



ESCUELA TÉCNICA SUPERIOR DE INGENIERÍA

GUIA DOCENTE

CURSO 2022-23

GRADO EN INGENIERÍA INFORMÁTICA

DATOS DE LA ASIGNATURA

Nombre:

MINERÍA DE DATOS

Denominación en Inglés:

Data Mining

Código:

606010307

Tipo Docencia:

Presencial

Carácter:

Optativa

Horas:

Totales

Presenciales

No Presenciales

Trabajo Estimado

150

60

90

Créditos:

Grupos Grandes

Grupos Reducidos

Aula estándar

Laboratorio

Prácticas de campo

Aula de informática

4

0

2

0

0

Departamentos:

TECNOLOGIAS DE LA INFORMACION

Áreas de Conocimiento:

CIENCIA DE LA COMPUTACION E INTELIG. ARTIFICIAL

Curso:

4º - Cuarto

Cuatrimestre

Primer cuatrimestre

DATOS DEL PROFESORADO (*Profesorado coordinador de la asignatura)

Nombre:	E-mail:	Teléfono:
* Luis Ignacio Lopez Gomez	nacho@dti.uhu.es	959 217 683
Maria Teresa Santos Rodriguez	tsantos@dti.uhu.es	

Datos adicionales del profesorado (Tutorías, Horarios, Despachos, etc...)**Luis Ignacio López Gómez:**

Despacho 142, ETSI, 1ª planta.

Teléfono 37683.

Horarios:

<http://www.uhu.es/etsi/informacion-academica/informacion-comun-todos-los-titulos/horarios-2/>

Tutorías: <https://guiadocente.uhu.es/tutoria/titulacion>

email: nacho@uhu.es

Teresa Santos Rodríguez:

Despacho 153, ETSI 1ª planta.

Teléfono: 959217679.

Horarios:

<http://www.uhu.es/etsi/informacion-academica/informacion-comun-todos-los-titulos/horarios-2/>

Tutorías: <https://guiadocente.uhu.es/tutoria/titulacion>

email: tsantos@dti.uhu.es

DATOS ESPECÍFICOS DE LA ASIGNATURA

1. Descripción de Contenidos:

1.1 Breve descripción (en Castellano):

- Conceptos básicos de minería de datos y extracción de conocimiento
- Almacenes de datos y preprocesamiento
- Clasificación y objetivo de las técnicas de minería de datos
- Estudio de los modelos de datos: estadísticos, basados en reglas, árboles de decisión, máquinas de soporte vectorial, etc.
- Evaluación y validación de modelos
- Minería de datos en sistemas con restricciones especiales y específicos, temporales, multimedia, web, etc.

1.2 Breve descripción (en Inglés):

- Basic concepts of data mining and knowledge extraction
- Data stores and preprocessing
- Classification and objective of data mining techniques
- Study of data models: statistical, rule-based, decision trees, support vector machines, etc.
- Evaluation and validation of models
- Data mining in systems with special and specific restrictions, temporary, multimedia, web, etc.

2. Situación de la asignatura:

2.1 Contexto dentro de la titulación:

Existen distintas asignaturas en el grado relacionadas con Minería de Datos (MD). Especialmente complementaria con ella, es la de Aprendizaje Automático que tiene la orientación a la parte algorítmica y de implementación (programación) de cada técnica que se utiliza en minería de datos. En MD, el enfoque está más orientado al uso de éstas técnicas, independientemente de la implementación particular. Es decir, consideraremos el conjunto de técnicas como una biblioteca de componentes reutilizables, cada uno de los cuales será seleccionada de acuerdo a las características de la tarea que se requiere resolver. En otros casos, esta asignatura puede servir de introducción a otras asignaturas del programa de posgrado. Asignatura de carácter práctico y de importancia emergente dentro del campo denominado Soft Computing. Hoy se ha de contemplar a la MD como una parte del análisis de datos dentro del contexto del denominado BIG DATA.

