



Universidad
de Alcalá

GUÍA DOCENTE

Fundamentos de la Ciencia de Datos

**Grado en
Ingeniería Informática**

Universidad de Alcalá

Curso Académico 2022/2023

4º Curso - 1^{er} Cuatrimestre

GUÍA DOCENTE

Nombre de la asignatura:	Fundamentos de la Ciencia de Datos
Código:	781006
Titulación en la que se imparte:	Grado en Ingeniería Informática
Departamento y Área de Conocimiento:	Ciencias de la Computación Ciencias de la computación e inteligencia artificial
Carácter:	Obligatoria
Créditos ECTS:	6.0
Curso y cuatrimestre:	4º Curso, 1^{er} Cuatrimestre
Profesorado:	Juan J. Cuadrado Gallego El listado completo actualizado de profesores para el curso 2021/2022 se podrá encontrar en la página web de la asignatura: https://www.uah.es/es/estudios/estudios-oficiales/grados/asignatura/Fundamentos-de-la-Ciencia-de-Datos-781006/
Horario de Tutoría:	Consultar al comienzo de la asignatura
Idioma en el que se imparte:	Español/English Friendly

1a. PRESENTACIÓN

En los planes de estudios el grado en Ingeniería Informática está estructurado en tres bloques de formación: Básica, Obligatoria y Optativa. Los bloques de formación básica y obligatoria cubren el cuerpo de conocimiento principal propuesto por los informes o guías curriculares: *Computing Curricula: Computer Science 2001*, *Computer Curricula Software Engineering 2004*, *Computing Curricula 2005* de ACM-IEEE, *Computing Engineering de 2004* y la *Guía IS de 2002* de la AIS. La materia *Fundamentos de la Ciencia de Datos* se encuadra dentro del bloque de Formación Obligatoria y consta de 6 ECTS, todos ellos se cursan en una única asignatura con el mismo nombre que la materia. La asignatura tiene por objetivo que el alumno adquiera los conocimientos correspondientes al cuerpo de conocimiento correspondiente a la *Ciencia de Datos*, en inglés *Data Science*.

El cuerpo de conocimiento utilizado para establecer los contenidos de la asignatura *Fundamentos de la Ciencia de Datos* ha sido el desarrollado en el proyecto EDISON, dicho proyecto fue financiado por la Unión Europea en 2015 para establecer los fundamentos de la *Data Science*, incluido la definición del cuerpo de conocimiento y el curriculum de la misma, cuyos resultados se publicaron en 2020 en el libro de la editorial Springer *The Data Science Framework: A View from the EDISON Project*. Para introducir sobre que versan dichos contenidos se pueden tomar las definiciones que se dan en el libro para la *Data Science* o *Ciencia de Datos* puede definirse como "*La Ciencia de Datos es una disciplina compleja que usa abstracciones y modelos conceptuales y matemáticos, y métodos estadísticos, junto con modernas herramientas computacionales para obtener conocimiento desde los datos (descubrir correlaciones y causas en los datos del dominio) para apoyar a la toma de decisiones tanto en la investigación científica como en las actividades de negocio*" y si tuviera que ser definida en una sola frase en el libro se define la *Ciencia de Datos* como "*Ciencia que estudia cómo obtener conocimiento desde los datos*".

Los conocimientos que los alumnos obtienen cursando la asignatura *Fundamentos de la Ciencia de Datos* son de la máxima actualidad e interés ya que les proporcionarán los fundamentos de la *Ciencia de Datos* que les permitan, mediante el desarrollo y la profundización de los mismos, convertirse en un Científico de Datos, perfil para el que, según las consultoras más prestigiosas del sector, se prevé que se crearán 1.2 millones de empleos en los próximos años.

Prerrequisitos y Recomendaciones

Para la asignatura *Fundamentos de la Ciencia de Datos* se recomienda haber superado las materias del grado básicas y obligatorias sobre Estadística, Matemáticas, Algoritmia y Programación.

1b. COURSE SUMMARY

In the study plans, the degree in Ingeniería Informática is structured in the three formative blocks: Basic, Mandatory and Optative. The basic and mandatory formative blocks cover the main body of knowledge proposed by the reports or curricular guides: *Computing Curricula: Computer Science 2001*, *Computer Curricula Software Engineering 2004*, *Computing Curricula 2005* from ACM-IEEE, *Computing Engineering of 2004* and the *2002 IS Guide* from the AIS. The matter *Basics of Data Science* has 6 ECTS credits and all of them are taught in a single subject with the same name as the matter. The subject has by objective that the student acquires the knowledge corresponding to the body of knowledge corresponding to *Data Science*.

