



## [ REVIEWED FOR YOU ]

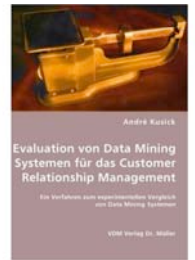
### mayato book recommendations in the area of business intelligence and customer relationship management

Title of book: **Evaluation von Data Mining Systemen für das Customer Relationship Management**  
[Evaluation of Data Mining Systems for Customer Relationship Management]

Author(s): **André Kusick**

Details: **in German / 124 pages / VDM Verlag Dr. Müller 2007 / Price €49**

Reviewed by: **Peter Neckel**



## 1 Recommendation

This book deals with the systematic comparison of data mining software and provides both theoretical considerations and impulses for practical implementation. However, the results of the comparison itself are of limited validity, because the study was carried out using software that was current in 2002 - which one wouldn't suspect of a book published in 2007. The market for analysis software and its application has developed apace in the past few years, making the results (as well as the summary and description of the methods used in the individual tools) incomplete from today's perspective. Nevertheless, the theoretical derivation and development of the comparison method remains useful, because it can serve as the basis for comparing other tools. Whether these insights are worth the princely sum of €49 is a decision that potential readers must make for themselves.

## 2 Summary of content

After a short introduction, the theoretical fundamentals of customer relationship management (CRM), data mining, and methods for evaluating software systems are presented (15 pages). Chapter 4 outlines how the study was designed (4 pages), before the main section of the book begins in chapter 5 with the implementation of the study itself (50 pages). The book closes with a conclusion and outlook (chapter 6, 2 pages), plus an extensive appendix, which includes, in table form, a summary of the methods used in the tools examined, and the test data records (38 pages).

### 3 Assessment criteria

**Content:** In his book, which is based on the author's final dissertation at the University of Koblenz-Landau in Germany, André Kusick has taken on the difficult task of comparing data mining software: Evaluating comprehensive software tools is generally a complex matter because, on the one hand, a multitude of criteria need to be considered and, on the other hand, the results depend to a great extent on each individual use case. In the special case of data mining tools, there is another serious problem: Absolute measures for the quality of a data mining analysis simply do not exist. Answering the question: "When is one customer segmentation better than another?" is not the only hard task. In practice, quality is multidimensional and is not just expressed in better results (difficult enough as this one dimension already is) but also in - for example - usability, computing time, system behavior in the case of very large quantities of data, the time required and the support provided for data preparation, the appropriate representation of the results, the options for exporting data, and the seamless integration of the insights gained into transactional systems. The author sets himself very high standards right from the start: He strives to achieve an "objective comparison" and for this purpose selects an experimental method that he enhanced - and which is, as he himself claims, "more objective than empirical methods." First, he sets the parameters for the object of the investigation. To this end, the three relations *industry - CRM requirement*, *CRM requirement - data mining task*, and *data mining task - data mining method* are contrasted in cross-classified tables, from which the data mining methods to be compared are then derived. The study centers on the telecommunications industry. The data that is used comprises synthetically generated test data records, publicly available test data (for example, from data mining competitions), and "real" data from the business world. The author compares the data mining tools SAS Enterprise Miner 4.1, IBM Intelligent Miner for Data 6.1, and SPSS Clementine 7.0 using seven types of method: association analysis, sequence analysis, cluster analysis, SOM, decision trees, neural networks, and logical regression. The tools are evaluated using an index that incorporates the quality of the results and the execution time of each algorithm in question. Ultimately, there is no clear winner: Each of the three tools takes the number one position for quality or execution time depending on the issue at hand.

**Readability:** Much of the book is easy to read - short and easily understandable sentences, a clear structure, and neat typography all contribute to this. However, in some places - especially the presentation of the analysis results - I would have liked to have had more detailed explanations. Furthermore, screenshots of the tools used were too few. The introductory section with its predominantly theoretical musings seems unbalanced in places, presumably because the original dissertation was too drastically shortened here.

**Practical use:** All in all, the book is of very little practical use. The fundamental structure of the comparison is logical and capable of development, but is at times not thoroughly thought-out. The evaluation restricts itself too narrowly to the analysis itself and to supposedly objective and easily measurable criteria such as the algorithm execution time or the accuracy of the results. Important aspects are thus neglected that have a considerable influence on practical deployability. Such aspects include - above all - the large area of data preparation as well as the integration of the results into transactional systems, and ultimately into business processes. What's more, the results of the tool comparison are now out-of-date: SAS Enterprise Miner 4.1 was launched back in 2001; SPSS Clementine 7.0 has been on version 12 since January 2008, and IBM retired the Intelligent Miner at version 8.1 in March 2006. Consequently, a large chunk of the appendix - for example, the summary and the description of the methods - also suffers from a lack of topicality. A current comparison would have to accommodate numerous new developments: Many new makers of data mining tools have appeared on the market, including some open source providers to be taken seriously. Fundamental data mining functions are now integrated into numerous BI suites (for example, Oracle Data Mining, Microsoft SQL Server Analysis Services, and SAP NetWeaver BI Data Mining). New types of method (such as support vector machines) or entirely new analysis approaches (such as self-acting data mining) would have had a great impact on the results. Nevertheless, the test data used deserves a positive mention: It is wide-ranging and comes from different sources, thus almost completely ruling out the possibility of a one-sided preference for one of the tools. However, the number of data records remains at a scale that does not enable conclusions to be drawn about system behavior for data quantities in the high gigabytes or even in terabyte territory (quite common these days) - and this is an important factor, especially in the telecommunications industry.