

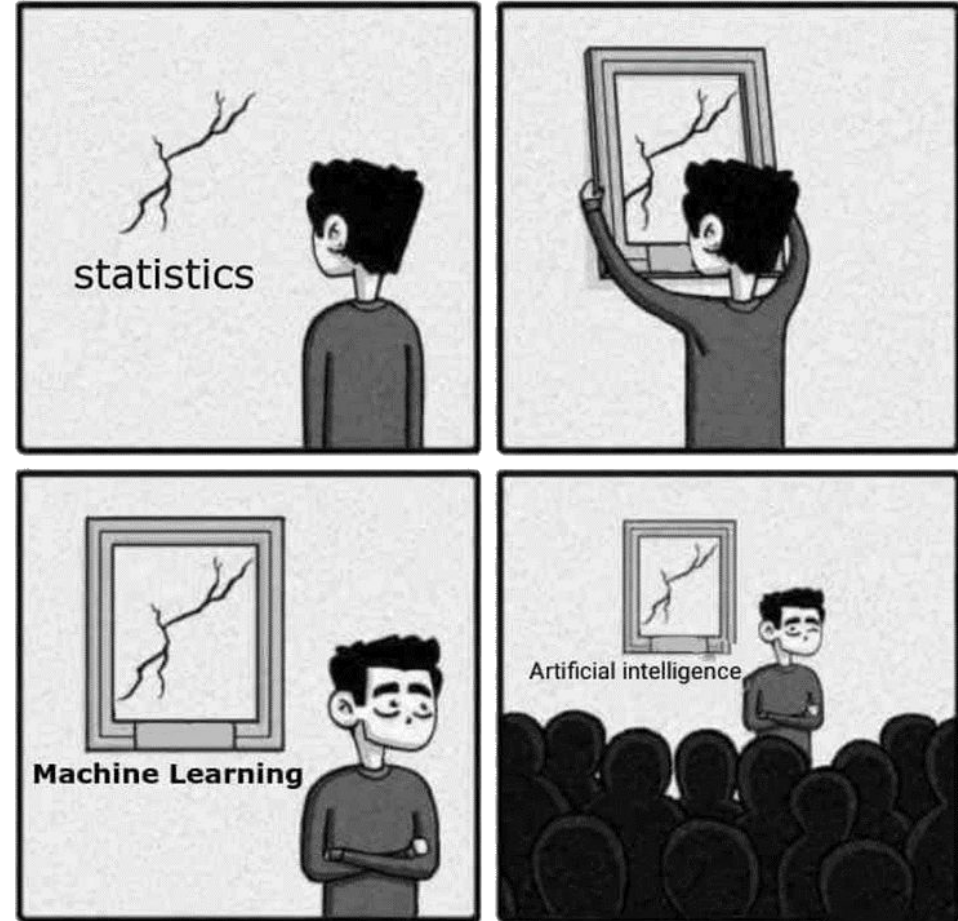


# Ciencia de datos

Agosto 2023

# Objetivo

Dar una introducción a los conceptos fundamentales de la Ciencia de Datos, su rol dentro de la toma de decisiones y sus casos de uso más comunes.



