

Artificial Intelligence  
and Machine Learning  
for Business

Third Edition


# 智能商业

人工智能、机器学习与大数据技术的制胜之道

(原书第3版)

[英]史蒂文·芬利 (Steven Finlay) 著

阿茹罕 姚旭 程国建 译

 机械工业出版社  
CHINA MACHINE PRESS

# Artificial Intelligence and Machine Learning for Business


Third Edition

# 智能 商业

人工智能、机器学习与大数据技术的制胜之道  
(原书第3版)

[英]史蒂文·芬利 (Steven Finlay) 著

阿茹罕 姚旭 程国建 译

 机械工业出版社  
CHINA MACHINE PRESS

人工智能 (AI) 和机器学习是当代主流的商业分析工具, 可应用于许多行业, 以增加利润、降低成本、挽救险局并改善客户体验。因此, 企业应该要了解如何使用这些工具, 并知道如何利用这些工具在与对手的竞争中获利。

本书引入了与这些主题 (AI、机器学习、大数据技术) 相关的概念, 并用非技术的语言降低了对技术术语的理解难度, 可为经理和商务人士提供简洁明快的入门知识。本书的重点是实际应用以及如何与技术专家 (数据科学家) 进行合作, 以最大限度地发挥这些技术的优势。

此第3版已经过大幅修订和更新, 增加了几个新的章节, 涵盖了比以往版本更广泛的主题, 但保留了原始版本的严肃风格。

Copyright © 2018 by Steven Finlay.

All rights reserved. No part of this book may be reproduced by any means whatsoever without written permission from the publisher. The Chinese edition Copyright © 2019 by China Machine Press

本书中文简体字版由机械工业出版社出版, 未经出版者书面允许, 本书的任何部分不得以任何方式复制或抄袭。

版权所有, 翻印必究。

北京市版权局著作权合同登记 图字: 01-2018-8097号。

## 图书在版编目 (CIP) 数据

智能商业: 人工智能、机器学习与大数据技术的制胜之道: 原书第3版 / (英) 史蒂文·芬利 (Steven Finlay) 著; 阿茹罕, 姚旭, 程国建译. —北京: 机械工业出版社, 2019.6

书名原文: Artificial Intelligence and Machine Learning for Business: A No-Nonsense Guide to Data Driven Technologies  
ISBN 978-7-111-62602-2

I . ①智… II . ①史… ②阿… ③姚… ④程… III . ①人工智能 - 应用 - 商业模式 - 研究 IV . ①F716

中国版本图书馆 CIP 数据核字 (2019) 第 080506 号

机械工业出版社 (北京市百万庄大街 22 号 邮政编码 100037)

责任编辑: 刘星宁 责任校对: 陈越

封面设计: 马精明 责任印制: 孙炜

保定市中国画美凯印刷有限公司印刷

2019年6月第1版第1次印刷

148mm × 210mm · 6.5 印张 · 132 千字

标准书号: ISBN 978-7-111-62602-2

定价: 49.00 元

凡购本书, 如有缺页、倒页、脱页, 由本社发行部调换

电话服务

网络服务

服务咨询热线: 010-88361066

机工官网: [www.cmpbook.com](http://www.cmpbook.com)

读者购书热线: 010-68326294

机工官博: [weibo.com/cmp1952](http://weibo.com/cmp1952)

金书网: [www.golden-book.com](http://www.golden-book.com)

封面无防伪标均为盗版

教育服务网: [www.cmpedu.com](http://www.cmpedu.com)

## 译者序

随着人工智能和机器学习（Artificial Intelligence & Machine Learning, AIML）技术不断地促进业务转型，更多的企业将越来越依赖于数据驱动技术（Data Driven Technologies, DDT）进行自动化数据分析。通过事前预测事件发生的可能性，可使企业能够主动采取行动来优化成果，而不是在事后对事件做出反应。通过AIML+DDT优化决策而付诸实施的行为结果反馈到预测模型和算法中，最终形成一个不断改进的自学系统。就业务分析和商业智能而言，人工智能（AI）的研究已经达到了某种程度的成熟，大规模的机器学习应用和小规模的深度学习应用已经开始进入市场。

企业使用AI、机器学习、数据挖掘、预测分析、深度学习和认知计算等技术，发现隐藏在大量杂乱数据中的特征模式和重要信息。与传统智能商业及相关方法不同，这些技术并不局限于结构化数据和小型数据集，相反主要面向大数据领域，其中主要处理对象是半结构化和非结构化数据，例如文本数据、传感器数据和社交媒体数据等。

