

1. DATOS GENERALES

 Asignatura: **MINERÍA DE DATOS**

 Código: **42348**

 Tipología: **OBLIGATORIA**

 Créditos ECTS: **6**

 Grado: **346 - GRADO EN INGENIERÍA INFORMÁTICA (AB)**

 Curso académico: **2018-19**

 Centro: **604 - ESCUELA SUPERIOR DE INGENIERÍA INFORMÁTICA (AB)**

 Grupo(s): **15**

 Curso: **4**

 Duración: **Primer cuatrimestre**

 Lengua principal de impartición: **Español**

 Segunda lengua: **Inglés**

 Uso docente de otras lenguas: **Material bibliográfico en inglés**

 English Friendly: **N**

Página web:

 Bilingüe: **N**

 Profesor: **JOSE ANTONIO GAMEZ MARTIN** - Grupo(s): **15**

Edificio/Despacho	Departamento	Teléfono	Correo electrónico	Horario de tutoría
ESII/1.C.13	SISTEMAS INFORMÁTICOS	2473	jose.gamez@uclm.es	Consultar http://esiab.uclm.es/pers.php?codpers=jgamez

2. REQUISITOS PREVIOS

Para cursar esta materia es aconsejable haber cursado los módulos de Formación Básica (Módulo I) y el módulo Común a la Rama de Informática (Módulo II).

Esta asignatura se apoya en las competencias y los conocimientos adquiridos en las asignaturas:

- Lógica
- Estadística
- Diseño de Algoritmos
- Sistemas inteligentes
- Sistemas Basados en el Conocimiento

3. JUSTIFICACIÓN EN EL PLAN DE ESTUDIOS, RELACIÓN CON OTRAS ASIGNATURAS Y CON LA PROFESIÓN

La asignatura se ubica en la intensificación de Computación, en el ámbito de las asignaturas de la Inteligencia Artificial. La Minería de datos, y el aprendizaje automático en general, están ligados al campo de la estadística y de la algorítmica, y abordan las técnicas para la extracción de conocimiento implícito en conjuntos de datos. En los últimos años, estas disciplinas están ganando importancia debido al incremento en la producción de datos -propiciado por fenómenos como por ejemplo el auge de internet o las redes sociales - o el desarrollo de nuevas técnicas para la obtención de información genética. Desde el punto de vista profesional, cada vez existe una mayor demanda de analistas de datos en ámbitos tan diversos como el marketing, el análisis de mercados, la seguridad, o la biología.

4. COMPETENCIAS DE LA TITULACIÓN QUE LA ASIGNATURA CONTRIBUYE A ALCANZAR

Competencias propias de la asignatura	
Código	Descripción
CM5	Capacidad para adquirir, obtener, formalizar y representar el conocimiento humano en una forma computable para la resolución de problemas mediante un sistema informático en cualquier ámbito de aplicación, particularmente los relacionados con aspectos de computación, percepción y actuación en entornos inteligentes.
CM7	Capacidad para conocer y desarrollar técnicas de aprendizaje computacional y diseñar e implementar aplicaciones y sistemas que las utilicen, incluyendo las dedicadas a extracción automática de información y conocimiento a partir de grandes volúmenes de datos.
INS5	Capacidad para argumentar y justificar lógicamente las decisiones tomadas y las opiniones.
UCLM3	Correcta comunicación oral y escrita.

5. OBJETIVOS O RESULTADOS DE APRENDIZAJE ESPERADOS

Resultados de aprendizaje propios de la asignatura	
Descripción	
Conocer y desarrollar técnicas de aprendizaje computacional, tanto supervisado como no supervisado, y diseñar e implementar aplicaciones y sistemas que las utilicen.	
Desarrollar e implementar un sistema de recuperación de la información de tamaño pequeño a mediano.	
Describir y aplicar las distintas fases del proceso de descubrimiento de extracción del conocimiento a partir de grandes volúmenes de datos.	
Resultados adicionales	
Descripción	
Obtener resultados concluyentes fruto del proceso de extracción del conocimiento y ser capaz de presentarlos y justificarlos	

6. TEMARIO

- **Tema 1:** Introducción a la minería de datos
- **Tema 2:** Proceso de descubrimiento de conocimiento a partir de datos.
- **Tema 3:** Validación y evaluación de modelos.
- **Tema 4:** Métodos basados en analogía (kNN)
- **Tema 5:** Predicción numérica: regresión.
- **Tema 6:** Árboles: clasificación y regresión.
- **Tema 7:** Clasificadores probabilísticos
- **Tema 8:** Reducción de la dimensionalidad.
- **Tema 9:** Multiclasificadores (ensembles)
- **Tema 10:** Redes neuronales.
- **Tema 11:** Clustering: Técnicas de agrupamiento/segmentación.
- **Tema 12:** Reglas de asociación.
- **Tema 13:** Detección de anomalías.

