

Datenqualität als Erfolgsfaktor im Business Analytics

Autorenvorstellung

Prof. Dr. *Mathias Klier* ist Inhaber der *Péter-Horváth*-Stiftungsprofessur für BWL mit dem Schwerpunkt betriebswirtschaftliches Informationsmanagement am Institut für Technologie- und Prozessmanagement (ITOP) der Universität Ulm, E-Mail: mathias.klier@uni-ulm.de

Prof. Dr. *Bernd Heinrich* ist Inhaber des Lehrstuhls für Wirtschaftsinformatik II an der Fakultät für Wirtschaftswissenschaften der Universität Regensburg, E-Mail: bernd.heinrich@wiwi.uni-regensburg.de

Zitation: Mathias Klier, Bernd Heinrich: Datenqualität als Erfolgsfaktor im Business Analytics. in: Controlling – Zeitschrift für erfolgsorientierte Unternehmenssteuerung, Vol. 28, No. 8-9, 2016, pp. 488-494, ISSN 0935-0381, DOI: 10.15358/0935-0381-2016-8-9-488.

Intro

„Garbage In, Garbage Out“ – dies gilt auch im Business Analytics und in der datengetriebenen Entscheidungsunterstützung. Vor diesem Hintergrund wird im vorliegenden Beitrag eine Metrik zur Messung der Aktualität von Daten vorgeschlagen. Die Anwendung und der praktische Nutzen des Einsatzes der Datenqualitätsmetrik werden am Fallbeispiel eines Automobilherstellers verdeutlicht.

Summary

“Garbage In, Garbage Out“ – this especially holds true for the fields of Business Analytics and data-driven decision support. Against this background, in this article we propose a probability-based metric to assess the currency of data. We demonstrate the application of the metric as well as its practical benefit using the case of a major German car manufacturer.

Keywords (deutsch): Aktualität, Business Analytics, Datenqualität, Messung, Metrik

Keywords (englisch): Assessment, Business Analytics, Currency, Data Quality, Metric

Zentrale Aussagen

- Im Business Analytics müssen die Messung und Sicherstellung der Datenqualität zentrale Bestandteile sein.
- Mithilfe der vorgeschlagenen Metrik kann die Aktualität von Datenwerten zielgerichtet und weitgehend automatisiert gemessen werden.
- Die Anwendung und den praktischen Mehrwert der Metrik verdeutlicht das Fallbeispiel eines großen deutschen Automobilherstellers.

1. Bedeutung der Datenqualität für Business Analytics

Im Zuge der Digitalisierung stehen Unternehmen – bspw. über soziale Medien und das Internet, aber insbesondere auch in unternehmensinternen Datenbanken – sehr umfangreiche und immer weiter wachsende Datenmengen („Big Data“) zur Verfügung. Für die Unternehmenssteuerung ergeben sich hieraus neue Chancen und Herausforderungen (*Internationaler Controller Verein*, 2014). So ermöglicht Business Analytics – in Form einer zielgerichteten und fundierten Analyse dieser Daten mittels statistischer und quantitativer Methoden und Modelle (vgl. *Davenport/Harris*, 2007) – eine verbesserte, datengetriebene Entscheidungsunterstützung und eröffnet damit großes Potenzial in den unterschiedlichsten Anwendungsbereichen.

Damit die abgeleiteten Ergebnisse und die darauf basierenden Entscheidungen valide und wertstiftend sind, sind die Messung und Sicherstellung der Qualität der zugrundeliegenden Daten unabdingbar. Gerade im Rahmen von Business Analytics ist der Datenqualität große Bedeutung beizumessen. Zum einen werden hier meist sehr komplexe Sachverhalte analysiert, die sich einfachen „Plausibilitätsüberprüfungen“ durch Menschen entziehen. Zum anderen geht die hohe normative Kraft quantitativer Modelle und abgeleiteter Ergebnisse häufig mit einer nicht zu unterschätzenden „(Leicht-)Gläubigkeit“ der Anwender einher. Darüber hinaus werden viele Analysen und Entscheidungen (teil-)automatisiert. Insgesamt führt dies dazu, dass im Business Analytics der Mensch als „Regulativ“ für eventuell fehlerbehaftete Prozesse und Ergebnisse an Wirksamkeit verliert und die Sicherstellung der Datenqualität und damit auch der Validität der Analysen und Ergebnisse umso wichtiger werden. Schizophren ist dabei der Umstand, dass die Messung und Sicherstellung der Datenqualität gerade diese Aspekte (d. h. die automatisierte Messung sehr komplexer Sachverhalte mittels quantitativer Modelle) ebenso adressieren muss. Gelingt dies nicht, wirkt die Datenqualität im Business Analytics als Hindernis und nicht als Enabler.

