



Introducción a la cátedra

Ing. Juan M. Rodríguez



Slack como canal de comunicación





Docentes

- Juan M. Rodríguez
- Karen Roberts
- Cecilia Azul Villanueva
- Carlos Marcelo Benítez



Colaboradores

- Santiago Locatelli
- Agustina Bocaccio
- Matías González
- Luis Paredes
- Ezequiel Vilaro
- Mateo Craviotto
- Nicolas Pinto
- Alejandro Paff
- Adrian Romero
- Jose Tomas Lorente
- Darius Maitia



Modalidad de cursada

- **Práctica:** Lunes 19:00 - 22:00 hs virtual
- **Teórica:** Miércoles 19:00 - 22:00 virtual



¿Cómo aprobar?

- 1 trabajo práctico durante la cursada (no hay reentrega)
- 1 trabajo práctico durante la cursada (una reentrega)
- Parcial. (2 recuperatorios)
- 1 trabajo práctico final, con posibilidad de ser entregado en cualquier fecha de coloquio. (2 posibilidades de reentrega)
- Defensa oral del último TP.



Grupos

- 4 alumnos máximo por grupo
- un colaborador asignado

Calendario

Semana	Fecha	Práctica virtual	Fecha	Teórica presencial
1	21/03/2022	- Presentación de la materia.	23/03/2022	-Pandas. - Visualización de datos.
2	28/03/2022	- Pandas - Presentación del TP1	30/03/2022	- Introducción a la ciencia de datos. - Métodos de regresión - Métodos de clasificación - Métodos de clusterización
3	04/04/2022	-Práctica ciencia de datos	06/04/2022	NLP - parte 1
4	11/04/2022	-Práctica NLP	13/04/2022	NLP - parte 2 (robutez y outliers)
5	18/04/2022	-Práctica 2 de NLP	20/04/2022	- Ingeniería de características
6	25/04/2022	- Vencimiento TP1 - Práctica Ingeniería de características	27/04/2022	- Árboles
7	02/05/2022		04/05/2022	- Presentación TP2 - Ensambls
8	09/05/2022	- Práctica Ensambls - Práctica Árboles	11/05/2022	- Redes neuronales parte 1
9	16/05/2022	- Práctica Redes neuronales	18/05/2022	
10	23/05/2022	- Teórica Redes neuronales parte 2	25/05/2022	
11	30/05/2022	- Vencimiento TP2 - Práctica redes neuronales 2	01/06/2022	- Redes neuronales profundas
12	06/06/2022		08/06/2022	- Parcial
13	13/06/2022	- Presentación del trabajo final - Reentrega TP2	15/06/2022	- SVM
14	20/06/2022		22/06/2022	- Reduccion de la dimensionalidad
15	27/06/2022	- Práctica de reducción de dimensionalidad - Práctica de SVM	29/06/2022	- Primer recuperatorio - Algoritmos genéticos
16	04/07/2022	- Práctica de algoritmos genéticos	06/07/2022	Consulta

* El 2do recuperatorio del parcial se tomará junto con la 1ra fecha de final (solo en la primera).

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:

