

Introducción al Big Data

Código: 104748
Créditos ECTS: 6

Titulación	Tipo	Curso	Semestre
2503873 Comunicación Interactiva	OB	3	2

La metodología docente y la evaluación propuestas en la guía pueden experimentar alguna modificación en función de las restricciones a la presencialidad que impongan las autoridades sanitarias.

Contacto

Nombre: Santiago Giraldo Luque

Correo electrónico: Santiago.Giraldo@uab.cat

Uso de idiomas

Lengua vehicular mayoritaria: español (spa)

Algún grupo íntegramente en inglés: No

Algún grupo íntegramente en catalán: No

Algún grupo íntegramente en español: No

Otras observaciones sobre los idiomas

El curso se realiza en castellano. Alguna parte de la asignatura puede impartirse en catalán o en inglés.

Prerequisitos

La asignatura no tiene requisitos obligatorios, pero se recomienda la superación previa de las siguientes asignaturas del grado:

Sistemas de información

Almacenamiento y Recuperación de Información

Servicios Web Avanzados

Objetivos y contextualización

El objetivo principal de la asignatura es introducir a los y las estudiantes en los conceptos básicos y en las prácticas principales del Big Data.

Asimismo, el curso tiene los siguientes objetivos específicos:

1. Desarrollar un conocimiento propedeútico para el desarrollo posterior de aplicaciones de Business Intelligence: el desarrollo de soluciones de big data para la inteligencia de negocio y su influencia en la toma de decisiones.
2. Introducir los conceptos de fuentes y tipos de datos (estructura, clasificación, integración y calidad).
3. Promover la exploración de las peticiones y el trabajo con fuentes de Open data.
4. Realizar las primeras aproximaciones al análisis de bases de datos en entornos de herramientas de trabajo con hojas de cálculo y otras herramientas prácticas.

Competencias

- Buscar, seleccionar y jerarquizar cualquier tipo de fuente y documento útil para la elaboración de mensajes, trabajos académicos, exposiciones, etc.
- Gestionar el tiempo de forma adecuada y ser capaz de planificar tareas a corto, medio y largo plazos.
- Que los estudiantes hayan desarrollado aquellas habilidades de aprendizaje necesarias para emprender estudios posteriores con un alto grado de autonomía.
- Que los estudiantes puedan transmitir información, ideas, problemas y soluciones a un público tanto especializado como no especializado.
- Que los estudiantes sepan aplicar sus conocimientos a su trabajo o vocación de una forma profesional y posean las competencias que suelen demostrarse por medio de la elaboración y defensa de argumentos y la resolución de problemas dentro de su área de estudio.
- Reconocer y planificar la infraestructura tecnológica necesaria para la creación, el almacenamiento, el análisis y la distribución de productos multimedia interactivos y del internet social.

Resultados de aprendizaje

1. Compartir las experiencias en grupo como forma de aprendizaje para trabajar posteriormente en grupos multidisciplinares.
2. Contrastar y verificar la veracidad de las informaciones aplicando criterios de valoración.
3. Describir la infraestructura necesaria para el almacenamiento del big data.
4. Diferenciar las variedades de tipos de arquitecturas existentes para trabajar con big data.
5. Diferenciar lo sustancial de lo relevante en todos los tipos de documentos de la asignatura.
6. Explicar las características de la infraestructura necesaria para la recuperación del big data.
7. Explicar la infraestructura necesaria para el tratamiento del big data.
8. Planificar y ejecutar proyectos académicos en el ámbito de Big data.
9. Presentar los trabajos de la asignatura en los plazos previstos y mostrando la planificación individual y/o grupal aplicada.
10. Solucionar problemas básicos en big data.

Contenido

Unidad 1. El Big Data: Introducción a la asignatura. Concepto del Big Data, sus procesos y sus características.

Unidad 2. Fuentes, captura y almacenamiento de datos: Presentación de fuentes de datos (principalmente fuentes abiertas), de los procesos de acceso y solicitudes de información pública y leyes de transparencia. Procesos de búsqueda, descarga y almacenamiento de diferentes tipos de datos (formatos).

Unidad 3. Procesamiento y análisis de datos: Manejo de herramientas y funciones de limpieza y análisis de datos para la toma de decisiones.

Unidad 4. Visualización de datos: Presentación de herramientas de visualización de datos para la presentación de informes orientados a la toma de decisión.

Unidad 5. Cartografía con datos: Presentación de diferentes herramientas y posibilidades de representación cartográfica de la información.

(*) El calendario detallado con el contenido de las diferentes sesiones se expondrá el día de presentación de la asignatura. Se colgará también en el Campus Virtual donde el alumnado podrá encontrar la descripción detallada de los ejercicios y prácticas, los diversos materiales docentes y cualquier información necesaria para el adecuado seguimiento de la asignatura. En caso de cambio de modalidad docente por razones sanitarias, el profesorado informará de los cambios que se producirán en la programación de la asignatura y en las metodologías docentes.

El contenido de esta asignatura será sensible a los aspectos relacionados con la perspectiva de género.

