从海量数据中挖掘出有用的知识是一件**有趣**的事儿
看上去晦涩难懂的算法，也有**接地气**的一面
只要找对学习方法和案例
数据挖掘与机器学习也可以**像听故事一样**有趣

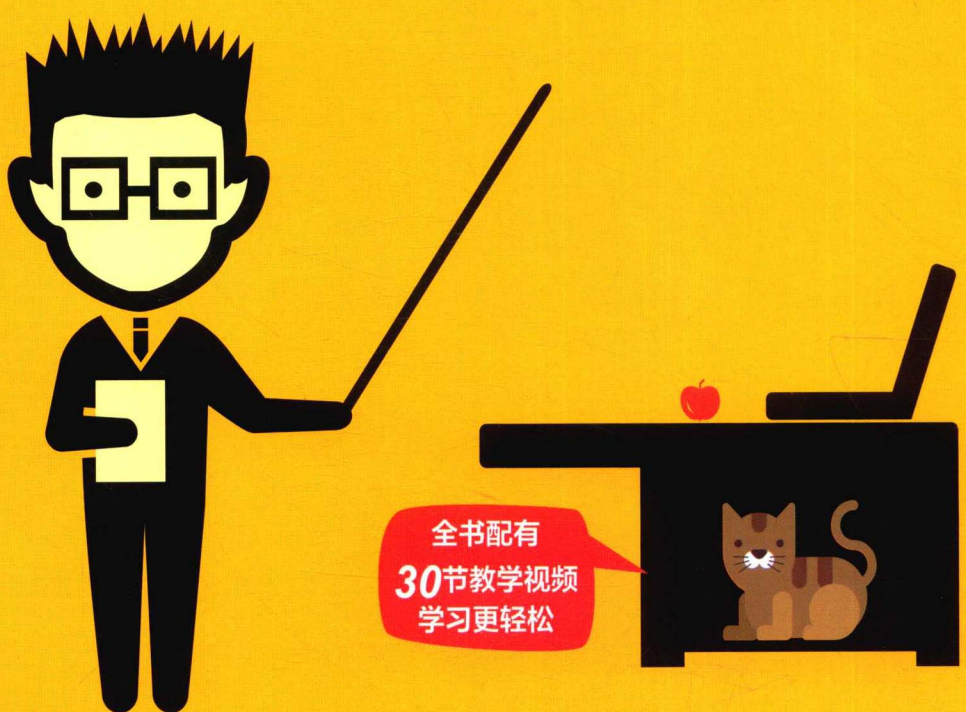
Broadview®
www.broadview.com.cn

小白学 数据挖掘与机器学习

18位业内专家联合力荐

SPSS Modeler案例篇

张浩彬 著



中国工信出版集团
電子工業出版社
PUBLISHING HOUSE OF ELECTRONICS INDUSTRY
http://www.phei.com.cn

小白学数据挖掘与机器学习

SPSS Modeler案例篇

张浩彬 著

電子工業出版社
Publishing House of Electronics Industry
北京•BEIJING

内 容 简 介

本书用生活中常见的例子、有趣的插图和通俗的语言，把看上去晦涩难懂的数据挖掘与机器学习知识以通俗易懂的方式分享给读者，让读者从入门学习阶段就发现，原来数据挖掘与机器学习不但有用，还很有趣。

本书以 IBM SPSS Modeler 作为案例实践工具，首先介绍了数据挖掘的基本概念及数据挖掘方法，然后介绍了 IBM SPSS Modeler 工具的基本使用、数据探索、统计检验、回归分析、分类算法、聚类算法、关联规则、神经网络以及集成学习。每一章都会以漫画形式介绍一些日常小例子并作为切入点，用通俗的语言介绍具体的算法理论，同时在每章最后都附上应用案例，让读者更轻松地阅读本书并掌握对应的算法和实践操作。

