

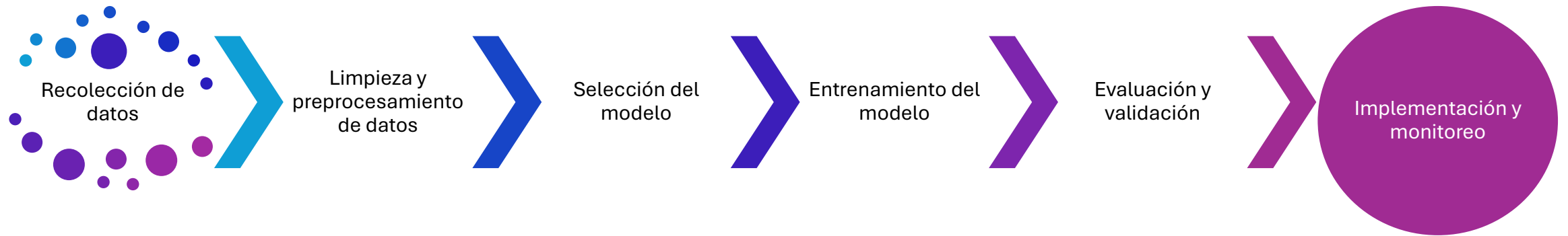
# Machine Learning



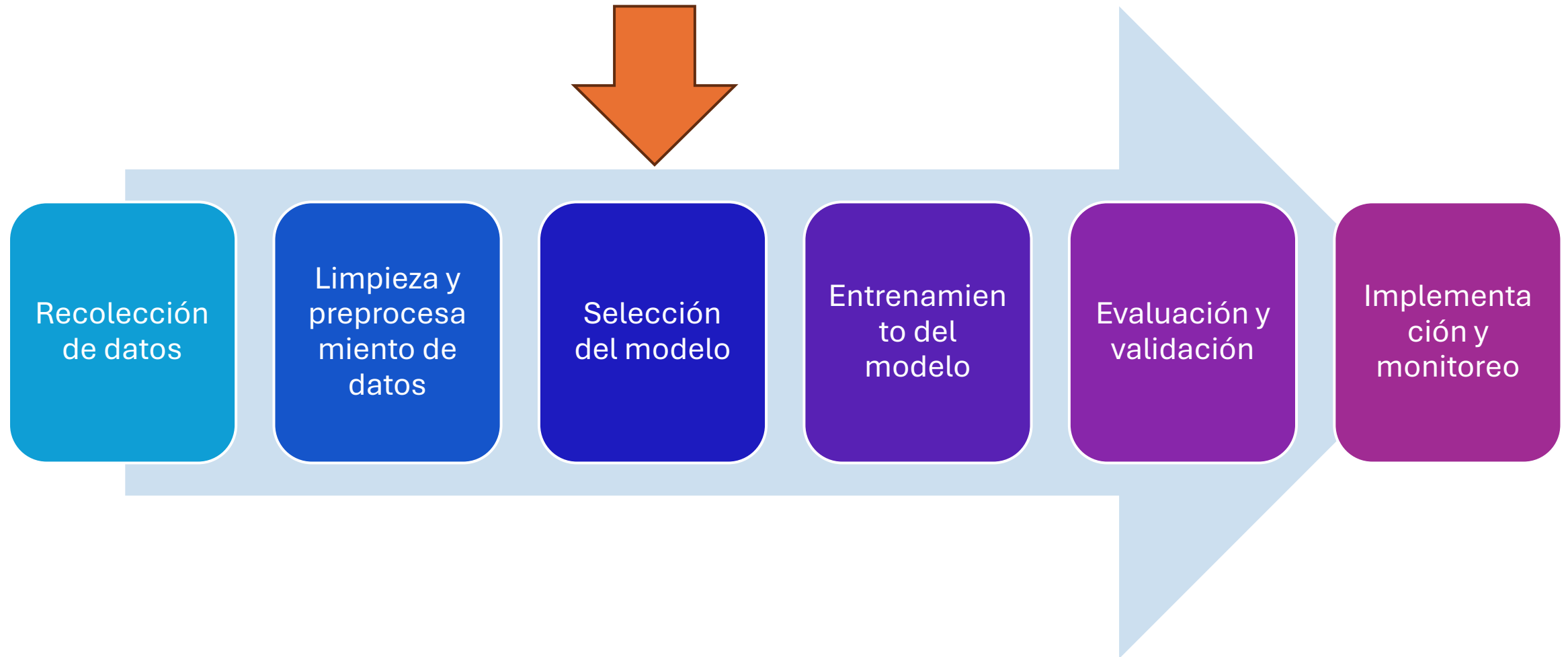
**Susana Medina Gordillo**

[susana.medina@correounivalle.edu.co](mailto:susana.medina@correounivalle.edu.co)

# Flujo de trabajo en Machine Learning



# Flujo de trabajo en Machine Learning



# **Introducción al Aprendizaje Supervisado**

# Introducción al Aprendizaje Supervisado

*Definición y Casos de Uso*

*Métricas de Evaluación para Clasificación y Regresión*

# ¿Qué es el Aprendizaje Supervisado?

El aprendizaje supervisado es un **tipo de aprendizaje automático** donde el algoritmo aprende a partir de un conjunto de datos **etiquetados**.