Die Empirie zeigt, dass die analysierten Daten häufig durch eine geringe Datenqualität charakterisiert sind – u. a. bedingt dadurch, dass Daten teilweise sehr schnell veralten (vgl. z. B. Heinrich et al., 2012). Eine Befragung durch *Experian Qas* (2013) ergab, dass inkorrekte und veraltete Daten die häufigsten Datenqualitätsdefekte darstellen und zu hohen Budgetüberschreitungen, sinkender Kundenzufriedenheit sowie Kundenbeschwerden und -abwanderungen führen. Bspw. gaben zwei Drittel der befragten Organisationen an, signifikante Datenqualitätsprobleme im Customer Relationship Management sowie bei der Analyse bzw. Verwertung von (großen) kundenbezogenen Datenmengen zu haben. Neben Volume, Variety und Velocity wird daher Veracity – also die Qualität der Daten – als vierte zentrale Dimension von Big Data bezeichnet (vgl. *IBM Institute for Business Value*, 2012). Im Business Analytics kommt gerade dieser Dimension hohe Bedeutung zu. Ist keine ausreichende Datenqualität gegeben, resultieren fehlerhafte Ergebnisse, die zu falschen Entscheidungen führen und mehr Schaden anrichten als Nutzen generieren (vgl. Abb. I).

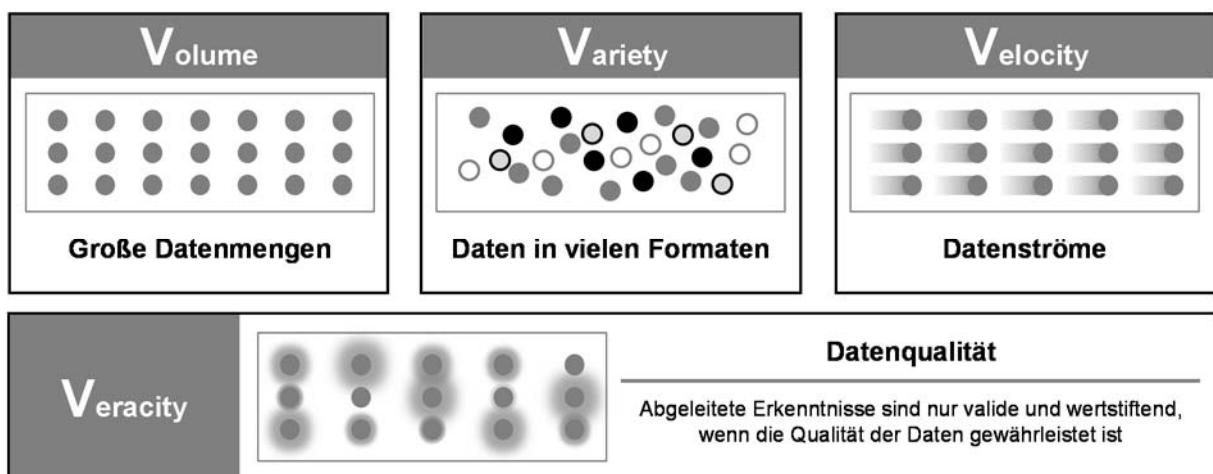


Abb. I: Charakteristika von Big Data (angelehnt an *IBM Institute for Business Value* (2012))

Deshalb widmet sich der vorliegende Beitrag der Fragestellung, wie die Datenqualität mittels Metriken adäquat gemessen und im Rahmen von Business Analytics zur Verbesserung der datengetriebenen Entscheidungsunterstützung berücksichtigt werden kann. Konkret werden dabei mit der Aktualität der Daten ein ausgewähltes Datenqualitätsmerkmal und mit der Möglichkeit einer fein granularen, zielorientierten Steuerung im Servicebereich eine exemplarische Anwendung fokussiert.

Der Beitrag ist wie folgt aufgebaut: Im nächsten Abschnitt werden das zugrunde gelegte Datenqualitätsverständnis und das fokussierte Datenqualitätsmerkmal Aktualität erläutert. Danach wird eine Metrik zur Messung der Aktualität von Daten vorgeschlagen. Anschließend werden die Anwendung der Metrik und der praktische Mehrwert des Einsatzes der Metrik am

Beispiel des Managements von Serviceaufträgen eines großen deutschen Automobilherstellers illustriert. Der letzte Abschnitt fasst die Ergebnisse zusammen.

2. Datenqualität und das Datenqualitätsmerkmal Aktualität

Im Bereich der Datenqualität werden mit Design- und Konformitätsqualität zwei Perspektiven unterschieden (vgl. z. B. *Heinrich/Helfert*, 2003). Dabei adressiert die Designqualität, inwiefern die nachgefragten Daten durch die spezifizierten Daten abgedeckt werden (vgl. Abb. 2). Das heißt, inwiefern bspw. alle Serviceauftragsattribute, welche für die geplanten Datenanalysen zur Unterstützung des Serviceauftragsmanagements benötigt werden (z. B. Servicestatus, Servicetyp, Auftraggeber), im Datenschema des Informationssystems tatsächlich erfasst sind. Zur Messung der Designqualität wird i. d. R. auf Informationsbedarfsanalysen und Befragungen der Datenanwender zurückgegriffen.

Demgegenüber repräsentiert die Konformitätsqualität, inwiefern die vorhandenen Daten mit den zugehörigen Realweltausprägungen korrespondieren (vgl. Abb. 2). Das heißt, inwiefern bspw. die gespeicherten Serviceauftragsdaten mit den tatsächlichen Auftragsdaten übereinstimmen. Die Konformitätsqualität kann anhand verschiedener Datenqualitätsmerkmale – wie z. B. Korrektheit, Vollständigkeit und Aktualität – beurteilt werden und wird im Weiteren fokussiert.

