



# Programa de Especialización:

## CIENCIA DE DATOS Y ANALISIS DE CONFIABILIDAD APLICADO AL SECTOR ELÉCTRICO

Octubre, 2025

# Docente:



## Angel Huamán Sarzo

Ingeniero Electricista | Magister en Gerencia e Ingeniería de Mantenimiento | Experto en Analítica y Ciencia de Datos aplicado al mantenimiento | Especialista en Evaluación del Mantenimiento en Red de Energía del Perú

LinkedIn





## CONTENIDO DEL PROGRAMA

MÓDULO I Machine learning aplicado al sector eléctrico (26H)

MÓDULO II Análisis de Confiabilidad aplicado al sector eléctrico (16H)

MÓDULO III Power BI aplicado al sector eléctrico (16H)

# Introducción a la ciencia de datos

## ¿Qué es y que no es ciencia de datos?

- Automatizar procesos o reportes
- Hacer cálculos determinísticos en Python o R
- Realizar gráficos avanzados usando Python o R
- Administrar y tratar grandes bases de datos
- Elaborar reportes descriptivos o dashboards sin generar valor predictivo
- Hacer análisis estadísticos o correlacional entre variables
- Combinar estadística, matemáticas, programación, conocimiento del negocio e interpretación de resultados.
- Realizar recolección, limpieza, análisis, modelado e interpretación de datos.
- Extraer conocimiento, patrones y predicciones para la toma de decisiones.
- Utilizar aprendizaje automático, minería de datos, estadística avanzada y visualización para entender datos complejos.
- Aplicar métodos reproducibles y rigurosos para validar resultados (ej. validación cruzada, métricas).



# Introducción a la ciencia de datos

## ¿Dónde puedo aplicar la Ciencia de Datos?

- ✓ Tesis de grado/Maestría/Doctorados
- ✓ Proyectos de Innovación aplicados en la industria eléctrica
- ✓ investigaciones para revistas indexadas
- ✓ Estudios o investigaciones Particulares

### Ejemplos:

- ✓ Predicción de la demanda energética
- ✓ Predicción del precio de la energía
- ✓ Detección de fraudes de energía
- ✓ Mantenimiento predictivo de equipos
- ✓ Diagnóstico de fallas en equipos transformadores, interruptores y demás equipos de potencia
- ✓ Otros

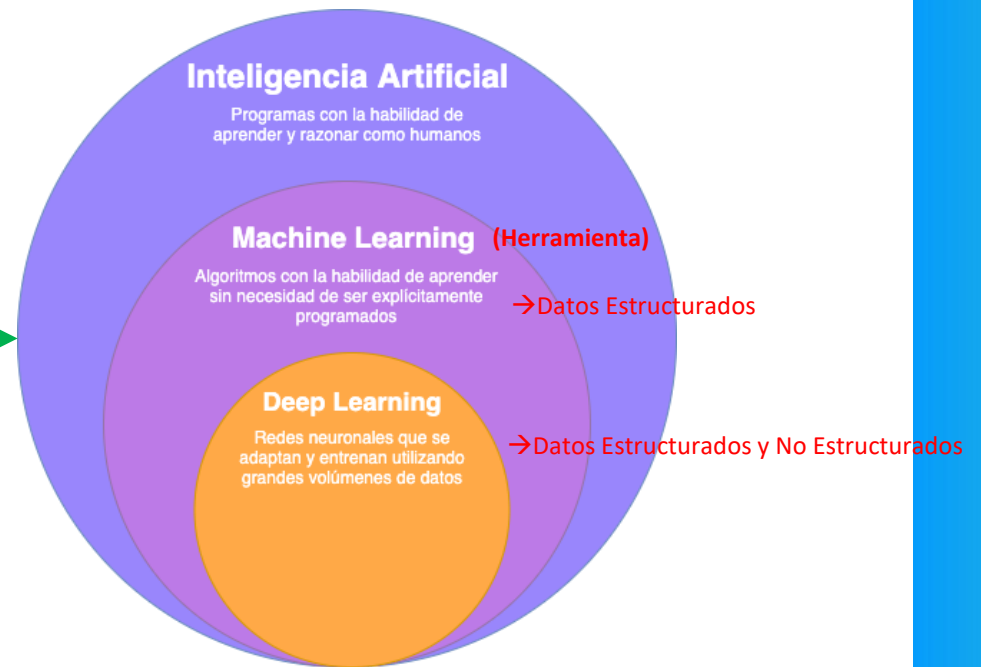
## ¿Qué perfil debo tener?

