



| Guía Docente          |  |                    |  |           |
|-----------------------|--|--------------------|--|-----------|
| Datos Identificativos |  |                    |  | 2024/25   |
| Asignatura (*)        | Aprendizaxe Automática a Gran Escala   |                    | Código   | 614G02032 |
| Titulación            | Grao en Ciencia e Enxeñaría de Datos   |                    |  |           |
| Descritores           |  |                    |  |           |
| Ciclo                 | Período  | Curso              | Tipo   | Créditos  |
| Grao                  | 1º cuatrimestre  | Cuarto             | Optativa   | 6         |
| Idioma                | Castelán   |                    |  |           |
| Modalidade docente    | Presencial   |                    |  |           |
| Prerrequisitos        |  |                    |  |           |
| Departamento          | Ciencias da Computación e Tecnoloxías da Información   |                    |  |           |
| Coordinación          | Pérez Sánchez, Beatriz   | Correo electrónico | beatriz.perezs@udc.es  |           |
| Profesorado           | Cancela Barizo, Brais<br>Morán Fernández, Laura<br>Pérez Sánchez, Beatriz  | Correo electrónico | brais.cancela@udc.es<br>laura.moranf@udc.es<br>beatriz.perezs@udc.es |           |
| Web                   | campusvirtual.udc.gal  |                    |  |           |
| Descrición xeral      | <p>Nesta materia abórdase a área da aprendizaxe automática en contornas onde existe unha gran cantidade de datos a analizar. Neste contexto xorden certas problemáticas que fan que moitos dos sistemas de aprendizaxe clásicos non sexan directamente aplicables por motivos de complexidade computacional. A aprendizaxe automática a escala trata dous ámbitos de escalabilidade diferentes. A primeira é o adestramento dun modelo con grandes conxuntos de datos, que necesita as funcionalidades da escalabilidade sobre unha agrupación industrial de computadores para realizar o adestramento. O segundo céntrase na posta en operación do modelo adestrado de maneira que se poida escalar para cumprir as necesidades das aplicacións que o consomen.</p> <p>Coa aprendizaxe automática a gran escala, o enfoque desprázase cara aos datos e a tarefa. O tempo dedicado á tarefa e os datos é significativo e, a miúdo, moito máis grande do previsto. O obxectivo desta materia é proporcionar unha visión clara sobre os conceptos do 'Big Data' no ámbito da aprendizaxe automática.</p> |                    |  |           |

| Competencias / Resultados do título |  |
|-------------------------------------|--|
| Código                              | Competencias / Resultados do título  |
| A24                                 | CE24 - Comprensión e dominio das principais técnicas básicas e avanzadas de aprendizaxe automática, incluíndo as dedicadas ao tratamento de grandes volumes de datos.  |
| A25                                 | CE25 - Capacidade para identificar a adecuación de cada unha das técnicas de aprendizaxe automática á resolución dun problema, incluíndo os aspectos relacionados coa súa complexidade computacional ou a súa capacidade explicativa, de acordo aos requisitos establecidos. |
| A26                                 | CE26 - Coñecemento das ferramentas informáticas actuais no campo da aprendizaxe automática, e capacidade para seleccionar a máis adecuada para a resolución dun problema.  |
| B2                                  | CB2 - Que os estudantes saiban aplicar os seus coñecementos ao seu traballo ou vocación dunha forma profesional e posúan as competencias que adoitan demostrarse por medio da elaboración e defensa de argumentos e a resolución de problemas dentro da súa área de estudo   |
| B3                                  | CB3 - Que os estudantes teñan a capacidade de reunir e interpretar datos relevantes (normalmente dentro da súa área de estudo) para emitir xuízos que inclúan unha reflexión sobre temas relevantes de índole social, científica ou ética                                    |
| B4                                  | CB4 - Que os estudantes poidan transmitir información, ideas, problemas e solucións a un público tanto especializado como non especializado  |
| B7                                  | CG2 - Elaborar adecuadamente e con certa orixinalidade composicións escritas ou argumentos motivados, redactar plans, proxectos de traballo, artigos científicos e formular hipóteses razoables.   |
| B8                                  | CG3 - Ser capaz de manter e estender formulacións teóricas fundadas para permitir a introdución e explotación de tecnoloxías novas e avanzadas no campo.   |
| B9                                  | CG4 - Capacidade para abordar con éxito todas as etapas dun proxecto de datos: exploración previa dos datos, preprocesado, análise, visualización e comunicación de resultados.  |



