HZ BOOKS 华章教育
ELSEVIER 爱思唯尔

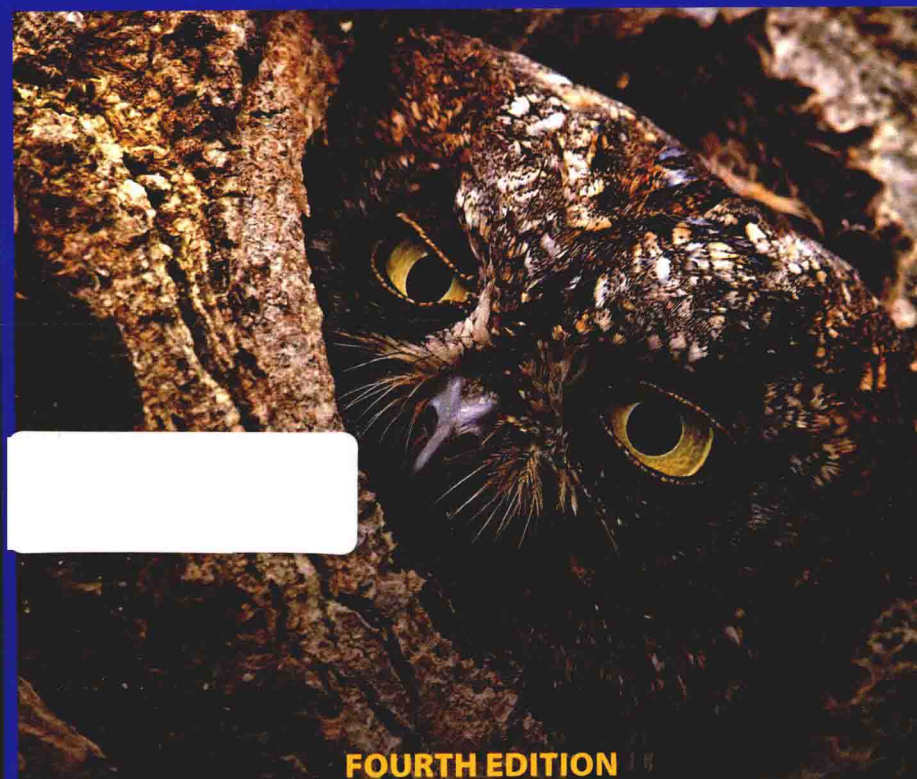
经 典 原 版 书 库

数据挖掘

实用机器学习工具与技术

[新西兰] 伊恩 H. 威腾 Ian H. Witten
埃贝·弗兰克 Eibe Frank
马克 A. 霍尔 Mark A. Hall
[加] 克里斯多夫 J. 帕尔 Christopher J. Pal 著

（英文版·第4版）



FOURTH EDITION

DATA MINING

Practical Machine Learning Tools and Techniques

Ian H. Witten • Eibe Frank • Mark A. Hall • Christopher J. Pal

MORGAN KAUFMANN
机械工业出版社 China Machine Press

经 典 原 版 书 库

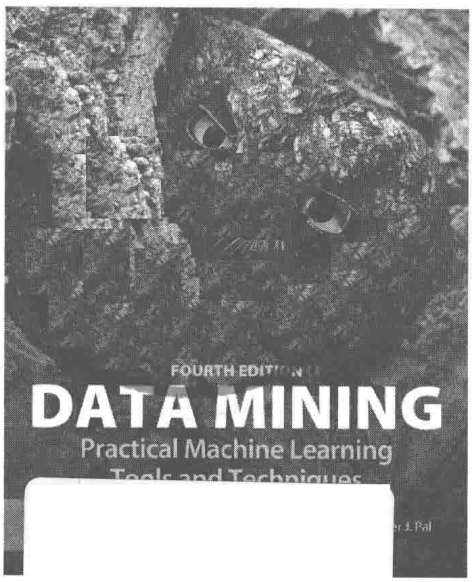
数据挖掘

实用机器学习工具与技术

（英文版·第4版）

Data Mining

Practical Machine Learning Tools and Techniques, Fourth Edition



[新西兰] 伊恩 H. 威腾 Ian H. Witten　埃贝·弗兰克 Eibe Frank　马克 A. 霍尔 Mark A. Hall　[加] 克里斯多夫 J. 帕尔 Christopher J. Pal　著

