



Universitat de Lleida

GUÍA DOCENTE
PROYECTO BIG DATA

Coordinación: COMAS RODRIGUEZ, CARLOS

Año académico 2022-23

Información general de la asignatura

Denominación	PROYECTO BIG DATA			
Código	103090			
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	COMAS RODRIGUEZ, CARLOS			
Departamento/s	INFORMATICA E INGENIERIA INDUSTRIAL			
Distribución carga docente entre la clase presencial y el trabajo autónomo del estudiante	30% tiempo sesiones presenciales (3 horas/semana) y el 70% restante se basa en trabajo autónomo del estudiante.			
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
COMAS RODRIGUEZ, CARLOS	carles.comas@udl.cat	3	
VIRGILI GOMA, JORDI	jordi.virgili@udl.cat	3	

Información complementaria de la asignatura

Para un mejor aprovechamiento de esta asignatura, se recomienda que el estudiante tenga conocimientos de programación en Python y que haya completado la asignatura Massive Data Processing.

Objetivos académicos de la asignatura

1. Comprender y aplicar técnicas estadísticas de minería de datos.
2. Utilizar adecuadamente los paquetes estadísticos para el análisis de datos.
3. Comprender y aplicar técnicas estadísticas relacionadas con el funcionamiento de grandes volúmenes de datos.
4. Utilizar adecuadamente los paquetes estadísticos para este tipo de explotación.
5. Proponer una adecuada visualización de la información analizada que facilita la comprensión.
6. Conocer las aplicaciones comunes en el ámbito de grandes volúmenes de datos y ser capaz de desarrollar soluciones para estos problemas.
7. Sistemáticamente desplegar el conjunto de técnicas comunes para resolver problemas grandes de datos.
8. Comunicar eficazmente los resultados del proyecto a los socios técnicos y clientes.

Competencias

Competencias específicas

Las competencias específicas para el ejercicio de la profesión de Ingeniero Informático, y a las que da acceso el Master en Ingeniería Informática de la UdL son:

*Capacidad para la integración de tecnologías, aplicaciones, servicios y sistemas propios de la Ingeniería Informática, con carácter generalista, y en contextos más amplios y multidisciplinares.

*Capacidad para la planificación estratégica, elaboración, dirección, coordinación, y gestión técnica y económica en los ámbitos de la ingeniería informática relacionados, entre otros, con: sistemas, aplicaciones, servicios, redes, infraestructuras o instalaciones informáticas y centros o factorías de desarrollo de software, respetando el adecuado cumplimiento de los criterios de calidad y medioambientales y en entornos de trabajo multidisciplinares.

*Capacidad para la dirección de proyectos de investigación, desarrollo e innovación, en empresas y centros tecnológicos, con garantía de la seguridad para las personas y bienes, la calidad final de los productos y su homologación.

*Capacidad para modelar, diseñar, definir la arquitectura, implantar, gestionar, operar, administrar y mantener aplicaciones, redes, sistemas, servicios y contenidos informáticos.

*Capacidad de comprender y saber aplicar el funcionamiento y organización de Internet, las tecnologías y

protocolos de redes de nueva generación, los modelos de componentes, software intermediario y servicios.

*Capacidad para asegurar, gestionar, auditar y certificar la calidad de los desarrollos, procesos, sistemas, servicios, aplicaciones y productos informáticos.

*Capacidad para diseñar, desarrollar, gestionar y evaluar mecanismos de certificación y garantía de seguridad en el tratamiento y acceso a la información en un sistema de procesamiento local o distribuido. *Capacidad para analizar las necesidades de información que se plantean en un entorno y llevar a cabo en todas sus etapas el proceso de construcción de un sistema de información.

*Capacidad para diseñar y evaluar sistemas operativos y servidores, y aplicaciones y sistemas basados en computación distribuida.

*Capacidad para comprender y poder aplicar conocimientos avanzados de computación de altas prestaciones y métodos numéricos o computacionales a problemas de ingeniería. Capacidad de diseñar y desarrollar sistemas, aplicaciones y servicios informáticos en sistemas empujados y ubicuos.

