

Introducción a la ciencia de datos para profesionales de la información: teoría y aplicaciones prácticas

Resumen

<i>Fechas</i>	14-18 Junio
<i>Modalidad docente</i>	Virtual, síncrono y asíncrono
<i>Medios</i>	Google Meet, Moodle y Telegram
<i>Horas</i>	15 h., 10 teóricas y 5 prácticas
<i>Evaluación</i>	Entrega de prácticas
<i>Emisión certificado</i>	Sí
<i>Número de profesores</i>	2
<i>Precio</i>	250 Euros
<i>Número de plazas</i>	50

Introducción

El big data ha traído consigo nuevas formas de trabajar debido a la dificultad de tratar con grandes volúmenes de datos, produciéndose constantemente cambios que nos obligan a mantenernos en continua formación. Este curso tiene como objetivo **ofrecer una introducción de los principales avances producidos en el ámbito de la ciencia de datos bajo una perspectiva práctica y considerando las múltiples herramientas disponibles**. Se aborda con ello los principales **entornos de trabajo** y elementos fundamentales que en ellos intervienen, las nuevas fuentes de información y que implican nuevas maneras de **recuperar datos y tratarlos**, y diferentes formas de **analizarlos y generar visualizaciones**. En este curso, por tanto, tratamos de ofrecer un punto de entrada a la ciencia de datos y recientes novedades, **estando dirigido a profesionales de la información que desean actualizar sus conocimientos y ponerlos al día**. Otro de nuestros objetivos es celebrar periódicamente este curso pero avanzando en los contenidos anteriores de forma que realmente podamos ofrecer un **programa de formación periódico y progresivo** centrado en las nuevas tendencias y métodos.

Método docente

Nuestro objetivo principal es ofertar un curso dinámico **centrado en clases virtuales emitidas por streaming**. Este curso se basa en lecciones magistrales donde podrás conocer e interactuar con los profesores. El curso cuenta con **contenidos teóricos distribuidos en tres bloques acompañados cada uno de ellos de sus correspondientes ejercicios prácticos tutelados**. En relación a las clases prácticas, en las mismas, el profesor expondrá la práctica y los ejercicios a realizar y durante el tiempo de la clase práctica se mantendrá el aula virtual abierta de manera que se puedan ir resolviendo las dudas que vayan surgiendo. No es necesario finalizar las prácticas durante la clase ya que una vez concluido el curso el alumno dispondrá de una semana para su entrega. Finalmente es importante indicar que **el curso se podrá seguir de manera síncrona**, es decir asistiendo a las clases en directo, **pero también podrá seguirse de manera asíncrona** por todos aquellos alumnos que no puedan seguir las emisiones. Asimismo el alumno podrá adoptar un modelo mixto, es decir seguir

algunas clases online y otras en diferido. En todos los casos el proceso de evaluación para la obtención del certificado será el mismo.

Medios

Las clases se emitirán por streaming a través de la **plataforma Google Meet** y quedarán asimismo registradas. El curso contará con una **plataforma moodle propia** en la cual los alumnos tendrán **acceso individualizado a los diferentes materiales** y recursos (vídeos, presentaciones, lecturas, enlaces y prácticas). Contaremos asimismo con un **foro dentro del moodle** y un **grupo de trabajo en Telegram** específico para la resolución rápida y dinámica de dudas.

Evaluación

Para la evaluación de los conocimientos adquiridos por el alumnado durante el curso se realizarán dos **bloques de ejercicios prácticos** que se podrán entregar hasta una semana después de la finalización del curso. Cada práctica cuenta un 50% de la puntuación final.

Certificados

Al finalizar el curso se entregará un certificado numerado de aprovechamiento indicándose las horas, la calificación y los contenidos impartidos

Inscripciones

El total de plazas que **ofertamos es de 50 a un precio de 250 euros** (IVA incluido). La plaza es personal e intransferible. **Las inscripciones se cerrarán el 10 de junio**. 48 horas antes del inicio del curso se entregarán a los inscritos las credenciales para el acceso a la plataforma moodle, asimismo se podrán solicitar pruebas de conexión a la plataforma de streaming para verificar su correcto funcionamiento. El pago se realizará a través de la plataforma online de pagos de ec3metrics mediante tarjeta de débito o crédito. En el

caso de no poder realizar el pago **mediante tarjeta** se podrá realizar **mediante transferencia bancaria**, en tal caso debe el alumno de ponerse en contacto con la organización del curso a través del correo investigacion@ec3metrics.com. Si es necesario **podremos emitir una factura para justificar la inscripción** ante sus instituciones.