|     |   |
|-----|---|
| B10 | CG5 - Ser capaz de traballar en equipo, especialmente de carácter multidisciplinar, e ser hábiles na xestión do tempo, persoas e toma de decisións.                                   |
| C1  | CT1 - 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. |
| C4  | CT4 - Valorar a importancia que ten a investigación, a innovación e o desenvolvemento tecnolóxico no avance socioeconómico e cultural da sociedade.                                   |

| Resultados da aprendizaxe   |                                     |   |          |
|---|-------------------------------------|---|----------|
| Resultados de aprendizaxe   | Competencias / Resultados do título |   |          |
| Identificar e saber afrontar os problemas máis frecuentes relacionados coa explosión de datos, coñecidos como as "Vs do Big Data" e como algúns deles afectan á aplicación de técnicas de Aprendizaxe Automática. | A25                                 | B2<br>B9<br>B10                         | C1<br>C4 |
| Coñecer os métodos e técnicas máis representativos e actuais de preprocesado de datos para tratar grandes volumes de datos.   | A24<br>A25<br>A26                   | B9                                      | C1       |
| Coñecer os métodos e técnicas máis representativas e actuais de Aprendizaxe Automática en contornas afectadas por problemas como o volume, a velocidade ou a privacidade dos datos.                               | A24<br>A25<br>A26                   | B3<br>B10                               | C1<br>C4 |
| Saber manexar as ferramentas e contornas de traballo máis actuais no ámbito da aprendizaxe automática para tratar grandes volumes de datos.   | A24                                 | B2<br>B4<br>B7                          | C1       |
| Coñecer técnicas para a representación visual de datos complexos e saber utilizar ferramentas de visualización de datos para poder comunicar eficazmente os resultados das análises realizadas.                   | A24<br>A26                          | B2<br>B3<br>B4<br>B7<br>B8<br>B9<br>B10 | C1       |
| Coñecer técnicas analíticas e escalables baseadas en grafos.  | A24<br>A26                          | B2<br>B8<br>B9<br>B10                   | C1       |

| Contidos   |   |
|--|---|
| Temas  | Subtemas  |
| Problemáticas da análise de datos en contornas "Big Data"; |   |
| Tratamento e visualización de grandes volumes de datos     | Técnicas de visualización<br>Técnicas de preprocesado de datos                          |
| Aprendizaxe distribuída                                    |   |
| Aprendizaxe federada                                       | Aprendizaxe na contorna<br>Preservación da privacidade                                  |
| Tratamento de datos en continuo                            | Aprendizaxe incremental<br>Aprendizaxe en tempo real<br>Problemas de cambio de concepto |
| Analítica de datos baseada en grafos                       |   |

| Planificación |
|---------------|
|---------------|



| Metodoloxías / probas    | Competencias / Resultados         | Horas lectivas (presenciais e virtuais) | Horas traballo autónomo | Horas totais |
|--------------------------|-----------------------------------|---|-------------------------|--------------|
| Solución de problemas    | A24 A25 A26 B2 B4<br>B7 B10 C1 C4 | 0                                       | 40                      | 40           |
| Proba obxectiva          | A24 A25 A26 B2 B3<br>B8 B9        | 3                                       | 0                       | 3            |
| Prácticas de laboratorio | A24 A25 A26 B2 B3<br>B7 B9 B10 C1 | 21                                      | 21                      | 42           |
| Sesión maxistral         | A25 A26 B2 B3 B8 B9               | 21                                      | 42                      | 63           |
| Atención personalizada   |                                   | 2                                       | 0                       | 2            |

