# Curso Actualízate – Machine Learning Gijón (Módulo 5)

Nombre: Daniel

Apellidos: Piñeros

Fecha: 07/07/2023

Responde a las siguientes preguntas. Justifica la respuesta.

1. ¿Qué es un ETL?

es un proceso(extraer,transformar,cargar por sus siglas en inglés) en el que

se busca optimizar la información que se va a obtener de distintas fuentes con el fin de analizarla

OK

ETL es un proceso, que proviene del anglicismo y acrónimo ***“extract, transform and load (extracción, transformación y carga)”***, utilizado en el campo de la informática con la idea de integrar y consolidar datos para un único sistema de destino.

En este proceso, se extraen datos de una o varias fuentes, las cuales pueden estar en diferentes formatos y almacenados en distintos tipos de repositorios. El objetivo principal del proceso es asegurar que ***los datos sean precisos, completos y coherentes entre sí antes de ser utilizados*** y tratados en análisis o replicados para el uso de aplicaciones.

***Una buena transformación de los datos***durante el proceso ETL ***debe incluir la conversión de los formatos, la eliminación de duplicados, la normalización, la aplicación de reglas y/o la agregación de nuevas variables.***

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

Python es un lenguaje de código abierto, con lo que cualquiera con conocimientos previos puede entender la base y modificarla según sus necesidades,también es un lenguaje multiplataforma, por lo que se puede implementar en distintos sistemas (hardware,software) sin que se tenga que modificar tanto la base del mismo, además de ser un lenguaje de alto nivel(lenguaje parecido al del ser humano), por lo que muchas de sus funciones son “intuitivas”. y es gratis, cualquiera puede tener acceso a él.

OK

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

-numéricos

-texto

-booleanos

En Python, existen tres tipos primitivos de datos:

● ***Numéricos,*** este tipo de dato representa números enteros sin decimales. Pueden ser positivos o negativos y no tienen límites de tamaño en Python***.*** Dentro de ellos contamos con,

○ ***enteros (int)sin decimales (1, 10, -15).*** Este tipo de dato representa números enteros sin decimales. Pueden ser positivos o negativos y no tienen límites de tamaño en python,

○ ***y flotantes (float) con decimales, (3.14, 2.718, -0.5).*** Este tipo de dato representa números con decimales. Los flotantes en python se escriben con un punto decimal y pueden ser positivos o negativos.

● ***Booleanos (bool),*** se representan con los valores, ***verdadero (True) y falso (False).*** Los booleanos son útiles en la programación para tomar decisiones y controlar el flujo del programa.

● ***String (str),*** este tipo de dato representa cadenas de texto. Las cadenas de texto en Python se escriben entre comillas simples o dobles, y pueden contener letras, números, signos de puntuación y espacios. Ejemplos de cadenas de texto son ***"hola", "Python me mola", "123"***.

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

Una de las estructuras de datos más compleja en Python es el ***diccionario (dictionary), que se utiliza para almacenar una colección de pares clave-valor***. Un diccionario en Python es una estructura de datos no ordenada y mutable que se compone de una serie de elementos, cada uno de los cuales tiene una clave única y un valor asociado.

1. Escribe la sintaxis para crear variables.

nombre\_variable = datos u otra variable? =D 🙂

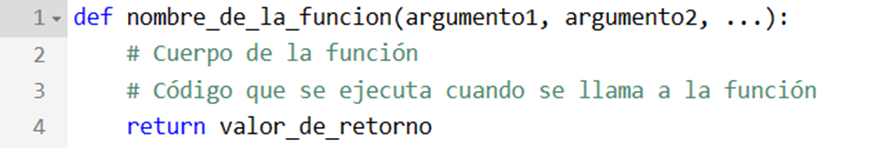


1. Escribe la sintaxis para crear funciones.

def n\_funcion(variable-columna):

condiciones para la funciones(if)

(return)



1. Escribe la sintaxis para llamar a variables.

print(variable)

OK

1. Escribe la sintaxis para llamar a funciones.

n\_funcion(argumento)

OK

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

panda: librería que facilita la manipulación de datos que se obtienen a través de tablas u hojas de cálculo, que además permite la exploración y análisis de forma mas “facil”

numpy: es la librería que contiene funciones matemáticas y de cálculo avanzado para extraer información y trabajar en distintas “dimensiones” en distintos tipos de datos.

OK

1. ¿Cómo representamos el valor vacío en Python?

~~none.nan~~

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?

interfaz en la que utilizar, visualizar y probar el código y las funciones aplicadas

Un IDE en Python ***es un programa que proporciona herramientas para facilitar el desarrollo de software***. En particular, el IDE de Python ofrece características como la edición de código, la depuración, la ejecución y la gestión de proyectos, todo en una sola interfaz gráfica de usuario.

1. ¿Qué es el CRISP-DM?

es un modelo guia de minería de datos

Es un modelo de proceso estándar para el desarrollo de proyectos de minería de datos y análisis de datos.

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

supervisado: cuando se le dan los datos de entrada y las respuestas para que efectúe análisis sobre los mismos, y así logre predicciones más acertadas

no supervisado: se dan los datos de entrada pero no la respuestas

reforzado: es cuando se le da al algoritmo las funciones /condiciones y el sistema aplica la información sin ningún patrón y aprende de sí mismo

semi supervisado: ~~se utiliza un modelo supervisado y una vez se obtienen respuestas, sobre las mismas se generan nuevas procesos y se agrupan de distintas formas para obtener información más detallada.~~

***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 Tensor Flow?

son bibliotecas de aprendizaje predictivo

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.

cuando se utiliza un algoritmo de predicción independiente que se relaciona con otros elementos independientes

*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.*

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

~~supervisado~~

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

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

falta o ausencia de datos además de la optimización, por que pueden ser o no válidos o representativos de lo que se quiere obtener

1. ***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.
2. ***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.
3. ***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.
4. ***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.
5. ¿Qué es el residuo en Machine Learning?

es la diferencia que hay entre el resultado obtenido y el resultado esperado