



Dr. Henning Baars ist Akademischer Oberrat am Lehrstuhl für Wirtschaftsinformatik 1 der Universität Stuttgart und Sprecher der Fachgruppe „Business Intelligence“ der Gesellschaft für Informatik.

Predictive Analytics in der IT-basierten Entscheidungsunterstützung – methodische, architektonische und organisatorische Konsequenzen

Henning Baars

Der folgende Beitrag argumentiert, dass es sich bei „Predictive Analytics“ weniger um einen neuartigen Ansatz als um eine besondere Schwerpunktsetzung beim Einsatz analytischer Verfahren handelt. Diese bringt besondere Anforderungen bei Methodenwahl, architektonischer Verankerung und organisatorischer Einbettung mit sich, die v. a. für das Controlling auf Funktionsbereichsebene relevant sind.

1. Was ist neu an Predictive Analytics?

Der Begriff „Predictive Analytics“ hat in den letzten Jahren eine erhebliche Prominenz erreicht. Versprochen werden zukunftsorientierte und handlungsleitende Analysen, die über eine deskriptive und vergangenheitsorientierte Erklärung historischer Daten hinausgehen und durch einen konkreten Handlungsbezug zu einem direkten Wertbeitrag führen. In der aktuellen Diskussion wird das Thema Predictive Analytics häufig in einen engen Bezug zu sogenannten „Big-Data“-Infrastrukturen für den Umgang mit großen, heterogenen und in kurzer Zeit zu verarbeitenden Datenbeständen gesetzt. Teilweise werden Predictive-Analytics-Ansätze sogar als konstituierender Bestandteil von Big Data gesehen (z. B. bei Russom, 2011, S. 7 f.).

Unabhängig von den zugrundeliegenden technischen Implementierungen stellt sich die Frage, ob sich die durchgeführten Analysen tatsächlich grundlegend von jenen unterscheiden, die im Umfeld der integrierten Management- und Entscheidungsunterstützung, also der „Business Intelligence“ (BI) (Kemper et al., 2010, S. 8 f.), mittlerweile seit Jahrzehnten zum Einsatz kommen. Insbesondere die unter der Überschrift „Data Mining“ diskutierten Anwendungen waren seit jeher zukunftsgerichtet: Zweck der meisten anspruchsvolleren Data-Mining-Anwendungen war und ist die Aufdeckung nicht trivialer Muster in größeren Datenbeständen. Und die damit gewonnenen Mo-

delle werden üblicherweise auch auf aktuelle und zukünftige Datenkonstellationen angewandt, sei es in der IT-gestützten Betrugserkennung („Fraud Detection“), im Umgang mit Kundenabwanderungen („Change-and-Turn“ bzw. kurz „Churn-Analysen“), in der Aufdeckung von Verkaufstrends, in der Sortiments- und Laden-Layouts-Gestaltung oder in der Steuerung von Logistikprozessen.

Andererseits muss eingeräumt werden, dass Data-Mining-Lösungen im Kontext der BI lange Zeit eher die Rolle von Nischenanwendungen zukam. Folgende Aussage des BI-Leiters eines größeren Herstellers chemiebasierter Konsumgüter aus dem Jahre 2005 steht hierfür stellvertretend: *„Data Mining ist für uns auf jeden Fall ein Zukunftsthema. Das war es auch schon immer. Und das wird auch vermutlich so bleiben“*. Zu viele Voraussetzungen auf der Ebene der verfügbaren Daten, der Qualifikation der Benutzer, der Beherrschung der Analysekomplexität und der Verwertbarkeit der Ergebnisse standen einem breiten Einsatz von Data-Mining-Analysen entgegen.

Hier hat jedoch ein tiefgreifender Wandel stattgefunden. Zu nennen sind die fortschreitende Digitalisierung von Wirtschaft und Gesellschaft, die Durchdringung der physischen Umwelt mit IT-Komponenten (Internet of Things, IoT) und damit eine zunehmende Verfügbarkeit relevanter Daten, leistungsstarke mobile Endgeräte, benutzerfreundlichere Werkzeuge auch für anspruchsvolle Analysen sowie Infrastrukturen, die einen großflächigen Einsatz von Data-Mining-

Stichwörter

- Advanced und Predictive Analytics
- Big Data
- Business Intelligence
- Data Mining
- Machine Learning

Modellen in „near realtime“ ermöglichen. Schließlich ist das Bewusstsein für die Potenziale der Datenanalyse gestiegen. Insbesondere eröffnen sich immer neue Möglichkeiten, die mit den Analysen gewonnenen Vorhersagen und Modelle zeitnah in primären Geschäftsprozessen zu verwerten, etwa durch eine unmittelbare Veränderung der Produktpräsentation in Online-Shops, durch mobile Steuerung von Vertriebs-, Service- oder Logistikprozessen oder den Eingriff in flexible Produktionsumgebungen.

Die explizite Verwendung des Adjektivs „predictive“ signalisiert darüber hinaus eine neue Schwerpunktsetzung. Ziel dieses Beitrags ist es, die resultierenden methodischen, architektonischen und organisatorischen Konsequenzen herauszuarbeiten sowie die Relevanz der Thematik für das Controlling auszuloten. Nach einer begrifflichen Klärung und einem Überblick über die Anwendungsbereiche und Methoden der Predictive Analytics werden diese Aspekte in den weiteren Kapiteln des Beitrags einzeln vertieft.

