

Markante Unterschiede im Funktionsspektrum von Data-Mining-Werkzeugen



Berlin, 16. Dezember 2009 — Das BI-Analysten- und Beratungshaus mayato untersucht in seiner diesjährigen Studie „Data Mining 2010“ 12 Data-Mining-Suiten und -Tools. Fünf Lösungen wurden umfassend unter die Lupe genommen. Besonderes Augenmerk legten die Analysten auf die jeweilige Umsetzung von immer häufiger genutzten Features wie Cross- und Upselling-Untersuchungen. Neben dem umfangreichen Praxistest wurde deshalb für jedes der fünf Werkzeuge eine detaillierte Funktionsübersicht mit dem Fokus auf Assoziations- und Sequenzanalysen erstellt. Resümee: Es gibt deutliche Unterschiede im Funktionsumfang. Zudem ist der Trend erkennbar, neben den Statistikexperten auch Fachabteilungsnutzer mit entsprechenden Bedienkonzepten einzubeziehen. Die Studie ist ab sofort zu einem Preis von 299 Euro für Unternehmen und 99 Euro für Studenten und Hochschulen unter www.mayato.com erhältlich.

Zu den untersuchten DM-Suiten gehören der SAS Enterprise Miner 6.1 und der StatSoft STATISTICA Data Miner 9. Als Open-Source-Anwendung wurde KNIME 2.0.3 (Uni Konstanz) einbezogen. Zudem kamen mit SAP BW 7.0 ein Business-Intelligence-Werkzeug und mit dem KXEN Analytic Framework 5.1.1 eine Lösung für Self-Acting Data Mining auf den Prüfstand. Im Praxistest wurden die Werkzeuge anhand eines großen Testdatensatzes von 1,8 Millionen Zeilen untersucht. Ausgehend von einer Fallstudie spielten die Analysten den gesamten Data-Mining-Prozess durch, einschließlich der Datenvorverarbeitung sowie Darstellung und Interpretation der Ergebnisse. Die Bewertung bezog sich auf Bedienbarkeit, Stabilität, Systemverhalten bei großen Datenmengen, Dokumentation und die Gesamteffizienz des Analyseprozesses. *„Uns hat überrascht, dass die Funktionsunterschiede der untersuchten, wenn auch nicht direkt vergleichbaren Werkzeuge doch recht groß sind. Manche Aufgaben wie die Sequenzanalyse lassen sich mit einigen Werkzeugen gar nicht erledigen, ohne dass dies aus der Produktbeschreibung ersichtlich ist“*, sagt Peter Neckel, Analyst und Leiter der Studie beim BI-Analysten- und Beratungshaus mayato.

Werkzeuge im Praxisvergleich

SAS Enterprise Miner und STATISTICA Data Miner zählten zu den Werkzeugen mit dem breitesten Funktionsumfang, hoher Ausführungsgeschwindigkeit und sicherem Umgang mit großen Datenmengen. Beide Tools überzeugten vor allem durch ausgereifte Funktionalität und die grafische Aufbereitung. Während das KXEN Analytic Framework mit der einsteigerfreundlichen Bedienung und seinem unerreicht schnellen, selbstentwickelten Assoziationsverfahren punktet, zeigte das Tool Schwächen in der grafischen Ergebnisauswertung.

Deutliche Unterschiede verzeichneten die BI-Experten beim Funktionsumfang der Werkzeuge. Das SAP BW und KNIME stellen beispielsweise keine Sequenzanalyse zur Verfügung, sodass etwa der zeitliche Abstand zwischen Kauftransaktionen nicht ausgewertet werden kann. Viele typische analytische Anwendungsszenarien im Umfeld des Cross- und Upselling, aber auch im Risikomanagement oder in der Betrugserkennung bei Banken und Versicherungen lassen sich dementsprechend mit diesen Tools nicht umsetzen. Bei der Ergebnisauswertung und -interpretation – gerade im Cross-/Upselling-Umfeld wichtig – ergaben sich im Praxistest ebenfalls Schwierigkeiten: Insbesondere bei großen Datenmengen ist die Unterstützung für den Anwender in manchen Fällen nicht ausreichend. KNIME liegt diesbezüglich im Test hinten, ebenso wie bei der Laufzeit. *„Bei der Datenvorbereitung lässt sich feststellen, dass die kostenlosen Open-Source-Tools einen sehr viel höheren Zeitaufwand erfordern“*, so Neckel.

