# Curso Hackea Tu Futuro – Machine Learning Madrid Nelso (Módulo 5)

😀 😃

Nombre: Cristina

Apellidos: Sanchez Muñoz

Fecha: 12/05/2023

Responde a las siguientes preguntas. Justifica las respuestas.

1. ¿Qué es un ETL?

Extraer, Transformar y Cargar. Proceso para mover y transformar datos desde diferentes fuentes de datos a un lugar centralizado. La información se extrae, se modifica y se carga en una base de datos para poder analizarla y tomar decisiones basadas en ella.

OK

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

Fácil de aprender: es una buena opción para principiantes y profesionales que quieren aprender un nuevo lenguaje de programación.

Bibliotecas y frameworks: Python tiene una gran cantidad de bibliotecas y frameworks disponibles, lo que hace que la programación sea más fácil y rápida.

Tipado débil: significa que una misma variable puede cambiar de tipo de datos en diferentes momentos de la ejecución del programa, lo que le da a Python una gran flexibilidad en el manejo de datos.

Multiplataforma: se puede ejecutar en diferentes sistemas operativos sin necesidad de hacer cambios en el código.

OK

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

* Numéricos: 8, 8.5 …
* Texto: ‘texto’
* Booleanos: True, False

OK

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

Valores nulos: (None) Ausencia de valor.

OK

1. Escribe la sintaxis para crear variables.

Nombre\_variable = valor

OK

1. Escribe la sintaxis para crear funciones.

def nombre\_funcion(parametro1, parametro2, ...):

instruccion1

instruccion2

return resultado

OK

1. Escribe la sintaxis para llamar a variables.

Simplemente se utiliza su nombre en el lugar donde se desea utilizar su valor.

Ej: nombre\_variable

OK

1. Escribe la sintaxis para llamar a funciones.

Nombre de la función seguido de paréntesis () Si la función recibe argumentos, se incluyen dentro de los paréntesis, separados por comas.

nombre\_funcion(argumento1, argumento2, …)

OK

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

* Pandas te ayuda a trabajar con datos en hojas de cálculo y otros formatos. También permite la limpieza de datos, manejo de valores faltantes y eliminar datos duplicados.
* NumPy te ayuda a trabajar con arreglos y matrices numéricas.

OK

1. ¿Cómo representamos el valor vacío en Python?

None

OK

1. ¿Qué es un IDE?

Entorno de desarrollo integrado. Es un software que proporciona un conjunto de herramientas integradas para facilitar el desarrollo de software. (Simplificar el proceso de desarrollo y mejorar la productividad de los programadores).

OK

1. ¿Qué es el CRISP-DM?

Proporciona una estructura sistemática para el análisis de datos y se compone de seis fases: comprensión del problema, comprensión de los datos, preparación de los datos, modelado, evaluación y ejecución.

OK

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

* Aprendizaje supervisado: para predecir resultados futuros en función de datos de entrada conocidos. Donde cada muestra está compuesta por una entrada y su correspondiente resultado deseado. Una vez entrenado, el modelo puede predecir el resultado para nuevas entradas desconocidas.
* Aprendizaje no supervisado: para la segmentación de datos o para reducir la dimensión de los datos. El modelo se entrena con datos sin etiquetar.
* Aprendizaje por refuerzo: un agente de inteligencia artificial interactúa con un entorno para aprender a tomar decisiones óptimas. El agente recibe recompensas o castigos en función de las acciones que realiza y utiliza estas recompensas para mejorar su comportamiento en el futuro.

OK, te falta el semisupervisado, el batch learning y el online learning

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?

Son librerías

* Prophet es una librería de código abierto desarrollada por Facebook para la predicción de series de tiempo.
* Scikit Learn es una librería de aprendizaje automático de código abierto para Python. Es una herramienta útil y popular para aplicar técnicas de aprendizaje automático a conjuntos de datos y resolver problemas de análisis de datos.
* Keras es una librería de código abierto para aprendizaje profundo en Python.
* TensorFlow es una biblioteca de código abierto para el aprendizaje automático y la inteligencia artificial desarrollada por Google.

OK

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

Es la tarea para predecir un valor numérico en función de uno o varios valores de entrada.

OK, básicamente es predecir números

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.

Selección inadecuada del modelo de aprendizaje automático para un problema específico, la falta de datos de calidad y cantidad suficiente, la falta de interpretación y explicación de los resultados obtenidos y la posibilidad de sobreajuste (overfitting) del modelo, lo que puede llevar a una mala generalización de los datos.

-Variables irrelevantes-Overfitting y Underfitting-

Datos insuficientes-Datos de mala calidad-

OK

1. ¿Qué es el residuo en Machine Learning?

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

OK