



Kernkompetenzzentrum
Finanz- & Informationsmanagement



Projektgruppe
Wirtschaftsinformatik

Diskussionspapier

Die Messung der Datenqualität im Controlling - Ein metrikbasierter Ansatz und seine Anwendung im Kundenwertcontrolling

von

Bernd Heinrich, Mathias Klier

in: Controlling & Management 53 (2009) 1, S. 34-42

WI-249

Universität Augsburg, D-86135 Augsburg
Besucher: Universitätsstr. 12, 86159 Augsburg
Telefon: +49 821 598-4801 (Fax: -4899)

Universität Bayreuth, D-95440 Bayreuth
Besucher: F.-v.-Schiller-Str. 2a, 95444 Bayreuth
Telefon: +49 921 55 - 4710 (Fax: - 844710)

www.fim-rc.de



Die Messung der Datenqualität im Controlling – ein metrikbasierter Ansatz und seine Anwendung im Kundenwertcontrolling

Von PD Dr. Bernd Heinrich und Dr. Mathias Klier

PD Dr. Bernd Heinrich; Dr. Mathias Klier; wissenschaftliche Mitarbeiter am Lehrstuhl für Betriebswirtschaftslehre, Wirtschaftsinformatik, Informations- & Finanzmanagement, Universität Augsburg, Universitätsstraße 16, 86135 Augsburg.

E-Mail: [{Bernd.Heinrich|Mathias.Klier} @wiwi.uni-augsburg.de](mailto:{Bernd.Heinrich|Mathias.Klier}@wiwi.uni-augsburg.de)

Telefonnummer: 0821/598-4137 bzw. 0821/598-4139 (Sekretariat)

Nutzwertkasten

- Im Controlling spielt der Einsatz moderner Anwendungssysteme zur Erhebung, Komprimierung und Aufbereitung von betriebswirtschaftlich relevanten Steuerungsdaten eine wichtige Rolle.
- Der Nutzen einer Versorgung von Entscheidungsträgern mit Kenngrößen steigt und fällt mit der Qualität der verarbeiteten Daten zur Ermittlung der Kenngrößen.
- Obwohl die Sicherstellung der Datenqualität für viele Unternehmen ein Problem darstellt, haben sich bisher keine Ansätze zur Messung der Datenqualität durchgesetzt.
- Im Beitrag werden zwei konkrete, wissenschaftlich fundierte Metriken für die Datenqualitätsmerkmale Vollständigkeit und Aktualität vorgeschlagen, die zu interpretierbaren Ergebnissen führen und in der Praxis anwendbar sind.
- Die Konstruktion der Datenqualitätsmetriken sowie deren Anwendung im Kundenwertcontrolling werden am Fallbeispiel eines Finanzdienstleisters illustriert.
- Auch bei anderen Unternehmen – z. B. aus der Mobilfunkbranche – konnten die vorgeschlagenen Metriken bereits erfolgreich angewendet werden.

Die Messung der Datenqualität im Controlling – ein metrikbasierter Ansatz und seine Anwendung im Kundenwertcontrolling

Einleitung

Eine Kernaufgabe im Controlling ist der Aufbau und der Betrieb eines bereichsinternen und -übergreifenden Berichtswesens zur Steuerung des Unternehmens (vgl. Eschenbach 1996, S. 91; Jung 2007, S. 12f.). Gerade beim Betrieb des Berichtswesens spielt – als eine logische Konsequenz – die regelmäßige Ermittlung betriebswirtschaftlich relevanter Steuerungsdaten in strukturierter und nachvollziehbarer Form eine wichtige Rolle. Der Einsatz moderner Anwendungssysteme zur Erhebung, Komprimierung und Aufbereitung dieser Daten ist dabei in der Diskussion unstrittig (vgl. Samtleben/Stadlbauer/Hess 2006, S. 86; Horváth 2006, S. 659). Als Beispiele hierfür sind Data-Warehouse-Systeme oder IT-gestützte Management-Cockpits zu nennen. Hintergrund für deren Einsatz ist nicht nur die Fülle zu verarbeitender Daten, sondern auch der Umstand, dass ein wesentlicher Teil der Berichtsdaten aus quantitativen Kenngrößen besteht. Deren IT-gestützte Bestimmung, d. h. die automatisier- und reproduzierbare Aggregation und Verrechnung vieler abteilungs-, prozess- oder projektspezifischer Einzelwerte, wird durch Softwareanwendungen i. d. R. erleichtert oder erst ermöglicht. Demzufolge kommt der Entwicklung und Nutzung von Anwendungssystemen im Controlling eine nicht zu unterschätzende Bedeutung in der Diskussion eines funktionierenden Berichtswesens zu.