2.2 Recomendaciones

Recomendaciones:

- Se deben tener los necesarios conceptos de aprendizaje automático, tanto a nivel simbólico como subsimbólico.
- Tener conocimientos de Bases de Datos, fundamentalmente las basadas en el modelo relacional de datos.
- El alumno debe tener conocimientos medios-avanzados de estadística, tanto la descriptiva como inferencial.
- Tener competencia en Álgebra Lineal.

3. Objetivos (Expresados como resultado del aprendizaje):

- Un acercamiento al campo para el estudio de tareas y técnicas, necesarias para la toma de decisiones de problemas complejos: problemas con solución inicial no algorítmica con datos estructurados, semiestructurados y no estructurados.
- Descripción del estado del arte: Big Data y nuevos enfoques para el tratamiento masivo de datos estructurados y no estructurados.

4. Competencias a adquirir por los estudiantes

4.1 Competencias específicas:

-

4.2 Competencias básicas, generales o transversales:

CB5: Que los estudiantes hayan desarrollado aquellas habilidades de aprendizaje necesarias para emprender estudios posteriores con un alto grado de autonomía

CG0: Capacidad de análisis y síntesis: Encontrar, analizar, criticar (razonamiento crítico), relacionar, estructurar y sintetizar información proveniente de diversas fuentes, así como integrar ideas y conocimientos.

CG02: Capacidad de comunicación oral y escrita en el ámbito académico y profesional con especial énfasis, en la redacción de documentación técnica.

CG03: Capacidad para la resolución de problemas.

CG04: Capacidad para tomar decisiones basadas en criterios objetivos (datos experimentales, científicos o de simulación disponibles) así como capacidad de argumentar y justificar lógicamente dichas decisiones, sabiendo aceptar otros puntos de vista.

CG05: Capacidad de trabajo en equipo.

CG07: Motivación por la calidad y la mejora continua, actuando con rigor, responsabilidad y ética profesional.

CG08: Capacidad para adaptarse a las tecnologías y a los futuros entornos actualizando las competencias profesionales.

CG09: Capacidad para innovar y generar nuevas ideas.

CG11: Respeto a los derechos fundamentales y de igualdad entre hombres y mujeres

CG06: Capacidad para el aprendizaje autónomo, así como iniciativa y espíritu emprendedor

CT6: Promover, respetar y velar por los derechos humanos, la igualdad sin discriminación por razón de nacimiento, raza, sexo, religión, opinión u otra circunstancia personal o social, los valores democráticos, la igualdad social y el sostenimiento medioambiental.

CT4: Capacidad de utilizar las Competencias Informáticas e Informacionales (CI2) en la práctica profesional.

CT3: Desarrollo de una actitud de indagación que permita la revisión y avance permanente del conocimiento.

5. Actividades Formativas y Metodologías Docentes

5.1 Actividades formativas:

- Sesiones de Teoría sobre los contenidos del Programa
- Sesiones Prácticas en Laboratorios Especializados o en Aulas de Informática
- Actividades Académicamente Dirigidas por el Profesorado: seminarios, conferencias, desarrollo de trabajos, debates, tutorías colectivas, actividades de evaluación y autoevaluación...

5.2 Metodologías Docentes:

- Desarrollo de Prácticas en Laboratorios Especializados o Aulas de Informática en grupos reducidos
- Clase Magistral Participativa
- Planteamiento, Realización, Tutorización y Presentación de Trabajos.
- Evaluaciones y Exámenes

5.3 Desarrollo y Justificación:

El desarrollo de la asignatura se basará en clases teóricas, prácticas al uso, y actividades

academicamente dirigidas.

Se planterá el ciclo de vida de un proyecto, para su análisis, implementación y discusión durante el desarrollo de la

asignatura.

En las clases prácticas se utilizará el entorno Weka para el desarrollo de los distintos trabajos. Así mismo, se presentarán

entornos para la introducción de las arquitecturas y software de Minería de Datos en el contexto de Big Data: HADOOP y

SPARK.

