

Guía docente

270954 - PODS - Ciencia de Datos Orientada a Procesos

Última modificación: 23/11/2023

Unidad responsable: Facultad de Informática de Barcelona
Unidad que imparte: 723 - CS - Departamento de Ciencias de la Computación.

Titulación: MÁSTER UNIVERSITARIO EN CIENCIA DE DATOS (Plan 2021). (Asignatura obligatoria).

Curso: 2023 **Créditos ECTS:** 6.0 **Idiomas:** Inglés

PROFESORADO

Profesorado responsable: CARLOS ESCOLANO PEINADO

Otros: Primer quadrimestre:
CARLOS ESCOLANO PEINADO - 11, 12

CAPACIDADES PREVIAS

Comprensión completa de la informática en general; buen dominio de varios lenguajes de programación; capacidad básica para formalizar matemáticamente cuestiones en ingeniería informática.

COMPETENCIAS DE LA TITULACIÓN A LAS QUE CONTRIBUYE LA ASIGNATURA

Específicas:

CE13. Identificar las principales amenazas en el ámbito de la ética y la privacidad de datos en un proyecto de ciencia de datos (tanto en el aspecto de gestión como de análisis de datos) y desarrollar e implantar medidas adecuadas para mitigar dichas amenazas.

CE5. Modelar, diseñar e implementar sistemas complejos de datos, incluyendo la visualización de datos

CE6. Diseñar el proceso de Ciencia de Datos y aplicar metodologías científicas para obtener conclusiones sobre poblaciones y tomar decisiones en consecuencia, a partir de datos estructurados o no estructurados y potencialmente almacenados en formatos heterogéneos.

CE7. Identificar las limitaciones impuestas por la calidad de datos en un problema de ciencia de datos y aplicar técnicas para disminuir su impacto

CE9. Aplicar métodos adecuados para el análisis de otro tipo de formatos, tales como procesos y grafos, dentro del ámbito de ciencia de datos

Genéricas:

CG2. Identificar y aplicar métodos de análisis, extracción de conocimiento y visualización de datos recogidos en formatos muy diversos.

CG3. Definir, diseñar e implementar sistemas complejos que cubran todas las fases en proyectos de ciencia de datos

Transversales:

CT1. ESPÍRITU EMPRENDEDOR E INNOVADOR: Conocer y entender la organización de una empresa y las ciencias que rigen su actividad; tener capacidad para entender las normas laborales y las relaciones entre la planificación, las estrategias industriales y comerciales, la calidad y el beneficio. Conocer y entender los mecanismos en que se basa la investigación científica, así como los mecanismos e instrumentos de transferencia de resultados entre los diferentes agentes socioeconómicos implicados en los procesos de I+D+i.

CT4. USO SOLVENTE DE LOS RECURSOS DE INFORMACIÓN: Gestionar la adquisición, la estructuración, el análisis y la visualización de datos e información en el ámbito de la especialidad y valorar de forma crítica los resultados de esta gestión.

CT5. TERCERA LENGUA: Conocer una tercera lengua, preferentemente el inglés, con un nivel adecuado oral y escrito y en consonancia con las necesidades que tendrán los titulados y tituladas.

Básicas:

CB10. Poseer y comprender conocimientos que aporten una base u oportunidad de ser originales en el desarrollo y/o aplicación de ideas, a menudo en un contexto de investigación.

CB6. Que los estudiantes sepan aplicar los conocimientos adquiridos y su capacidad de resolución de problemas en entornos nuevos o poco conocidos dentro de contextos más amplios (o multidisciplinares) relacionados con su área de estudio.

CB7. Que los estudiantes sean capaces de integrar conocimientos y enfrentarse a la complejidad de formular juicios a partir de una información que, siendo incompleta o limitada, incluya reflexiones sobre las responsabilidades sociales y éticas vinculadas a la aplicación de sus conocimientos y juicios

CB8. Que los estudiantes sepan comunicar sus conclusiones y los conocimientos y razones últimas que las sustentan a públicos especializados y no especializados de un modo claro y sin ambigüedades.

