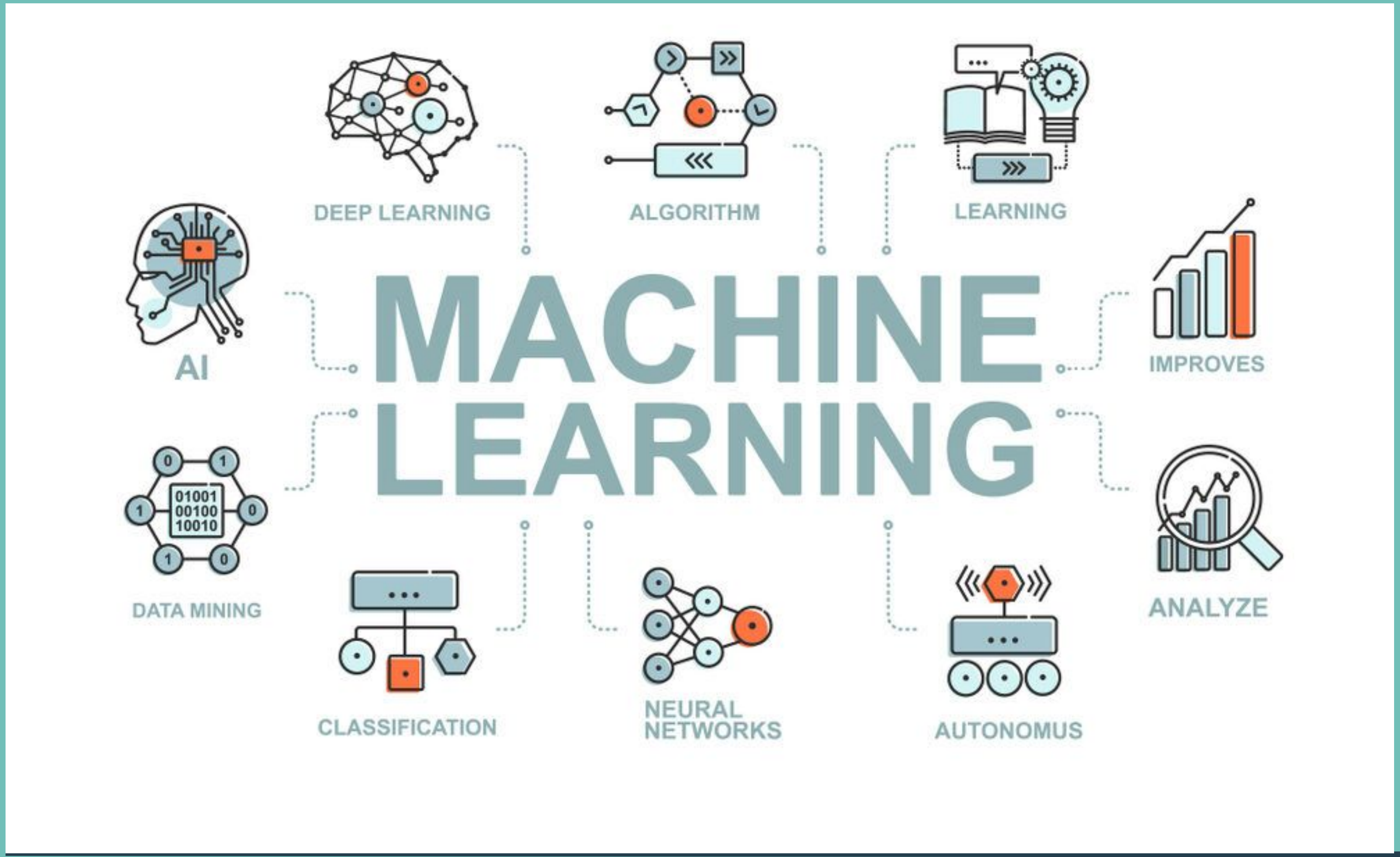
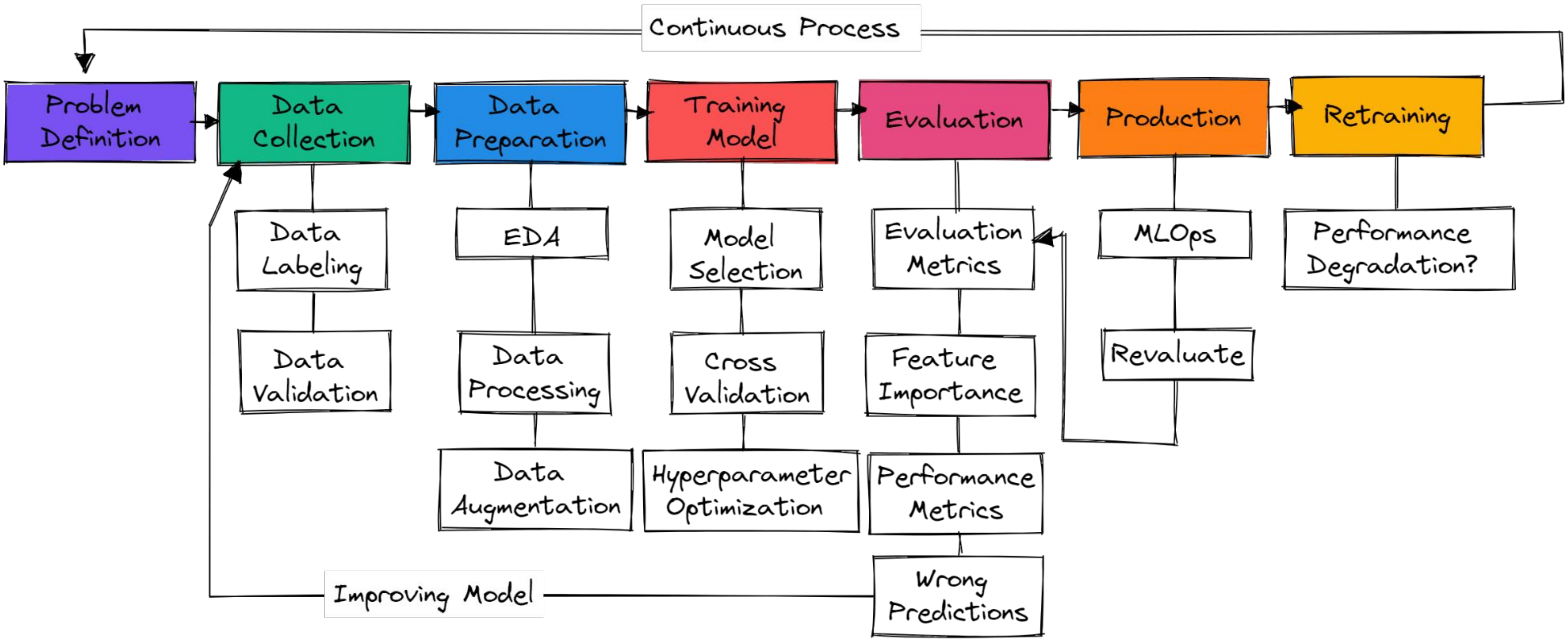

DEEP LEARNING
ALGORITHM
LEARNING
AI
MACHINE LEARNING
IMPROVES
DATA MINING
ANALYZE
CLASSIFICATION
NEURAL NETWORKS
AUTONOMUS

