

Ciencia de Datos para la toma de decisiones en Economía

Ignacio Sarmiento-Barbieri¹ Carlos Yanes-Guerra²

Curso en Maestría en Economía, 2022

E-mail (1): i.sarmiento@uniandes.edu.co

Website: <https://www.uninorte.edu.co>

E-mail (2): cayanes@uninorte.edu.co

Atención oficina: M 2:30-3:30pm

Días de clase: Viernes - Sábado

Oficina: D-215

Salón: LAB D1

Laboratorio: Economía Bloque D

Horas: V 2:30-6:25pm

Descripción del curso

El curso establecido para el **seminario de datos** de la Universidad del Norte, pretende brindar los lineamientos y las estrategias metodológicas para aquellas personas que tienen un alto interés por el análisis cuantitativo y desean tener un amplio espectro de metodologías para abordar un problema o pregunta de investigación.

El curso hace énfasis en la **aplicación** metodológica de las técnicas estadísticas dentro de la minería y análisis de datos. Además se resalta el desarrollo gráfico, el uso de las métricas de la estadística como a su vez de la parte inferencial. Todo lo anterior, con el objeto de contrastar ciertas hipótesis o preguntas y problemas cotidianos en que se envuelven los agentes económicos como los consumidores, las empresas y el mismo sector público.

Materiales requeridos

- Instalación del programa R y de Rstudio.
- Notas de clases del curso.
- Microsoft Excel, bloc de notas, entre otros.

Prerrequisitos

Aunque el curso será lo más autónomo posible, los estudiantes deberán estar familiarizados con conceptos de álgebra, cálculo y estadística. Se requiere tener cierta familiaridad con algún lenguaje de programación.

Objetivos del curso

El curso intenta:

1. Enseñar de forma practica el alcance de las metodologías para el análisis de datos para la comprensión sofisticada de los conceptos y métodos utilizados en el mercado; intentando conocer los alcances exactos y las posibles limitaciones de cada uno de los métodos.
2. Mostrar como recopilar datos relevantes, realizar el análisis de datos correspondiente utilizando métodos científicos y establecer conexiones adecuadas y potentes entre el análisis cuantitativo y los problemas del mundo real.
3. Mejorar la forma de realizar informes de análisis estructurado y lograr que este bien organizado para un estudio de caso del mundo real, incluyendo detalles reflexivos y convincentes.

Estructura del curso

Estructura de la clase

Las clases se imparten de forma magistral y con computadores a la mano. Se intenta que el estudiante una vez haya adquirido los conceptos claves dirija su atención a un caso de ejemplo y logre estipular la aplicación e interpretación de las mediciones requeridas.

Los temas de la clase incluyen tópicos en probabilidad, estadística, pruebas de hipótesis, regresión, la agrupación, árboles de decisión y pronósticos.

Contenido general

Los temas a abordar son los siguientes:

Semana	Fecha	Tema	Horas de Trabajo
1		Introducción al Aprendizaje Estadístico: Causalidad y predicción. Manipulación de datos en R	4
2		Regresión lineal: MCO. Propiedades numéricas. Sobreajuste. Métodos de remuestreo y validacion cruzada. Modelos lineales, linealizables, y no lineales.	5
3		Datos espaciales. Selección de modelos y regularización. Lasso y Ridge. Aplicaciones en inferencia causal.	9
4		Introducción a los Modelos de clasificación. Matriz de confusión, criterios de evaluación, ROC y AUC	9
5		Clusters y arboles de decisión. Bagging. Random forest	9
6		Trabajo final	4

Tareas y requerimientos

En esta parte, se definirán un paquete de lecturas para complementar lo aprendido en clases y la solicitud del desarrollo de los prácticos de la clase.

Textos de referencia

El curso no exige un solo libro para el desarrollo de este, sin embargo como forma opcional se puede optar por leer:

- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). An introduction to statistical learning
- Friedman, J., Hastie, T., & Tibshirani, R.(2001). The elements of statistical learning (Vol. 1, No. 10). New York: Springer series in statistics.
- Taddy, M. (2019). Business data science: Combining machine learning and economics to optimize, automate, and accelerate business decisions. McGraw Hill Professional.
- Ott, R. L., & Longnecker, M. T. (2015). An introduction to statistical methods and data analysis. Cengage Learning.
- Cochardt, A. (2019). Scott E. Page: The model thinker—what you need to know to make data work for you.
- Jank, Wolfgang. Business analytics for managers. Springer Science & Business Media, 2011.

Recursos online y Libros en línea (Bookdown)

- <https://education.rstudio.com/learn/>
- <https://bookdown.org/ndphillips/YaRrr/>

Laboratorios

Los laboratorios y/o prácticos deben ser subidos a la plataforma del curso para su respectiva evaluación en las fechas establecidas por el profesor.

Examen y proyecto final

Al finalizar el curso, el estudiante debe tener la capacidad de aplicar lo aprendido en un proyecto propio que cuente con al menos una de las metodologías y/o temáticas aquí abordadas.

Notas y calificaciones

La Universidad del Norte en sus normas estima que para la aprobación del curso el estudiante debe haber aprobado con nota mínima de 3.0. Sin embargo las actividades para aprobar el curso se explican a continuación:

- 40% De cada practico desarrollado del curso (20% cada uno).

- 15% Participación activa.
- 15% Propuesta de trabajo final.
- 30% Trabajo final escrito.

Las notas de cada actividad estará en la escala de $[0 \leq Nota \leq 5]$.

Políticas del curso

Durante las Clases

La asistencia es requerida. Aquel(los) estudiante(s) que por algún motivo deba(n) ausentarse, deberá(n) justificarlo mediante correo electrónico adjuntando la respectiva excusa del caso y debe(n) ponerse al día con lo visto. Los celulares solo están permitidos en modo silencioso y sin forma de interrumpir el desarrollo de la clase.

Reclamos

Todo estudiante que desee formular un reclamo sobre las calificaciones de cualquier evaluación, debería dirigirlo mediante escrito y/u oral, debidamente sustentado dentro de la socialización de las Notas en que se dan a conocer las calificaciones en cuestión. Al cabo de la consulta y sesión se dará respuesta al reclamo formulado y se informará al estudiante la decisión correspondiente, después de esto no habrá mas correcciones sin excepción. Si el estudiante considera que la decisión no corresponde a los criterios de evaluación, podrá solicitar la designación de un segundo calificador, mediante escrito debidamente sustentado, dirigido al Consejo de Facultad dentro de los ocho (8) días hábiles siguientes al conocimiento de la decisión.