Universidad de Navarra

Análisis estadístico de datos (Grupo A) (MBDS)

Guía docente 2025-26

PRESENTACIÓN

Breve descripción: En esta asignatura, los estudiantes desarrollarán competencias en la aplicación de algoritmos avanzados de análisis estadístico y técnicas de modelado predictivo, esenciales para resolver problemas complejos en Big Data. A través del estudio de casos prácticos, los alumnos adquirirán la experiencia necesaria para implementar estas herramientas en entornos de investigación o en la industria.

- **Titulación**: Máster Universitario en Ciencia de Datos Masivos / Big Data Science
- **Módulo**: Análisis de Datos
- **Materia**: Estadística
- **ECTS**: 8
- **Curso, semestre**: curso único del Máster, primer y segundo semestre
- **Carácter**: obligatoria
- **Profesor responsable**: Jesús López Fidalgo
- **Profesorado**: Montserrat-Ana Miranda Galcerán, Javier Marín Moreno, Luis Alberiko Gil-Alaña, Juncal Cuñado Eizaguirre
- **Idioma**: castellano
- **Aula, Horario**: ver cronograma

RESULTADOS DE APRENDIZAJE (Competencias)

Competencias Básicas y 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 interdisciplinares 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.

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.

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 mediante la detección de ruido y outliers, tratamiento de valores perdidos así como detectar posibles relaciones entre variables y sus distribuciones de probabilidad.

CE10 - Conocer los fundamentos matemáticos de los modelos estadísticos para una aplicación adecuada mediante la verificación de las hipótesis básicas del modelo o ajustando este a las condiciones de los datos reales.

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)

A4CTIVIDADES 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
- Exámenes prácticos de modelado

La distribución del tiempo será:

ACTIVIDADES FORMATIVAS	HORAS
Clases presenciales teóricas	20
Prácticas con ordenador	25
Trabajos dirigidos	75
Tutorías personalizadas	7
Estudio y trabajo personal	68
Pruebas presenciales de evaluación	5

EVALUACIÓN

La evaluación incluirá la asistencia a las sesiones, la calificación de las prácticas o actividades , la realización de un examen práctico de modelado y un examen general conjuntamente con las asignaturas de ML y DL a final de curso.

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): 40%
- Examen general (AED+ML+DL): 20%

EN LA CONVOCATORIA EXTRAORDINARIA

- Asistencia a clases, seminarios y clases prácticas: 10%
- Actividades individuales y/o en equipo (series temporales): 20%
- Examen (modelado): 50%
- Examen general (AED+ML+DL): 20%

HORARIOS DE ATENCIÓN

Contactar por correo electrónico:

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

BIBLIOGRAFÍA

Bibliografía Básica

- Hastie, T., Tibshirani, R., Friedman, J. (2009). The Elements of Stadistical Learning: Data Mining, Inference, and Prediction (2nd ed.). Springer New York, NY. https://doi.org/10.1007/978-0-387-84858-7
- Murphy, Kevin P. (2012). Machine learning: a probabilistic perspective. The MIT Press. ISBN: 9780262304320, 978-0-262-01802-9 Localízalo en la Biblioteca
- Lazzery, F. (2021). Machine Learning for Time Series Forecasting with Python (1st ed.). John Wiley & Sons, Inc. https://doi.org/10.1002/9781119682394

Bibliografía Complementaria

- Aggarwal C.C. (2024). Probability and Statistics for Machine Learning (1st. ed.). Springer. https://doi.org/10.1007/978-3-031-53282-5
- López Fidalgo J. (2015) 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. Localízalo en la Biblioteca