

09481: Inteligencia Artificial

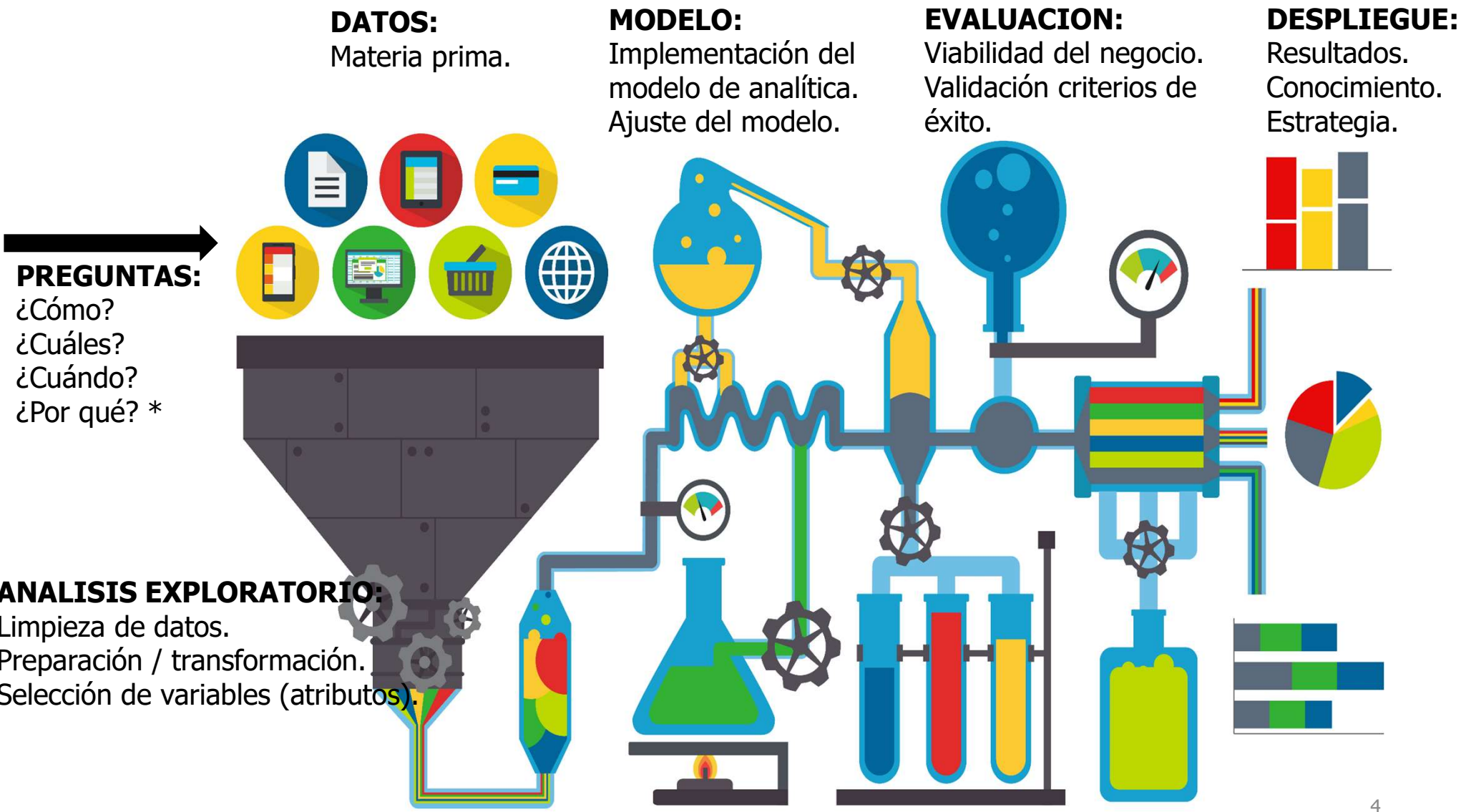
Profesor del curso: Breyner Posso, Ing. M.Sc.
e-mail: breyner.posso1@u.icesi.edu.co

Programa de Ingeniería de Sistemas.
Departamento TIC.
Facultad de Ingeniería.
Universidad Icesi.
Cali, Colombia.

Agenda

- Introducción
- Aprendizaje supervisado
- Underfitting y Overfitting.

Introducción



DATOS:

Materia prima.

MODELO:

Implementación del
modelo de analítica.
Ajuste del modelo.

EVALUACION:

Viabilidad del negocio.
Validación criterios de
éxito.

DESPLIEGUE:

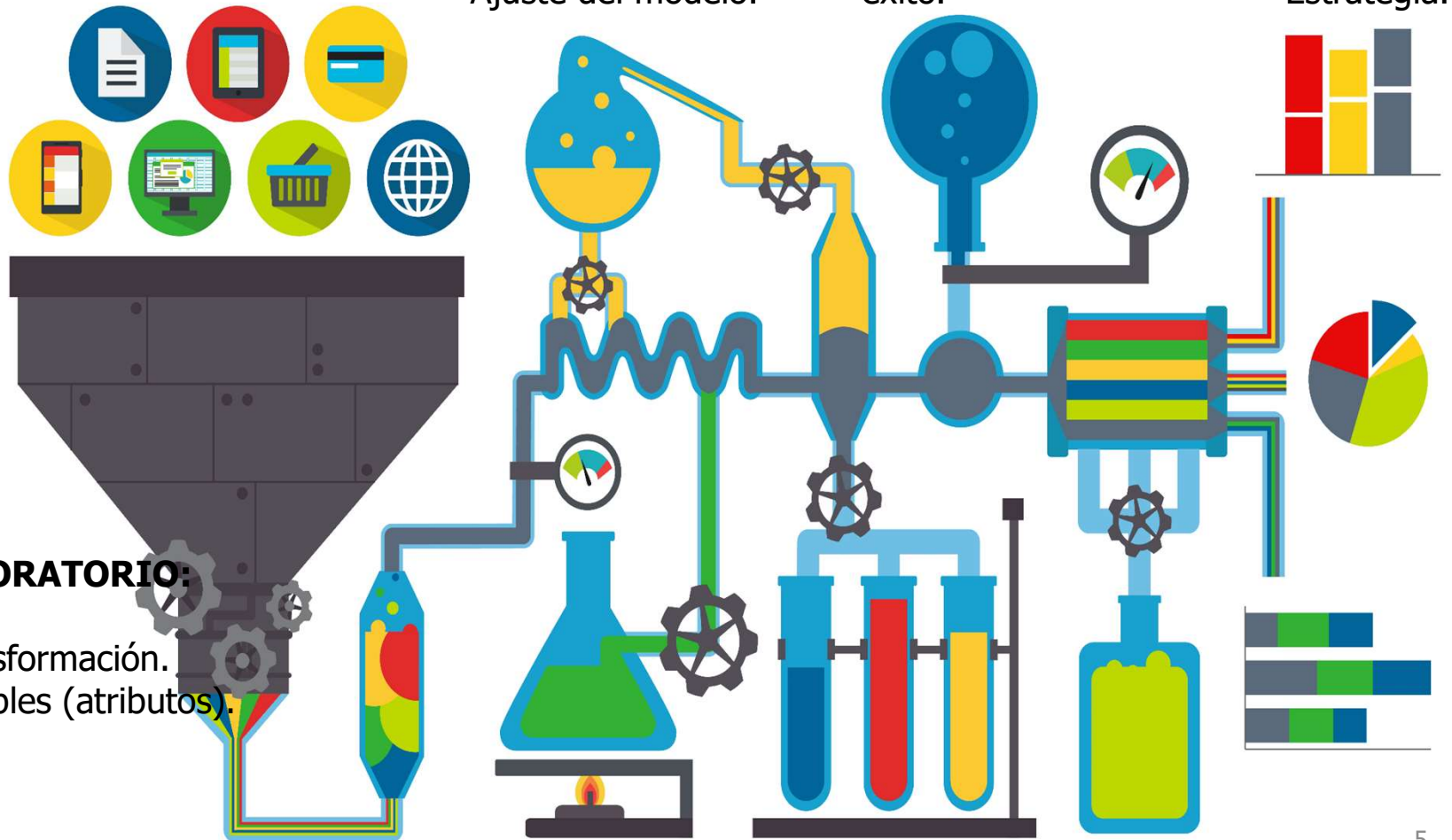
Resultados.
Conocimiento.
Estrategia.

PREGUNTAS:

¿Cómo?
¿Cuáles?
¿Cuándo?
¿Por qué? *

ANALISIS EXPLORATORIO:

Limpieza de datos.
Preparación / transformación.
Selección de variables (atributos).



PREGUNTAS:

¿Cómo?
¿Cuáles?
¿Cuándo?
¿Por qué? *

DATOS:

Materia prima.

MODELO:

Implementación del
modelo de analítica.
Ajuste del modelo.

EVALUACION:

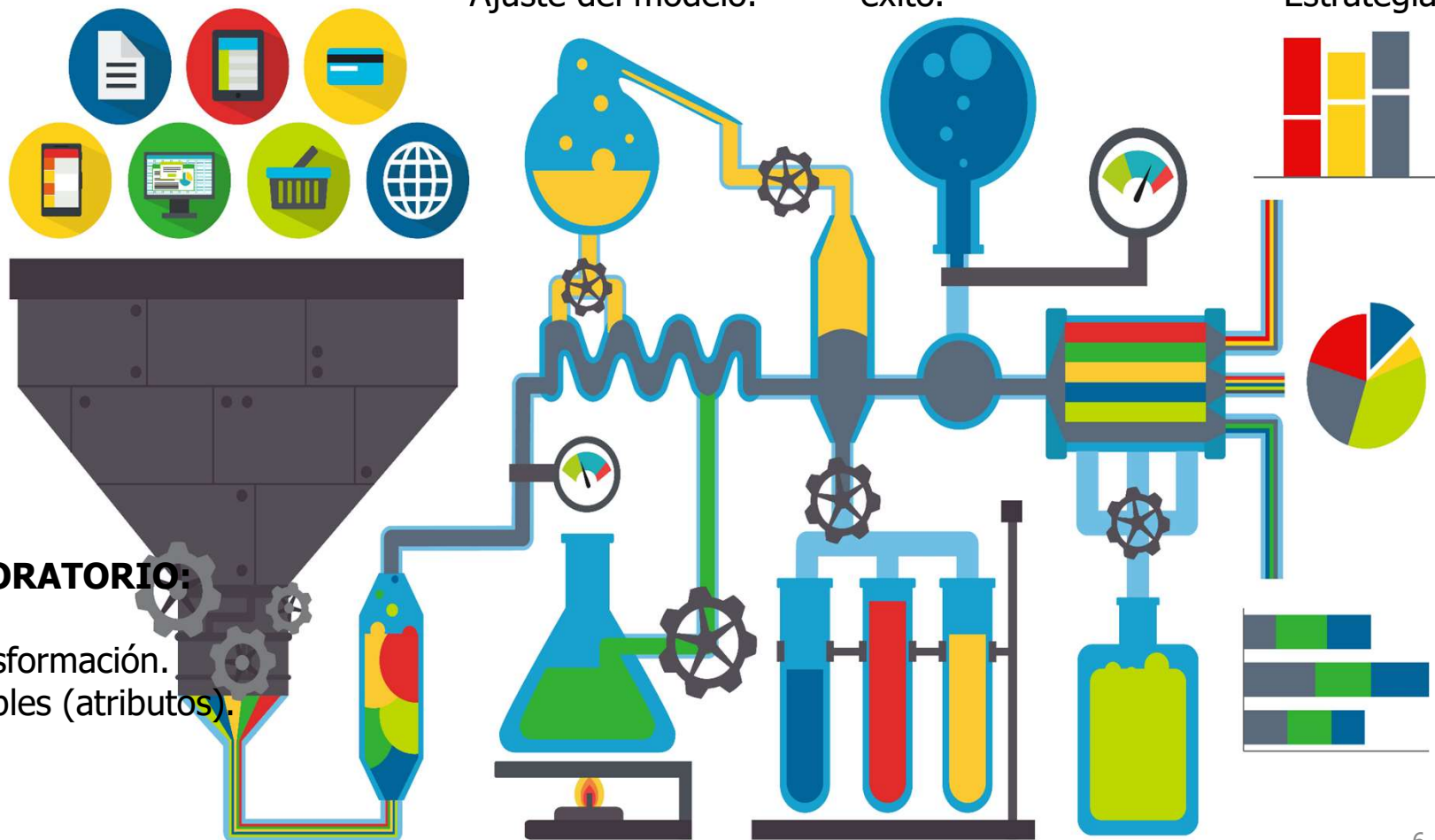
Viabilidad del negocio.
Validación criterios de
éxito.

