

Machine learning with RapidMiner Studio

Alfredo Daza Vergaray
David Santos Mauricio Sánchez



FONDO EDITORIAL
Universidad César Vallejo

Machine learning with RapidMiner Studio



FONDO EDITORIAL
Universidad César Vallejo

Machine learning with RapidMiner Studio

©Universidad César Vallejo, 2022

Edición y diseño: Fondo Editorial Universidad César Vallejo

Primera edición, octubre de 2022

Tiraje: 500 ejemplares

ISBN: 978-612-4435-88-1

Hecho el Depósito Legal en la Biblioteca Nacional del Perú N.º 2022-12059

Edición y diseño:

Universidad César Vallejo SAC

Av. Alfredo Mendiola 6232, Panamericana Norte, Los Olivos

Lima, Perú

El contenido de este libro fue validado a través de un proceso de evaluación de pares ciegos (double-blind peer review)

Todos los derechos reservados. La reproducción parcial o total de esta obra en cualquier tipo de soporte está prohibida sin la autorización expresa de los editores.

Machine learning with RapidMiner Studio

Alfredo Daza Vergaray
David Santos Mauricio Sánchez



FONDO EDITORIAL
Universidad César Vallejo

CONTENIDO

Resumen	21
Introducción	23
Capítulo 1: Conceptos básicos de minería de datos	25
1. Minería de datos	27
2. Procesos de minería de datos (kdd)	28
3. Metodología CRISP	29
4. Metodología SEMMA	31
5. Comparación entre CRISP y SEMMA	33
6. Modelo	34
7. Modelo híbrido	34
8. Predicción	35
9. Almacén de datos (datawarehouse)	35
10. Text mining	36
Definición	36
Etapas de text mining	36
Trabajos relacionado con text mining	36
11. Web mining	37
Definición	37
Etapas de web mining	38
Tipo de minería web	39
Trabajos de web mining con redes neuronales	39
12. Bigdata	40
Definición	40
Arquitectura de Big data	40
Formato de datos para Big data	41
Trabajos relacionados con Big data	42

CAPÍTULO 2: Técnicas y aplicación de la minería de datos 47

1. Técnicas y aplicación de la minería de datos	49
2. Métodos de minería de datos	49
Árboles de clasificación	49
Clúster	58
Reglas de asociación	58
Series de tiempo	59
Regresión	59
3. Tipos de problemas de minería de datos	59
4. Aplicación de la minería de datos	64
Investigación espacial	65
Medicina	65
Minería de datos en la educación	67
5. Herramientas top de minería de datos	83

CAPÍTULO 3: Presentación general del Software RapidMiner 91

1. Presentación general del software RapidMiner	93
Características de RapidMiner	93
2. Áreas de aplicación de RapidMiner	93
Sector público	94
CRM	94
Web mining	95
El descubrimiento de fármacos y la bioinformática	96

CAPÍTULO 4: Instalación y partes del Software RapidMiner 99

1. Instalación y partes del software RapidMiner	101
2. Partes del software RapidMiner	104
3. Operadores de RapidMiner	108

CAPÍTULO 5: Aplicaciones con diferentes algoritmos de Machine learning con RapidMiner **123**

1. Aplicaciones con diferentes algoritmos de Machine learning con RapidMiner	125
Casos aplicados a la minería de datos	125
Caso N° 1: Predicción de Juego de Golf	125
Caso N° 2: Predicción de Planta IRIS	129
Caso N° 3: Series de Tiempo con IPSA	136
Caso N° 4: Aplicando clusters con reducción de variables a la base de datos Iris	144
Caso N° 5: Aplicando clusters en user knowledge modeling data set	148
Caso N° 6: Matriz de Confusión	155
Caso N° 7: Redes Bayesianas con Pruebas del Modelo	158
Caso N° 8: Comparando resultados de modelos de aprendizaje a través del ROC	161
Caso N° 9: Integrando datos y luego aplicar árboles de decisión	165
Caso N° 10: Predicción de medicamento para determinadas alergias	170
Caso N° 11: Árboles de decisión para segmentación de mercados	176
Caso N° 12: Stacking y la mejora en resultados	178

CAPÍTULO 6: Aplicaciones con diferentes algoritmos de Machine learning con RapidMiner **183**

