

# ÍNDICE

<b>Introducción .....</b>	<b>XI</b>
<b>Capítulo 1. Técnicas de minería de datos y herramientas .....</b>	<b>1</b>
Clasificación de las técnicas de minería de datos y herramientas más comunes .....	1
Modelado originado por la teoría (técnicas explicativas o de la dependencia).....	3
Modelado originado por los datos (técnicas descriptivas o de la interdependencia) .....	4
El entorno de trabajo de IBM SPSS .....	5
Entorno de trabajo de IBM SPSS Modeler.....	10
Entorno de trabajo de SAS Enterprise Miner .....	12
Los entornos de trabajo de SAS y SAS Enterprise Guide .....	16
<b>Capítulo 2. Fase de selección en minería de datos. Herramientas .....</b>	<b>23</b>
La fase inicial en minería de datos: selección de la información.....	23
Selección mediante muestreo estadístico .....	25
Muestreo aleatorio simple con IBM SPSS .....	29
Muestreo estratificado con IBM SPSS .....	30
Estimaciones y cálculo de errores en el muestreo estratificado .....	40
Muestreo de conglomerados monoetápico y polietápico con IBM SPSS.....	42
Estimaciones y cálculo de errores en el muestreo de conglomerados .....	48
La fase de selección en IBM SPSS Modeler .....	49
Selección de datos con IBM SPSS Modeler .....	50
Muestreo con IBM SPSS Modeler .....	53
La fase de selección en SAS Enterprise Miner .....	54
Pestaña Data .....	56
Pestaña Variables .....	57
El perfil del objetivo (Target Profile).....	58
Especificar probabilidades a priori .....	59
Pestaña Interval Variables .....	60
Pestaña Class Variables .....	60
Pestaña Notes.....	61

Muestreo en SAS Enterprise Miner .....	61
Partición de datos en SAS Enterprise Miner.....	65
El nodo de selección de variables.....	68
Muestreo a través de SAS .....	71
Muestreo aleatorio simple .....	73
Muestreo aleatorio estratificado.....	75

**Capítulo 3. Fases de exploración y limpieza en minería de datos.**

<b>Herramientas .....</b>	<b>77</b>
Análisis exploratorio de datos .....	77
Limpieza de datos .....	79
Valores atípicos .....	79
Datos desaparecidos. Imputación .....	81
Análisis exploratorio gráfico y formal con IBM SPSS .....	85
Procedimiento Explorar .....	85
Contraste de aleatoriedad. Procedimiento Prueba de rachas.....	90
Normalidad. Test de ajuste a una distribución de frecuencias.	
Procedimiento prueba de Kolmogorov-Smirnov.....	93
Tratamiento de los datos ausentes con IBM SPSS. Imputación.....	94
Detección de datos atípicos con IBM SPSS.....	101
Detección de valores atípicos en IBM SPSS mediante gráficos de control ...	102
Detección de casos atípicos en IBM SPSS mediante gráficos de caja y bigotes..	104
Análisis exploratorio gráfico y formal con IBM SPSS Modeler .....	105
El nodo Gráfico .....	106
El nodo Distribución .....	108
El nodo Histograma .....	110
La fase de limpieza en IBM SPSS Modeler.....	111
El nodo Rellenar para imputación de datos missing.....	112
La fase de Exploración en Enterprise Miner.....	112
El nodo Explorador de distribuciones.....	113
El nodo de Exploración de Patrones.....	114
Tratamiento de datos atípicos con el nodo Filtro de Outliers de	
SAS Enterprise Miner.....	122
Opciones de filtrado automático .....	122
Observación y ajuste de los resultados del filtrado .....	123
El nodo Imputación de datos missing.....	125
La fase de Exploración en SAS.....	132

