

BIBLIOTECAS PARA EL ANÁLISIS DE DATOS

1. *Pandas* :

Ofrece estructuras de datos flexibles y herramientas para manipular datos tabulares y de series temporales de manera eficiente, lo que la convierte en una opción popular para el análisis de datos.

2. *NumPy* :

Ofrece un objeto de matriz multidimensional (ndarray) que es eficiente y rápido para realizar operaciones matemáticas en grandes conjuntos de datos, lo que la convierte en una herramienta fundamental para el análisis y procesamiento de datos.

3. *Matplotlib* :

Es una biblioteca de visualización de datos en 2D en Python. Permite crear gráficos estáticos, gráficos de barras, histogramas, diagramas de dispersión, etc.

4. *Seaborn* :

Es una biblioteca de visualización de datos basada en Matplotlib que proporciona una interfaz de alto nivel para crear gráficos estadísticos atractivos e informativos. Es especialmente útil para visualizar datos complejos.

5. *SciPy* :

Es una biblioteca de código abierto para matemáticas, ciencia e ingeniería que proporciona funcionalidades adicionales a NumPy. Incluye módulos para optimización, álgebra lineal, integración, interpolación, funciones especiales, etc.

6. *Scikit-learn* :

Es una biblioteca de aprendizaje automático de código abierto en Python que proporciona herramientas simples y eficientes para el análisis predictivo de datos. Incluye algoritmos de clasificación, regresión, agrupación, reducción de dimensionalidad, selección de modelos, entre otros.

7. *Statsmodels* :

Es una biblioteca de Python que proporciona clases y funciones para la estimación de modelos estadísticos, pruebas estadísticas y exploración de datos. Es útil para analizar regresión y modelos lineales.

8. *Pandas-profiling* :

Es una biblioteca que genera automáticamente informes exploratorios de datos a partir de DataFrames de Pandas. Proporciona una visión general rápida y completa de la estructura de los datos, las relaciones entre variables y la presencia de valores atípicos.

9. *Dask* :

Es una biblioteca de paralelización y programación distribuida en Python que permite el escalado de análisis de datos a conjuntos de datos más grandes que no caben en la memoria de una sola máquina. Permite realizar operaciones paralelas en Pandas, NumPy y otras estructuras de datos.

10. *TensorFlow y PyTorch* :

Estas son bibliotecas de aprendizaje profundo (deep learning) en Python. TensorFlow y PyTorch proporcionan herramientas para construir y entrenar modelos de redes neuronales profundas para tareas como clasificación, reconocimiento de imágenes, procesamiento del lenguaje natural, entre otros.

DEFINICION DE TÉRMINOS PARA EL ANÁLISIS DE DATOS

1. DataFrame :

Es una estructura de datos bidimensional, tabular y mutable en Pandas que se asemeja a una hoja de cálculo o una tabla de base de datos SQL. Está compuesto por filas y columnas etiquetadas, donde cada columna puede contener datos de diferentes tipos (numéricos, de cadena, booleanos, etc.). Es una de las estructuras de datos más utilizadas para el análisis de datos en Python.

2. DataTable :

Es una estructura de datos bidimensional en la biblioteca .NET que se utiliza para almacenar y manipular datos en forma de tabla. Es similar a un DataFrame en Pandas, pero en el contexto de .NET. Proporciona una forma de trabajar con datos de manera eficiente y ofrece funcionalidades para realizar consultas, filtrar datos, ordenar, entre otras operaciones comunes en el análisis de datos.

3. Series (para complementar la definición de DataFrame):

Es una estructura de datos unidimensional en Pandas que puede contener datos de cualquier tipo. Se puede pensar en una Series como una columna en un DataFrame. Cada elemento de una Series tiene una etiqueta o índice asociado que permite acceder a los datos de forma rápida y sencilla.

4. Array:

Es una estructura de datos que contiene una colección de elementos del mismo tipo, organizados en una o más dimensiones. En el contexto del análisis de datos, un Array NumPy (ndarray) es particularmente útil, ya que permite realizar operaciones matemáticas y estadísticas eficientes en grandes conjuntos de datos.

5. SQL (Structured Query Language):

Es un lenguaje de programación utilizado para administrar y manipular bases de datos relacionales. Permite realizar operaciones como consultas para recuperar datos, insertar nuevos datos, actualizar registros existentes y eliminar datos. El conocimiento de SQL es fundamental para trabajar con bases de datos en el análisis de datos.

6. JSON (JavaScript Object Notation):

Es un formato ligero de intercambio de datos usado para transmitir datos estructurados entre servidor y cliente en aplicaciones web. Es legible para los humanos y fácil de entender. En el análisis de datos, JSON se utiliza a menudo para almacenar y compartir datos de forma eficiente.

7. API (Interfaz de Programación de Aplicaciones):

Es un conjunto de reglas y protocolos que permite a diferentes aplicaciones comunicarse entre sí. Esencialmente, una API define cómo interactuar con un sistema de software determinado, especificando los tipos de solicitudes que se pueden realizar, cómo se deben estructurar esas solicitudes y qué tipos de respuestas se pueden esperar.