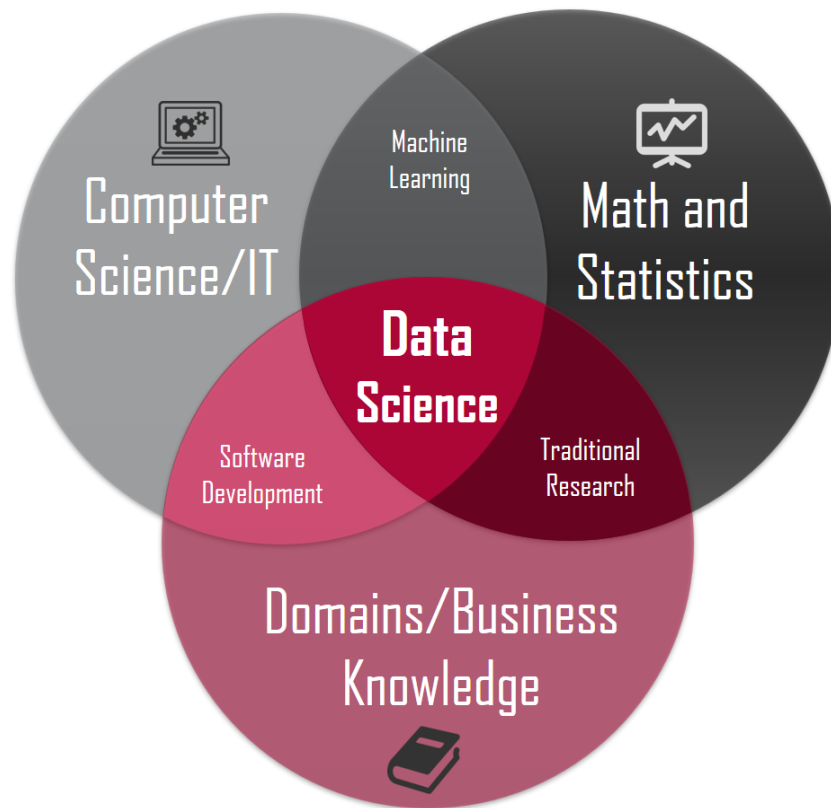
- Obtener, Interpretar, procesar y filtrar los datos.
- Llegar a conclusiones a partir de lo anterior.
- Construir soluciones para los problemas que se quiere solucionar.



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)

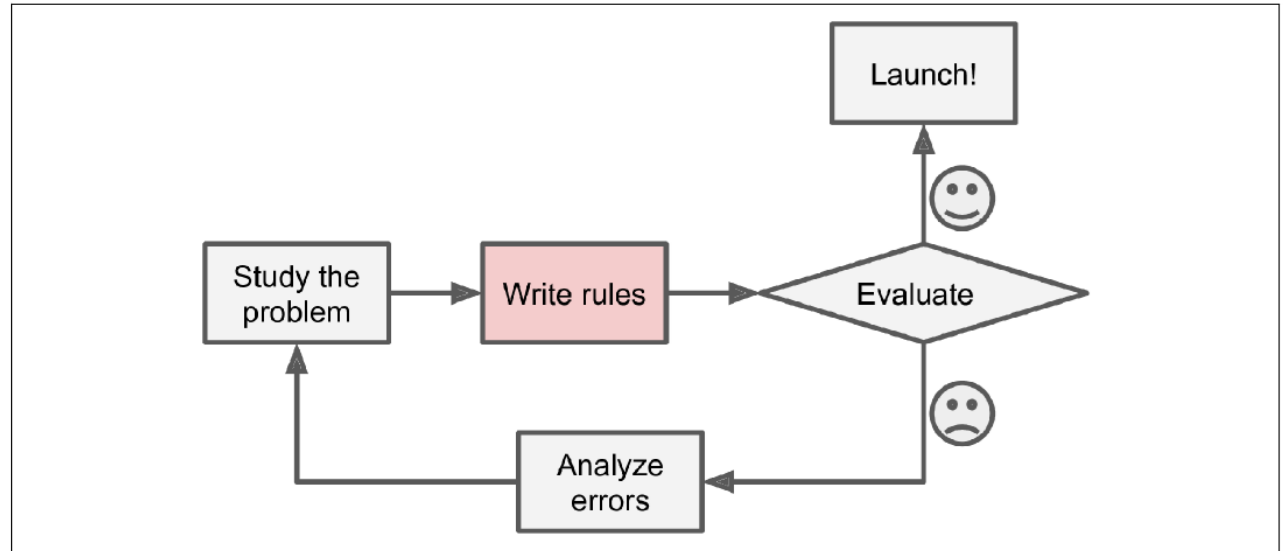
El aprendizaje automático es el campo de estudio que da a las computadoras la capacidad de aprender sin ser programadas de manera explícita.

