

**Proyecto/Guía docente de la asignatura**

Se debe indicar de forma fiel cómo va a ser desarrollada la docencia. Esta guía debe ser elaborada teniendo en cuenta a todos los profesores de la asignatura. Conocidos los espacios y profesorado disponible, se debe buscar la máxima presencialidad posible del estudiante siempre respetando las capacidades de los espacios asignados por el centro y justificando cualquier adaptación que se realice respecto a la memoria de verificación. Si la docencia de alguna asignatura fuese en parte online, deben respetarse los horarios tanto de clase como de tutorías). La planificación académica podrá sufrir modificaciones de acuerdo con la actualización de las condiciones sanitarias.

Asignatura	Minería de Datos		
Materia	Computación		
Módulo	Tecnologías Específicas		
Titulación	Grado en Ingeniería Informática, mención Computación		
Plan	541	Código	46970
Periodo de impartición	1er Cuatrimestre	Tipo/Carácter	OB (Mención CO)
Nivel/Ciclo	Grado	Curso	5º
Créditos ECTS	6		
Lengua en que se imparte	Español		
Profesor/es responsable/s	Carlos J. Alonso González Teodoro Calonge Cano		
Datos de contacto (E-mail, teléfono...)	calonso@infor.uva.es , 983 185602 teodoro@infor.uva.es 983 185603		
Departamento(s)	Informática (ATC, CCIA, LSI)		



1. Situación / Sentido de la Asignatura

1.1 Contextualización

La asignatura de Minería de Datos está programada como asignatura obligatoria en el Semestre 1 del 4º curso de la titulación del Grado en Ingeniería Informática, mención Computación.

La minería de datos es un término relativamente moderno que integra numerosas técnicas de análisis de datos y de extracción de modelos. El objetivo es ser capaces de extraer patrones, de describir tendencias y regularidades, de predecir comportamientos y, en general, de sacar partido a la información computerizada que nos rodea hoy en día, generalmente heterogénea y en grandes cantidades, permite a los individuos y a las organizaciones comprender y modelar de una manera más eficiente y precisa el contexto en el que deben actuar y tomar decisiones.

1.2 Relación con otras materias

Esta asignatura, tal y como está definida, tiene una relación clara con Técnicas de Aprendizaje Automático, y se asumirá que los alumnos han adquirido las competencias de esa asignatura.

De manera indirecta también tiene relación con Estructuras de Datos, Estadística, Bases de Datos y Programación.

1.3 Prerrequisitos

- Conocimientos de las técnicas de Aprendizaje Automático estudiadas en la asignatura de Técnicas de Aprendizaje Automático. En su defecto, conocimiento de los métodos básicos de inducción de clasificadores.
- Programación – Pues el alumno necesitará crear sus propios programas o funciones para manipular los datos, u obtener resultados.
- Estadística – Como herramienta para poder manipular y condensar los rasgos más relevantes de un conjunto de datos de tamaño grande, ya sea por el número de muestras, o por la dimensionalidad de dichas muestras.



2. Competencias

2.1 Generales

Código	Descripción
CT3	Comunicación oral y escrita en la lengua propia
	Resolución de problemas
G10	Toma de decisiones
G16	Capacidad de aplicar los conocimientos en la práctica
G21	Habilidad para trabajar de forma autónoma

2.2 Específicas

Código	Descripción
CO5	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 ambientes o entornos inteligentes.
CO6	Capacidad para desarrollar y evaluar sistemas interactivos y de presentación de información compleja y su aplicación a la resolución de problemas de diseño de interacción persona computadora.
CO7	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.



3. Objetivos

Código	Descripción
CO5.1	Comprender el ámbito de aplicación de las técnicas de minería de datos para la extracción y representación de conocimiento basada en datos.
CO7.1	Saber aplicar técnicas de selección, preparación y visualización de datos a problemas concretos y en dominios de interés práctico.
CO6.1	Emplear adecuadamente técnicas de evaluación y representación de resultados.
CO7.2	Ampliar las técnicas de aprendizaje automático conocidas y saber aplicarlas a problemas de minería de datos.
CO7.3	Usar eficazmente plataformas computacionales para resolver problemas concretos de minería de datos en áreas de interés práctico.





