

TALLER 3

Machine learning

El término *machine learning* engloba al conjunto de algoritmos que permiten identificar patrones presentes en los datos y crear con ellos estructuras (modelos) que los representan. Una vez que los modelos han sido generados, se pueden emplear para predecir información sobre hechos o eventos que todavía no se han observado. Es importante recordar que, los sistemas de *machine learning*, solo son capaces de memorizar patrones que estén presentes en los datos con los que se entrenan, por lo tanto, solo pueden reconocer lo que han visto antes. Al emplear sistemas entrenados con datos pasados para predecir futuros se está asumiendo que, en el futuro, el comportamiento será el mismo, cosa que no siempre ocurre.

Aunque con frecuencia, términos como *machine learning*, *data mining*, inteligencia artificial, *data science*... son utilizados como sinónimos, es importante destacar que los métodos de *machine learning* son solo una parte de las muchas estrategias que se necesita combinar para extraer información, entender y dar valor a los datos. El siguiente documento pretende ser un ejemplo del tipo de problema al que se suele enfrentar un analista: partiendo de un conjunto de datos más o menos procesado (la preparación de los datos es una etapa crítica que precede al *machine learning*), se desea crear un modelo que permita predecir con éxito el comportamiento o valor que toman nuevas observaciones.

Etapas de un problema de *machine learning*

El siguiente es un listado de las etapas que suelen formar parte de la mayoría de problemas de *machine learning*.

- **Definir el problema:** ¿Qué se pretende predecir? ¿De qué datos se dispone? o ¿Qué datos es necesario conseguir?
- Explorar y entender los datos que se van a emplear para crear el modelo.
- **Métrica de éxito:** definir una forma apropiada de cuantificar cómo de buenos son los resultados obtenidos.
- **Preparar la estrategia para evaluar el modelo:** separar las observaciones en un conjunto de entrenamiento, un conjunto de validación (este último suele ser un subconjunto del de entrenamiento) y un conjunto de test. Ninguna información del conjunto de test debe participar en el proceso de entrenamiento del modelo.
- **Preprocesar los datos:** aplicar las transformaciones necesarias para que los datos puedan ser interpretados por el algoritmo de *machine learning* seleccionado.
- **Ajustar un primer modelo capaz de superar unos resultados mínimos.** Por ejemplo, en problemas de clasificación, el mínimo a superar es el porcentaje de la clase mayoritaria (la moda).
- Gradualmente, mejorar el modelo optimizando sus hiperparámetros.
- Evaluar la capacidad del modelo final con el conjunto de test para tener una estimación de la capacidad que tiene el modelo cuando predice nuevas observaciones.

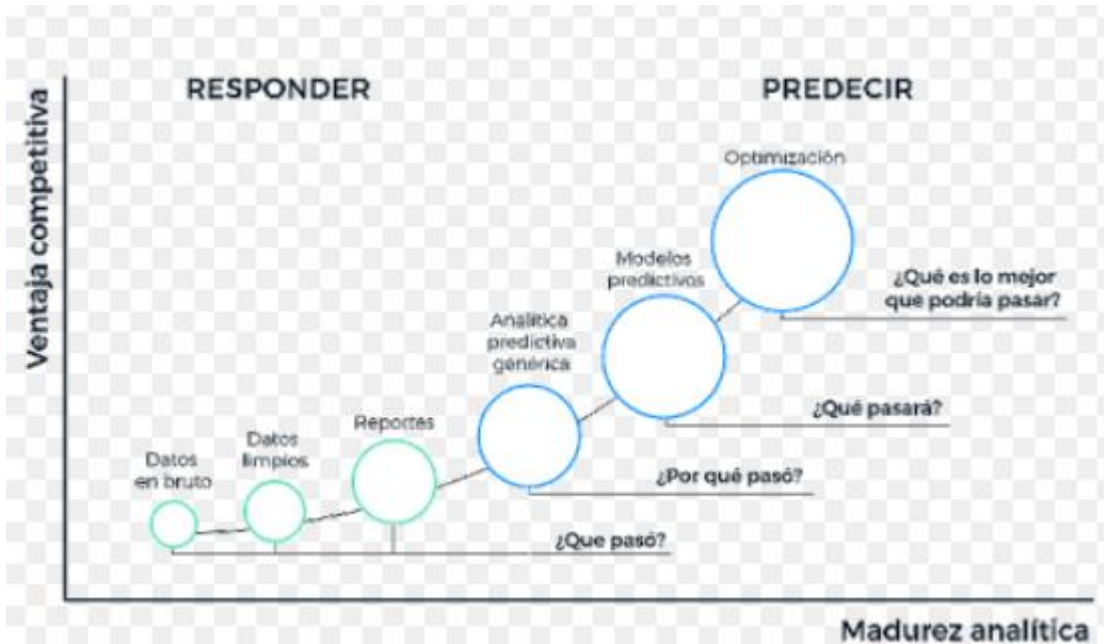
1. MODELOS MACHINE LEARNING

- Predicción
- Clasificación
- clustering

2. MODELOS

- Regresión lineal, múltiple, no lineal.
- Árboles de decisión KNN, logística.
- k medias o k means, jerárquico.

Modelo data Science



$$\text{Medida} = a \cdot \text{días} + b$$

a= pendiente o "slope"

b= intercepto

0,1

Practica 1

Utilizar el modelo lineal para relacionar dos variables, obteniendo su ecuación, el grado de ajuste de la misma con los datos y su representación gráfica.

1. Los pesos (en kg) y estaturas (en cm) de una muestra de 10 estudiantes universitarios son:

Peso	82	75	70	68	44	63	80	70	54	54
Altura	185	185	180	178	159	170	190	172	162	165

2. Se realiza un estudio para establecer una ecuación mediante la cual se pueda utilizar la concentración de estrona en saliva, para predecir la concentración del esteroide en plasma libre. Se extrajeron, en pg/ml, los siguientes datos de 14 varones sanos:

Saliva	7.4	7.5	8.5	9.0	9.0	11.0	13.0	14.0	14.5	16.0	17.0
	18.0	20.0	23.0								
Plasma	30.0	25.0	31.5	27.5	39.5	38.0	43.0	49.0	55.0	48.5	51.0
	64.5	63.0	68.0								

3. Hallar el coeficiente de correlación
4. hacer la predicción con el modelo cars que está incluido en r estudio
5. Elaborar el esquema de modelo de datos para el proyecto del semillero de investigación