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

| Metodoloxías             |  |
|--------------------------|--|
| Metodoloxías             | Descrición   |
| Solución de problemas    | Desenvolveranse exemplos e exercicios nos que o alumnado terá que aplicar os coñecementos teóricos da materia a casos concretos. Garantirase a interactividade, resolvendo as dúbidas plantexadas polo alumnado, animando a contrastar as súas solucións e expor cuestións relevantes. Parte dos problemas realizados serán avaliados. |
| Proba obxectiva          | Levarase a cabo unha avaliación da materia mediante unha proba que incluírá tanto preguntas sobre os contidos teóricos, como supostos prácticos e exercicios de aplicación relacionados cos distintos temas vistos na materia.   |
| Prácticas de laboratorio | Desenvolveranse exemplos e exercicios nos que o alumnado terá que aplicar os coñecementos teóricos da materia a casos concretos. Garantirase a interactividade, resolvendo as dúbidas plantexadas polo alumnado, animando a contrastar as súas solucións e expor cuestións relevantes. Parte dos problemas realizados serán avaliados. |
| Sesión maxistral         | Clases maxistras onde se expoñerán os conceptos teóricos da materia, sen perder nunca de vista exemplos de aplicación para motivar e contextualizar os contidos da materia. Fomentarase a interactividade en clase mediante a formulación de preguntas e utilizaranse distintos recursos como transparencias ou demostracións.         |

| Atención personalizada                            |  |
|---|--|
| Metodoloxías                                      | Descrición   |
| Solución de problemas<br>Prácticas de laboratorio | O desenvolvemento, tanto das clases maxistras como das de resolución de problemas e os laboratorios de prácticas, realizarase atendendo ao progreso do alumnado nas capacidades de comprensión e asimilación dos contidos impartidos. O avance xeral da clase compaxinarase cunha atención específica a quen presente maiores dificultades na tarefa da aprendizaxe e cun apoio adicional a quen amose maior desenvolvemento e desexe ampliar coñecementos.<br><br>A atención personalizada se realiza a través das seguintes canles:<br>- Correo-e: De uso para facer consultas de resposta curta ou solicitar encontros para resolver dúbidas.<br>- Teams: De uso nos horarios oficiais de titorías nos que se garante unha resposta inmediata. Non entanto, dado o elevado número de alumnado, recoméndase solicitude previa de titoría vía correo-e.<br>- Campus virtual: Poñeránse a disposición do alumnado "foros temáticos" que resolvan dúbidas xerais ligadas a actividades específicas como as prácticas. |

| Avaliación   |                           |            |               |
|--------------|---------------------------|------------|---------------|
| Metodoloxías | Competencias / Resultados | Descrición | Cualificación |
|              |                           |            |               |



|                          |                                   |   |    |
|--------------------------|-----------------------------------|---|----|
| Solución de problemas    | A24 A25 A26 B2 B4<br>B7 B10 C1 C4 | Valoraranse os resultados, forma e condicións de realización de diversos traballos puntuables que se detallarán durante o curso. Algunhas das sesións estarán dedicadas á revisión dos traballos entregados e se realizarán preguntas sobre os mesmos. A asistencia a estas sesións é de carácter obrigatoria, no caso contrario o efecto será o mesmo que a non entrega da práctica. | 40 |
| Proba obxectiva          | A24 A25 A26 B2 B3<br>B8 B9        | Realización obrigatoria. Avaliarase o dominio dos coñecementos teóricos e operativos da materia.  | 60 |
| Prácticas de laboratorio | A24 A25 A26 B2 B3<br>B7 B9 B10 C1 | Son obrigatorias para poder aprobar os traballos da solución de problemas e inflúen na cualificación final destes, pero non se puntúan á marxe da nota outorgada á solución de problemas.   | 0  |

### Observacións avaliación

Para poder aprobar a materia o estudante deberá cumprir os seguintes requisitos (puntuación entre 0 e 10 en todas as actividades):

- Lograr unha nota superior ou igual a 4 na proba mixta final realizada ao final do cuadrimestre.
- Lograr unha nota superior ou igual a 5 ao realizar a suma de todas as probas de avaliación.

