



Guía Docente				
Datos Identificativos				2018/19
Asignatura (*)	Intelixencia computacional para datos de alta dimensionalidade	Código	614522024	
Titulación	Mestrado Universitario en Bioinformática para Ciencias da Saúde			
Descritores				
Ciclo	Período	Curso	Tipo	Créditos
Mestrado Oficial	1º cuatrimestre	Segundo	Obrigatoria	3
Idioma	CastelánInglés			
Modalidade docente	Presencial			
Prerrequisitos				
Departamento	Computación			
Coordinación	Alonso Betanzos, Maria Amparo	Correo electrónico	amparo.alonso.betanzos@udc.es	
Profesorado	Alonso Betanzos, Maria Amparo Bolón Canedo, Verónica	Correo electrónico	amparo.alonso.betanzos@udc.es veronica.bolon@udc.es	
Web	moodle.udc.es			
Descrición xeral	Nesta materia traballarase nos fundamentos e aplicación práctica das bases de datos de alta dimensión e na aplicación de técnicas de minería de datos no ámbito da bioinformática			

Competencias do título	
Código	Competencias do título
A2	CE2 - Definir, avaliar e seleccionar a arquitectura e o software máis axeitado para resolver un problema no campo da Bioinformática
A3	CE3 - Analizar , deseñar , desenvolver, implementar , verificar e documentar solucións software eficientes sobre a base dun coñecemento adecuado das teorías, modelos e técnicas actuais no eido da Bioinformática
A4	CE4 - Capacidade para adquirir, obter, formalizar e representar o coñecemento humano nunha forma computable para a resolución de problemas mediante un sistema informático en calquera ámbito de aplicación, particularmente os relacionados con aspectos de computación, percepción e actuación en aplicacións Bioinformáticas
A6	CE6 ? Capacidade para identificar as ferramentas software e fontes de datos de bioinformática máis relevantes, e adquirir destreza no seu uso
B1	CB6 ? Posuír e comprender o coñecemento que fornecen unha base ou oportunidade de orixinalidade no desenvolvemento e / ou aplicación de ideas, a miúdo nun contexto de investigación.
B2	CB7 - Que os estudantes saiban aplicar os coñecementos adquiridos e a súa capacidade de resolución de problemas en contornas novas ou pouco coñecidas dentro de contextos máis amplos (ou multidisciplinares) relacionados coa súa área de estudo
B3	CB8 ? Que os estudantes sexan capaces de integrar coñecementos e xestionar a complexidade de formular xuízos en base a información que, sendo incompleta ou limitada, inclúa reflexións sobre as responsabilidades sociais e éticas relacionadas coa aplicación dos seus coñecementos e xuízos
B6	CG1 - Buscar e seleccionar a información útil necesaria para resolver problemas complexos, manexando con soltura as fontes bibliográficas do campo
B7	CG2 - Manter e estender enfoques teóricos fundados para permitir a introdución i explotación de tecnoloxías novas e avanzadas
C1	CT1 - Expresarse correctamente, tanto de xeito oral como escrito, nas linguas oficiais da comunidade autónoma
C3	CT3 - Utilizar as ferramentas básicas das tecnoloxías da información e as comunicacións (TIC) necesarias para o exercicio da súa profesión e para a aprendizaxe ao longo da súa vida
C6	CT6 - Valorar criticamente o coñecemento, a tecnoloxía e a información dispoñibles para resolver os problemas cos que deben enfrontarse

Resultados da aprendizaxe	
Resultados de aprendizaxe	Competencias do título



Coñecer e comprender os paradigmas e aspectos máis relevantes do tratamento de bases de datos de alta dimensión	AP2 AP3 AP4 AP6	BP1 BP2 BP3 BP6 BP7	CP1 CP3 CP6
Coñecer e saber aplicar os principais métodos de minería de datos; coñecer as plataformas e as paradigmas principais que se empregan no campo.	AP2 AP3 AP4 AP6	BP1 BP2 BP3 BP6 BP7	CP1 CP3 CP6

Contidos	
Temas	Subtemas
Introducción ao Big data.	Qué é Big Data Principais características do Big data Principais campos de aplicación
Minería de datos e alta dimensión	Analítica Big data Técnicas de preprocesado MapReduce
Modelos de programación Batch	Hadoop Resilient Distributed datasets Programación batch en Spark
Modelos de programación streaming	Conceptos básicos Kafka, Apache Storm, Spark streaming

