



PRESENTACIÓN

Breve descripción:

- **Titulación:** Máster Universitario en Ciencia de Datos Masivos / Big Data Science
- **Módulo/Materia:** Análisis de Datos
- **ECTS:** 8
- **Curso, semestre:** curso único del Máster, primer y segundo semestre
- **Carácter:** obligatoria
- **Profesorado:** Montserrat Miranda, Luis Alberiko Gil-Alaña, Juncal Cuñado, Elena Martín de Diego
- **Idioma:** castellano
- **Aula, Horario:** ver [cronograma](#)

COMPETENCIAS

Competencias Generales:

CG2 - Explorar y explotar tecnológica y estratégicamente los datos como un valor clave para diferentes empresas y organizaciones.

CG4 - Analizar los datos masivos y aportar medidas originales y novedosas para contribuir a la mejora de la competitividad de las empresas y otras instituciones.

CG5 - Analizar los datos que se generan en el día a día, extrayendo conocimiento de los mismos, realizando predicciones y transformándolos en productos y servicios utilizando las herramientas estadísticas de Data Science.

CG6 - Trabajar en equipos o grupos de investigación interdisciplinarios de forma eficaz y colaborativa.

CG7 - Conocer y entender las herramientas habituales que se utilizan en el análisis de grandes cantidades de datos.

CG8 - Saber aplicar los principios éticos relativos a la recogida, almacenamiento, y análisis de datos teniendo en cuenta las posibles discriminaciones directas o indirectas derivadas de la toma de decisiones.

Competencias Básicas:

CB6 - Poseer y comprender conocimientos que aporten una base u oportunidad de ser originales en el desarrollo y/o aplicación de ideas, a menudo en un contexto de investigación.

CB7 - Que los estudiantes sepan aplicar los conocimientos adquiridos y su capacidad de resolución de problemas en entornos nuevos o poco conocidos dentro de contextos más amplios (o multidisciplinares) relacionados con su área de estudio.

CB8 - Que los estudiantes sean capaces de integrar conocimientos y enfrentarse a la complejidad de formular juicios a partir de una información que, siendo incompleta o limitada, incluya reflexiones sobre las responsabilidades sociales y éticas vinculadas a la aplicación de sus conocimientos y juicios.



Universidad de Navarra

CB9 - Que los estudiantes sepan comunicar sus conclusiones y los conocimientos y razones últimas que las sustentan a públicos especializados y no especializados de un modo claro y sin ambigüedades.

CB10 - Que los estudiantes posean las habilidades de aprendizaje que les permitan continuar estudiando de un modo que habrá de ser en gran medida autodirigido o autónomo.

Competencias Específicas:

CE1 - Realizar un análisis exploratorio de datos.

CE2 - Preparar y limpiar los datos de modo eficiente mediante la corrección automatizada de errores de transcripción, tabulación, filtrado de datos, así como técnicas de componentes principales de reducción de la dimensionalidad.

CE10 - Conocer los fundamentos matemáticos de los modelos estadísticos para una aplicación de errores de transcripción, tabulación, filtrado de datos, así como técnicas de componentes principales y reducción de la dimensionalidad.

- CE4.1 - (Estadística) Programar con software estadístico libre R u otro similar y prácticas de cada contenido de esta materia con él.