- ✓ Estudiante de Ingeniería
- ✓ Egresado/bachiller
- ✓ Ingeniero junior/Seniors
- ✓ Interés por aprender y querer aplicar la ciencia de datos en el sector eléctrico o a fines

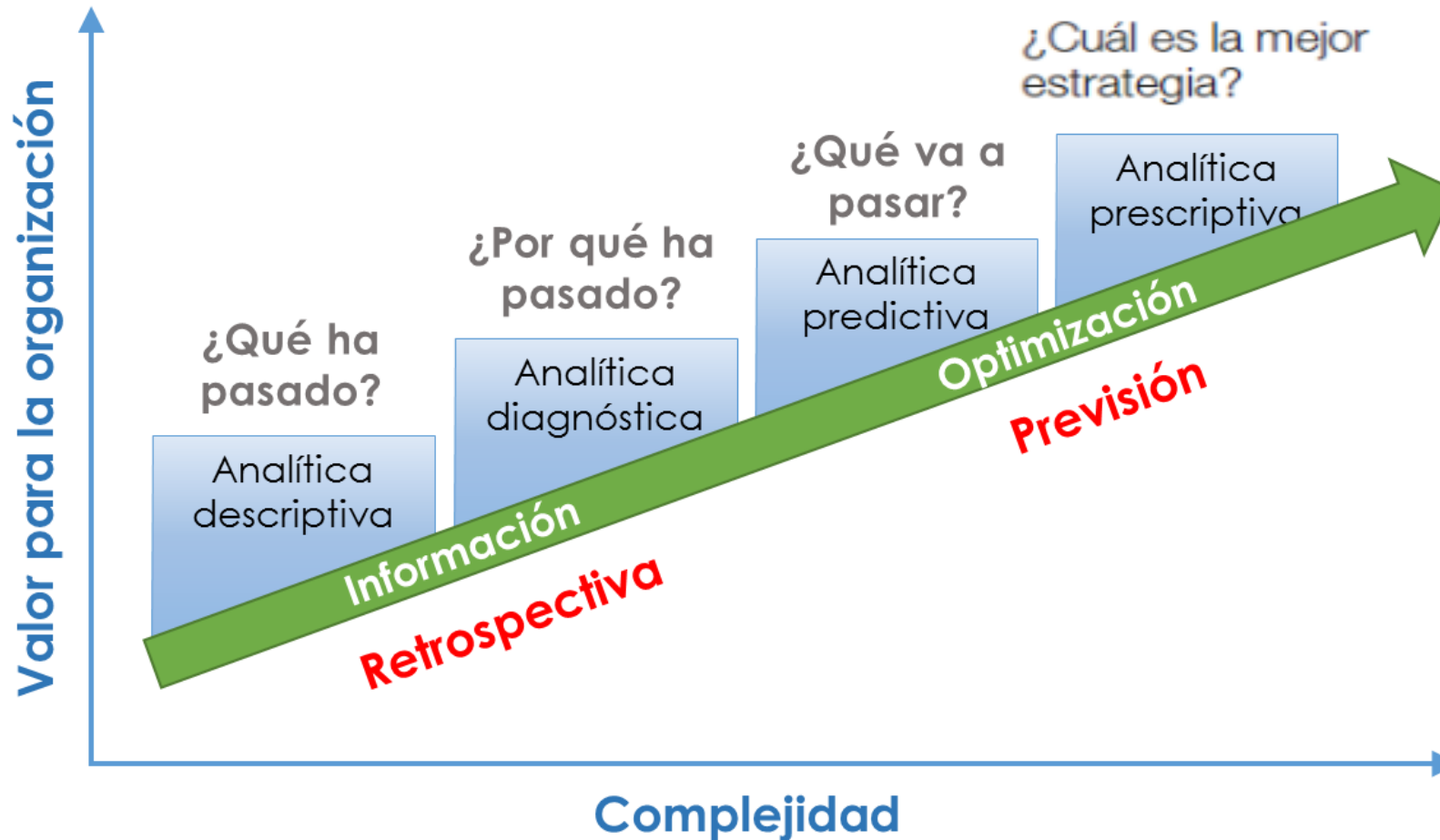


# Introducción a la ciencia de datos

## ¿Qué es Machine Learning?

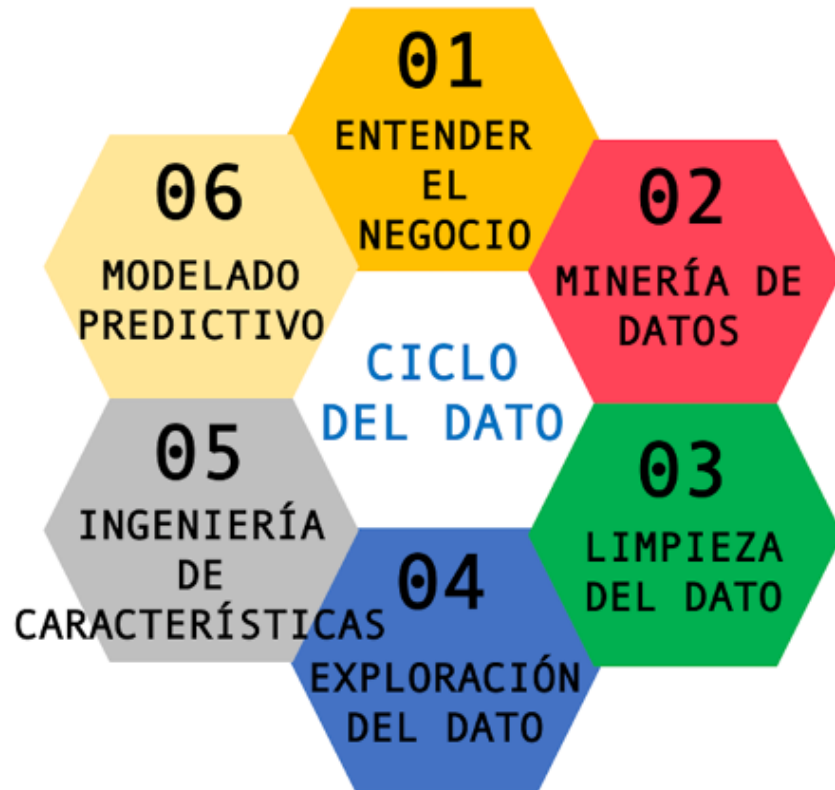