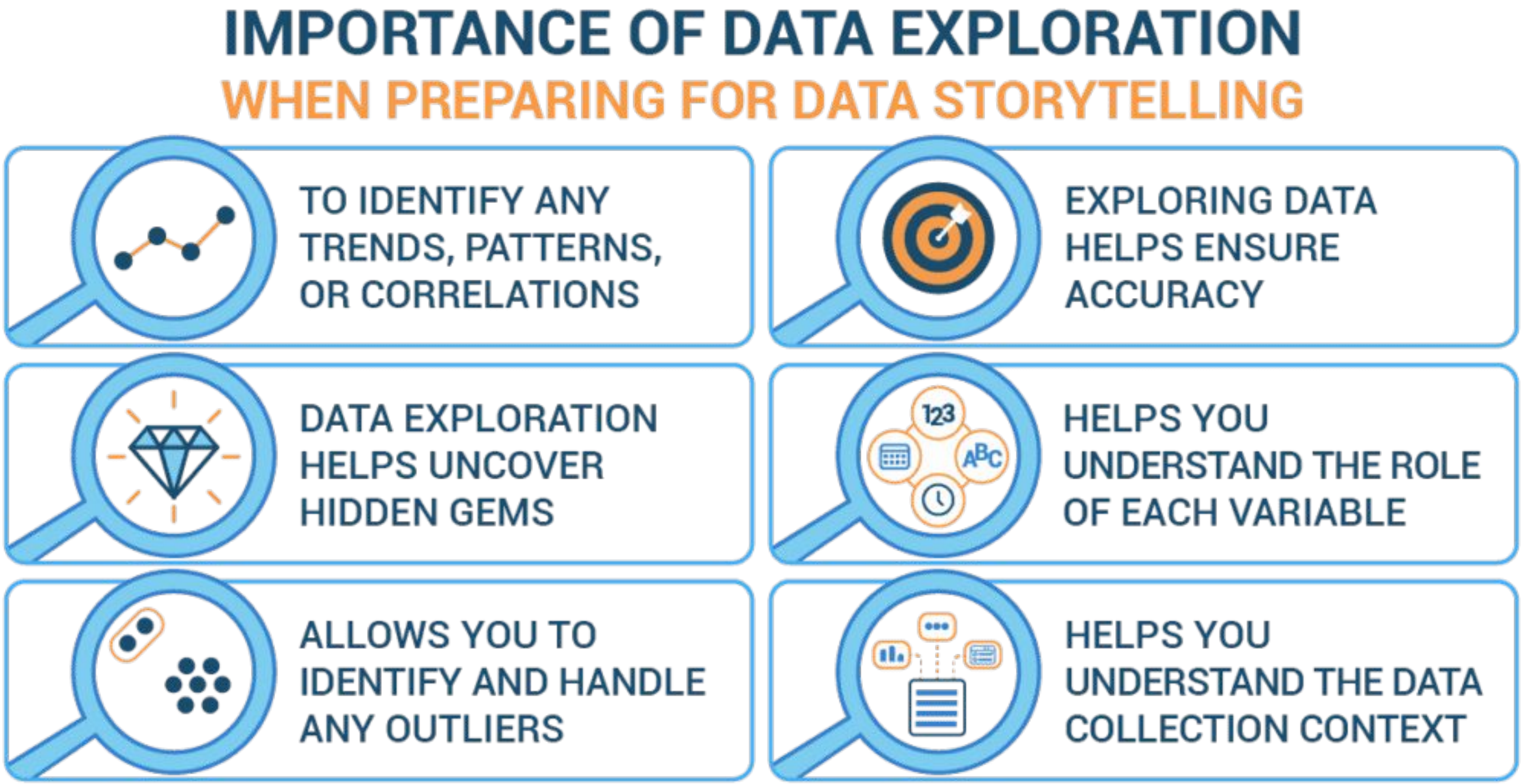
Resumen ML

ML Project



Continuous Process
Problem Definition
Data Collection
Data Preparation
Training Model
Evaluation
Production
Retraining
Data Labeling
Data Validation
EDA
Data Processing
Data Augmentation
Model Selection
Cross Validation
Hyperparameter Optimization
Evaluation Metrics
Feature Importance
Performance Metrics
Wrong Predictions
MLOps
Revaluate
Performance Degradation?
Improving Model

Data Exploration



Types of Machine Learning – At a Glance

Supervised Learning

• Makes machine Learn explicitly
• Data with clearly defined output is given
• Direct feedback is given
• Predicts outcome/future
• Resolves classification and regression problems

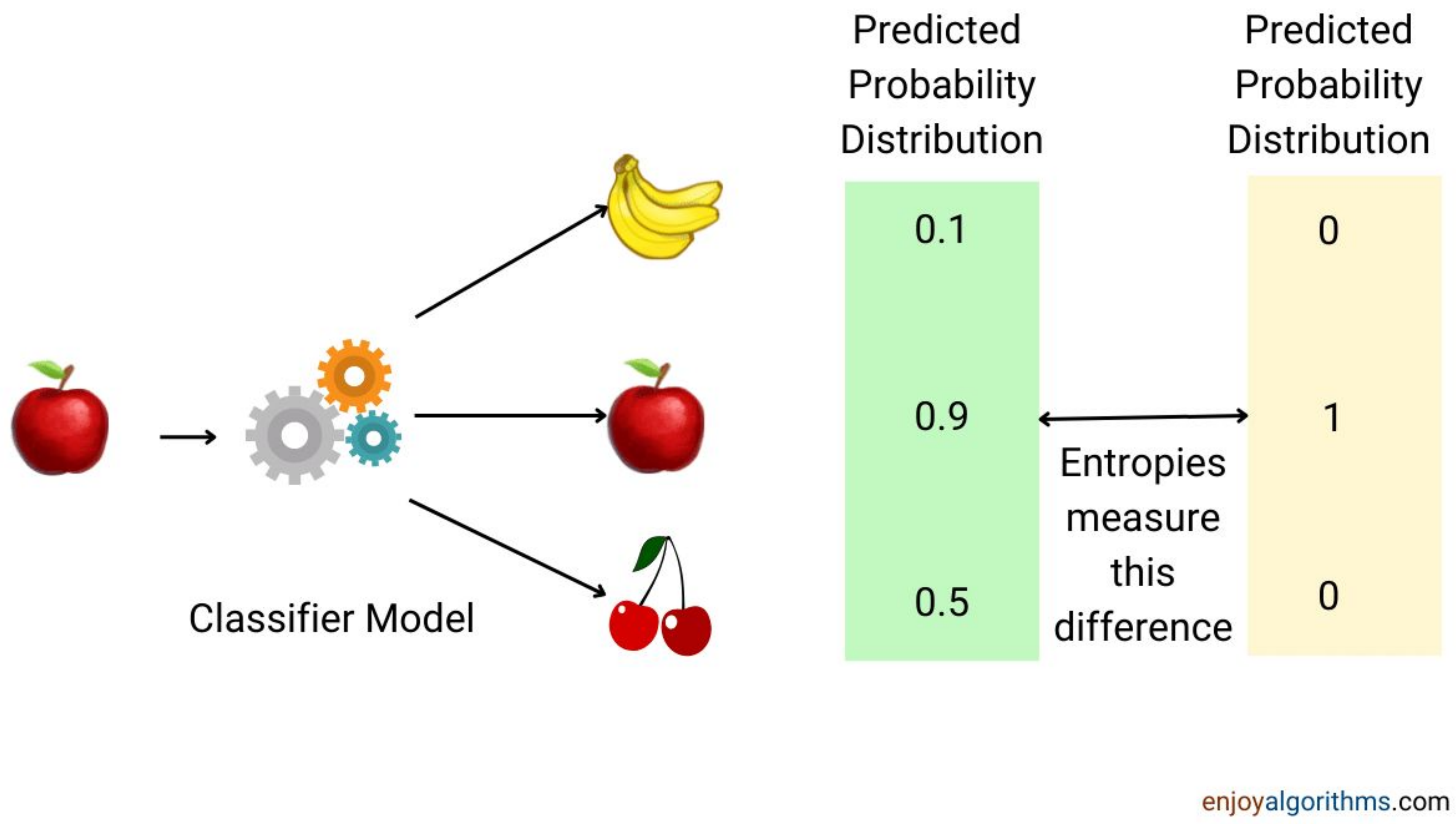
Unsupervised Learning

• Machine understands the data (Identifies patterns/structures)
•Evaluation is qualitative or indirect
• Does not predict/find anything specific

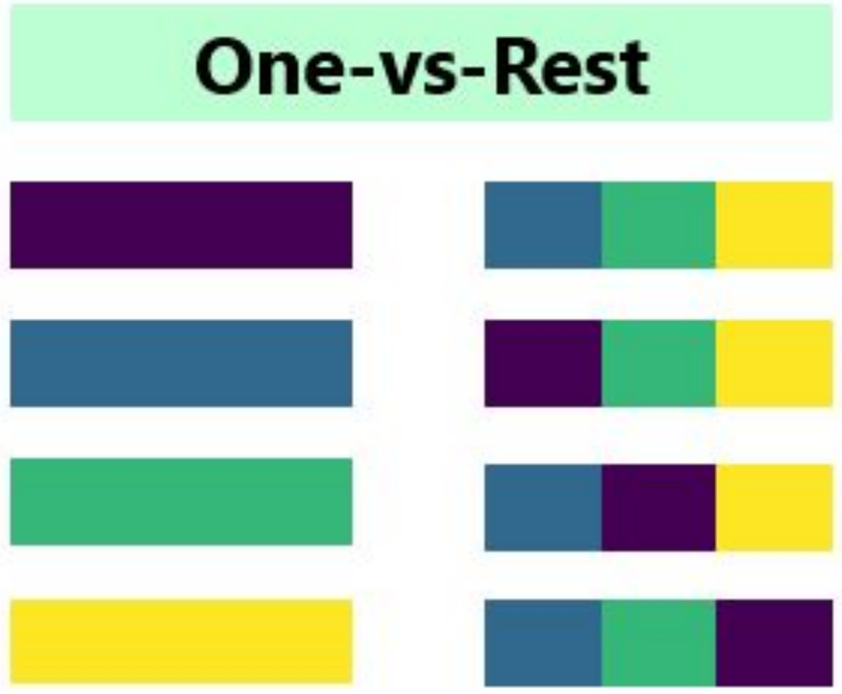
Reinforcement Learning

• An approach to AI
• Reward based learning
• Learning form +ve & +ve reinforcement
•Machine Learns how to act in a certain environment
• To maximize rewards