The body of knowledge used to establish the contents of the subject *Basics of Data Science* has been developed in the EDISON project, said project was founded by the European Union in 2015 to establish the foundations of *Data Science*, including the definition of the body of knowledge and its curriculum, the results of which were published in 2020 in the book from Springer publisher *The Data Science Framework: A View from the EDISON project*. To introduce what these contents are about, the

definitions given in the book and Data Science can be defined as "Data Science is a complex discipline that uses conceptual and mathematical abstractions and models, statistical methods, together with modern computational tools to obtain knowledge/derive insight from data to (uncover correlations and causations in business data) support decision making in scientific research and business activity" and if it was be defined only with one sentence: "Science that studies how to obtain knowledge from data.

The knowledge that students obtain attending the course Basics of Data Science are of the most current and interest since it will provide them, through the development and deepening of the same, to become a Data Scientist, a profile for which, according to the most prestigious consulting firms in the sector, it is expected 1.2 million jobs will be created in the coming years.

Prerequisites and Recomendations

For the subject Basics on Data Science it is recommended to have passed the subjects basic and mandatory of the degree on Statistics, Mathematics, Algorithm and Programming.

2. COMPETENCIAS

Competencias básicas, generales y transversales.

Esta asignatura contribuye a adquirir las siguientes competencias básicas, generales y transversales:

CB1 - Que los estudiantes hayan demostrado poseer y comprender conocimientos en un área de estudio que parte de la base de la educación secundaria general, y se suele encontrar a un nivel que, si bien se apoya en libros de texto avanzados, incluye también algunos aspectos que implican conocimientos procedentes de la vanguardia de su campo de estudio.

CB2 - 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.

CB3 - Que los estudiantes tengan la capacidad de reunir e interpretar datos relevantes (normalmente dentro de su área de estudio) para emitir juicios que incluyan una reflexión sobre temas relevantes de índole social, científica o ética.

CB4 - Que los estudiantes puedan transmitir información, ideas, problemas y soluciones a un público tanto especializado como no especializado.

CB5 - Que los estudiantes hayan desarrollado aquellas habilidades de aprendizaje necesarias para emprender estudios posteriores con un alto grado de autonomía.

CG8 - Conocimiento de las materias básicas y tecnologías, que capaciten para el aprendizaje y desarrollo de nuevos métodos y tecnologías, así como las que les doten de una gran versatilidad para adaptarse a nuevas situaciones.

CG9 - Capacidad para resolver problemas con iniciativa, toma de decisiones, autonomía y creatividad. Capacidad para saber comunicar y transmitir los conocimientos, habilidades y destrezas de la profesión de Ingeniero Técnico en Informática.

TRU1 - Capacidad de análisis y síntesis.

TRU2 - Comunicación oral y escrita.

TRU3 - Capacidad de gestión de la información.

TRU4 - Capacidad de aprendizaje autónomo.

TRU5 - Capacidad para trabajar en equipo.

Competencias Específicas

Esta asignatura proporciona la(s) siguiente(s) competencia(s) específica(s):

CC4 - Capacidad para conocer los fundamentos, paradigmas y técnicas propias de los sistemas inteligentes y analizar, diseñar y construir sistemas, servicios y aplicaciones informáticas que utilicen dichas técnicas en cualquier ámbito de aplicación.

CC5 - Capacidad para adquirir, obtener, formalizar y representar el conocimiento humano en una forma computable para la resolución de problemas mediante un sistema informático en cualquier ámbito de aplicación, particularmente los relacionados con aspectos de computación, percepción y actuación en ambientes o entornos inteligentes.

CC7 - Capacidad para conocer y desarrollar técnicas de aprendizaje computacional y diseñar e implementar aplicaciones y sistemas que las utilicen, incluyendo las dedicadas a extracción automática de información y conocimiento a partir de grandes volúmenes de datos.

Resultados de aprendizaje

Al terminar con éxito esta asignatura/enseñanza, los estudiantes serán capaces de:

RA1. Conocer las diferencias entre aprendizaje supervisado y no supervisado además de su contexto de aplicación.

RA2. Ser capaz de modelizar y resolver problemas de aprendizaje mediante paradigmas computacionales (supervisado y no supervisado).

