

商业分析工具+案例+实践指南

普
华
经
管

PEARSON

APPLIED BUSINESS ANALYTICS

大数据商业分析 整合大数据与业务流程 的高级商业分析指南

[美]纳撒尼尔·林 (Nathaniel Lin) ◎著
赵玲◎译

Integrating Business Process,
Big Data, and Advanced Analytics

从统揽全局的商业战略到面面俱到的运营细节
全方位掌控商业分析价值链

 中国工信出版集团

 人民邮电出版社
POSTS & TELECOM PRESS

APPLIED BUSINESS ANALYTICS

Integrating Business Process,
Big Data, and Advanced Analytics

大数据商业分析

整合大数据与业务流程
的高级商业分析指南

[美]纳撒尼尔·林 (Nathaniel Lin) ©著
赵玲◎译

人民邮电出版社
北京

图书在版编目 (C I P) 数据

大数据商业分析：整合大数据与业务流程的高级商业分析指南 / (美) 林 (Lin, N.) 著；赵玲译. — 北京：人民邮电出版社，2016. 4

ISBN 978-7-115-42025-1

I. ①大… II. ①林… ②赵… III. ①商业信息学
IV. ①F713. 51

中国版本图书馆CIP数据核字(2016)第053245号

内 容 提 要

在快速变化的时代，企业需要依靠与时俱进的商业分析和基于分析的决策来抓住未来的各种机遇，从而以快于竞争对手的速度适应商业发展趋势，成为市场的主导者而不是盲从者。

纳撒尼尔·林设计出了一个将商业分析融入商业流程中的架构，并剖析了具有实践性的分析工具和模型，以此帮助企业参透各类数据的复杂性并及时分析市场变化。本书不仅展示了一幅商业分析流程的全貌图，还阐述了发展和实现分析决策所需要的条件与过程。此外，本书还演示了数据驱动的变革如何推动组织发展和决策方式转变。本书在数据准备、数据分析和数据演绎几大方面，提供了丰富的案例研究以及有关KNIME分析工具的实操经验，这些内容可以引领企业决策者培养一种着眼于全局的思维，及时辨认商业机会并从容应对各种商业挑战。

本书对那些希望运用分析技术实现企业变革的商业人士尤为合适，同时适合各类企业和机构的决策者以及高等院校相关专业的师生阅读。

◆ 著 [美] 纳撒尼尔·林 (Nathaniel Lin)

译 赵 玲

责任编辑 姜 珊

执行编辑 田 甜

责任印制 焦志炜

◆ 人民邮电出版社出版发行

北京市丰台区成寿寺路 11 号

邮编 100164 电子邮件 315@ptpress.com.cn

网址 <http://www.ptpress.com.cn>

大厂聚鑫印刷有限责任公司印刷

◆ 开本：700×1000 1/16

印张：20

2016 年 4 月第 1 版

字数：151 千字

2016 年 4 月河北第 1 次印刷

著作权合同登记号 图字：01-2015-1887 号

定 价：65.00 元

读者服务热线：(010) 81055656 印装质量热线：(010) 81055316

反盗版热线：(010) 81055315

广告经营许可证：京东工商广字第 8052 号

本书的赞誉

当我第一次阅读纳撒尼尔·林的《大数据商业分析》的时候，我就希望向各位推荐这本书。纳撒尼尔对于分析决策的观点正中了一个独特的目标群体——那些还不理解如何使用数据来驱动商业的经理人。分析决策正处于交叉口的位置，发挥着如曼哈顿市中心交通警察一般的作用，使创新贯穿于整个组织成为可能。作为建立一支分析团队的重要参考书，这本书将是那些值得保留在你书桌上的书之一。这本书最让我喜爱的部分是高级分析层（具有直接的适用性）、案例研究（非常注重实践）以及KNIME的手把手实践操作教学（来自真实世界的经验）。

