



Guía docente

240346 - 240E0065 - Ciencia de Datos

Última modificación: 03/07/2024

Unidad responsable: Escuela Técnica Superior de Ingeniería Industrial de Barcelona
Unidad que imparte: 715 - EIO - Departamento de Estadística e Investigación Operativa.

Titulación: MÁSTER UNIVERSITARIO EN INGENIERÍA INDUSTRIAL (Plan 2014). (Asignatura optativa).
MÁSTER UNIVERSITARIO EN INGENIERÍA DE ORGANIZACIÓN (Plan 2021). (Asignatura obligatoria).
MÁSTER UNIVERSITARIO EN INVESTIGACIÓN EN INGENIERÍA MECÁNICA (Plan 2024). (Asignatura optativa).

Curso: 2024 **Créditos ECTS:** 5.0 **Idiomas:** Inglés

PROFESORADO

Profesorado responsable: Puig De Dou, Ignacio

Otros: Marta Castellano Palomino

REQUISITOS

Álgebra lineal y Cálculo Diferencial e Integral a nivel de los cursos del GETI.
Conocimientos de Probabilidad e Inferencia estadística a nivel de la asignatura de Estadística del GETI.
Fundamentos de Programación.

COMPETENCIAS DE LA TITULACIÓN A LAS QUE CONTRIBUYE LA ASIGNATURA

Específicas:

CEEORG2. Diseñar, desarrollar y aplicar métodos analíticos (métodos cuantitativos, estadística, modelos y herramientas de decisión) para la toma de decisiones estratégicas, tácticas y operativas en las organizaciones.

CEETI3. (CAST) Analitzar dades de gran volum mitjançant models estadístics i algorismes d'intel·ligència artificial. (Competència específica associada a l'especialitat en Tecnologies de la Informació per a la Indústria).

CEETI4. (CAST) Determinar la fiabilitat i seguretat en sistemes complexos mitjançant mètodes de manteniment predictiu i seguretat ciberfísica. (Competència específica associada a l'especialitat en Tecnologies de la Informació per a la Indústria).

Transversales:

CT4. USO SOLVENTE DE LOS RECURSOS DE INFORMACIÓN: Gestionar la adquisición, la estructuración, el análisis y la visualización de datos e información en el ámbito de especialidad, y valorar de forma crítica los resultados de dicha gestión.

Básicas:

CB 6. (CAST) Tenir i comprendre coneixements que aportin una base o oportunitat de ser originals en el desenvolupament i/o aplicació d'idees, sovint en un context d'investigació

CB 9. (CAST) Que els estudiants sàpiguen comunicar les seves conclusions i coneixements (i darrers raonaments que els sustentin), a públics especialitzats i no especialitzats de manera clara i sense ambigüitats.

CB10. (CAST) Que els estudiants poseixin les habilitats d'aprenentatge que els permetin continuar estudiant d'una manera d'una forma que haurà de ser en gran mesura autodirigit o autònom

METODOLOGÍAS DOCENTES

Presentación de fundamentos teóricos de las diferentes metodologías en clases magistrales de hora y media. Implementación práctica de las metodologías con ejemplos en sesiones en aula de ordenadores de hora y media.
Análisis y presentación de una problemática más compleja a través de un caso práctico



OBJETIVOS DE APRENDIZAJE DE LA ASIGNATURA

El alumno obtendrá un conocimiento de las características básicas de los modelos de aprendizaje estadístico y máquina y sabrá qué modelo emplear en función del problema a solucionar.

Entenderá las metodologías de entrenamiento y validación de los modelos de aprendizaje automático y será capaz de planificar de manera robusta el proceso de entrenamiento de un modelo.

Conocerá los diferentes modelos disponibles de aprendizaje automático y será capaz de implementarlos en casos reales.

Conocerá y sabrá emplear software R para implementar y validar modelos de aprendizaje automático.