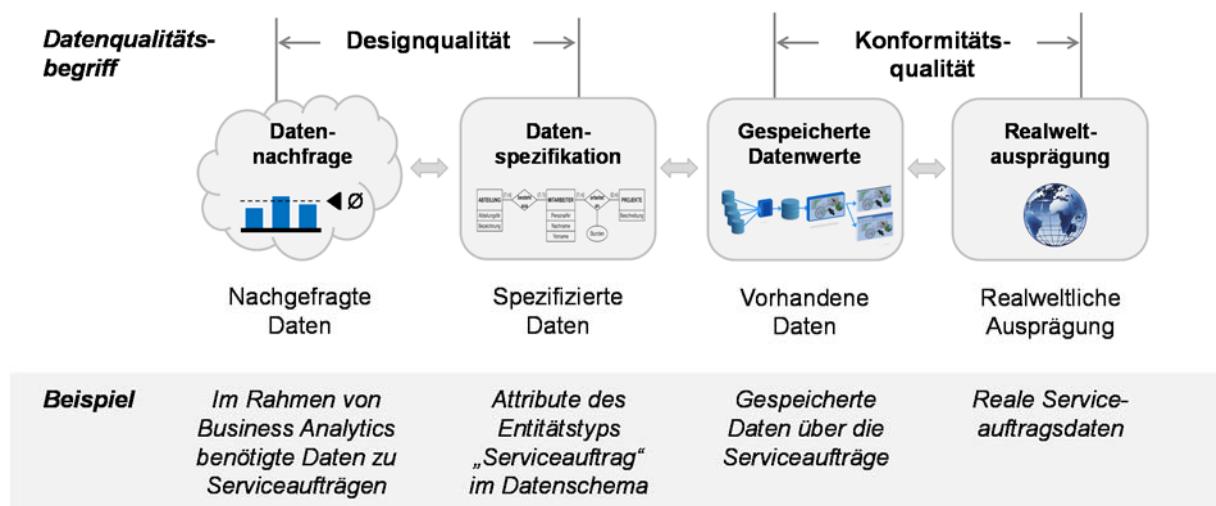


Abb. 2: Abgrenzung der beiden Datenqualitätsperspektiven

Unter Aktualität wird dabei die Eigenschaft der Gegenwartsbezogenheit eines Datenbestandes verstanden. Das heißt, inwiefern die gespeicherten Datenwerte nach wie vor den aktuellen Gegebenheiten in der Realwelt entsprechen und nicht bedingt durch zeitlichen Verfall veraltet

sind. Dem Datenqualitätsmerkmal Aktualität kommt in Wissenschaft und Praxis besondere Bedeutung zu. Zum einen analysiert die Aktualität den zeitlichen Verfall gespeicherter Datenwerte, was aus fachlicher und betriebswirtschaftlicher Perspektive hoch relevant ist. Zum anderen kommt der Aktualität gerade auch im Vergleich zum verwandten Datenqualitätsmerkmal Korrektheit eine besondere Relevanz zu, da sie eine weitgehend automatisierbare Messung zu geringerem Aufwand verspricht. So basiert die Messung der Aktualität im Gegensatz zur Korrektheit i. d. R. nicht auf einem Realweltabgleich in Form eines direkten Vergleichs jedes einzelnen gespeicherten Datenwerts mit seiner zugehörigen Realweltausprägung. Letzteres ist in der Unternehmenspraxis vielfach sehr aufwendig oder nicht praktikabel. Stattdessen wird mithilfe einer Schätzung ermittelt, inwiefern Datenwerte zum Bewertungszeitpunkt noch den realen Gegebenheiten entsprechen und nicht veraltet sind.

Damit adressiert das Datenqualitätsmerkmal Aktualität gerade auch die oben diskutierten Anforderungen im Business Analytics:

- Es wird – auch für komplexe Sachverhalte – eine Messung der Datenqualität unterstützt, die zu validen und für Menschen interpretierbaren Ergebnissen (im Sinne von Wahrscheinlichkeiten) führt und damit Plausibilitätsüberprüfungen ermöglicht.
- Die vorgeschlagene quantitative Metrik für Aktualität ist – und dies verdeutlichen die nachstehenden Ausführungen – in entscheidungstheoretische Modelle sinnvoll integrierbar.
- Die Metrik für Aktualität kann (teil-)automatisiert eingesetzt werden, um eine hohe Akzeptanz im Business Analytics zu ermöglichen.

3. Messung des Datenqualitätsmerkmals Aktualität mittels einer Metrik

Zur Messung der Aktualität von Daten wird im Folgenden eine wahrscheinlichkeitsbasierte Metrik vorgeschlagen (vgl. hierzu und im Weiteren *Heinrich et al., 2012*). Die Aktualität eines Datenwertes ω , bzw. das Ergebnis der Metrik, ist hierbei als Wahrscheinlichkeit dafür zu interpretieren, dass der Datenwert zum Bewertungszeitpunkt noch der zugehörigen Realweltausprägung entspricht und nicht veraltet ist.

In dieser Interpretation als Wahrscheinlichkeit liegt auch ein wesentlicher Vorteil der Metrik im Vergleich zu anderen Ansätzen, bei denen eine inhaltliche Interpretation der Metrikergebnisse nicht möglich bzw. nicht vorgesehen ist. So können die Ergebnisse der vorgeschlagenen Metrik methodisch fundiert in Erwartungswertkalküle von Entscheidungen

eingehen und damit entscheidungstheoretische Verfahren unterstützen. Zudem erlaubt die Metrik eine weitgehend automatisierbare Messung der Aktualität und kann datenattributspezifisch für den konkreten Anwendungsfall konfiguriert werden.

