



DIPLOMADO DATA SCIENCE

29

VERSIÓN

MACHINE LEARNING E INTELIGENCIA ARTIFICIAL, DEEP LEARNING

MODALIDAD ON-LINE

SOFTWARE: PYTHON, R, SPARK, SQL*

Nuevos contenidos sobre la Inteligencia Artificial para lenguaje natural de tipo ChatGPT

* Se incluye acceso a la plataforma cloud de Amazon Web Services (AWS).

** Todas las clases son de 3 horas y empiezan a las 19 hrs. en modalidad “online”

PRESENTACIÓN DEL DIPLOMADO

- ▶ Miércoles 30 de Septiembre de 2026.
- ▶ 19:00 hrs.

INICIO DE CLASES

- ▶ Lunes 5 de Octubre de 2026.
- ▶ 19:00 hrs.

FECHA LÍMITE PARA MATRICULARSE

- ▶ 29 de septiembre de 2026 o hasta completar cupo máximo.

PROFESORES

- ▶ **Hamdi Raissi**, PhD Universidad de Lille, Francia, Profesor Adjunto PUCV.
- ▶ **Mario Guzmán**, Magister en Estadística PUCV, Data Scientist y Profesor Agregado PUCV.
- ▶ **Patricio Videla**, Profesor Auxiliar PUCV y Jefe de Docencia del Instituto de Estadística.

HORARIO

- ▶ Lunes y Miércoles de 19:00 a 22:00 hrs.

CONTACTO

- ▶ diplomado.estadistica@pucv.cl

- ▶ El programa considera 96 hrs. Cronológicas.
- ▶ El programa requiere un número mínimo de participantes para dictarse.

CLASES

Octubre 2026	5	7	14	19	21	26	28		
Noviembre 2026	2	4	9	11	16	18	23	25	
Diciembre 2026	2	7	9	14	16	21	23	28	30
Enero 2027	4	6	11	13	18	20	25	27	

TEMARIO

TEMAS BÁSICOS

- 1. ESTADÍSTICA DESCRIPTIVA**
 - a. Estadísticas descriptivas y su visualización.
 - b. Tipos de variables en los datos.
- 2. TOMA DE DECISIÓN EN UN ENTORNO ALEATORIO**
 - a. Test estadístico.
 - b. Intervalos de confianza para pronósticos.
- 3. ANÁLISIS DE ASOCIACIÓN DE VARIABLES**
 - a. Estrategias para medir la correlación entre variables: Pearson, Spearman o Kendall? b. Modelos lineales simples:
 - b. Estimación MCO, Diagnóstico de bondad. Test de normalidad.
 - c. One way ANOVA y two way ANOVA, razón de correlación.
 - d. Variables categóricas en el modelo de regresión lineal.
- 4. MÉTODOS MULTIVARIADOS EN ESTADÍSTICA**
 - a. Análisis por componentes principales (ACP).

