



Guía Docente				
Datos Identificativos				2019/20
Asignatura (*)	Tecnoloxías de Xestión de Datos		Código	614493106
Titulación	Mestrado Universitario en Técnicas Estadísticas (Plan 2019)			
Descritores				
Ciclo	Período	Curso	Tipo	Créditos
Mestrado Oficial	1º cuatrimestre	Primeiro	Optativa	5
Idioma	Castelán			
Modalidade docente	Presencial			
Prerrequisitos				
Departamento	Enxeñaría de Computadores Matemáticas			
Coordinación	López Taboada, Guillermo	Correo electrónico	guillermo.lopez.taboada@udc.es	
Profesorado	Fernández Casal, Rubén	Correo electrónico	ruben.fcasal@udc.es	
	López Taboada, Guillermo		guillermo.lopez.taboada@udc.es	
Web	eio.usc.es/pub/mte			
Descrición xeral	O obxectivo deste curso é que calquera estudante, independentemente da súa formación académica, adquira un sólido coñecemento de tecnoloxías de xestión de bases de datos, tanto relacionais coma non relacionais. Así mesmo, buscarase a familiarización coas principais técnicas computacionais para a xestión práctica de datos masivos. Isto proporcionará a/a estudante unha gran autonomía cando se trata de procesar e estudar datos, independentemente do seu formato e orixe.			

Competencias / Resultados do título	
Código	Competencias / Resultados do título
A16	CE1 - Coñecer, identificar, modelar, estudar e resolver problemas complexos de estatística e investigación operativa, nun contexto científico, tecnolóxico ou profesional, xurdidos en aplicacións reais.
A17	CE2 ? Desenvolver autonomía para a resolución práctica de problemas complexos surtidos en aplicación reais e para a interpretación dos resultados cara á axuda na toma de decisións.
A18	CE3 - Adquirir coñecementos avanzados dos fundamentos teóricos subxacentes ás distintas metodoloxías da estatística e a investigación operativa, que permitan o seu desenvolvemento profesional especializado.
A21	CE6 - Adquirir coñecementos teórico-prácticos avanzados de distintas técnicas matemáticas, orientadas especificamente á axuda na toma de decisións, e desenvolver a capacidade de reflexión para avaliar e decidir entre distintas perspectivas en contextos complexos.
A23	CE8 - Adquirir coñecementos teórico-prácticos avanzados das técnicas destinadas á realización de inferencias e contrastes relativos a variables e parámetros dun modelo estatístico, e saber aplicarlos con autonomía suficiente nun contexto científico, tecnolóxico ou profesional.
A24	CE9 - Coñecer e saber aplicar con autonomía en contextos científicos, tecnolóxicos ou profesionais, técnicas de aprendizaxe automático e técnicas de análise de datos de alta dimensión (big data).
A25	CE10 - Adquirir coñecementos avanzados sobre metodoloxías para a obtención e o tratamento de datos derivados de distintas fontes, como enquisas, internet, ou entornos ?na nube&quot;.
B1	CB6 - Posuír e comprender coñecementos que acheguen unha base ou oportunidade de ser orixinais 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 ámbitos novos ou pouco coñecidos dentro de contextos máis amplos (ou multidisciplinares) relacionados coa súa área de estudo
B3	CB8 - Que os estudantes sexan capaces de integrar coñecementos e enfrontarse á complexidade de formular xuízos a partir dunha información que, sendo incompleta ou limitada, inclúa reflexións sobre as responsabilidades sociais e éticas vinculadas á aplicación dos seus coñecementos e xuízos
B4	CB9 - Que os estudantes saiban comunicar as súas conclusións e os coñecementos e razóns últimas que as sustentan a públicos especializados e non especializados dun modo claro e sen ambigüidades
B5	CB10 - Que os estudantes posúan as habilidades de aprendizaxe que lles permitan continuar estudando dun modo que haberá de ser en gran medida autodirixido ou autónomo.



