

Introducción a Machine Learning

**Licenciatura en Inteligencia Artificial y Ciencia de Datos, CUGDL,
Universidad de Guadalajara.**

Guadalajara, Jal., agosto de 2025

¿Qué es el aprendizaje automático?

Conceptos básicos

- Definición breve: "Disciplina que permite que las máquinas aprendan de los datos sin ser explícitamente programadas"
- Estos sistemas emplean métodos estadísticos y computacionales para "aprender" de la información proporcionada.
- Este aprendizaje se logra al procesar grandes cantidades de información y ajustar sus parámetros internos mediante diferentes técnicas, lo que permite mejorar su desempeño a medida que se exponen a más datos.



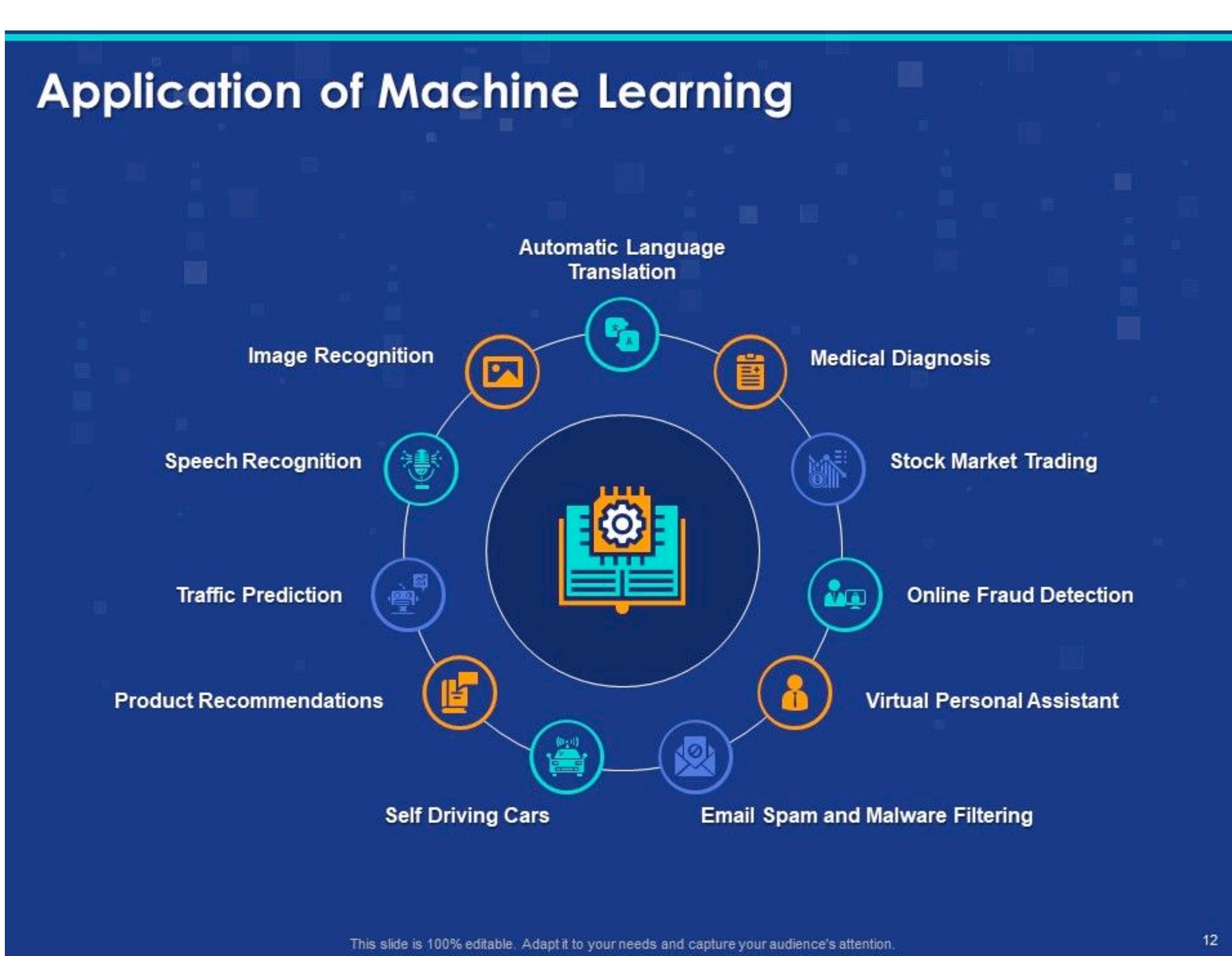
¿Por qué es importante?

