

Machine y deep learning en salud

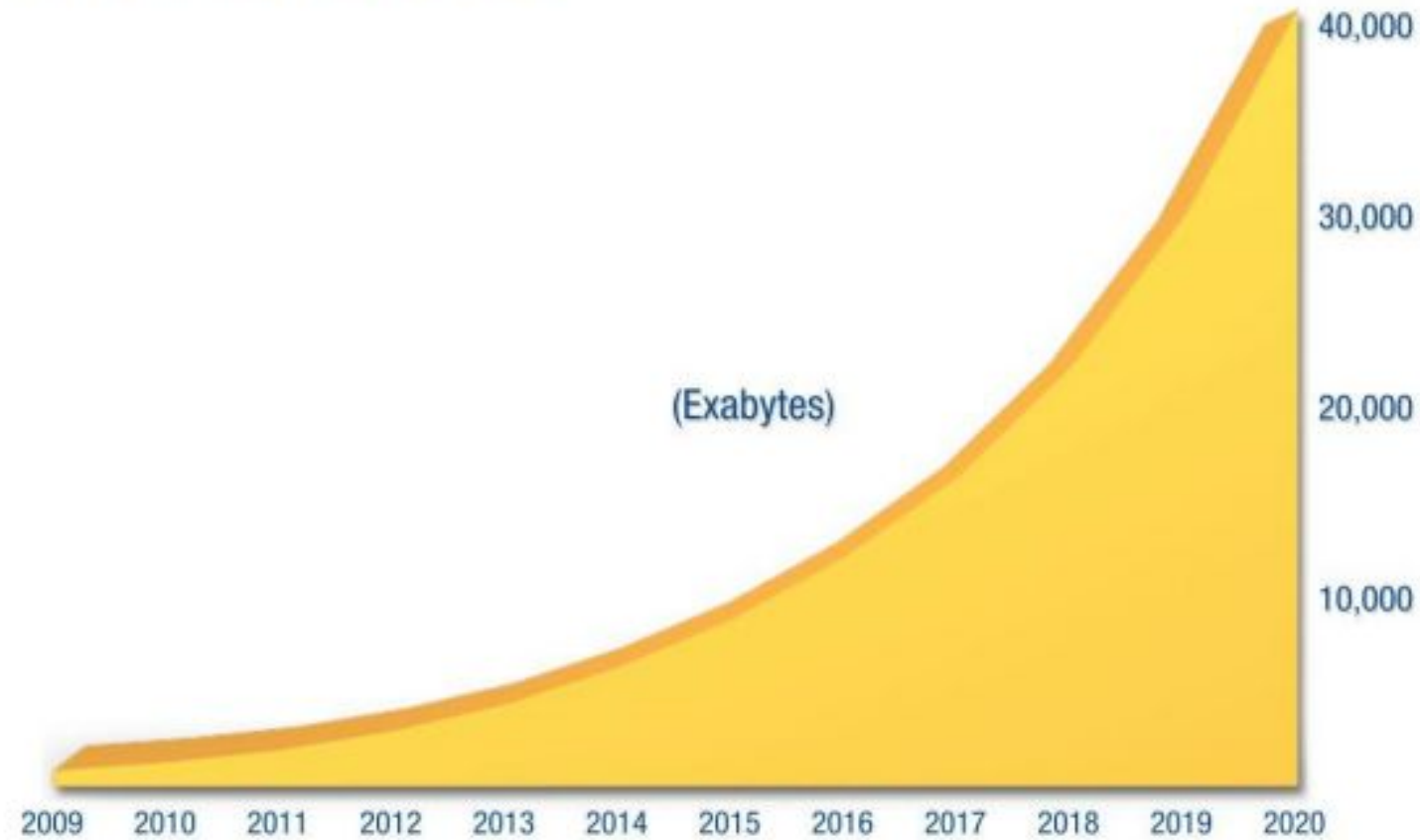
Fabián Torres Robles

Blanca Hilda Vázquez Gómez

Noviembre 2019

Panorama actual del análisis de datos

The Digital Universe: 50-fold Growth from the Beginning of 2010 to the End of 2020

