



1821 Universidad de Buenos Aires

Resolución Consejo Directivo

Número: RESCD-2026-1436-E-UBA-DCT#FCEN

CIUDAD DE BUENOS AIRES
Lunes 27 de Julio de 2026

Referencia: EX-2026-03594467- -UBA-DMESA#FCEN - POSGRADO - Sesión
20/07/2026

VISTO

La nota presentada por la Dirección del Departamento de Ecología, Genética y Evolución, mediante la cual eleva la información del curso de posgrado **Selección de Variables y Modelos Predictivos: Integración de Enfoques Estadísticos y de Aprendizaje Automático con Aplicaciones en Ciencias de la Vida** 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**

RESUELVE

ARTÍCULO 1º: Aprobar el nuevo curso de posgrado **Selección de Variables y Modelos Predictivos: Integración de Enfoques Estadísticos y de Aprendizaje Automático con Aplicaciones en Ciencias de la Vida** de 40 horas y 1 semana de duración, que será dictado por la Mg. Adriana Pérez y el Dr. Gerardo Cueto, con la colaboración de la Dra. Cecilia Bruno.

ARTÍCULO 2º: Aprobar el programa del curso de posgrado **Selección de Variables y Modelos Predictivos: Integración de Enfoques Estadísticos y de Aprendizaje Automático con Aplicaciones en Ciencias de la Vida** que como anexo forma parte de la presente Resolución, para su dictado durante el tercer bimestre de 2026.

ARTÍCULO 3º: Aprobar un puntaje máximo de dos (2) puntos para la Carrera de Doctorado.

ARTÍCULO 4º: Establecer un arancel de **CATEGORÍA BAJA**, estableciendo que dicho arancel estará sujeto a los descuentos y exenciones estipulados mediante la Resolución CD N.º 1072/19. Disponer que los fondos recaudados ingresen en la cuenta presupuestaria habilitada para tal fin, y sean utilizados de acuerdo a la Resolución 072/0.

ARTÍCULO 5º: 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 6º: 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 ECOLOGIA#FCEN y resérvese.

ANEXO

Selección de Variables y Modelos Predictivos: Integración de Enfoques Estadísticos y de Aprendizaje Automático con Aplicaciones en Ciencias de la Vida

PROGRAMA

Fundamentación:

En el análisis de datos supervisado en ciencias de la vida es frecuente trabajar con un número elevado de variables explicativas, muchas de ellas altamente correlacionadas, irrelevantes o redundantes. Este contexto plantea desafíos tanto para la inferencia como para la predicción, particularmente en términos de sobreajuste, multicolinealidad y pérdida de capacidad predictiva en datos nuevos.

En los últimos años, se ha incrementado el uso de métodos de aprendizaje automático para la construcción de modelos predictivos. Sin embargo, la evidencia empírica muestra que el desempeño de estos modelos depende críticamente de la estrategia de selección de variables empleada y de su interacción con el tipo de modelo ajustado.

Una distinción central es la que separa los modelos orientados a la predicción de aquellos orientados a la inferencia causal. En datos observacionales, propios de gran parte de las ciencias de la vida, esta distinción tiene consecuencias directas sobre la estrategia de modelización, la selección de variables y la interpretación de los resultados.

Este curso propone un abordaje integrador que combina métodos clásicos de modelización estadística con técnicas modernas de aprendizaje automático supervisado, poniendo énfasis en el proceso completo de construcción de modelos predictivos: selección de variables, ajuste, validación y evaluación, para problemas tanto de regresión continua como de clasificación binaria. Se hará especial hincapié en la comparación de estrategias de selección de variables (*stepwise*, filtros, algoritmos evolutivos, Boruta) y su impacto diferencial en modelos lineales (regresión lineal y logística) y no lineales (*random forest*), con implementación en R.

Objetivo general:

Desarrollar competencias para la construcción, validación y comparación de modelos predictivos en contextos de alta dimensionalidad, integrando métodos estadísticos y de aprendizaje automático.

Objetivos específicos:

- Comprender el marco conceptual de la modelización predictiva y sus diferencias con la inferencia estadística clásica.
- Conocer y aplicar métodos de selección de variables, evaluando críticamente sus supuestos, limitaciones y comportamiento en distintos contextos.
- Ajustar e interpretar modelos predictivos lineales y no lineales para respuestas continuas y binarias.
- Evaluar el desempeño predictivo mediante estrategias de validación adecuadas y métricas apropiadas al tipo de respuesta.
- Comparar combinaciones de métodos de selección y tipos de modelo, y fundamentar la elección de una estrategia en función del problema y los datos.

