



Tema 3: **Conceptos básicos de Aprendizaje de Máquina**

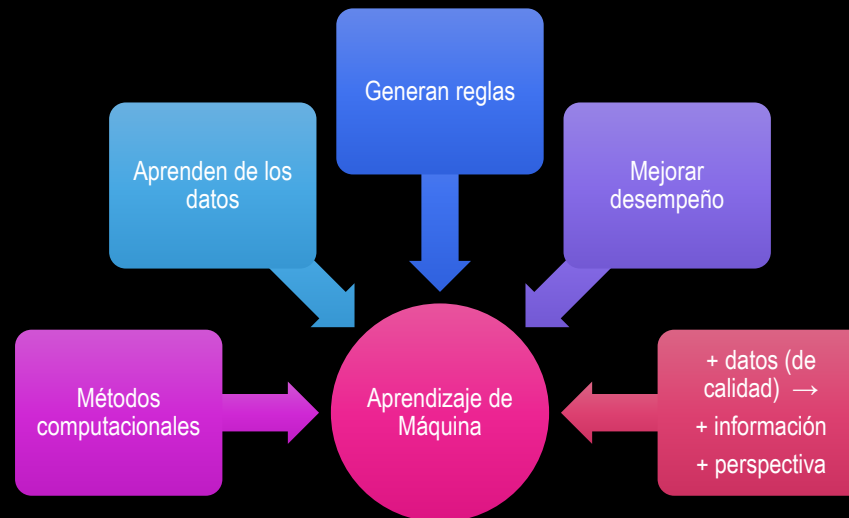
Parte 1

1. ¿Qué es Aprendizaje de Máquina?

TO PROVE YOU'RE A HUMAN,
CLICK ON ALL THE PHOTOS
THAT SHOW PLACES YOU
WOULD RUN FOR SHELTER
DURING A ROBOT UPRISING.



- **Métodos computacionales** para **aprender de datos** con el fin de **producir reglas para mejorar el desempeño** en alguna tarea o toma de decisión (González, 2018).
- Es una rama de la inteligencia artificial cuyo objetivo es **desarrollar técnicas que permitan a las computadoras aprender** (Mármol, 2018).
- Métodos computacionales que permiten que **un sistema, por sí mismo** —sin intervención humana y en forma automatizada—, **aprenda a descubrir patrones, tendencias y relaciones en los datos** y gracias a dicho conocimiento, **en cada interacción con datos nuevos se ofrecen mejores perspectivas** (SAS, 2020).



"Find a bug in a program, fix it, and the program will work today.
Show the program how to find and fix a bug, and the program will work forever."
- OliverG. Selfridge

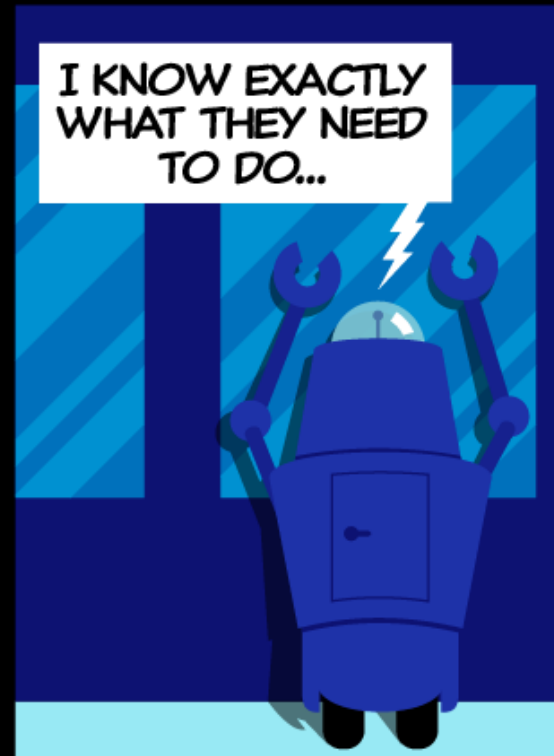
1. ¿Qué es Aprendizaje de Máquina?

¿Por qué usar ML?

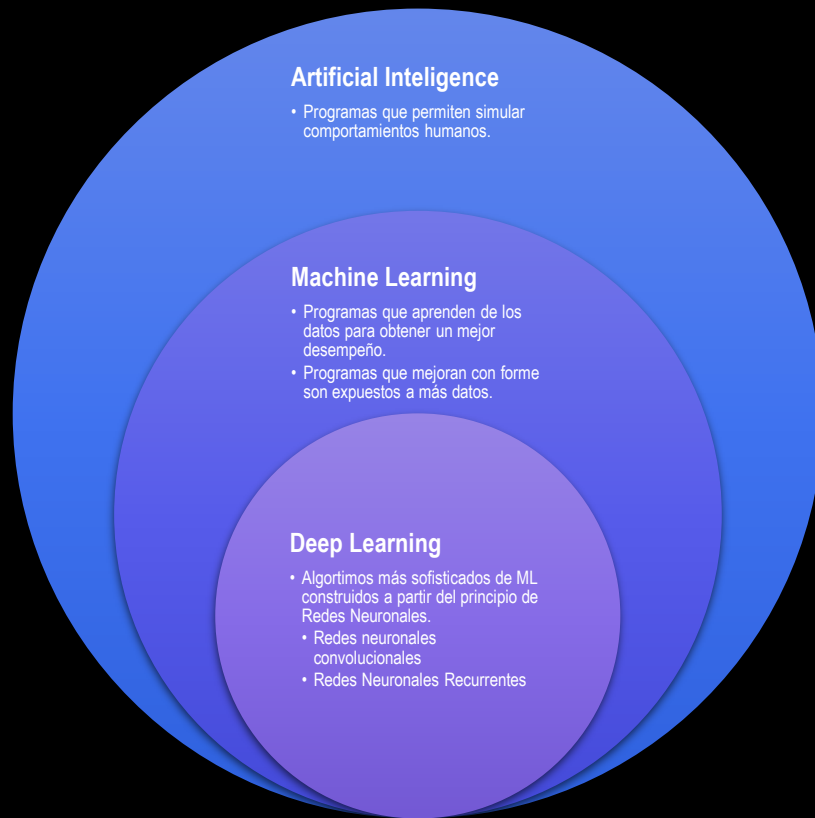
- Respuesta barata, rápida, automatizada, y con suficiente precisión.
- Superar el desempeño actual de los expertos o de reglas simples.
- Entender de manera más completa y sistemática el comportamiento de un fenómeno, identificando variables o patrones importantes.

