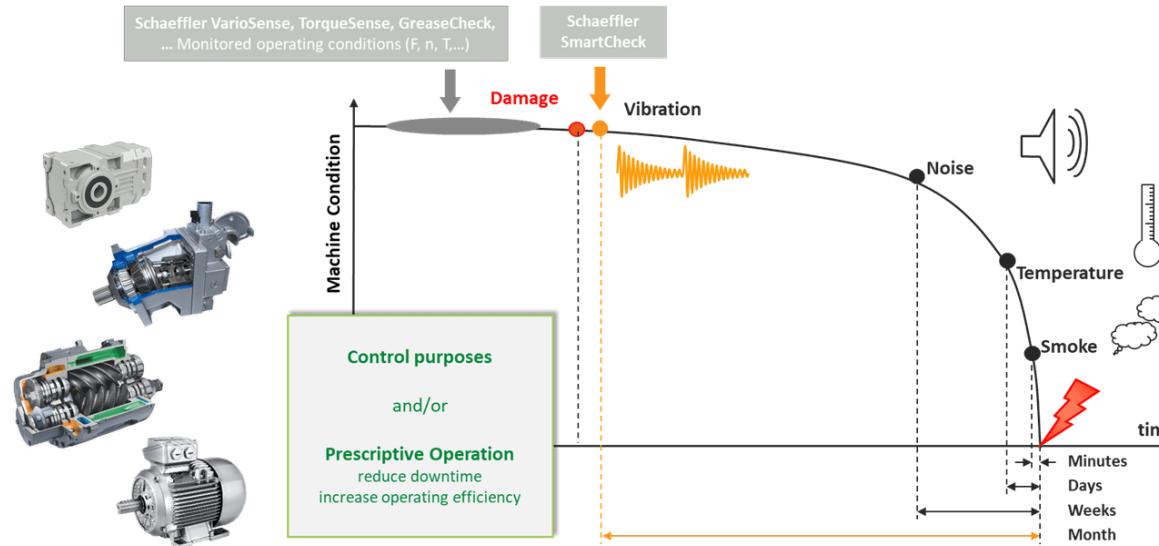


- El **mantenimiento reactivo** no depende en absoluto de los datos, sino que espera al fallo.
- El **mantenimiento preventivo** utiliza el tiempo o el uso para informar de su programa de mantenimiento, no son muchos datos - ¡pero es un comienzo!
- El **mantenimiento basado en la condición** examina una sola métrica y la compara con umbrales preestablecidos.
- El **mantenimiento predictivo** examina un puñado de datos y se basa en un análisis ligeramente mejor para detectar diversas tendencias en los datos y la salud del activo.
- El **mantenimiento prescriptivo** utiliza muchos conjuntos de datos y métricas y probablemente algunas técnicas de análisis propias para determinar la causa de un posible fallo.



