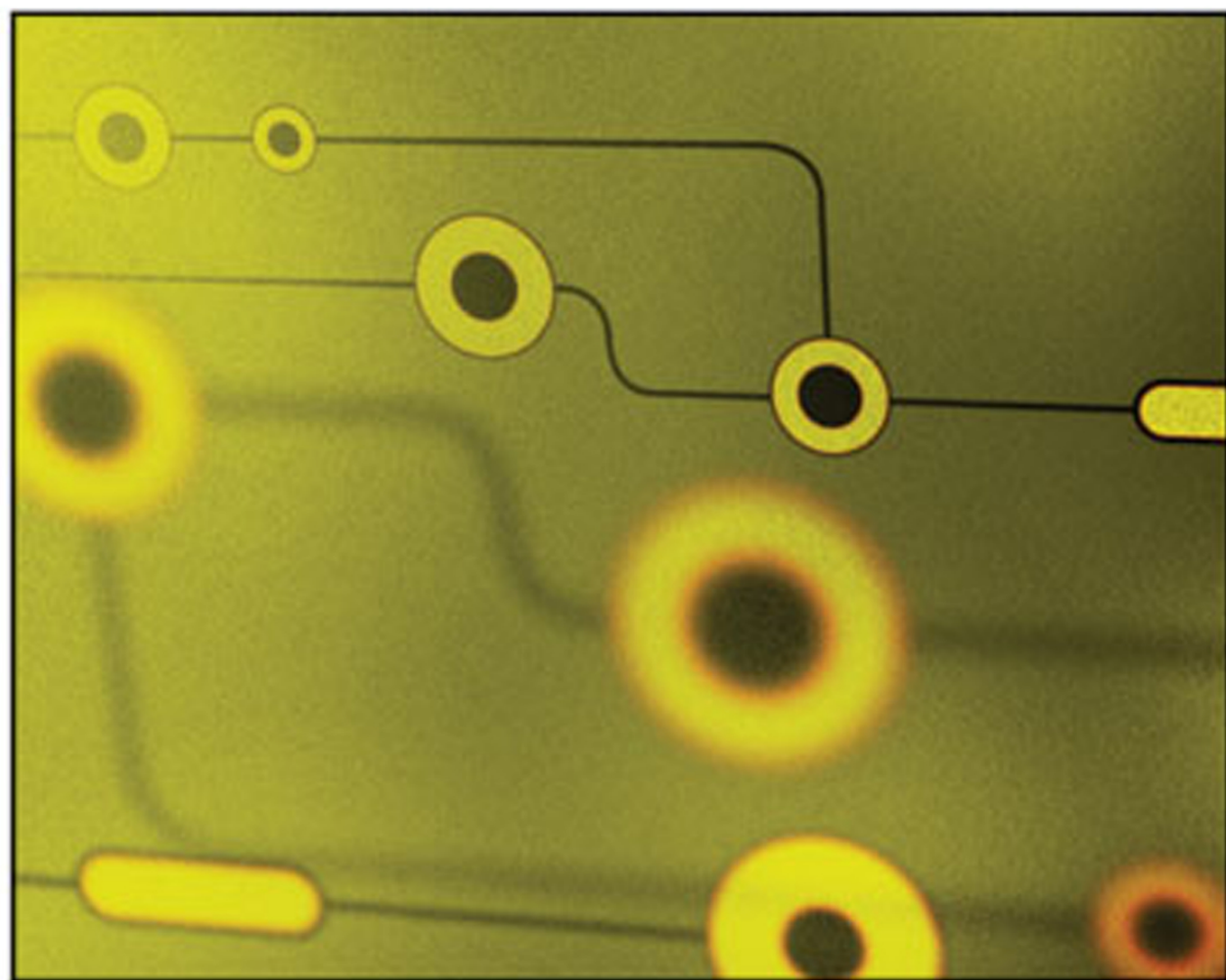


PREMIER REFERENCE SOURCE

# DATA MINING AND MEDICAL KNOWLEDGE MANAGEMENT

CASES AND APPLICATIONS



Petr Berka, Jan Rauch, & Djamel Abdelkader Zighed

# Data Mining and Medical Knowledge Management: Cases and Applications

Petr Berka

*University of Economics, Prague, Czech Republic*

Jan Rauch

*University of Economics, Prague, Czech Republic*

Djamel Abdelkader Zighed

*University of Lumiere Lyon 2, France*



MEDICAL INFORMATION SCIENCE REFERENCE

Hershey · New York

Director of Editorial Content: Kristin Klinger  
Managing Editor: Jamie Snavely  
Assistant Managing Editor: Carole Coulson  
Typesetter: Sean Woznicki  
Cover Design: Lisa Tosheff  
Printed at: Yurchak Printing Inc.

