

DIPLOMADO DATA SCIENCE

30

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

- Jueves 17 de Diciembre 2026.
- 19:00 hrs.

INICIO DE CLASES

- Lunes 21 de Diciembre 2026.
- 19:00 hrs.

FECHA LÍMITE PARA MATRICULARSE

- Martes 15 de diciembre 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

CLASES

Diciembre 2026	21	28	30					
Enero 2027	4	6	11	13	18			
Febrero 2027	8	10	15	17	22	24		
Marzo 2027	1	3	8	10	15	17	22	24
Abril 2027	5	7	12	14	19	21	26	28

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

TEMARIO

TEMAS BÁSICOS

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

TEMAS AVANZADOS

- MODELOS LINEALES MÚLTIPLES**
 - Estimación MCO, diagnóstico de bondad (t-test, test de Fisher) y tipos de predicción (individuo real e hipotético).
 - Modelos lineales simples: Identificación de pertinentes (Criterios de información, algoritmos de selección forward, stepwise y backward).
 - Problema de colinealidad y soluciones (regresión PCR, PLS, regresión Ridge, LASSO, LASSO por grupo).
 - Datos outliers (atípicos detección y diagnóstico (leverages, residuos studentizados, distancia de Cook). Solución con la estimación robusta.
 - Heteroscedasticidad y autocorrelación: diagnóstico (test de Durbin Watson, tests de Breusch-Pagan) y estimación MCG. Modelos AR, MA y ARMA ajustados a residuos.
 - Una extensión del modelo lineal al machine learning: la regresión SVM.
- MÉTODOS NUMÉRICOS DE ALTO NIVEL COMPUTACIONAL**
 - Introducción a EC2 de AWS.
 - Métodos bootstrap: paramétrico, no paramétrico y aplicados a modelos lineales.
- MODELOS PARA DATOS TEMPORALES**
 - Test de Box e identificación con criterios de información.
- Modelamiento univariado de datos temporales con modelos SARIMA.**
- Diagnóstico y predicción.**
- Regresión falsa y test de raíz unitaria.**
- INTRODUCCIÓN A SQL**
 - Comandos SQL y tipos de datos.
 - Modelos relacionales.
 - Rutinas de comandos en SQL Server.
 - Depuración de datos para resolución de problemas.
 - Conexión a SQL Server.
- INTRODUCCIÓN A SPARK**
 - Tratamiento de data frame.
 - Análisis descriptivo.
 - Categorización de bases.
 - Rutinas de Pyspark.
- ALGORITMO DE K-MEDIAS**
 - Medidas de similitudes.
 - Algoritmo K-medias.
 - Clustering Jerárquico.
 - Métricas de validación.
 - Aplicaciones.
- RANDOM FOREST Y ALGORITMO XGBOOST**
 - Introducción al Random Forest y XGBoost. Comparación de los dos enfoques.
 - Entrenamiento
 - Evaluación de out-of-bag error.
 - Evaluación del rendimiento
 - Aplicaciones.

- MODELO DE REGRESIÓN LOGÍSTICA**
 - Presentación del modelo e interpretación.
 - Ajuste del Modelo e interpretación de resultados.
 - Estudio de caso real en Python: evaluación y construcción.
- MÁQUINAS DE VECTORES DE SOPORTE**
 - Definición de hiperplano de separación.
 - Clasificador de margen máximo.
 - SVM para clasificador linealmente separable.
 - SVM para clasificador linealmente no separable.
 - Extensión de las máquinas de vectores de soporte.
 - Métricas de validación.
 - Aplicaciones.
- REDES NEURONALES**
 - Arquitectura de una red.
 - Perceptrón.
 - Función de activación.
 - Back-propagation.
 - Métricas de validación.
 - Aplicaciones.
- TEXT MINING**
 - Homologación de textos en base a cercanía de textos.
 - Arquitectura del web scraping.
 - Aplicaciones de web scraping y cercanía de textos en Python.

- MANEJO DE HERRAMIENTAS DE AWS**
 - Introducción a S3.
 - Gestión de permisos con IAM.
 - Redes virtuales en la nube VPC.
 - Introducción a SageMaker.
 - Rutinas de modelos de ML en SageMaker con Python.
- SISTEMAS DE RECOMENDACIÓN**
 - Filtros colaborativos.
 - Sistema basado en usuarios e ítems.
 - Aplicaciones de sistemas de recomendación.
- DEEP LEARNING**
 - Introducción al Deep Learning.
 - Redes convolucionales (CNN).
 - Arquitectura Alexnet.
 - Aplicaciones de CNN con framework torch en Python.
- INTELIGENCIA ARTIFICIAL PARA LENGUAJE NATURAL**
 - IA como modelos generativos.
 - LLM (Large Language Models) desde una perspectiva Estadística.
 - Características de un LLM: tamaño muestral, ventana de contexto, ingeniería de prompts.
 - Oportunidades, limitaciones y riesgos en el uso de LLM.
 - 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.