Ausdruck (1) zeigt die Basisdefinition der Metrik für Aktualität $Q_{Akt.}^\omega(t, \omega_1, \dots, \omega_n)$ für einen konkreten zu bewertenden Datenwert ω (z. B. den gespeicherten Status „offen“ für einen Serviceauftrag). Dabei repräsentiert der Parameter t das Alter des Datenwertes ω , das sich aus dem Zeitpunkt der Bewertung der Datenqualität und dem Zeitpunkt der Entstehung der zugehörigen Realweltausprägung ermitteln lässt (z. B. Zeit, die seit der Annahme des Serviceauftrags vergangen ist); die Parameter $\omega_1, \dots, \omega_n$ repräsentieren Zusatzdaten, die für die Messung der Aktualität des Datenwertes ω relevant sind, weil diese bspw. in Zusammenhang mit der unbekannten Gültigkeitsdauer T des Datenwertes ω stehen und somit Rückschlüsse auf diese zulassen (z. B. Priorität des Serviceauftrags, zuständiger Standort für den Serviceauftrag).

$$Q_{Akt.}^\omega(t, \omega_1, \dots, \omega_n) := P^\omega(T \geq t | W_1 = \omega_1, \dots, W_n = \omega_n) = 1 - F^\omega(t | \omega_1, \dots, \omega_n) \quad (1)$$

Die Metrik ist so definiert, dass die Ergebnisse $Q_{Akt.}^\omega(t, \omega_1, \dots, \omega_n)$ die Wahrscheinlichkeit dafür repräsentieren, dass der Datenwert ω zum Bewertungszeitpunkt noch den realen Gegebenheiten entspricht und nicht veraltet ist. Ausdruck (1) bildet dies mithilfe der bedingten Wahrscheinlichkeit $P^\omega(T \geq t | W_1 = \omega_1, \dots, W_n = \omega_n)$ ab, dass die Gültigkeitsdauer T größer oder gleich dem Alter t des Datenwertes ω ist. Die Zusatzdaten $\omega_1, \dots, \omega_n$ werden hierbei als Bedingung berücksichtigt. Auf Basis der zugehörigen Verteilungsfunktion $F^\omega(t | \omega_1, \dots, \omega_n)$ der Gültigkeitsdauer T des Datenwertes ω , ergibt sich das Metrikergebnis zu $1 - F^\omega(t | \omega_1, \dots, \omega_n)$.

Ausdruck (1) wird anhand eines kurzen Beispiels verdeutlicht: In einer Kundendatenbank wurde vor acht Semestern der Datenwert „Student“ als Berufsstatus einer Person, die sich zu diesem Zeitpunkt für ein Studium der Ingenieurwissenschaften an einer Universität immatrikuliert hat, gespeichert. Da die unbekannte Studiendauer und damit die Gültigkeitsdauer T des Datenwerts „Student“ von der Studienfachgruppe und der Hochschulart (Universität versus Fachhochschule) abhängt, werden die entsprechenden zur Person gespeicherten Daten bei der Messung der Aktualität als Zusatzdaten W_1 und W_2 berücksichtigt. Das Metrikergebnis $Q_{Akt.}^{\text{Student}}(8 \text{ Semester}, \text{Ingenieurwissenschaften}, \text{Univ.})$ ist als Wahrscheinlichkeit definiert, dass der Datenwert „Student“ noch die realen Gegebenheiten widerspiegelt. Dies entspricht der Wahrscheinlichkeit $P^{\text{Student}}(T \geq 8 \text{ Semester} | W_1 = \text{Ingenieurwissenschaften}, W_2 = \text{Univ.})$, dass die Studiendauer T

der Person größer oder gleich acht Semester ist. Dieser Wert wiederum lässt sich als $1 - F^{\text{"Student"}}(8 \text{ Semester} | \text{Ingenieurwissenschaften}, \text{Univ.})$ auf Basis der Verteilungsfunktion der Studiendauer ermitteln. Das linke Diagramm aus Abb. 3 illustriert die empirische Verteilungsfunktion (inkl. Abbrechern) für Studierende der Ingenieurwissenschaften an Universitäten. Das Metrikergebnis ergibt sich damit zu $1 - 0,30 = 0,70$. Das heißt, die Wahrscheinlichkeit, dass der Datenwert „Student“ noch nicht veraltet ist und die zugehörige Person nach wie vor studiert, beträgt 70%. Das rechte Diagramm aus Abb. 3 verdeutlicht zudem, wie sich das Metrikergebnis abhängig vom Alter t des Datenwerts „Student“ verändert.