CB9. Que los estudiantes posean las habilidades de aprendizaje que les permitan continuar estudiando de un modo que habrá de ser en gran medida autodirigido o autónomo.

METODOLOGÍAS DOCENTES

Sesiones teóricas que pueden incluir sesiones de resolución de problemas con o sin componente de programación, sesiones prácticas con software de ciencia de datos orientada a procesos de código abierto o comercial, desarrollo de un caso de estudio.

OBJETIVOS DE APRENDIZAJE DE LA ASIGNATURA

1. Conocer el conjunto de problemas teóricos y prácticos que constituyen la ciencia de datos orientada a procesos, y comprender los principales algoritmos para abordarlos: tanto a nivel conceptual como a nivel de su aplicación a través de algunas de las herramientas y librerías actuales.

2. Adquirir y demostrar capacidad para poner en práctica los conocimientos adquiridos durante el curso, y relacionarlos con las perspectivas organizativas y de equipo como un proyecto de ciencia de datos orientado a procesos que se ejecuta en una organización real.

HORAS TOTALES DE DEDICACIÓN DEL ESTUDIANTADO

Tipo	Horas	Porcentaje
Horas grupo grande	24,0	16.00
Horas grupo pequeño	24,0	16.00
Horas actividades dirigidas	6,0	4.00
Horas aprendizaje autónomo	96,0	64.00

Dedicación total: 150 h

CONTENIDOS

Modelos de procesos y datos de eventos

Descubrimiento automático de modelos de procesos

Verificación de conformidad entre modelos de procesos y datos de eventos

Mejora de procesos basada en la evidencia de los datos



Selección de aplicaciones y técnicas avanzadas

Metodología para proyectos de ciencia de datos orientados a procesos

ACTIVIDADES

Modelos de procesos y datos de eventos

Descripción:

En esta actividad se introducirán los modelos de procesos para especificar procesos en las organizaciones, y los datos que hablan de eventos que se originan en la ejecución de los procesos.

Objetivos específicos:

1

Competencias relacionadas:

CG3. Definir, diseñar e implementar sistemas complejos que cubran todas las fases en proyectos de ciencia de datos

CG2. Identificar y aplicar métodos de análisis, extracción de conocimiento y visualización de datos recogidos en formatos muy diversos.

CE6. Diseñar el proceso de Ciencia de Datos y aplicar metodologías científicas para obtener conclusiones sobre poblaciones y tomar decisiones en consecuencia, a partir de datos estructurados o no estructurados y potencialmente almacenados en formatos heterogéneos.

CE7. Identificar las limitaciones impuestas por la calidad de datos en un problema de ciencia de datos y aplicar técnicas para disminuir su impacto

CE13. Identificar las principales amenazas en el ámbito de la ética y la privacidad de datos en un proyecto de ciencia de datos (tanto en el aspecto de gestión como de análisis de datos) y desarrollar e implantar medidas adecuadas para mitigar dichas amenazas.

CE5. Modelar, diseñar e implementar sistemas complejos de datos, incluyendo la visualización de datos

CE9. Aplicar métodos adecuados para el análisis de otro tipo de formatos, tales como procesos y grafos, dentro del ámbito de ciencia de datos

CT4. USO SOLVENTE DE LOS RECURSOS DE INFORMACIÓN: Gestionar la adquisición, la estructuración, el análisis y la visualización de datos e información en el ámbito de la especialidad y valorar de forma crítica los resultados de esta gestión.

CT5. TERCERA LENGUA: Conocer una tercera lengua, preferentemente el inglés, con un nivel adecuado oral y escrito y en consonancia con las necesidades que tendrán los titulados y tituladas.