Metodología

La estructura de la asignatura, en la que se realizan diferentes actividades prácticas, busca la interiorización de competencias relacionadas con el manejo del Big Data (búsqueda, extracción, análisis y publicación de datos para la toma de decisiones). Su metodología es completamente práctica. A través de las actividades de laboratorio, de talleres y de la realización del trabajo final de curso, se evalúa tanto el componente teórico de la asignatura, como la aplicación práctica de los contenidos estudiados.

La evaluación continua de la asignatura, en la que se realizan prácticas puntuales y continuas de corta duración, permite realizar un seguimiento preciso sobre el aprendizaje y la progresión del estudiante.

Asimismo, se trabaja de forma progresiva sobre la adquisición de los conocimientos que, paso a paso, son involucrados en las prácticas siguientes.

La asignatura de Introducción al Big Data contempla tres tipos o categorías de actividades formativas evaluables:

Prácticas de laboratorio: trabajos individuales o en equipo en los que se realizan actividades prácticas con un entregable puntual con tiempo límite. Los estudiantes deben aplicar los conocimientos, distribuir el tiempo y preparar las entregas dentro del aula y en las horas destinadas a la práctica bajo la guía del profesor.

Seminarios: trabajos individuales o en equipo en los que se realizan actividades prácticas más extensas y con entregables abiertos a la creatividad de los estudiantes. No hay tiempos limitados al aula, pero sí fechas límites de entrega. Los estudiantes deben aplicar los conocimientos, distribuir el tiempo y preparar las entregas iniciando su trabajo dentro del aula, pero continuándolo bajo la modalidad de actividades supervisadas por el equipodocente.

Desarrollo del trabajo final de curso: ejercicio práctico de evaluación grupal en el que los estudiantes deben resolver, durante el desarrollo del curso, un problema de aplicación práctica vinculado al Big Data. Los estudiantes deben plantear el problema y realizar los cuatro procesos para brindar una propuesta de solución basada en grandes cantidades de datos: búsqueda, extracción, análisis y publicación de informe de datos que incluya una propuesta de decisión basada en la información recolectada y analizada.

Nota: se reservarán 15 minutos de una clase dentro del calendario establecido por el centro o por la titulación para que el alumnado rellene las encuestas de evaluación de la actuación del profesorado y de evaluación de la asignatura o módulo.

Actividades

Título	Horas	ECTS	Resultados de aprendizaje
Tipo: Dirigidas			
Clases teóricas	15	0,6	1, 2, 5, 4
Prácticas de laboratorio	33	1,32	1, 2, 3, 5, 4, 7, 6, 8, 9, 10
Tipo: Supervisadas			
Seminarios	10	0,4	1, 2, 3, 5, 4, 7, 6, 8, 9, 10
Tutorías	10	0,4	
Tipo: Autónomas			
Trabajo autónomo: Lecturas, preparación de trabajos y prácticas y estudio personal.	60	2,4	1, 2, 3, 5, 4, 7, 6, 8, 9, 10

Evaluación

Las actividades de Evaluación son:

Actividad A: Prácticas de laboratorio, que tienen un peso del 40% sobre la calificación final

Actividad B: Prácticas de aula, que tienen un peso del 30% sobre la calificación final

Actividad C: Entrega de trabajos, que tiene un peso del 30% sobre la calificación final

Para aprobar la asignatura, es necesario obtener una nota mínima aprobatoria (5,0) en cada una de las actividades.

RECUPERACIÓN:

En las tres últimas semanas del curso los estudiantes que no hayan superado la asignatura podrán presentarse a una prueba de síntesis de reevaluación que consistirá en un test teórico y en un ejercicio práctico. La condición obligatoria para poder optar a la recuperación de la asignatura es haber hecho, como mínimo, 2/3 del total de las prácticas del curso (actividades A, B y C) y haber obtenido una nota media igual o superior a 3,5 (y menor que 5) en todas las actividades de evaluación.

De acuerdo con los criterios señalados anteriormente, si un estudiante no realiza, al menos, el 66% de las prácticas de las actividades de evaluación se considerará como no evaluable de esta asignatura.

PLAGIO:

El estudiante que realice cualquier irregularidad (copia, plagio, suplantación de identidad...) se calificará con 0 este acto de evaluación. En caso de que se produzcan varias irregularidades, la calificación final de la asignatura será 0.

Actividades de evaluación

Título	Peso	Horas	ECTS	Resultados de aprendizaje
Entrega de trabajos	30%	6	0,24	1, 2, 3, 5, 4, 7, 6, 8, 9, 10
Prácticas de aula	30%	8	0,32	1, 2, 3, 5, 4, 7, 6, 8, 9, 10
Prácticas de laboratorio	40%	8	0,32	1, 2, 3, 5, 4, 7, 6, 8, 9, 10

Bibliografía

Alcalde, Ignasi. (2015). *Visualización de la información. De los datos al conocimiento*. Editorial UOC.

Bounegru, Liliana; Gary, Jonathan (Eds.). (2020). *The Data Journalism Handbook II. Towards a Critical Data Practice*. European Journalism Centre and Google News Initiative.

