

FACULTAD DE INGENIERÍA
MAGÍSTER EN DATA SCIENCE

A. Antecedentes Generales.

- | | |
|--|---|
| - Nombre de la asignatura | : Análisis de datos |
| - Tipo de asignatura (obligatoria, electiva, optativa) | : Obligatoria |
| - Pre – requisitos | : Python para data science y Estadísticas para data science |
| - Créditos | : 3 |
| - Ubicación dentro del plan de estudio (período académ.) | : Segundo trimestre |
| - Horas académicas directas teóricas | : 28 |
| - Horas académicas directas prácticas | : 0 |
| - Horas académicas de trabajo autónomo | : 84 |

B. Descripción de la asignatura.

- Esta asignatura presentará los conceptos básicos de pre-procesamiento y análisis descriptivo de datos. El objetivo principal es poder determinar cuáles datos son susceptibles de ser convertidos en información para apoyar la toma de decisiones, y separar el ruido de la señal. Es el primer paso en un proyecto de ciencia de datos. Los estudiantes aprenderán a identificar las problemáticas que presentan los datos desde el momento de su registro (por ej., error muestral, outliers), así como usar las herramientas necesarias para describirlos (por ej., distribuciones e histogramas), explorarlos (por ej., agrupar o filtrar bajo un criterio específico), y cruzarlos (por ej., utilizando otras fuentes). Asimismo, los estudiantes comprenderán que las etapas de este proceso no son lineales, sino que se benefician del diseño iterativo.
- Esta asignatura se enmarca en la línea de desarrollo de data science.
- Esta asignatura tributa, a través de sus resultados de aprendizaje, a las siguientes competencias del perfil de egreso del Magíster en Data Science:
 - Aplicar teorías, algoritmos, métodos, técnicas y herramientas básicas y avanzadas de Data Science para analizar, resolver y hacer una evaluación crítica de desafíos complejos e interdisciplinarios, utilizando datos internos y externos de las organizaciones.
 - Comunica efectivamente y argumenta sobre los resultados de su trabajo a públicos especializados y no especializados, de forma oral, escrita y visual, utilizando distintos medios y soportes.
 - Demuestra responsabilidad y comportamiento ético, cumpliendo los protocolos y normas que guían su desempeño, en las iniciativas de Data science.
 - Demuestra capacidad de aprendizaje continuo, mediante la aplicación de estrategias para utilizar nuevo conocimiento en data science en su ámbito de desempeño.

C. Objetivos de la asignatura (resultados de aprendizaje).

- Identificar las ventajas y desventajas de las herramientas computacionales utilizadas para el análisis de datos, utilizando lenguaje técnico afín.
- Recopilar y limpiar datos, en base a una propuesta de replicabilidad del proceso.
- Transformar y analizar datos, realizando preguntas clave para resolver problemas a partir del contexto en que se desarrollan.
- Modelar datos para extraer información y generar conclusiones basadas en evidencia.
- Identificar las buenas prácticas en el modelamiento de datos.

D. Contenidos.

1. Limpieza y estructura de datos.
 - a. Formateo de datos
 - b. Transformación de datos
 - c. ETL
2. Regresión y predicción.
 - a. Regresión lineal múltiple.
 - b. Predicción usando regresión y los peligros de la extrapolación.
 - c. Factores y variables categóricas en una regresión.
 - d. Multicolinealidad, variables de confusión e interacciones.
 - e. Diagnóstico de una regresión y supuestos (outliers, heterocedasticidad, no-normalidad, errores correlacionados y no-linealidad)
 - f. Sesgos en los análisis: Paradoja de Simpson, Paradoja de Berkson y Collider Bias.
3. Series de tiempo
 - a. Búsqueda y reorganización de datos de series de tiempo
 - b. Análisis de datos exploratorios para series temporales
 - i. Histogramas, gráfico de dispersión y métodos exploratorios para series de tiempo
 - ii. Estacionariedad y raíz unitaria
 - iii. Autocorrelación y correlaciones espurias en series de tiempo
 - c. Modelos estadísticos para series de tiempo
 - i. ¿Por qué no utilizar una regresión lineal?
 - ii. Modelos autorregresivos (AR), ARIMA y Autorregresión vectorial (VAR)
 - iii. Ventajas y desventajas de los métodos estadísticos para series de tiempo

E. Metodología de Enseñanza- aprendizaje.

Para asegurar que los estudiantes alcancen los resultados de aprendizaje propuestos, las estrategias seleccionadas son:

- Clases expositivo-participativas.
- Talleres prácticos de análisis de casos (uso de lenguaje de programación)

F. Evaluación.

El logro de los resultados de aprendizaje se evaluará de la siguiente forma:

- Para las diferentes instancias evaluativas se contará con una pauta de corrección con criterios claros y conocidos por los alumnos. El curso podrá ser evaluado con tareas, proyecto, certámenes y/o un examen final.

G. Bibliografía.

- Bibliografía Obligatoria.
 - Peter Bruce and Andrew Bruce. Practical Statistics for Data Scientists: 50 Essential Concepts 1st Edition. O'Reilly Media; 1 edition (May 28, 2017). ISBN-10: 1491952962. ISBN-13: 978-1491952962. Disponible en: [https://www.researchgate.net/profile/Janine-Zitianellis/post/Can anyone please suggest a books on machine learning using R Programming/attachment/613a5b83647f3906fc975a71/AS%3A1066204907204608%401631214467436/download/Practical+Statistics+for+Data+Scientists+50%2B+Essential+Concepts+Using+R+and+Python+by+Peter+Bruce%2C+Andrew+Bruce%2C+Peter+Gedeck.pdf](https://www.researchgate.net/profile/Janine-Zitianellis/post/Can_anyone_please_suggest_a_books_on_machine_learning_using_R_Programming/attachment/613a5b83647f3906fc975a71/AS%3A1066204907204608%401631214467436/download/Practical+Statistics+for+Data+Scientists+50%2B+Essential+Concepts+Using+R+and+Python+by+Peter+Bruce%2C+Andrew+Bruce%2C+Peter+Gedeck.pdf)
 - Aileen Nielsen. Practical Time Series Analysis. O'Reilly Media Inc.; (Oct 2019). ISBN: 9781492041658.
- Bibliografía Recomendada.
 - Cathy O'Neil, Rachel Schutt (2013). *Doing Data Science*. O'Reilly Media, Inc.
 - Grus, J. (2015). *Data Science from Scratch*. Chapter 10 and 19. O'Reilly Media, Inc.
 - Howard J. Seltman (2015). *Experimental Design and Analysis*. Available at <http://www.stat.cmu.edu/~hseltman/309/Book/Book.pdf>