



SAS UND IBM SPSS EIN VERGLEICH

Welche Analyzelösung eignet sich besser?

SAS und IBM sind zwei marktführende Anbieter im Bereich Analytics-Software. Traditionell mit unterschiedlichen Anwendungsfeldern nähern sie sich in ihren Funktionalitäten immer mehr an. Manche Konzerne haben sogar beide Lösungen im Einsatz. Lohnt es sich, zu konsolidieren?

Industrie 4.0, Digitalisierung und das Internet of Things beschäftigen insbesondere Fertigungsunternehmen, Maschinenbauer und Automobilhersteller. Smart Factorys, in denen alles vernetzt ist, Maschine mit Maschine kommuniziert, Prozesse zentral gesteuert werden oder sich selbst steuern, sind in Zukunft die Regel. Und während Konzerne heute Kennzahlen einzelner Geschäftsfelder isoliert betrachten, da Abhängigkeiten nicht offensichtlich sind, lassen sich künftig beispielsweise Daten eines konzerninternen Finanzinstituts für die Produktionsplanung zurate ziehen. Wenn die Bank verstärktes Interesse an der Finanzierung bestimmter Automodelle registriert, könnte sie diese Daten an die Produktion übergeben, die daraufhin

optimiert würde. Idealerweise könnte die Fertigung sogar auf dieselben Daten- und Analysensysteme zugreifen, um auf Basis relevanter Informationen aus Gegenwart und Vergangenheit für die Zukunft zu planen. Stichwort: Predictive Analytics.

Zwei derzeit führende Anbieter derartiger Lösungen sind SAS und IBM. Erstere stellen Kunden mit der SAS Intelligence Platform umfangreiche Predictive-Analytics-Lösungen zur Verfügung. IBM hat mit dem Kauf von SPSS das eigene Portfolio um leistungsfähige Analyse-Software ergänzt.

Bereits jetzt haben viele Unternehmen eine der Lösungen im Einsatz. SAS dominiert dabei traditionell im Finanzbereich, während SPSS verstärkt von den anderen Abteilungen bzw. Unternehmenszweigen genutzt wird. Manche Konzerne verwenden Software von beiden Anbietern in unterschiedlichen Bereichen. Daher drängt sich die Frage auf: Welches System kann mögliche Anforderungen am umfassendsten erfüllen?

Vergleich SAS und SPSS

Im Folgenden werden SAS Intelligence Platform 9.3 (kurz: SAS) und IBM SPSS Statistics 23, IBM SPSS Modeler 17, IBM SPSS Collaboration and Deployment Services (CADS) 7 (Kurz SPSS) anhand unterschiedlicher Parameter verglichen: Datenquellen und -haltung, Datenanalyse, Kosten und Integration von R. Darüber hinaus soll eine beispielhafte „Querintegration“ Klarheit bringen. Können typische Fragestellungen aus der Produktion in SAS abgebildet werden? Lassen sich charakteristische Bank-Fragestellungen mithilfe von SPSS analysieren?

Daten: Quellen und Haltung

Datenquellen

Sowohl SAS als auch SPSS unterstützen Daten aus Datenbanken. Die Ausnahme bildet PostgreSQL, das von SPSS nicht unterstützt wird. Auch ein Import aus Excel, XML, Hadoop, Geodaten sowie Text- bzw. CSV- oder SAV- und SAS-Dateien ist bei beiden Lösungsanbietern möglich. Zudem lassen sich Simulationen generieren und Daten manuell eingeben

Hinsichtlich der Datenquellen sind sich beide Lösungen sehr ähnlich. Bei der Datenhaltung gibt es allerdings Unterschiede.

Ausschließlich mit SAS können Anwender Daten aus SAP Business-Warehouse importieren. Im Gegenzug dazu ist ein Import aus IBM Cognos BI, IBM Cognos TM1 oder IBM Data Collection nur mit SPSS möglich. DA