——特雷莎·库什纳（Theresa Kushner），威睿公司（VMware）企业信息管理副总裁，思科公司（Cisco）前顾客智能部门总监

纳撒尼尔干了一件了不起的事，他捕捉到了驱动业务增长并实现高效的商业分析所需要的关键零件。本书是参与组织决策分析过程的高层管理

人员的必读书目。

——约翰·卡特 (John Carter), 嘉信理财集团 (Charles Schwab) 分析洞察和忠诚研究高级副总裁

作为分析研究的主流, 很多书研究了组织的分析实践, 然而《大数据商业分析》在两个方面显出了它的与众不同。首先, 它解释了为什么现实世界中的许多组织在将杠杆分析转换成实践的时候总是失败。其次, 通过持续的对话和修正, 它设计出了一个真正将分析融入商业流程中的架构, 而不只是孤立地介绍分析实验。如果你想要利用现今大数据和分析的杠杆作用实现组织转型和领导组织发展, 作为高瞻远瞩的企业领袖, 请将这本书的核心观点铭记心中, 你必将获益匪浅。

——普拉特·莫赫 (Prat Moghe), 卡泽纳公司 (Cazena) 首席执行官, IBM-Netezza 公司前高级副总裁

商业分析已经发展成一个备受企业和组织推崇和关注的领域, 如今, 它的成功应用更多地归功于执行艺术, 而不是科技。本书在弥补这一遗憾的道路上迈出了具有重要意义的一步, 它指导企业如何建立优秀的团队和设计业务流程, 使商业领袖做出成功的商业分析成为可能。商业分析本质上是一个生涩、复杂的技术性话题, 而纳撒尼尔没有将技术性的具体细节草草带过, 而是通过这本书, 展示了内容丰富且切实可行的实践过程。本书将分析技术回归到了本源——为相关的商业问题提供清晰的服务, 为分析项目提供必要的内容, 以及创造出有意义的价值。林博士在本书中不仅指导企业如何做分析, 还展示了一幅完整的分析组织的全貌图。同时, 他

阐述了发展和实现分析决策所需要的条件和过程，并演示了数据驱动的转变是如何推动组织发展的，实现了从以直觉为基础的决策方式向以客观证据为基础的理性决策方式的转型。本书囊括了真实世界的案例研究和精彩故事，对于任何一位想要运用分析技术实现组织变革的高层经理人来说，纳撒尼尔·林的《大数据商业分析》都会是一个可信赖的学习起点。

——亚当·费拉里（Adam Ferrari），克里斯皮里公司（Crisply）前首席技术官，甲骨文公司（Oracle）大数据战略顾问和产品发展部副总监，恩德卡科技公司（Endeca Technologies）前首席技术官

如果一位管理者想要建立数据驱动的企业文化，那么纳撒尼尔在本书中演示的清晰、透彻的分析过程和系统的方法论将让他受益匪浅。其中，分析生态系统这一章节对现代社会中不同团队的组织结构的比较，更是无与伦比的美妙。

——陈洁莹（Jane Chen），ADI 亚德诺半导体技术公司（Analog Devices）卓越分析中心主任

纳撒尼尔的这本书给我带来了积极的影响。这本书在商界具有独特的地位，它以其富有洞察力的观点脱颖而出，是一本平衡和综合了“行业圣经”和“技术指导手册”的好书。我非常欣赏纳撒尼尔深思熟虑的领导力和指导力。

本书由 70% 的关于商业和技术的讨论，加上 30% 的关于如何使系统模型更有效这两部分内容组成。在现今的世界中，分析决策最具挑战的部分仍然是商业驱动因素和技术细节之间的冲突和矛盾。这也是分析决策如此

大数据商业分析

整合大数据与业务流程的高级商业分析指南

重要的原因——它能使分析在商业实践中的效用最大化。如何平衡技术细节（怎么做）？为什么分析对你的成功至关重要？通过建立分析能力你能获得什么效果（期望成为的）？这些问题都是极具挑战性的。纳撒尼尔的个人故事、案例研究和真实世界的成功案例对理解这些问题有很大的帮助。我相信这本书对于新兴的分析领袖来说，将是伟大的参考指南。

——加里·曹（Gary Cao），康德乐公司（Cardinal Health, Inc.）信息服务和战略分析部副总监

推荐序

本书的出版在很多方面都符合当今时代的发展趋势，它是当今商业分析时代思潮的产物。如今，分析和大数据已经成为全世界公认的商业战略和商业竞争工具。事实上，许多调查显示，大部分经理人都赞同分析和大数据是商业活动中的重要工具和资源。