DESPLIEGUE:

Resultados.
Conocimiento.
Estrategia.

ANALISIS EXPLORATORIO:

Limpieza de datos.
Preparación / transformación.
Selección de variables (atributos).



DATOS:

Materia prima.

PREGUNTAS:

¿Cómo?
¿Cuáles?
¿Cuándo?
¿Por qué? *

ANÁLISIS EXPLORATORIO:

Limpieza de datos.
Preparación / transformación.
Selección de variables (atributos).

MODELO:

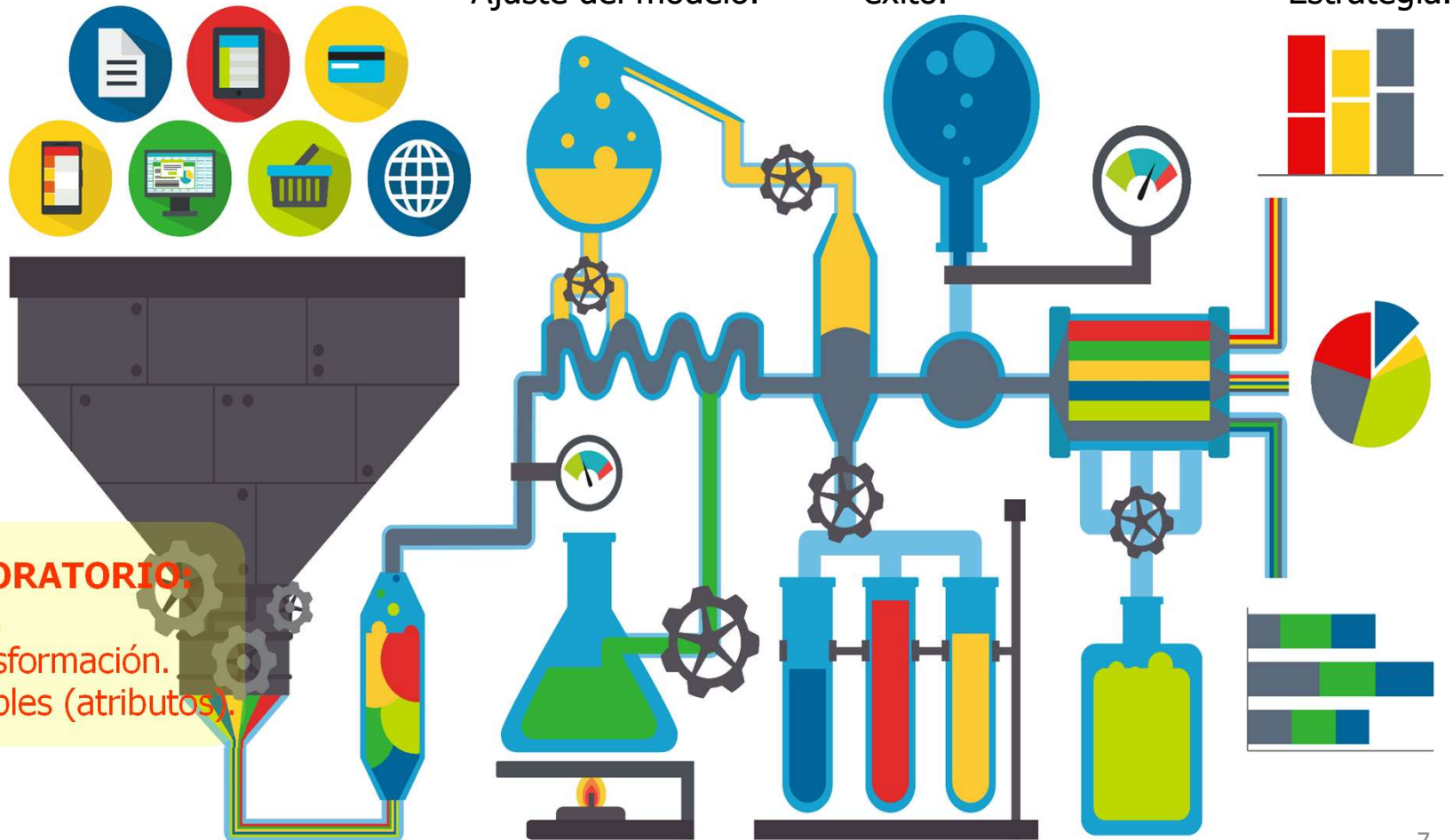
Implementación del
modelo de analítica.
Ajuste del modelo.

EVALUACION:

Viabilidad del negocio.
Validación criterios de
éxito.

DESPLIEGUE:

Resultados.
Conocimiento.
Estrategia.



DATOS:

Materia prima.

MODELO:

Implementación del
modelo de analítica.
Ajuste del modelo.

EVALUACION:

Viabilidad del negocio.
Validación criterios de
éxito.

DESPLIEGUE:

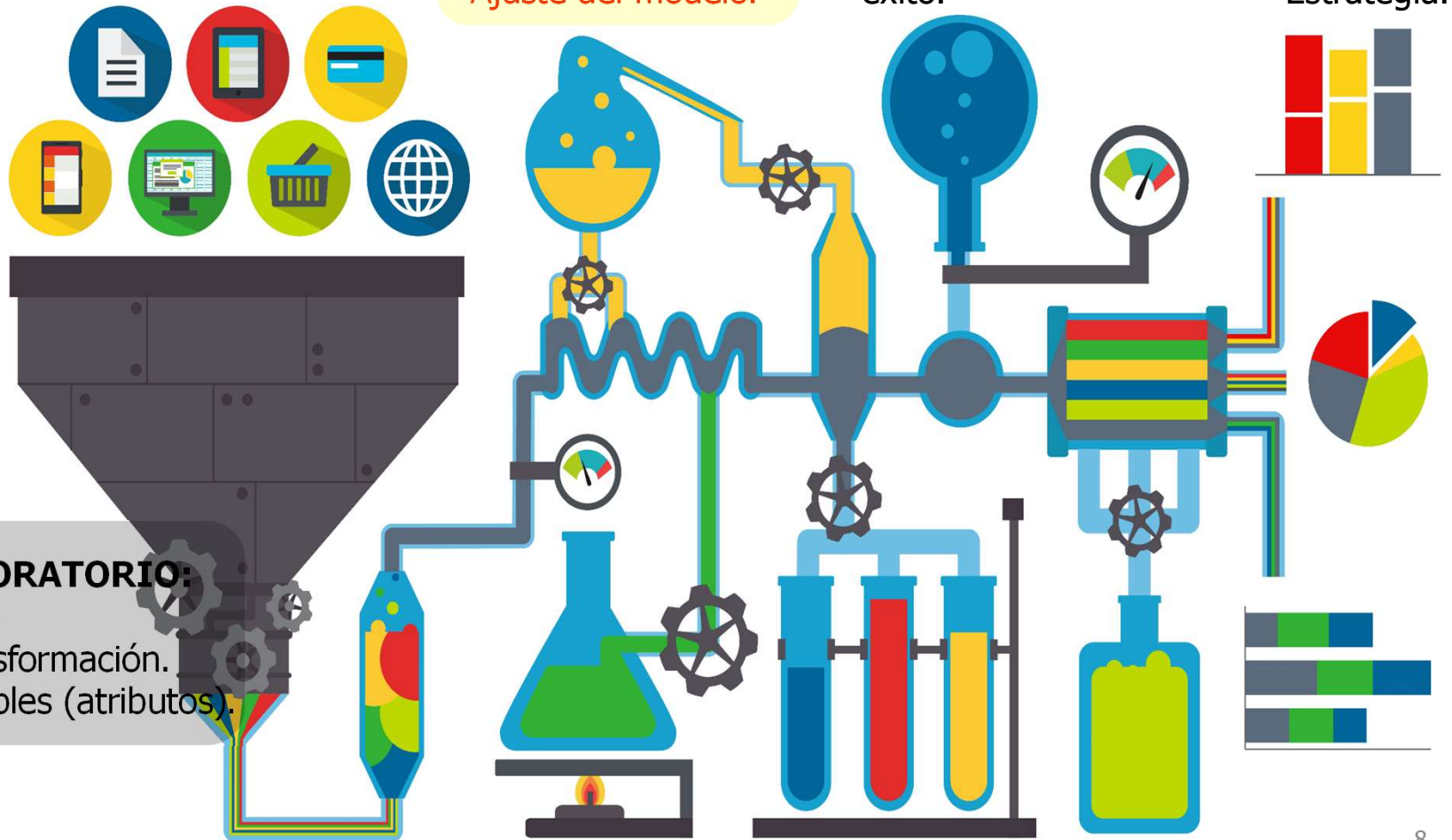
Resultados.
Conocimiento.
Estrategia.

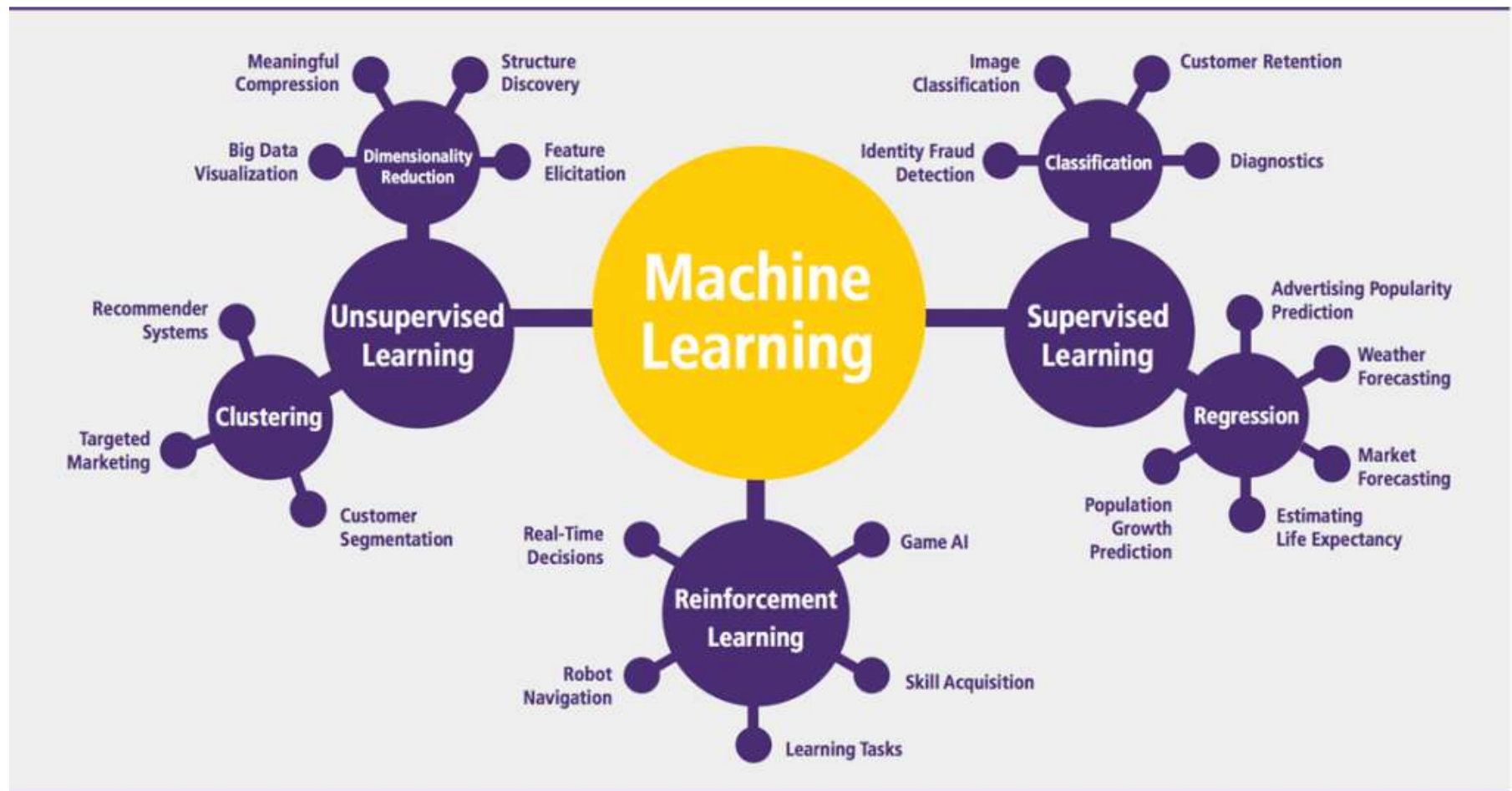
PREGUNTAS:

¿Cómo?
¿Cuáles?
¿Cuándo?
¿Por qué? *

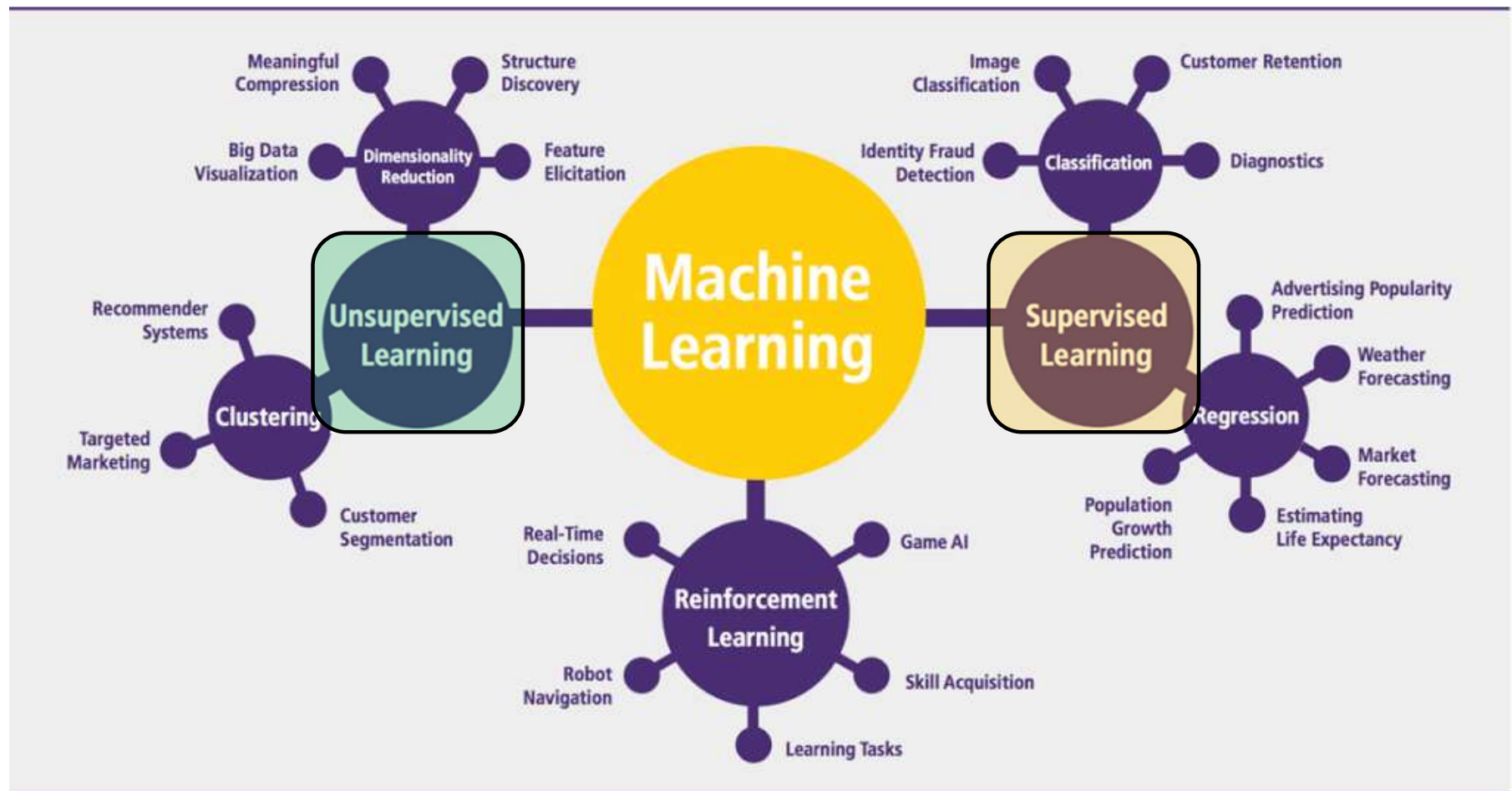
ANALISIS EXPLORATORIO:

Limpieza de datos.
Preparación / transformación.
Selección de variables (atributos).