机械工业出版社
China Machine Press

图书在版编目（CIP）数据

数据挖掘：实用机器学习工具与技术（英文版·第4版）/（新西兰）伊恩 H. 威腾（Ian H. Witten）等著. —北京：机械工业出版社，2017.4
（经典原版书库）
书名原文：Data Mining: Practical Machine Learning Tools and Techniques, Fourth Edition

ISBN 978-7-111-56527-7

I. 数… II. 伊… III. ① 数据采集－英文 ② 机器学习－英文 IV. ① TP274 ② TP181

中国版本图书馆 CIP 数据核字（2017）第 067529 号

本书版权登记号：图字：01-2017-0510

Elsevier (Singapore) Pte Ltd.
3 Killiney Road
#08-01 Winsland House I
Singapore 239519
Tel: (65) 6349-0200
Fax: (65) 6733-1817
First Published 2017

出版发行：机械工业出版社（北京市西城区百万庄大街 22 号 邮政编码：100037）

责任编辑：曲 熠	责任校对：殷 虹
印 刷：北京瑞德印刷有限公司	版 次：2017 年 4 月第 1 版第 1 次印刷
开 本：186mm×240mm 1/16	印 张：40.75
书 号：ISBN 978-7-111-56527-7	定 价：129.00 元

凡购本书，如有缺页、倒页、脱页，由本社发行部调换
客服热线：（010）88378991 88361066
购书热线：（010）68326294 88379649 68995259
投稿热线：（010）88379604
读者信箱：hzjsj@hzbook.com

出版者的话

文艺复兴以来，源远流长的科学精神和逐步形成的学术规范，使西方国家在自然科学的各个领域取得了垄断性的优势；也正是这样的优势，使美国在信息技术发展的六十多年间名家辈出、独领风骚。在商业化的进程中，美国的产业界与教育界越来越紧密地结合，计算机学科中的许多泰山北斗同时身处科研和教学的最前线，由此而产生的经典科学著作，不仅擘划了研究的范畴，还揭示了学术的源变，既遵循学术规范，又自有学者个性，其价值并不会因年月的流逝而减退。

近年，在全球信息化大潮的推动下，我国的计算机产业发展迅猛，对专业人才的需求日益迫切。这对计算机教育界和出版界都既是机遇，也是挑战；而专业教材的建设在教育战略上显得举足轻重。在我国信息技术发展时间较短的现状下，美国等发达国家在其计算机科学发展的几十年间积淀和发展的经典教材仍有许多值得借鉴之处。因此，引进一批国外优秀计算机教材将对我国计算机教育事业的发展起到积极的推动作用，也是与世界接轨、建设真正的世界一流大学的必由之路。

机械工业出版社华章公司较早意识到“出版要为教育服务”。自 1998 年开始，我们就将工作重点放在了遴选、移译国外优秀教材上。经过多年的不懈努力，我们与 Pearson，McGraw-Hill，Elsevier，MIT，John Wiley & Sons，Cengage 等世界著名出版公司建立了良好的合作关系，从他们现有的数百种教材中甄选出 Andrew S. Tanenbaum，Bjarne Stroustrup，Brian W. Kernighan，Dennis Ritchie，Jim Gray，Afred V. Aho，John E. Hopcroft，Jeffrey D. Ullman，Abraham Silberschatz，William Stallings，Donald E. Knuth，John L. Hennessy，Larry L. Peterson 等大师名家的一批经典作品，以“计算机科学丛书”为总称出版，供读者学习、研究及珍藏。大理石纹理的封面，也正体现了这套丛书的品位和格调。

