

ÍNDICE

<i>Introducción</i>	<i>XVII</i>
<i>Capítulo 1. Minería de datos: Conceptos, técnicas y sistemas</i>	<i>1</i>
Aproximación al concepto de minería de datos	1
El proceso de extracción del conocimiento	3
Técnicas de minería de datos.....	8
Sistemas de minería de datos	10
<i>Capítulo 2. Entorno de trabajo de SPSS Clementine</i>	<i>13</i>
Introducción a Clementine	13
Usando el ratón.....	16
Ayuda en Clementine	16
Panel de control en Clementine.....	18
Ejemplo de trabajo con Clementine	21
Insertar un nodo fuente (origen) de datos en el área de trabajo.....	22
Enlazar un nodo con una fuente de datos	23
Controlar la carga de datos con el <i>nodo Tabla</i>	25
Definir variables predictoras con el <i>nodo Tipo</i>	27
Utilizar un nodo de modelado	29
Ejecutar una ruta.....	29
Interpretar un modelo	32
Predecir con un modelo	34
Guardar un modelo	34
Nodos de orígenes de datos	35
Nodos de operaciones con registros	35

Nodos de operaciones con campos.....	36
Nodos para gráficos	37
Nodos para modelado	38
Nodos de salida	40
Capítulo 3. Entorno de trabajo de SAS Enterprise Miner.....	41
Introducción a SAS Enterprise Miner	41
Comenzando con SAS Enterprise Miner.....	43
Inicio de un proyecto nuevo	47
Menú principal de SAS Enterprise Miner	48
Ejemplo de trabajo con SAS Enterprise Miner	58
Leer ficheros y enlazarlos con Enterprise Miner mediante el <i>nodo</i> <i>Input Data Source</i>	58
Definir tipos de variables con el <i>nodo Input Data Source</i>	63
Enlace de nodos de un diagrama. El <i>nodo Data Partition</i>	65
Utilizar un nodo de modelado	67
Capítulo 4. Fase de selección en minería de datos.....	73
Selección en el proceso de extracción del conocimiento	73
Recopilación e integración de datos: <i>Data Warehouse</i>	74
<i>Data Warehouse</i> y <i>Data Mining</i>	77
Selección de datos mediante muestreo.....	78
Muestreo aleatorio simple	82
Muestreo estratificado	85
Muestreo sistemático.....	91
Muestreo unietápico de conglomerados	95
Muestreo bietápico de conglomerados	99
Muestreo polietápico de conglomerados	101
Diseños complejos: Bietápico con estratificación en primera etapa	101
Selección de números aleatorios: Método de Montecarlo.....	102
Selección de características relevantes.....	104
Análisis de correlaciones.....	105
Capítulo 5. Fase de selección en SAS Enterprise Miner y SPSS Clementine.....	109
La fase de selección en Enterprise Miner.....	109
El <i>nodo Fuente de Datos</i>	109
El <i>nodo Muestreo</i>	117
El <i>nodo de Partición de Datos</i>	122
El <i>nodo de Selección de Variables</i>	125
El <i>nodo de Series Temporales</i>	129

La fase de selección en SPSS Clementine.....	139
Importación de datos ASCII.....	140
Importación de datos de una fuente ODBC (Access, Excel, etc.)	140
Importación de datos de SPSS.....	143
Importación de datos de SAS	145
Selección de datos	148
Muestreo de datos	149
Capítulo 6. Fase de selección en SPSS Muestras Complejas y SAS Base.....	151
Técnicas de muestreo a través de SPSS	151
Diseños complejos y el asistente de muestreo.	
Creación de un nuevo plan de muestreo.....	152
Asistente de muestreo: modificar un plan existente	161
Asistente de muestreo: ejecutar un plan de muestreo dado.....	164
Preparación de una muestra compleja para su análisis:	
Creación de un nuevo plan de análisis	164
Preparación de una muestra compleja para su análisis.....	168
Cálculos en muestras complejas: Frecuencias, descriptivos, tablas de contingencia y razones	168
Selección de casos en SPSS	174
Selección de casos mediante criterios condicionales	174
Selección de fechas, horas y filas	175
Selección de una muestra aleatoria.....	175
Semilla de aleatorización.....	176
Operadores para la selección en SPSS	176
Operadores aritméticos.....	176
Operadores relacionales	177
Operadores lógicos.....	177
Funciones de generación de números aleatorios en SPSS.....	177
Selección de la información en SAS Base	180
Declarando valores perdidos con la sentencia MISSING	180
Seleccionando información por grupos: sentencia BY	180
Seleccionando variables de frecuencias: sentencia FREQ	182
Seleccionando variables de pesos: sentencia WEIGHT	183
Seleccionando variables de identificación: Sentencia ID.....	184
Operadores para la selección en SAS.....	184
Operadores aritméticos.....	185
Operadores de comparación	185
Operadores lógicos o booleanos.....	186
Operadores MIN, MAX y concatenación.....	187
Orden de evaluación de los operadores en las expresiones.....	188
Funciones de generación de números aleatorios en SAS	189
Cálculos con funciones en SAS.....	191

