



SECRETARÍA DE  
INNOVACIÓN

# CIENCIA DE DATOS



# Agenda

## Sesión 1/18

### **Introducción a la ciencia de datos**

- ¿Qué es la Ciencia de Datos?
- Datos, información, conocimiento y decisiones.
- Como medimos la información.
- Big Data y Ciencia de Datos.
- ¿Grandes datos y pobres conocimientos?

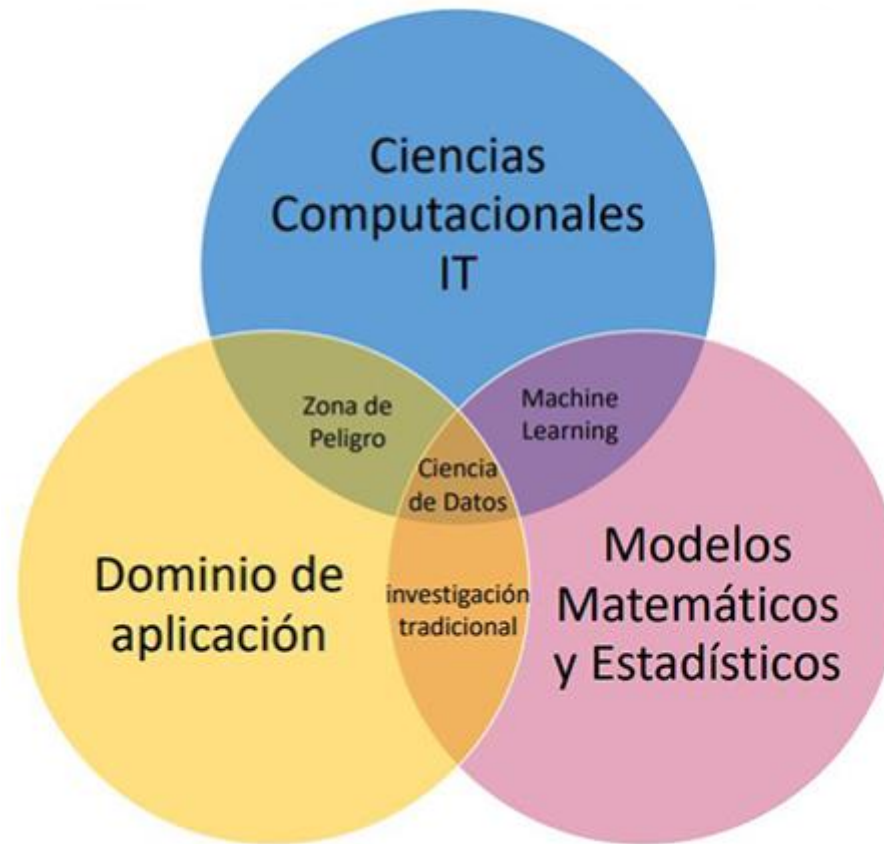
## ¿Qué es la Ciencia de Datos?

La ciencia de datos es un campo interdisciplinario que involucra métodos científicos, procesos y sistemas para extraer conocimiento o un mejor entendimiento de datos en sus diferentes formas, ya sea estructurados o no estructurados, lo cual es una continuación de algunos campos de análisis de datos como la **estadística**, la **minería de datos**, el **aprendizaje automático**, y la **analítica predictiva**.