HORAS TOTALES DE DEDICACIÓN DEL ESTUDIANTADO

Tipo	Horas	Porcentaje
Horas grupo pequeño	22,5	18.00
Horas grupo grande	22,5	18.00
Horas aprendizaje autónomo	80,0	64.00

Dedicación total: 125 h

CONTENIDOS

Introducción al Aprendizaje Máquina con R

Descripción:

Data Science en la Industria.

Conceptos básicos del Aprendizaje Máquina

Introducción a R:

- análisis descriptivo
- lenguaje y programación
- data wrangling

Dedicación: 20h

Grupo mediano/Prácticas: 10h

Aprendizaje autónomo: 10h

Aprendizaje Supervisado: el problema de Regresión y Clasificación

Descripción:

Regresión Multivariante Lineal y no Lineal

Regresión Logística como método de clasificación

Bondad de ajuste en modelos de clasificación.

Selección y ajuste de modelos. Remuestreo

Dedicación: 30h

Grupo mediano/Prácticas: 15h

Aprendizaje autónomo: 15h



Métodos de aprendizaje no supervisados

Descripción:

Técnicas de reducción de la dimensionalidad

Métodos de clústering basados en distancias: modelos jerárquicos y aglomerativos

Dedicación: 20h

Grupo mediano/Prácticas: 10h

Aprendizaje autónomo: 10h

Métodos de Aprendizaje Máquina Avanzados: basados en árboles y redes neuronales

Descripción:

Métodos basados en árboles: bagging, random forest y boosting.

Redes neuronales

Dedicación: 20h

Grupo mediano/Prácticas: 10h

Aprendizaje autónomo: 10h

SISTEMA DE CALIFICACIÓN

0.30 exámenes de evaluación continuada + 0.25 trabajo de fin de curso + 0.45 examen final

Habrán tres exámenes de evaluación continuada realizados en horario habitual de clase. La nota del examen final sustituirá la del parcial en caso de ser superior.

NORMAS PARA LA REALIZACIÓN DE LAS PRUEBAS.

Aplican las normas que establece la ETSEIB-UPC.

BIBLIOGRAFÍA

Básica:

- Gareth James, ... [et al.]. An Introduction to statistical learning : with applications in R [en línea]. 2nd ed. New York: Springer, 2021 [Consulta: 01/10/2024]. Disponible a:

<https://ebookcentral-proquest-com.recursos.biblioteca.upc.edu/lib/upcatalunya-ebooks/detail.action?pq-origsite=primo&docID=6686746>. ISBN 1071614177.

- Manly, Bryan F. J ; Navarro Alberto, Jorge A. Multivariate statistical methods : a primer. Fourth edition. Boca Raton: CRC Press, Taylor & Francis Group, [2017]. ISBN 9781498728966.

Complementaria:

- Hastie, T.; Tibshirani, R.; Friedman, J. The elements of statistical learning : data mining, inference and prediction [en línea]. 2nd ed. New York [etc.]: Springer, cop. 2009 [Consulta: 25/08/2022]. Disponible a: <https://link-springer-com.recursos.biblioteca.upc.edu/book/10.1007/978-0-387-84858-7>. ISBN 0387952845.

- Peña, Daniel. Análisis de datos multivariantes [en línea]. 1a ed. Madrid: McGraw-Hill/Interamericana de España, S.L., 2013 [Consulta: 22/10/2024]. Disponible a:

https://www-ingebook-com.recursos.biblioteca.upc.edu/ib/NPcd/IB_BooksVis?cod_primaria=1000187&codigo_libro=4203. ISBN 9788448191849.

- Bertsekas, Dimitri P; Tsitsiklis, John N. Introduction to probability : Dimitri P. Bertsekas and John N. Tsitsiklis. 2nd ed. Belmont, Massachusetts: Athena Scientific, cop. 2008. ISBN 9781886529236.



RECURSOS

Enlace web:

- <https://cdr-book.github.io/>. Recurso