COMENTARIOS ADICIONALES SOBRE EL TEMARIO

7. ACTIVIDADES O BLOQUES DE ACTIVIDAD Y METODOLOGÍA

Actividad formativa	Metodología	Competencias relacionadas	ECTS	Horas	Ev	Ob	Rec	Descripción
Enseñanza presencial (Teoría) [PRESENCIAL]	Método expositivo/Lección magistral	CM7 CM5	1.26	31.5	N	N	N	Se usará para introducir los principales conceptos de cada tema.
Prueba final [PRESENCIAL]	Pruebas de evaluación	INS5 CM7 UCLM3 CM5	0.1	2.5	S	S	N	Corresponde al examen oficial de la asignatura. De carácter individual.
Talleres o seminarios [PRESENCIAL]	Método expositivo/Lección magistral	CM7	0.08	2	N	N	N	Se realizará un seminario en laboratorio de las herramientas/APIs usadas para la implementación de las prácticas e ilustración de los casos de estudio (Python, Weka, etc). Corresponde a la práctica 0.
Enseñanza presencial (Prácticas) [PRESENCIAL]	Método expositivo/Lección magistral	CM7 CM5	0.06	1.5	N	N	N	Se usará la primera media hora de la primera sesión correspondiente a cada una de las tres prácticas obligatorias para describir el trabajo a realizar, el material disponible y el entregable a presentar.
Prácticas de laboratorio [PRESENCIAL]	Trabajo dirigido o tutorizado	CM7 CM5	0.66	16.5	N	N	N	Trabajo del alumno con supervisión del profesor en las prácticas de la asignatura.
Resolución de problemas o casos [PRESENCIAL]	Resolución de ejercicios y problemas	CM7 CM5	0.24	6	N	N	N	Resolución en clase de problemas y casos de estudio relacionados con los distintos temas estudiados.

Estudio o preparación de pruebas [AUTÓNOMA]	Trabajo autónomo	CM7 CM5	1.56	39	N	N	N	Se estima que el alumno dedicará aproximadamente 1.3 horas de estudio autónomo por cada hora de lección magistral recibida. Esto incluye la búsqueda y lectura de material adicional.
Otra actividad no presencial [AUTÓNOMA]	Prácticas	CM7 CM5	0.84	21	N	N	N	Se estima que el alumno dedicará 7 horas adicionales a las presenciales para completar las prácticas/proyectos de programación.
Elaboración de memorias de Prácticas [AUTÓNOMA]	Trabajo autónomo	INS5 CM7 UCLM3 CM5	0.36	9	S	S	S	Los alumnos han de plasmar en una memoria el trabajo realizado en la práctica, presentar de forma coherente los resultados y obtener las conclusiones de acuerdo a los objetivos inicialmente fijados. Se estiman unas tres horas por práctica. En caso de no superar la práctica, esta podrá volver a ser presentada coincidiendo (aproximadamente) con el examen final de la asignatura. Las prácticas se diseñan para ser realizadas y defendidas por parejas. No obstante en situaciones particulares y pactadas de antemano con el profesor, podrán realizarse de forma individual.
Elaboración de informes o trabajos [AUTÓNOMA]	Aprendizaje basado en problemas (ABP)	INS5 CM7 UCLM3 CM5	0.76	19	S	N	N	Se facilitarán 10 casos de estudio o ejercicios, uno por tema seleccionado que el alumno debe completar y que serán evaluados mediante la entrega de las mismas (tarea) o un cuestionario asociado a ellas. Este actividad es de carácter individual.
Pruebas on-line [AUTÓNOMA]	Resolución de ejercicios y problemas	INS5 CM7 UCLM3 CM5	0.08	2	S	N	N	El alumno dedicará unos 12 minutos a completar en CampusVirtual cada una de las tareas o cuestionarios asociados a los casos de estudio/ejercicios, en las que previamente habrá trabajado.
Total:			6	150				
Créditos totales de trabajo presencial: 2.4					Horas totales de trabajo presencial: 60			
Créditos totales de trabajo autónomo: 3.6					Horas totales de trabajo autónomo: 90			

Ev: Actividad formativa evaluable
Ob: Actividad formativa de superación obligatoria
Rec: Actividad formativa recuperable

