

经 典 原 版 书 库

# 商务智能

## 数据分析的管理视角

[美] 拉姆什·沙尔达 杜尔森·德伦 埃弗瑞姆·特班 著  
Ramesh Sharda Dursun Delen Efraim Turban

(英文版·第4版)

FOURTH EDITION

# BUSINESS INTELLIGENCE, ANALYTICS, AND DATA SCIENCE:

A Managerial  
Perspective

Ramesh Sharda  
Dursun Delen  
Efraim Turban

Pearson

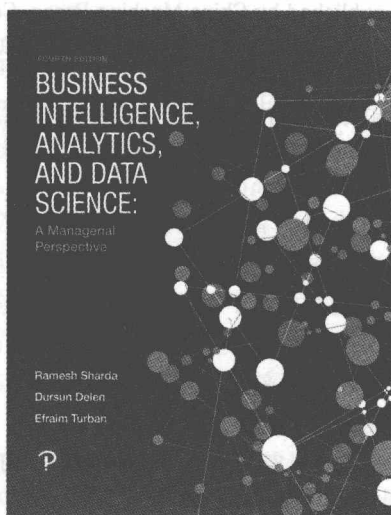
经 典 原 版 书 库

# 商务智能

## 数据分析的管理视角

(英文版·第4版)

*Business Intelligence,  
Analytics, and Data Science*  
A Managerial Perspective (4th Edition)



[美] 拉姆什·沙尔达 杜尔森·德伦 埃弗瑞姆·特班 著  
Ramesh Sharda      Dursun Delen      Efraim Turban



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 (英文版·第4版) / (美) 拉姆什·沙尔达 (Ramesh Sharda) 等著. —北京：机械工业出版社，2018.8

(经典原版书库)

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

ISBN 978-7-111-60785-4

I. 商… II. 拉… III. 数据处理 - 研究 - 英文 IV. TP274

中国版本图书馆 CIP 数据核字 (2018) 第 200278 号

本书版权登记号：图字 01-2018-4067

Authorized Reprint from the English language edition, entitled *Business Intelligence, Analytics, and Data Science: A Managerial Perspective*, Fourth Edition, ISBN 9780134633282, Ramesh Sharda, Dursun Delen, Efraim Turban, published by Pearson Education, Inc., Copyright © 2018, 2014, 2011, 2008 by Pearson Education, Inc., or its affiliates.

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.

English language edition published by China Machine Press, Copyright © 2018.

本书英文影印版由 Pearson Education Inc. 授权机械工业出版社独家出版。未经出版者书面许可，不得以任何方式复制或抄袭本书内容。

此影印版仅限于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) 销售发行。

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

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：张梦玲

责任校对：殷虹

印刷：北京市兆成印刷有限责任公司

版次：2018 年 9 月第 1 版第 1 次印刷

开本：186mm×240mm 1/16

印张：31.5

书号：ISBN 978-7-111-60785-4

定价：109.00 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

# 出版者的话

文艺复兴以来，源远流长的科学精神和逐步形成的学术规范，使西方国家在自然科学的各个领域取得了垄断性的优势；也正是这样的优势，使美国在信息技术发展的六十多年间名家辈出、独领风骚。在商业化的进程中，美国的产业界与教育界越来越紧密地结合，计算机学科中的许多泰山北斗同时身处科研和教学的最前线，由此而产生的经典科学著作，不仅擘划了研究的范畴，还揭示了学术的源变，既遵循学术规范，又自有学者个性，其价值并不会因年月的流逝而减退。

近年，在全球信息化大潮的推动下，我国的计算机产业发展迅猛，对专业人才的需求日益迫切。这对计算机教育界和出版界都既是机遇，也是挑战；而专业教材的建设在教育战略上显得举足轻重。在我国信息技术发展时间较短的现状下，美国等发达国家在其计算机科学发展的几十年间积淀和发展的经典教材仍有许多值得借鉴之处。因此，引进一批国外优秀计算机教材将对我国计算机教育事业的发展起到积极的推动作用，也是与世界接轨、建设真正的世界一流大学的必由之路。

