Top Quant
CHRD前海智库

零起点
Python足彩大数据与机器学习实盘分析

何海群 著

Win Or Home

率先导入机器学习模式
用专业的量化回溯手段，精准分析足彩赔率数据

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

零起点 Python足彩大数据与机器学习实盘分析

“零起点Python”系列丛书

本书继续保持了“零起点Python”系列丛书的一贯风格——简单实用。书中配备了大量的图表说明，没有枯燥的数学公式，只要懂Word、Excel，就能够轻松阅读全书。

▼IT零起点：不需要任何电脑编程基础，只要会打字、会使用Excel，就能看懂本书，利用本书配套的Python软件包，轻松学会如何利用Python对股票和足彩数据进行专业分析和量化投资分析。

▼投资零起点：无须购买任何专业软件，本书配套的zwPython软件包，采用开源模式，提供全功能、全免费的工业级数据分析平台。

▼配置零起点：所有软件、数据全部采用苹果“开箱即用”模式，绿色版本，无须安装，解压缩后即可直接运行系统。

▼理财零起点：不需要任何专业金融背景，采用通俗易懂的语言，配合大量专业的图表和实盘操作案例，轻松掌握各种量化投资策略。

▼数学零起点：全书没有任何复杂的数学公式，只有基本的加、减、乘、除，轻轻松松就能看懂全书。

本书学习资源

▼本书的读者QQ群是：264880547（Top极宽足彩大数据）。

▼有关的程序和数据下载，请浏览网站http://TopQuant.vip，在【下载中心】栏目下载。

▼本书在极宽量化社区设有专栏，读者有任何建议都可以在相关专栏发布信息，我们会第一时间进行反馈和答复。

上架建议：大数据&机器学习

ISBN 978-7-121-31074-4

9 787121 310744 >

定价：99.00元

策划编辑：黄爱萍（QQ：69476637）
责任编辑：徐津平
封面设计：李　玲

零起点
Python足彩大数据与机器学习实盘分析

何海群 著

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

内 容 简 介

本书使用 Python 编程语言、Pandas 数据分析模块、机器学习和人工智能算法对足彩大数据进行实盘分析，设计并发布了开源大数据项目 tfbDat 足彩数据包，汇总了 2010—2017 年全球近 7 万场足球比赛的赛事和赔率数据。此外，还介绍使用 Python 语言抓取网页数据、下载更新 tfbDat 足彩数据包、预测和分析比赛球队的取胜概率，同时提出了检测人工智能算法优劣的“足彩图灵”法则。

图书在版编目（CIP）数据

零起点 Python 足彩大数据与机器学习实盘分析 / 何海群著. —北京：电子工业出版社，2017.5
ISBN 978-7-121-31074-4

Ⅰ. ①零… Ⅱ. ①何… Ⅲ. ①软件工具－程序设计－应用－足球运动－彩票 Ⅳ. ①F719.52-39

中国版本图书馆 CIP 数据核字（2017）第 050287 号

策划编辑：黄爱萍
责任编辑：徐津平
印　　刷：三河市华成印务有限公司
装　　订：三河市华成印务有限公司
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：27.5　字数：523 千字
版　　次：2017 年 5 月第 1 版
印　　次：2017 年 5 月第 1 次印刷
定　　价：99.00 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。

前　言

从足彩到量化，再从量化到足彩。

生命总是在轮回中，不断完成自我突破与成长壮大。

本书是“Python 量化三部曲”的补充部分。

“Python 量化三部曲”

“Python 量化三部曲”包括：

- 《零起点 Python 大数据与量化交易》（入门课程）；
- 《零起点 Python 量化与机器学习实盘分析》（重点分析 Sklearn）；
- 《零起点 Python 量化与 TensorFlow 深度学习实盘分析》（重点分析 TensorFlow）。

此外还有两部补充作品：

- 《零起点 Python 足彩大数据与机器学习实盘分析》；
- 《零起点 Python 机器学习快速入门》。

较好的 Python 机器学习入门教程

本书中的机器学习算法章节，是目前较系统的 Python 机器学习入门教程，其特点如下。

- 独创的黑箱教学模式，全书无任何抽象理论和深奥的数学公式。
- 首次系统化融合 Sklearn 人工智能软件和 Pandas 数据分析软件，无需直接使用复杂的 Numpy 数学矩阵模块。
- 三位一体的课件模式：图书+开发平台+成套的教学案例。系统讲解，逐步深入。
- 业内第一个系统化的 Sklearn 函数和 API 中文文档，可作为案头工具书随时查阅。
- 基于 Sklearn+Pandas 模式，无需任何理论基础，全程采用 MBA 案例模式，懂 Excel 就可看懂。

