

Guía docente

Identificación de la asignatura

Asignatura / Grupo	11640 - Nuevas Tendencias en Minería de Datos / 1
Titulación	Máster Universitario en Análisis de Datos Masivos en Economía y Empresa
Créditos	6
Período de impartición	Segundo semestre
Idioma de impartición	Castellano

Profesores

Horario de atención a los alumnos

Profesor/a	Hora de inicio	Hora de fin	Día	Fecha inicial	Fecha final	Despacho / Edificio
Sebastián Massanet Massanet s.massanet@uib.es						Hay que concertar cita previa con el/la profesor/a para hacer una tutoría
Arnau Mir Torres arnau.mir@uib.es						Hay que concertar cita previa con el/la profesor/a para hacer una tutoría
Margarita María Lourdes Miró Julià margaret.miro@uib.es						Hay que concertar cita previa con el/la profesor/a para hacer una tutoría

Contextualización

Con un recorrido de más de dos décadas de investigación en minería de datos y sus aplicaciones, el futuro de esta disciplina es indiscutible. En este tiempo, la investigación en minería de datos ha permitido desarrollar técnicas y herramientas que facilitan el proceso de creación de conocimiento. Aunque en algunas áreas esta disciplina ha sido estudiada en profundidad, en otros campos su uso es aún incipiente. Sin embargo, como en cualquier área de investigación activa, aparecen nuevas tendencias en las tecnologías, métodos y aplicaciones que o bien mejoran el conocimiento establecido hasta el momento o bien responden a nuevos problemas que surgen constantemente. La asignatura "Nuevas tendencias en minería de datos", optativa englobada en el módulo "Herramientas en gestión y análisis inteligente de datos", pretende introducir las líneas actuales de investigación en minería de datos, analizando las técnicas, los modelos y los algoritmos desarrollados para resolver algunas situaciones problemáticas de interés en los últimos años.

Requisitos

Guía docente

Recomendables

Es aconsejable haber cursado las asignaturas Tecnologías para el análisis de datos masivos y Aprendizaje estadístico y toma de decisiones.

Competencias

Específicas

- * CE6: Saber sintetizar y describir una gran cantidad de datos seleccionando los estadísticos adecuados al tipo de variables y analizar las relaciones existentes entre ellas
- * CE7: Capacidad para modelar la dependencia entre una variable respuesta y varias variables explicativas, en conjuntos de datos complejos, mediante técnicas de regresión, e interpretando los resultados obtenidos.
- * CE8: Capacidad para analizar un conjunto dado de variables mediante técnicas de clasificación interpretando los resultados obtenidos.
- * CE15: Capacidad de modelizar fenómenos reales a través de vectores aleatorios y de aplicar las principales técnicas de análisis multivariante en el contexto de la industria y empresa.

Genéricas

- * CG2: Elaborar adecuadamente y con originalidad argumentos motivados y proyectos de trabajo, redactar planes, informes profesionales así como formular hipótesis y conjeturas razonables en su área de especialización.
- * CG4: Comprender y utilizar el lenguaje y las herramientas asociadas al análisis de datos para modelizar y resolver problemas complejos, reconociendo y valorando las situaciones y problemas susceptibles de ser tratados utilizando dichas herramientas y las técnicas asociadas
- * CG5: Conocer los modelos, métodos y técnicas relevantes en distintas áreas de aplicación de la Estadística participando en la creación de nuevas tecnologías que contribuyan al desarrollo de la Sociedad de la Información

Básicas

- * Se pueden consultar las competencias básicas que el estudiante tiene que haber adquirido al finalizar el máster en la siguiente dirección: http://estudis.uib.cat/es/master/comp_basiques/

Contenidos

Después de contextualizar la investigación en minería de datos y las nuevas tendencias existentes en el campo, se dividirán los contenidos de la asignatura en varios bloques temáticos correspondientes a algunas de las tendencias actuales en investigación en esta materia.

Contenidos temáticos

- Tema 1. Introducción a las líneas actuales en minería de datos
- Tema 2. Descubrimiento de subgrupos
- Tema 3. Clasificación no balanceada
- Tema 4. Detección de anomalías
- Tema 5. Reglas de asociación anómalas

Guía docente

Tema 6. Minería de datos en estructuras complejas: Graph Mining

Tema 7. Aprendizaje basado en múltiples instancias

Metodología docente

Se expondrá un breve contenido teórico de los temas a través de clases presenciales, siguiendo los libros de texto de referencia y utilizando, cuando sea conveniente, medios informáticos, que servirán para fijar los conocimientos necesarios para desarrollar las competencias previstas. Las clases presenciales de problemas permitirán a los estudiantes profundizar en los conceptos desarrollados. Por ello un buen aprendizaje de las técnicas en las clases prácticas presenciales establecidas será un objetivo esencial de la asignatura. Para alcanzar tal fin, los estudiantes dispondrán, vía la plataforma “Aula Digital” o a través de fotocopias, de aquel material docente que se estime oportuno y en particular de los correspondientes enunciados de problemas con objeto de poder trabajar en ellos con antelación.

