

# **Analytics in a Big Data World**

---

*The Essential Guide to Data Science  
and Its Applications*

**Bart Baesens**

**WILEY**

---

# Contents

**Preface** xiii

**Acknowledgments** xv

**Chapter 1 Big Data and Analytics** 1

- Example Applications 2
- Basic Nomenclature 4
- Analytics Process Model 4
- Job Profiles Involved 6
- Analytics 7
- Analytical Model Requirements 9
- Notes 10

**Chapter 2 Data Collection, Sampling,  
and Preprocessing** 13

- Types of Data Sources 13
- Sampling 15
- Types of Data Elements 17
- Visual Data Exploration and Exploratory  
Statistical Analysis 17
- Missing Values 19
- Outlier Detection and Treatment 20
- Standardizing Data 24
- Categorization 24
- Weights of Evidence Coding 28
- Variable Selection 29

Segmentation 32

Notes 33

### **Chapter 3 Predictive Analytics 35**

Target Definition 35

Linear Regression 38

Logistic Regression 39

Decision Trees 42

Neural Networks 48

Support Vector Machines 58

Ensemble Methods 64

Multiclass Classification Techniques 67

Evaluating Predictive Models 71

Notes 84

### **Chapter 4 Descriptive Analytics 87**

Association Rules 87

Sequence Rules 94

Segmentation 95

Notes 104

### **Chapter 5 Survival Analysis 105**

Survival Analysis Measurements 106

Kaplan Meier Analysis 109

Parametric Survival Analysis 111

Proportional Hazards Regression 114

Extensions of Survival Analysis Models 116

Evaluating Survival Analysis Models 117

Notes 117

### **Chapter 6 Social Network Analytics 119**

Social Network Definitions 119

Social Network Metrics 121

Social Network Learning 123

Relational Neighbor Classifier 124

Probabilistic Relational Neighbor Classifier	125
Relational Logistic Regression	126
Collective Inferencing	128
Egonets	129
Bigraphs	130
Notes	132

## **Chapter 7 Analytics: Putting It All to Work 133**

Backtesting Analytical Models	134
Benchmarking	146
Data Quality	149
Software	153
Privacy	155
Model Design and Documentation	158
Corporate Governance	159
Notes	159

## **Chapter 8 Example Applications 161**

Credit Risk Modeling	161
Fraud Detection	165
Net Lift Response Modeling	168
Churn Prediction	172
Recommender Systems	176
Web Analytics	185
Social Media Analytics	195
Business Process Analytics	204
Notes	220

## **About the Author 223**

## **Index 225**