

Curso Académico: ( 2023 / 2024 )

Fecha de revisión: 28-04-2023

Departamento asignado a la asignatura: Departamento de Estadística

Coordinador/a: MUÑOZ GARCIA, ALBERTO

Tipo: Optativa Créditos ECTS : 6.0

Curso : 4 Cuatrimestre :

**REQUISITOS (ASIGNATURAS O MATERIAS CUYO CONOCIMIENTO SE PRESUPONE)**

Métodos de Regresión y Análisis Multivariantes, ambas de tercer curso. Conocimiento del software estadístico R.

**OBJETIVOS**

1. Conocer y utilizar técnicas estadísticas avanzadas, con soporte informático de última generación.
2. Extraer y analizar información de grandes conjuntos de datos.
1. Capacidad de análisis y síntesis de la información en problemas de Data Mining.
2. Modelización y resolución de problemas prácticos.
3. Comunicación oral y escrita.

**DESCRIPCIÓN DE CONTENIDOS: PROGRAMA**

1. Introducción a Tidyverse
  - 1.1 Data wrangling
  - 2.2 Visualización de datos: ggplot2
  - 2.3 Agrupamiento e informes de datos.
2. Introducción al text mining.
  - 2.1 Principales conceptos.
  - 2.2 Nubes de palabras.
  - 2.3 Matriz de textos por documentos.
  - 2.4 Implementación en R y aplicaciones.
3. Visualización de datos: Escalamiento Multidimensional métrico, y biplots.
  - 3.1 Escalado multidimensional métrico.
  - 3.2 Biplots.
  - 3.2 Mapas perceptuales.
4. Análisis de cluster. Métodos jerárquicos, no jerárquicos (k-medias).
  - 4.1 Algoritmos de cluster jerárquicos bottom-up.
  - 4.2 k-medias y variantes.
5. Teoría de la información y árboles de clasificación.
  - 6.1 Nociones de teoría de la información.
  - 6.2 Teoría de los árboles de clasificación.
  - 6.3 Ejemplo real: credit scoring.
  - 6.4 Caso de estudio
6. Reglas de asociación.
  - 7.1 Conceptos básicos y algoritmos.
  - 7.2 Ejemplo completo con implementación en R.
  - 7.3 Caso de estudio.
7. Deep Learning.
  - 7.1 Support Vector Machines.
  - 7.2 Redes neuronales para clasificación.
  - 7.3 Redes neuronales para regresión.

8. Resolución de casos de estudio reales.

8.1 Prácticas con casos de estudio comprensivos de todas las técnicas estudiadas.

#### ACTIVIDADES FORMATIVAS, METODOLOGÍA A UTILIZAR Y RÉGIMEN DE TUTORÍAS

Teoría (4 ECTS). Clases teóricas con material de apoyo disponible en la Web. Prácticas (2 ECTS). Clases de resolución de problemas. Prácticas computacionales en aulas informáticas. Exposiciones orales y debates.

#### SISTEMA DE EVALUACIÓN

50% de la calificación: Examen final de carácter práctico, consistente en la elaboración de un caso de estudio individual.

10% de la calificación: Evaluación continua (\*).

40% de la calificación: entrega del trabajo final.

(\*) La evaluación continua consiste en la entrega de varias prácticas que se indicarán a lo largo del curso.

**Peso porcentual del Examen Final:** 50

**Peso porcentual del resto de la evaluación:** 50

#### BIBLIOGRAFÍA BÁSICA

- A.J. Izenman Modern Multivariate Statistical Techniques, Springer, 2008
- E. Alpaydin Introduction to Machine Learning, 2nd Edition, MIT Press, 2010
- T. Hastie, R. Tibshirani, J. Friedman The Elements of Statistical Learning, 2d Ed., Springer, 2009
- X. Wu The top ten algorithms in data mining, Chapman & Hall /CRC, 2009

#### BIBLIOGRAFÍA COMPLEMENTARIA

- I.H. Witten , E. Frank, M.A. Hall Data Mining. Practical Machine Learning Tools and Techniques, 3d Edition, Morgan Kaufmann, 2011
- John M. Chambers Software for Data Analysis. Programming with R., Springer, 2008
- Luis Torgo Data Mining with R, Chapman & Hall/CRC, 2001
- W.J. Braun, D.J. Murdoch A first course in statistical programming with R, Cambridge University Press, 2007

#### RECURSOS ELECTRÓNICOS BÁSICOS

- Rbloggers . Blogs de R: <https://www.r-bloggers.com/>