RA3. Ser capaz de explicar los modelos aplicados y resultados obtenidos del aprendizaje, incluyendo sus limitaciones y bondades.

RA4. Ser capaz de entender diferentes tipos de datos, problemas y posibles transformaciones para su aplicación tanto en el aprendizaje como en la evaluación de los resultados.

3. CONTENIDOS

Bloques de contenido	Total de clases, créditos u horas
Lección 1. Introducción a la Ciencia de Datos	4 h
Lección 2. Datos	12 h
Lección 3. Análisis de Asociación	8 h
Lección 4. Detección de Datos Anómalos	8 h
Lección 5. Análisis de Clasificación No Supervisada	8 h
Lección 6. Análisis de Clasificación Supervisada	10 h

4. METODOLOGÍAS DE ENSEÑANZA APRENDIZAJE. ACTIVIDADES FORMATIVAS

4.1. Distribución de créditos (especificar en horas)

Número de horas presenciales:	58 horas (50 horas de clase presencial +8 horas de evaluación)
Número de horas del trabajo propio del estudiante:	92 (Incluye horas de estudio, elaboración de actividades, preparación de exámenes)
Total horas	150

4.2. Estrategias metodológicas, materiales y recursos didácticos

Clases magistrales y expositivas, en combinación con prácticas en el laboratorio	<p>Las clases presenciales se realizarán en el aula y en el laboratorio. Habrá dos tipos de sesiones:</p> <ul style="list-style-type: none"> ■ Sin utilización de software: Se realizarán en el aula y se presentarán los contenidos teóricos de la asignatura, mediante herramientas de presentación de contenidos y se resolverán casos prácticos y ejercicios que complementarán los conocimientos teóricos. ■ Con utilización de software: Se realizarán en el laboratorio y se resolverán, mediante la utilización de software para Data Science, los mismos casos prácticos y ejercicios que se hayan resuelto en la clase previa en el aula son la utilización de software.
Trabajos en grupo y cooperativo	<p>La enseñanza práctica, mediante la resolución de cuestiones, ejercicios y prácticas (Pruebas de Laboratorio), tendrá un papel fundamental, como estrategia metodológica, en la impartición de la asignatura. En consecuencia se pedirá al alumno un alto grado de implicación y participación en la resolución de las cuestiones y ejercicios, de forma individual, pero también en grupo, sobre todo en la realización de las prácticas o pruebas de laboratorio, que se deberán resolver en grupo. Además, como parte de dicho trabajo, podrán realizarse presentaciones orales por los grupos de alumnos del trabajo realizado, a sus compañeros, del trabajo realizado por el grupo.</p>

Trabajo y estudio personal

Dentro del trabajo y estudio personal se realizarán las siguientes actividades:

- Estudio de los conceptos teóricos impartidos.
- Resolución de ejercicios. Esta será la actividad fundamental que el alumno deberá realizar como trabajo y estudio personal. Será muy importante que el alumno dedique suficiente esfuerzo al trabajo y estudio personal, ya que de su correcta ejecución dependerá en gran medida el aprovechamiento de las clases presenciales.
- Resolución de cuestiones con el entorno software utilizado.
- Realización de resúmenes textuales y esquemas gráficos sobre los conceptos impartidos.
- Consulta de fuentes y recursos bibliográficos físicos y electrónicos.

Tanto las clases presenciales como el trabajo autónomo estarán apoyados por tutorías, en el horario oficial establecido por la asignatura, que podrán realizarse tanto de forma individual como en grupo. Además, los estudiantes podrán solicitar tutorías fuera del horario oficial establecido para las mismas, en los casos en los que no hayan podido acudir a las mismas.

5. EVALUACIÓN: Procedimientos, criterios de evaluación y calificación

Preferentemente se ofrecerá a los alumnos un sistema de evaluación continua que tenga características de evaluación formativa de manera que sirva de realimentación en el proceso de enseñanza-aprendizaje por parte del alumno.

5.1. PROCEDIMIENTOS

La evaluación debe estar inspirada en los criterios de evaluación continua (Normativa de Evaluación de los Aprendizajes, NEA, art 3). No obstante, respetando la normativa de la Universidad de Alcalá se pone a disposición del alumno un proceso alternativo de evaluación final de acuerdo a la [Normativa de Evaluación de los Aprendizajes](#) según lo indicado en su Artículo 10, los alumnos tendrán un plazo de quince días desde el inicio del curso para solicitar por escrito al Director de la Escuela Politécnica Superior su intención de acogerse al modelo de evaluación no continua aduciendo las razones que estimen convenientes. La evaluación del proceso de aprendizaje de todos los alumnos que no cursen solicitud al respecto o vean denegada la misma se realizará, por defecto, de acuerdo al modelo de evaluación continua. El estudiante dispone de dos convocatorias para superar la asignatura, una ordinaria y otra extraordinaria.