Volumen

En el volumen de trabajo se contempla la posibilidad de cesión de horas de docencia a actividades comunes del máster: seminarios, charlas, etc. que sean propuestos por la dirección del máster.

Actividades de trabajo presencial (1,44 créditos, 36 horas)

Modalidad	Nombre	Tip. agr.	Descripción	Horas
Clases teóricas	Clases magistrales	Grupo grande (G)	Se expondrá un breve contenido teórico de los temas a través de clases presenciales, siguiendo los libros y trabajos de referencia y utilizando cuando sea conveniente medios informáticos, que servirán para fijar los conocimientos necesarios para desarrollar las competencias previstas. Se trabajarán las competencias específicas C6, CE7, CE8 y CE15; y las competencias genéricas CG2, CG4 y CG5.	18
Seminarios y talleres	Seminarios, clases con ordenador	Grupo mediano (M)	Exposición e introducción de herramientas informáticas adecuadas para el estudio y desarrollo de la materia. Se trabajarán las competencias específicas CE6, CE7, CE8 y CE15; y las competencias genéricas CG2, CG4 y CG5.	4
Seminarios y talleres	Seminarios y talleres de problemas y prácticas con ordenador	Grupo mediano (M)	Trabajos adicionales y/o resolución de problemas y prácticas con ordenador, conjuntamente profesor y alumnado en un ambiente de taller o seminario y en grupo mediano. Se trabajarán las competencias específicas CE6, CE7, CE8 y CE15; y las competencias genéricas CG2, CG4 y CG5.	6
Clases prácticas	Problemas y prácticas con ordenador evaluables	Grupo mediano (M)	Las clases presenciales de problemas y prácticas ante ordenador permitirán a los estudiantes profundizar en los conceptos desarrollados. Por ello un buen aprendizaje de las técnicas en las clases prácticas presenciales establecidas será un objetivo esencial de la asignatura. Se pretende preparar al alumno para que pueda afrontar la resolución de problemas por sí mismo o en grupo. Se trabajarán las competencias específicas CE6, CE7, CE8 y CE15; y las competencias genéricas CG2, CG4 y CG5.	6

Guía docente

Modalidad	Nombre	Tip. agr.	Descripción	Horas
Evaluación	Pruebas objetivas y/o escritas de desarrollo	Grupo grande (G)	Prueba escrita de mínimos sobre conceptos generales de la materia impartida, para evaluar la adquisición de conocimientos por los alumnos y el nivel de las competencias específicas y algunas genéricas. Se trabajarán las competencias específicas CE6, CE7, CE8 y CE15; y las competencias genéricas CG2, CG4 y CG5.	2

Al inicio del semestre estará a disposición de los estudiantes el cronograma de la asignatura a través de la plataforma UIBdigital. Este cronograma incluirá al menos las fechas en las que se realizarán las pruebas de evaluación continua y las fechas de entrega de los trabajos. Asimismo, el profesor o la profesora informará a los estudiantes si el plan de trabajo de la asignatura se realizará a través del cronograma o mediante otra vía, incluida la plataforma Aula Digital.

Actividades de trabajo no presencial (4,56 créditos, 114 horas)

Modalidad	Nombre	Descripción	Horas
Estudio y trabajo autónomo individual o en grupo	Estudio teórico y práctico	Estudio autónomo del alumno (individual y en grupo) dedicado tanto al estudio de la teoría como la resolución de problemas a lo largo del curso. Se trabajarán las competencias específicas CE6, CE7, CE8 y CE15; y las competencias genéricas CG2, CG4 y CG5.	80
Estudio y trabajo autónomo individual o en grupo	Preparación y entrega de problemas y trabajos	Estudio, análisis, redacción y entrega de la memoria de una práctica o trabajo de la asignatura más elaborado. Se trabajarán las competencias específicas CE6, CE7, CE8 y CE15; y las competencias genéricas CG2, CG4 y CG5.	34

Riesgos específicos y medidas de protección

Las actividades de aprendizaje de esta asignatura no conllevan riesgos específicos para la seguridad y salud de los alumnos y, por tanto, no es necesario adoptar medidas de protección especiales.

Evaluación del aprendizaje del estudiante

La evaluación del aprendizaje consiste en un trabajo final, una prueba escrita de conceptos mínimos y los talleres de prácticas evaluables en el aula de informática. El estudiante tendrá una calificación numérica entre 0 y 10 para cada una de las actividades evaluadas. La calificación global se calcula teniendo en cuenta los pesos asignados a las diferentes actividades de evaluación. Se supera la asignatura si la calificación global no es inferior a 5.