Aplicaciones

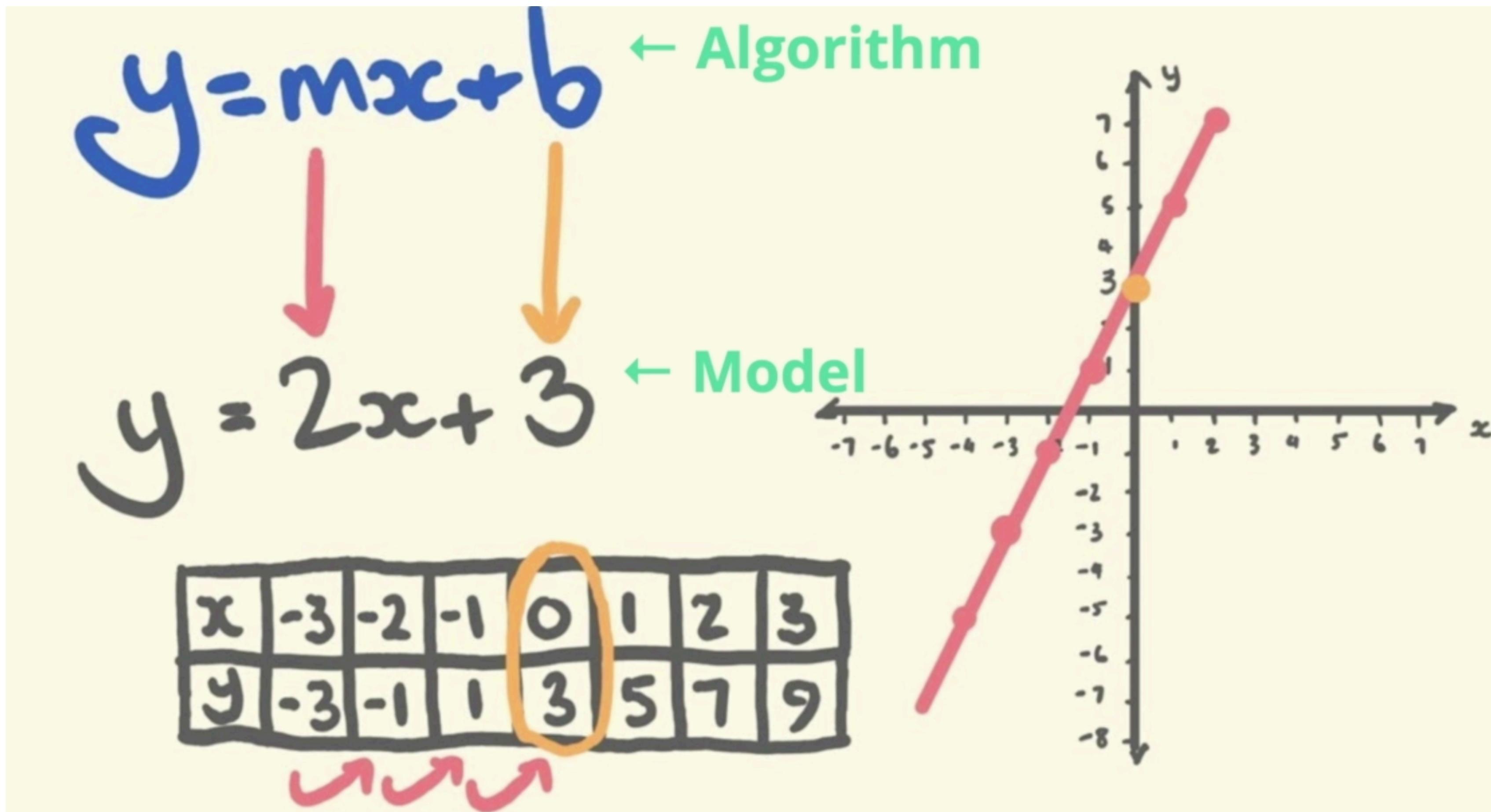
Está presente en muchas aplicaciones cotidianas.

- Recomendaciones (Netflix, Spotify, ...)
- Diagnóstico médico
- Predicción financiera
- Traducción automática

Automatiza decisiones basadas en datos.



Algoritmos y modelos



"Algorithm is to model as teaching is to doing"