Daten-Quelle	SAS	SPSS
Datenbanken	Ja	Ja*
Text / CSV-Dateien (Trennzeichengetrennt)	Ja	Ja
Textdateien mit fester Breite	Ja	Ja
Excel	Ja	Ja
XML	Ja	Ja
SAV-Datei (IBM SPSS Statistics)	Ja	Ja
IBM Cognos BI	Nein	Ja
IBM Cognos TM1	Nein	Ja
IBM Data Collection	Nein	Ja
SAS-Dateien	Ja	Ja
Hadoop etc...	SAS/ACCESS	Über Analytic Server
Geodaten (shp und dbf Dateien)	Bei SAS GIS bereits enthalten	Ja
SAP BusinessWarehouse	Ja	Nein
Simulationsgenerierung (generieren der Zufallszahlen entsprechend einer vorgegebenen Verteilung)	Ja	Ja
Manuelle Eingabe	Ja	Ja

*Außer PostgreSQL

Datenhaltung

SAS enthält eine eigene, interne Datenbank, die auf die Anforderungen der Informationsgewinnung ausgelegt ist.ⁱⁱ Dies ist bei SPSS nicht der Fall. Vielmehr ist der IBM SPSS Modeler dank Pushback-Funktionalität darauf ausgerichtet, Data-Mining und Datenvorbereitung direkt in der Datenbank vorzunehmen.ⁱⁱⁱ

Datenanalyse

Datenanalyse	SAS	SPSS
Univariate, deskriptive Analyse	Ja	Ja
Statistische Tests mit ein, zwei oder mehr unabhängigen bzw. abhängigen Variablen in unterschiedlichen Kombinationen	Ja	Ja
Lebensdauer- und Ereignisanalyse (Sterbetafel-Methode, Kaplan-Meier-Schätzer der Überlebenswahrscheinlichkeit, Cox-Regression)	Ja	Ja

Datenanalyse	SAS	SPSS
Screening von Feldern	Variable Selection (Knoten in SAS Enterprise Miner)	Merkmalauswahl (Knoten in IBM SPSS Modeller)
Entscheidungsbäume	Decision Tree (Knoten in SAS Enterprise Miner) Bietet ähnliche oder sogar erweiterte Funktionalitäten wie SPSS, erfordert aber mehr Kenntnis beim Anwender.	C5.0, Classification & Regression Tree, Chi Squared Automatic Interaction Detector, Quick unbiased efficient statistical tree (Knoten in IBM SPSS Modeller)
Cluster-Modelle	Cluster, SOM/Kohonen (Knoten in SAS Enterprise Miner)	K-Means, Kohonen, TwoStep, TwoStep-AS, Anomalie (Knoten in IBM SPSS Modeller)
Zeitreihenmodelle (einfach, lineare Trends nach Holt bzw. Brown, gedämpfter Trend, einfach saisonal, additives bzw. multiplikatives Winters-Modell)	Ja. Zusätzlich TS Similarity Analyse in SAS Enterprise Modeller zum Vergleich zweier Sequenzen	Ja. Zusätzlich Temporale kausale Modellierung sowie räumlich-zeitliche Vorhersage (spatio-temporal prediction)
Assoziationsregeln	Ja	Ja
Neuronale Netze	Neural Network, DM Neural und Auto Neural (Knoten in SAS Enterprise Modeller).	Neuronales Netz (Knoten in IBM SPSS Modeller)
Support Vector Machine	Ja	Ja
Nächste-Nachbarn-Modelle	Ja	Ja
Entity Analytics	Ja	Ja
Social Network	Nein	Ja
Text Mining	Ja	Ja

Kosten

SAS-Software ist nicht käuflich zu erwerben, sondern lediglich zu mieten, wobei die Firma ihren Kunden individuelle Angebote erstellt.^{iv} In der Regel beträgt die kürzeste Mietdauer ein Jahr. SAS bietet unterschiedliche Möglichkeiten der Implementierung. Viele Module sind bereits in der Cloud verfügbar oder werden dort künftig verfügbar sein.

Im Gegensatz dazu werden IBM SPSS Lizenzen gekauft.

Software	Basis	Premium
IBM SPSS Modeller ^v	ab ca. 4.200 EUR*	ab ca. 10.600 EUR*
IBM SPSS Statistics ^{vi}	ab 1.100 EUR*	ab ca. 7.300 EUR*

*Preise jeweils pro User, inkl. technischer Support für zwölf Monate. Weitere Preisinformationen auf der IBM Homepage.

Berücksichtigen Sie nicht nur die reinen Anschaffungskosten, sondern auch eventuell anfallende Schulungsaufwände.

