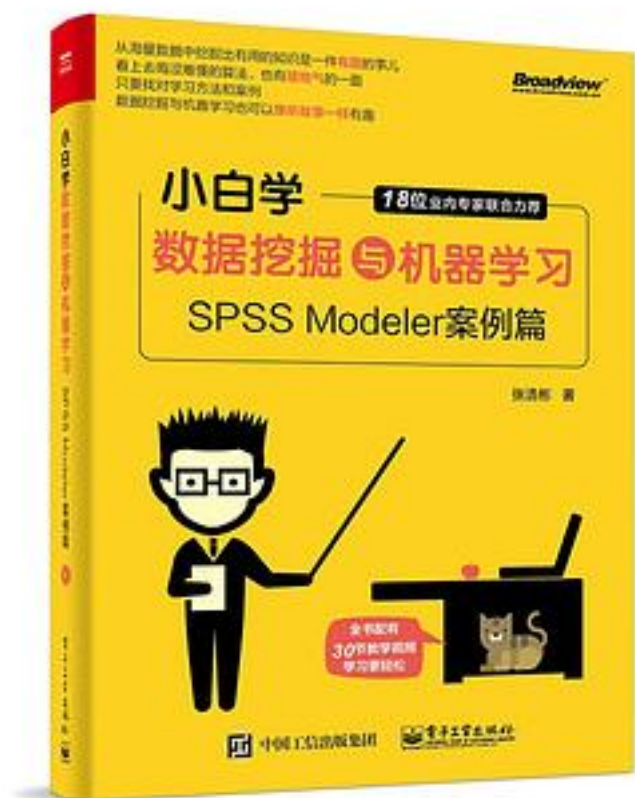


小白学数据挖掘与机器学习



[小白学数据挖掘与机器学习_下载链接1](#)

著者:张浩彬

出版者:电子工业出版社

出版时间:2018-7

装帧:平装

isbn:9787121338434

《小白学数据挖掘与机器学习——SPSS Modeler案例篇》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。

《小白学数据挖掘与机器学习——SPSS Modeler案例篇》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《小白学数据挖掘与机器学习——SPSS Modeler案例篇》并掌握对应的算法和实践操作。

《小白学数据挖掘与机器学习——SPSS Modeler案例篇》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。

作者介绍:

张浩彬，数据分析/数据挖掘专家，目前任职于国际商业机器（中国）有限公司认知计算部门，曾张浩彬，人称浩彬老撕，曾任IBM大中华区商业智能事业部SPSS分析工程师，认知解决方案事业部数据分析专家，现任广东柯内特环境科技有限公司首席数据科学家，致力机器学习及SPSS技术分享，专注于人工智能技术与应用。

微信公众号：探数寻理（wetalkdata）

目录: 第 1 章 数据挖掘那些事儿 \ 1	
1.1 当我们在谈数据挖掘时，其实在讨论什么 \ 2	
1.2 从 CRISP-DM 开启数据挖掘实践 \ 7	
第 2 章 数据挖掘之利器：SPSS Modeler \ 17	
2.1 SPSS Modeler 简介 \ 18	
2.2 SPSS Modeler 的下载与安装 \ 21	
2.3 SPSS Modeler 的主界面及基本操作 \ 23	
2.3.1 SPSS Modeler 主界面介绍 \ 23	
2.3.2 鼠标基本操作 \ 31	
2.4 将 SPSS Modeler 连接到服务器端 \ 31	
第 3 章 巧妇难为无米之炊：数据，数据！ \ 34	
3.1 数据的身份 \ 35	
3.1.1 变量的测量级别 \ 35	
3.1.2 变量的角色 \ 36	
3.2 数据的读取 \ 37	
3.2.1 读取 Excel 文件数据 \ 37	
3.2.2 读取变量文件数据 \ 38	
3.2.3 读取 SPSS Statistics (.sav) 文件数据 \ 40	
3.2.4 读取数据库数据 \ 42	
3.3 数据的基本设定 \ 45	
3.3.1 变量角色的设定 \ 45	
3.3.2 字段的筛选及命名 \ 46	
3.4 数据的集成 \ 47	
3.4.1 数据的变量集成：合并节点 \ 47	
3.4.2 数据的记录集成：追加节点 \ 50	
第 4 章 一点都不简单的描述性统计分析 \ 53	
4.1 分类变量的基本分析：“矩阵”节点 \ 54	
4.2 连续变量的基本分析：数据审核节点 \ 57	
4.2.1 连续变量基本分析指标介绍 \ 57	
4.2.2 “数据审核”节点 \ 63	