如果将大数据视为一种新的“石油”，那么AIML+DDT就

是“炼油技术”。通过分析数据而获得价值，将原始数据加工为信息和知识，进而转化为决策或行动，已成为推动企业数字化转型不可或缺的关键技术。从金融服务、保险、医疗保健再到能源等，各行各业都在通过 AI、机器学习与大数据分析实施业务转型。

智能商业相关的核心技术大致可以分为数据平台技术、数据整理技术、数据分析技术、数据交互技术、数据可视化技术等。与传统意义上的数据处理、数据分析相比，数据智能面临很多新挑战、新问题。

本书主要内容包括：什么是机器学习和 AI；企业使用 AI 的目的何在；预测模型究竟是什么样的；AI、机器学习和大数据之间的互动关系如何；应用 AI 所需的人员素质和工具类别；如何使用 AI 来改善业务流程和盈亏平衡点；在开发基于 AI 的解决方案时需要考虑的法律和道德问题；如何采用先进的机器学习方式来驱动 AI 应用程序的研发；目前机器学习和 AI 的局限性等。

本书篇幅短小精悍、内容翔实、面向实用，可作为企业管理人员、大专院校师生有关 AI、机器学习以及数据科学的入门读物。本书的翻译出版得到了机械工业出版社刘星宁编辑的推荐与支持，在此深表谢意。

译者

单程 56 分钟，往返 112 分钟，这是通勤列车从我的家乡普雷斯顿前往英国大城市曼彻斯特的时间。我发现这是一个理想的阅读时间。因此，我认为一本关于人工智能（AI）和机器学习的简洁读物，对于没有太多空闲时间的人来说利用这个时间段阅读非常合适。

对机器学习的理解很重要，因为它对我们生活的许多方面产生了巨大的影响。特别是，它推动“AI”在语言翻译、自主机器人和医疗诊断等许多应用领域呈现爆炸式增长。

AI 和机器学习也对许多日常业务产生直接影响。基于机器学习的自动化系统正在取代曾经由人类承担的众多任务。这使得拥有这些技术的组织比其竞争对手更具竞争优势，因为这样的系统可以提高效率并改善客户服务。

本书的第 3 版比上一版内容多很多，特别是有几个新的章节涵盖了比以前更广泛的主题。但是，我努力保留原作的“简洁”风格，因为这是读者喜欢它的一个关键原因。因此，本书现在可能需要花费超过 112 分钟才能读完（需要两次往返曼彻斯特的时间而不是一次），但是我希望您能找到合适的阅读时间。

## 作者简介

史蒂文·芬利（Steven Finlay）是一位数据科学家，在大规模数据环境中开发实用的“增值”机器学习解决方案已有 20 多年的经验。他拥有预测建模博士学位，并且是英国兰开斯特大学的荣誉研究员。

史蒂文之前曾受雇于英国十大银行之一，管理其信贷风险模型，为英国政府开发机器学习方法，并为多家咨询集团工作。他目前是英国计算机股份贷款服务公司 Computershare Loan Services (CLS) 的商务分析主管。

芬利博士出版了许多关于机器学习、人工智能（AI）和金融服务的实用性书籍。他的其他书籍（由 Palgrave Macmillan 出版）包括：

- 《预测分析、数据挖掘和大数据：神话、误解和方法》；
- 《信用评分、响应建模和保险评级：预测消费者行为的实用指南》；
- 《消费者信贷管理：理论与实践》；
- 《消费者信贷基础》。

# 目 录

译者序

前言

作者简介

1. 简介 / 001
2. 什么是机器学习和人工智能 (AI)? / 007
3. 预测模型生成的分数代表什么? / 021
4. 为什么使用机器学习? 它增加了什么价值? / 029
5. 机器学习是如何工作的? / 035
6. 使用预测模型做出决策 / 047
7. 这是记分卡, 那决策树是什么? / 053
8. 神经网络和深度学习 / 059
9. 无监督学习和强化学习 / 071



10.	如何建立预测模型？	/ 085
11.	实施机器学习	/ 101
12.	大数据与机器学习的关系	/ 113
13.	道德、法律和 GDPR	/ 119
14.	机器学习的最新前沿	/ 133
15.	我们何时可以购买无人驾驶汽车？	/ 143
16.	结束语	/ 153
	附录	/ 155
	附录 A	评估预测模型 / 155
	附录 B	更多信息及推荐阅读文献 / 163
	附录 C	机器学习和人工智能中的流行术语 / 169
	附录 D	业务成功清单 / 183
	注释	/ 187

