UNIVERSIDADE DA CORUÑA

Guia docente					
Datos Identificativos					2021/22
Asignatura (*)	Minería de Datos			Código	710G04030
Titulación	Grao en Xestión Dixital de Información e Documentación				
Descriptores					
Ciclo	Periodo	Curso		Tipo	Créditos
Grado	2º cuatrimestre	Tercero		Optativa	6
Idioma	CastellanoGallego				
Modalidad docente	Híbrida				
Prerrequisitos					
Departamento	Matemáticas				
Coordinador/a	Maseda Seco, Diego		Correo electrónico	d.maseda@udc.es	
Profesorado	Maseda Seco, Diego		Correo electrónico	d.maseda@udc.es	
Web					
Descripción general	El objetivo fundamental de esta asignatura es que el alumno conozca los conceptos fundamentales y los principales modelos de la minería de datos, y su aplicación en el campo de las ciencias de la información y documentación.				
Plan de contingencia	1. Modificaciones en los contenidos: No se modifican los contenidos. 2. Metodologías: Metodologías docentes que se mantienen: ? Prácticas a través de TIC. ? Sesión magistral. ? Prueba objetiva. ? Trabajos tutelados. ? Estudio de casos. Metodologías docentes que se modifican: 3. Mecanismos de atención personalizada al alumnado: Herramientas: Moodle, Microsoft Teams y correo electrónico. * Temporalización: Teams se utilizará como soporte para llevar a cabo las clases teóricas y prácticas, además de servir de medio para realizar las tutorías. Moodle se empleará para la publicación de contenidos y avisos, y para la evaluación de los alumnos (evaluación continua y examen). El correo electrónico servirá de medio para resolver dudas y para el intercambio de ficheros e información en general. 4. Modificaciones en la evaluación: Se mantendrán todas las metodologías y su ponderación en la nota global. El examen se realizará de manera no presencial en la modalidad síncrona. Observaciones de evaluación: No se modificará el peso que en la evaluación final tiene cada una de las metodologías. 5. Modificaciones de la bibliografía o webgrafía:				

Competencias / Resultados del título	
Código	Competencias / Resultados del título
A1	CE1 - Conocer y comprender los principios teóricos y metodológicos de la gestión de información y la documentación para aplicarlos en su actividad profesional
A4	CE4 - Dominar los fundamentos del comportamiento de los individuos en la búsqueda, recuperación y uso de la información teniendo como punto de referencia los aspectos de motivación, entorno y contexto
A8	CE8 - Dominar los diferentes métodos de representación de los datos, información y el conocimiento que garanticen su recuperación eficiente