Convocatoria ordinaria

Evaluación continua:

El proceso de evaluación continua utilizará diferentes estrategias y recogerá, a lo largo del cuatrimestre, evidencias que guardan relación con todo el proceso de enseñanza-aprendizaje durante la impartición de la asignatura. Dichas evidencias consistirán en las siguientes pruebas:

1. **Prueba de Evaluación Continua 1 (PEC1)** Consistirá en preguntas y ejercicios escritos basados

en los contenidos de las lecciones 1, 2 y 3; y podrá incluir también preguntas del entorno software utilizado para tratar esos temas. La prueba será escrita sin ordenador.

2. **Prueba de Evaluación Continua 2 (PEC2)**. Consistirá en preguntas y ejercicios escritos basados en los contenidos de las lecciones 4, 5 y 6; y podrá incluir también preguntas del entorno software utilizado para tratar esos temas. La prueba será escrita sin ordenador.
3. **Prueba de Laboratorio 1 (PL1)**. Consistirá en la realización de una práctica, que se realizará con el software de Data Science que se haya estudiado en la asignatura, y que estará basada en los contenidos visto en las lecciones 1, 2 y 3. La PL1 podrá tener asociada una prueba, oral o escrita, sobre los contenidos tratados en la misma, que se realizará de forma individual. Si no se aprobase dicha prueba no se consideraría la calificación obtenida en la PL1.
4. **Prueba de Laboratorio 2 (PL2)**. Consistirá en la realización de una práctica, que se realizará con el software de Data Science que se haya estudiado en la asignatura, y que estará basada en los contenidos visto en las lecciones 4, 5 y 6. La PL2 podrá tener asociada una prueba, oral o escrita, sobre los contenidos tratados en la misma, que se realizará de forma individual. Si no se aprobase dicha prueba no se consideraría la calificación obtenida en la PL2.

Sobre el desarrollo de las pruebas de evaluación se especifica la duración de cada una de las pruebas:

- PEC1 y PEC2 tendrán una duración máxima de 1 hora y 55 minutos y se realizarán en el aula, en dos fechas que se publicarán en la presentación de la asignatura.
- PL1 y PL2 serán realizadas durante las semanas en las que se estén estudiando las lecciones contenidas en cada una de ellas y se entragarán en la semana siguiente a la finalización de la impartición de las lecciones 3 y 6 respectivamente.

Evaluación mediante examen final:

En el caso de evaluación mediante examen final se establece la siguiente prueba:

Prueba Evaluación Final (PEF). Consistirá en preguntas sobre los contenidos de todos los temas de la asignatura y el software de Data Science utilizado. Tendrá lugar en la fecha fijada oficialmente por la dirección de los estudios para dicha prueba y tendrá una duración máxima de 2h y 50 minutos.

Convocatoria extraordinaria

La convocatoria extraordinaria tendrá el mismo formato y duración máxima que la evaluación final y tendrá lugar en la fecha fijada oficialmente por la dirección de los estudios para dicha prueba.

5.2. EVALUACIÓN

CRITERIOS DE EVALUACIÓN

Los criterios de evaluación que se aplicarán a la asignatura de Fundamentos de la Ciencia de Datos están vinculados a los resultados de aprendizaje que se ha definido para la misma en en la sección 2, y estarán orientados a recoger tanto el nivel de comprensión de los contenidos e ideas básicas como la capacidad de aplicar los conocimientos adquiridos a la resolución de problemas, sin y con ayuda de herramientas software. Además, se tendrá en cuenta la actitud de colaboración y proactiva en la realización de las tareas, tanto indivisules como de grupo, y el grado de interés mostrado en la realización y cumplimiento de las actividades de clase y de trabajo autónomo y colaborativo, por lo que será fundamental la entrega a tiempo y con el formato correcto, de todas las actividades de prácticas.