Al alumno a principio de curso se le entrega un material didáctico en el que se contempla cuatro apartados:

1.- Libros (Español e Ingles) de carácter introductorio de la materia.

2.- Transparencias de todos los capítulos y temas detallados en los contenidos de la materia.

3.- Herramienta de software (WEKA), para la realización de las prácticas y validaciones empíricas, tanto en clase como en

casa.

4.- Trabajos de Fin de Grados y TFM,s, realizados en distintas universidades españolas, con la idea para ejemplificar

conceptos, aplicaciones... así como inspiración y soporte para los trabajos academicamente dirigidos.

6. Temario Desarrollado

1.INTRODUCCIÓN

1.1.El concepto de Minería de Datos

1.2.La minería de datos y el proceso de descubrimiento de conocimiento a partir de datos

1.3.Relación con otras disciplinas

1.4.Aplicaciones

1.5.Fases del proceso de extracción de conocimiento a partir de datos

2.PREPARACIÓN DE DATOS

2.1.Consideraciones previas generales. Los almacenes de datos.

2.2.Técnicas sencillas de preprocesado

2.2.1.Compleción (datos faltantes)

2.2.2.Limpieza de errores

2.2.3.Transformación de atributos

2.2.4.Escalado

2.2.5.Discretización

2.2.6.Numerización

2.3.Técnicas de reducción de la dimensionalidad I: Análisis de Componentes Principales.

2.4.Técnicas de reducción de la dimensionalidad II: Métodos de Filtrado y Envoltura

3.TAREAS Y TÉCNICAS DE MINERÍA DE DATOS

3.1.Tareas en minería de datos.

3.2.Correspondencia entre métodos y tareas.

3.3.Caracterización de las técnicas de minería de datos.

3.4.Técnicas de Minería de Datos

3.4.1.Métodos estadísticos.

3.4.2.Reglas de asociación y dependencia.

3.4.3.Métodos Bayesianos.

3.4.4.Árboles de Decisión y sistemas de reglas.

3.4.5.Redes Neuronales Artificiales.

3.4.6.Máquinas de vectores soporte.

3.4.7.Extracción de conocimiento con algoritmos evolutivos y reglas difusas.

3.4.8.Métodos basados en casos y vecindad.

4.EVALUACIÓN

4.1.Consideraciones generales.

4.2.Técnicas básicas de evaluación de clasificadores

4.2.1.Medidas de la calidad de un clasificador: la tasa de errores

4.2.2.La descomposición del error en sesgo y varianza: el concepto de generalización

4.2.3.El sobreentrenamiento

4.2.4.Repetibilidad estadística: la validación cruzada.

4.3.Aspectos específicos de la evaluación de los diferentes clasificadores estudiados

4.4.Técnicas estadísticas de comparación de clasificadores

4.5.Medidas de calidad de agrupamiento

4.6.Interpretación, difusión y uso de modelos

5.IMPLANTACIÓN E IMPACTO DE LA MINERÍA DE DATOS

5.1.Implantación de un Programa de Minería de Datos (PMD) en una organización

5.1.1.Cuándo implantar un PMD: Necesidades y objetivos

5.1.2.Fases de un PMD: Estándar CRISP-DM

5.1.3.Integración de un PMD dentro de una organización

5.1.4.Recursos necesarios

5.2.Repercusiones y retos de la minería de datos

5.2.1.Impacto social

5.2.2.Cuestiones éticas y legales

5.2.3.Problemas y soluciones: Tendencias futuras

6.0 Nuevos enfoques. Big Data

7. Bibliografía

7.1 Bibliografía básica:

Bibliografía básica:

Manual básico: Introducción a la Minería de Datos de Hernández Orallo et al. Ed. Pearson, 2004.

Data Mining, Concepts and techniques. Jiawei Han et al. Morgan Kaufmann, 2012.

