

Escuela de Doctorado y Estudios de Posgrado

Máster Universitario en Ingeniería Informática

GUÍA DOCENTE DE LA ASIGNATURA :

**Análisis de datos masivos
(2020 - 2021)**

1. Datos descriptivos de la asignatura

Asignatura: Análisis de datos masivos	Código: 835941203
<ul style="list-style-type: none"> - Centro: Escuela de Doctorado y Estudios de Postgrado - Lugar de impartición: Escuela Superior de Ingeniería y Tecnología. Sección de Ingeniería Informática - Titulación: Máster Universitario en Ingeniería Informática - Plan de Estudios: 2018 (Publicado en) - Rama de conocimiento: Ingeniería y Arquitectura - Itinerario / Intensificación: - Departamento/s: <ul style="list-style-type: none"> Ingeniería Informática y de Sistemas - Área/s de conocimiento: <ul style="list-style-type: none"> Ciencia de la Computación e Inteligencia Artificial Lenguajes y Sistemas Informáticos - Curso: 1 - Carácter: Obligatoria - Duración: Segundo cuatrimestre - Créditos ECTS: 6,0 - Modalidad de impartición: Semipresencial - Horario: Enlace al horario - Dirección web de la asignatura: http://www.campusvirtual.ull.es - Idioma: Castellano e Inglés (0,30 ECTS en Inglés) 	

2. Requisitos para cursar la asignatura

3. Profesorado que imparte la asignatura

Profesor/a Coordinador/a: FRANCISCO CARMELO ALMEIDA RODRIGUEZ
- Grupo:
General <ul style="list-style-type: none"> - Nombre: FRANCISCO CARMELO - Apellido: ALMEIDA RODRIGUEZ - Departamento: Ingeniería Informática y de Sistemas - Área de conocimiento: Lenguajes y Sistemas Informáticos
Contacto <ul style="list-style-type: none"> - Teléfono 1: 922 31 81 73 - Teléfono 2: - Correo electrónico: falmeida@ull.es - Correo alternativo: - Web: http://www.campusvirtual.ull.es

Tutorías primer cuatrimestre:						
Desde	Hasta	Día	Hora inicial	Hora final	Localización	Despacho
Todo el cuatrimestre		Lunes	11:00	13:00	Edificio de Física y Matemáticas - AN.2B	82
Todo el cuatrimestre		Martes	11:00	13:00	Edificio de Física y Matemáticas - AN.2B	82
Todo el cuatrimestre		Miércoles	15:00	17:00	Edificio de Física y Matemáticas - AN.2B	82

Observaciones:

Tutorías segundo cuatrimestre:						
Desde	Hasta	Día	Hora inicial	Hora final	Localización	Despacho
Todo el cuatrimestre		Lunes	11:00	13:00	Edificio de Física y Matemáticas - AN.2B	82
Todo el cuatrimestre		Jueves	11:00	13:00	Edificio de Física y Matemáticas - AN.2B	82
Todo el cuatrimestre		Jueves	15:00	17:00	Edificio de Física y Matemáticas - AN.2B	82

Observaciones:

Profesor/a: JOSE MARCOS MORENO VEGA
- Grupo:
General - Nombre: JOSE MARCOS - Apellido: MORENO VEGA - Departamento: Ingeniería Informática y de Sistemas - Área de conocimiento: Ciencia de la Computación e Inteligencia Artificial

Contacto

- Teléfono 1: **922318175**
- Teléfono 2:
- Correo electrónico: **jmmoreno@ull.es**
- Correo alternativo:
- Web: **<http://www.campusvirtual.ull.es>**

Tutorías primer cuatrimestre:

Desde	Hasta	Día	Hora inicial	Hora final	Localización	Despacho
Todo el cuatrimestre		Lunes	09:00	12:00	Edificio de Física y Matemáticas - AN.2B	84
Todo el cuatrimestre		Viernes	09:00	12:00	Edificio de Física y Matemáticas - AN.2B	84

Observaciones: Las tutorías de los viernes de 9:00-12:00, serán en línea. Para llevar a cabo la tutoría en línea, se hará uso de algunas de las herramientas institucionales disponibles para ello, preferentemente Google Meet, con la dirección del correo xxxx@ull.edu.es.

Tutorías segundo cuatrimestre:

Desde	Hasta	Día	Hora inicial	Hora final	Localización	Despacho
Todo el cuatrimestre		Lunes	09:00	12:00	Edificio de Física y Matemáticas - AN.2B	84
Todo el cuatrimestre		Viernes	09:00	12:00	Edificio de Física y Matemáticas - AN.2B	84

Observaciones: Las tutorías de los viernes de 9:00-12:00, serán en línea. Para llevar a cabo la tutoría en línea, se hará uso de algunas de las herramientas institucionales disponibles para ello, preferentemente Google Meet, con la dirección del correo xxxx@ull.edu.es.

