



## [ FÜR SIE GELESEN ]

### mayato Buchempfehlungen im Umfeld Business Intelligence und Customer Relationship Management

Buchtitel: **Evaluation von Data Mining Systemen für das Customer Relationship Management**

Autor(en): **André Kusick**

Details: **Sprache Deutsch / 124 Seiten / VDM Verlag Dr. Müller 2007 / Preis 49 €**

Rezensent: **Peter Neckel**



## 1 Empfehlung

Das Buch hat den systematischen Vergleich von Data-Mining-Software zum Inhalt und liefert hierzu einige theoretische Überlegungen sowie konkrete Ansatzpunkte für die praktische Durchführung. Die Ergebnisse des Vergleichs an sich sind allerdings in ihrer Aussagekraft beschränkt, da die Untersuchung mit Software auf dem Versionsstand des Jahres 2002 durchgeführt wurde, was das Erscheinungsjahr des Buches (2007) nicht vermuten lässt. Der Markt für Analysesoftware und dessen Anwendung hat sich jedoch in den letzten Jahren zügig weiterentwickelt, so dass die Ergebnisse (und ebenso die Aufstellung und Beschreibung der Verfahren der einzelnen Tools) aus heutiger Sicht unvollständig sind. Die theoretische Herleitung und Entwicklung des Vergleichsverfahrens ist jedoch nach wie vor nützlich, da sie auch für andere Toolvergleiche als Grundlage dienen kann. Ob diese Erkenntnisse den stolzen Preis von 49 Euro wert sind, muss der potenzielle Leser individuell entscheiden.

## 2 Inhaltzusammenfassung

Nach einer kurzen Einleitung werden knapp die theoretischen Grundlagen des Customer Relationship Managements (CRM), des Data Mining, sowie zur Methodik der Evaluation von Softwaresystemen dargestellt (15 Seiten). In Kapitel 4 folgt eine kurze Darstellung der Konzeption der Studie (4 Seiten), bevor in Kapitel 5 der Hauptteil des Buches mit der Durchführung der Studie (50 Seiten) beginnt. Fazit und Ausblick (Kapitel 6, 2 Seiten) sowie ein recht umfangreicher Anhang mit u.a. einer tabellarischen Aufstellung der Verfahren der verwendeten Werkzeuge sowie der Testdatensätze (38 Seiten) schließen das Buch ab.

### 3 Bewertungskriterien

**Inhalt:** Im auf seiner Diplomarbeit an der Universität Koblenz-Landau basierenden Buch hat sich der Autor mit dem Vergleich von Data-Mining-Software ein schwieriges Thema vorgenommen: Evaluationen von umfangreichen Softwarewerkzeugen sind generell komplex, weil zum einen eine Vielzahl von Kriterien zu berücksichtigen sind, und zum anderen die Ergebnisse stark vom jeweiligen Anwendungsfall abhängen. Im speziellen Fall der Data-Mining-Werkzeuge kommt noch ein gravierendes Problem hinzu: Absolute Maßzahlen für die Qualität einer Data-Mining-Analyse gibt es nicht. Nicht nur lässt sich die Frage „Wann ist eine Kundensegmentierung besser als eine andere?“ schwer beantworten; die Qualität ist in der Praxis auch mehrdimensional und drückt sich nicht nur in besseren Ergebnissen aus (so schwer messbar schon diese eine Dimension ist), sondern z.B. auch im Bedienkomfort, in der Rechenzeit, im Systemverhalten bei sehr großen Datenmengen, im Zeitbedarf und der angebotenen Unterstützung für die Datenvorverarbeitung, in der angemessenen Repräsentation der Ergebnisse bis zu Exportmöglichkeiten und der nahtlosen Integration der gewonnenen Erkenntnisse in operative Systeme.