**Mantenimiento predictivo:** transformadores, cableado, aerogeneradores, etc.

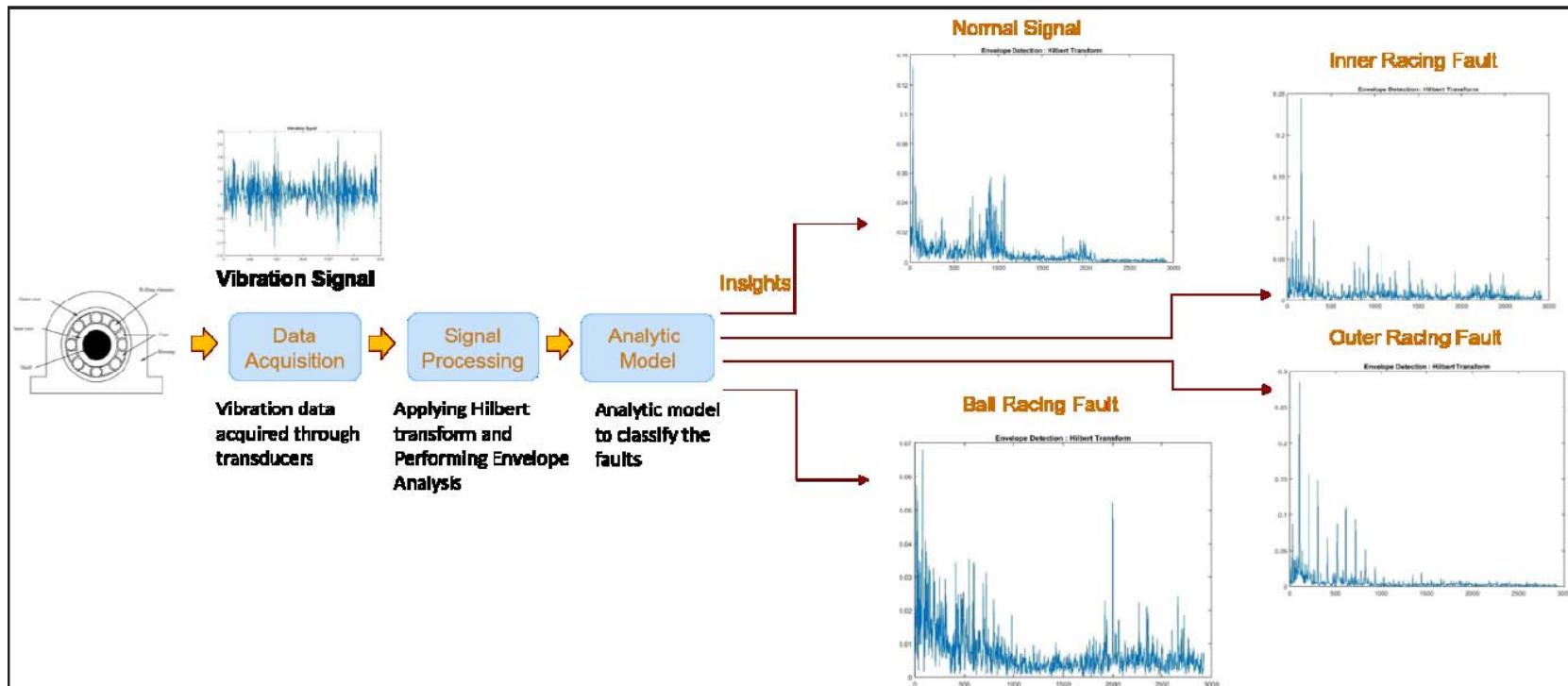
**Típicas medidas usadas:** vibración, ruido y temperatura.



The [WATEREYE project](#) tiene como objetivo desarrollar un conjunto de sensores, sistemas de monitorización, estrategias de diagnóstico y mantenimiento predictivo, y herramientas de apoyo a la toma de decisiones relacionadas con el funcionamiento y el mantenimiento de los parques eólicos marinos (componentes estructurales de las turbinas eólicas)



**Mantenimiento predictivo:** transformadores, cableado, aerogeneradores, etc.

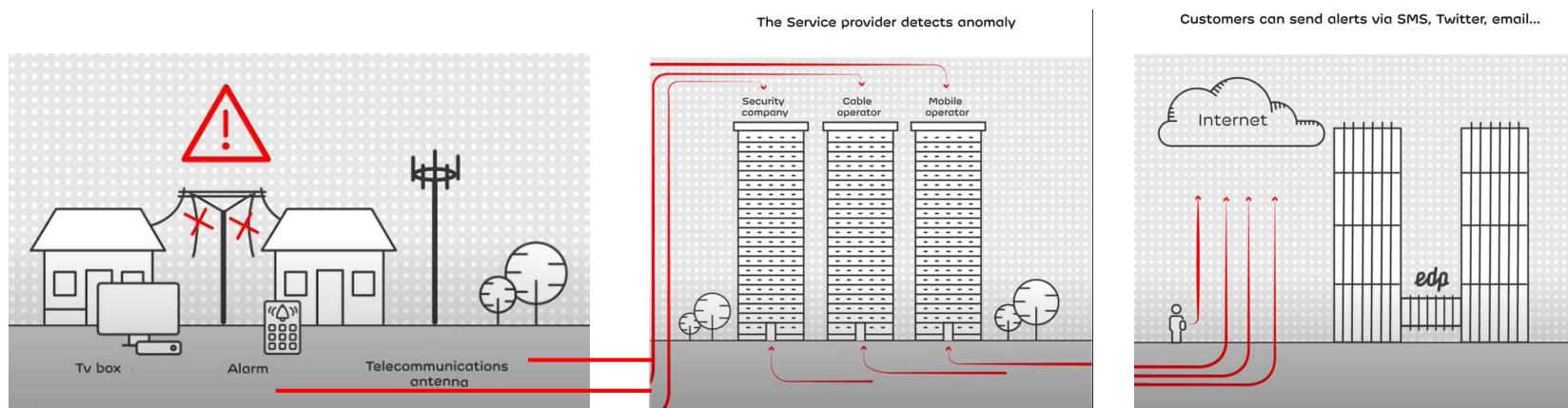


Gopi Krishna Durbhaka, Barani Selvaraj



## Detección de faltas en la red de baja tensión

- Mejora de la calidad del servicio prestado a los clientes
- Minimización de los tiempos de inactividad
- Automatización de los procesos que actualmente dependen de la intervención humana
- Gestión más eficiente del personal gracias a una mayor visibilidad de las causas de las anomalías del sistema.