Planificación				
Metodoloxías / probas	Competencias	Horas presenciais	Horas non presenciais / traballo autónomo	Horas totais
Sesión maxistral	A4 C1 C6	7	14	21
Solución de problemas	A25 A33 A41 B1 B6 C3	8	16	24
Traballos tutelados	A21 B3 B6 C1 C2 C3 C6	4	4	8
Seminario	A21 B1 B3 B6	4	4	8
Proba mixta	A2 A3 A4 A6 B1 B2 B3 B6 B7 C1 C3 C6	4	10	14
Atención personalizada		0	0	0

\*Os datos que aparecen na táboa de planificación son de carácter orientativo, considerando a heteroxeneidade do alumnado

Metodoloxías	
Metodoloxías	Descrición
Sesión maxistral	Empregada durante as clases presenciais teóricas para expor o núcleo básico de coñecementos que logo os alumnos terán que saber utilizar e ampliar nas prácticas, seminarios e traballos do curso
Solución de problemas	Emprego de técnicas de minería de datos en alta dimensión. Uso de paradigmas Big data Realización dunha práctica nunha plataforma específica de Big data
Traballos tutelados	Entrega dun breve traballo que discutirse na clase sobre algún aspecto concreto da materia.
Seminario	Exposición dun traballo específico de investigación que involucre tecnoloxías de alta dimensionalidade
Proba mixta	Realizarse ao final do cuadrimestre sobre os contidos tratados ao longo do curso.



## Atención personalizada

Metodoloxías	Descrición
Seminario Solución de problemas Traballos tutelados Proba mixta Sesión maxistral	No esquema de carácter práctico utilizado nesta materia, as tutorías resultan un recurso fundamental moi empregado polos alumnos, sobre todo debido á complexidade dalgúns conceptos da materia, en función das titulacións de entrada dos diferentes alumnos.  Os alumnos poden realizar dous tipos de tutorías: virtuais e presenciais. As primeiras poden utilizalas para dúbidas moi concretas de resposta rápida. As máis comúns iranse depositando nun apartado de %"Preguntas Frecuentes" que deberán consultar antes de enviar unha nova pregunta.

## Avaliación

Metodoloxías	Competencias	Descrición	Cualificación
Seminario	A21 B1 B3 B6		0
Traballos tutelados	A21 B3 B6 C1 C2 C3 C6	Nota correspondente á parte práctica da materia, que comprende tanto os desenvolvementos realizados sobre as plataformas, como os traballos entregados.	50
Proba mixta	A2 A3 A4 A6 B1 B2 B3 B6 B7 C1 C3 C6	Realizarase unha proba con cuestións relativas ás partes teóricas da materia	50
Sesión maxistral	A4 C1 C6		0

## Observacións avaliación

--

## Fontes de información

<b>Bibliografía básica</b>	<ul style="list-style-type: none"> <li>- Vladimir Bacvanski. (2015). Introduction to Big Data An Overview of Fundamental Big Data Concepts, Tools, Techniques and Practices.. O'Reilly Media</li> <li>- Venkat Ankam (2016.). Big Data Analytics. Packt Publishing</li> <li>- Tom White (2015). Hadoop: The Definitive Guide. O'Reilly Media</li> <li>- Thilina Gunarathne (2015). Hadoop MapReduce v2 Cookbook. Packt Publishing</li> <li>- Holden Karau, Andy Konwinski, Patrick Wendell, Matei Zaharia (2015). Learning Spark. O'Reilly Media</li> <li>- Sean T. Allen, Matthew Jankowski, and Peter Pathirana (2015). Storm Applied. . O'Reilly Media</li> </ul>
<b>Bibliografía complementaria</b>	

## Recomendacións

### Materias que se recomenda ter cursado previamente

Intelixencia computacional para bioinformática/614522012  
 Métodos estatísticos avanzados en bioinformática/614522009  
 Computación de altas prestacións en bioinformática/614522011  
 Introducción á programación/614522001  
 Fundamentos de intelixencia artificial/614522003

### Materias que se recomenda cursar simultaneamente

--

### Materias que continúan o temario

--

## Observacións

--



(\*)A Guía docente é o documento onde se visualiza a proposta académica da UDC. Este documento é público e non se pode modificar, salvo casos excepcionais baixo a revisión do órgano competente dacordo coa normativa vixente que establece o proceso de elaboración de guías