

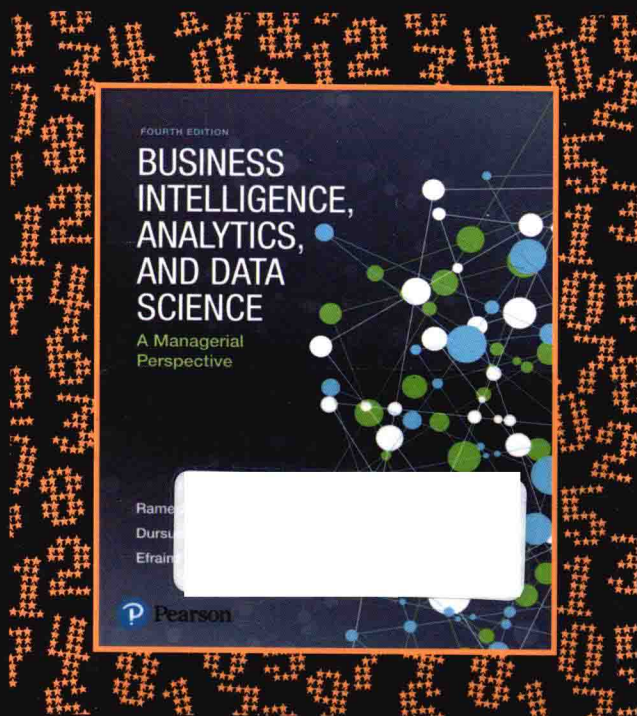
商务智能

数据分析的管理视角

(原书第4版)

[美] 拉姆什·沙尔达 (Ramesh Sharda) 杜尔森·德伦 (Dursun Delen) 埃弗雷姆·特班 (Efraim Turban) 著
俄克拉何马州立大学 俄克拉何马州立大学 夏威夷大学

赵卫东 译
复旦大学



BUSINESS INTELLIGENCE, ANALYTICS, AND DATA SCIENCE

A Managerial Perspective
Fourth Edition

BUSINESS INTELLIGENCE, ANALYTICS, AND DATA SCIENCE

A Managerial Perspective
Fourth Edition

商务智能 数据分析的管理视角

(原书第4版)

[美] 拉姆什·沙尔达 (Ramesh Sharda)
俄克拉何马州立大学

杜尔森·德伦 (Dursun Delen)
俄克拉何马州立大学

埃弗雷姆·特班 (Efraim Turban) 著
夏威夷大学

赵卫东 译
复旦大学



机械工业出版社
China Machine Press

图书在版编目 (CIP) 数据

商务智能: 数据分析的管理视角 (原书第 4 版) / (美) 拉姆什·沙尔达 (Ramesh Sharda), (美) 杜尔森·德伦 (Dursun Delen), (美) 埃弗瑞姆·特班 (Efraim Turban) 著; 赵卫东译. —北京: 机械工业出版社, 2018.4

(数据科学与工程丛书)

书名原文: Business Intelligence, Analytics, and Data Science: A Managerial Perspective, Fourth Edition

ISBN 978-7-111-59864-0

I. 商… II. ①拉… ②杜… ③埃… ④赵… III. 数据处理 - 研究 IV. TP274

中国版本图书馆 CIP 数据核字 (2018) 第 078504 号

本书版权登记号: 图字 01-2017-7521

Authorized translation from the English language edition, entitled Business Intelligence, Analytics, and Data Science: A Managerial Perspective, Fourth Edition, 9780134633282 by Ramesh Sharda, Dursun Delen, Efraim Turban, published by Pearson Education, Inc., Copyright © 2018.

All rights reserved. No part of this book may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording or by any information storage retrieval system, without permission from Pearson Education, Inc.

Chinese simplified language edition published by China Machine Press Copyright © 2018.

本书中文简体字版由 Pearson Education (培生教育出版集团) 授权机械工业出版社在中华人民共和国境内 (不包括香港、澳门特别行政区及台湾地区) 独家出版发行。未经出版者书面许可, 不得以任何方式抄袭、复制或节录本书中的任何部分。

本书封底贴有 Pearson Education (培生教育出版集团) 激光防伪标签, 无标签者不得销售。

本书主要介绍了商务智能一些核心技术的应用, 包括数据仓库、业务报表与企业绩效管理、数据挖掘、文本与 Web 分析、大数据分析等内容, 通过大量的案例和练习帮助读者轻松理解和掌握商务智能在业务决策中的应用。

本书可以作为管理科学、电子商务和企业管理等专业的 MBA、研究生以及本科生商务智能、决策支持系统等课程的教材, 也可以作为管理信息系统、商务数据分析等课程的辅助教材, 还适合从事企业信息管理、业务分析的人士阅读。