Anbieter fahren unterschiedlichste Ansätze

Nach wie vor gilt, dass der Anbietermarkt nicht nur breit gefächert ist, sondern auch eine große „Artenvielfalt“ an Lösungen umfasst. Für Anwender gestaltet sich die Auswahl daher relativ komplex. Verbundkaufanalysen à la Amazon („Kunden, die dieses Produkt angesehen haben, kauften auch:...“) sind mittlerweile nicht mehr nur im Versandhandel sondern auch bei Finanzdienstleistern und Versicherungen gängige Praxis für zielgerichtete Direktmarketingkampagnen und -aktionen.



Aber auch im Risikomanagement findet die erweiterte Variante (Sequenzanalyse) rege Anwendung. Umsetzung und Bedienbarkeit sehen allerdings bei allen Herstellern recht unterschiedlich aus. Im Praxistest fanden die Analysten heraus, dass sich die Bedienung dank grafischer Benutzeroberflächen grundsätzlich vereinfacht hat. Ein dominierendes Bedienkonzept lässt sich jedoch – wie schon in den letzten Jahren – nicht erkennen. Einige Hersteller bieten mittlerweile sogar mehrere Modi für unterschiedliche Anwendertypen an. Während die großen Suiten nach wie vor Spezialisten-Know-how verlangen, erleichtern Produkte wie die von StatSoft und KXEN auch Gelegenheitsanwendern den Einstieg in die explorative Datenanalyse, indem auf Wunsch über Assistenten eine feste Abfolge an Analyseschritten vorgegeben und die erforderlichen Eingaben dazu systematisch abgefragt werden.

Der Investitionsaufwand für Data-Mining-Suiten liegt abhängig von der Nutzeranzahl meist im sechsstelligen Euro-Bereich. Dennoch gibt es auch günstigere Angebote. Die Einzelplatzlizenz des STATISTICA Data Miner ist bei vollem Funktionsumfang beispielsweise bereits ab 20.000 Euro erhältlich. Für Open-Source-Lösungen entfällt der Anschaffungspreis; bei professioneller Nutzung sollten jährliche Supportgebühren im vierstelligen Eurobereich einkalkuliert werden.

Über mayato®:

Das BI-Analysten- und Beratungshaus mayato wurde im April 2007 mit der Vision der „perfekten Entscheidung“ gegründet. mayato deckt in drei Bereichen das Thema Business Intelligence komplett ab: „Think“ untersucht aktuelle Trends, Märkte, Methoden, Produkte und Technologien. In „Act“ werden integrierte und flexible Infrastrukturen für Business Analytics konzipiert und realisiert. Darüber hinaus unterstützen im Bereich „Analyze“ Statistik- und Data-Mining-Experten Konzerne bei der Analyse der Effizienz von Geschäftsprozessen mit Hilfe von innovativen Metriken. Als Partner mehrerer Softwareanbieter ist mayato grundsätzlich der Neutralität und in erster Linie der Qualität seiner eigenen Dienstleistungen verpflichtet. mayato berät vom Hauptsitz in Berlin sowie weiteren Standorten nationale und internationale Kunden vom Mittelständler bis zum Weltkonzern. Nähere Infos unter www.mayato.com.

Pressekontakt:

mayato® GmbH
Am Borsigturm 9
D-13507 Berlin

Georg Heeren
Tel. +49 171 481.8877
georg.heeren@mayato.com

Stemmermann – Text & PR
Hinter dem Rathaus 1
D-23966 Wismar

Tanja Stemmermann
Tel. +49 3841 22 43 14
info@stemmermann-pr.de