## **Del Dato a la Acción: “El Camino hacia la Analítica Avanzada”**



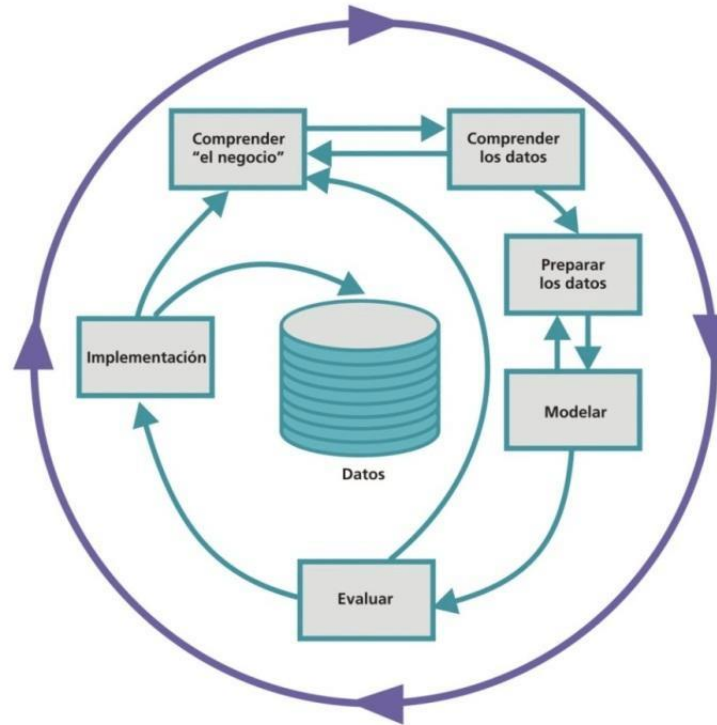
# Introducción a la ciencia de datos

## **Ciclo del Dato: Etapas del Proceso en Ciencia de Datos**



## **¿Existe alguna metodología?**

Metodología CRIS DM  
(Cross-Industry Standard Process for Data Mining)

