



# Introducción a la cátedra

Ing. Juan M. Rodríguez





## Slack como canal de comunicación





## Docentes

- Juan M. Rodríguez
- Azul Villanueva
- Marcelo Benítez



## Colaboradores

- Matías González
- Luis Paredes
- Adrian Romero
- Darius Maitia
- Ignacio Velasco
- Francisco Pereira
- Axel Perez Machado



## Modalidad de cursada

- Teórica: Martes 19:00 - 22:00 hs virtual
- Práctica: Jueves 19:00 - 22:00 hs virtual



## ¿Cómo aprobar?

- 2 trabajos prácticos durante la cursada (una reentrega por TP)
- Parcial (2 recuperatorios)
- Exámen Integrador (Coloquio)



## Grupos

- N alumnos máximo por grupo (a definir N)
- Un colaborador asignado





# Calendario

Semana	Fecha	Teórica (Martes)	Fecha	Práctica (Jueves)	
1	23/08/2022	Presentación de la materia. Visualización de datos.	25/08/2022	Introducción a Colab y maquinas Jupyter Pandas Librerías de visualización.	
2	30/08/2022	Introducción a la ciencia de datos. Métodos de regresión Métricas	01/09/2022	Introducción a la ciencia de datos. Métodos de regresión Métricas	
3	06/09/2022	Métodos de clasificación Métodos de agrupamiento (clustering) Metricas	08/09/2022	Métodos de clasificación Métodos de agrupamiento (clustering) Metricas	
4	13/09/2022	Ingeniería de características	15/09/2022	Ingeniería de características	
5	20/09/2022	Árboles de decisión	22/09/2022	Árboles de decisión con scikit-learn <b>Presentación TP1</b>	
6	27/09/2022	Reducción de la dimensionalidad	29/09/2022	Reducción de la dimensionalidad	
7	04/10/2022	k-nearest neighbors algorithm Support Vector Machines (SVM)	06/10/2022	k-nearest neighbors algorithm Support Vector Machines (SVM)	
8	11/10/2022	Ensamble de modelos	13/10/2022	Ensamble de modelos con scikit-learn y dmlc XGBoost	

## Herramientas y tecnologías a utilizar

- Python 3:
  - Tensorflow,
  - Keras,
  - Numpy,
  - Pandas



Keras



pandas

## Herramientas y tecnologías a utilizar



- Python 3:
  - PyTorch (a través de Yolo),
  - Yolo,
  - TPOT,
  - scikit-learn
- Weka



## Herramientas y tecnologías a utilizar

- Máquinas Jupyter
- [Google Colab](#)
- [Kaggle](#)



---

# Introducción a la materia



# Qué es un Data Scientist?

Es capaz de:

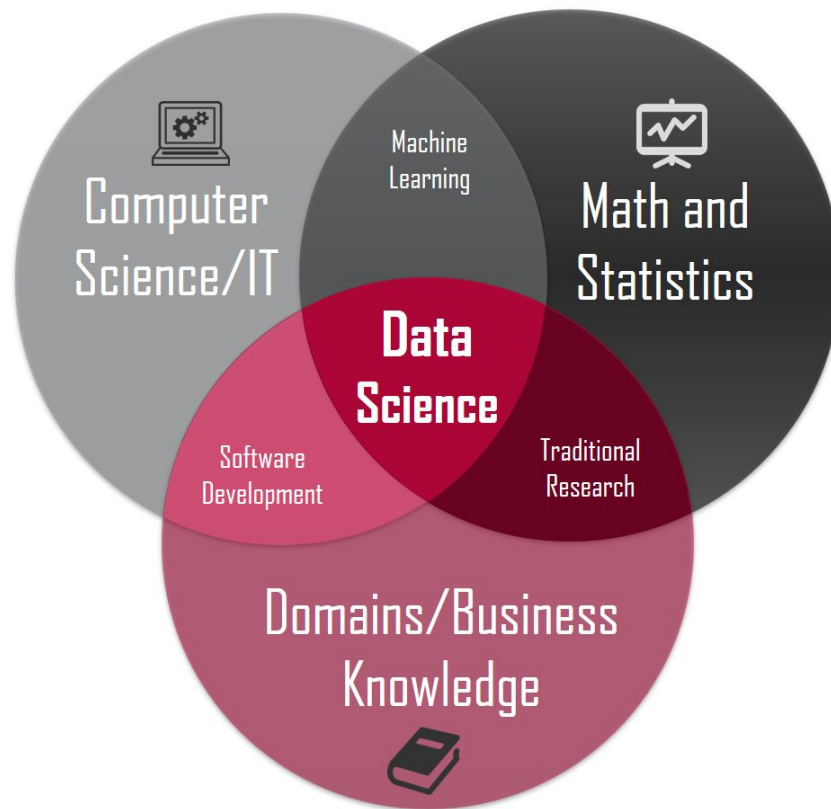
- Obtener, Interpretar, procesar y filtrar los datos.
- Llegar a conclusiones a partir de lo anterior.
- Construir soluciones para los problemas que se quiere solucionar.