Zusätzlich zu den reinen Lizenzkosten sind auch (Um-)Schulungsaufwände zu berücksichtigen. SPSS genießt den Ruf, leichter im Selbststudium erlernbar zu sein als SAS, da die Bedienung intuitiver ist. Hier könnten gegebenenfalls Kursgebühren gespart werden.

„SPSS Modeler bietet eine grafische Benutzeroberfläche zur einfachen Modellierung aktionsbasierter komplexer analytischer Auswertungen, die sich problemlos von Anwendern ohne Programmierkenntnisse nutzen lässt“, so Alexander Frank, Technical Sales Specialist/Business Analytics der IBM Software Group. „Dennoch bietet SPSS Modeler ein breites Spektrum an mathematischen Algorithmen, Datenvorbereitungs- und Integrationsmöglichkeiten, welche die Einbindung von Daten aus verschiedenen Systemen, Technologien und Umgebungen ermöglichen“, führt der IT-Experte weiter aus.

Grundsätzlich wird zu beiden Lösungen eine Vielzahl von Schulungen angeboten. Zudem sollten nicht nur Gebühren berücksichtigt werden, sondern auch die Zeit, die für das Einarbeiten in die jeweilige Software benötigt wird.

Querintegration

Traditionell ist SAS die bei Banken vorherrschende Analysesoftware. Laut SAS-eigenen Angaben nutzen 99 der weltweit Top-100-Banken die Software des Unternehmens. Eine typische Aufgabe von Finanzinstituten ist es, die Kreditwürdigkeit eines Kunden anhand unterschiedlicher Kriterien zu berechnen. SAS ermöglicht dies beispielsweise anhand der Entscheidungsbaum-Funktionalität. Auch mit SPSS lässt sich eine derartige Fragestellung beantworten – sogar mit der gleichen Funktionalität. Auch Cluster-Modelle zur Identifizierung von Zielgruppen lassen sich in beiden Lösungen auf die gleiche Weise erstellen.

Eine Anforderung an Analysesysteme seitens der Produktion ist zum Beispiel das Auslesen von Fehlerprotokollen, Händlerfeedback oder Reparaturberichten, um daraus verbesserte Prozesse abzuleiten. Dies lässt sich mit der Data- und Text-Mining-Funktionalität im IBM SPSS Modeler umsetzen.^{vii} Auch SAS bietet entsprechende Möglichkeiten, beispielsweise mit den Enterprise und Text Miners sowie Forecastmodulen.

Sowohl SAS als auch SPSS verfügen über spezielle Branchenlösungen. Dabei ist in der Regel die Anzahl der Modelle eingeschränkt und vorselektiert, was die Komplexität für den Anwender reduziert.

Integration von R

R ist eine Open-Source Analyselösung, die zunehmend an Beliebtheit gewinnt und kommerzielle Software wie SAS und SPSS teilweise im Funktionsumfang übertrifft. Zudem ist R flexibler. Mancherorts wird gefragt, wann sie SAS und SPSS den Rang ablauft. Während R bei Studierenden, Universitäten und Startups beliebt ist, da keine teuren Lizenzen angeschafft werden müssen, hat die Lösung sich bei Unternehmen bisher noch nicht durchsetzen können. Gründe dafür sind unter anderem, dass es keinen Support gibt und im Prinzip jeder ein R-Paket programmieren kann. Ob dieses dann so rigoros getestet wird, wie Funktionalitäten bei SAS bzw. SPSS ist nicht immer sicher.^{viii}

Nichtsdestotrotz reagieren SAS und SPSS auf die steigende Beliebtheit und die erweiterten Funktionalitäten von R. So können Daten zwischen SAS und R transferiert, R-Funktionen innerhalb von SAS aufgerufen und Ergebnisse von R nach SAS übermittelt werden. IBM SPSS Statistics enthält R-basierte Funktionserweiterungen und IBM SPSS Modeler beinhaltet einen R-Knoten, der es erfahrenen R-Nutzern ermöglicht, eigene R-Skripte einzugeben.

Die Open-Source-Statistiksoftware R wird immer wichtiger. Sowohl SAS als auch SPSS bieten Integrationsmöglichkeiten.