Ejemplos (¿Dónde aplicar ML?):

- **Determinar** si dar un préstamo o no (algoritmos, en vez de evaluaciones personales).
- **Reconocer** placas de un coche de una fotografía.
- **Detectar** llamados de ballenas en grabaciones de boyas.*
- **Estimar** la demanda de productos de una compañía.
- **Dividir** a los clientes de Netflix según sus gustos.
- **Recomendar** productos a clientes.



1. ¿Qué es Aprendizaje de Máquina?



Inteligencia Artificial:

La computadora muestra comportamientos “inteligentes”.

La capacidad de un agente informático para simular los procesos cognitivos que habitualmente se asocian a la inteligencia humana.

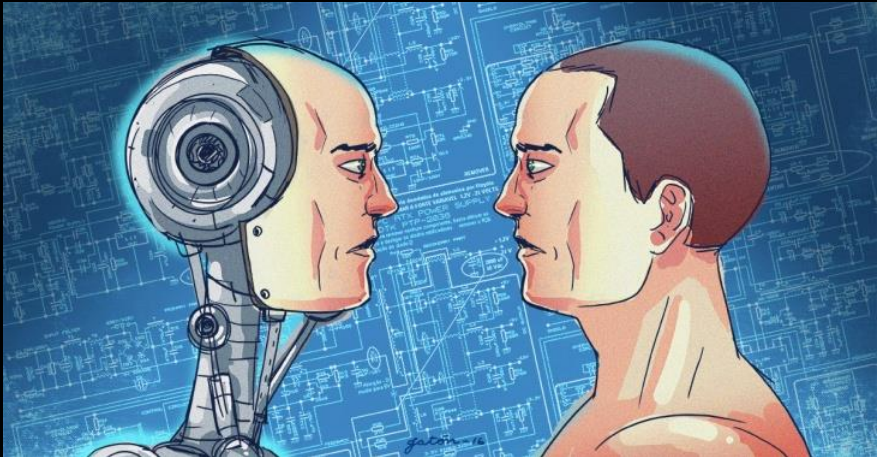
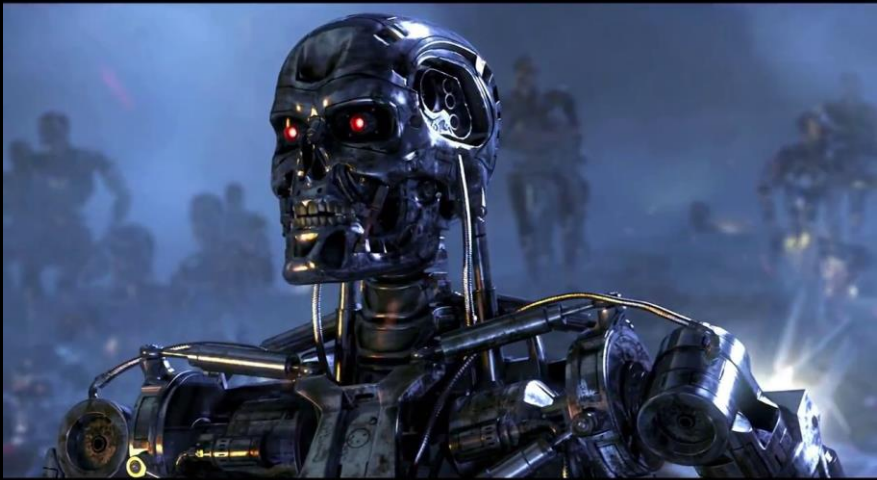
-Comportamiento X

Machine Learning:

Un sistema, por si mismo, aprenda de los datos a los que es expuesto, y genere reglas para obtener un mejor desempeño.

