



Universitat de Lleida

GUÍA DOCENTE
EXPLOTACIÓN DE DATOS

Coordinación: BEJAR TORRES, RAMON

Año académico 2022-23

Información general de la asignatura

Denominación	EXPLOTACIÓN DE DATOS			
Código	103089			
Semestre de impartición	1R Q(SEMESTRE) EVALUACIÓN CONTINUADA			
Carácter	Grado/Máster	Curso	Carácter	Modalidad
	Máster Universitario en Ingeniería Informática	2	OPTATIVA	Presencial
Número de créditos de la asignatura (ECTS)	6			
Tipo de actividad, créditos y grupos	Tipo de actividad	PRALAB	TEORIA	
	Número de créditos	3	3	
	Número de grupos	1	1	
Coordinación	BEJAR TORRES, RAMON			
Departamento/s	INFORMATICA E INGENIERIA INDUSTRIAL			
Distribución carga docente entre la clase presencial y el trabajo autónomo del estudiante	30% de horas presenciales (3 horas/semana) and 70% de horas de trabajo autónomo.			
Información importante sobre tratamiento de datos	Consulte este enlace para obtener más información.			
Idioma/es de impartición	Inglés			

Profesor/a (es/as)	Dirección electrónica\nprofesor/a (es/as)	Créditos impartidos por el profesorado	Horario de tutoría/lugar
BEJAR TORRES, RAMON	ramon.bejar@udl.cat	3	
MATEU PIÑOL, CARLOS	carles.mateu@udl.cat	3	

Información complementaria de la asignatura

El estudiante debe de tener conocimientos sólidos y experiencia previa sobre programación estructurada, orientada a objetos y funcional sobre el lenguaje python.

Objetivos académicos de la asignatura

1. Conocer las herramientas actuales de limpieza de datos y análisis de datos
2. Conocer los fundamentos para el desarrollo de procedimientos centrados en datos mediante herramientas de programación interactivas
3. Saber transformar datos brutos de cualquier fuente en datos coherentes para su análisis
4. Saber implementar y depurar procedimientos para la transformación de conjuntos de datos masivos utilizando aproximaciones Big Data
5. Adquirir un espíritu escéptico frente a conjuntos de datos para incentivar el análisis exploratorio utilizando herramientas informáticas
6. Adquirir las herramientas y conocimientos para el análisis descriptivo de conjuntos de datos potencialmente masivos e intratable
7. Conocer y utilizar los algoritmos de minería de datos más básicos para descubrir características relevantes de grandes conjuntos de datos
8. Conocer y utilizar algoritmos básicos y avanzados para el aprendizaje automático adecuados para aplicaciones con grandes cantidades de datos
9. Conocer los principios fundamentales de los sistemas de recomendación

Competencias

Competencias estratégicas de la Universidad de Lleida

- **UdL2:** Dominio de una lengua extranjera.

Competencias transversales de la EPS

- **EPS1:** Capacidad de planificar y organizar el trabajo personal.
- **EPS3:** Capacidad para transmitir información, ideas, problemas y soluciones a un público especializado y no especializado.

- **EPS4:** Capacidad para concebir, diseñar e implementar proyectos y / o contribuir a nuevas soluciones, utilizando herramientas de ingeniería.
- **EPS5:** Estar motivado por la calidad y la mejora constante.

Competencias generales

- **CG4:** Capacidad para modelar, calcular y simular matemáticamente en empresas tecnológicas y centros de ingeniería, especialmente en tareas de investigación, desarrollo e innovación en todos los campos relacionados con la ingeniería informática.
- **CG8:** Capacidad para aplicar los conocimientos adquiridos para resolver problemas en situaciones nuevas y poco familiares dentro de contextos más amplios y más multidisciplinares, y ser capaz de integrar este conocimiento.

Competencias básicas

- **CB3:** Los estudiantes son capaces de integrar el conocimiento y manejar la complejidad, y de formular juicios basados en información incompleta o limitada, incluyendo reflexionar sobre aspectos sociales y éticos.
- **CB4:** Los estudiantes pueden comunicar sus conclusiones y el conocimiento y fundamento que las sustenta, a los públicos especializados y no especializados de manera clara e inequívoca.

