

Curso Académico: (2022 / 2023)

Fecha de revisión: 16-05-2022

Departamento asignado a la asignatura: Departamento de Estadística

Coordinador/a: MINGUEZ SOLANA, ROBERTO

Tipo: Formación Básica Créditos ECTS : 6.0

Curso : 1 Cuatrimestre : 1

Rama de Conocimiento: Ingeniería y Arquitectura

OBJETIVOS

A la finalización del curso los estudiantes serán capaces de:

Conocer la importancia de la ciencia de datos en la sociedad del conocimiento actual.

Utilizar técnicas de visualización de datos para entender el problema al que se enfrentan un científico de datos, así como para presentar los resultados obtenidos.

Conocer cuando utilizar una técnica de análisis de datos supervisada y cuando una no supervisada.

Conocer algunas de las principales técnicas del análisis de datos, así como aplicaciones donde han sido utilizadas exitosamente.

Conocer los principales problemas con los que se puede encontrar un científico de datos y como enfrentarlos.

Conocer cuáles son las diferentes herramientas que utiliza un científico de datos.

Ser capaces de realizar sus primeros análisis de datos empleando R-Studio.

DESCRIPCIÓN DE CONTENIDOS: PROGRAMA

1. Importancia de la Ciencia de Datos
2. Introducción a R-Studio
3. Entendiendo los datos: Casos de estudio de análisis exploratorio de datos y técnicas de visualización I
4. Entendiendo los datos: Casos de estudio de análisis exploratorio de datos y técnicas de visualización II
5. Importancia de un buen diseño de experimentos y selección de las medidas de rendimiento: precisión, sensibilidad, especificidad. Sobreajuste
6. Introducción a la clasificación supervisada: casos de estudio sobre árboles de decisión y bosques aleatorios
7. Introducción a las técnicas no supervisadas: casos de estudio de técnicas de agrupación

ACTIVIDADES FORMATIVAS, METODOLOGÍA A UTILIZAR Y RÉGIMEN DE TUTORÍAS

La asignatura se imparte en 14 clases teórico-prácticas y 14 clases prácticas.

La asignatura es eminentemente práctica, y por ese motivo en las clases magistrales se explicarán los conceptos principales teóricos de la asignatura, pero también se pondrán en práctica con ejercicios con ordenador. Estos conceptos se profundizarán en las clases prácticas en las que se realizarán análisis adicionales de datos con ordenador.

Los estudiantes dispondrán además de tutorías donde pondrán resolver las dudas que tengan, tanto de las clases teóricas, prácticas o sobre los trabajos que tienen que entregar.

SISTEMA DE EVALUACIÓN

La nota final de la asignatura se obtendrá promediando la nota de la evaluación continua (40%) y la nota de examen final (60%).

La evaluación continua consistirá en la realización de dos trabajos en los que el estudiante deberá aplicar los diversos contenidos aprendidos en el curso. Estos trabajos consistirán de un análisis exploratorio de datos (15%) y de un análisis de datos desde un punto de vista de clasificación (25%). Se podrán realizar en parejas y deberán presentarse oralmente ante la clase.

Para superar la asignatura se deberá haber obtenido un 5 sobre 10 en el examen final.

Peso porcentual del Examen Final:	60
Peso porcentual del resto de la evaluación:	40

BIBLIOGRAFÍA BÁSICA

- PATHAK, Manas A. "Beginning Data Science with R", Springer, 2014

BIBLIOGRAFÍA COMPLEMENTARIA

- Bruce, P. C. & Bruce, A. Practical statistics for data scientists: 50 essential concepts., O'Reilly, 2017

- Irizarry, R. A. Introduction to data science: data analysis and prediction algorithms with R., CRC Press, 2020

- Peng, R. D. R programming for data science., Leanpub, 2016