“计算机科学丛书”的出版工作得到了国内外学者的鼎力相助，国内的专家不仅提供了中肯的选题指导，还不辞劳苦地担任了翻译和审校的工作；而原书的作者也相当关注其作品在中国的传播，有的还专门为其书的中译本作序。迄今，“计算机科学丛书”已经出版了近两百个品种，这些书籍在读者中树立了良好的口碑，并被许多高校采用为正式教材和参考书籍。其影印版“经典原版书库”作为姊妹篇也被越来越多实施双语教学的学校所采用。

权威的作者、经典的教材、一流的译者、严格的审校、精细的编辑，这些因素使我们的图书有了质量的保证。随着计算机科学与技术专业学科建设的不断完善和教材改革的逐渐深化，教育界对国外计算机教材的需求和应用都将步入一个新的阶段，我们的目标是尽善尽美，而反馈的意见正是我们达到这一终极目标的重要帮助。华章公司欢迎老师和读者对我们的工作提出建议或给予指正，我们的联系方法如下：

华章网站：www.hzbook.com

电子邮件：hzjsj@hzbook.com

联系电话：（010）88379604

联系地址：北京市西城区百万庄南街 1 号

邮政编码：100037

华章教育

华章科技图书出版中心

Preface

The convergence of computing and communication has produced a society that feeds on information. Yet most of the information is in its raw form: data. If *data* is characterized as recorded facts, then *information* is the set of patterns, or expectations, that underlie the data. There is a huge amount of information locked up in databases—information that is potentially important but has not yet been discovered or articulated. Our mission is to bring it forth.

Data mining is the extraction of implicit, previously unknown, and potentially useful information from data. The idea is to build computer programs that sift through databases automatically, seeking regularities or patterns. Strong patterns, if found, will likely generalize to make accurate predictions on future data. Of course, there will be problems. Many patterns will be banal and uninteresting. Others will be spurious, contingent on accidental coincidences in the particular dataset used. And real data is imperfect: some parts will be garbled, some missing. Anything that is discovered will be inexact: there will be exceptions to every rule and cases not covered by any rule. Algorithms need to be robust enough to cope with imperfect data and to extract regularities that are inexact but useful.

Machine learning provides the technical basis of data mining. It is used to extract information from the raw data in databases—information i.e., ideally, expressed in a comprehensible form and can be used for a variety of purposes. The process is one of abstraction: taking the data, warts and all, and inferring whatever structure underlies it. This book is about the tools and techniques of machine learning that are used in practical data mining for finding, and if possible describing, structural patterns in data.

As with any burgeoning new technology that enjoys intense commercial attention, the use of machine learning is surrounded by a great deal of hype in the technical—and sometimes the popular—press. Exaggerated reports appear of the secrets that can be uncovered by setting learning algorithms loose on oceans of data. But there is no magic in machine learning, no hidden power, no alchemy. Instead there is an identifiable body of simple and practical techniques that can often extract useful information from raw data. This book describes these techniques and shows how they work.

In many applications machine learning enables the acquisition of structural descriptions from examples. The kind of descriptions that are found can be used for prediction, explanation, and understanding. Some data mining applications focus on prediction: forecasting what will happen in new situations from data that describe what happened in the past, often by guessing the classification of new examples. But we are equally—perhaps more—interested in applications where the result of "learning" is an actual description of a structure that can be used to classify examples. This structural description supports explanation and understanding as well as prediction. In our experience, insights gained by the user are

of most interest in the majority of practical data mining applications; indeed, this is one of machine learning's major advantages over classical statistical modeling.

The book explains a wide variety of machine learning methods. Some are pedagogically motivated: simple schemes that are designed to explain clearly how the basic ideas work. Others are practical: real systems that are used in applications today. Many are contemporary and have been developed only in the last few years.

A comprehensive software resource has been created to illustrate the ideas in the book. Called the Waikato Environment for Knowledge Analysis, or WEKA[1] for short, it is available as Java source code at www.cs.waikato.ac.nz/ml/weka. It is a full, industrial-strength implementation of most of the techniques that are covered in this book. It includes illustrative code and working implementations of machine learning methods. It offers clean, spare implementations of the simplest techniques, designed to aid understanding of the mechanisms involved. It also provides a workbench that includes full, working, state-of-the-art implementations of many popular learning schemes that can be used for practical data mining or for research. Finally, it contains a framework, in the form of a Java class library, that supports applications that use embedded machine learning and even the implementation of new learning schemes.

