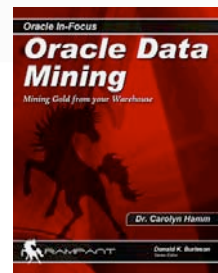




[ REVIEWED FOR YOU ]

**mayato book recommendations in the area of business intelligence and customer relationship management**

Title of book: **Oracle Data Mining: Mining Gold from your Warehouse**  
 Authors: **Dr. Carolyn K. Hamm**  
 Details: **in English / 260 pages / Rampant Techpress September 2007 / Price \$40**  
 Reviewed by: **Kannan Gopal**



**1 Recommendation**

Hamm's book is an easy-to-understand user guide simplifying initial use of the Oracle toolbox for data mining. Step by step, the reader is introduced to a multitude of detailed features. Many screenshots and practical examples make it the ideal companion for Oracle users who want to gain hands-on experience with the data mining tools in their database. The book, however, hardly ever looks beneath the hood of the powerful engines, nor does it explain major concepts of the Oracle Data Mining (ODM) architecture. It, therefore, always remains close to Oracle's online help and the available extension material. Some topics are not or barely covered: Predictive analytics – automated framework for data mining processes - is only briefly and superficially described. The Java Data Mining (JDM) compliant Java API is not mentioned in the book at all.

**2 Summary of content**

The author starts off with a short, one paragraph definition of what data mining is and then launches into detail about using the algorithms in the Oracle Data Miner (ODM). The book has seven chapters. Algorithms mainly for classification models available in the Oracle Data Miner are covered in chapters one to four. Chapter five provides details of tools available in ODM for preprocessing data. The next chapter, Predictive Analytics, covers an automated – and from the user perspective much simplified – approach to data mining also known as self-acting data mining. The final chapter offers a demo example of how to integrate data mining models into Oracle BI publisher, allowing the personalization, for example, of written communication to customers. There are two appendixes with information about installing Oracle Data Miner and script to create the ODM user.

### 3 Assessment criteria

**Content:** The author chose a very generic title “Oracle Data Mining: Mining Gold from your Warehouse,” which leaves a lot of room for expectations about content and topics. Readers looking for a description of tool features supporting their initial work with ODM will find their expectations fully met - if they already have a basic understanding of data mining and what it can be used for. The author provides a step-by-step introduction to model building with many screenshots and practical examples of how to build analytical models using ODM. The different steps in creating a data mining model and the use of a large number of detailed features are described based on popular sample data available on the Internet. About half of the book is dedicated to building classification models. Starting off with how to build a model in general, the author then covers four types of classification algorithm (naive Bayes, adaptive Bayes, decision tree, support vector machines) and two clustering algorithms (k means, O-Clustering). The concepts behind these algorithms are not explained and only a few hints are given about when to use which one. Hamm’s major recommendation is to use several algorithms and compare the results – an approach not unusual in data mining irrespective of the software used, but definitely not smart. Data mining is a highly labor-intensive process with much time spent on data preparation and parameter settings. Doing this for two or more algorithms just to throw away all but one result in the end is a waste of time and money. A better and systematic understanding of the limitations of each algorithm could potentially avoid considerable work and cost (and frustration). While building the above models, the author also explains the supporting features of ODM, such as attribute importance, publishing results to the Oracle database, and how to import a model into a different Oracle database. Chapter five gives an insight into the tools available in ODM for preprocessing data, without emphasizing sufficiently the impact that incomplete or incorrect preprocessing may have on a mining algorithm and the results of an analysis.

The chapter on predictive analytics will probably confuse many readers, as will Oracle’s choice of the rather vague term “predictive analytics” for something now better coined “self-acting data mining” (see the corresponding mayato white paper by Peter Neckel). Hamm does not explain clearly enough what is special about Oracle’s predictive analytics compared to the algorithms previously introduced. In addition, here, the reader would like to understand when to use predictive analytics and when it is better to use one of the other algorithms. The demo example on how to generate personalized letters with Oracle BI publisher gives a nice impression of how Oracle’s built-in data mining functionality can be used in deploying business intelligence applications. Just one example, however, is definitely not enough to inspire readers to invent their own scenarios. Furthermore, the tools, APIs, and so on available for application integration deserve at least a mention. Oracle is, for example, a driving force behind the Java Data Mining standard. Hamm has excellent personal contacts with the Oracle organization, so an outlook to Oracle’s 11g release issued in the meantime could have been expected. Many advanced data mining features that form part of this release do not appear in the book.

**Readability:** Despite the expert topic, the language used is easy to understand. The book gains a lot of volume (about one third of the overall 260 pages) from its many and relatively large black-and-white screenshots of the ODM tool. The reader is carried from feature to feature at user interface level and hardly ever has to stop in order to understand deeper underlying aspects. It is quite possible, therefore, to easily read the book in a day – unless the reader plays around with the software at the same time.

**Practical use:** The book is a very helpful user guide for data analysts taking their first steps in using the ODM tools. However, it remains at tool level and hardly provides a complete and conceptual overview of Oracle technology components and relevant business scenarios. Although it presents many features in detail, readers will find it extremely difficult to gain a more than superficial understanding of ODM, which would be necessary to fully exploit its potential. The information provides at least incremental value to a data mining expert, but hardly any to a data mining layman. The book examines only some details of classification and clustering algorithms. Association analysis and other useful algorithms provided in ODM are not covered at all.

In conclusion, Hamm’s book does not give much added value compared to the excellent material Oracle already provides on the Internet (<http://www.oracle.com/technology/products/bi/odm/index.html>). The value for many readers could have been increased by giving more space to explaining Oracle’s predictive analytics, because this approach does not require much user interaction and modeling work anyway. In the context of business applications, this self-acting data mining will gain ground in the coming years due to its cost efficiency and ease of use.