*Capacidad para aplicar métodos matemáticos, estadísticos y de inteligencia artificial para modelar, diseñar y desarrollar aplicaciones, servicios, sistemas inteligentes y sistemas basados en el conocimiento. *Capacidad para utilizar y desarrollar metodologías, métodos, técnicas, programas de uso específico, normas y estándares de computación gráfica. Capacidad para conceptualizar, diseñar, desarrollar y evaluar la interacción persona-ordenador de productos, sistemas, aplicaciones y servicios informáticos.

*Capacidad para la creación y explotación de entornos virtuales, y para la creación, gestión y distribución de contenidos multimedia.

Competencias transversales

Por otra parte la propia Universidad de Lleida y la Escuela Politécnica Superior establecen una serie de competencias transversales en el diseño del plan de estudios de todas las titulaciones de la Escuela que incluyen:

*Corrección en la expresión oral escrita. Dominio de una lengua extranjera.

*Dominio de las TIC.

*Respecto a los derechos fundamentales de igualdad entre hombres y mujeres, a la promoción de los Derechos Humanos y los valores propios de una cultura de paz y de valores democráticos.

*Capacidad de planificación y organización del trabajo personal.

*Capacidad de considerar el contexto socioeconómico así como los criterios de sostenibilidad en las soluciones de ingeniería.

*Capacidad de transmitir información, ideas, problemas y soluciones a un público tanto especializado como no especializado.

*Capacidad de concebir, diseñar e implementar proyectos y / o aportar soluciones nuevas, utilizando herramientas propias de la ingeniería.

*Tener motivación por la calidad y la mejora continua.

Perspectiva de género en la docencia.

Acciones básicas

•En la guía y material docente y en el aula, cuidar que el lenguaje sea inclusivo y no sexista.

•En el material docente, cuidar que las imágenes no perpetúen los estereotipos de género.

•En el material docente, cuidar que los ejemplos y ejercicios contrarresten estereotipos de género.

- En el material docente, cuidar que el contexto de los ejemplos y de los ejercicios cubran temáticas diversas.
- En la medida de lo posible incluir enunciados con relevancia social y/o de género.

Acciones más avanzadas

- En los proyectos, promover que se estudie algún aspecto con relevancia social y/o de género.
- Destacar explícitamente la relevancia social y/o de género en las actividades (proyectos, casos, prácticas).
- Contextualizar los enunciados de los exámenes para evidenciar la relevancia social y/o de género de la asignatura.
- Incorporar las variables “género” y “sexo” en el análisis (análisis estadísticos, diseño de soluciones, etc.).
- Incorporar a la Guía Docente objetivos relacionados con la relevancia social y/o de género.

Contenidos fundamentales de la asignatura

1. Introducción a los algoritmos PCA (Principal Component Analysis) y EM (Expectation-Maximization)
2. Principal Component Analysis
 1. Matrices de datos y espacios asociados
 2. Principal Component Analysis
 3. Interpretación y calidad de resultados PCA
3. Algoritmo Expectation-Maximization
 1. Maximum Likelihood Estimation (MLE)
 2. Algoritmo Expectation-Maximization (EM)
 3. EM para Missing Data
 4. Gaussian Mixture Model (EM Clustering)
4. Business Intelligence
 1. Apoyo a decisiones empresariales con Big Data
 2. Smart Data
 3. Visualización con matplotlib
 4. Databricks, Google Cloud,...
5. Exploración interactiva de Big Data
 1. Apache Spark para consultar y explorar interactivamente grandes volúmenes de datos heterogéneos
 2. SparkSQL
 3. Spark R y Exploratory Data Analysis

Ejes metodológicos de la asignatura

Todos los cursos del bloque de análisis del Big Datos (incluido éste), serán evaluados por un único, común, proyecto que involucra a todas las asignaturas (recopilación de datos, procesamiento, aprendizaje, estadística, visualización, etc.).