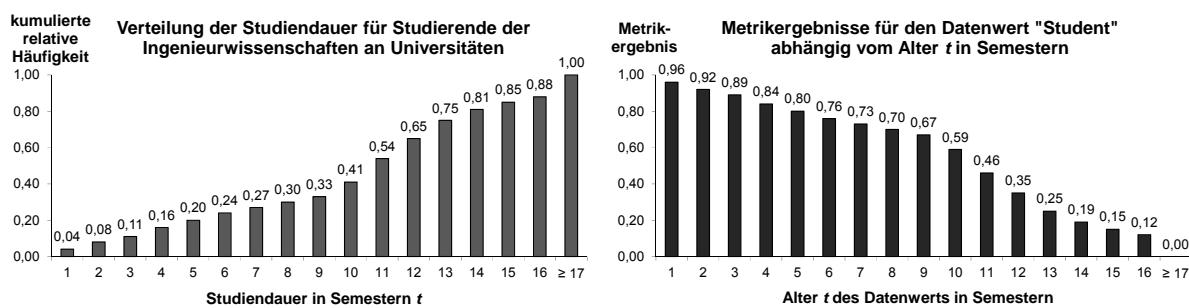


Abb. 3: Verteilung der Studiendauer sowie darauf basierende Metrikergebnisse im Beispiel

Für die Konfiguration und praktische Anwendung der Metrik ist es somit notwendig, die Verteilungsfunktion $F^{\omega}(t | \omega_1, \dots, \omega_n)$ zu ermitteln. Hierzu gibt es folgende Möglichkeiten:

- Analyse öffentlich zugänglicher Daten (z. B. des *Statistischen Bundesamts*)
- Analyse unternehmensinterner Daten (z. B. im Data Warehouse)
- Durchführung einer Studie (z. B. Befragung einer Stichprobe an Kunden)
- Heranziehen von Expertenschätzungen (z. B. für die Verfallrate von Daten)

Die Messung der Aktualität kann auf dieser Basis anhand obiger Metrik (vgl. Ausdruck (1)) und unter Einbeziehung der Metadaten bezüglich des Alters des entsprechenden Datenwertes ω i. d. R. automatisiert erfolgen (z. B. mittels *SQL DML Statements*).

4. Einsatz der Metrik bei einem großen deutschen Automobilhersteller

Im Weiteren werden am Fallbeispiel die Anwendung und der praktische Nutzen der Metrik verdeutlicht. Die Zahlen wurden aus Vertraulichkeitsgründen verändert und anonymisiert, wobei die grundsätzlichen Ergebnisse erhalten bleiben.

Beim betrachteten Unternehmen handelt es sich um einen großen deutschen Automobilhersteller. Interne Analysen hatten nicht nur Probleme bei der Bearbeitung von Serviceaufträgen, sondern auch bei der zugrunde liegenden Datenbasis verdeutlicht. Deshalb galt es – unter besonderer Berücksichtigung von Datenqualitätsaspekten – das Management von Ersatzteilaufträgen der angebundenen Händler durch eine fein granulare Steuerung zu verbessern. Die Ausführung der Ersatzteilaufträge wird dabei durch ein global eingesetztes, SAP-basiertes Supply Chain Management-System unterstützt. In diesem System werden monatlich mehr als 350.000 neue Ersatzteilaufträge einzelner Händler bearbeitet und gespeichert. Problematisch ist in diesem Zusammenhang, wenn bspw. der im System vorgehaltene Orderstatus veraltet ist. Um Probleme in den Serviceprozessen (z. B. Ersatzteil wird wegen veraltetem Status fälschlicherweise mehrmalig produziert oder bestellt) und der datengetriebenen Entscheidungsunterstützung (z. B. Kenngrößen zur Einhaltung von Service- und Lieferversprechen gegenüber Kunden wie großen Fuhrparkbetreibern) zu vermeiden, ist es unabdingbar, die Aktualität der gespeicherten Datenwerte zu messen und sicherzustellen. Im Folgenden wird die Konfiguration der Metrik am Beispiel der Ausprägung „offen“ des Attributs „Orderstatus“ illustriert.

Der Orderstatus „offen“ eines Ersatzteilauftrags kann seine Gültigkeit auf zwei Arten verlieren. Entweder der Auftrag wird erfolgreich abgeschlossen oder gestrichen. Beide Aspekte müssen bei der Ermittlung der Verteilungsfunktion $F^{\omega}(t | \omega_1, \dots, \omega_n)$ berücksichtigt werden, wobei auch relevante Zusatzdaten $\omega_1, \dots, \omega_n$ zu berücksichtigen sind. So haben Analysen ergeben, dass die Gültigkeitsdauer des Orderstatus signifikant von den im System vorgehaltenen Daten zu „Priorität“ („Eilbestellung“ vs. „Lagerbestellung“), „Auftragnehmer“ („Zentrale“ vs. „Händlerzentrum“), „Land des Auftragnehmers“ (Zentrale oder Händlerzentrum in Deutschland vs. Händlerzentrum in einem anderen Land) und „Land des Auftraggebers/Händlers“ abhängig ist. Die Verteilungsfunktion $F^{\omega}(t | \omega_1, \dots, \omega_n)$ konnte dabei auf Basis zuvor qualitätsgesicherter historischer Daten abgeleitet werden. Abb. 4 zeigt exemplarisch die kumulierten relativen Häufigkeiten für an die Zentrale gerichtete Eilbestellungen von Händlern in Deutschland bzw. Spanien.