出版发行: 机械工业出版社 (北京市西城区百万庄大街 22 号 邮政编码: 100037)

责任编辑: 陈佳媛

责任校对: 殷虹

印刷: 三河市宏图印务有限公司

版次: 2018 年 5 月第 1 版第 1 次印刷

开本: 185mm × 260mm 1/16

印张: 26.25

书号: ISBN 978-7-111-59864-0

定价: 119.00 元

凡购本书, 如有缺页、倒页、脱页, 由本社发行部调换

客服热线: (010) 88378991 88361066

投稿热线: (010) 88379604

购书热线: (010) 68326294 88379649 68995259

读者信箱: hzjsj@hzbook.com

版权所有·侵权必究

封底无防伪标均为盗版

本书法律顾问: 北京大成律师事务所 韩光 / 邹晓东

译者序

商务智能的概念从 20 世纪 90 年代中期正式提出，并在众多企业付诸实践，已经有 20 多年的历史。实践证明，作为企业新一代的决策支持利器，商务智能对提高各阶层的决策质量起到了非常重要的作用，因此也受到越来越多企业的重视。

进入大数据时代，数据科学作为一门独立的科学出现，并在商务中得到了越来越广泛的应用。在这种情况下，商务智能的相关支撑技术得到了飞速的发展。移动计算、云计算、物联网等技术的发展为商务智能提供了大量的业务数据。大数据技术、机器学习的大量深度应用，也提升了商务智能分析数据的类型、规模、速度和变现能力。随着数据分析技术与应用的发展，商务智能的内涵得到了扩充，从而被冠以商务数据分析、业务分析与优化、大数据分析等不同的名字。从本质上来看，商务智能应用涉及数据科学在企业商务领域的所有研究对象，其发展出现了许多特点：从传统的常规数据，到体量超常的大数据；从关系数据库管理系统中的结构化数据，扩展到 NoSQL 数据库管理的文本、图像、声音和视频等多媒体数据；从报表、在线分析处理、企业绩效管理、数据挖掘等传统分析方法，到风靡一时的深度学习；从集中式处理架构，到分布式计算模式。

当今的商务智能系统，已经不仅仅是简单提供数据仓库、报表、仪表盘等数据集成和信息展示的工具，而是借助可视化技术、Hadoop 分布式存储与处理、Spark 内存计算、深度学习等机器学习和大数据处理技术，对不同规模和类型的复杂数据进行快速、深度处理的企业商务分析系统。

数据分析师是 21 世纪最热门的职业之一。为了适应数据分析人才的培养，急需数据分析各类人才培养的相关教材。目前，市场上出现了很多介绍商务智能分析技术的专业书籍，促进了数据分析技术的应用。但对于企业应用人员和管理人员来说，一本入门的全面介绍商务智能的教材也是非常必要的。幸运的是，本书的 3 位作者为适应商务智能理论和应用的发展，在上一版的基础上，参阅了大量的文献，进行了改版，这将在很大程度上促进人们对商务智能的认识。

本版删除了上一版中过时的内容，根据商务智能的最新发展补充了很多新颖的内容，包括商务智能相关的概念、技术和应用案例等。全书以描述性分析、规范性分析和预测性分析为主线，分别介绍了这三类分析包括的内容。与数据分析技术类的书籍不同，全书娓

娓娓道来，深入浅出，不拘泥于复杂的算法和数据分析技术细节，使初步接触商务智能的读者也能理清商务智能与移动计算、云计算、物联网、大数据技术、机器学习和数据科学等众多概念的区别及联系，并从大量的实际案例中体会商务智能的功能和价值。

本书非常适合作为商学院、管理学院本科生、全日制研究生和 MBA 商务智能、商务数据分析、决策支持系统及管理信息系统等课程的教材或参考书，也适合对商务智能感兴趣的企业管理人员和企业业务人员参考。

在本书的翻译过程中，胡远文、陈子轩、赵洪博、朱荣斌、郑光键等同学也做了很多贡献。由于译者水平有限，书中难免存在翻译不当之处，欢迎读者批评指正。译者联系方式：wdzhao@fudan.edu.cn。