1. Aplicaciones con diferentes algoritmos de Machine learning con RapidMiner	185
Caso N° 13: Regresión lineal	185
Caso N° 14: Mostrar datos de sql serve en rapid miner	190
Caso N° 15: Regresión múltiple	198
Caso N° 16: Regresión Logística	203
Caso N° 17: Transformación de datos	209
Caso N° 18: Textminig en un archivo	217
Caso N° 19: Reglas de asociación	223
Caso N° 20: Prueba de significancia	231
Caso N° 21: Correlación	235
Caso N° 22: Análisis discriminante	243
2. Modelo	251
Caso N° 23: Caso de redes neuronales	255
Caso N° 24: Web -Textminig	263

ÍNDICE DE FIGURAS

Figura 1. Descripción general de los pasos que constituyen el proceso KDD	28
Figura 2. Esquema de los cuatro niveles de abstracción de la metodología	29
Figura 3. Fases del proceso de minería de datos, basado en la metodología CRISP-DM	30
Figura 4. Flujo del proceso de Minería de Datos	32
Figura 5. Proceso de clasificación y distribución de correo electrónico y correo SPAM	37
Figura 6. Etapas del web mining	38
Figura 7. Clasificación de tipo de minería web	39
Figura 8. Arquitectura de Big Data	41
Figura 9. Formato de datos para Big Data	42
Figura 10. Minería de datos	44
Figura 11. Árbol de decisión para jugar una partida de Golf	50
Figura 12. Esquema de árboles de decisión para determinar el tipo de planta iris	51
Figura 13. Algoritmos de aprendizaje de árboles de decisión	52
Figura 14. Algoritmo de árbol de decisión de la base de datos Iris	52
Figura 15. Partición cuadrícula del espacio obtenido por un árbol de Decisión	53
Figura 16. Funciones de impureza para criterios de partición	54
Figura 17. Esquema neuronal real y esquema de un perceptrón simple	55
Figura 18. Estructura de PMC	56
Figura 19. Manual de EDM	68
Figura 20. Árboles de Decisión	71
Figura 21. Modelo de redes neuronales	73
Figura 22. Análisis por Arboles de Decisión	75
Figura 23. Evaluación y comparación de modelo	76
Figura 24. Esquema de Decisión usada por el método propuesto	77
Figura 25. Herramientas más usadas de Minería de Datos	83
Figura 26. Cuadrante de Líderes en Plataforma Advanced Analytics	84
Figura 27. Big data predictive	85
Figura 28. Detección de fraude	94
Figura 29. Pirámide de porcentaje	95
Figura 30. Análisis de texto	95
Figura 31. Modelo de conglomerado y clasificación	96
Figura 32. Instalación del software Rapid Miner	101
Figura 33. Pasos de la instalación del Software RapidMiner	102
Figura 34. Pasos de la instalación del Software RapidMiner	102

Figura 35. Pasos de la instalación del Software RapidMiner	103
Figura 36. Pasos de la instalación del Software RapidMiner	103
Figura 37. Pasos de la instalación del Software RapidMiner	104
Figura 38. Pasos de la instalación del Software RapidMiner	104
Figura 39. Partes del Software Rapid Miner	105
Figura 40. Perspectiva de diseño del Software RapidMiner	106
Figura 41. Perspectiva de diseño del Software RapidMiner	106
Figura 42. Perspectiva de diseño del Software RapidMiner	106
Figura 43. Vista de operadores y repositorio	107
Figura 44. Operador aggregate	108
Figura 45. Operador filter examples	109
Figura 46. Operador select attributes	109
Figura 47. Operador relace missing values	110
Figura 48. Operador set role	110
Figura 49. Operador Append	111
Figura 50. Operador Join	111
Figura 51. Operador sort	111
Figura 52. Operador Nominal to binomial	112
Figura 53. Operador performance (classification)	112
Figura 54. Operador performance (regresión)	113
Figura 55. Operador ANOVA	113
Figura 56. Operador T-Test	113
Figura 57. Operador Validation	114
Figura 58. Operador create association rules	114
Figura 59. Operador FP - Growth	115
Figura 60. Operador apply model	115
Figura 61. Operador multiply	115
Figura 62. Operador Retrieve	116
Figura 63. Operador generate data	116
Figura 64. Operador clustering	116
Figura 65. Operador naive bayes	117
Figura 66. Operador linear regression	117
Figura 67. Operador decisión tree	118
Figura 68. Operador neural net	118
Figura 69. Operador support vector machine	118
Figura 70. Operador logistic regression	118
Figura 71. Operador correlation matrix	119
Figura 72. Operador windowing	119
Figura 73. Operador read document	119
Figura 74. Operador process document	119