A13	CE13 - Conocer y dominar las técnicas y normativas para la creación y autenticación, reunión, selección, organización, representación, preservación, recuperación, acceso, difusión e intercambio, y evaluación de los recursos y servicios de información
A20	CE20 - Dominar las bases para desarrollar actividades de investigación utilizando métodos y principios multidisciplinares
A21	CE21 - Poseer conocimientos de estadística y análisis cuantitativo de la información
A22	CE22 - Adquirir habilidades computacionales y de manejo de las nuevas TIC
B1	CB1 - Poseer y comprender conocimientos que aporten una base u oportunidad de ser originales en el desarrollo y/o aplicación de ideas, a menudo en un contexto de investigación
B2	CB2 - Aplicar los conocimientos adquiridos y su capacidad de resolución de problemas en entornos nuevos o poco conocidos dentro de contextos más amplios (o multidisciplinares) relacionados con su área de estudio
B3	CB3 - Ser capaz de integrar conocimientos y enfrentarse a la complejidad de formular juicios a partir de una información que, siendo incompleta o limitada, incluya reflexiones sobre las responsabilidades sociales y éticas vinculadas a la aplicación de sus conocimientos y juicios
B4	CB4 - Saber comunicar sus conclusiones -y los conocimientos y razones últimas que las sustentan- a públicos especializados y no especializados de un modo claro y sin ambigüedades
B5	CB5 - Poseer las habilidades de aprendizaje que les permitan continuar estudiando de un modo que habrá de ser en gran medida autodirigido o autónomo
B6	CG1 - Capacidad para la cooperación, el trabajo en equipo y el aprendizaje colaborativo
B7	CG2 - Capacidad de reflexión y razonamiento crítico
B8	CG3 - Capacidad de planificación, organización y gestión de recursos, información y operaciones
B9	CG4 - Capacidad de análisis, diagnóstico y toma de decisiones
B10	CG5 - Capacidad de trabajar en un contexto internacional y global
B11	CG6 - Capacidad de comprender la importancia, el valor y la función de la Gestión Digital de Información y Documentación en la actual sociedad de las TIC
C1	CT1 - Expresarse correctamente, tanto de forma oral como escrita, en las lenguas oficiales de la comunidad autónoma
C2	CT2 - Utilizar las herramientas básicas de las tecnologías de la información y las comunicaciones (TIC) necesarias para el ejercicio de su profesión y para el aprendizaje a lo largo de su vida
C3	CT3 - Desarrollarse para el ejercicio de una ciudadanía respetuosa con la cultura democrática, los derechos humanos y la perspectiva de género
C4	CT4 - Entender la importancia de la cultura emprendedora y conocer los medios al alcance de las personas emprendedoras
C5	CT5 - Adquirir habilidades para la vida y hábitos, rutinas y estilos de vida saludables
C6	CT6 - Desarrollar la capacidad de trabajar en equipos interdisciplinares o transdisciplinares, para ofrecer propuestas que contribuyan a un desarrollo sostenible ambiental, económico, político y social
C7	CT7 - Valorar la importancia que tiene la investigación, la innovación y el desarrollo tecnológico en el avance socioeconómico y cultural de la sociedad
C8	CT8 - Tener la capacidad de gestionar tiempos y recursos: desarrollar planes, priorizar actividades, identificar las críticas, establecer plazos y cumplirlos

Resultados de aprendizaje			
Resultados de aprendizaje	Competencias / Resultados del título		
Adquisición de habilidades para la selección, tratamiento, gestión y análisis de la 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

Elegir las técnicas cuantitativas adecuadas a los objetivos para tareas de investigación, administración y gestión.	A1 A8 A13 A20 A21 A22	B2 B4 B5 B6 B7 B8 B10 B11	C2 C4 C6 C8
Adquirir conocimientos de análisis de datos computacional, incluyendo programas como el software estadístico R.	A1 A4 A8 A13 A20 A21 A22	B1 B2 B3 B9 B11	C1 C3 C7
Conocimiento y habilidades para la aplicación de las principales técnicas de clasificación.	A1 A20 A21 A22	B1 B2 B3 B4 B5 B6 B7 B8 B9 B10 B11	C2 C4 C6 C7 C8
Conocimiento y habilidades para la aplicación de técnicas de regresión, detección de anomalías y series de tiempo.	A1 A20 A21 A22	B1 B2 B3 B4 B5 B6 B7 B8 B9 B10 B11	C2 C4 C6 C7 C8
Capacidad de análisis y de síntesis aplicada a la gestión y organización de la 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 la toma de decisiones a partir del análisis estadístico bases de datos complejas.	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

Contenidos	
Tema	**Subtema**
Introducción a la minería de datos.	Conceptos preliminares. Tipos de problemas en minería de datos: descripción, clasificación, predicción, clustering, detección de anomalías, etc. Tipos de aprendizaje: supervisado y no supervisado.
Métodos de clasificación no supervisada o clúster	Conceptos básicos. Métodos de clasificación jerárquica. Métodos de agrupamiento por particiones. Casos prácticos.
Métodos de clasificación supervisada.	Conceptos básicos. Modelos principales de clasificación supervisada o reconocimiento de patrones. Validación de modelos de clasificación (¿cómo de bien predicen?). Casos prácticos.
Métodos avanzados de regresión.	Introducción. Modelos de regresión univariantes y multivariantes. Selección de variables relevantes. Validación de modelos de regresión (¿cómo de bien se ajusta a los datos?, ¿cómo de bien hace predicciones?). Casos prácticos.
Series de tiempo	Conceptos básicos. Análisis descriptivo de series de tiempo. Uso práctico de los modelos de series de tiempo. Casos prácticos.
Técnicas estadísticas para minería de textos y recuperación de la información.	Conceptos básicos. Casos prácticos de aplicación de la minería de textos.