2. Verwandte Konzepte

Der Begriffsbestandteil „analytics“ wird oftmals sehr breit verstanden. Nicht we-

nige Autoren subsumieren hierunter jegliche Systeme zur Datenaufbereitung und -analyse, also auch solche für das Berichtswesen, für einfache multidimensionale Ad-hoc-Abfragen (Online Analytical Processing, OLAP) oder einfache deskriptive Statistiken (vgl. z. B. *Davenport/Harris*, 2007, S. 7 f.). Daneben findet sich auch ein Verständnis, das sich auf Analysen beschränkt, in denen weiterreichende formale Modelle zum Einsatz kommen. Dieses wird auch häufig durch den Zusatz „advanced“ hervorgehoben (vgl. *Bose*, 2009, S. 156). Jenes Verständnis entspricht dem älteren Konzept der „modellorientierten Entscheidungs- und Managementunterstützungssysteme“ (*Kemper et al.*, 2010, S. 110 f.). Es ist zu betonen, dass der Begriff „Advanced Analytics“ immer noch weit gefasst ist und u. a. auch Simulationsmodelle und spezifische Decision Support Systeme (DSS) (insbes. auf Basis von Optimierungs- und Heuristikmodellen) aus dem Feld des Operations Research beinhaltet. Die häufige Kombination zu „Advanced and Predictive Analytics“ zeigt die engen Bezüge zwischen den Konzepten (vgl. *Bose*, 2009, S. 155). Einige Autoren unterscheiden noch weiter zwischen „Predictive Analytics“ und „Prescriptive Analytics“ (präskriptive Analysen), wobei zu letzteren solche Methoden gezählt werden, die

über die Prognose hinaus konkrete Handlungsempfehlungen zum Ergebnis haben, welche in einigen Fällen auch automatisiert umgesetzt werden können (vgl. *Lustig et al.*, 2010, S. 12).

Wie bereits im ersten Kapitel diskutiert, besteht eine enge Verwandtschaft zwischen den Analytics-Begriffen und dem Data Mining. Beim Data Mining kommen statistische Methoden zur Mustererkennung zum Einsatz, wobei die Muster in Form formaler oder algorithmischer Regeln und Modelle vorliegen (*Han et al.*, 2011, S. 5 f.; *Witten et al.*, 2011, S. 5 ff.). Ein Beispiel ist ein statistisch gewonnenes Klassifikationsmodell zur Einteilung von Kunden in verschiedene Kundenkategorien. Vom Data Mining leiten sich auch die Varianten des „Text Mining“ (Erkennung von Mustern in Textdatenbeständen) oder des „Process Mining“ (Identifikation prozessualer Zusammenhänge) ab. Methodisch stehen bei den diversen Mining-Ansätzen Methoden des „Machine Learning“ im Zentrum. Dem Machine Learning werden Methoden zugerechnet, bei denen Modelle nicht explizit von einem Programmierer vorgegeben, sondern algorithmisch aus einem bestehenden Datenbestand „erlernt“ werden (vgl. *Domingos*, 2012, S. 78). Machine-Learning-Ansätze