Figura 75. Operador transform cases	120
Figura 76. Operador tokenize	120
Figura 77. Operador filter stopwords	120
Figura 78. Operador filter tokens	120
Figura 79. Interface de la herramienta RapidMiner	125
Figura 80. Panel de desarrollo conectado con la salida (res)	126
Figura 81. Base de datos en la herramienta de RapidMiner	126
Figura 82. Retrieve golf y res en la herramienta de RapidMiner	127
Figura 83. Modelo de árbol de decisión	127
Figura 84. Operador retrieve golf y res	128
Figura 85. Resultados de las reglas del modelo	128
Figura 86. Predicción de planta Iris	129
Figura 87. Modelo de Predicción de planta Iris	130
Figura 88. Nueva ventana en la herramienta de RapidMiner	130
Figura 89. Base de datos Iris	131
Figura 90. Panel de desarrollo en la base de datos Iris	131
Figura 91. Data view	132
Figura 92. Pestaña MetaDataView con datos en la etiqueta Name, Type y Range	132
Figura 93. 50 plantas del tipo Iris – setosa; Versicolor y Virginica	133
Figura 94. Modelo de clasificación (Decisión Tree)	134
Figura 95. Resultado de la ejecución del proyecto	134
Figura 96. Opción Text view	135
Figura 97. SPSS Clementine	135
Figura 98. Series Processing	136
Figura 99. Series de extensión	137
Figura 100. Árbol de operadores	137
Figura 101. Datos de entrenamiento	138
Figura 102. Datos de validación	138
Figura 103. Datos en el operador read csv	139
Figura 104. Operador set role	139
Figura 105. Operador window size	140
Figura 106. Operador apply móvil y forecasting permormance	140
Figura 107. Evaluación de los operadores	140
Figura 108. Pestaña Exampleset (windowing)	141
Figura 109. Modelo de Prediction tren accuracy	142
Figura 110. Datos en Data View o grafico en Plot View	142
Figura 111. Modelo para crear predicción	143
Figura 112. Resultados con nuevos datos (valor de cierre o y valor predicho)	143
Figura 113. Operador plot view	144
Figura 114. Panel de diseño seleccionando la base de datos Iris	145

Figura 115. Visualización de la información de la base de datos iris	145
Figura 116. Operador kmeans conectado al operador retrieve	146
Figura 117. Base de datos Iris conectado con el operador Kmeans	146
Figura 118. Pestaña clúster model	147
Figura 119. Información del registro en la opción folder view	147
Figura 120. Gráfico en la Pestaña exampleset (svd reduction)	148
Figura 121. Página Machine Learning Repository	148
Figura 122. Construcción del modelo	149
Figura 123. Operador read Excel	150
Figura 124. Operador kmeans conectado con el operador read Excel	150
Figura 125. Operador Singular Value Decomposition	151
Figura 126. Operador read Excel conectado con un operador aggregate	151
Figura 127. Opción edit list en un operador read Excel	151
Figura 128. Agrupación por atributos en el operador read Excel	152
Figura 129. Opción add filter en el operador filter examples	152
Figura 130. Pestaña ExampleSet(aggregate) con data de 4 clases: High, Low, Middle, Very_low	153
Figura 131. Información en modo gráfico de la data de 4 clases: High, Low, Middle, Very_low	153
Figura 132. Pestaña cluster models	154
Figura 133. Resultados del filtrado por el operador filter examples	154
Figura 134. Operador read Excel	155
Figura 135. Operador X-Validation conectado al operador read excel	155
Figura 136. Paneles en el operador X-Validation	156
Figura 137. Modelo K vecinos más cercanos en el panel de entrenamiento	156
Figura 138. Panel de evaluación Apply Model y Performance (classification)	157
Figura 139. Matriz de confusión con una precisión del 96%	157
Figura 140. Operador retrieve con la dirección de la base de datos golf	158
Figura 141. Modelo de clasificación Naive Bayes	158
Figura 142. Modelo entrenado con datos nuevos	159
Figura 143. Redes bayesianas con pruebas de modelo	159
Figura 144. Redes bayesianas con tipo de tarea de aprendizaje	159
Figura 145. Exactitud del modelo (64.29%)	160
Figura 146. Valor optimista del AUC (0.467)	160
Figura 147. Comparación de resultados de modelos de aprendizaje a través del ROC	161
Figura 148. Representación del desempeño de un modelo clasificador	162
Figura 149. Modelos representados por sus curvas ROC	162
Figura 150. Operador retrieve enlazado con el operador Compare Roc	163
Figura 151. Modelo de reglas y modelo de árboles de decisión	163
Figura 152. Modelo de reglas	164

