



Análisis de datos para la toma de decisiones

Profesora: Katherine Barquero Mejías



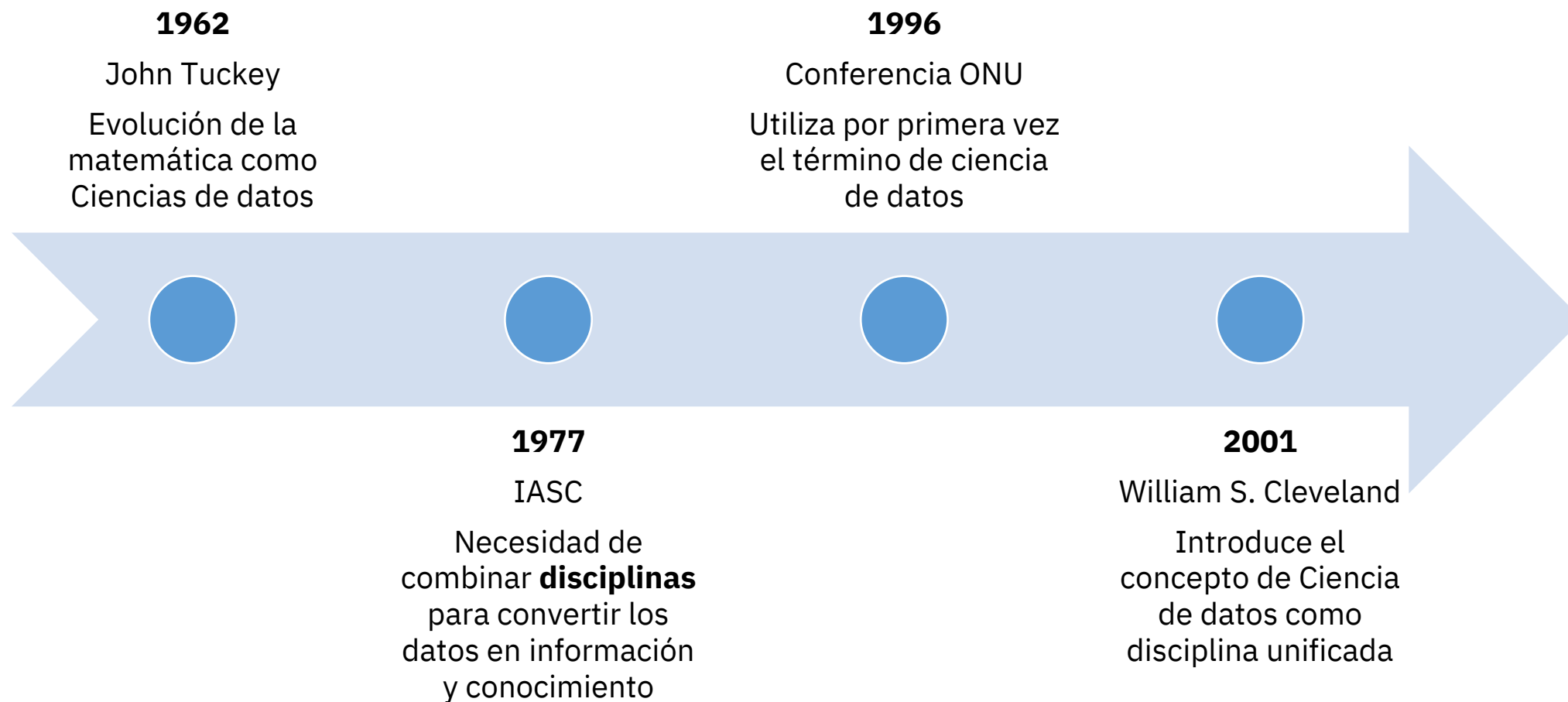
Clase 1: Introducción a la Ciencia de Datos



Agenda de la clase

**¿QUÉ ES LA CIENCIA DE DATOS?
¿QUÉ HAS ESCUCHADO SOBRE
ELLA?**

Origen de la Ciencia de Datos



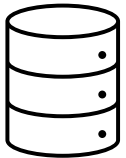
Antecedentes



1960 surgimiento
computadoras



1990 bases modernas
del internet.