机械工业出版社华章公司较早意识到“出版要为教育服务”。自1998年开始，我们就将工作重点放在了遴选、移译国外优秀教材上。经过多年的不懈努力，我们与Pearson, McGraw-Hill, Elsevier, MIT, John Wiley & Sons, Cengage等世界著名出版公司建立了良好的合作关系，从他们现有的数百种教材中甄选出Andrew S. Tanenbaum, Bjarne Stroustrup, Brian W. Kernighan, Dennis Ritchie, Jim Gray, Alfred V. Aho, John E. Hopcroft, Jeffrey D. Ullman, Abraham Silberschatz, William Stallings, Donald E. Knuth, John L. Hennessy, Larry L. Peterson等大师名家的一批经典作品，以“计算机科学丛书”为总称出版，供读者学习、研究及珍藏。大理石纹理的封面，也正体现了这套丛书的品位和格调。

“计算机科学丛书”的出版工作得到了国内外学者的鼎力相助，国内的专家不仅提供了中肯的选题指导，还不辞劳苦地担任了翻译和审校的工作；而原书的作者也相当关注其作品在中国的传播，有的还专门为其书的中译本作序。迄今，“计算机科学丛书”已经出版了近两百个品种，这些书籍在读者中树立了良好的口碑，并被许多高校采用为正式教材和参考书籍。其影印版“经典原版书库”作为姊妹篇也被越来越多实施双语教学的学校所采用。

权威的作者、经典的教材、一流的译者、严格的审校、精细的编辑，这些因素使我们的图书有了质量的保证。随着计算机科学与技术专业学科建设的不断完善和教材改革的逐渐深化，教育界对国外计算机教材的需求和应用都将步入一个新的阶段，我们的目标是尽善尽美，而反馈的意见正是我们达到这一终极目标的重要帮助。华章公司欢迎老师和读者对我们的工作提出建议或给予指正，我们的联系方式如下：

华章网站：[www.hzbook.com](http://www.hzbook.com)

电子邮件：[hzsj@hzbook.com](mailto:hzsj@hzbook.com)

联系电话：(010) 88379604

联系地址：北京市西城区百万庄南街1号

邮政编码：100037



华章科技图书出版中心

# 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](http://pearsonhighered.com/sharda))，还有一个作者维护的独立网站 ([dssbibook.com](http://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](http://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（[teradata universitynetwork.com](http://teradata.universitynetwork.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](http://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](http://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](http://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](http://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](http://pearsonhighered.com/sharda) 找到。TestGen 可从 Respondus 获取，也可以从 [www.respondus.com](http://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](http://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 and 问题的测试，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, Prasoon Mathur, Rupesh Agarwal, Shubham Singh, Nan Liang, Jacob Pearson, Kinsey Clemmer 和 Evan Murette（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, Universality 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：Axiom (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。

# 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、商务分析和管理信息超载的技术。

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) 斯派尔商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有关的研究项目, 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。

