

# Ejercicio 1

## ¿Qué es “Machine Learning”?

### Definiciones de Machine Learning:

#### IBM

Machine learning es una rama de la Inteligencia Artificial y de la Ciencia de la Computación al cual se centra en el uso de información y de algoritmos para imitar el proceso de aprendizaje del humano, mejorando gradualmente su precisión.

#### Chat Bing

El aprendizaje automático (AA) o aprendizaje de máquinas es el subcampo de las ciencias de la computación y una rama de la inteligencia artificial (IA), cuyo objetivo es desarrollar técnicas que permitan que las computadoras aprendan. El AA se enfoca en enseñar a las computadoras para que aprendan de los datos y mejoren con la experiencia —en lugar de ser explícitamente programadas para hacerlo

#### Iberdrola

El Machine Learning es una disciplina del campo de la Inteligencia Artificial que, a través de algoritmos, dota a los ordenadores de la capacidad de identificar patrones en datos masivos y elaborar predicciones (análisis predictivo). Este aprendizaje permite a los computadores realizar tareas específicas de forma autónoma, es decir, sin necesidad de ser programados.

### Pregunta 1

La diferencia entre machine learning e inteligencia artificial es que la inteligencia artificial es el concepto más amplio que se refiere a la capacidad de las máquinas de mostrar un comportamiento inteligente, mientras que el machine learning es una técnica o aplicación de la inteligencia artificial que se basa en el aprendizaje automático de los datos y la generación de predicciones, mejorando así las habilidades de las máquinas

### Pregunta 2

Los modelos de machine learning o aprendizaje automático son diseñados para ofrecer predicciones lo más certeras posible. Los modelos de estadística están creados para inferencia sobre relaciones entre variables. Las estadísticas se basan únicamente en espacios de probabilidad.

### Pregunta 3

El machine learning y el data mining son dos técnicas diferentes que se utilizan para analizar datos. Machine learning es una técnica que se utiliza para enseñar a las computadoras a aprender de los datos y mejorar automáticamente sin ser programadas explícitamente. Por otro lado, data mining es un proceso de descubrimiento de patrones en grandes conjuntos de datos.

#### **Pregunta 4**

Machine learning se utiliza en una amplia variedad de aplicaciones, incluyendo motores de búsqueda, diagnósticos médicos, detección de fraude en el uso de tarjetas de crédito, análisis de mercado para los diferentes sectores de actividad, clasificación de secuencias de ADN, reconocimiento del habla y del lenguaje escrito, juegos y robótica.

## **Ejercicio 2**

#### **Herramientas y plataformas de Machine Learning:**

Rapid Miner <https://rapidminer.com/>

- Es una plataforma de análisis que permite acelerar la creación, entrega y mantenimiento de analíticas predictivas de alto valor.

Weka

- Es un conjunto de Librerías para tareas de minería de datos. Las librerías pueden ser llamadas desde la interficie de weka o desde programación.

IBM Watson - <https://www.ibm.com/watson>

- Esta herramienta permite analizar e interpretar todo tipo de datos, tanto estructurados (bases de datos) como no estructurados (textos, e-mails, imágenes, audio, vídeo, tweets, mensajes) ya que es capaz de leer y entender el lenguaje natural.

H2O.ai: <https://h2o.ai/>

- H2O.ai es una plataforma de código abierto que ofrece una variedad de herramientas para el aprendizaje automático y análisis de datos a gran escala. Es conocida por su eficiencia y rendimiento en la gestión de grandes volúmenes de datos. H2O.ai permite a los usuarios construir modelos de aprendizaje automático de manera rápida y sencilla, lo que la convierte en una opción popular para aplicaciones empresariales y científicas.

PyTorch:

- (<https://www.iebschool.com/blog/herramientas-business-intelligence-2021-big-data/>)

- PyTorch es otra biblioteca de aprendizaje profundo de código abierto que ha ganado popularidad debido a su enfoque en la simplicidad y la flexibilidad. Es ampliamente utilizado por la comunidad de investigación debido a su facilidad para crear modelos complejos y experimentar con nuevas arquitecturas de redes neuronales. Con su modo de ejecución definido por el usuario (Eager mode), PyTorch permite a los desarrolladores depurar y ejecutar código de manera interactiva, lo que acelera el proceso de desarrollo y mejora la eficiencia.

Keras:

- (<https://www.iebschool.com/blog/herramientas-business-intelligence-2021-big-data/>)
- Keras es una biblioteca de aprendizaje profundo de alto nivel escrita en Python. Es conocida por su facilidad de uso y su enfoque en la experimentación rápida. Keras permite a los desarrolladores crear prototipos de modelos rápidamente y, a su vez, se integra perfectamente con TensorFlow y Microsoft Cognitive Toolkit (CNTK). Es una excelente opción para aquellos que desean comenzar rápidamente con el aprendizaje profundo sin sumergirse en detalles complejos.

Apache Spark MLlib:

- (<https://www.iebschool.com/blog/herramientas-business-intelligence-2021-big-data/>)
- Apache Spark MLlib es una biblioteca de aprendizaje automático escalable que se integra con el popular framework de procesamiento de datos Apache Spark. MLlib proporciona una amplia gama de algoritmos y herramientas para el procesamiento y análisis de grandes conjuntos de datos. Esta herramienta es ideal para entornos distribuidos y aplicaciones que requieren un alto rendimiento y una escalabilidad significativa.

## **Librerías en Python:**

Pandas

- Es una librería de python usada para trabajar con data sets. Tiene funcionalidad para analizar, limpiar, explorar y manipular datos.

Numpy

- Es una librería usada para trabajar con arrays, matrices y álgebra lineal.

Plotly

- Librería gráfica de python.

Scikit-learn

- Librería que contiene herramientas para el análisis de datos.

### Ejercicio 3

#### CRISP-DM

CRISP-DM (Cross-Industry Standard Process For Data Mining) es un metodo probado para orientar sus trabajo de mineria de datos. Cubre las fases de un proyecto, sus tareas respectivas y las relaciones entre estas tareas



Las fases son:

1. Comprensión del negocio
2. Comprensión de los datos
3. Preparación de los datos
4. Modelado
5. Evaluación
6. Despliegue

#### KDD

Es una metodología que propone 5 fases diferentes a CRISP-DM, las cuales son: selección, preprocesamiento, transformación, minería de datos y evaluación e implementación. Es un proceso iterativo e interactivo.

#### SEMMA

SEMMA propone las siguientes 5 fases: Sample (entrada de datos), Explore (exploración distribuida), Modify (transformación de variables), Model (Regresión, árboles, redes neuronales),

Assess (evaluación, medidas, reporte). La metodología es propuesta por SAS y la define como “proceso de selección, exploración, y modelamiento de grandes cantidades de datos para descubrir patrones de negocio desconocidos”.

## **Ejercicio 4**

## **Ejercicio 5**

### **Adult**

Determina si una persona gana más de 50 mil dólares al año, a partir de atributos como la edad, estado civil, educación, ocupación, etc.

Problema => Clasificación

Atributos:

- Age => edad.
- Workclass => tipo de trabajo.
- Education => tipo de educación.
- Marital-status: estado civil.
- Occupation: ocupación.
- Relationship: tipo de relación
- Sex: sexo
- capital-gain: capital ganado
- capital-loss: capital perdido
- hours-per-week: horas por semana
- native-country: País nativo
- Algoritmos =>
  - Xgboost Classification
  - Support vector Classification
  - Random Forest Classification
  - Neural Network Classification
  - Logistic Regression