- Mejorar comportamiento X

Lo que piensan todos vs. La realidad



Fitness tracking app Strava gives away location of secret US army bases

Data about exercise routes shared online by soldiers can be used to pinpoint overseas facilities

- Latest: Strava suggests military users 'opt out' of heatmap as row deepens

▲ A military base in Helmand Province, Afghanistan with route taken by joggers highlighted by Strava. Photograph: Strava Heatmap

| Human vs Machine: Lung Tumor | Human vs Machine: Pneumonia |
|---|--|
| <p>Chest X-Rays image the lungs, heart, blood vessels, and bones. AI has been used to read and understand them.</p> <p>Example: Lung Tumor</p> <p>Computers: Score: 0.291</p> <p>Doctors: 2/15 Detected</p> | <p>Chest X-Rays image the lungs, heart, blood vessels, and bones. AI has been used to read and understand them.</p> <p>Example: Pneumonia</p> <p>Computers: Score: 0.371</p> <p>Doctors: 0/15 Detected</p> |

Buzz

To infinity and beyond!

Buzz

Welcome! How can I help you today?

Consult with our experts

I want an app

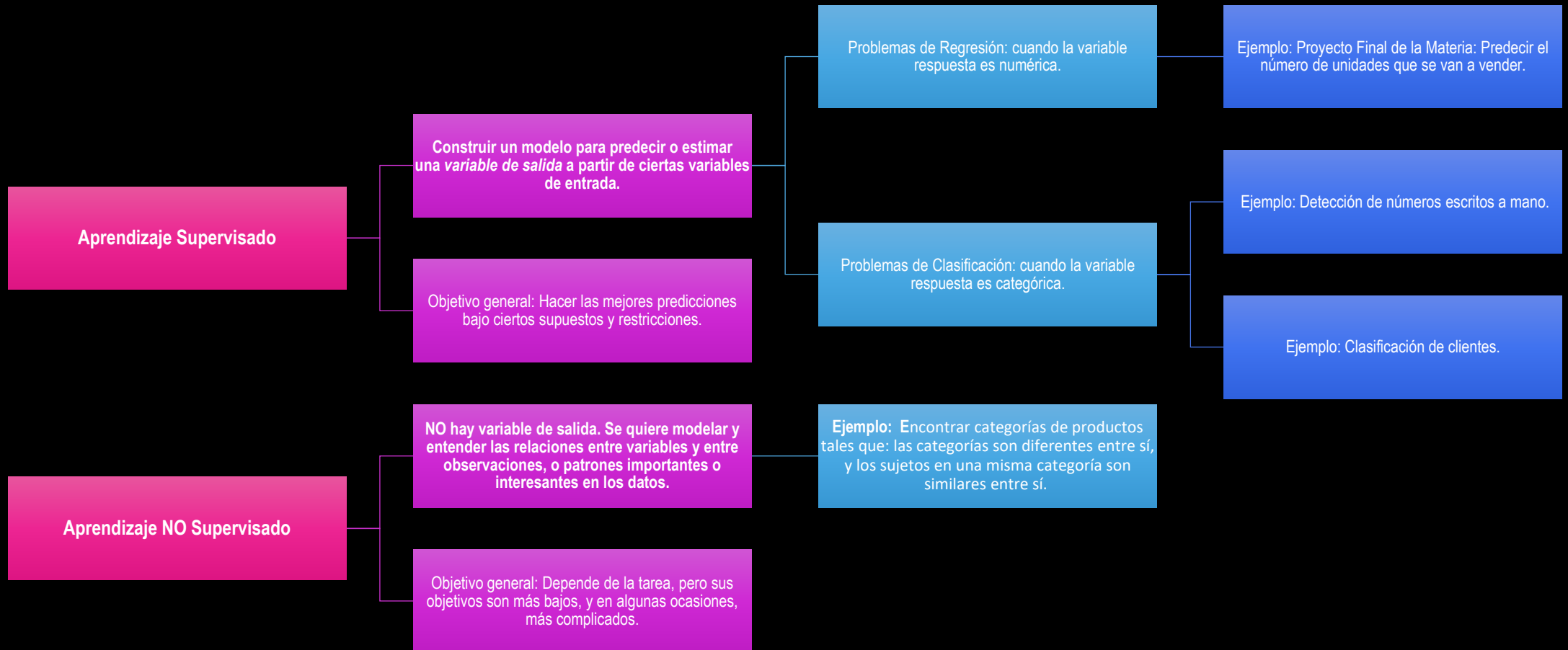
Looking for a cloud solution

Just browsing

Type your message and hit Enter

AI for Social Good

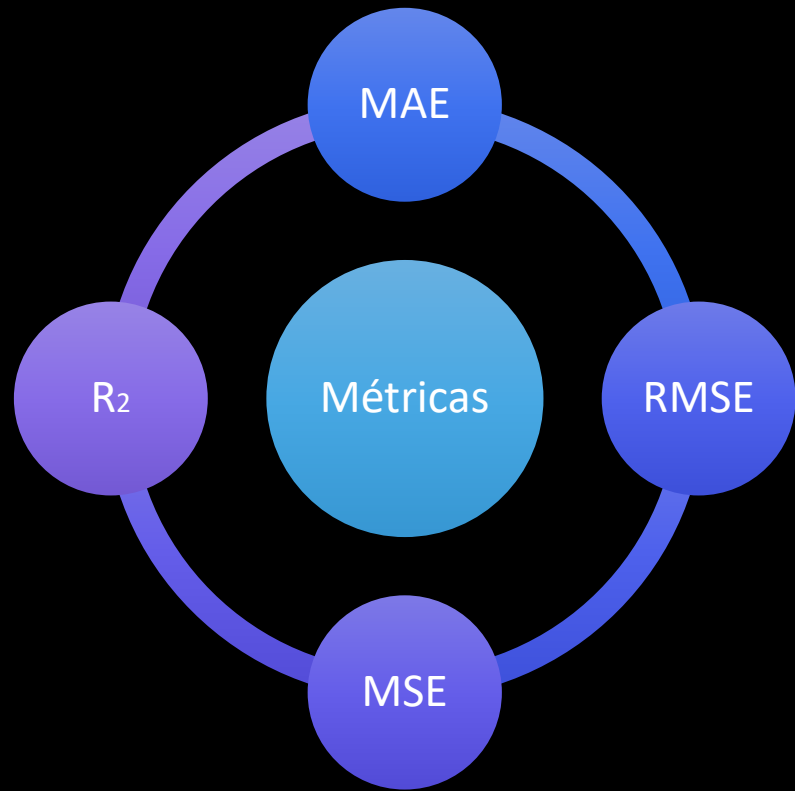
2. Aprendizaje supervisado Vs. Aprendizaje no supervisado