赵卫东

2017年9月 复旦大学

作者简介

Ramesh Sharda (威斯康星大学麦迪逊分校 MBA, 博士) 是康菲石油公司技术管理主席、俄克拉何马州立大学 (OSU) 斯派尔商学院管理科学与信息系统学系杰出教授。他参与创立并指导了俄亥俄州立大学商学博士课程, 共发表了大约 200 篇论文, 主要刊物包括《*Operations Research*》《*Management Science*》《*Information Systems Research*》《*Decision Support Systems*》和《*the Journal of MIS*》。他共同创办了决策支持系统和知识管理上的 AIS SIG (SIGDSA)。Sharda 博士担任多个编辑委员会委员, 包括《*Decision Sciences Journal*》《*Decision Support Systems*》和《*ACM Data Base*》。他撰写和编辑了几本教科书及研究书籍, 并与 Springer 合作编辑了系列书籍 (《*Integrated Series in Information Systems*》《*Operations Research/Computer Science Interface*》和《*Annals of Information Systems*》)。他目前也是 Teradata 大学网络执行总监, 研究方向是决策支持系统、商务分析和信息超载的技术。

Dursun Delen 博士 (俄克拉何马州立大学博士) 是斯派尔斯工商管理主席, Patterson 基金会商务分析基金会主席、卫生系统创新中心研究主任, 以及俄克拉何马州立大学 (OSU) 斯派尔学院管理科学与信息系统教授。在他的学术生涯之前, 他曾在得克萨斯大学车站的一家私人研究和咨询公司 Knowledge Based Systems 担任研究科学家 5 年。在此期间, 他领导了一些决策支持和其他由几个联邦机构资助的与信息系统的研究项目, 包括国防部 (DoD)、国家航空航天局 (NASA)、国家标准与技术研究所 (NIST)、弹道导弹防御组织 (BMDO) 和能源部 (DOE)。Delen 博士发表了 100 多篇同行评议文章, 其中有些出现在重要的杂志期刊中, 例如《*Decision Sciences*》《*Decision Support Systems*》《*Communications of the ACM*》《*Computers and Operations Research*》《*Computers in Industry*》《*Journal of Production Operations Management*》《*Artificial Intelligence in Medicine*》《*International Journal of Medical Informatics*》《*Expert Systems with Applications*》和《*IEEE Wireless Communications*》。他在商务分析、数据挖掘、文本挖掘、商务智能和决策支持系统的广泛领域撰写或者合作了七本书籍。他经常被邀请参加国内和国际会议, 主题演讲涉及与数据/文本挖掘、商务分析、决策支持系统、商务智能和

知识管理相关的主题。曾担任第四届网络计算和高级信息管理国际会议（2008年9月2日至2008年9月4日，韩国首尔）的大会共同主席，并定期在各种信息系统和分析会议上担任主席。他目前担任十几个学术期刊的主编、高级编辑、副主编或编委会成员。他的研究和教学方向是数据和文本挖掘、商务分析、决策支持系统、知识管理、商务智能和企业建模。

Efraim Turban（加州大学伯克利分校的MBA博士）是夏威夷大学太平洋信息系统管理研究所的访问学者。在此之前，他曾任职的大学包括香港城市大学、利哈伊大学、佛罗里达国际大学、加利福尼亚州立大学长滩分校、东伊利诺伊大学和南加州大学。Turban博士在《*Management Science*》《*MIS Quarterly*》和《*Decision Support Systems*》等领先期刊上发表了100多篇评论论文。他也是20本书的作者，其中包括《*Electronic Commerce : A Managerial Perspective*》和《*Information Technology for Management*》。他还是全球各大公司的顾问。Turban博士目前的兴趣方向领域是基于Web的决策支持系统、社会商务和协作决策。

前 言

分析已经成为这十年来的技术驱动力。诸如 IBM、SAP、SAS、Teradata、Oracle、Microsoft 和 Dell 等公司正在创建专注于分析的新型组织单元，帮助企业更加有效和高效地运营。决策者也正在使用更多的计算机化工具来支持他们的工作。消费者甚至都在直接或间接地使用分析工具来对购物、健康或医疗保健、旅游和娱乐等日常活动做出决定。商务智能和商务分析领域得到了迅速发展，它们更加专注于能从数据流中提取知识和洞察力的创新应用，以往这些数据流并没有被捕获，更不用说以任何重要的方式进行分析。分析技术在医疗保健、体育、旅游、娱乐、供应链管理、公用事业以及几乎所有可想象的行业都有新的应用，术语“分析”已成为主流。事实上，它已经演变成其他术语，如数据科学，其最新的形式演变是深度学习和物联网。

本书提供了从描述性分析（例如数据的性质、统计建模、数据可视化和商务智能）开始的商务分析连续统一体的管理视角，然后转向预测性分析（例如数据挖掘、文本/Web 挖掘、社交媒体挖掘），然后进行规范性分析（例如优化和模拟），最后两章分别讨论大数据的概念和工具以及商务分析未来趋势、隐私和管理思考。本书有一个原出版社支持的配套网站（pearsonhighered.com/sharda），还有一个作者维护的独立网站（dssbibook.com）。