# Contenido



**1** Ciencia de datos

**2** Inteligencia Artificial

**3** Niveles de madurez

**4** Toma de decisiones

**5** Casos de uso comunes

**6** Ecosistema de datos

**7**

# 01 ¿Que es la ciencia de datos?

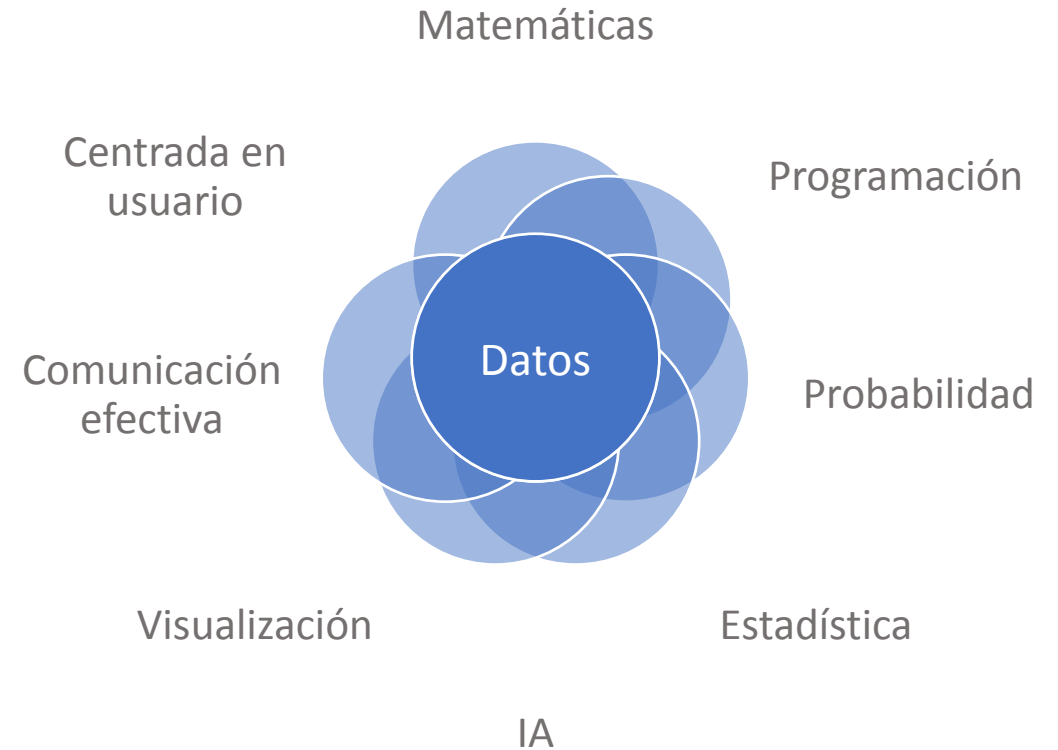


# Ciencia de datos

¿Qué es la ciencia de datos?

El objetivo de la ciencia de datos es abordar un problema de negocios y dar una solución basada en el análisis, visualización y/o modelado de datos con el fin:

- Aumento de la rentabilidad
- Ahorro de costos o tiempos
- Eficiencia operativa
- Evitar o reducir riesgos
- Reputación de marca



Elementos de la ciencia de datos



# Pensamiento Analítico

Habilidades analíticas para la empresa moderna impulsada por IA



El pensamiento analítico, base de lo que hoy llamamos enfoque orientado a los datos, puede entenderse como **la capacidad para resolver problemas de forma lógica y racional**. Implica el uso de herramientas para:

- Análisis estadístico y cuantitativo
- Modelos explicativos y predictivos
- Gestión basada en hechos
- Uso intensivo de datos para guiar decisiones y acciones

Defino el **razonamiento analítico** como la capacidad de **traducir problemas de negocio en soluciones prescriptivas**. Esto requiere tanto orientación a los datos como habilidades lógicas para encontrar soluciones efectivas.

# Perfil

## de un científico de datos



### Proactividad

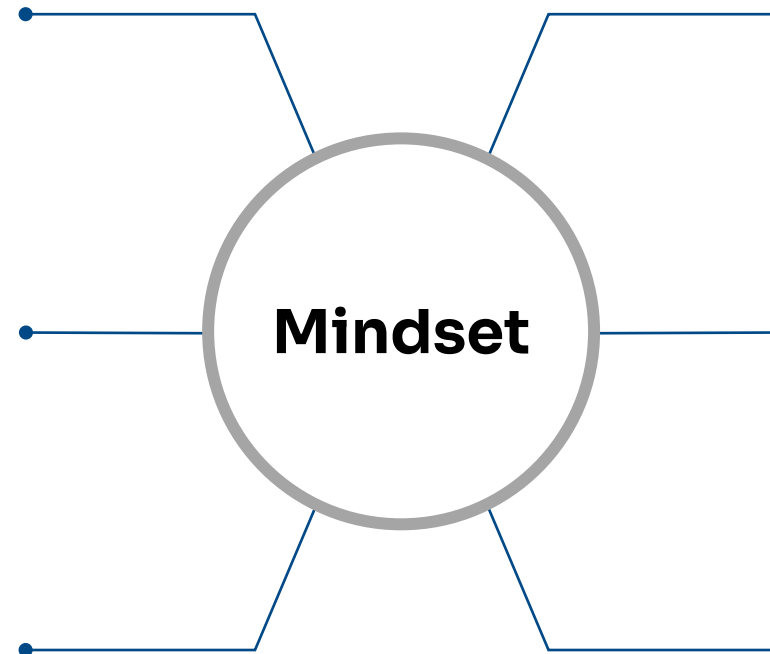
Identifica necesidades y es capaz de proponer alternativas de acción cuando es conveniente.