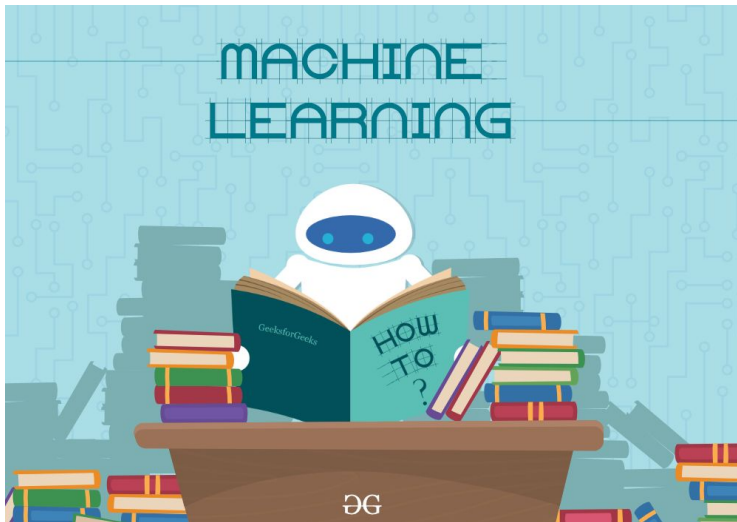


Source: IDC's Digital Universe Study, sponsored by EMC.

Definición del aprendizaje máquina

“Un programa aprende de la experiencia E con respecto a la tarea T y una medida de rendimiento P , si el rendimiento en T medido por P mejora con E ”

Tom M. Mitchell



Definición del aprendizaje máquina

*“Un programa aprende de la experiencia **E** con respecto a la tarea **T** y una medida de rendimiento **P**, si el rendimiento en **T** medido por **P** mejora con **E**”*



Ejemplo:

Supongamos que su programa de correo electrónico observa qué correos electrónicos usted marca como spam o no spam, y en base a eso aprende a cómo filtrar mejor el spam.

- ¿Cuál es la tarea **T** en este ejemplo?

Definición del aprendizaje máquina

*“Un programa aprende de la experiencia **E** con respecto a la tarea **T** y una medida de rendimiento **P**, si el rendimiento en **T** medido por **P** mejora con **E**”*



Ejemplo:

Supongamos que su programa de correo electrónico observa qué correos electrónicos usted marca como spam o no spam, y en base a eso aprende a cómo filtrar mejor el spam.

- ¿Cuál es la tarea **T** en este ejemplo?

