

Guía docente

230819 - DSBIO - Ciencia de Datos en Bioinformática y Biología Computacional

Última modificación: 24/03/2026

Unidad responsable:	Escuela Técnica Superior de Ingeniería de Telecomunicación de Barcelona
Unidad que imparte:	707 - ESII - Departamento de Ingeniería de Sistemas, Automática e Informática Industrial.
Titulación:	GRADO EN INGENIERÍA DE TECNOLOGÍAS Y SERVICIOS DE TELECOMUNICACIÓN (Plan 2015). (Asignatura optativa). GRADO EN CIENCIA E INGENIERÍA DE DATOS (Plan 2017). (Asignatura optativa).
Curso: 2026	Créditos ECTS: 6.0 Idiomas: Catalán

PROFESORADO

Profesorado responsable:	ALEXANDRE PERERA LLUNA
Otros:	Primer quadrimestre: ALEXANDRE PERERA LLUNA - 11

COMPETENCIAS DE LA TITULACIÓN A LAS QUE CONTRIBUYE LA ASIGNATURA

Genéricas:

12 CPE N1. CAPACIDAD PARA IDENTIFICAR, FORMULAR Y RESOLVER PROBLEMAS DE INGENIERÍA. Plantear y resolver problemas de ingeniería en el ámbito TIC. Desarrollar un método de análisis y solución de problemas sistemático, crítico y creativo.

11 CDIO N3. Capacidad para aplicar una visión integradora del ciclo de vida completo (concepción, diseño, implementación y operación) de un producto, proceso o servicio en el ámbito TIC: Identificar las necesidades del usuario y elaborar un conjunto de requerimientos del producto-proceso-servicio y unas especificaciones iniciales. Explorar las posibles soluciones y seleccionar la más adecuada. Llevar a cabo el proceso de diseño siguiendo una metodología estandarizada. Evaluar y proponer mejoras al diseño realizado. Tener en cuenta los aspectos económicos y sociales asociados al proyecto y al producto.

Transversales:

04 COE N3. COMUNICACIÓN EFICAZ ORAL Y ESCRITA - Nivel 3: Comunicarse de manera clara y eficiente en presentaciones orales y escritas adaptadas al tipo de público y a los objetivos de la comunicación utilizando las estrategias y los medios adecuados.

05 TEQ N1. TRABAJO EN EQUIPO - Nivel 1: Participar en el trabajo en equipo y colaborar, una vez identificados los objetivos y las responsabilidades colectivas e individuales, y decidir conjuntamente la estrategia que se debe seguir.

06 URI N3. USO SOLVENTE DE LOS RECURSOS DE INFORMACIÓN - Nivel 3: Planificar y utilizar la información necesaria para un trabajo académico (por ejemplo, para el trabajo de fin de grado) a partir de una reflexión crítica sobre los recursos de información utilizados.

07 AAT N2. APRENDIZAJE AUTÓNOMO - Nivel 2: Llevar a cabo las tareas encomendadas a partir de las orientaciones básicas dadas por el profesorado, decidiendo el tiempo que se necesita emplear para cada tarea, incluyendo aportaciones personales y ampliando las fuentes de información indicadas.

07 AAT N3. APRENDIZAJE AUTÓNOMO - Nivel 3: Aplicar los conocimientos alcanzados en la realización de una tarea en función de la pertinencia y la importancia, decidiendo la manera de llevarla a cabo y el tiempo que es necesario dedicarle y seleccionando las fuentes de información más adecuadas.

05 TEQ. TRABAJO EN EQUIPO: Ser capaz de trabajar como miembro de un equipo interdisciplinar ya sea como un miembro más, o realizando tareas de dirección con la finalidad de contribuir a desarrollar proyectos con pragmatismo y sentido de la responsabilidad, asumiendo compromisos teniendo en cuenta los recursos disponibles.

07 AAT. APRENDIZAJE AUTÓNOMO: Detectar deficiencias en el propio conocimiento y superarlas mediante la reflexión crítica y la elección de la mejor actuación para ampliar este conocimiento.

03 TLG. TERCERA LENGUA: Conocer una tercera lengua, que será preferentemente inglés, con un nivel adecuado de forma oral y por escrito y en consonancia con las necesidades que tendrán las tituladas y los titulados en cada enseñanza.

