

Modulo 2



LECCIÓN 2

Manejo de los Datos para aplicar ML

Manejo de los Datos para aplicar ML

Tiempo de ejecución: 2 horas

Objetivo de la Lección:

En esta lección, exploraremos las etapas fundamentales del manejo de datos en el contexto del Machine Learning. Desde la adquisición de datos hasta la creación de bases de datos estructuradas, abordaremos las técnicas esenciales para garantizar que los datos sean una base sólida para el desarrollo de modelos efectivos.

Agenda de la Lección:

1. Adquisición de Datos:

- Fuentes de Datos: Introducción a diversas fuentes de datos, como conjuntos de datos preexistentes, archivos CSV, bases de datos, y APIs. Analizaremos las consideraciones al seleccionar la fuente de datos más adecuada para un proyecto específico.
- Carga de Datos con Scikit-Learn: Utilización de funciones de Scikit-Learn para cargar conjuntos de datos. Exploraremos ejemplos prácticos de carga y visualización inicial de datos.
- Calidad y Relevancia: Discusión sobre la importancia de la calidad y relevancia de los datos. Estrategias para evaluar y mejorar la calidad de los datos adquiridos.

2. Creación de la Base de Datos a partir de los Datos Obtenidos:

- Limpieza y Preprocesamiento: Técnicas para limpiar y preprocesar datos, incluyendo la identificación y manejo de valores faltantes y duplicados.

- Transformación de Datos: Abordaremos la transformación de datos categóricos y numéricos para prepararlos para su uso en modelos de Machine Learning.
- Visualización Inicial: Exploraremos herramientas de visualización para comprender la distribución de datos y detectar patrones preliminares.

Actividades Prácticas:

Los participantes aplicarán los conceptos aprendidos utilizando Scikit-Learn para cargar un conjunto de datos real. Realizarán tareas de exploración y preprocesamiento, promoviendo la interactividad y la aplicación práctica de los conocimientos adquiridos.