本书的目的是向读者介绍这些通常被称为商务分析或数据科学的技术。本书介绍了这些系统的构建和使用方法的基本技术原理。我们遵循 EEE 方法介绍这些主题：展示、体验和探索（Exposure, Experience, and Exploration）。本书主要提供各种分析技术及其应用的展示，这将会激发学生去学习其他组织如何采用分析做出决策或者获得竞争优势。我们认为，这种对分析工作内容以及分析实现方式的展示是学习分析的关键组成部分。在描述这些技术时，我们还会介绍可用于开发此类应用的特定软件工具。但本书不限于任何一种软件工具，因此学生可以使用任何其他可用的软件工具来体验这些技术。每章都给出了特定的建议，学生和教师可以结合不同的数据分析软件工具阅读本书。本书的配套网站也包括特定的软件指导，学生可以通过多种方式获得运用这些技术的经验。最后，我们希望这些经验能够鼓励读者在各自不同的领域探究这些技术的潜力。为了便于学习，我们提供了一些指向 Teradata 大学网络和其他网站的练习以及

面向团队的练习。我们也会从本书的网站分布新的和创新性的应用。

本书第4版的具体改进主要集中在四个方面：重新编排、新的章节、内容更新和更加突出的重点。尽管发生了许多变化，但我们仍保持着全面性和用户友好性。最后，我们提供了更准确、更新的内容。接下来我们介绍第4版的变化。

第4版中有哪些更新变化

为了改进本书以反映商务分析的重点，意味着第4版需要进行重大重组。这一版围绕三种主要的商务分析（即描述性、预测性和规范性）组织。新版本补充了很多新的内容，删除了过时的内容，具体的变化如下。

新的组织形式：本书接受了 INFORMS 提出的三种类型的分析分类——描述性、预测性和规范性。第1章介绍了商务智能和商务分析，以及它们在许多行业中的应用。该章还包括分析生态系统的概述，以帮助用户探索在分析环境中参与和发展的所有不同方式。第2章对统计学、数据的重要性和描述性分析/可视化进行了概述。第3章介绍数据仓库和数据基础，包括更新的内容，特别是数据湖的概念。第4章涵盖了预测性分析。第5章将商务分析应用扩展到文本、Web 和社交媒体。第6章涉及规范性分析，特别是线性规划和模拟，这是本书中的全新内容。第7章介绍了大数据工具和平台。第8章对本书的内容进行了总结，涉及商务分析中的新兴趋势和主题，包括位置分析、物联网、基于云的分析 and 商务分析中的隐私与道德考虑。数据分析生态系统的讨论也提及了规范性分析。

新章节：以下章节被全新补充。

第2章：描述性分析 I：数据性质、统计建模和可视化

该章的目的是透彻了解数据的性质，这是任何分析研究的主要组成部分。然后，统计建模将作为描述性分析的一部分介绍。数据可视化已成为任何业务报表或描述性分析项目的热门部分，因此它在该章中会详细解释。该章还介绍了几个现实世界的案例（75%的新材料）。

第6章：规范性分析：优化与模拟

该章介绍规范性分析。重点介绍使用线性规划技术在 Excel 中进行优化建模，还介绍了模拟的概念。该章是作者关于决策支持系统的书第10版中两章内容的更新版本。对于这本书，这是一个全新的章节（99%的新材料）。

第8章：商务分析的未来趋势、隐私和管理思考

该章将考察几个新的社会现象，这些现象已经在改变，或者有可能改变分析领域，包括地理空间分析、物联网以及基于云分析的资料的重大更新。该章还更新了上一版中关于道德和隐私考虑的内容（70%的新材料）。

章节修订：其他的所有章节也已修改和更新，以下是其他章节的更改摘要。

第1章：商务智能、商务分析和数据科学概述

该章已经被重写并显著扩展，开篇小插曲涵盖了体育分析的多种应用，介绍了 INFORMS

提出的三种类型的商务分析：描述性、预测性和规范性分析。前面提到，这个分类用于指导本书的完整重组（之前的内容呈现了新的面貌）。然后，该章包括几个医疗保健和零售业分析的新例子。最后，该章以分析生态系统的扩展和更新内容结尾，为学生提供了商务分析和数据科学行业的广泛性（约 60% 的新材料）。