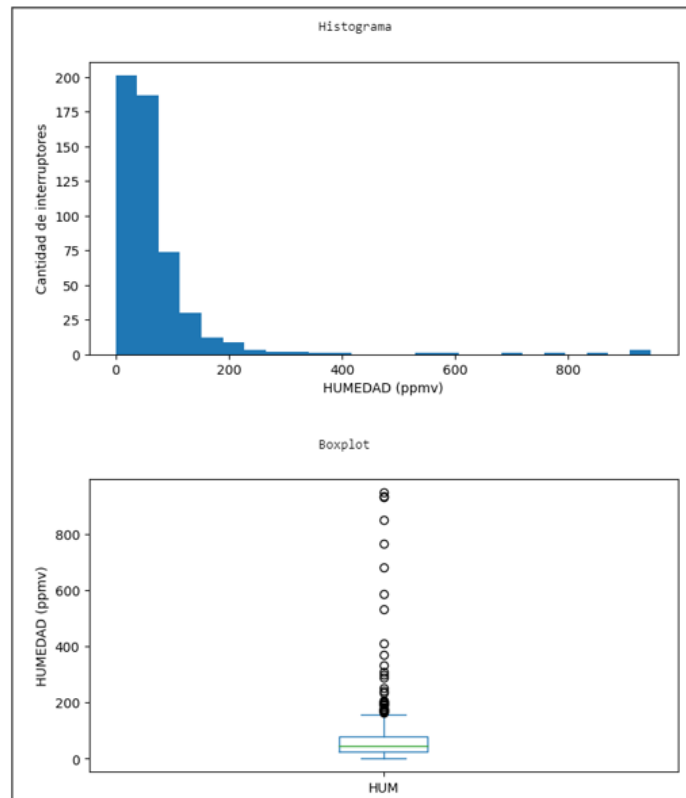


\*Es una metodología estándar y probada para guiar los proyectos de ciencia de datos, desde la comprensión del negocio hasta el despliegue final de los resultados

## Ingeniería de datos y análisis estadístico con Python

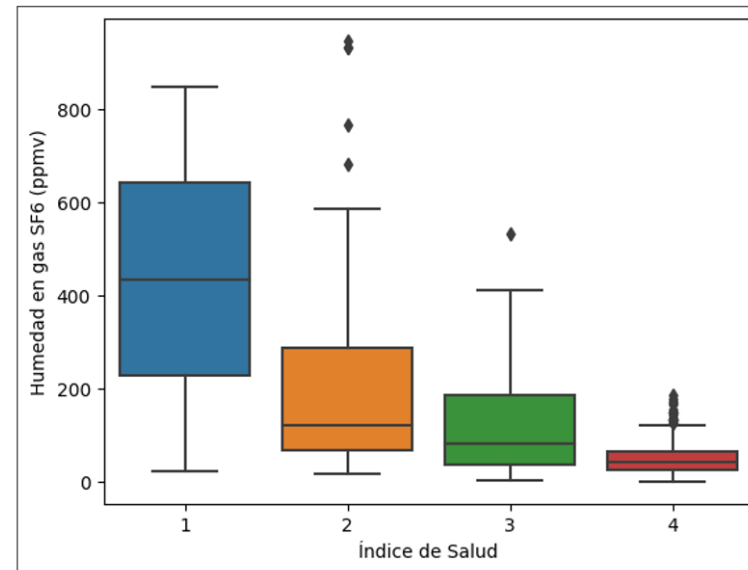
### 1. Análisis Univariante

Figura 23. Histograma y Boxplot de Variable: Humedad



### 2. Análisis Bivariante

Figura 30.. Analisis Bivariante: "Humedad" vs "Índice de Salud"



