

Table of Contents

Introduction 1

About This Book	1
Foolish Assumptions	2
Icons Used in This Book	2
Beyond the Book	3
Where to Go from Here	3

Part I: Getting Started with Predictive Analytics 5

Chapter 1: Entering the Arena 7

Exploring Predictive Analytics	7
Mining data	8
Highlighting the model	9
Adding Business Value	10
Endless opportunities	11
Empowering your organization	11
Starting a Predictive Analytic Project	13
Business knowledge	13
Data-science team and technology	14
The Data	15
Surveying the Marketplace	16
Responding to big data	16
Working with big data	17

Chapter 2: Predictive Analytics in the Wild 19

Online Marketing and Retail	21
Recommender systems	21
Implementing a Recommender System	23
Collaborative filtering	23
Content-based filtering	32
Hybrid recommender systems	36
Target Marketing	36
Targeting using predictive modeling	37
Uplift modeling	39
Predictive Analytics Fight Fraud and Crime	41
Content and Text Analytics	42

Chapter 3: Exploring Your Data Types and Associated Techniques	43
Recognizing Your Data Types	44
Structured and unstructured data.....	44
Static and streamed data	46
Identifying Data Categories	47
Attitudinal data	49
Behavioral data	49
Demographic data.....	50
Generating Predictive Analytics	50
Data-driven analytics.....	51
User-driven analytics.....	52
Connecting to Related Disciplines.....	53
Statistics.....	54
Data mining.....	54
Machine learning.....	55
Chapter 4: Complexities of Data	57
Finding Value in Your Data.....	58
Delving into your data.....	58
Data validity.....	59
Data variety	59
Constantly/Changing Data	60
Data velocity.....	60
High volume of data.....	60
Complexities in Searching Your Data	61
Keyword-based search.....	61
Semantic-based search	62
Differentiating Business Intelligence from Big-Data Analytics.....	63
Visualization of Raw Data	64
Identifying data attributes	64
Exploring data visualization	65
Tabular visualizations	66
Bar charts	67
Pie charts	67
Graph charts.....	68
Word clouds as representations.....	69
Line graphs	70
Flocking birds representation	70
Part II: Incorporating Algorithms in Your Models	73
Chapter 5: Applying Models	75
Modeling Data	76
Models and simulation.....	77
Categorizing models	79

Describing and summarizing data	80
Making better business decisions	81
Healthcare Analytics Case Studies	81
Google search queries as epidemic predictors	81
Cancer survivability predictors	82
Social and Marketing Analytics Case Studies	83
Tweets as predictors for the stock market	84
Target store predicts pregnant women	85
Twitter-based predictors of earthquakes	86
Twitter-based predictors of political campaign outcomes	87
Chapter 6: Identifying Similarities in Data	89
Explaining Data Clustering	90
Motivation.....	91
Converting Raw Data into a Matrix	93
Creating a matrix of terms in documents	93
Term selection	94
Identifying K-Groups in Your Data	95
K-means clustering algorithm	95
Clustering by nearest neighbors	100
Finding Associations Among Data Items	104
Applying Biologically Inspired Clustering Techniques.....	105
Birds flocking.....	106
Ant colonies.....	111
Chapter 7: Predicting the Future Using Data Classification	115
Explaining Data Classification.....	116
Lending.....	117
Marketing.....	118
Healthcare.....	119
What's next?	119
Introducing Data Classification to Your Business.....	119
Exploring the Data-Classification Process.....	122
Using Data Classification to Predict the Future.....	123
Decision trees.....	123
Support vector machine	126
Naïve Bayes classification algorithm	128
Neural networks.....	134
The Markov Model.....	136
Linear regression	141
Ensemble Methods to Boost Prediction Accuracy	142

