



Plan de estudios

Materia Data Mining I

Curso: 1

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

Introducción al data mining

- Extracción de conocimiento de los datos
- Aplicaciones
- Aprendizaje supervisado y no supervisado

Datos para la tarea de data mining

- Tipos de variables
- Preparación de los datos: limpieza, valores perdidos
- Reducción del número de atributos
- UCI repositorio de datos

Regresión lineal y logística

- Clase numérica frente a clase categórica
- Ecuación de la recta obtenida mediante mínimos cuadrados
- Ecuación mediante regresión logística
- Comparativa

Árboles de decisión: regresión y clasificación (N8, J48, C4.5)

- Terminología
- El algoritmo
- Usando la entropía para la selección de atributos
- Maximizando la ganancia de información
- Índice de Gini
- Estimando la exactitud en la predicción mediante uso de train y test
- Matriz de confusión: falsos positivos y positivos verdaderos
- Overfitting

Introducción a la clasificación: clasificador Naive Bayes y KNN

- Qué es un clasificador
- Naive Bayes: probabilidad a posteriori
- K-Nearest Neighbours (KNN): medidas de la distancia

Clustering

- Introducción
- Tipos de algoritmos: k-means y clustering jerárquico
- K-means
- Clustering jerárquico

Reglas de asociación

- Ejemplo de la cesta de la compra
- Transacciones e itemsets
- Soporte de un itemset
- Generando reglas de asociación
- El algoritmo A priori

Sistemas de Recomendación

- Objetivos de un sistema recomendador
- Modelos básicos para un algoritmo recomendador
- Evaluación de un sistema recomendador

- El problema del arranque en frío
- Comparación de recomendadores
- Otros algoritmos de recomendación

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

- Bramer, M (2016). Principles of Data Mining. 3ª edición. Reino Unido. Springer
- Aggarwal, C (2016). Recommender Systems. Suiza. Springer