
Contents



Forward xiii

Preface xv

Acknowledgments xix

Introduction 1

Big Data Timeline 5

Why This Topic Is Relevant Now 8

Is Big Data a Fad? 9

Where Using Big Data Makes a Big Difference 12

Part One The Computing Environment23

Chapter 1 Hardware 27

Storage (Disk) 27

Central Processing Unit 29

Memory 31

Network 33

Chapter 2 Distributed Systems 35

Database Computing 36

File System Computing 37

Considerations 39

Chapter 3 Analytical Tools 43

Weka 43

Java and JVM Languages 44

R 47

Python 49

SAS 50

| | |
|--|------------|
| Part Two Turning Data into Business Value | 53 |
| Chapter 4 Predictive Modeling | 55 |
| A Methodology for Building Models | 58 |
| sEMMA | 61 |
| Binary Classification | 64 |
| Multilevel Classification | 66 |
| Interval Prediction | 66 |
| Assessment of Predictive Models | 67 |
| Chapter 5 Common Predictive Modeling Techniques | 71 |
| RFM | 72 |
| Regression | 75 |
| Generalized Linear Models | 84 |
| Neural Networks | 90 |
| Decision and Regression Trees | 101 |
| Support Vector Machines | 107 |
| Bayesian Methods Network Classification | 113 |
| Ensemble Methods | 124 |
| Chapter 6 Segmentation | 127 |
| Cluster Analysis | 132 |
| Distance Measures (Metrics) | 133 |
| Evaluating Clustering | 134 |
| Number of Clusters | 135 |
| K-means Algorithm | 137 |
| Hierarchical Clustering | 138 |
| Profiling Clusters | 138 |
| Chapter 7 Incremental Response Modeling | 141 |
| Building the Response Model | 142 |
| Measuring the Incremental Response | 143 |
| Chapter 8 Time Series Data Mining | 149 |
| Reducing Dimensionality | 150 |
| Detecting Patterns | 151 |
| Time Series Data Mining in Action: Nike+ FuelBand | 154 |
| Chapter 9 Recommendation Systems | 163 |
| What Are Recommendation Systems? | 163 |
| Where Are They Used? | 164 |

| | |
|--|-----|
| How Do They Work? | 165 |
| Assessing Recommendation Quality | 170 |
| Recommendations in Action: SAS Library | 171 |

| | |
|---|------------|
| Chapter 10 Text Analytics | 175 |
| Information Retrieval | 176 |
| Content Categorization | 177 |
| Text Mining | 178 |
| Text Analytics in Action: Let's Play <i>Jeopardy!</i> | 180 |

Part Three Success Stories of Putting It All Together 193

| | |
|---|------------|
| Chapter 11 Case Study of a Large U.S.-Based Financial Services Company | 197 |
| Traditional Marketing Campaign Process | 198 |
| High-Performance Marketing Solution | 202 |
| Value Proposition for Change | 203 |
| Chapter 12 Case Study of a Major Health Care Provider | 205 |
| CAHPS | 207 |
| HEDIS | 207 |
| HOS | 208 |
| IRE | 208 |
| Chapter 13 Case Study of a Technology Manufacturer | 215 |
| Finding Defective Devices | 215 |
| How They Reduced Cost | 216 |
| Chapter 14 Case Study of Online Brand Management | 221 |
| Chapter 15 Case Study of Mobile Application Recommendations | 225 |
| Chapter 16 Case Study of a High-Tech Product Manufacturer | 229 |
| Handling the Missing Data | 230 |
| Application beyond Manufacturing | 231 |
| Chapter 17 Looking to the Future | 233 |
| Reproducible Research | 234 |
| Privacy with Public Data Sets | 234 |
| The Internet of Things | 236 |

| | |
|------------------------------------|-----|
| Software Development in the Future | 237 |
| Future Development of Algorithms | 238 |
| In Conclusion | 241 |

| | |
|-------------------------|------------|
| About the Author | 243 |
|-------------------------|------------|

| | |
|-----------------|------------|
| Appendix | 245 |
|-----------------|------------|

| | |
|-------------------|------------|
| References | 247 |
|-------------------|------------|

| | |
|--------------|------------|
| Index | 253 |
|--------------|------------|