



Von Big Data zu Smart Data

Intelligente Daten für Geschäftsentscheidungen

Das Thema Big Data wird bereits seit Jahren in der Fachpresse als Trend bezeichnet. Dabei sind große Datenmengen schon seit geraumer Zeit ein Thema. Allerdings hat sich in der Handhabung von Big Data einiges getan. Der Anspruch, große Datenmengen mithilfe von IT-Unterstützung intelligent und effizient zu nutzen, ist deutlich gestiegen.

Um ein verwertbares Ergebnis zu erhalten, war es früher notwendig, große Analysten-Teams zu beauftragen, um einen Zusammenhang zwischen den verschiedenen Parametern der im Tagesgeschäft gesammelten Daten herzustellen. Mittlerweile übernehmen kleine Expertengruppen mit Unterstützung einer geeigneten Softwarelösung solche Auswertungen. Durch den Einsatz von Big Data werden – neben dem Potenzial in der Umsatzsteigerung – weitere Vorteile in Unternehmensteilen ersichtlich, die mit der IT oder dem Handel im geringen Maße zu tun haben. In der medizinischen Branche wird es beispielsweise möglich, durch die Sammlung von Krankheitsverläufen bei Patientendaten Gemeinsamkeiten zu erkennen. Zusammen mit den persönlichen Gewohnheiten des Patienten, wie sportliche oder berufliche Aktivitäten oder Trink- und Essverhalten, können effektive Therapien entwickelt werden.

Smart Data & Big Data. Big Data definiert sich aus der Gesamtheit aller gesammelten Daten. Ohne vorheriges Data Assessment, also einer detaillierten Datenauswertung und Datenklassifizierung, ist es schwierig, die relevanten Daten und Informationen für die Weiternutzung zu erhalten. Eine willkürliche Datenhaltung führt außerdem zu zusätzlichen Kosten im Betrieb einer Big-Data-Infrastruktur.

Daten werden als wichtiges und kostbares Asset für die Zukunft angesehen. Wichtig ist hier, die Datenmengen vorher zu analysieren und das Sammelverhalten anzupassen, um nur die notwendigen Informationen im Speicher zu halten. Eine wesentliche Voraussetzung bei der Datenauswahl ist die Festlegung von Umfang und Validität der Daten sowie die Abschätzung des unternehmerischen Mehrwerts, der erzielt werden kann. Demnach lässt sich aus dem Roh-

stoff Big Data durch Weiterentwicklung Smart Data generieren. Smart Data setzt sich aus den gesammelten Daten des Datenspeichers verknüpft mit dem unternehmerischen Nutzen, der Datenqualität, Semantik und dem Datenschutz zusammen. Durch Smart Data gewinnen Unternehmen einen intelligenten Datensatz mit Metainformationen, die sicher genutzt werden können.

Vorhersagen mit Smart Data Analytics. Der Begriff Business Intelligence (BI) wird oftmals in Zusammenhang mit Big Data oder Smart Data genannt und umfasst in der Regel die Planung und das Controlling des laufenden Geschäfts anhand von Kennzahlen. Zusätzlich lassen sich bei BI die Geschäftsperformance messen und ein zielgruppenspezifisches Reporting generieren. In der Praxis wird leider häufig das Potenzial nicht erkannt oder berücksichtigt, darüber hinaus auch mit den Ergebnissen von Smart Data Analytics zu arbeiten. Bei Einsatz von Predictive Analytics verbergen sich beispielsweise weitere Potenziale, um das eigene Business zu optimieren und die vorliegenden Daten für geschäftsrelevante Vorhersagen zu nutzen. Eine große Rolle spielen dabei computergestützte Softwarelösungen, die Datenanalysten optimal unterstützen. Machine Learning beziehungsweise maschinelles Lernen sind in diesem Zusammenhang zu nennen. Machine Learning bezeichnet die künstliche Generierung von Intelligenz aus historischen Erfahrungswerten. Die eingesetzte Softwarelösung lernt aus diversen Beispielen aus der Vergangenheit und verallgemeinert diese auf zukünftige Vorgänge.