全书内容循序渐进，完整覆盖了数据挖掘与机器学习的主要知识点，适合数据挖掘与机器学习入门读者阅读。

图书在版编目（CIP）数据

小白学数据挖掘与机器学习. SPSS Modeler 案例篇 / 张浩彬著. —北京：电子工业出版社，2018.7
ISBN 978-7-121-33843-4

Ⅰ. ①小… Ⅱ. ①张… Ⅲ. ①数据采集②机器学习 Ⅳ. ①TP274②TP181

中国版本图书馆 CIP 数据核字(2018)第 048262 号

策划编辑：王　静
责任编辑：王　静
印　　刷：三河市君旺印务有限公司
装　　订：三河市君旺印务有限公司
出版发行：电子工业出版社
　　北京市海淀区万寿路 173 信箱　邮编 100036
开　　本：787×980　1/16　印张：14.5　字数：298 千字
版　　次：2018 年 7 月第 1 版
印　　次：2018 年 7 月第 1 次印刷
定　　价：79.00 元

凡所购买电子工业出版社图书有缺损问题，请向购买书店调换。若书店售缺，请与本社发行部联系，联系及邮购电话：（010）88254888，88258888。

质量投诉请发邮件至 zlts@phei.com.cn，盗版侵权举报请发邮件至 dbqq@phei.com.cn。

本书咨询联系方式：010-51260888-819，faq@phei.com.cn。

专家推荐语

从事 SPSS Modeler 软件开发十年来，我第一次看到以独到的应用场景设置，图文并茂的形式介绍统计分析技术和数据挖掘方法的书。此书让初学者和行业应用人员从枯燥的公式中体验到数据挖掘的乐趣，在 AI 大行其道的今天，让读者如虎添翼。

王俊波 申有软件科技（上海）有限公司创始人 IBM 全球分析软件实验室前总经理

从机器学习到人工智能，往往让人第一个联想到的就是晦涩难懂。但这本书通俗易懂，轻松有趣，更重要的是做到了理论和实践兼备。现在的你不一定是专业人士，但要想了解什么是真正的机器学习和数据挖掘，相信此书能让你对算法理论及工具使用有更深入的理解。

刘胜利 IBM 大中华区大数据分析与认知产品技术总监

在大数据时代，我们早已看到数据分析技术在科技公司手中展现的非常魅力，而对我个人来说，更加重要的技能是能否在面对“大数据”时，使用一些“利器”去驾驭它。如果说算法和工具是驾驭“大数据”的利器，那么此书就是关于驾驭利器的“秘籍”。最为难得的是，这本“秘籍”在做到了系统化的同时，还做到了通俗化、趣味化，相信这将给各位读者带来不一样的感受。

华明胜 埃森哲数字服务大中华区董事总经理

IBM SPSS Modeler 在 IBM 业务分析产品家族内被定义为实现预测性分析的工具，通过图形化的“拖拉拽”形式，实现机器学习或深度学习、文本分析和地理空间分析，它能建议用

户使用合适的算法实现业务价值，尤其适合业务部门的人使用。在具体实现业务预测场景时，仍需要对算法实现细节有一定了解，庆幸浩彬老师写了这本好玩、易懂的书。他利用大量业余时间，怀着对机器学习极大的热情，从场景和工具层面给予数据小白以无私指导，从而能够帮助有类似需求的读者快速入门，熟练掌握机器学习这个利器，实现个人价值增值。

何军 IBM 南区大数据及认知计算技术负责人

浩彬在过去的几年完成了许多基于数据挖掘和分析的商业落地项目，涉及领域包括用户洞察、绩效预测和分析等。这两年，他又聚焦环保行业，在环境监管领域大数据分析方向深耕细作，沉淀积累良多。我也一直希望他能把其中精华部分加以提炼、升华，与大家一起分享。我相信此书仅仅是这一系列的开始，期待未来有更多的精彩内容分享。