Planificación				
Metodologías / pruebas	**Competencias / Resultados**	**Horas lectivas (presenciales y virtuales)**	**Horas trabajo autónomo**	**Horas totales**
Sesión magistral	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
Estudio de casos	A1 A8 A21 B2 B3 B4 B5 B6 B7 B8 B9 C1 C8	7	7	14

Trabajos tutelados	B2 B4 B5 B6 B8 B9 B10 C1 C3 C5 C6 C8	1	101	102
Prueba objetiva	A21 B1 B2	1	0	1
Atención personalizada		1	0	1
(*)Los datos que aparecen en la tabla de planificación són de carácter orientativo, considerando la heterogeneidad de los alumnos				

Metodologías	
Metodologías	**Descripción**
Sesión magistral	Serán sesiones expositivas en las que se introducirán y describirán los diversos temas de la materia, mediante presentaciones (usando los adecuados medios audiovisuales) que incluirán teoría y ejemplos.
Prácticas a través de TIC	Se desarrollarán clases prácticas mediante software estadístico, en las que se introducirá su programación y aplicación a partir de casos reales y simulados.
Estudio de casos	Se aplicarán las técnicas estadísticas impartidas en la materia para la resolución de ejercicios y casos de estudio reales y simulados en el ámbito de la gestión digital de la información.
Trabajos tutelados	Se realizarán trabajos individuales y/o en grupo, tutelados por los docentes de la materia, en los que se abordará la resolución, mediante la aplicación de técnicas estadísticas y el software R, de ejercicios prácticos o de casos de estudio particulares relacionados con el ámbito de la comunicación y de las ciencias de la información. También se podrá realizar un estudio de revisión acerca de un tema concreto de la materia o del software utilizado. Los trabajos podrán ser propuestos por los docentes o por los propios alumnos (las propuestas se tendrán en cuenta o no siempre según el criterio del docente).
Prueba objetiva	Consistirá en una prueba tipo test sobre los contenidos impartidos en la materia.

Atención personalizada	
Metodologías	**Descripción**
Prácticas a través de TIC Estudio de casos Sesión magistral Trabajos tutelados	En las clases magistrales se fomentará en todo momento el debate entre los alumnos y entre los alumnos y el profesor. Para la resolución de problemas será importante atender personalmente a los alumnos ante las posibles dudas que puedan surgir. Esta atención servirá también, por una parte, al profesor para detectar posibles problemas en la metodología utilizada para impartir la asignatura y, por otra, a los alumnos para consolidar conocimientos teóricos y para expresar sus inquietudes acerca de la asignatura. Será también fundamental a atención personalizada al estudiante durante las clases de prácticas TIC, sobre todo hasta que no se familiarice con el software estadístico a utilizar, además de en la resolución de casos de estudio.

Evaluación			
Metodologías	**Competencias / Resultados**	**Descripción**	**Calificación**
Prácticas a través de TIC	A13 A22 B11 C2	Se valorará la asistencia y/o el desempeño de los alumnos con el software estadístico.	20
Prueba objetiva	A21 B1 B2	Prueba tipo test consistente en un número de preguntas entre 10 y 20 con 3 respuestas posibles.	40
Trabajos tutelados	B2 B4 B5 B6 B8 B9 B10 C1 C3 C5 C6 C8	Se realizarán trabajos individuales y/o en grupo, tutelados por los docentes de la materia, en los que se abordará la resolución, mediante la aplicación de técnicas estadísticas y el software R, de ejercicios prácticos o de casos de estudio particulares relacionados con el ámbito de la comunicación y de las ciencias de la información. También se podrá realizar un estudio de revisión acerca de un tema concreto de la materia o del software utilizado. Los trabajos podrán ser propuestos por los docentes o por los propios alumnos (las propuestas serán tenidas en cuenta o no siempre según el criterio del docente).	40

