



Diplomado en Machine Learning y Analítica de Datos en Python

El diplomado en *Machine Learning y Analítica de Datos en Python* ofrece una introducción a los conceptos fundamentales y herramientas claves para el diseño de soluciones basadas en el análisis descriptivo y predictivo de datos. El diplomado está diseñado específicamente para atender las necesidades de la empresa EPM y capacitar a su personal en los elementos necesarios para comenzar a desarrollar soluciones basadas en análisis de datos que respondan a las necesidades de la empresa. El diplomado realiza un recorrido por los componentes fundamentales de los sistemas basados en Machine Learning, a través de un enfoque práctico que le permitirá a los asistentes aplicar cada uno de los conceptos y técnicas discutidos en las diferentes sesiones.

Objetivo

Proveer a los asistentes con los conocimientos y aplicaciones de herramientas básicas del análisis de datos en el lenguaje de programación Python, partiendo del análisis exploratorio de los mismos y llegando al planteamiento de métodos/algoritmos de aprendizaje automático básicos, con el fin de hacer inferencias y predicciones necesarias para el entendimiento de los fenómenos de interés de EPM y guía en la posterior toma de decisiones.

Contenido

El diplomado está dividido en tres módulos organizados de la siguiente manera:

Módulo 1: Análisis estadístico y manipulación de datos en python

Duración 20h.

- Introducción a Python y al entorno de programación
 - El lenguaje de programación Python y la herramienta Jupyter Notebook
 - Operadores y estructuras básicas
 - Para no perder el foco de la formación, el resto de componentes del lenguaje se irán presentando en el transcurso del Diplomado.
- Exploración de problemas y proyectos de EPM
 - Taller para analizar los problemas de interés para abordar en el diplomado
- Carga y manipulación de datos
 - Importación/exportación de datos
 - Nombramiento de variables
 - Tipo de datos
 - Transformación de variables
 - Ordenamiento, fusión, agregación y filtro de datos
 - Conversión de datos (codificación de variables)



- **Análisis exploratorio de datos**
 - Estadística descriptiva
 - Tablas de frecuencia
 - Correlaciones
 - Missing data
- **Visualización de datos**
 - Histograma
 - Gráficos de densidad
 - Gráficos de correlación
 - Gráficos box-plot y bagplot
 - Gráficos de violín
 - Gráficos de dispersión

Módulo 2: Fundamentos de Machine Learning

Duración 44 h.

- **Introducción al Machine Learning**
 - Modelos derivados de los datos
 - IA vs ML vs DL vs Big Data vs Data Science
 - Tipos de aprendizaje: supervisado, no supervisado y de refuerzo
 - Tareas de aprendizaje: clasificación, regresión, clustering, regresión ordinal, etc.
 - Procesos de anotación de datos
- **Exploración de problemas y proyectos de EPM**
 - Taller para analizar qué tareas objetivo, tipos de modelos, volumetrías de datos disponibles son abordables a través de métodos de ML
- **Flujo de trabajo del usuario de métodos de ML**
 - Concepto del proceso iterativo y experimental
 - Complejidad de datos vs. complejidad de modelos
 - Construcción de evidencia y confiabilidad de desempeño métodos de ML
 - Métricas y consideraciones de evaluación de desempeño. sobreajuste, sesgo, curvas de aprendizaje
 - Despliegue, monitoreo y recalibración de modelos
- **Flujo de trabajo en el diseño de métodos de ML**
 - Planteamiento de la función de pérdida
 - Proceso de optimización y requerimientos computacionales
 - Ejemplo con regresión logística y regresión lineal
 - Expansión de características y regularización.
- **Algoritmos supervisados**
 - Árboles de decisión.
 - Métodos de ensemble. Bagging, Random Forests y Boosting
 - Support Vector Machines
 - Redes neuronales
- **Algoritmos no supervisados**
 - Clustering



- Estimación de densidad
 - Extracción de características
 - Detección de anomalías
 - Compleción de matrices
- Abordaje con ML de problemas de EPM
 - Taller para iniciar la implementación con métodos de ML de problemas seleccionados de EPM
- Gestión de proyectos ML y conformación de equipos:
 - Métricas de ML vs métricas de negocio.
 - Perfiles profesionales: data scientist, ML engineer, data engineer, data labeller, data product manager, etc.
 - Dinámica y best practices de la comunidad ML (Kaggle, etc.)
 - Procesos de adopción de ML e IA en las organizaciones.

Módulo 3: Introducción a TensorFlow

Duración 20h.

- Fundamentos
 - TensorFlow APIs, Sesiones, Variables, Operadores
 - Grafo Computacional y computación simbólica
 - Funciones de costo y optimización
- Construcción de modelos y optimización
 - Redes Neuronales
 - Algoritmos de optimización
 - API funcional y de secuencia
- Convolutional Neural Networks
 - Transfer learning
 - Arquitecturas de redes
- Aspectos adicionales:
 - Tensor Board
 - Estimator API
 - Data pipelines
 - Deployment: TFServe, TFLight

Módulo 4: Análisis de series de tiempo

Duración 36h.