T: es la tarea de clasificar el correo como spam o no spam

E: el programa observa tus etiquetas de spam o no spam

P: es el número de correos correctamente clasificados como spam / no spam

Panorama actual del análisis de datos

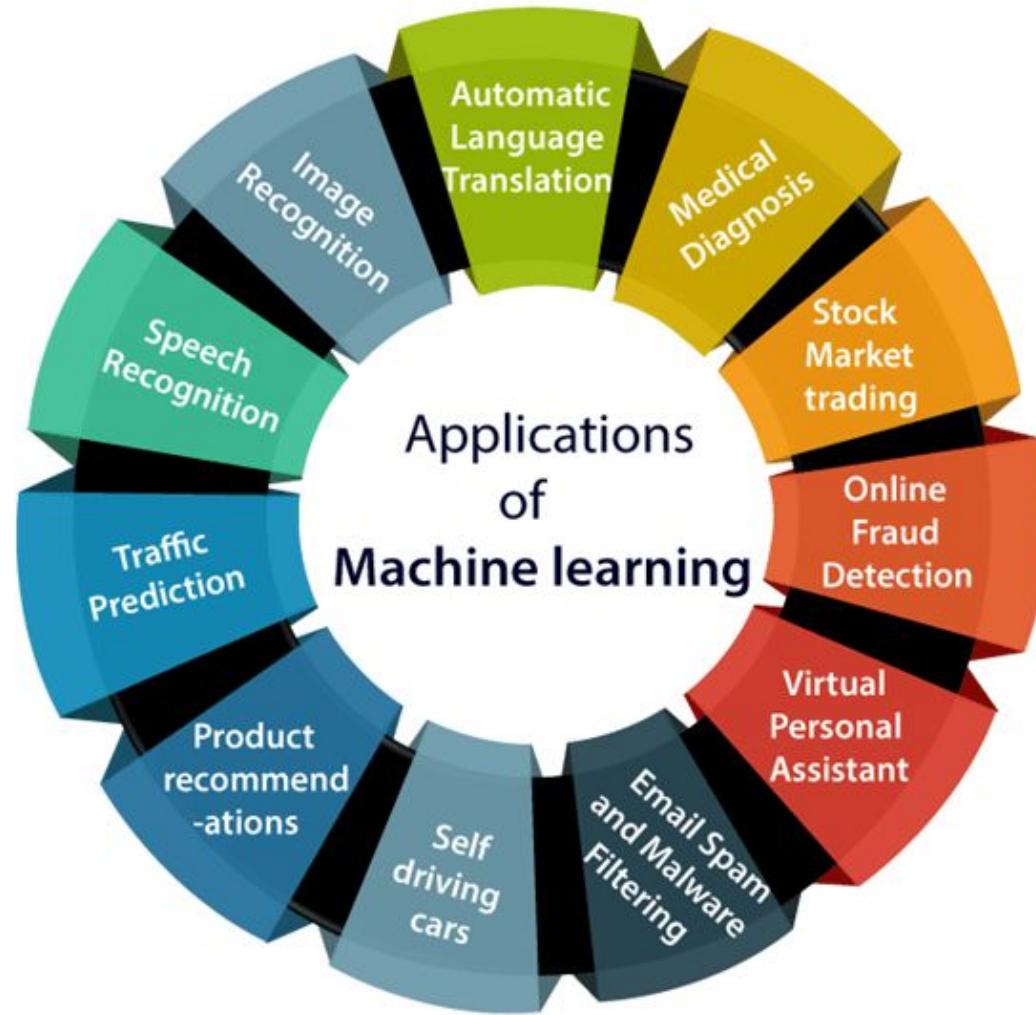
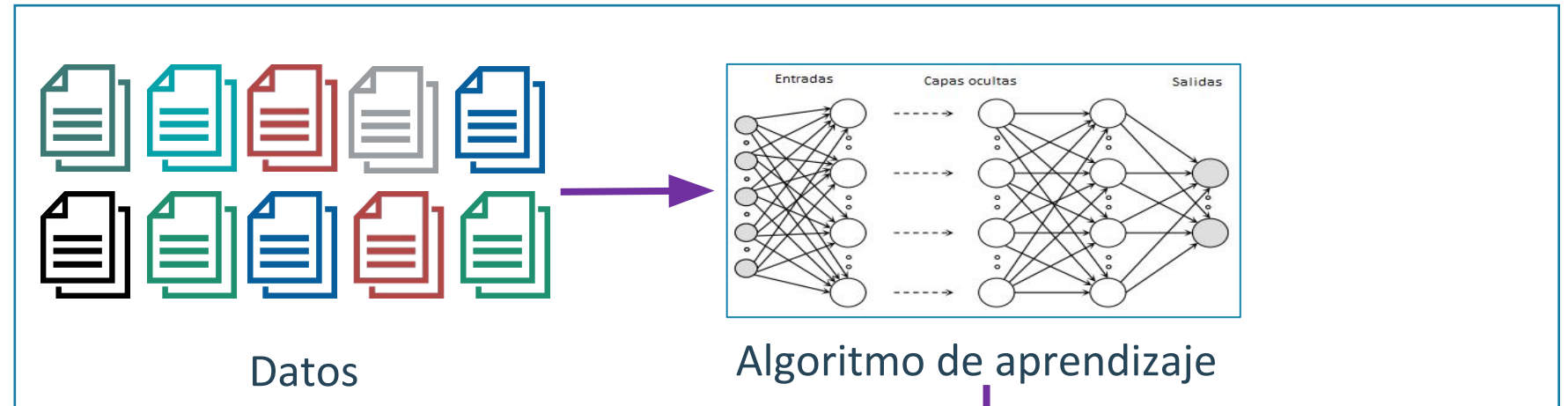


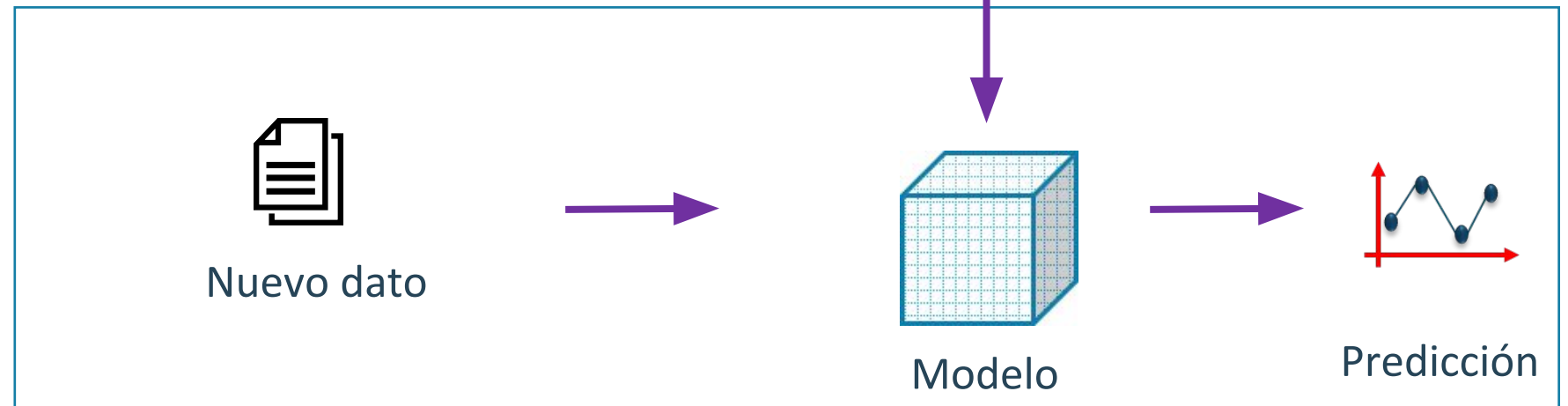
Imagen tomada de <https://data-flair.training/blogs/python-machine-learning-tutorial/>