Observaciones evaluación

Primera oportunidad

Se realizará una prueba de respuesta múltiple de 10 a 20 preguntas que representa el 40% de la nota. Por otra parte, la evaluación continua constará de la asistencia y/o entrega de prácticas relacionadas con el aprendizaje y aplicación del software estadístico R para la resolución de problemas en el campo de la gestión digital de la información (20% de la nota global), además de la entrega de uno o varios trabajos de aplicación de la estadística para la resolución de casos de estudio en documentación digital (alternativamente podrán ser trabajos de revisión o ampliación de la materia) que representa el 40% de la nota total.

Segunda oportunidad

En la evaluación de la segunda oportunidad se seguirá el mismo criterio que en la primera.

Convocatoria adelantada

Todas las observaciones previas son aplicables a los estudiantes que soliciten la convocatoria adelantada del examen.

Calificación de no presentado

En cualquiera de las dos oportunidades anuales figurará un NO PRESENTADO en aquellos casos en los que el alumnado no acuda al examen oficial de la materia.

Estudiante con reconocimiento de dedicación a tiempo parcial y dispensa académica de exención de asistencia

En el caso del alumnado con reconocimiento de dedicación a tiempo parcial y dispensa académica de exención de asistencia que decida no asistir a clases, éste será evaluado en las dos oportunidades como el resto del alumnado que se encuentra en una situación similar.

La realización fraudulenta de las pruebas o actividades de evaluación implicará directamente la calificación de suspenso "0" en la materia en la convocatoria correspondiente, invalidando así cualquier calificación obtenida en todas las actividades de evaluación de cara a la convocatoria extraordinaria.

Fuentes de información	
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
Complementária	

Recomendaciones
Asignaturas que se recomienda haber cursado previamente
Ciencia de Datos/710G04026 Fundamentos Estadísticos/710G04040
Asignaturas que se recomienda cursar simultáneamente
Asignaturas que continúan el temario
Otros comentarios

Para ayudar a conseguir un entorno inmediato sostenible y cumplir con el objetivo de la acción número 5: ?Docencia e investigación saludable y sustentable ambiental y social? del "Plan de Acción Green Campus Ferrol":1.- La entrega de los trabajos documentales que se realicen en esta materia: 1.1. Se solicitará en formato virtual y/o soporte informático. 1.2. Se realizará a través de Moodle, en formato digital sin necesidad de imprimirlos 1.3. De realizarse en papel: -No se emplearán plásticos. - Se realizarán impresiones a doble cara. - Se empleará papel reciclado. - Se evitará la impresión de borradores. 2.- Se debe hacer un uso sostenible de los recursos y la prevención de impactos negativos sobre el medio natural. 3.- Se debe tener en cuenta a importancia de los principios éticos relacionados con los valores de la sostenibilidad en los comportamientos personales y profesionales. 4.- Según se recoge en las distintas normativas de aplicación para la docencia universitaria se deberá incorporar la perspectiva de género en esta materia (se usará lenguaje no sexista, se utilizará bibliografía de autores de ambos sexos, se propiciará la intervención en clase de alumnos y alumnas?).5.- Se trabajará para identificar y modificar perjuicios y actitud sexistas, y se influirá en el entorno para modificarlos y fomentar valores de respeto e igualdad. 6. Se deberán detectar situaciones de discriminación por razón de género y se propondrán acciones y medidas para corregirlas. 7. Se facilitará la plena integración del alumnado que por razón físicas, sensoriales, psíquicas o socioculturales, experimenten dificultades a un acceso idóneo, igualitario y provechoso a la vida universitaria.

(*) La Guía Docente es el documento donde se visualiza la propuesta académica de la UDC. Este documento es público y no se puede modificar, salvo cosas excepcionales bajo la revisión del órgano competente de acuerdo a la normativa vigente que establece el proceso de elaboración de guías