Supervised-Classification

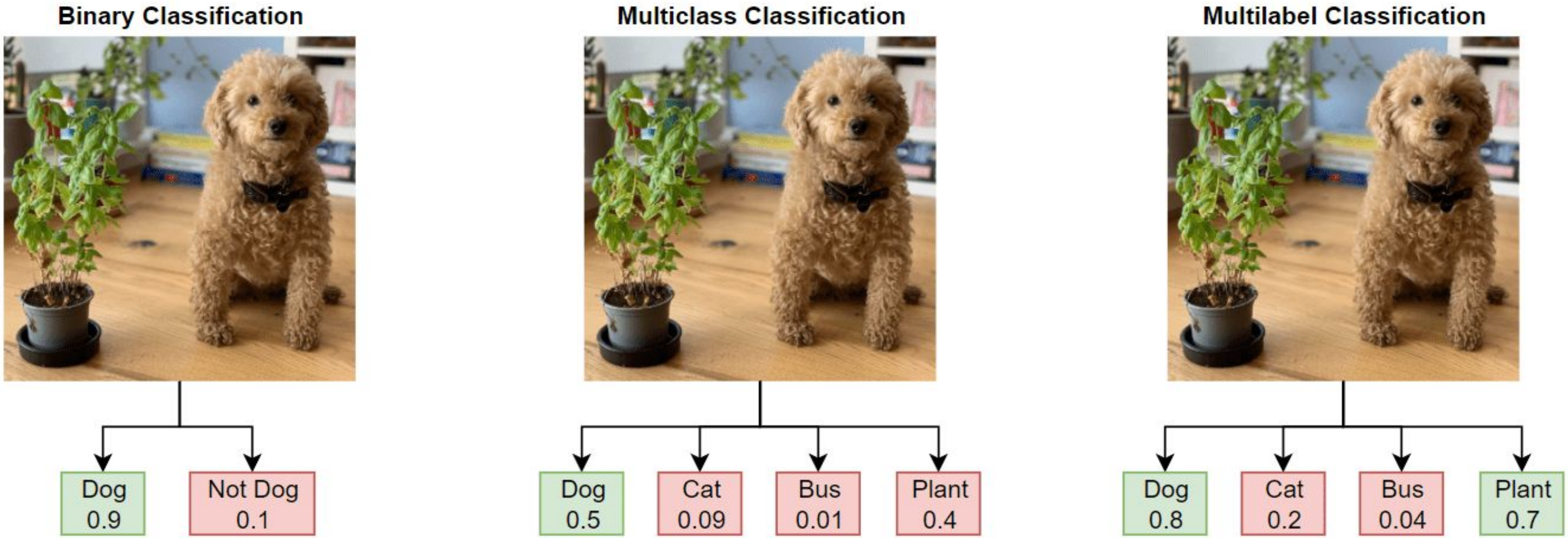


Supervised-Multi Classification

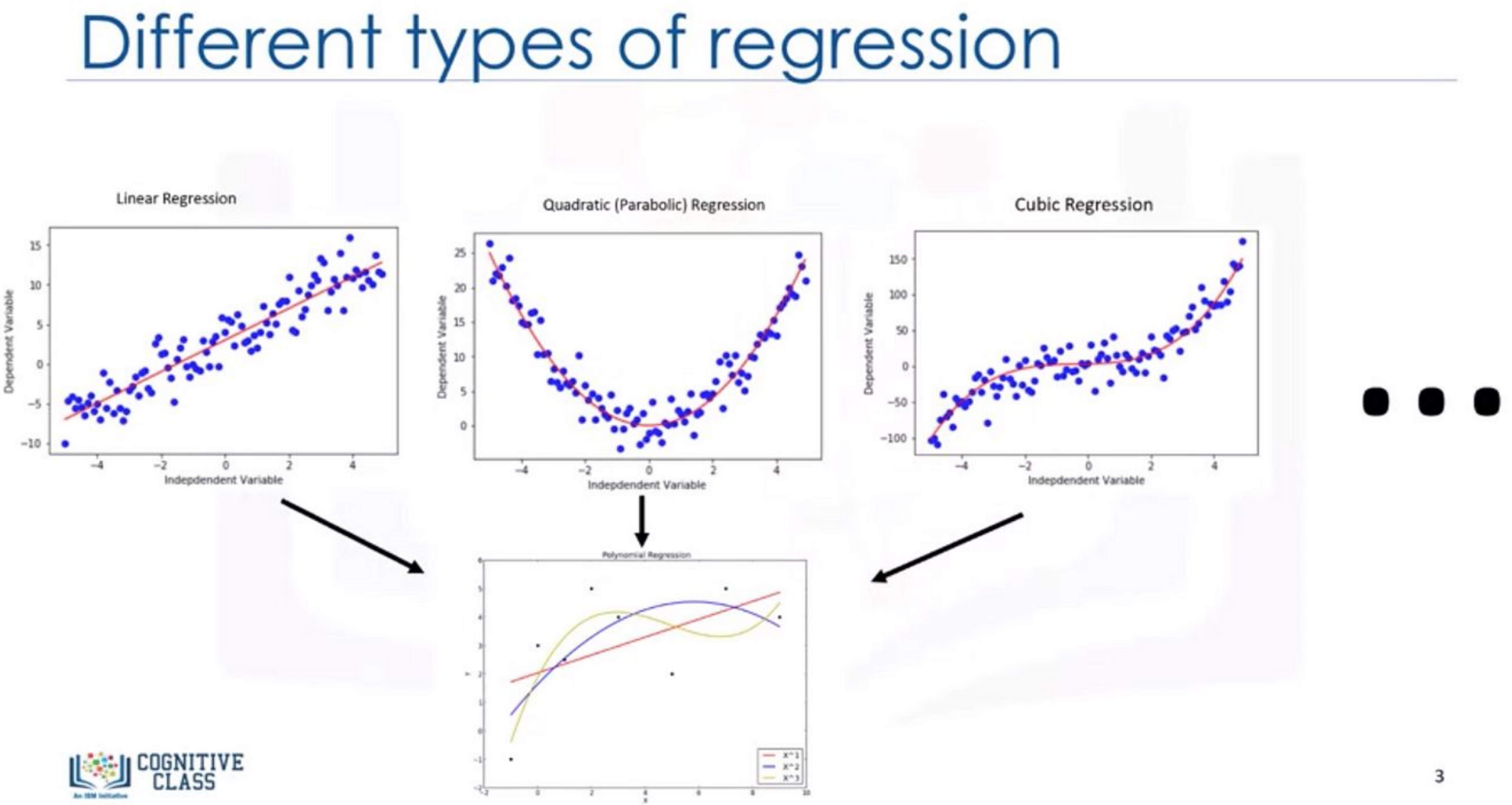


Train one binary model for **each class**, with remaining classes treated as aggregates. Also known as **One-vs-All**. Commonly used with logistic regression.