Ecuación

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots$$

Optimización

$$e = \sum (y - \hat{y})^2$$

Algoritmo

$$\beta_1 = \frac{\sigma_{xy}}{\sigma_x^2}$$

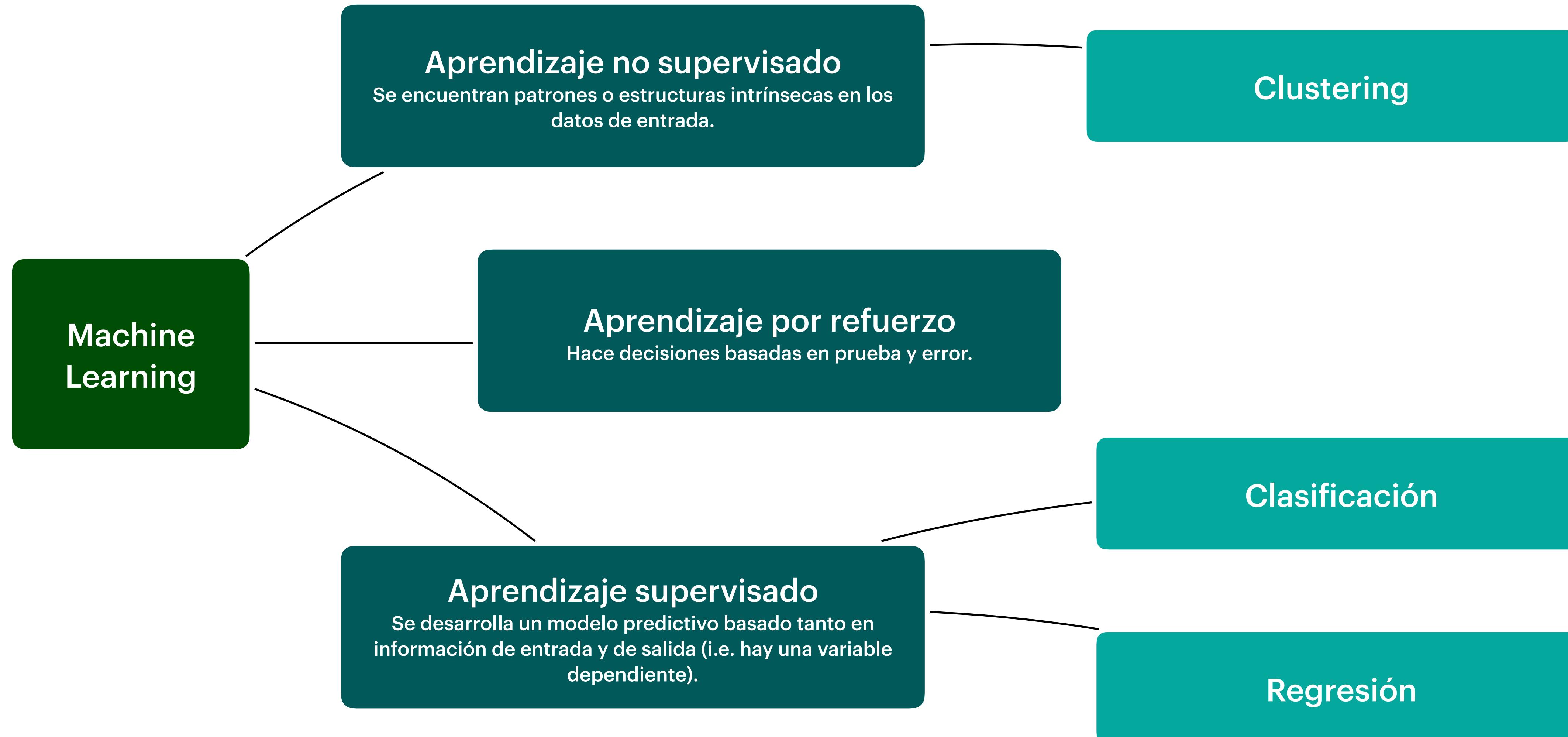
$$\beta_0 = \bar{y} - \beta_1 \bar{x}$$

Modelo

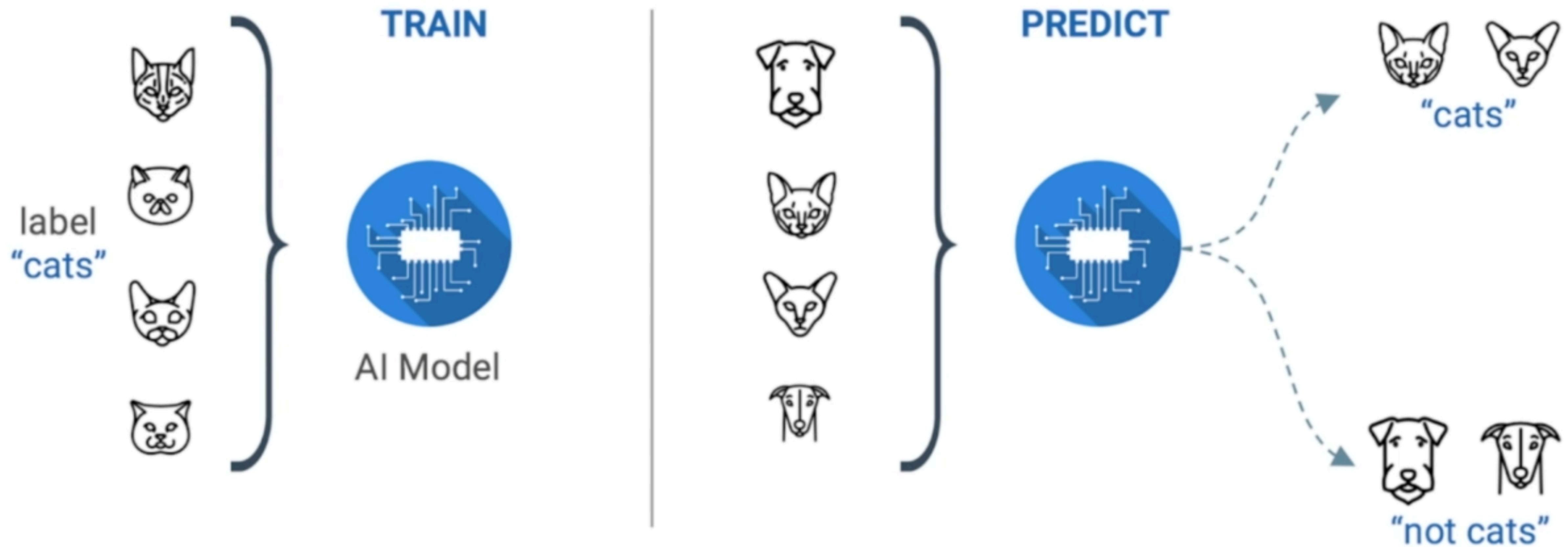
$$\text{precio} = \beta_0 + \beta_1 \text{material}$$