Published in the United States of America by  
Information Science Reference (an imprint of IGI Global)  
701 E. Chocolate Avenue, Suite 200  
Hershey PA 17033  
Tel: 717-533-8845  
Fax: 717-533-8661  
E-mail: [cust@igi-global.com](mailto:cust@igi-global.com)  
Web site: <http://www.igi-global.com/reference>

and in the United Kingdom by  
Information Science Reference (an imprint of IGI Global)  
3 Henrietta Street  
Covent Garden  
London WC2E 8LU  
Tel: 44 20 7240 0856  
Fax: 44 20 7379 0609  
Web site: <http://www.eurospanbookstore.com>

Copyright © 2009 by IGI Global. All rights reserved. No part of this publication may be reproduced, stored or distributed in any form or by any means, electronic or mechanical, including photocopying, without written permission from the publisher.

Product or company names used in this set are for identification purposes only. Inclusion of the names of the products or companies does not indicate a claim of ownership by IGI Global of the trademark or registered trademark.

#### Library of Congress Cataloging-in-Publication Data

Data mining and medical knowledge management : cases and applications / Petr Berka, Jan Rauch, and Djamel Abdelkader Zighed, editors.  
p. ; cm.

Includes bibliographical references and index.

Summary: "This book presents 20 case studies on applications of various modern data mining methods in several important areas of medicine, covering classical data mining methods, elaborated approaches related to mining in EEG and ECG data, and methods related to mining in genetic data"--Provided by publisher.

ISBN 978-1-60566-218-3 (hardcover)

1. Medicine--Data processing--Case studies. 2. Data mining--Case studies. I. Berka, Petr. II. Rauch, Jan. III. Zighed, Djamel A., 1955- [DNLM: 1. Medical Informatics--methods--Case Reports. 2. Computational Biology--methods--Case Reports. 3. Information Storage and Retrieval--methods--Case Reports. 4. Risk Assessment--Case Reports. W 26.5 D2314 2009]

R858.D33 2009

610.0285--dc22

2008028366

#### British Cataloguing in Publication Data

A Cataloguing in Publication record for this book is available from the British Library.

All work contributed to this book is new, previously-unpublished material. The views expressed in this book are those of the authors, but not necessarily of the publisher.

*If a library purchased a print copy of this publication, please go to <http://www.igi-global.com/agreement> for information on activating the library's complimentary electronic access to this publication.*

## Editorial Advisory Board

Riccardo Bellazzi, *University of Pavia, Italy*  
Radim Jiroušek, *Academy of Sciences, Prague, Czech Republic*  
Katharina Morik, *University of Dortmund, Germany*  
Ján Paralič, *Technical University, Košice, Slovak Republic*  
Luis Torgo, *LIAAD-INESC Porto LA, Portugal*  
Blaž Župan, *University of Ljubljana, Slovenia*