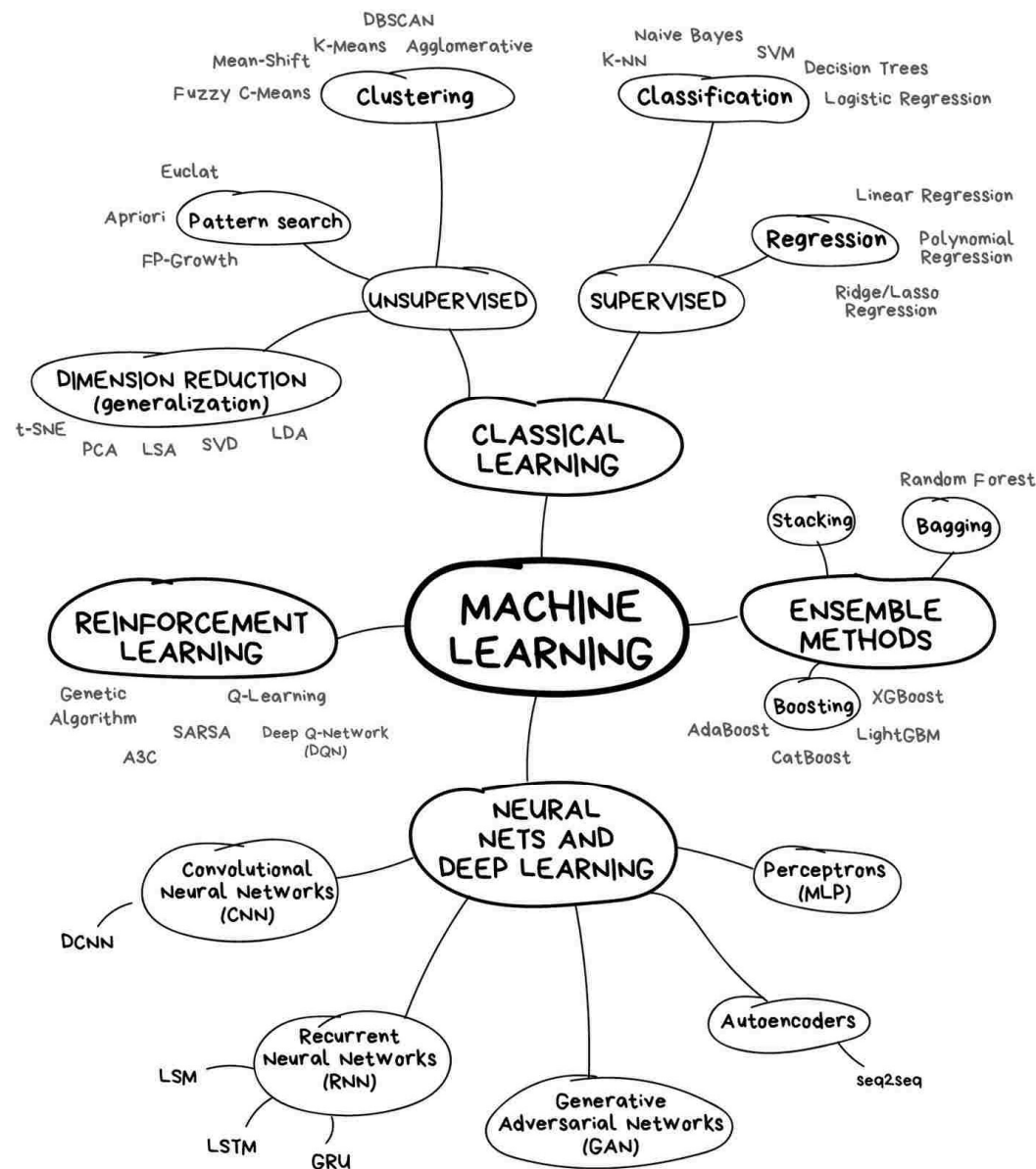


Desde el punto de vista de las aplicaciones

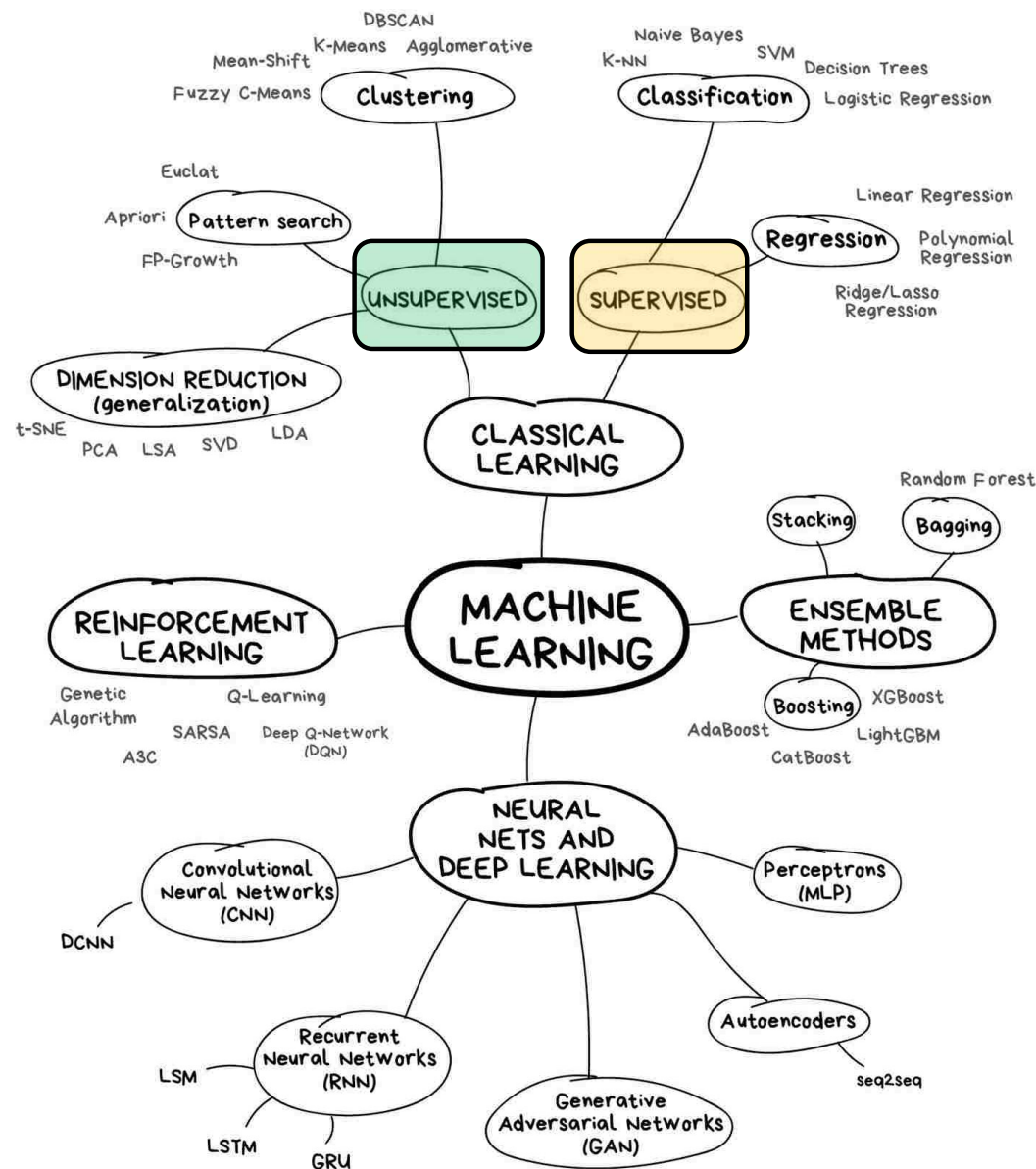


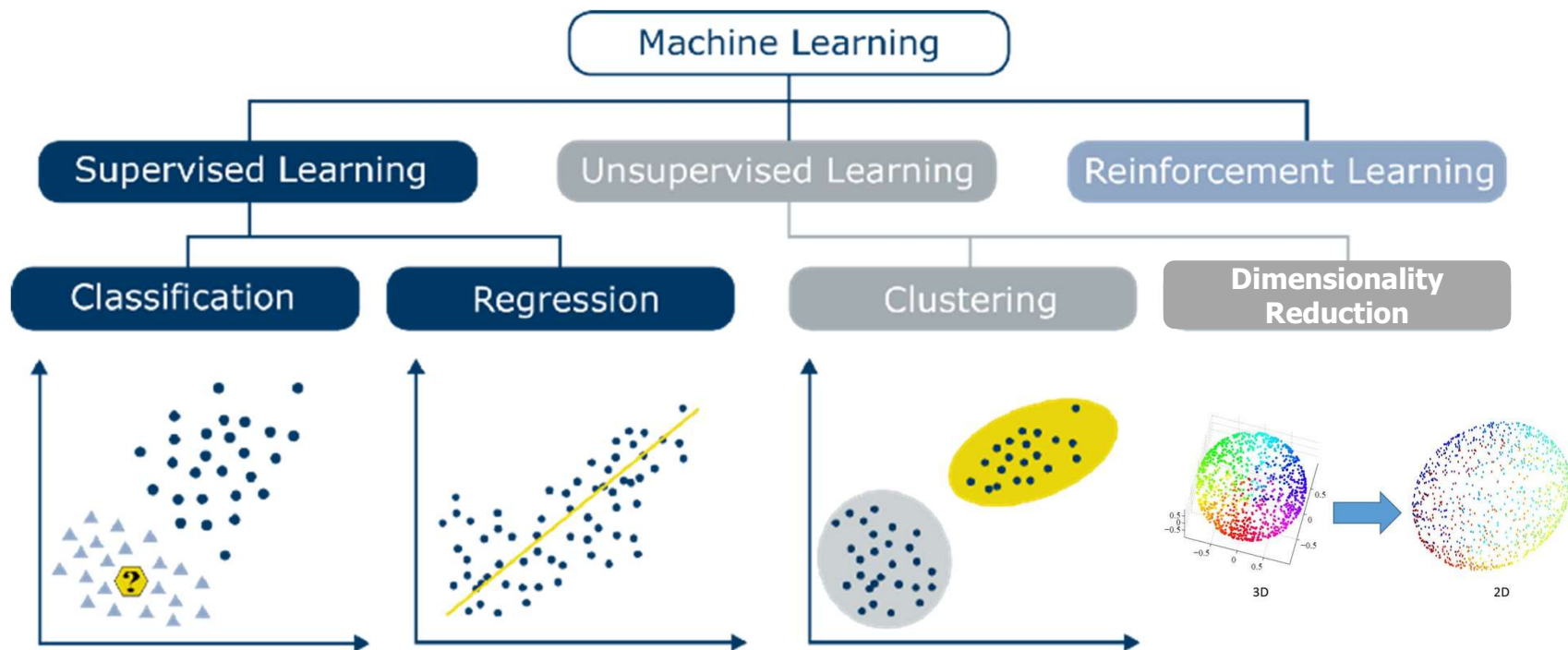
Desde el punto de vista de las aplicaciones

Desde el punto de los algoritmos



Desde el punto de los algoritmos





DATOS:

Materia prima.

MODELO:

Implementación del
modelo de analítica.
Ajuste del modelo.

EVALUACION:

Viabilidad del negocio.
Validación criterios de
éxito.

DESPLIEGUE:

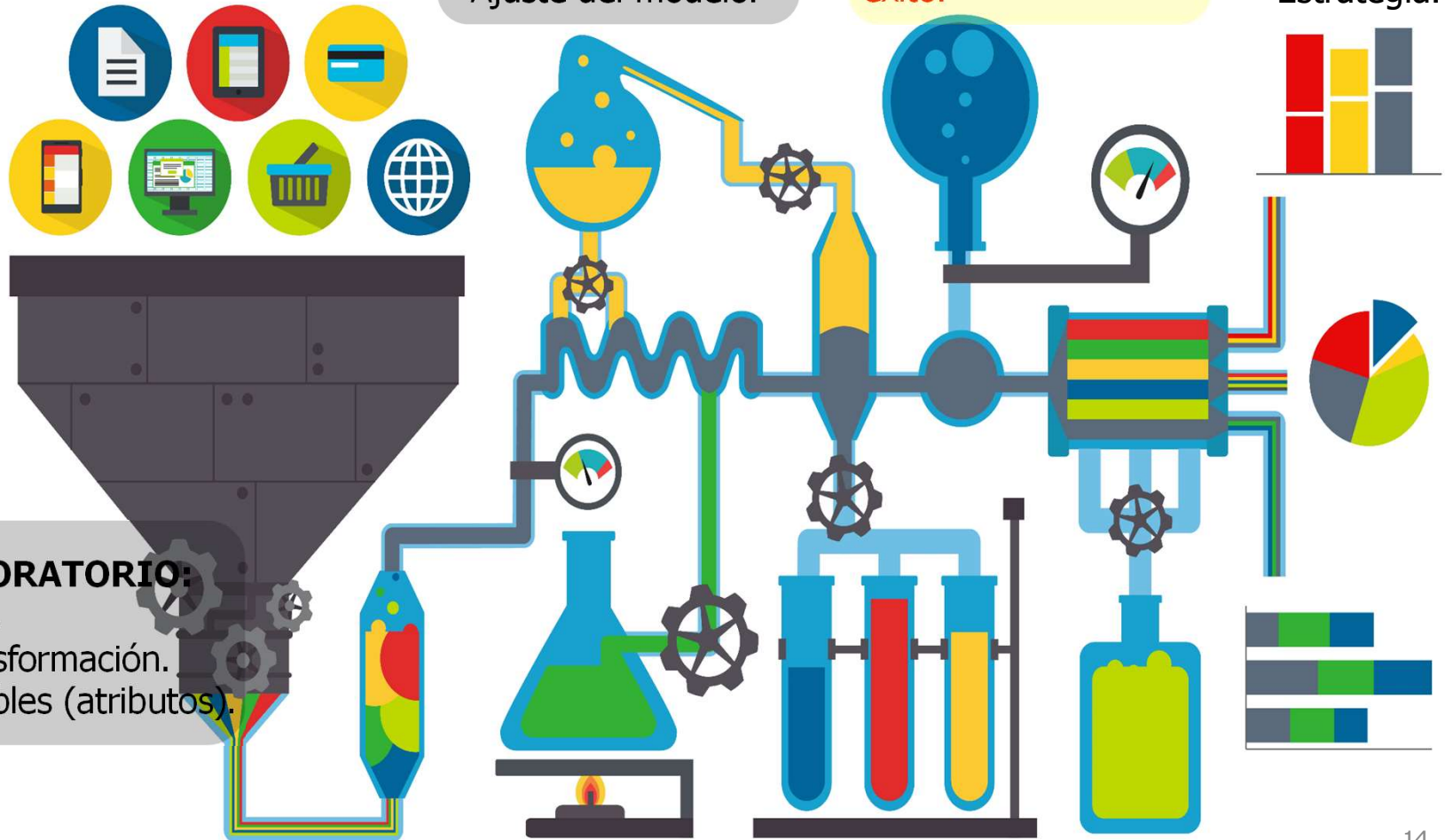
Resultados.
Conocimiento.
Estrategia.

PREGUNTAS:

¿Cómo?
¿Cuáles?
¿Cuándo?
¿Por qué? *

ANALISIS EXPLORATORIO:

Limpieza de datos.
Preparación / transformación.
Selección de variables (atributos).



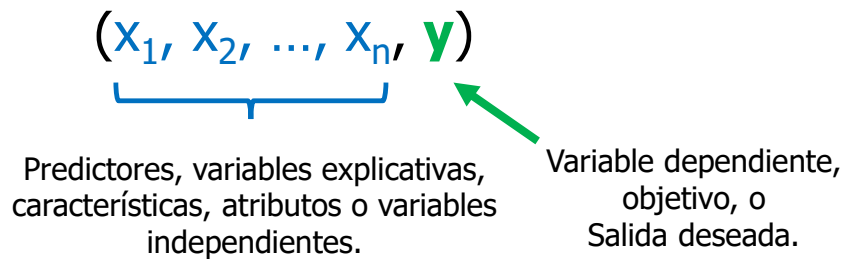
Métricas de evaluación

- Se requiere evaluar la calidad de los modelos de aprendizaje automático.
- Existen diferentes aspectos a tener en cuenta:
 - Desempeño de la predicción.
 - Simplicidad.
 - Interpretabilidad.
 - Tiempo de aprendizaje o de predicción.
 - Escalabilidad (esto es importante en problemas de *big data*).

Aprendizaje supervisado

Aprendizaje supervisado

- Conjunto de datos **etiquetados** con una clase o valor:



- **Objetivo:** predecir un valor (*Regresión*) o predecir una clase (*Clasificación*).

