

ANX-PR/CL/001-02
GUÍA DE APRENDIZAJE

ASIGNATURA

Minería de datos

CURSO ACADÉMICO - SEMESTRE

2015-16 - Segundo semestre

Datos Descriptivos

Nombre de la Asignatura	Minería de datos
Titulación	10II - Grado en Ingeniería Informatica
Centro responsable de la titulación	E.T.S. de Ingenieros Informaticos
Semestre/s de impartición	Octavo semestre
Materia	Optatividad
Carácter	Optativa
Código UPM	105000169
Nombre en inglés	Data Mining

Datos Generales

Créditos	3	Curso	4
Curso Académico	2015-16	Período de impartición	Febrero-Junio
Idioma de impartición	Castellano	Otros idiomas de impartición	

Requisitos Previos Obligatorios

Asignaturas Previas Requeridas

El plan de estudios Grado en Ingeniería Informatica no tiene definidas asignaturas previas superadas para esta asignatura.

Otros Requisitos

El plan de estudios Grado en Ingeniería Informatica no tiene definidos otros requisitos para esta asignatura.

Conocimientos Previos

Asignaturas Previas Recomendadas

Probabilidades y estadística I

Otros Conocimientos Previos Recomendados

El coordinador de la asignatura no ha definido otros conocimientos previos recomendados.

Competencias

CG-1/21 - Capacidad de resolución de problemas aplicando conocimientos de matemáticas, ciencias e ingeniería.

CG-13/CE55 - Capacidad de comunicarse de forma efectiva con los compañeros, usuarios (potenciales) y el público en general acerca de cuestiones reales y problemas relacionados con la especialización elegida.

CG-19 - Capacidad de usar las tecnologías de la información y la comunicación.

CG-2/CE45 - Capacidad para el aprendizaje autónomo y la actualización de conocimientos, y reconocimiento de su necesidad en el área de la informática.

CG-5 - Capacidad de gestión de la información.

CG-6 - Capacidad de abstracción, análisis y síntesis

Ce 12/16 - Conocer los campos de aplicación de la informática, y tener una apreciación de la necesidad de poseer unos conocimientos técnicos profundos en ciertas áreas de aplicación; apreciación del grado de esta necesidad en, por lo menos, una situación.

Ce 14/15 - Conocer el software, el hardware y las aplicaciones existentes en el mercado, así como el uso de sus elementos, y capacidad para familiarizarse con nuevas aplicaciones informáticas.

Ce 17 - Conocer los temas informáticos avanzados de modo que permita a los alumnos vislumbrar y entender las fronteras de la disciplina, por medio de la inclusión de experiencias de aprendizaje que dirigen a los alumnos desde los temas elementales a los temas avanzados o los temas de los que se nutren los novísimos desarrollos.

Ce 44 - Conocimiento de tecnologías punteras relevantes y su aplicación.

Resultados de Aprendizaje

RA277 - Dado un problema real elegir la tecnología informática existente en el mercado mas apropiada para su solución y diseñar su desarrollo e integración, analizando la viabilidad de su solución, lo que se puede y no se puede conseguir a través del estado actual de desarrollo de la tecnología usada, y lo que se espera que avance en el futuro.

RA278 - Desarrollar la solución matemática y algorítmica mas apropiada a un problema informático que requiera un tratamiento especialmente complejo, analizando y exponiendo su viabilidad.

RA280 - Obtención de las competencias lingüísticas comunicativas (comprensión, expresión, etc.) habladas y escritas en entornos académicos/profesionales nacionales/internacionales.

RA281 - Obtención de las técnicas necesarias para la realización de un informe o memoria sobre un trabajo realizado en un entorno socio-lingüístico nacional/internacional.

Profesorado

Profesorado

Nombre	Despacho	e-mail	Tutorías
Larrañaga Mugica, Pedro Maria (Coordinador/a)	2208	pedro.larranaga@upm.es	M - 15:00 - 19:00 V - 15:00 - 17:00
Bielza Lozoya, Maria Concepcion	2210	c.bielza@upm.es	M - 15:00 - 19:00 V - 15:00 - 17:00
Serrano Fernandez, Emilio	2201	emilio.serrano@upm.es	L - 15:00 - 16:00 M - 11:00 - 13:00 X - 15:00 - 16:00

Nota.- Las horas de tutoría son orientativas y pueden sufrir modificaciones. Se deberá confirmar los horarios de tutorías con el profesorado.

Descripción de la Asignatura

La Minería de Datos trata de construir sistemas informáticos cuando no existe experiencia humana o cuando no es fácilmente explicable. Así, los datos se transforman en conocimiento. Entre las muchas aplicaciones exitosas pueden citarse el reconocimiento del habla o de texto manuscrito, navegación autónoma de robots, recuperación de información documental, filtrado cooperativo, sistemas de diagnóstico, análisis de microarrays de ADN, etc.