(Arthur Samuel, 1959)

Machine Learning (Aprendizaje automático)

Filtro de SPAM

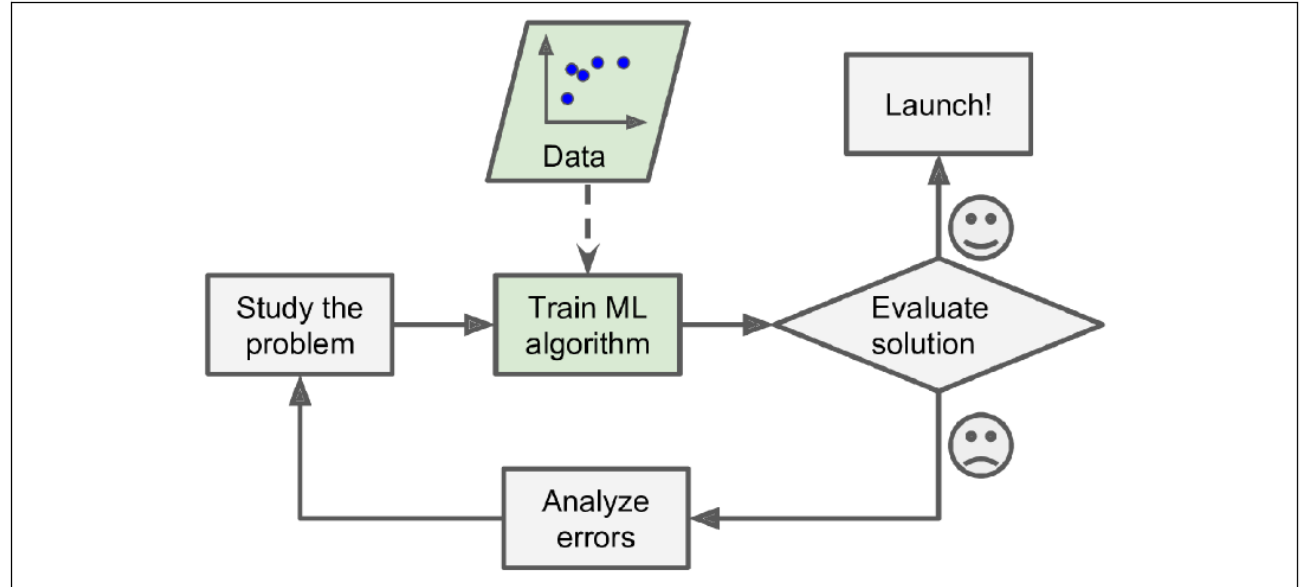
Enfoque tradicional



Machine Learning (Aprendizaje automático)

Filtro de SPAM

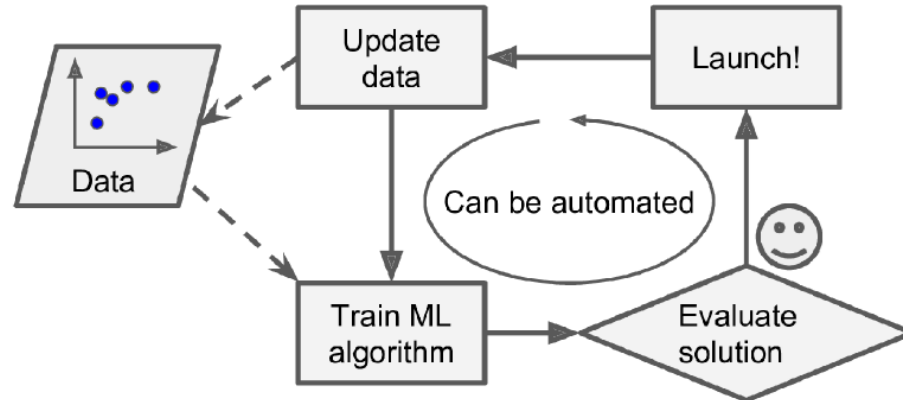
Enfoque ML



Machine Learning (Aprendizaje automático)