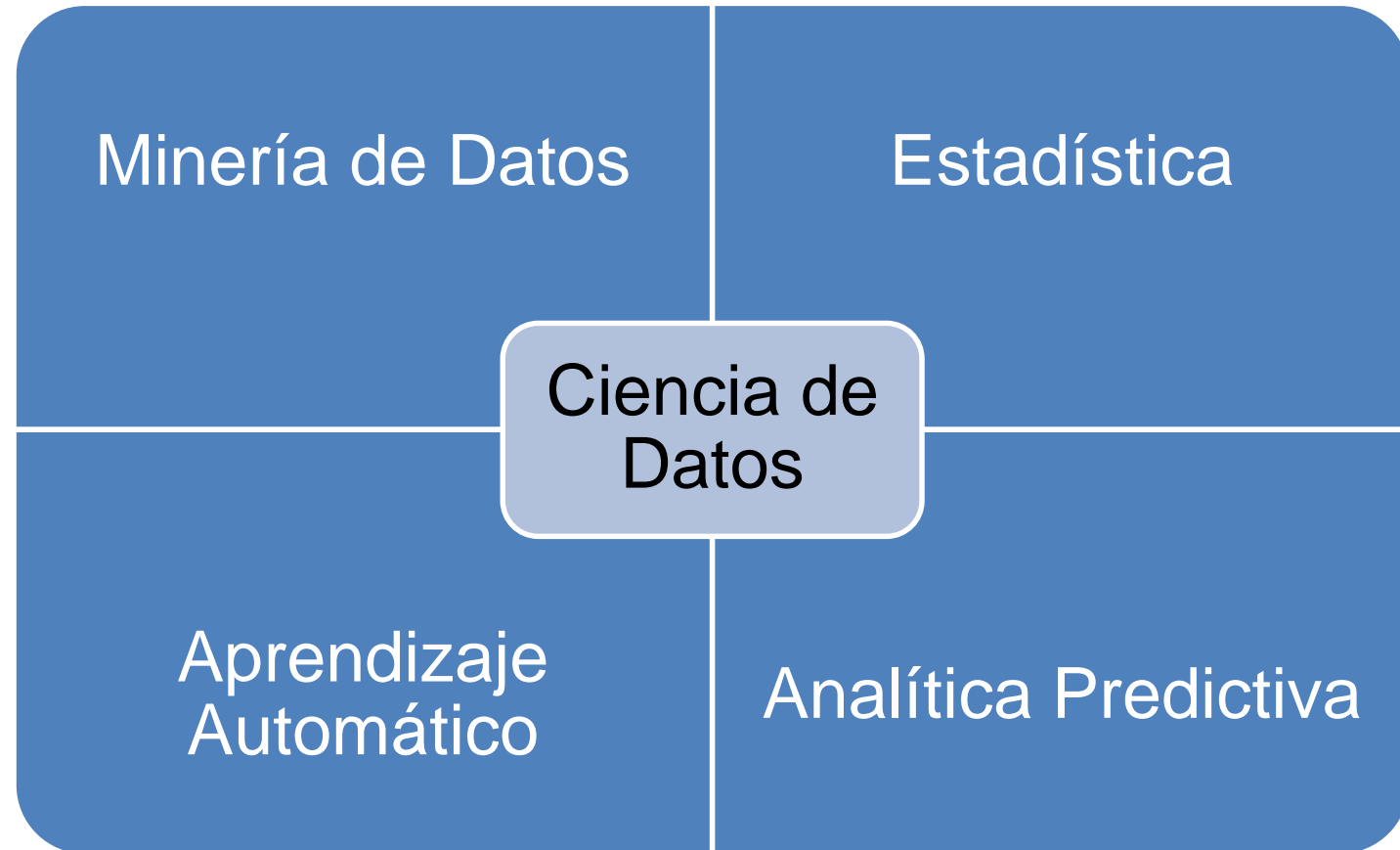
## ¿Qué es la Ciencia de Datos?

También se define La ciencia de datos como "un concepto para unificar estadísticas, análisis de datos, aprendizaje automático, y sus métodos relacionados, a efectos de comprender y analizar los fenómenos reales", empleando técnicas y teorías extraídas de muchos campos dentro del contexto de las matemáticas, la estadística, la ciencia de la información, y la informática.

## ¿Cómo se nutre la ciencia de datos?



¿Dónde encontramos la ciencia de datos?



