

# Introducción a la IA /LM

Se va a trabajar bastante por arriba, pero cuando sea necesario se va a ondar en el problema y su resolución.

**Es necesario** poder reconocer donde están los datos **Basura** (saber cuáles son y el alcance de los mismos)

**Video: como arruinar nuestra empresa con Machine Learning**

14/10	parcial
28/11	ultimo parcial y portafolio

## Dinámica de trabajo

- Aprendizaje basado en equipos
- preparación previa básica
- breves introducciones conceptuales
- ejercicios de aplicación - modelos/ prototipos
- ejercicios domiciliarios - evaluación de modelos, programas en python
- generación incremental del portafolios y POSTER

## Evaluación

- Componentes:
  - parciales (2) 40%
  - rats 15
  - portafolio y POSTER 35 (15% y 20%)
  - Ejercicios Domiciliarios 5%

- Evaluacion de pares 5%

## Unidad Temática 1 - Introducción

Conocer conceptos de IA, marcos de trabajo y enfoque en la industria.

### Resultados

- Identificar tecnicas y herramientas dpara el tratamiento del aprendizaje automático. Ventajas y desventajas.
- Comprenden, describir y aplicar un proceso de estandar de DataScience / Machine Learning
- Tipos de algoritmos de ML existentes y sus aplicaciones
- Comenzar a utilizar una herramienta tipica industrial de modelado y ejecucion de ML
- Utilizar planillas electronicas para analisis estadisticos basicos.

### Enfoque

- Sistemas aplicados (como informaticos tratar de ver mas a fondo).
- Algoritmos como “caja negra”, más adelante analisis y ajuste de parametros
- problemas reales cominmente atacods con tecnicas de ML
- Integracion en aplicaciones de SW

## Que es machine learning?

What is machine learning? (<https://www.ibm.com/topics/machine-learning>)

El aprendizaje automático es una rama de [la inteligencia artificial \(IA\)](#) y la informática que se centra en el uso de datos y algoritmos para imitar la forma en que los humanos aprenden, mejorando gradualmente su precisión.

IBM tiene una rica [historia](#) con el aprendizaje automático. A uno de los suyos, Arthur Samuel, se le atribuye haber acuñado el término "aprendizaje automático" con su [investigación](#) (PDF, 481 KB) (el enlace reside fuera de IBM) en torno al juego de damas. Robert Nealey, el autoproclamado maestro de las damas, jugó en una computadora IBM 7094 en 1962 y perdió contra la computadora. En comparación con lo que se puede hacer hoy, esta hazaña parece trivial, pero se considera un hito importante en el campo de la inteligencia artificial.

Durante las últimas dos décadas, los avances tecnológicos en almacenamiento y poder de procesamiento han permitido algunos productos innovadores basados en el aprendizaje automático, como el motor de recomendación de Netflix y los autos sin conductor.

El aprendizaje automático es un componente importante del creciente campo de la ciencia de datos. Mediante el uso de métodos estadísticos, los algoritmos se entrenan para hacer clasificaciones o predicciones y descubrir información clave en proyectos de minería de datos. Estos conocimientos posteriormente impulsan la toma de decisiones dentro de las aplicaciones y los negocios, lo que idealmente impacta en las métricas de crecimiento clave. A medida que el big data continúa expandiéndose y creciendo, la demanda del mercado de científicos de datos aumentará. Se les pedirá que ayuden a identificar las preguntas comerciales más relevantes y los datos para responderlas.

