Curso Actualízate – Machine Learning Sevilla (Módulo 5)

Nombre: Juan Miguel

Apellidos: Caballero Mesa

Fecha: 17 de febrero de 2023

Responde a las siguientes preguntas. Justifica la respuesta.

1. ¿Qué es un ETL?

Extract-Transform-Load.

Es un proceso que permite extraer y mover datos de múltiples fuentes, reformatearlos y/o limpiarlos, y cargarlos en otra base de datos para analizarlos.

OK

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

Lenguaje de alto nivel y fácil de aprender.

Multiplataforma, válido para todos los sistemas operativos.

Interpretado, el código fuente no se compila a un archivo ejecutable antes de su ejecución. En su lugar, el intérprete de Python ejecuta el código línea por línea en tiempo real.

Multiparadigma, soporta varios enfoques de programación, incluyendo programación orientada a objetos, programación funcional y programación estructurada.

OK

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

Numéricos, números enteros, decimales… 3 4.75 -10.0

Booleanos, Verdadero o Falso.

String, cadena de texto entre comillas. “HoLa 45”

OK

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

Diccionarios: estructura de datos que se utiliza para almacenar una colección de elementos clave-valor. {"nombre": "Juan", "edad": 30, "ciudad": "Madrid"}

OK

1. Escribe la sintaxis para crear variables.

nombre\_de\_variable = valor\_de\_variable

OK

1. Escribe la sintaxis para crear funciones.

def nombre\_de\_funcion(argumentos):

# Cuerpo de la función

return resultado

OK

1. Escribe la sintaxis para llamar a variables.

print(nombre\_de\_variable)

OK

1. Escribe la sintaxis para llamar a funciones.

nombre\_de\_funcion(argumentos)

OK

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

Pandas: librería que se utiliza para el análisis y manipulación de datos estructurados. Esta librería es muy útil para la limpieza y preparación de datos antes de realizar el análisis, y también para el análisis exploratorio de datos y la visualización.

Numpy: librería que se utiliza principalmente para trabajar con arrays numéricos y matrices. Ofrece una amplia variedad de funciones para realizar operaciones matemáticas y estadísticas en estos tipos de datos.

OK

1. ¿Cómo representamos el valor vacío en Python?

None

OK

1. ¿Qué es un IDE?

Herramienta de software que se utiliza para desarrollar programas de ordenador de manera más eficiente y productiva.

OK

1. ¿Qué es el CRISP-DM?

CRISP-DM estándar de procesos para proyectos dedicados a minería de datos

OK

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

Supervisado: se proporciona al modelo un conjunto de ejemplos de entrada y salida, y se entrena el modelo para que aprenda a predecir la salida correcta a partir de la entrada.

No supervisado: trabaja con datos no etiquetados, lo que significa que no hay una salida conocida para una entrada dada. En cambio, el modelo busca patrones y estructuras en los datos sin una guía previa.

Reforzado: No necesita datos, analiza comportamientos. Si el modelo falla, es corregido por el humano y aprende.

Semisupervisado: comienza con un aprendizaje no supervisado, haciendo clusters. Luego se combina con un aprendizaje supervisado para decirle que output ponerle a cada cluster. Ejemplo: reconocimiento facial.

OK

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

Aprendizaje supervisado por batch learning

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

Prophet se utiliza para pronósticos de series temporales.

Scikit Learn para aprendizaje automático supervisado y no supervisado.

Keras para aprendizaje profundo de alto nivel, y TensorFlow para construir modelos de aprendizaje automático en general.

OK

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

Es un tipo de tarea de aprendizaje supervisado en la que se busca predecir un valor numérico continuo a partir de una serie de variables de entrada.

OK básicamente regresión es predecir números.

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

Regresión lineal simple, se busca modelar la relación lineal entre una variable de entrada y una variable de salida numérica.

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

Datos de mala calidad

Variables irrelevantes

Underfitting: Tenemos pocos datos y, o el modelo no aprende con tan pocos datos o aprende mal.

Overfitting: Entrenar al modelo con todos los datos y quedarse sin datos para testear.

OK

1. ¿Qué es el residuo en Machine Learning?

Se refiere a la diferencia entre el valor real de la variable de destino y la predicción hecha por el modelo.

OK