随着这个共识得到广泛认可，目前存在的问题是，许多经理人和组织仍然缺乏让分析为己所用的技能和知识。纳撒尼尔·林的这本书恰恰回答了这个问题。如他所说，许多书籍致力于开发经理人的热情，让他们对那些介绍深奥、晦涩的分析方法和分析技术细节的教科书产生好奇乃至兴奋的情绪，却没有介绍太多商业方面的内容。而有些经理人已经相信分析的重要性，但还不具备足够的知识和能力亲自将分析付诸实践，对这部分人而言，本书恰好是那些极端书籍之间的一个过渡。

这个过渡很重要，实际上也是必不可少的。在错综复杂的企业环境中，要在分析方法和业务模式之间划出一条清晰的分界线非常困难。因为商业

即分析，分析即商业。本书致力于让普通商业人士创造和运用分析方法，乃至高级分析方法。详细的建模和数据分析的重要性也许将继续提升，甚至成为商业之船的“篙”，但是正如纳撒尼尔·林和一些定量分析专家所提到的，在框架构建和数据演绎的分析工作中，真正重要的仍然是决策者之间的高度合作。

这本书符合现代分析环境的一个表现是，它强调了商业活动和分析的过程。纳撒尼尔·林正确地指出，分析正在日益深入到我们的关键业务中。从这个意义上来说，分析在某种程度上变得隐形。事实上，高德纳咨询公司“无所不在却隐于无形的先进分析技术”命名为“2015年对企业组织而言最重要的十大策略科技趋势之一”。这个无形性越发凸显出管理人员对分析的理解的重要性。如果你的业务在很大程度上依赖于基础结构流程和基层技术的分析，那么就需要有人能够持续地提出关于基础假设、方法和结果的明智的问题。

为了加强流程方面的培训，本书从数据准备、数据分析到数据演绎，按照分析流程的逻辑组织了许多话题和观点。有关流程的思维方式十分符合如今的分析环境。在很长一段时间里，分析已经被追捧为一个缓慢的、一次性的、艰辛的过程，似乎难以制度化。但是，今时今日许多公司正在开始创造“分析工厂”，强调分析的速度、规模和可重复性。分析工厂的意义是将分析解剖成一个具体流程，这个流程包含清晰的步骤、接口和模块。这正是本书的焦点所在，它使用开放源代码分析工具 KNIME 的案例作为支持，而 KNIME 本身就是按照流程步骤构建的。如果我们采纳纳撒尼尔·林的基于流程思维的模型和分析，那么我们的分析将促进生产率的提升并产生更大的影响。

这本书和现代主流分析环境之间的最后一个联系是聚焦于大数据和小数据这两个方面的内容。现在的大部分书籍是基于传统的小数据分析或大数据分析的，但很少有书籍能够兼顾两者并跨越两者之间的断层。纳撒尼尔·林游刃有余地游荡于众多数据类型之间，包括传统数据分析（如营销倾向型模型）、社交媒体，以及线上产品评价的文本数据等。在我看来，用不了太长时间，所谓的大数据思维毫无疑问将消失在时代的洪流中，但是，我们仍将确定无疑地需要继续处理大容量的、未组织的原始数据。无论如何，这个世界正逐渐产生和整合各种类型的数据。阅读本书后，你将为这个逐渐形成的新世界做好准备。纳撒尼尔·林提供了一系列恰当和有价值的数据元素；同时，一个组织在不同业务领域里可能运用到的典型分析均被囊括在了本书中。

本书不仅仅是时代的产物，它还具有极其广泛的适用范围，几乎涉及了所有领域的商业分析。当你获得这本书的时候，可能迫不及待地想要快速地阅读它，然后将它好好地保存起来，作为你在往后数年里遇到各种分析情况时必备的参考书之一。本书的观点既古典又现代，在很长一段时间里，这些观点都不会过时或脱离人们的需要。

托马斯·达文波特（Thomas Davenport）

巴布森学院（Babson College）特聘教授

前言

我不得不承认，这本书的撰写过程比我原来设想的要困难得多。经过数十年的分析实践以及在建立和领导分析团队之中的磨炼，我以为要将以前学到的和做过的东西总结起来会非常容易。然而，我发现自己就像小伙子遇到了心仪的姑娘一样，总是踌躇不前、思前想后。我总是在想，是不是涉及了太多的技术细节，或者是太少了？在我看来，后者更糟，因为细节的匮乏可能让商业读者无法亲自将分析过程付诸实践。可是，在书中写太多的技术细节又会阻碍商业读者把这本书读下去。