B17	CG1 - Coñecer, comprender e saber aplicar os principios, metodoloxías e novas tecnoloxías na estatística e a investigación operativa en contextos científico/académicos, tecnolóxicos ou profesionais especializados e multidisciplinares, así como adquirir as destrezas e competencias descritas nos obxectivos xerais do título.
B18	CG2 - Desenvolver autonomía para identificar, modelar e resolver problemas complexos da estatística e da investigación operativa en contextos científico/académicos, tecnolóxicos ou profesionais especializados e multidisciplinares.
B19	CG3 - Desenvolver a capacidade para realizar estudos e tarefas de investigación e transmitir os resultados a públicos especializados, académicos e xeneralistas.
B20	CG4 - Integrar coñecementos avanzados e enfrontarse á toma de decisións a partir de información científica e técnica.
B21	CG5 - Desenvolver a capacidade de aplicación de algoritmos e técnicas de resolución de problemas complexos no eido da estatística e a investigación operativa, manexando o software especializado axeitado.
C11	CT1 - Desenvolver firmes capacidades de razoamento, análise crítica e autocrítica, así como de argumentación e de síntese, contextos especializados e multidisciplinares.
C12	CT2 - Desenvolver destrezas avanzadas no manexo de Tecnoloxías da Información e a Comunicación (TIC), tanto para a obtención de información como para a difusión do coñecemento, nun ámbito científico/académico, tecnolóxico ou profesional especializado e multidisciplinar.
C13	CT3 - Ser capaz de resolver problemas complexos en novos escenarios mediante a aplicación integrada dos coñecementos.
C14	CT4 - Desenvolver unha sólida capacidade de organización e planificación do estudo, asumindo a responsabilidade do seu propio desenvolvemento profesional, para a realización de traballos en equipo e de xeito autónomo.
C15	CT5 - Desenvolver capacidades para o aprendizaxe e a integración no traballo en equipos multidisciplinares, nos ámbitos científico/académico, tecnolóxico e profesional.

## Resultados da aprendizaxe

Resultados de aprendizaxe	Competencias / Resultados do título		
Manexar de forma autónoma e solvente o software necesario para acceder a conxuntos de datos en contornas profesionais e/ou na nube.	AM16 AM17 AM23 AM24 AM25	BP5 BP17 BP19 BP20 BP21	CP12
Saber xestionar conxuntos de datos masivos nunha contorna multidisciplinar que permita a participación en proxectos profesionais complexos que requieran o uso de técnicas estatísticas.	AM18 AM21	BP1 BP2 BP3 BP4 BP18	CP11 CP13 CP14 CP15
Saber relacionar o software de deseño e xestión de bases de datos con especificamente implementado para a análise de datos.	AM16 AM17 AM21 AM24 AM25	BP17 BP18 BP21	CP12 CP13

## Contidos

Temas	Subtemas
1. Introducción a linguaxe SQL	1.1 Bases de datos relacionais 1.2 Sintaxe SQL 1.3 Conexión con bases de datos dende R
2. Introducción a tecnoloxías NoSQL	2.1 Conceptos e tipos de bases de datos NoSQL (documental, columnar, clave/valor e de grafos) 2.2 Conexión de R a NoSQL



3. Tecnoloxías para o tratamento de datos masivos	3.1 Tecnoloxías Big Data (Hadoop, Spark, Hive, Rspark, Sparklyr) 3.2 Visualización e xeración de cadros de mando 3.3 Introdución a análise de datos masivos
---	---

Planificación				
Metodoloxías / probas	Competencias / Resultados	Horas lectivas (presenciais e virtuais)	Horas traballo autónomo	Horas totais
Sesión maxistral	A16 A18 A21 A23	20	28	48
Prácticas de laboratorio	A16 A18 A24 A25 B3 B2 C13 C12	7	10.5	17.5
Solución de problemas	A17 A16 B2 C13	0	28	28
Seminario	A17 A24 A25 B1 B17 B19 B21 C15	7	10.5	17.5
Traballos tutelados	A16 A17 A21 A23 A24 A25 B1 B2 B3 B4 B5 B17 B18 B19 B20 B21 C11 C12 C13 C14 C15	1	9.5	10.5
Proba mixta	A16 A18 A21 A23 A24 A25 B1 B2 B3 B4 B17 B18 C13	3	0	3
Atención personalizada		0.5	0	0.5

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

Metodoloxías	
Metodoloxías	Descrición
Sesión maxistral	Sesións expositivas nas que presentarán conceptos e/ou procedementos, proporcionando información básica necesaria para comprender unha perspectiva teórica ou un procedemento práctico, promovendo a participación dos alumnos.
Prácticas de laboratorio	Sesións interactivas de prácticas de laboratorio (laboratorio de informática) ou resolución de problemas, onde os profesores apoiarán e supervisarán a posta en práctica dos coñecementos adquiridos polos estudantes.
Solución de problemas	Traballo persoal dos alumnos na realización de exercicios prácticos e resolución de problemas a partir das clases expositivas e das prácticas de laboratorio.
Seminario	Técnica de traballo en grupo que ten como finalidade o estudo intensivo dun tema. Caracterízase pola discusión, a participación, a elaboración de documentos e as conclusións ás que teñen que chegar todos os compoñentes do grupo.
Traballos tutelados	Actividades de aprendizaxe autónoma, nas que os profesores guiarán a realización de traballos individuais ou en pequenos grupos de alumnos.
Proba mixta	Proba que integra preguntas tipo de probas de ensaio e preguntas tipo de probas obxectivas. En canto a preguntas de ensaio, recolle preguntas abertas de desenvolvemento. Ademais, en canto a preguntas obxectivas, pode combinar preguntas de resposta múltiple, de ordenación, de resposta breve, de discriminación, de completar e/ou de asociación.

