

1 Información del equipo pedagógico y horario atención a estudiantes

Profesor: Ignacio Sarmiento-Barbieri (i.sarmiento@uniandes.edu.co)

- Horario Clase: Sábados 8:00 a.m. – 10:50 a.m.
- Web del Curso: Bloque Neón
- Horario de atención a estudiantes: hacer cita en [este link](#)

Profesor Complementario: Gustavo Adolfo Castillo Álvarez (ga.castillo@uniandes.edu.co)

- Horario Clase: Viernes, 6:00 p.m. – 7:20 p.m.
- Horario de atención a estudiantes: hacer cita en [este link](#)

2 Descripción del curso

Este es un curso con un enfoque especial en herramientas relevantes para economistas y ciencias sociales. Está destinado a estudiantes interesados en investigación aplicada y/o análisis de datos grandes y no estructurados. Problemas de predicción e inferencia, con especial énfasis en inferencia causal, atraviesan transversalmente al curso.

Mediante una combinación de contenido asincrónico, clases sincrónicas magistrales y complementarias, talleres grupales, y quices los estudiantes adquirirán las herramientas estadísticas y computacionales necesarias para responder varias preguntas en economía y en una gran cantidad de subcampos en investigación aplicada. Se hará énfasis especial en el análisis de datos reales, y la aplicación de metodologías específicas; ejemplos incluyen encuestas de hogares, precios de propiedades, datos de internet y redes sociales.

Prerrequisitos:

Microeconomía 3 y econometría 2 o microeconometría aplicada. Se recomienda haber cursado econometría avanzada del PEG. Se necesita experiencia básica en manejo de datos y en software “R” o “Python”. El curso se basará principalmente en “R”.

Los estudiantes que se inscriban en este curso virtual deben contar con una cámara web funcional y mantenerla encendida durante las sesiones. Asimismo, deben contar con un computador compatible con LockDown Browser y Respondus Monitor, ya que las evaluaciones se realizarán mediante este sistema de supervisión. La falta de disponibilidad de estos recursos no será una excusa válida. En caso de no tener disponible computador personal, recuerde que puede solicitar equipos en préstamo en las salas de computo de la universidad.

3 Resultados de aprendizaje

- Aplicar las técnicas provenientes de la ciencia de datos, la ciencia computacional, y la estadística usando la visión de la economía para resolver problemas puntuales identificados en el contexto.
- Contrastar los distintos algoritmos y su conveniencia para contestar preguntas económicas y sociales con base en criterios relacionados con la naturaleza de problemas económicos y sociales.
- Implementar procesos técnicos para el manejo cuantitativo de datos que surgen de distintas fuentes: páginas web, encuestas, geoespaciales, texto, etc, para resolver problemas económicos y sociales
- Generar conclusiones y recomendaciones sobre preguntas relevantes a las ciencias sociales por medio del manejo, análisis y síntesis de bases de datos con gran número de observaciones y variables.
- Aplicar el software R y su ecosistema para análisis estadístico, de big data y machine learning.

4 Cronograma

Duración	Tema	¿Qué estudiaremos?
~ 2 semanas	Introducción a big data y machine learning	Presentaremos los objetivos del curso y discutiremos los fundamentos del aprendizaje de máquinas desde una perspectiva predictiva. Exploraremos ejemplos motivadores y el flujo de trabajo en proyectos de ML aplicados a economía.
~ 2 semanas	Sobreajuste, validación cruzada y bootstrap. Adquisición de datos web	Estudiaremos cómo evaluar modelos mediante técnicas como validación cruzada y bootstrap. Veremos cómo prevenir el sobreajuste y exploraremos herramientas básicas para la adquisición de datos desde la web (scraping, APIs).

Continúa en la siguiente página

Duración	Tema	¿Qué estudiaremos?
~ 2 semanas	Selección de modelos y regularización	Abordaremos la comparación y selección de modelos predictivos, e introduciremos técnicas de regularización como Ridge y Lasso, útiles para mejorar la generalización y seleccionar variables.
~ 2 semanas	Clasificación	Analizaremos modelos de clasificación, comenzando con el modelo Logit y métricas específicas (precisión, recall, F1, AUC). Discutiremos el desbalance de clases y técnicas para enfrentarlo.
~ 2 semanas	Árboles, bosques y boosting	Estudiaremos métodos basados en árboles de decisión, Bagging y Bosques Aleatorios. Luego nos enfocaremos en Boosting y variantes como Gradient Boosting y XGBoost.
~ 2 semanas	Redes neuronales y aprendizaje profundo	Introduciremos redes neuronales de una y múltiples capas. Analizaremos técnicas de entrenamiento (backpropagation), regularización (dropout, batch norm) y su aplicación en problemas económicos.
~ 2 semanas	Datos espaciales, texto como datos y Super Learners	Exploraremos el análisis de datos espaciales y la validación cruzada espacial. Luego nos centraremos en el concepto de Super Learners como método de ensamblado.
~ 1 semana	Texto como Datos y Aprendizaje No Supervizado	En este módulo estudiaremos técnicas básicas de procesamiento de texto como datos y finalizaremos con técnicas de reducción de dimensión no supervisadas.