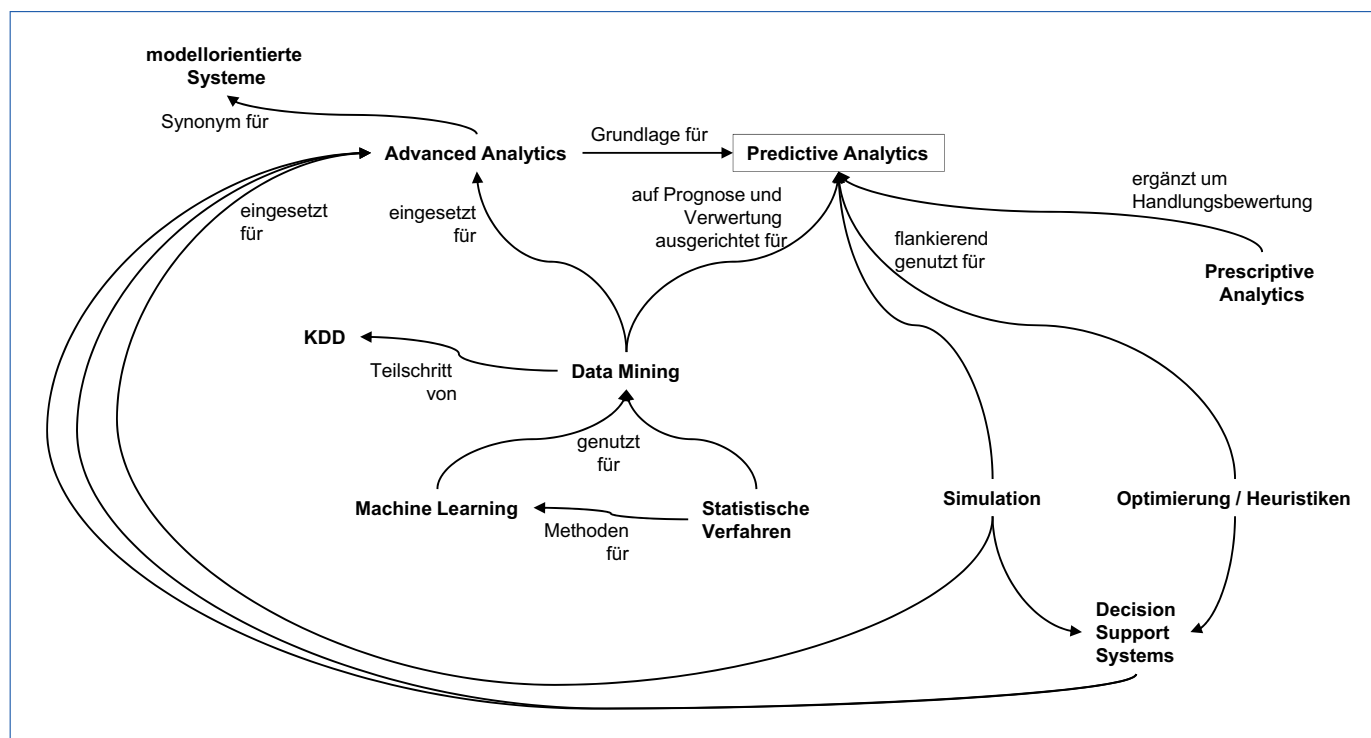


Abb. 1: Begriffsnetzwerk Predictive Analytics

haben sich in vielerlei Anwendungsdomänen bewährt – von der Internetsuche oder der Generierung von Produktvorschlägen in Online-Shops über die Anlagen- und Maschinenwartung bis zur Bild- und Sprachverarbeitung.

Den Anwendungskontext des Data Mining bildet die „Knowledge Discovery in Databases“ (KDD) ab, also der Prozess zur datenbankbasierten Wissensgenerierung, in den das Data Mining als einzelne Phase eingebettet ist (vgl. *Fayyad et al.*, 1996, S. 28 ff.). Sowohl beim Data Mining als auch bei KDD deutet bereits die Wortwahl auf einen explorativen Charakter hin. Aus betrieblicher Sicht steht jedoch i. d. R. die weitere Verwertung der gewonnenen Regeln und Modelle im Zentrum, d. h. deren Anwendung auf neue Datensätze zur Vorhersage zukünftiger Entwicklungen. Insofern trifft der Begriff „Predictive Analytics“ den intendierten Zweck besser als „Data Mining“.

Abb. 1 stellt die Begriffe und ihre Zusammenhänge im Überblick als Begriffsnetzwerk dar.

3. Anwendungen, Methoden und methodische Voraussetzungen bei den Benutzern

Predictive-Analytics-Methoden können für ein breites Spektrum an Anwendungen eingesetzt werden. In der Literatur finden sich eine Reihe spektakulärer Beispiele für innovative Einsätze, die oft mit dem Vorsatz „Predictive“ gekennzeichnet werden. So wurde das „Predictive Policing“ als Vorhersage von Straftaten und -orten (vgl. *Pearsall*, 2010) intensiv in der Presse behandelt. Auch die proaktive Lieferung von Produkten in kundennahe Lager vor einer tatsächlichen Bestellung wird gerne zitiert. Eine Marktstudie des *The Data Warehousing Institute (TDWI)* aus dem Jahr 2014 zeigt allerdings, dass sich die Hauptanwendungsgebiete nicht wesentlich von denen unterscheiden, in denen bereits bisher Data Mining eingesetzt wurde: Die Liste der elf häufigsten Einsatzbereiche umfasst ausschließlich Data-Mining-Anwendungen wie die Generierung von Produktempfehlungen für Cross- und Upselling, Analyse des Kundenbindungsverhaltens

(Churn-Analyse u. ä.), Betrugserkennung oder die Vorhersage von Zahlungsausfällen (vgl. Abb. 2).

Neue Anwendungen entstehen derzeit insbesondere im Umfeld des Internet of Things und dessen Einsatz in Produktion und Logistik, die in Deutschland unter der Schlagzeile „Industrie 4.0“ diskutiert werden. Eine wichtige Anwendung in diesem Kontext ist die „Predictive Maintenance“, d. h. die datengestützte und proaktive Wartung.

Hinsichtlich der eingesetzten Methoden dominieren bislang überwiegend „klassische“ Ansätze aus dem Data Mining und hier besonders solche aus dem Machine Learning (vgl. *Halper*, 2014, S. 18; zur folgenden Einteilung vgl. *Hippner/Wilde*, 2001, S. 96 ff.). Aufgrund ihrer prognostischen Ausrichtung unmittelbar relevant sind Wirkungsanalysen und Klassifikationsanalysen. Bei Wirkungsanalysen wird von gegebenen Merkmalen auf zu prognostizierende Variablen geschlossen. Dies beinhaltet insbes. auch Zeitreihenanalysen. Konkrete Verfahren sind u. a. Regressionsverfahren, Künstliche Neuronale Netze (KNN), Box-Jenkins-Verfahren