Avaliación no caso da convocatoria adiantada

- No caso de que o alumnado solicite e asista á convocatoria adiantada, o 50% da súa nota será a proba mixta (exame final) e o outro 50% o traballo tutelado. O traballo tutelado deberá entregarse como data límite o mesmo día do exame oficial da convocatoria adiantada. Para aprobar a materia, o alumnado debe cumprir os requisitos mencionados anteriormente.

Sobre a responsabilidade

compartida dos traballos en grupo

- Nas

actividades que se levan a cabo en grupos, todas as persoas que conforman o grupo de traballo serán responsables solidarias da súa realización e entrega, así como das consecuencias que se deriven do incumprimento das

normas de autoría do mesmo.

- O traballo entregado deberá ser orixinal do grupo. De acordo ao Artigo 11, Apartado 4 b), do Regulamento disciplinar do estudantado da UDC, a entrega de traballos non orixinais ou con partes duplicadas (sexa por copias entre compañeiros/as, por obtención doutras fontes, etc.) conllevará a calificación de SUSPENSO con 0 puntos nesa convocatoria completa tanto para o grupo que presente material copiado como para o grupo que o facilitase.

Matrícula a tempo parcial

-

O alumnado matriculado a tempo parcial terá que entregar as actividades avaliadas nas condicións e prazos específicos que se establecerán. Será obriga do alumnado comunicar a súa situación ao profesorado.

Non presentado

- Únicamente terá a condición de ?Non presentado? quen non entregue actividade algunha nin conorra á proba obxectiva.

NOTA: Todos os aspectos relacionados con ?dispensa académica?, ?dedicación ao estudo?, ?permanencia? e ?fraude académica? rexeranse de acordo coa normativa académica vixente da UDC.

### Fontes de información

|                                    |   |
|------------------------------------|---|
| <b>Bibliografía básica</b>         | - Bill Chambers and Matei Zaharia (2018). Spark : the definitive guide :big data processing made simple . Sebastopol, CA : O'Reilly Media, Inc.<br>- Jules S. Damji, Brooke Wenig, Tathagata das, and Denny Lee (2020). Learning Spark : Lightning-fast big data analytics. Sebastopol, CA : O'Reilly |
| <b>Bibliografía complementaria</b> |   |

### Recomendacións

Materias que se recomenda ter cursado previamente



Procesamento Paralelo/614G02023

Aprendizaxe Automática III/614G02026

Aprendizaxe Automática I/614G02019

Aprendizaxe Automática II/614G02021

### Materias que se recomenda cursar simultaneamente

### Materias que continúan o temario

### Observacións

\* No caso de discrepancia da información proporcionada polas guías docentes en galego e castelán, prevalece o documento correspondente ao idioma institucional (galego)

#### RECOMENDACIÓNS

#### EN IGUALDADE DE XÉNERO E DIVERSIDADE - Segundo

se recolle nas distintas normativas de aplicación para a docencia universitaria

deberase incorporar a perspectiva de xénero nesta materia, polo que os traballos

entregados polo estudiantado e o material preparado polo profesorado deben usar

linguaxe non sexista. - Fomentaranse valores de respecto e igualdade,

evitando deste xeito prexuízos e actitudes sexistas e discriminatorias.- Facilitarase a

integración do alumnado que por razón físicas, sensoriais, psíquicas ou

socioculturais, experimenten dificultades a un acceso axeitado, igualitario e

proveitoso á vida universitaria.

#### RECOMENDACIÓNS EN

SOSTIBILIDADE E MEDIO AMBIENTE- Débese facer un uso sostible dos

recursos e a prevención de impactos negativos sobre o medio natural.- Débese ter en conta a importancia

dos principios éticos relacionados cos valores da sostibilidade nos

comportamentos persoais e profesionais.

(\*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