

Guía Docente: Análisis inteligente en biomedicina, biotecnología, salud y actividades deportivas

DATOS GENERALES	
Facultad	Facultad de Ciencias y Tecnología
Titulación	Máster en Análisis Inteligente de Datos (Big Data)
Plan de estudios	2018
Materia	Aplicaciones del análisis inteligente de datos masivos
Carácter	Obligatorio
Período de impartición	Segundo Trimestre
Curso	Primero
Nivel/Ciclo	Máster
Créditos ECTS	6
Lengua en la que se imparte	Castellano
Prerrequisitos	No se prevén requisitos previos; por tanto, los requisitos serán los propios del título.

DATOS DEL PROFESORADO			
Profesor Responsable	Pedro Manuel Martínez García	Correo electrónico	pedromanuel.martinez@ui1.es
Área		Facultad	Facultad de Ciencias y Tecnología
Perfil Profesional 2.0	LinkedIn Twitter		
Profesor	Victor Martinez Martinez	Correo electrónico	victor.martinez.martinez@ui1.es
Área		Facultad	Facultad de Ciencias y Tecnología
Perfil Profesional 2.0	ResearchGate LinkedIn		

CONTEXTUALIZACIÓN Y JUSTIFICACIÓN DE LA ASIGNATURA

Asignaturas de la materia	<ul style="list-style-type: none"> • Análisis de Redes Sociales y Colaborativas • Inteligencia de Negocio y Decisiones Estratégicas • Análisis inteligente en biomedicina, biotecnología, salud y actividades deportivas • Análisis Inteligente aplicado a la Industria Digital
Contexto y sentido de la asignatura en la titulación y perfil profesional	<p>El creciente desarrollo de las técnicas de secuenciación masiva de ADN ha dado lugar a la producción de ingentes cantidades de información biológica, requiriéndose un conocimiento específico tanto de la tipología de la información generada como de los métodos computacionales necesarios para la apropiada extracción de información relevante. Por otro lado, los ámbitos deportivo y de salud no son ajenos a la transformación digital que experimentamos en la actualidad, generando a su vez enormes cantidades de datos. Es por ello que en el contexto de las Ciencias Experimentales la demanda de profesionales capaces de gestionar dicha información crece en consonancia con la exponencial producción de la misma.</p> <p>Esta asignatura tiene por objeto instruir al alumnado en la adquisición, procesado y análisis de datos generados tanto en el ámbito biotecnológico como en el deportivo. Los contenidos que se imparten se relacionan con otras materias de la titulación como <i>Análisis estadístico y visualización de datos</i>, así como con materias propias de la <i>Ingeniería Informática</i>, en especial la programación en lenguaje R.</p>

COMPETENCIAS QUE ADQUIERE EL ESTUDIANTE Y RESULTADOS DE APRENDIZAJE

<p>Competencias de la asignatura</p>	<ul style="list-style-type: none"> • CB6: 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: Saber aplicar los conocimientos adquiridos y ser capaz de resolver problemas en entornos nuevos o poco conocidos dentro de contextos más amplios (o multidisciplinares) relacionados con su área de estudio. • CB8: Ser capaz 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: Ser capaz de transmitir sus conclusiones, y los conocimientos y fundamentos que las sustentan, a públicos especializados y no especializados de un modo claro y sin ambigüedades. • CB10: 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. • CG1: Conocer la legislación en materia de datos personales, privacidad y derechos fundamentales de las personas, interoperabilidad, propiedad intelectual y reutilización de datos. • CG2: Ser capaz de permanecer eficaz dentro de un medio cambiante, así como a la hora de enfrentarse con nuevas tareas, retos y personas • CG5: Diseñar y desarrollar la implementación y puesta en marcha de proyectos de bigdata en diferentes áreas de aplicación social y profesional. • CG6: Ser capaz de integrarse en equipos de trabajo o investigación multidisciplinares de manera eficaz y colaborativa. • CE02: Identificar y utilizar herramientas software especializadas para el tratamiento de grandes volúmenes de datos en distintos contextos. • CE03: Saber diseñar y desarrollar soluciones en lenguajes y entornos de programación especializados en big data. • CE04: Comprender y saber aplicar y analizar técnicas estadísticas avanzadas y herramientas para el análisis y tratamiento de datos masivos. • CE06: Comprender y utilizar técnicas avanzadas de visualización de datos y de experiencia de usuario para el diseño e implementación de interfaces adaptadas al usuario en los procesos de análisis de bigdata en distintos contextos. • CE07: Saber realizar el análisis y seguimiento de la estrategia en una empresa. Comprender el concepto de cuadro de mando, así como los indicadores que lo componen. • CE09: Conocer y aplicar análisis inteligente de datos masivos en casos reales de estudio, predicción y toma de decisiones en las áreas de biomedicina, bioquímica y biotecnología.
<p>Resultados de aprendizaje de la asignatura</p>	<ul style="list-style-type: none"> • Conocer las características de la tipología de los datos en biomedicina, biotecnología y actividades físico-deportivas. • Realizar la extracción y filtrado de datos adecuados a su formato utilizando herramientas apropiadas. • Comprender, diseñar, desarrollar y evaluar la realización de un caso práctico de aplicación inteligente de análisis biomédico. • Comprender, diseñar, desarrollar y evaluar la realización de un caso práctico de aplicación inteligente de análisis biotecnológico. • Comprender, diseñar, desarrollar y evaluar la realización de un caso práctico de aplicación inteligente de análisis de actividades físico-deportivas.