### Inteligencia de negocio

Se asegura de que cada pieza de análisis funcione adecuadamente para atender las demandas y objetivos de su organización.

### Creatividad

Investiga a profundidad las distintas caras de los datos para construir soluciones innovadoras a problemas complejos.



### Comunicación efectiva

Explica con claridad los modelos que construye y escucha con atención la retroalimentación de su trabajo.

### Trabajo en equipo

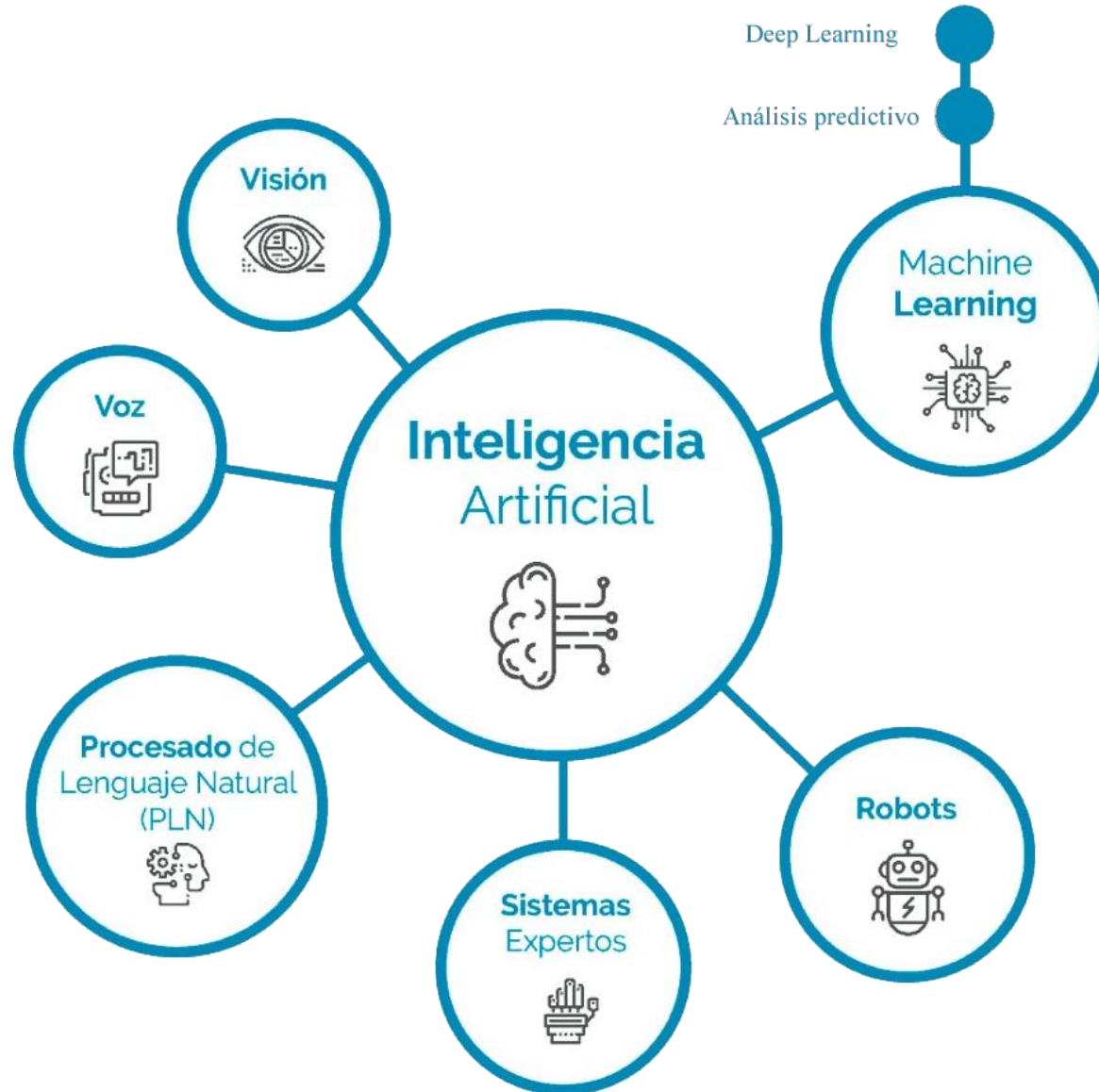
Coopera con otros profesionales del área para construir soluciones integrales.