"**Etiquetado**" significa que conocemos la respuesta correcta (la "etiqueta") para cada ejemplo en los datos de entrenamiento.

El **objetivo** es que el algoritmo aprenda a mapear las entradas a las salidas correctas, de manera que pueda **predecir la etiqueta** para datos nuevos y desconocidos.



# **Tipos de Problemas de Aprendizaje Supervisado**

# Clasificación

Identificar si un correo electrónico es spam o no.

Clasificar imágenes de animales (perro, gato, pájaro)

Diagnosticar enfermedades a partir de síntomas

El objetivo es **asignar una etiqueta** a una **entrada**, eligiéndola de un conjunto de categorías predefinidas.



# Regresión

Predecir el precio de una casa en función de sus características.

Pronosticar la temperatura del día siguiente.

Estimar las ventas de un producto.

El objetivo es predecir un **valor numérico continuo**

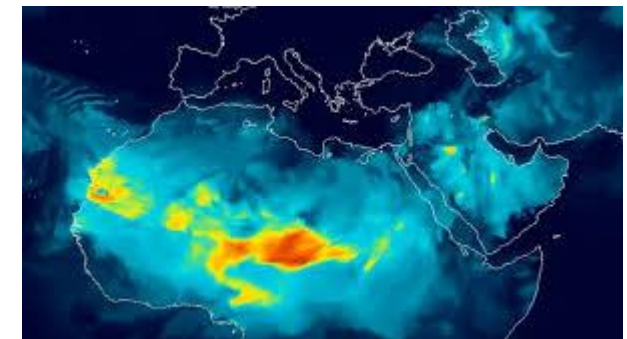
# Clasificación: Casos de uso

- ✓ **Medicina:** Diagnóstico de enfermedades, detección de riesgo de readmisión hospitalaria.
- ✓ **Finanzas:** Evaluación de riesgo crediticio, detección de fraude.
- ✓ **Marketing:** Segmentación de clientes, recomendación de productos.



# Regresión: Casos de uso

- ✓ **Economía:** Predicción de indicadores económicos, análisis de tendencias.
- ✓ **Ingeniería:** Predicción de fallas en equipos, optimización de procesos.
- ✓ **Medio Ambiente:** Predicción de la calidad del aire, modelado del cambio climático.



# Métricas de Evaluación para Clasificación

## Exactitud (Accuracy)

- Proporción de predicciones correctas.

## Precisión (Precision)

- Proporción de verdaderos positivos entre los positivos predichos.

## Recall (Sensibilidad)

- Proporción de verdaderos positivos entre los positivos reales.

## F1-score

- Media armónica de precisión y recall.

## Curva ROC y AUC

- Muestran el rendimiento del clasificador a diferentes umbrales.

# Métricas de Evaluación para Regresión

**Error Cuadrático Medio  
(MSE)**

- Promedio de los errores al cuadrado.

**Raíz del Error Cuadrático  
Medio (RMSE)**

- Raíz cuadrada del MSE.

**Error Absoluto Medio  
(MAE)**

- Promedio de los errores absolutos.

**R-cuadrado ( $R^2$ )**

- Proporción de la varianza explicada por el modelo.



# scikit-learn

Machine Learning in Python

Getting Started

Release Highlights for 1.6

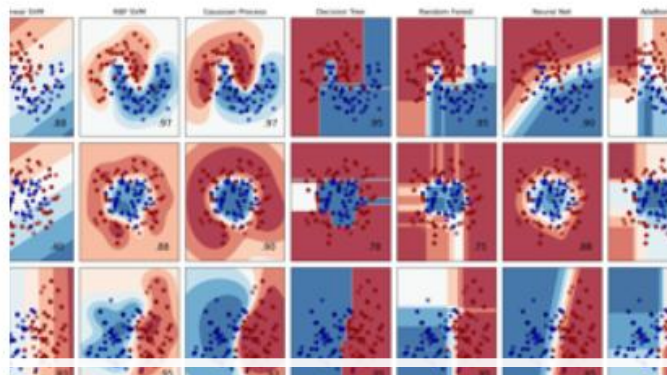
- Simple and efficient tools for predictive data analysis
- Accessible to everybody, and reusable in various contexts
- Built on NumPy, SciPy, and matplotlib
- Open source, commercially usable - BSD license

## Classification

Identifying which category an object belongs to.