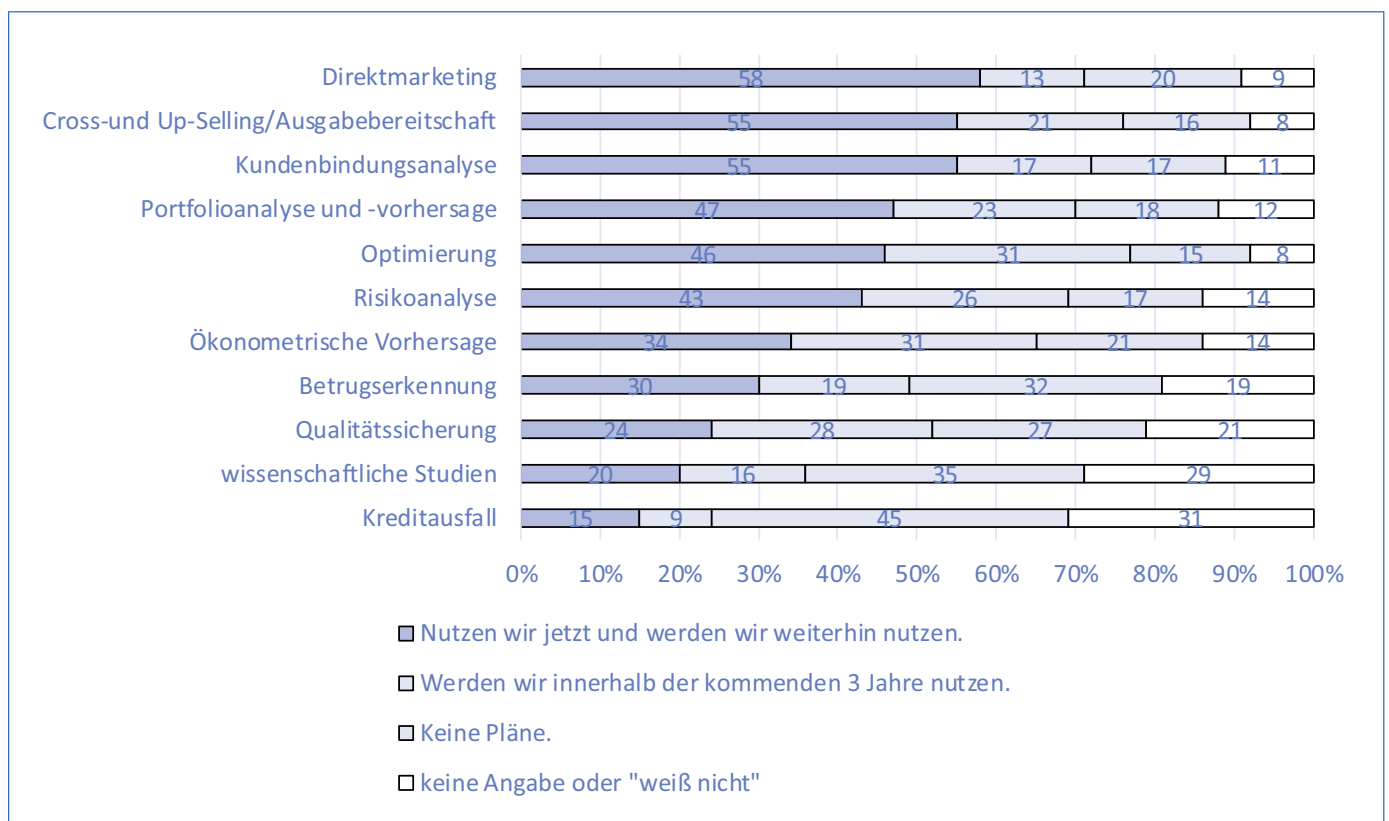


Abb. 2: Antworten auf die Frage „Wofür wird Predictive Analytics in Ihrem Unternehmen genutzt?“ (n=126) (übersetzt übernommen von *Halper*, 2014, S. 8)

ren oder genetische Algorithmen. Klassifikationsanalysen ordnen Objekte aufgrund ihrer Eigenschaften in vorgegebene Klassen ein (z. B. A-, B- oder C-Kunde). Zu den eingesetzten Verfahren zählen die logistische Regression, Klassifikationsbäume, Support-Vektor-Maschinen (SVM), Bayessche Netze, KNN oder genetische Algorithmen.

Diese unmittelbar prädikativ orientierten Methoden können auf solchen aufsetzen, die erst einmal deskriptiver Natur sind, insbes. Assoziationsanalysen, bei denen Abhängigkeiten zwischen Objekten identifiziert werden (v. a. bei Warenkorbanalysen, die Auskunft darüber geben, welche Produkte zusammen gekauft werden) und Clusteranalysen (Gruppenbildung). Derartige Verfahren können zum besseren Verständnis der Daten genutzt und später als Grundlage etwa für Klassifikationen genutzt werden.

Neuere Ansätze kombinieren mehrere Verfahren, um deren jeweilige Schwächen zu kompensieren und ein „Überlernen“ zu vermeiden, d. h. eine zu spezifische Ausrichtung auf den Lerndatenbestand, anhand derer das Modell „trainiert“ wird. Dies wird als „Ensemble Learning“ bezeichnet (vgl. Witten et al., 2011, S. 351 ff.).

Im Laufe der Zeit ist eine unübersehbare Menge an Verfahren und Verfahrensvarianten für das Machine Learning entstanden. Kommerzielle Data-Mining-Tools führen die Benutzer hier mittlerweile eng und intuitiv und schlagen automatisch Verfahren anhand der erzielten Ergebnishüte vor. De facto ist jedoch immer noch ein Verständnis für die Methoden und Parameter notwendig, schon allein, um die Modelle sinnvoll kalibrieren zu können. Zudem setzen neue Data-Mining-Verfahren sowie deren Einsatz in den massiv parallelen Big-Data-Umgebungen dennoch den Einsatz anspruchsvollerer Werkzeuge wie *R* oder *Octave* (oder gar Programmierung) voraus, die sich eher an IT- und Statistik-versierte Anwender richten. Auch werden für Predictive-Analytics Methoden eingesetzt, die gänzlich andere Werkzeuge voraussetzen, etwa wenn Netzwerk- oder Prozessstrukturanalysen für Predictive-Analytics-Anwendungen durchgeführt werden.