Fazit

Sowohl SAS als auch SPSS bieten umfangreiche Analysemöglichkeiten und Funktionalitäten, die unterschiedlichen Branchen und Geschäftsfeldern wertvolle Informationen liefern können. In den letzten Jahren hat sich der Funktionsumfang beider Lösungen zunehmend angeglichen. Insbesondere durch die Übernahme von SPSS seitens IBM hat sich das Leistungsspektrum der Software stetig erhöht. Funktionen, die SAS bietet und die in SPSS selbst nicht verfügbar sind, werden von Lösungen weiterer IBM Tochterfirmen abgedeckt. Dabei werden die einzelnen Produkte so weiterentwickelt, dass sie vollständig untereinander kompatibel sind und sich harmonisch ergänzen. Von IBM getätigte Investitionen in Business Intelligence, Analytics und Optimierung zahlen sich aus, da auch SPSS von der im Mutterkonzern vorhandenen IT-Expertise und -Know-how profitiert.

In Zukunft liegt der Fokus von SPSS voraussichtlich auf industrie- und branchenspezifischen Lösungen, in denen die Analysesoftware ein elementarer Bestandteil ist: Counter Fraud, Predictive Maintenance and Quality Optimization, Predictive Customer Insight und neue Branchenlösungen für beispielsweise Retail, Insurance und Banking.

Aufgrund des sehr ähnlichen Leistungsumfanges der beiden Tools stellen mitt-

lerweile die jeweiligen Anschaffungs-, Nutzungs- und Supportkosten zentrale Kriterien für Unternehmen dar. Daher gilt es, Unternehmensanforderungen klar zu definieren. Die Anzahl der Mitarbeiter, die das Tool nutzen werden, vorhandenes IT-Know-how, bestehende IT-Infrastruktur, Datenquellen und Analysetools müssen betrachtet werden, um eine Empfehlung für die eine oder andere Software zu geben. Erst auf dieser Basis lässt sich prüfen, ob auf eines der beiden Systeme konsolidiert werden kann oder sogar sollte.

Gerne beraten wir Sie individuell, welche Analyzelösung für Sie die richtige ist. Wir unterstützen Sie mit Analyse, Geschäftsprozessberatung, Anforderungsmanagement, Technologieberatung, Fach- und IT-Konzepten sowie IT-Architektur. Auch beim Rollout und im laufenden Betrieb können Sie auf unsere Unterstützung zählen.

www.sulzer.de/highlights/business-analytics/

Über die Autorin

Marina Reznic ist Senior IT-Consultant Business Analytics bei der Sulzer GmbH in München. Der mittelständische Full-Service-Anbieter von Unternehmens- und IT-Beratung wurde 1978 gegründet und zählt namhafte Automobilhersteller zu seinen Kunden. In Deutschland verfügt die Sulzer GmbH über Standorte in München, Ingolstadt und Stuttgart. Darüber hinaus ist das Unternehmen in Paramus (USA) vertreten.

Sie erreichen Frau Reznic unter +49 89 31858 165 oder per E-Mail an

Marina.Reznic@sulzer.de

© Sulzer GmbH, Oktober 2015

- ⁱ Vgl. z.B. Forrester Wave™ Big Data Predictive Analytics Solutions, Q2 2015
- ⁱⁱ Vgl. <http://de.saswiki.org/wiki/SAS-Software>, letzter Zugriff 07.09.2015
- ⁱⁱⁱ Vgl. http://www-01.ibm.com/support/knowledgecenter/SS3RA7_15.0.0/com.ibm.spss.modeler.help/sql_overview.htm?lang=de letzter Zugriff 07.09.2015
- ^{iv} INWT Statistics beziffert die Kosten auf ca. 5.500 EUR, Stand Oktober 2010. Letzter Zugriff 08.09.2015
- ^v <http://www.ibm.com/marketplace/cloud/spss-modeler/de/en-de?step=Plan> letzter Zugriff 08.09.2015
- ^{vi} <http://www.ibm.com/marketplace/cloud/spss-statistics/de/en-de?step=Plan> letzter Zugriff 08.09.2015
- ^{vii} <http://www-03.ibm.com/software/businesscasestudies/no/no/corp?synkey=C832269U02087M63>. letzter Zugriff 09.09.2015
- ^{viii} Vgl. Z.B. die Diskussion bei Quora „How long before R overtakes SAS and SPSS?“ <http://www.quora.com/How-long-before-R-overtakes-SAS-and-SPSS> letzter Zugriff am 07.09.2015 6