# ¿Qué es la Ciencia de Datos?

**La minería de datos o exploración de datos** es un campo de la estadística y las ciencias de la computación referido al proceso que intenta descubrir patrones en grandes volúmenes de conjuntos de datos.

**La estadística** es la rama de las matemáticas que estudia la variabilidad, colección, organización, análisis, interpretación, y presentación de los datos, así como el proceso aleatorio que los genera siguiendo las leyes de la probabilidad.



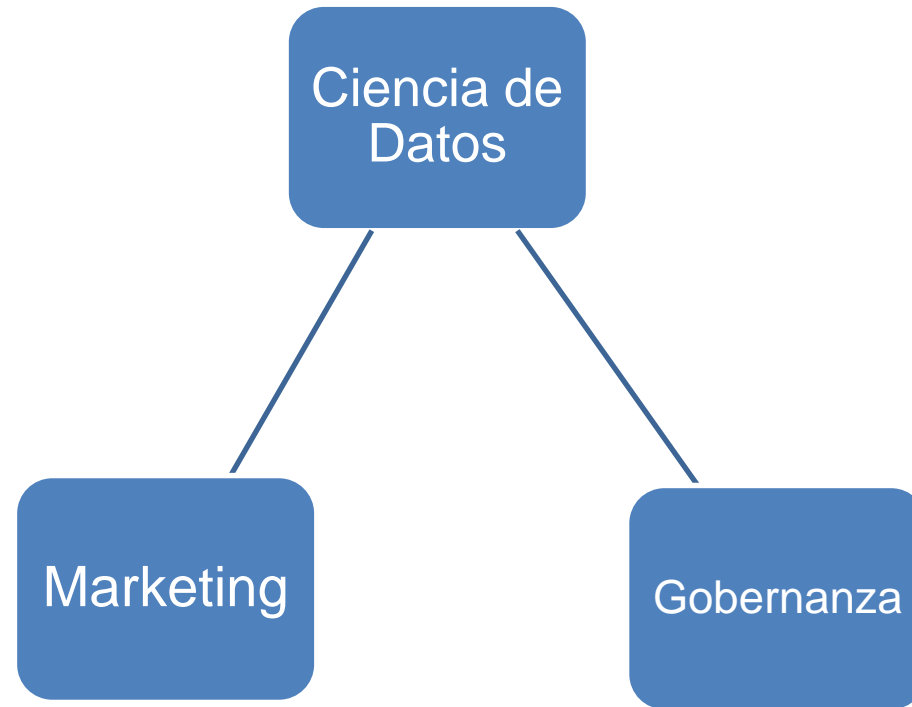
# ¿Qué es la Ciencia de Datos?

**El aprendizaje automático (machine learning)** es el subcampo de las ciencias de la computación y una rama de la inteligencia artificial, cuyo objetivo es desarrollar técnicas que permitan que las computadoras aprendan.

**El análisis predictivo** agrupa una variedad de técnicas estadísticas de modelización, aprendizaje automático y minería de datos que analiza los datos actuales e históricos reales para hacer predicciones acerca del futuro o acontecimientos no conocidos.

# ¿Qué es la Ciencia de Datos?

## Aplicaciones mas comunes



# ¿Datos, información, conocimiento y decisiones?

Dato.

En informática, los datos son representaciones simbólicas (vale decir: numéricas, alfabéticas, algorítmicas, etc.) de un determinado atributo o variable cualitativa o cuantitativa, o sea: la descripción codificada de un hecho empírico, un suceso, una entidad.

La información es un conjunto organizado de datos relevantes para uno o más sujetos que extraen de él un conocimiento.



¿Datos, información, conocimiento y decisiones?

¿Qué es el conocimiento?

Resulta sumamente complicado definir el conocimiento o establecer sus límites conceptuales. La mayoría de las aproximaciones a qué cosa es, dependen siempre de la perspectiva filosófica y teórica que uno posea, dado que existen conocimientos relacionados a todas las ramas del saber humano, y también a todas las áreas de la experiencia.