Capítulo 7. Fase de exploración en minería de datos	193
Exploración en el proceso de extracción del conocimiento	193
Análisis exploratorio	194
Herramientas de exploración visual	194
Histograma de frecuencias	195
Diagrama de tallo y hojas	196
Gráfico de caja y bigotes	198
Gráfico múltiple de caja y bigotes	199
Gráfico de simetría	201
Gráfico de dispersión	203
Gráficos para variables cualitativas	205
Herramientas de exploración formal	207
Contrastes de la bondad de ajuste a una distribución: Test de la Chi-cuadrado....	208
Contraste de Kolmogorov-Smirnov Lilliefors de la bondad de ajuste a una distribución	209
Estadísticos robustos de centralización	211
Estadísticos robustos de dispersión	212
Estadísticos robustos de asimetría y curtosis	214
Contrastes de aleatoriedad	216
Transformaciones de las variables	220
Supuestos subyacentes en las técnicas de minería de datos	221
Normalidad	221
Heteroscedasticidad	225
Multicolinealidad	227
Autocorrelación	227
Linealidad	228
Un ejemplo	230
 Capítulo 8. Fase de exploración en SAS Enterprise Miner y SPSS Clementine	 239
La fase de exploración en Enterprise Miner	239
El <i>nodo Explorador de distribuciones</i>	239
El <i>nodo Multigráficos</i>	243
El <i>nodo de exploración de patrones</i>	250
La fase de exploración en SPSS Clementine	266
El <i>nodo Gráfico</i>	267
El <i>nodo Distribución</i>	270
El <i>nodo Histograma</i>	271
El <i>nodo Malla</i>	273
El <i>nodo Malla Direccional</i>	274
El <i>nodo Gráfico Múltiple</i>	275
El <i>nodo Recolectar</i>	276

Capítulo 9. Fase de exploración en SPSS y SAS	277
Análisis exploratorio de datos con SPSS. <i>Procedimiento Explorar</i>	277
Gráficos de análisis exploratorio con SPSS	282
Tipos de gráficos	282
Histogramas.....	283
Gráficos de normalidad	283
Gráficos de caja y bigotes	286
Gráficos de dispersión.....	288
Gráficos interactivos dinámicos de análisis exploratorio con SPSS	290
Creación interactiva de gráficos a partir de tablas.....	297
Gráficos interactivos de caja y bigotes.....	298
Histogramas interactivos.....	299
Diagramas interactivos de dispersión.....	301
Análisis exploratorio formal con SPSS	303
Contraste de aleatoriedad. <i>Procedimiento Prueba de rachas</i>	303
Contraste de ajuste a una distribución de frecuencias. <i>Procedimiento Prueba de Kolmogorov-Smirnov</i>	304
Análisis exploratorio de los datos con SAS Base. <i>Procedimiento Univariate</i>	305
Gráficos de análisis exploratorio con SAS.....	318
Gráficos exploratorios de alta resolución. <i>Procedimiento GCHART</i>	318
Gráficos exploratorios de mapas: <i>Procedimiento GMAP</i>	322
Gráficos exploratorios de caja y bigotes: <i>Procedimiento BOXPLOT</i>	328
 Capítulo 10. Fases de limpieza y transformación de datos	 333
Limpieza y transformación de datos en el proceso de extracción del conocimiento	333
Valores atípicos (<i>Outliers</i>)	333
Información faltante (<i>Datos missing</i>).....	337
Soluciones para los datos ausentes: Supresión de datos e imputación de información faltante.....	343
Transformación de datos	346
Transponer, fusionar, agregar, segmentar y ordenar archivos.....	346
Ponderar casos y categorizar y numerizar variables.....	347
Pareamiento o <i>matching</i>	348
Transformación de datos mediante técnicas de reducción de la dimensión.....	349
Componentes principales	350
Análisis factorial.....	357