Aprendizaje supervisado

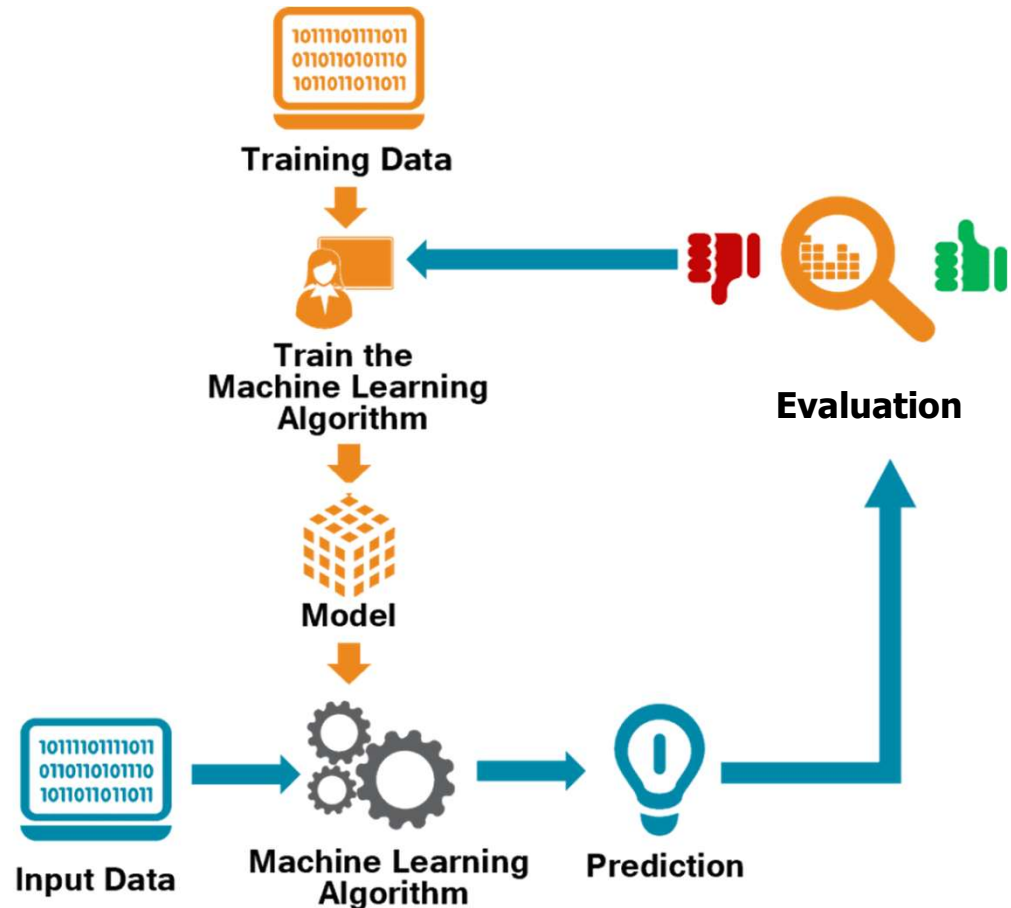
- Conjunto de datos **etiquetados** con una clase o valor:

$(x_1, x_2, \dots, x_n, y)$

Predictores, variables explicativas, características, atributos o variables independientes.

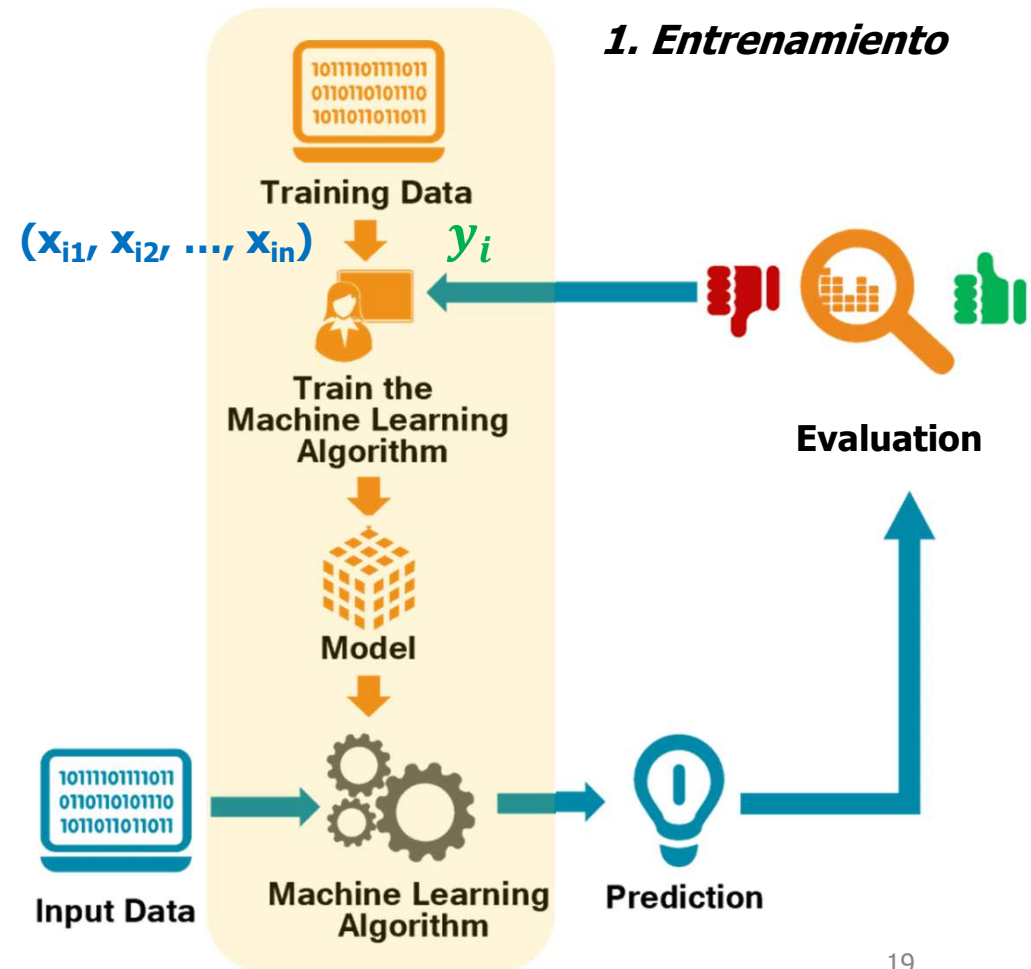
Variable dependiente, objetivo, o Salida deseada.

- Objetivo:** predecir un valor (*Regresión*) o predecir una clase (*Clasificación*).



Aprendizaje supervisado

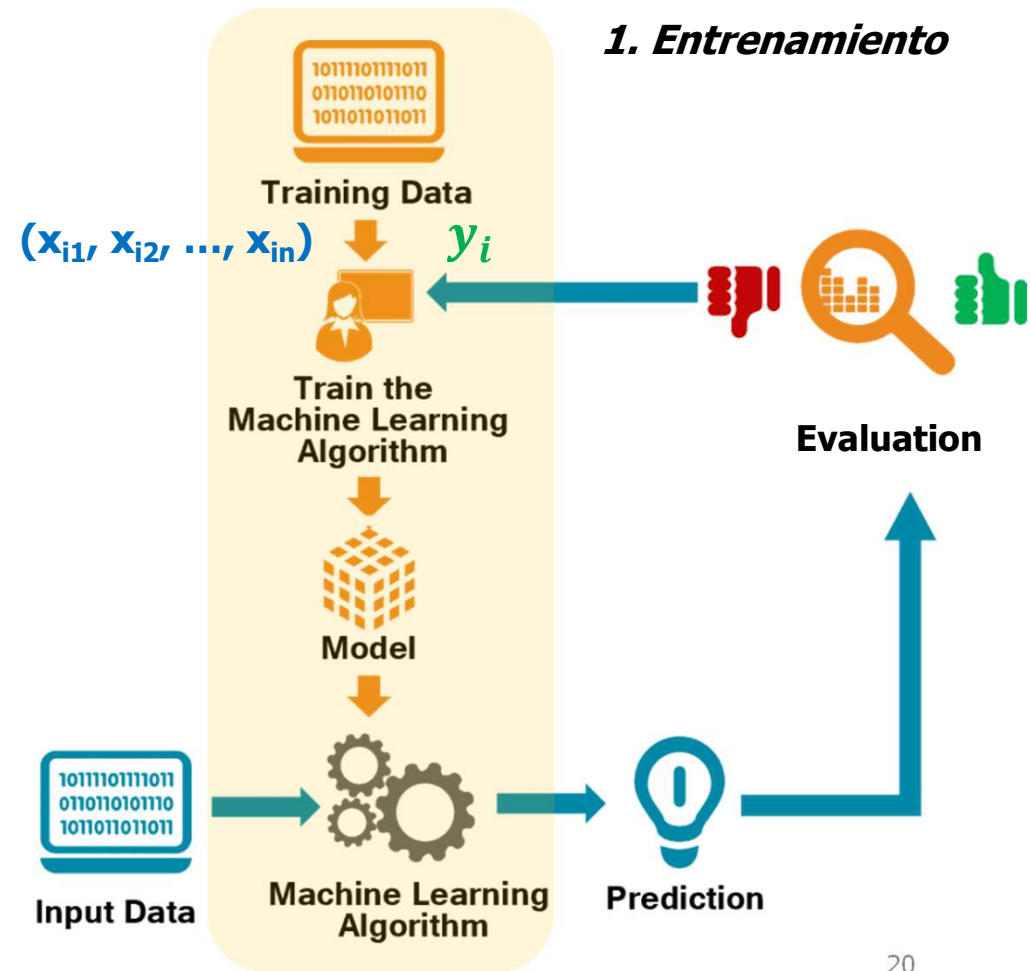
Fase 1: Entrenamiento



Aprendizaje supervisado

Fase 1: Entrenamiento

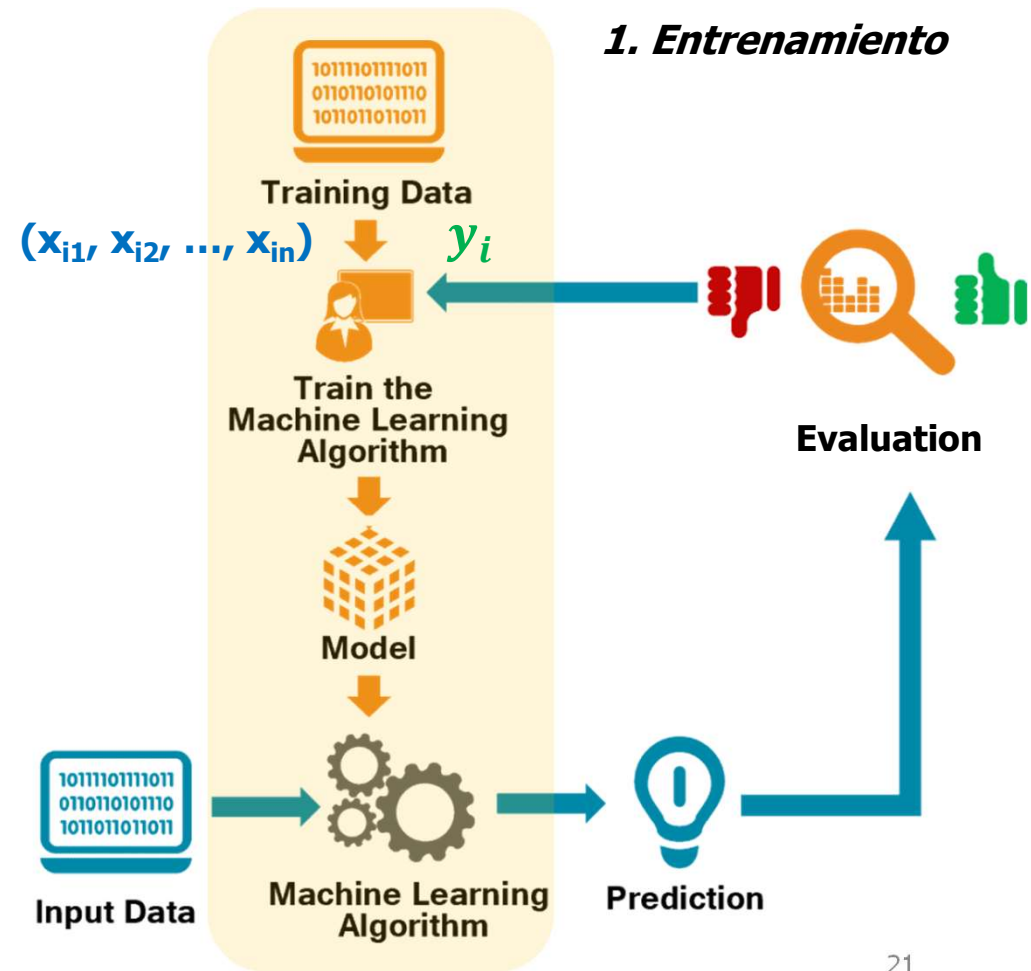
a. ¿Clasificación o regresión?