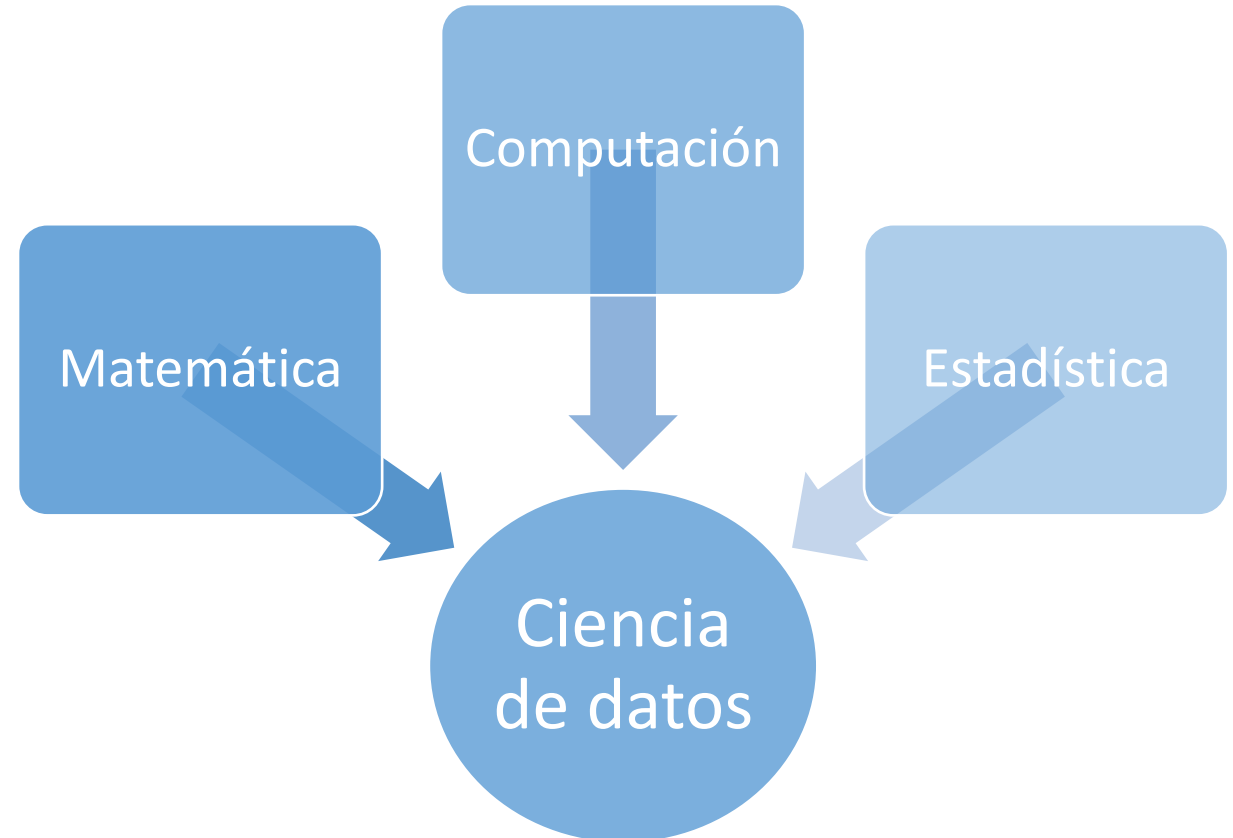
cantidad de datos
almacenados excedía
la capacidad de
procesamiento

Su evolución originó los trabajadores del conocimiento, ahora conocidos como **científicos de datos**.