**Proyecto SINAPSE (EDP):** Una plataforma para informar de las anomalías en la red eléctrica.

**Objetivo:** reestablecer la electricidad lo más rápido posible. ([Vídeo](#))



## Detección de fraude

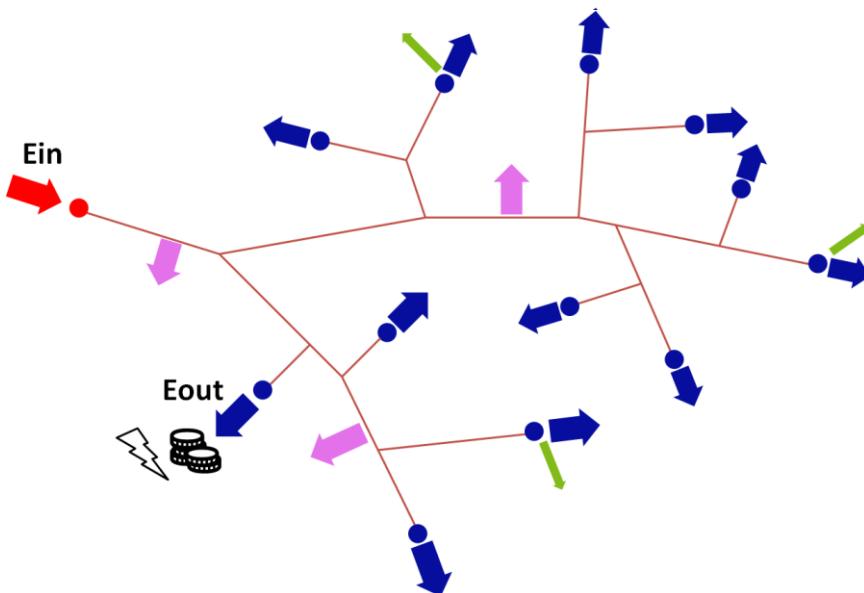
Las pérdidas de electricidad en los niveles de distribución abarcan tanto las pérdidas técnicas como las no técnicas (NTL). Las primeras se producen por el efecto Joule, mientras que las NTL se refieren a la electricidad consumida pero no facturada.





## Detección de fraude

- Meter measuring energy entering the grid
- Meters measuring energy consumed



**Pérdidas en la red de distribución = TL + NTL**

- **TL (Technical losses):** Pérdidas variables, pérdidas fijas y pérdidas del servicio de red (consumos no contratados de los equipos de red)
- **NTL (Non-technical losses):** Problemas de equipos de red, problemas de información de red y problemas de procesamiento de datos de energía

**Conexión directa**  
**Meter Tampering**



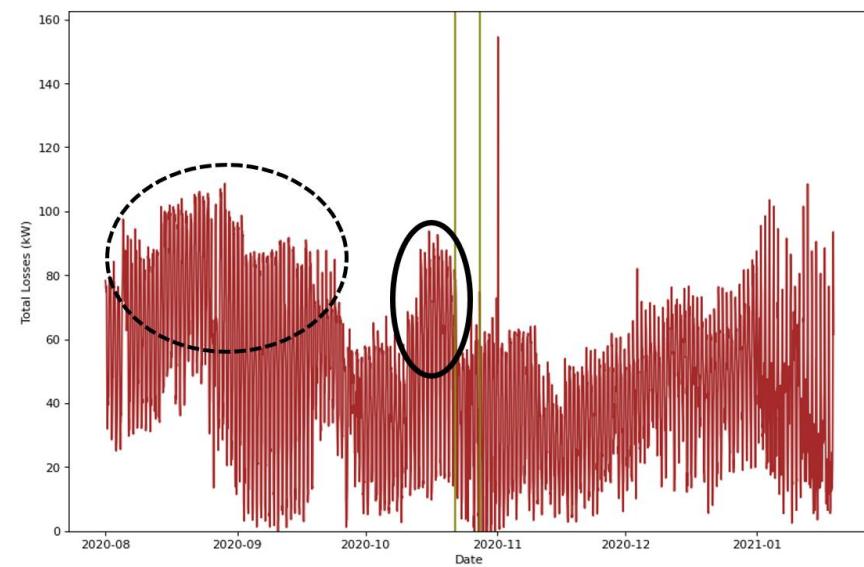
## Detección de fraude

### Conexión directa

- Tipología de fraude más común en la actualidad
- Diferentes naturalezas de fraude (plantación de cannabis, minería de bitcoin, ocupación, imposibilidad de pagar facturas altas...)
- Más fácil de detectar in situ por los operadores (aunque a veces se oculta la conexión)
- Problema: este consumo no es registrado por ningún dispositivo de medición
- Posibles enfoques: Analizar la curva de LNT // Analizar la caída de tensión a lo largo de la red.

