



| Guía Docente          |  |                    |                             |          |
|-----------------------|--|--------------------|-----------------------------|----------|
| Datos Identificativos |  |                    |                             | 2022/23  |
| Asignatura (*)        | Enxeñaría de Datos   | Código             | 614544002                   |          |
| Titulación            | Máster Universitario en Intelixencia Artificial  |                    |                             |          |
| Descritores           |  |                    |                             |          |
| Ciclo                 | Período  | Curso              | Tipo                        | Créditos |
| Mestrado Oficial      | 1º cuatrimestre  | Primeiro           | Obrigatoria                 | 3        |
| Idioma                | Inglés   |                    |                             |          |
| Modalidade docente    | Presencial   |                    |                             |          |
| Prerrequisitos        |  |                    |                             |          |
| Departamento          | Ciencias da Computación e Tecnoloxías da Información   |                    |                             |          |
| Coordinación          | Bernardo Roca, Guillermo de  | Correo electrónico | guillermo.debernardo@udc.es |          |
| Profesorado           | Bernardo Roca, Guillermo de  | Correo electrónico | guillermo.debernardo@udc.es |          |
| Web                   |  |                    |                             |          |
| Descrición xeral      | O obxectivo da materia é a introdución dos aspectos básicos da enxeñaría de datos, fundamentalmente no ámbito do Big Data. As competencias adquiridas permitirán a análise e a xestión eficiente de información heteroxénea, tanto estruturada como non estruturada, dentro do desenvolvemento de aplicacións de IA, alí onde os métodos tradicionais mostren a súa insuficiencia. |                    |                             |          |

| Competencias / Resultados do título |  |
|-------------------------------------|--|
| Código                              | Competencias / Resultados do título  |
| A17                                 | CE16 - coñecemento do proceso e as ferramentas para o procesamento e preparación de datos dende a súa adquisición ou extracción, limpeza, transformación, carga, organización e acceso   |
| B2                                  | CG02 - Abordar con éxito todas as etapas dun proxecto de Intelixencia Artificial   |
| B3                                  | CG03 - Buscar e seleccionar a información útil necesaria para resolver problemas complexos, manexando con soltura as fontes bibliográficas do campo  |
| B4                                  | CG04 - Elaborar axeitadamente e con certa orixinalidade composicións escritas ou argumentos motivados, redactar plans, proxectos de traballo, artigos científicos e formular hipóteses razoables no campo  |
| B5                                  | CG05 - Traballar en equipo, especialmente de carácter multidisciplinar, e ser hábiles na xestión do tempo, persoas e toma de decisións   |
| B6                                  | CB01 - Poseer e comprender coñecementos que aporten unha base ou oportunidade de ser orixináis no desenvolvemento e/ou aplicación de ideas, a miúdo nun contexto de investigación  |
| B7                                  | CB02 - Que os estudantes saiban aplicar os coñecementos adquiridos e posúan capacidade de resolución de problemas en entornos novos ou pouco coñecidos dentro de contextos máis amplos (ou multidisciplinares) relacionados coa su área de estudo  |
| B8                                  | CB03 - Que os estudantes sexan capaces de integrar coñecementos e enfrentarse á complexidade de formular xuízos a partiren dunha información que, sendo incompleta ou limitada, inclúa reflexións sobre as responsabilidades sociais e éticas vinculadas á aplicación dos seus coñecementos e xuízos |
| C3                                  | CT03 - 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  |
| C7                                  | CT07 - Desenvolver a capacidade de traballar en equipos interdisciplinares ou transdisciplinares, para ofrecer propostas que contribúan a un desenrolo sostible ambiental, económico, político e social  |
| C8                                  | CT08 - Valorar a importancia que ten a investigación, a innovación e o desenrolo tecnolóxico no avance socioeconómico e cultural da sociedade  |
| C9                                  | CT09 - Ter a capacidade de xestionar tempos e recursos: desenvolver plans, priorizar actividades, identificar as críticas, establecer prazos e cumprilos   |