在几轮修订以后，我希望我已经取得了一个比较好的平衡，实现了给商业读者写一本商业分析实操指南书这个最初的写作目标。然而，如果你还是在阅读本书的过程中发现太专业或者还不够专业的内容，我衷心恳求你的谅解。我给你的建议是，请简单地略过那部分内容，直接进入到一个部分。我确信这本书里有足够多的例子和商业案例，可以让任何一位胸怀抱负、想要提升分析技能的严谨的商业读者有所收获。

为什么写一本关于分析的书

虽然市面上有很多书籍的标题都带有“分析”两个字，但是一项简单的调查显示，大部分书籍可以粗略地分为两类。一类书主要推广“商业如何受益于分析”这一观点，这类书常常由“传道者”们撰写，内容主要是宣传分析的好处；另一类书则是“宽客”（Quants, 又称量化投资者）写给其他分析型“宽客”看的。然而，很少有书是介于两者之间，并在两者之间架设桥梁的。读者有一种切实存在的需求，那就是希望有一本商业分析的实操指南书，既能解决现实中的商业疑难问题，又易于商业读者理解、练习和使用。我经常听到商业读者感叹：“阅读了那些宣传分析力量的书之后，我被说服了，可是没有人能告诉我怎么做和如何真正地将它运用到我的实际工作中，我想要的结论并不是重新回到学校去攻读一个分析和统计学的高等学位！”如果你有相同的看法，那么这本书正是为你写的。

本书结构安排

本书的前四章介绍了分析的基础内容，包括数据、分析工具（从简单的到前沿的）以及它们与各种商业流程之间的关系。随后，本书将继续介绍如何在商业中嵌入分析，以及如何将现有的或新的战略流程中的发现整合到分析中。各章节的具体安排如下。

- 第1章“导论”，是关于分析的介绍。本章展示了分析是数据的精炼，它将彻底改造商业活动模式，正如石油的精炼在20世纪掀起的变革。本章还介绍了书中涉及的一些概念和定义、潜在的应用领域、一些重要的学习经验、分析在商业背景下如何定位，以及分析

是如何被整合到商业流程中的。

- 第2章“了解你的原料——大数据和小数据”。本章讨论了你需要知道的成功进行商业分析所涉及的每一个基本原料。这部分内容界定了大数据，以及各种可能有用的商业分析数据类型之间的不同。本章还介绍了数据格式，低结构化或非结构化数据对商业洞察与日俱增的重要性，以及处理这些数据的方法。
- 第3章“数据管理——数据整合、数据质量和数据治理”。本章介绍了企业管理和分析数据所需要掌握的一切：处理不同数据的流程、数据安全和隐私管理的重要性、好数据和坏数据的对比、数据质量和延时性，以及超越内部数据源的方式。
- 第4章“掌握工具——分析方法和工具”。本章讨论了常用的分析模型和工具，并提供了一些常用的分析建模方法的实操练习，帮助你看见和感受分析工具在真实的商业案例中是如何运作的。这部分内容包括基本模型，比如回归、集聚分析、决策树，以及较复杂的大数据分析，比如文本挖掘和基于真实数据的情感分析。案例包括亚马逊精美食品（Amazon Fine Food）和猫途鹰旅行社区（TripAdvisor Hotels）的在线评论。本章还介绍了运用集成回归树模型（Ensemble Regression Tree）进行价值预测的方法。
- 第5章“分析决策流程和分析决策者”。本章解释了传统决策和分析决策过程的不同之处。以研究传统决策过程为何常常出现各种弊端为开端，本章描述了分析决策的过程，即商业分析流程（Business Analytics Process, BAP），并介绍了分析决策者是如何在管理体制的“交叉口”发挥其作用，并确保商业分析流程取得成功的。本章最后一部分内容指出了如何成为、定义、招募和留住分析决策者。