## Ingeniería de datos y análisis estadístico con Python

### 3. Análisis de matriz de correlaciones

Figura 35. Matriz de Correlaciones con el metodo de Spearman



### 4. Determinación de variables principales

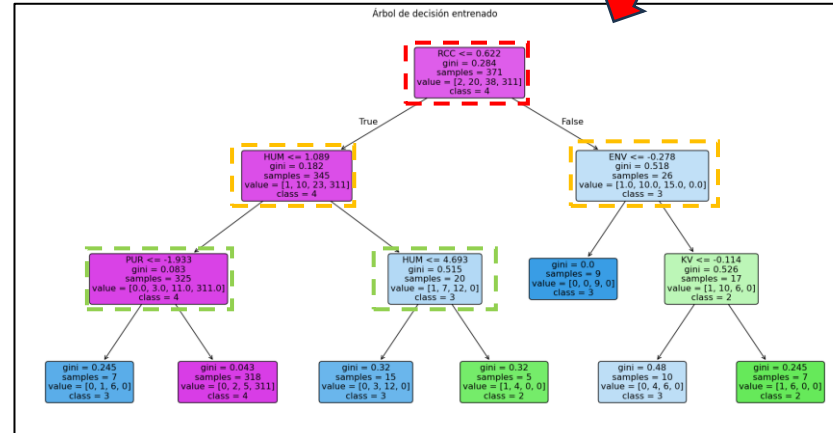
Correlaciones entre variables predictoras y la variable objetivo:

```
HUM 0.522575
RCC 0.478248
PUR 0.390355
FAB 0.286841
ENV 0.233256
SO2 0.190911
KV 0.106007
dtype: float64
```

Importancia de variables:

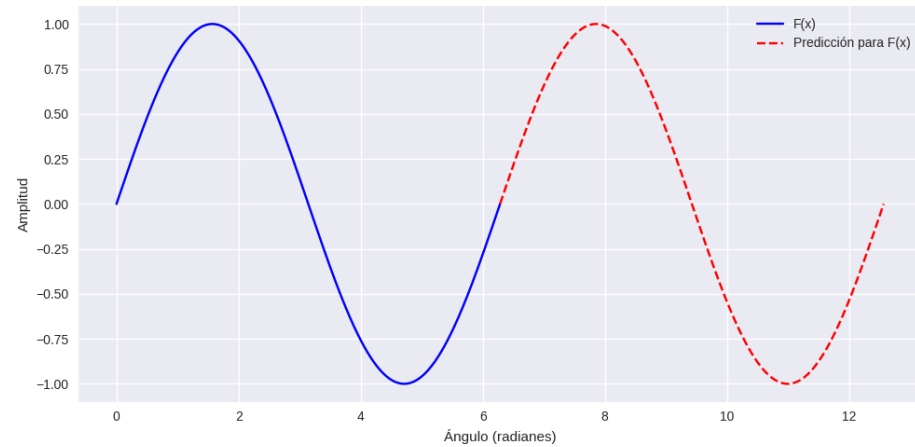
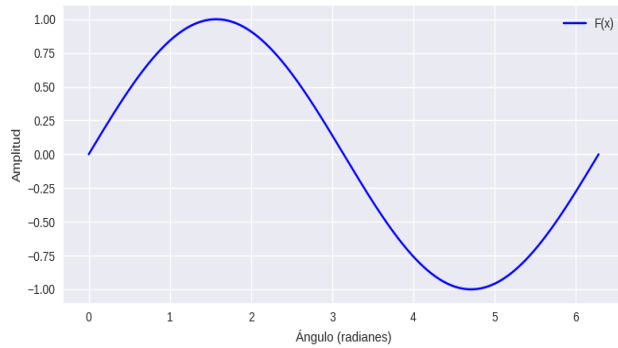
	variable	importancia	importancia_acumulada
5	RCC	0.358130	0.358130
2	HUM	0.321078	0.679208
3	PUR	0.121163	0.800371
4	SO2	0.073428	0.873799
6	ENV	0.059756	0.933555
1	FAB	0.048685	0.982240
0	KV	0.017760	1.000000