### Ejemplo:

- Reducción de  $\approx 20$  kW después de la inspección.
- Ciclo similar (y sospechoso) hace 2-3 meses.
- Gran variabilidad, probablemente debida a la diferencia de consumo entre el día y la noche



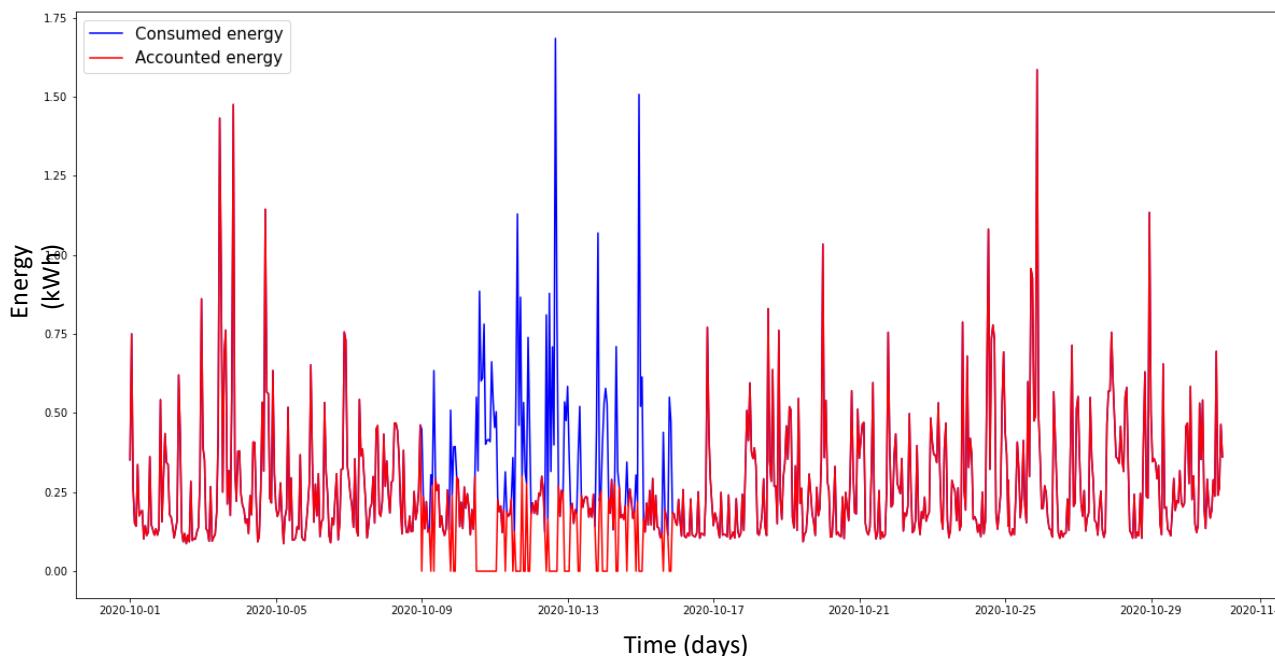
Total Losses curve



## Detección de fraude

### Meter Tampering

- Tendencia al alza debido a la implantación de contadores inteligentes
- Los cambios en el consumo se registran en el SM, pero son más difíciles de detectar por el operador.
- Posibles enfoques: Agrupación de patrones cortos // Agrupación de todos los SM
- Problema: los cambios sutiles son difíciles de detectar.

