

Año académico	2019-20
Asignatura	10721 - Introducción a la Minería de Datos. Extracción de Información Oculta ...
Grupo	Grupo 1

Identificación de la asignatura

Asignatura / Grupo	10721 - Introducción a la Minería de Datos. Extracción de Información Oculta ... / 1
Titulación	Máster U. en Gestión de Recursos Humanos. Interv. Psicológica y Pedagógica
Créditos	5
Período de impartición	Primer 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
Rafael Jiménez López <i>Responsable</i> rafa.jimenez@uib.es	14:00	15:00	Martes	17/02/2020	17/07/2020	A-213 (Ed. Guillem Cifre de Colonya)

Contextualización

Esta asignatura semestral, de formación común y obligatoria, está enmarcada en el Plan de *Máster en Gestión de Recursos Humanos. Intervención psicológica y pedagógica* de la Universitat de les Illes Balears y pertenece al Área de Metodología de las Ciencias del Comportamiento. La asignatura consta de 5 créditos (125 horas), de los cuales, 1.2 créditos son presenciales (30 horas) y el resto no presenciales (95 horas).

La materia integrada en la asignatura pertenece al módulo de *Metodología de investigación e intervención en las organizaciones*, al que pertenecen también las asignaturas *Usos y abusos de la estadística* y *Técnicas psicométricas de validación de constructo*.

La asignatura proporciona un conjunto de resultados de aprendizaje que se concretan en los siguientes puntos:

- * Conocer y utilizar la minería de datos para descubrir patrones y relaciones en los datos que queden recogidos en modelos de conocimiento válidos para describir la realidad bajo estudio y/o realizar predicciones.
- * Ser capaz de realizar análisis exploratorios de los datos para identificar variables importantes en la predicción del suceso de interés, determinar la utilidad de generar nuevas variables, y obtener información que afecta a la calidad de los datos (detección de datos irrelevantes, valores omitidos, valores alejados y contradicciones).
- * Capacidad para depurar los datos: tratamiento de valores omitidos, tratamiento de valores alejados, resolución de contradicciones.
- * Preparación de los datos para el modelado: selección definitiva de variables, selección de registros (muestras independientes, balanceado), generación de nuevas variables predictoras, transformación de variables.
- * Aplicación de técnicas de modelado adecuadas en función del tipo de problema formulado.
- * Planificar y desarrollar la estrategia de obtención del modelo de conocimiento. Explorar modelos alternativos en búsqueda del que mejor resuelva el problema planteado.
- * Conocer los fundamentos matemático-estadísticos de algunas de las principales técnicas de minería de datos.
- * Aplicar correctamente las técnicas de minería de datos utilizando software profesional al uso, y ser capaz de interpretar adecuadamente los resultados obtenidos.

Guía docente

* Escoger los contenidos fundamentales obtenidos de la aplicación de la minería de datos para su uso y difusión en informes científicos.

Esta asignatura proporciona, por tanto, una serie de herramientas que facilitan la labor del profesional en la toma de decisiones dentro de su campo de intervención, puesto que un análisis adecuado de los datos provenientes de su campo de trabajo le permitirá detectar patrones y regularidades, y en definitiva reducir incertidumbre a la hora de adoptar una decisión sobre qué medidas de intervención pueden ser más adecuadas.

Requisitos

Recomendables

La asignatura tiene como requisito recomendable el conocimiento de las bases de la metodología de investigación y del análisis de datos: ser capaz de diseñar una investigación, ser capaz de reconocer las distintas escalas de medida, y elaborar y reunir los datos de investigación (en un paquete estadístico), y ser capaz de realizar análisis descriptivos de los datos.