3. Estadística tradicional Vs. Aprendizaje de Máquina

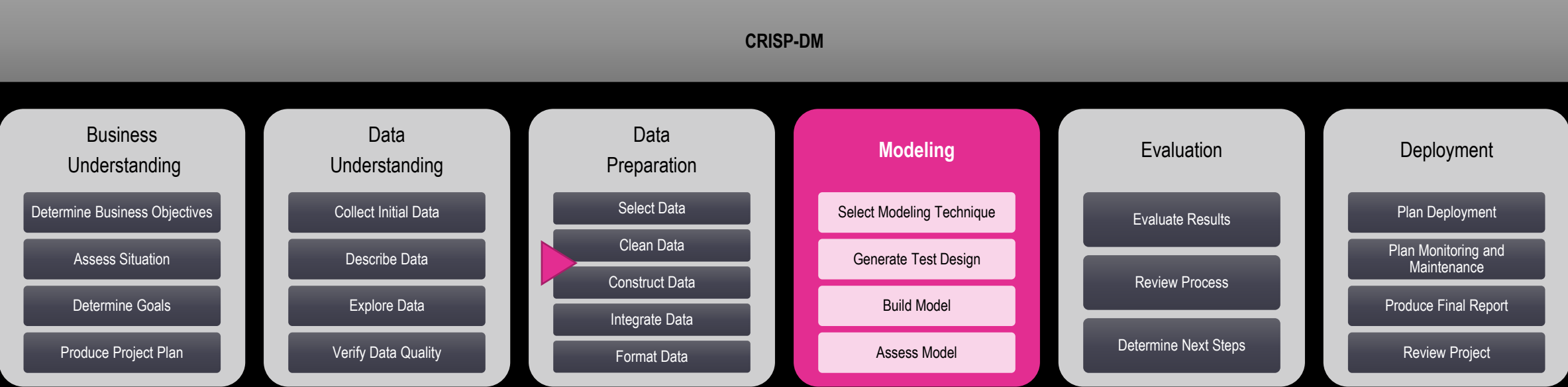
| Aprendizaje de Máquina | Análisis Estadísticos Tradicionales |
|--|---|
| Algoritmo que puede aprender de los datos sin depender de la programación manual basada en reglas. | Relaciones entre variables en los datos en forma de ecuaciones matemáticas. |
| Estimación – Aprendizaje | Estimación - Reglas |
| Instrucciones (reglas) aprendidas. | Instrucciones explícitamente programadas |
| Generalmente se aplica a conjuntos de datos de alta dimensión, cuantos más datos tenga, más exacta será su predicción. | Datos con dimensión no tan alta. |
| Software (Enfocado en programación y analítica) | Software (Enfocado en analítica) |
| Modelos más complejos (reglas) | Modelos más sencillos (reglas) |
| Busca que las predicciones sean lo más certero posible | Modelos simples que expliquen el por qué de los fenómenos |