## List of Reviewers

Ricardo Bellazzi, *University of Pavia, Italy*  
Petr Berka, *University of Economics, Prague, Czech Republic*  
Bruno Crémilleux, *University Caen, France*  
Peter Eklund, *Umeå University, Umeå, Sweden*  
Radim Jiroušek, *Academy of Sciences, Prague, Czech Republic*  
Jiří Kléma, *Czech Technical University, Prague, Czech Republic*  
Mila Kwiatkowska, *Thompson Rivers University, Kamloops, Canada*  
Martin Labský, *University of Economics, Prague, Czech Republic*  
Lenka Lhotská, *Czech Technical University, Prague, Czech Republic*  
Ján Paralič, *Technical University, Kosice, Slovak Republic*  
Vincent Pisetta, *University Lyon 2, France*  
Simon Marcellin, *University Lyon 2, France*  
Jan Rauch, *University of Economics, Prague, Czech Republic*  
Marisa Sánchez, *National University, Bahía Blanca, Argentina*  
Ahmed-El Sayed, *University Lyon 2, France*  
Olga Štěpánková, *Czech Technical University, Prague, Czech Republic*  
Vojtěch Svátek, *University of Economics, Prague, Czech Republic*  
Arnošt Veselý, *Czech University of Life Sciences, Prague, Czech Republic*  
Djamel Zighed, *University Lyon 2, France*

# Table of Contents

Foreword .....	xiv
Preface .....	xix
Acknowledgment .....	xxiii

## Section I Theoretical Aspects

### Chapter I

Data, Information and Knowledge.....	1
<i>Jana Zvárová, Institute of Computer Science of the Academy of Sciences of the Czech Republic v.v.i., Czech Republic; Center of Biomedical Informatics, Czech Republic</i>	
<i>Arnošt Veselý, Institute of Computer Science of the Academy of Sciences of the Czech Republic v.v.i., Czech Republic; Czech University of Life Sciences, Czech Republic</i>	
<i>Igor Vajda, Institutes of Computer Science and Information Theory and Automation of the Academy of Sciences of the Czech Republic v.v.i., Czech Republic</i>	

### Chapter II

Ontologies in the Health Field.....	37
<i>Michel Simonet, Laboratoire TIMC-IMAG, Institut de l'Ingénierie et de l'Information de Santé, France</i>	
<i>Radja Messai, Laboratoire TIMC-IMAG, Institut de l'Ingénierie et de l'Information de Santé, France</i>	
<i>Gayo Diallo, Laboratoire TIMC-IMAG, Institut de l'Ingénierie et de l'Information de Santé, France</i>	
<i>Ana Simonet, Laboratoire TIMC-IMAG, Institut de l'Ingénierie et de l'Information de Santé, France</i>	

### Chapter III

Cost-Sensitive Learning in Medicine.....	57
<i>Alberto Freitas, University of Porto, Portugal; CINTESIS, Portugal</i>	
<i>Pavel Brazdil, LIAAD - INESC Porto L.A., Portugal; University of Porto, Portugal</i>	
<i>Altamiro Costa-Pereira, University of Porto, Portugal; CINTESIS, Portugal</i>	

<b>Chapter IV</b>	
Classification and Prediction with Neural Networks.....	76
<i>Arnošt Veselý, Czech University of Life Sciences, Czech Republic</i>	

<b>Chapter V</b>	
Preprocessing Perceptrons and Multivariate Decision Limits.....	108
<i>Patrik Eklund, Umeå University, Sweden</i>	
<i>Lena Kallin Westin, Umeå University, Sweden</i>	

## **Section II**

### **General Applications**

<b>Chapter VI</b>	
Image Registration for Biomedical Information Integration.....	122
<i>Xiu Ying Wang, BMIT Research Group, The University of Sydney, Australia</i>	
<i>Dagan Feng, BMIT Research Group, The University of Sydney, Australia; Hong Kong Polytechnic University, Hong Kong</i>	

<b>Chapter VII</b>	
ECG Processing.....	137
<i>Lenka Lhotská, Czech Technical University in Prague, Czech Republic</i>	
<i>Václav Chudáček, Czech Technical University in Prague, Czech Republic</i>	
<i>Michal Huptych, Czech Technical University in Prague, Czech Republic</i>	

<b>Chapter VIII</b>	
EEG Data Mining Using PCA.....	161
<i>Lenka Lhotská, Czech Technical University in Prague, Czech Republic</i>	
<i>Vladimír Krajča, Faculty Hospital Na Bulovce, Czech Republic</i>	
<i>Jitka Mohylová, Technical University Ostrava, Czech Republic</i>	
<i>Svojmil Petránek, Faculty Hospital Na Bulovce, Czech Republic</i>	
<i>Václav Gerla, Czech Technical University in Prague, Czech Republic</i>	