Árbol de decisión entrenado



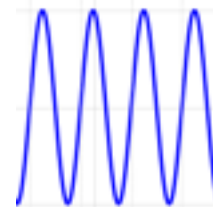
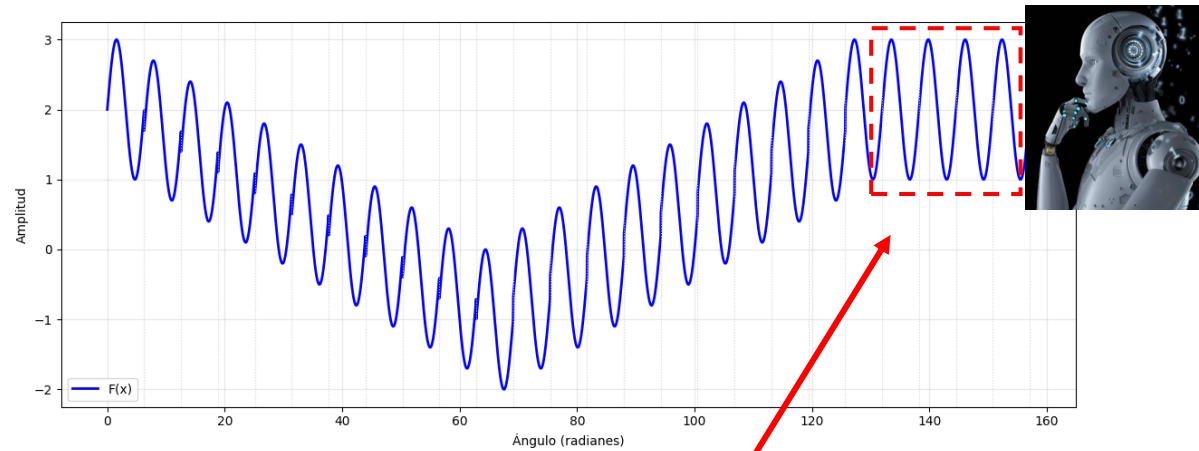
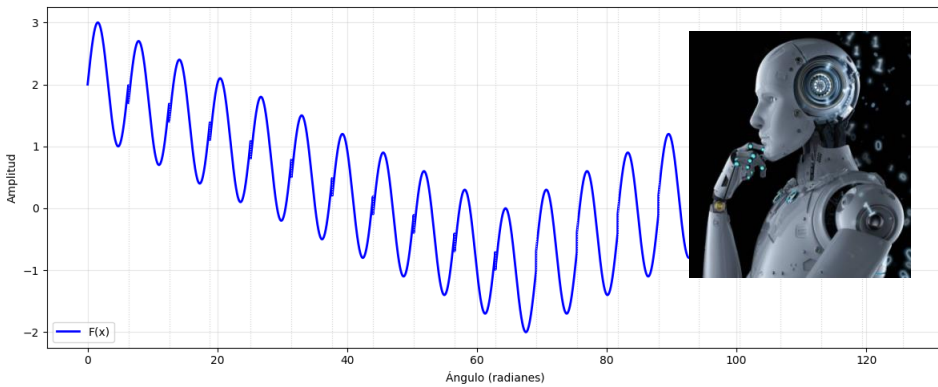
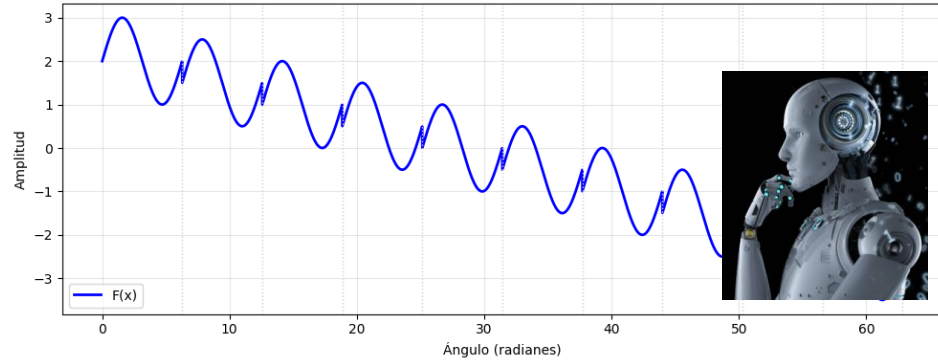
# Introducción a la ciencia de datos