8. CRITERIOS DE EVALUACIÓN Y VALORACIONES

Sistema de evaluación	Valoraciones		Descripción
	Estudiante presencial	Estud. semipres.	
Prueba final	40.00%	0.00%	(ESC) En la convocatoria ordinaria esta prueba tendrá un valor del 40% de la nota final de la asignatura. Entrega individual.
Resolución de problemas o casos	15.00%	0.00%	(ESC) Este 15% de la nota se dividirá entre los 10 casos de estudio/ejercicios proporcionados, en función de la complejidad de las mismas y de la corrección de los ejercicios entregados. Entrega individual.
Elaboración de memorias de prácticas	15.00%	0.00%	(INF) Es obligatoria la entrega de las tres memorias de prácticas. El alumno puede obtener hasta un 5% del total por memoria en función de su corrección. Entrega por grupo de prácticas.
Realización de prácticas en laboratorio	15.00%	0.00%	(LAB) Se supervisará la eficiencia y eficacia del código generado para las tres prácticas planteadas. Entrega por grupo de prácticas.
Otro sistema de evaluación	15.00%	0.00%	(PRES) Cada grupo de prácticas deberá presentar su práctica al profesor y responder a las preguntas realizadas. Entrevista por grupo de prácticas, con preguntas individuales.
Total:	100.00%	0.00%	

Criterios de evaluación de la convocatoria ordinaria:

- Superar la parte de teoría con nota ≥ 4.5 .
- Superar cada práctica individualmente con nota ≥ 5 .
- La entrega de las hojas de problemas/cuestionarios no es obligatoria, aunque si muy recomendable, ya que además del valor formativo que tienen, en caso contrario la nota máxima sería un 8.5.

- Si teoría y prácticas (todas) se han superado, entonces la nota será $0.4 \cdot \text{teoría} + 0.45 \cdot \text{prácticas} + 0.15 \cdot \text{problemas}$.
- Si no se superan ambas partes la nota será No Presentado si no se ha realizado el examen de teoría y mínimo(4.0, nota-teoría) si el examen si ha sido realizado.

- Se guarda la teoría para el extraordinario.
- Se guardan las prácticas (individualmente) para el extraordinario.

Originalidad: La entrega de cualquier ejercicio (examen, memoria de prácticas, código, problemas, etc.) supone implícitamente la declaración de originalidad por parte de los autores, por lo que en caso de detección de plagio, copia, etc. se iniciarán las medidas disciplinarias oportunas.

Particularidades de la convocatoria extraordinaria:

- No se consideran los problemas/cuestionarios de la convocatoria ordinaria.
- La nota se obtiene como la media entre teoría y prácticas una vez superadas ambas partes por separado. En caso contrario la nota será No Presentado si no se ha realizado el examen de teoría y mínimo(4.0, nota-teoría) si el examen si ha sido realizado.

Particularidades de la convocatoria especial de finalización:

- No se consideran los problemas/cuestionarios de la convocatoria ordinaria.
- La nota se obtiene como la media entre teoría y prácticas una vez superadas ambas partes por separado. En caso contrario la nota será No Presentado si no se ha realizado el examen de teoría y mínimo(4.0, nota-teoría) si el examen si ha sido realizado.

9. SECUENCIA DE TRABAJO, CALENDARIO, HITOS IMPORTANTES E INVERSIÓN TEMPORAL

No asignables a temas	
Horas	Suma horas
Prueba final [PRESENCIAL][Pruebas de evaluación]	2.5
Talleres o seminarios [PRESENCIAL][Método expositivo/Lección magistral]	1.5
Enseñanza presencial (Prácticas) [PRESENCIAL][Método expositivo/Lección magistral]	1.5
Prácticas de laboratorio [PRESENCIAL][Trabajo dirigido o tutorizado]	16.5
Resolución de problemas o casos [PRESENCIAL][Resolución de ejercicios y problemas]	7.5
Otra actividad no presencial [AUTÓNOMA][Prácticas]	21
Elaboración de memorias de Prácticas [AUTÓNOMA][Trabajo autónomo]	9
Elaboración de informes o trabajos [AUTÓNOMA][Aprendizaje basado en problemas (ABP)]	19
Pruebas on-line [AUTÓNOMA][Resolución de ejercicios y problemas]	1

Tema 1 (de 13): Introducción a la minería de datos	
Actividades formativas	Horas
Enseñanza presencial (Teoría) [PRESENCIAL][Método expositivo/Lección magistral]	1.5
Estudio o preparación de pruebas [AUTÓNOMA][Trabajo autónomo]	1.3
Comentario: Se introducirá en que consiste al minería de datos y sus principales aplicaciones.	