# 目 录

<b>第 1 章 商务智能、商务分析和数据科学概述</b> .....	3
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 .....	4
1.2 变化的商业环境，决策支持和商务分析的需求变化 .....	11
1.3 计算机决策支持向商务分析和数据科学的演变 .....	13
1.4 商务智能框架 .....	15
商务智能的定义 .....	16
商务智能的简史 .....	16
商务智能的架构 .....	16
商务智能的起源和驱动力 .....	16
<b>应用案例 1.1 Sabre 通过仪表盘和分析帮助客户</b> .....	18
商务智能多媒体应用 .....	19
事务处理与分析处理 .....	19
进行适当的规划并与企业战略保持一致 .....	20
实时按需的 BI 实现 .....	21
开发或获取 BI 系统 .....	21
合理性和成本效益分析 .....	22
安全和隐私保护 .....	22
系统和应用集成 .....	22
1.5 商务分析概述 .....	22
描述性分析 .....	24
<b>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</b> .....	24
<b>应用案例 1.3 西门子通过使用数据可视化降低成本</b> .....	25
预测性分析 .....	25
<b>应用案例 1.4 运动损伤分析</b> .....	26
规范性分析 .....	26
<b>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</b> .....	27
商务分析应用于不同领域 .....	27
商务分析或数据科学 .....	28
1.6 所处领域的商务分析实例 .....	29
应用于医疗保健的商务分析——Humana 实例 .....	29
零售价值链中的商务分析 .....	33
1.7 大数据分析简介 .....	35
什么是大数据 .....	35
<b>应用案例 1.6 CenterPoint Energy 使用实时大数据分析改善客户服务</b> .....	37
1.8 商务分析生态系统概述 .....	37
数据生成基础设施提供商 .....	39
数据管理基础设施提供商 .....	39
数据仓库提供商 .....	40
中间件提供商 .....	40
数据服务提供商 .....	40
专注于商务分析的软件开发人员 .....	41
应用开发人员：特定行业或一般行业 .....	42
商务分析行业分析师和有影响力者 .....	43
学术机构和认证机构 .....	44
监管者和政策制定者 .....	45
分析用户组织 .....	45
本书计划 .....	46
资源、链接和 Teradata 大学网络连接 .....	47