¿Cómo funciona?

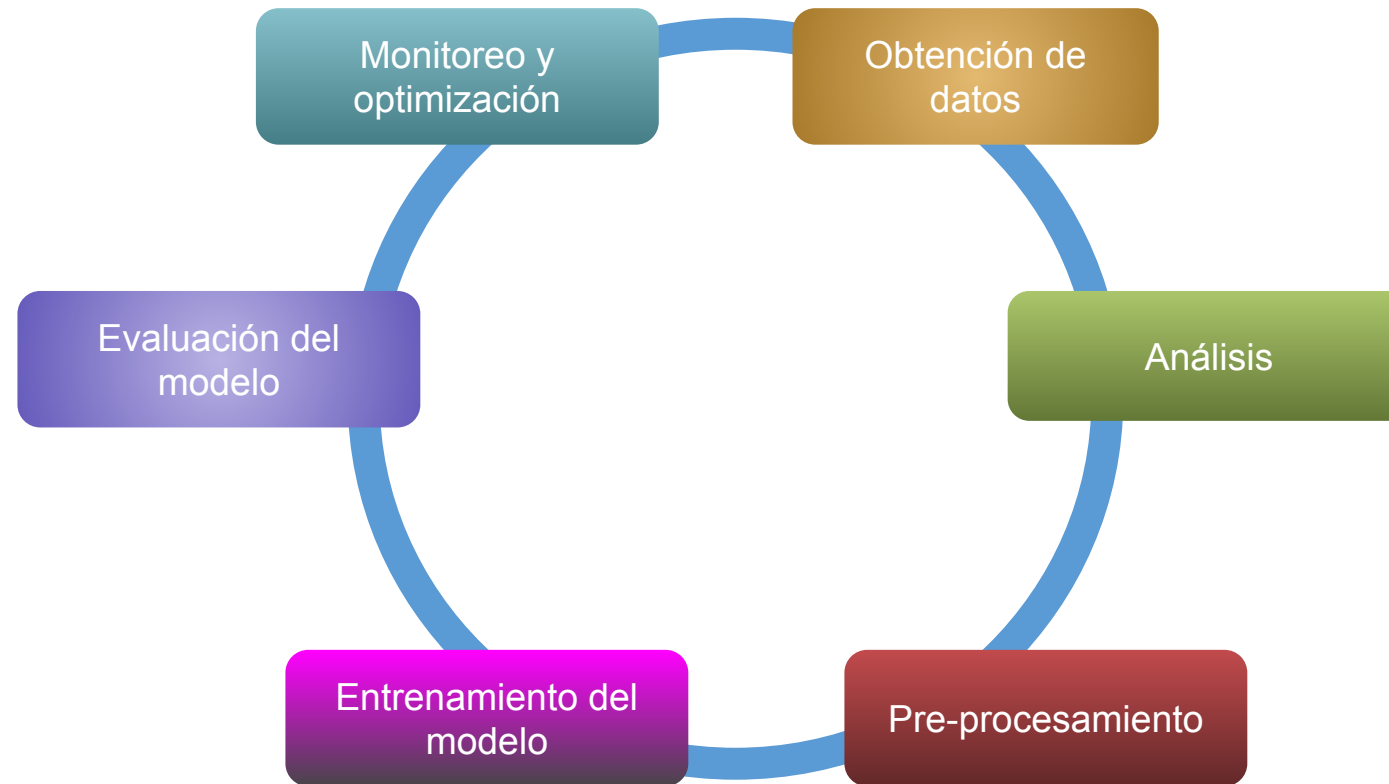
Entrenamiento



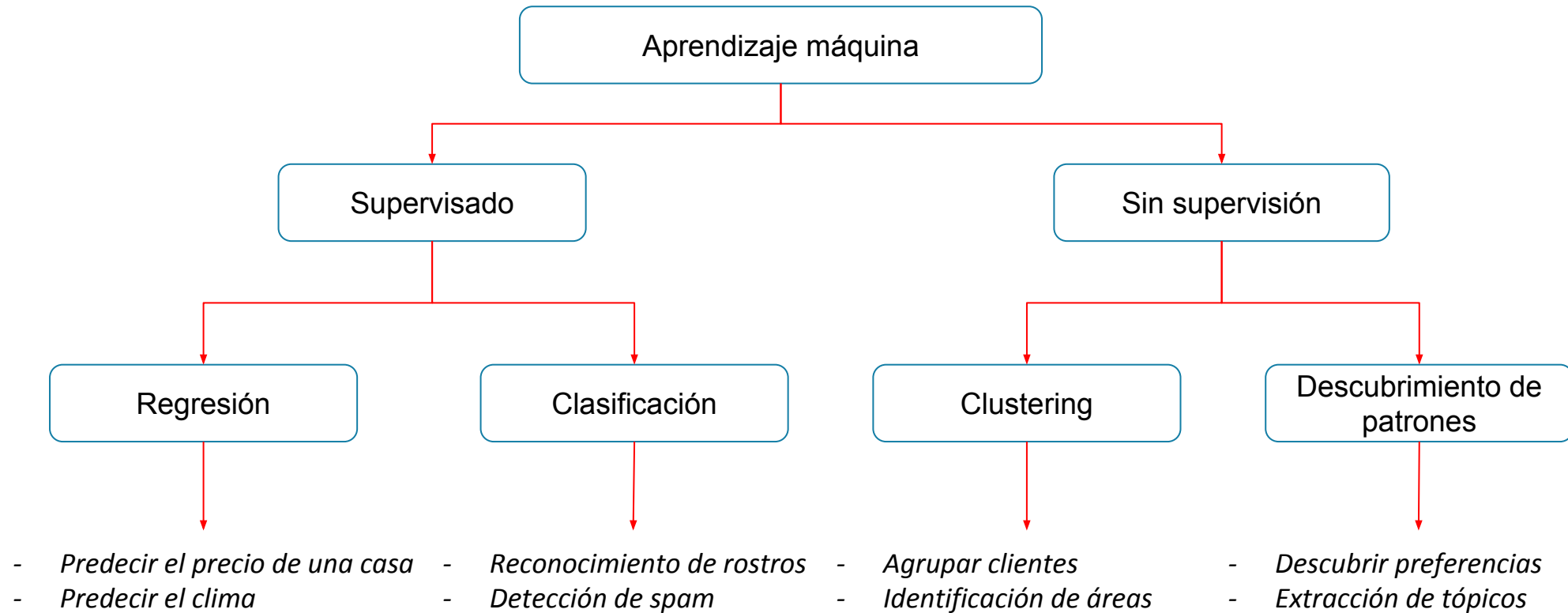
Inferencia



Ciclo de desarrollo de aplicaciones basadas en aprendizaje máquina



Tipos de algoritmos



Ejemplo de clasificación

Clasificar subespecies de la flor Iris basado en el ancho y largo de su pétalo

Ancho	Largo	Especie
