



# Probabilidad, estadística y machine learning en ciencia de datos

**ÁREA: DESARROLLO DE SOFTWARE**

## PRESENTACIÓN

En el contexto actual, el análisis de datos se ha convertido en una herramienta esencial para la toma de decisiones en diversos sectores como finanzas, salud, tecnología y manufactura. La creciente cantidad de datos generados diariamente requiere profesionales capacitados en técnicas avanzadas y especializadas de análisis para interpretar y utilizar esta información de manera efectiva. Este curso responde a la necesidad de formar expertos en ciencia de datos, capaces de aplicar métodos predictivos, estadísticos y de machine learning utilizando herramientas de programación para resolver problemas complejos y optimizar procesos en el mundo real.

## PERFIL DE INGRESO

Este curso está dirigido a las personas interesadas en profundizar sus conocimientos a un nivel avanzado y especializado en el uso de las herramientas de Ciencia de Datos, con enfoque práctico. Profesionales en áreas de ciencia de datos, inteligencia artificial, estadísticas, ingeniería o tecnologías de la información, así como en industrias donde la toma de decisiones basada en datos y los modelos predictivos son fundamentales. Estudiantes en disciplinas afines interesados en prepararse para roles especializados en análisis de datos y machine learning. Así como investigadores que deseen aplicar análisis cuantitativo avanzado en sus proyectos. Se requiere haber acreditado o demostrar conocimientos y habilidades equivalentes al curso de Introducción a la ciencia de datos utilizando PYTHON.

## OBJETIVO

El participante aplicará, analizará y evaluará técnicas de alto nivel y especializadas en ciencia de datos, análisis predictivo, probabilidad, estadística y machine learning, utilizando modelos y herramientas de programación para diseñar, implementar y optimizar soluciones a problemas complejos en distintos ámbitos.

## TEMARIO

1. PROBABILIDAD PARA CIENCIA DE DATOS
  - 1.1 Fundamentos de probabilidad
  - 1.2 Distribuciones de probabilidad
  - 1.3 Distribución Binomial, Poisson y Normal
  - 1.4 Medidas probabilísticas avanzadas
  - 1.5 Simulaciones probabilísticas
2. ESTADÍSTICA PARA CIENCIA DE DATOS
  - 2.1 Estadística descriptiva avanzada
  - 2.2 Estadística inferencial avanzada
  - 2.3 Análisis multivariante
  - 2.4 Modelos estadísticos avanzados
3. ANÁLISIS PREDICTIVO UTILIZANDO REGRESIÓN
  - 3.1 Introducción al análisis predictivo
  - 3.2 Modelos predictivos vs. descriptivos
  - 3.3 Tipos de regresión y sus aplicaciones
  - 3.4 Regresión lineal múltiple avanzada
  - 3.5 Métricas de evaluación e interpretación
  - 3.6 Regresión regularizada
4. *MACHINE LEARNING* AVANZADO SUPERVISADO Y NO SUPERVISADO
  - 4.1 *Machine learning* avanzado supervisado
  - 4.2 Ensamblados (*Random Forest, Gradient Boosting, XGBoost*)
  - 4.3 Máquinas de soporte vectorial (SVM)
  - 4.4 *Machine learning* avanzado no supervisado
  - 4.5 Modelos jerárquicos de *clustering*
5. INTRODUCCIÓN A REDES NEURONALES BÁSICAS
  - 5.1 Construcción de modelos en *Python*
  - 5.2 Uso de *Keras* y *TensorFlow*
6. OPTIMIZACIÓN Y EVALUACIÓN DE MODELOS
  - 6.1 Técnicas de optimización de modelos
  - 6.2 Métricas avanzadas de evaluación
  - 6.3 Optimización
  - 6.4 Automatización de *pipelines*

## PERFIL DE EGRESO

Al finalizar este curso, los participantes habrán adquirido conocimientos sólidos y habilidades en técnicas

avanzadas de regresión, probabilidad y estadística descriptiva e inferencial. Estarán capacitados para interpretar y modelar datos de manera efectiva, así como para implementar modelos predictivos avanzados, tanto supervisados como no supervisados. Además, contarán con experiencia en el uso de herramientas modernas de *machine learning*, incluyendo ensambles, máquinas de soporte vectorial y *clustering* jerárquico. Podrán diseñar y construir redes neuronales básicas utilizando *Python* y bibliotecas especializadas como *TensorFlow* y *Keras*. Asimismo, estarán preparados para optimizar y evaluar modelos analíticos mediante métricas avanzadas, aplicando sus conocimientos en la resolución de problemas complejos mediante proyectos prácticos con *datasets* reales en diversos contextos profesionales.

## REQUISITOS ACADÉMICOS

Se requieren conocimientos y habilidades en:

### 1. Estadística descriptiva:

- Medidas de tendencia central (media, mediana, moda).
- Medidas de dispersión (varianza, desviación estándar, rango intercuartílico).
- Representaciones gráficas de datos (histogramas, diagramas de caja, gráficos de dispersión).

### 2. Estadística inferencial:

- Conceptos de población y muestra.
- Teorema central del límite.

### 3. Regresión

- Regresión lineal simple.

## DURACIÓN

30 horas.

## RECURSOS INFORMÁTICOS

- Cuenta de correo electrónico.
- Conexión a internet.

## HARDWARE

- Procesador Intel Core I3 o superior.
- Memoria RAM instalada de 4GB como mínimo recomendable.
- Pantalla con una resolución de 1024x768 o superior.
- Cámara.
- Micrófono.
- Bocinas.

## SOFTWARE

- Microsoft Windows 10 o superior.
- Navegador de Internet (Chrome, Edge, Firefox, etc.), es recomendable utilizar las versiones más actuales.
- Python Pip 3.9 o superior.
- PyCharm Professional 2023.
- Anaconda 2023.11 o posterior.
- Jupyter (versión más reciente).
- Adobe Acrobat Reader u otro software libre para abrir los archivos PDF.
- Instalar la aplicación de Zoom, la cual se puede descargar de: <https://zoom.us/support/download>

## BIBLIOGRAFÍA

- Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction (2nd ed.)*. Springer.
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). *An Introduction to Statistical Learning: With Applications in R (2nd ed.)*. Springer.
- McKinney, W. (2022). *Python for Data Analysis: Data Wrangling with pandas, NumPy, and Jupyter (3rd ed.)*. O'Reilly Media.
- Bruce, P., Bruce, A. & Gedeck, P. (2020). *Practical Statistics for Data Scientists, 2nd Edition*. O'Reilly Media, Inc.
- Ozdemir, S. (2024). *Principles of Data Science - Third Edition*. Packt Publishing.
- Pierson, L. (2024). *Data Science Essentials For Dummies*. For Dummies.
- VanderPlas, J. (2022). *Python Data Science Handbook, 2nd Edition*. O'Reilly Media, Inc.
- Géron, A. (2022). *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow, 3rd. Edition*. O'Reilly Media, Inc.
- Kapoor, A., Gulli, A. & Pal, S. (2022). *Deep Learning with TensorFlow and Keras - Third Edition*. Packt Publishing.
- Schuler, N. (2023). *Statistics and Mathematics for Data Science and Data Analytics*. Packt Publishing. 11 hrs. 23 min.
- Krohn, J. (2022). *The Essential Machine Learning Foundations: Math, Probability, Statistics, and Computer Science (Video Collection)*. Pearson. 28 hrs. 12 min.
- Alby, T. (2024). *Data science in practice*. CRC Press.
- Kalita, K. (2024). *Fundamentals of data science: theory and practice*. Academic Press.

Febrero 2025

101531/E

