大数据技术与应用丛书

# 数据挖掘与机器学习

## 基础及应用

许桂秋 吴丽镐 张文明 主 编
龙法宁 王子琦 王珂 朱琳玲 徐强 副主编

中国工信出版集团
人民邮电出版社
POSTS & TELECOM PRESS


《大数据导论（通识版）》

《数据采集及预处理基础与应用》

《分布式数据库基础与应用》

**《数据挖掘与机器学习基础及应用》**

《数据可视化基础与应用》

《大数据处理技术基础与应用（Hadoop+Spark)》

《Python编程基础与应用——任务式案例教程》

分类建议：计算机 / 大数据
人民邮电出版社网址：www.ptpress.com.cn

ISBN 978-7-115-64576-0

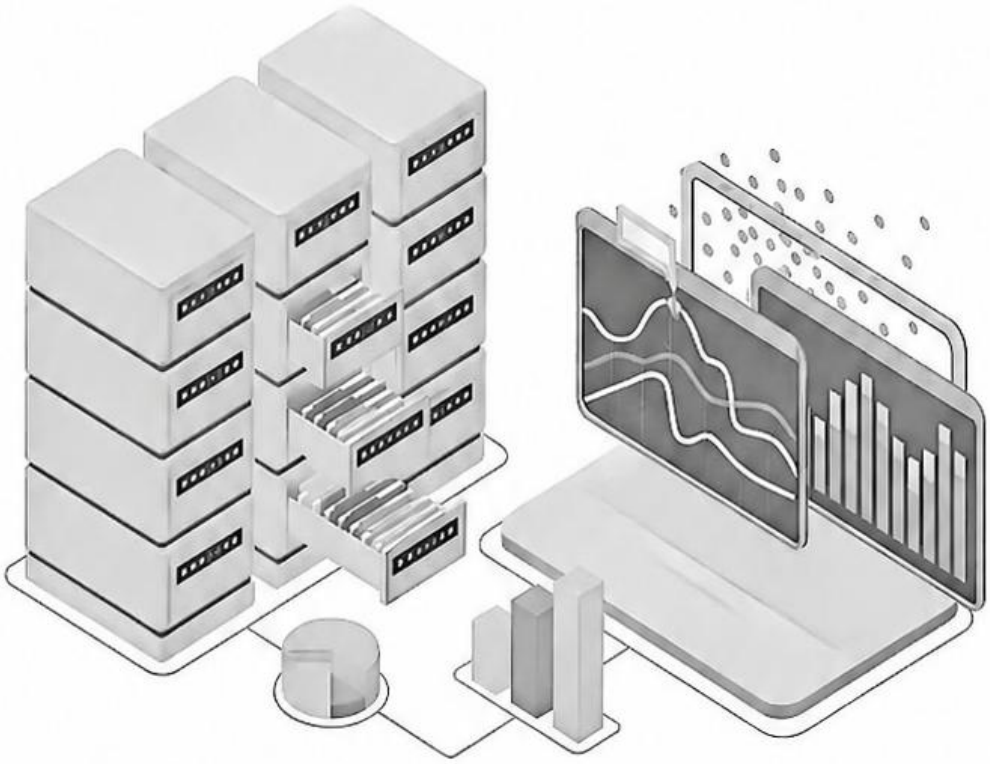
定价：69.80 元

大数据技术与应用丛书

# 数据挖掘与机器学习

## 基础及应用

许桂秋 吴丽楠 张文明◎主 编
龙法宇 王子茹 王珂 朱琳玲 徐强◎副主编



人民邮电出版社
北京

图书在版编目（CIP）数据

数据挖掘与机器学习基础及应用 / 许桂秋，吴丽镐，张文明主编. — 北京 ：人民邮电出版社，2024. 8.
（大数据技术与应用丛书）. — ISBN 978-7-115-64576-0

Ⅰ. TP274；TP181

中国国家版本馆 CIP 数据核字第 2024CN3518 号

## 内 容 提 要

这是一本全面介绍数据挖掘与机器学习的大数据专业类图书，阅读本书可以提升读者对大数据分析与挖掘的认知及动手能力。本书共 10 章，由浅入深地讲解数据挖掘与机器学习的基本概念与流程、相关算法与实现工具。全书理论与实践相结合，既有新技术的深度，也有行业应用的广度，使读者可以全面了解数据挖掘与机器学习相关技术。

本书可以作为高等学校计算机、数据科学与大数据技术等相关专业“机器学习”或者“数据挖掘”课程的教材，也可作为从事机器学习与数据挖掘、数据分析相关工作的技术人员的参考书。

♦ 主　　编　许桂秋　吴丽镐　张文明
副 主 编　龙法宁　王子琦　王　珂　朱琳玲　徐　强
责任编辑　张晓芬
责任印制　马振武

♦ 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　https://www.ptpress.com.cn
三河市祥达印刷包装有限公司印刷

♦ 开本：787×1092　1/16
印张：16.5　　2024 年 8 月第 1 版
字数：381 千字　　2024 年 8 月河北第 1 次印刷

定价：69.80 元

**读者服务热线：(010)81055493　印装质量热线：(010)81055316**
**反盗版热线：(010)81055315**

# 前言

数据挖掘与机器学习是计算机科学和人工智能学科中一个非常重要的研究领域，也是一些交叉学科的重要支撑技术。在过去的数十年中，互联网以及各种信息系统产生了大量数据。数据的爆炸性增长激起了人们对新技术和自动化工具的需求，以便于帮助我们将海量数据转换成信息和知识。数据挖掘与机器学习作为一种前沿的分析工具，引起了产业界和学术界的广泛关注，并快速成为计算机研究领域的一个热点。

本书是数据挖掘与机器学习的入门级图书，主要描述数据挖掘与机器学习的相关定义与算法，包含数据挖掘与机器学习中各种模式的概念和实现方式、算法的具体使用技巧等。本书可帮助读者了解数据挖掘与机器学习的内容，并让读者能根据书中提供的案例完成数据挖掘的各种模式操作。

全书共 10 章，可分为 4 个部分。

第一部分是基础概述，包括第 1 章和第 2 章。第 1 章阐述数据挖掘与机器学习的发展历史、基本概念、算法分类、一般流程，以及主要的应用领域；第 2 章详细介绍实现数据挖掘与机器学习的基本工具，例如 Numpy、pandas、Matplotlib 和 scikit-learn 数据科学分析库，以及它们的使用方法。

第二部分是数据挖掘与机器学习的算法使用，包括第 3～7 章。这 5 章详细介绍数据挖掘与机器学习中回归、分类、聚类等算法与应用，以及关联规则与协同过滤，最后介绍特征工程、降维与超参数调优的内容。

第三部分是进阶部分，包括第 8 章和第 9 章。这两章通过图像分类之猫狗识别和基于 NLTK 实现文本数据处理这两个案例介绍文本与图像的相关分析方法。

第四部分主要介绍深度学习的相关内容，只包括第 10 章。这章以 Fashion MNIST 数据集为例，介绍基于深度学习的图像处理方法以及模型的搭建、优化、保存等处理过程。

由于编者水平有限，书中难免存在不足之处，恳请广大读者批评指正。

编者

2024 年 6 月

# 目录