## 第2章 描述性分析 I：数据的性质、统计建模与可视化 ..... 53

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 ..... 54

2.2 数据的性质 ..... 57

2.3 数据的简单分类 ..... 61

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 ..... 63

2.4 数据预处理的艺术与科学 ..... 65

应用案例 2.2 通过数据驱动分析提高学生保留率 ..... 68

2.5 用于商务分析的统计建模 ..... 74

用于描述性分析的描述性统计 ..... 75

集中趋势度量（也可以称作位置或中心度量） ..... 76

算术平均数 ..... 76

中位数 ..... 77

众数 ..... 77

离散趋势度量（也可称为散布或分散度量） ..... 77

极差 ..... 78

方差 ..... 78

标准差 ..... 78

平均绝对偏差 ..... 78

四分位数与四分位距 ..... 78

箱图 ..... 79

分布形状 ..... 80

技术洞察 2.1 如何使用 Microsoft Excel 进行描述性统计 ..... 82

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 ..... 84

2.6 用于推断性统计的回归建模 ..... 86

我们如何开发线性回归模型 ..... 87

我们如何知道模型是否足够好 ..... 88

什么是线性回归最重要的假设 ..... 89

逻辑回归 ..... 90

应用案例 2.4 预测 NCAA 碗赛结果 ..... 91

时间序列预测 ..... 96

2.7 业务报表 ..... 98

应用案例 2.5 纸质报表的洪水结束于 FEMA ..... 100

2.8 数据可视化 ..... 101

数据可视化简史 ..... 101

应用案例 2.6 Macfarlan Smith 使用 Tableau Online 提高运营绩效洞察 ..... 103

2.9 不同类型的图表 ..... 106

基本图表 ..... 106

专用图表 ..... 107

你应该使用哪种图表 ..... 108

2.10 可视化分析的兴起 ..... 110

技术洞察 2.2 商务智能和分析平台的 Gartner 魔力象限 ..... 111

可视化分析 ..... 112

高效率可视化分析环境 ..... 112

技术洞察 2.3 通过数据和可视化讲出色的故事 ..... 113

2.11 信息仪表盘 ..... 117

应用案例 2.7 Dallas Cowboys 使用 Tableau 与 Teknion 获得高分 ..... 118

仪表盘设计 ..... 119

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 ..... 119

在仪表盘中要寻找的内容 ..... 121

仪表盘设计的最佳实践 ..... 121

符合行业标准的基准关键绩效指标 ..... 121

使用情境的元数据包装仪表盘指标 ..... 121

符合行业标准的基准关键绩效指标 ..... 121

使用情境的元数据包装仪表盘指标 ..... 121

使用情境的元数据包装仪表盘指标 ..... 121

使用情境的元数据包装仪表盘指标 ..... 121

通过可用性专家验证仪表盘设计 .....	122	3.6 数据仓库的开发 .....	150
把传输到仪表盘的报警或		应用案例 3.3 使用 Teradata 分析的	
异常按优先级划分并排序 .....	122	SAP 解决方案加速大	
以业务用户评论丰富仪表盘 .....	122	数据交付 .....	151
从三个不同层次呈现信息 .....	122	数据仓库开发方法 .....	153
使用仪表盘设计原则选取		数据仓库开发的其他思考 .....	156
正确的视觉构造 .....	122	数据仓库中的数据表示 .....	156
为指导性分析做准备 .....	122	技术洞察 3.1 托管数据仓库 .....	157
		数据仓库中的数据分析师 .....	158
<b>第 3 章 描述性分析 II: 商务智能和</b>		OLAP 和 OLTP .....	158
<b>数据仓库</b> .....	127	OLAP 操作 .....	159
3.1 开篇小插曲: 利用商务智能和		3.7 数据仓库的实施问题 .....	160
数据仓库定位税务欺诈 .....	128	大型数据仓库和可扩展性 .....	162
3.2 商务智能与数据仓库 .....	130	应用案例 3.4 EDW 帮助连接密歇根	
什么是数据仓库 .....	131	的政府机构 .....	163
数据仓库的历史视角 .....	132	3.8 数据仓库管理、安全问题和未来	
数据仓库的特征 .....	133	趋势 .....	164
数据集市 .....	134	数据仓库的未来 .....	165
操作数据存储 .....	135	技术洞察 3.2 数据湖 .....	166
企业数据仓库 .....	135	3.9 企业绩效管理 .....	170
元数据 .....	135	闭环 BPM 周期 .....	171
应用案例 3.1 更好的数据计划: 著名		应用案例 3.5 AAPR 改造其 BI 基础	
的 TELCO 公司使用数		设施, 并在三年内达到	
据仓库和商务分析技术		347% 的投资回报 .....	173
在竞争激烈的行业中保		3.10 绩效度量 .....	175
持领先地位 .....	135	关键绩效指标 .....	176
3.3 数据仓库过程 .....	137	绩效度量系统 .....	177
3.4 数据仓库架构 .....	139	3.11 平衡记分卡 .....	177
可选的数据仓库架构 .....	142	四个视角 .....	177
哪种架构最好 .....	144	平衡在平衡记分卡中的意义 .....	179
3.5 数据集成以及提取、转换和加载		3.12 作为绩效度量系统的六西格玛 .....	179
过程 .....	145	DMAIC 绩效模型 .....	180
数据集成 .....	146	平衡记分卡与六西格玛 .....	180
应用案例 3.2 BP Lubricants 公司		有效的绩效度量 .....	181
实现 BIGS 成功 .....	146	应用案例 3.6 Expedia.com 的客户	
提取、转换和加载 .....	148	满意度记分卡 .....	182

## 第4章 预测性分析 I: 数据挖掘的过程、方法和算法

- 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

189

- 4.2 数据挖掘概念与应用
- 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

193

194

定义、特征和优势

数据挖掘原理

196

197

- 应用案例 4.2 戴尔在 21 世纪通过分析保持敏捷和高效
- 数据挖掘与统计学

203

4.3 数据挖掘应用

203

- 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

205

4.4 数据挖掘过程

206

步骤 1: 业务理解

207

步骤 2: 数据理解

208

步骤 3: 数据准备

208

步骤 4: 建模

209

- 应用案例 4.4 数据挖掘有助于癌症研究
- 步骤 5: 测试和评估
- 步骤 6: 部署
- 其他数据挖掘标准化过程和方法

209

212

212

4.5 数据挖掘方法

215

分类

215

估计分类模型的准确度

216

- 应用案例 4.5 InfluenceHealth 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
- 数据挖掘聚类分析

223

225

关联规则挖掘

227

4.6 数据挖掘软件工具

231

- 应用案例 4.6 数据挖掘来到好莱坞:

预测电影的商业成功

233

- 4.7 数据挖掘的隐私问题、谬误和隐患

237

- 应用案例 4.7 预测客户购买模式——

Target 公司的故事

238

数据挖掘谬误和隐患

238

## 第5章 预测性分析 II: 文本、Web 以及社交媒体分析

- 5.1 开篇小插曲: 《Jeopardy》上的人机大战: Watson 的故事

247

5.2 文本分析与文本挖掘概述

251

技术洞察 5.1 文本挖掘术语

253

- 应用案例 5.1 保险集团通过文本挖掘解决方案来加强风险管理

254

5.3 自然语言处理

255

- 应用案例 5.2 AMC Networks 正在使用分析为多渠道世界中的广告商捕获新的观众、预测评级和增加价值

257

5.4 文本挖掘应用

261

营销应用

261

安全应用

261

应用案例 5.3 挖掘谎言

262

生物学应用

264

学术应用

266

- 应用案例 5.4 将客户带入质量方程:

联想使用分析重新构思设计

266

5.5 文本挖掘过程

268

任务 1: 建立语料库

269

任务 2: 创建词项 - 文档矩阵	269	社交网络分析	304
任务 3: 提取知识	271	社交网络分析指标	305
应用案例 5.5 使用文本挖掘研究		应用案例 5.8 Tito 的伏特加酒通过	
文献综述	273	可靠的社交策略建立	
5.6 情感分析	276	品牌忠诚度	305
应用案例 5.6 创造独特的数字体验		社交媒体分析	309
来享受温网中的精彩		人们如何使用社交媒体	310
瞬间	277	度量社交媒体的影响	311
情感分析应用	280	社交媒体分析的最佳实践	311
情感分析过程	282	第 6 章 规范性分析: 优化和模拟	319
极性识别方法	284	6.1 开篇小插曲: 费城学区通过	
使用词典	284	使用规范性分析找到巴士路	
使用训练文档集	285	线的最佳解决方案	320
识别句子和短语的语义倾向	286	6.2 基于模型的决策	322
识别文档的语义倾向	286	规范性分析模型案例	322
技术洞察 5.2 用于预测性文本挖掘		应用案例 6.1 通过决策支持系统	
和情感分析的大规模		向埃克森美孚下游	
文本数据集	286	进行最佳运输	323
5.7 Web 挖掘概述	287	识别问题和环境分析	324
Web 内容挖掘和 Web 结构挖掘	289	模型类别	324
5.8 搜索引擎	291	应用案例 6.2 Ingram Micro 使用	
搜索引擎剖析	292	商务智能应用进行	
搜索引擎优化	294	定价决策	325
搜索引擎优化方法	295	6.3 用于决策支持的数学模型的	
技术洞察 5.3 最流行的排名前 15 的		结构	328
搜索引擎 (2016 年 8 月)	296	决策支持数学模型的组件	329
应用案例 5.7 理解为什么客户放弃		数学模型的结构	329
购物车使销售额增加		6.4 确定性、不确定性以及风险	
了 1000 万美元	297	分析	330
5.9 Web 使用挖掘 (Web 分析)	298	确定性下的决策	330
Web 分析技术	299	不确定性下的决策	331
Web 分析指标	300	风险分析	331
网站可用性	300	6.5 使用电子表格进行决策建模	331
流量来源	301	应用案例 6.3 美国航空使用成本	
访客特征	302	建模来评估出货路	
转化统计	302	线的不确定性	332
5.10 社交分析	304		

