

29526 - Minería de datos

Información del Plan Docente

Año académico: 2023/24

Asignatura: 29526 - Minería de datos

Centro académico: 175 - Escuela Universitaria Politécnica de La Almunia

Titulación: 625 - Graduado en Ingeniería de Datos en Procesos Industriales

Créditos: 6.0

Curso: 3

Periodo de impartición: Primer semestre

Clase de asignatura: Obligatoria

Materia:

1. Información básica de la asignatura

La Minería de Datos es una disciplina que se centra en el procesamiento de datos obtenidos de sistemas de información complejos, con el objetivo de extraer conocimientos relevantes y útiles para la toma de decisiones en diversos contextos.

Esta área requiere la combinación de técnicas y metodologías provenientes de diferentes campos, como el diseño de almacenes de datos, la verificación de la calidad de los datos, el análisis multivariante de datos, la visualización de datos, la modelización estadística y el aprendizaje automático (machine learning). Estas disciplinas se integran para hacer frente a la complejidad inherente de la minería de datos.

2. Resultados de aprendizaje

- Aprender las bases necesarias para extraer información y visualizar un conjunto de datos.
- Identificar los pasos necesarios para un análisis de datos.
- Modelar las asociaciones entre variables con técnicas lineales y no lineales.
- Validar y comparar modelos mediante técnicas de remuestreo o segmentación de datos.

3. Programa de la asignatura

- Análisis exploratorio de un conjunto de datos: Calidad de datos. Resumen y visualización de datos. Imputación de datos perdidos.
- Aprendizaje supervisado: Modelos de clasificación y predicción. Validación de modelos. Regresión lineal, modelos lineales generalizados. Modelos lineales mixtos. Árboles de clasificación y regresión, modelos aditivos de árboles (random forest, bagging y boosting). Redes neuronales. Regresión lineal regularizada. Métodos LASSO.
- Aprendizaje no supervisado: análisis de componentes principales, análisis factorial, algoritmo K-means, análisis cluster.

4. Actividades académicas

La asignatura consiste en un total de 4 horas de clase presencial por semana a lo largo de un período de 15 semanas durante el cuatrimestre. Todas las clases se llevan a cabo en el aula de informática y se enfocan en impartir conceptos teóricos que se refuerzan con actividades prácticas de análisis y modelización de bases de datos utilizando programas de análisis estadístico.

5. Sistema de evaluación

Sistema de evaluación continua:

Pruebas escritas (60%): los estudiantes realizarán 2 pruebas escritas a lo largo del curso sobre modelización de datos con aprendizaje supervisado y no supervisado. Estas pruebas se realizarán en el aula de informática y conllevarán el análisis de bases de datos, su modelización e interpretación de resultados.

Controles participativos (20%): los estudiantes realizarán 2 controles participativos a lo largo del curso sobre análisis multivariante de un conjunto de datos y modelización multivariante en aprendizaje supervisado. Estas pruebas se realizarán en el aula de informática y conllevarán el análisis de bases de datos, su modelización e interpretación de resultados.

Caso práctico (20%): los estudiantes formarán grupos para analizar un caso práctico de modelización estadística a lo largo del curso que deberán defender mostrando el resultado final de su estudio.

Sistema de evaluación global:

Se realizará una única prueba escrita valorada en el total del 100% de la puntuación que versará sobre los contenidos de la asignatura. Esta prueba se realizará en el aula de informática y conllevará el análisis de bases de datos, su modelización e interpretación de resultados.