



FACULTAD DE CIENCIAS
FÍSICAS Y MATEMÁTICAS
UNIVERSIDAD DE CHILE



Clase I: Introducción

FM849 Proyecto de Ciencia de Datos: Inteligencia Artificial (IA) y sus aplicaciones

Máximo Flores Valenzuela (mflores@dcc.uchile.cl)

Universidad de Chile • 2 de agosto de 2025

Información general

- Clases todos los sábados de agosto (9:00 a 10:30 ① y 11:00 a 12:30 ②).
 - En el bloque ① usualmente son las cátedras. El ② se usará para los laboratorios, salvo el día sábado 9 de agosto.
- Las clases y laboratorios se harán en la sala **B213**. Ésta se encuentra en el segundo piso del edificio norte de Beauchef 851.
- Cada uno/a de ustedes tendrá un computador con un usuario y una clave única para trabajar en el proyecto.
- En el transcurso del curso, se subirá al menos 1 video donde se explicarán los fundamentos de las tecnologías de punta de hoy en día para quienes estén interesados/as. No es obligatorio verlos.

Equipo docente

Profesor de Cátedra: Máximo Flores Valenzuela

- Participante de la EdV en varias oportunidades: 2019, 2020, 2021 como estudiante, y 2023 como Prof. Auxiliar.
- En proceso de cursar el doble título de Magíster en Ciencia de Datos con Ingeniería Civil en Computación.
- Actualmente estoy trabajando part-time en TOCTOC (toctoc.com). Aplico los conocimientos que veremos en este curso.



Equipo docente

Profesor Auxiliar: Héctor Jiménez Orellana

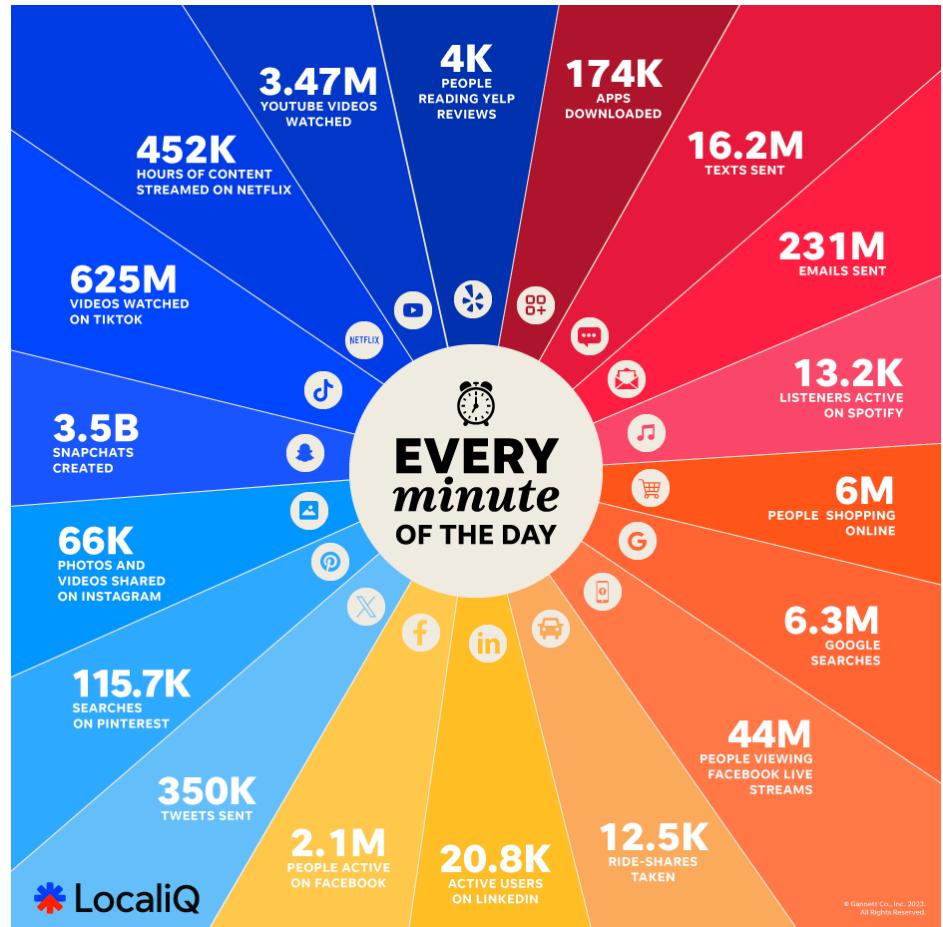
- Ingeniero Civil con Mención en Ingeniería Hidráulica, Sanitaria y Ambiental.
- Magíster en Ciencias de la Ingeniería, Mención Recursos y Medio Ambiente Hídrico.
- Estudiante del Magíster en Ciencias, Mención Computación.
- Trabajó en 2 proyectos asociados a videojuegos: de Steam y Sokoban.