### Pensamiento crítico

Filtra información a través de su conocimiento sobre las tendencias en el área, formula preguntas y resuelve problemas de forma objetiva

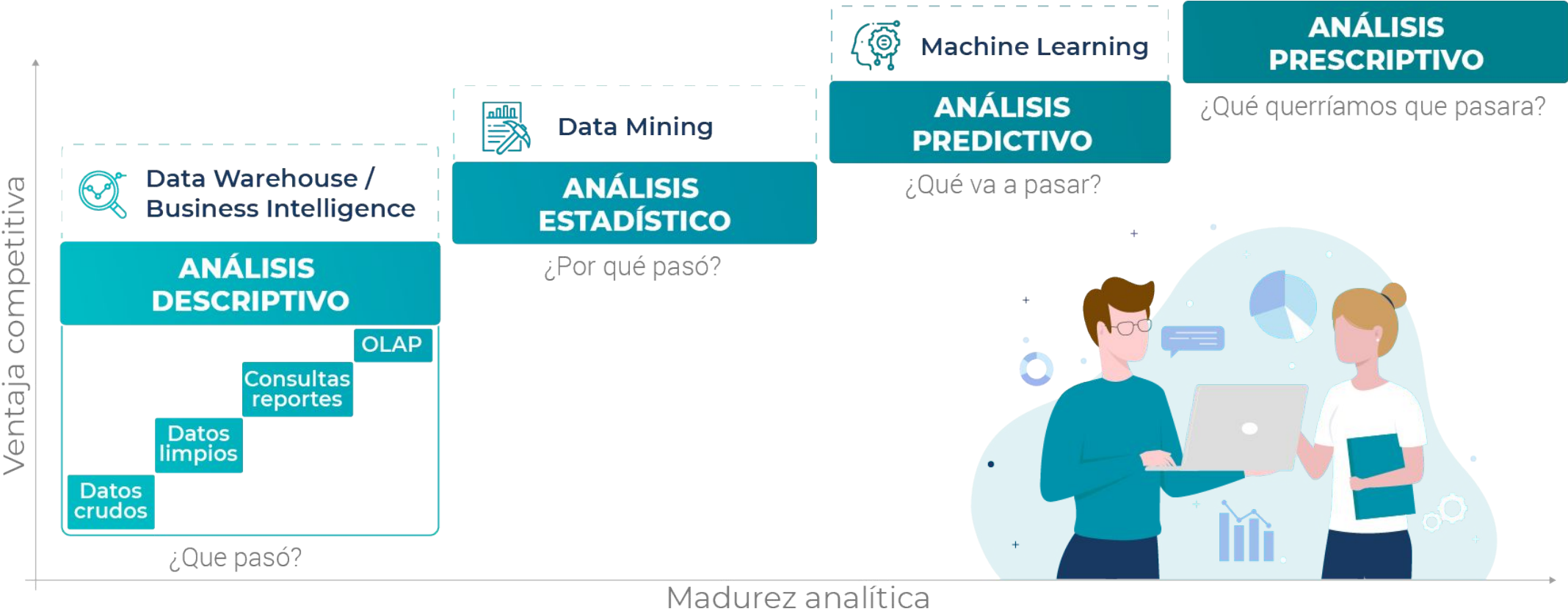
# Inteligencia Artificial

Campo de estudio

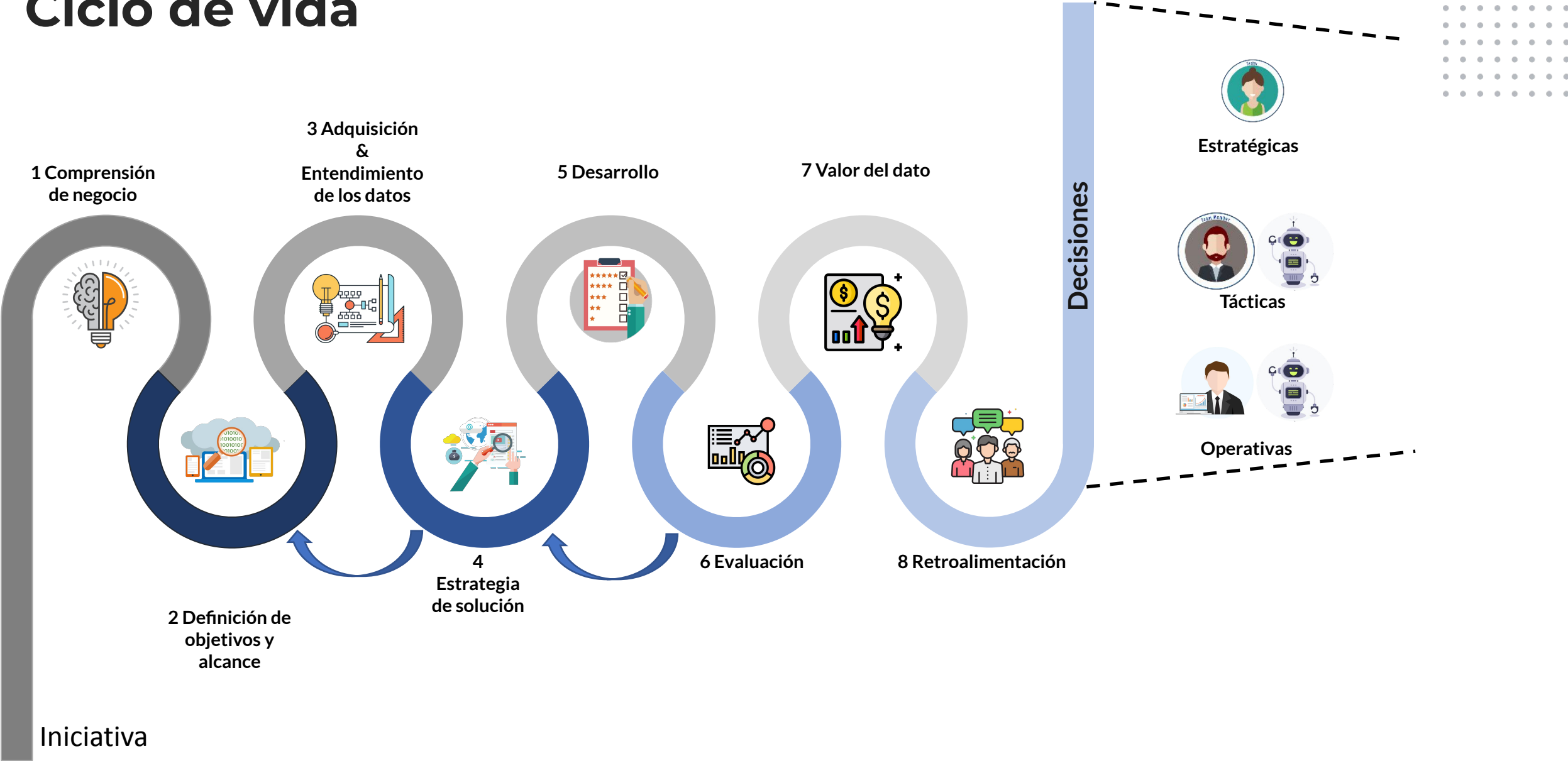




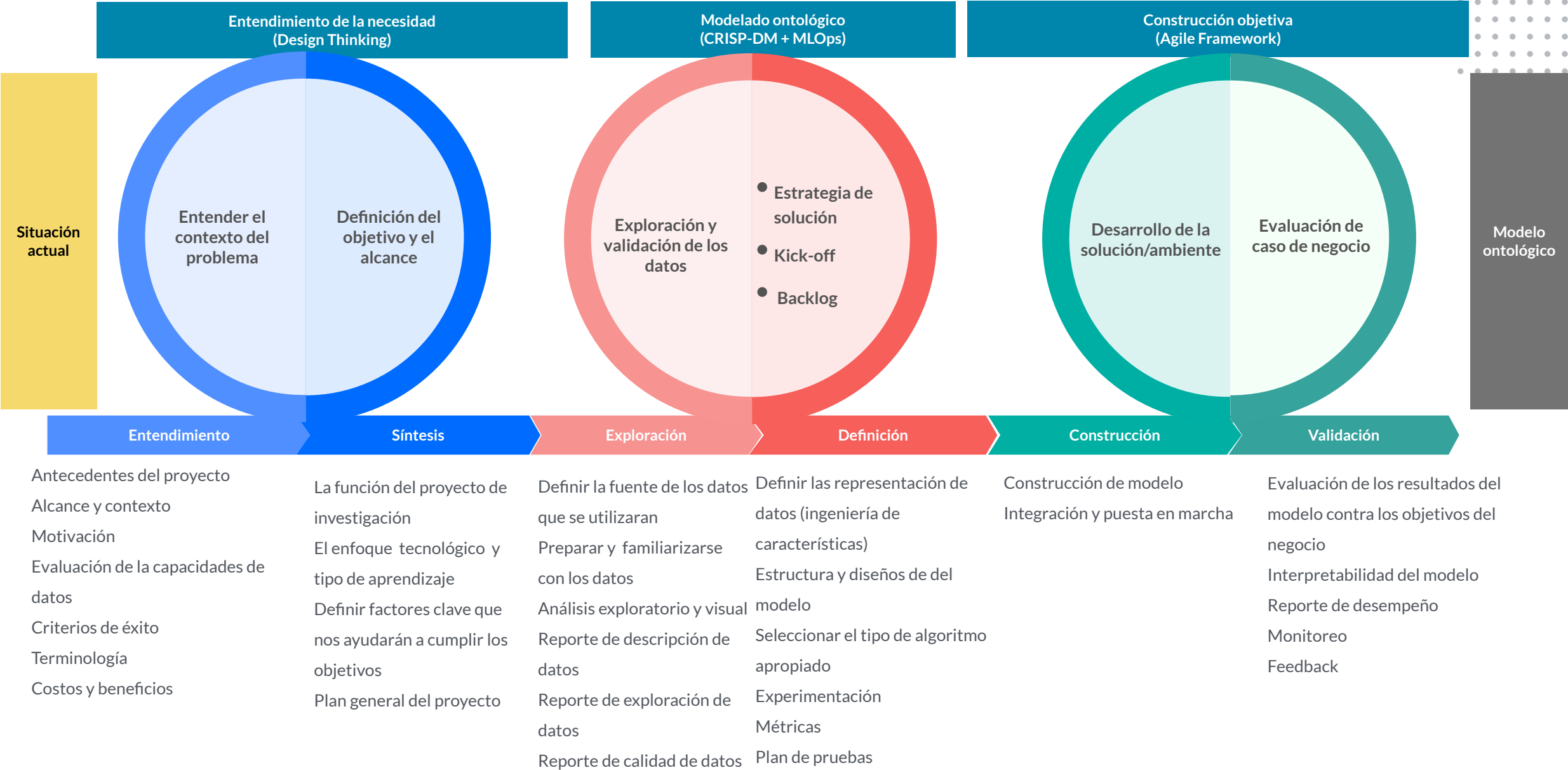
# Niveles de madurez



# Ciclo de vida



# Metodología



# Casos de uso

## Perfilado de comportamiento

- Venta cruzada, captación y retención de clientes
- Sistema de recomendaciones
- Next best action

- Clientes con comportamiento similares para ofrecer servicios y productos
- Personalización de productos y servicios
- Satisfacción de clientes

## Automatización

- Chatbots y voice bots
- People analytics

- Asistentes virtuales que puedan ofrecer servicios y que estén integrador a los sistemas transaccionales
- Reducir la necesidad de proporcionar los mismos datos una y otra vez

## Gestión de riesgo

- Sistemas antifraude y ciberseguridad
- Falsificación
- Detección de armas y comportamiento sospechoso

- Robustecer los sistemas antifraude para pagos con TDC
- Conducta delictiva y zonas de alerta de crimen
- Acceso a sistema o edificios y autenticadores

