

Contents

Preface	ix
About This Book	xi
About These Authors	xv
Acknowledgments	xvii
Chapter 1: Decision Trees—What Are They?	1
Introduction	1
Using Decision Trees with Other Modeling Approaches.....	4
Why Are Decision Trees So Useful?	6
Level of Measurement	9
Chapter 2: Descriptive, Predictive, and Explanatory Analyses	15
Introduction	16
The Importance of Showing Context	17
Antecedents	18
Intervening Factors.....	19
A Classic Study and Illustration of the Need to Understand Context	19
The Effect of Context	21
How Do Misleading Results Appear?.....	22
Automatic Interaction Detection	24
The Role of Validation and Statistics in Growing Decision Trees	28
The Application of Statistical Knowledge to Growing Decision Trees.....	30
Significance Tests	30
Validation to Determine Tree Size and Quality.....	34
What Is Validation?.....	35
Pruning.....	38

Machine Learning, Rule Induction, and Statistical Decision Trees	43
Rule Induction.....	44
Rule Induction and the Work of Ross Quinlan	48
The Use of Multiple Trees	49
A Review of the Major Features of Decision Trees.....	50
Roots and Trees	50
Branches.....	50
Similarity Measures	51
Recursive Growth.....	51
Shaping the Decision Tree.....	51
Deploying Decision Trees	51
A Brief Review of the SAS Enterprise Miner ARBORETUM Procedure	52
Chapter 3: The Mechanics of Decision Tree Construction.....	55
The Basics of Decision Trees.....	55
Step 1—Preprocess the Data for the Decision Tree Growing Engine.....	57
Step 2—Set the Input and Target Modeling Characteristics.....	59
Targets.....	60
Inputs.....	61
Step 3—Select the Decision Tree Growth Parameters.....	64
Step 4—Cluster and Process Each Branch-Forming Input Field.....	66
Clustering Algorithms	69
The Kass Merge-and-Split Heuristic	76
Dealing with Missing Data and Missing Inputs in Decision Trees	76
Step 5—Select the Candidate Decision Tree Branches	79
Step 6—Complete the Form and Content of the Final Decision Tree	94
Switching Targets	105
Example of Multiple Target Selection Using the Home Equity Demonstration Data.....	106
Synergy, Functionality, and the Wisdom of the End User.....	114
Chapter 4: Business Intelligence and Decision Trees.....	117
Introduction	117
A Decision Tree Approach to Cube Construction	120
Multidimensional Cubes and Decision Trees Compared: A Small Business Example	121
Multidimensional Cubes and Decision Trees: A Side-By-Side Comparison	126
The Main Difference between Decision Trees and Multidimensional Cubes.....	128
Regression as a Business Tool	128
Decision Trees and Regression Compared	129

Multidimensional Analysis with Trees.....	134
An Example with Multiple Targets	135
Chapter 5: Theoretical Issues in the Decision Tree Growing Process.....	145
Introduction	145
Crafting the Decision Tree Structure for Insight and Exposition.....	146
Conceptual Model.....	147
Predictive Issues: Accuracy, Reliability, Reproducibility, and Performance.....	154
Choosing the Right Number of Branches.....	157
Perspectives on Selection Bias	158
Potential Remedies to Variable Selection Bias	159
Multiple Decision Trees	171
Ensembles.....	172
Chapter 6: The Integration of Decision Trees with Other Data Mining	
Approaches	187
Introduction	187
Decision Trees in Stratified Regression.....	188
Time-Ordered Data	189
Decision Trees in Forecasting Applications.....	191
Decision Trees in Variable Selection.....	194
Decision Tree Results.....	197
Interactions.....	198
Cross-Contributions of Decision Trees and Other Approaches	199
Decision Trees in Analytical Model Development.....	200
The Use of Decision Trees in Rule Induction	204
Iterative Removal of Observations	205
Conclusion	216
Business Intelligence.....	217
Data Mining	217
Glossary	219
References	233
Index	239

viii *Contents*