The objective of this book is to introduce the tools and techniques for machine learning that are used in data mining. After reading it, you will understand what these techniques are and appreciate their strengths and applicability. If you wish to experiment with your own data, you will be able to do this easily with the WEKA software. But WEKA is by no means the only choice. For example, the freely available statistical computing environment R includes many machine learning algorithms. Devotees of the Python programming language might look at a popular library called *scikit-learn*. Modern "big data" frameworks for distributed computing, such as Apache Spark, include support for machine learning. There is a plethora of options for deploying machine learning in practice. This book discusses fundamental learning algorithms without delving into software-specific implementation details. When appropriate, we point out where the algorithms we discuss can be found in the WEKA software. We also briefly introduce other machine learning software for so-called "deep learning" from high-dimensional data. However, most software-specific information is relegated to appendices.

The book spans the gulf between the intensely practical approach taken by trade books that provide case studies on data mining and the more theoretical, principle-driven exposition found in current textbooks on machine learning. (A brief description of these books appears in the *Further reading* section at the end of chapter: What's it all about?) This gulf is rather wide. To apply machine learning techniques productively, you need to understand something about how

[1]Found only on the islands of New Zealand, the *weka* (pronounced to rhyme with "Mecca") is a flightless bird with an inquisitive nature.

they work; this is not a technology that you can apply blindly and expect to get good results. Different problems yield to different techniques, but it is rarely obvious which techniques are suitable for a given situation: you need to know something about the range of possible solutions. And we cover an extremely wide range of techniques. We can do this because, unlike many trade books, this volume does not promote any particular commercial software or approach. We include a large number of examples, but they use illustrative datasets that are small enough to allow you to follow what is going on. Real datasets are far too large to show this (and in any case are usually company confidential). Our datasets are chosen not to illustrate actual large-scale practical problems, but to help you understand what the different techniques do, how they work, and what their range of application is.

The book is aimed at the technically aware general reader who is interested in the principles and ideas underlying the current practice of machine learning. It will also be of interest to information professionals who need to become acquainted with this new technology, and to all those who wish to gain a detailed technical understanding of what machine learning involves. It is written for an eclectic audience of information systems practitioners, programmers, consultants, developers, data scientists, information technology managers, specification writers, patent examiners, curious lay people—as well as students and professors—who need an easy-to-read book with lots of illustrations that describes what the major machine learning techniques are, what they do, how they are used, and how they work. It is practically oriented, with a strong "how to" flavor, and includes algorithms, and often pseudo-code. All those involved in practical data mining will benefit directly from the techniques described. The book is aimed at people who want to cut through to the reality that underlies the hype about machine learning and who seek a practical, nonacademic, unpretentious approach. In most of the book we have avoided requiring any specific theoretical or mathematical knowledge. However, recognizing the growing complexity of the subject as it matures, we have included substantial theoretical material in Chapter 9, Probabilistic methods, and Chapter 10, Deep learning, because this is necessary for a full appreciation of recent practical techniques, in particular deep learning.

The book is organized in layers that make the ideas accessible to readers who are interested in grasping the basics, as well as to those who would like more depth of treatment, along with full details on the techniques covered. We believe that consumers of machine learning need to have some idea of how the algorithms they use work. It is often observed that data models are only as good as the person who interprets them, and that person needs to know something about how the models are produced to appreciate the strengths, and limitations, of the technology. However, it is not necessary for all users to have a deep understanding of the finer details of the algorithms.

