

GUÍA DOCENTE

Curso 2012-2013

1. DESCRIPCIÓN DE LA ASIGNATURA

Grado:	Ingeniería Informática en Sistemas de Información
Doble Grado:	
Asignatura:	Minería de Datos
Módulo:	Modulo 8
Departamento:	Deporte e Informática
Año académico:	2012/2013
Semestre:	Segundo Semestre
Créditos totales:	6
Curso:	3º
Carácter:	Optativo
Lengua de impartición:	Español

Modelo de docencia:	C1	
a. Enseñanzas Básicas (EB):		50%
b. Enseñanzas de Prácticas y Desarrollo (EPD):		50%
c. Actividades Dirigidas (AD):		

GUÍA DOCENTE

Curso 2012-2013

2. EQUIPO DOCENTE

2.1. Responsable de la asignatura

JESÚS SALVADOR AGUILAR RUIZ

2.2. Profesores

Nombre:	Jesús Salvador Aguilar Ruiz
Centro:	Escuela Politécnica Superior
Departamento:	Deporte e Informática
Área:	Lenguajes y Sistemas Informáticos
Categoría:	Profesor Titular de Universidad
Horario de tutorías:	Concertar por email
Número de despacho:	2.B.30
E-mail:	aguilar@upo.es
Teléfono:	954977940

GUÍA DOCENTE

Curso 2012-2013

3. UBICACIÓN EN EL PLAN FORMATIVO

3.1. Descripción de los objetivos

La asignatura de minería de datos se apoya fundamentalmente en tres pilares: estadística, inteligencia artificial y aprendizaje automático. En la asignatura se pretende mostrar al alumno las técnicas capaces de extraer conocimiento útil a partir de los datos, por lo que algunas nociones estadísticas y derivadas de la inteligencia artificial son necesarias. No obstante, la minería de datos ha evolucionado lo suficiente como para que el peso de la asignatura lo lleve en mayor medida el área del aprendizaje automático. En cualquier caso, unos fundamentos estadísticos básicos y un conocimiento de las técnicas heurísticas clásicas son recomendables y, por tanto, se incluyen dentro del temario de la asignatura.

Fijados los soportes de la asignatura, podríamos decir que el objetivo principal de la asignatura podría resumirse en: que el alumno conozca y sepa aplicar las técnicas de extracción de conocimiento a partir de datos. Esta definición, un tanto general, puede desglosarse en los siguientes puntos:

- Mostrar los conocimientos básicos de estadística.
- Mostrar las técnicas que mejoran la calidad de los datos iniciales.
- Mostrar las técnicas que extraen conocimiento a partir de los datos: fundamento teórico y aplicación práctica.
- Mostrar las técnicas que presentan el conocimiento extraído en forma legible por el no experto.
- Aprender a utilizar una herramienta de extracción de conocimiento de libre distribución, así como las técnicas de minería de datos implementadas.
- Aprender a desarrollar una planificación metódica de aplicación de técnicas para la resolución de problemas particulares.

Estos conocimientos teóricos deben completarse con los prácticos, que se adquieren en las clases de laboratorio mediante la utilización de una herramienta de minería de datos de libre distribución (WEKA).

Un objetivo final de la asignatura puede y debe de ser orientar al alumno en la realización de un Trabajo Fin de Grado. Los alumnos que se matriculan en la asignatura se encuentran habitualmente en la tesitura de elegir un tema para su trabajo, y en este sentido los trabajos propuestos de la asignatura pueden ser el germen de un trabajo más ambicioso que permita al alumno profundizar en un tema concreto de la minería de datos al tiempo que completa su formación con la realización del mismo.

GUÍA DOCENTE

Curso 2012-2013

Finalmente, se ha de mencionar que un objetivo colateral en la planificación de la asignatura es despertar el interés por esta temática, dado que el grupo de investigación al que pertenece el candidato dedica un gran esfuerzo en esta línea y, por tanto, una manifiesta apuesta por la capacidad de formación.

3.2. Aportaciones al plan formativo

La asignatura de Minería de Datos es una asignatura de 6 créditos, enmarcada en el Módulo 8 de Complementos Optativos Específicos de Sistemas de Información.