Neben der softwaretechnischen Unterstützung ist auch die Qualität der Inputdaten (gespeichert in Datenbanken) zur Ermittlung von Kenngrößen ein wichtiges Thema. Dies ist wenig überraschend, da der Nutzen einer Versorgung von Entscheidungsträgern mit Kenngrößen mit der Vollständigkeit, Korrektheit, Konsistenz oder Aktualität der eingehenden Inputdaten steigt bzw. fällt (vgl. Fisher/Chengalur-Smith/Ballou 2003, S. 170; Ballou/Tayi 1999, S. 73; Al-Hakim 2007, S. 172; Even/Shankaranarayanan 2007, S. 75). Letztgenannte Eigenschaften sind als Datenqualitätsmerkmale bekannt (vgl. z. B. Wang/Storey/Firth 1995, S. 632; Rohwedder/Kasten/Malzahn/Piro/Schmid 2008, S. 28ff.). Für viele Unternehmen stellt dabei die Sicherstellung der Datenqualität (DQ) ein grundsätzliches Problem dar (vgl. Helfert 2000, S. 13; Ballou/Wang/Pazer/Tayi 1998). So sind beispielsweise durchschnittlich zwischen 15% und 20% der Datenwerte in typischen Kundendatenbanken fehlerhaft (English 1999, S. 9). Eine andere Studie des Data Warehousing Institute besagt, dass 53% der Unternehmen Mängel bei der DQ als Grund für Verluste, Probleme oder zusätzliche Kosten anführen. Darüber hinaus geben 75% der Befragten einer internationalen Studie zum Thema DQ an, bereits we-

sentliche Entscheidungen aufgrund fehlerhafter Daten falsch getroffen zu haben (Harris Interactive 2006). Deshalb wenden sie bis zu 30% ihrer Arbeitszeit auf, um Daten auf deren Qualität hin zu überprüfen (Harris Interactive 2006). Vor diesem Hintergrund ist es durchaus verwunderlich, dass sich bisher in Wissenschaft und Praxis noch keine Ansätze zur Messung der DQ durchgesetzt haben, um dem Thema DQ – insbesondere auch im Berichtswesen – in ausreichendem Maße Rechnung tragen zu können.

Zusammengefasst ist somit die Frage zu beantworten, wie DQ adäquat gemessen und wie die Ergebnisse dieser Messung sowie darauf basierende DQ-Maßnahmen im Berichtswesen berücksichtigt werden können. Dadurch soll transparent gemacht werden, welche Qualität die einer Kenngröße zugrunde liegenden Daten aufweisen und wie belastbar die Kenngröße demnach selbst ist. Dies soll die Entscheidungsunterstützung des Managements verbessern.

Der Beitrag ist wie folgt aufgebaut: Im nächsten Kapitel werden das zugrunde gelegte Qualitätsverständnis und die hier betrachteten DQ-Merkmale Vollständigkeit und Aktualität dargestellt. Letztgenannte haben – wie sich später herausstellen wird – besondere Bedeutung, wenn es darum geht, die Qualität der einer Kenngröße zugrunde liegenden Inputdaten zu messen. Im darauf folgenden Kapitel werden Metriken zur Messung dieser DQ-Merkmale vorgeschlagen und diskutiert. Danach illustriert ein Fallbeispiel aus der Finanzdienstleistungsbranche die Anwendbarkeit und den praktischen Mehrwert der Metriken, bevor das letzte Kapitel die Ergebnisse zusammenfasst.