# ¿Datos, información, conocimiento y decisiones?

La toma de decisiones basada en el conocimiento

La gestión del conocimiento ayuda a incrementar y mejorar las competencias, así como a lograr mayor eficiencia en los procesos y las operaciones. Al pasar del dato a la comprensión de la información, es decir, al conocimiento se tienen las condiciones para mejorar la toma de decisiones.

## ¿Como medimos la información?

Así como usamos medidas para saber cuánto pesan o miden las cosas, también hay unidades de medida que te permiten calcular la capacidad de almacenamiento de información o procesamiento de datos.

Nombre	Símbolo	Representa
byte	B	8 bits
kilobyte	KB	1 000 B
megabyte	MB	1 000 000 B
gigabyte	GB	1 000 000 000 B
terabyte	TB	1 000 000 000 000 B
petabyte	PB	1 000 000 000 000 000 B
exabyte	EB	1 000 000 000 000 000 000 B
zettabyte	ZB	1 000 000 000 000 000 000 000 B
yottabyte	YB	1 000 000 000 000 000 000 000 000 B
saganbyte	SB	1 000 000 000 000 000 000 000 000 000 B
jotabyte	JB	1 000 000 000 000 000 000 000 000 000 000 B

## ¿Como medimos la información?

Todos los que tenemos estanterías repletas de libros, gavetas llenas de memorias externas, DVD, discos duros y cajas donde no cabe una fotografía más entendemos bien el problema: almacenar información a largo plazo es un dolor de cabeza cada vez más desafiante, por no decir caro.

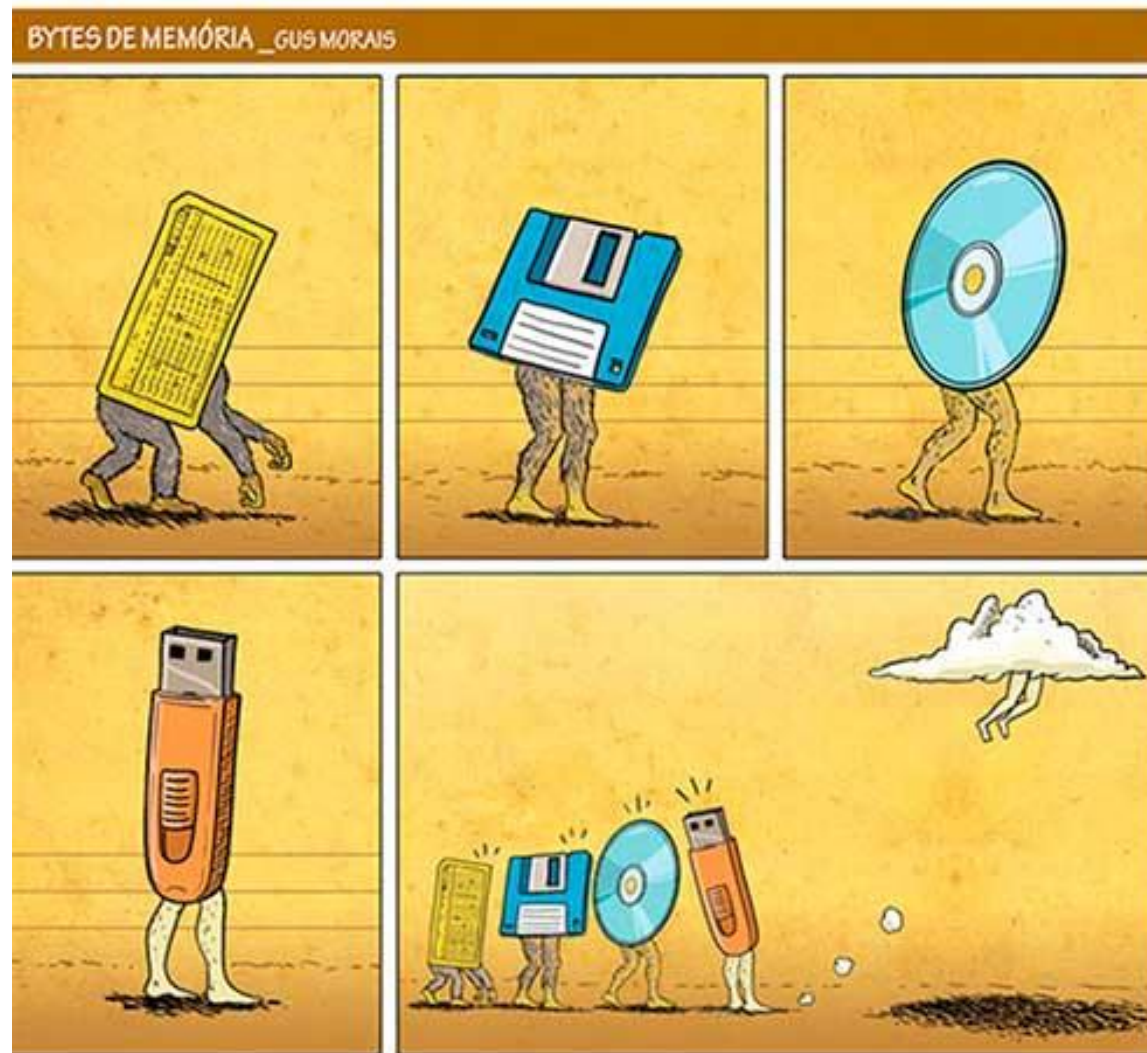
No importa si se trata de las fotos del paseo a la playa, la tesis de doctorado o los trillones de datos que arroja una investigación científica, la cuestión se pone cada vez peor. En este momento hay en el planeta Tierra alrededor de 10 billones de gigabytes de datos digitales, y todos los días, los humanos producimos correos electrónicos, fotos, tuits y otros archivos que suman otros 2,5 millones de gigabytes.