# 1. 简介

**您**有智能手机或信用卡吗？您是从超市买东西吗？玩电脑游戏吗？您是否受雇或使用医疗保健？如果这些问题的答案是肯定的，那么人工智能（AI）和机器学习将以某种方式对您的生活产生影响。这是因为它们是企业的主要工具，利用它们拥有的关于您的数据，来决定它们将如何对待您。它们告知企业您在不同情况下的行为方式，以及它们应该如何对待您以便最大化其（有时是您的）目标。这些技术现在几乎用于各行各业，通过“AI”机器和计算机界面改进流程并增强人们的日常体验。亚马逊的 Echo、苹果的 Siri 和谷歌翻译只是三个众所周知的软件产品，体现了这些技术带来的好处。

如今，许多产品和服务也是自适应的。它们根据个人用户的行为调整响应。电视和音乐流媒体服务学会识别您喜欢的内容，并向您推荐您必然会感兴趣的内容。改变您听音乐的类型，它们的推荐也会改变。

同样，您也可以买一些供暖系统，这些系统可以学会在最合适的时候打开暖气，这样您就不必费心了，同时优化能源使用，减少开支。

这些都是 AI 发挥作用的进一步例子。

这本简洁的书提供了 AI 和机器学习的管理（即非技术和非复杂公式的）概述，它们是什么以及如何使用它们。不需要任何预先的知识，换句话说，如果您能读、能写、能做基本的算术（有一点算术，但没那么多），那么您掌握这本书的内容没有问题。

现在有一个问题要问：我为什么需要知道这些事情？一个原因是出于个人原因。基于机器学习的智能决策系统被企业广泛用于决定如何对待您、您的朋友和您的家人。他们用这些工具来决定您会得到一份好工作还是一份差工作，您应该被排在队伍的前面还是后面，您是否会受到税务审计，或者是否会被当作刑事案件的嫌疑人。因此，了解这些事情并不是一个坏事，这样您就能理解为什么一个组织可能用这种方式对待您，而不是以另一种方式。

学习 AI 的另一个原因，也是本书主要关注的一个原因，它现在是一种主流的商业工具。不久之前，AI 还是少数技术专家的领域，他们主要在学术界、金融服务业或大型营销部门工作。如今，无论您从事什么行业，基于机器学习的 AI 应用都可以在各种各样的商业活动中找到。这涵盖了从员工审查、回答客户询问、目标营销到生产线上的机器人、仓库管理和客户配送等方方面面。因此，AI 在许多领域支持或取代了人类的专业技能。例如：替代了

保险人员设定保费，帮助人力资源部门决定雇佣/解雇人员，当客户走进您的商店时自动识别他们的身份，以及在诊断癌症和心脏病等疾病时给予医生支持。

AI时代已经到来。这不是一时的狂热，它会一直持续下去。那些用它来解决业务问题、提高效率和降低成本的组织将从竞争对手的损失中获益。

这并不意味着您要学习技术专家（数据科学家）所知道的东西。然而，了解什么是AI和机器学习，以及如何使用它们来帮助您的企业提供更好的产品和服务，这将对我们来说是非常有益的。最重要的是，为了有效地利用这些工具，他们需要把注意力集中在商业目标上，以便解决企业面临的具体问题。

另一方面，如果您恰好是一个经常杂耍般地引用各种眼花缭乱的公式，认为关于AI啥都懂的数学奇才，那么本书也会对您有一些价值。不过，其实他们如果真对AI什么都不懂，本书价值或许更大。为什么呢？因为如果您所关注的只是理论方面，一旦在满是雷区的真实世界需要您提供实用的解决方案，您就会有遭遇一筹莫展困境的风险。这是一个充满社会、伦理和政治问题的真实世界。再加上越来越多的隐私和数据保护立法，可能会破坏您的解决方案，无论它们在数学上多么美好。如果不考虑这些“软问题”，好一点的情况是您开发的解决方案无法部署。最糟糕的情况是，您设计了一个基于AI的系统，因为它歧视少数族裔、女性或其他群体，而让您上了法庭。也许您可以略过前面几章，但您

一定要读后面几章。

为了最大限度地利用 AI，数据科学家需要与商业用户接触，以了解他们的需求。数据科学家还需要了解企业的文化及其采用新思想、技术和工作实践的方法。客户所在地区的法律和监管问题也需要适当考虑。无论一个解决方案在硬件和软件方面多么优秀，如果它与企业的业务目标和操作流程不一致，那么它就是在浪费时间和金钱。

许多解决方案供应商可以用他们的花哨技术和最新术语迷惑您，这些术语通常只是对去年的技术进行了重新包装，添加了新的花样。然而，创造价值的供应商会花时间了解您和您的企业的工作方式。然后，他们将决定是否使用以及如何使用他们的解决方案来改进您的工作，并使用简单的语言向您解释这一点，您可以理解，而无需求助于维基百科（Wikipedia）。