Aprendizaje supervisado

Fase 1: Entrenamiento

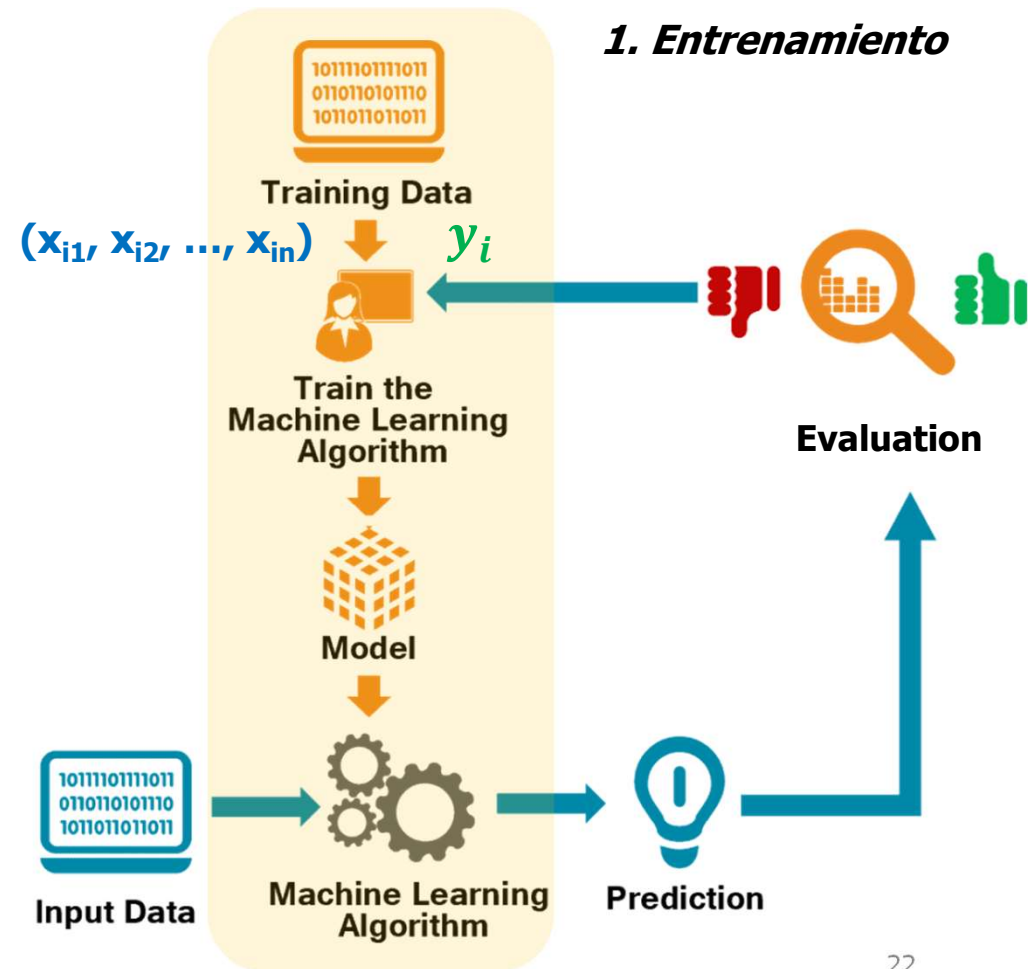
- a. ¿Clasificación o regresión?
- b. ¿Qué atributos voy a utilizar?



Aprendizaje supervisado

Fase 1: Entrenamiento

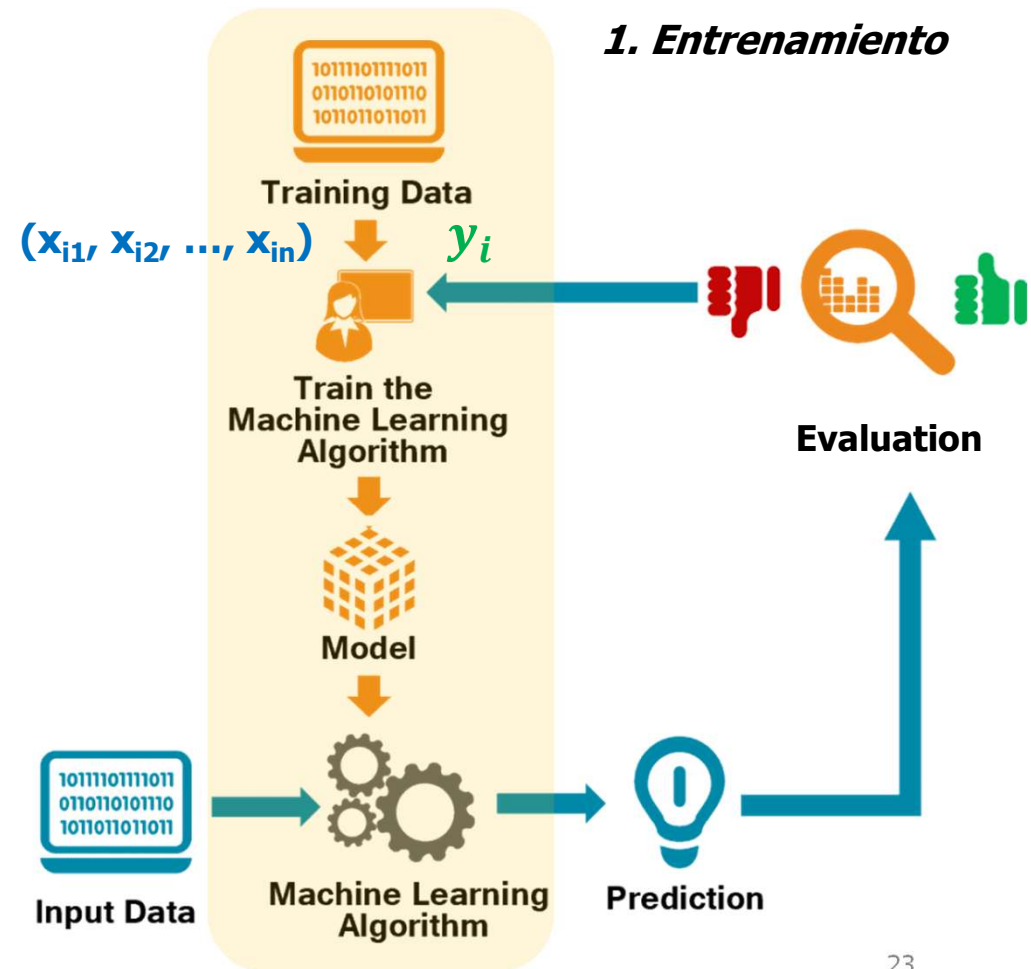
- a. ¿Clasificación o regresión?
- b. ¿Qué atributos voy a utilizar?
- c. ¿Cómo particionar los datos?
 - ¿Cuántos para entrenamiento?
 - ¿Cuántos para validación?
 - ¿Cuántos para prueba?
 - ¿Se escogen aleatoriamente?
 - ¿El dataset es desbalanceado?
 - ...



Aprendizaje supervisado

Fase 1: Entrenamiento

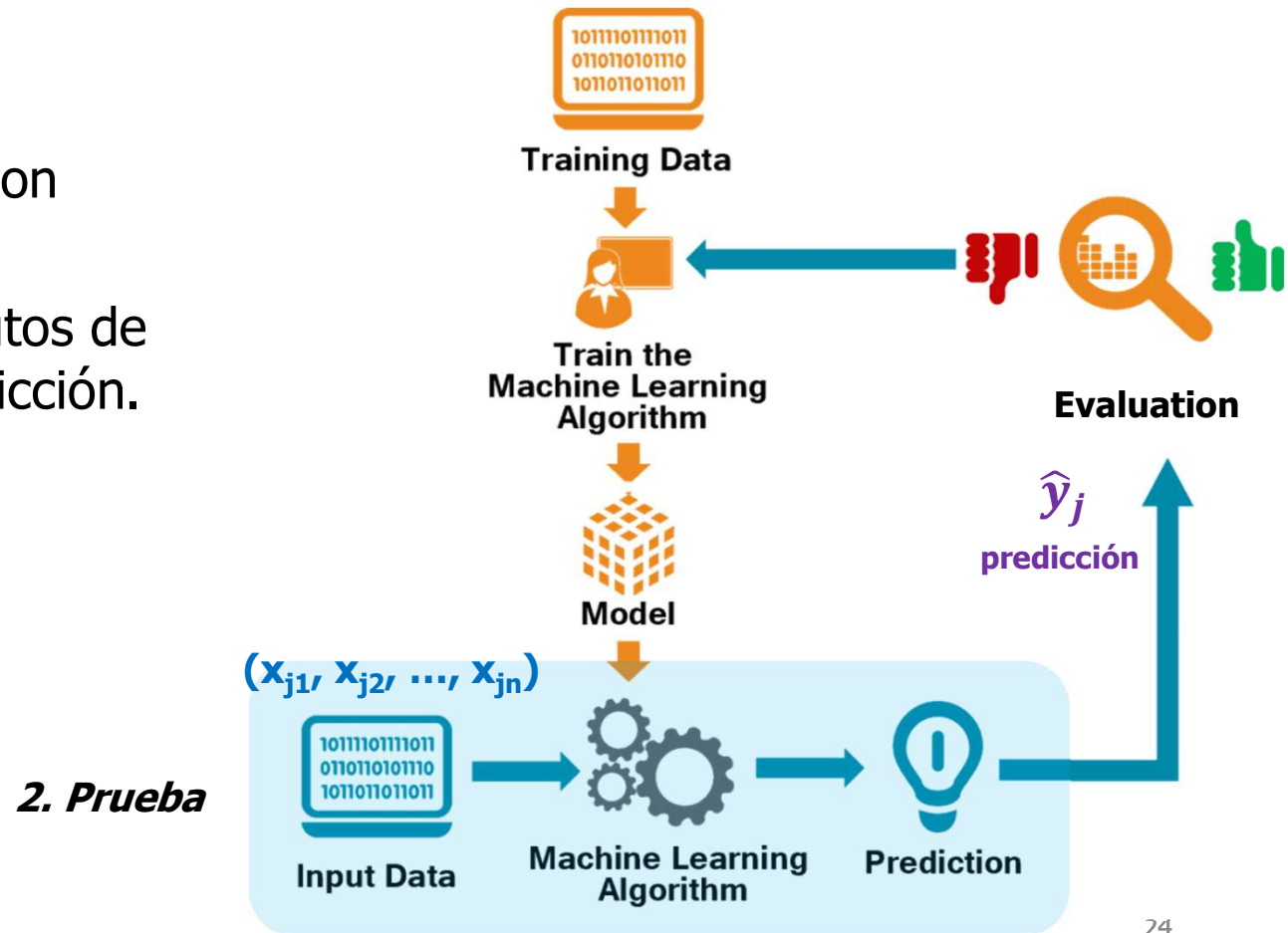
- a. ¿Clasificación o regresión?
- b. ¿Qué atributos voy a utilizar?
- c. ¿Cómo particionar los datos?
- d. ¿Qué modelo voy a usar?
 - ¿Cómo ajusto el valor de los hiperparámetros?