PROGRAMACION DE CONTENIDOS

Breve descripción de la asignatura

En esta asignatura se verá entre otros:

- Tipología de las fuentes de datos en biomedicina, biotecnología, salud y actividades deportivas
- Disponibilidad y acceso a datos biomédicos
- Herramientas de tratamiento masivo de datos
 - Biotecnología.
 - Biomedicina
 - Salud.
 - Actividades deportivas
- Dispositivos móviles en biomedicina, biotecnología, salud y actividades deportivas.
- Casos prácticos:
 - Biomedicina
 - Biotecnología
 - Actividad deportiva

Contenidos

1. Introducción a la tipología de datos biotecnológicos

- Nociones de biología molecular
 - Vida y células
 - Biomoléculas
 - Proteínas
 - Ácidos nucleicos
 - Genética básica
- Secuenciación de ADN
 - Formato FASTQ
 - Formato FASTA
- Tratamiento de secuencias biológicas en

2. Análisis de secuencias biológicas

- Comparación de secuencias
- Alineamiento de secuencias
 - Alineamiento pareado
 - Sistemas de puntuación: La matriz de sustitución
 - Métodos de alineamiento
 - Alineamiento múltiple
 - Algoritmos progresivos
 - Ensamblado y mapeo

- Conceptos básicos
- Ensamblado
- Mapeo
- Alineamiento de secuencias en R
 - Alineamiento pareado: Función *pairwiseAlignment* de *Biostrings*
 - Alineamiento múltiple: La librería *msa*

3. Bases de datos biológicas

- BBDD de secuencias: GenBank, Uniprot
- BBDD específicas: Flybase, TAIR, SGD, MGD, IMG, PFAM, JASPAR
- Proyectos colaborativos de colección de datos
- BBDD bibliográficas: PubMed
- Acceso a bases de datos biológicas mediante R
 - Acceso a GenBank
 - Librería *genbankr*
 - Librería *rentrez*
 - Acceso a genomas completos y sus anotaciones
 - Librerías *TxDb*
 - Librerías *EnsDb*
 - Librerías *BSgenome*
 - Acceso a JASPAR
 - Librerías *TFBSTools* y *JASPAR2018*

4. Introducción a *sports analytics* (analítica deportiva)

- Casos prácticos aplicados de Sports Analytics en el deporte profesional.
- Futuro en Sports Analytics.

5. Análisis de datos deportivos

- Tipología de análisis dentro del mundo deportivo.
- Proceso analítico riguroso para conseguir "Ganar".

6. Bases de datos deportivas

- Bases de Datos Deportivas.
- Librerías de R aplicadas al mundo deportivo.

METODOLOGÍA

Actividades formativas

El proceso de enseñanza-aprendizaje de la asignatura está estructurado en base a 6 unidades didácticas, para cada una de las cuales el alumno dispone del material de estudio en formato interactivo así como de material complementario a modo de blogs, vídeos, etc.