- 第 5 章 何为足够大的差异：常用的统计检验 \ 67
 - 5.1 假设检验 \ 68
 - 5.1.1 假设检验的基本原理 \ 68
 - 5.1.2 假设检验的一般步骤 \ 69
 - 5.2 连续变量与分类变量之间的关系：t 检验 \ 70
 - 5.2.1 两组独立样本均值比较 \ 71
 - 5.2.2 两组配对样本均值比较 \ 72
 - 5.2.3 使用 t 检验的前提条件 \ 73
 - 5.2.4 案例：使用均值比较分析电信客户的流失情况 \ 73
 - 5.3 两个连续变量之间的关系：相关分析 \ 75
 - 5.3.1 相关分析理论 \ 76
 - 5.3.2 案例：使用相关分析研究居民消费水平与国内生产总值的相关关系 \ 77
 - 5.4 两个分类变量之间的关系：卡方检验 \ 80
 - 5.4.1 卡方检验的原理 \ 80
 - 5.4.2 卡方检验的前提条件 \ 82
 - 5.4.3 案例：使用卡方检验研究两个分类字段之间的关系 \ 82
- 第 6 章 从身高和体重的关系谈起：回归分析 \ 84
 - 6.1 一元线性回归分析 \ 85
 - 6.1.1 分析因变量与自变量的关系，构建回归模型 \ 85
 - 6.1.2 估计模型系数，求解回归模型 \ 87
 - 6.1.3 对模型系数进行检验，确认模型有效性 \ 88
 - 6.1.4 拟合优度检验，判断模型解释能力 \ 89
 - 6.1.5 借助回归模型进行预测 \ 90
 - 6.2 多元线性回归分析 \ 90
 - 6.2.1 估计模型系数，求解回归模型 \ 91
 - 6.2.2 对模型参数进行检验，确认模型有效性 \ 92
 - 6.2.3 拟合优度检验，判断模型解释能力 \ 94
 - 6.2.4 模型的变量选择 \ 95
 - 6.3 使用线性回归分析的注意事项 \ 97
 - 6.4 案例：使用回归分析研究影响房屋价格的重要因素 \ 98
- 第 7 章 回归岂止这么简单：回归模型的进一步扩展 \ 102
 - 7.1 曲线回归 \ 103
 - 7.2 Logistic 回归 \ 110
 - 7.2.1 Logistic 回归理论 \ 110
 - 7.2.2 案例：使用 Logistic 回归模型分析个人收入水平影响因素 \ 112
- 第 8 章 模型评估那些事儿：过拟合与欠拟合 \ 117
 - 8.1 过拟合与欠拟合 \ 118
 - 8.2 留出法与交叉验证 \ 122
 - 8.2.1 留出法与分层抽样 \ 122
 - 8.2.2 交叉验证 \ 124
- 第 9 章 从看电影的思考到决策树的生成 \ 126
 - 9.1 决策树概述 \ 127
 - 9.2 决策树生成 \ 129
 - 9.2.1 从 ID3 算法到 C5.0 算法 \ 131
 - 9.2.2 CART 算法 \ 134
 - 9.3 决策树的剪枝 \ 136
 - 9.3.1 预剪枝策略 \ 137
 - 9.3.2 后剪枝策略 \ 137
 - 9.3.3 代价敏感学习 \ 138
 - 9.4 案例：用决策树分析客户违约情况 \ 140
 - 9.5 关于信息熵的扩展 \ 147
- 第 10 章 人工神经网络：从人脑神经元开始 \ 151
 - 10.1 从人脑神经元到人工神经网络 \ 152
 - 10.2 感知机 \ 154