# Data Scientist != Data Engineer

- Considerable solapamiento entre las habilidades y responsabilidades.
- Hay una importante diferencia en el enfoque.
- El ingeniero de datos se enfoca en la **creación de la infraestructura y arquitectura para la generación, soporte y extracción de los datos. (Big Data)**
- El científico de datos se enfoca en la interpretación y análisis de los datos mediante el análisis estadístico y matemático.

# Dominios de conocimiento





# Habilidades técnicas

- Data mining



- Visualización de datos



- Matemática y Estadística

- Machine Learning



- Plataformas: Linux, AWS, Google Cloud, etc.

---

¿Qué es *machine learning*?



# Machine Learning (Aprendizaje automático)

El aprendizaje automático es la ciencia (y el arte) de programar computadoras para que aprendan a partir de datos.

(Aurélien Géron, 2019)



# Machine Learning (Aprendizaje automático)

Se dice que un programa de computadoras aprende de la experiencia  $E$ , respecto de una tarea  $T$  y una medida de rendimiento  $R$ , si su rendimiento en  $T$ , medido por  $R$ , mejora la experiencia  $E$ .

(Tom Mitchell, 1997)



# Machine Learning (Aprendizaje automático)

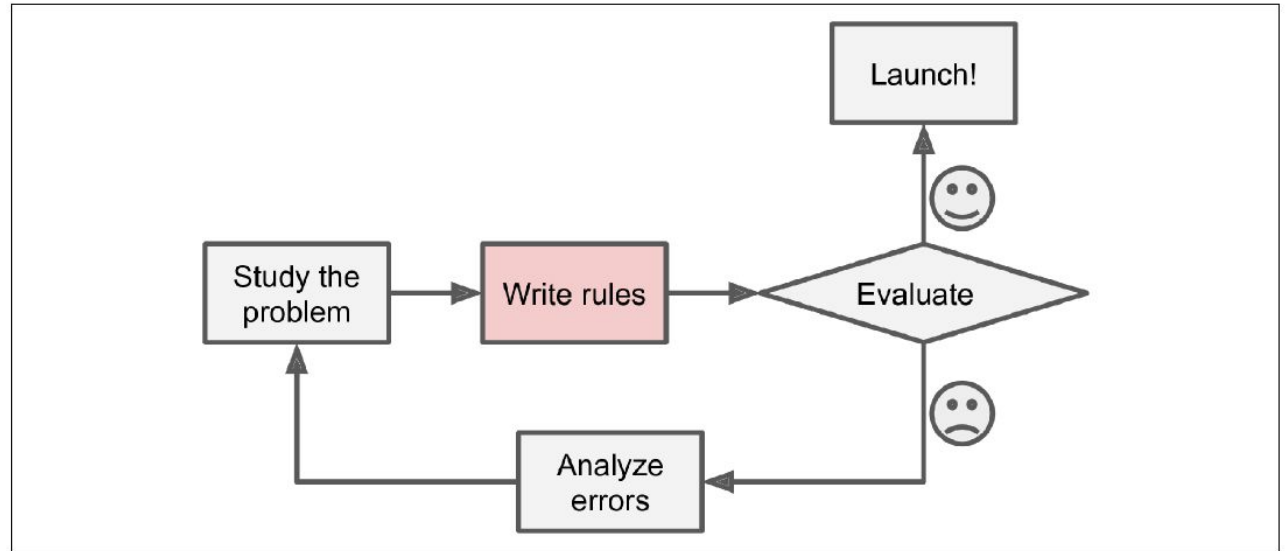
El aprendizaje automático es el campo de estudio que da a las computadoras la capacidad de aprender sin ser programadas de manera explícita.