CB10. Poseer y comprender conocimientos que aporten una base u oportunidad de ser originales en el desarrollo y/o aplicación de ideas, a menudo en un contexto de investigación.

CB7. Que los estudiantes sean capaces de integrar conocimientos y enfrentarse a la complejidad de formular juicios a partir de una información que, siendo incompleta o limitada, incluya reflexiones sobre las responsabilidades sociales y éticas vinculadas a la aplicación de sus conocimientos y juicios

CB9. Que los estudiantes posean las habilidades de aprendizaje que les permitan continuar estudiando de un modo que habrá de ser en gran medida autodirigido o autónomo.

Dedicación: 24h 54m

Grupo grande/Teoría: 5h

Grupo pequeño/Laboratorio: 4h

Aprendizaje autónomo: 15h 54m

Descubrimiento automático de modelos de procesos

Descripción:

En esta actividad se introducirán diversas técnicas que extraen modelos de procesos en diversos formalismos a partir de datos de eventos.

Objetivos específicos:

1

Competencias relacionadas:

CG3. Definir, diseñar e implementar sistemas complejos que cubran todas las fases en proyectos de ciencia de datos

CG2. Identificar y aplicar métodos de análisis, extracción de conocimiento y visualización de datos recogidos en formatos muy diversos.

CE6. Diseñar el proceso de Ciencia de Datos y aplicar metodologías científicas para obtener conclusiones sobre poblaciones y tomar decisiones en consecuencia, a partir de datos estructurados o no estructurados y potencialmente almacenados en formatos heterogéneos.

CE7. Identificar las limitaciones impuestas por la calidad de datos en un problema de ciencia de datos y aplicar técnicas para disminuir su impacto

CE13. Identificar las principales amenazas en el ámbito de la ética y la privacidad de datos en un proyecto de ciencia de datos (tanto en el aspecto de gestión como de análisis de datos) y desarrollar e implantar medidas adecuadas para mitigar dichas amenazas.

CE5. Modelar, diseñar e implementar sistemas complejos de datos, incluyendo la visualización de datos

CE9. Aplicar métodos adecuados para el análisis de otro tipo de formatos, tales como procesos y grafos, dentro del ámbito de ciencia de datos

CT4. USO SOLVENTE DE LOS RECURSOS DE INFORMACIÓN: Gestionar la adquisición, la estructuración, el análisis y la visualización de datos e información en el ámbito de la especialidad y valorar de forma crítica los resultados de esta gestión.

CT5. TERCERA LENGUA: Conocer una tercera lengua, preferentemente el inglés, con un nivel adecuado oral y escrito y en consonancia con las necesidades que tendrán los titulados y tituladas.

CB10. Poseer y comprender conocimientos que aporten una base u oportunidad de ser originales en el desarrollo y/o aplicación de ideas, a menudo en un contexto de investigación.

CB7. Que los estudiantes sean capaces de integrar conocimientos y enfrentarse a la complejidad de formular juicios a partir de una información que, siendo incompleta o limitada, incluya reflexiones sobre las responsabilidades sociales y éticas vinculadas a la aplicación de sus conocimientos y juicios

CB9. Que los estudiantes posean las habilidades de aprendizaje que les permitan continuar estudiando de un modo que habrá de ser en gran medida autodirigido o autónomo.

Dedicación: 28h

Grupo grande/Teoría: 6h

Grupo pequeño/Laboratorio: 6h

Aprendizaje autónomo: 16h

Verificación de conformidad de modelos de procesos y datos de eventos

Descripción:

En esta actividad se introducirán algoritmos para establecer la relación entre el comportamiento modelado y observado de un proceso.

Objetivos específicos:

1

Competencias relacionadas:

CG3. Definir, diseñar e implementar sistemas complejos que cubran todas las fases en proyectos de ciencia de datos

CG2. Identificar y aplicar métodos de análisis, extracción de conocimiento y visualización de datos recogidos en formatos muy diversos.