*“combina la **estadística**, las **matemáticas**, la **programación** y las habilidades para la solución de problemas con la captura datos, y es capaz de llevar a cabo las actividades propias de limpieza, preparación e integración de datos con la intención de **localizar patrones en la información** analizada (Davenport y Patil 2012)*

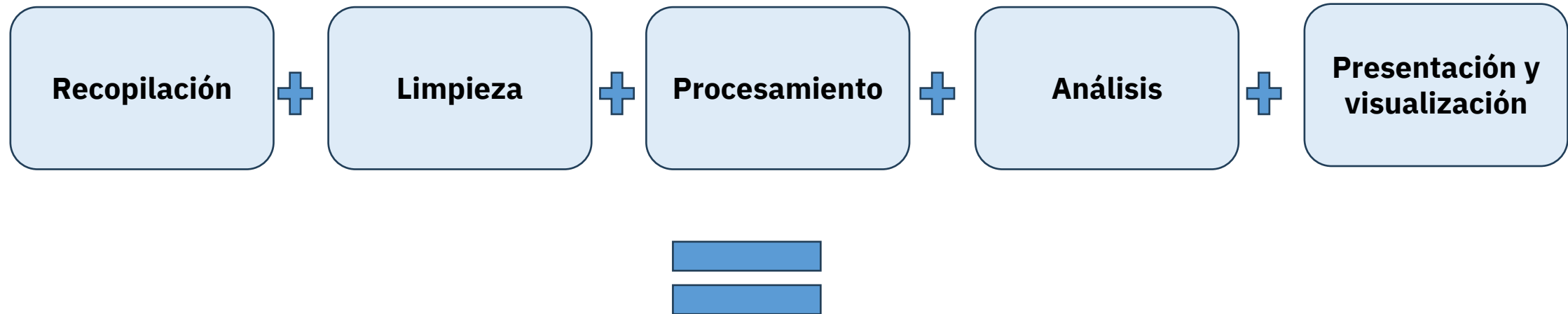
Ciencia de datos

- Es uno de los principales campos en la **actualidad**
- Uso del **método científico**
- Campo interdisciplinario:



Ciencia de datos

- Abarca todo el ciclo de vida de los datos



**Toma de decisiones
informada basada en
evidencia**

Ciencia de datos, Análisis de datos y Estadística

Ciencia de datos

- Gran volumen de datos
- Abarca todo el ciclo de vida de los datos
- Procesos automatizados
- Buscan la clasificación y predicción