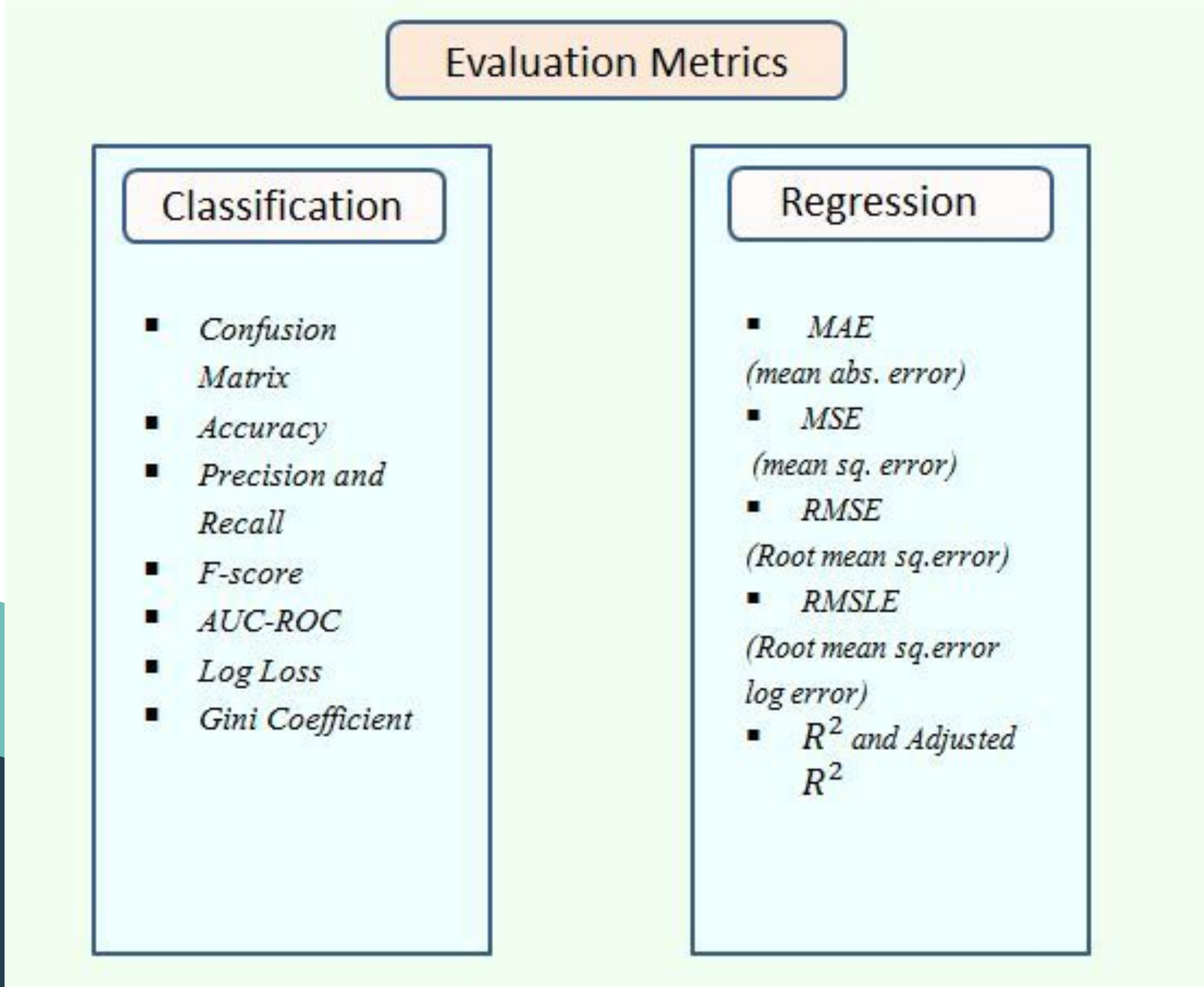
Supervised-Classification



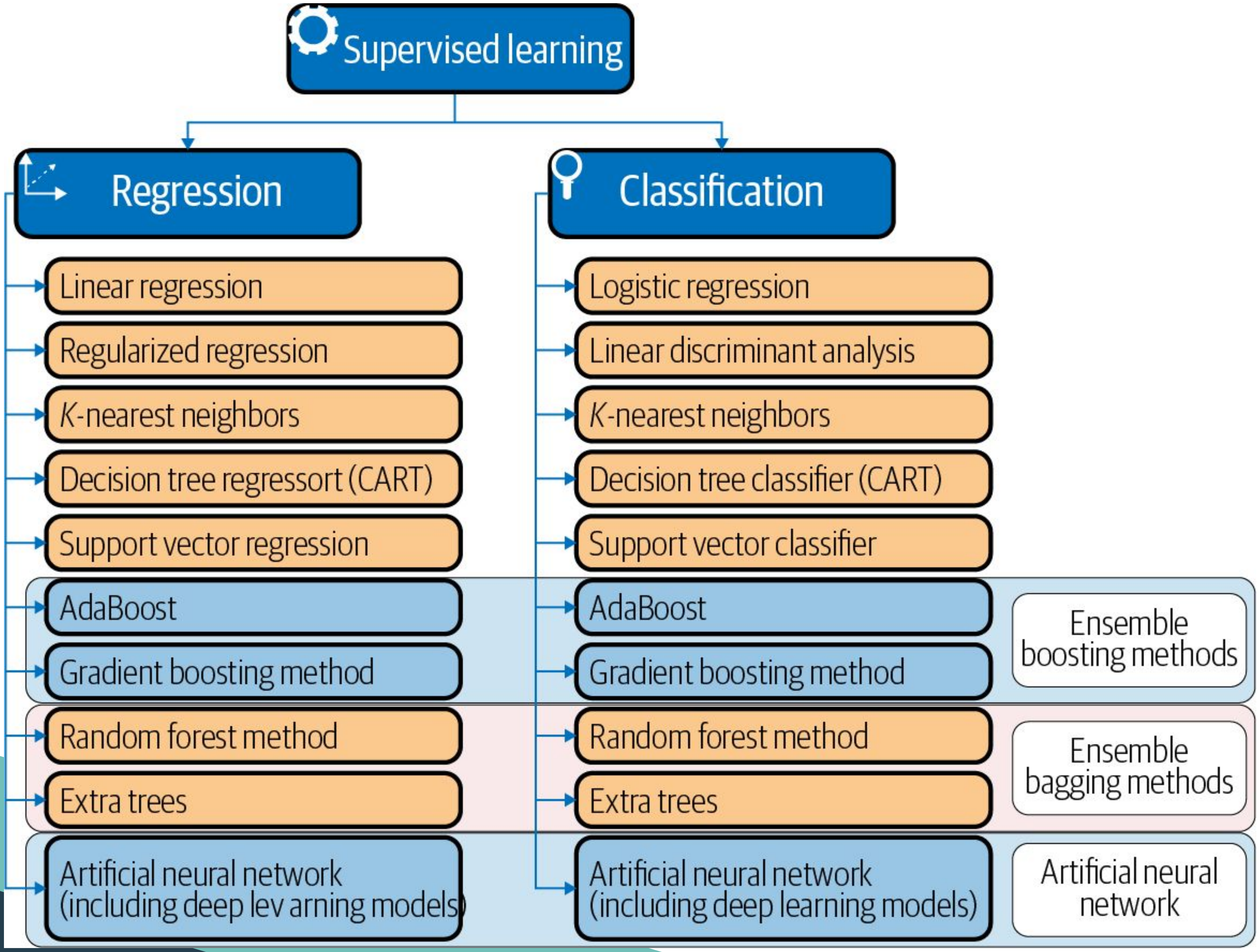
Supervised-Regression

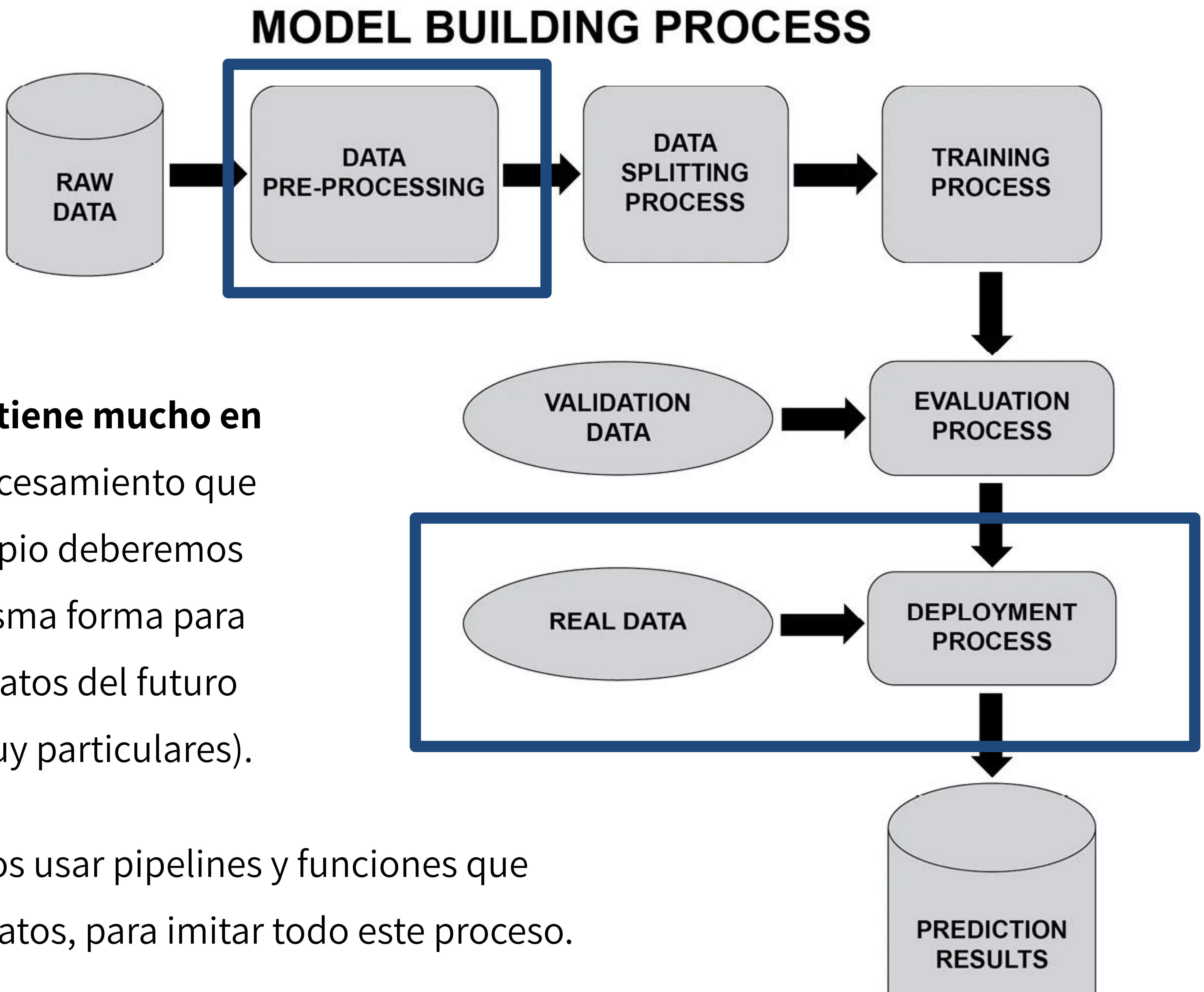


Supervised-Metrics



Supervised Models

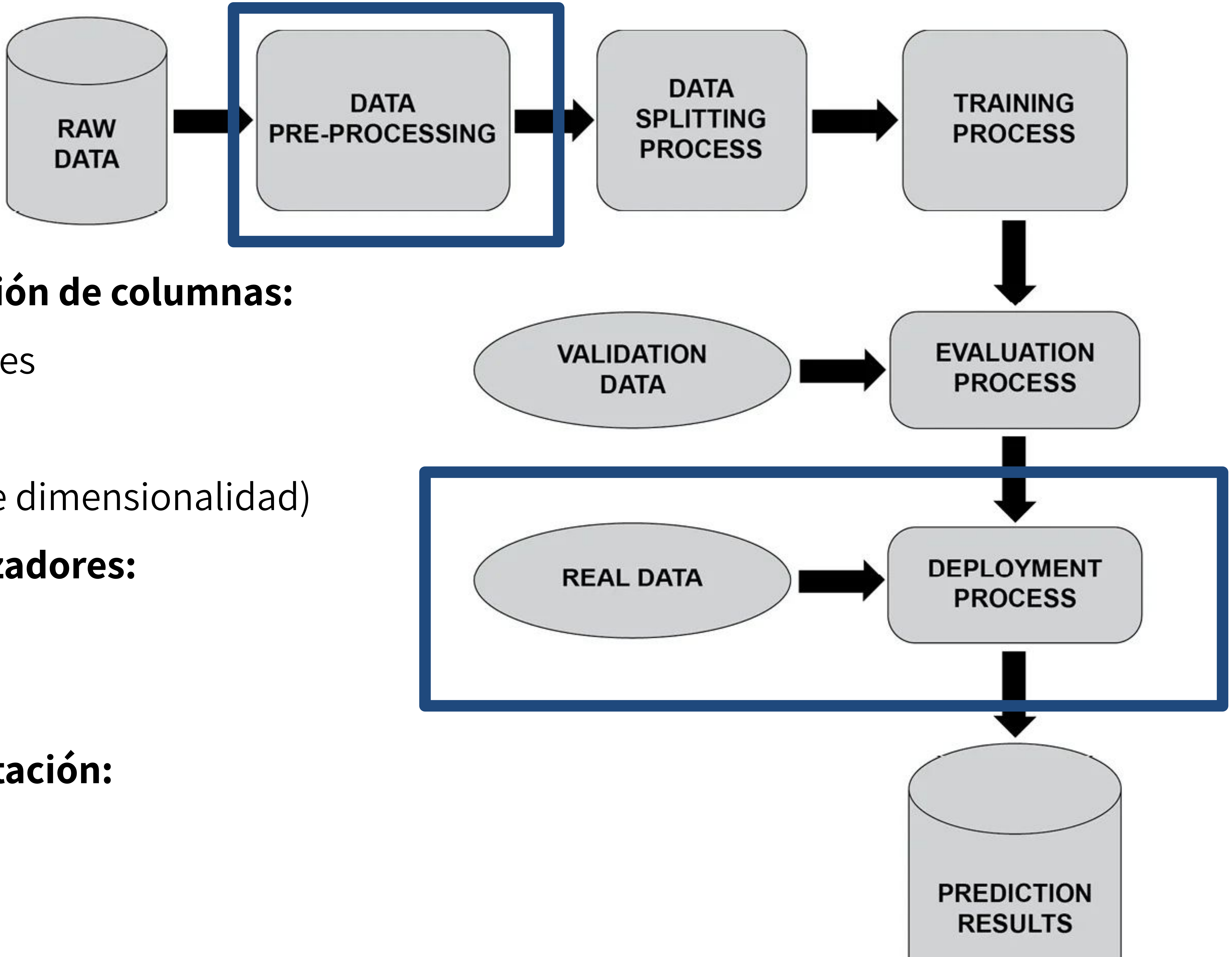




Estas dos partes tiene mucho en común. El preprocesamiento que hacemos al principio deberemos aplicarlo de la misma forma para cualquier set de datos del futuro (excepto casos muy particulares).

Por eso intentamos usar pipelines y funciones que transforman los datos, para imitar todo este proceso.

MODEL BUILDING PROCESS



Objetos de creación de columnas:

PolynomialFeatures

OneHotEncoder

PCA (reducción de dimensionalidad)

Objetos normalizadores:

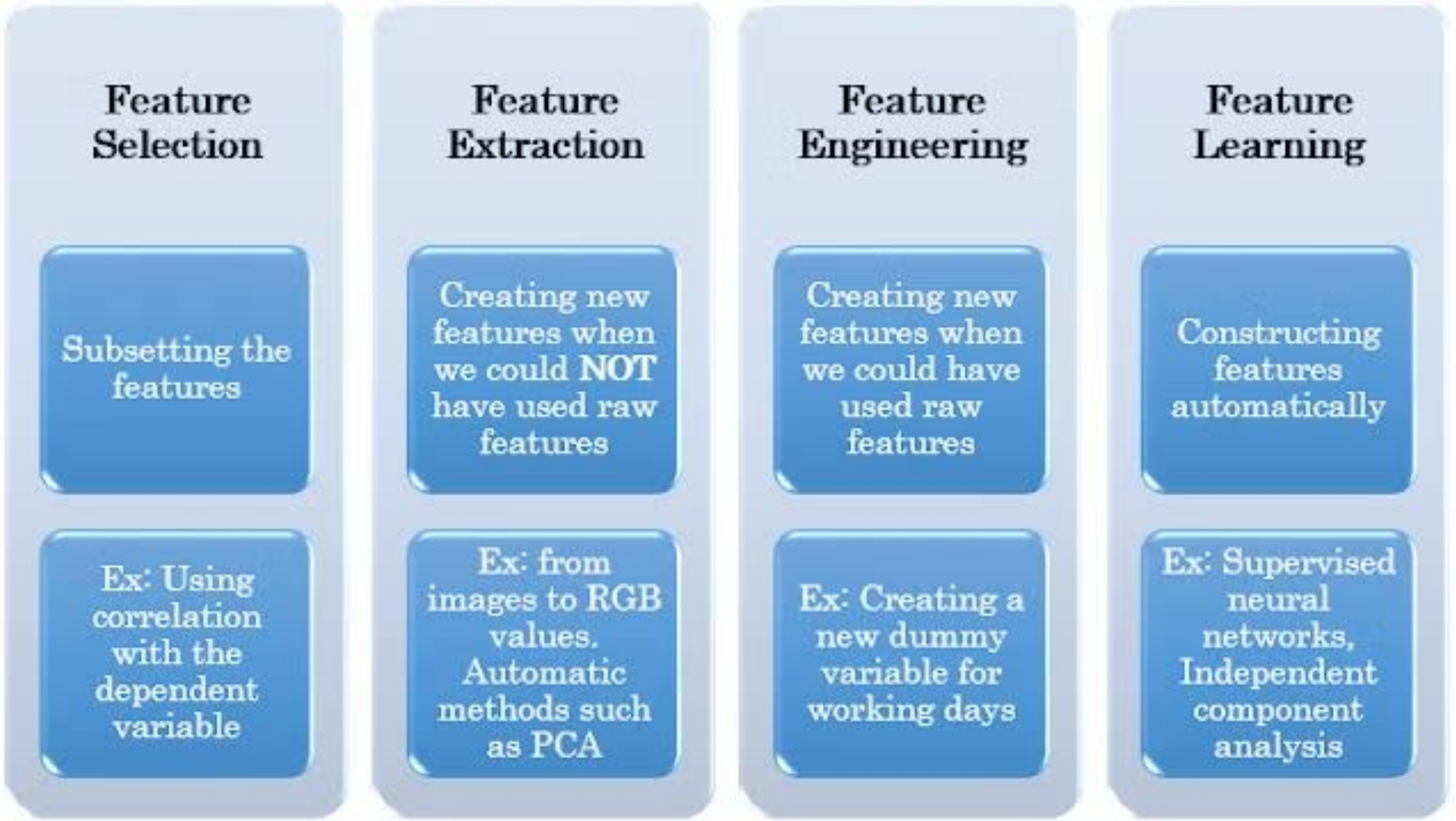
StandardScaler

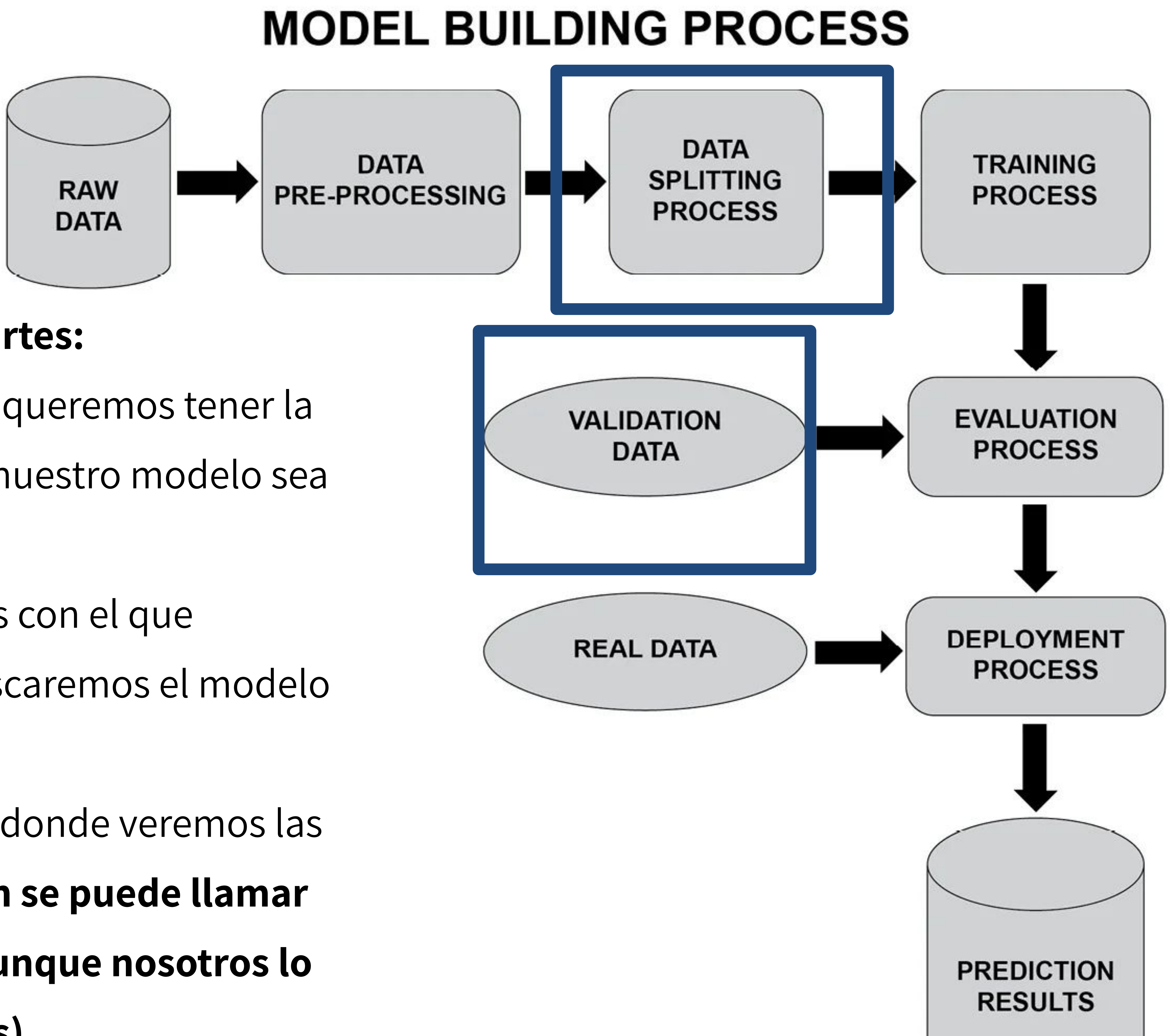
MinMaxScaler

Objetos de imputación:

SimpleImputer

KNNImputer


Feature Selection
Subsetting the features
Ex: Using correlation with the dependent variable
Feature Extraction
Creating new features when we could NOT have used raw features
Ex: from images to RGB values. Automatic methods such as PCA
Feature Engineering
Creating new features when we could have used raw features
Ex: Creating a new dummy variable for working days
Feature Learning
Constructing features automatically
Ex: Supervised neural networks, Independent component analysis



Dividimos en 2 partes:

¿Por que? Porque queremos tener la seguridad de que nuestro modelo sea estable.

