

Contents

Part I INTRODUCTION	
<hr/>	
1 Introduction.....	3
What is Data Mining?	5
What is Needed to Do Data Mining.....	5
Business Data Mining.....	7
Data Mining Tools	8
Summary.....	8
2 Data Mining Process.....	9
CRISP-DM	9
Business Understanding.....	11
Data Understanding	11
Data Preparation	12
Modeling	15
Evaluation	18
Deployment.....	18
SEMMA.....	19
Steps in SEMMA Process.....	20
Example Data Mining Process Application.....	22
Comparison of CRISP & SEMMA.....	27
Handling Data.....	28
Summary.....	34
<hr/>	
Part II DATA MINING METHODS AS TOOLS	
<hr/>	
3 Memory-Based Reasoning Methods.....	39
Matching	40
Weighted Matching.....	43
Distance Minimization.....	44
Software	50
Summary.....	50
Appendix: Job Application Data Set.....	51

4 Association Rules in Knowledge Discovery.....	53
Market-Basket Analysis.....	55
Market Basket Analysis Benefits.....	56
Demonstration on Small Set of Data	57
Real Market Basket Data	59
The Counting Method Without Software	62
Conclusions.....	68
5 Fuzzy Sets in Data Mining.....	69
Fuzzy Sets and Decision Trees	71
Fuzzy Sets and Ordinal Classification	75
Fuzzy Association Rules.....	79
Demonstration Model	80
Computational Results.....	84
Testing	84
Inferences.....	85
Conclusions.....	86
6 Rough Sets	87
A Brief Theory of Rough Sets	88
Information System.....	88
Decision Table	89
Some Exemplary Applications of Rough Sets.....	91
Rough Sets Software Tools.....	93
The Process of Conducting Rough Sets Analysis.....	93
1 Data Pre-Processing.....	94
2 Data Partitioning.....	95
3 Discretization.....	95
4 Reduct Generation	97
5 Rule Generation and Rule Filtering.....	99
6 Apply the Discretization Cuts to Test Dataset.....	100
7 Score the Test Dataset on Generated Rule set (and measuring the prediction accuracy)	100
8 Deploying the Rules in a Production System	102
A Representative Example.....	103
Conclusion	109
7 Support Vector Machines	111
Formal Explanation of SVM.....	112
Primal Form	114

Dual Form	114
Soft Margin	114
Non-linear Classification	115
Regression.....	116
Implementation	116
Kernel Trick	117
Use of SVM – A Process-Based Approach	118
Support Vector Machines versus Artificial Neural Networks	121
Disadvantages of Support Vector Machines.....	122
8 Genetic Algorithm Support to Data Mining	125
Demonstration of Genetic Algorithm	126
Application of Genetic Algorithms in Data Mining	131
Summary	132
Appendix: Loan Application Data Set.....	133
9 Performance Evaluation for Predictive Modeling	137
Performance Metrics for Predictive Modeling	137
Estimation Methodology for Classification Models	140
Simple Split (Holdout).....	140
The k -Fold Cross Validation.....	141
Bootstrapping and Jackknifing	143
Area Under the ROC Curve.....	144
Summary	147

Part III APPLICATIONS

10 Applications of Methods.....	151
Memory-Based Application.....	151
Association Rule Application	153
Fuzzy Data Mining	155
Rough Set Models.....	155
Support Vector Machine Application	157
Genetic Algorithm Applications.....	158
Japanese Credit Screening	158
Product Quality Testing Design.....	159
Customer Targeting	159
Medical Analysis	160

Predicting the Financial Success of Hollywood Movies	162
Problem and Data Description	163
Comparative Analysis of the Data Mining Methods	165
Conclusions.....	167
Bibliography	169
Index	177



<http://www.springer.com/978-3-540-76916-3>

Advanced Data Mining Techniques

Olson, D.L.; Delen, D.

2008, XII, 180 p. 21 illus., Softcover

ISBN: 978-3-540-76916-3