Datenqualitätsverständnis und Datenqualitätsmerkmale

Um zu verdeutlichen, welche Art von Datenqualität hier gemessen werden soll, ist es in einem ersten Schritt notwendig, das zugrunde liegende Datenqualitätsverständnis zu erläutern. Grundsätzlich werden im Qualitätsmanagement mit **Design- und Konformitätsqualität** zwei Perspektiven unterschieden (vgl. z. B. Helfert/Heinrich 2003; Juran 1998; Teboul 1991). Dabei bezeichnet die **Designqualität** den Grad der Übereinstimmung zwischen den Anforderungen der Datenanwender und der entsprechenden Repräsentation in Form einer Spezifikation (bspw. zu berechnende Kenngröße oder Datenschemata einer Datenbank). Darauf abzielende Ansätze der Bedarfsanalyse betrachten bspw., welche Kenngrößen in einen Bericht zweckorientiert eingehen sollen, ob deren eindeutige Interpretation bzw. Auslegbarkeit beim Management gewährleistet ist und wie diese dargestellt werden sollen, um ein schnelles Verständnis der Aussage der Kenngrößen zu ermöglichen.

Die **Konformitätsqualität**, die hier im Vordergrund steht, drückt dagegen aus, in welchem Ausmaß diese Spezifikation durch die in der Datenbank tatsächlich vorhandenen Daten ein-

gehalten wird. Ist demnach eine Kenngröße als Teil eines Berichts definiert worden, betrachtet die Konformitätsqualität, ob die zur Berechnung der Kenngröße notwendigen Inputdatenwerte bspw. überhaupt vollständig und aktuell vorliegen. Finden nämlich z. B. an Stelle von fehlenden oder veralteten Daten Schätzwerte (bspw. Durchschnittswerte) bei der Ermittlung von Kenngrößen – wie dem Kundenwert oder von Prozesskosten – Verwendung und wird dies in den Berichten an das Management nicht transparent gemacht, so kann dadurch die Entscheidungsunterstützung wesentlich beeinträchtigt werden.

Im Hinblick auf die Messung der DQ ist die Unterscheidung in Design- und Konformitätsqualität auch deshalb wichtig, da die größtenteils subjektive Einschätzung der Übereinstimmung des Datenbedarfs des Anwenders mit der spezifizierten Datenmenge (Designqualität) von der Analyse der Übereinstimmung von spezifiziertem und tatsächlich vorhandenem Datenangebot (Konformitätsqualität) getrennt wird. Letztgenannte Analyse soll dabei nachvollziehbar und reproduzierbar sein.

Auf Basis dieser Unterscheidung lassen sich im Berichtswesen unterschiedliche DQ-Merkmale grundsätzlich voneinander abgrenzen und deren jeweilige Ansatzpunkte für eine verbesserte Entscheidungsunterstützung verdeutlichen (vgl. Abbildung 1).