(Arthur Samuel, 1959)

# Machine Learning (Aprendizaje automático)

Filtro de SPAM

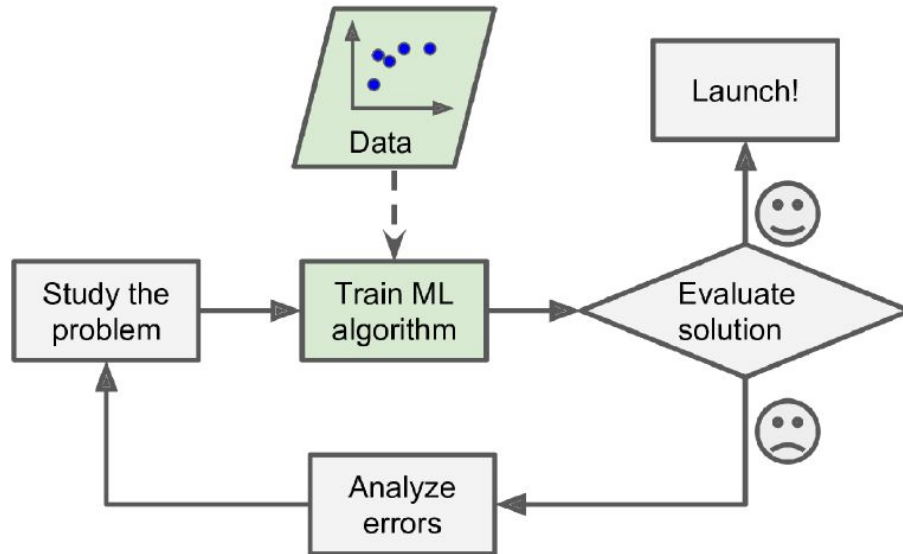
Enfoque tradicional



# Machine Learning (Aprendizaje automático)

Filtro de SPAM

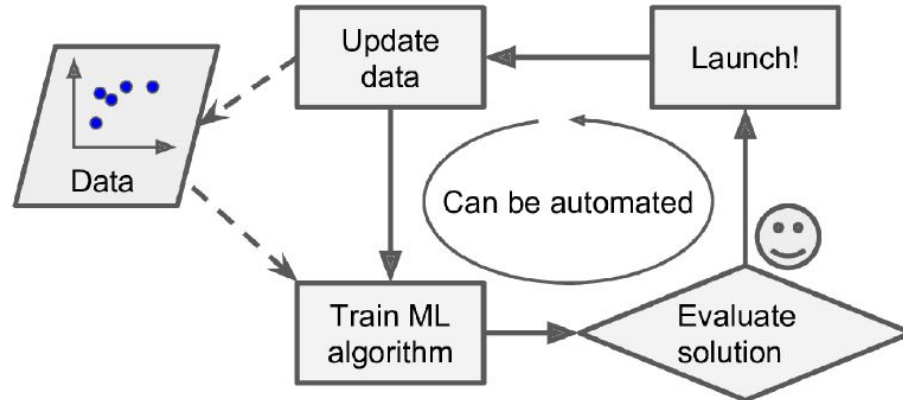
Enfoque ML



# Machine Learning (Aprendizaje automático)

Filtro de SPAM - enfoque ML

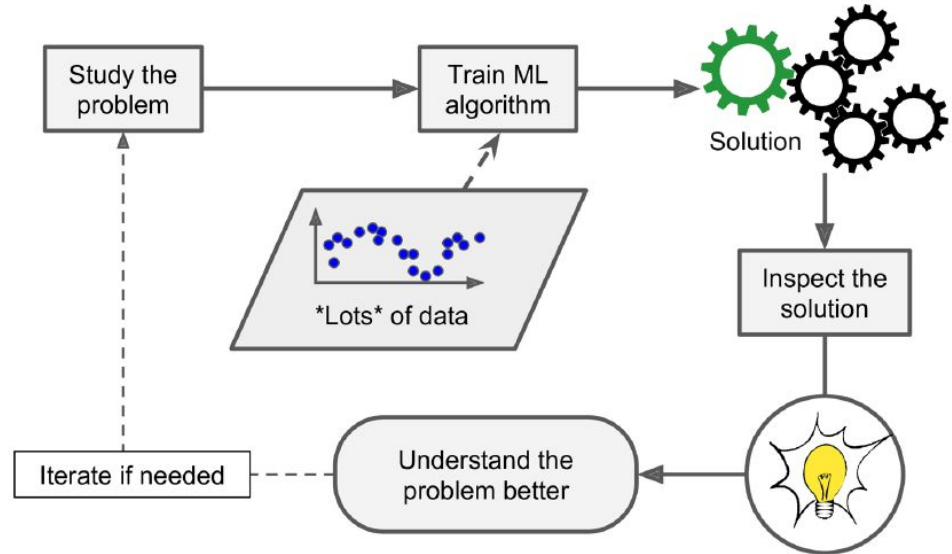
Adaptación al cambio de forma automática





