

Año académico	2016-17
Asignatura	11548 - Análisis de Datos Masivos
Grupo	Grupo 1, 2S
Guía docente	A
Idioma	Castellano

Identificación de la asignatura

Asignatura	11548 - Análisis de Datos Masivos
Créditos	0,72 presenciales (18 horas) 2,28 no presenciales (57 horas) 3 totales (75 horas).
Grupo	Grupo 1, 2S (Campus Extens)
Período de impartición	Segundo semestre
Idioma de impartición	Castellano

Profesores

Profesor/a	Horario de atención a los alumnos					
	Hora de inicio	Hora de fin	Día	Fecha inicial	Fecha final	Despacho
Ricardo Alberich Martí r.alberich@uib.es	08:15	09:30	Lunes	15/09/2016	20/07/2017	156 Anselm Turmeda primer piso
Arnau Mir Torres arnau.mir@uib.es	09:00	10:00	Lunes	01/09/2016	31/01/2017	D-212
Pedro Antonio Palmer Rodríguez pere.palmer@uib.es	17:30	18:30	Viernes	01/09/2016	03/03/2017	D142
	11:30	12:30	Viernes	01/09/2016	30/06/2017	D142
Jairo Enrique Rocha Cárdenas jairo@uib.es	Hay que concertar cita previa con el/la profesor/a para hacer una tutoría					

Contextualización

A nadie se le escapa que la humanidad se dirige hacia un mundo en el que la información se está convirtiendo en un bien que cada vez es más voluminoso. Cualquier búsqueda que se realiza en Internet requiere el análisis de una ingente cantidad de información. La consulta de contenidos en las llamadas redes sociales puede fácilmente generar suficiente información como para que sea realmente difícil no tan solo almacenarla sino incluso poder determinar qué parte es realmente útil, o qué cosas están relacionadas entre sí, y en definitiva qué opción se debe escoger de entre todas las posibilidades que se puedan plantear.

Los grandes volúmenes de información requieren un tratamiento específico debido obviamente a la cantidad de datos a considerar pero también a la existencia de deficiencias en esos datos así como a la falta de coherencia en la información. Por ello se hace necesario utilizar algoritmos diseñados especialmente para poder obtener los resultados deseados. Dichos algoritmos se basan en parte en técnicas derivadas del análisis estadístico de los datos y también de las técnicas de minería de datos.

Está asignatura es una continuación de Gestión y Almacenamiento de Datos Masivos. Usando los algoritmos ya descritos en la asignatura previa, se hará énfasis en las búsquedas en internet, en las páginas web de conexión o de información, en la detección de spam, y en el análisis de estilos de vida de usuarios de internet con el fin

Guía docente

de recomendar a estos posibles consumidores algunos productos comerciales. La técnica del agrupamiento según diferentes criterios se estudiará en profundidad.

Requisitos

Recomendables

Haber aprobado la asignatura Gestión y Almacenamiento de Datos Masivos.

Competencias

Específicas

- * CE10 - Capacidad para comprender y poder aplicar conocimientos avanzados de computación de altas prestaciones y métodos numéricos o computacionales a problemas de ingeniería..
- * CE12 - 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..

Genéricas

- * CG4 - Capacidad para el modelado matemático, cálculo y simulación en centros tecnológicos y de ingeniería de empresa, particularmente en tareas de investigación, desarrollo e innovación, desarrollo e innovación en todos los ámbitos relacionados con la Ingeniería en informática.
- * CG8 - Capacidad para la aplicación de los conocimientos adquiridos y de resolver problemas en entornos nuevos o poco conocidos dentro de contexto más amplios y multidisciplinares, siendo capaces de integrar estos conocimientos..
- * CG9 - Capacidad para comprender y aplicar la responsabilidad ética, la legislación y la deontología profesional de la actividad de la profesión de Ingeniero en Informática..
- * CB6 - Poseer y comprender conocimientos que aporten una base u oportunidad de ser originales en el desarrollo y/o aplicación de ideas, a menudo en un contexto de investigación.
- * CB8 - Que los estudiantes sean capaces de integrar conocimientos y enfrentarse a la complejidad de formular juicios a partir de una información que, siendo incompleta o limitada, incluya reflexiones sobre las responsabilidades sociales y éticas vinculadas a la aplicación de sus conocimientos y juicios.
- * CB9 - Que los estudiantes sepan comunicar sus conclusiones y los conocimientos y razones últimas que las sustentan a públicos especializados y no especializados de un modo claro y sin ambigüedades.
- * CB10 - Que los estudiantes posean las habilidades de aprendizaje que les permitan continuar estudiando de un modo que habrá de ser en gran medida autodirigido o autónomo..

