



centro adscrito a:



UNIVERSITAT POLITÈCNICA  
DE CATALUNYA  
BARCELONATECH

# **GRADO EN CIENCIAS Y TECNOLOGÍAS APLICADAS AL DEPORTE Y AL ACONDICIONAMIENTO FÍSICO**

**GUÍA DOCENTE de DATA SCIENCE:  
PROGRAMACIÓN Y ANÁLISIS DE DATOS  
CUARTO CURSO, PRIMER CUATRIMESTRE**

**2022-2023**



## DATOS GENERALES

Nombre Asignatura:	DATA SCIENCE: PROGRAMACIÓN Y ANÁLISIS DE DATOS
Código:	801757
Curso:	2022-23
Titulación:	Grado en Ciencias y Tecnologías Aplicadas al Deporte y al Fitness
N.º de créditos (ECTS):	6
Requisitos:	Sin requisitos previos
Ubicación en el plan de estudios:	Cuarto curso, primer cuatrimestre
Fecha de la última revisión:	3 de octubre de 2022
Profesor Responsable:	Ray G. Butler

## 1 DESCRIPCIÓN GENERAL

Durante el transcurso de un simple partido de fútbol o de baloncesto se generan varios millones de datos. Información que, en muchos casos, ni siquiera el ojo humano no es capaz de percibir pero que, con una adecuada gestión y análisis, pueden dar lugar a ventajas competitivas claras.

Un contexto tan complejo como este requiere de la conjunción de diferentes áreas de conocimiento para dar respuestas a los profesionales del deporte: la *estadística* –para diseñar de forma eficiente los estudios y analizar de forma inteligente los datos-, la *programación informática* –para la organización del sistema de información y la automatización de los procesos-, la *comunicación* –para poder construir historias basadas en el datos que generen impacto real en la organización- y, por supuesto, la *ciencia del deporte* para poder darle sentido a todo este conocimiento dentro de un área de negocio tan específica.

En esto consiste la ciencia de datos: un nuevo perfil profesional en enorme auge a nivel internacional, especialmente en el sector deportivo, que permitirá al alumno aportar un gran valor añadido a la organización.

En esta asignatura aprenderemos, de forma eminentemente práctica, a diseñar estudios analíticos avanzados, construir sistemas de información que los soporten, recoger, integrar, explorar y preparar los datos, generar modelos numéricos de clasificación y regresión (aprendizaje de máquina) y, finalmente, reportar los resultados de forma efectiva.

## 2 OBJETIVOS

La asignatura y sus resultados previstos responden a estos objetivos:

- Comprender la relación entre los conceptos básicos del área: de las matemáticas y la estadística al aprendizaje de máquina y la inteligencia artificial.
- Diseñar estudios de analítica avanzada de datos efectivos para el sector del deporte.
- Conocer las técnicas más apropiadas para la preparación adecuada de los datos.

- Aprender a escoger y manejar las librerías y algoritmos de análisis avanzado más adecuados en cada caso.
- Construir, interpretar y mejorar la calidad de los modelos numéricos avanzados.
- Desarrollar las habilidades de comunicación de impacto basada en datos.
- Identificar los posibles caminos en el desarrollo profesional dentro de esta área.

### 3 CONTENIDOS

## BLOQUE 1. INTRODUCCIÓN A LA CIENCIA DE DATOS

---

### TEMA 1: LA CIENCIA DE DATOS

#### Resultados del aprendizaje

El estudiante después de estudiar el capítulo y realizar los ejercicios, será capaz de:

- Entender la relación entre la inteligencia artificial, el aprendizaje de máquina, la estadística y las matemáticas.
- Definir con precisión las características de un científico de datos.
- Enumerar los elementos clave del diseño de un estudio de analítica avanzada de datos en el deporte.

#### Contenido

- 1.1 Necesidad de la ciencia de datos en el deporte y en el fitness.
- 1.2 Qué es un científico de datos.
- 1.3 Posibilidades de desarrollo profesional.
- 1.4 Diseño de un proyecto de ciencia de datos.

## BLOQUE 2. INGENIERÍA DE DATOS

---

### TEMA 2: SISTEMAS DE INFORMACIÓN Y BASES DE DATOS

#### Resultados del aprendizaje

El estudiante después de estudiar el capítulo y realizar los ejercicios, será capaz de:

- Diseñar el sistema de información apropiado para un proyecto de ciencia de datos
- Construir la base de datos óptima para el sistema de información
- Cargar, consultar y manipular la información de la base de datos

#### Contenido

- 2.1 Elementos de un sistema de información óptimo.
- 2.2 Creación de bases de datos: de Excel a MySQL.
- 2.3 Consultas de datos con SQL.

