



Guía Docente				
Datos Identificativos				2021/22
Asignatura (*)	Minaría de Datos	Código	710G04030	
Titulación				
Descritores				
Ciclo	Período	Curso	Tipo	Créditos
Grao	2º cuatrimestre	Terceiro	Optativa	6
Idioma	CastelánGalego			
Modalidade docente	Híbrida			
Prerrequisitos				
Departamento	Matemáticas			
Coordinación	Maseda Seco, Diego	Correo electrónico	d.maseda@udc.es	
Profesorado	Maseda Seco, Diego	Correo electrónico	d.maseda@udc.es	
Web				
Descrición xeral	O obxectivo fundamental desta materia é que o alumno coñeza os conceptos fundamentais e os principais modelos da minería de datos, e a súa aplicación no campo das ciencias da información e documentación.			
Plan de continxencia	<p>1. Modificacións nos contidos: Non se modifican os contidos.</p> <p>2. Metodoloxías: Metodoloxías docentes que se manteñen: ? Prácticas a través de TIC. ? Sesión maxistral. ? Proba obxectiva. ? Traballos tutelados. ? Estudo de casos.</p> <p>Metodoloxías docentes que se modifican:</p> <p>3. Mecanismos de atención personalizada ao alumnado: Ferramentas: Moodle, Microsoft Teams e correo electrónico. * Temporalización: Teams utilizarase como soporte para levar a cabo as clases teóricas e prácticas, ademais de servir de medio para realizar as titorías. Moodle empregarase para a publicación de contidos e avisos, e para a avaliación dos alumnos (avaliación continua e exame). O correo electrónico servirá de medio para resolver dúbidas e para o intercambio de ficheiros e información en xeral.</p> <p>4. Modificacións na avaliación: Manteranse todas as metodoloxías e a súa ponderación na nota global. O exame realizarase de maneira non presencial na modalidade síncrona. Observacións de avaliación: Non se modificará o peso que na avaliación final ten cada unha das metodoloxías.</p> <p>5. Modificacións da bibliografía ou webgrafía:</p>			

Competencias / Resultados do título	
Código	Competencias / Resultados do título

Resultados da aprendizaxe	
Resultados de aprendizaxe	Competencias / Resultados do título



Adquisición de habilidades para a selección, tratamento, xestión e análise da información mediante técnicas de minería de datos.	A8 A13 A20 A21 A22	B2 B4 B5 B6 B7 B8 B10 B11	C2 C4 C5 C6 C8
Elixir as técnicas cuantitativas adecuadas aos obxectivos para tarefas de investigación, administración e xestión.	A1 A8 A13 A20 A21 A22	B2 B4 B5 B6 B7 B8 B10 B11	C2 C4 C6 C8
Adquirir coñecementos de análises de datos computacional, incluíndo programas como o software estatístico R.	A1 A4 A8 A13 A20 A21 A22	B1 B2 B3 B9 B11	C1 C3 C7
Coñecemento e habilidades para a aplicación das principais técnicas de clasificación.	A1 A20 A21 A22	B1 B2 B3 B4 B5 B6 B7 B8 B9 B10 B11	C2 C4 C6 C7 C8
Coñecemento e habilidades para a aplicación de técnicas de regresión, detección de anomalías e series de tempo.	A1 A20 A21 A22	B1 B2 B3 B4 B5 B6 B7 B8 B9 B10 B11	C2 C4 C6 C7 C8



Capacidade de análise e de síntese aplicada á xestión e organización da información.	A1 A4 A8 A13 A20 A21 A22	B1 B2 B3 B4 B5 B6 B7 B8 B9 B10 B11	C1 C2 C3 C4 C5 C6 C7 C8
Adquisición de habilidades para a toma de decisións a partir da análise estatística bases de datos complexas.	A1 A4 A8 A13 A20 A21 A22	B1 B2 B3 B4 B5 B6 B7 B8 B9 B10 B11	C1 C2 C3 C4 C5 C6 C7 C8

