# Del Excel al Machine Learning: Análisis de Datos Inteligente

## Descripción del curso

Este curso práctico te lleva paso a paso desde el uso de Excel para analizar datos hasta la aplicación de modelos de Machine Learning con Python. Aprenderás a explorar, limpiar y visualizar información de manera inteligente, sin necesidad de conocimientos previos de programación avanzada. Al finalizar, podrás automatizar análisis, crear reportes dinámicos y aplicar técnicas de aprendizaje automático en tus propios proyectos.

## Objetivo general

Capacitar a los participantes en el uso integrado de Excel y Python para el análisis de datos y la aplicación de técnicas básicas de Machine Learning, orientadas a la toma de decisiones basadas en evidencia.

## Objetivos específicos

• Utilizar funciones de Excel para preprocesamiento, análisis exploratorio y generación de base de datos para procesamiento de Python con técnicas de máquina de aprendizaje (Machine Learning).

• Importar y procesar datos con Python usando pandas y numpy.

• Aplicar modelos básicos de Machine Learning con scikit-learn.

• Automatizar reportes y análisis combinando Excel y Python.

## Contenidos por sesión

### Sesión 1. Introducción al análisis de datos inteligente (4 h)

* Machine Learning y la importancia del preprocesamiento de bases de datos.
* Tipos de datos, variables y estructuras, configuración de datasets.
* Modelos básicos: regresión lineal y clasificación.
* Conceptos clave: features, target, entrenamiento y validación.
* Evaluación del modelo: precisión, R² y matriz de confusión.
* Técnicas de limpieza de datos: (IQR: Interquartil Range Method y Tukey).
* El flujo de trabajo: desde Excel hasta modelos predictivos.

### Sesión 2. Excel para análisis de datos (4 h)

* Repaso de funciones, tablas dinámicas, filtros y segmentación avanzada.
* Funciones estadísticas y lógicas aplicadas (CUARTIL.EX, CONTAR.SI, SI.CONJUNTO, MIN.SI.CONJUNTO, MAX.SI.CONJUNTO, PROMEDIO.SI, CONTAR.SI.CONJUNTO).
* Limpieza de datos con método estadístico (Tukey) y manual.
* Exportación de datos limpios a CSV.

Desarrollo Plus Demostrativo: Visualización de Datos en Power BI.

### Sesión 3. Python para preprocesamiento de datos con Numpy y Matplotlib y machine learning con Scikit-learning (4 h)

* Google Colab para operación de Python, entorno de trabajo y librerías: numpy, matplotlib, scikit-learn.
* Lectura y escritura de archivos Excel/CSV.
* Limpieza automática de datos: valores nulos, duplicados y tipos de variables.
* Visualización básica con gráficos de dispersión, histogramas y boxplots.
* Features, target, entrenamiento y validación.
* Evaluación del modelo: precisión, R² y matriz de confusión.
* Exportación de resultados y gráficos a Excel.

### Sesión 4. Automatización y reportes inteligentes (4 h)

* Uso de openpyxl y pandas.ExcelWriter para automatizar reportes.
* Creación de dashboards simples con resultados del modelo.
* Ejemplo práctico de flujo completo: carga de datos, modelado y visualización.
* Buenas prácticas para compartir y documentar resultados.

## Cronograma sugerido

|  |  |  |
| --- | --- | --- |
| Semana | Tema | Actividades principales |
| 1 | Introducción al análisis de datos inteligente | Exploración de dataset y conceptos básicos. |
| 2 | Excel para análisis de datos | Análisis descriptivo y limpieza de datos en Excel. |
| 3 | Python para preprocesamiento, análisis y machine learning | Preprocesamiento, visualización e implementación de un modelo simple predictivo con Python. |
| 4 | Automatización y reportes | Integración Excel–Python y generación de reportes. |

## Bibliografía y recursos digitales

• McKinney, W. (2022). \*Python for Data Analysis\*. O’Reilly Media.

• Géron, A. (2023). \*Hands-On Machine Learning with Scikit-Learn and TensorFlow\*. O’Reilly Media.

• Documentación oficial de pandas: https://pandas.pydata.org/docs/

• Documentación oficial de scikit-learn: https://scikit-learn.org/stable/

• Portal de datos abiertos: https://www.kaggle.com/datasets

## Duración total del curso

El curso tiene una duración total de 16 horas, distribuidas en cuatro sesiones teórico-prácticas de 4 horas cada una. Cada sesión combina la explicación conceptual con ejercicios guiados y actividades aplicadas en Excel y Python.

## Distribución horaria por sesión

|  |  |  |  |
| --- | --- | --- | --- |
| Sesión | Tema | Duración | Distribución sugerida |
| 1 | Introducción al análisis de datos inteligente | 4 horas | 4 h teoría y conceptos + 1 h con dataset. |
| 2 | Excel avanzado para análisis de datos | 4 horas | 1 h repaso de funciones + 3 h ejercicios en Excel |
| 3 | Python para exploración y preprocesamiento de datos | 4 horas | 4 h práctica en Jupyter Notebook en Google Colab |
| 5 | Automatización y reportes inteligentes | 4 horas | 1 h conceptos + 3 h práctica de integración Excel–Python |