

Table of Contents

Dedication	iii
Table of Contents.....	v
Acknowledgements.....	xi
SECTION ONE: Data Mining Basics	1
Chapter One: Introduction to Data Mining and CRISP-DM.....	3
Introduction.....	3
A Note About Tools	4
The Data Mining Process	5
Data Mining and You.....	11
Chapter Two: Organizational Understanding and Data Understanding.....	13
Context and Perspective	13
Learning Objectives.....	14
Purposes, Intents and Limitations of Data Mining	15
Database, Data Warehouse, Data Mart, Data Set...?	15
Types of Data	19
A Note about Privacy and Security.....	20
Chapter Summary	21
Review Questions	22
Exercises.....	22
Chapter Three: Data Preparation	25
Context and Perspective	25
Learning Objectives.....	25
Collation	27

Data Scrubbing.....	28
Hands on Exercise.....	29
Preparing RapidMiner, Importing Data, and.....	30
Handling Missing Data	30
Data Reduction	46
Handling Inconsistent Data	50
Attribute Reduction.....	52
Chapter Summary	54
Review Questions	55
Exercise	55
SECTION TWO: Data Mining Models and Methods	57
Chapter Four: Correlation	59
Context and Perspective	59
Learning Objectives.....	59
Organizational Understanding.....	59
Data Understanding	60
Data Preparation	60
Modeling.....	62
Evaluation	63
Deployment	65
Chapter Summary	67
Review Questions	68
Exercise	68
Chapter Five: Association Rules	73
Context and Perspective	73
Learning Objectives.....	73
Organizational Understanding.....	73

Data Understanding	74
Data Preparation	76
Modeling.....	81
Evaluation	84
Deployment	87
Chapter Summary	87
Review Questions	88
Exercise	88
Chapter Six: k-Means Clustering	91
Context and Perspective	91
Learning Objectives.....	91
Organizational Understanding.....	91
Data UnderstanDing	92
Data Preparation	92
Modeling.....	94
Evaluation	96
Deployment	98
Chapter Summary	101
Review Questions	101
Exercise	102
Chapter Seven: Discriminant Analysis.....	105
Context and Perspective	105
Learning Objectives.....	105
Organizational Understanding.....	106
Data Understanding	106
Data Preparation	109
Modeling	114

Evaluation	118
Deployment	120
Chapter Summary	121
Review Questions	122
Exercise	123
Chapter Eight: Linear Regression	127
Context and Perspective	127
Learning Objectives.....	127
Organizational Understanding.....	128
Data Understanding	128
Data Preparation	129
Modeling	131
Evaluation	132
Deployment	134
Chapter Summary	137
Review Questions	137
Exercise	138
Chapter Nine: Logistic Regression.....	141
Context and Perspective	141
Learning Objectives.....	141
Organizational Understanding.....	142
Data Understanding	142
Data Preparation	143
Modeling	147
Evaluation	148
Deployment	151
Chapter Summary	153

Review Questions	153
Exercise	154
Chapter Ten: Decision Trees	157
Context and Perspective	157
Learning Objectives.....	157
Organizational Understanding.....	158
Data Understanding	159
Data Preparation	161
Modeling	166
Evaluation.....	169
Deployment	171
Chapter Summary	172
Review Questions	172
Exercise	173
Chapter Eleven: Neural Networks	175
Context and Perspective	175
Learning Objectives.....	175
Organizational Understanding.....	175
Data Understanding	176
Data Preparation	178
Modeling	181
Evaluation	181
Deployment	184
Chapter Summary	186
Review Questions	186
Exercise	187
Chapter Twelve: Text Mining	189

Context and Perspective	189
Learning Objectives.....	189
Organizational Understanding.....	190
Data Understanding	190
Data Preparation.....	191
Modeling	202
Evaluation	203
Deployment	213
Chapter Summary	213
Review Questions	214
Exercise	214
SECTION THREE: Special Considerations in Data Mining	217
Chapter Thirteen: Evaluation and Deployment.....	219
How Far We've Come	219
Learning Objectives.....	220
Cross-Validation	221
Chapter Summary: The Value of Experience.....	227
Review Questions	228
Exercise	228
Chapter Fourteen: Data Mining Ethics	231
Why Data Mining Ethics?	231
Ethical Frameworks and Suggestions.....	233
Conclusion	235
GLOSSARY and INDEX.....	237
About the Author	251