Este curso expone varios métodos de Minería de Datos desde el punto de vista práctico. El objetivo es enfrentarse a un conjunto de datos sobre los que construir modelos clasificatorios con la ayuda del software WEKA.

Temario

1. Introducción a la minería de datos
 - 1.1. Reconocimiento de patrones
 - 1.2. Preproceso de datos
2. Clasificación supervisada
 - 2.1. Métodos de evaluación
 - 2.2. Vecinos más cercanos
 - 2.3. Clasificadores Bayesianos
 - 2.4. Regresión logística
 - 2.5. Árboles de clasificación
 - 2.6. Inducción de reglas
 - 2.7. Redes neuronales
 - 2.8. Máquinas de vector soporte
 - 2.9. Selección de variables
 - 2.10. Metaclasificadores
3. Clasificación no supervisada
 - 3.1. Métodos particionales
 - 3.2. Clasificación ascendente jerárquica
 - 3.3. Clustering probabilista

Cronograma

Horas totales: 32 horas

Horas presenciales: 32 horas (41%)

Peso total de actividades de evaluación continua:
100%

Peso total de actividades de evaluación sólo prueba final:
100%

Semana	Actividad Presencial en Aula	Actividad Presencial en Laboratorio	Otra Actividad Presencial	Actividades Evaluación
Semana 1	Clases teóricas del tema 1 Duración: 02:00 LM: Actividad del tipo Lección Magistral			
Semana 2	Clases teóricas del tema 1 Duración: 01:00 LM: Actividad del tipo Lección Magistral	Clases prácticas del tema 1 Duración: 01:00 PL: Actividad del tipo Prácticas de Laboratorio		
Semana 3	Clases teóricas del tema 2.1 Duración: 01:00 LM: Actividad del tipo Lección Magistral	Clases prácticas del tema 2.1 Duración: 01:00 PL: Actividad del tipo Prácticas de Laboratorio		
Semana 4	Clases teóricas del tema 2.2 Duración: 01:00 LM: Actividad del tipo Lección Magistral	Clases prácticas del tema 2.2 Duración: 01:00 PL: Actividad del tipo Prácticas de Laboratorio		
Semana 5	Clases teóricas del tema 2.3 Duración: 01:00 LM: Actividad del tipo Lección Magistral	Clases prácticas del tema 2.3 Duración: 01:00 PL: Actividad del tipo Prácticas de Laboratorio		
Semana 6	Clases teóricas del tema 2.4 Duración: 01:00 LM: Actividad del tipo Lección Magistral	Clases prácticas del tema 2.4 Duración: 01:00 PL: Actividad del tipo Prácticas de Laboratorio		
Semana 7	Clases teóricas del tema 2.5 Duración: 01:00 LM: Actividad del tipo Lección Magistral	Clases prácticas del tema 2.5 Duración: 01:00 PL: Actividad del tipo Prácticas de Laboratorio		
Semana 8	Clases teóricas del tema 2.6 Duración: 01:00 LM: Actividad del tipo Lección Magistral	Clases prácticas del tema 2.6 Duración: 01:00 PL: Actividad del tipo Prácticas de Laboratorio		
Semana 9	Clases teóricas del tema 2.7 Duración: 01:00 LM: Actividad del tipo Lección Magistral	Clases prácticas del tema 2.7 Duración: 01:00 PL: Actividad del tipo Prácticas de Laboratorio		
Semana 10	Clases teóricas del tema 2.8 Duración: 01:00 LM: Actividad del tipo Lección Magistral	Clases prácticas del tema 2.8 Duración: 01:00 PL: Actividad del tipo Prácticas de Laboratorio		
Semana 11	Clases teóricas del tema 2.9 Duración: 01:00 LM: Actividad del tipo Lección Magistral	Clases prácticas del tema 2.9 Duración: 01:00 PL: Actividad del tipo Prácticas de Laboratorio		
Semana 12	Clases teóricas del tema 2.10 Duración: 01:00 LM: Actividad del tipo Lección Magistral	Clases prácticas del tema 2.10 Duración: 01:00 PL: Actividad del tipo Prácticas de Laboratorio		

Semana 13				<p>Entrega y exposición oral del trabajo individual sobre clasificación supervisada</p> <p>Duración: 02:00</p> <p>PI: Técnica del tipo Presentación Individual</p> <p>Evaluación continua</p> <p>Actividad presencial</p>
Semana 14	<p>Clases teóricas del tema 3.1 y 3.2</p> <p>Duración: 01:00</p> <p>LM: Actividad del tipo Lección Magistral</p>	<p>Clases prácticas del tema 3.1 y 3.2</p> <p>Duración: 01:00</p> <p>PL: Actividad del tipo Prácticas de Laboratorio</p>		
Semana 15	<p>Clases teóricas del tema 3.2 y 3.3</p> <p>Duración: 01:00</p> <p>LM: Actividad del tipo Lección Magistral</p>	<p>Clases prácticas del tema 3.2 y 3.3</p> <p>Duración: 01:00</p> <p>PL: Actividad del tipo Prácticas de Laboratorio</p>		
Semana 16				<p>Entrega y exposición oral del trabajo individual sobre clasificación no supervisada</p> <p>Duración: 02:00</p> <p>PI: Técnica del tipo Presentación Individual</p> <p>Evaluación continua</p> <p>Actividad presencial</p>
Semana 17				<p>Entrega y exposición oral del trabajo individual sobre clasificación supervisada y no supervisada</p> <p>Duración: 02:00</p> <p>PI: Técnica del tipo Presentación Individual</p> <p>Evaluación sólo prueba final</p> <p>Actividad presencial</p>

Nota.- El cronograma sigue una planificación teórica de la asignatura que puede sufrir modificaciones durante el curso.