We address this situation by describing machine learning methods at successive levels of detail. The book is divided into two parts. Part I is an introduction to machine learning for data mining. The reader will learn the basic ideas, the

topmost level, by reading the first three chapters. Chapter 1, What's it all about?, describes, through examples, what machine learning is, where it can be used; it also provides actual practical applications. Chapter 2, Input: concepts, instances, attributes, and Chapter 3, Output: knowledge representation, cover the different kinds of input and output—or *knowledge representation*—that are involved. Different kinds of output dictate different styles of algorithm, and Chapter 4, Algorithms: the basic methods, describes the basic methods of machine learning, simplified to make them easy to comprehend. Here the principles involved are conveyed in a variety of algorithms without getting involved in intricate details or tricky implementation issues. To make progress in the application of machine learning techniques to particular data mining problems, it is essential to be able to measure how well you are doing. Chapter 5, Credibility: evaluating what's been learned, which can be read out of sequence, equips the reader to evaluate the results that are obtained from machine learning, addressing the sometimes complex issues involved in performance evaluation.

Part II introduces advanced techniques of machine learning for data mining. At the lowest and most detailed level, Chapter 6, Trees and rules, and Chapter 7, Extending instance-based and linear models, expose in naked detail the nitty-gritty issues of implementing a spectrum of machine learning algorithms, including the complexities that are necessary for them to work well in practice (but omitting the heavy mathematical machinery that is required for a few of the algorithms). Although many readers may want to ignore such detailed information, it is at this level that full working implementations of machine learning schemes are written. Chapter 8, Data transformations, describes practical topics involved with engineering the input and output to machine learning—e.g., selecting and discretizing attributes. Chapter 9, Probabilistic methods, and Chapter 10, Deep learning, provide a rigorous account of probabilistic methods for machine learning and deep learning respectively. Chapter 11, Beyond supervised and unsupervised learning, looks at semisupervised and multi-instance learning, while Chapter 12, Ensemble learning, covers techniques of "ensemble learning," which combine the output from different learning techniques. Chapter 13, Moving on: applications and beyond, looks to the future.

The book describes most methods used in practical machine learning. However, it does not cover reinforcement learning because it is rarely applied in practical data mining; nor genetic algorithm approaches because these are really just optimization techniques that are not specific to machine learning; nor relational learning and inductive logic programming because they are not very commonly used in mainstream data mining applications.

An Appendix covers some mathematical background needed to follow the material in Chapter 9, Probabilistic methods, and Chapter 10, Deep learning. Another Appendix introduces the WEKA data mining workbench, which provides implementations of most of the ideas described in Parts I and II. We have done this in order to clearly separate conceptual material from the practical aspects of

how to use it. At the end of each chapter in Parts I and II are pointers to related WEKA algorithms. You can ignore these, or look at them as you go along, or skip directly to the WEKA material if you are in a hurry to get on with analyzing your data and don't want to be bothered with the technical details of how the algorithms work.

UPDATED AND REVISED CONTENT

We finished writing the first edition of this book in 1999, the second and third in 2005 and 2011 respectively, and now, in May 2016, are just polishing this fourth edition. How things have changed over the past couple of decades! While the basic core of material remains the same, we have made the most of opportunities to update it and add new material, and as a result the book has doubled in size to reflect the changes that have taken place. Of course, there have also been errors to fix, errors that we had accumulated in our publicly available errata file (available through the book's home page at http://www.cs.waikato.ac.nz/ml/weka/book.html).

SECOND EDITION

The major change in the second edition of the book was a separate part at the end of the book that included all the material on the WEKA machine learning workbench. This allowed the main part of the book to stand alone, independent of the workbench. At that time WEKA, a widely used and popular feature of the first edition, had just acquired a radical new look in the form of an interactive graphical user interface—or rather, three separate interactive interfaces—which made it far easier to use. The primary one is the "Explorer," which gives access to all of WEKA's facilities using menu selection and form filling. The others are the Knowledge Flow interface, which allows you to design configurations for streamed data processing, and the Experimenter, with which you set up automated experiments that run selected machine learning algorithms with different parameter settings on a corpus of datasets, collect performance statistics, and perform significance tests on the results. These interfaces lower the bar for becoming a practitioner of machine learning, and the second edition included a full description of how to use them.

It also contained much new material that we briefly mention here. We extended the sections on rule learning and cost-sensitive evaluation. Bowing to popular demand, we added information on neural networks: the perceptron and the closely related Winnow algorithm; the multilayer perceptron and backpropagation algorithm. Logistic regression was also included. We described how to implement nonlinear decision boundaries using both the kernel perceptron and

radial basis function networks, and also included support vector machines for regression. We incorporated a new section on Bayesian networks, again in response to readers' requests and WEKA's new capabilities in this regard, with a description of how to learn classifiers based on these networks, and how to implement them efficiently using AD trees.

