

Contents

	About This Book	ix
	Contributing Authors	xi
CHAPTER 1	What's It All About?	1
1.1	Data Mining and Machine Learning.....	1
1.2	Simple Examples: The Weather Problem and Others	7
1.3	Fielded Applications	20
1.4	Machine Learning and Statistics	27
1.5	Generalization as Search.....	28
1.6	Data Mining and Ethics.....	32
1.7	Resources	34
CHAPTER 2	Data Acquisition and Integration	37
2.1	Introduction	37
2.2	Sources of Data	37
2.3	Variable Types.....	39
2.4	Data Rollup	41
2.5	Rollup with Sums, Averages, and Counts	48
2.6	Calculation of the Mode	49
2.7	Data Integration	50
CHAPTER 3	Data Preprocessing	57
3.1	Why Preprocess the Data?.....	58
3.2	Descriptive Data Summarization	61
3.3	Data Cleaning.....	72
3.4	Data Integration and Transformation	78
3.5	Data Reduction	84
3.6	Data Discretization and Concept Hierarchy Generation	98

3.7	Summary	108
3.8	Resources	109
CHAPTER 4	Physical Design for Decision Support, Warehousing, and OLAP	113
4.1	What Is Online Analytical Processing?	113
4.2	Dimension Hierarchies	116
4.3	Star and Snowflake Schemas	117
4.4	Warehouses and Marts	119
4.5	Scaling Up the System	122
4.6	DSS, Warehousing, and OLAP Design Considerations	124
4.7	Usage Syntax and Examples for Major Database Servers	125
4.8	Summary	128
4.9	Literature Summary	129
	Resources	129
CHAPTER 5	Algorithms: The Basic Methods	131
5.1	Inferring Rudimentary Rules	132
5.2	Statistical Modeling	136
5.3	Divide and Conquer: Constructing Decision Trees	144
5.4	Covering Algorithms: Constructing Rules	153
5.5	Mining Association Rules	160
5.6	Linear Models	168
5.7	Instance-Based Learning	176
5.8	Clustering	184
5.9	Resources	188
CHAPTER 6	Further Techniques in Decision Analysis	191
6.1	Modeling Risk Preferences	191
6.2	Analyzing Risk Directly	198
6.3	Dominance	200
6.4	Sensitivity Analysis	205
6.5	Value of Information	215
6.6	Normative Decision Analysis	220
CHAPTER 7	Fundamental Concepts of Genetic Algorithms	221
7.1	The Vocabulary of Genetic Algorithms	222
7.2	Overview	230
7.3	The Architecture of a Genetic Algorithm	241
7.4	Practical Issues in Using a Genetic Algorithm	285

	7.5	Review	290
	7.6	Resources	290
CHAPTER 8		Data Structures and Algorithms for Moving Objects Types	293
	8.1	Data Structures.....	293
	8.2	Algorithms for Operations on Temporal Data Types	298
	8.3	Algorithms for Lifted Operations.....	310
	8.4	Resources	319
CHAPTER 9		Improving the Model.....	321
	9.1	Learning from Errors.....	323
	9.2	Improving Model Quality, Solving Problems	343
	9.3	Summary	395
CHAPTER 10		Social Network Analysis.....	397
	10.1	Social Sciences and Bibliometry	398
	10.2	PageRank and Hyperlink-Induced Topic Search	400
	10.3	Shortcomings of the Coarse-Grained Graph Model	410
	10.4	Enhanced Models and Techniques.....	416
	10.5	Evaluation of Topic Distillation	424
	10.6	Measuring and Modeling the Web	430
	10.7	Resources	440
		Index	443