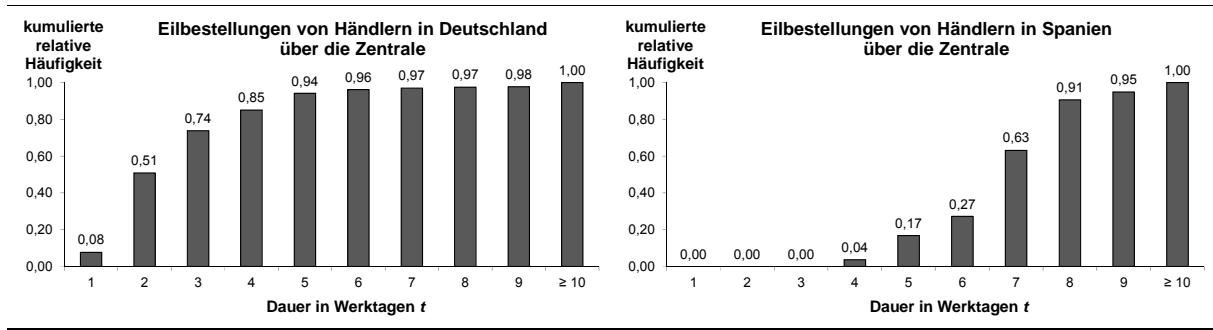


Abb. 4: Empirische Verteilungsfunktionen zur Metrikkonfiguration – Vergleich Eilbestellungen von Händlern in Deutschland versus Spanien

Mithilfe der ermittelten empirischen Verteilungsfunktion $F^\omega(t | \omega_1, \dots, \omega_n)$ können basierend auf Ausdruck (1) datenwertspezifisch in Abhängigkeit des Alters t des Datenwertes ω und der zugehörigen Zusatzdaten $\omega_1, \dots, \omega_n$ die Metrikergebnisse $Q_{Akt.}^\omega(t, \omega_1, \dots, \omega_n)$ berechnet werden. So ergibt sich bspw. für eine Eilbestellung eines Händlers in Deutschland, die über die Zentrale abgewickelt werden soll und vor fünf Werktagen angelegt wurde (d. h. $t = 5$), ein Wert von $1 - 0,94 = 0,06$ (vgl. Abb. 4, linkes Diagramm). Das heißt, die Wahrscheinlichkeit, dass der entsprechende Auftrag tatsächlich noch offen ist, beträgt lediglich 6%. Demgegenüber liefert die Metrik für den Fall, dass ein entsprechender Auftrag von einem Händler in Spanien erfolgte, einen Wert von $1 - 0,17 = 0,83$ (vgl. Abb. 4, rechtes Diagramm). Die Wahrscheinlichkeit ist in Spanien demnach fast 14-mal höher.

Abb. 5 zeigt die Metrikergebnisse für ausgewählte Datensätze mit $t = 5$. Hierbei wird deutlich, welche wichtige Rolle Zusatzdaten spielen.

ID	Order-status	Priorität	Auftragnehmertyp	Land des Auftragnehmers	Land des Händlers	Alter [Werktagen]	Metrikergebnis
A	offen	Eil-bestellung	Zentrale	Deutschland	Deutschland	5	0,06
B	offen	Lager-bestellung	Zentrale	Deutschland	Deutschland	5	0,20
C	offen	Eil-bestellung	Händlerzentrum	Deutschland	Deutschland	5	0,00
D	offen	Eil-bestellung	Zentrale	Deutschland	Spanien	5	0,83
E	offen	Lager-bestellung	Zentrale	Deutschland	Spanien	5	0,87
F	offen	Eil-bestellung	Händlerzentrum	Spanien	Spanien	5	0,05

Abb. 5: Ausgewählte Datensätze und zugehörige Metrikergebnisse

Auf Basis der Verteilungsfunktion können die Metrikergebnisse automatisiert berechnet werden. Das Alter der Datenwerte geht hierbei jeweils als Differenz aus dem Bewertungszeitpunkt und dem individuellen Zeitpunkt der Annahme des Auftrags ein; die Zusatzdaten sind ohnehin direkt im System gespeichert. Technisch wurde die Metrik in einer Advanced Business Application Programming (ABAP)/4-Komponente „Data Quality Checker“ umgesetzt und in das *SAP*-basierte Supply Chain Management-System integriert.

Die Einsatzszenarien für die Metrik sind vielfältig. Im Berichtswesen können etablierte Kennzahlen wie die durchschnittliche Bearbeitungszeit von Serviceaufträgen mit einer Qualitätskennzahl versehen werden, um Entscheidungen besser zu unterstützen (für ein Beispiel aus dem Kundenwertcontrolling vgl. Heinrich/Klier, 2009). Zudem führt die Berücksichtigung der Datenqualität in den Datenanalysemethoden im Business Analytics zu validen Ergebnissen und einer verbesserten Entscheidungsqualität. Im Folgenden wird diese fein granulare Steuerung für den Kontext des Managements von Ersatzteilaufträgen detaillierter beleuchtet. Konkret führen inaktuelle Daten zum Status von Ersatzteilaufträgen einerseits dazu, dass die Rechnungsstellung nicht oder verspätet erfolgt. Andererseits werden entsprechende Ersatzteillieferungen nicht oder zeitlich verzögert durchgeführt, was zu Ineffizienzen in den Serviceprozessen und letztendlich zu Unzufriedenheit der Kunden führt. Abb. 6 zeigt exemplarisch für einen Zeitraum von drei Wochen die Metrikergebnisse für insgesamt 476.999 Ersatzteilaufträge, die noch als „offen“ gespeichert sind. Dabei wurden die Aufträge jeweils in Abhängigkeit vom Metriergebnis zu Gruppen zusammengefasst. Das heißt, jeder Auftrag wurde einem der Intervalle $[0; 0,1]$, $]0,1; 0,2]$, ..., $]0,9; 1,0]$ zugeordnet und die relativen Häufigkeiten der Gruppenzuordnungen wurden ausgewiesen.

