Universidad de Navarra

Preparación y limpieza de datos (Grupo A) (MBDS)
Guía docente 2026-27

PRESENTACIÓN

Breve descripción: En esta asignatura, los estudiantes aprenderán a realizar el análisis exploratorio de los datos identificando aquellos valores que requieran tratamiento para asegurar la calidad y fiabilidad del análisis estadístico. Los estudiantes desarrollarán competencias técnicas para aplicar métodos de limpieza y transformación de los datos, garantizando la integridad del conjunto de datos y evitando que anomalías o vacíos comprometan la validez de los resultados.

- **Titulación**: Máster Universitario en Ciencia de Datos Masivos / Big Data Science
- **Módulo**: Análisis de Datos
- **Materia**: Estadística
- **ECTS**: 2
- **Curso, semestre**: curso único del Máster, primer y segundo semestre
- **Carácter**: obligatoria
- **Profesor responsable:** Jesús López Fidalgo
- **Profesorado**: Juan José Fernández Tebar
- **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.

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

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

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.

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 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.

Universidad de Navarra

PROGRAMA

1. Análisis exploratorio de datos
2. Pre procesado de datos
3. Detección de ruido y de outliers
4. Tratamiento de valores perdidos
5. Tratamiento del problema no balanceado
6. Datos estructurados y no estructurados
7. Evaluación de las distribuciones de las variables
8. Selección de características
9. Pipelines para limpieza y tratamiento de datos

ACTIVIDADES FORMATIVAS

ACTIVIDADES FORMATIVAS	HORAS
Clases presenciales teóricas	10
Prácticas con ordenador	10
Trabajos dirigidos	10
Estudio y trabajo personal	17
Tutorías personalizadas	3

EVALUACIÓN

La nota final será un promedio con las siguientes ponderaciones

EN LA CONVOCATORIA ORDINARIA

- Intervención en clases, seminarios y clases prácticas: 15%
- Resolución de ejercicios con el ordenador: 25%
- Trabajos individuales y/o en equipo: 60%

EN LA CONVOCATORIA EXTRAORDINARIA

- Intervención en clases, seminarios y clases prácticas: 15%
- Resolución de ejercicios con el ordenador:10%
- Trabajos individuales y/o en equipo: 85%

HORARIOS DE ATENCIÓN

Contactar por correo electrónico:

- jfernandezt@external.unav.es

Universidad de Navarra

BIBLIOGRAFÍA

Bibliografía básica:

Van der Loo, M., & De Jonge, E. (2018). Statistical data cleaning with applications in R. John Wiley & Sons, Incorporated. Localízalo en la Biblioteca (Versión electrónica)

Reyes, J. M. M. (2017). Introduction to Data Science for Social and Policy Research. Cambridge University Press. Localízalo en la Biblioteca (Versión electrónica)

Buttrey, S., & Whitaker, L. R. (2017). A Data Scientist's Guide to Acquiring, Cleaning, and Managing Data in R. Wiley. Localízalo en la Biblioteca (Versión electrónica)

Bibliografía complementaria:

Cody, R. (2017). *Cody's data cleaning techniques using SAS*. SAS Institute.