1.4	0.2	Setosa
1.7	0.4	Setosa
1.5	0.1	Setosa
⋮	⋮	⋮
4.7	1.4	Versicolor
4.5	1.5	Versicolor
3.3	1.0	Versicolor
⋮	⋮	⋮



Iris setosa



Iris versicolor



Iris virginica



Características o atributos Respuesta o etiqueta
(x) (y)

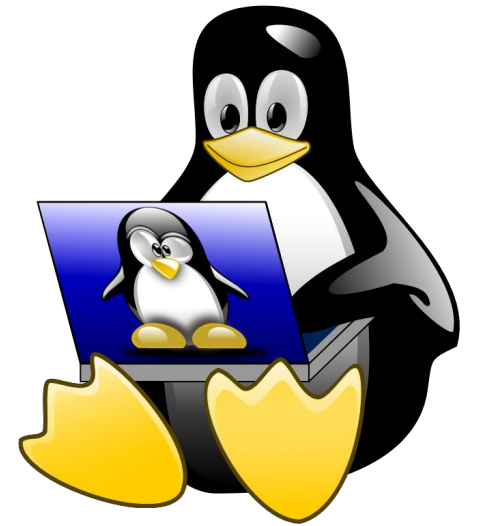
¿Por qué desarrollar modelos predictivos en el área clínica?