CE6. Diseñar el proceso de Ciencia de Datos y aplicar metodologías científicas para obtener conclusiones sobre poblaciones y tomar decisiones en consecuencia, a partir de datos estructurados o no estructurados y potencialmente almacenados en formatos heterogéneos.

CE7. Identificar las limitaciones impuestas por la calidad de datos en un problema de ciencia de datos y aplicar técnicas para disminuir su impacto

CE13. Identificar las principales amenazas en el ámbito de la ética y la privacidad de datos en un proyecto de ciencia de datos (tanto en el aspecto de gestión como de análisis de datos) y desarrollar e implantar medidas adecuadas para mitigar dichas amenazas.

CE5. Modelar, diseñar e implementar sistemas complejos de datos, incluyendo la visualización de datos

CE9. Aplicar métodos adecuados para el análisis de otro tipo de formatos, tales como procesos y grafos, dentro del ámbito de ciencia de datos

CT4. USO SOLVENTE DE LOS RECURSOS DE INFORMACIÓN: Gestionar la adquisición, la estructuración, el análisis y la visualización de datos e información en el ámbito de la especialidad y valorar de forma crítica los resultados de esta gestión.

CT5. TERCERA LENGUA: Conocer una tercera lengua, preferentemente el inglés, con un nivel adecuado oral y escrito y en consonancia con las necesidades que tendrán los titulados y tituladas.

CB10. Poseer y comprender conocimientos que aporten una base u oportunidad de ser originales en el desarrollo y/o aplicación de ideas, a menudo en un contexto de investigación.

CB7. Que los estudiantes sean capaces de integrar conocimientos y enfrentarse a la complejidad de formular juicios a partir de una información que, siendo incompleta o limitada, incluya reflexiones sobre las responsabilidades sociales y éticas vinculadas a la aplicación de sus conocimientos y juicios

CB9. Que los estudiantes posean las habilidades de aprendizaje que les permitan continuar estudiando de un modo que habrá de ser en gran medida autodirigido o autónomo.

Dedicación: 28h

Grupo grande/Teoría: 6h

Grupo pequeño/Laboratorio: 6h

Aprendizaje autónomo: 16h

Mejora de procesos basada en la evidencia de los datos

Descripción:

En esta actividad se presentarán técnicas para utilizar datos de eventos para proyectar y mejorar modelos de procesos y registros de eventos.

Objetivos específicos:

1

Competencias relacionadas:

CG3. Definir, diseñar e implementar sistemas complejos que cubran todas las fases en proyectos de ciencia de datos

CG2. Identificar y aplicar métodos de análisis, extracción de conocimiento y visualización de datos recogidos en formatos muy diversos.

CE6. Diseñar el proceso de Ciencia de Datos y aplicar metodologías científicas para obtener conclusiones sobre poblaciones y tomar decisiones en consecuencia, a partir de datos estructurados o no estructurados y potencialmente almacenados en formatos heterogéneos.

CE7. Identificar las limitaciones impuestas por la calidad de datos en un problema de ciencia de datos y aplicar técnicas para disminuir su impacto

CE13. Identificar las principales amenazas en el ámbito de la ética y la privacidad de datos en un proyecto de ciencia de datos (tanto en el aspecto de gestión como de análisis de datos) y desarrollar e implantar medidas adecuadas para mitigar dichas amenazas.

CE5. Modelar, diseñar e implementar sistemas complejos de datos, incluyendo la visualización de datos

CE9. Aplicar métodos adecuados para el análisis de otro tipo de formatos, tales como procesos y grafos, dentro del ámbito de ciencia de datos

CT4. USO SOLVENTE DE LOS RECURSOS DE INFORMACIÓN: Gestionar la adquisición, la estructuración, el análisis y la visualización de datos e información en el ámbito de la especialidad y valorar de forma crítica los resultados de esta gestión.