Programa analítico

- Unidad 1: Introducción a la modelización predictiva.
 - Inferencia y predicción.
 - Predicción vs. causalidad en datos observacionales.
 - Tipos de problema supervisado: respuesta continua y respuesta binaria.
 - Error de predicción, compromiso sesgo-varianza.
 - Sobreajuste y generalización.
 - Alta dimensionalidad y colinealidad entre variables predictoras.
- Unidad 2: Selección de variables.
 - Métodos clásicos: *forward*, *backward*, *stepwise*.
 - Criterios de información (AIC).
 - Diagnóstico de multicolinealidad: factor de inflación de la varianza (VIF).
 - Estrategias combinadas (*Step* + VIF + valor p).
 - Métodos de regularización: Ridge, Lasso.
 - Métodos basados en aprendizaje automático: *random forest*, Boruta.
- Unidad 3: Modelos predictivos.
 - Modelos lineales.
 - Regresión lineal y logística.
 - Modelos no lineales.
 - Árboles de decisión: estructura, criterios de partición, sobreajuste.
 - *Bagging* y ensamblado.
 - Importancia de variables, *tuning*.
 - Comparación de modelos lineales vs. no lineales: supuestos, flexibilidad, interpretabilidad.
- Unidad 4: Validación y evaluación del desempeño predictivo.

- Estrategias de validación.
 - Separación entrenamiento/prueba.
 - Validación cruzada (*k-fold*, *k-fold* repetida).
 - *Bootstrap* para estimación del error de predicción.
 - Métricas: RMSE, MAE, *Accuracy*, sensibilidad y especificidad, Curvas ROC y AUC, etc.
- Unidad 5: Integración: selección de variables, modelos y estrategias.
 - Comparación de combinaciones: selección × modelo.
 - Impacto de la estrategia de selección en modelos lineales y no lineales.
 - Parsimonia vs complejidad en función del objetivo del análisis.
 - Interpretabilidad vs capacidad predictiva.
 - Discusión de casos aplicados.

Modalidad de dictado:

El curso se desarrollará íntegramente en modalidad presencial en laboratorio, con un enfoque teórico-práctico. Los distintos contenidos se introducirán a partir de casos de estudio provenientes de las ciencias de la vida, que servirán como eje para presentar y discutir los diferentes abordajes de modelización predictiva. En cada tema se analizarán las preguntas de investigación que pueden abordarse con cada metodología, los fundamentos conceptuales de los métodos, sus supuestos, las condiciones para una correcta aplicación y las principales limitaciones e implicancias para la interpretación de los resultados. Asimismo, se discutirán criterios para la selección de estrategias analíticas en función de las características de los datos y de los objetivos del estudio. Las actividades prácticas consistirán en la implementación de los métodos utilizando el software R, aplicados tanto a bases de datos provistas por los docentes como, cuando sea posible, a datos aportados por los participantes. Se promoverá el análisis crítico de los resultados obtenidos y la comparación de diferentes estrategias de selección de variables, algoritmos predictivos y procedimientos de validación. Se destinarán espacios para la discusión grupal de los proyectos de investigación de los participantes.

Modalidad de evaluación:

La evaluación final consistirá en el análisis de un conjunto de datos aplicando los métodos desarrollados durante el curso. Los participantes deberán presentar un informe breve y realizar una exposición oral en la que describan la estrategia analítica implementada, los resultados obtenidos y la justificación de las decisiones tomadas.

BIBLIOGRAFIA

- Gareth, J., Witten, D., Hastie, T., & Tibshirani, R. (2013). An Introduction to Statistical Learning. In An Introduction to Statistical Learning. Springer.
- Heinze G, Wallisch C, Dunkler D. (2018). Variable selection – A review and recommendations for the practicing statistician. Biometrical Journal, 60(3), 431–449.
- Suarez F., Bruno C., Giannini Kurina F., Giménez Pecci M.P., Rodríguez Pardina P., Balzarini M. (2023). Marriage between variable selection and prediction methods to model plant disease risk. European Journal of Agronomy (151).

Digitally signed by MARTI Marcelo Adrian
Date: 2026.07.24 13:35:42 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 14:07:34 ART
Location: Ciudad Autónoma de Buenos Aires

Guillermo Alfredo Duran
Decano
Decanato
Facultad de Ciencias Exactas y Naturales