The previous 5 years (1999–2004) had seen great interest in data mining for text, and this was reflected in the introduction of string attributes in WEKA, multinomial Bayes for document classification, and text transformations. We also described efficient data structures for searching the instance space: *k*D-trees and ball trees for finding nearest neighbors efficiently, and for accelerating distance-based clustering. We described new attribute selection schemes such as race search and the use of support vector machines; new methods for combining models such as additive regression, additive logistic regression, logistic model trees, and option trees. We also covered recent developments in using unlabeled data to improve classification, including the cotraining and co-EM methods.

THIRD EDITION

For the third edition, we thoroughly edited the second edition and brought it up to date, including a great many new methods and algorithms. WEKA and the book were closely linked together—pretty well everything in WEKA was covered in the book. We also included far more references to the literature, practically tripling the number of references that were in the first edition.

As well as becoming far easier to use, WEKA had grown beyond recognition over the previous decade, and matured enormously in its data mining capabilities. It incorporates an unparalleled range of machine learning algorithms and related techniques. The growth has been partly stimulated by recent developments in the field, and is partly user-led and demand-driven. This puts us in a position where we know a lot about what actual users of data mining want, and we have capitalized on this experience when deciding what to include in this book.

Here are a few of the highlights of the material that was added in the third edition. A section on web mining was included, and, under ethics, a discussion of how individuals can often be "reidentified" from supposedly anonymized data. Other additions included techniques for multi-instance learning, new material on interactive cost-benefit analysis, cost-complexity pruning, advanced association rule algorithms that use extended prefix trees to store a compressed version of the dataset in main memory, kernel ridge regression, stochastic gradient descent, and hierarchical clustering methods. We added new data transformations: partial least squares regression, reservoir sampling, one-class learning, decomposing multiclass classification problems into ensembles of nested dichotomies, and calibrating class probabilities. We added new information on ensemble learning techniques: randomization vs. bagging, and rotation forests. New sections on data stream learning and web mining were added as well.

FOURTH EDITION

One of the main drivers behind this fourth edition was a desire to add comprehensive material on the topic of deep learning, a new development that is essentially enabled by the emergence of truly vast data resources in domains like image and speech processing, and the availability of truly vast computational resources, including server farms and graphics processing units. However, deep learning techniques are heavily based on a potent mix of theory and practice. And we had also received other requests asking us to include more, and more rigorous, theoretical material.

This forced us to rethink the role of theory in the book. We bit the bullet and added two new theoretically oriented chapters. Chapter 10, Deep learning, covers deep learning itself, and its predecessor, Chapter 9, Probabilistic methods, gives a principled theoretical development of probabilistic methods that is necessary to understand a host of other new algorithms. We recognize that many of our readers will not want to stomach all this theory, and we assure them that the remainder of the book has intentionally been left at a far simpler mathematical level. But this additional theoretical base puts some key material in the hands of readers who aspire to understand rapidly advancing techniques from the research world.

Developments in WEKA have proceeded apace. It now provides ways of reaching out and incorporating other languages and systems, such as the popular R statistical computing language, the Spark and Hadoop frameworks for distributed computing, the Python and Groovy languages for scripting, and the MOA system for stream-oriented learning—to name but a few. Recognizing that it is not possible, and perhaps not desirable, to document such a comprehensive and fast-evolving system in a printed book, we have created a series of open online courses, *Data Mining with Weka*, *More Data Mining with Weka*, and *Advanced Data Mining with Weka*, to accompany the book (at https://weka.waikato.ac.nz).

The fourth edition contains numerous other updates and additions, and far more references to the literature. But enough of this: dive in and see for yourself.

