

Guía docente

Identificación de la asignatura

Asignatura / Grupo	20582 - Análisis de Datos / 9
Titulación	Doble titulación: Grado en Matemáticas y Grado en Ingeniería Telemática - Quinto curso Grado en Matemáticas - Cuarto curso
Créditos	6
Período de impartición	Primer semestre
Idioma de impartición	Castellano

Profesores

Horario de atención a los alumnos

Profesor/a	Hora de inicio	Hora de fin	Día	Fecha inicial	Fecha final	Despacho / Edificio
Ricardo Alberich Martí <i>Responsable</i> r.alberich@uib.es	11:30	12:30	Martes	17/02/2020	16/07/2020	Despacho 169 primera planta Anselm Turmeda
Irene María García Mosquera irene.garcia@uib.es	13:00	14:30	Viernes	03/03/2020	25/06/2020	AT, D-122

Contextualización

Desde el siglo pasado el desarrollo de las tecnologías de la información ha estado generando grandes cantidades de datos. Estos datos, que suelen carecer de estructura, pueden contener información relevante para la toma de decisiones en diversas áreas de conocimiento y/o aplicaciones a la industria y a los negocios. Es por esto que a la actividad tradicional de un matemático dedicado a la estadística se han hecho imprescindible conocimientos adicionales creando la figura del científico de datos.

Así el analista de datos necesita ahora:

- * Capturar, tratar y dotar de estructura a los datos de diversas fuentes utilizando cualquier herramienta informática que sea necesaria.
- * Transformar la información a un modelo de datos que esté preparado para poder ser procesado de forma automática. Hacer el análisis exploratorio de datos, incluyendo su imprescindible visualización.
- * Modelización y puesta en producción de las inferencias, clasificaciones, discriminaciones, predicciones que pueda facilitar el modelo.
- * Comunicación de los resultados, interpretación y puesta en producción.

Así que además de los conocimientos de estadística, se necesitan conocimientos informáticos que posibiliten la manipulación y visualización de resultados.

Este curso presenta al estudiante de matemáticas las herramientas y técnicas modernas para manipular, limpiar, visualizar, analizar y modelar datos multivariantes de gran volumen y formato, incluyendo una introducción a datos de texto y aquellos tomados en el tiempo.

Guía docente

Durante el curso se trabajará con ejemplos reales provenientes de diversas áreas de conocimiento y se espera que el estudiante aprenda a redactar y presentar un informe profesional junto con el modelo de datos y los algoritmos con el objetivo de hacer además desarrollos reproducibles.

Esta asignatura es obligatoria y se imparte en el primer semestre del cuarto curso de Grado de Matemáticas.

Requisitos

Haber cursado con aprovechamiento: Probabilidades II (20320), Inferencia Estadística (20326), Álgebra Lineal II (20311), Cálculo Diferencial en Diversas Variables (20312) y Cálculo Integral en Diversas Variables (20318).

Dominio básico de herramientas de informática.

Recomendables

Dominio a nivel intermedio del programa R.

Competencias

Específicas

- * E40: Desarrollar la capacidad de identificar y describir matemáticamente un problema, de estructurar la información disponible y de seleccionar un modelo matemático adecuado para su resolución.
- * E41: Capacidad de realizar las diferentes etapas en el proceso de modelado matemático: planteamiento del problema, experimentación/pruebas, modelo matemático, simulación/programa, discusión de los resultados y refinamiento/replanteamiento del modelo.
- * E44: Capacidad de manejar, sintetizar, mostrar e interpretar desde el punto de vista de la estadística descriptiva conjuntos de datos.

Genéricas

- * TG2: Desarrollar capacidades de análisis y síntesis, de organización y planificación, y de toma de decisiones.
- * TG3: Capacidad para comunicarse de manera oral o escrita con personas con diferentes niveles de conocimientos en matemáticas.