Atención personalizada	
Metodoloxías	Descrición



Prácticas de laboratorio Seminario Solución de problemas Traballos tutelados	Atención personalizada (ou en pequeno grupo / grupo moi reducido) nestas metodoloxías, tanto na aula como nas horas de titoría, para resolver as dúbidas que poidan xurdir no desenvolvemento do proceso de ensino/aprendizaxe e que non foron resoltas de forma eficaz previamente.
---	--

Avaliación			
Metodoloxías	Competencias / Resultados	Descrición	Cualificación
Prácticas de laboratorio	A16 A18 A24 A25 B3 B2 C13 C12	Evaluación das prácticas de laboratorio levadas a cabo polos estudantes.	30
Proba mixta	A16 A18 A21 A23 A24 A25 B1 B2 B3 B4 B17 B18 C13	O exame da materia avaliará os seguintes aspectos:  Conceptos da materia: Dominio dos coñecementos teóricos e operativos da materia.  Asimilación práctica de materia: Asimilación e comprensión dos coñecementos teóricos e operativos da materia.	60
Traballos tutelados	A16 A17 A21 A23 A24 A25 B1 B2 B3 B4 B5 B17 B18 B19 B20 B21 C11 C12 C13 C14 C15	Evaluación dos traballos tutelados desenvolto polos estudantes.	10

Observacións avaliación
Para poder aprobar a materia na primeira oportunidade será necesario obter como mínimo o 30% da nota máxima da suma das prácticas de laboratorio+traballos tutelados, e, igualmente, o 30% da nota máxima final da Proba Mixta (exame), e ter unha nota total (prácticas de laboratorio + traballos tutelados + proba mixta) igual ou superior ao 50% da nota máxima. Na segunda oportunidade soamente se poderá recuperar a nota da proba mixta. As notas de prácticas e de traballo tutelados serán as obtidas durante o curso. Para os alumnos que utilicen a oportunidade adiantada de decembro utilizaranse as notas de prácticas e problemas que obtiveran no seu último curso. Nesta oportunidade só será necesario para aprobar obter unha nota total igual ou superior ao 50% da nota máxima. Unha vez que un estudante é avaliado nunha práctica de laboratorio ou nun traballo tutelado implica que será calificado. Por tanto, a calificación "Non Presentado" non é posible unha vez que unha práctica/traballo ten sido avaliado.

Fontes de información	
<b>Bibliografía básica</b>	<ul style="list-style-type: none"> <li>- Daroczi, G. (2015). Mastering Data Analysis with R. Packt Publishing</li> <li>- Golemund, G. y Wickham, H. (2016). R for Data Science. <a href="https://r4ds.had.co.nz/">https://r4ds.had.co.nz/</a> &amp; O'Reilly</li> <li>- Silberschatz, A., Korth, H. y Sudarshan, S. (2014). Fundamentos de Bases de Datos. Mc Graw Hill</li> <li>- Rubén Fernández Casal (2019). Ayuda y Recursos para el Aprendizaje de R. <a href="https://rubenfcasal.github.io/post/ayuda-y-recursos-para-el-aprendizaje-de-r/">https://rubenfcasal.github.io/post/ayuda-y-recursos-para-el-aprendizaje-de-r/</a></li> </ul>
<b>Bibliografía complementaria</b>	<ul style="list-style-type: none"> <li>- Wes McKinney (2017). Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython. O'Reilly (2ª ed.)</li> <li>- Tom White (2015). Hadoop: The Definitive Guide. O'Reilly (4ª ed.)</li> <li>- Alex Holmes (2014). Hadoop in practice. Manning (2ª ed.)</li> <li>- Centro de Supercomputación de Galicia (2019). Servicio de Big Data del CESGA. <a href="https://bigdata.cesga.es/">https://bigdata.cesga.es/</a></li> </ul>

Recomendacións
<b>Materias que se recomenda ter cursado previamente</b>
<b>Materias que se recomenda cursar simultaneamente</b>



Materias que continúan o temario
Observacións
Recomendacións para o estudo da materiaDebido ao forte compoñente práctico é recomendable ir facendo as actividades prácticas e traballos academicamente dirixidos de forma regular ao longo do cuadrimestre.As ferramentas software utilizadas nesta materia son xeralmente open-source ou teñen licenza gratuita para estudantes.&nbsp;
(*)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