4. Contextualización de la asignatura en el plan de estudio

Bloque formativo al que pertenece la asignatura: **Tecnologías Informáticas**
 Perfil profesional:

5. Competencias

Generales

CO1 - Capacidad para proyectar, calcular y diseñar productos, procesos e instalaciones en todos los ámbitos de la Ingeniería Informática

CO3 - Capacidad para dirigir, planificar y supervisar equipos multidisciplinares

CO4 - Capacidad para el modelado matemático, cálculo y simulación en centros tecnológicos y de ingeniería de empresa, particularmente en tareas de investigación, desarrollo e innovación en todos los ámbitos relacionados con la Ingeniería en Informática

CO7 - Capacidad para la puesta en marcha, dirección y gestión de procesos de fabricación de equipos informáticos, con garantía de la seguridad para las personas y bienes, la calidad final de los productos y su homologación

CO8 - Capacidad para la aplicación de los conocimientos adquiridos y de resolver problemas en entornos nuevos o poco conocidos dentro de contextos más amplios y multidisciplinarios, siendo capaces de integrar estos conocimientos

Específicas

SL3 - Comprender, diseñar, evaluar y aplicar tecnologías, herramientas, recursos, estándares en el marco del desarrollo de software y su implantación en sistemas y entornos libres.

TI_7 - Capacidad para comprender y poder aplicar conocimientos avanzados de computación de altas prestaciones y métodos numéricos o computacionales a problemas de ingeniería

TI_9 - 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

6. Contenidos de la asignatura

Contenidos teóricos y prácticos de la asignatura

- Introducción al problema del análisis de datos a gran escala: desafíos actuales, tendencias y aplicaciones.
- Algoritmos para el análisis de datos a gran escala: algoritmos de minería y aprendizaje automático que se han desarrollado específicamente para hacer frente a grandes conjuntos de datos.
- Tecnologías para la gestión de datos a gran escala: paradigma Map-Reduce y sistemas de soporte.
- Visualización de resultados.
- Software libre en el análisis de datos masivos.

Actividades a desarrollar en otro idioma

En virtud de lo dispuesto en la normativa autonómica (decreto 168/2008, 22 de julio) un 5% del contenido será impartido en inglés. En la guía docente de cada asignatura se indicarán las actividades formativas a impartir en inglés y el método de evaluación del contenido correspondiente.

7. Metodología y volumen de trabajo del estudiante

Descripción

* Enseñanza expositiva: clases teóricas donde el profesor expone los conceptos teóricos de la asignatura y clases prácticas o de problemas donde se estudiarán ejemplos prácticos. Se espera que tanto las clases teóricas como las prácticas sean participativas

* Tutorías individuales presenciales o virtuales a través del portal de la asignatura.

* Realización de trabajos/proyectos propuestos por el profesor que serán de tipo práctico y podrán realizarse dentro o fuera del aula virtual, y que serán el material utilizado para la evaluación continua.

* Dentro de las actividades on-line que se realizan en la asignatura, vamos a distinguir dos modalidades:

a) Online asíncrona en las que las comunicaciones profesor/alumno se realizan mediante herramientas de mensajería (correo electrónico, foros de discusión, ...). En estas actividades, se propondrá un aspecto teórico/práctico de la asignatura que el alumno deberá abordar haciendo uso del material suministrado por el profesor. Mediante sistemas de mensajería se establecerá el canal para la resolución de dudas o discusiones y se propondrá el entregable asociado (máquina virtual configurada, cuestionario,...)

b) Online síncrona en las que se fija un horario en el que se establece una comunicación directa profesor/alumno usando herramientas como chats y videoconferencias

Actividades formativas en créditos ECTS, su metodología de enseñanza-aprendizaje y su relación con las competencias que debe adquirir el estudiante

Actividades formativas	Horas presenciales	Horas de trabajo autónomo	Total horas	Relación con competencias
Clases teóricas	13,00	0,00	13,0	[TI_9], [TI_7], [SL3], [CO8], [CO4]
Clases prácticas (aula / sala de demostraciones / prácticas laboratorio)	7,50	7,50	15,0	[TI_9], [TI_7], [SL3], [CO7], [CO3]
Realización de seminarios u otras actividades complementarias	2,70	9,30	12,0	[CO7], [CO1]
Realización de trabajos (individual/grupal)	0,00	24,00	24,0	[TI_9], [TI_7], [SL3], [CO3]
Estudio/preparación de clases teóricas	0,00	25,00	25,0	[TI_9], [TI_7], [SL3], [CO4]
Estudio/preparación de clases prácticas	0,00	11,00	11,0	[TI_9], [TI_7], [SL3], [CO7], [CO4], [CO3]
Preparación de exámenes	0,00	10,00	10,0	[TI_9], [TI_7], [SL3]
Realización de exámenes	5,00	0,00	5,0	[TI_9], [TI_7], [SL3]
Preparación de informes u otros trabajos.	0,00	20,00	20,0	[TI_9], [TI_7], [SL3], [CO1]