<https://datajournalism.com/read/handbook/two>

Bounegru, Liliana; Chambers, Lucy; Gary, Jonathan (Eds.) (2012). *The Data Journalism Handbook: How Journalists Can Use Data to Improve the News*, O'Reilly Media. <https://datajournalism.com/read/handbook/one>

Bradshaw, Paul. (2017). *Scraping for Journalists. How to grab information from hundreds of sources, put it in data you can interrogate - and still hit deadlines (2nd edition)*. Leanpub

Bradshaw, Paul. (2019). *Finding Stories in Spreadsheets. Recipes for interviewing data - and getting answers*. Leanpub

Bradshaw, Paul., Maseda, Bárbara. (2015). *Periodismo de datos: Un golpe rápido. Cómo entrar, obtener los datos, escabullirse con la noticia... ¡Y asegurarse de que nadie salga herido!* Leanpub.

Cairo, Alberto. (2011). *El arte funcional: Infografía y visualización de información*. Alamut.

Cairo, Alberto. (2016). *The Truthful Art: Data, charts, and maps for communication*. New Riders.

Cairo, Alberto. (2017). ¿Visualización de datos: una imagen puede valer más que mil números, pero no siempre más que mil palabras. *El profesional de la información*, 26(6), 1025-1028.

Carlberg, Conrad. (2011). *Análisis estadístico con Excel*. Anaya.

CARTO (2018). The Top Trends in Data Visualization for 2018. *Medium*.

<https://medium.com/@carto/the-top-trends-in-data-visualization-for-2018-54911e875375>

Charte Ojeda, Francisco (2016). *Excel 2016*. Anaya.

Ferrer-Sapena, Antonia; Sánchez-Pérez, Enrique. (2013). Open data, big data: ¿Hacia dónde nos dirigimos? *Anuario ThinkEPI*, 7, 150-156.

Fuchs, Christian. (2017). "Dallas Smythe Today - The Audience Commodity, the Digital Labour debate, Marxist Political Economy and Critical Theory. Prolegomena to a Digital Labour Theory of Value". En: Fuchs, C., Mosco, V. (Eds.). *Marx and the Political Economy of the Media*. Haymarket Books. pp. 522-599.

Giraldo-Luque Santiago; Fernández-Rovira Cristina (2021) Economy of Attention: Definition and Challenges for the Twenty-First Century. En: Park S.H., Gonzalez-Perez M.A., Floriani D.E. (Eds.). *The Palgrave Handbook of Corporate Sustainability in the Digital Era*. Palgrave Macmillan. pp. 283-305.

Greene, Derek. (2014). Practical Social Network Analysis With Gephi. [Practical Social Network Analysis With Gephi · Derek Greene](#)

Greene, Derek; Cunningham, Pàdraig (2013). Producing a Unified Graph Representation from Multiple Social Network Views. Proc. ACM Web Science'13

Jauset, J.A. (2007). Estadística para periodistas, publicitarios y comunicadores (1 ed.). Catalunya: UOC.

Kauffmann, Erick; Peral, Jesús; Gil, David; Ferrández, Antonio; Sellers, Ricardo; Mora Higinio (2020). A framework for big data analytics in commercial social networks: A case study on sentiment analysis and fake review detection for marketing decision-making, *Industrial Marketing Management*, 90, 523-537.

Mayer-Schönberger, Viktor; Cukier, Kenneth (2013). *Bigdata. La revolución de los datos masivos*. Turner.

Moore, David.S. (2005). *Estadística aplicada básica*. Antoni Bosch.

O'Neil, Cathy. (2017). *Armas de destrucción matemática. Cómo el Big Data aumenta la desigualdad y amenaza la democracia*. Capitan Swing.

Patino, Bruno (2020). *La civilización de la memoria de pez. Pequeño tratado sobre el mercado de la atención*. Alianza.

Mierzejewski, Mikołaj; Lee, Maggie; Grün, Gianna-Carina; Bradshaw, Paul; Panuccio, Erika; Aldhous, Peter (2018). Data scraping for stories. Conversations with Data: #12.

<https://datajournalism.com/read/newsletters/data-scraping-for-stories>

Tascón, Mario (2013). Introducción. Big Data. Pasado, presente, futuro. *Telos: Cuadernos de comunicación e innovación*, 95, 47-50.

Turing, Alan. (1974). *¿Puede pensar una máquina?* Universidad de Valencia.

Software

Al tratarse de una asignatura completamente práctica el software requerido es el habitual para la realización de las tareas de captura, tratamiento y análisis de información en diferentes formatos.

En concreto, se requieren de las siguientes herramientas:

Software de edición de textos: Word o similar

Software de análisis de datos: Excel o similar

Software de visualización de datos: Infogram - Datawrapper - CARTO - Gephi

Como la asignatura puede realizar algunas sesiones virtuales, los estudiantes deben tener acceso también a un ordenador con cámara y micrófono y conexión a Internet para poder participar de las sesiones on-line.