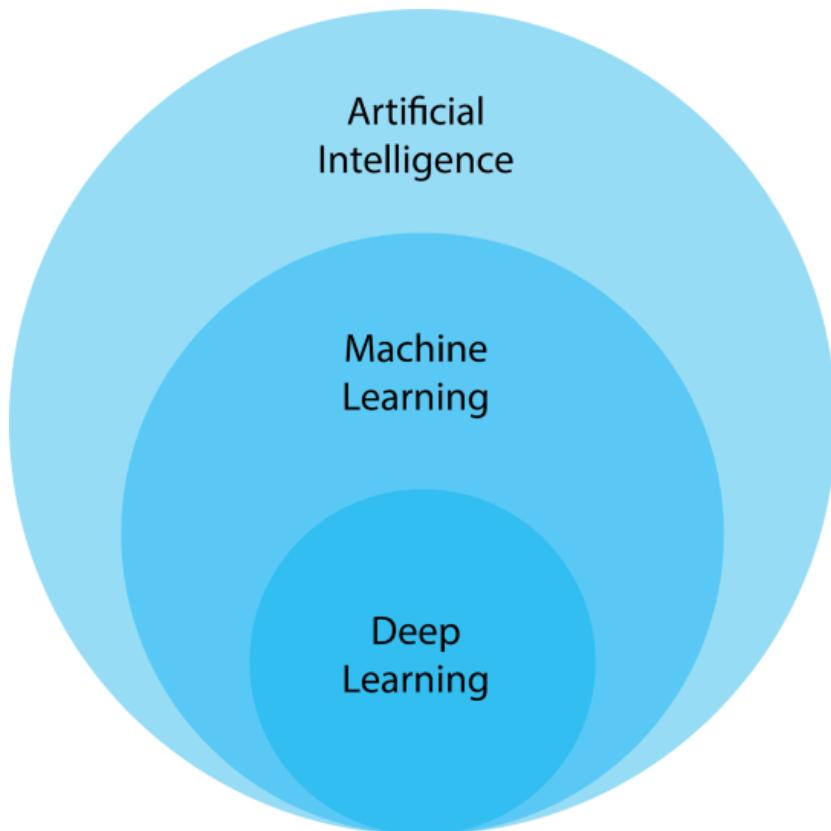
¿De qué se trata el curso?

En esta era digital, es fundamental poseer conocimientos sobre **Ciencia de Datos**. No sólo para aplicarlo en trabajos, sino para adquirir una visión más crítica del mundo.

«*Sin datos, eres solo otra persona con una opinión*»,
de W. Edwards Deming.



¿De qué se trata el curso?



El **aprendizaje de máquinas** (ML) es un área que usa algoritmos y conceptos estadísticos sobre datos para realizar una determinada tarea.

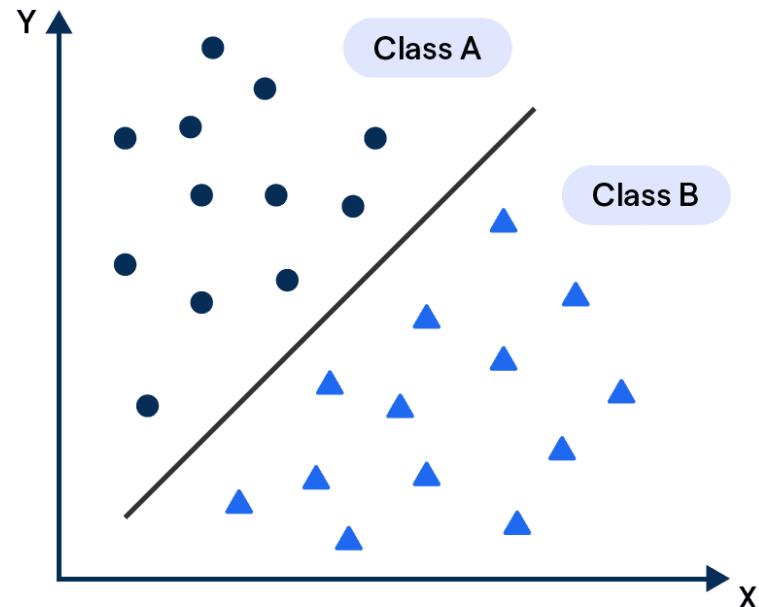
- Generalmente, dicha tarea es **clasificar** o **agrupar**.
- Muchos de los algoritmos de hoy en día usan como base algoritmos de ML.