El módulo ofrece formación optativa complementaria en sistemas de información, incorporando temáticas emergentes como la computación ubicua –o inteligencia ambiental-, la minería de datos, la administración electrónica, los sistemas en tiempo real y los métodos para abordar una auditoría de sistemas de información.

Además de los resultados del aprendizaje directamente derivados de las competencias, se espera que los estudiantes hayan potenciado:

- la capacidad de análisis y síntesis,
- la habilidad para aplicar conocimientos de la ingeniería en otros contextos,
- la capacidad para redactar un informe técnico.

3.3. Recomendaciones o conocimientos previos requeridos

No hay.

GUÍA DOCENTE

Curso 2012-2013

4. COMPETENCIAS

4.1 Competencias de la Titulación que se desarrollan en la asignatura

4.1.1 Que los estudiantes tengan la capacidad de reunir e interpretar datos relevantes (normalmente dentro de su área de estudio) para emitir juicios que incluyan una reflexión sobre temas relevantes de índole social, científica o ética. (Competencia EB3)

4.1.2 Que los estudiantes puedan transmitir información, ideas, problemas y soluciones a un público tanto especializado como no especializado. (Competencia EB4)

4.1.3 Capacidad para planificar, concebir, desplegar y dirigir proyectos, servicios y sistemas informáticos en todos los ámbitos, liderando su puesta en marcha y su mejora continua y valorando su impacto económico y social. (Competencia EC02)

4.1.4 Habilidad de expresión oral y escrita, en español y/o inglés. (Competencia TI1)

4.1.5 Facilidad de trabajo en grupo multidisciplinar. (Competencia TP1)

4.1.6 Respeto en las relaciones interpersonales. (Competencia TP2)

4.2. Competencias del Módulo que se desarrollan en la asignatura

4.2.1 Conocimiento y aplicación de los principios fundamentales y técnicas básicas de los sistemas inteligentes y su aplicación práctica. (Competencia G6)

4.2.2 Capacidad de integrar soluciones de Tecnologías de la Información y las Comunicaciones y procesos empresariales para satisfacer las necesidades de información de las organizaciones, permitiéndoles alcanzar sus objetivos de forma efectiva y eficiente, dándoles así ventajas competitivas. (Competencia ET1)

4.3. Competencias particulares de la asignatura

4.3.1 Comparar y contrastar diferentes perspectivas de la minería de datos, tal como aparecen en la investigación y en la industria.

GUÍA DOCENTE

Curso 2012-2013

- 4.3.2 Explicar el papel de búsqueda de asociaciones en los datos.
- 4.3.3 Caracterizar los tipos de patrones que pueden ser descubiertos con técnicas de minería de datos.
- 4.3.4 Evaluar cuestiones metodológicas subyacentes de la aplicación de la minería de datos.
- 4.3.5 Identificar y caracterizar fuentes de ruido, redundancia y outliers en los datos. Explicar en qué consisten los algoritmos evolutivos y comparar su efectividad con técnicas clásicas de búsqueda.
- 4.3.6 Explicar cómo simulated annealing puede ser utilizado para reducir la complejidad de la búsqueda y comparar su metodología con las técnicas clásicas de búsqueda.
- 4.3.7 Aplicar técnicas de búsqueda local a problemas representativos.
- 4.3.8 Explicar las diferencias entre el aprendizaje supervisado y no supervisado.
- 4.3.9 Implementar algoritmos simples para aprendizaje supervisado y no supervisado.
- 4.3.10 Determinar la adecuación de una técnica a un problema particular.
- 4.3.11 Comparar y contrastar las siguientes técnicas, proporcionando ejemplos de cuándo una estrategia es superior a otra: árboles de decisión y redes neuronales.
- 4.3.12 Caracterizar el estado del arte del aprendizaje automático, incluyendo sus ventajas e inconvenientes.
- 4.3.13 Explicar el algoritmo del vecino más cercano y su papel dentro del aprendizaje automático.
- 4.3.14 Explicar el caso del sobreajuste, junto con técnicas para detectar y controlar el problema.