Básicas

- * Se pueden consultar las competencias básicas que el estudiante tiene que haber adquirido al finalizar el grado en la siguiente dirección: http://estudis.uib.cat/es/grau/comp_basiques/

Contenidos

La asignatura está compuesta por dos bloques:

Guía docente

- * En el primero se hará una introducción a la ciencia de datos: manipulación de volúmenes "masivos" de datos. Dado un conjunto de datos desestructurado o con estructura parcial, los transformaremos en otro conjunto de datos preparado para su procesamiento.
- * En el segundo bloque, daremos una introducción a la estadística multivariante y algunas técnicas de modelización. Por último, se hará breve introducción al tratamiento de series de datos en el tiempo.

Contenidos temáticos

Bloque I. Introducción a la ciencia de datos

- * Herramientas para el tratamiento de datos avanzado.
- * Visualización, transformación, análisis exploratorio, manipulación y limpieza de datos.
- * Tipos especiales de datos: minería de texto.

Bloque II. Análisis multivariado de datos

- * Introducción a series temporales
- * Distribuciones multivariantes
- * Inferencia multivariante
- * Escalado multidimensional
- * Clasificación automática
- * Técnicas de reducción de la factorialidad

Metodología docente

Actividades de carácter presencial:

Asistir a las clases de teoría, realización de talleres y proyectos, presentar resultados.

Actividades de trabajo no presencial:

Resolución y preparación de la exposición de los ejercicios, de las presentaciones, lectura de la bibliografía recomendada, búsqueda en la web de material relevante, experimentación informática, elaboración de programas e informes asociados al proyecto y su presentación.

Con el propósito de favorecer el trabajo del alumno, se ha solicitado que la asignatura forme parte de Aula Digital, dedicada a la enseñanza flexible y a distancia y que incorpora el uso de la telemática en la enseñanza universitaria. Mediante esta plataforma, el alumno tendrá a su disposición un calendario con noticias de interés, documentos electrónicos, problemas y proyectos propuestos tanto para el trabajo individual como en grupo y maneras de comunicarse con los profesores de la asignatura y sus compañeros de clase.

Actividades de trabajo presencial (2,4 créditos, 60 horas)

Modalidad	Nombre	Tip. agr.	Descripción	Horas
Clases teóricas	Introducción temas propuestos	Grupo grande (G)	Introducción teórica a los temas del curso. Se presentarán los conceptos, resultados y procedimientos básicos de la asignatura, así como ejemplos detallados de aplicaciones mediante la exposición magistral. Se trabajarán todas las competencias específicas y la genérica TG2.	15
Seminarios y talleres	Exposición de trabajos	Grupo mediano (M)	Los alumnos harán una exposición de sus trabajos tanto e individual como en grupos.	10
Clases prácticas	Talleres de prácticas	Grupo mediano (M)	Los alumnos resolverán, redactarán y expondrán la solución de ejercicios. Se trabajarán todas las competencias específicas y genéricas.	31



Guía docente

Modalidad	Nombre	Tip. agr.	Descripción	Horas
Evaluación	Exámenes	Grupo grande (G)	Para cada bloque se hará un examen final de bloque. Se evaluará el nivel de asimilación de las competencias.	4

Al inicio del semestre estará a disposición de los estudiantes el cronograma de la asignatura a través de la plataforma UIBdigital. Este cronograma incluirá al menos las fechas en las que se realizarán las pruebas de evaluación continua y las fechas de entrega de los trabajos. Asimismo, el profesor o la profesora informará a los estudiantes si el plan de trabajo de la asignatura se realizará a través del cronograma o mediante otra vía, incluida la plataforma Aula Digital.