第 3 章：描述性分析 II：商务智能和数据

这是一个旧章节，但插入了一些新的小节（例如数据湖）和新的案例（约 30% 的新材料）。

第 4 章：预测性分析 I：数据挖掘过程、方法和算法

这是一个旧章节，但是补充了新的内容、流程和一些新的案例（约 20% 的新材料）。

第 5 章：预测性分析 II：文本、Web 和社交媒体分析

这是一个旧章节，但是添加了一些新的内容组织、流程以及新的案例（约 25% 的新材料）。

第 7 章：大数据概念和分析，这是上一版的第 6 章，更新了一个新的开篇小插曲和案例，介绍了 Teradata Aster 以及替代数据的新材料（约 25% 的新材料）。

□ **改版作者团队**：基于以前的版本作者（Turban, Sharda, Delen 和 King）的优秀内容，本版本主要由 Ramesh Sharda 和 Dursun Delen 进行修订。Ramesh 和 Dursun 都在分析行业中工作已久，拥有行业和研究经验。

□ **实时更新的网站**：本书的使用者可以访问一个网站，其中包含与新故事、软件、教程，甚至与本书涵盖主题相关的 YouTube 视频的链接。该网站可在 dssbibook.com 上访问。

□ **修订和更新的内容**。几乎所有的章节都有基于最近的故事和事件的新的开篇小插曲。此外，整本书中的应用案例已更新，包含了特定技术和模型的最新应用示例。在整本书中增加了新的网站链接。我们还删除了许多较旧的产品链接和引用。最后，大多数章节都有新的练习、网络任务和问题讨论。

□ **与 Teradata 大学网络 (TUN) 的链接**。大部分章节包括了与 TUN (teradatauniversitynetwork.com) 的新链接。

□ **书名**。已经很明显，这本书的书名和重点已经发生了很大变化。

□ **软件支持**。TUN 网站免费提供软件支持。它还提供了免费数据挖掘和其他软件的链接。此外，该网站还提供使用此类软件的练习。

补充包：www.pearsonhighered.com/sharda

提供全面灵活的技术支持包，以提高教学和学习经验。以下教师和学生补充包可在本书的网站 pearsonhighered.com/sharda 上获得。

□ **教师手册**。教师手册 (Instructor's Manual) 包括整个课程每章的学习目标和每章的结尾处的问题与练习的答案以及教学建议 (包括项目说明)。教师手册可在 Pearsonhighered.com/sharda 的 secure faculty 部分获得。

□ **测试项目文件和 TestGen 软件**。测试项目文件 (the Test Item File) 是判断题、多项选

择题、填空题和叙述题的综合集合。测试项目文件在 Microsoft Word 和 TestGen 中可用。Pearson Education 的测试生成软件可从 www.pearsonhighered.com/irc 获得。该软件与 PC/MAC 兼容，并预装了所有的测试项目文件问题。你可以手动或随机查看测试问题并进行拖放以创建测试，也可以根据需要添加或修改测试问题。我们的 TestGens 可以被转换为在 BlackBoard、WebCT、Moodle、D2L 和 Angel 中使用。这些转换可以在 pearsonhighered.com/sharda 找到。TestGen 可从 Respondus 获取，也可以从 www.respondus.com 上找到。

□ PowerPoint 幻灯片。PowerPoint 幻灯片可用于演示和构建内容中的关键概念。教师可以从 Pearson-highered.com/sharda 下载 PowerPoint 幻灯片。

致谢

自本书第 1 版出版以来，许多人提出了建议和批评。也有几十名学生参加了各个章节、软件和方法的测试，并协助收集材料。列举参与这个项目的所有人是不可能的，但我们要感谢所有人。有些人也做出了重大贡献，他们值得特别感谢。

首先，我们感谢那些为第 1 版到第 3 版提供正式评审的人的辛苦工作：

Ann Aksut, Central Piedmont Community College
 Bay Arinze, Drexel University
 Andy Borchers, Lipscomb University
 Ranjit Bose, University of New Mexico
 Marty Crossland, MidAmerica Nazarene University
 Kurt Engemann, Iona College
 Badie Farah, Eastern Michigan University
 Gary Farrar, Columbia College
 Jerry Fjermestad, New Jersey Institute of Technology
 Christie M. Fuller, Louisiana Tech University
 Martin Grossman, Bridgewater State College
 Jahangir Karimi, University of Colorado, Denver
 Huei Lee, Eastern Michigan University
 Natalie Nazarenko, SUNY Fredonia
 Joo Eng Lee-Partridge, Central Connecticut State University
 Gregory Rose, Washington State University, Vancouver
 Khawaja Saeed, Wichita State University
 Kala Chand Seal, Loyola Marymount University
 Joshua S. White, PhD, State University of New York Polytechnic Institute
 Roger Wilson, Fairmont State University
 Vincent Yu, Missouri University of Science and Technology
 Fan Zhao, Florida Gulf Coast University