# Machine Learning (Aprendizaje automático)

Pueden ayudarnos a aprender

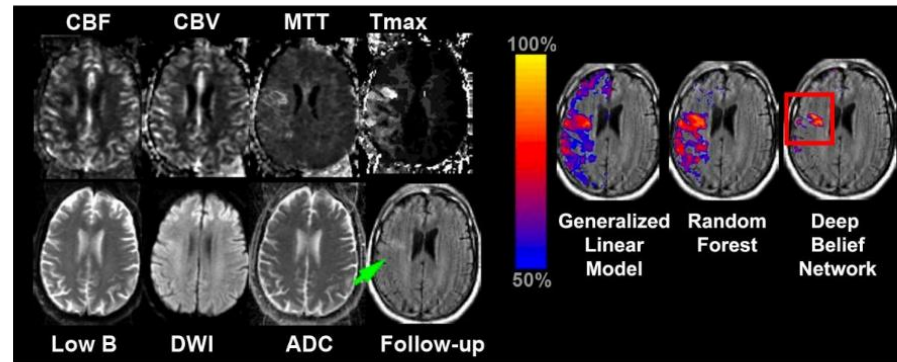


---

¿Para qué se usa?

# Áreas de aplicación: Salud

- Pre-diagnosticos en base a historial clínico
- Detección temprana de tumores
- Generación de nuevas drogas
- Test de [coronavirus a través de la voz?](#)
- Psiquiatría (depresión "pura" y psicosis "pura")
- Epidemiología