¿Qué nos depara el futuro del almacenamiento digital?





¿Qué nos depara el futuro del almacenamiento digital?



## Big Data y Ciencia de Datos

El término "big data" se refiere a los datos que son tan grandes, rápidos o complejos que es difícil o imposible procesarlos con los métodos tradicionales. El acto de acceder y almacenar grandes cantidades de información para la analítica ha existido desde hace mucho tiempo. Pero el concepto de big data cobró impulso a principios de la década de 2000 cuando el analista de la industria, Doug Laney, articuló la definición actual de grandes datos como las tres V:

## Big Data y Ciencia de Datos

Volumen: Las organizaciones recopilan datos de diversas fuentes, como transacciones comerciales, dispositivos inteligentes (IO), equipo industrial, vídeos, medios sociales y más. En el pasado, su almacenamiento habría sido un problema - pero el almacenamiento más barato en plataformas como los data lakes y el Hadoop han aliviado la carga.

# Big Data y Ciencia de Datos

Velocidad: Con el crecimiento del Internet de las Cosas, los datos llegan a las empresas a una velocidad sin precedentes y deben ser manejados de manera oportuna. Las etiquetas RFID, los sensores y los medidores inteligentes están impulsando la necesidad de manejar estos torrentes de datos en tiempo casi real.

# Big Data y Ciencia de Datos

La Ciencia de Datos es una herramienta para abordar Big Data y luego extraer información de él.

Primero la Ciencia de Datos reúnen conjuntos de datos de múltiples disciplinas y los compila. Después de esto, aplican machine learning, análisis predictivo y sentimental. Luego se acentúan en un punto donde se pueda derivar algo. Por último, se extrae la información útil de ella.

# ¿Grandes datos y pobres conocimientos?

El hombre inteligente aprende de sus  
propios errores, el sabio aprende de los  
errores de los demás.

# RESUMEN DE SESIÓN



SECRETARÍA DE  
INNOVACIÓN