En las actividades evaluables correspondientes a competencias de aprendizaje y conocimientos para comprender ámbitos de aplicación (CE1-15), se considerará la asistencia y participación en actividades propuestas por la dirección del máster.

Guía docente

Fraude en elementos de evaluación

De acuerdo con el artículo 33 del Reglamento Académico, "con independencia del procedimiento disciplinario que se pueda seguir contra el estudiante infractor, la realización demostrablemente fraudulenta de alguno de los elementos de evaluación incluidos en guías docentes de las asignaturas comportará, a criterio del profesor, una minusvaloración en su calificación que puede suponer la calificación de «suspense 0» en la evaluación anual de la asignatura".

Problemas y prácticas con ordenador evaluables

Modalidad	Clases prácticas
Técnica	Informes o memorias de prácticas (no recuperable)
Descripción	Las clases presenciales de problemas y prácticas ante ordenador permitirán a los estudiantes profundizar en los conceptos desarrollados. Por ello un buen aprendizaje de las técnicas en las clases prácticas presenciales establecidas será un objetivo esencial de la asignatura. Se pretende preparar al alumno para que pueda afrontar la resolución de problemas por sí mismo o en grupo. Se trabajarán las competencias específicas CE6, CE7, CE8 y CE15; y las competencias genéricas CG2, CG4 y CG5.
Criterios de evaluación	Desarrollo y entrega individual de pequeños trabajos y/o problemas resueltos por los alumnos. Se evaluarán las competencias específicas CE6, CE7, CE8 y CE15; y las competencias genéricas CG2, CG4 y CG5.

Porcentaje de la calificación final: 30%

Pruebas objetivas y/o escritas de desarrollo

Modalidad	Evaluación
Técnica	Pruebas de respuesta larga, de desarrollo (no recuperable)
Descripción	Prueba escrita de mínimos sobre conceptos generales de la materia impartida, para evaluar la adquisición de conocimientos por los alumnos y el nivel de las competencias específicas y algunas genéricas. Se trabajarán las competencias específicas CE6, CE7, CE8 y CE15; y las competencias genéricas CG2, CG4 y CG5.
Criterios de evaluación	Prueba escrita de mínimos sobre conceptos generales de la materia impartida, para evaluar la adquisición de conocimientos por los alumnos. Se evaluarán las competencias específicas CE6, CE7, CE8 y CE15; y las competencias genéricas CG2, CG4 y CG5.

Porcentaje de la calificación final: 40%

Guía docente

Preparación y entrega de problemas y trabajos

Modalidad	Estudio y trabajo autónomo individual o en grupo
Técnica	Trabajos y proyectos (no recuperable)
Descripción	Estudio, análisis, redacción y entrega de la memoria de una práctica o trabajo de la asignatura más elaborado. Se trabajarán las competencias específicas CE6, CE7, CE8 y CE15; y las competencias genéricas CG2, CG4 y CG5.
Criterios de evaluación	Redacción y entrega de la memoria de una práctica o trabajo de la asignatura más elaborado de manera individual o en grupo. Se evaluarán las competencias específicas CE6, CE7, CE8 y CE15; y las competencias genéricas CG2, CG4 y CG5.

Porcentaje de la calificación final: 30%

Recursos, bibliografía y documentación complementaria

Bibliografía básica

- * F. Herrera, C. J. Carmona, P. González, M. J. del Jesús: An overview on subgroup discovery: foundations and applications. *Knowl. Inf. Syst.* 29(3): 495-525 (2011)
- * A. Fernández, S. García, M. Galar, R. C. Prati, B. Krawczyk, F. Herrera: Learning from Imbalanced Data Sets. Springer 2018.
- * V. Chandola, A. Banerjee, V. Kumar: Anomaly detection: A survey. *ACM Comput. Surv.* 41(3) (2009)
- * M. D. Ruiz, M. J. Martín-Bautista, D. Sánchez, M. A. Vila Miranda, M. Delgado: Anomaly detection using fuzzy association rules. *IJESDF* 6(1): 25-37 (2014)
- * D. Chakrabarti, C. Faloutsos: Graph Mining: Laws, Tools, and Case Studies. In *Synthesis Lectures on Data Mining and Knowledge Discovery Series*. Morgan & Claypool Publishers (2012)
- * J. Amores. Multiple Instance Classification: Review, Taxonomy and Comparative Study. *Artificial Intelligence* 201: 81-105 (2013)
- * Samatova, N.F., Hendrix, W., Jenkins, J., Padmanabhan, K., Chakraborty, A. Practical Graph Mining with R. *Data Mining and Knowledge Discovery Series*. Chapman & Hall/CRC, 2014.

Otros recursos

Mediante la plataforma de tele-educación "Aula Digital" en el espacio web dedicado a la asignatura, el alumno tendrá a su disposición una serie de recursos de interés para su formación, como documentos electrónicos sobre la materia elaborados por el profesorado responsable de la asignatura y enlaces a Internet.