### TEMA 3: ACCESO Y PREPARACIÓN DE DATOS

#### Resultados del aprendizaje

El estudiante después de estudiar el capítulo y realizar los ejercicios, será capaz de:

- Conectar con las fuentes de datos deportivas más importantes.
- Identificar y eliminar la información innecesaria.
- Identificar y corregir datos erróneos y outliers.
- Transformar adecuadamente las escalas de los datos.
- Aplicar estrategias adecuadas para los datos perdidos (missing data).
- Enriquecer los datos: feature engineering e integración de datos.

## Contenido

3.1 Preprocesado de datos con Python.

3.2 Integración de datos: concepto y automatización con Python.

## **BLOQUE 3. MODELOS DE APRENDIZAJE AUTOMÁTICO**

---

### **TEMA 4: EXPLORACIÓN DE DATOS**

#### **Resultados del aprendizaje**

El estudiante después de estudiar el capítulo y realizar los ejercicios, será capaz de:

- Extraer información estadística clave de los datos.
- Diseñar y construir dashboards para la exploración visual de los datos.
- Aplicar técnicas de exploración automatizada de datos.

#### **Contenido**

4.1 Estadística exploratoria.

4.2 Exploración visual de datos: dashboards.

4.3 Exploración automatizada de datos: AutoDiscovery.

### **TEMA 5: CLASIFICACIÓN, REGRESIÓN y CLUSTERING**

#### **Resultados del aprendizaje**

El estudiante después de estudiar el capítulo y realizar los ejercicios, será capaz de:

- Escoger el mejor modelo para dar solución a las necesidades del proyecto.
- Identificar y estructurar los datos para las etapas de entrenamiento, validación y testeo.
- Construir, evaluar la calidad y mejorar los modelos construidos.

#### **Contenido**

5.1 Modelos

5.2 Tipos de modelos: clasificación, regresión y clustering.

5.3 Etapas del proceso: entrenamiento, validación y testeo.

5.4 Los árboles de clasificación Random Forest.

5.5 Regresión lineal y logística.

5.6 Clustering jerárquico y k-means.

5.7 Indicadores de calidad de los modelos: sensibilidad, especificidad y curvas ROC.

## METODOLOGÍA DE ENSEÑANZA Y APRENDIZAJE

Se basa en clases prácticas desarrolladas con los equipos y herramientas informáticas del aula (o del alumno) y complementadas con la exposición de los contenidos teóricos fundamentales.

## EVALUACIÓN

De acuerdo con el Plan Bolonia, el modelo premia el esfuerzo constante y continuado del estudiantado. Un 40% de la nota se obtiene de la evaluación continua de las actividades dirigidas y el 60% porcentaje restante, del examen final presencial. El examen final tiene dos convocatorias.

La nota final de la asignatura (NF) se calculará a partir de la siguiente fórmula:

- $NF = \text{Nota Examen Final} \times 60\% + \text{Nota Evaluación Continuada} \times 40\%$
- Nota mínima del examen final para calcular la NF será de 40 puntos sobre 100.
- La asignatura queda aprobada con una NF igual o superior a 50 puntos sobre 100.

Tipo de actividad	Descripción	% Evaluación	
<b>Entregas:</b>			<b>40%</b>
	ENTREGA #1: TRABAJO EN EQUIPO (DATOS PREPROCESADOS)	25%	
	ENTREGA #2: TRABAJO EN EQUIPO (EXPLORACIÓN Y MODELOS FINALES)	75%	
<b>Examen final</b>			<b>60%</b>
	Examen final	100%	

## BIBLIOGRAFÍA BÁSICA

- Miller, T.W. (2015). *Sports analytics and data science: winning the game with methods and models*. FT Press.
- Haslwanter (2016). *An Introduction to Statistics with Python: With Applications in the Life Sciences (Statistics and Computing)*. Springer.
- Raschka y Mirjalili (2019). *Python Machine Learning: With Applications in the Life Sciences (Statistics and Computing)*. Springer.

## BIBLIOGRAFÍA ADICIONAL

- Ortega, Ortiz y Artés (2009). *Manual de estadística aplicada a las ciencias de la actividad física y el deporte*. Murcia: Diego Marín.

## WEBGRAFÍA / BLOGS / ETC.

- Data Science Central: <https://www.datasciencecentral.com/>
- Towards Data Science - The Data Science Process: <https://towardsdatascience.com/the-data-science-process-a19eb7ebc41b>
- KDNuggets: <https://www.kdnuggets.com/>
- Barça Innovation Hub : <https://barcainnovationhub.com/es/que-se-necesita-aprender-para-trabajar-en-football-analytics/>