Trabajos y proyectos	1,80	13,20	15,0	[TI_9], [TI_7], [SL3], [CO3], [CO1]
Total horas	30,00	120,00	150,00	
Total ECTS			6,00	

8. Bibliografía / Recursos

Bibliografía Básica

José Hernández Orallo, M.José Ramírez Quintana, Cesar Ferri Ramírez.

Introducción a la Minería de Datos

. Editorial Pearson, 2004

Ethem Alpaydin,

Introduction to Machine Learning

, The MIT Press, 2004

Pang-Ning Tan, Michael Steinbach, Anuj Karpatne, Vipin Kumar,

Introduction to Data Mining

, Addison-Wesley, 2006

Charu. C. Aggarwal,

Data Mining.

The Textbook

. Springer, 2015

Bibliografía Complementaria

Ian H. Witten, Eibe Frank, Mark A. Hall, Christopher J. Pal,

Data Mining

.

Practical Machine Learning Tools and Techniques

, Morgan Kaufmann, 2017

Foster Provost, Tom Fawcett, Data Science for Business, O'Reilly Media, 2013

Michael R. Berthold, Christian Borgelt, Frank Höppner, Frank Klawonn,

Guide to Intelligent Data Analysis

, Springer, 2010

Yanchang Zhao,

R and Data Mining: Examples and Case Studies,

Academic Press, 2012

Michael J. Crawley, The R Book, Wiley, 2013

Otros Recursos

R project

kaggle

Google colab

Python

Scipy

Apache Spark

9. Sistema de evaluación y calificación

Descripción

El sistema de evaluación, siguiendo las directrices especificadas en la Memoria del Título, será de 20% evaluación de la teoría (pruebas de respuesta corta y pruebas de desarrollo) y 80% evaluación de prácticas, trabajos y proyectos (memorias y trabajo realizado).

Es obligatorio asistir a clases y hacer uso de los foros y tutorías tanto presenciales como on-line.

Para la primera convocatoria de esta asignatura se aplicará un esquema de Evaluación Continua que combina las siguientes actividades:

- 1.- Prácticas individuales a desarrollar en los laboratorios frente al ordenador y online a través del campus virtual.
- 2.- Prácticas en grupo para desarrollar pequeños proyectos de datos masivos.
- 3.- Actividades online de tipo teórico y/o práctico.
- 4.- Exámen Final de la asignatura.

La Calificación Final (CF) de la asignatura en este primer llamamiento se obtiene a partir de una Calificación de Prácticas, Trabajos y Proyectos (CPTP) y la Calificación de Teoría. (CTeoria):

- 1.- CPTP – Calificación de Prácticas, Trabajos y Proyectos [valor numérico entre 0 y 10]. Esta nota se obtiene de la media ponderada de las calificaciones de las prácticas de laboratorio, trabajos individuales y proyectos. Estas actividades tendrán carácter individual y/o grupal.
- 2.- CTeoria – Calificación de Teoría [valor numérico entre 0 y 10]. Esta nota se obtiene de cuestionarios y/o exámenes realizados de forma individual.

En ningún caso las notas (CTeoria, CPTP) se guardarán de un curso a otro.

Así pues, la Calificación Final (CF) de la Evaluación Continua será:

$$CF = 20\% \text{ CTeoría} + 80\% \text{ CPTP}$$

La ponderación anterior se aplicará sólo en el caso de haber superado las dos partes, es decir, obtener una evaluación superior a 5 en CTeoría y CPTP.

La estrategia evaluativa se detalla en la tabla que aparece a continuación. En ella se establecen los criterios de evaluación de las competencias que se desarrollan en esta asignatura, así como la ponderación de los mismos dentro de los distintos tipos de calificación descritos en el párrafo anterior. En ningún caso se guardarán actividades para siguiente cursos académicos.

Las actividades desarrolladas en inglés se evalúan a través de la misma metodología que el resto de actividades, teniendo en cuenta que el inglés pudiera ser utilizado en la propia evaluación, tanto a la hora de formular las preguntas y/o ejercicios como a la hora de responderlas/resolverlos por parte del alumnado.

