Data Mining With R

داده کاوی با R

Ali Ghavami
Somayeh Ghavami

سرشناسه	:	قوامی، علی، ۱۳۵۷ - Ghavami, Ali
عنوان و نام پدید آور	:	Data mining with R/Ali Ghavami, Somayeh Ghavami.
مشخصات نشر	:	تهران: ندای کارآفرین ، ۲۰۱۳م=۱۳۹۲.
مشخصات ظاهری	:	۱۶۰ ص؛ ۱۷×۲۴ س م.
شابک	:	978 - 600 - 93625 - 1 - 6
وضعیت فهرست نویسی	:	فیپا.
یادداشت	:	انگلیسی.
آواتویسی عنوان	:	دیتا ماینینگ —
موضوع	:	داده‌کاوی.
موضوع	:	داده‌کاوی— نرم‌افزار.
شناسه افزوده	:	قوامی، سمیه، ۱۳۵۶ -
شناسه افزوده	:	Ghavami, Somayeh
رده بندی کنگره	:	QA۷۶/۹د۲ ق ۱۳۹۲
رده بندی دیویی	:	۰۰۶/۳۱۲
شمارهٔ کتابشناسی ملی	:	۳۱۳۰۲۷۱

انتشارات ندای کارآفرین

داده کاوی به کمک R

مؤلفان: علی قوامی،سمیه قوامی

طراح جلد: مرتضی توحیدی

لیتوگرافی، چاپ و صحافی: ستاره نو

شمارگان: ۱۰۰۰ نسخه

نوبت و تاریخ چاپ: چاپ اول اردیبهشت ۱۳۹۲

شابک: ۶ - ۱ - ۹۳۶۲۵ - ۶۰۰ - ۹۷۸

ISBN: 978 - 600 - 93625 - 1 - 6

قیمت: ۴۰۰۰۰ریال

TABLE OF CONTENTS

PREFACE

The main goal of this book is to introduce the reader to the use of R as a tool for performing data mining. R is a freely downloadable language and environment for statistical computing and graphics. Its capabilities and the large set of available packages make this tool an excellent alternative to the existing (and expensive!) data mining tools.

One of the key issues in data mining is size. A typical data mining problem involves a large database from where one seeks to extract useful knowledge. In this book we will use MySQL as the core database management system. MySQL is also freely available for several computer platforms. This means that you will be able to perform "serious" data mining without having to pay any money at all. Moreover, we hope to show you that this comes with no compromise in the quality of the obtained solutions. Expensive tools do not necessarily mean better tools! R together with MySQL forms a pair very hard to beat as long as you are willing to spend some time learning how to use them. We think that it is worthwhile, and we hope that you are convinced as well at the end of reading this book.

The goal of this book is not to describe all facets of data mining processes. Many books exist that cover this area. Instead we propose to introduce the reader to the power of R and data mining by means of several case studies. Obviously, these case studies do not represent all possible data mining problems that one can face in the real world. Moreover, the solutions we describe cannot be taken as complete solutions. Our goal is more to introduce the reader to the world of data mining using R through practical examples. As such our analysis of the cases studies has the goal of showing examples of knowledge extraction using R, instead of presenting complete reports of data mining case studies. They should be taken as examples of possible paths in any data mining project and can be used as the basis for developing solutions for the reader's data mining projects. Still, we have tried to cover a diverse set of problems posing different challenges in terms of size, type of data, goals of analysis and tools that are necessary to carry out this analysis.

We do not assume any prior knowledge about R. Readers that are new to R and data mining should be able to follow the case studies. We have tried to make the different case studies self-contained in such a way that the reader can start anywhere in the document. Still, some basic R functionalities are introduced in the first, simpler, case studies, and are not repeated, which means that if you are new to R, then you should at least start with the first case studies to get acquainted with R. Moreover, the first chapter provides a very short introduction to R basics, which may facilitate the understanding of the following chapters.

We also do not assume any familiarity with data mining or statistical techniques. Brief introductions to different modeling approaches are provided as they are necessary in the case studies. It is not an objective of this book to provide the reader with full information

on the technical and theoretical details of these techniques. Our descriptions of these models are given to provide basic understanding on their merits, drawbacks and analysis objectives. Other existing books should be considered if further theoretical insights are required. At the end of some sections we provide "Further readings" pointers for the readers interested in knowing more on the topics. In summary, our target readers are more users of data analysis tools than researchers or developers. Still, we hope the latter also find reading this book useful as a form of entering the "world" of R and data mining.

Ali Ghavami

http://www.future-gate.ir *a_ghavami@future-gate.ir*
a_ghavami@pnu.ac.ir

Somayeh Ghavami

s.ghavami977@gmail.com