Competencias

Específicas

- * Conocer los fundamentos matemático-estadísticos de algunas de las principales técnicas estadísticas, numéricas y gráficas (competencia 17)
- * Escoger los contenidos estadísticos fundamentales para su uso y difusión en informes profesionales o literatura científica (competencia 18)
- * Conocer los fundamentos matemático-estadísticos de algunas de las principales técnicas de minería de datos (competencia 19)
- * Escoger los contenidos fundamentales obtenidos de la aplicación de la minería de datos para su uso y difusión en informes científicos (competencia 20)
- * Desarrollar la capacidad de resolver problemas con autonomía y siguiendo un adecuado proceso de toma de decisiones (competencia 23)
- * Aplicar correctamente las técnicas de análisis de datos, y ser capaz de interpretar adecuadamente los resultados obtenidos (competencia 38)
- * Capacidad para diagnosticar y resolver problemas (competencia b2)
- * Interpretar y manejar las tecnologías de información y comunicación (TIC) relativas al ámbito de estudio y desarrollo profesional (competencia b7)
- * Capacidad de comunicación oral y escrita (redacción de informes o documentos) (competencia c1)
- * Capacidad para respetar las aportaciones de los demás (competencia c3)
- * Capacidad para elaborar informes orales y escritos (competencia c7)
- * Capacidad para trabajar en equipo, coordinar actuaciones y colaborar eficazmente con otros/as, incluyendo equipos interdisciplinarios (competencia c8)
- * Saber ajustarse a las obligaciones deontológicas del desempeño profesional (competencia d6)

Genéricas

- * Desarrollar la capacidad de aplicar los conocimientos a la práctica y de adaptarse a situaciones y contextos distintos, transfiriendo aprendizajes (competencia 15)

Guía docente

- * Aplicar la capacidad de resolver problemas con autonomía y siguiendo un adecuado proceso de toma decisiones (competencia 17)
- * Capacidad para utilizar e incorporar adecuadamente en las diferentes actividades las tecnologías de la información y la comunicación (competencia 19)

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

Contenidos temáticos

Unidad Didáctica 1. Introducción a la minería de datos

- 1.1. Definición, origen y evolución de la minería de datos
- 1.2. Conceptos básicos de minería de datos
- 1.3. Ámbitos de aplicación de la minería de datos

Unidad Didáctica 2. Etapas en la minería de datos

- 2.1. Formulación del problema
- 2.2. Acceso a base de datos y selección de variables relevantes
- 2.3. Exploración y preprocesamiento de datos
- 2.4. Modelado predictivo/descriptivo
- 2.5. Evaluación de modelos y selección del modelo definitivo

Unidad Didáctica 3. Técnicas descriptivas

- 3.1. Técnicas de exploración de datos
- 3.2. Modelado descriptivo
 - 3.2.1. *Técnicas de agrupamiento*
 - 3.2.2. *Reglas de asociación*
- 3.3. Interpretación de resultados

Unidad Didáctica 4. Técnicas predictivas

- 4.1. Modelado predictivo
 - 4.1.1. *Redes neuronales artificiales*
 - 4.1.2. *Árboles de decisión*
 - 4.1.3. *K-nearest neighbor*
 - 4.1.4. *Naive Bayes*
 - 4.1.5. *Técnicas estadísticas clásicas*
- 4.2. Evaluación del rendimiento del modelo

Año académico	2019-20
Asignatura	10721 - Introducción a la Minería de Datos. Extracción de Información Oculta ...
Grupo	Grupo 1

4.3. Interpretación de resultados

Metodología docente

En este apartado se describen las actividades de trabajo presencial y no presencial (o autónomo) previstas en la asignatura.

Actividades de trabajo presencial (1,2 créditos, 30 horas)

Modalidad	Nombre	Tip. agr.	Descripción	Horas
Clases teóricas	Clases magistrales	Grupo grande (G)	Mediante el método expositivo, el profesor establecerá los fundamentos teóricos, así como la ejemplificación práctica de las técnicas y los procedimientos de minería de datos de las unidades didácticas que componen la materia. Además, se proporcionará información, para cada unidad didáctica, sobre el método de trabajo aconsejable y el material didáctico que deberá utilizar el alumnado para preparar de forma autónoma los contenidos. Las clases teóricas constan de 6 sesiones (sesiones de 2 a 4 horas).	14
Clases prácticas	Prácticas presenciales	Grupo grande (G)	Mediante el método de resolución de ejercicios y problemas, el alumnado pondrá en práctica los procedimientos y técnicas de minería de datos expuestos en las clases teóricas. Además, dichas prácticas se realizarán en aula informática, para poder hacer uso de aplicaciones apropiadas para la minería de datos. Las clases prácticas constan de 5 sesiones (sesiones de 2 a 4 horas).	12
Evaluación	Prácticas evaluativas	Grupo grande (G)	Se realizarán dos prácticas evaluativas que permitirán valorar si el alumnado conoce y sabe aplicar correctamente los procedimientos y técnicas que forman parte de la materia.	4

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 (3,8 créditos, 95 horas)