Tipos de aprendizaje

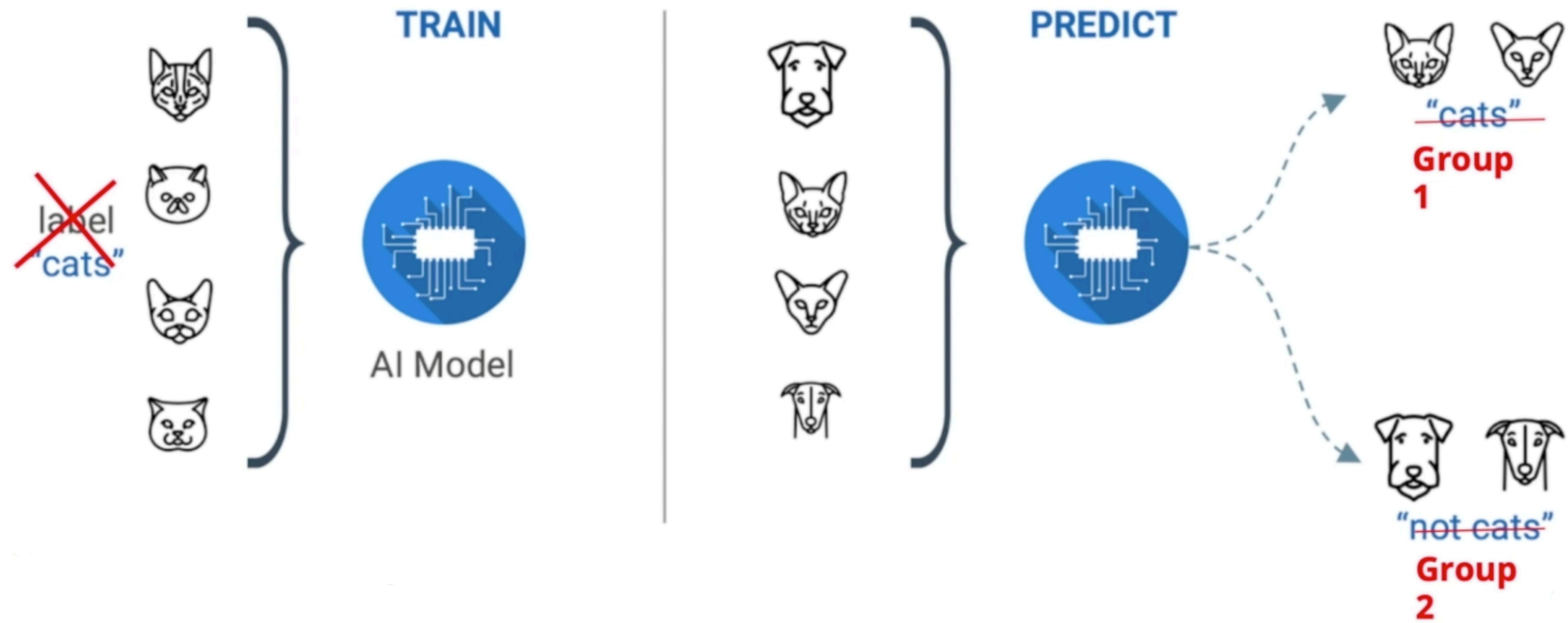
Supervisado, No supervisado, Por refuerzo



Aprendizaje automático supervisado



Aprendizaje automático no supervisado



Clasificación vs Regresión



Regression

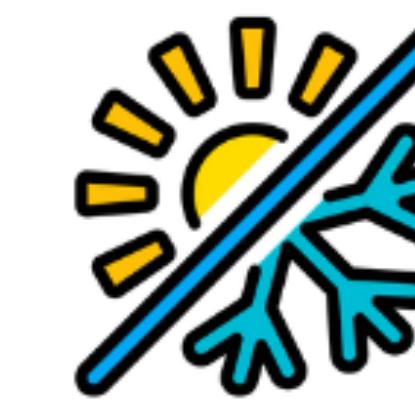


What will be the temperature tomorrow?

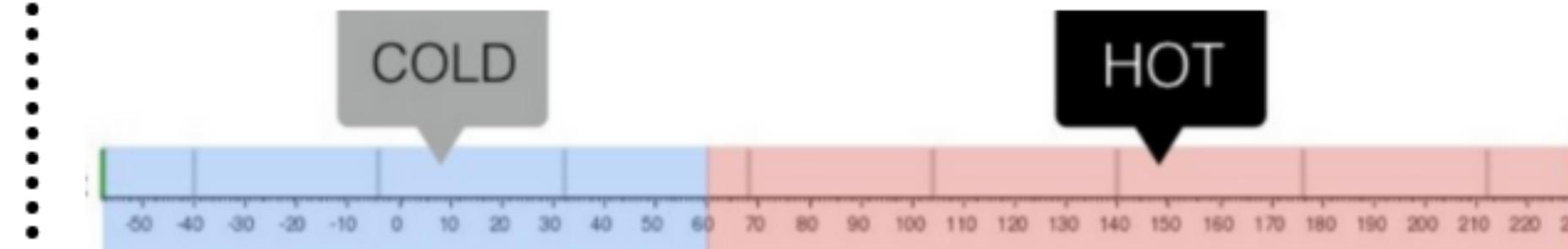


Fahrenheit

Classification



Will it be hot or cold tomorrow?



Fahrenheit

Las técnicas de clasificación predicen respuestas categóricas, mientras que las técnicas de regresión predicen una respuesta continua.