## ¿Como funciona el algoritmo de un modelo predictivo?



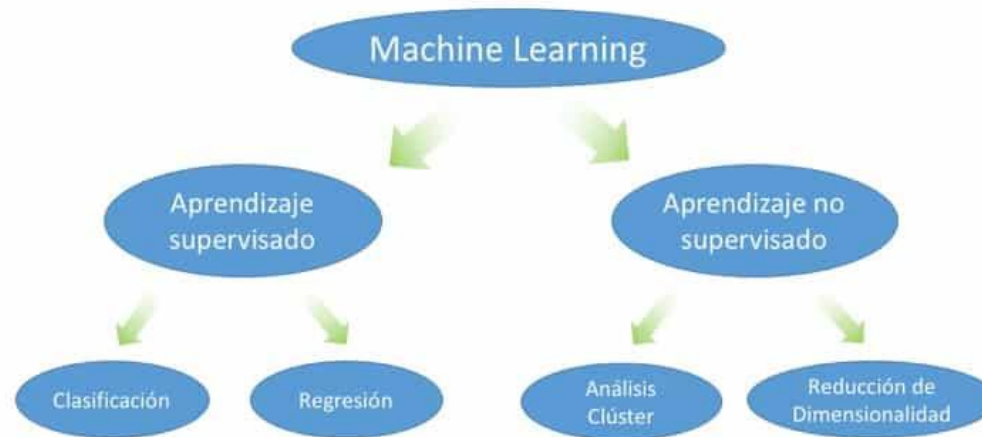
# Introducción a la ciencia de datos

## Importancia de los datos de entrenamiento en modelos predictivos

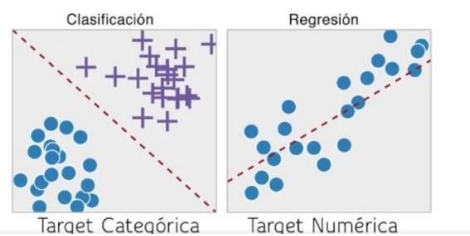


# Introducción a la ciencia de datos

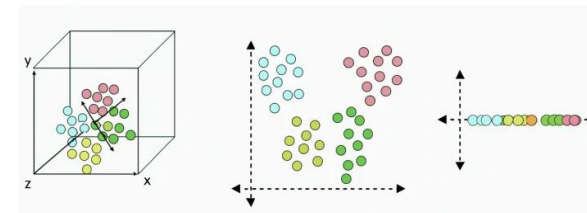
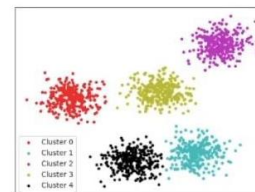
## Tipos de aprendizaje en Machine Learning



Se conoce la variable target



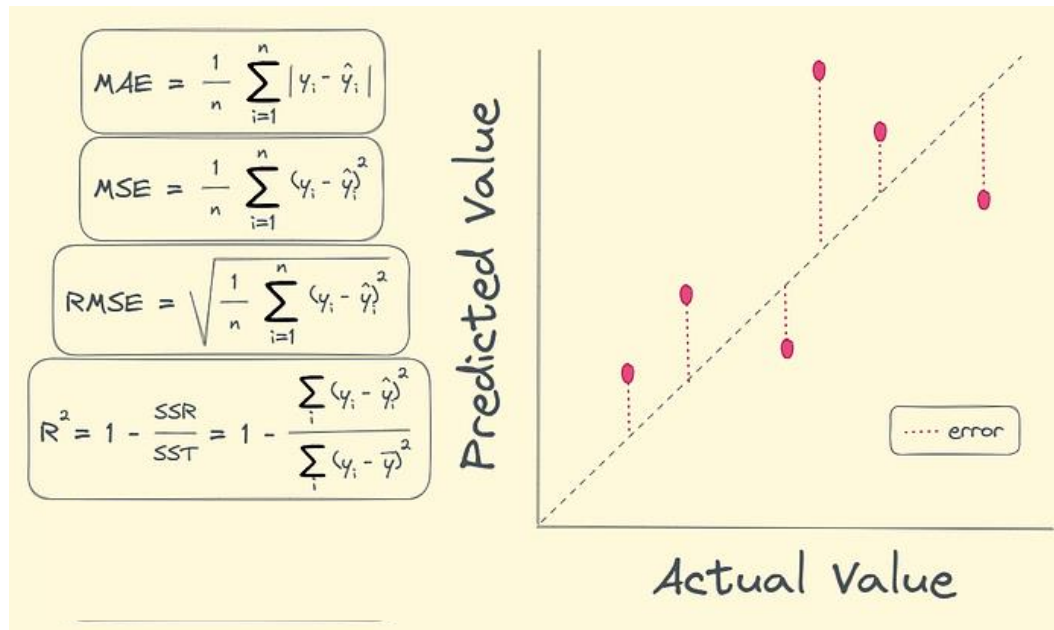
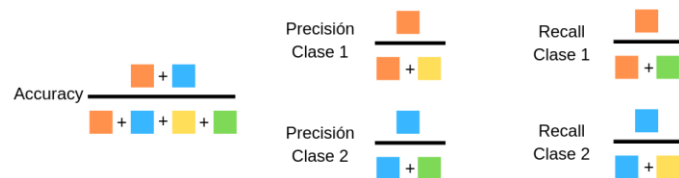
No se conoce la variable target