4. Contenidos y/o bloques temáticos

Bloque 1: Introducción a la Minería de Datos

Carga de trabajo en créditos ECTS: 0.2

a. Contextualización y justificación

Introducir el proceso de minería de datos identificando sus etapas principales.

b. Objetivos de aprendizaje

Comprender el ámbito de aplicación de las técnicas de minería de datos para la extracción y representación de conocimiento basada en datos.

c. Contenidos

Tema 1: Introducción. Etapas del proceso de minería de datos.

d. Métodos docentes

Ver apartado 5

e. Plan de trabajo

Presentación en aula y trabajo individual

f. Evaluación

Ver apartado 7.

g. Material docente

Esta sección será utilizada por la Biblioteca para etiquetar la bibliografía recomendada de la asignatura (curso) en la plataforma Leganto, integrada en el catálogo Almena y a la que tendrán acceso todos los profesores y estudiantes. Es fundamental que las referencias suministradas este curso estén actualizadas y sean completas. Los profesores tendrán acceso, en breve, a la plataforma Leganto para actualizar su bibliografía recomendada ("Listas de Lecturas") de forma que en futuras guías solamente tendrán que poner el enlace permanente a Leganto, el cual también se puede poner en el Campus Virtual.

g.1 Bibliografía básica

H. Witten, E. Frank and M. A. Hall, C. J. Pall. Data Mining: Practical Machine Learning Tools and Techniques. Morgan Kaufmann, 4th edition, 2017.

Basilio Sierra Araujo. Aprendizaje Automático: conceptos básicos y avanzados. Pearson Prentice Hall, 2006.

Gregory Piatetsky. CRISP-DM, still the top methodology for analytics, data mining, or data science projects. <http://www.kdnuggets.com/2014/10/crisp-dm-top-methodology-analytics-data-mining-data-science-projects.html>.
. Último acceso Julio 2020



g.2 Bibliografía complementaria

J. Han, M. Kamber and J. Pei. Data Mining: Concepts and Techniques. Morgan Kaufmann, 3rd edition, 2012.

CRISP-DM 1.0. Step-by-step data mining guide. The CRISP-DM consortium, 2000. Retrieved from <https://the-modeling-agency.com/crisp-dm.pdf> . Último acceso Julio 2020

g.3 Otros recursos telemáticos (píldoras de conocimiento, blogs, videos, revistas digitales, cursos masivos (MOOC), ...)

h. Recursos necesarios

Proyector de video en aula, notas de la asignatura y recursos bibliográficos.

Aula virtual Escuela Ingeniería Informática.

i. Temporalización

CARGA ECTS	PERIODO PREVISTO DE DESARROLLO
0.2 ECTS	Semana 1



Bloque 2: Preprocesamiento y Evaluación

Carga de trabajo en créditos ECTS: 2.3

a. Contextualización y justificación

Se presentan aquí las técnicas básicas de las etapas iniciales de la ingeniería de datos: preparación de datos y selección de características.

Conocer los métodos principales de evaluación de clasificadores en varios contextos.

b. Objetivos de aprendizaje

Saber aplicar técnicas de selección, preparación y visualización de datos a problemas concretos y en dominios de interés práctico.

Emplear adecuadamente técnicas de evaluación y representación de resultados.

Usar eficazmente plataformas computacionales para resolver problemas concretos de minería de datos en áreas de interés práctico.

c. Contenidos

Tema 2: Preparación y selección de datos.

Tema 3: Selección de características y reducción de la dimensionalidad.

Tema 3: Evaluación de resultados.

d. Métodos docentes

Ver apartado 5

e. Plan de trabajo

Presentación en aula y trabajo individual.

Realización de prácticas individuales y en grupo

f. Evaluación

Ver apartado 7.

g Material docente

Esta sección será utilizada por la Biblioteca para etiquetar la bibliografía recomendada de la asignatura (curso) en la plataforma Leganto, integrada en el catálogo Almena y a la que tendrán acceso todos los profesores y estudiantes. Es fundamental que las referencias suministradas este curso estén actualizadas y sean completas. Los profesores tendrán acceso, en breve, a la plataforma Leganto para actualizar su bibliografía recomendada ("Listas de Lecturas") de forma que en futuras guías solamente tendrán que poner el enlace permanente a Leganto, el cual también se puede poner en el Campus Virtual.

g.1 Bibliografía básica



H. Witten, E. Frank and M. A. Hall, C. J. Pall. Data Mining: Practical Machine Learning Tools and Techniques. Morgan Kaufmann, 4th edition, 2017.