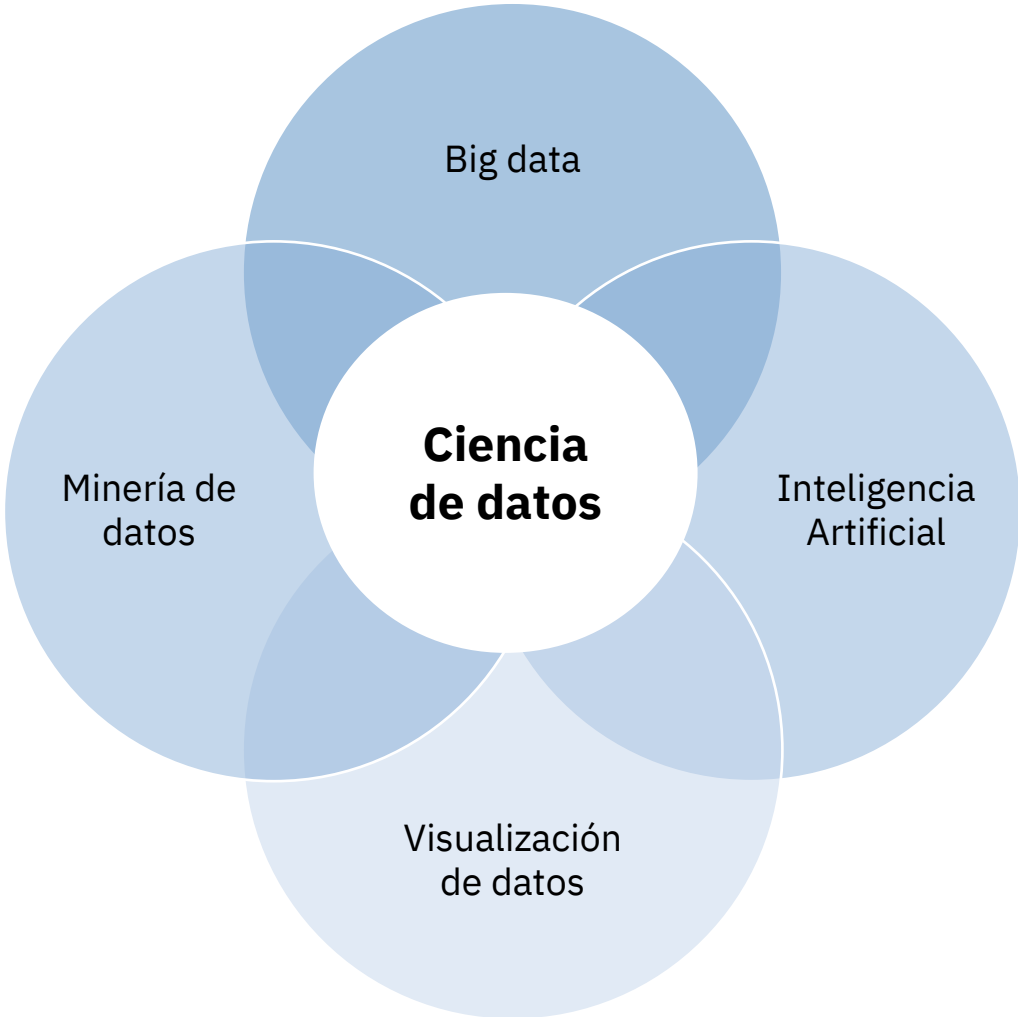
Análisis de datos

- No siempre se trabaja con alto volumen de datos
- Son una parte específica de la ciencia de datos
- No son procesos automatizados

Estadística

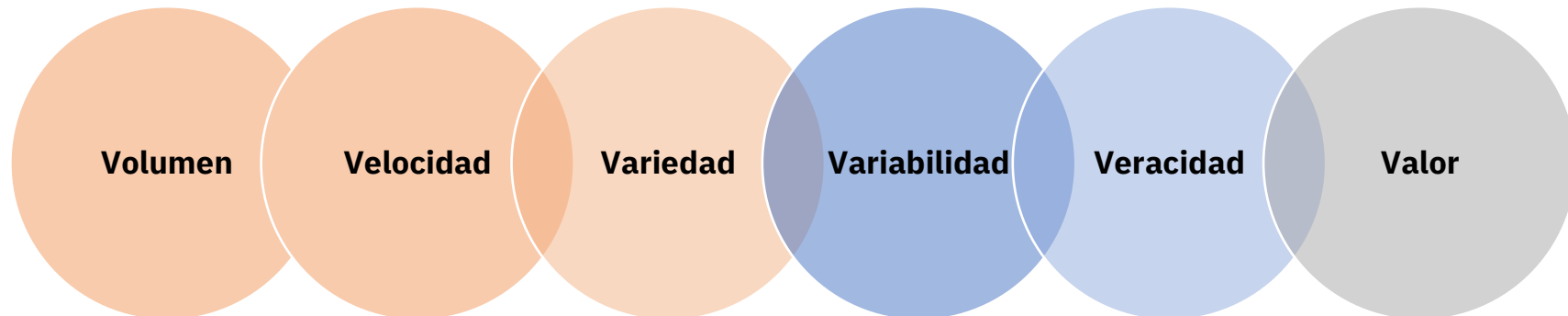
- Trabaja con muestras no con toda la información
- Supone que los datos se comportan de acuerdo a ciertos supuestos (distribuciones de probabilidad)
- Análisis descriptivos y estadística inferencial

Áreas involucradas en la Ciencia de Datos



Big Data

- ❖ Terminó que describe el **gran volumen de datos** – estructurados y no estructurados – que abundan en las organizaciones.
- ❖ Sus dimensiones gigantescas, velocidades tan altas o estructuras tan complejas no pueden ser procesados eficientemente mediante enfoques convencionales
- ❖ **Busca proveer información que mejores las decisiones y acciones estrategias en los negocios o en la política**
- ❖ **Un proceso de BIG DATA cumple con las siguientes V's**