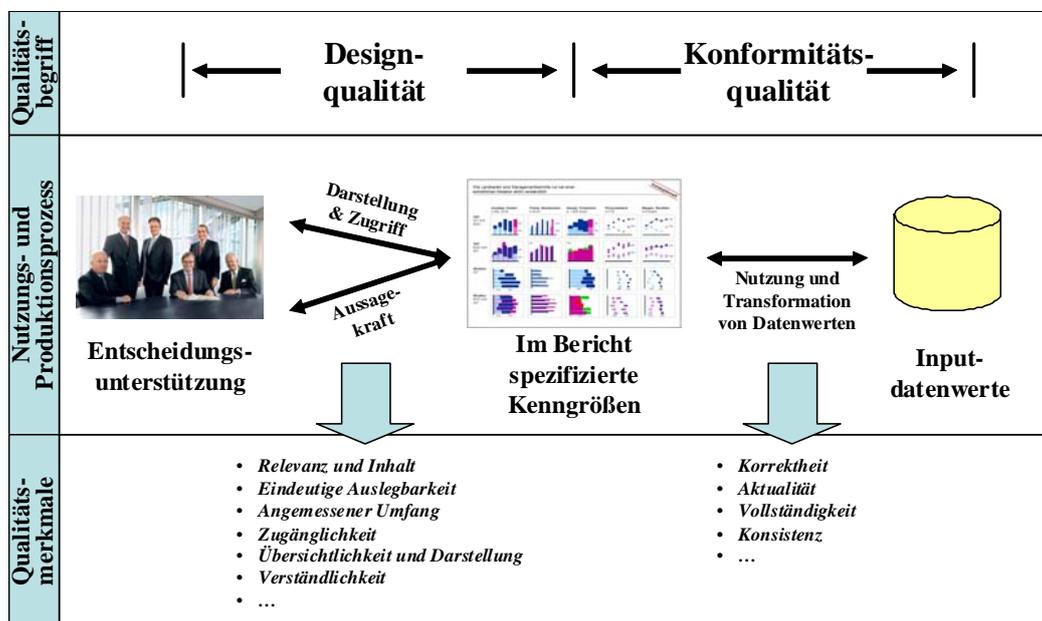


Abbildung 1: Abgrenzung unterschiedlicher DQ-Begriffe und -Merkmale

Der Bereich der Designqualität wird in einer Vielzahl von Veröffentlichungen adressiert. Folglich existieren dazu auch fundierte bzw. praxiserprobte Ansätze zur Messung der zugehörigen DQ-Merkmale. Hier sind bspw. die am Massachusetts Institute of Technology (MIT) entwickelte AIM Quality (AIMQ)-Methode (vgl. z. B. Lee/Strong/Kahn/Wang 2002),

der prozessorientierte Ansatz von Redman (vgl. Redman 1996) oder die Goal-Question-Metric-Method (vgl. Bobrowski/Marre/Yankelevich 1999) zu nennen. Auch werden im Zusammenhang mit der Aufbereitung von Daten in einem Management-Cockpit oftmals eine übersichtliche Darstellung und eine möglichst zeitnahe Bereitstellung der Berichte und damit der Kenngrößen thematisiert. Im Gegensatz dazu werden der Bereich der Konformitätsqualität und gerade deren Messung mittels Metriken noch relativ wenig problematisiert. Deswegen betrachtet der vorliegende Beitrag die DQ-Merkmale Vollständigkeit und Aktualität und zeigt den praktischen Nutzen von deren Messung im Rahmen der Entscheidungsunterstützung auf.

Datenqualitätsmetriken für Vollständigkeit und Aktualität

Im Folgenden werden Metriken für die DQ-Merkmale Vollständigkeit und Aktualität vorgestellt.

Metrik für das Datenqualitätsmerkmal Vollständigkeit

Unter Vollständigkeit wird hier die Eigenschaft verstanden, dass die Inputdatenwerte in der Datenbank mit Werten belegt sind, die (semantisch) vom Wert *NULL* abweichen. *NULL* ist dabei kein erforderlicher oder definierter Datenwert, sondern lediglich ein Platzhalter für die Nichtbefüllung. Die Metrik für Vollständigkeit $Q_{Vollst.}(w)$ wird wie folgt definiert, wobei w einen Datenwert in der Datenbank symbolisiert:

Fehler! Es ist nicht möglich, durch die Bearbeitung von Feldfunktionen Objekte zu erstellen.

Die Qualität eines Datenwerts wird also hinsichtlich Vollständigkeit mit dem Minimalwert von null bewertet, falls das entsprechende Datenattribut nicht befüllt ist oder einen zu *NULL* (semantisch) äquivalenten (Default-)Wert enthält (z. B. Dummy-Wert). Ansonsten ergibt sich der Wert der Metrik zu eins.

Probleme bei der Metrik können dann auftreten, wenn ein Datenwert nicht aus Mangel an verfügbaren Daten mit *NULL* belegt ist, sondern, weil der entsprechende Wert tatsächlich nicht existiert (z. B. Name eines stellvertretenden Projektleiters, falls für kleine Projekte kein Stellvertreter vorgesehen ist). In diesem Fall wäre das entsprechende Attribut mit dem Wert *NULL* in der Tat richtig belegt und die Bewertung hinsichtlich Vollständigkeit müsste den Wert eins und nicht den Wert null liefern. Schwierigkeiten dieser Art können umgangen werden, indem Indikatoren dafür eingeführt werden, dass der entsprechende Wert tatsächlich nicht existiert. So kann bspw. das Attribut *Name des stellvertretenden Projektleiters* (automatisiert) mit „nicht erforderlich“ belegt werden, falls das Projekt aufgrund seines Volumens als

klein eingestuft wird. Somit ist der Datenwert ungleich *NULL* und die obige Metrik liefert den korrekten Wert eins. Vor der ersten Anwendung der Metrik müssen die Daten somit hinsichtlich der vorgestellten Problematik untersucht und ggf. auftretende Schwachstellen beseitigt werden. Hierbei wäre bspw. an eine Vervollständigung des Datenbestands mit Indikatoren zu denken, sofern dies technisch sowie fachlich möglich und sinnvoll ist und keine Seiteneffekte auf Ergebnisse anderer Metriken besitzt.