Algoritmos más comunes

Classification

Logistic Regression
Binomial logistic regression
Multinomial logistic regression
Decision tree classifier
Random forest classifier
Gradient-boosted tree classifier
Multilayer perceptron classifier
Linear Support Vector Machine
One-vs-Rest classifier
Naive Bayes

Regression

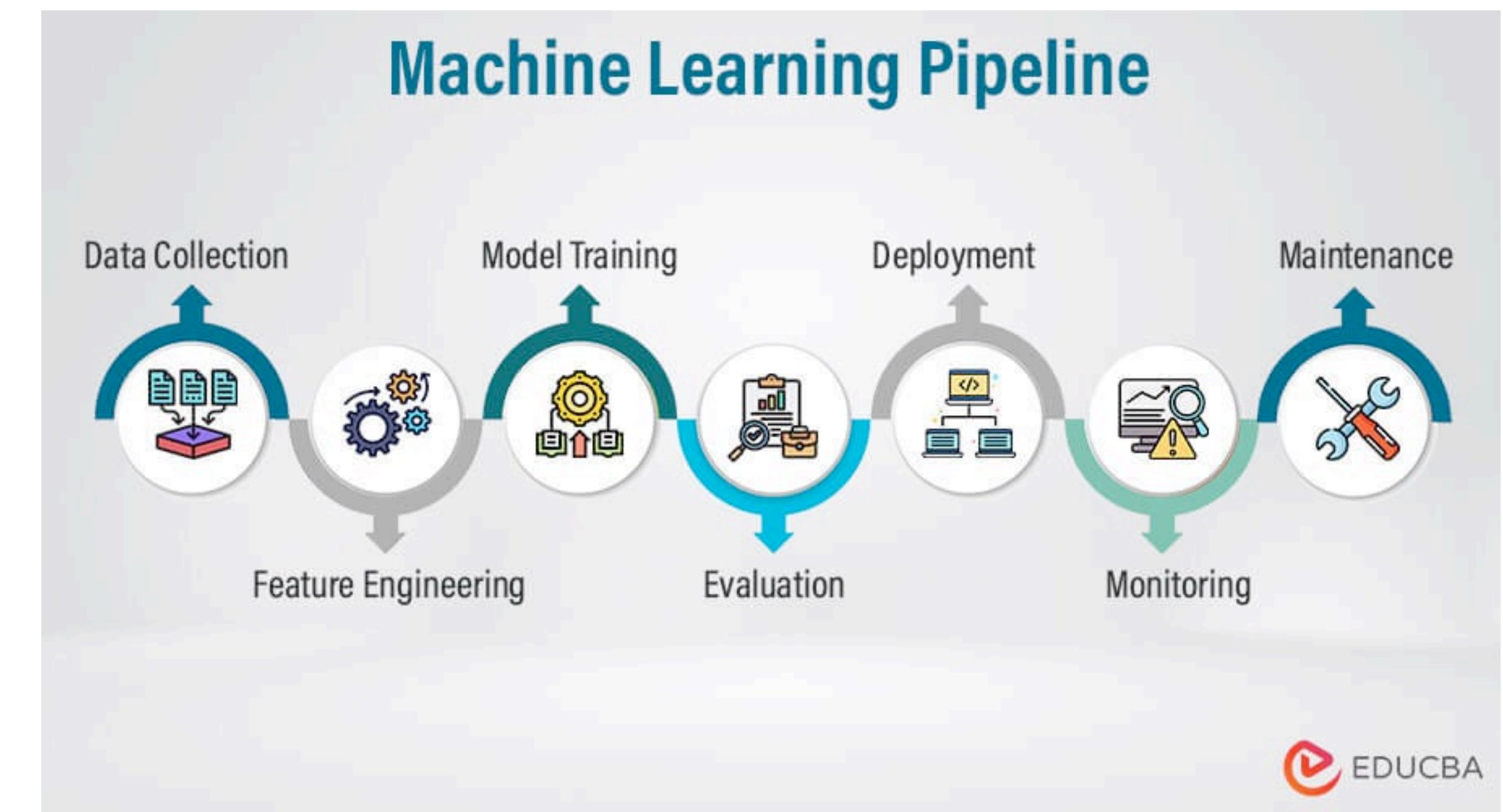
Linear regression
Generalized linear regression
Decision tree regression
Random forest regression
Gradient-boosted tree regression