7.2 Bibliografía complementaria:

Bibliografía complementaria:

Data Mining: Practical Machine Learning Tools and Techniques (The Morgan Kaufmann Series in Data Management

Systems). Ian H. Witten y Eibe Frank, 2011.

Data Mining: A Heuristic Approach. Hussein A. Abbass, Ruhul A. Sarker and Charles S. Newton.
University of New South
Wales, Australia (2002).

8. Sistemas y criterios de evaluación

8.1 Sistemas de evaluación:

- Examen de Teoría/Problemas
- Defensa de Prácticas
- Defensa de Trabajos e Informes Escritos

8.2 Criterios de evaluación relativos a cada convocatoria:

8.2.1 Convocatoria I:

Calificación ponderada (para todas las convocatorias) con los porcentajes: examen teórico-práctico (50%), prácticas de laboratorio(40%), defensa de trabajos (10%). Es necesario aprobar con un cinco cada apartado evaluable, para el cálculo de la calificación final.

El examen teórico constará de tres partes. La primera (correspondiente a 1/3 de la nota) versará sobre preguntas teóricas, sin cortapisas en la longitud de las respuestas. Otro tercio del examen (1/3 de la nota), versará sobre problemas/cuestiones de carácter práctico en donde se deberá utilizar algún tipo de cálculo matemático: Álgebra Lineal, Probabilidad... El último tercio consistirá en cuestiones de respuesta corta. La duración del examen será de dos horas y media. Con este examen se da cumplimiento a los objetivos CG0y CG4. La evaluación de las prácticas va dirigida a dar cumplimiento del objetivo CG8. Los trabajos académicamente dirigidos se evaluarán en las dos últimas sesiones de prácticas. La motivación, justificación y defensa del trabajo dirigido, será tanto oral cómo por escrito. Con este trabajo se tratará de consolidar el objetivo CG9.

La duración del examen será dos horas y media. No se necesitará ningún material extra para la realización del examen. La calificación global será la suma ponderada de los tres apartados, con los índices de ponderación aplicados en la evaluación continua. Es necesario aprobar los tres apartados, para superar la asignatura. Estos criterios de evaluación única, serán aplicables en su caso, a las convocatoria III y extraordinaria.

La matrícula de honor será para el alumno que obtenga un sobresaliente (10) en su calificación global. En caso de empate, se optará por el alumno con el mejor trabajo académicamente dirigido.

Es necesario aprobar los tres apartados, para superar la asignatura.

8.2.2 Convocatoria II:

En esta convocatoria y, en el caso de no haber superado el trabajo o los ejercicios de prácticas, estos se presentarán y defenderán por escrito. Los demás criterios serán los especificados en la convocatoria de Febrero/Junio.

8.2.3 Convocatoria III:

En esta convocatoria y, en el caso de no haber superado el trabajo o los ejercicios de prácticas, estos se presentarán y defenderán por escrito. Los demás criterios serán los especificados en la

convocatoria de Febrero/Junio.

8.2.4 Convocatoria extraordinaria:

En esta convocatoria y, en el caso de no haber superado el trabajo o los ejercicios de prácticas, estos se presentarán y defenderán por escrito. Los demás criterios serán los especificados en la convocatoria de Febrero/Junio.

8.3 Evaluación única final:

8.3.1 Convocatoria I:

En el caso de evaluación única, el alumno además del examen teórico/práctico, anteriormente descrito, deberá haber

realizado las prácticas de laboratorio y presentar en el acto académico del examen, el trabajo academicamente dirigido.

La motivación, justificación y defensa del trabajo dirigido, será por escrito.

La duración del examen será de dos horas y media.

No se necesitará ningún material extra para la realización del examen. La calificación global será la suma de los tres apartados, con los índices de ponderación aplicados a la evaluación continua. Es necesario aprobar los tres apartados, para superar la asignatura.