Um Predictive Analytics und Machine Learning zusammen mit Smart Data einsetzen zu können, kommen Methoden von Data Mining, Text Mining und Forecasting

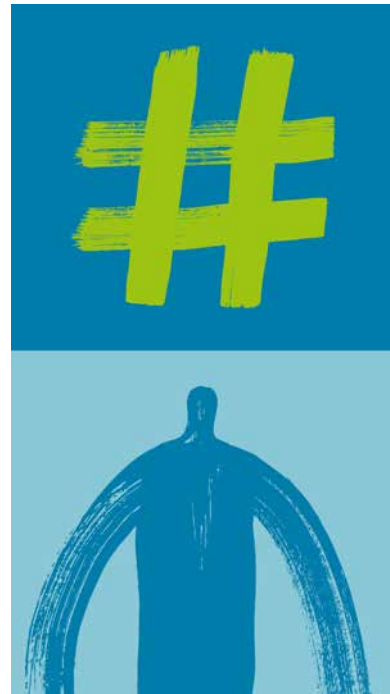
zum Einsatz. In einem ersten Schritt werden die eingesetzten Daten verknüpft und verschiedene Muster ermittelt. Anschließend erfolgt die Ermittlung mathematischer Verfahren unter Verwendung historischer Datensätze, um entsprechende Prognosen zu berechnen. Das entwickelte Modell wird dann auf aktuelle Daten angewendet, mit dem Ziel, zukünftige Ereignisse vorherzusagen. Die Subjektivität der getroffenen Annahmen wird dadurch auf ein Minimum reduziert und stützt sich auf datengestützte Erkenntnisse.

Machine Learning und Analytics als nächste Stufe von Big Data.

Um Machine Learning vorteilhaft einsetzen zu können, ist es zwingend erforderlich, Datenerfassungsstrategien über alle im Unternehmen befindlichen Datenspeicher und die gesamte IT-Infrastruktur zu entwickeln. Somit werden Maschinendaten mit strukturierten Daten verknüpft und auf die gesammelte Datenbasis angewendet. Zudem sind für die Erkennung von Mustern und der Generierung von Ergebnissen mathematische Algorithmen notwendig. Die Modelle, die eingesetzt werden, umfassen das überwachte Lernen (Supervised Learning) sowie das unüberwachte Lernen (Unsupervised Learning).

Beim Supervised Learning werden im Vorfeld Beispielmuster definiert und spezifiziert, um Informationen passend den Modellgruppen zuordnen zu können. Beispielsweise ist die automatische E-Mail-Spam-Erkennung ein Ergebnis des überwachten Lernens. Ob die Spam-Vorhersage richtig war, wird nach dem Vorgang durch einen Analysten oder Trainer bewertet. Für die Zukunft weiß das System nun, dass dieser Absender potenziell Spammessages versendet und verschiebt die betroffenen Mails, nach Anpassung der Parameter der angewendeten Methode, automatisch in den Spamordner. Das Unsupervised Learning hingegen lernt komplett automatisiert und selbstständig. Ziel dieses Ansatzes ist es, aus den gesammelten Daten bisher unbekannte Muster zu erkennen. Ähnliche Muster werden durch Segmentierung in Gruppen klassifiziert und anschließend, möglichst ohne Informationsverluste, komprimiert, um eine bessere Übersicht zu gewährleisten. Diese Methode benötigt in der Regel eine sehr große Datenmenge, um aussagekräftige Prognosen erstellen zu können.

Weitere Unterkategorien des überwachten Lernens sind das teilüberwachte Lernen, das bestärkende Lernen sowie das aktive Lernen. Das teilüberwachte Lernen stellt eine Kombination aus überwachtem und unüberwachtem Lernen dar und wird eingesetzt, wenn nur für einen Teil der Eingaben die möglichen Ergebnisse bekannt sind. Beim bestärkenden Lernen lernt der Algorithmus durch Belohnung und Bestrafung, wie in auftretenden Situationen zu handeln ist. Das beste Beispiel für bestärkendes Lernen ist das menschliche Lernen. Die Modelle des bestärkenden Lernens versuchen das natürliche Lernverhalten nachzubilden. Im aktiven Lernen erfolgt durch den Algorithmus eine Vorauswahl relevanter Fragen mit einer hohen Ergebnisrelevanz, um die Anzahl der Fragen zu minimieren und bietet



! Durch Smart Data und Data Analytics sollen Menschen bei wichtigen Entscheidungen unterstützt werden. Visualisierung hilft hierbei, die Daten bildlich verständlich darzustellen.

dem Algorithmus die Möglichkeit, für gewisse Eingangsdaten die gewünschten Ergebnisse selbst zu ermitteln.