ACKNOWLEDGMENTS

Writing the acknowledgments is always the nicest part! A lot of people have helped us, and we relish this opportunity to thank them. This book has arisen out of the machine learning research project in the Computer Science Department at the University of Waikato, New Zealand. We have received generous encouragement and assistance from the academic staff members early on in that project: John Cleary, Sally Jo Cunningham, Matt Humphrey, Lyn Hunt, Bob McQueen, Lloyd Smith, and Tony Smith. We also benefited greatly from interactions with staff members who arrived later: Michael Mayo and Robert Durrant. Special thanks go to Geoff Holmes, who led the project for many years, and Bernhard Pfahringer, who had significant input into many different aspects of the WEKA software. All who have worked on the machine learning project here have contributed to our thinking: we would particularly like to mention early students Steve Garner,

Stuart Inglis and Craig Nevill-Manning for helping us to get the project off the ground in the beginning when success was less certain and things were more difficult.

The WEKA system that illustrates the ideas in this book forms a crucial component of it. It was conceived by the authors and designed and implemented principally by Eibe Frank, Mark Hall, Peter Reutemann, and Len Trigg, but many people in the machine learning laboratory at Waikato made significant contributions. Since the first edition of the book the WEKA team has expanded considerably: so many people have contributed that it is impossible to acknowledge everyone properly. We are grateful to Chris Beckham, for contributing several packages to WEKA, Remco Bouckaert for his Bayes net package and many other contributions, Lin Dong for her implementations of multi-instance learning methods, Dale Fletcher for many database-related aspects, Kurt Driessens for his implementation of Gaussian process regression, James Foulds for his work on multi-instance filtering, Anna Huang for information bottleneck clustering, Martin Gütlein for his work on feature selection, Kathryn Hempstalk for her one-class classifier, Ashraf Kibriya and Richard Kirkby for contributions far too numerous to list, Nikhil Kishore for his implementation of elastic net regression, Niels Landwehr for logistic model trees, Chi-Chung Lau for creating all the icons for the Knowledge Flow interface, Abdelaziz Mahoui for the implementation of K*, Jonathan Miles for his implementation of kernel filtering, Stefan Mutter for association rule mining, Malcolm Ware for numerous miscellaneous contributions, Haijian Shi for his implementations of tree learners, Marc Sumner for his work on speeding up logistic model trees, Tony Voyle for least-median-of-squares regression, Yong Wang for Pace regression and the original implementation of M5′, Benjamin Weber for his great unification of WEKA parsing modules, and Xin Xu for his multi-instance learning package, *JRip*, logistic regression and many other contributions. Our sincere thanks go to all these people for their dedicated work, and also to the many contributors to WEKA from outside our group at Waikato.

Tucked away as we are in a remote (but very pretty) corner of the southern hemisphere, we greatly appreciate the visitors to our department who play a crucial role in acting as sounding boards and helping us to develop our thinking. We would like to mention in particular Rob Holte, Carl Gutwin, and Russell Beale, each of whom visited us for several months; David Aha, who although he only came for a few days did so at an early and fragile stage of the project and performed a great service by his enthusiasm and encouragement; and Kai Ming Ting, who worked with us for 2 years on many of the topics in this book, and helped to bring us into the mainstream of machine learning. More recent visitors included Arie Ben-David, Carla Brodley, Gregory Butler, Stefan Kramer, Johannes Schneider, Jan van Rijn and Michalis Vlachos, and many others who have given talks at our department. We would particularly like to thank Albert Bifet, who gave us detailed feedback on a draft version of the third edition—most of which we have incorporated.

Students at Waikato have played a significant role in the development of the project. Many of them are in the above list of WEKA contributors, but they have also contributed in other ways. In the early days, Jamie Littin worked on ripple-down rules and relational learning. Brent Martin explored instance-based learning and nested instance-based representations. Murray Fife slaved over relational learning, Nadeeka Madapathage investigated the use of functional languages for expressing machine learning algorithms. Kathryn Hempstalk worked on one-class learning and Richard Kirkby on data streams. Gabi Schmidberger worked on density estimation trees, Lan Huang on concept-based text clustering, and Alyona Medelyan on keyphrase extraction. More recently, Felipe Bravo has worked on sentiment classification for Twitter, Mi Li on fast clustering methods, and Tim

Leathart on ensembles of nested dichotomies. Other graduate students have influenced us in numerous ways, particularly Gordon Paynter, YingYing Wen, and Zane Bray, who have worked with us on text mining, and Quan Sun and Xiaofeng Yu. Colleagues Steve Jones and Malika Mahoui have also made far-reaching contributions to these and other machine learning projects. We have also learned much from our many visiting students from Freiburg, including Nils Weidmann.