In der Praxis ist es i. d. R. notwendig, die Metrikergebnisse für einzelne Datenwerte zu aggregieren. Hier ist bspw. daran zu denken, dass in die Berechnung eines Kundenwerts als Kenngröße eine Reihe einzelner Datenwerte, wie z. B. Alter, Berufsstatus und Fachrichtung eines Kunden eingehen. Um dabei eine Aussage über die Qualität (hier Vollständigkeit) der zugrunde liegenden Datenbasis zur Berechnung des Kundenwerts treffen zu können, ist eine Aggregation der Metrikergebnisse für die einzelnen Datenwerte erforderlich.

Beispielhaft wird dies im Folgenden – basierend auf den obigen Ausführungen – für die Aggregation mehrerer Metrikergebnisse formuliert. Sei hierbei T ein Tupel mit den Datenwerten $T.A_1, \dots, T.A_{|A|}$ für die Datenattribute $A_1, \dots, A_{|A|}$ und $g_i \in [0; 1]$ die relative Wichtigkeit von A_i in Bezug auf Vollständigkeit. Dann ergibt sich unter Verwendung der obigen Metrik für die Datenwertebene die Metrik für die Aggregation als gewichtetes arithmetisches Mittel:

Fehler! Es ist nicht möglich, durch die Bearbeitung von Feldfunktionen Objekte zu erstellen.

Die Vollständigkeit eines Tupels wird demnach basierend auf der Vollständigkeit der enthaltenen Datenwerte berechnet. Dabei ist es möglich, diese je nach Anwendungskontext mit Gewichtungen g_i zu versehen, da die einzelnen Datenwerte hier von unterschiedlicher Bedeutung sein können. So sind je nach verwendetem Ansatz zur Kundenbewertung z. B. für dessen Berechnung Attribute wie das *Alter* besonders relevant, wohingegen bspw. das *Geschlecht* von untergeordneter Bedeutung ist.

Ist eine weitere Aggregation über Tupel hinweg durchzuführen – bspw. um die Vollständigkeit der Datenbasis zur Berechnung des Werts des gesamten Kundenstamms zu ermitteln – so kann dies grundsätzlich analog zu oben erfolgen. Hierbei ist wiederum zu entscheiden, ob die einzelnen Tupel unterschiedlich gewichtet werden oder ob diese (sinnvoller Weise) ungewichtet bleiben sollen. Die Messung der Vollständigkeit mit Hilfe der Metrik kann i. d. R. mittels entsprechender SQL-Abfragen und bei Bedarf für den kompletten Datenbestand (bspw. Kundenbestand) erfolgen. Ist die Metrik für das DQ-Merkmal Vollständigkeit noch

vergleichsweise einfach, so wird im nächsten Abschnitt eine Metrik für das DQ-Merkmal Aktualität vorgestellt, die auf wahrscheinlichkeitstheoretischen Überlegungen basiert.