Estrategia Evaluativa

Tipo de prueba	Competencias	Criterios	Ponderación
Pruebas de respuesta corta	[TI_9], [TI_7], [SL3]	Adecuación de las respuestas dadas a las pruebas y claridad de redacción de las mismas.	5,00 %
Pruebas de desarrollo	[TI_9], [TI_7], [SL3]	Adecuación de las respuestas dadas a los ejercicios y claridad de redacción de las mismas. Uso adecuado de las herramientas de análisis y visualización. Coherencia de las conclusiones obtenidas	15,00 %
Trabajos y proyectos	[TI_9], [TI_7], [SL3], [CO8], [CO7], [CO4], [CO3], [CO1]	<ul style="list-style-type: none"> • Defensa oral • Estructura de la exposición, lenguaje empleado y respuesta a las preguntas. • Análisis y visualización de los datos • Grado de conocimiento adquirido en el manejo de las herramientas para el tratamiento inteligente de datos y su visualización. • Rigor en el análisis de los datos y coherencia de las conclusiones. 	70,00 %
Informes memorias de prácticas	[TI_9], [CO7], [CO3]	<ul style="list-style-type: none"> • Estructura, calidad y claridad de la redacción, fuentes consultadas, rigor en el análisis y visualización de los datos y coherencia de las conclusiones. 	10,00 %

10. Resultados de Aprendizaje

El alumnado se familiarizará con los conceptos fundamentales de administración y análisis de datos a gran escala. Será capaz de reconocer los desafíos a los que se enfrentan las aplicaciones que tratan con volúmenes muy grandes de datos, así como de proponer soluciones escalables para ellos. Además, será capaz de integrar las tecnologías para la gestión de datos a gran escala en contextos más amplios y multidisciplinares.

11. Cronograma / calendario de la asignatura

Descripción

-El cronograma que se presenta a continuación describe las actividades que se van a desarrollar durante el cuatrimestre en el que se imparte la asignatura. Sin embargo, esta planificación puede ser modificada si así lo demanda el desarrollo de la asignatura.

En dicho cronograma se presentan actividades que son presenciales (A) y otras que se desarrollan de forma online utilizando las herramientas TIC disponibles. Dentro de estas actividades "online", se indicará con "C" aquellas actividades online asíncronas y con "B", aquellas que sean síncronas.

Debido al carácter semipresencial del máster habrá semanas en las que las clases prácticas, trabajos y proyectos se desarrollarán en formato no-presencial según lo dispuesto en el horario de clases del máster.

Segundo cuatrimestre					
Semana	Temas	Actividades de enseñanza aprendizaje	Horas de trabajo presencial	Horas de trabajo autónomo	Total
Semana 1:	Introducción al análisis de datos masivos	Clase teórica. Laboratorio online (C)	2.00	8.00	10.00
Semana 2:	Modelos de clasificación	Clase teórica. Laboratorio presencial (A)	2.00	7.00	9.00
Semana 3:	Modelos de clasificación	Clase teórica. Laboratorio online (B)	2.00	7.00	9.00
Semana 4:	Modelos de agrupamiento	Clase teórica. Laboratorio online (C)	2.00	8.00	10.00
Semana 5:	Modelos de agrupamiento	Clase teórica. Laboratorio presencial (A)	2.00	8.00	10.00
Semana 6:	Reglas de asociación	Clase teórica. Laboratorio online (B)	2.00	7.00	9.00
Semana 7:	Detección de anomalías	Clase teórica. Laboratorio online (C)	3.00	7.00	10.00

Semana 8:	Gestión de datos a gran escala: el paradigma Map-Reduce	Clase teórica. Laboratorio presencial (A)	2.00	8.00	10.00
Semana 9:	Sistemas de soporte a la gestión de datos a gran escala	Clase teórica. Laboratorio online (B)	2.00	7.00	9.00
Semana 10:	Sistemas de soporte a la gestión de datos a gran escala	Clase teórica. Laboratorio online (C)	2.00	8.00	10.00
Semana 11:	Técnicas y herramientas para la visualización de datos	Clase teórica. Laboratorio presencial (A)	2.00	7.00	9.00
Semana 12:	Técnicas y herramientas para la visualización de datos	Clase teórica. Laboratorio online (B)	2.00	8.00	10.00
Semana 13:	Técnicas y herramientas para la visualización de datos	Clase teórica. Laboratorio online (C)	2.00	7.00	9.00
Semana 14:	Herramientas basadas en software libre para el análisis de datos masivos	Clase teórica. Laboratorio presencial (A)	3.00	7.00	10.00
Semana 15 a 17:	Herramientas basadas en software libre para el análisis de datos masivos	Laboratorio presencial (A)	0.00	16.00	16.00
Total			30.00	120.00	150.00