Mediante un conjunto de actividades (calificables y no calificables) tutorizadas por el profesor en todo el proceso de trabajo, se pretende transformar los contenidos en aprendizajes. Dichas actividades se estructuran en los siguientes bloques:

- **Foros de debate**

Se incluyen foros para promover el pensamiento crítico y el contraste de puntos de vista del alumnado sobre temas de actualidad en el tratamiento de datos biotecnológicos. Dado el creciente desarrollo de las técnicas de secuenciación de ADN, el buen uso de los datos generados ha abierto nuevos debates sobre el tratamiento de la información genómica personalizada y las consecuencias que puede tener el acceso público y privado a los bancos de datos generados. Con los foros de debate se pretende que el alumnado tome parte activa en la discusión sobre este nuevo tipo de problemática de ámbito bioético.

- **Contenidos prácticos**

Los contenidos teóricos de cada unidad didáctica se complementan con la correspondiente aplicación práctica de los mismos. En general se hará uso del lenguaje de programación R para ilustrar los conceptos teóricos desarrollados en el material de estudio mediante ejemplos concretos de procesamiento y análisis de datos genómicos. En ocasiones se usarán también programas bioinformáticos específicos de análisis de datos.

- **Ejercicios prácticos**

El alumnado deberá realizar una serie de ejercicios prácticos relacionados con los contenidos prácticos de las unidades didácticas. Se deberá usar para ello las distintas librerías del lenguaje de programación R o alternativos que se han ilustrado en los contenidos.

- **Cuestionarios**

A través de preguntas tipo test, los cuestionarios servirán al alumnado como herramienta de autoevaluación de los contenidos expuestos en las diferentes unidades didácticas.

- **Videotutorías**

Sesiones de videotutoría *online* para mostrar casos reales aplicados del contenido de la asignatura, así como poder aclarar con el docente dudas de la asignatura. Se realizarán sesiones en vivo que se grabarán para su posterior consulta dentro del aula por parte de

los alumnos. Este tipo de actividades son no calificables.

- **Estudio de Caso**

Presentación de una situación motivadora que introduzca de manera atractiva y sugerente en una parcela de conocimiento, se plantea una posibilidad que pueda darse en la realidad en torno al tipo de conocimientos propios de la unidad didáctica. La presentación del caso al alumnado se asocia como una serie concatenada de preguntas, se le sugieren consultas, se le suministran textos, imágenes, gráficos... con datos suficientes como para que pueda ofrecer una solución o llegar a unas conclusiones lógicas.

- **Trabajo de investigación**

Los alumnos deberán realizar una documentación e investigación acerca de un tema propuesto por el docente. Como resultado del mismo, se deberá realizar una memoria que contenga los resultados de la investigación desarrollada.

- **Foros de dudas**

Se incluyen foros en cada unidad didáctica para que los alumnos puedan trasladar sus dudas y ponerlas en común con el resto de integrantes del aula. Este tipo de actividades, al igual que las videotutorías, son no calificables.

EVALUACIÓN

Sistema evaluativo

En caso de que la situación sanitaria impida la realización presencial de los exámenes con todas las garantías, la Universidad Isabel I celebrará dichas pruebas en modalidad online. Para la realización de dichos exámenes, la universidad incorporará la herramienta de proctoring a nuestra plataforma tecnopedagógica, con el objetivo de garantizar los procesos de autenticación del alumno, como el control del entorno durante el desarrollo de las pruebas de evaluación. A su vez, la Universidad Isabel I pondrá a disposición del alumnado una Unidad de Exámenes Online específica para ofrecer apoyo técnico durante todo el proceso y así solventar todas las incidencias que se puedan presentar.

El sistema de evaluación se basará en una selección de las pruebas de evaluación más adecuadas para el tipo de competencias que se trabajen. El sistema de calificaciones estará acorde con la legislación vigente (*Real Decreto 1125/2003, de 5 de septiembre, por el que se establece el sistema europeo de créditos y el sistema de calificaciones en las titulaciones universitarias de carácter oficial y de validez en todo el territorio nacional*).

El sistema de evaluación de la Universidad Isabel I queda configurado de la siguiente manera:

Sistema de evaluación convocatoria ordinaria

Opción 1. Evaluación continua

Los estudiantes que opten por esta vía de evaluación deberán realizar el **seguimiento de la evaluación continua (EC)** y podrán obtener hasta un **60 %** de la calificación final a través de las actividades que se plantean en la evaluación continua.