<b>Capítulo 4. Fase de transformación en minería de datos. Herramientas .....</b>	<b>135</b>
Transformación de datos .....	135
Transponer, fusionar, agregar, segmentar y ordenar archivos.....	136
Ponderar casos y categorizar y numerizar variables.....	137
Pareamiento o <i>matching</i> .....	137
Transformación de datos mediante técnicas de reducción de la dimensión ...	139
Análisis factorial y componentes principales .....	139
Contrastes en el modelo factorial .....	141
Rotación de los factores .....	143
Interpretación gráfica de los factores y puntuaciones factoriales.....	144
Técnicas de reducción de la dimensión en IBM SPSS.....	146
Análisis factorial y componentes principales en IBM SPSS.....	147
Transformación de datos en IBM SPSS.....	153
Transponer, fusionar, agregar y segmentar archivos. <i>Matching</i> .....	153
Ponderar casos .....	160
Transformación de datos y técnicas de reducción de la dimensión en IBM SPSS Modeler .....	161
<i>Matching</i> mediante el nodo Combinar.....	161
Ajuste de registros con el nodo Equilibrar.....	163
Estadísticos por subgrupos con el nodo Agregar .....	164
Remover duplicados con el nodo Distinguir .....	165
Concatenación de archivos con el nodo Añadir.....	166
Transformación de variables con el nodo Derivar .....	167
Análisis factorial y componentes principales con el nodo Factor/PCA.....	169
Transformación de datos y reducción de la dimensión en SAS Enterprise Miner .	177
El nodo Transformación de variables .....	177
Componentes principales con el nodo Exploración de patrones.....	181
Componentes principales y análisis factorial en SAS y SAS Enterprise Guide ..	188
<b>Capítulo 5. Fase de análisis en minería de datos. Técnicas predictivas. Herramientas .....</b>	<b>197</b>
Fase de análisis en minería de datos.....	197
Modelos de regresión múltiple y de elección discreta. Modelos Logit y Probit. Segmentación.....	199
Modelos de elección discreta binaria: modelo lineal de probabilidad y regresión logística binaria .....	200
Modelos de elección múltiple: modelo Logit Multinomial.....	202

IBM SPSS y los modelos de variable dependiente limitada. Aplicaciones a la segmentación.....	202
IBM SPSS y la regresión logística binaria .....	203
IBM SPSS y el modelo Probit.....	210
IBM SPSS y el modelo Logit Multinomial .....	214
SAS y los modelos de elección discreta .....	220
SAS y la regresión logística. PROC LOGISTIC .....	220
SAS y el modelo Probit: procedimiento Probit .....	227
El modelo de análisis discriminante como técnica de clasificación y segmentación.....	230
IBM SPSS y el análisis discriminante.....	236
SAS y el análisis discriminante: PROC DISCRIM .....	244
Ejemplo de análisis discriminante con SAS .....	245
SAS y el análisis discriminante paso a paso: PROC STEPDISC y ejemplo práctico.....	248
Técnicas predictivas con SAS Enterprise Miner.....	253
Modelo de regresión múltiple con el nodo Regression.....	253
El nodo Regression: modelo lineal general GLM.....	260
Modelos de elección discreta Logit y Probit con el nodo Regression.....	271
Modelos predictivos con IBM SPSS Modeler.....	273
Modelo de regresión múltiple con el nodo Regresión Lineal .....	274
Modelos de elección discreta con el nodo Regresión Logística .....	281
<b>Capítulo 6. Modelos predictivos con árboles de decisión. Herramientas.....</b>	<b>285</b>
Introducción a los árboles de decisión .....	285
Características de los árboles de decisión.....	287
Tipos de árboles de decisión .....	289
Árboles CHAID .....	289
Árboles CART .....	290
Árboles QUEST.....	291
IBM SPSS y los árboles de decisión.....	292
Creación de un árbol de decisión: método CHAID .....	294
Métodos CRT y QUEST. Poda de árboles.....	300
Árboles de decisión y análisis de riesgos.....	304
El proceso de creación de árboles de decisión con IBM SPSS Modeler.....	307
Situar los datos en el área de trabajo con un nodo de origen de datos .....	308
Enlace del origen de datos con la fuente de datos .....	309
Visualización de los datos con el nodo Tabla.....	310