我们也感谢那些对本书以及我们关于决策支持系统的书——《Business Intelligence and Analytics: Systems for Decision Support》(第 10 版, Pearson Education, 2013) 进行正式评审的人的辛苦工作。

第二，有几个人为本书或支持材料提供了材料。Teradata 的 Susan Baskin 和 David Schrader

博士为本书确定新的 TUN 和 Teradata 内容提供了特别的帮助，并为其确定了许可。Dave Schrader 博士为本书写了开篇小插曲。这个小插曲还得到了奥本大学的 Ashish Gupta 博士和田纳西州 - 查塔努加大学的 Gary Wilkerson 开发的材料的帮助，它将为分析提供一个很好的介绍。我们也非常感谢 INFORMS 允许我们从 *Interfaces* 引用一些内容。以下人员在我们更新这本书的时候也提供了帮助：Pankush Kalgotra, Prason Mathur, Rupesh Agarwal, Shubham Singh, Nan Liang, Jacob Pearson, Kinsey Clemmer 和 Evan Murllette (都来自俄克拉何马州立大学)，非常感激他们对本版的帮助。Teradata Aster 团队，特别是 Mark Ott，为第 7 章的开篇小插曲提供了材料。第 7 章中的 Aster 材料改编自 John Thuma 和 Greg Bethardy 开发的培训指南。Humana Corporation 首席信息官 Brian LeClaire 博士提供了他和他的团队在 Humana 开发的几项真实的医疗保健案例。vCreaTek 的 Abhishek Rathi 贡献了他在零售行业的分析愿景。感谢 Rick Wilson 博士在 Excel 中优秀的教学和线性编程技能的表现。Matt Turck 同意让我们采用他的物联网生态系统材料。Ramesh 还感谢女儿 Ruchy Sharda Sen 提供的文字编辑工作。此外，我们以前的博士生和研究同事以多种直接或间接的方式为本书提供了内容或建议和支持：

Asil Oztekin, University of Massachusetts-Lowell
 Enes Eryarsoy, Sehir University
 Hamed Majidi Zolbanin, Ball State University
 Amir Hassan Zadeh, Wright State University
 Supavich (Fone) Pengnate, North Dakota State University
 Christie Fuller, Boise State University
 Daniel Asamoah, Wright State University
 Selim Zaim, Istanbul Technical University
 Nihat Kasap, Sabanci University

第三，对于上一版，我们非常感谢 Dave King (JDA Software Group, Inc.) 的贡献。上一版的其他主要贡献者包括 J. Aronson (佐治亚大学)，他是我们的共同作者，贡献了数据仓库章节。Mike Goul (亚利桑那州立大学) 主要在第 1 章做出了贡献；以及 T. P. Liang (台湾“中山大学”) 在以前的版本中为神经网络提供了材料。Judy Lang 与我们所有人合作，完成编辑任务，并在第 1 版的整个项目中指导我们。

第四，几家供应商通过为以前的版本提供案例研究和演示软件与我们合作：Acxiom (Little Rock, Arkansas), California Scientific Software (Nevada City, California), Cary Harwin of Catalyst Development (Yucca Valley, California), IBM (San Carlos, California), DS Group, Inc. (Greenwich, Connecticut), Gregory Piatetsky-Shapiro of KDnuggets.com, Gary Lynn of NeuroDimension Inc. (Gainesville, Florida), Palisade Software (Newfield, New York), Promised Land Technologies (New Haven, Connecticut), Salford Systems (La Jolla, California), Sense Networks (New York, New York), Gary Miner of StatSoft, Inc. (Tulsa, Oklahoma), Ward Systems Group, Inc. (Frederick, Maryland), Idea Fisher Systems, Inc. (Irving, California), Wordtech Systems (Orinda, California)。

第五，特别感谢 Teradata 大学网络，特别感谢项目总监 Susan Baskin。Hugh Watson，他创

办了 TUN 网络；以及 Michael Goul、Barb Wixom 和 Mary Gros，他们鼓励将本书与 TUN 网络相连，并为本书提供有用的材料。

最后，感谢 Pearson 团队的 Samantha Lewis，他们与我们合作完成了修订，感谢生产团队的 Ann Pulido、Revathi Viswanathan，以及 Cenveo 的工作人员，他们将手稿变成了一本书。