Train: Set de datos con el que entrenaremos/buscaremos el modelo

Test: Set de datos donde veremos las métricas **(También se puede llamar validation data aunque nosotros lo llamamos al revés)**

MODEL BUILDING PROCESS

Training process:

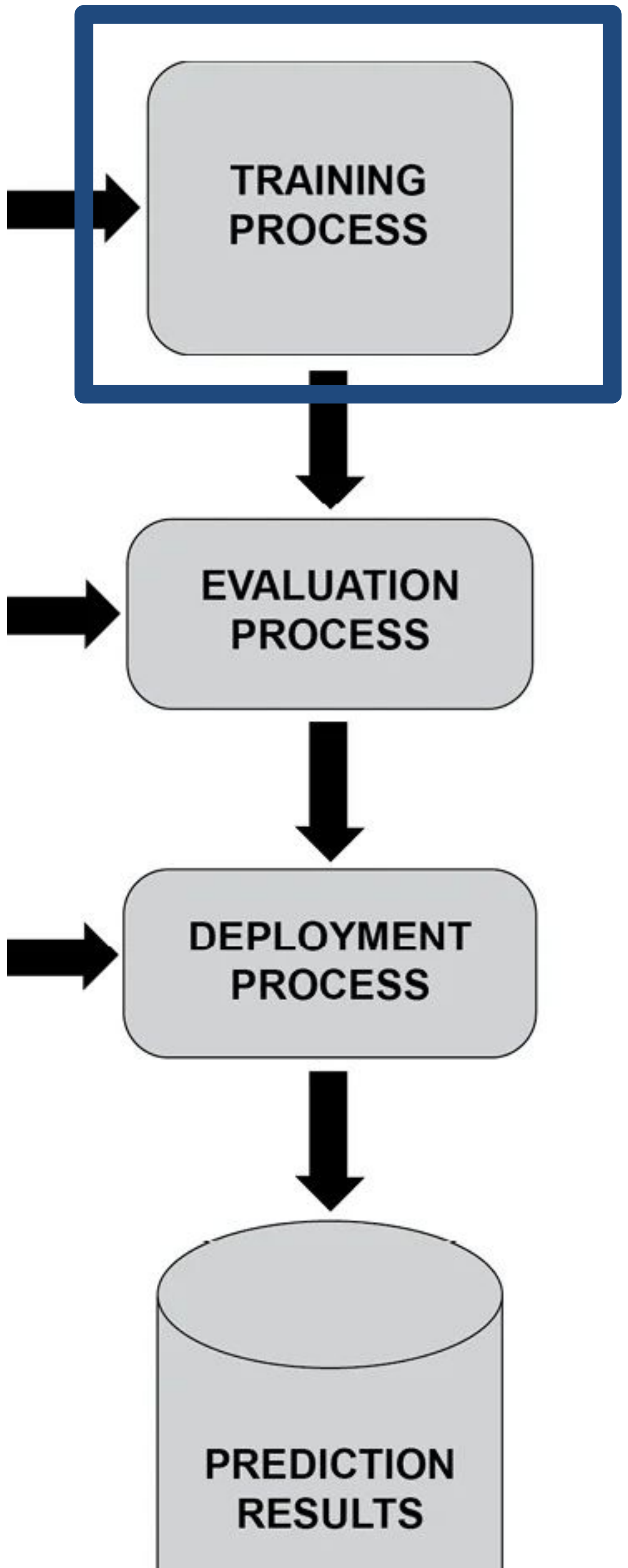
Aquí nosotros hacemos la elección de un tipo de modelo, dependiendo del tipo de problema. Asumimos que elegimos un XGBoost. **Entrenamos un modelo a partir de enseñarles unas X y una Y asociadas.**

Búsqueda de hiperparametros (get_params y grid_searchCV):

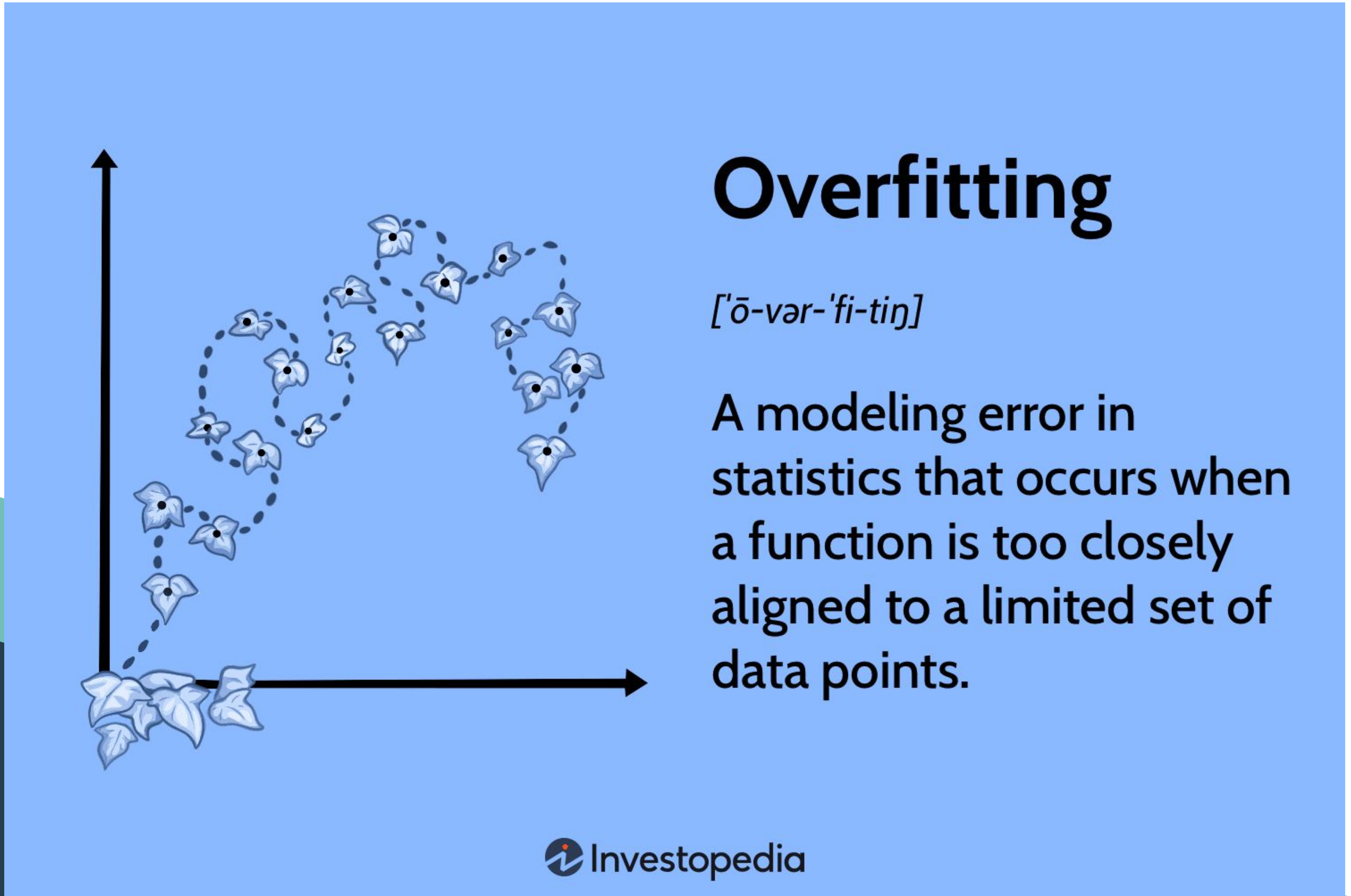
Definimos un conjunto de hiperparametros del modelo, como el max_depth. Probaremos cada combinación de candidatos y elegiremos la mejor