Clustering

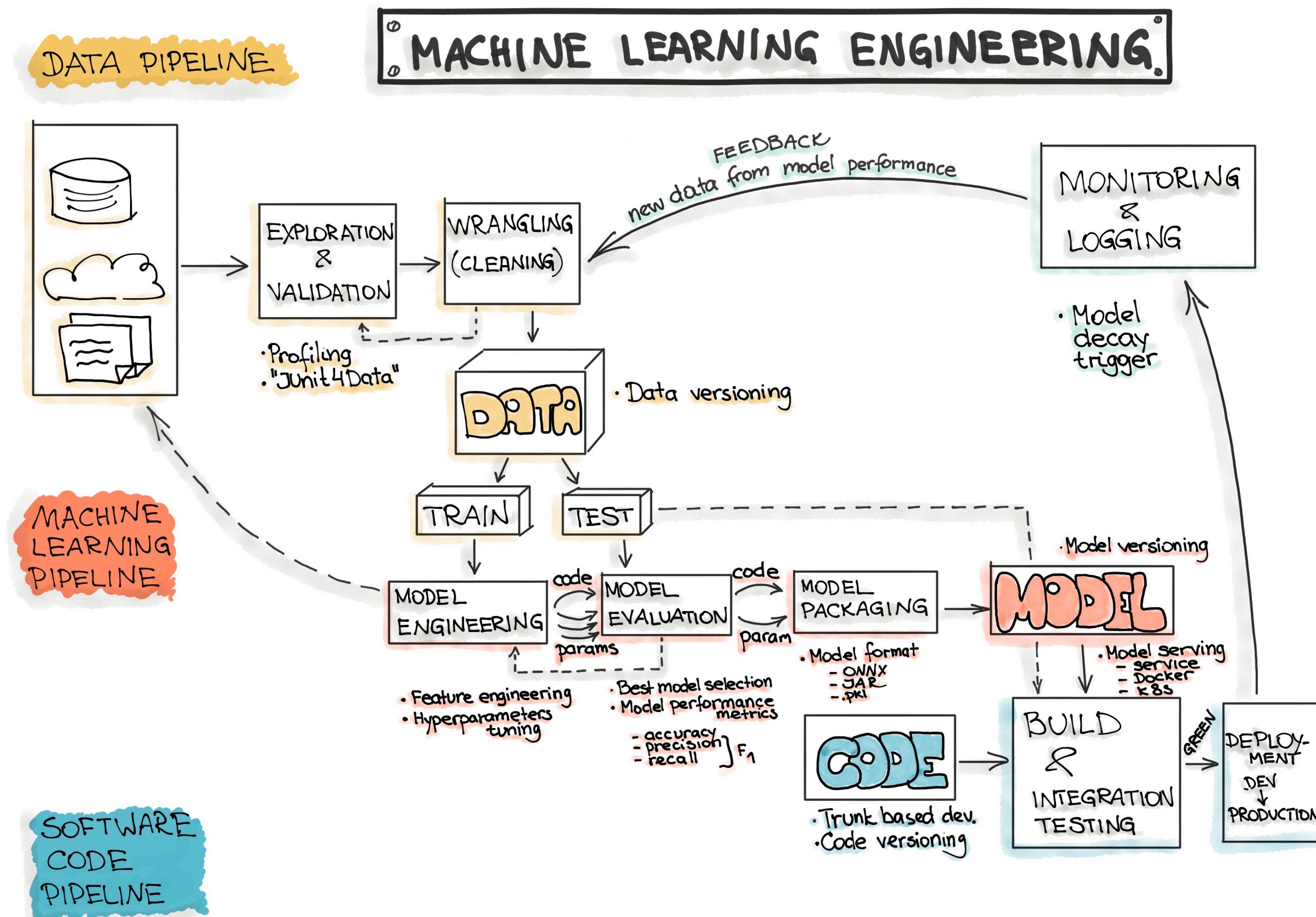
K-means
Latent Dirichlet allocation (LDA)
Bisecting k-means
Gaussian Mixture Model (GMM)

Componentes de un problema de ML

- Conjunto de datos (dataset).
- Características (features).
- Variable objetivo (target).
- Modelo
- Fase de entrenamiento y prueba
- Evaluación
- Despliege