Big Data

- ❖ **Volumen:** Tamaño. Las organizaciones recopilan datos de diversas fuentes, como transacciones comerciales, dispositivos inteligentes (IO), equipo industrial, vídeos, medios sociales y más.
- ❖ **Variabilidad:** Flujo de datos impredecibles. Cambian a menudo y producen cargas inesperadas de información. Trending topics, temas virales
- ❖ **Veracidad:** Calidad y confiabilidad de los datos. Se refiere a como las organizaciones garantizan procesos transparentes: vienen de fuente confiables, y han sido recopilados y almacenados de forma segura.
- ❖ **Velocidad:** Producción en tiempo real. Gracias al Internet de las cosas las organizaciones enfrentan avalancha de datos.
- ❖ **Valor:** Utilidad y beneficio de analizar la información

Big Data: ejemplo

- ❖ **Volumen:** Más de 3 mil millones de usuarios: interacciones y publicaciones que realizan multiplican la información
- ❖ **Variiedad:** Textos, emojis, fotos, videos, interacciones sociales, ubicaciones geográficas
- ❖ **Veracidad:** Algunos perfiles saltan las políticas de privacidad y pueden contener información errónea o fraudulenta (perfiles falsos)

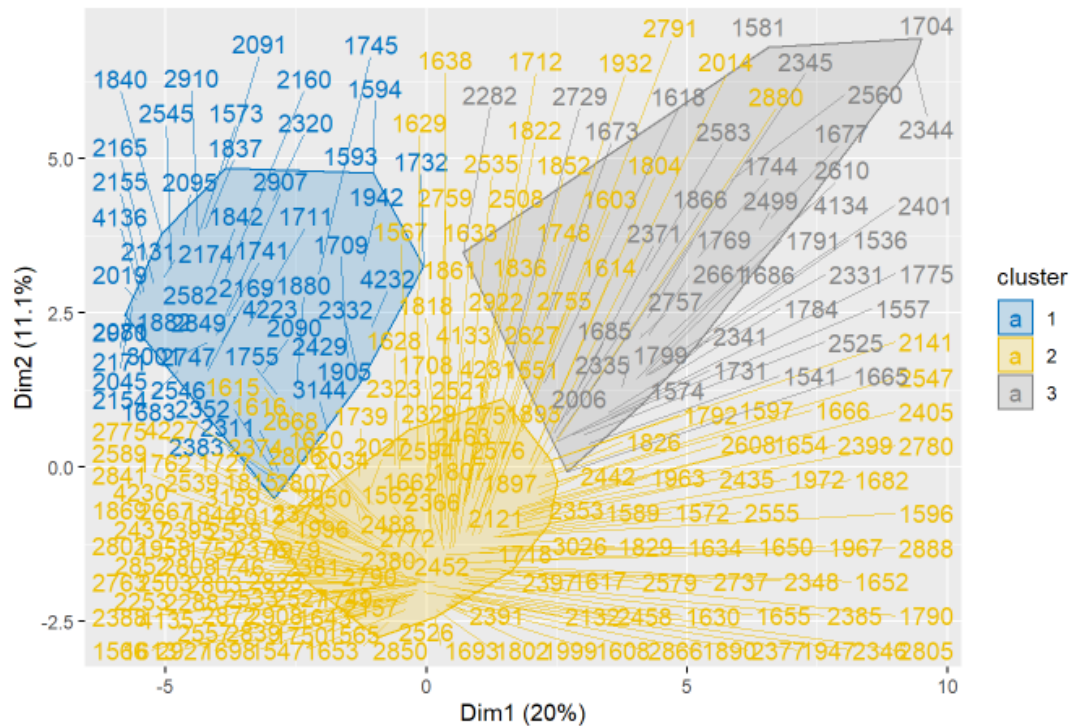


- ❖ **Velocidad:** Información se produce y almacena en tiempo real
- ❖ **Variabilidad:** los temas virales o trendings topics cambian impredeciblemente
- ❖ **Valor:** Publicidad, prensa e información, intereses políticos

Data Mining

- Análisis de grandes volúmenes de datos, principalmente estructurados.
- Se enfoca en la **detección de patrones o relaciones de los datos** para **clasificar o predecir**