Modalidad	Nombre	Descripción	Horas
Estudio y trabajo autónomo individual	Preparación de las unidades didácticas	Después de la exposición por parte del profesor en las clases magistrales, el alumnado deberá profundizar en la materia. Para facilitar esta tarea, se indicará, para cada unidad didáctica, los manuales que se han de consultar.	25
Estudio y trabajo autónomo individual	Resolución de las prácticas individuales	Se propondrán una serie de prácticas de trabajo autónomo a lo largo del curso, consistentes en la resolución de una serie de ejercicios. Las soluciones de los ejercicios propuestos se proporcionarán en las clases	50

Año académico	2019-20
Asignatura	10721 - Introducción a la Minería de Datos. Extracción de Información Oculta ...
Grupo	Grupo 1

Modalidad	Nombre	Descripción	Horas
		prácticas presenciales, y también estarán disponibles en el <i>Aula Digital</i> de la asignatura.	
Estudio y trabajo autónomo en grupo	Informe de prácticas: minería de datos sobre una matriz	Para desarrollar esta actividad, los estudiantes deberán organizarse en grupos* de 2 ó 3 personas. A partir de una matriz de datos proporcionada por el profesor, el grupo deberá aplicar de forma práctica los procedimientos y técnicas desarrollados en las sesiones teóricas y prácticas. El grupo deberá entregar, en formato electrónico, un informe del trabajo desarrollado. *La posibilidad de realizar el trabajo de manera individual solo se aceptará en casos excepcionales, debidamente justificados.	20

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

Los resultados de aprendizaje planificados serán valorados mediante la aplicación de una serie de procedimientos de evaluación. En el presente apartado se describe para cada procedimiento de evaluación los criterios de evaluación y su peso en la calificación global de la asignatura.

El estudiante obtendrá una calificación numérica entre 0 y 10 puntos para cada actividad evaluativa, la cual será ponderada según su peso a fin de obtener dicha calificación global de la asignatura. Para superar la asignatura, el alumno deberá cumplir necesariamente dos condiciones:

- 1) Obtener un promedio mínimo de 5 puntos sobre 10 en las dos prácticas evaluativas.
- 2) Obtener un mínimo de 5 puntos sobre 10 mediante la suma ponderada de todas las actividades (prácticas evaluativas + informe de prácticas).

Los estudiantes que no alcancen la calificación mínima de 5 puntos en las prácticas evaluativas (60% de la nota) podrán hacerlo en el período de recuperación a través de una práctica global (semana del 3 al 7 de febrero de 2020), con independencia de que hayan aprobado alguna de las dos prácticas evaluativas. La puntuación obtenida en esta práctica de recuperación se conmutará por el promedio de los dos prácticas evaluativas. La calificación final se calculará de nuevo a partir de la suma ponderada de los elementos evaluativos realizados, sustituyendo el promedio de las dos prácticas evaluativas por la nota obtenida en la práctica de recuperación.

Para superar la asignatura, el estudiante que haya realizado la práctica de recuperación deberá cumplir necesariamente las dos condiciones anteriormente establecidas (mínimo de 5 puntos sobre 10 en la práctica de recuperación, y un mínimo de 5 puntos sobre 10 mediante la suma ponderada de todas las actividades realizadas).

Si el estudiante no obtiene al menos una puntuación promedio de 5 entre las dos prácticas evaluativas o no obtiene una puntuación mínima de 5 en el examen de recuperación, y el cálculo ponderado de la nota final iguala o excede los 5 puntos, le corresponderá una calificación de “suspense” con una nota global de 4,5 puntos (reglamento académico, artículo 26, punto 7)

Año académico	2019-20
Asignatura	10721 - Introducción a la Minería de Datos. Extracción de Información Oculta ...
Grupo	Grupo 1

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".

Prácticas evaluativas

Modalidad	Evaluación
Técnica	Otros procedimientos (recuperable)
Descripción	Se realizarán dos prácticas evaluativas que permitirán valorar si el alumnado conoce y sabe aplicar correctamente los procedimientos y técnicas que forman parte de la materia.
Criterios de evaluación	Adecuación de los procedimientos aplicados para resolver los ejercicios propuestos y exactitud de los resultados obtenidos. Dominio de los conceptos consustanciales a los contenidos de la asignatura. Cada práctica evaluativa tendrá una duración máxima de 2 horas, y se realizará en aula informática (uso de aplicaciones apropiadas para la minería de datos). El criterio numérico de evaluación se adjuntará con el enunciado de la prueba. La práctica evaluativa 1 (representa un 30% de la calificación final) se ha fijado en el cronograma evaluativo el día 14 de octubre de 2019, y la práctica evaluativa 2 (representa un 30% de la calificación final) el día 4 de noviembre de 2019, ambas en horario de tarde.

