# Curso Actualízate – Machine Learning Gijón (Módulo 5)

Nombre: Hugo

Apellidos: Arbesú Valles

Fecha: 07/07/2023

Responde a las siguientes preguntas. Justifica la respuesta.

1. ¿Qué es un ETL?

Es un proceso de transformación de datos para adecuarlos a un mejor uso según las necesidades propias o del propio proceso para el que se usarán, las siglas corresponden a Extract(captura de datos), Transform(transformación de los datos, ejemplo eliminación de nulos, etc) y Load(carga de los datos transformados)

OK

1. Enumera las características de Python explicando cada una de ellas (menciona 4 al menos).
2. Lenguaje de alto nivel: su sintaxis está cerca del lenguaje humano
3. Bibliotecas disponibles: existen muchas bibliotecas de las que podemos usar las funciones que nos interesen para problemas ya resueltos
4. Código abierto: el código es accesible por dentro para los usuarios
5. Interpretado: la ejecución del código se hace línea a línea sin crear previamente una traducción completa previa a lenguaje máquina

OK

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

Tipo número: ejemplo int, float

Representan valores numéricos con los que se puede operar matemáticamente, int representa los enteros y float los reales

Tipo texto: string

Representan valores que son texto, para diferenciarlos de texto que represente un nombre de variable van entre comillas

Tipo lógico: bool

Representan la posibilidad de dos estados: True y False

OK

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

Lista: tipo de estructura que se compone de una colección de varios valores ordenados, dichos valores pueden ser de distintos tipos entre sí, incluso otra lista puede formar parte de una lista

OK

1. Escribe la sintaxis para crear variables.

nombre\_variable = valor

OK

1. Escribe la sintaxis para crear funciones.

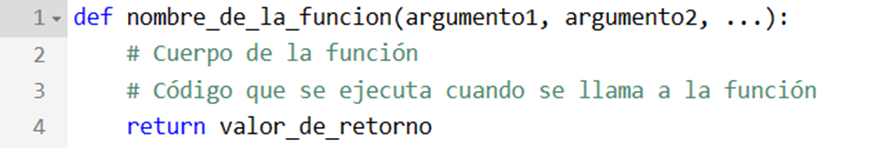
def nombre\_funcion(variable1, variable2,...):

#código usando variable1, variable2,...

#resultado/s salida(print, return,...)

#código,…

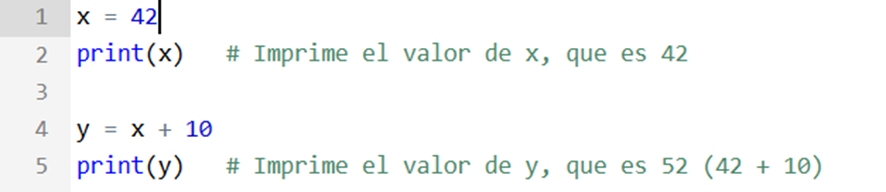
OK



1. Escribe la sintaxis para llamar a variables.

variable\_creada=...(variable\_llamada)...

OK



1. Escribe la sintaxis para llamar a funciones.

variable/s\_salida=nombre\_funcion(variable/s\_entrada)

OK



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

Pandas: sirve para la manipulación de datos con objetivo de hacer selecciones sobre datos mayores(como dataframes), añadir nuevos datos, modificar los existentes, etc

Numpy: está más enfocada en operaciones numéricas para hacer cálculos sobre datos en formato número

OK

1. ¿Cómo representamos el valor vacío en Python?

variable\_numero=None indica que queremos no tener valor ni tipo

variable\_cadena=’”” indica que queremos una cadena vacía

variable\_lista=[] indica que queremos una lista vacía

OK

El valor vacío se representa mediante la palabra clave None. None se utiliza para indicar que una variable o expresión no tiene un valor válido o que una función no devuelve ningún valor. Ej, ***x = None.***

1. ¿Qué es un IDE?

Es un entorno, un programa, de desarrollo de código, proporciona las herramientas para crear tus propios programas

OK

1. ¿Qué es el CRISP-DM?

Es el proceso por etapas de análisis de datos, desde la captación de los datos, su depuración, aplicación del modelo de machine learning, etc

OK

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

-Machine learning clásico: basado en predicciones sobre datos numéricos

-Deep learning: tipo de machine learning que opera con datos más complejos como imágenes, texto, etc, y que usa modelos de predicción basado en algoritmos de redes neuronales

-Tipo supervisado: el algoritmo sabe los tipos de salida en los que debe clasificar los datos

-Tipo no supervisado: el algoritmo crea sus propios tipos de clasificación en función de patrones que detecte

-Clustering: tipo de no supervisado, consiste en segmentar los datos

-Semisupervisado: el algoritmo crea al principio sus propias categorias de clasificación y luego le decimos que reasignación debe hacer

-Por refuerzo: el algoritmo no usa datos de entrada, se le dan una reglas y a partir de ellas crea sus propios datos basándose en qué es éxito y en qué es fracaso

OK

***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?

Con machine learning tradicional

Predicciones numéricas con regresiones lineales

Predicciones de categoría de éxito o fracaso en función de datos numéricos

Y supervisado por batch learning.

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

Prophet: biblioteca de Python para hacer predicciones sobre datos numéricos según el patrón de los datos antiguos

Scikit Learn: biblioteca de Python para hacer predicciones de clasificación

Keras: biblioteca de deep learning

Tensorflow: biblioteca de deep learning que permite operar con más detalle en en los algortimos

OK

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

Es una técnica matemática para estimar el valor de una variable en función de otra. Dados unos valores de relación entre la variable a predecir y la otra variable podemos “crear” una ley de comportamiento para extrapolar esa ley y estimar valores futuros

OK

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

La regresión lineal simple es la más sencilla, se basa en que la relación de la variable a predecir y la otra variable es una línea recta

OK

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

Disponer de pocos datos: si tenemos pocos datos de entrada no tendremos un comportamiento entre las variables realmente fiable

Capacidad de cómputo: el entrenamiento de modelos muy complejos que sean muy ambiciosos, como las redes neuronales que pretendan imitar la mente humana, puede requerir gran capacidad de computación, una posible solución puede estar implementándose actualmente en base a la computación cuántica

OK

***Datos de mala calidad:*** Si los datos de entrada contienen errores o inconsistencias, puede ser difícil para un modelo de aprendizaje automático aprender patrones significativos.

***Variables irrelevantes:*** Si los datos de entrada contienen características irrelevantes o redundantes, puede ser difícil para un modelo de aprendizaje automático aprender patrones significativos y puede llevar a una disminución en el rendimiento del modelo.

***Underfitting:*** Si el modelo tiene una capacidad limitada o hay muy pocos datos de entrenamiento, puede ser difícil para el modelo aprender patrones significativos y hacer predicciones precisas.

***Overfitting:*** Si el modelo tiene demasiada capacidad o se entrena con demasiados datos, puede ser difícil para el modelo generalizar y hacer predicciones precisas en nuevos datos.

1. ¿Qué es el residuo en Machine Learning?

Es la diferencia entre el valor predicho por el modelo y el valor real observado

OK