Clasificación



Predicción

Entrenamiento					
id	edad	estado_civil	ingreso	incumplimiento	
1	Cliente 1	35	casado	1000	no
2	Cliente 2	30	casado	2500	no
3	Cliente 4	30	casado	500	si
4	Cliente 5	25	soltero	800	si
5	Cliente 7	20	casado	2500	si
6	Cliente 9	30	casado	800	no
7	Cliente 10	25	casado	2000	no

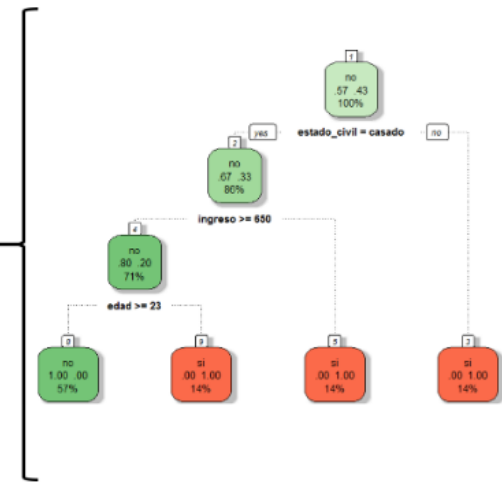
Númérica

Categoría

Númérica

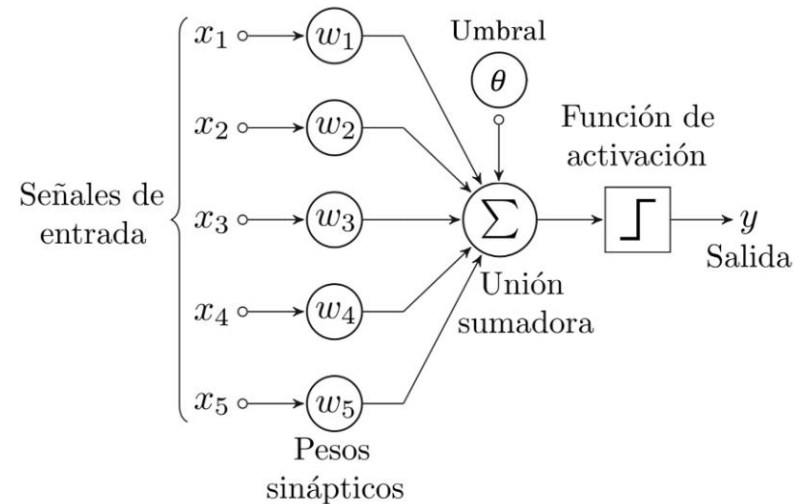
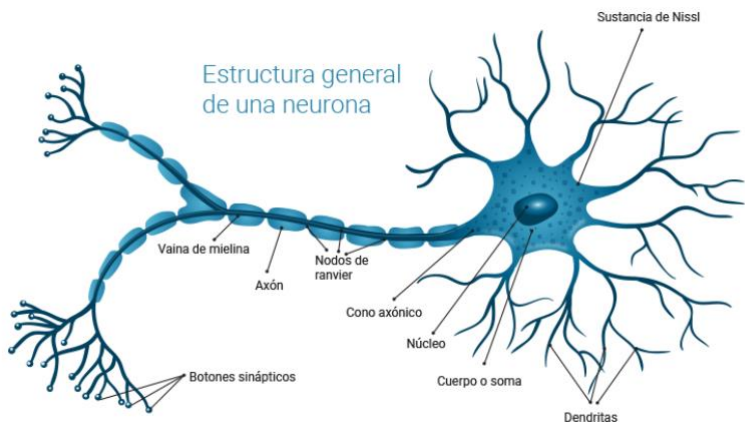
Clase