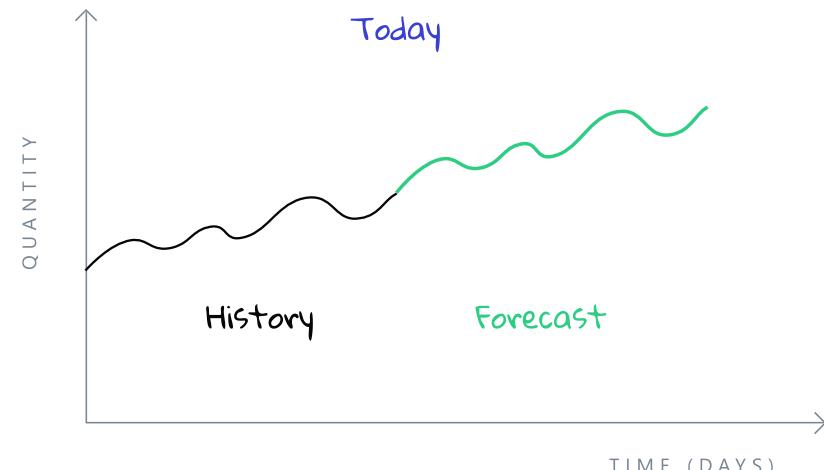




## Predicción (forecasting)

Generación, demanda, flexibilidad, previsión de precios en diferentes horizontes temporales

Forecasting horizon	Time interval
Real-time	$t \leq 1 \text{ h}$
Short-term	$1 \text{ h} < t < 1 \text{ week}$
Medium-term	$1 \text{ week} < t < 12 \text{ months}$
Long-term	$t > 1 \text{ year}$





## Predicción (forecasting)

Generación, demanda, flexibilidad, previsión de precios en diferentes horizontes temporales

### Key parameters for forecasting of wind and solar generation

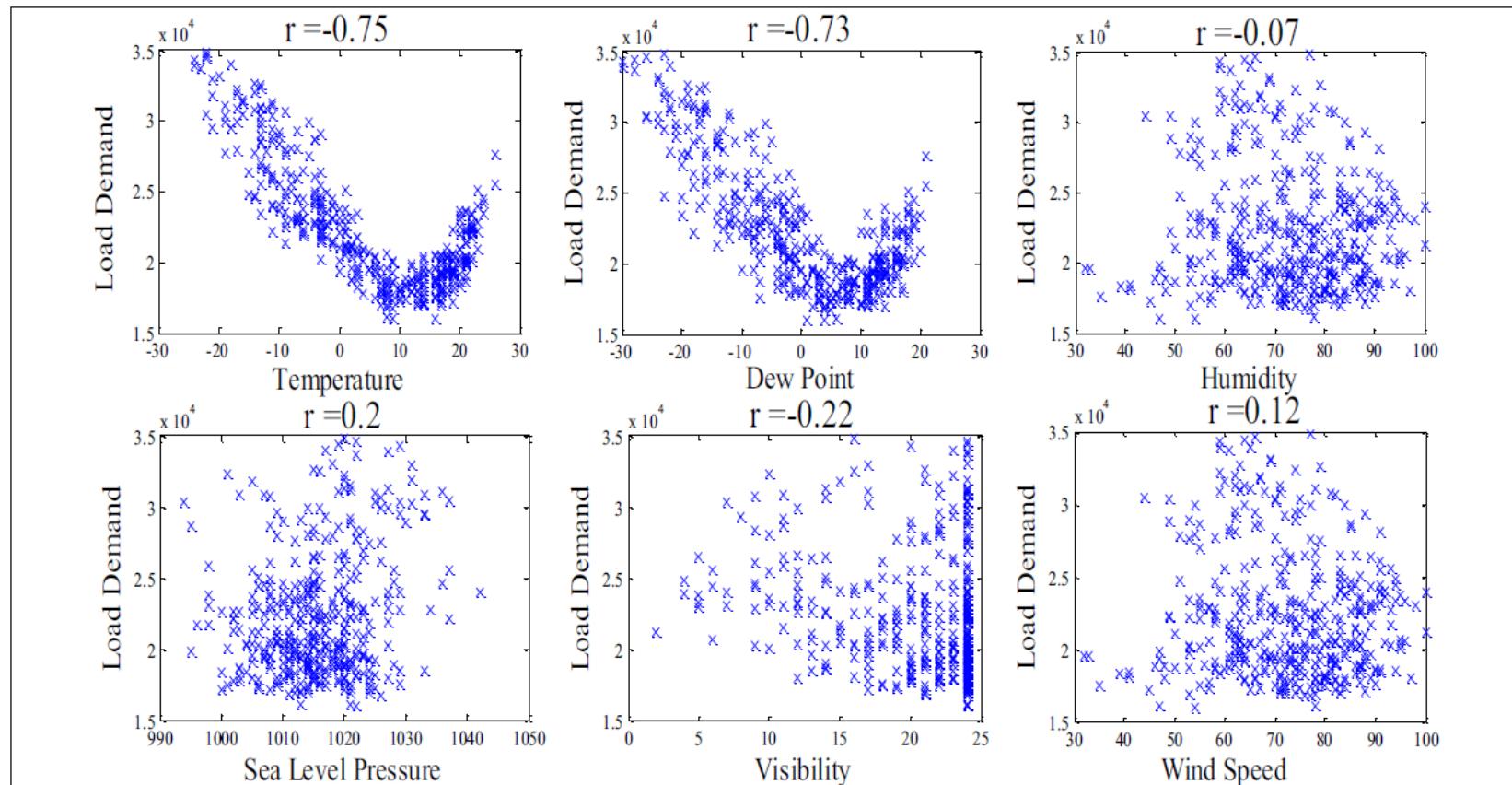
- Power output (real power output)
- Available active power (power output due to meteorological conditions)
- Information on the scheduled availability of wind farms and solar plants in terms of effective installed power, scheduled maintenance and known outages of machines
- Information on current availability
- Information on scheduled curtailment
- Information on currently activated curtailment by grid operators or despatch centres
- Meteorological data that measures the available resource:
  - Wind farms: Wind speed and direction close to hub height
  - Solar plants: Solar irradiation