应用案例 6.4 宾夕法尼亚州收养交易 所使用电子表格模型更 好地匹配儿童与家庭	333	传统模拟的不足	359
应用案例 6.5 Metro Meals on Wheels Treasure Valley 使用 Excel 来寻找最佳的 配送路线	334	视觉交互式模拟	359
6.6 数学规划优化	336	视觉交互式模型和决策支持系统	360
应用案例 6.6 混合整数规划模型 帮助田纳西大学医 学中心调度医生	337	模拟软件	360
线性规划模型	338	应用案例 6.9 通过 RFID 改进作业 车间调度决策：基于 模拟的评估	361
线性规划中的建模：一个例子	339	第 7 章 大数据概念和工具	369
实现	344	7.1 开篇小插曲：使用大数据方法 分析电信公司的客户流失	370
6.7 多目标、敏感性分析、假设分 析和目标寻求	346	7.2 大数据的定义	373
多目标	346	大数据的特征定义	374
敏感性分析	347	技术洞察 7.1 数据规模越来越大， 越来越大	375
假设分析	348	应用案例 7.1 市场分析或预测的替 代数据	377
目标寻求	348	7.3 大数据分析的基础	378
6.8 基于决策表和决策树的决策 分析	349	大数据分析解决的业务问题	381
决策表	350	应用案例 7.2 前五大投资银行实现 了真正的单一来源	382
决策树	351	7.4 大数据技术	383
6.9 模拟概论	352	MapReduce	383
模拟的主要特征	352	为什么使用 MapReduce	385
应用案例 6.7 乙型肝炎干预的模 拟效果	353	Hadoop	385
模拟的优点	354	Hadoop 如何工作	385
模拟的缺点	355	Hadoop 技术组件	386
模拟方法	355	Hadoop 的优缺点	387
模拟类型	356	技术洞察 7.2 关于 Hadoop 的一些 神秘事实	388
蒙特卡罗模拟	357	NoSQL	389
离散事件模拟	358	应用案例 7.3 eBay 的大数据解决 方案	390
应用案例 6.8 Cosan 使用模拟改善 其可再生能源供应链	358	应用案例 7.4 了解 Twitter 上医疗 保健信息的质量和 可靠性	392
6.10 视觉交互式模拟	359	7.5 大数据和数据仓库	393

Hadoop 用例	393	应用案例 8.2 Rockwell Automation	
数据库用例	394	监控昂贵的石油和天	
灰色区域	395	然气探测资产	421
Hadoop 和数据仓库共存	396	IoT 技术基础设施	422
7.6 大数据供应商和平台	397	RFID 传感器	422
IBM Info Sphere BigInsights	398	雾计算	425
应用案例 7.5 使用社交媒体预测流		IoT 平台	426
感活动	400	应用案例 8.3 Pitney Bowes 与 General	
Teradata Aster	401	Electric IoT 平台合作优	
应用案例 7.6 从电子病历数据仓库		化生产	426
分析疾病模式	402	IoT 初创生态系统	427
技术洞察 7.3 如何在大数据上取得		物联网的管理注意事项	428
成功	405	8.3 云计算和商务分析	429
7.7 大数据和流分析	406	数据即服务 (DaaS)	431
流分析与永久分析	408	软件即服务 (SaaS)	432
关键事件处理	408	平台即服务 (PaaS)	432
数据流挖掘	408	基础设施即服务 (IaaS)	432
7.8 流分析的应用	409	云计算的基本技术	433
电子商务	409	云部署模型	433
电信	409	主要分析云平台提供商	434
应用案例 7.7 Salesforce 正在使用		分析即服务 (AaaS)	435
流数据提升客户价值	410	代表性的分析即服务产品	435
执法和网络安全	411	使用云基础设施的说明性	
电力行业	411	分析应用	436
金融服务	411	8.4 为组织提供基于位置的分析	441
健康科学	411	地理空间分析	441
政府	412	应用案例 8.4 Great Clips 采用空间	
第 8 章 商务分析的未来趋势、		分析来减少位置决策	
隐私和管理思考	417	的时间	443
8.1 开篇小插曲: 传感器数据分析		应用案例 8.5 Starbucks 利用 GIS 和	
帮助西门子避免火车故障	418	分析在全球各地扩展	444
8.2 物联网	419	实时位置智能	445
应用案例 8.1 SilverHook 汽艇使用		应用案例 8.6 Quiznos 吸引顾客	
实时数据分析通知参		策略	446
赛者和赛艇迷	420	消费者的分析应用	446
		8.5 合法性、隐私和道德问题	448
		法律问题	448