# 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 avanzadas de Excel para el análisis exploratorio de datos.

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

• Visualizar información de manera efectiva mediante gráficos y dashboards.

• 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

• Qué es el análisis de datos y cómo se relaciona con Machine Learning.

• El flujo de trabajo: desde Excel hasta modelos predictivos.

• Tipos de datos, variables y estructuras.

• Herramientas gratuitas necesarias y entorno de trabajo.

### Sesión 2. Excel avanzado para análisis de datos

• Tablas dinámicas, filtros y segmentación avanzada.

• Funciones estadísticas y lógicas aplicadas (PROMEDIO.SI, BUSCARX, CORREL).

• Limpieza de datos, detección de errores y normalización manual.

• Exportación de datos limpios a CSV.

### Sesión 3. Python para exploración y preprocesamiento de datos

• Introducción a pandas, numpy y matplotlib.

• 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.

### Sesión 4. Fundamentos de Machine Learning con Scikit-learn

• Conceptos clave: features, target, entrenamiento y validación.

• Modelos básicos: regresión lineal y clasificación.

• Evaluación del modelo: precisión, R² y matriz de confusión.

• Exportación de resultados y gráficos a Excel.

### Sesión 5. Automatización y reportes inteligentes

• 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 | Exploración de dataset y conceptos básicos. |
| 2 | Excel avanzado | Análisis descriptivo y limpieza de datos en Excel. |
| 3 | Python para análisis | Preprocesamiento y visualización de datos con Python. |
| 4 | Machine Learning básico | Implementación de un modelo predictivo simple. |
| 5 | Automatización y reportes | Integración Excel–Python y generación de reportes. |

## Rúbricas de evaluación

|  |  |  |
| --- | --- | --- |
| Criterio | Peso | Descripción |
| Participación y asistencia | 20% | Intervención activa en las sesiones y cumplimiento de actividades. |
| Ejercicios prácticos semanales | 50% | Desarrollo correcto de ejercicios y análisis de resultados. |
| Evaluación final de comprensión | 30% | Resolución de cuestionario integrador sobre los temas vistos. |

## 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 20 horas, distribuidas en cinco 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 | 1 h teoría + 3 h práctica 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 | 1 h conceptos + 3 h práctica en Jupyter Notebook |
| 4 | Fundamentos de Machine Learning con Scikit-learn | 4 horas | 1.5 h teoría + 2.5 h práctica de modelado |
| 5 | Automatización y reportes inteligentes | 4 horas | 1 h conceptos + 3 h práctica de integración Excel–Python |

## Rúbricas de evaluación (curso de 20 horas)

|  |  |  |
| --- | --- | --- |
| Criterio | Peso | Descripción |
| Participación y asistencia | 20% | Intervención activa en las sesiones, puntualidad y cumplimiento de actividades. |
| Ejercicios prácticos semanales | 50% | Desarrollo correcto de ejercicios en Excel y Python, demostrando comprensión del flujo de análisis. |
| Evaluación final de comprensión | 30% | Resolución de un cuestionario integrador sobre los temas vistos en las cinco sesiones. |