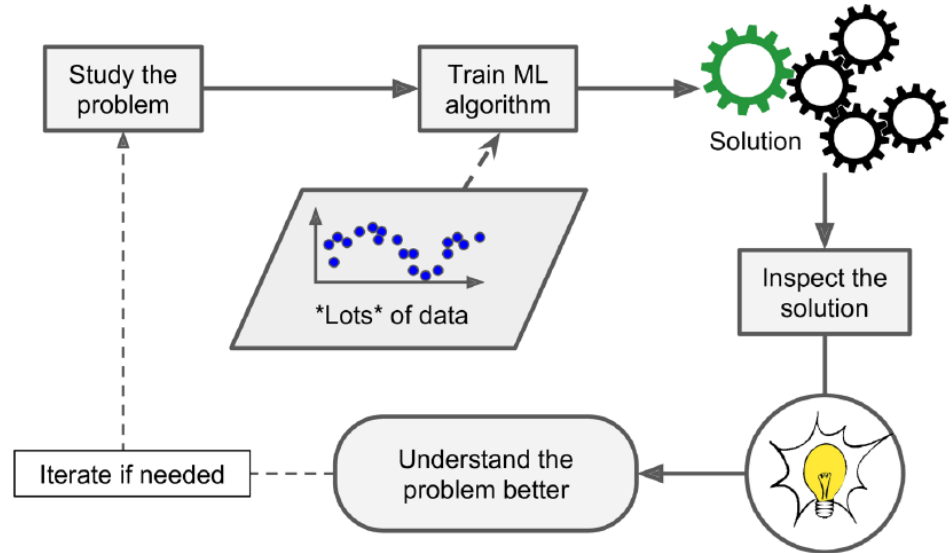
Filtro de SPAM - enfoque ML

Adaptación al cambio de forma automática



Machine Learning (Aprendizaje automático)

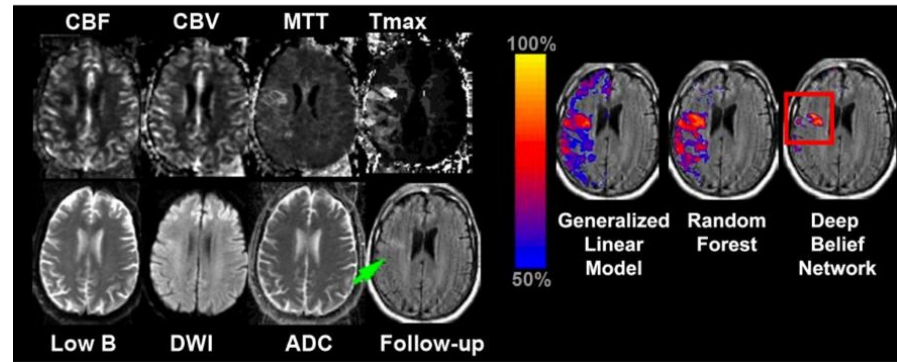
Pueden ayudarnos a aprender



¿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



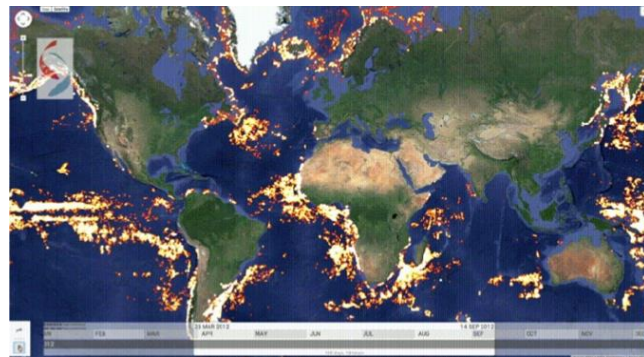
Áreas de aplicación: Gaming y Videos

- 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



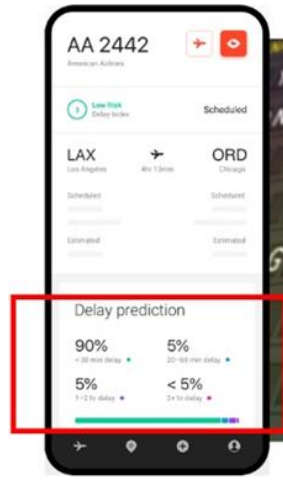
Áreas de aplicación: Energía

- 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](#)