Des Weiteren ist der Analyseprozess selbst voraussetzungsvoll, was sich auch an Industrie-Vorgehensmodellen für das

Data Mining wie CRISP-DM (Cross-Industry-Standard for Data Mining) zeigt (vgl. Shearer, 2000). Hier sind lange Phasen für die Datengewinnung, -interpretation und -bereinigung vorgesehen, die erfahrungsgemäß sehr zeitaufwändig sind und Erfahrung erfordern. Schließlich muss die Güte der Modelle laufend überwacht und diese müssen ggf. neu erstellt werden. Bei der Angriffs- und Betrugserkennung ist es beispielsweise so, dass Betrüger ihr Vorgehen an die Modelle anpassen. In einem Marktkontext verschieben sich einzelne Kundengruppen oder die Wettbewerbssituation und bei einer Maschine ändert sich der Wartungsbedarf mit veränderten Produktmischen oder Materialeigenschaften.

4. Predictive Analytics: Methodische Besonderheiten

Die gezielte Ausrichtung auf die zukunfts- und verwertungsorientierte Perspektive der Predictive Analytics führt zu einigen methodischen Anforderungen, insbes. 1. die Ergänzung von Sensitivitäts- und Konfidenzinformationen und 2. die Erweiterung des Machine-Learning-orientierten Methodenportfolios aus dem Data Mining um Optimierungs-, Heuristik- und Simulations-Modelle aus dem DSS-Bereich.

1. Der erste Punkt erschließt sich von selbst: Je weiter die Prognose reicht, je unsicherer sie ist und je relevanter diese wird, desto wichtiger wird es, dem Entscheider neben dem Ergebnis (etwa einer Ausfallprognose) Kennzahlen zur Modellgüte (z. B. Sensitivität und Spezifität) zu liefern (vgl. Bruckhaus, 2007). Deren korrekte Interpretation setzt allerdings ebenfalls wieder ein hohes Maß an Vertrautheit des Benutzers mit der Thematik voraus.
2. Bei den in Kapitel 3 aufgeführten Machine-Learning-Methoden handelt es sich um generische Ansätze, die nicht oder nur sehr eingeschränkt auf vorhandenes domänenspezifisches Wissen zurückgreifen. So ist es aus Sicht eines Machine-Learning-Klassifikationsalgorithmus unerheblich, ob die Input-Parameter zur Einordnung von Krankheitsbildern, von Webseitenutzern oder von Energieverbrauchern genutzt werden. Der Benutzer hinterlegt die ihm bekannten Zusammenhänge nicht explizit im System,

sondern bereitet lediglich in der Vergangenheit gesammelte Daten vor. Letztlich ist die Modellgenerierung aus Benutzersicht somit eine „Black Box“. Dies gilt bei den meisten Methoden auch für die Funktionsweise des resultierenden Modells. Während der Aufbau eines Klassifikationsbaums noch nachvollziehbar ist, gilt dies für ein KNN oder eine SVM nicht mehr.

Machine-Learning-Methoden unterscheiden sich somit grundlegend von spezifischen DSS-Ansätzen für Optimierungen und Simulation, wie sie etwa für Fragen der Produktion, der Logistik oder auch des Marketing entwickelt wurden. Hier werden explizit Zusammenhänge und Restriktionen formalisiert, etwa Kapazitätsrestriktionen und ingenieurwissenschaftliche Abhängigkeiten bei einem Optimierungsmodell für die Produktion oder Prozesszusammenhänge für ein Simulationsmodell für Transportnetzwerke. Es handelt sich insofern gewissermaßen um „White-Box-Modelle“, deren Inhalte das Wissen des Modellierers aufgreifen.

Obgleich sich die („Black-Box“-)Machine-Learning-Modelle in vielen Fällen gegenüber den domänenspezifischen („White-Box“-)Modellen als überlegen erwiesen haben, kann dies ihre grundlegende Beschränkung der Nichtberücksichtigung kontextspezifischer Zusammenhänge nicht aus dem Weg räumen.

Bei einem prädikativen Einsatz kann dies schnell relevant werden, v. a. wenn die vergangenheitsorientierten Daten strukturell nicht die zukünftige Situation abbilden können oder zu wenig vergleichbar sind. Ein Beispiel für den ersten Fall sind absehbare Veränderungen in der Alterspyramide, die ein mit Vergangenheitsdaten gefüttertes Modell nicht berücksichtigen kann (vgl. Koschtial/Felden, 2015, S. 22). Ein Beispiel für den zweiten Fall sind Produkte wie Bücher mit geringer Auflage, deren Nachfrage oft von sehr spezifischen Faktoren abhängig ist, die bei jedem Buchtitel individuell ausfallen. In diesem Fall kann sich eine scheinbare Datenflut für den Gesamtbuchbestand bei Betrachtung des einzelnen Titels schnell als Datenwüste erweisen.

Letztendlich empfiehlt es sich hier, die Machine-Learning-Methoden durch Simulationsmethoden oder explizite