Además, deberán realizar un **examen final presencial (EX)** que supondrá el **40 %**

restante. Esta prueba tiene una parte dedicada al control de la identidad de los estudiantes que consiste en la verificación del trabajo realizado durante la evaluación continua y otra parte en la que realizan diferentes pruebas teórico-prácticas para evaluar las competencias previstas en cada asignatura.

Para la aplicación de los porcentajes correspondientes, el estudiante debe haber obtenido una nota mínima de un 4 en cada una de las partes de las que consta el sistema de evaluación continua.

Se considerará que el estudiante supera la asignatura en la convocatoria ordinaria por el sistema de evaluación continua, siempre y cuando al aplicar los porcentajes correspondientes se alcance una calificación mínima de un 5.

Opción 2. Prueba de evaluación de competencias

Los estudiantes que opten por esta vía de evaluación deberán realizar una **prueba de evaluación de competencias (PEC)** y un **examen final presencial (EX)**.

La **PEC** se propone como una prueba que el docente plantea con el objetivo de evaluar en qué medida el estudiante adquiere las competencias definidas en su asignatura. Dicha prueba podrá ser de diversa tipología, ajustándose a las características de la asignatura y garantizando la evaluación de los resultados de aprendizaje definidos. Esta prueba supone el 50 % de la calificación final.

El **examen final presencial**, supondrá el **50 %** de la calificación final. Esta prueba tiene una parte dedicada al control de la identidad de los estudiantes que consiste en la verificación del seguimiento de las actividades formativas desarrolladas en el aula virtual y otra parte en la que realizan diferentes pruebas teórico-prácticas para evaluar las competencias previstas en cada asignatura.

Al igual que con el sistema de evaluación anterior, para la aplicación de los porcentajes correspondientes el estudiante debe haber obtenido una puntuación mínima de un 4 en cada una de las partes de las que consta la opción de prueba de evaluación de competencias.

Se considerará que el estudiante supera la asignatura en la convocatoria ordinaria por el sistema de la prueba de evaluación de competencias siempre y cuando al aplicar los porcentajes correspondientes se alcance una calificación mínima de un 5.

Características de los exámenes

Los exámenes constarán de 30 ítems compuestos por un enunciado y cuatro opciones de respuesta, de las cuales solo una será la correcta. Tendrán una duración de 90 minutos y la calificación resultará de otorgar 1 punto a cada respuesta correcta, descontar 0,33 puntos por cada respuesta incorrecta y no puntuar las no contestadas. Después, con el resultado total, se establece una relación de proporcionalidad en una escala de 10.

Sistema de evaluación convocatoria extraordinaria

Todos los estudiantes, independientemente de la opción seleccionada, que no superen las pruebas evaluativas en la convocatoria ordinaria tendrán derecho a una convocatoria extraordinaria.

La convocatoria extraordinaria completa consistirá en la realización de una **prueba de**

evaluación de competencias que supondrá el **50 %** de la calificación final y un **examen final presencial** cuya calificación será el **50 %** de la calificación final.

Para la aplicación de los porcentajes correspondientes, el estudiante debe haber obtenido una nota mínima de un 4 en cada una de las partes de las que consta el sistema de evaluación de la convocatoria extraordinaria.

Los estudiantes que hayan suspendido todas las pruebas evaluativas en convocatoria ordinaria (evaluación continua o prueba de evaluación de competencias y examen final) o no se hayan presentado deberán realizar la convocatoria extraordinaria completa, como se recoge en el párrafo anterior.

En caso de que hayan alcanzado una puntuación mínima de un 4 en alguna de las pruebas evaluativas de la convocatoria ordinaria (evaluación continua o prueba de evaluación de competencias y examen final), se considerará su calificación para la convocatoria extraordinaria, debiendo el estudiante presentarse a la prueba que no haya alcanzado dicha puntuación o que no haya realizado.

En el caso de que el alumno obtenga una puntuación que oscile entre el 4 y el 4,9 en las dos partes de que se compone la convocatoria ordinaria (EC o PEC y examen), solo se considerará para la convocatoria extraordinaria la nota obtenida en la evaluación continua o prueba de evaluación de competencias ordinaria (en función del sistema de evaluación elegido), debiendo el alumno realizar el examen extraordinario para poder superar la asignatura.

Al igual que en la convocatoria ordinaria, se entenderá que el alumno ha superado la materia en convocatoria extraordinaria si, aplicando los porcentajes correspondientes, se alcanza una calificación mínima de un 5.