**Applications:** Spam detection, image recognition.

**Algorithms:** [Gradient boosting](#), [nearest neighbors](#), [random forest](#), [logistic regression](#), and [more...](#)



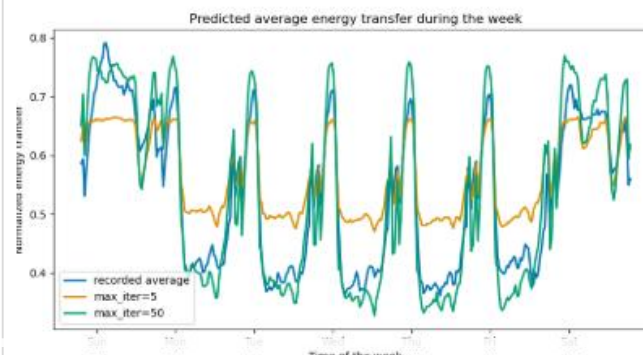
Examples

## Regression

Predicting a continuous-valued attribute associated with an object.

**Applications:** Drug response, stock prices.

**Algorithms:** [Gradient boosting](#), [nearest neighbors](#), [random forest](#), [ridge](#), and [more...](#)



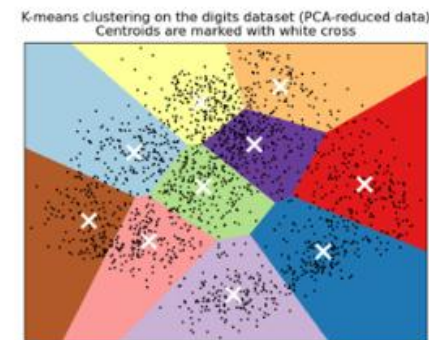
Examples

## Clustering

Automatic grouping of similar objects into sets.

**Applications:** Customer segmentation, grouping experiment outcomes.

**Algorithms:** [k-Means](#), [HDBSCAN](#), [hierarchical clustering](#), and [more...](#)



Examples

## Dimensionality reduction

Reducing the number of random variables to consider.

**Applications:** Visualization, increased efficiency.

## Model selection

Comparing, validating and choosing parameters and models.

## Preprocessing

Feature extraction and normalization.

**Applications:** Transforming input data such as text for



# Consideraciones Importantes

La elección de la métrica de evaluación depende del problema específico y de los objetivos del modelo.

Es importante evaluar el modelo en datos nuevos y no vistos durante el entrenamiento para asegurar que generalice bien.

A menudo, es útil utilizar múltiples métricas para obtener una visión completa del rendimiento del modelo.

# Conclusiones...

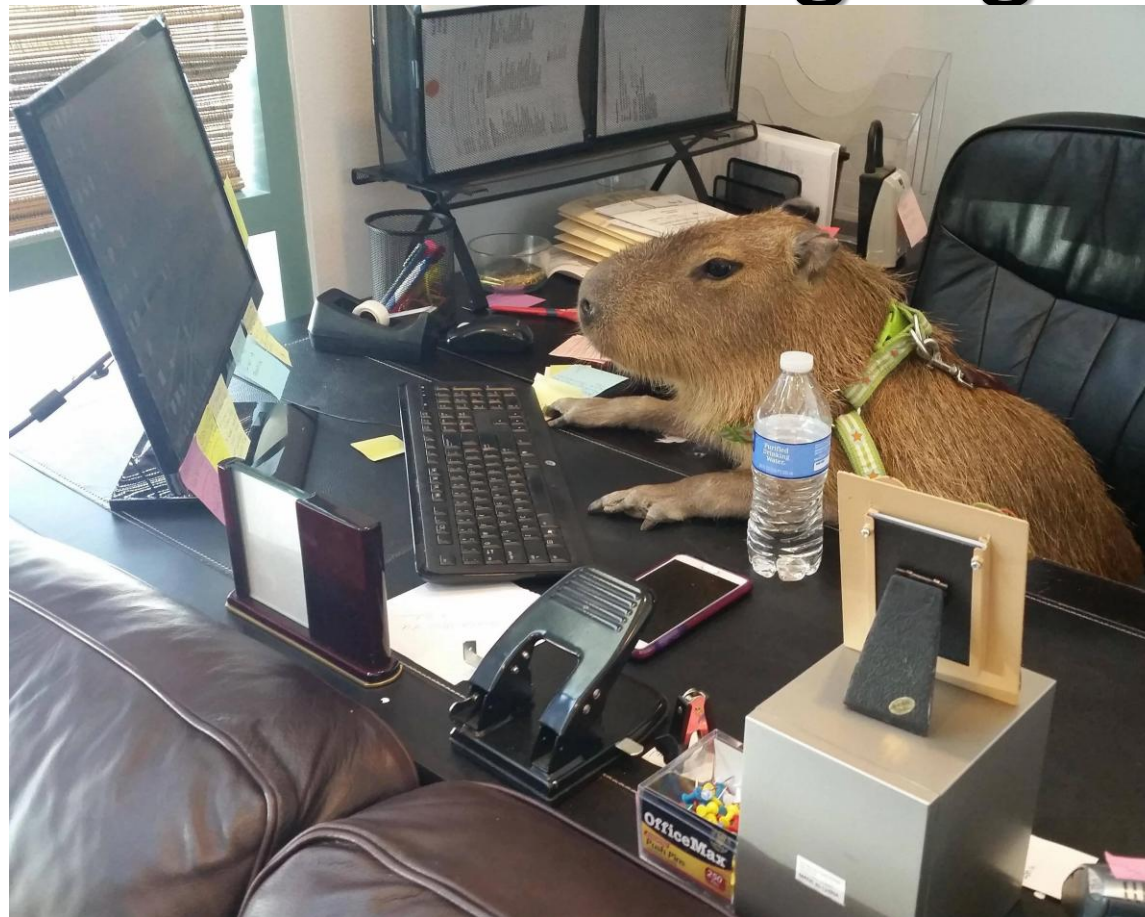
El aprendizaje supervisado es una herramienta poderosa para resolver problemas de clasificación y regresión.

