

# Diplomado en Ciencia de Datos y Aprendizaje de Máquinas

## JUSTIFICACIÓN

El aprendizaje de máquinas combina las ciencias de la computación y la estadística, buscando producir métodos que permitan a las máquinas aprender por sí mismas a partir de la experiencia y hacer predicciones basadas en ella. Los sistemas de aprendizaje automático desempeñan un papel importante en muchas aplicaciones, utilizándose cada vez más en diversos campos como la agricultura, la biología, la industria, la medicina, la robótica y la investigación científica en general. Por ello, esta área es de enorme importancia en la formación profesional de una amplia variedad de profesiones ya que permite al estudiante comprender, analizar, utilizar y desarrollar técnicas de aprendizaje automático en procesos de modelado y análisis de datos. Así, el alumno conocerá y comprenderá los fundamentos del aprendizaje automático, sus características y conceptos básicos, y será capaz de emplearlas en diversos campos para el análisis de datos.

## DIRIGIDO A

- Profesionales con conocimientos de programación que participen en el análisis de datos y toma de decisiones.
- Investigadores científicos y tecnológicos.
- Desarrolladores de aplicaciones de software basadas en aprendizaje de máquinas.

## RESULTADOS DE APRENDIZAJE

- Instalar y utilizar herramientas en Python necesarias para el manejo de datos y aprendizaje de máquinas.
- Utilizar métodos de ETL para la limpieza y manejo eficiente de datos.
- Conocer y utilizar métodos para la determinación de las variables más discriminantes mediante métodos de selección de descriptores.
- Conocer las bases de las técnicas más conocidas del aprendizaje de máquinas para su uso adecuado de acuerdo a sus características.
- Implementar métodos de aprendizaje de máquinas en Python empleando métodos supervisados o no supervisado de acuerdo a las características de los datos para su uso práctico.
- Conocer y utilizar métricas para la evaluación de los resultados obtenidos empleando técnicas de aprendizaje de máquinas.
- Optimización de modelos mediante el ajuste de parámetros para obtener un mejor rendimiento.

## METODOLOGÍA

- Clase magistral para la enseñanza de las bases del conocimiento teórico de la ciencia de datos y el aprendizaje de máquinas.
- Aprendizaje basado en casos.
- Aprendizaje basado en proyectos.
- Aprendizaje colaborativo.

## CONTENIDO

**MÓDULO 1.** Introducción al aprendizaje de máquinas. Tipos de métodos de aprendizaje. Conceptos fundamentales de la teoría de decisión, SkLearn.

**MÓDULO 2.** ETL, limpieza de datos, manejo de datos faltantes.

**MÓDULO 3.** Clasificador bayesiano, análisis de muestras, determinación de muestras más discriminantes, LDA, PCA, curvas ROC, matriz de confusión, clasificadores no supervisados: K-means, silueta, modelos gaussianos mixtos, expectación maximización, AIC, BIC.

**MÓDULO 4.** Aproximación no paramétrica: métodos basados en árboles. Árboles de decisión, bosques aleatorios, GB decisión trees, XGBoost, variación y sesgo, vecinos más próximos. Ventanas de Parzen, mejoras en velocidad: K-D tree, local sensitive hashing, lista invertida.

**MÓDULO 5.** Clasificación por maximización de márgenes. Clasificador por centroides, clasificador pasivo-agresivo, convergencia, SVM, kernels.

**MÓDULO 6.** Regresión. Regresión lineal, regresión multivariable, regresión polinómica, gradiente descendente, regresión logística.

**MÓDULO 7.** Redes neuronales, introducción, neurona, perceptrón, backpropagation, sintonización, aprendizaje profundo, redes neuronales convolucionales.

## EXPERTOS FACILITADORES

### MANUEL GUILLERMO FORERO VARGAS.

Ingeniero electrónico de la Universidad Javeriana, Magister en ingeniería eléctrica en la especialidad de bioingeniería y control de la Universidad de Los Andes, Master en imagenología médica y Doctor en ingeniería biomédica de la Université de Technologie de Compiègne (Francia), postdoctorado en imágenes de microscopía en el Centro Superior de Investigaciones Científicas de Madrid (España). Ha sido docente (profesor asociado) del programa de Ingeniería de Sistemas de la Universidad Nacional de Colombia, profesor visitante de la Universidad de Salamanca (España), profesor de la Facultad de Ingeniería Electrónica de la Universidad Escuela Colombiana de Ingeniería, investigador de las Universidades de Birmingham y Cardiff (Reino Unido), actualmente es decano e investigador de la Facultad de Ingeniería de la Universidad de Ibagué. Experiencia posdoctoral de más de 25 años como investigador en las áreas de procesamiento de imágenes digitales y aprendizaje de máquinas, habiendo participado como conferencista invitado en congresos internacionales en Chile, Perú, México, España, Portugal y publicado 96 artículos reconocidos en Scopus. Actualmente colabora con grupos de investigación de Italia, Francia, Reino Unido, Perú y Colombia en el desarrollo de métodos de procesamiento de imágenes y aprendizaje de máquinas para su uso en el análisis de muestras biológicas.

## MODALIDAD

### REMOTO

## DURACION DEL PROGRAMA

### 90 HORAS