El nodo Tipo y las variables de predicción .....	311
Nodo de modelado para crear el árbol.....	313
Trabajar con árboles .....	313
Interpretación del árbol.....	316
Predicción con un modelo de árbol .....	317
Guardar un modelo .....	317
Nodos para árboles de decisión con IBM SPSS Modeler .....	317
El nodo Crear C5.0 .....	318
El nodo Árbol C&R .....	319
El proceso de creación de árboles de decisión con SAS Enterprise Miner .....	320
Nodo Input Data Source para leer ficheros de datos y enlazarlos con SAS Enterprise Miner .....	320
Nodo Input Data Source y los tipos de variables .....	324
Enlace de nodos de un diagrama. El nodo Data Partition.....	325
Nodo de modelado para crear el árbol.....	328
Profundizando en el nodo Tree.....	332
<b>Capítulo 7. Técnicas descriptivas de minería de datos. Análisis clúster y segmentación. Herramientas.....</b>	<b>343</b>
El análisis clúster como técnica de clasificación y segmentación.....	343
Clústers jerárquicos, secuenciales, aglomerativos y exclusivos (S.A.H.N.) ...	345
El dendograma en el análisis clúster jerárquico .....	346
Análisis clúster no jerárquico.....	346
IBM SPSS y el análisis clúster jerárquico .....	349
IBM SPSS y el análisis clúster no jerárquico .....	355
SAS y el análisis clúster jerárquico .....	360
Procedimiento ACECLUS.....	360
Procedimiento CLÚSTER .....	361
Procedimiento TREE .....	363
SAS y el análisis clúster no jerárquico .....	365
Análisis clúster con Enterprise Miner. El nodo Clustering.....	371
Análisis clúster con IBM SPSS Modeler .....	377
El nodo Entrenar K-medias: clúster no jerárquico .....	378
El nodo Clúster bietápico: clúster jerárquico.....	382

<b>Capítulo 8. Redes neuronales. Herramientas.....</b>	<b>383</b>
Introducción .....	383
Redes neuronales con SAS Enterprise Miner .....	383
Optimización y ajuste de modelos con redes: nodo Neural Network .....	384
Análisis en componentes principales a través de redes neuronales:	
nodo Princomp/ Dmneural.....	406
Predicción y análisis discriminante a través de redes neuronales:	
nodo Two Stage Model.....	412
Análisis clúster con redes neuronales: nodo SOM/Kohonen .....	418
Redes neuronales con IBM SPSS Modeler.....	426
Nodo Entrenar red.....	427
Análisis clúster con redes neuronales: nodo Entrenar Kohonen .....	430
 <b>Capítulo 9. Fase de evaluación. Comparación de modelos. Herramientas....</b>	 <b>433</b>
Fase de evaluación en SAS Enterprise Miner. Nodo Assessment.....	433
Trabajo con el nodo de evaluación .....	434
El fichero de datos con las predicciones .....	434
Models.....	434
Options .....	437
Reports .....	438
Output .....	438
Gráficos del nodo Assessment .....	438
Lift (Gráfico de ganancias) .....	438
Vista de los datos relativos a los gráficos de ganancias.....	441
DIAGNOSTIC (Gráfico de clasificación).....	442
THRESHOLD-BASED (Gráfico de clasificación basado en el umbral) .....	443
Curvas ROC .....	445
Scatter Plot (Gráfico de dispersión).....	446
Ejemplo de combinación de modelos de árboles con otros modelos .....	446
Capacidad predictiva de un modelo .....	449
Selección de modelos. Ejemplo con IBM SPSS .....	450
Curvas ROC. Ejemplo con IBM SPSS .....	452
 <b>Índice analítico.. .....</b>	 <b>455</b>