Source: GIZ



## Predicción (forecasting)

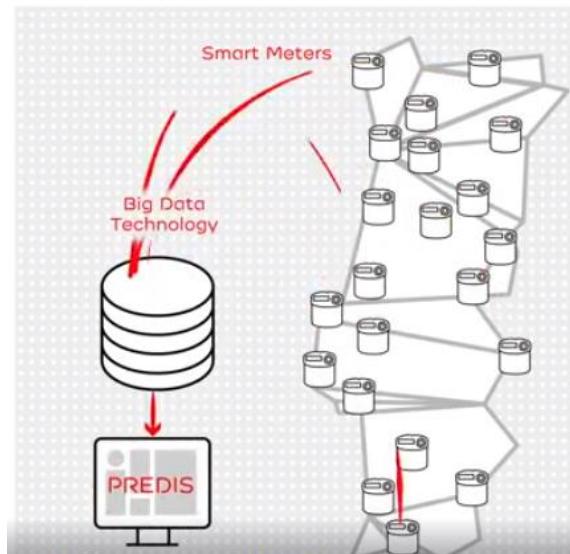
Variables exógenas para predicción de demanda



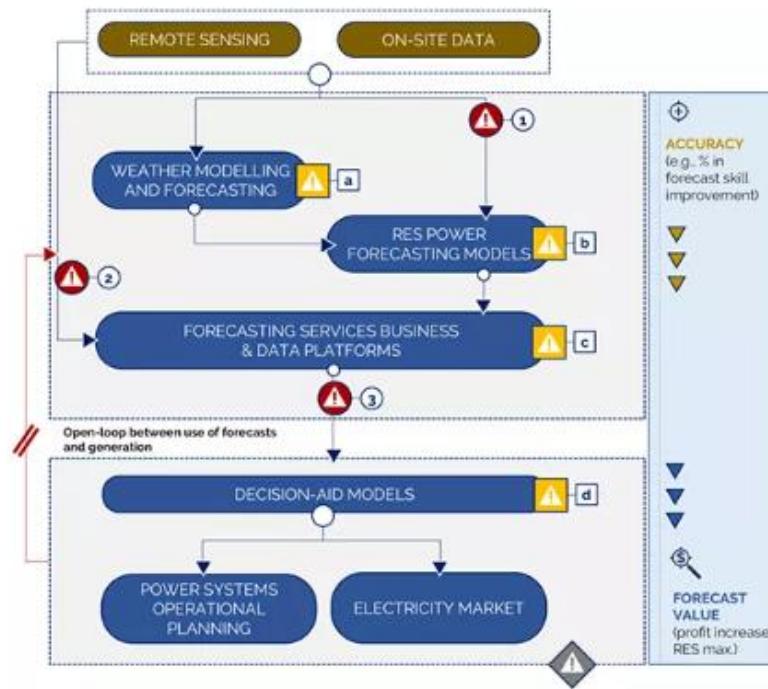


## Predicción (forecasting)

[PREDIS Project](#): Predicción de demanda y generación desagregada en tiempo real.