BIBLIOGRAFÍA Y OTROS RECURSOS

Bibliografía básica

- Durbin, R., Eddy, S., Krogh, A. y Mitchison, G. (1998). *Biological Sequence Analysis: Probabilistic Models of Proteins and Nucleic Acids*. Cambridge University Press, 356 pp. doi: 10.1017/CBO9780511790492.
- Miller TW. (2015) *Sports Analytics and Data Science: Winning the Game with Methods and Models*. FT Press Analytics.

Bibliografía complementaria

- Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K. y Walter, P. (2008). *Molecular biology of the cell*, 5th edition. Biochemistry and Molecular Biology Education. doi: 10.1002/bmb.20192.
- Diniz WJ y Canduri F. (2017). *Bioinformatics: an overview and its applications*.
- Heather, J.M. y Chain, B. (2016). The sequence of sequencers: The history of sequencing DNA. *Genomics*, 107(1): 1-8. doi: 10.1016/j.ygeno.2015.11.003.
- Luscombe NM, Greenbaum D y Gerstein M. (2001). What is bioinformatics? An introduction and overview.
- Martin L. (2016) *Sports Performance Measurement and Analytics: The Science of Assessing Performance, Predicting Future Outcomes, Interpreting Statistical Models, and Evaluating the Market Value of Athletes*. Pearson FT Press.
- Zhang, H. (2016). Overview of Sequence Data Formats. *Methods in Molecular Biology*, 418:3-17. doi: 10.1007/978-1-4939-3578-9_1.
- Zou, D., Ma, L., Yu, J. y Zhang, Z. (2015). Biological databases for human research. *Genomics Proteomics Bioinformatics*, 13(1): 55–63. doi: 10.1016/j.gpb.2015.01.006.

Otros recursos

- NBA Fantasy Dashboard
<http://dataremixed.com/2014/01/an-nba-fantasy-dashboard/>
- Open Data: [Link Web](#)
 - 7 Fuentes de datos gratuitas para el Big Data Analytics que todo el mundo debería conocer: [Link Web](#)
- Putting Data to Work: Lessons from "Moneyball" (Paul DePodesta) [Link web](#)
- Video completo de la analítica avanzada y pionera realizada por el FC Barcelona:
 - https://www.youtube.com/watch?v=6_JnO3LrkhI
- Forbes Releases 2020 List Of The World's Most Valuable Sports Teams [Link Web](#)
- Ejemplo de apuestas deportivas:
<https://noeliagorod.com/2018/06/26/2018-world-cup-predictions-using-different-methodology>
- ¿Quién va a ganar el próximo mundial de fútbol?
<https://noeliagorod.com/2018/04/26/quien-va-a-ganar-el-proximo-mundial-de-futbol/>
- Campeona del mundo 200 mariposa. Felicidades Mireia Belmonte. Data Science y la Natación <https://noeliagorod.com/2017/07/28/campeona-del-mundo-200-mariposa-felicidades-mireia-belmonte-data-science-y-la-natacion/>
- El Big Data esta cambiando el Baloncesto <https://noeliagorod.com/2017/03/10/el-big-data-esta-cambiando-el-baloncesto/>
- Un ejemplo de reconocimiento facial aplicado en un evento (no deportivo, pero si de entretenimiento): Taylor Swift tracked stalkers with facial recognition tech at her concert.
<https://noeliagorod.com/2019/01/11/taylor-swift-tracked-stalkers-with-facial-recognition-tech-at-her-concert/>
- Video completo de la analítica avanzada y pionera realizada por el FC Barcelona:
 - https://www.youtube.com/watch?v=6_JnO3LrkhI
- Blog @noeliagorod: www.noeliagorod.com

Librerías de R Bioconductor:

<http://bioconductor.org/packages/release/bioc/html/Biostrings.html>

<https://bioconductor.org/packages/release/bioc/vignettes/ShortRead/inst/doc/Overview.pdf>

<https://bioconductor.org/packages/release/bioc/vignettes/GenomicRanges/inst/doc/GenomicRangesIntroduction.html>

<https://bioconductor.org/packages/release/bioc/vignettes/GenomicFeatures/inst/doc/GenomicFeatures.pdf>

<https://bioconductor.org/packages/release/bioc/vignettes/msa/inst/doc/msa.pdf>

<https://www.rdocumentation.org/packages/seqinr/versions/3.4-5>