Elegimos un método de validación (grid_searchCV):

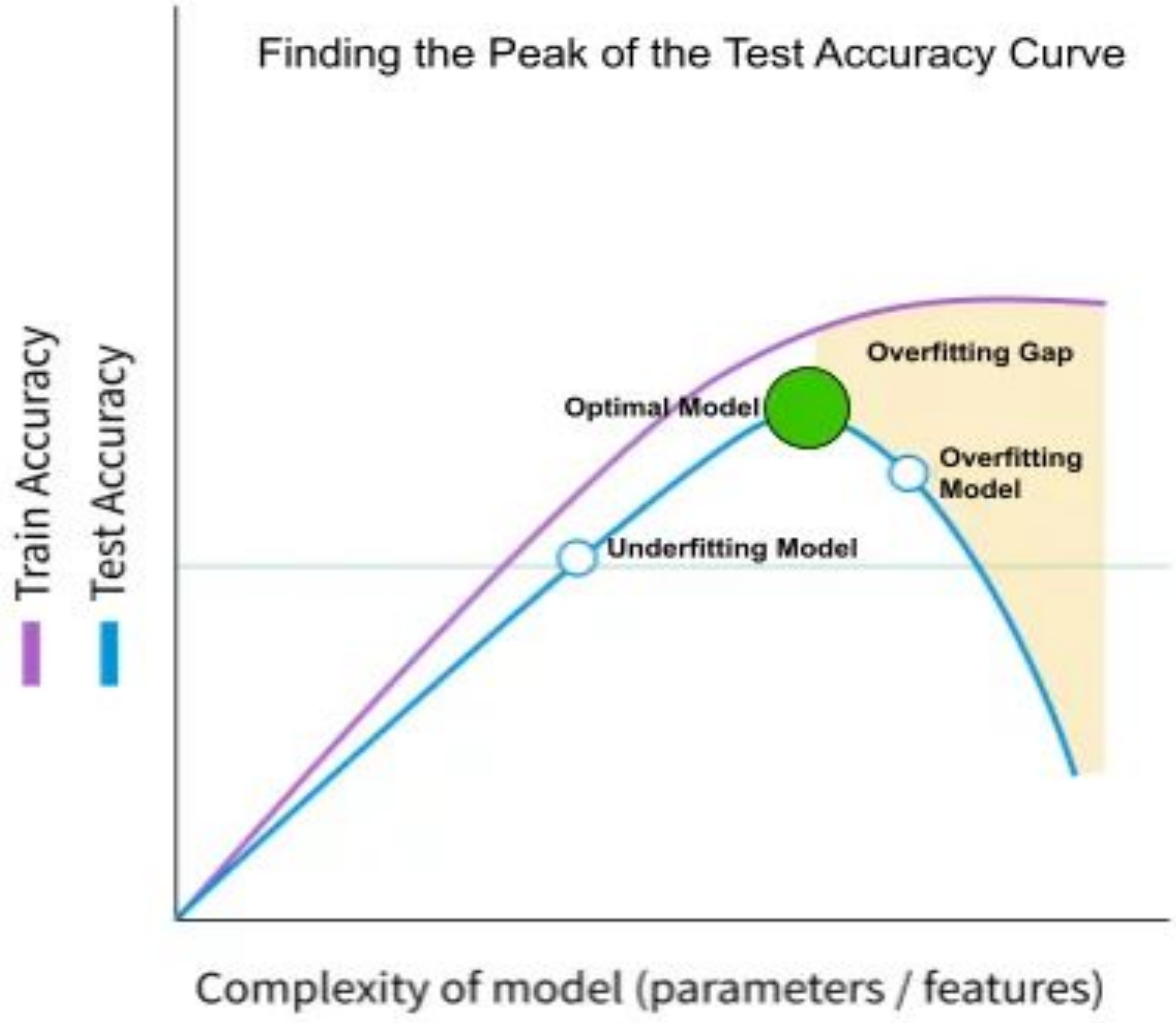
Cross-validation: En vez de tener una métrica de cada combinación tendremos K (5 normalmente, hacemos 5 modelos entrenando con 4 partes del train y testeando con 1 parte del train). Con estas K métricas elegimos el mejor modelo.