METODOLOGÍAS DOCENTES

El curso está estructurado en tres etapas diferentes. En la primera etapa (sesiones teóricas):

- Clase de inicio (2 horas). Explicación de la dinámica de la clase. Introducción del primer tema
- Las clases posteriores incluirán conferencias introductorias y conferencias avanzadas sobre el uso de bioinformática, bases de datos en línea, análisis de conjuntos de datos, R científica e interfaz con bases de datos y visualización. Se pedirá a los estudiantes que lean y comprendan artículos sobre un análisis específico en el área de bioinformática. Las clases serán principalmente teóricas, entremezcladas con sesiones prácticas interactivas dirigidas a la programación de pequeñas tareas en R mientras se acercan gradualmente a los flujos de trabajo de análisis de datos completos bioinformáticos.
- Este conjunto combinará sesiones sincrónicas y material en línea en forma de videos, el esquema de aprendizaje generalmente adoptará aulas invertidas.

La segunda sección (sesiones de laboratorio):

- Consistirá en sesiones de laboratorio guiadas, realizadas en grupo. Las sesiones de laboratorio cubrirán ejemplos prácticos de los diferentes aspectos del curso. Estas sesiones se realizarán en línea con sesiones de preguntas y respuestas sincrónicamente en horarios de laboratorio.

La tercera etapa:

- Las clases contendrán conferencias magistrales y la actividad estará enfocada al desarrollo de un hackathon orientador.
- Este proyecto de hackathon implica el diseño del enfoque programático en un ámbito biomédico a través de la síntesis y análisis de un conjunto de datos biomédicos. En este proyecto de trabajo en equipo, los estudiantes aprenderán, utilizarán las técnicas de minería de datos introducidas con la integración de R, para extraer información relevante de la base de datos, incluyendo estadísticas y visualización.
- Esta tercera etapa se concibe como un sistema de aprendizaje bajo demanda, donde los estudiantes comprenderán en profundidad el material didáctico y completarán los conocimientos requeridos para el proyecto asignado a cada equipo.
- La tercera etapa finaliza con una presentación conjunta con los resultados del hackathon. Esta presentación se realizará presencial o online según el índice de riesgo sars-cov-2.

OBJETIVOS DE APRENDIZAJE DE LA ASIGNATURA

- Obtener conocimientos generales sobre los principios de la biología, la bioinformática y la ciencia clínica básica
- Comprender los problemas relacionados con las bases de datos de bioinformática y su estructura
- Adquirir fluidez en el software R como lenguaje de programación de ciencias de datos biomédicos
- Adquirir habilidades para el análisis estadístico de una enfermedad, comprendiendo el papel de diferentes factores
- Aprender a implementar las mejores prácticas para el análisis de datos y su gestión
- Adquirir habilidades de acceso programático a bases de datos biológicas mediante R
- Demostrar mediante un proyecto la comprensión y análisis de conjuntos de datos de marcadores biológicos
- Aprender a explicar detalladamente información compleja, incluidas datos técnicos y resultados, a un público amplio
- Aprender la capacidad de trabajar bien en equipo y comunicarse eficazmente tanto en lugares verbales como escritos
- Evitar las creencias conductuales y aprender a actuar y comprender la realidad a partir de evidencia
- Reforzar el pensamiento profesional y crítico
- Realizar investigación de una manera ética que mantenga la seguridad y la privacidad de los datos

HORAS TOTALES DE DEDICACIÓN DEL ESTUDIANTADO

Tipo	Horas	Porcentaje
Horas aprendizaje autónomo	98,0	65.33
Horas grupo grande	26,0	17.33
Horas grupo pequeño	26,0	17.33

Dedicación total: 150 h

CONTENIDOS

Ciencia de datos en bioinformática y biología computacional

Descripción:

1. Introducción.
2. Introducción a la genética
 1. La diversidad de vida y material genético
2. Genes
3. Estructura del ADN
4. Transferencia de información entre el ADN y las proteínas.
5. ¿Cómo se analiza el ADN?
6. Fuentes de variabilidad genética. Entre individuos, entre especies
3. Fuentes de datos y algoritmos
 1. Organización y niveles de abstracción
 2. Bases de datos de genética
 3. Ontologías genéticas
 4. Lenguajes de programación con orientación bio.
5. Fuentes de algoritmos
4. Análisis de secuencia
 1. Medidas de distancia entre secuencias
 2. Modelos por alineación.
 3. Needleman-Wunsch
 4. Algoritmos de Smith-Waterman, BLAST
 5. Alineación múltiple. Algoritmos e interpretación
 5. Análisis de la expresión genética
 1. Tecnologías para la medición de la expresión genética
- una. Micro-matrices
- B. Secuenciación de alto rendimiento
 2. Flujo de trabajo de análisis de datos para datos de expresión génica
 3. Introducción al reconocimiento de patrones para la expresión génica
 4. Aprendizaje supervisado y no supervisado
 5. Visualización avanzada
 6. Reducción de dimensionalidad
 7. Métodos de asociación
6. Análisis funcional
 1. Introducción al análisis funcional
 2. Análisis de sobrerepresentación
 3. GSEA
 4. Métodos basados en redes
 5. Aplicaciones en otras ómicas (metabolómica)
 7. Análisis de estructuras proteicas
 1. Estructura primaria
 2. Estructura secundaria y terciaria o 3D
 3. Predicción de estructuras secundarias
 4. Predicción de estructuras terciarias
 5. Análisis de estructuras proteicas
 8. Propuesta de hackathon

Actividades vinculadas:

1. Introducción a R para bioinformática
2. Ensembl y NCBI. Acceso a bases de datos biomédicas con R y bioconductor
3. Homología
4. Introducción a R (III). Reconocimiento de patrones, análisis supervisado y no supervisado
5. Introducción a R (IV). Reconocimiento de patrones, análisis supervisado y no supervisado (si es necesario)
6. Análisis de expresión génica con R y Bioconductor
7. Enriquecimiento funcional

Dedicación: 55h

Grupo grande/Teoría: 28h

Grupo pequeño/Laboratorio: 16h
Actividades dirigidas: 11h

SISTEMA DE CALIFICACIÓN

Cualquier acto de fraude académico, plagio o uso o mera tenencia al alcance de medios no autorizados en cualquier actividad de evaluación comportará la calificación de cero (0) en la prueba o entrega afectada. Además, de acuerdo con la normativa de la Universidad, la posible derivación de los hechos para la apertura de un expediente disciplinario implicará que la asignatura quede en el estado provisional de "pendiente de evaluación" hasta la resolución del expediente. La gestión de estas incidencias se lleva a cabo de acuerdo con el [Marco de actuación para la integridad académica en la evaluación de la UPC](#).

La calificación se construirá a partir de una evaluación continua, que incluye:

- 60% para pequeñas asignaciones individuales e informe de laboratorio (aproximadamente 2 asignaciones más laboratorios)
- 40% calificado de una tarea final individual, dedicada al análisis de un conjunto de datos de biomarcadores, presentación de resultados y presentación.

BIBLIOGRAFÍA

Básica:

- Jones, Neil C.; Pevzner, Pavel A. An Introduction to bioinformatics algorithms [en línea]. Cambridge, MA: MIT Press, cop. 2004 [Consulta: 30/06/2021]. Disponible a: <https://ebookcentral.proquest.com/lib/csuc-ebooks/detail.action?docID=3338847>. ISBN 0262101068.
- Durbin, Richard. Biological sequence analysis : probabilistic models of proteins and nucleic acids [en línea]. Cambridge [etc.]: Cambridge University Press, 1998 [Consulta: 30/06/2021]. Disponible a: <https://ebookcentral.proquest.com/lib/csuc-ebooks/detail.action?docID=320915>. ISBN 0521629713.
- Gentleman, Robert. Bioinformatics and computational biology solutions using R and Bioconductor [en línea]. New York: Springer Science+Business Media, cop. 2005 [Consulta: 30/06/2021]. Disponible a: <http://link.springer.com/openurl?genre=book&isbn=978-0-387-25146-2>. ISBN 0387251464.
- Wickham, Hadley; Golemund, Garrett. R for data science : import, tidy, transform, visualize, and model data. Beijing: O'Reilly, 2016. ISBN 9781491910399.