- Introducción a los conceptos básicos de series de tiempo
 - Dependencia temporal
 - Estacionariedad
 - Regularidades empíricas de las series de tiempo
- Medidas relacionadas con las series de tiempo
 - Función de media
 - Función de autocovarianza



- Funciones de autocorrelación y de autocorrelación parcial
- Problemas de series de tiempo en el contexto de EPM
- Modelos básicos de series de tiempo
 - Modelos estacionarios autorregresivos -AR-
 - Modelos de medias móviles -MA-
 - Modelos estacionarios autorregresivos de medias móviles -ARMA-
 - Modelos no estacionarios -ARIMA-
 - Aplicaciones (identificación, estimación y diagnóstico)
- Series de tiempo no convencionales
 - Clasificación y anotación de señales
 - Procesamiento de Texto
- Métodos de ML aplicados a series de tiempo
 - Estrategia rolling window
 - Estrategia walk-forward
 - Redes neuronales recurrentes
 - Arquitecturas de redes recurrentes en TensorFlow: one-to-many, many-to-many, sequence-to-sequences.

Metodología

El diplomado se desarrollará en sesiones magistrales en las cuales se introducirán los conceptos básicos y se discutirá la intuición detrás de las diferentes metodologías y técnicas expuestas, además de la presentación de ejemplos prácticos y la solución de ejercicios usando la herramienta jupyter y el lenguaje de programación python. Los asistentes deberán desarrollar algunos talleres propuestos y podrán aplicar las habilidades adquiridas en problemas de particular interés para la empresa, siempre y cuando se cuente con la información necesaria para tal fin.

Consideraciones adicionales

El diplomado se llevará a cabo en las instalaciones EPM. Durante el desarrollo del diplomado cada asistente debe contar con un computador personal para el seguimiento de las sesiones. Las herramientas informáticas que serán usadas durante el curso son de acceso libre. La organización del diplomado pone a disposición de los asistentes un monitor que los apoyará en las labores de instalación y configuración de las herramientas necesarias para el desarrollo de las prácticas, así como en el apoyo durante la realización de talleres y prácticas.

Profesores

- **Julián D. Arias Londoño**, Ingeniero Electrónico, Master en Ingeniería - Automatización Industrial de la Universidad Nacional de Colombia, Doctor en Ingeniería de Sistemas de la Universidad Politécnica de Madrid, España y Doctor en Ingeniería - Automática de la Universidad Nacional de Colombia. Profesor Asociado del Departamento de Ingeniería de Sistemas de la Facultad de ingeniería de la Universidad de Antioquia. Áreas de interés y trabajo académico: aprendizaje automático (Machine Learning), aprendizaje Bayesiano, Deep Learning y



In2Lab
Intelligent Information
Systems Lab | I+D+i



**UNIVERSIDAD
DE ANTIOQUIA**

Facultad de Ingeniería

procesamiento digital de señales aplicados al procesamiento de voz, procesamiento de lenguaje natural, detección de fallas en diferente tipo de sistemas y clasificación de secuencias.

E-mail: julian.ariasl@udea.edu.co

Página personal: <https://sites.google.com/site/juliandariaslondono/>

- **Santiago A. Gallón.** Economista y Master en Economía de la Universidad de Antioquia, y Doctor en Matemáticas Aplicadas (área de probabilidad y estadística) del Instituto de Matemáticas Aplicadas, Universidad de Toulouse III -Paul Sabatier-, Toulouse-Francia. Actualmente, Profesor Asociado del Departamento de Matemáticas y Estadística de la Facultad de Ciencias Económicas de la Universidad de Antioquia. Áreas de interés y trabajo académico: análisis de datos funcionales, aprendizaje automático estadístico, estadística de alta dimensión, análisis de series de tiempo y econometría aplicada, principalmente en economía, finanzas, educación superior y mercado de energía.

E-mail: santiago.gallon@udea.edu.co

Página personal: <https://santiagogallongomez.wordpress.com/>

- **Raúl Ramos Pollán,** Profesor e investigador de la Universidad de Antioquia en las áreas de Inteligencia Artificial y Ciencia de Datos. PhD en Ingeniería Informática por la Universidad de Oporto, Portugal, en análisis de imágenes biomédicas, ha desarrollado su carrera en la industria y en la academia, como Director del Centro de Cómputo del CETA-CIEMAT (Extremadura, España, www.ceta-ciemat.es), Ingeniero de Software en el Centro Europeo para la Física de Partículas (www.cern.ch, Ginebra, Suiza), Arquitecto Java para Sun Microsystems Suiza (www.oracle.com) y cofundador de Pildo Labs (www.pildo.com), una PYME en el sector aeronáutico y de software crítico basada en Barcelona, antes de establecerse en Colombia en 2012, inicialmente como investigador invitado en la Universidad Nacional de Colombia en Bogotá en el grupo MindLab en analítica de imágenes, en la Universidad Industrial de Santander desde 2013 y en la Universidad de Antioquia desde 2018.

E-mail: raul.ramos@udea.edu.co

Página personal: <https://sites.google.com/site/rulixrp>

Horario: Miércoles 14:00 - 18:00 y Viernes 8:00 - 12:00