龙力辉 广东柯内特环境科技有限公司大数据研究院 院长

本书以通俗易懂的语言，形象生动、生活化的案例，让数据挖掘这门数据科学变得更加有趣味。系统化的知识，强大的 SPSS Modeler，又让专业的数据挖掘技术变得有理可依，有器可行。知识在于学以致用，浩彬写的这本书可以帮助你解决数据挖掘工作中的难题。

梁勇 天善智能创始人

这是一本很有心思的机器学习入门书籍，不仅涵盖了主流的机器学习算法，而且在每章的开篇，都以一个日常生活中常见的例子作为引子，再将专业的算法理论娓娓道来，最后则以一个实战例子作为结束，不禁让人有一种想要一口气读完的冲动。

黄志洪 炼数成金创始人

在我学习数据分析那些枯燥的理论公式时，是多么希望有一本书能通过有趣、生动的案例来解惑。很多人在数据挖掘上一而再地弃“坑”，就是因为数理型的内容难以“啃”下去。浩彬老师这本书充满了趣味性，相信会给读者带来全新的体验，在 SPSS Modeler 的学习道路上打下坚实的基础。

秦路 资深互联网数据分析师

浩彬不仅精通统计分析，还擅长授业解惑；不仅是 SPSS 高手，还有讲好故事和画好漫画的本事。市面上的数据挖掘和机器学习书籍很多，一般来说，这些书要么偏重理论，要么偏重编程实现：前者适合于学术研究型读者，后者适合于程序员。然而，越来越多的企业需要数据分析人员能够准确地应用理论和快捷地使用工具进行商业分析，与理论和编程比起来，对商业和数据的深刻理解反而成为核心。在本书中，浩彬从常见的商业分析需求出发，由浅入深地讲解了数据分析和数据挖掘的核心理论，并演示了如何使用强大且方便的 SPSS Modeler 将这些理论应用到商业分析中。本书行文轻松、欢快，图文并茂，相信数据分析人员可以借助本书，快速领悟数据挖掘和机器学习的要素、步骤和技能。对于学术型或工程型读者，其实也可以通过本书学习如何更好地“讲课”，以及更好地从商业分析的视角应用数据挖掘。

郭鹏程 山东财经大学金融数学系

本书轻松、诙谐，又入木三分，把数据挖掘讲得如此有趣味，也唯有浩彬老师了。在白描中解释概念，在嬉戏中探索原理，想来也是我们理工科读者的一大幸事。

邹伟 人工智能专家，睿客邦 CEO

我一直认为 IBM SPSS Modeler 是数据建模非常重要的利器和里程碑，特别是随着人工智能领域中机器学习的突破，如今，我们需要把更多的关注力放在业务理解和价值上，而 SPSS Modeler 的简单，恰好形成它独特的优势——能让业务最快速地通过数据获取到价值。我认识浩彬是在 IBM 中国，这里是 SPSS Modeler 非常不错的实践地，有大量的行业用户，有独特专有的 IBM 内部文档，再加上浩彬所具有的传道授业的天资，在离开 IBM 不久就形成了这本浅显易懂的佳作，让希望进入人工智能这个行业的业务专家摇身一变成为今天最火的职业——数据科学家，而不必在算法的围墙外苦苦挣扎。

廖显 华为 GTS AMS 人工智能主任架构师/人工智能业务转型项目技术顾问，
IBM 大中华区云与认知技术生态前首席架构师

企业数字化转型浪潮的必然结果是，未来企业的核心竞争力建筑于数据资产之上，而数据资产化和变现价值的挖掘又离不开更多数据科学家的培养与成长。很高兴看到这样一本兼具专业性和实践性的大数据基础读本，有别于一般同类书过于强调理论讲述或工具操作的定位，本