TEMAS AVANZADOS

- 1. MODELOS LINEALES MÚLTIPLES**
 - a. Estimación MCO, diagnóstico de bondad (t-test, test de Fisher) y tipos de predicción (individual y del fenómeno estudiado).
 - b. Identificación de pertinentes (Criterios de información, algoritmos de selección forward, stepwise y backward).
 - c. Problema de colinealidad y soluciones (regresión PCR, PLS, regresión Ridge, LASSO, LASSO por grupo).
 - d. Datos outliers (atípicos detección y diagnóstico (leverages, residuos studentizados, distancia de Cook). Solución con la estimación M.
 - e. Heteroscedasticidad y autocorrelación: diagnóstico (test de Durbin Watson, tests de Breusch-Pagan) y estimación MCG. Modelos AR, MA y ARMA ajustados a residuos.
 - f. Una extensión del modelo lineal al machine learning: la regresión SVM.
- 2. MÉTODOS NUMÉRICOS DE ALTO NIVEL COMPUTACIONAL**
 - a. Introducción a EC2 de AWS.
 - b. Métodos bootstrap: paramétrico, no paramétrico y aplicados a modelos lineales.
- 3. MODELOS PARA DATOS TEMPORALES**
 - a. Identificación: Autocorrelaciones (ACF), Autocorrelaciones parciales (PACF), Criterios de información.
 - b. Modelamiento univariado de datos temporales con modelos SARIMA.
 - c. Diagnóstico y predicción.
 - d. Regresión falsa y test de raíz unitaria.
- 4. INTRODUCCIÓN A SQL**
 - a. Comandos SQL y tipos de datos.
 - b. Modelos relacionales.
 - c. Rutinas de comandos en SQL Server.
 - d. Depuración de datos para resolución de problemas.
 - e. Conexión a SQL Server.
- 5. INTRODUCCIÓN A SPARK**
 - a. Tratamiento de data frame.
 - b. Análisis descriptivo.
 - c. Categorización de bases.
 - d. Rutinas de Pyspark.
- 6. ALGORITMO DE K-MEDIAS**
 - a. Medidas de similitudes.
 - b. Algoritmo K-medias.
 - c. Clustering Jerárquico.
 - d. Métricas de validación.
 - e. Aplicaciones.
- 7. RANDOM FOREST Y ALGORITMO XGBOOST**
 - a. Introducción al Random Forest y XGBoost. Comparación de los dos enfoques.
 - b. Entrenamiento
 - c. Evaluación de out-of-bag error.
 - d. Evaluación del rendimiento
 - e. Aplicaciones.
- 8. MODELO DE REGRESIÓN LOGÍSTICA**
 - a. Presentación del modelo e interpretación.
 - b. Ajuste del Modelo e interpretación de resultados.
 - c. Estudio de caso real en Python: evaluación y construcción.
- 9. MÁQUINAS DE VECTORES DE SOPORTE**
 - a. Definición de hiperplano de separación.
 - b. Clasificador de margen máximo.
 - c. SVM para clasificador linealmente separable.
 - d. SVM para clasificador linealmente no separable.
 - e. Extensión de las máquinas de vectores de soporte.
 - f. Métricas de validación.
 - g. Aplicaciones.
- 10. REDES NEURONALES**
 - a. Arquitectura de una red.
 - b. Perceptrón.
 - c. Función de activación.
 - d. Back-propagation.
 - e. Métricas de validación.
 - f. Aplicaciones.
- 11. TEXT MINING**
 - a. Homologación de textos en base a cercanía de textos.
 - b. Arquitectura del web scraping.
 - c. Aplicaciones de web scraping y cercanía de textos en Python.
- 12. MANEJO DE HERRAMIENTAS DE AWS**
 - a. Introducción a S3.
 - b. Gestión de permisos con IAM.
 - c. Redes virtuales en la nube VPC.
 - d. Introducción a SageMaker.
 - e. Rutinas de modelos de ML en SageMaker con Python.
- 13. SISTEMAS DE RECOMENDACIÓN**
 - a. Filtros colaborativos.
 - b. Sistema basado en usuarios e ítems.
 - c. Aplicaciones de sistemas de recomendación.
- 14. DEEP LEARNING**
 - a. Introducción al Deep Learning.
 - b. Redes convolucionales (CNN).
 - c. Arquitectura Alexnet.
 - d. Aplicaciones de CNN con framework torch en Python.
- 15. INTELIGENCIA ARTIFICIAL PARA LENGUAJE NATURAL**
 - a. IA como modelos generativos.
 - b. LLM (Large Language Models) desde una perspectiva Estadística.
 - c. Características de un LLM: tamaño muestral, ventana de contexto, ingeniería de prompts.
 - d. Oportunidades, limitaciones y riesgos en el uso de LLM.
 - e. Caso de uso: usando un modelo de tipo ChatGPT.

* No se necesita conocimientos previos de los software dado que una introducción será hecha para cada software ocupado. Los códigos listos para el uso y comentados en la clase.

** Datos reales o simulados.