Porcentaje de la calificación final: 60% con calificación mínima 5

Informe de prácticas: minería de datos sobre una matriz

Modalidad	Estudio y trabajo autónomo en grupo
Técnica	Informes o memorias de prácticas (no recuperable)
Descripción	Para desarrollar esta actividad, los estudiantes deberán organizarse en grupos* de 2 ó 3 personas. A partir de una matriz de datos proporcionada por el profesor, el grupo deberá aplicar de forma práctica los procedimientos y técnicas desarrollados en las sesiones teóricas y prácticas. El grupo deberá entregar, en formato electrónico, un informe del trabajo desarrollado. *La posibilidad de realizar el trabajo de manera individual solo se aceptará en casos excepcionales, debidamente justificados.
Criterios de evaluación	Adecuación de los procedimientos aplicados para resolver los ejercicios propuestos y exactitud de los resultados obtenidos. Adecuación de las interpretaciones y conclusiones establecidas en función de los resultados obtenidos. Dominio de los conceptos consustanciales a los contenidos de la asignatura. La fecha límite de entrega de dicha actividad evaluativa se ha fijado en el cronograma evaluativo el día 22 de noviembre de 2019.

Porcentaje de la calificación final: 40%

Recursos, bibliografía y documentación complementaria

Bibliografía básica

* Aggarwal, C. (2015). *Data Mining: The Textbook*. New York: Springer.



Año académico	2019-20
Asignatura	10721 - Introducción a la Minería de Datos. Extracción de Información Oculta ...
Grupo	Grupo 1

- * Dasu, T., & Johnson, T. (2003). *Exploratory Data Mining and Data Cleaning*. John Wiley & Sons.
- * Dunham, M.H. (2003). *Data mining introductory and advanced topics*. Upper Saddle River, NJ: Pearson Education.
- * Hernández, J., Ramírez, M.J., & Ferri, C. (2004). *Introducción a la Minería de Datos*. Madrid: Pearson Educación, S.A.
- * Jiménez, R. (2010). *Introducción a la minería de datos (data mining): extracción de información oculta en los datos*. Manual del curso.
- * Larose, D.T. (2005). *Discovering Knowledge in Data: An Introduction to Data Mining*. Hoboken, NJ: Wiley.
- * Pérez, M. (2014). *Minería de datos a través de ejemplos*. Madrid: RC Libros.
- * Smith, C. (2017). *Decision Trees and Random Forests: A Visual Introduction For Beginners*. Vancouver, CA: Blue Windmill Media.

Bibliografía complementaria

- * Daza, A. (2017). *Data Mining. Minería de datos*. Barcelona: Marcombo.
- * Han, J., & Kamber, M. (2000). *Data Mining: Concepts and Techniques*. San Francisco: Morgan Kaufmann.
- * Hand, D., Mannila, H., & Smyth, P. (2001). *Principles of Data Mining*. Cambridge, MA: The MIT Press.
- * Kantardzic, M. (2003). *Data Mining: Concepts, Models, Methods, and Algorithms*. New York: Wiley.
- * Makhabel, B. (2015). *Learning Data Mining with R*. Birmingham, UK: Packt Publishing.
- * Palmer, A. y Jiménez, R. (1999). *Introducción al Sistema Estadístico SPSS para Windows*. Palma de Mallorca: Edicions UIB. Col·lecció Materials Didàctics, 61. Sèrie de Metodologia. Universitat de les Illes Balears.
- * Tan, P.-N., Steinbach, M., & Kumar, V. (2006). *Introduction to Data Mining*. Boston: Pearson Education.
- * Two Crows Corporation (1999). *Introduction to Data Mining and Knowledge Discovery* (3ª ed.). Maryland: Two Crows Corporation.
- * Ye, N. (Ed.) (2003). *The Handbook of Data Mining*. Mahwah, NJ: Lawrence Erlbaum Associates.

Otros recursos

Se proporcionará al estudiante un compendio de recursos telemáticos de interés para su formación (aplicaciones informáticas, documentos electrónicos, tutoriales, bases de datos, etc.). Estos recursos suponen un complemento importante para el aprendizaje de los contenidos de la asignatura.