



Ciencia de Datos: Un Enfoque Práctico en la Era del Big Data

(7ª edición – Julio 2021)

<http://sci2s.ugr.es/CienciaDatosBigData>

Introducción

Ciencia de datos es un área de trabajo interdisciplinar que incluye procesos para recopilar, preparar, analizar, visualizar y modelizar datos que permitan generar conocimiento útil para comprender problemas complejos y ayudar en la toma de decisiones. Estos datos con frecuencia son no estructurados y heterogéneos. En muchas ocasiones, se trata de grandes volúmenes de datos que por su complejidad y diversidad requiere de arquitecturas y técnicas innovadoras para extraer conocimiento relevante: es el conocido *big data*. Ciencia de datos es un campo emergente con una alta aplicabilidad en ciencias de la salud, marketing, negocios, mercados financieros, transporte, comunicaciones, redes sociales, etc.

Como indica la consultora Gartner (la más prestigiosa en tecnologías de la información), los científicos de datos no son analistas de negocio tradicionales, son profesionales con la rara capacidad para obtener modelos matemáticos a partir de datos que generan beneficios empresariales claros y contundentes. Así, cada vez más se exigen profesionales con habilidades en campos como informática, matemáticas, estadística o negocios que dominen las nuevas tecnologías y sepan gestionar datos. Las empresas de todos los sectores están adoptando cada vez más la ciencia de datos, de modo que la demanda de expertos en este sector es enorme; así lo refleja un estudio del [MIT Sloan Management Review \(2015\)](#), está considerada como una de las mejores oportunidades laborales de los próximos años (no en vano, Glassdoor la define como la [mejor profesión en 2016](#)). Catalogada por el Harvard Business Review como la profesión '[más sexy del siglo XXI](#)' (2012). Según un [estudio](#) basado en información de LinkedIn (2015), el número de profesionales en ciencia de datos se ha duplicado en los últimos cuatro años. Otro estudio de [Burtch Works \(2015\)](#) reconoce el impacto positivo en el salario al incluir conocimientos de ciencia de datos. El último informe sobre el futuro del trabajo elaborado por el [Foro Económico Mundial \(2018\)](#) sitúa a los analistas y científicos de datos entre los perfiles laborales que más crecimiento de demanda experimentarán hasta 2022.

Objetivos

Los planes de formación universitaria reglada difícilmente reaccionan a oportunidades laborales emergentes. Además, se tiende a delimitar fronteras que dificultan el desarrollo de especialidades híbridas. Este curso pretende iniciar al alumno en el campo de ciencia de datos, sirviendo así de puente entre diversas disciplinas y ayudando a completar la formación universitaria con una orientación eminentemente práctica. El curso se compone de 30 horas lectivas telepresenciales repartidas en 13 horas de conceptos teóricos y fundamentos y otras 17 horas de prácticas con *software* especializado y datos de casos reales.

La teoría incluye visualización de datos, técnicas de clasificación básicas (árboles de decisión, redes neuronales...) y avanzadas (máquinas de soporte vectorial, *ensemble learning*, *deep learning*...), preprocesado (eliminación de ruido, imputación de valores perdidos, reducción de datos...), aprendizaje no supervisado (agrupamiento y reglas de asociación), aprendizaje incremental y minería de flujo de datos, *big data* y sus paradigmas y, finalmente, experiencias reales de ciencia de datos en la empresa. La práctica introduce al alumno en herramientas *software* tales como KNIME y Python y arquitecturas *big data* como Spark. También se adquirirá experiencia en competiciones con problemas reales (Kaggle, DrivenData...).

A quién va dirigido

Las personas que se dedican a la ciencia de datos se conocen como científicos de datos, que no es más que una mezcla de matemáticos, estadísticos, informáticos y creativos con habilidades para recopilar, procesar y extraer valor de las diversas y extensas bases de datos; imaginación para comprender, visualizar y comunicar sus conclusiones a los no científicos de datos; y capacidad para crear soluciones basadas en datos que aumentan los beneficios, reducen los costos y ayudan a construir un mundo mejor.

El curso se orienta a estudiantes de grado, máster y profesionales con formación previa principalmente en informática, matemáticas, estadística e ingeniería que busquen completar su formación como científico de datos. No obstante, la presentación de los fundamentos teóricos y el uso de *software* especializado se impartirán de forma apropiada para atender a las diferentes necesidades y formación previa del alumnado. Ciencia de datos es una disciplina que se nutre de experiencias y formaciones diversas, de forma que el curso aprovechará la variedad de intereses y capacidades del alumnado.

Equipo docente

El profesorado lo componen docentes e investigadores universitarios séniores y jóvenes del área de Ciencias de la Computación e Inteligencia Artificial de la Universidad de Granada. Se trata de personal altamente especializado en ciencia de datos con excelentes trayectorias en investigación. En el área de Ingeniería y Ciencias de la Computación, la Universidad de Granada está considerada según el prestigioso *ranking* [ARWU 2019 de Shanghái](#) entre las 15 mejores de Europa y primera de España.