书更为注重结合实例，图文并茂地展现逻辑原理与方法，适合帮助更为多样背景的朋友踏上数据科学家的探索与成长之路。

华晓亮 华兴力拓创始合伙人，开创消费大数据驱动时尚企业新零售成长的领域专家

一直很敬佩浩彬老师的专业性，并期待着老师的新书，但是看到书稿的时候还是很惊叹，数据挖掘的内容竟然被浩彬老师以这么生动易懂的方式表达出来，再结合可视化的挖掘工具 SPSS Modeler 的案例，对想要学习数据挖掘的读者来说，简直可以算完美了。希望老师的新书能带领更多的同学加入数据挖掘的领域，一起见证大数据的价值。

李双 一起大数据站长

本书图文并茂，以通俗易懂的语言讲解数据挖掘与机器学习的理论知识，并以图示帮助读者理解，让数据小白能快速理解各种算法背后的原理。本书选用 SPSS Modeler 这个图形化数据挖掘工具，快速实现各种算法及模型，减少大量编写代码的工作，让读者可以更专注数据本身及模型结论。

谢佳标 平安寿险 AI 智能平台团队资深数据挖掘工程师

本书对数据挖掘与机器学习的基本理论、方法和实践案例进行了通俗、趣味性的介绍，融入了作者多年的实战经验。有助于初入或即将进入数据科学行业的朋友，快速将业务、思路、分析技术融会贯通，是一本极好的数据科学工具参考书。

黄小伟 与度科技联合创始人

浩彬老师站在数据小白的角度，以一种幽默风趣和通俗易懂的文风介绍数据挖掘专业知识和如何用 SPSS Modeler 软件完成数据挖掘任务。我相信，每位数据人通过阅读本书，对数据挖掘是什么，为什么用数据挖掘，以及如何做数据挖掘这些问题一定会有新的认识，也会有新的收获。

陆勤 数据人网

本书有趣而又不失专业性，通过配图和故事情节来帮助读者学习和理解，同时又有来龙去脉及 SPSS Modeler 实现的详细讲解，是一本很好的入门书籍。

栗超 百分点集团资深数据挖掘工程师

SPSS 封装了大量成熟的算法，使它成为新人们上手数据挖掘最方便的工具。浩彬老师生动、有趣地讲解了算法原理，经他指点，读者可以深度掌握算法的核心知识。浩彬老师与 SPSS 的完美结合，便有了这本适合新人上手、老人进阶的《小白学数据挖掘与机器学习 SPSS Modeler 案例篇》。新人可以从中快速掌握数据挖掘的基本方法及操作指南。老手们可以深度学习算法原理，夯实基础。想步入人工智能时代的大门，看这一本书就够了。

接地气的陈老师 知乎大 V

前言

浩彬老撕（作者网名），一个有趣的人。

数据挖掘与机器学习，一件好玩的事情。

IBM SPSS Modeler，一套有用的工具。

在日常生活和工作中，笔者经常会遇到有朋友面带难色地咨询：怎么做数据挖掘?怎么学习数据挖掘？笔者发现，大家都认识到，在这个大数据时代，数据挖掘是一项非常有用的技能，但与此同时，他们往往又会觉得学习数据挖掘与机器学习非常难，因为必须要花费大量的时间去重新学习数学知识以及各种编程技能。

对于这些困难，笔者当然理解，而且，随着大数据的兴起，市面上也出现了越来越多关于数据挖掘与机器学习方面的书籍。这些书籍固然都写得很好，但是很多都是一上来就介绍统计理论和模型算法，未免又增加了初学者的畏难情绪。

就笔者看来，从海量数据中挖掘出有用的知识本来是一件很好玩的事情，而且看上去晦涩难懂的算法，其实也有接地气的一面，只要找对学习方法和案例，数据挖掘与机器学习也可以像听故事一样有趣。也是基于这一点，笔者开始了个人公众号以及本书的写作，希望可以用生活中一些常见的例子和一些有趣的插图及通俗的语言故事，把这些看上去晦涩的数据挖掘与机器学习知识以通俗易懂的方式分享给读者，希望让读者从入门学习阶段就发现，原来数据挖掘与机器学习这件事情不但有用，而且还真的有趣。

