

Contents

Acknowledgments — VII

About the Author — IX

Introduction — 1

Part I: Big Data Analytics

Chapter 1

Data Analytics Overview — 13

- 1.1 Data Analytics Definition — 13**
- 1.2 The Distinction between BI and Analytics — 14**
- 1.3 Why Advanced Data Analytics? — 16**
- 1.4 Analytics Platform Framework — 17**
- 1.5 Data Connection Layer — 19**
- 1.6 Data Management Layer — 20**
- 1.7 Analytics Layer — 25**
- 1.8 Presentation Layer — 29**
- 1.9 Data Analytics Process — 30**

Chapter 2

Basic Data Analysis — 33

- 2.1 KPIs, Analytics and Business Optimization — 33**
- 2.2 Key Considerations in Data Analytics Reports — 34**
- 2.3 The Four Pillars of a Real World Data Analytics Program — 35**
- 2.4 The Eight Axioms of Big Data Analytics — 39**
- 2.5 Basic Models — 41**
- 2.6 Complexity of Data Analytics — 42**
- 2.7 Introduction to Data Analytics Methods — 43**
- 2.8 Statistical Models — 44**
- 2.9 Predictive Analytics — 45**
- 2.10 Advanced Analytics Methods — 45**

Chapter 3

Data Analytics Process — 49

- 3.1 A Survey of Data Analytics Process — 49**
- 3.2 KDD—Knowledge Discovery Databases — 52**
- 3.3 CRISP-DM Process Model — 54**
- 3.4 The SEMMA Process Model — 56**

- 3.5 Microsoft TDSP Framework — 57**
- 3.6 Data Analytics Process Example—Predictive Modeling Case Study — 59**

Part II: Advanced Analytics Methods

Chapter 4

Natural Language Processing — 65

- 4.1 Natural Language Processing (NLP) — 65**
- 4.2 NLP Capability Maturity Model — 69**
- 4.3 Introduction to Natural Language Processing — 70**
- 4.4 NLP Techniques—Topic Modeling — 72**
- 4.5 NLP—Names Entity Recognition (NER) — 73**
- 4.6 NLP—Part of Speech (POS) Tagging — 74**
- 4.7 NLP—Probabilistic Context-Free Grammars (PCFG) — 77**
- 4.8 NLP Learning Method — 78**
- 4.9 Word Embedding and Neural Networks — 79**
- 4.10 Semantic Modeling Using Graph Analysis Technique — 79**
- 4.11 Putting It All Together — 82**

Chapter 5

Quantitative Analysis—Prediction and Prognostics — 85

- 5.1 Probabilities and Odds Ratio — 86**
- 5.2 Additive Interaction of Predictive Variables — 87**
- 5.3 Prognostics and Prediction — 87**
- 5.4 Framework for Prognostics, Prediction and Accuracy — 88**
- 5.5 Significance of Predictive Analytics — 89**
- 5.6 Prognostics in Literature — 89**
- 5.7 Control Theoretic Approach to Prognostics — 91**
- 5.8 Artificial Neural Networks — 94**

Chapter 6

Advanced Analytics and Predictive Modeling — 97

- 6.1 History of Predictive Methods and Prognostics — 97**
- 6.2 Model Viability and Validation Methods — 99**
- 6.3 Classification Methods — 100**
- 6.4 Traditional Analysis Methods vs. Advanced Analytics Methods — 100**
- 6.5 Traditional Analysis Overview: Quantitative Methods — 101**
- 6.6 Regression Analysis Overview — 101**
- 6.7 Cox Hazard Model — 103**
- 6.8 Correlation Analysis — 104**