成功的 AI 是双向的。数据科学家需要了解您的企业以及您的企业是做什么的，您也需要了解一些有关 AI 和机器学习的知识。如果没有这种共识，您或您的企业就不可能获得 AI 所带来的全部益处。

好的。那么，您会从这本书中学到什么呢？我们将在下面的章节中讨论的主题是：

- 机器学习和 AI 是什么？
- 企业使用 AI 的目的。
- 预测模型是什么样的。

- AI、机器学习和“大数据”之间的关系。
- 应用 AI 所需的人员和工具。
- 如何使用 AI 来改善业务流程和盈亏平衡点。
- 在开发基于 AI 的解决方案时需要考虑的法律和道德问题，这些解决方案将用于制定有关人员的决策。
- 如何应用先进的机器学习形式来驱动 AI 应用程序，如目标识别和语言翻译。
- 目前机器学习和 AI 的局限性。

推荐阅读的文献和常见机器学习 / AI 术语表分别在附录 B 和 C 中给出。

机器学习是利用数学程序（算法）来分析数据，目的是发现不同数据项之间的有用模式（关系或相关性）。一旦确定了这些关系，就可以使用它们来预测新案例出现时的行为。本质上，这类似于人们的学习方式。我们观察周围发生的事情，从我们的经验中得出关于世界如何运转的结论。然后，我们将学到的知识来帮助我们在处理出现的新情况。我们经历和学习得越多，我们做决定的能力就越强。

机器学习的一个应用是目标识别。我们的目标是开发一种系统，能够从展示给系统的图像中识别日常物体。用于开发物体识别系统的数据包括椅子、椅子、洗衣机等不同物体的图片。向机器学习算法展示的每张图片都被贴上标签，以识别其中包含的对象类型。对于每种类型的对象，可能有成百上千种不同的图像，从不同的角度表示该对象的不同形式。（您会惊讶于一把椅子有这么多的形式！）



## 2. 什么是机器学习和人工智能 (AI) ?

**机**器学习是利用数学程序（算法）来分析数据，目的是发现不同数据项之间的有用模式（关系或相关性）。一旦确定了这些关系，就可以使用它们来推断新案例出现时的行为。本质上，这类似于人们的学习方式。我们观察周围发生的事情，从我们的经验中得出关于世界如何运转的结论。然后，我们应用学到的知识来帮助我们处理出现的新情况。我们经历和学习得越多，我们做决定的能力就越强。

机器学习的一个应用是目标识别。我们的目标是开发一种系统，能够从展示给系统的图像中识别日常物体。用于开发物体识别系统的数据包括椅子、雨伞、洗衣机等不同物体的图片。向机器学习算法展示的每张图片都被贴上标签，以识别其中包含的对象类型。对于每种类型的对象，可能有成百上千种不同的图像，从不同的角度表示该对象的不同形式。（您会惊讶于一把伞有这么多的形式！）



通过分析不同的图像，机器学习算法识别出某些对象与某些特征（模式）相关联。椅子往往有突起（腿）来自一个扁平的，通常是方形的底座（座位）。它们还可以通过靠背与底座来识别。洗衣机往往是带有旋钮的立方体，几乎从来不是粉色或橙色的。（如果您在什么地方能买到粉色的洗衣机，一定要告诉我！）同样的，雨伞又长又细（合上的时候），通常是黑色的，但也不总是这样。

机器学习最常见的，也可以说是第一个应用是预测。预测是根据您目前拥有的信息，使用机器学习来确定一些您目前不知道的东西，人们发现有关行为和结果之间关系的模式，这通常与人们过去的行为以及他们接下来的行为有关。确定了存在的关系之后，就可以根据一个人目前的状态来预测他未来的行为。如果您给我一个人们之前购买历史的样本，我可以利用机器学习来识别他们购买行为的模式。然后我可以使用这些模式来预测人们接下来可能会购买的商品；也就是说，未来的购买是我想预测的结果。这使我能够针对这些特定的产品，为它们量身定制促销优惠。

使用机器学习预测有时被称为预测模型或预测分析（PA）。事实上，预测分析是机器学习的一个常见应用，许多人（正确或错误地）经常使用这两种术语。

通常人们应用机器学习来预测个体未来的行为，但是机器学习还可以应用于其他情况和问题。您所需要的只是一些您想要确定（预测）的未知事件或事物，这些可能是过去的、现在的或未来的。为了收集证据（数据），医生检查病人，并询问他们的症