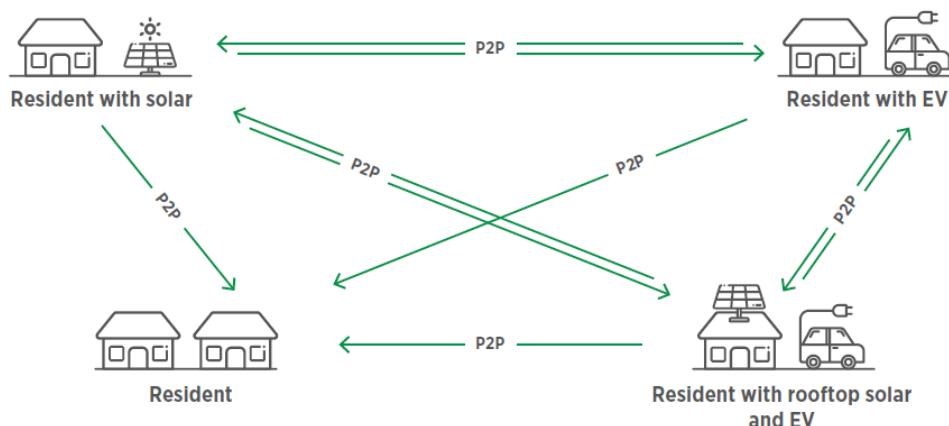
[SMART4RES](#): proponer un conjunto de herramientas de previsión de generación de energía renovable.





## P2P Trading and cyber security

**P2P Trading:** modelo de negocio, basado en una plataforma interconectada, que sirve de mercado donde consumidores y productores se "encuentran" para comerciar directamente con la electricidad, sin necesidad de intermediarios.



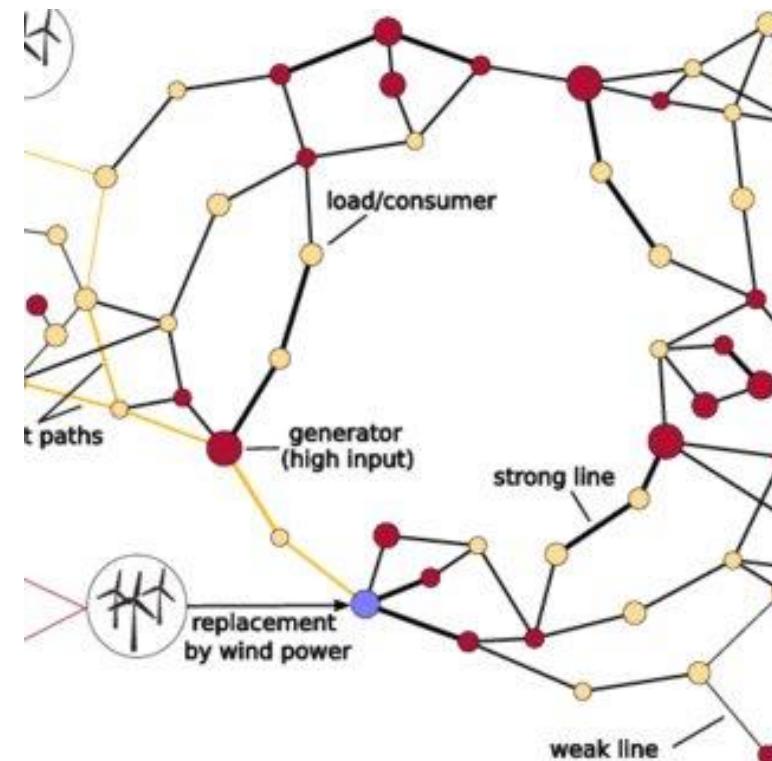
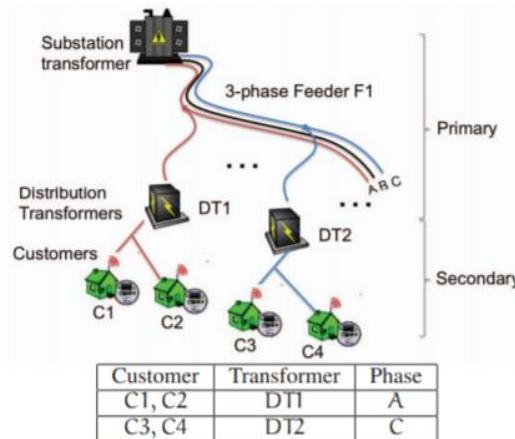
Source: Liu *et al.*, 2019, Note: The direction of the arrow indicates the accounting and transactions flow directions

[TDX-Assit](#): diseñar y desarrollar nuevos métodos y herramientas para un **intercambio seguro y confiable de datos** entre el operador del sistema de transmisión y el operador de los sistemas de distribución.





