

Introduction to Data Mining with Case Studies

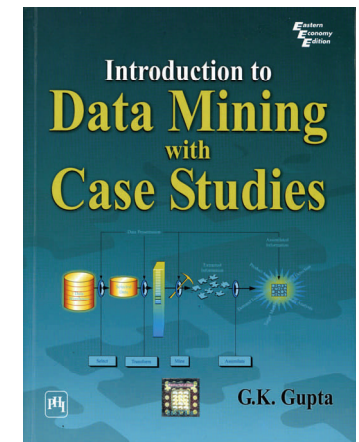
G.K. Gupta

THE BOOK

The field of data mining provides techniques for automated discovery of most valuable information from the accumulated data of computerized operations of enterprises. This book offers a clear and comprehensive introduction to both data mining theory and practice. It is written primarily as a textbook for students of computer science, management, computer applications and information technology.

The book ensures that students are exposed to all major data mining techniques without requiring a strong mathematical background. To enhance the understanding of the concepts introduced, and to show how the techniques described in the book are used in practice, each chapter is followed by one or two case studies that have been published in scholarly journals. Most case studies deal with real business problems (for example, marketing, e-commerce, CRM). Studying the case studies provides the reader with a deep insight into data mining techniques.

The book also provides many examples, end-of-chapter exercises and a good list of references and Web resources especially those which are easy to understand and useful for students.



Contact the Author: gkgupta@acm.org
URL: <http://www.csse.monash.edu.au/~gopal>

Some powerpoint slides are available at
<http://www.csse.monash.edu.au/~gopal/Teaching/Datamining/index.html>

ABOUT THE AUTHOR

Professor Gupta has recently retired after a thirty five year distinguished career in academic management as well as in information technology education, research and professional activities. He started his academic career at Monash University and subsequently moved to James Cook University as Foundation professor and Head to establish a new department of computer science. He later served as Dean of the School of Information Technology at Bond University, Deputy Dean and then Acting Dean of the Faculty of Information Technology at Monash University.

Professor Gupta is currently Honorary Professor of Computer Science at Monash University, Clayton, Australia. He received his PhD from Monash University, his Master's in computer science from the University of Waterloo and his Bachelor's degree in engineering from the University of Roorkee (now the Indian Institute of Technology Roorkee).

Professor Gupta is a Fellow of the Association of Computing Machinery (ACM), a Fellow of the Australian Computer Society (ACS) and a Senior Member of the IEEE.

2006 • 476 pages • 17.8 x 23.5 cm • ISBN-81-203-3053-6

Prentice-Hall of India Private Limited

Regd. Office: M-97, Connaught Circus, New Delhi-110001, INDIA
Phones: 23411779, 23415326, 23418078 • Fax: 011-23417179

Email: phi@phindia.com • Website: www.phindia.com
To purchase outside India: www.amazon.co.uk

Contents

Chapter 1	INTRODUCTION
Case Study	Data Mining Techniques for Optimizing Inventories for Electronic Commerce
Case Study	Crime Data Mining: A General Framework and Some Examples

The introductory chapter explains data mining and the data mining process. A number of application areas of data mining are discussed and brief descriptions of case studies from different application areas are presented.

Chapter 2	ASSOCIATION RULE MINING
Case Study	Mining Customer Value: From Association Rule to Direct Marketing

This chapter presents techniques for *association rules mining* or *market basket analysis* which involves searching for interesting customer habits by looking at associations in customer data at a supermarket. The technique has many other applications including applications in marketing, medicine, classification and finance.

Chapter 3	CLASSIFICATION
Case Study	KDD Insurance Risk Assessment: A Case Study
Case Study	A Data Mining Approach for Retailing Bank Customer Attrition Analysis

This chapter on classification deals with one of the most widely used data mining techniques. The technique is called supervised classification since it is assumed that the classes in which data will be divided have been decided a priori. Classification then consists of training the system so that when a new object is presented to the system it is able to assign the object to one of the existing classes. The widely used decision tree approach is described.

Chapter 4	CLUSTER ANALYSIS
Case Study	Efficient Clustering of Very Large Document Collections

This chapter is about unsupervised classification or cluster analysis. In contrast to supervised classification described in Chapter 3, in cluster analysis the user does not know what classes or clusters exist in the data and the aim is to group the given data into meaningful groups. A number of cluster analysis methods including the K-Means method are described.

Chapter 5	WEB DATA MINING
Case Study	Lessons and Challenges from Mining Retail E-Commerce Data

This chapter introduces the relatively new area called Web mining in which data mining techniques are used to find interesting and potentially useful knowledge from Web data including hyperlink structures of the Web and the Web log data. Web mining includes Web content mining, Web structure mining and Web usage mining. All three are described.

Chapter 6	SEARCH ENGINES
Case Study	The Anatomy of a Large-Scale Hypertextual Web Search Engine

This chapter deals with search engines that are designed to retrieve pages when a user wishes to find information of interest on the Web. Various aspects of search engines including page ranking and Web crawlers are discussed.

Chapter 7	DATA WAREHOUSING
Case Study	Data Warehouse Governance: Best Practices at Blue Cross and Blue Shield of North Carolina

This chapter is devoted to issues related to an overall enterprise view of the data and data warehousing. It has been reported that several years ago Coca Cola could not quickly determine how many bottles were produced in a day since the information was distributed in more than 20 different computer systems in different locations. Creating a separate database that only stores information that is of interest to the management is discussed.

Chapter 8	ONLINE ANALYTICAL PROCESSING (OLAP)
Case Study	Discovering Web Access Patterns and Trends by Applying OLAP and Data Mining Technology on Web Logs

This chapter is about OLAP (Online Analytical Processing). OLAP may be viewed as using a database like using a spreadsheet in which data is presented in a two-dimensional table and all the summaries of columns and rows are easy to obtain. In contrast to a spreadsheet, OLAP can deal with many dimensions providing management with a variety of summaries and aggregations to assist decision making.

Chapter 9	INFORMATION PRIVACY AND DATA MINING
Case Study	Privacy Conflicts in CRM Services for Online Shops: A Case Study

This chapter is devoted to the issue of information privacy in data mining. It is shown that the information privacy principles developed some 20-30 years ago are not particularly effective in dealing with the privacy concerns that are being raised about increased use of data mining as a result of the 9/11 terrorist attacks.