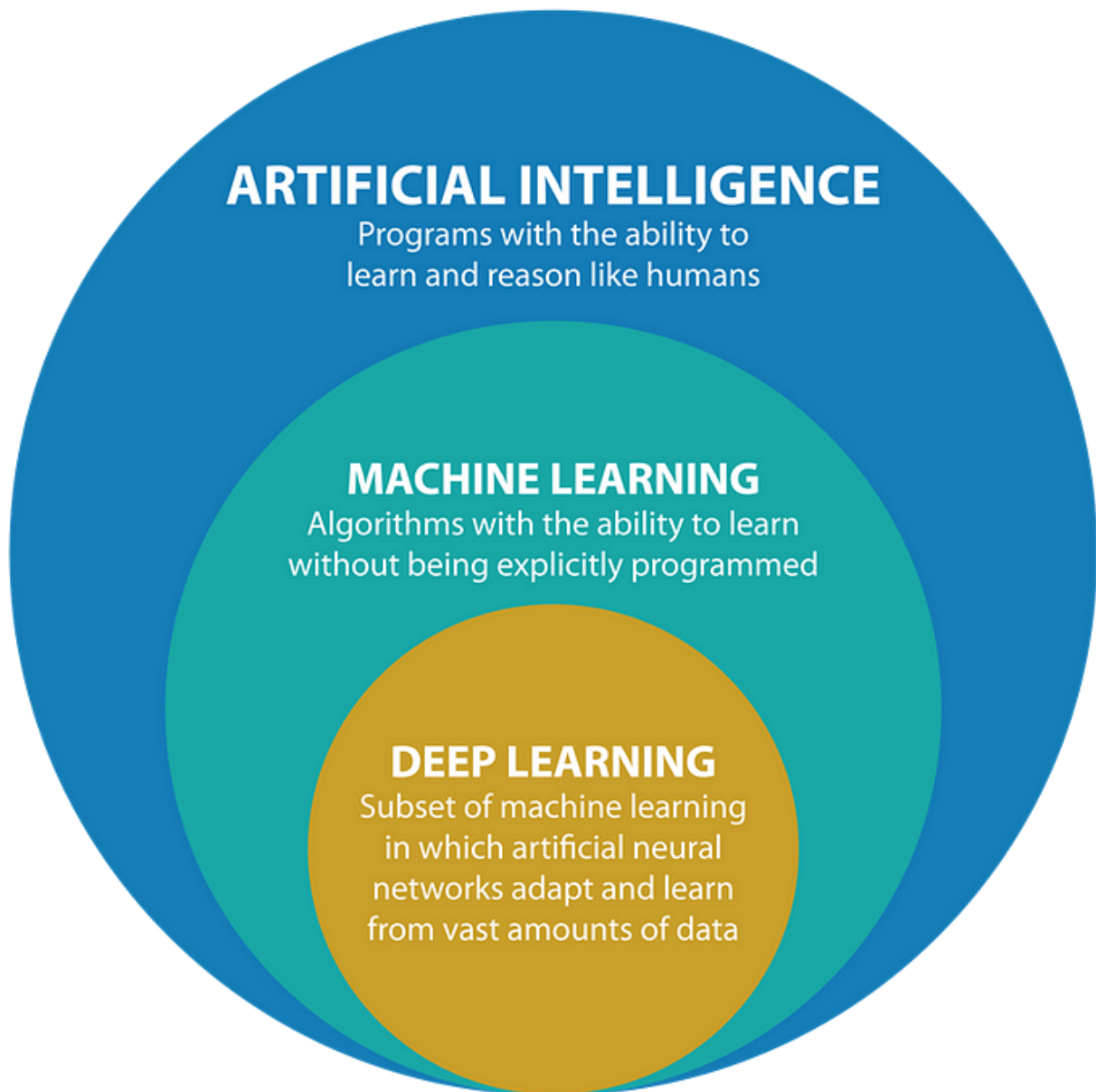
Los algoritmos de aprendizaje automático generalmente se crean utilizando marcos que aceleran el desarrollo de soluciones, como TensorFlow y PyTorch.

## Diferencias entre la Inteligencia Artificial y el Machine Learning

La [inversión en inteligencia artificial y \*machine learning\*](#) crecerá de los 12.000 millones de dólares de 2017 hasta **57.600 millones en 2021**. Aunque estas tecnologías emergentes van de la mano, suelen confundirse con regularidad y utilizarse como sinónimos. Aquí te explicaremos por qué no deben utilizarse indistintamente.

### AI ≠ ML: destacando diferencias y similitudes

La inteligencia artificial es un área de la computación fundada oficialmente como una disciplina en 1956, en la Universidad de Dartmouth, en Nuevo Hampshire. A medida que avanza la tecnología y nacen múltiples derivados, los especialistas han hecho un esfuerzo en delimitarla o diferenciarla de sus vertientes. La IA como tecnología — o conjunto de tecnologías y herramientas — tiene por objeto la creación de máquinas inteligentes que funcionan y reaccionan como seres humanos. El reconocimiento de voz y los asistentes virtuales son ejemplos recientes de inteligencia artificial aplicada. McKinsey define a la **inteligencia artificial** como “**la capacidad de una máquina para realizar funciones cognitivas que asociamos a la mente humana**, como percibir, razonar, aprender, interactuar con el entorno y resolver problemas o incluso utilizar la creatividad”.