Figura 153. Ingreso del modelo SVM y Regresión logística	164
Figura 154. Curvas ROCs	165
Figura 155. Ingreso del operador generate	166
Figura 156. Operador append	166
Figura 157. Integrando datos al operador X- Validation	167
Figura 158. Operador X- Validation	167
Figura 159. Ingreso de operadores	168
Figura 160. 34 datos generados	168
Figura 161. Árbol de decisión	169
Figura 162. Reglas del árbol de decisión	169
Figura 163. Pasos para llegar al modelo	170
Figura 164. Elección de operadores para datos de prueba	171
Figura 165. Unión de operadores Append y read csv	171
Figura 166. Operador Multiply	172
Figura 167. Modelo de clasificación de árboles de decisión conectado con el operador Multiply	172
Figura 168. Operador Apply Model conectado con los modelos entrenados	173
Figura 169. Rendimiento del modelo de árboles de decisión agregando el operador performance	173
Figura 170. Operador Multiply conectado a operador Apply Model	174
Figura 171. Reglas del modelo	174
Figura 172. Generación del Árbol de decisión	175
Figura 173. Pestaña performance Vector	175
Figura 174. Operador Generate Direct Mailing Data colocado en el atributo numbers examples	176
Figura 175. Operador Split Validation conectado con el operador Generate Direct Mailing	176
Figura 176. Ingreso del operador Decisión Tree como del operador Apply Model y el Operador Performance	177
Figura 177. Árbol de decisión para segmentación de mercado	177
Figura 178. Pestaña Performance Vector	178
Figura 179. Esquema de modelo stacking combinado	178
Figura 180. Operador Generate Data con 1000 ejemplos numéricos	179
Figura 181. Operador Stacking con 3 modelos de entrenamiento	179
Figura 182 Generación de valores de predicción	180
Figura 183. Modelo Stacking con operador generate date conectado con el operador Apply Model	180
Figura 184. Modelos agregados con el operador Performance	181
Figura 185. Resultados de los modelos en la pestaña stacking	181
Figura 186. Pestaña Performance Vector	182

Figura 187. Valores del Modelo Stacking	182
Figura 188. Modelo de regresión lineal para predecir altura	185
Figura 189. Modelo de regresión lineal	186
Figura 190. Operador set role enlazado con el operador read excel	186
Figura 191. Modelo de regresión lineal conectado al operador set role	186
Figura 192. Operador read Excel para predecir las respectivas alturas	187
Figura 193. Operador apply model enlazado con el operador regresión lineal y read Excel	187
Figura 194. Pestaña Example Set con valores seleccionados	188
Figura 195. Altura dada por la fórmula	188
Figura 196. Pestaña Example set (read excel 2)	189
Figura 197. Gráfico generado con los valores predichos	190
Figura 198. Datos de SQL Server	190
Figura 199. Logueado usando usuario y clave	191
Figura 200. Base de datos - IRIS	191
Figura 201. Datos de 150 filas correspondiente a tres tipos de plantas	192
Figura 202. Software RapidMiner	192
Figura 203. Creación de un nuevo proyecto "laboratorio_conexion_sqlserver"	193
Figura 204. Operador Read Data base	193
Figura 205. Datos del operador read data base	194
Figura 206. Ingreso de los valores en el parámetro connection	194
Figura 207. Parámetro query colocando código	195
Figura 208. Nodo read data base	195
Figura 209. Información extraída desde SQL server al RapidMiner	196
Figura 210. Nodo setrol, neural net, Apply Model, performance	197
Figura 211. Precisión del modelo	197
Figura 212. Pestaña de la red neuronal	198
Figura 213. Hardware data set	199
Figura 214. Operador Read csv	200
Figura 215. Atributo select attributes conectado con el operador read scv	200
Figura 216. Selección de atributos	200
Figura 217. Operador de regresión lineal conectado con el operador select attributes	201
Figura 218. Copia del operador select attributes conectado con el operador read csv	201
Figura 219. Operador Apply Model conectado con el modelo regresión lineal	202
Figura 220. Pestaña regresión lineal	202
Figura 221. Resultados de la pestaña Example Set (select attributes 2)	203
Figura 222. Operador read Excel con el archivo de datos "\\laboratorio\ practica_16_regresion_logisyica\ telco"	204
Figura 223. Operador select attributes conectado con el operador read excel	204
Figura 224. Selección de atributos age, income, marital y reside	204