Aprendizaje supervisado

Fase 2: Prueba

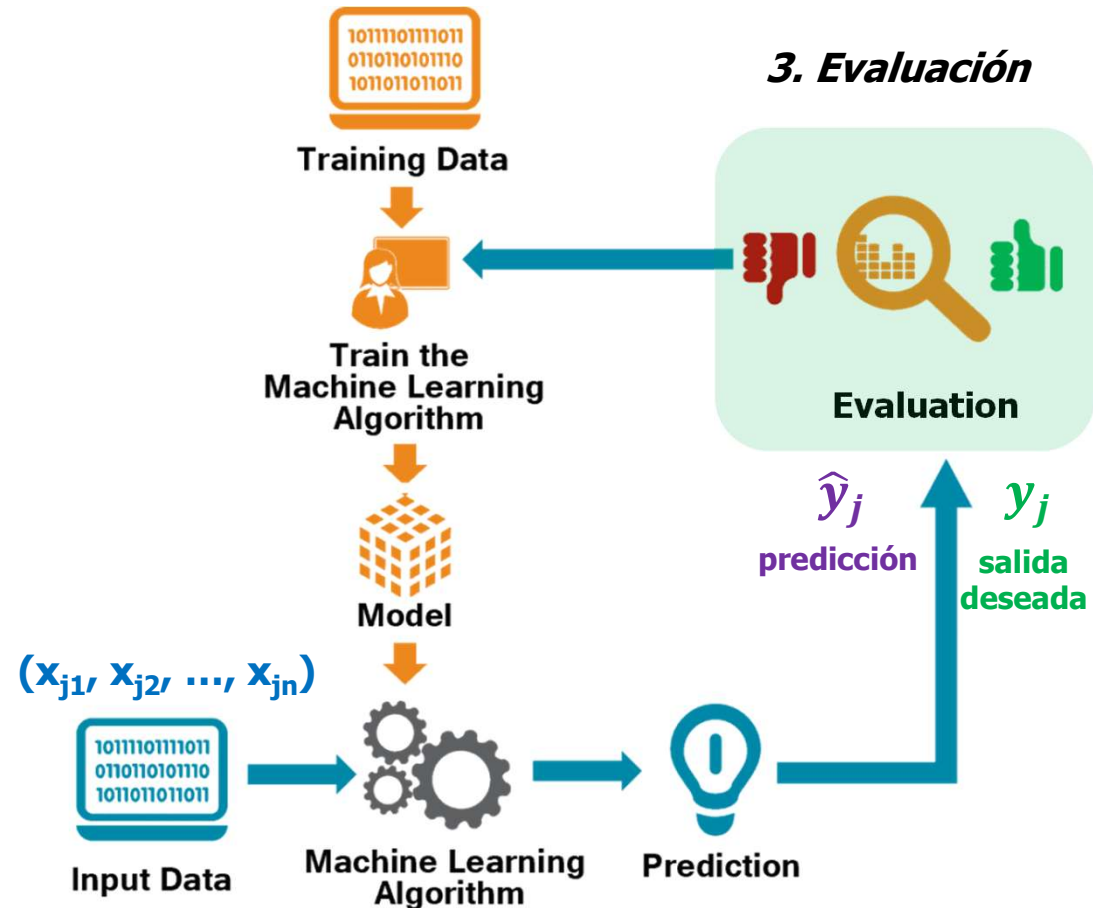
- Se utilizan datos que no fueron usados para entrenamiento.
- Para cada conjunto de atributos de entrada, se genera una predicción.



Aprendizaje supervisado

Fase 3: Evaluación

Se compara la salida deseada y_j con la predicción \hat{y}_j obtenida.

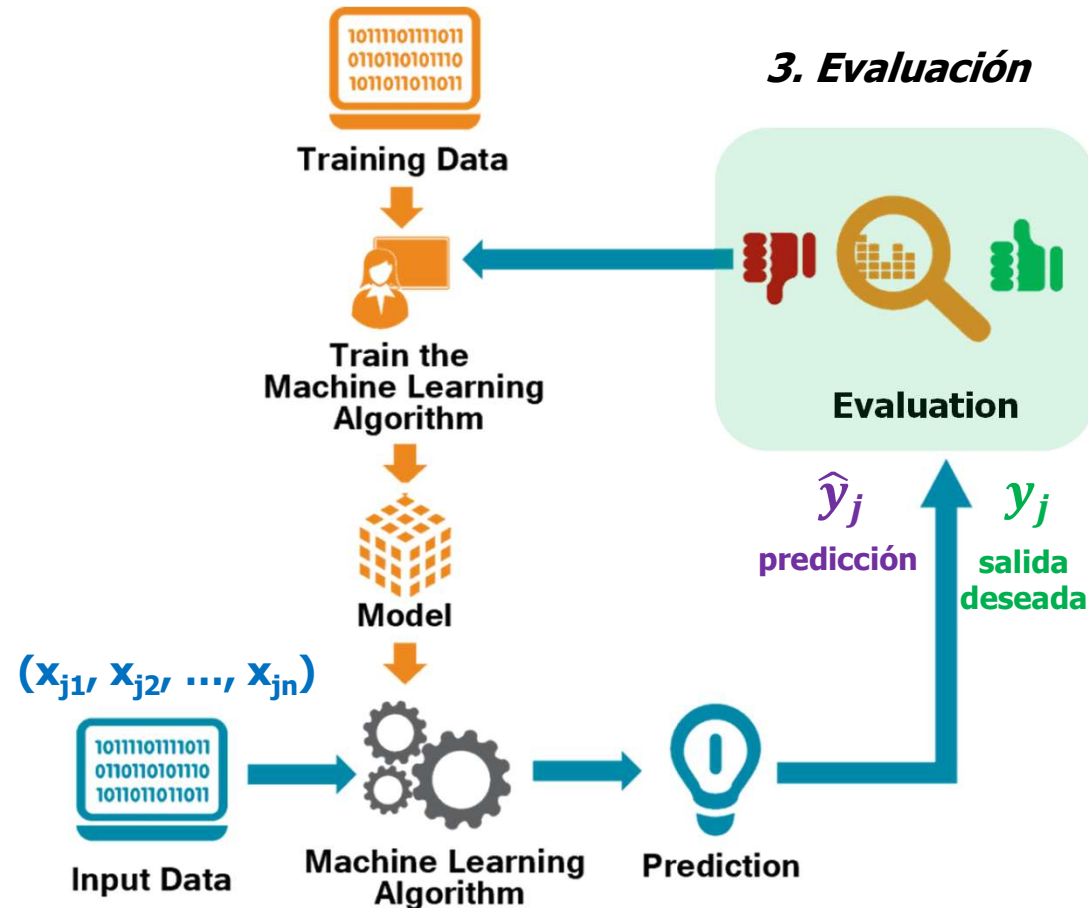


Aprendizaje supervisado

Fase 3: Evaluación

a. ¿Qué métricas voy a utilizar?

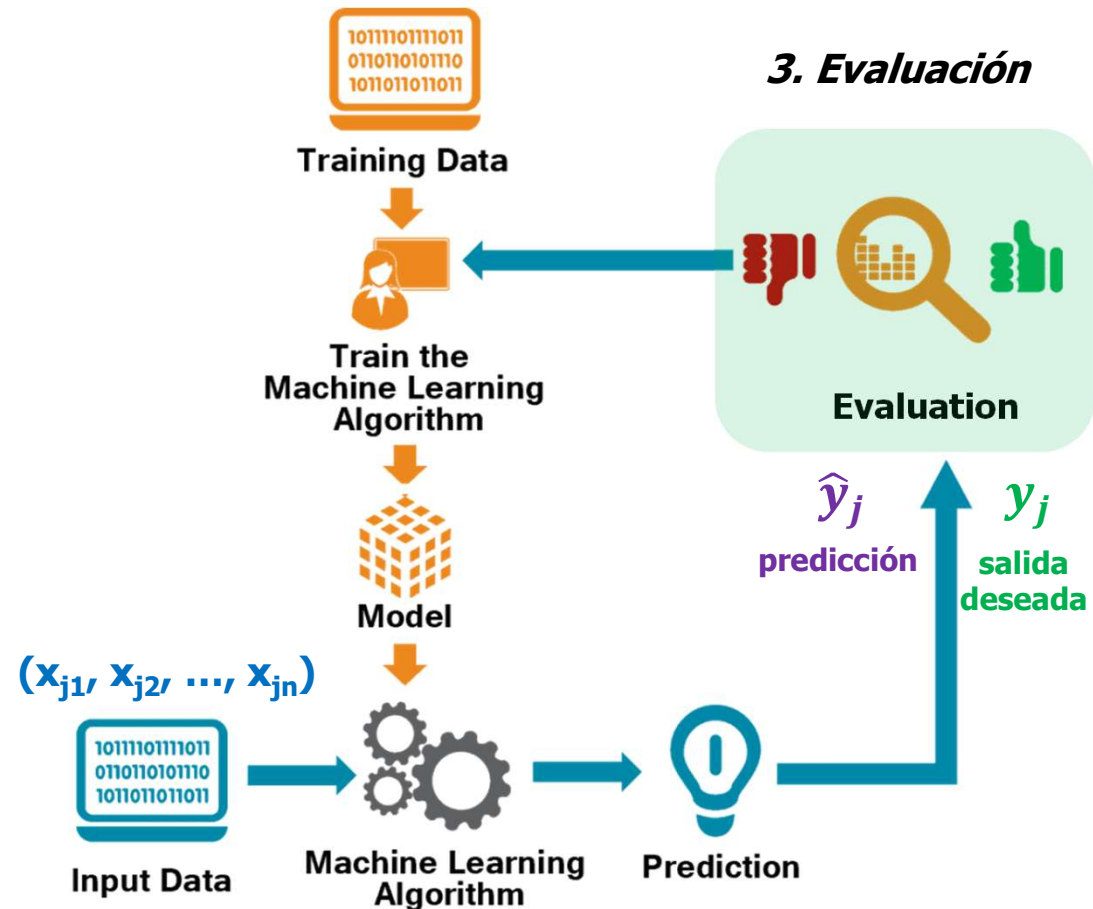
- Si es un modelo de regresión se pueden usar unas métricas.
- Si es un modelo de clasificación se pueden usar otras métricas.