Mithilfe integrierter Daten, effizientem Einsatz von Analysten und Machine Learning können Ergebnisse mit wertvollem Wissen generiert werden. Die vorhandenen Daten lassen sich unter anderem mit Daten verknüpfen, die nicht unbedingt im eigenen Datenpool vorhanden sind und extern beschafft werden. Beispiel sind hier Verkehrs-, Wetter- oder andere vom Geschäft unabhängige Daten.

Visualisierung der Ergebnisse. Durch Smart Data und Data Analytics sollen Menschen bei wichtigen Entscheidungen unterstützt werden. Visualisierung hilft hierbei, die Daten bildlich verständlich darzustellen. Eine wesentliche Rolle spielt bei der Unterstützung die Aufarbeitung der Ergebnisse aus dem Gesamtprozess von Smart Data Analytics. Ohne die zusammengefasste Darstellung können Analysten die große Anzahl verschiedener Datensätze und -formate nicht ohne Weiteres erfassen und somit auch keine Erkenntnisse für das Business oder Antworten auf Ad-hoc-Fragen finden. Zudem

wird es ohne einen Einblick in die Daten hinter der Visualisierung schwieriger, die Datenqualität zu kontrollieren. Mit aussagekräftigen Grafiken ist es möglich, Daten verständlich aufzubereiten und somit zusätzlich zur Optimierung anstehender Entscheidungen beizutragen. Aktuelle Softwarelösungen aus dem Bereich Data Analytics bieten hierfür geeignete Darstellungsmöglichkeiten unterschiedlichster Erkenntnisse.

Fazit: Smart Data statt Big Data. In Zeiten knapper Personalressourcen und schnell wachsender Datenmengen durch die steigende Anzahl datengenerierender Systeme verändert sich die Art und Weise, wie Unternehmen in Zukunft agieren. Bei der Optimierung von geschäftsrelevanten Prozessen spielt das Thema Big Data eine äußerst wichtige Rolle. Um einen Nutzen aus den generierten Daten zu ziehen, müssen diese allerdings zunächst intelligent miteinander verknüpft werden. Es ist zudem sehr wichtig, sich Zeit für die Themen rund um Datenhaltung, -klassifizierung und -aufbereitung zu nehmen. Die Komplexität und der Aufwand von Analysen im Big-Data-Umfeld dürfen nicht außer Acht gelassen werden. Fortgeschrittenes Daten-Management und Daten-Assessment bilden zwingende Voraussetzungen für die effektive Nutzung von Daten. Durch die Verknüpfung von Daten wird aus Big Data Smart Data, ein wertvolles Asset, das zur

Findung von geschäftsrelevanten Optimierungsmöglichkeiten genutzt werden kann. Die neuen technischen Möglichkeiten in diesem Umfeld können zusätzlich mit computergestützten Softwarelösungen beschleunigt werden. Machine Learning bietet hier die Möglichkeit, durch Erkennung von Anomalien und Mustern schnell und automatisiert unternehmerische Erkenntnisse aus den Daten zu erhalten. Ausschlaggebend ist hier, Trends und Ausreißer zu erkennen und mithilfe von Analysten richtig zu deuten.

Controlware unterstützt Sie bei der erfolgreichen Implementierung von Smart-Data-Lösungen als kompetenter Partner rund um Data Analytics. Unsere Experten erarbeiten gemeinsam mit Ihnen in einem individuellen Data Assessment Strategien zur optimalen Nutzung Ihrer Daten. Darüber hinaus erhalten Sie Empfehlungen zur Use-Case-Entwicklung oder Visualisierung & Interpretation aller relevanten Daten. Nutzen Sie unsere Expertise, um einen Mehrwert für Ihr Business zu generieren! ■



Christopher Gasteier,
Solution Manager IT-Management,
Controlware GmbH
www.controlware.de