<b>Chapter IX</b>	
Generating and Verifying Risk Prediction Models Using Data Mining.....	181
<i>Darryl N. Davis, University of Hull, UK</i>	
<i>Thuy T.T. Nguyen, University of Hull, UK</i>	

## Chapter X

Management of Medical Website Quality Labels via Web Mining..... 206

*Vangelis Karkaletsis, National Center of Scientific Research “Demokritos”, Greece*

*Konstantinos Stamatakis, National Center of Scientific Research “Demokritos”, Greece*

*Pythagoras Karampiperis, National Center of Scientific Research “Demokritos”, Greece*

*Martin Labský, University of Economics, Prague, Czech Republic*

*Marek Růžička, University of Economics, Prague, Czech Republic*

*Vojtěch Svátek, University of Economics, Prague, Czech Republic*

*Enrique Amigó Cabrera, ETSI Informática, UNED, Spain*

*Matti Pöllä, Helsinki University of Technology, Finland*

*Miquel Angel Mayer, Medical Association of Barcelona (COMB), Spain*

*Dagmar Villarroel Gonzales, Agency for Quality in Medicine (AquMed), Germany*

## Chapter XI

Two Case-Based Systems for Explaining Exceptions in Medicine..... 227

*Rainer Schmidt, University of Rostock, Germany*

## Section III Specific Cases

## Chapter XII

Discovering Knowledge from Local Patterns in SAGE Data..... 251

*Bruno Crémilleux, Université de Caen, France*

*Arnaud Soulet, Université François Rabelais de Tours, France*

*Jiří Kléma, Czech Technical University, in Prague, Czech Republic*

*Céline Hébert, Université de Caen, France*

*Olivier Gandrillon, Université de Lyon, France*

## Chapter XIII

Gene Expression Mining Guided by Background Knowledge..... 268

*Jiří Kléma, Czech Technical University in Prague, Czech Republic*

*Filip Železný, Czech Technical University in Prague, Czech Republic*

*Igor Trajkovski, Jožef Stefan Institute, Slovenia*

*Filip Karel, Czech Technical University in Prague, Czech Republic*

*Bruno Crémilleux, Université de Caen, France*

*Jakub Tolar, University of Minnesota, USA*

## Chapter XIV

Mining Tinnitus Database for Knowledge..... 293

*Pamela L. Thompson, University of North Carolina at Charlotte, USA*

*Xin Zhang, University of North Carolina at Pembroke, USA*

*Wenxin Jiang, University of North Carolina at Charlotte, USA*

*Zbigniew W. Ras, University of North Carolina at Charlotte, USA*

*Pawel Jastreboff, Emory University School of Medicine, USA*

<b>Chapter XV</b>	
Gaussian-Stacking Multiclassifiers for Human Embryo Selection.....	307
<i>Dinora A. Morales, University of the Basque Country, Spain</i>	
<i>Endika Bengoetxea, University of the Basque Country, Spain</i>	
<i>Pedro Larrañaga, Universidad Politécnica de Madrid, Spain</i>	
<b>Chapter XVI</b>	
Mining Tuberculosis Data.....	332
<i>Marisa A. Sánchez, Universidad Nacional del Sur, Argentina</i>	
<i>Sonia Uremovich, Universidad Nacional del Sur, Argentina</i>	
<i>Pablo Acrogliano, Hospital Interzonal Dr. José Penna, Argentina</i>	
<b>Chapter XVII</b>	
Knowledge-Based Induction of Clinical Prediction Rules.....	350
<i>Mila Kwiatkowska, Thompson Rivers University, Canada</i>	
<i>M. Stella Atkins, Simon Fraser University, Canada</i>	
<i>Les Matthews, Thompson Rivers University, Canada</i>	
<i>Najib T. Ayas, University of British Columbia, Canada</i>	
<i>C. Frank Ryan, University of British Columbia, Canada</i>	
<b>Chapter XVIII</b>	
Data Mining in Atherosclerosis Risk Factor Data.....	376
<i>Petr Berka, University of Economics, Prague, Czech Republic; Academy of Sciences of the</i>	
<i>Czech Republic, Prague, Czech Republic</i>	
<i>Jan Rauch, University of Economics, Praague, Czech Republic; Academy of Sciences of the</i>	
<i>Czech Republic, Prague, Czech Republic</i>	
<i>Marie Tomečková, Academy of Sciences of the Czech Republic, Prague, Czech Republic</i>	
<b>Compilation of References</b> .....	398
<b>About the Contributors</b> .....	426
<b>Index</b> .....	437