Ian Witten would like to acknowledge the formative role of his former students at Calgary, particularly Brent Krawchuk, Dave Maulsby, Thong Phan, and Tanja Mitrovic, all of whom helped him develop his early ideas in machine learning, as did faculty members Bruce MacDonald, Brian Gaines, and David Hill at Calgary, and John Andreae at the University of Canterbury.

Eibe Frank is indebted to his former supervisor at the University of Karlsruhe, Klaus-Peter Huber, who infected him with the fascination of machines that learn. On his travels Eibe has benefited from interactions with Peter Turney, Joel Martin, and Berry de Bruijn in Canada; Luc de Raedt, Christoph Helma, Kristian Kersting, Stefan Kramer, Ulrich Rückert, and Ashwin Srinivasan in Germany.

Mark Hall thanks his former supervisor Lloyd Smith, now at Missouri State University, who exhibited the patience of Job when his thesis drifted from its original topic into the realms of machine learning. The many and varied people who have been part of, or have visited, the machine learning group at the University of Waikato over the years deserve a special thanks for their valuable insights and stimulating discussions.

Chris Pal thanks his coauthors for the invitation to help write this fourth edition; and his family for accommodating the extra time spent writing. He thanks Polytechique Montréal for the sabbatical that allowed him to travel to New Zealand, leading to this fruitful collaboration; and the University of Waikato Department of Computer Science for hosting him. He also thanks his many mentors, academic family, coauthors, and colleagues over the years, whose perspectives have been enriching and have influenced the presentation here, including: Brendan Frey, Geoff Hinton, Yoshua Bengio, Sam Roweis, Andrew McCallum, and Charles Sutton among many others. Particular thanks goes to Hugo Larochelle for his pedagogical perspectives on deep learning. Chris also thanks his friends and colleagues at the Montréal Institute for Learning Algorithms, the Theano development team and all his current and former students for helping to create such a great environment for machine learning research. Particular thanks goes to Chris Beckham who provided excellent feedback on earlier drafts of the new chapters in this edition.

Charlie Kent and Tim Pitts of Morgan Kaufmann have worked hard to shape this book, and Nicky Carter, our production editor, has made the process go very smoothly. We would like to thank the librarians of the Repository of Machine Learning Databases at the University of California, Irvine, whose carefully collected datasets have been invaluable in our research.

Our research has been funded by the New Zealand Foundation for Research, Science and Technology and the Royal Society of New Zealand Marsden Fund. The Department of Computer Science at the University of Waikato has generously supported us in all sorts of ways, and we owe a particular debt of gratitude to Mark Apperley for his enlightened leadership and warm encouragement. Part of the first edition was written while both authors were visiting the University of Calgary, Canada, and the support of the Computer Science department there is gratefully acknowledged—as well as the positive and helpful attitude of the long-suffering students in the machine learning course, on whom we experimented. Part of the second edition was written at the University of Lethbridge in Southern Alberta on a visit supported by Canada's Informatics Circle of Research Excellence.

Last, and most of all, we are grateful to our families and partners. Pam, Anna and Nikki were all too well aware of the implications of having an author in the house ("not again!") but let Ian go ahead and write the book anyway. Julie was always supportive, even when Eibe had to burn the midnight oil in the machine learning lab, and Immo and Ollig provided exciting diversions. Bernadette too was very supportive, somehow managing to keep the combined noise output of Charlotte, Luke, Zach, Kyle, and Francesca to a level that allowed Mark to concentrate. Between us we hail from Canada, England, Germany, Ireland, New Zealand, and Samoa: New Zealand has brought us together and provided an ideal, even idyllic, place to do this work.

Contents