我们要感谢所有为本书出版做出贡献的个人和公司。若没有他们的帮助，本书是不可能出现的。

目 录

译者序	1.4.9 开发或获取 BI 系统	17
作者简介	1.4.10 合理性和成本效益分析	17
前言	1.4.11 安全和隐私保护	17
第 1 章 商务智能、商务分析和数据科学概述	1.4.12 系统和应用集成	18
1.1 开篇小插曲：体育分析——学习和了解商务分析应用的一个令人兴奋的前沿	1.5 商务分析概述	18
1.2 变化的商业环境，决策支持和商务分析的需求变化	1.5.1 描述性分析	19
1.3 计算机决策支持向商务分析和数据科学的演变	应用案例 1.2 Silvaris 通过可视化分析和实时报表功能提升业务	19
1.4 商务智能框架	应用案例 1.3 西门子通过使用数据可视化降低成本	20
1.4.1 商务智能的定义	1.5.2 预测性分析	20
1.4.2 商务智能的简史	应用案例 1.4 运动损伤分析	21
1.4.3 商务智能的架构	1.5.3 规范性分析	21
1.4.4 商务智能的起源和驱动力	应用案例 1.5 特种钢筋公司使用商务分析确定可承诺交付日期	22
应用案例 1.1 Sabre 通过仪表盘和分析帮助客户	1.5.4 商务分析应用于不同领域	22
1.4.5 商务智能多媒体应用	1.5.5 商务分析或数据科学	23
1.4.6 事务处理与分析处理	1.6 所处领域的商务分析实例	23
1.4.7 进行适当的规划并与企业战略保持一致	1.6.1 应用于医疗保健的商务分析——Humana 实例	24
1.4.8 实时按需的 BI 实现	1.6.2 零售价值链中的商务分析	27
	1.7 大数据分析简介	29

应用案例 1.6 CenterPoint Energy 使用 实时大数据分析改善客 户服务	30	2.5.1 用于描述性分析的描述性 统计	62
1.8 商务分析生态系统概述	31	2.5.2 集中趋势度量 (也可以称作位置 或中心度量)	62
1.8.1 数据生成基础设施提供商	32	2.5.3 算术平均数	62
1.8.2 数据管理基础设施提供商	32	2.5.4 中位数	63
1.8.3 数据仓库提供商	33	2.5.5 众数	63
1.8.4 中间件提供商	33	2.5.6 离散趋势度量 (也可称为散布或 分散度量)	63
1.8.5 数据服务提供商	33	2.5.7 极差	64
1.8.6 专注于商务分析的软件 开发者	34	2.5.8 方差	64
1.8.7 应用开发者: 特定行业或 一般行业	35	2.5.9 标准差	64
1.8.8 商务分析行业分析师和有 影响力者	36	2.5.10 平均绝对偏差	64
1.8.9 学术机构和认证机构	37	2.5.11 四分位数与四分位距	64
1.8.10 监管者和政策制定者	37	2.5.12 箱图	65
1.8.11 分析用户组织	37	2.5.13 分布形状	66
1.9 本书计划	38	技术洞察 2.1 如何使用 Microsoft Excel 进行描述性统计	67
1.10 资源、链接和 Teradata 大学网络 连接	39	应用案例 2.3 Cary 镇使用商务分析来 分析传感器数据、评估 需求和检测问题	70
第 2 章 描述性分析 I: 数据的性质、 统计建模与可视化	44	2.6 用于推断性统计的回归建模	71
2.1 开篇小插曲: SiriusXM 利用 数据驱动的营销吸引新一代的 移动消费者	44	2.6.1 我们如何开发线性回归模型	72
2.2 数据的性质	47	2.6.2 我们如何知道模型是否 足够好	73
2.3 数据的简单分类	50	2.6.3 什么是线性回归最重要的 假设	74
应用案例 2.1 医疗器械公司节省开支的 同时确保产品质量	52	2.6.4 逻辑回归	74
2.4 数据预处理的艺术与科学	54	应用案例 2.4 预测 NCAA 碗赛结果	75
应用案例 2.2 通过数据驱动分析提高学 生保留率	56	2.6.5 时间序列预测	79
2.5 用于商务分析的统计建模	61	2.7 业务报表	80
		应用案例 2.5 纸质报表的洪水结束于 FEMA	82
		2.8 数据可视化	83

