# Curso Actualízate – Machine Learning Gijón (Módulo 5)

Nombre: Héctor

Apellidos: Espina Antuña

Fecha: 07/07/2023

Responde a las siguientes preguntas. Justifica la respuesta.

1. ¿Qué es un ETL?

Es un proceso para el tratamiento de los datos que consiste en extraer los datos que nos interesan, transformarlos y hacer modificaciones de errores y unificación de criterios y por último volver a cargarlos en una base de datos

OK

1. Enumera las características de Python explicando cada una de ellas (menciona 4 al menos).

-Código abierto: su código es accesible y puede ser modificado

- Lenguaje de alto nivel con una sintaxis clara y sencilla que facilita el aprendizaje

- Gran cantidad de librerías de terceros disponibles

-Comunidad activa que garantiza un soporte continuo, abundantes recursos de aprendizaje y contribuciones constantes

OK

1. ¿Cuáles son los tipos primitivos en Python y qué valores pueden contener cada uno de ellos? Pista: Son 3 tipos.

- Tipo booleano: puede contener los valores True y False, el 1 hace referencia a True y el 0 a False

-Tipo string: Son cadenas de texto o caracteres que van siempre entre comillas

-Tipo numérico: son números que pueden ser de tipo entero o flotante(decimal)

OK

1. Menciona alguna estructura de datos más compleja que los tipos primitivos que conozcas.

-Tipo lista: son listas de elementos que pueden ser de distinto tipo cada uno de ellos  
-También estarían las tuplas, sets o diccionarios

OK

1. Escribe la sintaxis para crear variables.

nombre\_variable = valor\_variable

OK

1. Escribe la sintaxis para crear funciones.

def fun(argumento):

condiciones de la función

OK

1. Escribe la sintaxis para llamar a variables.

nombre\_variable

OK

1. Escribe la sintaxis para llamar a funciones.

fun(argumento)

OK

1. Explica con tus palabras para qué sirven las librerías: Pandas y Numpy.

La librería Pandas nos permite transformar datos de manera sencilla  
La librería Numpy es un librería para cuestiones matemáticas que también nos permite trabajar con vectores y matrices

OK

1. ¿Cómo representamos el valor vacío en Python?

None

OK

1. ¿Qué es un IDE?

Entorno de desarrollo integrado. Aplicación de software para programar

OK

1. ¿Qué es el CRISP-DM?

Es un modelo de minería de datos que consta de 6 etapas: Comprensión del negocio, Comprensión de los datos, Preparación de los datos, Modelado, Evaluación y Despliegue

OK

1. Explica cada uno de los tipos de Machine Learning.

**Supervisado**: se le dan datos de entrada y datos de salida  
**No supervisado**: se le dan datos de entrada pero no datos de salida  
**Reforzado**: no se le dan datos de entrada ni de salida, solo las instrucciones y se le va diciendo si lo hace bien o mal

**Semi-Supervisado**: Empieza con el no supervisado y clasifica en clusters, luego un humano le dice si está bien o mal para reagrupar categorías.

***Aprendizaje supervisado:*** ***Conjunto de datos etiquetados***, es decir, que ***se tiene información de salida para cada ejemplo de entrada***. El objetivo es entrenar un modelo que pueda predecir correctamente las etiquetas para nuevos datos de entrada. Algunos ***ejemplos*** de aplicaciones de aprendizaje supervisado incluyen la ***clasificación de imágenes o el reconocimiento de voz***.

***Aprendizaje no supervisado:*** En este tipo de ML no se cuenta con datos etiquetados, el objetivo es ***encontrar patrones y relaciones en los datos sin información previa sobre las etiquetas.*** Algunos ejemplos de aplicaciones de aprendizaje no supervisado incluyen ***la segmentación de clientes, la agrupación de datos y la reducción de la dimensionalidad.***

***Aprendizaje reforzado:*** En este tipo de ML no se necesita datos, analiza comportamientos. Algunos ***ejemplos*** de aplicaciones de aprendizaje por refuerzo ***incluyen la robótica, los videojuegos y la optimización de sistemas de control***.

***Aprendizaje semi-supervisado:*** *Es un enfoque híbrido que combina elementos del aprendizaje supervisado y no supervisado. En este tipo de aprendizaje, el modelo se entrena con un conjunto de datos que contiene tanto ejemplos etiquetados como no etiquetados.*

*La idea es utilizar la información disponible en los datos no etiquetados para mejorar el rendimiento del modelo. Algunos* ***ejemplos serían el reconocimiento facial o la clasificación de documentos.***

1. ¿Con qué tipo de Machine Learning hemos estado trabajando nosotros?

Supervisado por batch learning.

1. ¿Qué es el Prophet? ¿Y Scikit Learn, Keras y Tensorflow?

Es un modelo predictivo. Prophet se utiliza para regresiones lineales  
El resto son librerías de python sobre Machine Learning

Prophet es una ***biblioteca de Python****,* que se utiliza para ***realizar análisis de series temporales*** y hacer pronósticos en el futuro.

Scikit-Learn es una ***biblioteca de aprendizaje que*** ***proporciona una amplia variedad de algoritmos de aprendizaje automático supervisado y no supervisado***, herramientas de preprocesamiento de datos y utilidades para evaluar modelos de aprendizaje automático.

Keras es una ***biblioteca de alto nivel*** ***que proporciona una interfaz para construir y entrenar modelos de aprendizaje profundo***. Keras se centra en la facilidad de uso, la modularidad y la extensibilidad, lo que la hace muy popular entre la comunidad de aprendizaje profundo. ***Tanto Prophet, Scikit-Learn como Keras son de código abierto***.

1. ¿Qué significa en ML regresión? Responde brevemente.

Es un tipo de tarea en el que el objetivo es predecir valores numéricos de interés sobre unos datos de entrada

OK

1. ¿Cuál es el tipo de predicción más sencilla, pero a la vez la más usada?

La regresión lineal simple

1. ¿Cuáles son los problemas principales del Machine Learning? Justifica tu respuesta.

Mala calidad de los datos. Si los datos son de mala calidad el modelo no va a poder predecir con exactitud los resultados y el error será más grande.  
  
Sesgos en los datos: Los datos que introducimos en el modelo para la predicción pueden estar sesgados por comportamientos humanos.

OK

1. ¿Qué es el residuo en Machine Learning?

Se refiere a la diferencia entre el valor que predice un modelo y el valor real.

OK