Capítulo 11. Las fases de limpieza y transformación de datos en SAS Enterprise Miner y SPSS Clementine.....	365
Las fases de limpieza y transformación de datos en Enterprise Miner	365
El <i>nodo</i> Transformación de variables	365
El <i>nodo</i> Asignación de atributos	371
Tratamiento de los datos atípicos con el <i>nodo</i> Filtro de Outliers	378
El <i>nodo</i> Imputación de datos missing.....	384
El <i>nodo</i> Exploración de patrones para Componentes Principales	393
Las fases de limpieza y transformación de datos en Clementine	400
El <i>nodo</i> Seleccionar.....	402
El <i>nodo</i> Muestra para procesos de muestreo.....	404
El <i>nodo</i> Combinar para procesos de <i>matching</i>	405
El <i>nodo</i> Equilibrar.....	407
El <i>nodo</i> Ordenar.....	408
El <i>nodo</i> Agregar para calcular estadísticos por subgrupos	409
El <i>nodo</i> Distinguir.....	411
El <i>nodo</i> Añadir para concatenación de archivos	411
El <i>nodo</i> Filtrar.....	412
El <i>nodo</i> Derivar para transformación de variables.....	413
El <i>nodo</i> Tipo para asignar atributos a variables.....	415
El <i>nodo</i> Rellenar para imputación de datos <i>missing</i>	416
El <i>nodo</i> Factor/PCA para Análisis Factorial y Componentes Principales ...	417
 Capítulo 12. Fases de limpieza y transformación de datos en SPSS y SAS ..	427
Técnicas de reducción de la dimensión en SPSS Base.....	427
Componentes principales con SPSS.....	428
Análisis factorial con SPSS	439
Transformación de datos en SPSS Base	447
Transformación de valores de datos	447
Remodificación de variables	449
Ordenar casos	451
Transponer, fusionar, agregar y segmentar archivos. <i>Matching</i>	451
Ponderar casos	458
Categorizar variables: Categorizador visual.....	459
Asignar rangos a casos y tipificar variables	462
SPSS y el análisis de datos <i>missing</i> . Imputación.....	463
Reemplazar valores perdidos.....	469
Detección de valores atípicos en SPSS	470
Detección de casos atípicos mediante gráficos de control	470
Detección de casos atípicos mediante gráficos de caja y bigotes.....	472
Técnicas de reducción de la dimensión en SAS STAT.....	475

Componentes principales en SAS. <i>Procedimiento PRINCOMP</i> y <i>Procedimiento FACTOR</i>	475
Análisis factorial en SAS. <i>Procedimiento FACTOR</i>	482
Transformación de datos en SAS Base	487
Operaciones con ficheros: Concatenación y <i>Matching</i>	487
Actualizando ficheros de datos SAS	489
Añadir información. <i>Procedimiento APPEND</i>	491
Tipificación de datos: <i>Procedimiento STANDARD</i>	494
 Capítulo 13. Fase de minería de datos. Técnicas predictivas de modelización ...	497
 Técnicas de minería de datos propiamente dichas	497
Técnicas predictivas para la modelización	498
Modelo de regresión múltiple	504
Estimación del modelo lineal de regresión múltiple	505
Estimación del modelo, contrastes e intervalos de confianza a través del cálculo matricial	506
Análisis de la varianza en el modelo de regresión múltiple	507
Predicciones	510
Análisis de los residuos	511
Técnicas de selección en el modelo de regresión	512
Modelos de elección discreta	513
Modelos de elección discreta binaria: Modelo lineal de probabilidad y regresión logística binaria	514
Modelos de elección múltiple: Modelo Logit Multinomial	519
Modelo lineal general de regresión múltiple (GLM)	521
Clasificación ad hoc: Análisis discriminante	521
Hipótesis en el modelo discriminante	522
Estimación del modelo discriminante	523
Clasificación mediante el modelo discriminante	525
 Capítulo 14. Técnicas predictivas de modelización con SAS Enterprise Miner y SPSS Clementine	529
 Técnicas predictivas de modelización con SAS Enterprise Miner	529
El <i>nodo Regression</i> : Modelo de regresión múltiple	530
El <i>nodo Regression</i> : Modelo lineal general GLM	538
El <i>nodo Regression</i> : Modelo de elección discreta Logit y Probit	551
Técnicas predictivas de modelización con SPSS Clementine	554
El <i>nodo Regresión Lineal</i> : Modelo de regresión múltiple	555
El <i>nodo Regresión Logística</i> : Modelos de elección discreta	561