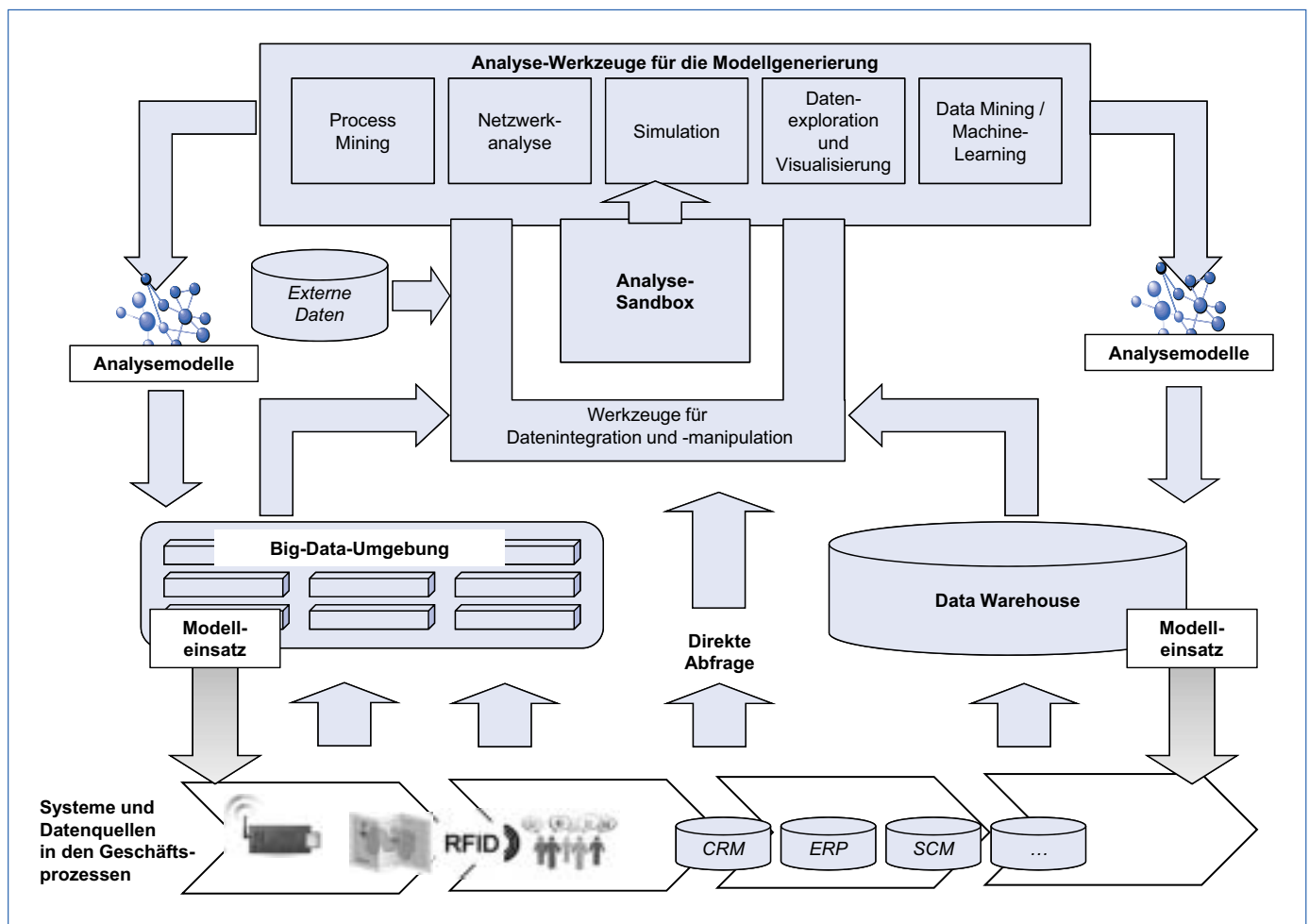


Abb. 3: Architekturelle Einbindung von Predictive-Analytics-Lösungen

(Abkürzungen: RFID = Radio Frequency Identification/funkgestützte Identifikation, CRM = Customer Relationship Management, SCM = Supply Chain Management, ERP = Enterprise Resource Planning)

Regeln oder Faktoren zu ergänzen oder zu ersetzen (vgl. Koschtial/Felden, 2015, S. 22 ff.).

5. Predictive Analytics: Architektonische Besonderheiten

Die architektonischen Besonderheiten ergeben sich aus dem Erfordernis separierter Analyseumgebungen, dem Bedarf hochwertiger Daten für die Modellgenerierung sowie aus der Verknüpfung der Modellausführung mit der Initiierung von Aktionen. Diese resultieren in Anforderungen zum Aufbau von Analyse-Sandboxes sowie zur Integration in die Systemwelten für Business Intelligence und Big Data. Abb. 3 gibt hierzu einen Überblick.

Die Modellbildung selbst erfordert eine von den produktiv genutzten Datenbanken (etwa für das Reporting) isolierte

Datenhaltung in Form einer sog. „Analyse-Sandbox“. In dieser werden interne und externe Daten für die Analyse und damit die Modellgenerierung gesammelt. Neben den diskutierten Werkzeugen für Data Mining sowie ggf. Prozess- und/oder Netzwerkanalysen, Simulationen und weiterer Analysesoftware benötigt der Modellersteller außerdem Werkzeuge zur Datenintegration und -manipulation (für die Integration neuer Daten in die Sandbox) sowie für Datenexploration und -visualisierung (zum Verständnis, zur Interpretation und zur Beurteilung der Daten und Modelle).

Aus der Business Intelligence können üblicherweise zentrale unternehmensinterne Kennzahlen bezogen werden, v. a. für die betriebswirtschaftliche Bewertung von Sachverhalten. Hervorzuheben sind hierbei die anwendungsübergreifend integrierten, dauerhaften und konsistenten Datenhaltungssysteme der BI, die „Data

Warehouses“ (DWHs). Ein systematisch betriebenes DWH, das durch ein von der zuständigen BI-Einheit betriebenes Datenqualitätsmanagement flankiert wird, sorgt für hochwertige Kennzahlen und reduziert den in Kapitel 3 angesprochenen Aufwand für die Datenintegration und -aufbereitung bei der Modellgenerierung. Darüber hinaus bieten aktuelle BI-Infrastrukturen auch Möglichkeiten zum Modellimport und zur Modellausführung. Mit den Modellen können relevante Datenkonstellationen im DWH (oder in einer anderen Datenbank oder einem operativen System) zur Initiierung von konkreten Aktionen genutzt werden, etwa zur Sperrung bestimmter Zahlungsoptionen für Kunden, deren Verhalten gemäß Modell auf ein problematisches Zahlungsverhalten hindeutet („Operationalisierung“ des Modells).