Flujo de trabajo típico

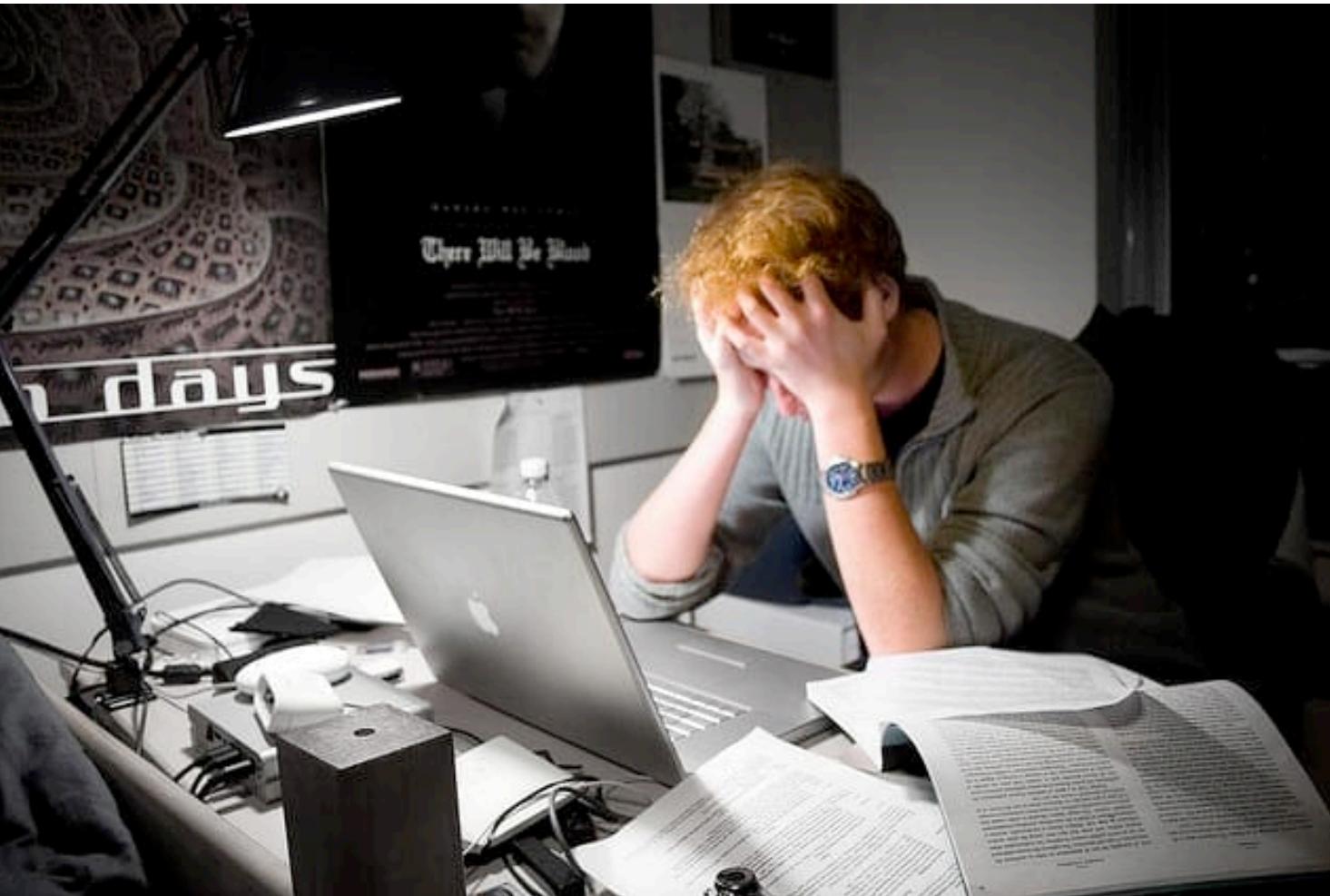


- Recolección de datos
- Limpieza y preprocesamiento
- División en conjuntos de entrenamiento y prueba
- Entrenamiento del modelo
- Evaluación
- Ajuste
- Implementación

Problemas comunes

Cosas que iremos abordando durante el curso

- ¿Cómo evitar que el modelo memorice todo? **(Overfitting)**
- ¿Cómo medir el rendimiento de un modelo?
- ¿Cómo elegir el modelo correcto?
- ¿Qué hacer si tenemos pocos datos?



Herramientas

Lenguaje: Python

Librerías: numpy, matplotlib, seaborn, pandas, scikit-learn

Entorno: Jupyter Notebook
(via Anaconda o Google Colab)

Con enfoque práctico desde el principio.



BEST PYTHON LIBRARIES FOR MACHINE LEARNING

TensorFlow

Keras

matplotlib

SciPy

pandas

PyTorch

orange

NumPy

scikit-learn

theano

¿Qué no es Machine Learning?



Magic ?