Imagen tomada de <https://southwesthealthcollaborative.org/workgroups/emergency-department-utilization/>

- Recursos limitados en áreas clínicas
- Decisiones críticas
- Cada minuto cuenta

It's time to code



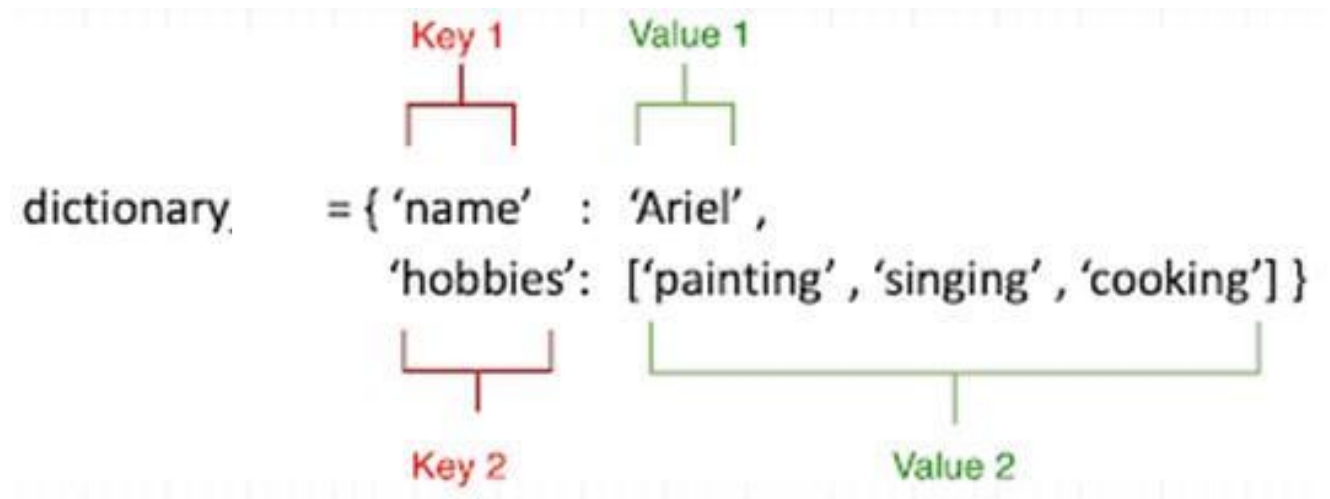
En esta URL puedes encontrar todo el material necesario para el taller:

https://github.com/blancavazquez/DAMA_EHR19

Nota: Antes del curso descarga los datos e instala las librerías indicadas

Estructuras de datos en python

- Diccionario

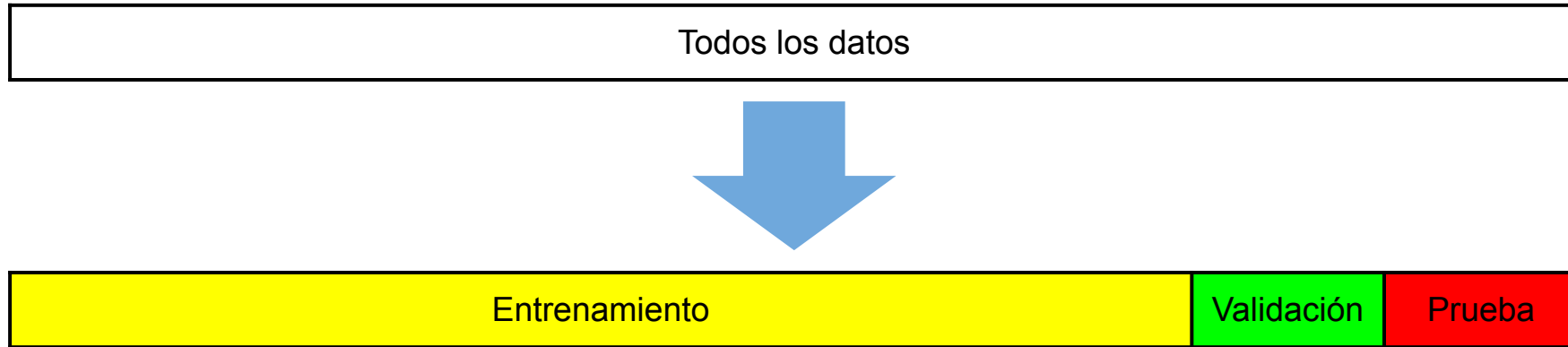


- Lista

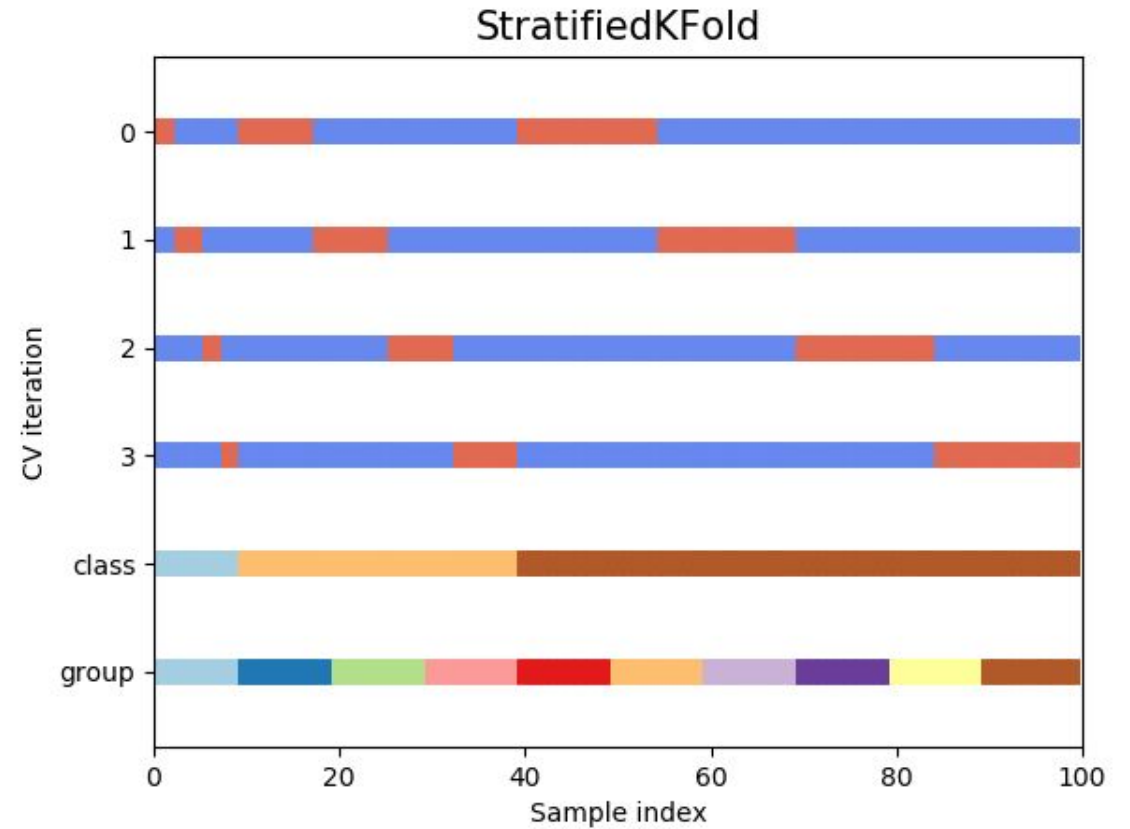
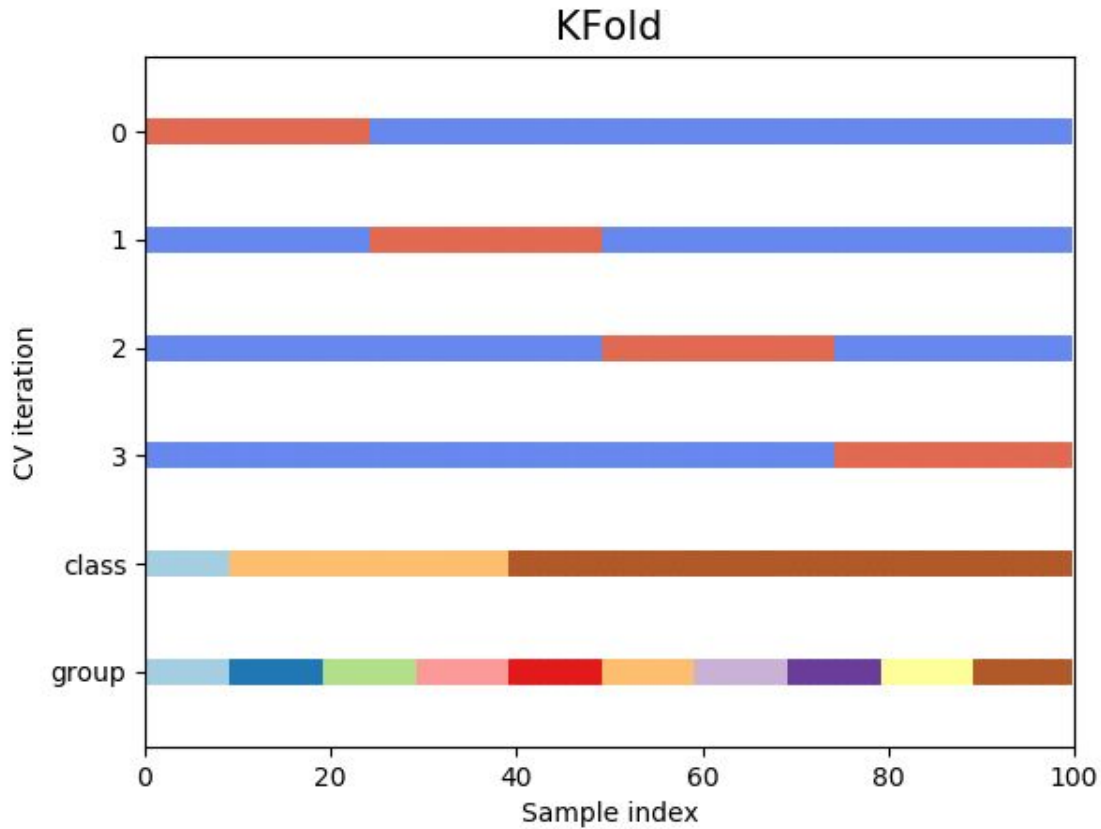
1, 2, 3, 4, 5, 6, 7, 8