Eine besondere Herausforderung stellt die zeitnahe und parallele Modellanwen-

dung auf eine große Zahl an zu berücksichtigenden Objekten dar, etwa, wenn ein neues Klassifikationsmodell kurzfristig auf den gesamten Kundenbestand angewendet werden soll. Noch problematischer ist der Umgang mit Datenströmen („Streaming Data“), z. B. bei einer Folge von Benutzeraktivitäten auf einer Webseite. Gerade im IoT- und Industrie-4.0-Umfeld kommen derartige Datenströme zudem z. T. simultan von einer großen Zahl an Quellen (vgl. *Chen et al.*, 2014, S. 177 ff.). In derartigen Umgebungen ist es vorgesehen, dass eine Vielzahl an „smarten“ Sensoren kontinuierlich und eigenständig Messwerte über das Netz sendet.

Für die genannten Fälle bieten sich Lösungen aus dem Big-Data-Umfeld an, die z. B. eine Hauptspeicherbasierte Ausführung von Modellen auf vielen parallelen Rechnern erlauben oder die auf den Umgang mit Streaming-Daten ausgelegt sind (vgl. *Williams et al.*, 2014). Die Big-Data-Systeme können auch für die Vorhaltung und Near-realtime-Analyse unstrukturierter Daten eingesetzt werden, deren Einsatz für Predictive-Analytics-Anwendungen zunehmend diskutiert wird. Aktuell finden sich v. a. Beispiele im Kontext der Sicherheit (etwa Erkennung auffälligen Verhaltens von Personen in Videoübertragungen). Es ist aber zu vermuten, dass entsprechende Methoden auch in betriebliche Kontexte Einzug halten werden, etwa bei der Maschinen- und Anlagenüberwachung.

Insgesamt kann konstatiert werden, dass eine Ausrichtung auf Predictive Analytics stärker als früher bei Analyseaufgaben üblich mit infrastrukturellen Aufgaben einhergeht, insbes. durch den implizierten Verwertungszusammenhang für die gewonnenen Modelle. Dies erhöht die Anforderungen an die Benutzer zusätzlich.

6. Predictive Analytics: Organisatorische Aspekte

Wie in den Kapiteln 3 bis 5 bereits ausgeführt, hängt der Einsatz von Predictive Analytics von zahlreichen Voraussetzungen ab. Idealerweise liegen bei den Benutzern statistisch-methodische Kenntnisse für die Anwendung, Kalibrierung und Anpassung der Methoden vor. Aufgrund der Zukunftsorientierung und der Grenzen des Black-Box-Machine-Learning-

Ansatzes sollten die Benutzer auch umfangreiches Wissen zur Anwendungsdomäne mitbringen und ggf. Simulationsmodelle beherrschen. Gleichzeitig sollten sie in der Lage sein, sich in den BI- und Big-Data-Systemwelten zu bewegen, um relevante Quelldaten abzugreifen und spezifische Methoden einzusetzen. Der Modelleinsatz schließlich bedingt oft auch Programmierkenntnisse. Somit ist das Berufsbild eines „Data Analyst/Data Scientist“ umrissen (vgl. *Dhar*, 2013, S. 69), das in dieser Breite kaum voll von einer einzelnen Person auszufüllen sein wird – insbesondere nicht in den anwendenden Fachabteilungen. Dies führt zum Bedarf für zentrale Data-Science-Teams, die als externe oder interne Dienstleister auftreten können und die von der BI-Einheit, der IT oder den Fachabteilungen für Predictive-Analytics-Aufgaben gebucht werden können.

7. Bedeutung für das Controlling und Ausblick

Besondere Relevanz hat Predictive Analytics auf der Ebene der Funktionsbereiche, in denen primäre Geschäftsprozesse geplant, überwacht und gesteuert werden. Beispiele sind die Maschinenwartung im Produktionscontrolling, die Überwachung von Warenflüssen im Logistikcontrolling oder der Umgang mit Kunden im Vertriebscontrolling.

Die Modelle und Systeme sind dabei einerseits selbst Gegenstand des Controllings, da sie in ihrer Effektivität und Effizienz eng überwacht werden müssen. Gerade bei automatisierten Reaktionen können falsch kalibrierte Modelle schnell erheblichen Flurschaden anrichten, etwa durch die Vergabe zu hoher Rabatte, die Ablehnung guter Kunden oder das Übersehen eines kritischen Maschinendefektes.

Andererseits verändert sich auch die Rolle des Controllers. Die Veränderung kann in zwei Richtungen laufen: Der Controller kann sich entweder in Richtung des Data Scientist entwickeln, der sein Steuerungswissen in die Modellgenerierung und -verankerung einbringt und so die Automatisierung standardisierbarer Analyseaufgaben vorantreibt. Oder aber er nimmt verstärkt die Rolle eines „Exception-Handlers“ ein, der in Grenz- oder Ausnahmesituationen für die Interpretation zuständig ist, beispielsweise, wenn

die Klassifikation eines Kunden unklar ist oder das System einen kritischen Trend vorhersagt, der einer tieferen Analyse bedarf. Beide Rollen sind mit steigenden Anforderungen an den Controller verbunden.

