



- 1. Código:** 14021 **Nombre:** Proyecto III, análisis de datos
- 2. Créditos:** 6,00 **--Teoría:** 3,00 **--Prácticas:** 3,00 **Carácter:** Obligatorio
- Titulación:** 189-Grado en Ciencia de Datos
- Módulo:** 2-Materias obligatorias **Materia:** 10-Metodología y proyecto
- Centro:** E.T.S. DE INGENIERÍA INFORMÁTICA
- 3. Coordinador:** Hernández Orallo, José
- Departamento:** SISTEMAS INFORMÁTICOS Y COMPUTACIÓN

4. Bibliografía

Managing data science : effective strategies to manage data science projects and build a sustainable team	Dubovikov, Kirill
Data science for business	Provost, Foster Fawcett, Tom
Creating a data-driven organization : practical advice from the trenches	Anderson, Carl
Managing Your Data Science Projects [electronic resource] : Learn Salesmanship, Presentation, and Maintenance of Completed Models	de Graaf, Robert
Doing data science	Schutt, Rachel O'Neil, Cathy
Building machine learning powered applications : going from idea to product	Ameisen, Emmanuel
Computer age statistical inference : algorithms, evidence, and data science / [electronic resource]	Efron, Bradley Hastie, Trevor
Value proposition design : how to create products and services customers want : get started with--	Osterwalder, Alexander Pigneur, Yves Bernarda, Greg Smith, Alan

5. Descripción general de la asignatura

Objetivos de la asignatura

Los alumnos, organizados en equipo, realizarán las últimas fases de un proyecto de análisis de datos (modelado descriptivo y predictivo y evaluación) sobre conjuntos de datos.

1. El contexto del análisis
 - Identificación de oportunidades y problemas y sus criterios de éxito
 - Planteamiento de la solución trasladando desde el nivel de dominio al nivel de análisis
 - Selección de costes, riegos y trade-offs
2. Puesta en común del trabajo y los objetivos a alcanzar
 - Identificar las tareas (regresión, clasificación, agrupamiento, ..)
 - Establecimiento de métricas (error, costes, ...)
3. Realizar un modelo básico para cada una de las tareas
4. Mejora iterativa de los modelos básicos
 - Ampliación de los modelos (añadiendo características)
 - Propuesta de modelos alternativos
 - Variación de parámetros, ajuste de técnicas, etc.
 - Propuesta de monitorización y mantenimiento del modelo
5. Presentación, visualización y defensa de la solución propuesta
 - Visualización de los datos y de los resultados de los modelos.
 - Elaboración de informes / presentación para gerencia o cliente.
 - Defensa de la propuesta en público.

Contextualización de la asignatura

Este es la tercera asignatura de proyecto de la asignatura, y llega en un momento en el que están más maduros y con experiencia de los otros dos grandes proyectos en grupo que han hecho en cursos anteriores. En este se les exige ser más autónomos y centrarse más en el análisis y en el valor de los datos. Como en todas las asignaturas de proyecto, se intenta que el trabajo esté relacionado más con las asignaturas de tercero, aunque puedan utilizar lo aprendido en asignaturas de primero y segundo.

6. Conocimientos recomendados

- (13997) Fundamentos de organización de empresas
- (14010) Modelos descriptivos y predictivos I
- (14011) Modelos descriptivos y predictivos II
- (14012) Comportamiento económico y social

Document signat electrònicament per Documento firmado electrónicamente por Electronically signed document by	UNIVERSITAT POLITÈCNICA DE VALÈNCIA	Data/Fecha/Date 02/06/2023	1 / 4	
Autenticitat verificable mitjançant Codi Segur Verificació Autenticidad verificable mediante Código Seguro Verificación Original document can be verified by Secure Verification Code		ALU6MIAX0FH https://sede.upv.es/eVerificador		



6. Conocimientos recomendados

- (14013) Economía digital
- (14016) Infraestructura para el procesamiento de datos
- (14017) Seguridad de los datos
- (14019) Proyecto I, comprensión de datos
- (14020) Proyecto II, integración y preparación de datos
- (14022) Visualización
- (14023) Modelado y simulación continuos
- (14024) Modelado discreto y teoría de la información
- (14027) Gestión de datos
- (14028) Evaluación, despliegue y monitorización de modelos
- (14029) Lenguaje natural y recuperación de la información
- (14030) Representación del conocimiento y razonamiento
- (14039) English B2 - Upper Intermediate English for Data Science

Por el tipo de asignatura, se recomiendan los conocimientos de asignaturas de los dos cursos anteriores, y especialmente las metodologías y dinámicas de trabajo en equipo de las asignaturas de proyecto de I y II. Además, al tratarse de una asignatura integradora de todo lo visto en el tercer curso, se utilizarán los conocimientos y habilidades de las asignaturas ya cursadas en el cuatrimestre 3A. De la misma forma, se avanzará de manera simultánea a las asignaturas del cuatrimestre 3B, de modo que se vayan incorporando a Proyectos III las competencias adquiridas en el resto de asignaturas del cuatrimestre.