## Plan de inversión y de equipos en la red.



["Machine learning for inferring phase connectivity in distribution networks"](#),  
Sambaran Bandyopadhyay et al .  
(TOPOLOGY service)



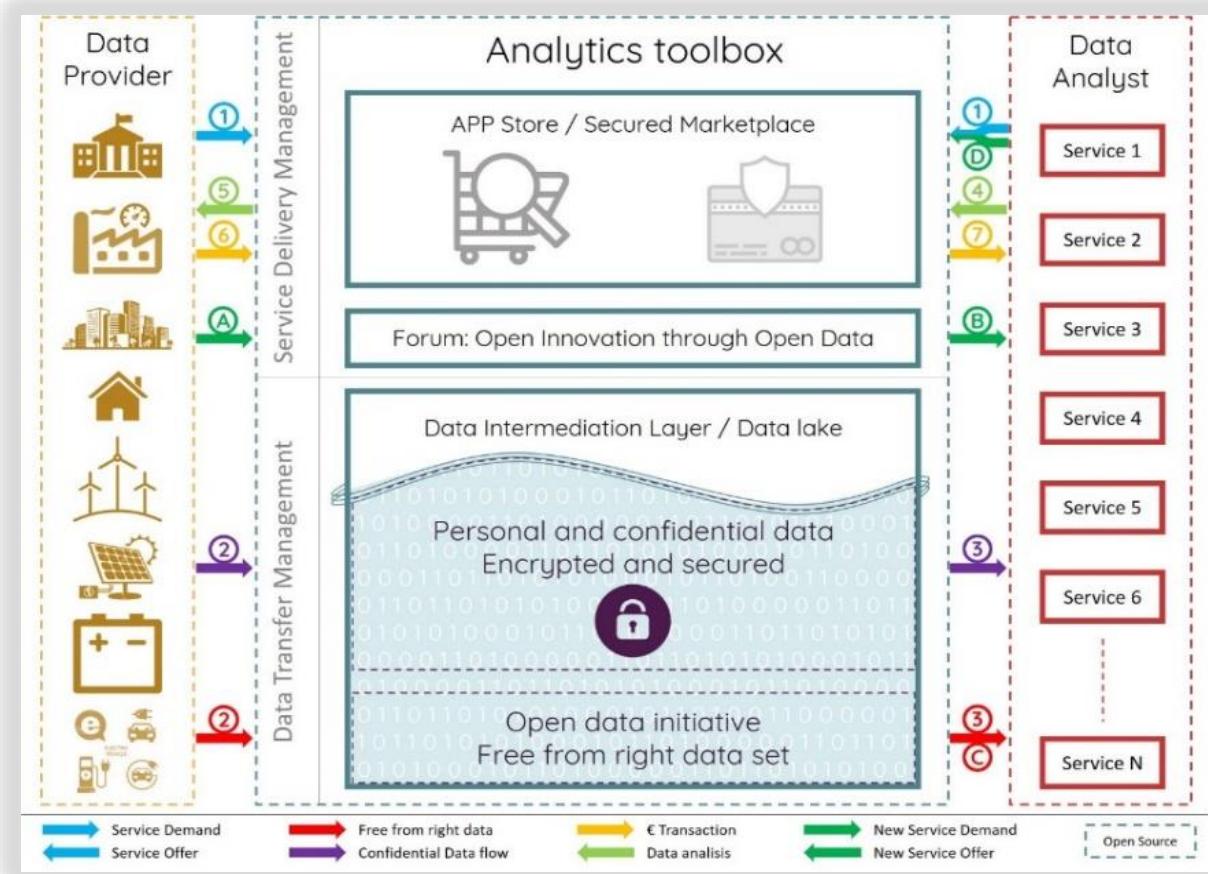
# BD4OPEM: Big data for Open Innovation Energy Marketplace

## BD4OPEM H2020

### AI based services

- Congestions identification
- Grid planning
- Predictive maintenance
- EMS
- Flexibility services
- V2G services
- Fraud detection

3 DSOs (ESTABANELL, ELCE, OEDAS), Bornhom island (NUVVE), Brussels Health Campus (VUB)





# Bibliografía recomendada

1. Glosario sobre aprendizaje automático  
<https://developers.google.com/machine-learning/glossary>
2. *Artificial intelligence techniques for enabling Big Data services in distribution networks: A review*, Sara Barja et al., Renewable and Sustainable Energy Reviews, 2021
3. *Big data analytics for future electricity grids*, Mladen Kezunovica et al, Electric Power Systems Research, 2020