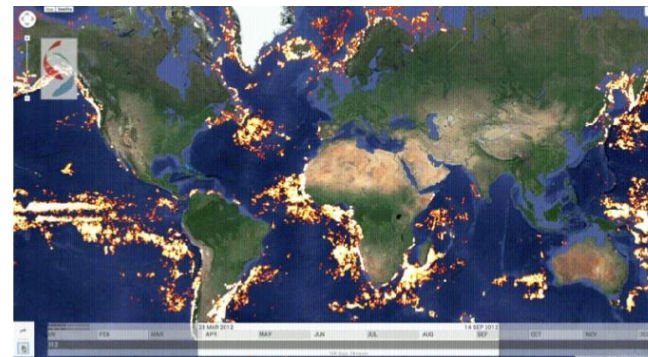
# Áreas de aplicación: Gaming y Videos

- Aumentar la conexión emocional (*engagement*) de los jugadores
- Monetización del juego
- Mejorar la calidad visual de los juegos
- Hacer jugadores virtuales de distintas dificultades
- Elegir el [mejor thumbnail](#) para un video



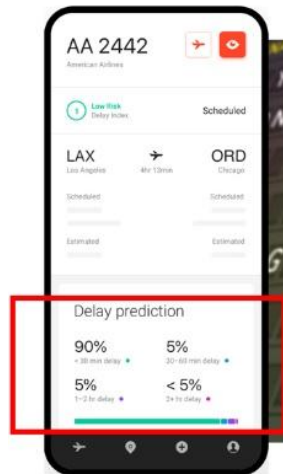
# Áreas de aplicación: Energía

- Proporcionar costos e inversiones óptimos
- Minimizar riesgos
- Descubrir nuevas fuentes de energía
- Optimizar y eficientizar el abastecimiento
- Maximizar extracción/generación



# Áreas de aplicación: Turismo

- Crear perfiles de viajeros en base a redes sociales y otras fuentes
- Optimización de rutas aéreas
- Búsqueda de hospedajes, actividades, destinos
- Optimización de costos

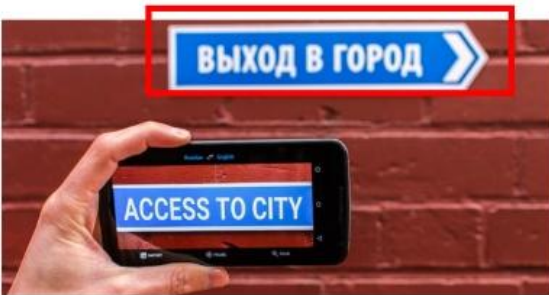
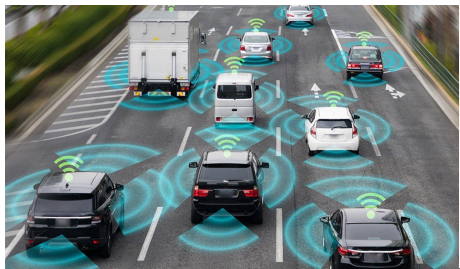


# Áreas de aplicación: Seguridad

- Detección de objetos en imagen.
  - Personas
  - Fallas en tuberías
  - Deteccion de humo o incendios