| Resultados da aprendizaxe   |      |     |                                     |
|---|------|-----|-------------------------------------|
| Resultados de aprendizaxe   |      |     | Competencias / Resultados do título |
| Desenvolver a capacidade de analizar e modelar datos para o seu procesado en sistemas intelixentes. | AM16 | BM6 | CM3                                 |
|   |      | BM7 | CM9                                 |



|  |      |                          |                   |
|--|------|--------------------------|-------------------|
| Coñecer e comprender o proceso de extracción, limpeza, transformación, carga e procesado de datos. | AM16 | BM2<br>BM3<br>BM8        | CM3<br>CM9        |
| Coñecer e saber utilizar bases de datos multidimensionais e de tipo NoSQL.                         |      | BM3<br>BM4<br>BM7        | CM8               |
| Coñecer os fundamentos de data lakes e data warehouses.  |      | BM2<br>BM5<br>BM7<br>BM8 | CM3<br>CM7<br>CM8 |

| Contidos  |   |
|---|---|
| Temas   | Subtemas  |
| Conceptos e fundamentos de Enxeñaría de datos                     | Conceptos e definicións básicas, problemas de carga eficiente en escenarios Big Data, almacenamento de datos masivos e acceso aos mesmos. |
| Técnicas de limpeza e preparación de datos.                       | Técnicas máis comúns.<br>Definición de fluxos de procesamento.<br>Medidas de calidade.  |
| Estruturas avanzadas e almacéns de datos eficientes para Big Data | Data warehouses e BD multidimensionais, Data lakes, Bases de Datos NoSQL.   |

| Planificación          |                             |   |                         |              |
|------------------------|-----------------------------|---|-------------------------|--------------|
| Metodoloxías / probas  | Competencias / Resultados   | Horas lectivas (presenciais e virtuais) | Horas traballo autónomo | Horas totais |
| Sesión maxistral       | B4 B5 C3 C9                 | 12                                      | 0                       | 12           |
| Proba práctica         | A17 B2 B5 B7 C3             | 8                                       | 0                       | 8            |
| Solución de problemas  | A17 B2 B4 B7 C7 C9          | 0                                       | 50                      | 50           |
| Traballos tutelados    | A17 B2 B3 B6 B7 B8<br>C7 C8 | 5                                       | 0                       | 5            |
| Atención personalizada |                             | 0                                       |                         | 0            |

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

| Metodoloxías          |  |
|-----------------------|--|
| Metodoloxías          | Descrición   |
| Sesión maxistral      | O profesorado presenta un tema ao alumnado co obxectivo de facilitar un conxunto de información con alcance concreto.  |
| Proba práctica        | O profesorado da materia expón ao alumnado un problema ou problemas de carácter práctico cuxa resolución require a comprensión e aplicación dos contidos teórico-prácticos incluídos nos contidos da materia.<br>O alumnado pode traballar a solución aos problemas expostos de forma individual ou en grupos. |
| Solución de problemas | O profesorado expón ao alumnado un traballo cuxo alcance e obxectivos requiren que sexa traballado polos alumnos de forma autónoma, aínda que coa tutela do profesorado da materia.  |
| Traballos tutelados   | Exponse ao alumnado proxectos prácticos cuxo alcance require que se lle dedique un parte importante da dedicación total do alumno á materia. Ademais, polo alcance dos traballos a realizar, requírese que o alumnado aplique competencias de xestión ademais de competencias de índole técnica.               |

|                        |
|------------------------|
| Atención personalizada |
|------------------------|