Contidos	
Temas	Subtemas
Introducción á minería de datos.	Conceptos preliminares. Tipos de problemas en minería de datos: descrición, clasificación, predición, clustering, detección de anomalías, etc. Tipos de aprendizaxe: supervisado e non supervisado.
Métodos de clasificación non supervisada ou clúster	Conceptos básicos. Métodos de clasificación xerárquica. Métodos de agrupamento por particións. Casos prácticos.
Métodos de clasificación supervisada.	Conceptos básicos. Modelos principais de clasificación supervisada ou recoñecemento de patróns. Validación de modelos de clasificación (como de ben predín?). Casos prácticos.
Métodos avanzados de regresión.	Introdución. Modelos de regresión univariantes e multivariantes. Selección de variables relevantes. Validación de modelos de regresión (como de ben se axusta aos datos?, como de ben fai predicións?). Casos prácticos.
Series de tempo	Conceptos básicos. Análise descritivo de series de tempo. Uso práctico dos modelos de series de tempo. Casos prácticos.
Técnicas estatísticas para minería de textos e recuperación da información.	Conceptos básicos. Casos prácticos de aplicación da minería de textos.



Planificación

Metodoloxías / probas	Competencias / Resultados	Horas lectivas (presenciais e virtuais)	Horas traballo autónomo	Horas totais
Sesión maxistral	A1 A4 A8 A20 A21 B1 B3 B7 C4 C7	19	0	19
Prácticas a través de TIC	A13 A22 B11 C2	13	0	13
Estudo de casos	A1 A8 A21 B2 B3 B4 B5 B6 B7 B8 B9 C1 C8	7	7	14
Traballos tutelados	B2 B4 B5 B6 B8 B9 B10 C1 C3 C5 C6 C8	1	101	102
Proba obxectiva	A21 B1 B2	1	0	1
Atención personalizada		1	0	1

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

Metodoloxías

Metodoloxías	Descrición
Sesión maxistral	Serán sesións expositivas nas que se introducirán e describirán os diversos temas da materia, mediante presentacións (usando os adecuados medios audiovisuais) que incluírán teoría e exemplos.
Prácticas a través de TIC	Desenvolveranse clases prácticas mediante software estatístico, nas que se introducirá a súa programación e aplicación a partir de casos reais e simulados.
Estudo de casos	Aplicaranse as técnicas estatísticas impartidas na materia para a resolución de exercicios e casos de estudo reais e simulados no ámbito da xestión dixital da información.
Traballos tutelados	Realizaranse traballos individuais e/ou en grupo, tutelados polos docentes da materia, nos que se abordará a resolución, mediante a aplicación de técnicas estatísticas e o software R, de exercicios prácticos ou de casos de estudo particulares relacionados co ámbito da comunicación e das ciencias da información. Tamén se poderá realizar un estudo de revisión acerca dun tema concreto da materia ou do software utilizado. Os traballos poderán ser propostos polos docentes ou polos propios alumnos (as propostas teranse en conta ou non sempre segundo o criterio do docente).
Proba obxectiva	Consistirá nunha proba tipo test sobre os contidos impartidos na materia.

Atención personalizada

Metodoloxías	Descrición
Prácticas a través de TIC	Nas clases maxistras se fomentará en todo momento o debate entre os alumnos e entre os alumnos e o profesor. Para a resolución de problemas será importante atender personalmente aos alumnos ante as posibles dúbidas que poidan xurdir.
Estudo de casos	Esta atención servirá tamén, por unha parte, ao profesor para detectar posibles problemas na metodoloxía utilizada para impartir a asignatura e, por outra, aos alumnos para consolidar coñecementos teóricos e para expresar as súas inquedanzas acerca da asignatura. Será tamén fundamental a atención personalizada ao estudante durante as clases de prácticas TIC, sobre todo ata que non se familiarice co software estatístico a utilizar, ademais de na resolución de casos de estudo.
Sesión maxistral	
Traballos tutelados	

Avaliación

Metodoloxías	Competencias / Resultados	Descrición	Cualificación
Prácticas a través de TIC	A13 A22 B11 C2	Valorarase a asistencia e/ou o desempeño dos alumnos co software estatístico.	20
Proba obxectiva	A21 B1 B2	Proba tipo test consistente nun número de preguntas entre 10 e 20 con 3 respostas posibles.	40



Traballos tutelados	B2 B4 B5 B6 B8 B9 B10 C1 C3 C5 C6 C8	Realizaranse traballos individuais e/ou en grupo, tutelados polos docentes da materia, nos que se abordará a resolución, mediante a aplicación de técnicas estatísticas e o software R, de exercicios prácticos ou de casos de estudo particulares relacionados co ámbito da comunicación e das ciencias da información. Tamén se poderá realizar un estudo de revisión acerca dun tema concreto da materia ou do software utilizado. Os traballos poderán ser propostos polos docentes ou polos propios alumnos (as propostas serán tidas en conta ou non sempre segundo o criterio do docente).	40
---------------------	---	---	----