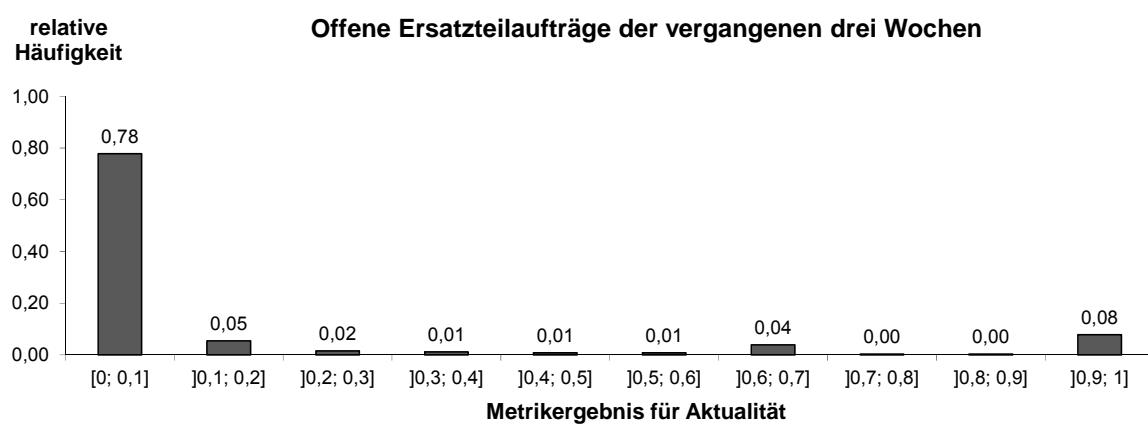


Abb. 6: Metrikergebnisse für im System noch als „offen“ gespeicherte Ersatzteilaufträge

Die Ergebnisse sind alarmierend: Für 78% der 476.999 Ersatzteilaufträge ist das Metrikergebnis kleiner oder gleich 0,1. Das heißt, die Wahrscheinlichkeit dafür, dass die Ersatzteilaufträge tatsächlich noch nicht abgeschlossen oder storniert wurden, beträgt jeweils höchstens 10%, obwohl diese als “offen” gespeichert sind. Lediglich für 8% der Datenwerte ist das Metrikergebnis hoch.

Auf Basis dieser Analyse wurden die Aufträge mit einem Metrikergebnis kleiner oder gleich 0,1 zusammen mit den Fachabteilungen per Stichprobe genauer untersucht. Hier wurden zwei Kategorien unterschieden: (1) Ersatzteilaufträge, für die der gespeicherte Orderstatus „offen“ in der Tat noch aktuell ist und (2) Ersatzteilaufträge, für die der gespeicherte Orderstatus „offen“ nicht mehr aktuell ist. Im Ergebnis wurde festgestellt, dass nur 6,6% der Ersatzteilaufträge auf Kategorie (1) entfallen. Insofern fällt der Anteil der Aufträge mit einem Metrikergebnis kleiner oder gleich 0,1, die tatsächlich noch offen waren (Kategorie (1)), genau in das Intervall [0; 0,1] für die ermittelten Metrikergebnisse. Dies unterstreicht die Validität der Metrik und ihre Interpretation als Wahrscheinlichkeit.

Mit 93,4% ist der Großteil der Aufträge Kategorie (2) zuzurechnen und spiegelt Datenqualitätsprobleme wider. Als Ursachen für veraltete Datenwerte zum Orderstatus wurden insbesondere Systemfehler, technische Probleme und Erfassungsfehler identifiziert. Die systembedingten Fehler gingen primär auf fehlende oder falsche Statusupdates in Folge fehlerhafter Schnittstellen zwischen dem Supply Chain Management-System und noch nicht abgeschalteten Altsystemen zurück. Temporäre Serverausfälle und Unterbrechungen der Verbindungen zu den Partnersystemen sind exemplarisch für technische Probleme; Fehler bei der Datenerfassung kommen primär bei einem Wechsel des Verantwortungsbereichs zustande.

Neben der Behebung der identifizierten Ursachen einer unverhältnismäßig verzögerten Bearbeitung der Aufträge (vgl. Kategorie (1)) sowie der Datenqualitätsprobleme (vgl. Kategorie (2)) wurde ein kennzahlenbasiertes Monitoring zur Verbesserung des Managements der Serviceaufträge etabliert. So soll eine fein granulare Steuerung auf Basis der Metrik künftig dazu beitragen, Aufträge effizient und verlässlich abzuwickeln. Konkret sollen die Metrikergebnisse nicht nur im Berichtswesen berücksichtigt werden, um Transparenz bezüglich der Qualität der Daten zu schaffen. Vielmehr sollen die Metrikergebnisse eingesetzt werden, um ein fortlaufendes Monitoring durchzuführen und Aufträge auszusteuern, deren Status mit einer Wahrscheinlichkeit von mindestens 80% nicht mehr aktuell ist. So sollen “red flagged” Aufträge identifiziert und adressiert werden, um Folgen schlechter Datenqualität zu vermeiden.