- 第6章“商业流程和分析”。本章讨论了哪些商业流程能从分析中受益，包括今天以分析为动力的企业资源规划（Enterprise Resource Planning, ERP）、客户关系管理（Customer Relationship Management, CRM）、供应链管理（Supply Chain Management, SCM）、财务管理（Financial Management）、产品生命周期管理（Product Lifecycle Management, PLM）和人力资本管理（Human Capital Management）。
- 第7章“通过辨认模式发现商业机会”。本章解释了为什么现有的划分消费者细分市场的方式还不够。本章将告诉你如何利用高级分析工具来发现影响商业行为的模式、事件、趋势、话题、偏好和其他重要因素。本章还讨论了模式检测，通过从购物篮分析（Market Basket Analysis）中出现的观点和消费者细分市场的人物形象特征，以及消费者群体行为，帮助营销人员设计更加有效的客户接触策略。
- 第8章“了解不可知之事”。本章讨论了普遍断定为不可知和未知的（Unknowable or Unique, UU）话题和问题。本章内容包括，在拥有极少或有限数据的情况下，如何实时预测个人行为，以及确定复杂环境中的因果关系及杠杆设置。本章提供了现实生活中的商业案例和真实的分析流程，举例说明了开明的企业如何在这些明显的不可知事件中了解和利用分析，从而在获得竞争优势。本章还介绍了一种新奇的方法来预测消费者的消费容量，以及如何使用这种方法形成有效的战略。
- 第9章“商业分析流程实例——分析企业”。这部分内容用正确的材料和工具来回答除了客户关系管理以外的重要商业问题，客户关系管理的相关问题将在第10章进行详细讨论。我鼓励你下载和修改书中提供的工作流程，尝试进行实际操作并将其运用到你自己的需求中。

- 第 10 章“商业分析流程实例——分析客户关系管理”。这部分内容介绍了如何使用正确的材料和工具来回答关于客户关系管理的重要商业问题。本章提供了如何深入了解消费者的建议——他们是谁？他们的价值观、侧面介绍、人物形象特征是什么？在哪里和如何找到他们？以及他们的需要和需求行为是什么？所有这一切都是通过分析获得的。本章提出了一些重要的观点，如消费者购买偏好和模式、忠诚度和即将发生的变革，以及获得新顾客的方式，要获得以上信息，企业可以利用当下的社交网络、移动设备和其他媒体，从而实现成本效益。我再次鼓励你尝试填写包含分析模块的商业分析流程并下载相关的工作流程，尝试使用你自己的数据并应用到实际业务中。
- 第 11 章“分析竞争力和生态系统”。本章讨论了企业现在的竞争情况，以及如何将竞争提升到更高层次。本章还介绍了一个健康的分析生态系统的重要性。本章提供了一个企业分析战略规划的例子，内容包括组织结构、人才管理、人员的角色和责任、运行一个有效的分析组织的方法，以及如何在分析生态系统中利用外部资源和合作伙伴。
- 第 12 章“总结以及现在怎么办”。本章是对相关知识和建议的总结，它可以指导你接下来应该做些什么，从而实现分析竞争优势并制定成熟的战略路线图。本书还附上了一个推荐清单，帮助读者在各种领域获得分析竞争优势。
- 附录包含着重要内容，但是对于本书的主体而言太过详细而不宜放入其中。这部分内容包括 KNIME 分析工具的基本知识和一些关于 KNIME 关键环节的描述。你可以访问 KNIME.com 和 [YouTube](https://www.youtube.com/channel/UC8v11111111111111111111) 网站，通过观看入门指南来获得更多的信息。

本书的作用

阅读本书之后，你应该能够：

- 对整个商业分析过程具有一个充分的理解；
- 掌握商业分析的专业词汇并了解商业分析模型运作的基本知识；
- 研究商业分析模型以及能够修正或建立现实生活中复杂的分析模型；
- 结合商业和分析过程，深入挖掘或整合任何级别的工作流程，紧密跟踪分析过程中的数据、结果和见解；
- 与分析团队进行更好的交流和沟通；
- 成为快速建模研究分析团队中卓有贡献的成员，能判断商业发展方向；
- 从看起来无关紧要的分析细节中提炼出可付诸实践的商业观点；
- 确保商业活动的重点与最终交付结果之间的平衡，并最终成为一位高效的分析领导者和决策者。

最后，祝你阅读愉快！

目录

第1章 导论 ↘ 1

新的石油——原始数据 ↘ 3

分析 10 大商业问题 ↘ 7

重要的学习经验 ↘ 14

连接分析和商业 ↘ 15

小结 ↘ 19

第2章 了解你的原料——大数据和小数据 ↘ 21

垃圾进，垃圾出 ↘ 23

数据和大数据 ↘ 24

数据类型 ↘ 26

数据格式 ↘ 41

小结 ↘ 45