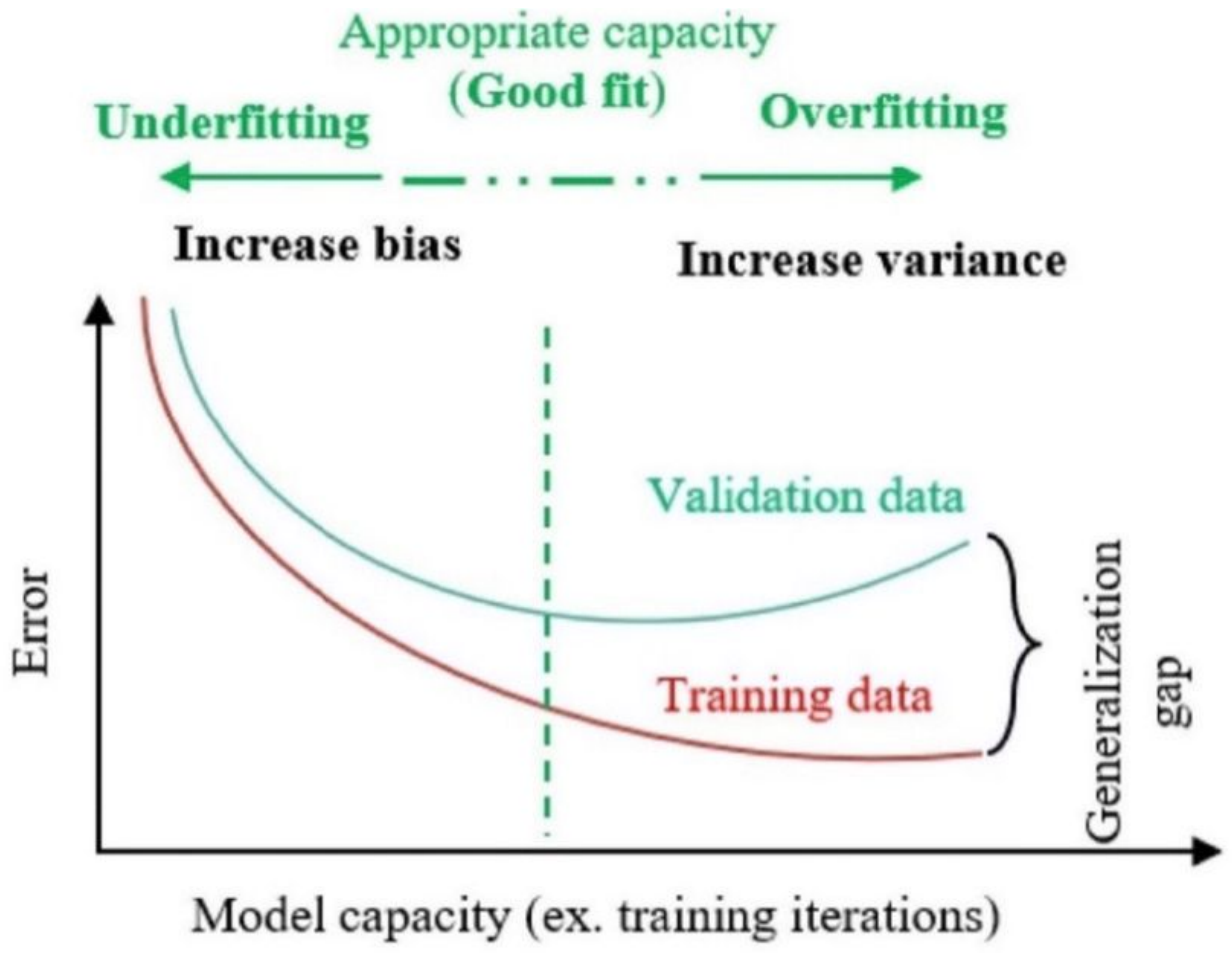


Overfitting



Overfitting

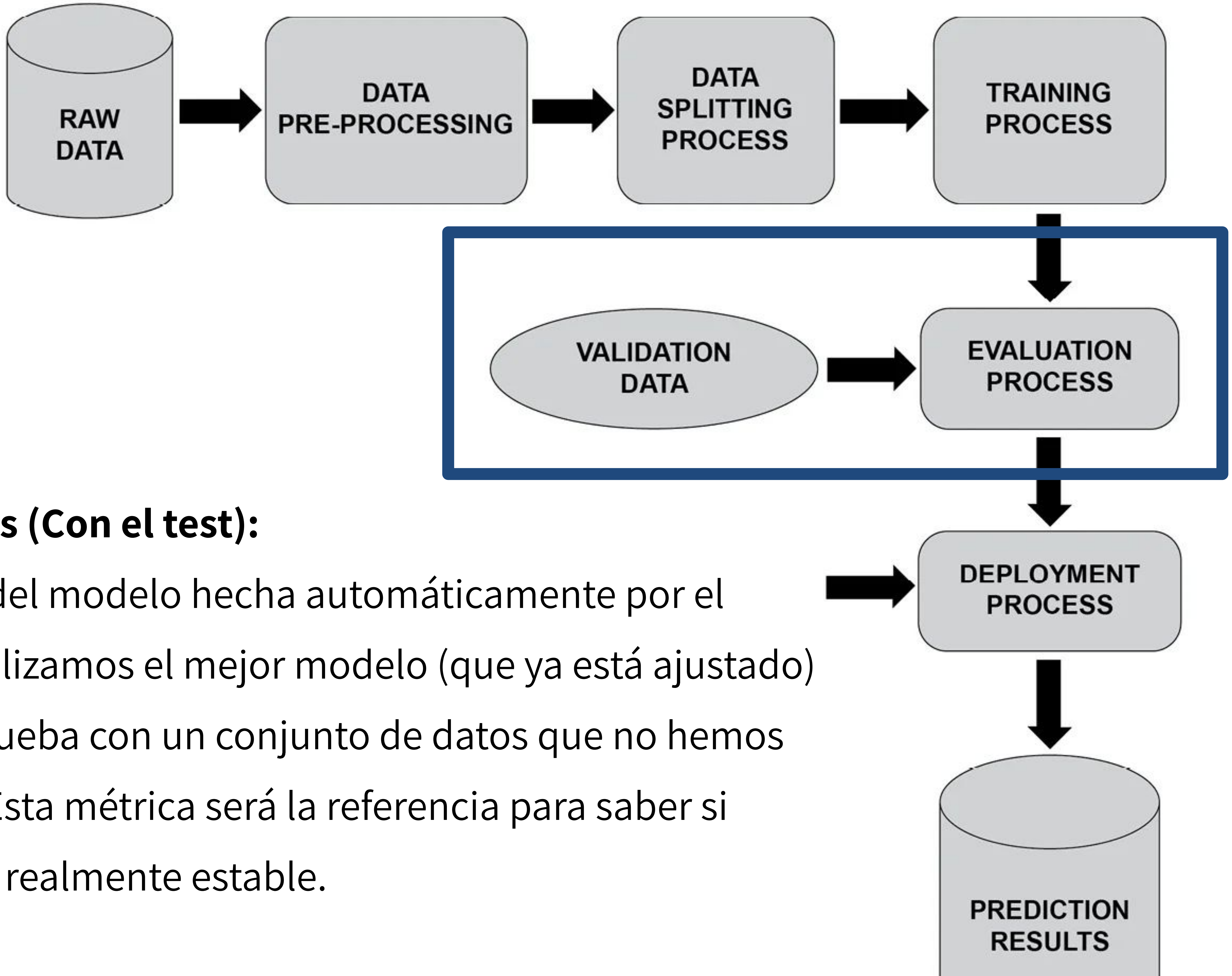




Training loss	Validation loss	Situation	Solution
High	High	Underfitting	Increase capacity
Low	High	Overfitting	Decrease capacity (shrink or regularization techniques)
Low	Low	Good fit	Run test
High	Low	Unlikely	Debug

Avoid Overfitting

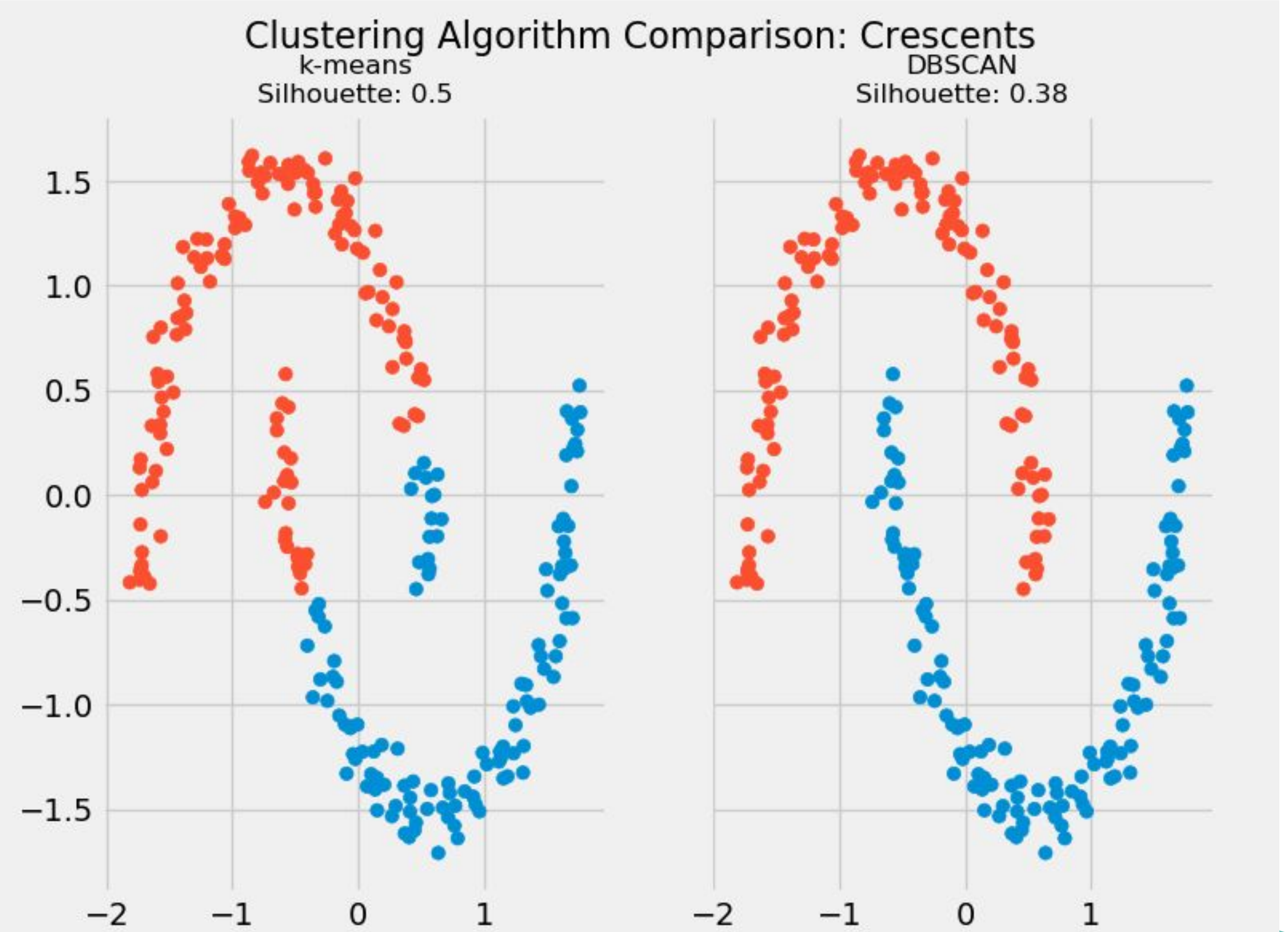
MODEL BUILDING PROCESS



Evaluation process (Con el test):

Ya con la elección del modelo hecha automáticamente por el **grid_searchCV**, utilizamos el mejor modelo (que ya está ajustado) para una última prueba con un conjunto de datos que no hemos utilizado todavía. Esta métrica será la referencia para saber si nuestro modelo es realmente estable.

Unsupervised-Models



Unsupervised-Models