Aprendizaje supervisado

Fase 3: Evaluación

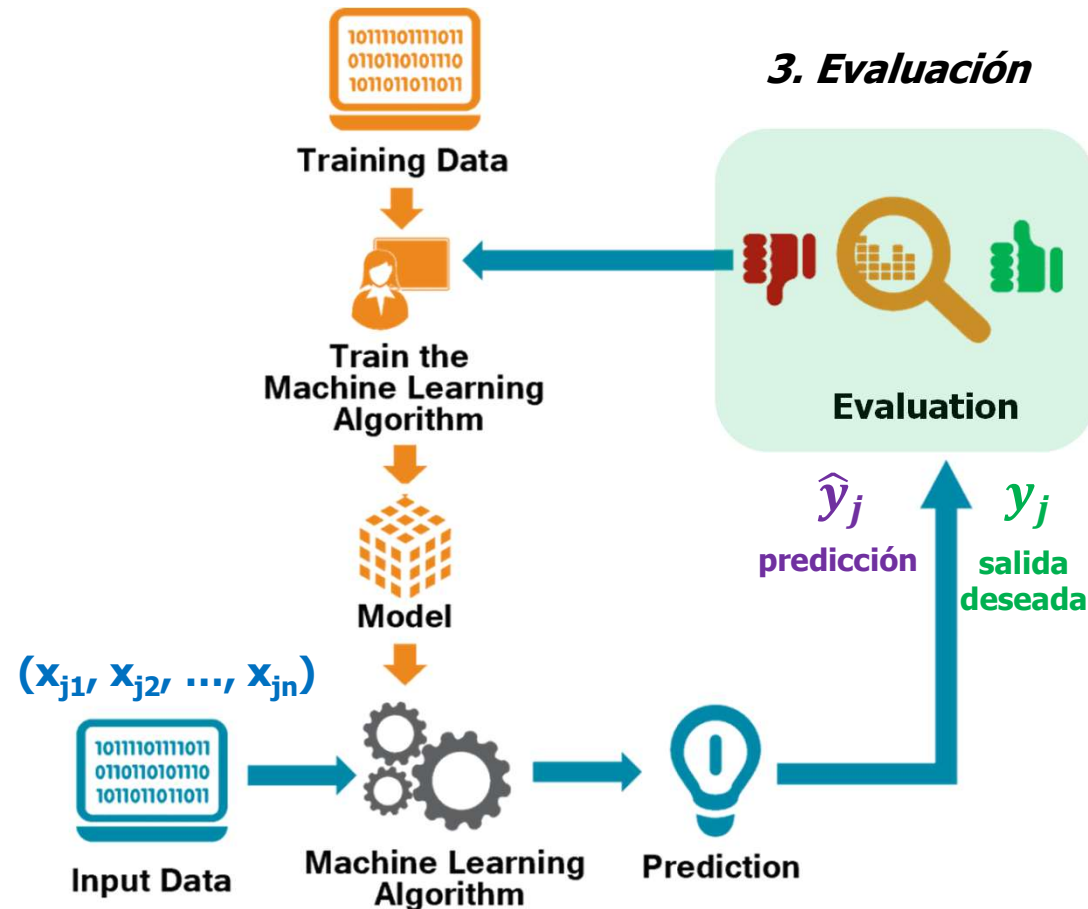
- a. ¿Qué métricas voy a utilizar?
- b. ¿Cuáles son los criterios para considerar que los resultados son exitosos?



Aprendizaje supervisado

Fase 3: Evaluación

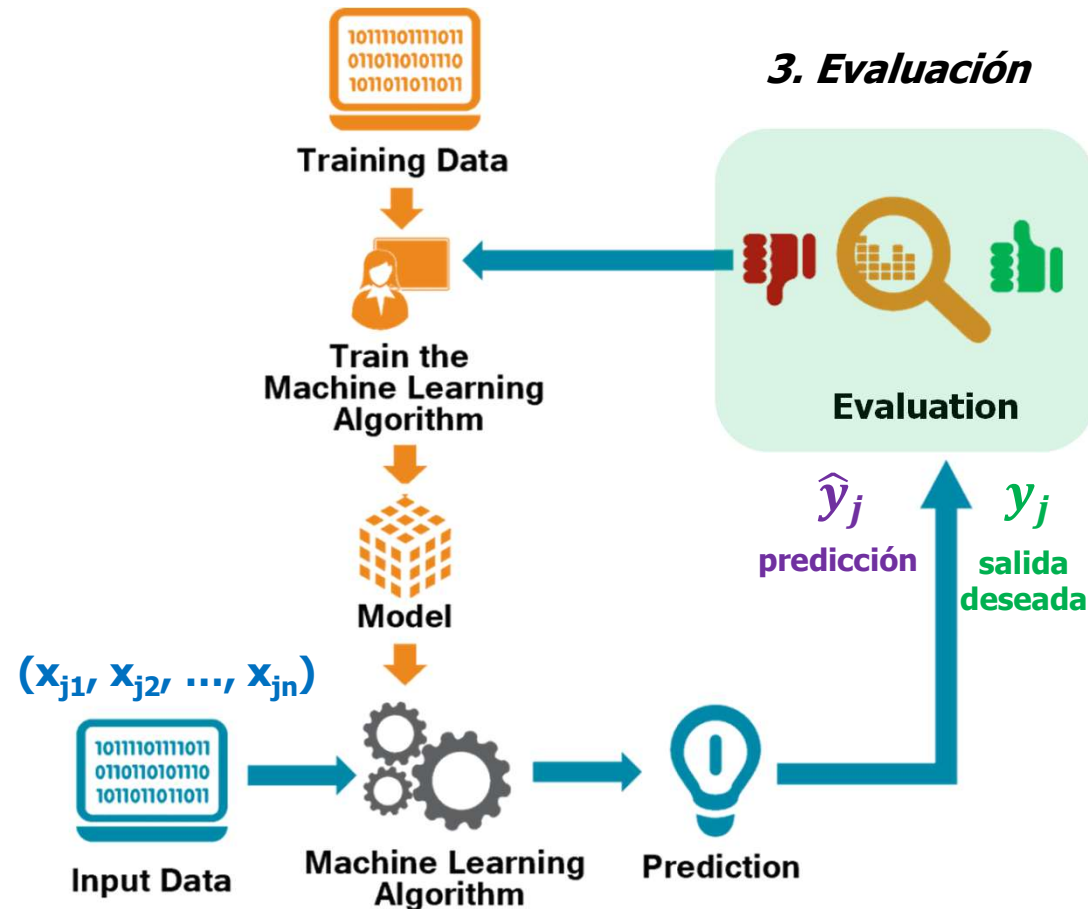
- a. ¿Qué métricas voy a utilizar?
- b. ¿Cuáles son los criterios para considerar que los resultados son exitosos?
- c. ¿Se presenta underfitting (sub entrenamiento) u overfitting (sobre entrenamiento)?



Aprendizaje supervisado

Fase 3: Evaluación

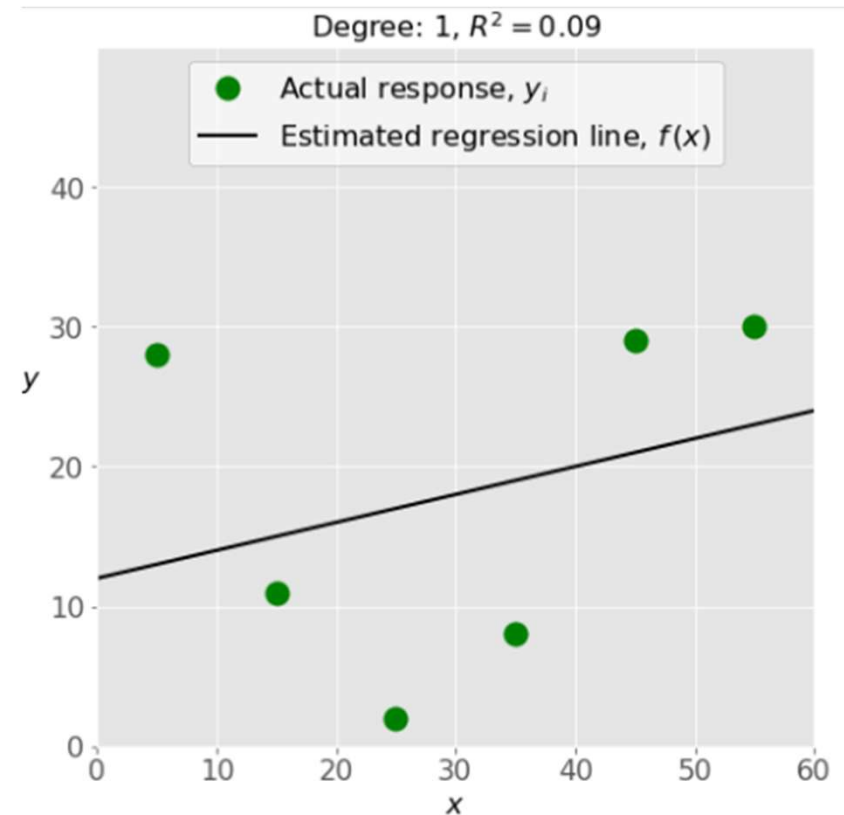
- ¿Qué métricas voy a utilizar?
- ¿Cuáles son los criterios para considerar que los resultados son exitosos?
- ¿Se presenta underfitting (sub entrenamiento) u overfitting (sobre entrenamiento)?
- Si el modelo no es exitoso, podemos modificar los hiperparámetros o usar otro modelo.



Underfitting y Overfitting

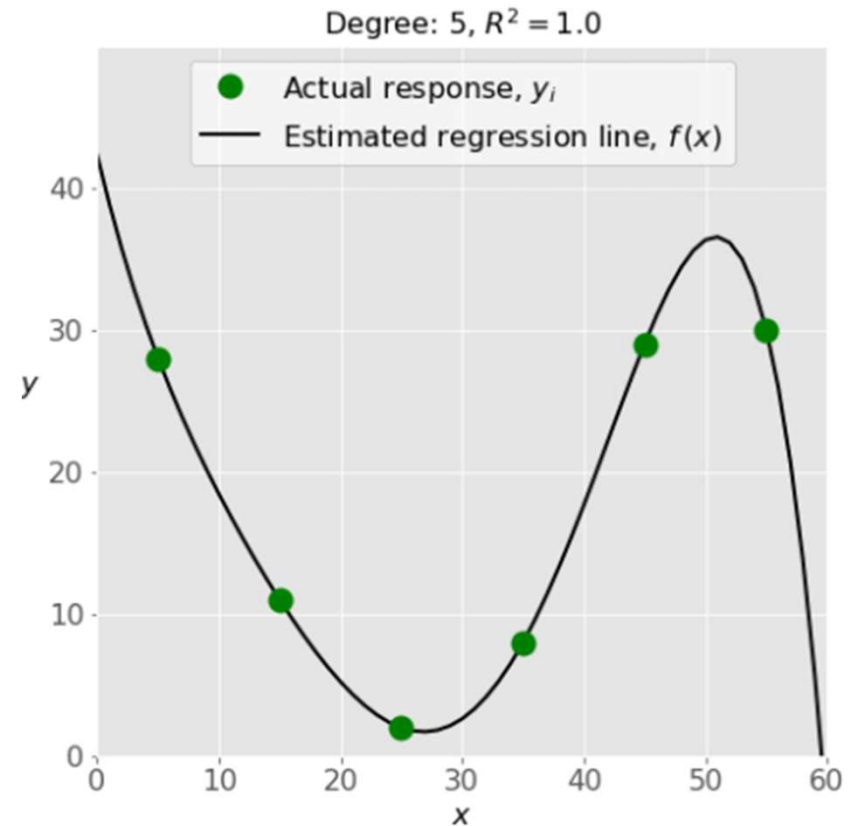
Underfitting and Overfitting

Underfitting: Ocurre cuando el modelo no puede capturar la relación entre la variable objetivo y los predictores. Generalmente, se detecta cuando se obtiene un error alto para los datos de entrenamiento y también para los de prueba.



Underfitting and Overfitting

Overfitting: Ocurre cuando el modelo captura demasiado bien la relación entre la variable objetivo y los predictores en el conjunto de entrenamiento. Los modelos complejos que tienen muchos términos tienden al overfitting. Generalmente, se detecta cuando se obtiene un error muy bajo para los datos de entrenamiento y alto para los de prueba.



Lecturas recomendadas

Supervised Machine Learning

<https://www.javatpoint.com/supervised-machine-learning>

Types of Machine Learning Algorithms You Should Know

<https://towardsdatascience.com/types-of-machine-learning-algorithms-you-should-know-953a08248861>

Referencias

- *Introduction to Statistical Learning with Applications in R (ISLR)*, G. James, D. Witten, T. Hastie & R. Tibshirani, 2014.
- *Data Mining (4th Edition)*, Ian Witten, Eibe Frank, Mark A. Hall & Christopher J. Pal, Elsevier, 2016.
- *Machine Learning*, Tom M. Mitchell, McGraw-Hill, 1997.
- *Data Science for Business*, Foster Provost & Tom Fawcett, O'Reilly, 2013.