5. Fazit

Unternehmen stehen im Zeitalter der Digitalisierung große, multi-dimensionale Datenmengen („Big Data“) zur Verfügung, die im Business Analytics genutzt werden können. Speziell im Controlling können diese eine verbesserte, datengetriebene Entscheidungsunterstützung ermöglichen. Dabei gilt es zu beachten, dass die im Business Analytics abgeleiteten Ergebnisse und die darauf basierenden Entscheidungen nur dann valide und wertstiftend sind, wenn auch die Qualität der zugrundeliegenden Daten gewährleistet ist. Vor diesem Hintergrund wurde im Beitrag die Fragestellung aufgegriffen, wie Datenqualität fundiert gemessen und berücksichtigt werden kann, um die Entscheidungsunterstützung zu verbessern. Mithilfe der vorgeschlagenen Metrik kann die Aktualität von Datenwerten zielgerichtet und weitgehend automatisiert gemessen werden. Zudem können die Metrikergebnisse aufgrund ihrer Interpretation als Wahrscheinlichkeit methodisch fundiert in Erwartungswertkalküle und entscheidungstheoretische Verfahren eingehen. Die praktischen Einsatzmöglichkeiten der Metrikergebnisse sind vielfältig und reichen von der Anreicherung etablierter Kennzahlen im Berichtswesen um Qualitätsinformationen bis hin zur Integration in Datenanalysemethoden und Prognosemodelle zur Verbesserung der Entscheidungsqualität. Die Anwendung und der praktische Nutzen der Metrik wurden am Beispiel eines großen deutschen Automobilherstellers verdeutlicht. Hier wurde die Datenqualitätsmetrik erfolgreich eingesetzt, um das Management von Serviceaufträgen zu verbessern und eine feingranulare Steuerung zu ermöglichen. Dennoch gilt es kritisch festzuhalten, dass neben Aktualität noch weitere Datenqualitätsmerkmale existieren. Auch wenn Aktualität aus fachlicher und betriebswirtschaftlicher Perspektive zentrale Bedeutung besitzt, entspricht dies dennoch einer Partialansicht. Deshalb gilt es in Wissenschaft und Praxis, auch für weitere Datenqualitätsmerkmale wie Vollständigkeit, Korrektheit und Konsistenz geeignete Metriken zu entwickeln und einzusetzen. Dies ist umso herausfordernder, da im Business Analytics spezielle Anforderungen gelten: nämlich die automatisierte Messung der Datenqualität in sehr komplexen Sachverhalten mittels quantitativer Modelle, die ihrerseits wieder in entscheidungsorientierte Kalküle integrierbar sind. Gleichwohl ist die Messung und Sicherstellung der Qualität der zugrundeliegenden Daten unabdingbar, damit die zweifellos wertvollen Verfahren des Business Analytics auch Akzeptanz in der Praxis finden.

Literatur

Davenport, T. H./Harris, J. G., Competing on analytics: The new science of winning, Boston 2007.

Heinrich, B./Klier, M., Die Messung der Datenqualität im Controlling, in: Zeitschrift für Controlling & Management, 53 Jg. (2009), H. 1, S. 34-42.

Heinrich, B./Klier, M./Görz, Q., Ein metrikbasierter Ansatz zur Messung der Aktualität von Daten in Informationssystemen, in: Zeitschrift für Betriebswirtschaft, 82. Jg. (2012), H. 11, S. 1193-1228.

Heinrich, B./Helfert, M., Nützt Datenqualität wirklich im CRM? – Wirkungszusammenhänge und Implikationen, in: Uhr, W./Esswein, W./Schoop, E. (Hrsg.), Wirtschaftsinformatik 2003/Band II: Medien – Märkte – Mobilität, Heidelberg 2003, S. 231–253.

IBM Institute for Business Value, Analytics: Big Data in der Praxis – Wie innovative Unternehmen ihre Datenbestände effektiv nutzen, <http://www-935.ibm.com/services/de/gbs/thoughtleadership/GBE03519-DEDE-00.pdf>, Stand: 2012.

Internationaler Controller Verein, Big Data – Potenzial für den Controller, Dream Car der Ideenwerkstatt im ICV 2014, https://www.icv-controlling.com/fileadmin/Assets/Content/AK/Ideenwerkstatt/Files/ICV_Ideenwerkstatt_DreamCar-Bericht_BigData.pdf, Stand: 2014.

Experian QAS, The Data Advantage: How Accuracy Creates Opportunity. London 2013.

Implikationen für die Praxis

- Ergebnisse und Entscheidungen im Business Analytics sind nur valide und wertstiftend, wenn auch die Qualität der zugrundeliegenden Daten gewährleistet ist.
- Die praktische Anwendung der vorgeschlagenen Datenqualitätsmetrik für Aktualität gewährleistet eine verbesserte, datengetriebene Entscheidungsunterstützung.
- Die Einsatzmöglichkeiten der Metrikergebnisse reichen von der Anreicherung von Kennzahlen im Berichtswesen um Qualitätsinformationen bis hin zur Integration in Datenanalysemethoden und Prognosemodelle zur Verbesserung der Entscheidungsqualität.