6.9	Non-linear Correlation —	107
6.10	Kaplan-Meier Estimate of Survival Function —	107
6.11	Handling Dirty, Noisy and Missing Data —	109
6.12	Data Cleansing Techniques —	111
6.13	Analysis of Variance (ANOVA) and MANOVA —	115
6.14	Advanced Analytics Methods At-a-Glance —	116
6.15	LASSO, L1 and L2 Norm Methods —	117
6.16	Kalman Filtering —	118
6.17	Trajectory Tracking —	118
6.18	N-point Correlation —	118
6.19	Bi-partite Matching —	119
6.20	Mean Shift and K-means Algorithm —	120
6.21	Gaussian Graphical Model —	120
6.22	Parametric vs. Non-parametric Methods —	121
6.23	Non-parametric Bayesian Classifier —	122
6.24	Machine Learning —	123
6.25	Geo-spatial Analysis —	123
6.26	Logistic Regression or Logit —	125
6.27	Predictive Modeling Approaches —	125
6.28	Alternate Conditional Expectation (ACE) —	126
6.29	Clustering vs. Classification —	126
6.30	K-means Clustering Method —	127
6.31	Classification Using Neural Networks —	128
6.32	Principal Component Analysis —	129
6.33	Stratification Method —	130
6.34	Propensity Score Matching Approach —	131
6.35	Adherence Analysis Method —	133
6.36	Meta-analysis Methods —	133
6.37	Stochastic Models—Markov Chain Analysis —	134
6.38	Handling Noisy Data—Kalman Filters —	135
6.39	Tree-based Analysis —	135
6.40	Random Forest Techniques —	137
6.41	Hierarchical Clustering Analysis (HCA) Method —	141
6.42	Outlier Detection by Robust Estimation Method —	144
6.43	Feature Selection Techniques —	144
6.44	Bridging Studies —	145
6.45	Signal Boosting and Bagging Methods —	145
6.46	Generalized Estimating Equation (GEE) Method —	146
6.47	Q-Q Plots —	146
6.48	Reduction in Variance (RIV) —Intergroup Variation —	146
6.49	Coefficient of Variation (CV)—Intra Group Variation —	147

Chapter 7

Ensemble of Models: Data Analytics Prediction Framework — 149

- 7.1 Ensemble of Models — 149
- 7.2 Artificial Neural Network Models — 150
- 7.3 Analytic Model Comparison and Evaluation — 151

Chapter 8

Machine Learning, Deep Learning—Artificial Neural Networks — 155

- 8.1 Introduction to ANNs — 155
- 8.2 A Simple Example — 157
- 8.3 A Simplified Mathematical Example — 160
- 8.4 Activation Functions — 161
- 8.5 Why Artificial Neural Network Algorithms — 163
- 8.6 Deep Learning — 164
- 8.7 Mathematical Foundations of Artificial Neural Networks — 164
- 8.8 Gradient Descent Methods — 165
- 8.9 Neural Network Learning Processes — 167
- 8.10 Selected Analytics Models — 171
- 8.11 Probabilistic Neural Networks — 172
- 8.12 Support Vector Machine (SVM) Networks — 175
- 8.13 General Feed-forward Neural Network — 177
- 8.14 MLP with Levenberg-Marquardt (LM) Algorithm — 181

Chapter 9

Model Accuracy and Optimization — 185

- 9.1 Accuracy Measures — 187
- 9.2 Accuracy and Validation — 187
- 9.3 Vote-based Schema — 192
- 9.4 Accuracy-based Ensemble Schema — 192
- 9.5 Diversity-based Schema — 193
- 9.6 Optimization-based Schema — 194

**Part III: Case Study—Prediction and Advanced Analytics
in Practice**

Chapter 10

**Ensemble of Models—Medical Prediction Case Study: Data Types, Data
Requirements and Data Pre-Processing — 197**

- 10.1 How Much Data Is Needed for Machine Learning? — 198
- 10.2 Learning Despite Noisy Data — 198

10.3	Pre-processing and Data Scaling —	199
10.4	Data Acquisition for ANN Models —	201
10.5	Ensemble Models Case Study —	202

Appendices

Appendix A: Prognostics Methods —	213
--	------------

Appendix B: A Neural Network Example —	216
---	------------

Appendix C: Back Propagation Algorithm Derivation —	218
--	------------

Appendix D: The Oracle Program —	220
---	------------

References —	223
---------------------	------------

Index —	229
----------------	------------