Tom M. Mitchell. Machine Learning. McGraw-Hill, 1997.

g.2 Bibliografía complementaria

J. Han, M. Kamber and J. Pei. Data Mining: Concepts and Techniques. Morgan Kaufmann, 3rd edition, 2012.

Ethem Alpaydin. Introduction to Machine Learning. Third edition. The MIT Press, 2014

Pang-Ning Tan, Michael Steinbach, and Vipin Kumar. Introduction to Data Mining. Addison Wesley, 2006.

Arlot, S. and Celisse, A. A survey of cross-validation procedures for model selection. Statistics Surveys, 4, 40-79, 2010. DOI:10.1214/09-SS054. Disponible en: http://www.di.ens.fr/willow/pdfs/2010_Arlot_Celisse_SS.pdf

David Page Evaluating Machine Learning Methods. <http://pages.cs.wisc.edu/~dpage/cs760/evaluating.pdf>. Último acceso Marzo 2019.

g.3 Otros recursos telemáticos (píldoras de conocimiento, blogs, videos, revistas digitales, cursos masivos (MOOC), ...)

h. Recursos necesarios

Proyector de video en aula, notas de la asignatura y recursos bibliográficos.

Aula virtual Escuela Ingeniería Informática.

Laboratorio de informática con software libre: Weka y Python fundamentalmente.

i. Temporalización

CARGA ECTS	PERIODO PREVISTO DE DESARROLLO
0.4 ECTS: Preparación y selección de datos.	Semanas 1 y 2
1.2 ECTS: Selección de características y reducción de la dimensionalidad.	Semana 2, 3 y 4
0.7 ECTS: Evaluación de resultados.	Semana 4 y 5.

Bloque 3: Extracción de conocimiento

Carga de trabajo en créditos ECTS: 3.4

a. Contextualización y justificación

Se presentan aquí técnicas avanzadas de minería de datos que no se han estudiado en cursos anteriores.

b. Objetivos de aprendizaje

Ampliar las técnicas de aprendizaje automático conocidas y saber aplicarlas a problemas de minería de datos.
Usar eficazmente plataformas computacionales para resolver problemas concretos de minería de datos en áreas de interés práctico.

c. Contenidos

Tema 4: Multiclasificadores.

Tema 5: Clustering

Tema 6: Redes recurrentes.

d. Métodos docentes

Ver apartado 5.

e. Plan de trabajo

Presentación en aula y trabajo individual.
Realización de prácticas individuales y en grupo

f. Evaluación

Ver apartado 7.

g Material docente

Esta sección será utilizada por la Biblioteca para etiquetar la bibliografía recomendada de la asignatura (curso) en la plataforma Leganto, integrada en el catálogo Almena y a la que tendrán acceso todos los profesores y estudiantes. Es fundamental que las referencias suministradas este curso estén actualizadas y sean completas. Los profesores tendrán acceso, en breve, a la plataforma Leganto para actualizar su bibliografía recomendada ("Listas de Lecturas") de forma que en futuras guías solamente tendrán que poner el enlace permanente a Leganto, el cual también se puede poner en el Campus Virtual.

g.1 Bibliografía básica

H. Witten, E. Frank and M. A. Hall, C. J. Pall. Data Mining: Practical Machine Learning Tools and Techniques. Morgan Kaufmann, 4th edition, 2017.

M. Kantardzic. Data Mining: Concepts, Models, Methods and Algorithms, 2nd Edition. Wiley 2011.

L. Kuncheva, Combining pattern classifiers, Second edition. Wiley, 2014.



g.2 Bibliografía complementaria

Christopher M. Bishop. Pattern Recognition and Machine Learning. Springer, 2006

Tan, Steinbach, Kumar - Introduction to Data Mining, 2004

J. Han, M. Kamber and J. Pei. Data Mining: Concepts and Techniques. Morgan Kaufmann, 3rd edition, 2012.

Ethem Alpaydin. Introduction to Machine Learning. Third edition. The MIT Press, 2014

Pang-Ning Tan, Michael Steinbach, and Vipin Kumar. Introduction to Data Mining. Addison Wesley, 2006.