Red, Green, Yellow

Conjuntos de datos



División del set de entrenamiento



Técnicas de muestreo de datos

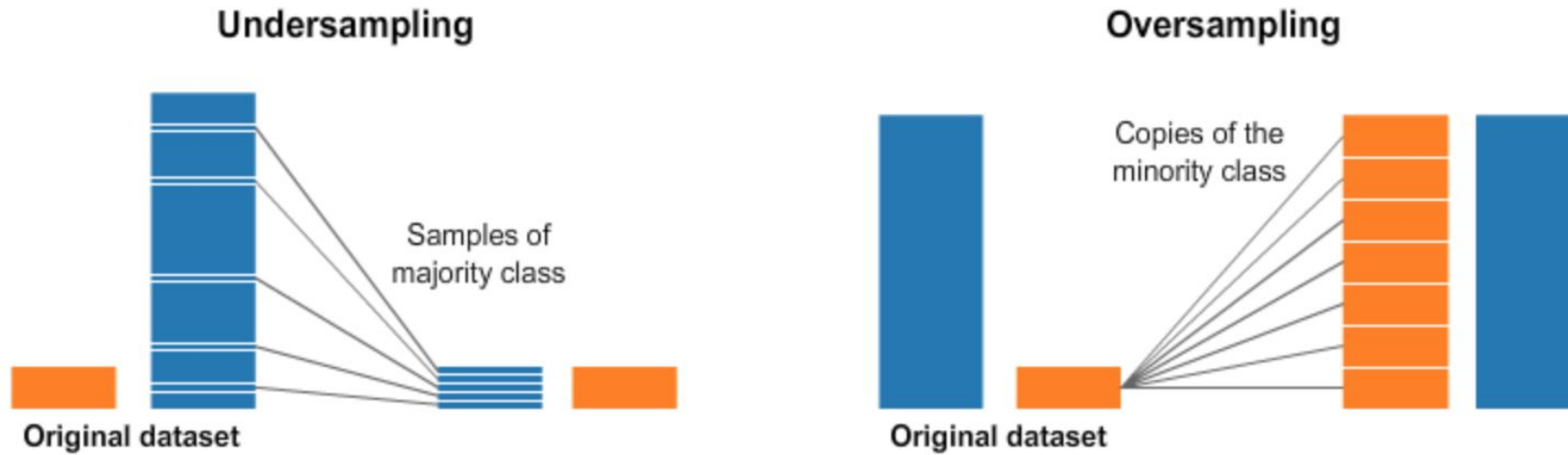


Imagen tomada de <https://towardsdatascience.com/having-an-imbalanced-dataset-here-is-how-you-can-solve-it-1640568947eb>

Material extra

- Métricas para medir el rendimiento del modelo
https://scikit-learn.org/stable/modules/model_evaluation.html#scoring-parameter
- Funciones para dividir los datos de entrenamiento:
- https://scikit-learn.org/stable/auto_examples/model_selection/plot_cv_indices.html#sphx-glr-auto-examples-model-selection-plot-cv-indices-py
- C
- C
- C