Insgesamt bleibt festzuhalten, dass die Akzentuierung einer zukunfts- und verwertungsorientierten Ausrichtung von Datenanalysen verschärfte Anforderungen an Methoden, Architekturen und die Organisation stellt. Die Möglichkeiten sind hierbei noch lange nicht ausgeschöpft und viele relevante Technologien etwa für Prozess- und Netzwerkanalysen oder auch für Simulationen sind noch nicht in integrierten Umgebungen enthalten. Andererseits bedeutet dies, dass hier noch Differenzierungspotenzial besteht – ein Differenzierungspotenzial, das angesichts der voranschreitenden Digitalisierung von Wirtschaft und Gesellschaft für viele Unternehmen auch ein strategisches ist.

Keywords

- Advanced and Predictive Analytics
- Big Data
- Business Intelligence
- Data Mining
- Machine Learning

Summary

The following contribution argues that „predictive analytics“ is less a novel approach than a new focus for the application of analytical methods. This comes with special requirements regarding method selection, architectural integration, and organizational implementation that are of particular relevance for management control on business function level.

Literatur

- Bose, R.*, Advanced analytics: opportunities and challenges, in: *Industrial Management and Data Systems*, 109. Jg. (2009), H. 2, S. 155–172.
- Bruckhaus, T.*, The business impact of predictive analytics, in: *Zhu, Q./Davidson, I.* (Hrsg.), *Knowledge Discovery and Data Mining. Challenges and Realities*, Hershey, New York 2007, S. 114–138.
- Davenport, T. H./Harris, J. G.*, *Competing on Analytics – The New Science of Winning*, Boston 2007.
- Dhar, V.*, Data Science and Prediction, in: *Communications of the ACM*, 56. Jg. (2013), H. 12, S. 64–73.

Domingos, P., A few useful things to know about machine learning, in: Communications of the ACM, 55. Jg. (2012), H. 10, S. 78–87.

Fayyad, U./Piatetsky-Shapiro, G./Smyth, P., The KDD Process for Extracting Useful Knowledge from Volumes of Data, in: Communications of the ACM, 39. Jg. (1996), H. 11, S. 27–34.

Halper, F., Predictive Analytics for Business Advantage, in: TDWI Best Practices Report, 1. Quartal 2014.

Han, J./Kamber, M./Pei, J., Data Mining – Concepts and Techniques, 3. Aufl., Amsterdam u. a. 2011.

Hippner, H./Wilde, K. D., Der Prozess des Data Mining im Marketing, in: Hippner, H./Küsters, U./Meyer, M./Wilde, K. D. (Hrsg.), Handbuch Data Mining im Marketing, Wiesbaden 2001, S. 21–91.

Kemper, H. G./Baars, H./Mehanna, W., Business Intelligence – Grundlagen und praktische Anwendungen, Wiesbaden 2010.

Koschtiel, C./Felden, C., On the Way from a Knowledge Discovery in Databases to a Predictive Analytics, in: Mayer, J. H./Quick, R. (Hrsg.), Business Intelligence for New-Generation Managers, Heidelberg u. a. 2015, S. 17–26.

Lustig, I./Dietrich, B./Johnson, C./Dziekan, C., The Analytics Journey – An IBM view of the structured data analysis landscape: descriptive, predictive and prescriptive analytics, in: Analytics, 2010, S. 11–18.

Russom, P., Big data analytics, in: TDWI Best Practices Report, 1. Quartal 2011.

Shearer, C., The CRISP-DM Model: The New Blueprint for Data Mining, in: Journal of Data Warehousing, 4 Jg. (2000), H. 5, S. 13–22.

Witten, I. H./Frank, E./Hall, M. A., Data Mining – Practical Machine Learning Tools and Techniques, 3. Aufl., Amsterdam u. a. 2011.

Literaturtipps aus dem [Online-Archiv](#) der CONTROLLING:

- Michael Fahn und Oliver Köhler, Aufbau eines strategischen IT-Controllings zur Unterstützung übergreifender IT-Steuerung im Lufthansa-Konzern, Ausgabe 10/2008, S. 535–542.

Das neue revolutionäre Management-System.



2016. Rund 200 Seiten. Gebunden ca. € 24,90
ISBN 978-3-8006-5087-3 | Neu im Februar 2016

Portofrei geliefert: vahlen.de/15478815

Holacracy

ist eine der faszinierendsten Managementphilosophien seit Six Sigma. Doch anders als Six Sigma und weitere autoritäre, Top-down orientierte Strukturen, versetzt Holakratie jeden Mitarbeiter in einer Organisation in die Lage, nach bestimmten Prinzipien selbstständig Entscheidungen zu treffen. Das neue Buch zeigt die Praktiken auf, holakratische Strukturen unternehmensweit oder in ausgewählten Abteilungen zu implementieren.

Brian Robertson

ist US-amerikanischer Unternehmer. Er gründete in den 1990er Jahren »Ternary Software«. Hier beschäftigte er sich zum ersten Mal intensiv mit alternativen Organisationsformen. Nach dem Verkauf des Unternehmens startete er mit »HolacracyOne« ein Beratungsunternehmen, das Organisationen hilft, eine holakratische Kultur zu implementieren.

Erhältlich im Buchhandel oder bei: vahlen.de | Verlag Franz Vahlen GmbH
80791 München | bestellung@vahlen.de | Preise inkl. MwSt. | 165340

Vahlen