$f: D \rightarrow C$

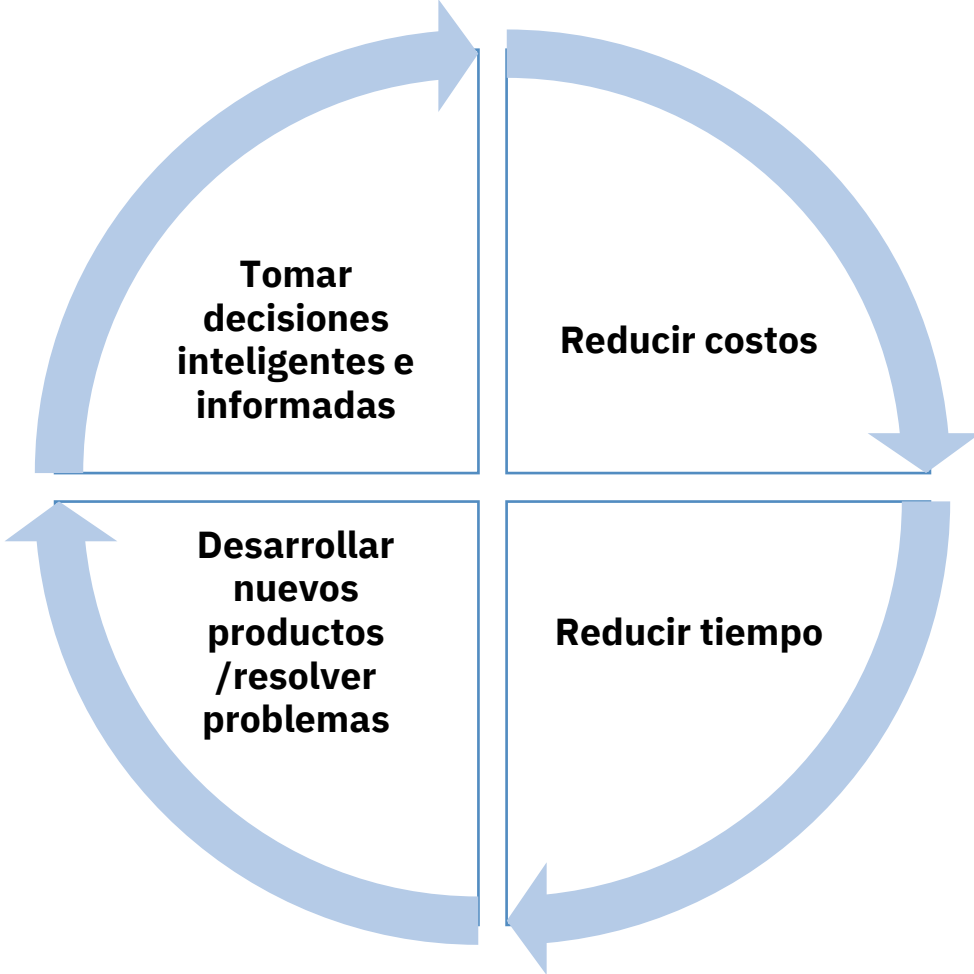


Inteligencia artificial

- ❖ Capacidad de las máquinas para usar algoritmos, aprender de los datos y utilizar lo aprendido en la toma de decisiones tal y como lo haría un ser humano
- ❖ A diferencia de los humanos, las máquinas no requieren descanso y pueden analizar grandes volúmenes de información
- ❖ IA está siendo utilizada para ayudar a las personas en diversas tareas
- ❖ Algunos ejemplos: chat GPT, asistentes virtuales, reconocimiento facial, diagnóstico médico
- ❖ **Una de las técnicas usadas es el machine learning**



Importancia de la Ciencia de datos



La importancia no es el volumen de información, sino, qué se hace con ella

Ética y manejo de datos sensibles

Privacidad y protección de datos

- Manejo de datos seguros
- Anonimización de información
- Protocolo de almacenamiento seguro
- Ataques cibernéticos

Volvemos al problema inicial, el problema no es el volumen de los datos, sino que se hace con ellos

Puede favorecer la toma de decisiones informada o ser empleada para manipulación: si los datos y los modelos de ciencia de datos no son utilizados de manera ética, pueden generar sesgos, o influir en la percepción de las personas con fines malintencionados

**¿Se implementan protocolos en
sus organizaciones para el
manejo y resguardo de datos?
¿Cuáles?**

Muchas gracias