Tema 2 (de 13): Proceso de descubrimiento de conocimiento a partir de datos.	
Actividades formativas	Horas
Enseñanza presencial (Teoría) [PRESENCIAL][Método expositivo/Lección magistral]	1.5
Estudio o preparación de pruebas [AUTÓNOMA][Trabajo autónomo]	2.6
Comentario: Se describirá el proceso completo de extracción de conocimiento a partir de bases de datos. Se comentará la metodología CRISP-DM, en particular algunas tareas de preprocesamiento como la discretización o la imputación de valores perdidos.	

Tema 3 (de 13): Validación y evaluación de modelos.	
Actividades formativas	Horas
Enseñanza presencial (Teoría) [PRESENCIAL][Método expositivo/Lección magistral]	3
Estudio o preparación de pruebas [AUTÓNOMA][Trabajo autónomo]	2.6
Comentario: Se introducirán las principales técnicas de validación y medidas de error. Se usarán clasificadores y estudiados para ilustrar los conceptos. Se introducirán problemas particulares para ilustrar la necesidad de otras medidas de	

evaluación, p.e. problemas con desequilibrio entre clases para introducir el análisis ROC.

Tema 4 (de 13): Métodos basados en analogía (kNN)

Actividades formativas	Horas
Enseñanza presencial (Teoría) [PRESENCIAL][Método expositivo/Lección magistral]	3
Estudio o preparación de pruebas [AUTÓNOMA][Trabajo autónomo]	3.9
Comentario: En este tema se estudiará el método de los vecinos más cercanos. Se verán las aplicaciones a clasificación y regresión.	

Tema 5 (de 13): Predicción numérica: regresión.

Actividades formativas	Horas
Enseñanza presencial (Teoría) [PRESENCIAL][Método expositivo/Lección magistral]	2
Estudio o preparación de pruebas [AUTÓNOMA][Trabajo autónomo]	3.9
Comentario: Introducción a la predicción numérica. Técnicas basadas en análisis de regresión. Método de optimización basado en gradiente.	

Tema 6 (de 13): Árboles: clasificación y regresión.

Actividades formativas	Horas
Enseñanza presencial (Teoría) [PRESENCIAL][Método expositivo/Lección magistral]	2
Estudio o preparación de pruebas [AUTÓNOMA][Trabajo autónomo]	2.6
Comentario: Se repasarán los árboles de clasificación (id3-c4.5) y se estudiará su adaptación a los árboles de regresión y de modelos de regresión.	

Tema 7 (de 13): Clasificadores probabilísticos

Actividades formativas	Horas
Enseñanza presencial (Teoría) [PRESENCIAL][Método expositivo/Lección magistral]	3
Estudio o preparación de pruebas [AUTÓNOMA][Trabajo autónomo]	3.9
Comentario: En este tema se revisará el modelo de clasificación probabilística Naive Bayes y se estudiará la regresión logística y modelos de tipo semi-naive Bayes como TAN o KDB.	

Tema 8 (de 13): Reducción de la dimensionalidad.

Actividades formativas	Horas
Enseñanza presencial (Teoría) [PRESENCIAL][Método expositivo/Lección magistral]	2
Estudio o preparación de pruebas [AUTÓNOMA][Trabajo autónomo]	2.6
Comentario: Se estudiarán técnicas de reducción de la dimensionalidad (variables) mediante la transformación y selección de variables. Estas técnicas son genéricas pero de especial interés para los métodos probabilísticos (naive Bayes, regresión logística, TAN, AODE, etc...).	

Tema 9 (de 13): Multiclasificadores (ensembles)

Actividades formativas	Horas
Enseñanza presencial (Teoría) [PRESENCIAL][Método expositivo/Lección magistral]	3
Estudio o preparación de pruebas [AUTÓNOMA][Trabajo autónomo]	3.9
Comentario: En este tema se verá como el uso de múltiples modelos y su posterior combinación/agregación suele mejorar la eficacia/acierto en el aprendizaje supervisado.	

Tema 10 (de 13): Redes neuronales.

Actividades formativas	Horas
Enseñanza presencial (Teoría) [PRESENCIAL][Método expositivo/Lección magistral]	3.5
Estudio o preparación de pruebas [AUTÓNOMA][Trabajo autónomo]	2.6
Comentario: Explicaremos un modelo de carácter bio-inspirado con aplicación a clasificación, regresión y con ciertas modificaciones a clustering.	

Tema 11 (de 13): Clustering: Técnicas de agrupamiento/segmentación.