Figura 225. Operador multiply	205
Figura 226. Operador de correlation matrix	205
Figura 227. Operador X-Validation	206
Figura 228. Ingreso del operador apply model y performance	206
Figura 229. Pestaña correlation matrix	207
Figura 230. Relación que hay entre las variables	207
Figura 231. Pestaña Example Set	207
Figura 232. Pestaña permormance vector	208
Figura 233. Pestaña w-logistic	208
Figura 234. Import CSV file	209
Figura 235. Búsqueda de la ruta del archivo	209
Figura 236. Separación a través de coma	210
Figura 237. Data import wizard	210
Figura 238. Colocación del nombre " base transformación"	211
Figura 239. Ventana con datos	211
Figura 240. Operador replace missing values	212
Figura 241. Colocación de parámetros	212
Figura 242. Modificación de datos	213
Figura 243. Operador Filter Examples	213
Figura 244. Colocación de parámetros	213
Figura 245. Resultados del Example Set	214
Figura 246. Operador replace	214
Figura 247. Colocación de parámetros	215
Figura 248. Valores reemplazados	215
Figura 249. Selección de atributos agregando al operador select attributes	216
Figura 250. Selección de parámetros	216
Figura 251. Resultados finales	217
Figura 252. Operador read document	217
Figura 253. Selección de archivo de texto	218
Figura 254. Operador process document conectado con read documet	218
Figura 255. Operador process document	219
Figura 256. Operador "trasformer cases"	219
Figura 257. Operador "tokenize"	219
Figura 258. Operador "filter stopwords"	220
Figura 259. Operador filter tokens	220
Figura 260. Tokens basados en su longitud	220
Figura 261. Resultado del proyecto	221
Figura 262. Operador Genearte n-grams	221
Figura 263. Tokens con una longitud máxima	222
Figura 264. Parámetro string	222

Figura 265. Resultados del proyecto	223
Figura 266. Operador read csv	225
Figura 267. Ejecución del proyecto	226
Figura 268. Parámetro attribute filter type	226
Figura 269. Opción select attributes	227
Figura 270. Conversión de todos los atributos	227
Figura 271. Cambio de integer a binomial	228
Figura 272. Operador FP – Growth colocando el valor de min support	228
Figura 273. Resultados del proyecto	229
Figura 274. Operador créate association	229
Figura 275. Resultados del proyecto	230
Figura 276. Cambio del valor de confianza en el operador create association rules	230
Figura 277. Operador generate date(ExampleSetGenerator)	231
Figura 278. Operador multiply	232
Figura 279. Operador X - Validation	232
Figura 280. Operador support vector machine	232
Figura 281. Operador X-Validation conectado con el operador Multiply	233
Figura 282. Operador regresión lineal en el panel de entrenamiento	233
Figura 283. Operador ANOVA y operador T-Test	234
Figura 284. Valores medios reales probablemente diferentes en Anova Test	234
Figura 285. Operador T-Test	235
Figura 286. Operador read csv	237
Figura 287. Ejecución del proyecto	237
Figura 288. Opción estadística	238
Figura 289. Puerto mat conectado a un puerto res	238
Figura 290. Resultados de la correlation matrix	239
Figura 291. Valores de atributo en la misma dirección (correlación positiva)	239
Figura 292. Variables se mueven en sentidos contrarios (correlación negativa)	240
Figura 293. Fuerza de correlación	240
Figura 294. Colocación de parámetros	242
Figura 295. Cambio del parámetro plotter(grafico) Scatter y colocación de parámetros	242
Figura 296. Datos de entrenamiento	245
Figura 297. Selección de la columna de separación (coma)	246
Figura 298. Colocación de etiqueta label	246
Figura 299. Conexión del operador read csv con la salida res	247
Figura 300. Opción statistics	247
Figura 301. Valores que no corresponden a los datos analizados	248
Figura 302. Valores que no corresponden al campo	248

