Contents

List of Figures    xv

Foreword   xxiii

Preface   xxv

Acknowledgments    xxix

Chapter 1 Fraud: Detection, Prevention, and Analytics!    1

   Introduction    2
   Fraud!    2
   Fraud Detection and Prevention    10
   Big Data for Fraud Detection    15
   Data-Driven Fraud Detection    17
   Fraud-Detection Techniques    19
   Fraud Cycle    22
   The Fraud Analytics Process Model    26
   Fraud Data Scientists    30
      A Fraud Data Scientist Should Have Solid Quantitative Skills    30
      A Fraud Data Scientist Should Be a Good Programmer    31
      A Fraud Data Scientist Should Excel in Communication and Visualization Skills    31
      A Fraud Data Scientist Should Have a Solid Business Understanding    32
      A Fraud Data Scientist Should Be Creative    32
      A Scientific Perspective on Fraud    33
   References    35

Chapter 2 Data Collection, Sampling, and Preprocessing    37

   Introduction    38
   Types of Data Sources    38
   Merging Data Sources    43
   Sampling    45
   Types of Data Elements    46
Visual Data Exploration and Exploratory Statistical Analysis 47
Benford's Law 48
Descriptive Statistics 51
Missing Values 52
Outlier Detection and Treatment 53
Red Flags 57
Standardizing Data 59
Categorization 60
Weights of Evidence Coding 63
Variable Selection 65
Principal Components Analysis 68
RIDITs 72
PRIDIT Analysis 73
Segmentation 74
References 75

Chapter 3 Descriptive Analytics for Fraud Detection 77
Introduction 78
Graphical Outlier Detection Procedures 79
Statistical Outlier Detection Procedures 83
  Break-Point Analysis 84
  Peer-Group Analysis 85
  Association Rule Analysis 87
Clustering 89
  Introduction 89
  Distance Metrics 90
  Hierarchical Clustering 94
  Example of Hierarchical Clustering Procedures 97
  k-Means Clustering 104
  Self-Organizing Maps 109
  Clustering with Constraints 111
  Evaluating and Interpreting Clustering Solutions 114
One-Class SVMs 117
References 118

Chapter 4 Predictive Analytics for Fraud Detection 121
Introduction 122
Target Definition 123
Linear Regression 125
Logistic Regression 127
  Basic Concepts 127
  Logistic Regression Properties 129
  Building a Logistic Regression Scorecard 131
Chapter 5  Social Network Analysis for Fraud Detection  207
Networks: Form, Components, Characteristics, and Their Applications  209
Social Networks  211
Network Components  214
Network Representation  219
Is Fraud a Social Phenomenon? An Introduction to Homophily  222
Impact of the Neighborhood: Metrics  227
Neighborhood Metrics  228
Centrality Metrics  238
Collective Inference Algorithms  246
Featurization: Summary Overview  254
Community Mining: Finding Groups of Fraudsters  254
Extending the Graph: Toward a Bipartite Representation  266
Multipartite Graphs  269
Case Study: Gotcha!  270
References  277

Chapter 6  Fraud Analytics: Post-Processing  279
Introduction  280
The Analytical Fraud Model Life Cycle  280
Model Representation  281
Traffic Light Indicator Approach  282
Decision Tables  283
Selecting the Sample to Investigate  286
Fraud Alert and Case Management  290
Visual Analytics  296
Backtesting Analytical Fraud Models  302
Introduction  302
Backtesting Data Stability  302
Backtesting Model Stability  305
Backtesting Model Calibration  308
Model Design and Documentation  311
References  312

Chapter 7  Fraud Analytics: A Broader Perspective  313
Introduction  314
Data Quality  314
Data-Quality Issues  314
Data-Quality Programs and Management  315
Privacy  317
The RACI Matrix  318
Accessing Internal Data  319
CONTENTS

Label-Based Access Control (LBAC) 324
Accessing External Data 325
Capital Calculation for Fraud Loss 326
   Expected and Unexpected Losses 327
   Aggregate Loss Distribution 329
   Capital Calculation for Fraud Loss Using Monte Carlo Simulation 331
An Economic Perspective on Fraud Analytics 334
   Total Cost of Ownership 334
   Return on Investment 335
In Versus Outsourcing 337
Modeling Extensions 338
   Forecasting 338
   Text Analytics 340
The Internet of Things 342
Corporate Fraud Governance 344
References 346

About the Authors 347

Index 349