## Explotar los datos

- Búsqueda
- Auditar calidad de servicios

- Búsqueda en lenguaje natural
- Visión 360
- Respuestas en tiempo real

# Ecosistemas de datos

Para tener cuidado



---

Data Consulting & Strategy

---

---

Data Infrastructure

---

---

Data Architecture

---

---

Data Engineering & Big Data

---

---

Data Analytics

---

---

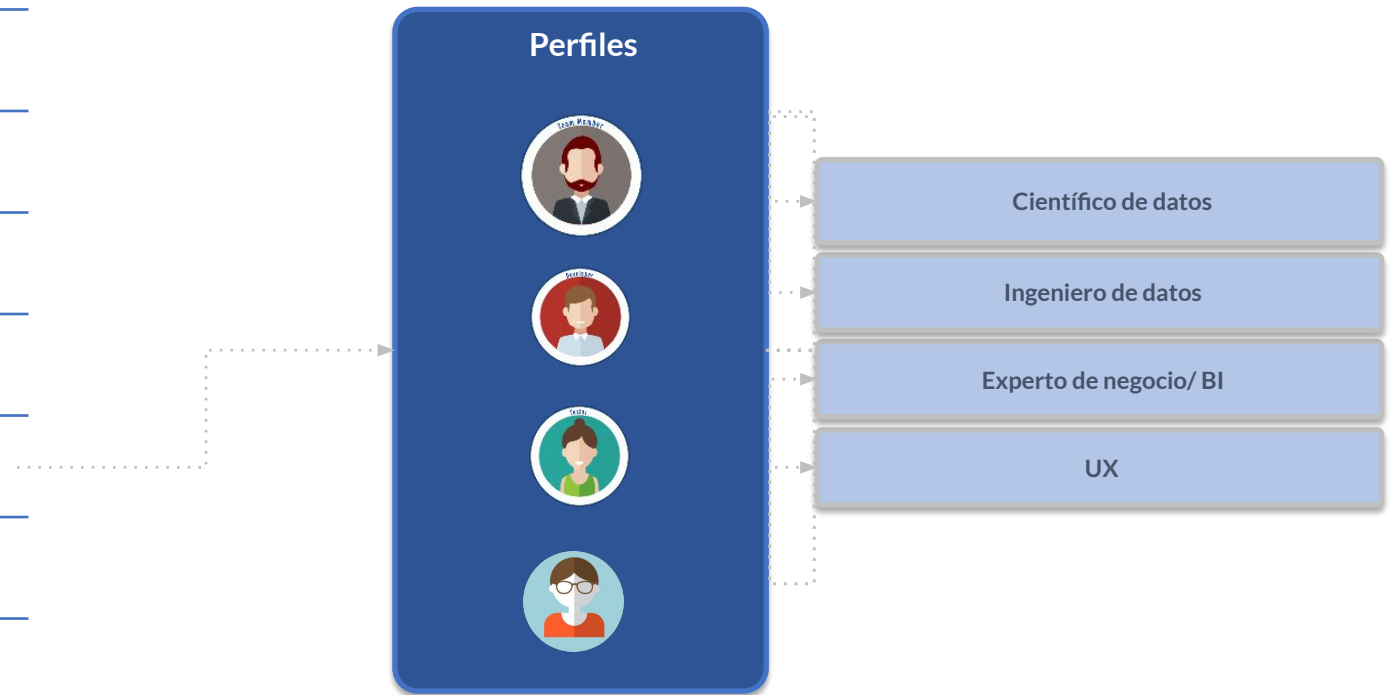
DaaS & Data Virtualization

---

---

Data Governance

---





**Gracias por tu tiempo**

# Modelo de predicción de severidad de fallas



Universo de variables

Modelo de clasificación

Predicción

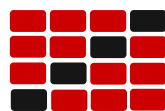
## Información Interna

### Variables

- ✓ Location
- ✓ Event type
- ✓ Log feature
- ✓ Volume
- ✓ Resource type
- ✓ Severity type



Se utiliza modelos de bosques aleatorio para predecir la severidad de las fallas del servicio de red



■ test  
■ train

Se utiliza una búsqueda de malla y stratified 10-cross fold validation para la selección de hiperparámetros



Rendimiento

El modelo tiene un rendimiento de 0.64383 sobre la métrica log loss

0

No hay falla

1

Pocas fallas

2

Muchas fallas