这些内容采用独创的黑箱模式和 MBA 案例教学机制，结合一线实战案例，介绍 Sklearn 人工智能模块库和常用的机器学习算法。

进一步学习

本书的读者如果有兴趣可以进一步学习“Python 量化三部曲”的内容，虽然“Python 量化三部曲”的内容是以金融量化分析为主，但基本原理都是相通的，本质上都是数据分析，只是数据源不同，一个是金融数据，另一个是足彩赔率数据。

对于“Python 量化三部曲”的读者而言，本书也有很大的价值，特别是对于入门的读者。

本书有多章关于网络爬虫的内容和具体案例，讲解了提取网络数据的方法，为希望自己编写自动交易程序的读者提供了一种基于 Web 的操作接口。

网络资源

本书的读者 QQ 群是：264880547（Top 极宽足彩大数据）。

本书有关的程序和数据下载，请浏览网站：TopQuant.vip 极宽量化社区。网站【下载中心】有最新的程序和数据下载地址。

本书在 TopQuant.vip 极宽量化社区设有专栏，对本书和足彩有任何建议的读者，请在社区相关专栏发布信息，笔者会在第一时间进行反馈和答复。

“零起点 Python”系列丛书

本书继续保持了“零起点 Python”系列丛书的一贯风格，简单实用，书中配有大量图表说明，没有一条数学公式，普通读者只要懂 Word 和 Excel，就能够轻松阅读全书。

- IT 零起点，不需要任何电脑编程基础，会打字、会 Excel 就能看懂本书，利用本书配套的 Python 软件包，轻松学会利用 Python 对股票、足彩数据进行专业分析、量化投资分析。
- 投资零起点，无须购买任何专业软件，本书配套的 zwPython 软件包采用开源模式，提供 100%全功能、全免费的工业级数据分析平台。
- 配置零起点，所有软件、数据全部采用苹果“开箱即用”模式，绿色版本，无须安装，解压即可，直接运行系统。
- 理财零起点，不需要任何专业金融背景，使用通俗易懂的语言，配合大量专业图表和实盘操作案例，轻松掌握各种量化投资策略。
- 数学零起点，全书没有任何复杂的数学公式，只有最基本的加、减、乘、除，易于理解。

关于足彩的几个误区

近年来，大数据产业、人工智能行业风生水起，可国内足彩大数据的专业研究还处于冷门和偏门，这其中有很多关于足彩领域的误区。

很多主流学者，到现在都看不起足彩，认为是赌博，这个我们不讨论。最近国家级别的彩票大数据研究中心已经正式成立了，资本的力量是无穷的。

为此，笔者做了一个简单的小结，以正视听。

- 足彩虽然容易与赌球混淆，但却是最好的大数据研究对象，没有之一。
- 微软、百度和谷歌等公司目前都有专业团队在做足彩大数据研究，并定期发布结果。
- 足彩相当于十倍配资的股票。
- 国内足彩赔率的确很低，差不多是全球最低的，比欧洲平均低 10%左右。

- 彩票和股票的发明人据说都是同一个英国爵士。
- 必胜足彩交易所成立当年获得了英国 MBA 商业创新大奖。
- 高盛公司多年前就开始进行足彩套利业务，维基百科中有介绍。
- 极宽黑天鹅（红牛吧）足彩是业内首家公开进行实盘测试的足彩大数据模型。
- 黑天鹅算法在业内率先以“盈利率”而不是“胜率”测试足彩算法模型。

致谢

虽然很多网友在笔者博客中留言，建议笔者早日完成这本书的写作，但本书的创作和正式出版还是经历了许多波折。

如今本书终于出版，在此，要特别感谢电子工业出版社的黄爱萍和戴新编辑，感谢她们在选题策划和稿件整理方面所做的大量工作。

同时，在本书创作过程中，极宽开源量化团队和培训班的全体成员提出很多宝贵的意见，并对部分课件程序做了中文注解。

特别是吴娜、余勤、王硕三位同学，为极宽开源量化文库和 zwQuant 开源量化软件编写文档，并在团队成员管理方面做了大量工作，对他们的付出表示感谢。

何海群（字王）

北京极宽科技有限公司 CTO

2017 年 3 月 25 日

目　录