Figura 303. Filtro de valores menores o iguales que 100	249
Figura 304. Filtro de valores mayores o iguales que 3	249
Figura 305. Obtención de los 482 registros	249
Figura 306. Operador linear discriminaNt analysis para crear el modelo	250
Figura 307. Operador set rol para cambiar el valor a la función label	250
Figura 308. Resultados del campo objetivo	251
Figura 309. Tipo de separación (coma)	252
Figura 310. Operador read cvs conectado a la salida res	252
Figura 311. Valores inconsistentes en el campo de decisión making	253
Figura 312. Copia de los operadores filters examples	253
Figura 313. Operador apply model conectado a la salida res	254
Figura 314. Ejecución del proyecto con los valores creados	254
Figura 315. Opción Statistics mostrando campos con valor predicho	255
Figura 316. Operador read csv para data de entrenamiento y data de prueba	259
Figura 317. Ejecución del proyecto	259
Figura 318. Operadores set rol y selección de opciones	260
Figura 319. Operador set role conectado al modelo de redes neuronales	260
Figura 320. Operador neural net para entrenar el modelo	261
Figura 321. Generación de una red neuronal con una capa de entrada, capa oculta y capa de salida	261
Figura 322. Operador apply model con conexiones	262
Figura 323. Ejecución del proyecto con valores nuevos	262
Figura 324. Operador read excel	263
Figura 325. Selección del archivo Excel	263
Figura 326. Análisis de texto (maoxian)	264
Figura 327. Análisis de texto(stocktwits)	264
Figura 328. Análisis de texto (Neural market trends)	265
Figura 329. Análisis de texto(stocktwits)	265
Figura 330. Operador get pages	266
Figura 331. Operador data to documents generando documentos de los valores	266
Figura 332. Operador process documents creando vectores de palabras de un texto	267
Figura 333. Operador "transformer cases"	267
Figura 334. Operador "tokenize"	267
Figura 335. Operador "filters stopwords"	268
Figura 336. Operador "filter tokens"	268
Figura 337. Longitud de los tokens	268
Figura 338. Resultado al ejecutar el proyecto	269
Figura 339. Operador generate n-grams	269
Figura 340. Creación de tokens con una longitud máxima de 2	270

Figura 341. Operador filters token con el parámetro string	270
Figura 342. Selección de las palabras que contienen la cadena "-pr"	271
Figura 343. Operador data to similar	271
Figura 344. Distancia entre cada página web	272

ÍNDICE DE TABLAS

Tabla 1. Metodología Sema	32
Tabla 2. Comparación entre Crisp y semma	33
Tabla 3. Minería de Datos	64
Tabla 4. Lista de Atributos por Hipótesis Planteadas	69
Tabla 5. Reporte de técnicas y precisión usadas en la literatura	71
Tabla 6. Los resultados del valor r y la mejor precisión	74
Tabla 7. Resumen de resultados de análisis de errores	74
Tabla 8. Variables Identificadas	74
Tabla 9. Atributos de estudiantes usadas para el entrenamiento	78
Tabla 10. Atributos de estudiantes usadas para el estudio	79
Tabla 11. Predicción de retención de estudiantes en ingeniería	80
Tabla 12. Representación Y, X, W y Z	198

RESUMEN

El machine learning consiste en el análisis de datos para obtener conocimiento importante que ha sido extraído a partir de la información de grandes volúmenes de datos. El objetivo principal del libro es instruir al lector en esta disciplina y guiarlo en el desarrollo de modelos descriptivos y predictivos que faliciten la toma de decisiones en una empresa, haciendo uso de la herramienta Rapid Miner Studio, uno de los programas más usados hoy en día para este fin.

Entre los principales temas abordados en esta obra se hallan las metodologías y las técnicas de este campo, la característica del Rapid Miner Studio, las opciones de la interfaz y los pasos importantes para su instalación y configuración. Además, se proponen 24 casos prácticos que son resueltos de manera didáctica con técnicas como árboles de decisiones, redes neuronales, los clústers, las series temporales, las reglas de asociación y dependencia, la validación de datos erróneos y la integración y partición de datos.

Este libro está dirigido a todos aquellos que necesiten efectuar toma de decisiones a partir del análisis de grandes volúmenes de datos. Por tal motivo representa un aporte valioso para los estudiantes y profesionales de Ingeniería de Sistemas, Ingeniería Industrial, Estadística, Administración de Empresas, Marketing, Medicina, Educación, entre otras áreas.