CT5. TERCERA LENGUA: Conocer una tercera lengua, preferentemente el inglés, con un nivel adecuado oral y escrito y en consonancia con las necesidades que tendrán los titulados y tituladas.

CB10. Poseer y comprender conocimientos que aporten una base u oportunidad de ser originales en el desarrollo y/o aplicación de ideas, a menudo en un contexto de investigación.

CB7. Que los estudiantes sean capaces de integrar conocimientos y enfrentarse a la complejidad de formular juicios a partir de una información que, siendo incompleta o limitada, incluya reflexiones sobre las responsabilidades sociales y éticas vinculadas a la aplicación de sus conocimientos y juicios

CB9. Que los estudiantes posean las habilidades de aprendizaje que les permitan continuar estudiando de un modo que habrá de ser en gran medida autodirigido o autónomo.

Dedicación: 24h

Grupo grande/Teoría: 4h

Grupo pequeño/Laboratorio: 4h

Aprendizaje autónomo: 16h



Selección de aplicaciones y técnicas avanzadas

Descripción:

Varias técnicas para resolver problemas de ciencias de datos orientados a procesos de la vida real

Objetivos específicos:

1, 2

Competencias relacionadas:

CG3. Definir, diseñar e implementar sistemas complejos que cubran todas las fases en proyectos de ciencia de datos

CG2. Identificar y aplicar métodos de análisis, extracción de conocimiento y visualización de datos recogidos en formatos muy diversos.

CE6. Diseñar el proceso de Ciencia de Datos y aplicar metodologías científicas para obtener conclusiones sobre poblaciones y tomar decisiones en consecuencia, a partir de datos estructurados o no estructurados y potencialmente almacenados en formatos heterogéneos.

CE7. Identificar las limitaciones impuestas por la calidad de datos en un problema de ciencia de datos y aplicar técnicas para disminuir su impacto

CE13. Identificar las principales amenazas en el ámbito de la ética y la privacidad de datos en un proyecto de ciencia de datos (tanto en el aspecto de gestión como de análisis de datos) y desarrollar e implantar medidas adecuadas para mitigar dichas amenazas.

CE5. Modelar, diseñar e implementar sistemas complejos de datos, incluyendo la visualización de datos

CE9. Aplicar métodos adecuados para el análisis de otro tipo de formatos, tales como procesos y grafos, dentro del ámbito de ciencia de datos

CT1. ESPÍRITU EMPRENDEDOR E INNOVADOR: Conocer y entender la organización de una empresa y las ciencias que rigen su actividad; tener capacidad para entender las normas laborales y las relaciones entre la planificación, las estrategias industriales y comerciales, la calidad y el beneficio. Conocer y entender los mecanismos en que se basa la investigación científica, así como los mecanismos e instrumentos de transferencia de resultados entre los diferentes agentes socioeconómicos implicados en los procesos de I+D+i.

CT4. USO SOLVENTE DE LOS RECURSOS DE INFORMACIÓN: Gestionar la adquisición, la estructuración, el análisis y la visualización de datos e información en el ámbito de la especialidad y valorar de forma crítica los resultados de esta gestión.

CT5. TERCERA LENGUA: Conocer una tercera lengua, preferentemente el inglés, con un nivel adecuado oral y escrito y en consonancia con las necesidades que tendrán los titulados y tituladas.

CB6. Que los estudiantes sepan aplicar los conocimientos adquiridos y su capacidad de resolución de problemas en entornos nuevos o poco conocidos dentro de contextos más amplios (o multidisciplinares) relacionados con su área de estudio.

CB10. Poseer y comprender conocimientos que aporten una base u oportunidad de ser originales en el desarrollo y/o aplicación de ideas, a menudo en un contexto de investigación.