Nota 2.- Para poder calcular correctamente la dedicación de un alumno, la duración de las actividades que se repiten en el tiempo (por ejemplo, subgrupos de prácticas") únicamente se indican la primera vez que se definen.

Actividades de Evaluación

Semana	Descripción	Duración	Tipo evaluación	Técnica evaluativa	Presencial	Peso	Nota mínima	Competencias evaluadas
13	Entrega y exposición oral del trabajo individual sobre clasificación supervisada	02:00	Evaluación continua	PI: Técnica del tipo Presentación Individual	Sí	60%	3 / 10	CG-6, Ce 44, Ce 14/15, CG-13/CE55, Ce 12/16, CG-1/21, Ce 17
16	Entrega y exposición oral del trabajo individual sobre clasificación no supervisada	02:00	Evaluación continua	PI: Técnica del tipo Presentación Individual	Sí	40%	3 / 10	CG-6, CG-5, CG-2/CE45, CG-13/CE55
17	Entrega y exposición oral del trabajo individual sobre clasificación supervisada y no supervisada	02:00	Evaluación sólo prueba final	PI: Técnica del tipo Presentación Individual	Sí	100%	5 / 10	Ce 14/15, CG-6, Ce 44, CG-13/CE55, Ce 12/16, CG-5, CG-1/21, CG-2/CE45, Ce 17

Criterios de Evaluación

La calificación vendrá dada a partir de la evaluación de:

1. El trabajo escrito sobre clasificación supervisada (tema 2) proporcionado por el alumno y su exposición oral en el aula. Se utilizará el software WEKA. Este trabajo tendrá un peso del 60%.
2. El trabajo escrito sobre clasificación no supervisada (tema 3) proporcionado por el alumno y su exposición oral en el aula. Se utilizará el software WEKA. Este trabajo tendrá un peso del 40%.

Estos pesos se utilizan para la nota final siempre que el alumno se presente a ambas partes y obtenga una nota mínima de 3 sobre 10 en cada parte. Si un alumno no se presenta a una de las dos partes, deberá presentar esta parte en la prueba final de esta convocatoria. Si habiéndose presentado a ambas partes una de ellas no superara el mínimo de 3 (su nota es inferior a 3), se presentará de nuevo a esta parte en Julio. Si, en cambio, ambas son superiores o iguales a 3, pero la nota final no resulta un aprobado (mayor o igual que 5 sobre 10), el alumno tendrá que presentarse en Julio a una de las partes o a las dos, lo que él decida.

Si el estudiante decide realizar sólo una prueba final, se efectuará en la fecha de examen final fijada por jefatura de estudios y consistirá (como en la evaluación continua) en las dos exposiciones orales y sus respectivos trabajos escritos realizados de forma individual, evaluándose con los mismos pesos que en la evaluación continua. La solicitud para hacer sólo una prueba final se tiene que realizar por email al coordinador de la asignatura antes de la primera entrega.

No podrá haber dos alumnos que utilicen los mismos datos y con el mismo objetivo.

Está terminantemente prohibido copiar o dejar copiar. Su detección implicará el suspenso inmediato y la repetición de los trabajos con datos diferentes.

Para la convocatoria extraordinaria de Julio las evaluaciones serán análogas (dos exposiciones orales y sus respectivos trabajos escritos realizados de forma individual).

Recursos Didácticos

Descripción	Tipo	Observaciones
Libro 1	Bibliografía	E. Alpaydin. Introduction to Machine Learning. MIT Press
Libro 3	Bibliografía	J. Hernández-Orallo, M.J. Ramírez, C. Ferri. Introducción a la Minería de Datos. Pearson Educación. 2004.
Libro 4	Bibliografía	L. Kuncheva. Combining Pattern Classifiers. Wiley. 2004
Libro 7	Bibliografía	I. Witten, E. Frank. Data Mining. Morgan Kaufmann. 2ª ed. 2005.
software	Recursos web	Software WEKA: http:// www.cs.waikato.ac.nz/ml/weka
Repositorios de datos	Recursos web	- http://www.inside-r.org/howto/finding-data-internet - http://archive.ics.uci.edu/ml/ - https://www.quora.com/Where-can-I-find-large-datasets-open-to-the-public