Actividades formativas	Horas
Enseñanza presencial (Teoría) [PRESENCIAL][Método expositivo/Lección magistral]	3
Estudio o preparación de pruebas [AUTÓNOMA][Trabajo autónomo]	2.6
Comentario: En este tema se estudian los paradigmas de clustering y segmentación, sin duda los de mayor uso en aprendizaje no supervisado. Se trata de extraer patrones entre los registros o instancias del conjunto de datos.	

Tema 12 (de 13): Reglas de asociación.

Actividades formativas	Horas
Enseñanza presencial (Teoría) [PRESENCIAL][Método expositivo/Lección magistral]	2
Estudio o preparación de pruebas [AUTÓNOMA][Trabajo autónomo]	3.9
Comentario: En este tema se estudia uno de los paradigmas considerados en principio como de minería de datos: el descubrimiento de reglas de asociación. Ahora se trata de aprender relaciones (reglas) entre las variables a partir de amplias colecciones de transacciones.	

Tema 13 (de 13): Detección de anomalías.

Actividades formativas	Horas
Enseñanza presencial (Teoría) [PRESENCIAL][Método expositivo/Lección magistral]	2
Estudio o preparación de pruebas [AUTÓNOMA][Trabajo autónomo]	2.6
Comentario: En este tema se aborda como detectar anomalías o patrones extraños mediante el aprendizaje de modelos que representan la mayoría de los patrones, y clasificando como anomalías aquellos datos que pertenecen con escasa probabilidad al modelo aprendido.	


Actividad global

Actividades formativas	Suma horas
Enseñanza presencial (Teoría) [PRESENCIAL][Método expositivo/Lección magistral]	31.5
Prueba final [PRESENCIAL][Pruebas de evaluación]	2.5
Talleres o seminarios [PRESENCIAL][Método expositivo/Lección magistral]	1.5
Enseñanza presencial (Prácticas) [PRESENCIAL][Método expositivo/Lección magistral]	1.5
Prácticas de laboratorio [PRESENCIAL][Trabajo dirigido o tutorizado]	16.5
Resolución de problemas o casos [PRESENCIAL][Resolución de ejercicios y problemas]	7.5
Estudio o preparación de pruebas [AUTÓNOMA][Trabajo autónomo]	39
Otra actividad no presencial [AUTÓNOMA][Prácticas]	21
Elaboración de memorias de Prácticas [AUTÓNOMA][Trabajo autónomo]	9
Elaboración de informes o trabajos [AUTÓNOMA][Aprendizaje basado en problemas (ABP)]	19
Pruebas on-line [AUTÓNOMA][Resolución de ejercicios y problemas]	1
Total horas:	150

Comentarios generales sobre la planificación:

Esta planificación es ORIENTATIVA, pudiendo variar a lo largo del periodo lectivo en función de las necesidades docentes, festividades, o por cualquier otra causa imprevista. La planificación semanal de la asignatura podrá encontrarse de forma detallada y actualizada en la plataforma Campus Virtual (Moodle). Las actividades presenciales se organizan en tres clases de 1.5 horas por semana. Las clases concretas a usar para cubrir los 6 créditos (60 horas presenciales) se anunciarán oportunamente en CampusVirtual. Aunque la asignatura tiene un horario de mañana, las actividades de evaluación o recuperación de clases (teóricas o prácticas) podrían planificarse, excepcionalmente, en horario de tarde (de 16h a 20:30h).

10. BIBLIOGRAFÍA, RECURSOS

Autor/es	Título	Libro/Revista Población Editorial	ISBN	Año	Descripción	Enlace Catálogo Web biblioteca
Basilio Sierra	Manuales de Python. APRENDIZAJE AUTOMÁTICO: CONCEPTOS BÁSICOS Y AVANZADOS: ASPECTOS PRÁCTICOS UTILIZANDO EL SOFTWARE WEKA	Prentice-Hall	9788483223185	2006		
García, Salvador, Luengo, Julián, Herrera, Francisco	Data Preprocessing in Data Mining	Springer	978-3-319-10246-7	2015		
José Hernández Orallo, M. José Ramírez Quintana, César Ferrí Ramírez	INTRODUCCIÓN A LA MINERÍA DE DATOS	Pearson	84 205 4091 9	2004		
Pang-Ning Tan, Michael Steinbach, and Vipin Kumar	Introduction to Data Mining	Addison-Wesley Longman Publishing Co	0321321367	2005		
Witten, Frank & Hall	Data Mining: Practical Machine Learning Tools and Techniques	Morgan & Kaufmann	978-0-12-374856-0	2011		
Xindong Wu, Vipin Kumar	The Top Ten Algorithms in Data Mining	Chapman and Hall/CRC	9781420089646	2009		