1- tienen en común:

aprendizaje

análisis de datos

Difieren en:

la inteligencia Artificial

## machine learning vs data mining

*La ML es aprendizaje automatico y data mining es buscar patrones de comportamiento y conocimiento de datos.*

## Definiciones

AI: habilitar a las maquinas a pensar como humanos

ML: entrenar a las maquinas como aprender.

DL: (deep Learning) Subconjunto de ML para casos especificos.

## Plataformas y herramientas

- R - python con SciKitLearn
- RapidMiner
- Microsoft Azure ML Studio
- Knime
- Weka
- Keras, Tensor Flow

Estas herramientas se actualizan constantemente, cambian y se crean nuevas.

## Plataformas y herramientas

<https://www.simplilearn.com/best-machine-learning-tools-article>

1. **Microsoft Azure Machine Learning, permite construir, entrenar e implementar modelos de AI**

2. **IBM Watson**, utiliza ML y DL para produccion de modelos
3. **Google TensorFlow**, libreria de codigo abierto para programacion de flujos de datos
4. **Amazon Machine Learning**, variedad enorme de herramientas y
5. **OpenNN**

<https://www.javatpoint.com/machine-learning-tools> (fofo busqueda)

## Proceso de DataScience

posible proceso sistematico, en 5 pasos:

1. definir el problema
2. preparar los datos, proceso tedioso que incluye integrar datos que vienen de formas diversas o incluso sin formatos.
3. elegir los algoritmos más apropiados, la estadística es parte esencial del modelo. En base a su comportamiento estadísticos. poder definir la herramienta utilizada en cuanto al tipo de problema que se nos presenta.
4. mejorar los resultados, tener la capacidad de poder tener la mejor version de herramientas y saber que datos arrojan y como funcionan. Tecnicas de mejorar la performance del resultado.
5. presentar los resultados, poder presentar los resultados de la mejor manera posible, haciendolos entendibles.

## Proceso...

1. Conocimiento previo
2. preparacion
3. modelado
4. aplicacion

## 5. conocimiento

Entender y comprender el problema planteado y los datos obtenidos para trabajar (saber si son basura o no).

Preparacion de los datos. Escalas, tipos, etc...

Datos de entrenamiento → construccion de modelos (algoritmos)

Datos de test → Aplicacion de modelo y evaluacion de la performance

Despliegue aplicado a una aplicación de un cliente.

Conocimiento y acciones (deteccion de fraudes, entre otros)

## Proceso CRISP-DM

- Comprension del negocio, conocer de que estamos hablando ( no existe ML aplicada a todo) necesitamos conocimiento de lo que estamos haciendo.
- Comprension de datos (estudios de datos, extremos, casos)
- Preparación de datos, contar con un modelo de datos que permita realizar las operaciones necesarias.
- Modelado, construcción de un algoritmo que resuelva el problema.
- Evaluación, etapa de testeo y performance.
- Despliegue, aplicar la solución en alguna aplicación o plataforma.

## Tipos de algoritmos y sus aplicaciones

(machine learning algorithms cheat sheet)

Azure ML Algorithms CheatSheet

Regresión || two-class clasificacion || deteccion de anomalias || multiclass clasificacion



*Ver que tipo de problema se me presenta para ver que tipo de algoritmo debo utilizar.*

Ej. Prediccion de estados de tiempo. Procesos en una fabrica, temperaturas, presiones, etc.

Puedo predecir una falla para programar una parada en la planta industrial.

## **Clustering**

permite agrupar datos en funcion de las características de sus atributos.

Ej. tiempo en la cantina, dedicacion en tareas, tiempo de estudio, etc para poder mejorar los servicios que se ofrecen.

## **Aprendizaje supervisado clasificacion**

identificar a que categoria pertenece una muestra.

ej. clasificacion binaria a partir de un modelo lineal

ej clasificacion binaria mediante “kernel trick” (SVM) clasificacion no lineal

ej. clasificacion basada en arbol de decision, recorrido segun al arbol. Algoritmo de caja Blanca (facil de explicar)

## **Aprendizaje supervisado regresion**

Predecir un valor numerico (variable dependiente a partir de variables independientes.

Ej. Regresion mediante polinomios. (Error cuadratico medio)

## **Conjunto de datos**

- Fuentes
- características deseables
- tipos de problemas a resolver

Complemento de excel Data Analysis Pack