

CONTENTS

Preface x

01 Who should read this book? 1

Who should read this book 2

Where we find text 3

Sense and sensibility in thinking about text 6

7A few places we will not be going 17

Where we will be going from here 18

Summary 19

References 20

02 Getting ready: capturing, sorting, sifting, stemming and matching 23

What we need to do with text 24

Ways of corralling words 31

Summary 44

References 47

03 In pictures: word clouds, wordles and beyond 49

Getting words into a picture 50

The many types of pictures and their uses 51

Clustering words 53

Applications, uses and cautions 69

Summary 70

References 71

04 Putting text together: clustering documents using words 73

Where we have been and moving on to documents 74

Clustering and classifying documents 82

Clustering documents	86
Document classification	102
Summary	113
References	116

05 In the mood for sentiment (and counting) 119

Basics of sentiment and counting	120
Counting words	120
Understanding sentiment	126
Summary	139
References	141

06 Predictive models 1: having words with regressions 145

Understanding predictive models	146
Starting from the basics with regression	148
Rules of the road for regression	153
Divergent roads: regression aims and regression uses	156
Practical examples	158
Summary	168
References	170

07 Predictive models 2: classifications that grow on trees 173

Classification trees: understanding an amazing analytical method	174
Seeing how trees work, step by step	174
CHAID and CART (and CRT, C&RT, QUEST, J48 and others)	192
Summary: applications and cautions	193
References	195

08 Predictive models 3: all in the family with Bayes Nets 197

What are Bayes Nets and how do they compare with other methods?	197
---	-----

Our first example: Bayes Nets linking survey questions and behaviour	214
Using a Bayes Net with text	219
Bayes Net software: welcome to the thicket	225
Summary, conclusions and cautions	226
References	229

09 Looking forward and back 231

Where we may be going	232
What role does text analytics play?	235
Summing up: where we have been	237
Software and you	240
In conclusion	243
References	245

Glossary 247

Index 251