GUÍA DOCENTE

Curso 2012-2013

5. CONTENIDOS DE LA ASIGNATURA (TEMARIO)

1. Visión general: conceptos básicos.
 - 1.1. ¿Qué es la minería de datos?
 - 1.2. Proceso KDD. Fases:
 - 1.3. Definiciones básicas:
 - 1.4. Calidad de los datos:
 - 1.5. Fundamentos estadísticos:
2. Aprendizaje automático.
 - 2.1. Clasificación del aprendizaje: supervisado/ no supervisado.
 - 2.2. Tareas de minería de datos.
 - 2.3. Descripción de técnicas que no generan modelo de conocimiento: vecinos más cercanos, regresión, redes neuronales.
 - 2.4. Tipos de modelo de conocimiento: reglas (decisión y asociación) y árboles (decisión y regresión).
 - 2.5. Heurísticas de búsqueda.
3. Preprocesamiento.
 - 3.1. Técnicas para el tratamiento de valores omisos.
 - 3.2. Técnicas para el tratamiento de outliers.
 - 3.3. Técnicas para el tratamiento del ruido.
 - 3.4. Normalización.
 - 3.5. Estandarización.
 - 3.6. Técnicas de discretización.
 - 3.7. Técnicas de transformación
4. Editado.
 - 4.1. Vecinos más cercanos: k-NN.
 - 4.2. CNN
 - 4.3. Ibk
 - 4.4. EPO
5. Selección de Atributos.
 - 5.1. Clasificación: filter/wrapper.
 - 5.2. χ^2
 - 5.3. Entropía
 - 5.4. Relief/ReliefF
 - 5.5. Correlation-based Feature Selection (CFS)
 - 5.6. SOAP: proyecciones
6. Clustering.
 - 6.1. Clasificación de técnicas de clustering
 - 6.2. Clustering jerárquico
 - 6.3. Representación de clusters
7. Clasificación.

GUÍA DOCENTE

Curso 2012-2013

- 7.1. Vecinos más cercano.
- 7.2. J48.
- 8. Reglas de Asociación.
 - 8.1. Conceptos básicos: regla de asociación, soporte, confianza.
 - 8.2. Algoritmo Apriori.
- 9. Algoritmos Evolutivos.
 - 9.1. Esquema canónico de algoritmo evolutivo.
 - 9.2. Operadores genéticos.
 - 9.3. La optimización de funciones.
 - 9.4. Ideas generales sobre las estrategias evolutivas, la programación genética y los algoritmos de estimación de distribuciones.

GUÍA DOCENTE

Curso 2012-2013

6. METODOLOGÍA Y RECURSOS

Se distinguen dos tipos de clases: clases teóricas o Enseñanzas Básicas (EB) y clases prácticas de laboratorios o Enseñanzas Prácticas de Desarrollo (EPD).

En las clases teóricas, la exposición de cada uno de los temas comienza con una breve introducción donde se sitúa el tema en el contexto de la asignatura y se presenta una síntesis de su contenido, prosigue con el desarrollo del mismo y finaliza con unas conclusiones y un resumen de las principales ideas. Por otra parte, a lo largo de la clase se plantea a los alumnos una serie de actividades para hacerles participar y comprobar el grado de asimilación del tema. También se les suministran referencias bibliográficas y direcciones de Internet donde pueden encontrar información ampliada sobre el mismo tema. Las clases se imparten en el aula de teoría mediante una presentación realizada con un ordenador conectado al cañón de proyección.

El objetivo de las clases de laboratorio es aplicar los conocimientos explicados en las clases de teoría a la resolución de problemas de minería de datos, que pueden ser pequeñas implementaciones o, en mayor medida, análisis de datos existentes con las técnicas explicadas. Las clases se impartirán en un aula de laboratorio donde cada alumno tendrá a su disposición un ordenador. El profesor utilizará su ordenador para impartir las explicaciones que los alumnos podrán seguir en la pantalla del aula o, en el caso de la explicación de la herramienta, en su propio ordenador siguiendo las instrucciones del profesor.

Las tutorías desempeñan un papel importante en la asignatura, sobre todo en lo que respecta al trabajo de curso. Ya antes de que los alumnos comiencen a realizar su trabajo el profesor deberá orientarles en la elección del mismo, proporcionándoles detalles acerca de los distintos trabajos propuestos y eventualmente estudiando las propuestas presentadas por ellos.