5 Referencias

- Békés, G., & Kézdi, G. (2021). Data analysis for business, economics, and policy. Cambridge University Press.
- Berk, R. A. (2008). Statistical learning from a regression perspective (Vol. 14). New York: Springer.
- Davidson, R., & MacKinnon, J. G. (2004). Econometric theory and methods
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). An introduction to statistical

learning (ISLR)

- 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.
- Chernozhukov, V., Hansen, C., Kallus, N., Spindler, M., & Syrgkanis, V. (2024). Applied causal inference powered by ML and AI. arXiv preprint arXiv:2403.02467.
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep learning. MIT press.

6 Metodología

La metodología del curso combina clases magistrales virtuales, sesiones complementarias virtuales, actividades asincrónicas, quices, talleres grupales y presentaciones.

El objetivo de las clases magistrales es presentar los principales conceptos, modelos y metodologías del aprendizaje de máquinas, tanto desde una perspectiva teórica como a través de aplicaciones prácticas que ilustran los conceptos. La presentación del contenido se realizará mediante diapositivas y cuadernos interactivos.

Las clases complementarias acompañarán el trabajo en los talleres y permitirán resolver inquietudes específicas sobre el material del curso. En estas sesiones se desarrollarán ejemplos adicionales, se profundizará en la implementación de algoritmos y se brindará espacio para la discusión y el trabajo colaborativo.

Se espera que los estudiantes completen las actividades asincrónicas antes de cada clase magistral, estudien las lecturas asignadas disponibles en la página del curso y repliquen las aplicaciones presentadas. Estas actividades están diseñadas para apoyar el aprendizaje progresivo y preparar a los estudiantes para las evaluaciones individuales y grupales.

Para un desarrollo exitoso del curso se espera que los estudiantes asistan a todas las clases sincrónicas con cámara encendida y participen activamente en las discusiones.

7 Evaluaciones

- Quices. Los estudiantes tendrán 8 quices individuales virtuales que evaluarán el aprendizaje individual. El peso de los quices será de 4%, con la nota del quiz más baja sera eliminada, para un peso total de 28%.
- Los estudiantes realizarán trabajos prácticos grupales. Habrá 3 talleres durante el cursado. Los talleres constarán de dos partes: presentaciones y repositorio público en GitHub. Cualquier cambio a la conformación de equipos deberá ser comunicada al profesor magistral y al profesor complementario con **72 horas** de antelación de la entrega.

Las presentaciones y el link al repositorio público en GitHub serán entregados vía Bloque Neón. Se espera que todos los miembros hagan contribuciones al repositorio del taller y ser

completamente reproducible. La calificación del taller se verá reducida si no hay evidencia de contribución de todos los miembros.

Cada taller tendrá un peso total de 24%. De ese total, el repositorio representará un 10%, mientras que la presentación oral tendrá un peso del 14%. La presentación se realizará durante la clase magistral correspondiente y será parte de la discusión colectiva de los resultados del taller.

Table 1: Puntajes

	Puntaje Individual	Puntaje Total	Fecha <i>tentativa</i>
Quices		28%	
Quiz 0*	0%		Febrero 2, 2026
Quiz 1	4%		Febrero 2, 2026
Quiz 2	4%		Febrero 16, 2026
Quiz 3	4%		Marzo 2, 2026
Quiz 4	4%		Marzo 16, 2026
Quiz 5	4%		Abril 13, 2026
Quiz 6	4%		Abril 27, 2026
Quiz 7	4%		Mayo 11, 2026
Quiz 8	4%		Mayo 25, 2026
Talleres		72%	
Taller 1	24%		Febrero 21, 2026
Taller 2	24%		Abril 25, 2026
Taller 3	24%		Mayo 16, 2026

Nota: * *Opcional*

Nota bene: El objetivo de permitir el trabajo en equipos es fomentar la discusión y colaboración en el proceso de aprendizaje, no simplemente dividir las tareas. Se espera que cada miembro del equipo contribuya de manera significativa en cada parte de cada taller. Además, **cada estudiante es responsable de todo el contenido del taller** independientemente de cómo se organicen para trabajar en equipo.

Sistema de aproximación de notas definitiva

Las calificaciones definitivas de las materias serán numéricas de uno cinco (1,50) a cinco (5,00), en unidades, décimas y centésimas. La calificación aprobatoria mínima será de tres (3,0). En este curso se aproximará la nota a la centésima más cercana. Por ejemplo, si el cálculo del cómputo es 3.245, la nota final se aproximará a 3.25; si el resultado del cálculo es 2.994 la nota final será de 2.99

8 Asistencias

Se espera que los estudiantes asistan a todas las clases. Si los asistentes faltan a más del 20% sin excusa válida de las clases sincrónicas en las que nos reuniremos se penalizará hasta un 10% la nota final del curso.

9 Políticas generales de los cursos de Economía y fechas importantes

Los estudiantes deben consultar [este enlace](#), donde se encuentran las reglas sobre asistencia a clase, excusas válidas, fraude académico y faltas disciplinarias, reclamos, políticas de bienestar y fechas importantes del semestre.