Básicas

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

Contenidos

Guía docente

Contenidos temáticos

1. Tecnologías de motores de búsqueda
Google's PageRank, Detección de Spam, HITS (Hubs and Authorities)
2. Reglas de asociación
Frequent-itemset, market-basket y el algoritmo A-Priori
3. Algoritmos de agrupamiento
Clustering de grandes conjuntos de datos multidimensionales: jerárquico, K-means, CURE, de flujo de datos, paralelismo.
4. Sistemas de recomendación
Ofertas de artículos a lectores de periódicos, etc, basado en predicción de los intereses de los consumidores. Matriz de utilidad, recomendaciones basadas en contenido, filtraje cooperativo.

Metodología docente

Actividades de trabajo presencial

Modalidad	Nombre	Tip. agr.	Descripción	Horas
Clases teóricas	Clase	Grupo grande (G)	Sesiones expositivas, explicativas y/o demostrativas de contenidos (las presentaciones pueden ser a cargo del profesor o de los estudiantes).	12
Seminarios y talleres	Clases especiales	Grupo mediano (M)	Sesiones monográficas supervisadas con participación compartida (profesores, estudiantes o expertos).	2
Clases prácticas	Laboratorio	Grupo grande (G)	Prácticas en el laboratorio.	2
Evaluación	Examen parcial	Grupo grande (G)	Se ha de demostrar la habilidad para implementar o resolver individualmente problemas.	2

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 Campus Extens.

Actividades de trabajo no presencial

Modalidad	Nombre	Descripción	Horas
Estudio y trabajo autónomo individual	estudio individual	Estudiar de libros, preparar exámenes, trabajo en biblioteca, lecturas complementarias, hacer problemas y ejercicios, preparación de seminarios, lecturas, investigaciones, trabajos, memorias, obtención y análisis de datos, etc. para aprender lo visto en clase o exponer o entregar en clase.	40

Guía docente

Modalidad	Nombre	Descripción	Horas
Estudio y trabajo autónomo en grupo	Trabajo en grupo	Las mismas actividades que la modalidad anterior, pero realizadas de forma conjunta, incluye, además, la organización de las reuniones, y la generación de los resultados de la reunión.	17

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

Clases especiales

Modalidad	Seminarios y talleres
Técnica	Pruebas orales (no recuperable)
Descripción	Sesiones monográficas supervisadas con participación compartida (profesores, estudiantes o expertos).
Criterios de evaluación	Calidad y claridad de la presentación

Porcentaje de la calificación final: 5%

Examen parcial

Modalidad	Evaluación
Técnica	Pruebas de respuesta larga, de desarrollo (recuperable)
Descripción	Se ha de demostrar la habilidad para implementar o resolver individualmente problemas.
Criterios de evaluación	Precisión, corrección y eficiencia de las respuestas.

Porcentaje de la calificación final: 50%

estudio individual

Modalidad	Estudio y trabajo autónomo individual
Técnica	Trabajos y proyectos (no recuperable)
Descripción	Estudiar de libros, preparar exámenes, trabajo en biblioteca, lecturas complementarias, hacer problemas y ejercicios, preparación de seminarios, lecturas, investigaciones, trabajos, memorias, obtención y análisis de datos, etc. para aprender lo visto en clase o exponer o entregar en clase.
Criterios de evaluación	Precisión, corrección y eficiencia de las respuestas a los problemas asignados

Porcentaje de la calificación final: 25%



Guía docente

Trabajo en grupo

Modalidad	Estudio y trabajo autónomo en grupo
Técnica	Informes o memorias de prácticas (no recuperable)
Descripción	Las mismas actividades que la modalidad anterior, pero realizadas de forma conjunta, incluye, además, la organización de las reuniones, y la generación de los resultados de la reunión.
Criterios de evaluación	Calidad y claridad del informe escrito
Porcentaje de la calificación final:	20%

Recursos, bibliografía y documentación complementaria

Bibliografía básica

- A. Rajaraman, J. Leskovec, J. Ullman. *Mining of Massive Datasets*. <http://infolab.stanford.edu/~ullman/mmds/book.pdf>
- Zhou, Y. *R and Data Mining: Examples and Case Studies*. <http://www.RDataMining.com>