4. Importancia de las Métricas

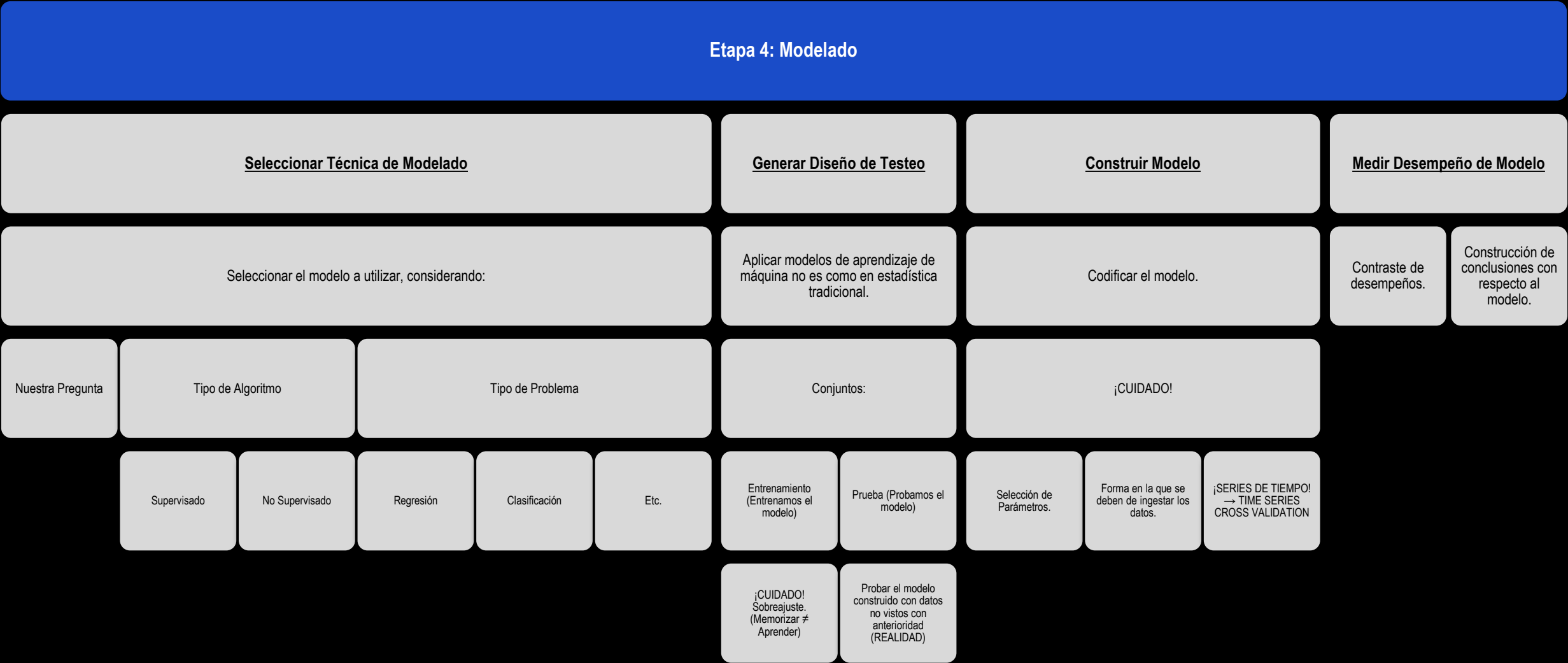


- Medir el desempeño.
- Contraste Modelo actual vs. Modelo propuesto.
- Resultados: interpretables y entendibles.