Competencias específicas del grado

- **CE1:** Capacidad para entender y aplicar conocimientos avanzados en computación de alto rendimiento y métodos numéricos y computacionales a problemas de ingeniería.
- **CE4:** Capacidad para modelar, diseñar, definir la arquitectura, implantar, gestionar, operar, administrar y mantener aplicaciones, redes, sistemas, servicios y contenidos informáticos.

Contenidos fundamentales de la asignatura

1. Extracción de información y aprendizaje automático
 1. Extracción de elementos distintivos y frecuentes. Extracción de subconjuntos frecuentes
 2. Análisis y reducción de la dimensionalidad
 3. Regresión lineal y logística con SGD
 4. Clasificadores Naive Bayes
 5. Clustering
 6. Sistemas de recomendación
2. Redes neuronales y Deep Learning
 1. Introducción a las redes neuronales
 2. Redes neuronales profundas
 3. Redes Convolucionales
 4. Redes Recurrentes
3. Control: Reinforcement Learning.

Ejes metodológicos de la asignatura

Cada semana, cada estudiante recibirá:

- Tres horas de asistencia. Las clases magistrales se llevan a cabo mediante explicaciones teóricas acompañadas de ejemplos ilustrativos en la primera parte, finalizando con ejercicios prácticos en la segunda parte. Como material de apoyo de la clase, seguiremos las diapositivas o cuadernos de python del curso.
- Otros materiales de apoyo para seguir el tema de una manera de no asistencia.

La evaluación es continua a lo largo del semestre y consta de cuatro partes diferentes:

- Dos prácticas: Ampliación de la aplicación bigdata comenzó en el tema anterior "Masivo de procesamiento de datos" con redes neuronales y herramientas de datos grandes.
- Dos informes y presentaciones orales sobre la integración de redes neuronales y herramientas de minado de datos en su aplicación bigdata.

Debido al COVID-19 algunas clases serán en formato no presencial, es decir, mediante videos y videoconferencia.

Plan de desarrollo de la asignatura

Semanas:

1. Mining frequent items/item sets.
2. Mining distinct elements.
3. Dimensionality reduction.
4. Linear and Logistic regression.
5. Naive Bayes classifiers.
6. Clustering - crisp and probabilistic.
7. Recommender systems.
8. Work on Data mining projects.
9. Work on Big data application.
10. Neural Networks.
11. Feed forward neural networks.
12. Deep Learning.
13. Convolutional Neural Networks.
14. Convolutional Neural Networks II.
15. Recursive Neural Networks.
16. Control: Reinforcement Learning.
17. Work on Deep Learning projects.
18. Work on Big Data Applications.
19. Final presentations of big data projects.

Debido al COVID-19 algunas clases serán en formato no presencial, es decir, mediante videos y videoconferencias.

Sistema de evaluación

La evaluación de este curso se basa en una evaluación continua.

Actividades de evaluación	%	Fechas	O/V (1)	I/G (2)
<i>Data mining application (3)</i>	40%	End Sem.	M	G
<i>Presentación de la aplicación (4)</i>	10%	Middle Sem.	M	G
Deep learning exercises	25%	Middle Sem.	M	G
Data mining excises	25%	Middle Sem.	M	G

(1) Mandatory / Voluntary

(2) Individual / Group

(3) Habrá una revisión del trabajo realizado por cada miembro del grupo

(4) Cada miembro del grupo debe participar por igual en la presentación y responder a las preguntas del evaluador

Bibliografía y recursos de información

- Wes McKinney. Python for Data Analysis: Data Wrangling with Pandas, NumPy, and Ipython. O'Really, 2012
- Holden Karau, Andy Konwinski, Patrick Wendell & Matei Zaharia. Learning Spark. O'Really, 2015
- Jure Leskovec, Anand Rajaraman & Jeffrey David Ullman. Mining of Massive Datasets. Cambridge University Press , 2014 (Podéis encontrar una copia gratuita en <http://www.mmds.org/> , también se encuentra disponible en la biblioteca).