

Data Mining and Predictive Analysis

INTELLIGENCE GATHERING
AND CRIME ANALYSIS



Colleen McCue

Contents

Foreword	xiii
Preface	xv
Introduction	xxv
Introductory Section	I
I Basics	3
1.1 Basic Statistics	3
1.2 Inferential versus Descriptive Statistics and Data Mining	4
1.3 Population versus Samples	4
1.4 Modeling	6
1.5 Errors	7
1.6 Overfitting the Model	14
1.7 Generalizability versus Accuracy	14
1.8 Input/Output	17
1.9 Bibliography	18
2 Domain Expertise	19
2.1 Domain Expertise	19
2.2 Domain Expertise for Analysts	20
2.3 Compromise	22
2.4 Analyze Your Own Data	24
2.5 Bibliography	24
3 Data Mining	25
3.1 Discovery and Prediction	27
3.2 Confirmation and Discovery	28
3.3 Surprise	30

3.4	Characterization	31
3.5	“Volume Challenge”	32
3.6	Exploratory Graphics and Data Exploration	33
3.7	Link Analysis	37
3.8	Nonobvious Relationship Analysis (NORA)	37
3.9	Text Mining	39
3.10	Future Trends	40
3.11	Bibliography	40
Methods		43
4	Process Models for Data Mining and Analysis	45
4.1	CIA Intelligence Process	47
4.2	CRISP-DM	49
4.3	Actionable Mining and Predictive Analysis for Public Safety and Security	53
4.4	Bibliography	65
5	Data	67
5.1	Getting Started	69
5.2	Types of Data	69
5.3	Data	70
5.4	Types of Data Resources	71
5.5	Data Challenges	82
5.6	How Do We Overcome These Potential Barriers?	87
5.7	Duplication	88
5.8	Merging Data Resources	89
5.9	Public Health Data	90
5.10	Weather and Crime Data	90
5.11	Bibliography	91
6	Operationally Relevant Preprocessing	93
6.1	Operationally Relevant Recoding	93
6.2	Trinity Sight	94
6.3	Duplication	100
6.4	Data Imputation	100
6.5	Telephone Data	101

6.6	Conference Call Example	103
6.7	Internet Data	110
6.8	Operationally Relevant Variable Selection	111
6.9	Bibliography	114
7	Predictive Analytics	117
7.1	How to Select a Modeling Algorithm, Part I	117
7.2	Generalizability versus Accuracy	118
7.3	Link Analysis	119
7.4	Supervised versus Unsupervised Learning Techniques	119
7.5	Discriminant Analysis	121
7.6	Unsupervised Learning Algorithms	122
7.7	Neural Networks	123
7.8	Kohonen Network Models	125
7.9	How to Select a Modeling Algorithm, Part II	125
7.10	Combining Algorithms	126
7.11	Anomaly Detection	127
7.12	Internal Norms	127
7.13	Defining “Normal”	128
7.14	Deviations from Normal Patterns	130
7.15	Deviations from Normal Behavior	130
7.16	Warning! Screening versus Diagnostic	132
7.17	A Perfect World Scenario	133
7.18	Tools of the Trade	135
7.19	General Considerations and Some Expert Options	137
7.20	Variable Entry	138
7.21	Prior Probabilities	138
7.22	Costs	139
7.23	Bibliography	141
8	Public Safety—Specific Evaluation	143
8.1	Outcome Measures	144
8.2	Think Big	149
8.3	Training and Test Samples	153
8.4	Evaluating the Model	158
8.5	Updating or Refreshing the Model	161
8.6	Caveat Emptor	162
8.7	Bibliography	163

9 Operationally Actionable Output	165
9.1 Actionable Output	165
 Applications	 175
10 Normal Crime	177
10.1 Knowing Normal	178
10.2 "Normal" Criminal Behavior	181
10.3 Get to Know "Normal" Crime Trends and Patterns	182
10.4 Staged Crime	183
10.5 Bibliography	184
 11 Behavioral Analysis of Violent Crime	 187
11.1 Case-Based Reasoning	193
11.2 Homicide	196
11.3 Strategic Characterization	199
11.4 Automated Motive Determination	203
11.5 Drug-Related Violence	205
11.6 Aggravated Assault	205
11.7 Sexual Assault	206
11.8 Victimology	208
11.9 Moving from Investigation to Prevention	211
11.10 Bibliography	211
 12 Risk and Threat Assessment	 215
12.1 Risk-Based Deployment	217
12.2 Experts versus Expert Systems	218
12.3 "Normal" Crime	219
12.4 Surveillance Detection	219
12.5 Strategic Characterization	220
12.6 Vulnerable Locations	222
12.7 Schools	223
12.8 Data	227
12.9 Accuracy versus Generalizability	228
12.10 "Cost" Analysis	229
12.11 Evaluation	229

12.12 Output	231
12.13 Novel Approaches to Risk and Threat Assessment	232
12.14 Bibliography	234
Case Examples	237
13 Deployment	239
13.1 Patrol Services	240
13.2 Structuring Patrol Deployment	240
13.3 Data	241
13.4 How To	246
13.5 Tactical Deployment	250
13.6 Risk-Based Deployment Overview	251
13.7 Operationally Actionable Output	252
13.8 Risk-Based Deployment Case Studies	259
13.9 Bibliography	265
14 Surveillance Detection	267
14.1 Surveillance Detection and Other Suspicious Situations	267
14.2 Natural Surveillance	270
14.3 Location, Location, Location	275
14.4 More Complex Surveillance Detection	282
14.5 Internet Surveillance Detection	289
14.6 How To	294
14.7 Summary	296
14.8 Bibliography	297
Advanced Concepts and Future Trends	299
15 Advanced Topics	301
15.1 Intrusion Detection	301
15.2 Identify Theft	302
15.3 Syndromic Surveillance	303
15.4 Data Collection, Fusion and Preprocessing	303
15.5 Text Mining	306
15.6 Fraud Detection	308

15.7 Consensus Opinions	310
15.8 Expert Options	311
15.9 Bibliography	312
16 Future Trends	315
16.1 Text Mining	315
16.2 Fusion Centers	317
16.3 “Functional” Interoperability	318
16.4 “Virtual” Warehouses	318
16.5 Domain-Specific Tools	319
16.6 Closing Thoughts	319
16.7 Bibliography	321
Index	323