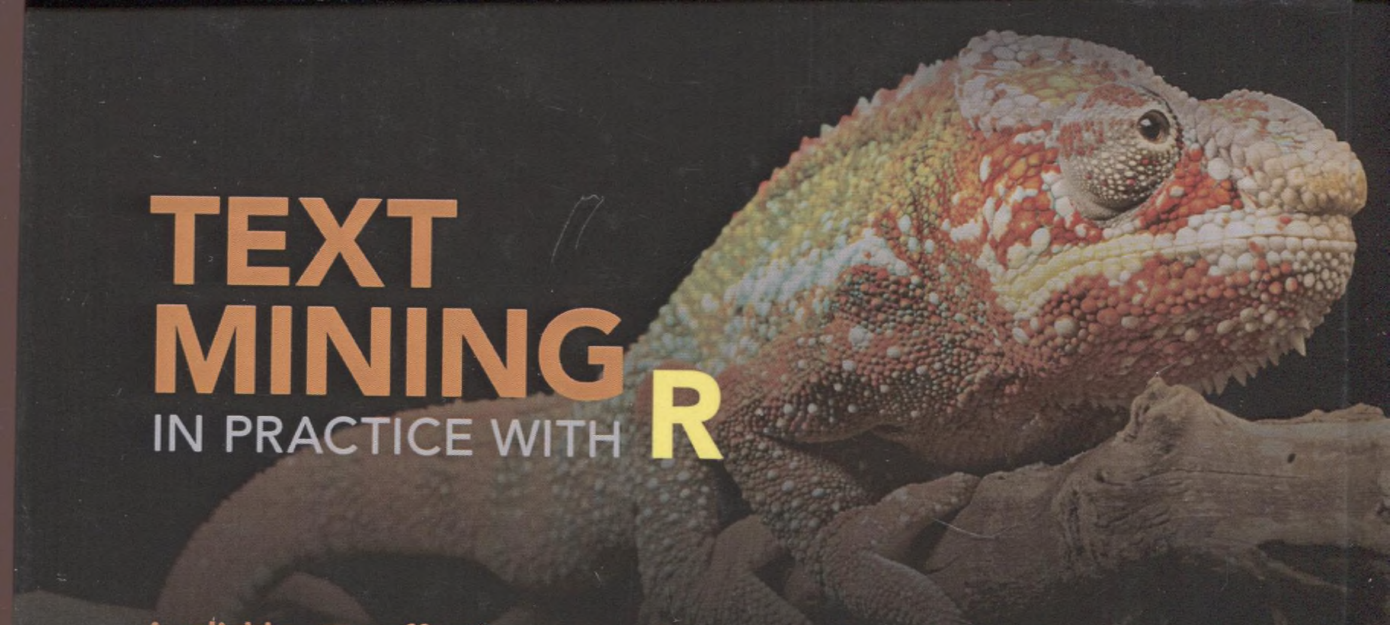

TEXT MINING IN PRACTICE WITH R

A reliable, cost-effective approach to extracting priceless business information from all sources of text

Excavating actionable business insights from data is a complex undertaking, and that complexity is magnified by an order of magnitude when the focus is on documents and other text information. This book takes a practical, hands-on approach to teaching you a reliable, cost-effective approach to mining the vast, untold riches buried within all forms of text using R.

Author Ted Kwartler clearly describes all of the tools needed to perform text mining and shows you how to use them to identify practical business applications to get your creative text mining efforts started right away. With the help of numerous real-world examples and case studies from industries ranging from healthcare to entertainment to telecommunications, he demonstrates how to execute an array of text mining processes and functions, including sentiment scoring, topic modelling, predictive modelling, extracting clickbait from headlines, and more. You'll learn how to:

- Identify actionable social media posts to improve customer service
- Use text mining in HR to identify candidate perceptions of an organisation, match job descriptions with resumes, and more
- Extract priceless information from virtually all digital and print sources, including the news media, social media sites, PDFs, and even JPEG and GIF image files
- Make text mining an integral component of marketing in order to identify brand evangelists, impact customer propensity modelling, and much more

Most companies' data mining efforts focus almost exclusively on numerical and categorical data, while text remains a largely untapped resource. Especially in a global marketplace where being first to identify and respond to customer needs and expectations imparts an unbeatable competitive advantage, text represents a source of immense potential value. Unfortunately, there is no reliable, cost-effective technology for extracting analytical insights from the huge and ever-growing volume of text available online and other digital sources, as well as from paper documents—until now.

TED KWARTLER is a data science instructor at DataCamp.com. He has worked in analytical and executive roles at DataRobot, Liberty Mutual Insurance and Amazon.com.

Cover Design: Wiley
Cover Image: © ChrisPole/Gettyimages

www.wiley.com

WILEY

Also available as an e-book

ISBN 978-1-119-28201-3

Contents