

## Ejercicio 1

¿Qué es el machine learning?

- 1- “Es una disciplina científica del ámbito de la inteligencia artificial que crea sistemas que aprenden automáticamente. Aprender quiere decir identificar patrones complejos en una cantidad gigante de datos.” <https://cleverdata.io/que-es-machine-learning-big-data/>
- 2- “Es una disciplina del campo de la inteligencia artificial que, a través de algoritmos, dota a los ordenados de la capacidad de identificar patrones en datos masivos y elaborar predicciones (análisis predictivo)”.  
<https://www.iberdrola.com/innovacion/machine-learning-aprendizaje-automatico>
- 3- “Se refiere a la capacidad de una máquina equipada con inteligencia artificial de administrar, pero, sobre todo, de aprender por si misma a través de procesos sistemáticos y algoritmos.” <https://www.atinternet.com/es/glosario/machine-learning/>

¿Qué tiene en común y en qué se diferencia de “Inteligencia artificial”?

Como bien describe la definición (2), el machine learning es una disciplina de la inteligencia artificial. La inteligencia artificial en si misma significa la capacidad de una computadora o de un objeto inanimado de poder realizar procedimientos de forma autónoma al igual que podría hacerlo un ser humano pero usualmente con mayor efectividad y velocidad. El machine learning en cambio, hace referencia, no solo a la capacidad de realizar procesos de las computadoras, sino que además, ajustar la forma de realizar esos procedimientos en base a resultados de forma de mejorarlos en próximas iteraciones, es decir, “aprendiendo”.

¿Qué tiene en común y en que se diferencia de “Análisis estadístico”?

Ambos tienen en común el análisis de las variables para generar ecuaciones matemáticas o relaciones de tal forma de predecir resultados en base a un conjunto de entrada de datos. La diferencia mayor está en que el análisis estadístico es estático y este realiza estimaciones en lugar de aprendizaje como lo hace el machine learning, por lo que a la larga, el machine learning logra incorporar esos conocimientos de estadística y otros métodos de relacionamiento para mejorar sus resultados a medida que se realizan más iteraciones y se agregan más datos.

¿ Como se diferencia de data mining?

Van de la mano, el data mining se trata más que nada de la extracción y recopilado de los datos, la búsqueda en particular de relaciones entre grandes cantidad de variables para llegar a conclusiones sobre la correlación entre los datos. El machine learning también incluye estos comportamientos en sus algoritmos, pero además logra ver más allá de estos datos y realizar

otros tipos de predicciones. Es resumen, data mining realiza conclusiones sobre los datos que haya recolectado y machine learning, además de hacer esas conclusiones, usa algoritmos para mejorarlas continuamente.

#### ¿ En que se aplica el machine learning?

El machine learning tiene una gran cantidad de usos, entre ellos, se destacan:

- Reconocimiento de objetos o seres vivos en imágenes
- Reconocimiento de fraude en bancos
- Chatbots para que las empresas no tengan que hablar con los clientes directamente.
- Etc.

## Ejercicio 2 – Herramientas

- Rapidminer ( <https://rapidminer.com/>) : Es una herramienta para en análisis y minería de datos (Data mining). Se destaca por su potencia a la hora de presentar subrutinas o “bloques” donde realizar cálculos y predicciones con la data presentada en él. Entre los algoritmos que más se destacan posee:
  - Generado de gráficas 3D para visualización de data
  - Ajustado de variables para analizar una como salida del resto
  - Filtrado automático de outliers
  - Limpieza de datos, Normalización
- Python: El lenguaje de programación más utilizado para machine learning es Python. Este lenguaje es el preferido por toda empresa relacionada a este campo debido a su optimización y gran cantidad de operaciones para este tipo de tareas. Asimismo, existen una gran cantidad de herramientas y librerías creadas por la comunidad y por empresas grandes como Google capaces de realizar la mayoría del trabajo para el análisis de datasets, lo cual hace simple utilizarlo para amateurs en el área.
- Excel: Originalmente creado con la intención de contener grandes volúmenes de datos de forma ordenada en grillas, se ha visto utilizado bastante en los últimos años para tareas pequeñas de machine learning y data mining gracias a las poderosas herramientas y operaciones estadísticas que posee de fábrica. Asimismo, la curva de aprendizaje es bastante sencilla y es una herramienta del Stack de Office, manejada por una gran cantidad de personas.

## Ejercicio 3 – Procesos de “Data science”

### CRISP-DM

Cross Industry Standard Process for Data Mining, es un método probado para orientar trabajos de minería de datos (metodología. Se divide en seis fases principales:

- Comprensión del negocio: Tener conciencia del contexto.
- Comprensión de los datos: Saber que estamos viendo y como interpretarlo
- Preparación de los datos: Normalización, eliminación de outliers, etc.

- Fase de modelado: Modelar los métodos para llegar a la salida deseada
- Evaluación: Evaluar el modelo generado, ver si cumple con nuestras expectativas del negocio y en caso contrario, ajustarlo.
- Implantación: Liberar la salida, una vez que tengamos todos los parámetros de la forma deseada, ya sea con gráficos, descripciones o tablas.

Otros procesos como CRISP-DM son:

- Team Data Science Process (TDSP) de Microsoft, una alternativa más moderna que Crisp, que se basa en la utilización de tecnologías más actuales y con una documentación mucho más rica (proveída por el propio Microsoft).
- IBM-ASUM-DM: Se basa mismo en la tecnología de CRISP pero busca ser más especializada en su utilización con el Stack de herramientas de machine learning de IBM.

## Ejercicio 5

Dataset seleccionado: <https://archive.ics.uci.edu/ml/datasets/Automobile>

Este dataset lista varias características de un auto (cantidad de ruedas, velocidad, tipo de combustible, tipo de motor, peso, altura, etc), en búsqueda de determinar que tan arriesgado es viajar en él (o lo que es lo mismo, que tan seguro es el auto).

Cantidad de parámetros: 26

Parámetros con más correlación:

- Precio
- Marca del auto ( toyota, alfa romeo, etc)
- Peso
- Tracción (Trasera, delantera, 4x4)
- Peso
- Tipo de cuerpo (Sedan, hatchback, wagon).