应用案例 2.6 Macfarlan Smith 使用 Tableau Online 提高运营绩效洞察	85	2.11.10 使用仪表盘设计原则选取正确的视觉构造	102
2.9 不同类型的图表	87	2.11.11 为指导性分析做准备	102
2.9.1 基本图表	88	第 3 章 描述性分析 II: 商务智能和数据仓库	109
2.9.2 专用图表	88	3.1 开篇小插曲: 利用商务智能和数据仓库定位税务欺诈	109
2.9.3 你应该使用哪种图表	89	3.2 商务智能与数据仓库	111
2.10 可视化分析的兴起	91	3.2.1 什么是数据仓库	112
技术洞察 2.2 商务智能和分析平台的 Gartner 魔力象限	91	3.2.2 数据仓库的历史视角	112
2.10.1 可视化分析	93	3.2.3 数据仓库的特征	114
技术洞察 2.3 通过数据和可视化讲出色的故事	93	3.2.4 数据集市	115
2.10.2 高效率可视化分析环境	95	3.2.5 操作数据存储	115
2.11 信息仪表盘	97	3.2.6 企业数据仓库	115
应用案例 2.7 Dallas Cowboys 使用 Tableau 与 Teknion 获得高分	98	3.2.7 元数据	116
2.11.1 仪表盘设计	99	应用案例 3.1 更好的数据计划: 著名的 TELCO 公司使用数据仓库和商务分析技术在竞争激烈的行业中保持领先地位	116
应用案例 2.8 可视化分析帮助能源供应商实现更好的连接	100	3.3 数据仓库过程	117
2.11.2 在仪表盘中要寻找的内容	101	3.4 数据仓库架构	119
2.11.3 仪表盘设计的最佳实践	101	3.4.1 可选的数据仓库架构	121
2.11.4 符合行业标准的基准关键绩效指标	101	3.4.2 哪种架构最好	123
2.11.5 使用情境的元数据包装仪表盘指标	101	3.5 数据集成以及提取、转换和加载过程	124
2.11.6 通过可用性专家验证仪表盘设计	102	3.5.1 数据集成	124
2.11.7 把传输到仪表盘的报警或异常按优先级划分并排序	102	应用案例 3.2 BP Lubricants 公司实现 BIGS 成功	124
2.11.8 以业务用户评论丰富仪表盘	102	3.5.2 提取、转换和加载	126
2.11.9 从三个不同层次呈现信息	102	3.6 数据仓库的开发	128
		应用案例 3.3 使用 Teradata 分析的 SAP 解决方案加速大数据交付	128

3.6.1 数据仓库开发方法	131	4.2 数据挖掘概念与应用	165
3.6.2 数据仓库开发的其他思考	133	应用案例 4.1 Visa 利用预测性分析和 数据挖掘提升用户体验, 同时减少欺诈行为	166
3.6.3 数据仓库中的数据表示	133	4.2.1 定义、特征和优势	167
技术洞察 3.1 托管数据仓库	133	4.2.2 数据挖掘原理	169
3.6.4 数据仓库中的数据分析师	135	应用案例 4.2 戴尔在 21 世纪通过分析 保持敏捷和高效	169
3.6.5 OLAP 和 OLTP	135	4.2.3 数据挖掘与统计学	173
3.6.6 OLAP 操作	135	4.3 数据挖掘应用	173
3.7 数据仓库的实施问题	136	应用案例 4.3 预测性分析和数据挖掘 有助于阻止对恐怖分子 的资助	175
应用案例 3.4 EDW 帮助连接密歇根的 政府机构	138	4.4 数据挖掘过程	176
3.8 数据仓库管理、安全问题和未来 趋势	140	4.4.1 步骤 1: 业务理解	176
技术洞察 3.2 数据湖	141	4.4.2 步骤 2: 数据理解	177
3.9 企业绩效管理	145	4.4.3 步骤 3: 数据准备	177
应用案例 3.5 AAPR 改造其 BI 基础 设施, 并在三年内达到 347% 的投资回报	146	4.4.4 步骤 4: 建模	178
3.10 绩效度量	149	应用案例 4.4 数据挖掘有助于癌症 研究	178
3.10.1 关键绩效指标	149	4.4.5 步骤 5: 测试和评估	180
3.10.2 绩效度量系统	150	4.4.6 步骤 6: 部署	180
3.11 平衡记分卡	150	4.4.7 其他数据挖掘标准化过程和 方法	181
3.11.1 四个视角	150	4.5 数据挖掘方法	182
3.11.2 平衡在平衡记分卡中的 意义	151	4.5.1 分类	182
3.12 作为绩效度量系统的六西格玛	152	4.5.2 估计分类模型的准确度	183
3.12.1 DMAIC 绩效模型	152	应用案例 4.5 Influence Health 使用 先进的预测性分析来 关注影响人们医疗保 健决策的因素	189
3.12.2 平衡记分卡与六西格玛	153	4.5.3 数据挖掘聚类分析	191
3.12.3 有效的绩效度量	153	4.5.4 关联规则挖掘	192
应用案例 3.6 Expedia.com 的客户满意 度记分卡	154	4.6 数据挖掘软件工具	195
第 4 章 预测性分析 I: 数据挖掘的 过程、方法和算法	162		
4.1 开篇小插曲: 迈阿密达德警察 部门应用预测性分析预测和打 击犯罪	162		