Una vez adjudicados los trabajos, el profesor deberá facilitar a cada alumno la documentación adecuada para su trabajo en particular y atender las dudas que le vayan surgiendo durante el desarrollo del mismo. Al ser cada trabajo distinto de los demás, esto supone una atención personalizada a los alumnos y una gran exigencia al profesor. En este sentido resulta muy interesante potenciar las tradicionales tutorías con los nuevos medios como el correo electrónico o los grupos de usuarios que permitan una comunicación rápida y flexible entre el profesor y los alumnos, y entre los propios alumnos.

Todas estas actividades formativas estarán además complementadas con soporte a través de herramientas Web como la herramienta de Aula Virtual WebCT.

GUÍA DOCENTE

Curso 2012-2013

7. EVALUACIÓN

CONVOCATORIA DE JUNIO

La evaluación de la asignatura se acogerá al modelo de evaluación continua y sólo será aplicable para la convocatoria de junio (1ª convocatoria de curso). Aquellos estudiantes que no superen la asignatura en dicha convocatoria, se podrán evaluar en la convocatoria de julio como se describe más adelante.

Un examen escrito que permitirá evaluar la asimilación de los conocimientos impartidos en las clases de teoría. Consistirá en una serie de preguntas que podrán versar sobre conceptos teóricos o sobre la aplicación de estos conceptos a situaciones prácticas. Será una prueba más bien corta en la que no se incidirá en detalles muy concretos sino, antes al contrario, en ideas fundamentales repetidas una y otra vez a lo largo del curso. La calificación obtenida en el examen escrito ponderará un 20% en la nota final de la asignatura.

Además del examen escrito el alumno deberá realizar un trabajo que podrá escoger de entre una lista de trabajos propuestos por el profesor al principio del curso o bien será un trabajo propuesto por el propio alumno y que cuente con el visto bueno del profesor.

Los trabajos podrán ser de dos tipos:

1. *Teóricos*: trabajos de investigación y documentación sobre temas relacionados con la minería de datos.
2. *Prácticos*: desarrollo de una aplicación práctica mediante alguna de las herramientas utilizadas en las clases de laboratorio u otra similar.

La lista de trabajos propuestos se hará pública a principios del cuatrimestre, y los alumnos elegirán entre uno de ellos, que será realizado por un grupo de dos alumnos. Al final los alumnos presentarán, junto con el software desarrollado, un informe de 10 páginas que versará sobre el trabajo realizado, a modo de artículo, que recogerá un breve estado del arte de la temática elegida, la motivación, la propuesta per se, los experimentos realizados y la documentación bibliográfica. Este trabajo tendrá una valoración del 80% de la nota final.

Para la realización de cualquier prueba evaluable no se permitirá el uso o consulta de documentación, salvo indicación expresa del profesor/a.

GUÍA DOCENTE

Curso 2012-2013

CONVOCATORIA DE JULIO

Para la convocatoria de julio, la asignatura se evaluará mediante pruebas escritas correspondientes a Enseñanzas Básicas y Enseñanzas Prácticas y de Desarrollo que englobarán todos los contenidos de la asignatura. Las notas obtenidas en las pruebas evaluables durante la evaluación continua no se guardarán para convocatorias futuras.

8. BIBLIOGRAFÍA GENERAL

- Han, J., Kamber, M. *Data Mining: Concepts and Techniques*. Morgan Kaufman Publishers, 2000. ISBN 1558604898.
- Mitchell, T. *Machine Learning*. McGraw-Hill Science/Engineering/Math, 1997. ISBN 0070428077.
- Witten, I.H., Frank, E. *Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations*. Morgan Kaufmann Publishers, 1999. ISBN 1558605525
- Liu, H. *Feature Extraction, Construction and Selection: A Data Mining Perspective*. Kluwer Academic Publishers, 1998. ISBN 0792381963
- Michalewicz, Z. *Genetic Algorithms + Data Structures = Evolution Programs*. Springer-Verlag, 1997. ISBN 3540606769
- Soukup, T. *Visual Data Mining: Techniques and Tools for Data Visualization and Mining*. John Wiley & Sons, 2002. ISBN 0471149993