Part III: Developing a Roadmap..... 145

Chapter 8: Convincing Your Management to Adopt Predictive Analytics	147
Making the Business Case	148
Benefits to the business	149
Gathering Support from Stakeholders	155
Working with your sponsors	156
Getting business and operations buy-in	158
Getting IT buy-in.....	160
Rapid prototyping.....	164
Presenting Your Proposal	164
Chapter 9: Preparing Data	167
Listing the Business Objectives	168
Identifying related objectives.....	168
Collecting user requirements.....	169
Processing Your Data.....	169
Identifying the data.....	170
Cleaning the data	171
Generating any derived data	172
Reducing the dimensionality of your data.....	173
Structuring Your Data	174
Extracting, transforming and loading your data.....	174
Keeping the data up to date	175
Outlining testing and test data.....	176
Chapter 10: Building a Predictive Model	177
Getting Started	178
Defining your business objectives.....	179
Preparing your data.....	180
Choosing an algorithm	182
Developing and Testing the Model.....	183
Developing the model	184
Testing the model	184
Evaluating the model.....	186
Going Live with the Model.....	187
Deploying the model	187
Monitoring and maintaining the model.....	188
Chapter 11: Visualization of Analytical Results.....	189
Visualization As a Predictive Tool.....	189
Why visualization matters	190
Getting the benefits of visualization.....	191
Dealing with complexities.....	192

Evaluating Your Visualization	193
How relevant is this picture?	193
How interpretable is the picture?	193
Is the picture simple enough?	194
Does the picture lead to new insights?	194
Visualizing Your Model's Analytical Results.....	194
Visualizing hidden groupings in your data.....	195
Visualizing data classification results	196
Visualizing outliers in your data	197
Visualization of Decision Trees.....	197
Visualizing predictions.....	199
Other Types of Visualizations in Predictive Analytics	200
Bird-flocking behavior data visualization	201

Part IV: Programming Predictive Analytics..... 205

Chapter 12: Creating Basic Prediction Examples 207

Installing the Software Packages	207
Installing Python	208
Installing the machine-learning module.....	208
Installing the dependencies.....	212
Preparing the Data.....	216
Getting the sample dataset.....	216
Labeling your data	216
Making Predictions Using Classification Algorithms	218
Creating a supervised learning model with SVM.....	218
Creating a supervised learning model with logistic regression...	227
Comparing two classification models	231

Chapter 13: Creating Basic Examples of Unsupervised Predictions 233

Getting the Sample Dataset	233
Using Clustering Algorithms to Make Predictions	234
Comparing two clustering models.....	235
Creating an unsupervised learning model with K-means	236
Creating an unsupervised learning model with DBSCAN	245

Chapter 14: Predictive Modeling with R 249

Programming in R.....	250
Installing R	251
Installing RStudio	251
Getting familiar with the environment	251
Learning just a bit of R	253



Making Predictions Using R	257
Predicting using regression.....	257
Using classification to predict.....	266
Chapter 15: Avoiding Analysis Traps.....	275
Data Challenges	276
Outlining the limitations of the data	277
Dealing with extreme cases (outliers)	279
Data smoothing.....	282
Curve fitting.....	286
Keeping the assumptions to a minimum	289
Analysis Challenges.....	290
Supervised analytics	290
Relying on only one analysis	291
Describing the limitations of the model.....	292
Avoiding non-scalable models	293
Scoring your predictions accurately.....	294
Chapter 16: Targeting Big Data.....	295
Major Technological Trends in Predictive Analytics.....	296
Exploring predictive analytics as a service.....	296
Aggregating distributed data for analysis	297
Real-time data-driven analytics.....	298
Applying Open-Source Tools to Big Data	299
Apache Hadoop.....	299
Apache Mahout	302
Building a Rapid Prototype of Your Predictive Analytics Model	303
Prototyping for predictive analytics	303
Testing your predictive analytics model.....	306
<i>Part V: The Part of Tens</i>	<i>307</i>
Chapter 17: Ten Reasons to Implement Predictive Analytics.....	309
Outlining Business Goals	310
Knowing Your Data.....	310
Organizing Your Data	311
Satisfying Your Customers	312
Reducing Operational Costs.....	314
Increasing Returns on Investments (ROI)	314
Increasing Confidence.....	315
Making Informed Decisions	316
Gaining Competitive Edge	317
Improving the Business	318

Chapter 18: Ten Steps to Build a Predictive Analytic Model	319
Building a Predictive Analytics Team	320
Getting business expertise on board	320
Firing up IT and math expertise.....	321
Setting the Business Objectives	322
Preparing Your Data.....	322
Sampling Your Data.....	323
Avoiding “Garbage In, Garbage Out”	324
Keeping it simple isn’t stupid.....	324
Data preparation puts the good stuff in.....	324
Creating Quick Victories.....	325
Fostering Change in Your Organization.....	326
Building Deployable Models	327
Evaluating Your Model.....	328
Updating Your Model.....	329
 Index	 331