La asignatura utiliza inglés instrumental, maximizando la interacción entre profesores y alumnos en ese idioma, con material y trabajo en inglés. La evaluación y preguntas de los alumnos pueden ser en inglés, castellano o valenciano.

7. Resultados

Resultados fundamentales

CB02(GE) 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.

CB04(GE) Que los estudiantes puedan transmitir información, ideas, problemas y soluciones a un público tanto especializado como no especializado.

CE01(ES) Detectar nuevas oportunidades de negocio para mejorar el funcionamiento de cualquier tipo de organización, a partir de los datos generados en su entorno socioeconómico, cuantificando el valor de éstos e identificando los agentes interesados.

CE04(ES) Aplicar metodologías y técnicas estadísticas y de aprendizaje automático para el procesamiento y análisis de datos.

CE10(ES) Comunicar de forma efectiva el proceso de extracción de conocimiento a partir de los datos, elaborar informes y emitir recomendaciones para la toma de decisiones.

CG04(GE) Integrarse en un equipo multidisciplinar en el marco de un proyecto de ciencia de datos.

CE12(ES) Aplicar los aspectos éticos, legales y normativos relacionados con el tratamiento de los datos y la explotación del conocimiento obtenido.

CE18(ES) Diseñar, implementar y evaluar las soluciones algorítmicas más apropiadas para procesar datos de forma eficiente.

CG01(GE) Concebir, diseñar y utilizar aplicaciones para extraer conocimiento a partir de distintas fuentes de datos.

CG02(GE) Desarrollar, gestionar y liderar proyectos que utilicen las técnicas, herramientas y metodologías propias de la ciencia de datos en distintos ámbitos de aplicación.

CG03(GE) Planificar los procesos necesarios para la extracción de conocimiento a partir de distintas fuentes de datos.

CE11(ES) Seleccionar y utilizar las técnicas y herramientas de visualización más adecuadas para representar datos y resultados en un contexto determinado.

Competencias transversales

(1) Compromiso social y medioambiental

- Actividades desarrolladas relacionadas con la adquisición de la competencia

Se realiza un proyecto en el que se deben establecer los costes, los riesgos y los compromisos.

En el plan de proyecto los alumnos han de realizar una evaluación de costes-beneficios, evaluación de riesgos y mitigaciones, con una evaluación de impactos medioambientales, de protección de datos, éticos y legales. Además se enmarcará el proyecto en el contexto de los Objetivos del Desarrollo Sostenible, indicando su eventual contribución a nivel de objetivo y de meta.

- Criterios de evaluación

En los hitos del plan del proyecto se examinarán los informes de evaluación de impacto medioambiental y de protección

Document signat electrònicament per Documento firmado electrónicamente por Electronically signed document by	UNIVERSITAT POLITÈCNICA DE VALÈNCIA	Data/Fecha/Date 02/06/2023	2 / 4	
Autenticitat verificable mitjançant Codi Segur Verificació Autenticidad verificable mediante Código Seguro Verificación Original document can be verified by Secure Verification Code	ALU6MIAX0FH https://sede.upv.es/eVerificador			



7. Resultados

Competencias transversales

de datos, así como la cobertura de riesgos y sus mitigaciones. En concreto, se evaluará si el alumno es capaz de contemplar las responsabilidades éticas y el impacto medioambiental y profesional del proyecto y articular las medidas para evitarlos o paliarlos.

(2) Innovación y creatividad

- Actividades desarrolladas relacionadas con la adquisición de la competencia
Se debe determinar cuál es el valor añadido del proyecto, y que aporta respecto a resultados o servicios ya existentes.

- Criterios de evaluación

En los hitos del proyecto se examina si el alumno ha comprobado el objetivo del propuesto no ha sido realizado ya, qué hace distintivo el proyecto, mirando otras propuestas e ideas en la literatura o en el mercado. Se valoran elementos como el "value proposition" en el "business canvas" o qué novedades aporta en caso de proyectos de investigación o de enfoque social.

(3) Trabajo en equipo y liderazgo

- Actividades desarrolladas relacionadas con la adquisición de la competencia

Trabajar en equipo para la realización de un proyecto de ciencia de datos orientado al análisis de datos.

Se configura un equipo de trabajo que determina roles, programa tareas y carga de trabajo entre los miembros del equipo, estableciendo una programación, reglas de funcionamiento y mecanismos de resolución de conflictos.

- Criterios de evaluación

Para la evaluación de esta competencia se utilizarán rúbricas de evaluación por el profesor y de co-evaluación entre alumnos para determinar cada miembro del equipo:

- Se ha coordinado bien, balanceado la carga y compartido las responsabilidades.
- Ha participado activamente en las reuniones del equipo, no sólo asistiendo sino proponiendo ideas y soluciones.
- Ha realizado las tareas asignadas dentro del equipo en el plazo fijado.

8. Unidades didácticas

1. Unit 1. The context of a data analysis project: Opportunities, problems and success criteria
2. Unit 2 Profiles of data science projects: business, social and research
3. Unit 3 Task identification and model building
4. Unit 4 Iterative improvement
5. Unit 5 Exploitation and value assessment
6. Unit 6 Presentation and report

9. Método de enseñanza-aprendizaje

Se utilizará el inglés como lengua vehicular en la mayoría de actividades y el material se dará en este idioma.

