

ÍNDICE DE CONTENIDO

Índice de contenido	VII
Prefacio.....	XIII
Motivación y objetivos del libro.....	XIII
Destinatarios	XV
Organización e itinerarios.....	XVI
Terminología.....	XVIII
Agradecimientos	XVIII

PARTE I: INTRODUCCIÓN

Capítulo 1 ¿Qué es la minería de datos?	3
1.1 Nuevas necesidades.....	3
1.2 El concepto de minería de datos. Ejemplos	5
1.3 Tipos de datos.....	9
1.4 Tipos de modelos	12
1.5 La minería de datos y el proceso de descubrimiento de conocimiento en bases de datos.....	13
1.6 Relación con otras disciplinas.....	14
1.7 Aplicaciones.....	16
1.8 Sistemas y herramientas de minería de datos	18
Capítulo 2 El proceso de extracción de conocimiento	19
2.1 Las fases del proceso de extracción de conocimiento.....	19
2.2 Fase de integración y recopilación.....	21
2.3 Fase de selección, limpieza y transformación	22
2.4 Fase de minería de datos.....	24
2.5 Fase de evaluación e interpretación.....	35
2.6 Fase de difusión, uso y monitorización.....	39

PARTE II: PREPARACIÓN DE DATOS

Capítulo 3 Recopilación. Almacenes de datos	43
3.1 Introducción.....	44
3.2 Necesidad de los almacenes de datos	46
3.3 Arquitectura de los almacenes de datos	49
3.4 Carga y mantenimiento del almacén de datos.....	59
3.5 Almacenes de datos y minería de datos	62
Capítulo 4 Limpieza y transformación.....	65
4.1 Introducción.....	66
4.2 Integración y limpieza de datos	67
4.3 Transformación de atributos. Creación de características.....	78
4.4 Discretización y numerización	89
4.5 Normalización de rango: escalado y centrado.....	93
4.6 Otras transformaciones	94
Capítulo 5 Exploración y selección	97
5.1 Introducción. El contexto de la vista minable	97
5.2 Exploración mediante visualización.....	103
5.3 Sumarización, descripción, generalización y pivotamiento.....	107
5.4 Selección de datos	112
5.5 Lenguajes, primitivas e interfaces de minería de datos.....	125

PARTE III: TÉCNICAS DE MINERÍA DE DATOS

Capítulo 6 El problema de la extracción de patrones	137
6.1 Introducción.....	137
6.2 Tareas y métodos	139
6.3 Minería de datos y aprendizaje inductivo.....	148
6.4 El lenguaje de los patrones. Expresividad.....	154
6.5 Breve comparación de métodos	161
Capítulo 7 Modelización estadística paramétrica.....	165
<i>Tomàs Aluja y Pedro Delicado</i>	
7.1 Concepto de modelización estadística	166
7.2 Modelo de regresión.....	167
7.3 Modelos de regresión sobre componentes incorrelacionados	183
7.4 Modelos de regresión con variables categóricas.....	185
7.5 Análisis de los residuos.....	187
7.6 Ejemplo: aplicación a los datos SERVO	190
7.7 Modelos lineales generalizados	194
7.8 Análisis discriminante.....	203
7.9 Sistemas, aplicabilidad y recomendaciones de uso.....	211

Capítulo 8 Modelización estadística no paramétrica.....	213
<i>Pedro Delicado y Tomàs Aluja</i>	
8.1 Introducción.....	213
8.2 Regresión no paramétrica	215
8.3 Discriminación no paramétrica	229
8.4 Conclusiones, aplicabilidad y sistemas.....	236
Capítulo 9 Reglas de asociación y dependencia	237
9.1 Introducción.....	237
9.2 Reglas de asociación	239
9.3 Reglas de dependencias	243
9.4 Reglas de asociación multinivel	247
9.5 Reglas de asociación secuenciales.....	249
9.6 Aprendizaje de reglas de asociación con sistemas de minería de datos.....	252
Capítulo 10 Métodos bayesianos	257
<i>José A. Gámez Martín, Ismael García Varea y José M. Puerta Callejón</i>	
10.1 Introducción.....	257
10.2 Teorema de Bayes e hipótesis MAP.....	259
10.3 Naïve Bayes.....	260
10.4 Redes bayesianas.....	263
10.5 Aprendizaje de redes bayesianas	266
10.6 Clasificadores basados en redes bayesianas.....	271
10.7 Tratamiento de datos desconocidos	275
10.8 Sistemas	278
Capítulo 11 Árboles de decisión y sistemas de reglas	281
11.1 Introducción.....	281
11.2 Sistemas por partición: árboles de decisión para clasificación.....	283
11.3 Sistemas de aprendizaje de reglas por cobertura.....	287
11.4 Poda y reestructuración.....	290
11.5 Árboles de decisión para regresión, agrupamiento o estimación de probabilidades	293
11.6 Aprendizaje de árboles de decisión híbridos	295
11.7 Adaptación para grandes volúmenes de datos.....	295
11.8 Sistemas, aplicabilidad y recomendaciones de uso	297
Capítulo 12 Métodos relacionales y estructurales.....	301
12.1 Introducción.....	301
12.2 Programación lógica y bases de datos.....	304
12.3 Programación lógica inductiva.....	306
12.4 Programación lógica inductiva y minería de datos	312
12.5 Otros métodos relacionales y estructurales	317
12.6 Sistemas	325

