Contents

Acknowledgments xi

Preface xiii

Introduction xvii

Glossary xxi

Part 1. Towards an Understanding of Big Data: Are You Ready? 1

Chapter 1. From Data to Big Data: You Must Walk Before You Can Run 3

1.1. Introduction 3
1.2. No analytics without data 4
1.2.1. Databases 5
1.2.2. Raw data. 5
1.2.3. Text 6
1.2.4. Images, audios and videos 6
1.2.5. The Internet of Things 6
1.3. From bytes to yottabytes: the data revolution 7
1.4. Big data: definition 10
1.5. The 3Vs model 12
1.6. Why now and what does it bring? 15
1.7. Conclusions 19

Chapter 2. Big Data: A Revolution that Changes the Game 21

2.1. Introduction 21
2.2. Beyond the 3Vs 22
2.3. From understanding data to knowledge 24
2.4. Improving decision-making 27
2.5. Things to take into account 31
2.5.1. Data complexity 31
2.5.2. Data quality: look out! Not all data are the right data 32
2.5.3. What else?…Data security 33
2.6. Big data and businesses 34
2.6.1. Opportunities 34
2.6.2. Challenges. 36
2.7. Conclusions 40

Part 2. Big Data Analytics: A Compilation of Advanced Analytics Techniques that Covers a Wide Range of Data 41

Chapter 3. Building an Understanding of Big Data Analytics 43

3.1. Introduction 43
3.2. Before breaking down the process… What is data analytics? 44
3.3. Before and after big data analytics 47
3.4. Traditional versus advanced analytics: What is the difference? 49
3.5. Advanced analytics: new paradigm 52
3.6. New statistical and computational paradigm within the big data context 54
3.7. Conclusions 58

Chapter 4. Why Data Analytics and When Can We Use It? 59

4.1. Introduction 59
4.2. Understanding the changes in context 60
4.3. When real time makes the difference 63
4.4. What should data analytics address?. 64
4.5. Analytics culture within companies 68
4.6. Big data analytics application: examples 71
4.7. Conclusions 75

Chapter 5. Data Analytics Process: There's Great Work Behind the Scenes 77

5.1. Introduction 77
5.2. More data, more questions for better answers 78
5.2.1. We can never say it enough: "there is no good wind for those who don't know where they are going" 78
5.2.2. Understanding the basics: identify what we already know and what we have yet to find out 79
5.2.3. Defining the tasks to be accomplished 80
5.2.4. Which technology to adopt? 80
5.2.5. Understanding data analytics is good but knowing how to use it is better! (What skills do you need?) 81
5.2.6. What does the data project cost and how will it pay off in time? 82
5.2.7. What will it mean to you once you find out? 82
5.3. Next steps: do you have an idea about a "secret sauce"? 83
5.3.1. First phase: find the data (data collection) 84
5.3.2. Second phase: construct the data (data preparation) 85
5.3.3. Third phase: go to exploration and modeling (data analysis) 85
5.3.4. Fourth phase: evaluate and interpret the results (evaluation and interpretation). 86
5.3.5. Fifth phase: transform data into actionable knowledge (deploy the model) 87
5.4. Disciplines that support the big data analytics process 88
5.4.1. Statistics 88
5.4.2. Machine learning. 88
5.4.3. Data mining 89
5.4.4. Text mining 90
5.4.5. Database management systems 90
5.4.6. Data streams management systems 91
5.5. Wait, it's not so simple: what to avoid when building a model? 91
5.5.1. Minimize the model error. 94
5.5.2. Maximize the likelihood of the model 95
5.5.3. What about surveys? 95
5.6. Conclusions 99

Part 3. Data Analytics and Machine Learning: the Relevance of Algorithms 101

Chapter 6. Machine Learning: a Method of Data Analysis that Automates Analytical Model Building 103

6.1. Introduction 103
6.2. From simple descriptive analysis to predictive and prescriptive analyses: what are the different steps? 104
6.3. Artificial intelligence: algorithms and techniques 106
6.4. ML: what is it? 109
6.5. Why is it important? 113
6.6. How does ML work? 116
6.6.1. Definition of the business need (problem statement) and its formalization 117
6.6.2. Collection and preparation of the useful data that will be used to meet this need. 117
6.6.3. Test the performance of the obtained model. 118
6.6.4. Optimization and production start. 118
6.7. Data scientist: the new alchemist. 120
6.8. Conclusion 122

Chapter 7. Supervised versus Unsupervised Algorithms: a Guided Tour 123

7.1. Introduction 123
7.2. Supervised and unsupervised learning 124
7.2.1. Supervised learning: predict, predict and predict! 124
7.2.2. Unsupervised learning: go to profiles search! 127
7.3. Regression versus classification 129
7.3.1. Regression. 130
7.3.2. Classification 133
7.4. Clustering gathers data. 141
7.4.1. What good could it serve? 141
7.4.2. Principle of clustering algorithms 144
7.4.3. Partitioning your data by using the K-means algorithm . . 148
7.5. Conclusion 151

Chapter 8. Applications and Examples 153

8.1. Introduction 153
8.2. Which algorithm to use? 153

8.2.1. Supervised or unsupervised algorithm: in which case do we use each one? 154
8.2.2. What about other ML algorithms? 157
8.3. The duo big data/ML: examples of use 161
8.3.1. Netflix: show me what you are looking at and I'll personalize what you like 162
8.3.2. Amazon: when AI comes into your everyday life 165
8.3.3. And more: proof that data are a source of creativity 168
8.4. Conclusions 171

Bibliography 173

Index. 181