PROGRAMA

1. Test de hipótesis

1.1 Tipos de error, p-valor y reglas de decisión

1.2 Test T de Student

1.3 Pruebas de Normalidad y medidas de homogeneidad de la varianza

1.4 Tipos de contraste de hipótesis. Potencia estadística y determinación del tamaño muestral

1.5 Pruebas paramétricas vs no paramétricas

1.6 Test de ANOVA (Analysis of Variance) y análisis post-hoc

2. Técnicas de regresión lineal múltiple

2.1 Modelos de regresión lineal

2.2 Regresión lineal univariable. Validación de un modelo de regresión

2.3 Regresión lineal multivariable

2.4 Multicolinealidad

2.5 Métodos no supervisado y supervisados de selección de variables predictoras

2.6 Técnicas de regularización

2.7 Regresión lineal aplicada a funciones no lineales



2.3 Regresión lineal multivariable

3. Series temporales

3.1 Modelos de series temporales estacionarias

3.2 Contrastes de raíz unitaria

3.3 Contrastes de cointegración

3.4 Modelización de series temporales económicas

3.5 Autoregresiones y procesos de media móvil

3.6 Estimación, test de hipótesis y selección del modelo

3.7 Estacionalidad

3.8 Predicción

3.9 Series temporales no estacionarias

3.10 Dominio de frecuencias y procesos de memoria larga

4. Técnicas de clústering

4.1 Introducción a las técnicas de clústering

4.2 Análisis de disimilaridad

4.3 Métricas de validación

4.3 El algoritmo de k-Means y sus variaciones: k-Medoids, Fuzzy y Weighted k-Means

4.4 Clústering jerárquico

4.5 DBSCAN

5. Técnicas de clasificación

5.1 Introducción a las técnicas de clasificación

5.2 Métricas de clasificación y curvas características.

5.3 Regresión logística

5.4 Naive Bayes

5.5 K-Primeros vecinos (KNN)

5.6 Clasificadores basados en árboles. Reglas de partición y métricas específicas. Árboles de decisión y Random Forest.



5.7 Clasificador de soporte vectorial (SVC)

ACTIVIDADES FORMATIVAS

En la asignatura habrá:

- Sesiones teórico-prácticas
- Ejercicios y problemas que deberán entregarse en las fechas que indique el profesor
- Un trabajo en grupo aplicando las técnicas aprendidas
- Examen práctico de modelado

La distribución del tiempo será:

ACTIVIDADES FORMATIVAS	HORAS
Clases presenciales teóricas	25
Prácticas con ordenador	25
Trabajos dirigidos	70
Estudio y trabajo personal	65
Pruebas presenciales de evaluación	5

EVALUACIÓN

La evaluación incluirá la asistencia a las sesiones, la calificación de las prácticas (actividades realizadas dentro o fuera de clase) y la realización del examen.

La nota final de la asignatura será un promedio con las ponderaciones siguientes

EN LA CONVOCATORIA ORDINARIA

- Asistencia a clases, seminarios y clases prácticas: 10%
- Actividades individuales y/o en equipo (series temporales): 30%
- Examen (modelado): 60%

EN LA CONVOCATORIA EXTRAORDINARIA

- Asistencia a clases, seminarios y clases prácticas: 10%
- Actividades individuales y/o en equipo (series temporales): 20%
- Examen (modelado): 70%

HORARIOS DE ATENCIÓN



Contactar por correo electrónico:

- emartindedi@unav.es
- montse@unav.es
- jcunado@unav.es
- alana@unav.es

BIBLIOGRAFÍA

Bibliografía Básica

- Martínez González MA, Jokin de Irala Estévez, Faulin Fajardo FJ. Bioestadística amigable. Madrid: Elsevier, 2014. [Localízalo en la Biblioteca](#) (Versión electrónica)
- Peña Sánchez de Rivera, D. Fundamentos de estadística. Alianza Editorial, 2008. [Localízalo en la Biblioteca](#)
- Peña Sánchez de Rivera, D. Regresión y Diseño de Experimentos. Alianza Editorial, 2008. [Localízalo en la Biblioteca](#)
- James, Gareth, [et al.]. An Introduction to Statistical Learning: With Applications in R [Electrónico]. New York, NY : Springer New York, 2013. [Localízalo en la Biblioteca](#)
- Davidson-Pilon, Cameron. Bayesian methods for hackers: probabilistic programming and Bayesian inference. New York : Addison-Wesley, [2016]. [Localízalo en la Biblioteca](#)
- Hastie, T., Tibshirani, R., Friedman, J. The Elements of Statistical Learning [Recurso electrónico]: Data Mining, Inference, and Prediction. New York: Springer, 2009. [Localízalo en la Biblioteca](#)
- Murphy, Kevin P. Machine learning: a probabilistic perspective. Cambridge, Mass.; London: MIT Press, c2012. [Localízalo en la Biblioteca](#)

Bibliografía Complementaria

- Ricardo Cao Abad (2017). Ingenuas reflexiones de un estadístico en la era del Big Data. Boletín de Estadística e Investigación Operativa 33(3), 295-321.
- León P. Bioética y explotación de grandes conjuntos de datos. En: Carnicero J. y Rojas D. (Coordinadores). La explotación de datos de salud: Retos, oportunidades y límites. Pamplona: Sociedad Española de Informática de la Salud, 2016. <http://www.seis.es>.
- López Fidalgo J. El azar no existe: tratamiento para el manipulador patológico, apto solamente para gente que piensa, (lea el prospecto antes de comenzar el tratamiento). Ediciones Electolibris, 2015. [Localízalo en la Biblioteca](#)