Comprender las métricas de evaluación es fundamental para seleccionar y optimizar modelos.

El aprendizaje supervisado tiene una amplia gama de aplicaciones en diversos campos de la ingeniería informática.

# Ejercicio práctico

[colab.research.google.com](https://colab.research.google.com)



# Ejercicio práctico: Google Colaboratory (*Colabs*)

- Página oficial: <https://colab.google/>
- Abrir Colab (incluye tutorial): <https://colab.research.google.com/>
- Guía para EDA: [https://colab.research.google.com/github/Tanu-N-Prabhu/Python/blob/master/Exploratory\\_data\\_Analysis.ipynb](https://colab.research.google.com/github/Tanu-N-Prabhu/Python/blob/master/Exploratory_data_Analysis.ipynb)
- Guía / tutorial para Selección de características con **scikit-learn**: <https://www.datacamp.com/tutorial/feature-selection-python>

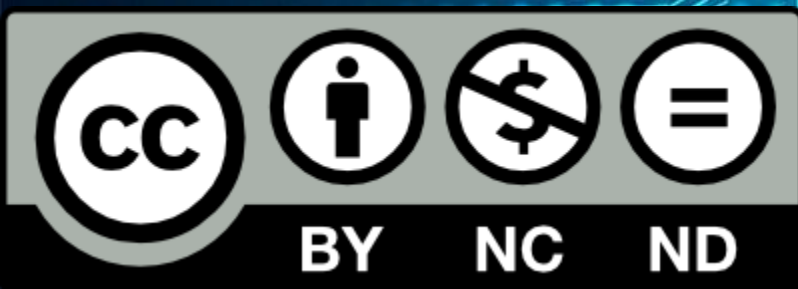


# Referencias

- “scikit-learn: Machine Learning in Python”. Consultado: el 20 de febrero de 2025. [En línea]. Disponible en: <https://scikit-learn.org/stable/>
- Imagen de Features. Consultado: el 20 de febrero de 2025. [En línea]. Disponible en: <https://themanoftalent.medium.com/feature-selection-9b1609f1f6b0>
- Imagen de Transforming variables. Consultado: el 20 de febrero de 2025. [En línea]. Disponible en: <https://www.datasklr.com/ols-least-squares-regression/transforming-variables>
- “A Probabilistic Algorithm to Reduce Dimensions: t — Distributed Stochastic Neighbor Embedding (t-SNE)”. Consultado: el 20 de febrero de 2025. [En línea]. Disponible en: <https://pub.towardsai.net/a-probabilistic-algorithm-to-reduce-dimensions-t-distributed-stochastic-neighbor-embedding-23ff457fbc8a>
- Información e ideas presentadas basadas en el conocimiento general de modelos de lenguaje de IA. Gemini 2.9 Flash. Consultado: el 20 de febrero de 2025. [En línea].



# Machine Learning



**Susana Medina Gordillo**

[susana.medina@correounivalle.edu.co](mailto:susana.medina@correounivalle.edu.co)