Calificación será ponderada con los porcentajes: examen teórico-práctico (50%), prácticas de laboratorio(40%), defensa de trabajos (10%). Es necesario aprobar con un cinco cada apartado evaluable, para el cálculo de la calificación final.

Es necesario aprobar los tres apartados, para superar la asignatura.

8.3.2 Convocatoria II:

En este caso de evaluación única, el alumno además del examen teórico/práctico, anteriormente descrito y, en el caso de no haber superado el trabajo o los ejercicios de prácticas, estos se presentarán y defenderán por escrito. La duración del examen será de dos horas y media.

No se necesitará ningún material extra para la realización del examen. La calificación global será la suma de los tres apartados, con los índices de ponderación aplicados a la evaluación continua. Es necesario aprobar los tres apartados, para superar la asignatura.

Calificación será ponderada con los porcentajes: examen teórico-práctico (50%), prácticas de laboratorio(40%), defensa de trabajos (10%). Es necesario aprobar con un cinco cada apartado evaluable, para el cálculo de la calificación final.

Es necesario aprobar los tres apartados, para superar la asignatura.

8.3.3 Convocatoria III:

En este caso de evaluación única, el alumno además del examen teórico/práctico, anteriormente descrito y, en el caso de no haber superado el trabajo o los ejercicios de prácticas, estos se presentarán y defenderán por escrito. La duración del examen será de dos horas y media.

No se necesitará ningún material extra para la realización del examen. La calificación global será la suma de los tres apartados, con los índices de ponderación aplicados a la evaluación continua. Es necesario aprobar los tres apartados, para superar la asignatura.

Calificación será ponderada con los porcentajes: examen teórico-práctico (50%), prácticas de laboratorio(40%), defensa de trabajos (10%). Es necesario aprobar con un cinco cada apartado evaluable, para el cálculo de la calificación final.

Es necesario aprobar los tres apartados, para superar la asignatura.

8.3.4 Convocatoria Extraordinaria:

En este caso de evaluación única, el alumno además del examen teórico/práctico, anteriormente descrito y, en el caso de no haber superado el trabajo o los ejercicios de prácticas, estos se presentarán y defenderán por escrito. La duración del examen será de dos horas y media.

No se necesitará ningún material extra para la realización del examen. La calificación global será la suma de los tres apartados, con los índices de ponderación aplicados a la evaluación continua. Es necesario aprobar los tres apartados, para superar la asignatura.

Calificación será ponderada con los porcentajes: examen teórico-práctico (50%), prácticas de laboratorio(40%), defensa de trabajos (10%). Es necesario aprobar con un cinco cada apartado evaluable, para el cálculo de la calificación final.

Es necesario aprobar los tres apartados, para superar la asignatura.

9. Organización docente semanal orientativa:							
Fecha	Grupos Grandes	G. Reducidos				Pruebas y/o act. evaluables	Contenido desarrollado
		Aul. Est.	Lab.	P. Camp	Aul. Inf.		
19-09-2022	1	0	0	0	0.5		TEMA 1
26-09-2022	3	0	0	0	1.5		TEMA 2
03-10-2022	3	0	0	0	1.5		TEMA 2
10-10-2022	3	0	0	0	1.5		TEMA 2
17-10-2022	3	0	0	0	1.5		TEMA 3
24-10-2022	3	0	0	0	1.5		TEMA 3
31-10-2022	3	0	0	0	1.5		TEMA 3
07-11-2022	3	0	0	0	1.5		TEMA 4
14-11-2022	3	0	0	0	1.5		TEMA 4
21-11-2022	3	0	0	0	1.5		TEMA 4
28-11-2022	3	0	0	0	1.5		TEMA 5
05-12-2022	3	0	0	0	1.5		TEMA 5
12-12-2022	3	0	0	0	1.5		TEMA 5
19-12-2022	3	0	0	0	1.5		TEMA 6
09-01-2023	0	0	0	0	0		
TOTAL	40	0	0	0	20		