CB7. Que los estudiantes sean capaces de integrar conocimientos y enfrentarse a la complejidad de formular juicios a partir de una información que, siendo incompleta o limitada, incluya reflexiones sobre las responsabilidades sociales y éticas vinculadas a la aplicación de sus conocimientos y juicios

CB8. Que los estudiantes sepan comunicar sus conclusiones y los conocimientos y razones últimas que las sustentan a públicos especializados y no especializados de un modo claro y sin ambigüedades.

CB9. Que los estudiantes posean las habilidades de aprendizaje que les permitan continuar estudiando de un modo que habrá de ser en gran medida autodirigido o autónomo.

Dedicación: 24h

Grupo grande/Teoría: 4h

Grupo pequeño/Laboratorio: 4h

Aprendizaje autónomo: 16h



Metodología para proyectos de ciencia de datos orientada a procesos

Descripción:

Descripción general de cómo gestionar un proyecto PODS

Objetivos específicos:

2

Competencias relacionadas:

CT1. ESPÍRITU EMPRENDEDOR E INNOVADOR: Conocer y entender la organización de una empresa y las ciencias que rigen su actividad; tener capacidad para entender las normas laborales y las relaciones entre la planificación, las estrategias industriales y comerciales, la calidad y el beneficio. Conocer y entender los mecanismos en que se basa la investigación científica, así como los mecanismos e instrumentos de transferencia de resultados entre los diferentes agentes socioeconómicos implicados en los procesos de I+D+i.

CB6. Que los estudiantes sepan aplicar los conocimientos adquiridos y su capacidad de resolución de problemas en entornos nuevos o poco conocidos dentro de contextos más amplios (o multidisciplinares) relacionados con su área de estudio.

CB8. Que los estudiantes sepan comunicar sus conclusiones y los conocimientos y razones últimas que las sustentan a públicos especializados y no especializados de un modo claro y sin ambigüedades.

Dedicación: 21h

Grupo grande/Teoría: 2h

Grupo pequeño/Laboratorio: 3h

Aprendizaje autónomo: 16h

SISTEMA DE CALIFICACIÓN

La evaluación de la asignatura consiste en dos elementos: exámen final (60%), prácticas de curso (40%).

El examen final contendrá preguntas y problemas sobre los contenidos teóricos que se explican en las clases de teoría.

Las prácticas de curso serán prácticas guiadas que se realizarán durante el curso sobre diversas herramientas y plataformas de minería de procesos. Las prácticas podrán hacerse por parejas o individualmente.

BIBLIOGRAFÍA

Básica:

- Aalst, Wil van der. Process mining : data science in action [en línea]. 2nd ed. Berlin: Springer, 2016 [Consulta: 10/01/2024]. Disponible a : <https://ebookcentral-proquest-com.recursos.biblioteca.upc.edu/lib/upcatalunya-ebooks/detail.action?pq-origsite=primo&docID=4505537>. ISBN 9783662498514.

- Carmona Vargas, Josep; Van Dongen, Boudewijn; Solti, Andreas; Weidlich, Matthias. Conformance checking : relating processes and models. Cham: Springer International Publishing, 2018. ISBN 9783319994130.

Complementaria:

- Dumas, M.; La Rosa, M.; Mendling, J.; Reijers, H.A. Fundamentals of business process management. 2nd ed. Springer, 2018. ISBN 9783662565087.

- Reinkemeyer, Lars. Process mining in action : principles, use cases and outlook [en línea]. Cham: Springer, 2020 [Consulta: 10/01/2024]. Disponible a : <https://ebookcentral-proquest-com.recursos.biblioteca.upc.edu/lib/upcatalunya-ebooks/detail.action?pq-origsite=primo&docID=6134217>. ISBN 9783030401726.

- Weske, Mathias. Business process management: concepts, languages, architectures. 3rd ed. Berlin: Springer, 2019. ISBN 9783662594346.

- Wolfgang Reisig. Understanding petri nets. Berlin: Springer, 2013. ISBN 9783642332777.



RECURSOS

Enlace web:

- <https://www.tf-pm.org/>