¿De qué se trata el curso?

La **clasificación** es la tarea de separación de entidades con características, asignándoles una etiqueta (aprendizaje supervisado). En la imagen, las etiquetas son «A» y «B», también llamadas clases.

- ¡Ojo !! Los problemas no están necesariamente restringidos a dos clases.

Classification Algorithm



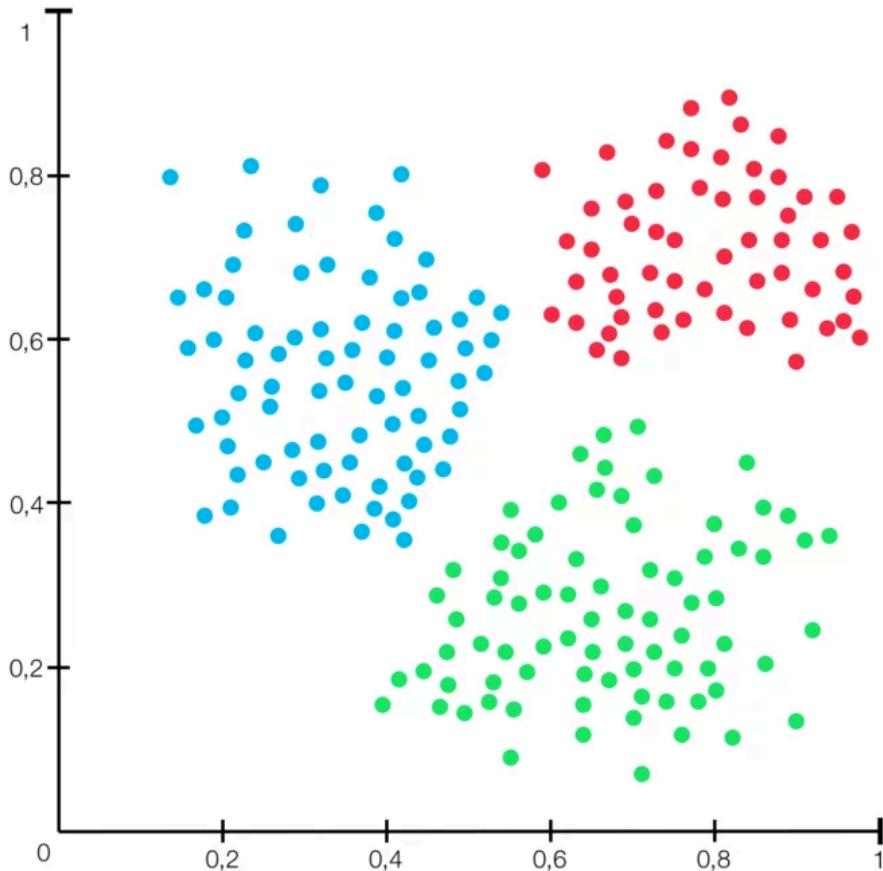
■ ¿De qué se trata el curso?

Ejemplos de **clasificación** (algunos requieren estrategias más sofisticadas...):

- Dada la información sobre un perfil de Instagram (nombre de usuario, antigüedad de la cuenta, interacciones), reportar si es un bot o no. **B**
- Dado un texto clínico, deducir sobre qué área médica se está hablando. **M**
- Dado el audio de una canción en formato .mp4, reportar a qué género pertenece. **M**
- ¡Y mucho más...! El límite es su imaginación, y obviamente los recursos de procesamiento.