Los estudiantes trabajarán en este proyecto desde el principio hasta los últimos cursos. Durante los cursos regulares, se introducirán diferentes temas, mostrando su relación con el proyecto común y cómo todos los temas encajan entre sí para crear una tarea o un proyecto complejo del mundo real.

Los tres cursos de formación de Big Data Analytics utilizarán la misma configuración de base tecnológica:

- Python como lenguaje de programación de base.
- Hadoop / Spark (con Java, si es necesario)
- Si bien durante los cursos se introducirán otros paquetes tecnológicos:

- Scala
- NodeJS
- Etc.

Plan de desarrollo de la asignatura

Semana	Descripción	Actividades Presenciales	Trabajo Autónomo Alumno
1	Introducción a CPA y algoritmo EM	Presentación asignatura Clase magistral y participativa	Estudio y resolución de ejercicios
2	Data Matrices y Associated Spaces y PCA	Clase magistral y participativa	Estudio y resolución de ejercicios
3	Bank Holiday (Sant Miquel)		
4	Interpretación y calidad de resultados de PCA	Clase magistral y participativa	Estudio y resolución de ejercicios
5	Maximum Likelihood Estimation (MLE)	Clase magistral y participativa	Estudio y resolución de ejercicios
6	Expectation-Maximization (EM) Algorithm	Clase magistral y participativa	Estudio y resolución de ejercicios
7	Presentaciones Orales	Presentaciones CPA y EM	Estudio y resolución de ejercicios Desarrollo proyecto
8	Apache Spark Intro y Demos Revisión Big Data Project	Clase magistral y participativa Presentaciones Proyectos	Casos de estudio Desarrollo proyecto
9	Big Data Exploration	Clase magistral y participativa	Desarrollo proyecto
10	Big Data Exploration	Clase magistral y participativa	Desarrollo proyecto
11	Big Data Exploration	Clase magistral y participativa	Desarrollo proyecto
12	Big Data Exploration	Clase magistral y participativa	Desarrollo proyecto
13	Business Intelligence	Clase magistral y participativa	Desarrollo proyecto
14	Business Intelligence	Clase magistral y participativa	Desarrollo proyecto
15	Business Intelligence	Presentaciones Proyectos	Desarrollo proyecto

Sistema de evaluación

La evaluación de esta asignatura se basa en la evaluación continua. En función de la situación sanitaria, algunas de estas actividades podrían realizarse como una actividad en el aula o virtualmente utilizando herramientas del Campus Virtual.

ID	Actividades evaluación	%	Fechas	Obligatoria	I/G (1)
OP1	Presentación Oral Resolución de problema práctico (CPA)	25%	Semana 7	Sí	Grupo
OP2	Presentación Oral Resolución de problema práctico (EM)	25%	Semana 7	Sí	Grupo
PP	Documentación Escrita y Presentación Oral Planificación Proyecto	15%	Semana 10	Sí	Grupo
PD	Documentación Escrita Entregable Proyecto	15%	Semana 15	Sí	Grupo
OP3	Presentación Oral Proyecto	20%	Semana 15	Sí	Grupo

(1) Individual / Grupo

La nota final se calculará en base a la siguiente fórmula:

$$\text{Nota Final} = 0,25 \cdot \text{OP1} + 0,25 \cdot \text{OP2} + 0,15 \cdot \text{PP} + 0,15 \cdot \text{PD} + 0,2 \cdot \text{OP3}$$

Bibliografía y recursos de información

[Kar15] Holden Karau, Andy Konwinski, Patrick Wendell, Matei Zaharia, "Learning Spark: Lightning-Fast Big Data Analysis", O'Reilly, 2015

[Ryz15] Sandy Ryza, Uri Laserson, Sean Owen, Josh Wills, "Advanced Analytics with Spark: Patterns for Learning from Data at Scale", O'Reilly, 2015

[Bae14] Bart Baesens, "Analytics in a Big Data World: The Essential Guide to Data Science and its Applications"