Observacións avaliación

Primeira oportunidade

Realizarase unha proba de resposta múltiple de 10 a 20 preguntas que representa o 40% da nota. Por outra banda, a avaliación continua constará da asistencia e/ou entrega de prácticas relacionadas coa aprendizaxe e aplicación do software estatístico R para a resolución de problemas no campo da xestión dixital da información (20% da nota global), ademais da entrega dun ou varios traballos de aplicación da estatística para a resolución de casos de estudo en documentación dixital (alternativamente poderán ser traballos de revisión ou ampliación da materia) que representa o 40% da nota total.

Segunda oportunidade

Na avaliación da segunda oportunidade se seguirá o mesmo criterio que na primeira.

Convocatoria adiantada

Todas as observacións previas son aplicables aos estudantes que soliciten a convocatoria adiantada do exame.

Calificación de non presentado

En calquera das dúas oportunidades anuais figurará un NON PRESENTADO naqueles casos nos que o alumnado non acuda ó exame oficial da materia.

Estudiante con recoñecemento de dedicación a tempo parcial e dispensa académica de exención de asistencia

No caso do alumnado con recoñecemento de dedicación a tempo parcial e dispensa académica de exención de asistencia que decida non asistir a clases, este será avaliado nas dúas oportunidades como o resto do alumnado que se atopa nunha situación similar.

A realización fraudulenta das probas ou actividades de avaliación implicará directamente a calificación de suspenso "0" na materia na convocatoria correspondente, invalidando así calquera calificación obtida en todas as actividades de avaliación de cara á convocatoria extraordinaria.

Fontes de información

Bibliografía básica

- Williams, G. (2011). Data mining with Rattle and R: The art of excavating data for knowledge discovery. . Springer Science & Business Media.
- Cirillo, A. (2017). R Data Mining: Implement Data Mining Techniques Through Practical Use Cases and Real-world Datasets. . Packt Publishing.
- Silge, J. y Robinson, D. (2017). Text Mining with R: A Tidy Approach. O'Reilly
- Jockers, M.L. (2014). Text Analysis with R for Students of Literature. Springer

Bibliografía complementaria

Recomendacións

Materias que se recomenda ter cursado previamente

Ciencia de Datos/710G04026

Fundamentos Estatísticos/710G04040

Materias que se recomenda cursar simultaneamente

Materias que continúan o temario

Observacións



Para axudar a conseguir unha contorna sostible e cumprir co obxectivo da acción número 5: ?Docencia e investigación saudable e sustentable ambiental e social? do "Plan de Acción Green Campus Ferrol": 1.- A entrega dos traballos documentais que se realicen nesta materia: 1.1. Solicitarase en formato virtual e/ou soporte informático. 1.2. Realizarase a través de Moodle, en formato dixital sen necesidade de imprimilos 1.3. De se realizar en papel: -Non se empregarán plásticos. - Realizaranse impresións a dobre cara. - Empregarase papel reciclado. - Evitarase a impresión de borradores. 2.- Débese facer un uso sostible dos recursos e a prevención de impactos negativos sobre o medio natural. 3.- Débese ter en conta a importancia dos principios éticos relacionados cos valores da sostenibilidade nos comportamentos persoais e profesionais. 4.- Segundo se recolle nas distintas normativas de aplicación para a docencia universitaria deberase incorporar a perspectiva de xénero nesta materia (usarase linguaxe non sexista, utilizarase bibliografía de autores de ambos os sexos, propiciarse a intervención en clase de alumnos e alumnas?).5.- Traballarase para identificar e modificar prexuizos e actitudes sexistas, e influirase na contorna para modificalos e fomentar valores de respecto e igualdade. 6. Deberanse detectar situacións de discriminación por razón de xénero e proponerse accións e medidas para corrixilas. 7. Facilitarase a plena integración do alumnado que por razón físicas, sensoriais, psíquicas ou socioculturais, experimenten dificultades a un acceso axeitado, igualitario e proveitoso á vida universitaria.

(*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