



Plan de estudios

Materia

Data Mining II

Curso: 2

Créditos: 5 ECTS

Idioma: Castellano

Competencias

Competencias básicas:

CB1. Que los estudiantes hayan demostrado poseer y comprender conocimientos en un área de estudio que parte de la base de la educación secundaria general, y se suele encontrar a un nivel que, si bien se apoya en libros de texto avanzados, incluye también algunos aspectos que implican conocimientos procedentes de la vanguardia de su campo de estudio

CB2. Que los estudiantes sepan aplicar sus conocimientos a su trabajo o vocación de una forma profesional y posean las competencias que suelen demostrarse por medio de la elaboración y defensa de argumentos y la resolución de problemas dentro de su área de estudio.

Competencias Generales / Personales:

CG1. Pensamiento analítico y capacidad para dar respuesta a retos complejos

Competencias Específicas / Profesionales:

CE2. Dispone de las habilidades de programación necesarias para abordar todo el ciclo de vida del dato

CE3. Dispone de habilidades estadísticas y matemáticas aplicadas a la ciencia de datos

CE4. Es capaz de formular una problemática relativa al mundo de los datos en términos de modelización

Resultados de aprendizaje

RA1. Es capaz de analizar, de manera lógica y desde distintas vertientes y disciplinas, los retos complejos a los que se enfrenta. Para ello, identifica cuáles son los aspectos clave e importantes del problema, los analiza y, tras una reflexión crítica y razonada, propone, de manera argumentada, diferentes alternativas/soluciones creativas y transformadoras.

RA9. Disponer de las habilidades de programación necesarias para abordar todo el ciclo de vida del dato

RA10. Disponer de habilidades estadísticas y matemáticas aplicadas a la ciencia de datos

RA11. Ser capaz de formular una problemática relativa al mundo de los datos en términos de modelización

Contenidos

Series temporales

- Función de autocorrelación (ACF)
- Modelos de Autorregresión (AR) y Medias Móviles (MA)
- Modelos ARMA
- Modelos ARIMA
- Auto ARIMA

Clasificación supervisada

- Imbalanced learning
- Regresión logística
- Naïve Bayes
- Support Vector Machines
- Validación de modelos
- Selección de modelos

Regresión

- Regresión lineal simple y múltiple
- Lasso & Elastic Net
- Principal Components Regression
- Generalized Linear Models
- Validación de modelos
- Selección de modelos

Clustering

- K-Means
- Mean Shift
- Clustering jerárquico
- DBSCAN
- OPTICS
- Gaussian Mixtures Models
- Validación y selección de modelos

Modelos de ensemble learning

- Bagging
- Boosting
- Gradient Boosting Methods

Selección automática de modelos (AutoML)

- TPOT

Optimización mediante algoritmos evolutivos

- Evolutionary Computation
- Ant Colony Optimization
- Genetic Algorithms
- Differential Evolution
- Coevolution

Actividades formativas

Las acciones formativas planificadas para este módulo son las siguientes:

- Aprendizaje basado en retos (2 ECTS)
- Talleres (0,5 ECTS)
- Recursos online (0,5 ECTS)
- Comunidades de aprendizaje (1 ECTS)
- Trabajo individual (1 ECTS)

Sistema de evaluación

La evaluación será mediante el sistema de evaluación continua a través de la cual se proporciona una información constante, tanto a los/las profesores como a los/las estudiantes, del proceso de aprendizaje a lo largo del período académico:

- Las actividades formativas de presentación de conocimientos y estudio individual se podrán evaluar con pruebas orales y/o escritas que corresponderán como máximo al 60% de la nota final.
- Las acciones formativas dirigidas a la adquisición de las competencias prácticas de las asignaturas se evaluarán a través de la realización de diferentes actividades (trabajos, casos, retos, etc.) correspondiendo como mínimo a un 40% de la nota final.

Los detalles de evaluación y calificación se harán explícitos en la planificación docente anual de las materias en función de los profesores responsables y de los condicionantes de cada curso.

Bibliografía

- Box, G. E. P., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). Time series analysis: Forecasting and control (5th ed). Hoboken, New Jersey: John Wiley & Sons
- Brockwell, P. J., & Davis, R. A. (2016). Introduction to time series and forecasting (3rd ed). New York, USA: Springer
- Albon C. (2018). Machine Learning with Python Cookbook: Practical Solutions from Preprocessing to Deep Learning. O'Reilly Media, Inc. ISBN: 9781491989388
- McKinney W. (2017). Python for Data Analysis (2nd Edition). O'Reilly Media, Inc. ISBN: 9781491957660
- Eiben A.E., Smith J.E. (2015). Introduction to Evolutionary Computing (2nd Edition). Springer-Verlag Berlin Heidelberg. ISBN: 9783662448731