10.3 人工神经网络 \ 159
10.3.1 隐藏层的作用 \ 159
10.3.2 人工神经网络算法 \ 160
10.4 案例：利用人工神经网络分析某电信运营商的客户流失情况 \ 164
第 11 章 物以类聚，人以群分：聚类分析 \ 172
11.1 聚类思想的概述 \ 173
11.2 聚类方法的关键：距离 \ 175
11.3 K-Means 算法 \ 176
11.3.1 K-Means 算法原理 \ 176
11.3.2 轮廓系数 (Silhouette coefficient) \ 177
11.4 案例：利用 K-Means 算法对不同型号汽车的属性进行聚类分群研究 \ 179
第 12 章 啤酒+尿布=关联分析？ \ 186
12.1 一个关于关联分析的传说 \ 187
12.2 关联分析的基本概念 \ 188
12.3 关联规则的有效性指标 \ 190
12.4 Apriori 算法 \ 192
12.4.1 生成频繁项集 \ 193
12.4.2 生成关联规则 \ 195
12.5 案例：利用 Apriori 算法对顾客的个人信息及购买记录进行关联分析 \ 195
第 13 章 三个臭皮匠，赛过诸葛亮：集成学习算法 \ 199
13.1 集成学习算法概述 \ 200
13.2 3 种不同的集成学习算法 \ 201
13.2.1 Bagging 算法 \ 201
13.2.2 Boosting 算法 \ 203
13.2.3 随机森林 \ 204
13.3 集成学习算法实践 \ 205
13.3.1 Bagging 算法和 Boosting 算法 \ 205
13.3.2 随机森林 \ 211
13.3.3 集成学习算法结果比较 \ 214
• • • • • ([收起](#))

[小白学数据挖掘与机器学习_下载链接1](#)

标签

通俗易懂

数据挖掘

很棒的书，推荐！

内容丰富

实践者解答

大数据

统计

数据科学

评论

大神出的书，行过路过都不要错过～

小白表示，真的能看懂～

很好的入门书籍，讲解深入浅出，对很多概念豁然开朗

从关注微信公众号开始向作者学习数据挖掘，特别通俗易懂，适合新手。

开始是从“探数寻理”的公众号开始关注的，刚开始觉得那些图片好可爱，读下去发现文章总结的很好。后来听说出书了，自己本身学统计学的，就买了一本看看，发现书本比公众号整理得更系统，例子也有代表性且通俗易懂，对日常工作有指导作用。真的挺不错，值得推荐！！！

很不错的入门数据挖掘书籍，很多新手难懂的问题讲的很透彻，非常适用于入门爱好数据挖掘的同学学习。

浩彬老撕的这本小白都能看懂的数据挖掘书，从基本统计的描述分析到高大上的建模算法，由浅入深，抽丝剥茧，让我这个小白竟然也能看懂啦。能把数据挖掘用这种幽默易懂的语言，讲出来的，非浩彬老撕莫属啦，很棒很棒，感谢老师。希望以后多出这种系

列丛书，期待期待。

从微信公众号到天善智能的课程、再到这本小白学数据挖掘与机器学习图书，一直都是浩彬老撕的粉丝，给本书点赞，很适合对数据挖掘感兴趣的小伙伴们学习嗷～～

通俗易懂，深入浅出！还有37个视频讲解！！！实用的工具书！数据分析、数据挖掘必备～

非常适合作为数据挖掘数据分析的入门书籍，除了带你对spss modeler工具的使用到了解并应用现阶段较火的几大挖掘算法，如果想要深究每个算法原理还有公式推导和讲解可以辅助学习，是从入门到基础巩固的首选，感觉可以重复几遍，每遍都有新知识新体会

最大特色应该就是通俗易懂了，第一次见数据挖掘的书能以这种方式呈现，而且也有算法推导，解答了之前不少疑惑。买的时候没留意，收到后发现带了30节免费的视频课程，算是意外之喜，点赞。

很不错的入门数据挖掘书籍，很多新手难懂的问题讲的很透彻，非常适用于入门爱好数据挖掘的同学学习

[小白学数据挖掘与机器学习_下载链接1](#)

书评

作者在IBM工作过，专业水平很高，在业内非常出名，一直在阅读作者的公众号，非常喜欢，终于出书了，书非常赞，甚至超出期待！！！！内容深入浅出，是基于SPSS Modeler的案例分析，书中通过大量图片解释，另菜鸟读也毫无压力（配套的视频一定要看）。统计学是当今并且在将来是非...

本书对基于SPSS

Modeler对数据分析的过程以及主流的数据挖掘算法进行了深入浅出的介绍。数据分析是不是只能很晦涩难懂？是不是只有数不尽的公式？不是！

我是个数据小白，怎么办？一步步上手呀！

从一个新人的角度对数据分析中遇到的要点、难点进行了清晰的讲述，而图文并茂...

[小白学数据挖掘与机器学习_下载链接1](#)