本书采用 IBM SPSS Modeler（以下简称 SPSS Modeler）作为案例实践工具。SPSS Modeler 是业界公认的数据挖掘利器，它依据 CRISP-DM 方法论，内置了丰富的数据挖掘算法，同时作为一款以“图形化语法”的数据挖掘工具，它的最大优点就是在保证专业性的同时，很好地兼顾了易用性，相信读者使用 SPSS Modeler 作为数据挖掘与机器学习入门工具，将能够很快掌握实际的应用技巧。

本书特色

本书从结构上看，首先介绍了数据挖掘的基本概念以及数据挖掘方法论，接下来介绍了 SPSS Modeler 工具的基本使用、数据探索、统计检验、回归分析、分类算法、聚类算法、关联规则、神经网络以及集成学习。全书内容循序渐进，完整覆盖了数据挖掘与机器学习的主要知识点。

特别地，在每一章中都会以漫画形式介绍一些日常小例子作为切入点，并用通俗的语言为读者介绍具体的算法理论，同时在每章最后都附上应用案例，希望以这样的形式帮助读者更轻松地阅读本书并掌握对应的算法和实践操作。

致谢

感谢图标网站 http://www.easyicon.net/以及 http://pictogram2.com/提供的原始素材，本书的插图大部分来源于对这些原始素材的再创作。

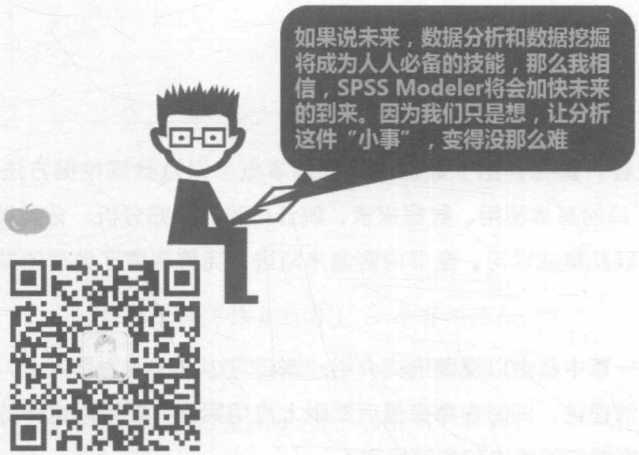
感谢公众号“探数寻理”的读者的关注与支持。感谢 IBM 大中华区分析事业部周伟珠等多位同事的帮助和建议，是你们的建议让本书变得更加完善。感谢柯内特环保大数据研究院院长龙力辉等多位书评作者，感谢你们能够在百忙之中抽出时间阅读书稿，并提出宝贵的建议。感谢电子工业出版社博文视点王静老师的大力支持和辛勤工作，让本书能够顺利出版。最后感谢我的家人和徐小白同学，也因为你们的支持和理解，本书才能顺利出版。

联系方式和电子文件资源

由于笔者水平有限，本书难免会出现一些纰漏和不足之处，恳请各位读者批评、指正。如

果有任何意见和想法，欢迎扫描下方的二维码或在微信中搜索“wetalkdata”，关注“探数寻理”公众号，与笔者进行互动沟通，衷心感谢各位读者的意见和建议。

读者可以通过关注公众号，回复“SPSS”获取软件试用版下载链接以及回复“案例数据”获取本书所有章节对应的数据文件，以及数据模型文件。



作　者

读者服务

轻松注册成为博文视点社区用户（www.broadview.com.cn），扫码直达本书页面。

- 提交勘误：您对书中内容的修改意见可在 提交勘误 处提交，若被采纳，将获赠博文视点社区积分（在您购买电子书时，积分可用来抵扣相应金额）。
- 交流互动：在页面下方 读者评论 处留下您的疑问或观点，与我们和其他读者一同学习交流。

页面入口：http://www.broadview.com.cn/33843

目录