L. Breiman. Bagging predictors. Machine Learning, 24(2):123–140, 1996.

L. Rokach. Pattern classification using ensemble methods. Word Scientific Publishing, 2010.

D. Wolpert. Stacked generalization. Neural networks, 5:241–260, 1992.

g.3 Otros recursos telemáticos (píldoras de conocimiento, blogs, videos, revistas digitales, cursos masivos (MOOC), ...)

h. Recursos necesarios

Proyector de video en aula, notas de la asignatura y recursos bibliográficos.

Aula virtual Escuela Ingeniería Informática.

Laboratorio de informática con software libre: Weka y Python fundamentalmente.

i. Temporalización

CARGA ECTS	PERIODO PREVISTO DE DESARROLLO
0.7 ECTS: Multiclasificadores	Semanas 6 y 7
1.2 ECTS: Clustering	Semanas 7 a 9
1.6 ECTS: Redes recurrentes	Semanas 10 a 13



5. Métodos docentes y principios metodológicos

Actividad	Metodología
Clase de teoría	<ul style="list-style-type: none">• Clase magistral participativa• Estudio de casos en aula• Resolución de problemas
Clase práctica	<ul style="list-style-type: none">• Clase magistral participativa• Realización de prácticas, cerradas y abiertas.
Seminarios	<ul style="list-style-type: none">• Talleres de aprendizaje
Tutoría	<ul style="list-style-type: none">• Discusión de los contenidos teóricos y de las prácticas



6. Tabla de dedicación del estudiante a la asignatura

ACTIVIDADES PRESENCIALES o PRESENCIALES A DISTANCIA ⁽¹⁾	HORAS	ACTIVIDADES NO PRESENCIALES	HORAS
Clases teórico-prácticas (T/M)	30	Estudio y trabajo autónomo individual	30
Clases prácticas de aula (A)		Estudio y trabajo autónomo grupal	60
Laboratorios (L)	20		
Prácticas externas, clínicas o de campo			
Seminarios (S)	6		
Tutorías grupales (TG)			
Evaluación	4		
Total presencial	60	Total no presencial	90
TOTAL presencial + no presencial			150

(1) Actividad presencial a distancia es cuando un grupo sigue una videoconferencia de forma síncrona a la clase impartida por el profesor para otro grupo presente en el aula.

7. Sistema y características de la evaluación

Criterio: cuando al menos el 50% de los días lectivos del cuatrimestre transcurran en normalidad, se asumirán como criterios de evaluación los indicados en la guía docente. Se recomienda la evaluación continua ya que implica minimizar los cambios en la agenda.

INSTRUMENTO/PROCEDIMIENTO	PESO EN LA NOTA FINAL	OBSERVACIONES
Examen práctico	65%	A realizar en el laboratorio durante la misma jornada que el examen final.
Examen tipo test	35%	El día del examen final

CRITERIOS DE CALIFICACIÓN

- **Convocatoria ordinaria, 1ª convocatoria:**
 - La nota final será el resultado de la suma ponderada del examen práctico y el examen tipo test con los porcentajes indicados en la tabla anterior.
 - No obstante, para superar la asignatura es preciso que en ambas partes (práctico y examen tipo test) se obtenga un **mínimo de 3 puntos sobre 10**.
 - Adicionalmente, la entrega de las prácticas, que es una actividad optativa, se valorará sobre un punto, que se añadirá a la nota final, **siempre que ésta sea igual o superior a 4.5 puntos**.
- **Convocatoria ordinaria, 2ª convocatoria:**
 - Los mismos criterios que en la primera ordinaria.
 - Solo se conserva la entrega de prácticas de la primera convocatoria.
- **Convocatoria extraordinaria:**



- La nota final será el resultado de la suma ponderada del examen práctico y el examen tipo test con los porcentajes indicados en la tabla anterior.
- No obstante, para superar la asignatura es preciso que en ambas partes (práctico y final escrito) se obtenga un **mínimo de 3 puntos sobre 10**.
- Al no tener prácticas asociadas, en esta convocatoria no se considera la entrega de prácticas.

8. Consideraciones finales