# Detailed Table of Contents

<b>Foreword</b> .....	xiv
<b>Preface</b> .....	xix
<b>Acknowledgment</b> .....	xxiii

## **Section I** **Theoretical Aspects**

This section provides a theoretical and methodological background for the remaining parts of the book. It defines and explains basic notions of data mining and knowledge management, and discusses some general methods.

### **Chapter I**

Data, Information and Knowledge.....	1
<i>Jana Zvárová, Institute of Computer Science of the Academy of Sciences of the Czech Republic v.v.i., Czech Republic; Center of Biomedical Informatics, Czech Republic</i>	
<i>Arnošt Veselý, Institute of Computer Science of the Academy of Sciences of the Czech Republic v.v.i., Czech Republic; Czech University of Life Sciences, Czech Republic</i>	
<i>Igor Vajda, Institutes of Computer Science and Information Theory and Automation of the Academy of Sciences of the Czech Republic v.v.i., Czech Republic</i>	

This chapter introduces the basic concepts of medical informatics: data, information, and knowledge. It shows how these concepts are interrelated and can be used for decision support in medicine. All discussed approaches are illustrated on one simple medical example.

### **Chapter II**

Ontologies in the Health Field .....	37
<i>Michel Simonet, Laboratoire TIMC-IMAG, Institut de l'Ingénierie et de l'Information de Santé, France</i>	
<i>Radja Messai, Laboratoire TIMC-IMAG, Institut de l'Ingénierie et de l'Information de Santé, France</i>	
<i>Gayo Diallo, Laboratoire TIMC-IMAG, Institut de l'Ingénierie et de l'Information de Santé, France</i>	
<i>Ana Simonet, Laboratoire TIMC-IMAG, Institut de l'Ingénierie et de l'Information de Santé, France</i>	

This chapter introduces the basic notions of ontologies, presents a survey of their use in medicine, and explores some related issues: knowledge bases, terminology, information retrieval. It also addresses the issues of ontology design, ontology representation, and the possible interaction between data mining and ontologies.

### **Chapter III**

Cost-Sensitive Learning in Medicine..... 57

*Alberto Freitas, University of Porto, Portugal; CINTESIS, Portugal*

*Pavel Brazdil, LIAAD - INESC Porto L.A., Portugal; University of Porto, Portugal*

*Altamiro Costa-Pereira, University of Porto, Portugal; CINTESIS, Portugal*

Health managers and clinicians often need models that try to minimize several types of costs associated with healthcare, including attribute costs (e.g. the cost of a specific diagnostic test) and misclassification costs (e.g. the cost of a false negative test). This chapter presents some concepts related to cost-sensitive learning and cost-sensitive classification in medicine and reviews research in this area.

### **Chapter IV**

Classification and Prediction with Neural Networks..... 76

*Arnošt Veselý, Czech University of Life Sciences, Czech Republic*

This chapter describes the theoretical background of artificial neural networks (architectures, methods of learning) and shows how these networks can be used in medical domain to solve various classification and regression problems.

### **Chapter V**

Preprocessing Perceptrons and Multivariate Decision Limits..... 108

*Patrik Eklund, Umeå University, Sweden*

*Lena Kallin Westin, Umeå University, Sweden*

This chapter introduces classification networks composed of preprocessing layers and classification networks, and compares them with “classical” multilayer perceptrons on three medical case studies.

## **Section II General Applications**

This section presents work that is general in the sense of a variety of methods or variety of problems described in each of the chapters.

### **Chapter VI**

Image Registration for Biomedical Information Integration..... 122

*Xiu Ying Wang, BMIT Research Group, The University of Sydney, Australia*

*Dagan Feng, BMIT Research Group, The University of Sydney, Australia; Hong Kong Polytechnic University, Hong Kong*