# Más y más aplicaciones

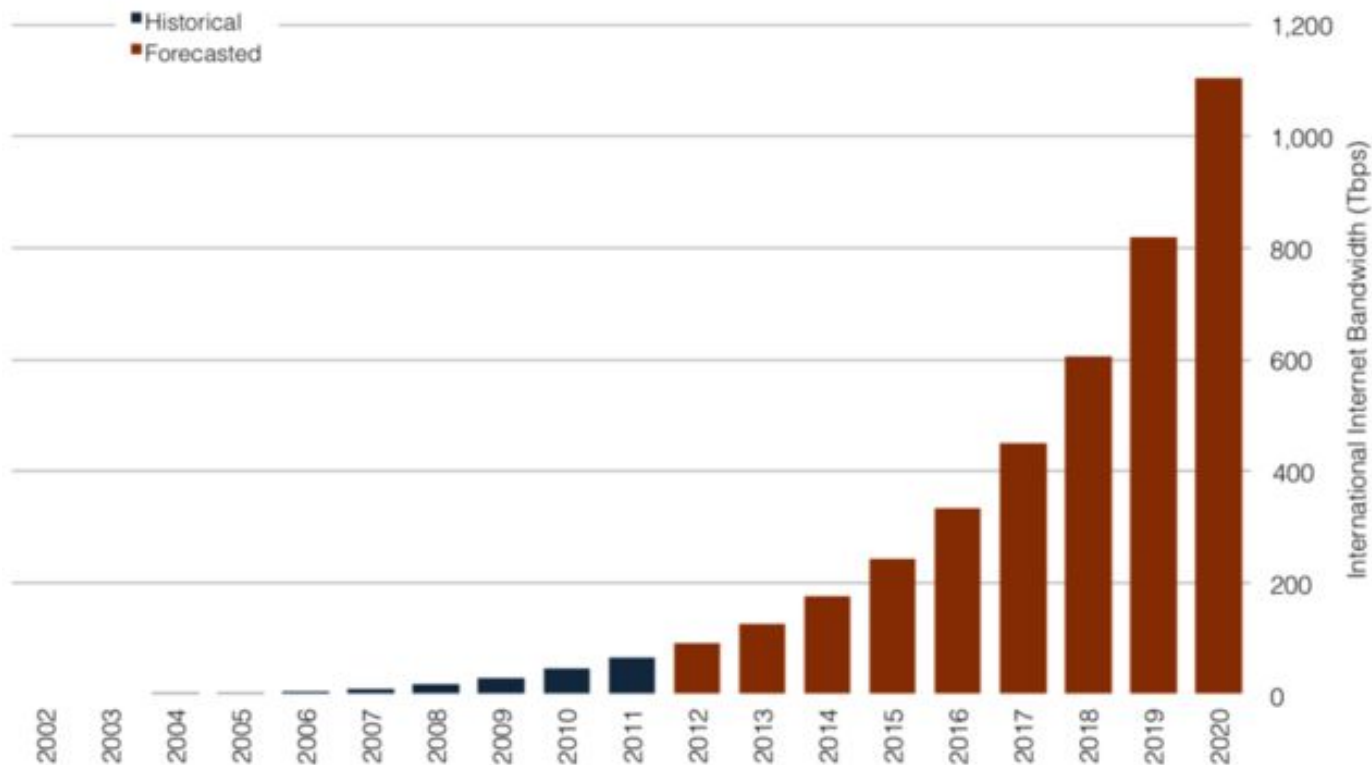




# ¿Por qué ahora?

## Uso de Internet

Used International Bandwidth, 2002-2020

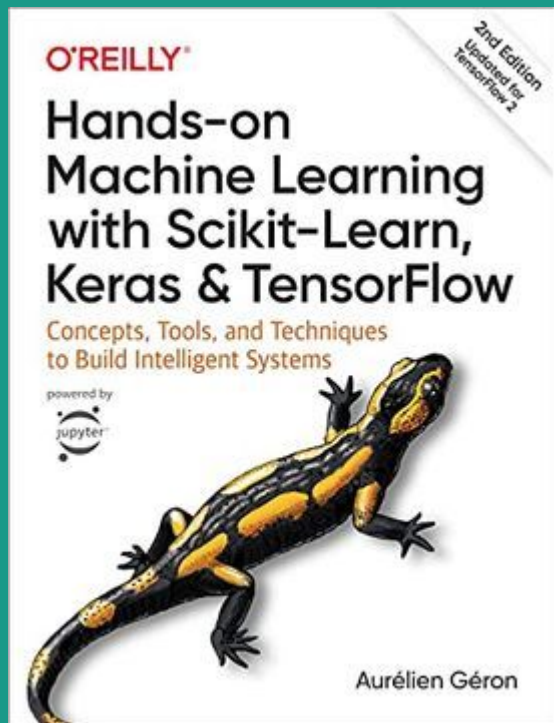


# Metodología



# Bibliografía

---



Aurélien Géron

También en español

