

Macroentrenamiento en **MeIA** **20**  
**Inteligencia**  
**Artificial** **25**

Módulo 2 - Especialización

Minería de Datos con Python y Altair AI Studio

Presentada por: Eduardo Espinosa Avila

Junio 2025

## Contenido

Título del tutorial . . . . .	2
Temario detallado del tutorial . . . . .	2
Descripción . . . . .	2
Temario detallado . . . . .	3
Requisitos para los participantes . . . . .	4
Metodología de enseñanza y evaluación . . . . .	4
Calendario tentativo para la impartición del módulo . . . . .	4
Datos del instructor . . . . .	5

## Título del tutorial: Minería de Datos con Python y Altair AI Studio®

### Temario detallado del tutorial

#### Descripción:

Este tutorial es importante debido a que proporciona herramientas para profundizar en el camino de conocimiento de la minería de datos con Python, así como conocimientos de algunos modelos de aprendizaje de gran uso en la actualidad para extraer resultados significativos a partir de los datos. Además, introducirá a las personas participantes en el uso de flujos de trabajo automatizados para procesos de minería de datos y el uso de una herramienta de desarrollo sin código para crear modelos de forma rápida para poder aplicarse a problemas de diversa naturaleza; entre ellos: procesamiento de lenguaje natural, visión por computadora, medicina, etc.

## Temario detallado:

### Módulo 1 – Herramientas de Python para Minería de Datos

- Pandas
- Numpy
- Matplotlib
- SciKitLearn
- MLXtend

### Módulo 2 – Aprendizaje no Supervisado

- Detección básica de anomalías y valores atípicos
  - Varianza y *boxplot*
- Análisis de agrupamientos
  - K-Medias, K-Modas, Jerárquico por aglomeración, Mezclas Gaussianas, DBSCAN.
- Métricas de evaluación
- Análisis de componentes principales
  - Básica y por núcleos.

### Módulo 3 – Aprendizaje Supervisado

- Regresión lineal y polinomial
- Perceptrón
- Regresión logística
- Máquinas de soporte vectorial
- k-vecinos: clasificación y regresión
- Naive Bayes
- XGBoost: clasificación y regresión
- Métricas de evaluación de modelos

### Módulo 4 – Mejores prácticas para evaluar modelos y ajustar hiperparámetros

- Introducción a *pipelines* para automatizar flujos de trabajo
- Evaluación del rendimiento de un modelo con validación cruzada
- Mejora de algoritmos usando curvas de aprendizaje y validación
  - Diagnosticando problemas de sesgo y varianza con curvas de aprendizaje
  - Enfrentando sub y sobreajuste con curvas de validación
- Ajuste fino de modelos de aprendizaje automático
  - Ajuste de hiperparámetros con búsqueda de malla
- Selección de algoritmos con validación cruzada anidada

### Módulo 5 – Introducción al uso de Altair AI Studio®

- Introducción
- Creación de un proyecto
- Réplica de algún ejemplo hecho con Python con ayuda de Altair AI Studio®

## Requisitos para los participantes

Se utilizará Google Colab® para el desarrollo de los ejemplos; se requiere una cuenta de gmail, en caso de necesitar alguna biblioteca adicional se puede instalar durante el mismo taller.

Altair AI Studio® brinda una licencia de uso gratuito con la limitación de uso para no más de 10,000 registros; se puede descargar y generar la licencia en:

<https://docs.rapidminer.com/latest/studio/installation/index.html>

<https://community.altair.com/discussion/33624/steps-to-install-and-activate-altair-ai-studio-for-educational-use-from-rapidminer-com-with-rapidminer-licensing-personal-edition>

## Metodología de enseñanza y evaluación

La metodología a seguir es aprender haciendo, presentando la parte teórica de los algoritmos y modelos, para posteriormente realizar al menos un ejemplo práctico de uso del mismo.

He aplicado esta metodología tanto en mis cursos regulares como en cursos extracurriculares y diplomados dando muy buenos resultados en la mayoría de los casos.

Respecto a la evaluación se pueden realizar ejercicios y cuestionarios dentro de las sesiones síncronas; además de realizar miniproyectos para reforzar lo aprendido.

## Calendario tentativo para la impartición del módulo

Lunes 9 de junio:

- Módulo 1: 1 hora
- Módulo 2 hasta agrupamientos y sus métricas: 3 horas

Martes 10 de junio:

- Módulo 2, Análisis de componentes principales 1 hora
- Módulo 3 hasta regresión logística: 3 horas

Miércoles 11 de junio:

- Módulo 3: 4 horas

Jueves 12 de junio:

- Módulo 4: 4 horas

Viernes 13 de junio:

- Módulo 5: 4 horas

Nota: Este calendario es flexible, en caso de ser necesario, se puede adaptar reduciendo el número de horas del módulo 5

## Datos del instructor

En caso de seleccionar la modalidad por equipo, indicando quién es el representante del equipo.

- Eduardo Espinosa Avila.
- Universidad Nacional Autónoma de México.
- Doctor en Ciencias (Computación).
- Inteligencia artificial, aprendizaje automático.