Actividades de trabajo no presencial (3,6 créditos, 90 horas)

Modalidad	Nombre	Descripción	Horas
Estudio y trabajo autónomo individual o en grupo	Tareas personales o en grupo	Estudio individual y resolución de ejercicios. Realización de los trabajos propuestos, tanto individual como en grupo. El alumno debe dedicar tiempo tanto al estudio de la teoría como a la resolución de problemas y/o proyectos a lo largo del curso. Se trabajarán todas las competencias específicas y genéricas.	90

Riesgos específicos y medidas de protección

Las actividades de aprendizaje de esta asignatura no conllevan riesgos específicos para la seguridad y salud de los alumnos y, por tanto, no es necesario adoptar medidas de protección especiales.

Evaluación del aprendizaje del estudiante

Fraude en elementos de evaluación

De acuerdo con el artículo 33 del Reglamento Académico, "con independencia del procedimiento disciplinario que se pueda seguir contra el estudiante infractor, la realización demostrablemente fraudulenta de alguno de los elementos de evaluación incluidos en guías docentes de las asignaturas comportará, a criterio del profesor, una minusvaloración en su calificación que puede suponer la calificación de «suspense 0» en la evaluación anual de la asignatura".

Exposición de trabajos

Modalidad	Seminarios y talleres
Técnica	Trabajos y proyectos (no recuperable)
Descripción	Los alumnos harán una exposición de sus trabajos tanto individual como en grupos.
Criterios de evaluación	Las presentaciones de las actividades realizadas en el bloque I, incluyendo la del proyecto de este bloque contará un 10%.

Guía docente

Las presentaciones de las actividades realizadas en el bloque II, incluyendo la del proyecto de este bloque contará un 20%.

Porcentaje de la calificación final: 30%

Talleres de prácticas

Modalidad	Clases prácticas
Técnica	Informes o memorias de prácticas (no recuperable)
Descripción	Los alumnos resolverán, redactarán y expondrán la solución de ejercicios. Se trabajarán todas las competencias específicas y genéricas.
Criterios de evaluación	Los talleres realizados en el bloque 1 tendrán una ponderación del 10%, mientras que los del bloque 2 contarán un 20%.

Porcentaje de la calificación final: 30%

Exámenes

Modalidad	Evaluación
Técnica	Pruebas de respuesta larga, de desarrollo (recuperable)
Descripción	Para cada bloque se hará un examen final de bloque. Se evaluará el nivel de asimilación de las competencias.
Criterios de evaluación	El examen del bloque I contará un 10%. El examen del bloque II contará un 30% y podrá contener material del bloque I. Para aprobar la asignatura se debe obtener como mínimo de 4 en cada examen de cada bloque.

Porcentaje de la calificación final: 40% con calificación mínima 4

Recursos, bibliografía y documentación complementaria

Adicional a los libros recomendados, los profesores irán publicando en Aula Digital los apuntes y listas de problemas de cada tema.

Bibliografía básica

Para el bloque I de la asignatura:

R for Data Science. Garrett Grolemund, Hadley Wickham. Editorial O'Reilly, primera edición. January 2017. Este libro puede consultarse de manera gratuita en <http://r4ds.had.co.nz/>. Si se quiere una copia en papel debe comprarse en amazon.

Para el bloque II de la asignatura:

Nuevos Métodos de Análisis Multivariante. Carles M. Cuadras. CMC Editions, Barcelona. 2014. Revisión 2018 disponible en <http://www.ub.edu/stat/personal/cuadras/cuad2.html>

Bibliografía complementaria

Para el bloque I de la asignatura:

Modern Data Science with R. Benjamin S. Baumer, Daniel T. Kaplan y Nicholas J. Horton. Chapman & Hall / CRC. Texts in Statistical Science Series. 2017.

Para el bloque II de la asignatura:

Análisis Multivariante de Datos. Daniel Peña. Mc. Graw Hill. Febrero 2002.





Guía docente

Mastering Data Analysis with R. Gergely Daróczi. PACKT open source. 2015

Otros recursos

R para Data Science de cero a experto con ejercicios reales. Curso en la plataforma Udemy del Profesor Juan Gabriel Gomila Salas. 2018.

