Preface

"He who defends everything, defends nothing."  
— Frederick the Great

# Machine learning is a very broad topic. The following quote sums it up nicely: *The first problem facing you is the bewildering variety of learning algorithms available.* *Which one to use? There are literally thousands available, and hundreds more are published* *each year.* (Domingo, P., 2012.) It would therefore be irresponsible to try and cover everything in the chapters that follow because, to paraphrase Frederick the Great, we would achieve nothing. With this constraint in mind, we hope to provide a solid foundation of algorithms and business considerations that will allow the reader to walk away and, first of all, take on any machine learning tasks with complete confidence, and secondly, be able to help themselves in figuring out other algorithms and topics. Essentially, if this course significantly helps you to help yourself, then I would consider this a victory. Don't think of this course as a destination but rather, as a path to self-discovery. What this learning path covers

Module 1, *R Machine Learning By Example*, Data science and machine learning are some of the top buzzwords in the technical world today. From retail stores to Fortune 500 companies, everyone is working hard to make machine learning give them data-driven insights to grow their businesses. With powerful data manipulation features, machine learning packages, and an active developer community, R empowers users to build sophisticated machine learning systems to solve real-world data problems. This module takes you on a data-driven journey that starts with the very basics of R and machine learning and gradually builds upon the concepts to work on projects that tackle real-world problems.

Module 2, *Machine Learning with R*, Machine learning, at its core, is concerned with the algorithms that transform information into actionable intelligence. This fact makes machine learning well-suited to the present-day era of big data. Without machine learning, it would be nearly impossible to keep up with the massive stream of information. Given the growing prominence of R—a cross-platform, zero-cost statistical  
programming environment—there has never been a better time to start using  
machine learning. R offers a powerful but easy-to-learn set of tools that can  
assist you with finding data insights. By combining hands-on case studies with the essential theory that you need to understand how things work under the hood, this book provides all the knowledge that you will need to start applying machine learning to your own projects. .

Module 3 *Mastering Machine Learning with R*, The world of R can be as bewildering as the world of machine learning! There is seemingly an endless number of R packages with a plethora of blogs, websites, discussions, and papers of various quality and complexity from the community that supports R. This is a great reservoir of information and probably R's greatest strength, but I've always believed that an entity's greatest strength can also be its greatest weakness. R's vast community of knowledge can quickly overwhelm and/or sidetrack  
you and your efforts. Show me a problem and give me ten different R programmers  
and I'll show you ten different ways the code is written to solve the problem. As I've  
written each chapter, I've endeavored to capture the critical elements that can assist  
you in using R to understand, prepare, and model the data. I am no R programming  
expert by any stretch of the imagination, but again, I like to think that I can provide a  
solid foundation herein. Another thing that lit a fire under me to write this book was an incident that happened in the hallways of a former employer a couple of years ago. My team had an IT contractor to support the management of our databases. As we were walking and chatting about big data and the like, he mentioned that he had bought a book about machine learning with R and another about machine learning with Python. He  
stated that he could do all the programming, but all of the statistics made absolutely  
no sense to him. I have always kept this conversation at the back of my mind throughout the writing process. It has been a very challenging task to balance the technical and theoretical with the practical. One could, and probably someone has, turned the theory of each chapter to its own book. I used a heuristic of sorts to aid me in deciding whether a formula or technical aspect was in the scope, which was would this help me or the readers in the discussions with team members and business leaders? If I felt it might help, I would strive to provide the necessary details. I also made a conscious effort to keep the datasets used in the practical exercises large enough to be interesting but small enough to allow you to gain insight without becoming overwhelmed. This book is not about big data, but make no mistake about it, the methods and concepts that we will discuss can be scaled to big data. In short, this module will appeal to a broad group of individuals, from IT experts seeking to understand and interpret machine learning algorithms to statistical gurus desiring to incorporate the power of R into their analysis. However, even those that are well-versed in both IT and statistics—experts if you will—should be able to pick up quite a few tips and tricks to assist them in their efforts.

# What you need for this learning path

This software applies to all the chapters of the book:  
• Windows / Mac OS X / Linux  
• R 3.2.0 (or higher)  
• RStudio Desktop 0.99 (or higher)  
For hardware, there are no specific requirements, since R can run on any PC that has  
Mac, Linux, or Windows, but a physical memory of minimum 4 GB is preferred to  
run some of the iterative algorithms smoothly.

# Who this learning path is for

Aimed for intermediate-to-advanced people (especially data scientist) who are already into the field of data science

# Reader feedback

Feedback from our readers is always welcome. Let us know what you think about this book—what you liked or disliked. Reader feedback is important for us as it helps us develop titles that you will really get the most out of.

To send us general feedback, simply e-mail feedback@packtpub.com, and mention the book's title in the subject of your message.

If there is a topic that you have expertise in and you are interested in either writing or contributing to a book, see our author guide at www.packtpub.com/authors.

# Customer support

Now that you are the proud owner of a Packt book, we have a number of things to help you to get the most from your purchase.

## Downloading the example code

You can download the example code files for this book from your account at http://www.packtpub.com. If you purchased this book elsewhere, you can visit http://www.packtpub.com/support and register to have the files e-mailed directly to you.

You can download the code files by following these steps:

1. Log in or register to our website using your e-mail address and password.
2. Hover the mouse pointer on the SUPPORT tab at the top.
3. Click on Code Downloads & Errata.
4. Enter the name of the book in the Search box.
5. Select the book for which you're looking to download the code files.
6. Choose from the drop-down menu where you purchased this book from.
7. Click on Code Download.

You can also download the code files by clicking on the Code Files button on the book's webpage at the Packt Publishing website. This page can be accessed by entering the book's name in the Search box. Please note that you need to be logged in to your Packt account.

Once the file is downloaded, please make sure that you unzip or extract the folder using the latest version of:

* WinRAR / 7-Zip for Windows
* Zipeg / iZip / UnRarX for Mac

7-Zip / PeaZip for Linux

The code bundle for the book is also hosted on GitHub at https://github.com/PacktPublishing/repository-name. We also have other code bundles from our rich catalog of books and videos available at https://github.com/PacktPublishing/. Check them out!

## Errata

Although we have taken every care to ensure the accuracy of our content, mistakes do happen. If you find a mistake in one of our books—maybe a mistake in the text or the code—we would be grateful if you could report this to us. By doing so, you can save other readers from frustration and help us improve subsequent versions of this book. If you find any errata, please report them by visiting http://www.packtpub.com/submit-errata, selecting your book, clicking on the Errata Submission Form link, and entering the details of your errata. Once your errata are verified, your submission will be accepted and the errata will be uploaded to our website or added to any list of existing errata under the Errata section of that title.

To view the previously submitted errata, go to https://www.packtpub.com/books/content/support and enter the name of the book in the search field. The required information will appear under the Errata section.

## Piracy

Piracy of copyrighted material on the Internet is an ongoing problem across all media. At Packt, we take the protection of our copyright and licenses very seriously. If you come across any illegal copies of our works in any form on the Internet, please provide us with the location address or website name immediately so that we can pursue a remedy.

Please contact us at copyright@packtpub.com with a link to the suspected pirated material.

We appreciate your help in protecting our authors and our ability to bring you valuable content.

## Questions

If you have a problem with any aspect of this book, you can contact us at questions@packtpub.com, and we will do our best to address the problem.