# Introducción a la ciencia de datos

## ¿ Como medimos la efectividad de un modelo Predictivo?

	Predicción Clase 1	Predicción Clase 2
Valor real Clase 1	Aciertos True Positive Clase 1	Fallos False Positive Clase 2
Valor real Clase 2	Fallos False Positive Clase 1	Aciertos True Positive Clase 2



# Introducción a la ciencia de datos

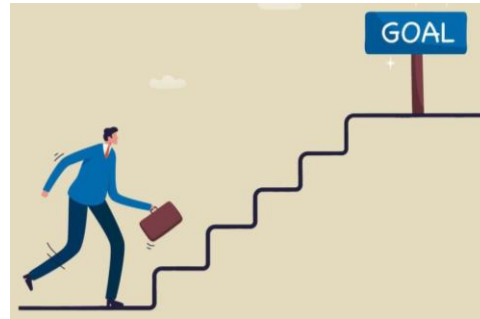
## ¿Que lograras al finalizar el programa?

- ✓ Dominar el uso de editores de código como "VS Code" y "Google Colab" aplicados a la ciencia de datos
- ✓ Conocer la estructura completa de un proyecto de ciencia de datos
- ✓ Desarrollar proyectos de ciencia de datos desde cero
- ✓ Comprender las métricas de efectividad de un modelo predictivo y saber como mejorar su rendimiento
- ✓ Auditar proyectos de ciencia de datos elaborados por terceros



# Introducción al mantenimiento y generalidades

## ¿Como se va lograr el objetivo?



1. Introducción al lenguaje Python (2H)
2. Ingeniería de datos y análisis estadístico con Python (4H)
  - ✓ Tratamiento de datos nulos y datos atípicos
  - ✓ Análisis Univariante de variables predictoras
  - ✓ Análisis Bivariante entre variables predictoras y variable objetivo
  - ✓ Análisis de matriz de correlaciones entre variables
  - ✓ Métodos para determinación de variables principales
3. Algoritmos de Machine Learning Supervisados (4H)
  - ✓ Clasificación: Binaria y Multiclase
  - ✓ Regresión: lineal, lineal múltiple y no lineal
4. Algoritmos de Machine Learning No Supervisados (4H)
  - ✓ Reducción de dimensionalidad
  - ✓ Clusterización y Jerarquización:
  - ✓ Detección de anomalías
5. Algoritmos de Deep Learning en Series de Tiempo (4H)
  - ✓ GRU (Gated Recurrent Unit)
  - ✓ LSTM (Long Short-Term Memory)
  - ✓ CNN (Convolutional Neural Network)
6. Técnicas de búsqueda de Hiperparametros (2H)
  - ✓ GridSearch
  - ✓ RandomizedSearch
  - ✓ Bayesian Optimization
7. Evaluación del desempeño de modelos predictivos (2H)
  - ✓ Matriz de confusión
  - ✓ Precisión, exactitud, sensibilidad y F1-Score,
  - ✓ ROC, AUC y Curva de aprendizaje
  - ✓ R2 Score, MAE (Mean Absolute Error), MSE (Mean Squared Error), RMSE (Root Mean Squared Error) y MBD (Mean Bias Deviation)
8. Proyecto integrador 1 (2H)
  - ✓ Desarrollo de Modelo predictivo con algoritmos supervisados para diagnosticar la condición de un transformador
9. Proyecto integrador 2 (2H)
  - ✓ Desarrollo de Modelo predictivo con algoritmos de series de tiempo para pronosticar la demanda eléctrica

# GRACIAS

