ISBN: 9788183713736

Data Warehousing and Mining

Prof. Dr. M. N. Viswanathan

Dean

Department of Computer Science and Engineering
and Information Technology

Park College of Engineering and Technology
Kaniyur
Coimbatore - 641659

SCITECH PUBLICATIONS (INDIA) PVT. LTD.

CHENNAI

www.scitechpublications.com

ISBN: 9788183713736

Preface

Objective of the book

There are many reasons for the need to acquire knowledge of Data Warehousing and Mining. The interest is impart fundamental knowledge of Data Warehousing and Mining and applications. ***Data Warehousing and Mining*** is a comprehensive textbook covering the syllabi of most of the Indian universities.

The Material and order of presentation

Chapter 1 deals with	***Introduction***.
Chapter 2 deals with	***Data Preprocessing, Language, Architectures, Concept Description***.
Chapter 3 deals with	***Data Warehouse and OLAP Technology***.
Chapter 4 deals with	***Data Cube Computation and Data Generalization***.
Chapter 5 deals with	***Mining Frequent Patterns, Associations and Correlations***.
Chapter 6 deals with	***Classification and Prediction***.
Chapter 7 deals with	***Cluster Analysis***.
Chapter 8 deals with	***Mining Stream, Time Series, and Sequence Data***.
Chapter 9 deals with	***Graph Mining, Social Network Analysis and Multidirectional Data Mining***.
Chapter 10 deals with	***Mining Object, Spatial, Multimedia, Text, and Web Data***.
Chapter 11 deals with	***Applications and Trends in Data Mining***.

This book is well written to fulfill the aim and objective of the course. It is hoped that this book will be found more useful by both students and teachers.

Scitech Publications deserve all praise for having brought out this book in its excellent from in record time.

Any suggestions for improvement in the utility of the book will be very much appreciated.

March, 2010 **Dr. M. N. Viswanathan**
Coimbatore.

ISBN: 9788183713736

Contents

Chapter 1 Introduction 1.1 - 1.13

1.1 Motivation for Data Mining 1.1
1.2 What is Data Mining? 1.1
1.3 Types of Data for Mining 1.4
1.4 What kinds of patterns can be mined? 1.6
1.5 Are all patterns interesting? 1.9
1.6 Classification of Data Mining Systems 1.10
1.7 Data Mining Tasks 1.10
1.8 Integration of Data Mining System with Data Base or Data Warehouse System 1.11
1.9 Major Issues in Data Mining 1.12

Chapter 2 Data Preprocessing, Language, Architectures, Concept Description 2.1 - 2.15

2.1 Why Preprocess the Data 2.2
2.2 Data Summarization 2.2
2.3 Data Cleaning 2.4
2.4 Data Integration and Transformation 2.6
2.5 Data Reduction 2.8
2.6 Data Discretization and Hierarchy Generation 2.12

Chapter 3 Data Warehouse and OLAP Technology 3.1 - 3.14

3.1 What is a Data Warehouse? 3.1
3.2 Multidimensional Data Model 3.4
3.3 Data Warehouse Architecture 3.7
3.4 Data Warehouse Implementation 3.10
3.5 From Data Warehouse to Data Mining 3.11

Chapter 4 Data Cube Computation and Data Generalization 4.1 - 4.8

4.1 Methods for Data Cube Computation 4.1
4.2 Data Cube and OLAP Technology 4.5
4.3 Attribute - Oriented Induction - An Alternative Method for Data Generalization and Concept Description. 4.6

ISBN: 9788183713736

Chapter 5 Mining Frequent Patterns, Associations and Correlations 5.1 - 5.14

5.1 Basic Concepts 5.1
5.2 Efficient and Scalable Frequent Itemset Mining Methods 5.5
5.3 Mining of Various Kinds of Association Rules 5.11
5.4 From Association Mining to Correlation Analysis 5.12
5.5 Constraint Based Association Mining 5.13

Chapter 6 Classification and Prediction 6.1 - 6.31

6.1 Classification and Prediction Issues 6.2
6.2 Classification by Decision Tree Induction 6.4
6.3 Bayesian Classification 6.9
6.4 Rule Based Classification 6.12
6.5 Classification by Back Propagation 6.13
6.6 Support Vector Machines 6.18
6.7 Classification by Association Rule Analysis 6.19
6.8 Lazy Learners 6.19
6.9 Other Classification Methods 6.20
6.10 Prediction 6.22
6.11 Accuracy and Error Measures 6.24
6.12 Evaluation of Accuracy 6.24
6.13 Increasing the Accuracy - Ensemble Methods 6.25
6.14 Model Selection 6.26

Chapter 7 Cluster Analysis 7.1 - 7.16

7.1 What is Cluster Analysis? 7.1
7.2 Types of Data in Cluster Analysis 7.3
7.3 Major Clustering Methods 7.5
7.4 Partitioning Methods 7.5
7.5 Hierarchical Methods 7.6
7.6 Density Based Methods 7.7
7.7 Grid Based Methods 7.8
7.8 Model Based Methods 7.9
7.9 Clustering High Dimensional Data 7.12
7.10 Constraint Based Clustering Analysis 7.13
7.11 Outlier Analysis 7.14

ISBN: 9788183713736

Chapter 8 Mining Stream, Time Series, and Sequence Data 8.1 - 8.12

8.1 Mining Data Streams 8.2
8.2 Mining Time Series Data 8.6
8.3 Mining Sequence Patterns in Transactional Data Base 8.8
8.4 Mining Sequence Patterns in Biological Data 8.9

Chapter 9 Graph Mining, Social Network Analysis and Multidirectional Data Mining 9.1 - 9.8

9.1 Graph Mining 9.1
9.2 Social Network Analysis 9.4

Chapter 10 Mining Object, Spatial, Multimedia, Text, and Web Data 10.1 - 10.12

10.1 Multidimensional Analysis and Descriptive Mining of Complex Data Objects 10.1
10.2 Spatial Data Mining 10.3
10.3 Multimedia Data Mining 10.5
10.4 Text Mining 10.7
10.5 Mining the World Wide Web 10.9

Chapter 11 Applications and Trends in Data Mining 11.1 - 11.11

11.1 Data Mining Applications 11.1
11.2 Data Mining System Products and Research Prototypes 11.5
11.3 Additional Themes on Data Mining 11.7
11.4 Social Impacts of Data Mining 11.8
11.5 Trends in Data Mining 11.10

Appendix - A A.1 - A.9
Appendix - B B.1 - B.1
Appendix - C C.1 - C.3
Model Questions MQ.1 - MQ.7
Glossary G.1 - G.3
Index I.1 - I.3