



1821 Universidad de Buenos Aires

Resolución Consejo Directivo

Número: RESCD-2026-1433-E-UBA-DCT#FCEN

CIUDAD DE BUENOS AIRES
Lunes 27 de Julio de 2026

Referencia: EX-2026-03433016- -UBA-DMESA#FCEN - POSGRADO – Sesión
20/07/2026

VISTO

La nota presentada por la Dirección del Departamento de Física, mediante la cual eleva la información del curso de posgrado **Laboratorio Avanzado de Datos** (DOC8801018) para el año 2026,

CONSIDERANDO

lo actuado por la Comisión de Doctorado,

lo actuado por este Cuerpo en la sesión realizada el día 20 de julio de 2026,

en uso de las atribuciones que le confiere el Artículo 113° del Estatuto Universitario,

**EL CONSEJO DIRECTIVO DE LA FACULTAD
DE CIENCIAS EXACTAS Y NATURALES**

R E S U E L V E

ARTÍCULO 1°: Aprobar el dictado del curso de posgrado **Laboratorio Avanzado de Datos** (DOC8801018) de 108 horas y 16 semanas de duración, que será dictado por el Dr. Enzo Rodolfo Tagliazucchi.

ARTÍCULO 2°: Aprobar el programa del curso de posgrado **Laboratorio Avanzado de Datos** (DOC8801018) que como anexo forma parte de la presente Resolución, para su dictado durante el segundo cuatrimestre de 2026.

ARTÍCULO 3°: Aprobar un puntaje máximo de cinco (5) puntos para la Carrera de Doctorado.

ARTÍCULO 4°: Establecer un arancel de **CATEGORÍA NULA**.

ARTÍCULO 5°: Disponer que, de no mediar modificaciones en el programa, la carga horaria y el arancel, el presente Curso de Posgrado tendrá una vigencia de cinco (5) años a partir de la fecha de la presente Resolución.

ARTÍCULO 6°: Comuníquese a todos los Departamentos Docentes, a la Dirección de Estudiantes y Graduados, a la Biblioteca de la FCEyN y a la Secretaría de Posgrado con copia del programa incluida. Cumplido, pase a FISICA#FCEN y resérvese.

ANEXO

Laboratorio Avanzado de Datos (DOC8801018)

PROGRAMA

Objetivos:

El objetivo general es proporcionar conocimientos teóricos y prácticos sobre el análisis de datos, con aplicaciones a física experimental, computacional, y problemas interdisciplinarios.

La asignatura tiene como objetivo familiarizar a los estudiantes en el ciclo necesario para responder una pregunta basándose en el análisis de datos:

- Identificar si un problema puede ser resuelto mediante el análisis de datos disponibles
- Implementar herramientas para obtener estos datos, procesarlos y estandarizarlos
- Desarrollar herramientas estadísticas y de aprendizaje automático para modelar la relación entre variables
- Identificar la evidencia disponible en los datos y elaborar informes escritos y orales que sintetizen los resultados del análisis

Programa analítico:

- *Primera parte:* análisis exploratorio de datos numéricos.

Introducción a la ciencia de datos. Consideraciones prácticas y éticas. Repaso Python, Google Colab, herramientas para la materia. Distintos tipos de datos en ciencia de datos.

Procesamiento y preparación de datos numéricos. Estandarización, normalización, detección y remoción de outliers, datos faltantes. Estadística descriptiva. Medias, medianas, cuartiles, desvío estándar, error estándar. Distribuciones. Rudimentos de probabilidad.

Visualización de datos. Tipos de plots básicos y cuándo usarlos: scatter, barras, boxplots, violinplots, tortas, y otros.

- *Segunda parte:* modelos de regresión.

Regresión lineal, cuadrados mínimos. Introducción a scikit-learn. Regresión lineal, cuadrados mínimos. Regresión lineal, polinomios. Concepto de overfiteado. Regresión logística. Clasificación vs. regresión. Regresión lineal y logística como algoritmos de ML. Conceptos básicos de ML: función de costo, optimización, gradient descent

- *Tercera parte:* modelos de clasificación.

Test set-train set. Cross-validation. Métricas para la medición de performance en regresión y clasificación. Label shuffling. Clasificador lineal. Clasificación basada en instancias (KNN) Árboles de decisión. Ensembles de árboles aleatorios (random forest) Support vector machines. Kernels.

- *Cuarta parte:* clustering y reducción de la dimensionalidad.

Clustering de datos y reducción de la dimensionalidad (ejemplo y análisis: PCA)
Clustering de datos y reducción de la dimensionalidad (ejemplo y análisis: kmeans, kmeans jerárquico)

- *Quinta parte:* obtención de datos, datos no numéricos.

Uso de APIs, obtención de keys, requests, distintos ejemplos Scrapeo de páginas web. Preprocesado y limpieza de datos no numéricos, principalmente texto. Preprocesado de datos de texto (segunda parte): matriz de frecuencia, TF-IDF, nube de palabras, análisis de emotividad.

- *Sexta parte:* introducción al procesamiento del lenguaje natural.

Clasificación de documentos (KNN, Bayes) Detección de tópicos (LSA, NMF, LDA) Ejemplos de adquisición y visualización de otros tipos de datos (datos geoespaciales, visualización).

BIBLIOGRAFIA

- VanderPlas, J. (2016). Python data science handbook: Essential tools for working with data.” O’Reilly Media, Inc.”.
- Wilke, C. O. (2019). Fundamentals of data visualization: a primer on making informative and compelling figures. O’Reilly Media.
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An introduction to statistical learning (Vol. 112, p. 18). New York: springer.
- Kelleher, J. D., & Tierney, B. (2018). Data science. MIT Press.
- Hastie, T., Tibshirani, R., Friedman, J. H., & Friedman, J. H. (2009). The elements of statistical learning: data mining, inference, and prediction (Vol. 2, pp. 1-758). New York: springer.
- Bishop, C. M., & Nasrabadi, N. M. (2006). Pattern recognition and machine learning (Vol. 4, No. 4, p. 738). New York: springer.

Digitally signed by MARTI Marcelo Adrian
Date: 2026.07.24 13:36:21 ART
Location: Ciudad Autónoma de Buenos Aires

Marcelo Marti
Secretario
Secretaría de Posgrado
Facultad de Ciencias Exactas y Naturales

Digitally signed by DURAN Guillermo Alfredo
Date: 2026.07.27 14:06:38 ART
Location: Ciudad Autónoma de Buenos Aires

Guillermo Alfredo Duran
Decano
Decanato
Facultad de Ciencias Exactas y Naturales