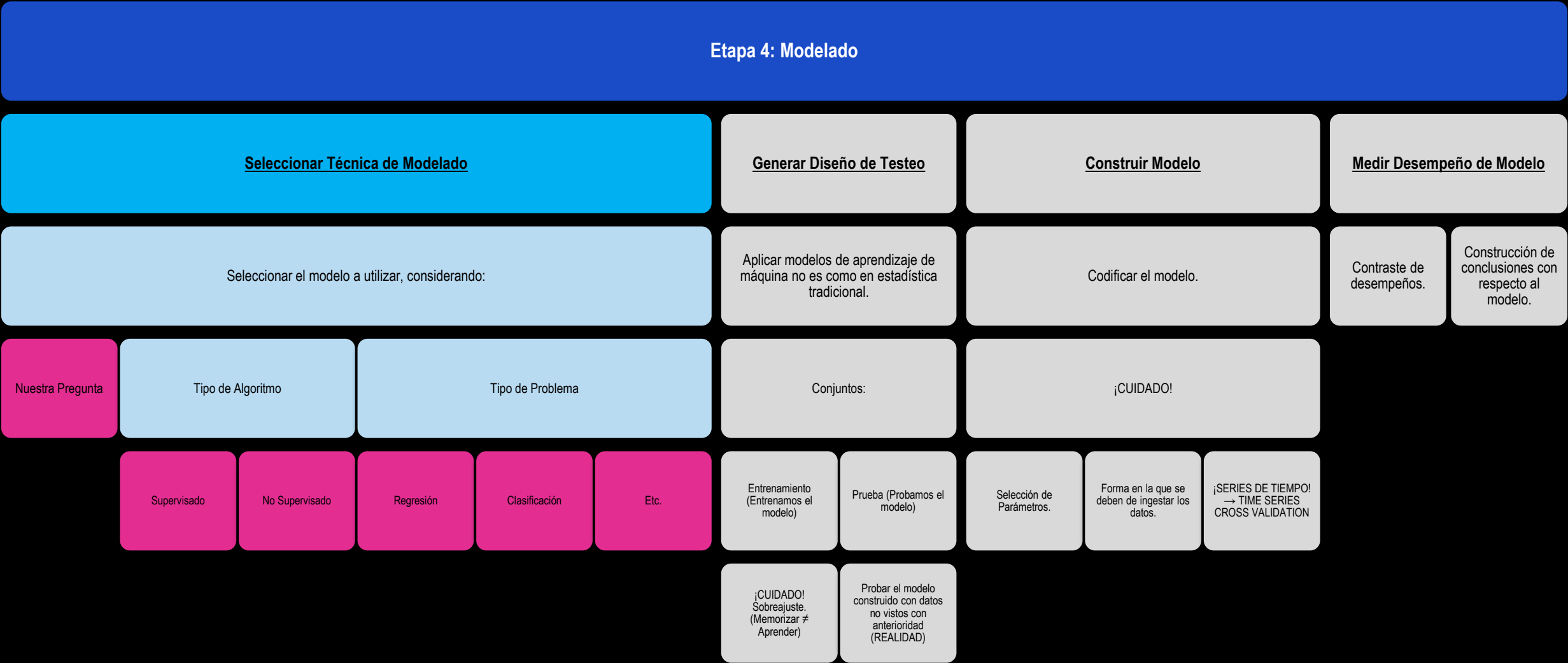
Descripción y Construcción de Modelos



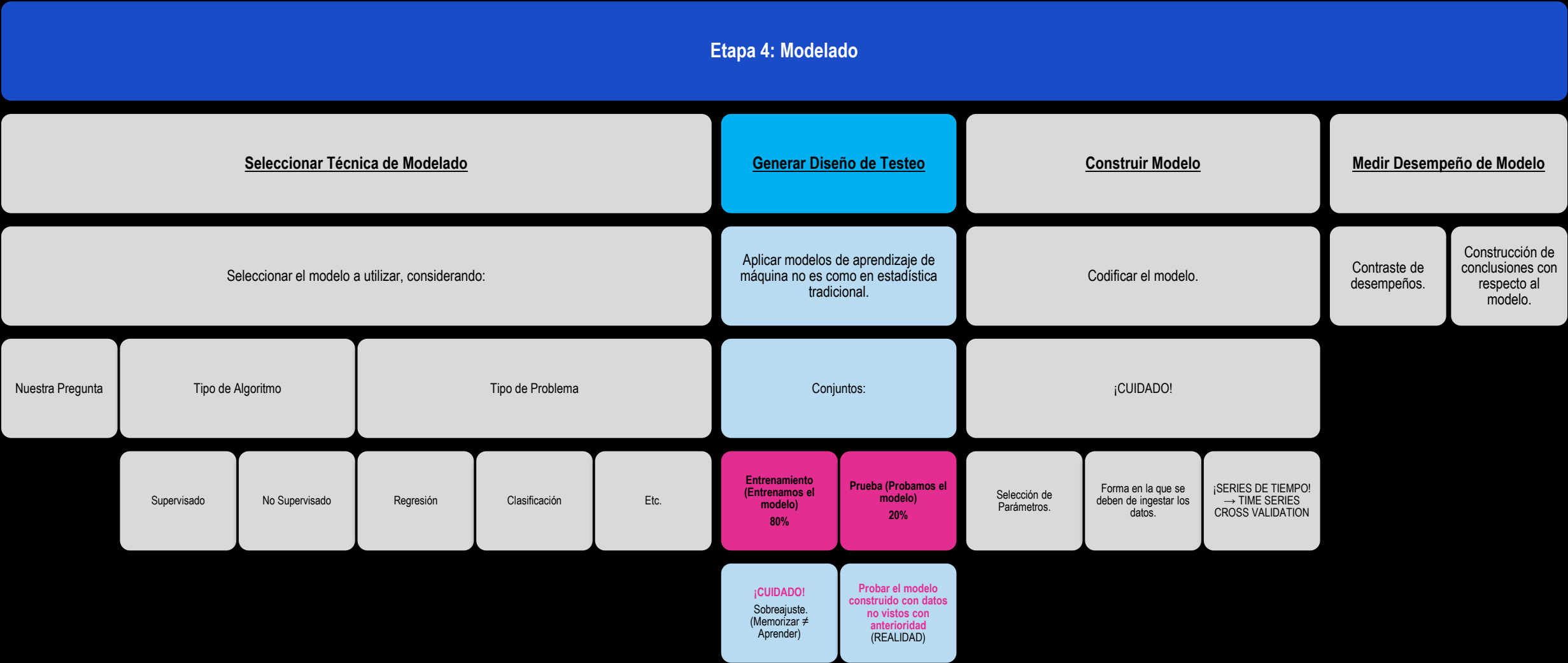
Descripción y Construcción de Modelos



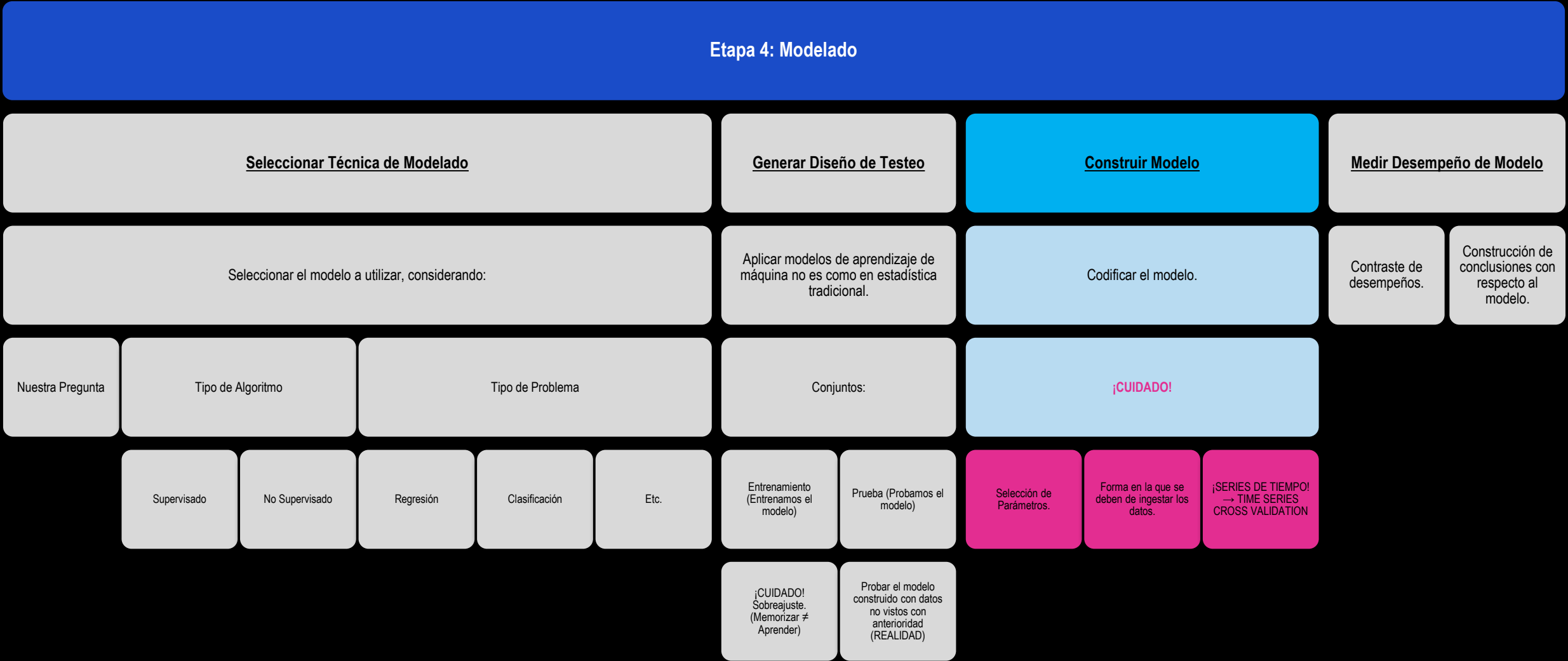
Descripción y Construcción de Modelos



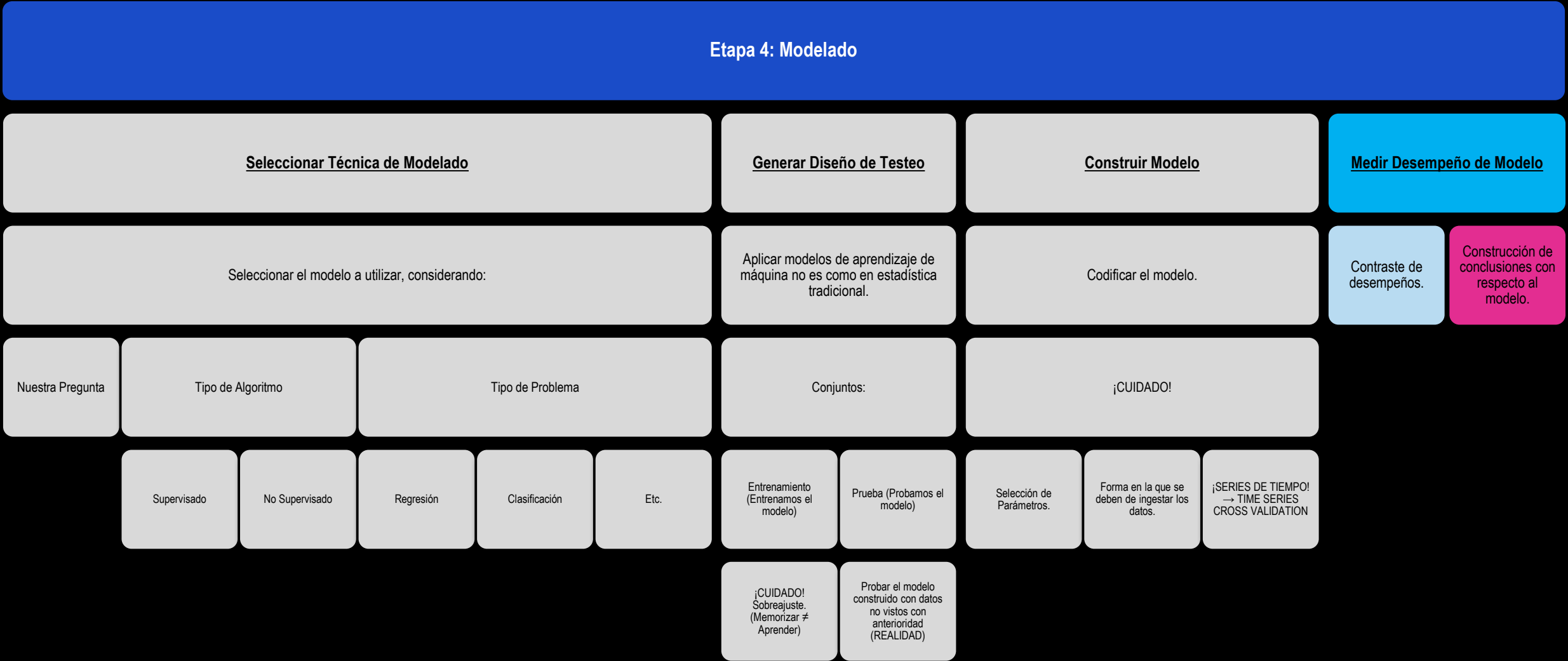
Descripción y Construcción de Modelos



Descripción y Construcción de Modelos



Descripción y Construcción de Modelos



Descripción y Construcción de Modelos

Para construir propuestas de modelos que permitan lidiar mejor con la problemática de la empresa, es necesario desglosarlo de tal manera que se pueda entender mejor y así su planteamiento sea más sencillo (estructura anterior).

Proceso generador de datos (modelo teórico) [3].