Der Autor legt sich selbst gleich zu Anfang die Messlatte hoch: Er strebt einen „objektiven Vergleich“ an und wählt dazu die von ihm weiterentwickelte experimentelle Methode – nach eigener Aussage „objektiver als empirische Verfahren“. Zunächst wird der Untersuchungsgegenstand eingegrenzt; hierzu werden die drei Relationen Branche – CRM-Fragestellung, CRM-Fragestellung – Data Mining Aufgabe, Data Mining Aufgabe – Data Mining Methode in Kreuztabellen gegenübergestellt, anhand derer die zu vergleichenden Data-Mining-Verfahren abgeleitet werden. Dazu werden Fragestellungen aus der Telekommunikationsbranche gewählt. Als Daten kommen synthetisch generierte Testdatensätze, öffentlich zugängliche Testdaten (z.B. aus Data-Mining-Wettbewerben) sowie „echte“ Daten aus der Unternehmenspraxis zum Einsatz. Der Autor vergleicht die Data-Mining-Werkzeuge „SAS Enterprise Miner 4.1“, „IBM Intelligent Miner for Data 6.1“ sowie „SPSS Clementine 7.0“ anhand von sieben Verfahrensklassen: Assoziationsanalyse, Sequenzanalyse, Clusteranalyse, SOM, Entscheidungsbäume, Neuronale Netze, Log. Regression. Die Bewertung der Werkzeuge erfolgt anhand eines Indexes, in den die Qualität der Ergebnisse sowie die Ausführungszeit des jeweiligen Algorithmus einfließt. Ein eindeutiges Endergebnis gibt es nicht – jedes der drei Werkzeuge liegt bei unterschiedlichen Fragestellungen bei der Qualität oder der Ausführungszeit vorne.

**Lesbarkeit:** Das Buch ist über weite Strecken leicht lesbar – kurze, verständliche Sätze, klare Gliederung und eine übersichtliche Typografie tragen dazu bei. An manchen Stellen, insbesondere bei der Darstellung der Analyseergebnisse, würde man sich allerdings ausführlichere Erläuterungen wünschen. Auch Screenshots der verwendeten Tools werden zu spärlich gezeigt. Der Anfangsteil mit vorwiegend theoretischen Überlegungen wirkt in Teilen unausgewogen; vermutlich wurde hier der Diplomarbeitstext zu stark gekürzt.

**Praxistauglichkeit:** Insgesamt betrachtet ist die Praxistauglichkeit des Buches stark eingeschränkt. Der grundlegende Aufbau des Vergleichsszenarios ist schlüssig und ausbaufähig, allerdings an manchen Stellen nicht ganz durchdacht. Die Evaluation beschränkt sich zu stark nur auf die Analyse selbst und vermeintlich objektive und gut messbare Kriterien wie Laufzeit des Algorithmus oder Genauigkeit der Ergebnisse. Dabei kommen wichtige Aspekte zu kurz, die erheblichen Einfluss auf die Einsetzbarkeit in der Praxis haben. Dazu gehören vor allem der große Bereich der Datenvorverarbeitung sowie die Integration der Ergebnisse in operative Systeme und damit in die Geschäftsprozesse. Zudem sind die Ergebnisse des Toolvergleichs mittlerweile veraltet: Der verwendete SAS Enterprise Miner 4.1 kam bereits im Jahr 2001 auf den Markt; SPSS Clementine 7.0 trägt seit Januar 2008 die Versionsnummer 12, und den Intelligent Miner hat die IBM in der letzten Version 8.1 bereits im März 2006 vom Markt genommen. Auch große Teile des Anhangs, z.B. die Aufstellung und Erläuterung der Verfahren, leiden somit an mangelnder Aktualität. Ein aktueller Vergleich müsste zahlreichen neuen Entwicklungen Rechnung tragen: Eine Vielzahl neuer Hersteller von Data-Mining-Werkzeugen haben den Markt betreten, darunter auch einige Ernst zu nehmende Open-Source Anbieter. Grundlegende Data-Mining-Funktionen finden sich ebenso in zahlreichen BI-Suiten (z.B. Oracle Data Mining, MS-SQL-Server Analysis Services, SAP Netweaver BI Data Mining). Auch neue Verfahrensklassen (etwa Support Vector Machines) oder gänzlich neue Analyseansätze (z.B. Self-Acting Data Mining) hätten starken Einfluss auf die Ergebnisse. Positiv sind die verwendeten Testdaten zu erwähnen: Es kommen Daten aus einem breiten Spektrum und aus unterschiedlichen Quellen zum Einsatz, was eine einseitige Bevorzugung eines Werkzeugs nahezu ausschließt. Die Anzahl der Datensätze bewegt sich allerdings in Größenordnungen, die keine Rückschlüsse auf das Systemverhalten bei Datenmengen im (mittlerweile durchaus üblichen) hohen Gigabyte- oder gar Terabytebereich zulassen, was gerade in der Telekommunikationsbranche als ein wichtiger Faktor anzusehen ist.