(**B** := clasificación binaria; **M** := clasificación multiclas)

¿De qué se trata el curso?



El **agrupamiento** (o *clustering*) es la tarea de encontrar grupos donde los elementos dentro de cada uno comparten características.

A diferencia de la clasificación, acá los datos no están etiquetados, entonces se habla de aprendizaje no supervisado.

■ ¿De qué se trata el curso?

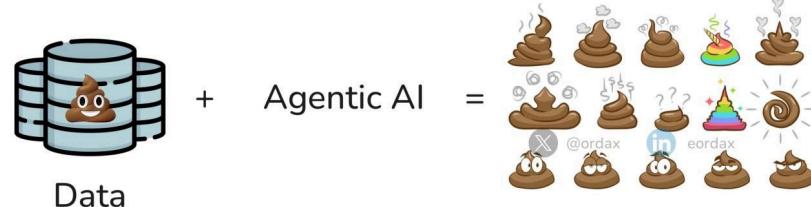
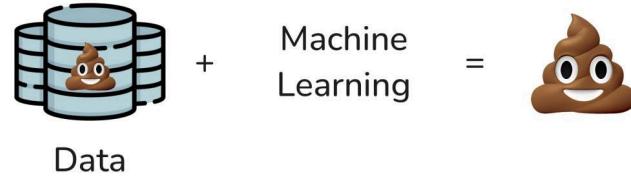
Ejemplos de **agrupamiento**:

- Dada una página de venta de productos, agrupar a los clientes según sus preferencias para mostrar anuncios personalizados.
- Detectar comunidades en redes sociales de acuerdo a las interacciones entre determinadas cuentas.
- Detectar documentos que tratan tópicos similares analizando su contenido.

¿De qué se trata el curso?

En la práctica, completar estas tareas no es tan fácil. Los *datasets* en la generalidad de los casos vienen con problemas de formato, datos incompletos, redundantes, etc.

Veremos la importancia de realizar un análisis exploratorio de los datos antes de aplicar un modelo de IA.



■ ¿De qué se trata el curso?

Como valor agregado, buscamos responder las siguientes interrogantes:

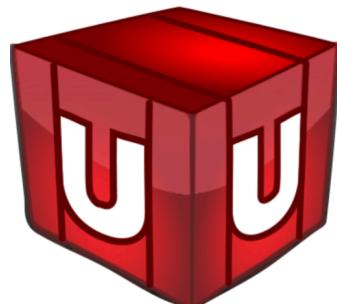
- ¿Qué debo considerar para levantar una problemática?
- ¿...y para la propuesta de una solución?
- ¿Cómo puedo transmitir de manera efectiva mis ideas considerando el contexto de una presentación?

Existen estándares, pero no son reglas ineludibles...

Comunicación

El canal oficial de comunicación del curso es la página web U-Cursos (u-cursos.cl).

- Se subirá todo el material asociado: presentaciones, enlaces de interés, videos asíncronos, notas, avisos, etc.
- Puede ser difícil de usar al principio... me pueden contactar por correo electrónico mflores@dcc.uchile.cl y los/as orientaré.



Metodología de evaluación

El curso tiene 4 sesiones prácticas y una presentación final.

- Las sesiones prácticas son los días 2, 16, 23 y 30 de agosto. La presentación final es este último día.
¡Ojo !! El día 9 de agosto hay sólo clases de cátedra.
- En cada sesión práctica se evaluarán avances en el proyecto. Para tener un 7.0 basta con realizar avances significativos.
- La nota final (NF) del curso se calculará con la siguiente fórmula:

$$NF = 80\% \cdot NP + 20\% \cdot NL$$

donde NP es la nota del proyecto, y NL la nota de laboratorios, calculada como el promedio de las notas de los avances en sesiones prácticas.

Metodología de evaluación

- En algunas cátedras se realizarán actividades con bonificación por participación. Se usará la plataforma Menti (menti.com).
- Está permitido el uso de modelos grandes de lenguaje (p. ej., GPT) para apoyarse en la implementación del proyecto. Sin embargo, **toda** decisión debe poder ser explicada por Uds.
 - ▶ En la presentación final habrá un tiempo de discusión para poder hacer preguntas sobre los proyectos, que también será evaluado.

¡A revisar el tutorial de la clase!