UD	TA	SE	PA	PL	PC	PI	EVA	TP	TNP	TOTAL HORAS
1	--	10,00	--	5,00	--	--	--	15,00	10,00	25,00
2	--	10,00	--	5,00	--	--	0,50	15,50	10,00	25,50
3	--	2,00	--	5,00	--	--	0,50	7,50	20,00	27,50
4	--	2,00	--	5,00	--	--	--	7,00	20,00	27,00
5	--	4,00	--	5,00	--	--	--	9,00	20,00	29,00
6	--	2,00	--	5,00	--	--	7,00	14,00	20,00	34,00
TOTAL HORAS	--	30,00	--	30,00	--	--	8,00	68,00	100,00	168,00

UD: Unidad Didáctica. TA: Teoría de Aula. SE: Seminario. PA: Práctica de Aula. PL: Práctica de Laboratorio. PC: Práctica de Campo. PI: Práctica de Informática. EVA: Actividades de Evaluación. TP: Trabajo Presencial. TNP: Trabajo No Presencial.

10. Evaluación

Descripción

Descripción	Nº Actos	Peso (%)
(05) Trabajos académicos	2	30
(16) Evaluación con participación del estudiantado	2	20
(09) Proyecto	1	50

The qualification will be based on a team project that will be developed during the term. It is composed of the results of two milestones (M1 and M2), one final evaluation (G1) (affected by the intra-team co-evaluation C1), as well as the inter-team co-evaluation C2.

- Two milestones (whose dates will be announced), with 15% of the final mark for each of them, i.e., M1 (15%), M2 (15%), with

Document signat electrònicament per Documento firmado electrónicamente por Electronically signed document by	UNIVERSITAT POLITÈCNICA DE VALÈNCIA	Data/Fecha/Date 02/06/2023	3 / 4	
Autenticitat verificable mitjançant Codi Segur Verificació Autenticidad verificable mediante Código Seguro Verificación Original document can be verified by Secure Verification Code	ALU6MIAX0FH	https://sede.upv.es/eVerificador		



10. Evaluación

a public rubric for each of them, which will also evaluate the application of the ideas and recommendations seen during the seminars. M1 will be mostly oral and M2 will be mostly written. FEEDBACK AND RESIT OPTION (RECOVERY): There will be continuous feedback for both milestones. M1 has no resit. For M2, once finished and evaluated, there will be a period of 10 days to improve the grade of M2 by making amendments.

- The final oral presentation (G1) and associated material (final report) represents 60% of the final mark. This is not additive, as 50% (G1) + 10% (C1) suggested by the percentage breakout but multiplicative. This means that the grade of this G1 will be multiplied by a coefficient C1, which may go between 0 and 1.5, but will be 1 if the student contributes to the group in the proportion and attitude that is expected. This C1 is co-evaluated by the other team mates through the intra-team co-evaluation questionnaire. There will be public rubrics for G1 and C1, which will also evaluate the application of the ideas and recommendations seen during the seminars. FEEDBACK AND RESIT OPTION (RECOVERY): The presentation has a pre-evaluation, which produces a grade and feedback for the students. After that, in a few weeks, there will be a final evaluation that serves as resit. No further option for resit will be allowed beyond the final presentation.

- The inter-team evaluation where each group is evaluated by the rest of the class (C2: 10%), using a questionnaire with a finite number of points to assign among the other groups, with a score that is obtained as a median of all peer evaluations. There's no resit for this part of the evaluation.

The project will be done in teams, whose composition will follow some methodology for maximising team diversity. The teams will find or be assigned some data to exploit (from repositories, open data, companies, etc.) in such a way that they figure out how an existing procedure can be optimised after the extraction of descriptive or predictive models, or they will innovate a new data-based product that can be developed from the data.

Two or three weeks before the end of the term, the students will be allowed to ask for a timeslot for a rehearsal (a pre-evaluation). All students have to perform this pre-evaluation, and they will be given a valid grade, jointly with their co-evaluations. Only if this pre-evaluation is made, will the team still have the option to re-sit for the presentation of the project on the dates originally scheduled for the final evaluation.

Resits (recovery): the first paragraphs have explained how --and when-- to resit for the final presentation G1 (and associated report) and how --and when-- to remake the milestone M2 in case the students want to improve the grades. No other resit option will be offered in any case.

Students with "DISPENSA": They will be able to work asynchronously and remotely for the project (but also in a team). They will have to deliver the same milestones and presentations.

The final grade for all students just adds this value to the project result:

$$\text{Grade (0-10)} = M1 (0-1.5) + M2 (0-1.5) + G1 (0-6) * C1 (0-1.5) + C2 (0-1)$$

While the course uses English instrumentally for all the activities and interactions, and evaluations are encouraged in this language, students can also ask to be evaluated in Spanish or Valencian.

11. Porcentaje máximo de ausencia

<u>Actividad</u>	<u>Porcentaje</u>	<u>Observaciones</u>
Teoría Seminario	50	
Práctica Laboratorio	50	