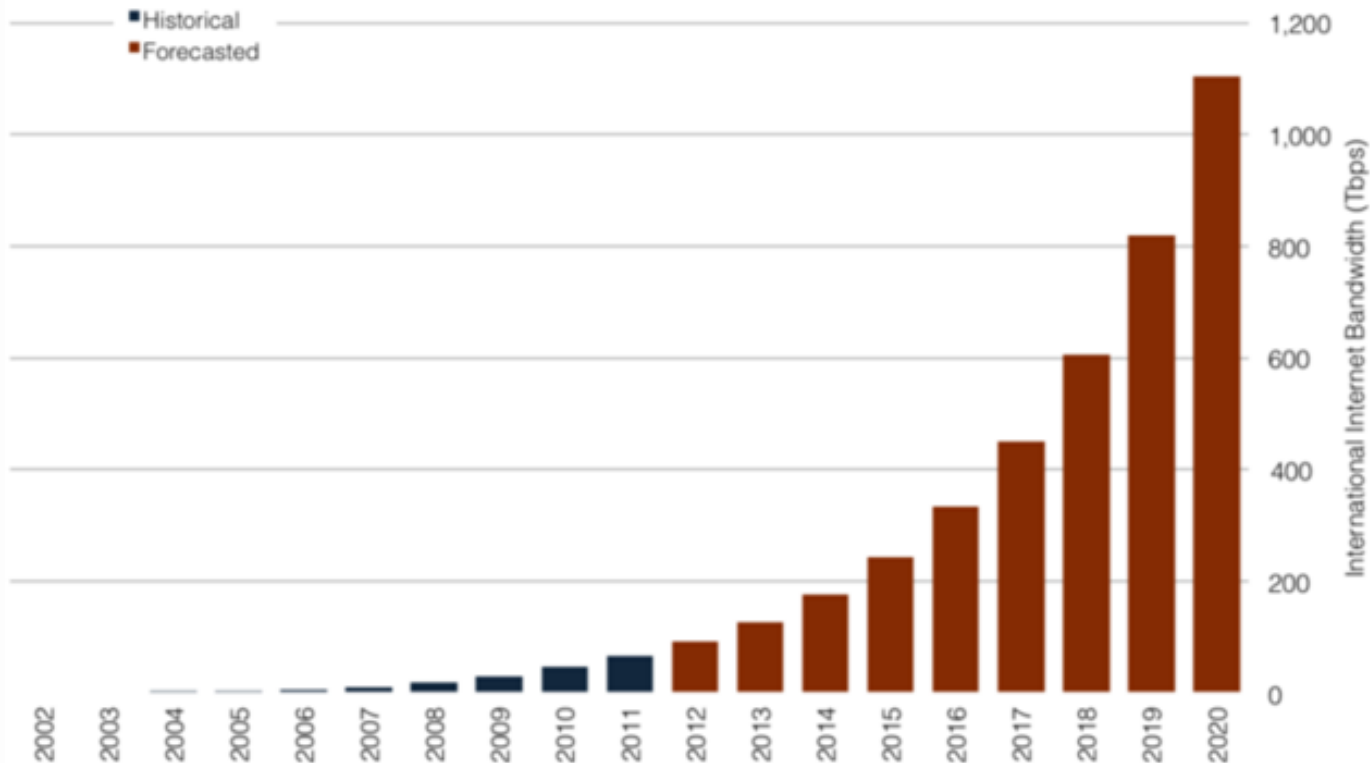
Más y más aplicaciones



¿Por qué ahora?

Uso de Internet

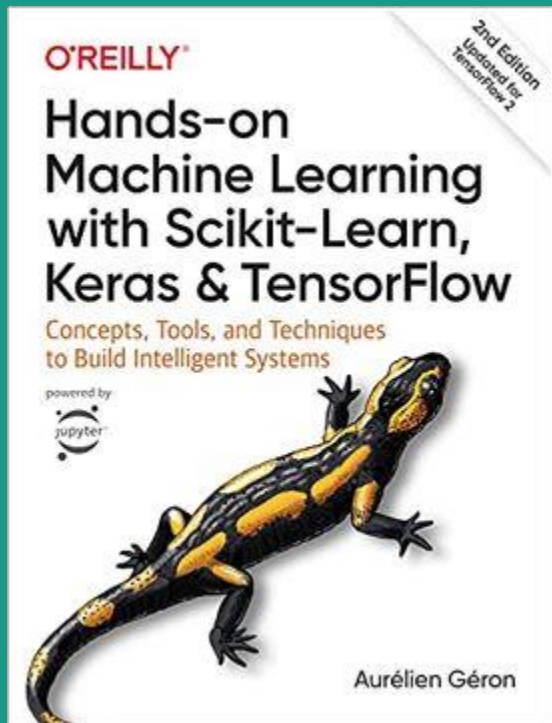
Used International Bandwidth, 2002-2020



Metodología



Bibliografía



Aurélien Géron

También en español