IMPORTANTE: Una vez realizado el pago envíanos un correo con el justificante de pago y tus datos personales (nombre, apellidos, DNI o equivalente e institución) al correo: investigacion@ec3metrics.com

Profesorado

El curso Introducción a la ciencia de datos para profesionales de la información está impartido por diferentes profesionales, investigadores y profesores especializados y activos en ciencia de datos. Conoce a los profesores:

- **Arroyo-Machado, Wenceslao** – [Más Info](#) – [Google Scholar Profile](#). Graduado en Información y Documentación y máster en Ciencia de Datos e Ingeniería de Computadores, con especialidad en Ciencia de Datos, en la UGR. Actualmente estudiante con el apoyo de una FPU. Su actividad científica se centra en técnicas de Big Data aplicadas a la informetría.
- **Mariottini, Sara** - [Más Info](#) - Graduada en Sociología en la Università di Roma "Sapienza", Máster en Ciencias Sociales Aplicadas en la misma y Máster en Ciencia de Datos e Ingeniería de Computadores, con especialidad en Ciencia de Datos, en la UGR. Actualmente estudiante con FPI en el Departamento de Información y Documentación de la UGR. Su actividad científica y interés se enfocan en la aplicación y desarrollo de técnicas de Big Data aplicadas al procesamiento del lenguaje natural (NLP).

Programa

Bloque I: Introducción a la ciencia de datos (3 h.)

- Introducción a la ciencia de datos. Wenceslao Arroyo y Sara Mariottini (2 horas)
- Ejercicios prácticos: 1 hora

Bloque II: Datos, recuperación y tratamiento (6 h.)

- Minería de datos, del Big Data al Small Data. Sara Mariottini (2 hora)
- Extracción de datos: APIs y scraping. Wenceslao Arroyo (2 horas)
- Ejercicios prácticos: 2 horas

Bloque III: Técnicas de análisis y visualización (6 h.)

- Procesamiento del Lenguaje Natural. Sara Mariottini (2 horas)
- Visualizaciones. Wenceslao Arroyo (2 horas)
- Ejercicios prácticos: 2 horas

	Lunes 14 Junio	Martes 15 Junio	Miércoles 16 Junio	Jueves 17 Junio	Viernes 18 Junio
	Presentación				
9:30-11:30	Bloque I Introducción	Bloque II Minería de datos	Bloque II Extracción de datos	Bloque III PLN	Bloque III Visualizaciones
11:30-12:00	DESCANSO	DESCANSO	DESCANSO	DESCANSO	DESCANSO
12:00-13:00	Preparación del entorno	PRÁCTICA 1	PRÁCTICA 1	PRÁCTICA 2	PRÁCTICA 2

Cuestionario Clausura

Resúmenes

Introducción a la ciencia de datos. Son muchas las novedades que ha traído consigo la ciencia de datos en los últimos años, cambiando por completo la forma de trabajar. En este curso hacemos una introducción a la ciencia de datos, repasando algunos de los conceptos clave y principales tendencias. De igual manera, se van a repasar las herramientas y los elementos más destacados que intervienen en dichos entornos de trabajo.

Minería de datos, del Big Data al Small Data. Uno de los procesos que más tiempo conlleva y que resulta clave para el análisis de los datos está en la lectura y preprocesamiento de los mismos. Es por ello que aquí veremos cómo realizar una correcta importación de datos a nuestro entorno de trabajo, desde Excel a Python, teniendo en cuenta los principales y más comunes problemas para ello. De igual manera, se presentarán las transformaciones más destacadas para realizar el preprocesamiento de los datos con el fin de garantizar su calidad.

Extracción de datos: APIs y scraping. Además de nuevas formas de tratar los datos, también han surgido nuevas formas de acceder a estos. Una de las principales y más eficaces herramientas para la recopilación de datos son las APIs (Application Programming Interfaces). De igual manera, también existe la posibilidad de recuperar datos rastreando de manera automatizada sitios web, el denominado web scraping. En este curso se llevará a cabo un acercamiento teórico y práctico a ambos procesos de recuperación de datos mediante R, Python y plugins de navegadores web.

Procesamiento del Lenguaje Natural. En este curso se va a realizar una introducción práctica a la minería de textos. Se introducirá el paquete de procesamiento de texto en Python (NLTK), así como el software Knime. De igual manera, se dará al alumnado las herramientas necesarias para la visualización gráfica de los datos de texto en Python.

Visualizaciones. Son muchas las tipologías de visualizaciones que existen para representar los datos, siendo un aspecto clave escoger la que mejor se ajusta a cada caso. En este curso se introducirán los aspectos fundamentales para trabajar con visualizaciones a partir de datos en bruto en R y Tableau.