- Y es la variable que se busca predecir;
- X es una variable, o conjunto de variables, que se espera que pueda mejorar la predicción de Y ;
- Y y X están relacionadas de la siguiente manera: $Y = f(X) + \varepsilon$

En problemas de Aprendizaje de Máquina, f no se conoce, por ende, hay que estimarla \hat{f} con ayuda de los datos (aprender de los datos).

- Dividir los datos en conjuntos de entrenamiento y de prueba.
- Revisar su desempeño considerando métricas que permita comparar el valor real con el de las predicciones [4].

La idea general de un algoritmo de aprendizaje de máquina es aprender de una muestra de entrenamiento y así generar una \hat{f} que permita hacer predicciones. $\hat{Y} = \hat{f}(X)$

Actividad 6: Visualización Proyecto – EDA

En clase se ha discutido la importancia de hacer un Análisis Exploratorio de los Datos. En esta actividad el equipo deberá de construir un Análisis Exploratorio de los datos que permita dar más conocimiento y contexto de la situación a la que se están enfrentando como equipo.

Requisitos:

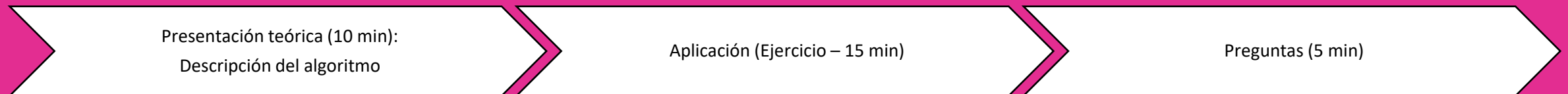
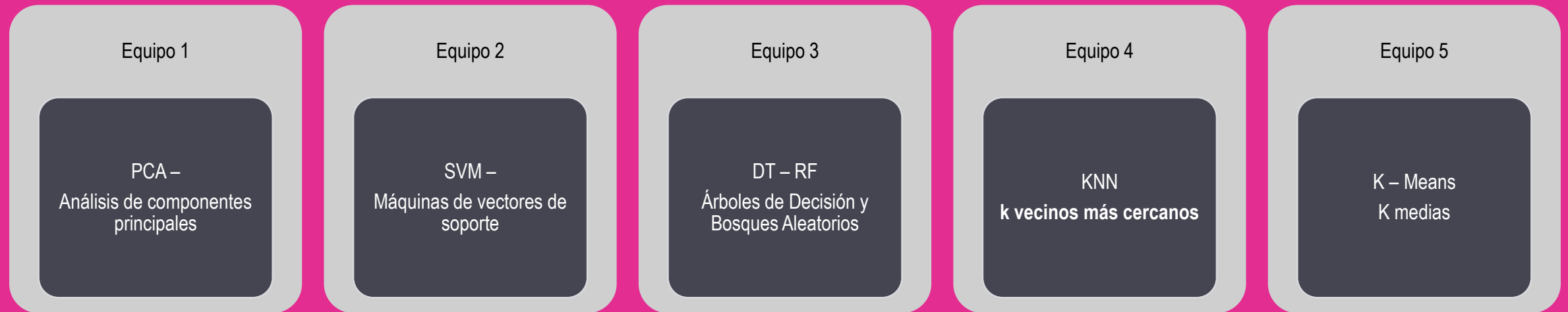
- Mínimo 3 gráficas distintas (Gráfica de barras, Boxplot, Mapas, Líneas, Puntos, etc.) que representen información relevante de los datos.
- Descripción de las gráficas:
 - ¿Qué se puede observar?
 - ¿Por qué es importante ese análisis?
- Documentación (Historia, comentarios en código, resultados, etc.).

Entregable:

- Reporte en formato pdf.
- Actividad en equipo.
- Subir a CANVAS (Fecha límite: 21 de septiembre 15:55 hrs.) .

Actividad 6: Visualización Proyecto – EDA

Actividad 7: Presentación Algoritmos de Aprendizaje de Máquina



- Regresión / Clasificación
- Funcionamiento
- Ventajas / Desventajas

- R / Python
- Ejercicio sencillo para que la clase pueda seguirlo
- Datos pequeños – Compartirlos (DRIVE)

Bibliografía

- SAS. (2020). *Machine Learning, una expresión de la Inteligencia Artificial*. 10 de septiembre de 2020. de SAS Sitio web: https://www.sas.com/es_mx/whitepapers/local/machine-learning.html?gclid=CjwKCAjwnef6BRAGeiwAgv8mQZl1qtq3ZaUfJ_C2b1YTzQ_dYiEhUpxITdCEA5kiNZ7KTTZrRThfYBoCb5QQAvD_BwE
- González, L. (2018). ¿Qué es aprendizaje de máquina (machine learning)?. 26 de septiembre de 2020, de Github Sitio web: <https://github.com/felipegonzalez/aprendizaje-maquina-mcd-2018>
- González, L. (2018). Aprendizaje Supervisado: Proceso Generador de Datos (Modelo teórico). 26 de septiembre de 2020, de Github Sitio web: <https://felipegonzalez.github.io/aprendizaje-maquina-mcd-2018/introduccion.html#aprendizaje-supervisado-1l>
- Mármol, J. (2018). Introducción. 15 de agosto de 2020, de Github Sitio web: <https://github.com/AnaLuisaMasetto/intro-to-data-science-2018>