Capítulo 15. Técnicas predictivas de modelización con SAS y SPSS	565
El modelo lineal general con SAS. <i>Procedimiento GLM</i>	565
Modelos del análisis de la varianza y la covarianza con SAS.....	571
Modelo de elección discreta en SAS.....	574
Modelo Logit: <i>Procedimiento LOGISTIC</i>	574
Modelo Probit: <i>Procedimiento PROBIT</i>	579
SAS y el análisis discriminante: <i>Procedimiento DISCRIM</i>	581
El modelo lineal general con SPSS. <i>Procedimiento MLG Multivariante</i>	585
Modelo de elección discreta en SPSS.....	593
Modelo Logit: <i>Procedimiento LOGISTICA MULTINOMIAL</i>	593
Modelo Probit: <i>Procedimiento PROBIT</i>	599
SPSS y el análisis discriminante.....	601
Capítulo 16. Técnicas descriptivas y predictivas de clasificación.	
Clusters y árboles de decisión	609
El análisis cluster como técnica descriptiva de clasificación.....	609
Medidas de similitud.....	610
Técnicas en el análisis <i>cluster</i>	614
Clusters jerárquicos, secuenciales, aglomerativos y exclusivos (S.A.H.N.).....	616
El dendograma en el análisis cluster jerárquico.....	617
Análisis cluster no jerárquico.....	617
Los árboles de decisión como técnica predictiva de clasificación.....	621
Características de los árboles de decisión.....	622
Herramientas para el trabajo con árboles de decisión.....	626
Árboles CHAID.....	627
Árboles CART.....	628
Árboles QUEST.....	630
Análisis de conglomerados y árboles de decisión como métodos de segmentación.....	631
Capítulo 17. Clusters y árboles de decisión con SAS Enterprise Miner y SPSS Clementine	633
Análisis <i>cluster</i> con Enterprise Miner. El <i>nodo Clustering</i>	633
Árboles de decisión con Enterprise Miner. El <i>nodo Tree</i>	641
Entrenamiento interactivo (<i>Interactive Training</i>).....	652
Análisis <i>cluster</i> con SPSS Clementine.....	656
El <i>nodo Entrenar K-medias: Cluster</i> no jerárquico.....	656
El <i>nodo Cluster Bietápico: Cluster</i> jerárquico.....	661
Árboles de decisión con SPSS Clementine.....	662
El <i>nodo Crear C5.0</i>	662
El <i>nodo Árbol C&R</i>	664

Capítulo 18. Clusters y árboles de decisión con SAS y SPSS	665
SPSS y el análisis <i>cluster</i> jerárquico	665
SPSS y el análisis <i>cluster</i> no jerárquico	671
SAS y el análisis <i>cluster</i> jerárquico	675
Procedimiento <i>ACECLUS</i>	675
Procedimiento <i>CLUSTER</i>	677
Procedimiento <i>TREE</i>	678
SAS y el análisis <i>cluster</i> no jerárquico	681
Árboles de decisión (o clasificación) con SPSS	687
Creación de un árbol de decisión: Método CHAID	689
Métodos CRT y QUEST. Poda de árboles	695
Capítulo 19. Redes neuronales	699
Descripción de una red neuronal	699
Definición	699
Función de salida y funciones de transferencia o activación	701
Redes neuronales y ajuste de modelos de regresión	703
Aprendizaje en las redes neuronales	704
Funcionamiento de una red neuronal	707
El algoritmo de aprendizaje Retropropagación (<i>Back-Propagation</i>)	708
Análisis discriminante a través del Perceptrón	709
Análisis de series temporales mediante redes neuronales	713
Análisis de componentes principales con redes neuronales	715
Clustering mediante redes neuronales	717
Capítulo 20. Redes neuronales con SAS Enterprise Miner y SPSS Clementine	721
Redes neuronales con SAS Enterprise Miner	721
Optimización y ajuste de modelos con redes: <i>Nodo Neural Network</i>	722
Análisis en componentes principales a través de redes neuronales: <i>Nodo Princomp/Dmneural</i>	745
Predicción y análisis discriminante a través de redes neuronales: <i>Nodo Two Stage Model</i>	751
Análisis <i>cluster</i> con redes neuronales: <i>Nodo SOM/Kohonen</i>	756
Redes neuronales con SPSS Clementine	765
<i>Nodo Entrenar red</i>	765
<i>Nodo Entrenar Kohonen</i>	769
<i>Nodo Entrenar K-medias</i>	771
Índice alfabético	775