Equipo docente: Jorge Casillas (coordinador)
Alberto Fernández, Salvador García, Diego J. García, Julián Luengo, Daniel Molina

Contenidos de Teoría (13h)

1. Ciencia de datos, analítica avanzada y *big data* (1h) – **Jorge Casillas**
2. Análisis exploratorio de datos: visualización (1h) – **Jorge Casillas**
3. Fundamentos de clasificación: árboles de decisión, *lazy*, RNA, bayesianos, evaluación (2h) – **Salvador García**
4. Preprocesamiento: selección y procesado de instancias y características, tratamiento del ruido (2h) – **Salvador García**
5. Clasificación avanzada: SVM, *ensemble learning*, problemas no balanceados, *deep learning* (2,5h) – **Alberto Fernández**
6. Segmentación y relaciones: *clustering* y reglas de asociación (2h) – **Jorge Casillas**
7. Aprendizaje incremental y *data stream mining* (1h) – **Jorge Casillas**
8. *Big data*: fundamentos y paradigmas (1,5h) – **Alberto Fernández**

Contenidos de Prácticas (17h)

1. KNIME (5,5h): predicción fundamental – **Julián Luengo**
2. Python para ciencia de datos (6,5h): visualización y predicción avanzada – **Daniel Molina**
3. Spark y MLLib (5h): *big data* – **Diego J. García Gil**

Planificación

Julio 2021	
Sesión 1 - miércoles, 30/06/2021 15:30-16:00 – Presentación del curso 16:00-17:00 (1h) – Teoría 1 17:00-18:00 (1h) – Teoría 2 18:00-18:30 (30m) – <i>descanso</i> 18:30-20:30 (2h) – Teoría 3	Sesión 5 – viernes, 09/07/2021 16:00-18:00 (2h) – Teoría 6 18:00-18:30 (30m) – <i>descanso</i> 18:30-19:30 (1h) – Teoría 7
Sesión 2 - viernes, 02/07/2021 16:00-18:00 (2h) – Prácticas 1 18:00-18:30 (30m) – <i>descanso</i> 18:30-20:00 (1,5h) – Prácticas 1	Sesión 6 - lunes, 12/07/2021 15:30-18:00 (2,5h) – Prácticas 2 18:00-18:30 (30m) – <i>descanso</i> 18:30-20:30 (2h) – Prácticas 2
Sesión 3 - lunes, 05/07/2021 16:00-18:00 (2h) – Teoría 4 18:00-18:30 (30m) – <i>descanso</i> 18:30-20:30 (2h) – Prácticas 1	Sesión 7 - miércoles, 14/07/2021 16:00-17:30 (1,5h) – Teoría 8 17:30-18:00 (30m) – <i>descanso</i> 18:00-20:30 (2,5h) – Prácticas 3
Sesión 4 – miércoles, 07/07/2021 15:30-18:00 (2,5h) – Teoría 5 18:00-18:30 (30m) – <i>descanso</i> 18:30-20:30 (2h) – Prácticas 2	Sesión 8 - viernes, 16/07/2021 16:00-18:30 (2,5h) – Prácticas 3 18:30-19:00 (30m) – <i>descanso</i> 19:00-19:30 (30m) – Kaggle/DrivenData 19:30-19:40 (10m) – Clausura

Asistencia: Obligatorio asistir al 80% de las clases

Evaluación: Respuesta a preguntas de teoría (modalidad telemática asincrónica) y competición internacional en una plataforma de ciencia de datos ([Kaggle](#) o [DrivenData](#))

Lugar de celebración

ETS de Ingenierías Informática y de
Telecomunicación – Universidad de Granada

<http://etsiit.ugr.es>

<https://goo.gl/maps/8N1zEYajC9k> 

Las prácticas se desarrollarán en una sala
de ordenadores con 8 GB de RAM e Intel® Core™ i5



Número de plazas: 30

Precio: 110 € (10% de becas)

Más información: Centro Mediterráneo, <https://cemed.ugr.es/curso/21on07/>

Web del curso: <http://sci2s.ugr.es/CienciaDatosBigData>



Reconocimiento de créditos optativos pendiente de aprobación. En la anterior edición fue así: 3 ECTS en todos los Grados de la Facultad de Comunicación y Documentación; 2 ECTS en todos los Grados de la Facultad de Ciencias; 3 ECTS en todos los Grados de la ETS Ing. Caminos, Canales y Puertos; 3 ECTS en todos los Grados de la ETS Ing. Informática y de Telecomunicación; 3 ECTS en todos los Grados impartidos en la ETS Ing. Edificación