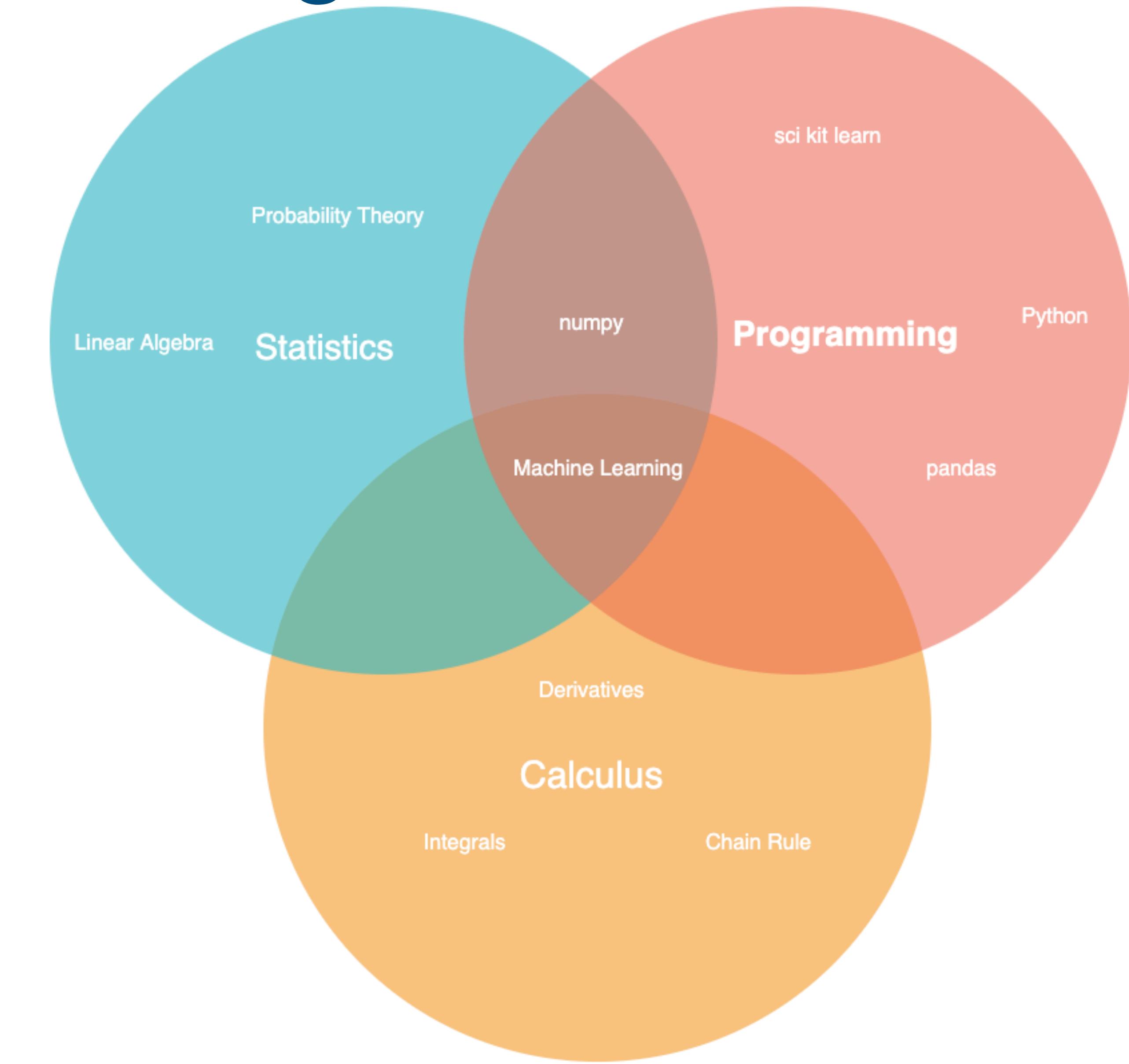
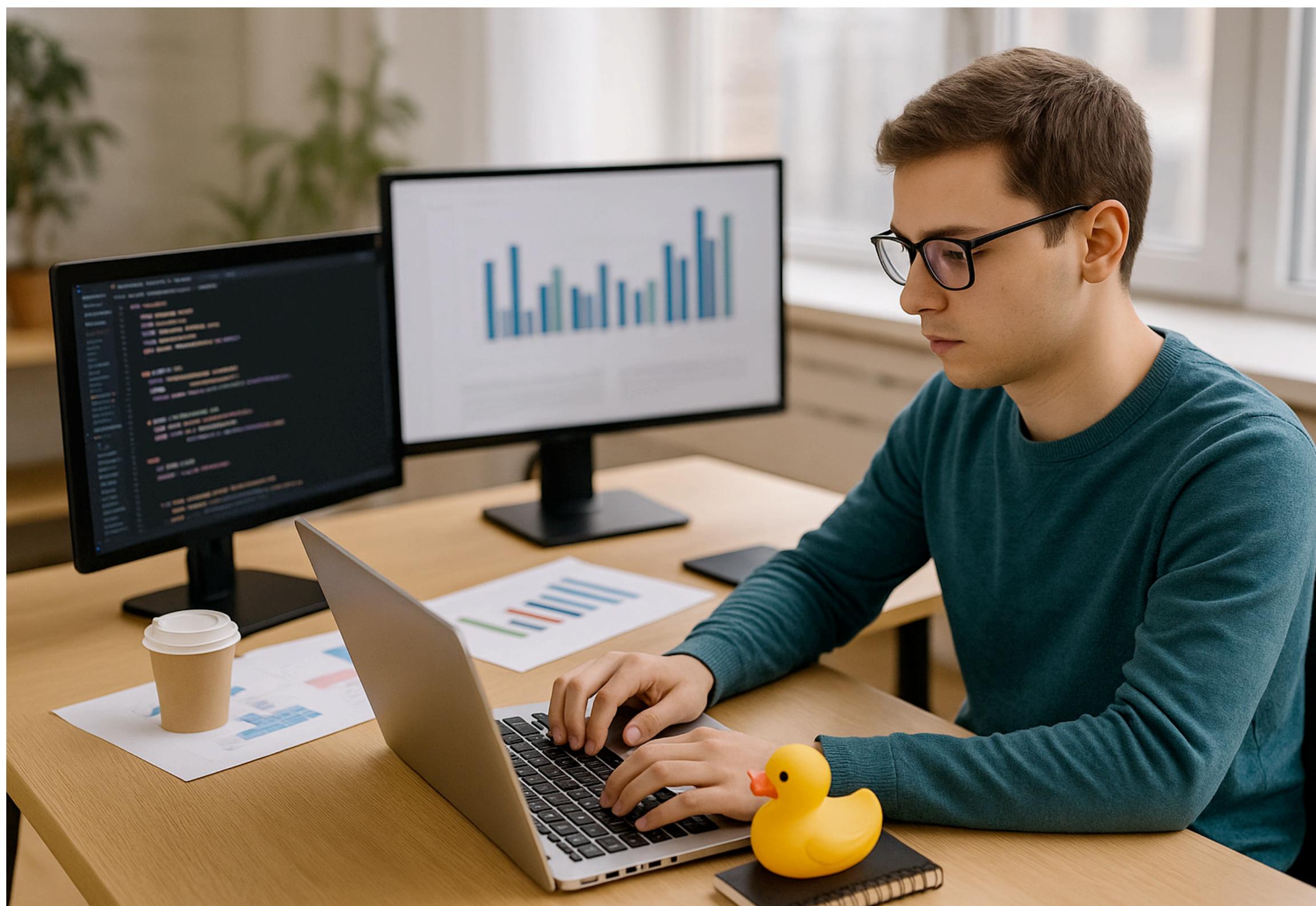
human-level
thinking ?



self-aware
intelligence?

- No es magia.
- No reemplaza el razonamiento humano.
- No funciona bien con datos mal recolectados o sin contexto.

¿Qué es Machine Learning?



Lo que vamos a aprender

- Aplicar modelos de ML supervisados y no supervisados.
- Evaluar estos modelos con métricas estadísticas relevantes.
- Entender cuándo usar cada enfoque.
- Introducción al enfoque bayesiano
- **No se trata de saberlo todo desde el inicio, sino la disposición para aprender con curiosidad**