Se utilizarán los siguientes porcentajes para cada uno de los scriterios para la evaluación de la asignatura, relacionados con los resultados del aprendizaje:

CE1. Dominio y aplicación de los contenidos: 70%

CE2. Resolución mediante herramientas software de problemas prácticos: 25%

CE3. Interés y proceder en la asignatura: 5%

INSTRUMENTOS DE EVALUACIÓN

Esta sección resume los instrumentos de calificación que serán aplicados a cada uno de los criterios de Evaluación.

CE1. PEC1 y PEC2

CE2. PL1 y PL2

CE3. Interés y proceder en las clases en el aula y en el laboratorio

CRITERIOS DE CALIFICACIÓN

En la convocatoria **ordinaria–evaluación continua** la relación entre las competencias, resultados del aprendizaje, criterios e instrumentos de evaluación, es la siguiente.

Competencia	Resultado de Aprendizaje	Criterio de Evaluación	Instrumento de evaluación	Peso en la calificación
CG8, CG9, CB1-CB5, TRU1-TRU5, CC4-CC7	RA1-RA4	CE1, CE2, CE3	PEC1	35%
CG8, CG9, CB1-CB5, TRU1-TRU5, CC4-CC7	RA1-RA4	CE1, CE2, CE3	PEC2	35%
CG8, CG9, CB1-CB5, TRU1-TRU5, CC4-CC7	RA1-RA4	CE2, CE3	PL1	12,5%
CG8, CG9, CB1-CB5, TRU1-TRU5, CC4-CC7	RA1-RA4	CE2, CE3	PL2	12,5%
TR4, TR5		CE3	Clases y Pruebas	5%

Si la calificación de alguna PEC es inferior a 1.2 sobre 3.5 no se podrá compensar con el resto de calificaciones.

Se otorgará la calificación de "No presentado" al alumno que habiendo optado por el procedimiento de evaluación continua, cumpla alguno de los siguientes requisitos:

- Cuando el alumno haya incumplido al menos la asistencia al 80% de las clases de teoría y problemas.
- Cuando el alumno no se haya presentado a las pruebas de evaluación continua.
- Cuando el alumno no haya entregado las pruebas de laboratorio.

En la convocatoria **ordinaria–evaluación final** la relación entre las competencias, resultados del aprendizaje, criterios e instrumentos de evaluación, es la siguiente.

Competencia	Resultado de Aprendizaje	Criterio de Evaluación	Instrumento de evaluación	Peso en la calificación
CG8, CG9, CB1-CB5, TRU1-TRU5, CC4-CC7	RA1-RA4	CE1, CE2, CE3	PEF	100%

Convocatoria extraordinaria

En el caso de la convocatoria extraordinaria se mantendrán los mismos porcentajes que se han establecido en el caso de la evaluación mediante examen final.

6. BIBLIOGRAFÍA

6.1. Bibliografía básica

- J.J. Cuadrado-Gallego y Y. Demchenko, The Data Science Framework: A View from the EDISON Project. Springer, 2020. ISBN: 978-3-030-51023-7. Introduce el marco de la Ciencia de Datos y permite al lector conocer desde que abarca el cuerpo de conocimiento de la misma hasta que salidas profesionales tiene.
- J. Han et al., Data Mining: Concepts and Techniques. Morgan Kaufman, 2012. ISBN: 978-0-12-381479-1. Se trata de uno de los libros de referencia más utilizados para el aprendizaje sobre Data Analytics, permite al lector introducirse en todos los conceptos que configuran la materia.
- M. Crawley, The R Book. John Wiley & Sons, 2007. Un libro muy interesante para profundizar en los conceptos de R. Puede accederse on-line desde la biblioteca de la Universidad de Alcalá.

6.2. Bibliografía complementaria

- J. Leskovec, Mining of Massive Datasets. Cambridge University Press, 2011. Un libro muy interesante para profundizar en los conceptos de Data Analytics.
- P. Teetor, R Cookbook. O'Reilly Media, 2011. ISBN: 978-0-596-80915-7. Libro muy útil para introducirse en el entorno R. Puede accederse on-line desde la biblioteca de la Universidad de Alcalá.

NOTA INFORMATIVA

La Universidad de Alcalá garantiza a sus estudiantes que, si por exigencias sanitarias las autoridades competentes impidieran la presencialidad total o parcial de la actividad docente, los planes docentes alcanzarían sus objetivos a través de una metodología de enseñanza-aprendizaje y evaluación en formato online, que retornaría a la modalidad presencial en cuanto cesaran dichos impedimentos.