Capítulo 13	Redes neuronales artificiales.....	327
	<i>Emilio Corchado y Colin Fyfe</i>	
13.1	Introducción.....	327
13.2	El aprendizaje en las redes neuronales artificiales	330
13.3	Aprendizaje supervisado en RNA.....	330
13.4	Aprendizaje no supervisado en RNA	343
13.5	Sistemas, aplicabilidad y recomendaciones de uso.....	351
Capítulo 14	Máquinas de vectores soporte.....	353
	<i>Xavier Carreras, Lluís Màrquez y Enrique Romero</i>	
14.1	Introducción.....	353
14.2	Máquinas de vectores soporte para clasificación binaria	356
14.3	Justificación teórica.....	367
14.4	Aplicaciones de las máquinas de vectores soporte	367
14.5	Extensiones y temas avanzados	375
14.6	Paquetes <i>software</i> y recomendaciones de uso.....	378
Anexo.	Optimización con restricciones lineales	381
Capítulo 15	Extracción de conocimiento con algoritmos evolutivos y reglas difusas.....	383
	<i>María José del Jesus, Pedro González y Francisco Herrera</i>	
15.1	Introducción.....	383
15.2	Computación evolutiva.....	385
15.3	Algoritmos evolutivos para la extracción de conocimiento	389
15.4	Lógica difusa.....	403
15.5	Lógica difusa en minería de datos	405
15.6	Sistemas evolutivos difusos en minería de datos	409
15.7	Ejemplos.....	412
15.8	Sistemas <i>software</i>	417
15.9	Conclusiones.....	418
Capítulo 16	Métodos basados en casos y en vecindad	421
	<i>Pedro Isasi</i>	
16.1	Introducción.....	421
16.2	Técnicas para agrupamiento.....	428
16.3	Técnicas para clasificación.....	440
16.4	Métodos de vecindad con técnicas evolutivas	448
16.5	Otros métodos y aplicabilidad	455

PARTE IV: EVALUACIÓN, DIFUSIÓN Y USO DE MODELOS

Capítulo 17	Técnicas de evaluación	461
17.1	Introducción.....	461
17.2	Evaluación de clasificadores.....	462
17.3	Evaluación de modelos de regresión.....	476

17.4 Comparación de técnicas de aprendizaje.....	477
17.5 Evaluación basada en complejidad de la hipótesis. El principio MDL.....	477
17.6 Evaluación de modelos de agrupamiento.....	480
17.7 Evaluación de reglas de asociación.....	481
17.8 Otros criterios de evaluación.....	482
Capítulo 18 Combinación de modelos.....	485
18.1 Introducción.....	485
18.2 Métodos de construcción de multclasificadores.....	487
18.3 Métodos de fusión.....	492
18.4 Métodos híbridos.....	494
Capítulo 19 Interpretación, difusión y uso de modelos.....	503
19.1 Introducción.....	503
19.2 Extracción de reglas comprensibles.....	504
19.3 Visualización posterior.....	506
19.4 Intercambio y difusión de modelos: estándares de representación.....	510
19.5 Integración con la toma de decisiones.....	512
19.6 Actualización y revisión de modelos.....	520

PARTE V: MINERÍA DE DATOS COMPLEJOS

Capítulo 20 Minería de datos espaciales, temporales, secuenciales y multimedia.....	525
20.1 Introducción.....	525
20.2 Minería de datos espaciales.....	526
20.3 Minería de datos temporales.....	531
20.4 Extracción de patrones secuenciales.....	536
20.5 Minería de datos multimedia.....	539
Capítulo 21 Minería de web y textos.....	545
21.1 Introducción.....	545
21.2 Minería web.....	548
21.3 Minería del contenido de la web.....	551
21.4 Minería de la estructura de la web.....	560
21.5 Minería de uso web.....	563
21.6 Sistemas de minería de web y textos.....	568

PARTE VI: IMPLANTACIÓN E IMPACTO DE LA MINERÍA DE DATOS

Capítulo 22 Implantación de un programa de minería de datos.....	573
22.1 Introducción.....	573
22.2 ¿Cuándo empezar? Necesidades y objetivos de negocio.....	575
22.3 Formulación del programa: fases e implantación.....	580
22.4 Integración con las herramientas y proyectos de la organización.....	586

XII Introducción a la Minería de Datos

22.5 Recursos necesarios	590
Capítulo 23 Repercusiones y retos de la minería de datos	597
23.1 Impacto social de la minería de datos	597
23.2 Cuestiones éticas y legales	599
23.3 Escalabilidad. Minería de datos distribuida	601
23.4 Tendencias futuras	605

APÉNDICES

Apéndice [A] Sistemas y herramientas de minería de datos	609
<i>Vicent Estruch Gregori</i>	
Librerías	609
Suites	611
Herramientas específicas	621
Apéndice [B] Datos de ejemplo	625
Tabla resumen	627
Referencias bibliográficas	629
Índice analítico	651