| Metodoloxías                                 | Descrición   |
|--|--|
| Traballos tutelados<br>Solución de problemas | <p>Traballos tutelados (aprendizaxe baseada en proxectos):</p> <p>Exponse ao alumnado un escenario de traballo, real ou ficticio, que presenta unha determinada problemática. O alumnado debe aplicar os coñecementos teórico-prácticos da materia para buscar unha solución á cuestión ou cuestións expostas. Como norma xeral, o estudo de casos realizarase en grupos. Os distintos grupos de traballo exporán e porán en común as súas solucións.</p> <p>Resolución de problemas de forma autónoma</p> <p>O profesorado atenderá ao alumnado en sesións de tutorías individualizadas dedicadas ao desenvolvemento do proxecto.</p> |

| Avaliación            |                             |   |               |
|-----------------------|-----------------------------|---|---------------|
| Metodoloxías          | Competencias / Resultados   | Descrición  | Cualificación |
| Traballos tutelados   | A17 B2 B3 B6 B7 B8<br>C7 C8 | Levarase a cabo a defensa da solución achegada por parte do alumno ante o profesor ou unha presentación oral da solución desenvolvida.  | 30            |
| Proba práctica        | A17 B2 B5 B7 C3             | Realizaranse distintas probas de avaliación, orientadas especialmente a avaliar a comprensión dos coñecementos expostos nas clases de teoría e/ou prácticas.  | 30            |
| Solución de problemas | A17 B2 B4 B7 C7 C9          | A avaliación do traballo autónomo tutelado levará a cabo mediante a entrega dun informe e unha defensa na que o alumnado explica a súa proposta e conclusións ante o profesorado, ou mediante unha presentación oral da solución ante a aula. | 40            |

| Observacións avaliación  |
|--|
| <p><b>PRIMEIRA CONVOCATORIA e SEGUNDA CONVOCATORIA [Asistentes e Non Asistentes]</b></p> <p>Cualificación final = 0,30 * Probas periódicas + 0,40 * Resolución de problemas de forma autónoma + 0,30 * Traballos tutelados.</p> <p>Os estudantes con dispensa académica realizarán as mesmas probas e entregas que os asistentes.</p> <p><b>PROCESO DE CUALIFICACIÓN DE ACTAS</b></p> <p>Para superar a materia en calquera convocatoria, a calificación final debe ser igual ou superior a 5, debéndose obter como mínimo un 5 (sobre 10) en cada unha das partes.</p> <p><b>DATAS DE AVALIACIÓN</b></p> <p>As datas de realización das probas periódicas e entregas de traballos publicaranse na primeira semana de clases na plataforma de teledocencia.</p> <p><b>OUTRAS CONSIDERACIÓNS</b></p> <p>No caso de detectar plaxio en algún dos traballos (teóricos o práctico), a cualificación final da materia será de Suspenso (0) e o feito será comunicado a Dirección do Centro para que tome as medidas oportunas.</p> <p>Ante calquera contradición que se puidera dar entre as distintas versións da guía, debido a algún erro de tradución, a versión que prevalecerá é a versión en inglés.</p> |

| Fontes de información              |   |
|------------------------------------|---|
| <b>Bibliografía básica</b>         | <ul style="list-style-type: none"> <li>- Sadalage, Fowler (2012). NoSQL Distilled: A Brief Guide to the Emerging World of Polyglot Persistence. Addison-Wesley</li> <li>- Avi Silberschatz, Henry F. Korth, S. Sudarshan (2010). Database System Concepts. McGraw-Hill</li> <li>- Ihab F. Ilyas, Xu Chu, (2019). Data Cleaning. Association for Computing Machinery. ACM</li> <li>- Alex Gorelik (). The Enterprise Big Data Lake: Delivering the Promise of Big Data and Data Science. O'Reilly</li> </ul> |
| <b>Bibliografía complementaria</b> | - Matt Casters, Roland Bouman, Jos van Dongen (2013). Pentaho Kettle Solutions: Building Open Source ETL Solutions with Pentaho Data Integration. Wiley   |



| Recomendacións                                    |
|---|
| Materias que se recomenda ter cursado previamente |
| 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