Metrik für das Datenqualitätsmerkmal Aktualität

Unter Aktualität wird hier die Eigenschaft der Gegenwartsbezogenheit eines Inputdatenwerts (für eine Kenngröße) verstanden, d. h., inwiefern die in der Datenbank erfassten Werte den aktuellen, tatsächlichen Gegebenheiten entsprechenden und nicht veraltet sind. Die Überprüfung basiert dabei – im Gegensatz zum DQ-Merkmal Korrektheit – auf einer Schätzung und nicht auf einer tatsächlichen Überprüfung. Letzteres wäre in vielen Fällen gar nicht durchführbar oder zu kostenintensiv (bspw. wäre eine Überprüfung aller gespeicherten Kundendaten vor der Kundenwertberechnung unrealistisch). Deshalb wird eine wahrscheinlichkeitstheoretische Betrachtung zugrunde gelegt (vgl. hierzu im Detail Heinrich/Kaiser/Klier 2007a, Heinrich/Kaiser/Klier 2007b). Aktualität kann dabei als jene Wahrscheinlichkeit interpretiert werden, mit welcher die untersuchten Datenwerte noch aktuell sind. In dieser Interpretierbarkeit liegt auch der Vorteil der entwickelten Metrik im Vergleich zu existierenden Metriken, bei denen eine inhaltliche, eindeutige Interpretation der resultierenden Werte nicht möglich bzw. nicht vorgesehen ist. Die Metrik für Aktualität wird im Weiteren nur für Datenwerte vorgestellt, ist jedoch – analog zu oben – auch für eine Aggregation definiert.

Sei A ein Datenattribut (bspw. *Berufsstatus* zur Berechnung des Kundenwerts), w ein entsprechender Datenwert (bspw. „Student“) und $Alter(w,A)$ das Alter des Datenwertes, das sich aus dem Zeitpunkt der Messung und dem Zeitpunkt der Datenerfassung errechnen lässt. Des Weiteren sei $Verfall(A)$ die (ggf. empirisch ermittelte) Verfallsrate von Werten des Datenattributs A . Dann stellt sich die Aktualitätsmetrik für Datenwerte wie folgt dar:

Fehler! Es ist nicht möglich, durch die Bearbeitung von Feldfunktionen Objekte zu erstellen.

Unter der Annahme, dass die Gültigkeitsdauer der zugrunde liegenden Datenwerte exponentialverteilt mit dem Parameter $Verfall(A)$ ist, stellt der Wert $Q_{Akt.}(w,A)$ dabei die Wahrscheinlichkeit dar, mit welcher der vorliegende Datenwert w noch den aktuellen Gegebenheiten entspricht. Bei der Exponentialverteilung handelt es sich um eine typische Lebensdauerverteilung, die sich insbesondere im Rahmen der Qualitätssicherung bewährt hat (für andere Verteilungsannahmen vgl. Heinrich/Kaiser/Klier 2007a).

Bei Attributen wie z. B. *Geburtsdatum* oder *Geburtsort*, die sich in der Realwelt nicht ändern, gilt $Verfall(A)=0$ und die Metrik für Aktualität ergibt sich somit zu eins:

Fehler! Es ist nicht möglich, durch die Bearbeitung von Feldfunktionen Objekte zu erstellen.

Zudem wird die Aktualität von Datenwerten, die zum Betrachtungszeitpunkt neu erfasst werden – d. h. $Alter(w,A)=0$ – ebenfalls mit eins bewertet:

Fehler! Es ist nicht möglich, durch die Bearbeitung von Feldfunktionen Objekte zu erstellen.

Die erneute Erfassung eines Datenwertes wird dabei als Aktualisierung eines bereits vorhandenen Datenwertes verstanden.

Insgesamt ist festzuhalten, dass das Metrikergebnis und damit auch die DQ für ein bestimmtes, festes Alter umso geringer sind, je höher beim entsprechenden Attribut die Verfallsrate ist. Umgekehrt nimmt bei zunehmendem Alter die Wahrscheinlichkeit, dass der entsprechende Datenwert noch gültig ist, und somit das Metrikergebnis für die Aktualität ab.

Für die praktische Anwendung der Metrik ist es notwendig, für ein Datenattribut den Parameter $Verfall(A)$ der Wahrscheinlichkeitsverteilung festzulegen. Dieser ist als Verfallsrate zu verstehen und gibt an, welcher Datenanteil bezogen auf das jeweilige Attribut innerhalb einer Zeiteinheit inaktuell wird. Eine Verfallsrate von 0,2 drückt bspw. aus, dass im Laufe einer Periode von 100 Datenwerten des entsprechenden Attributs im Durchschnitt 20 Werte inaktuell werden. Dabei kann entweder auf Erfahrungswerte, statistische Werte (bspw. veröffentlichte Raten des Statistischen Bundesamts für Studienabgänger bzw. -abbrecher als Grundlage zur Schätzung der Verfallsrate des Werts „Student“ des Attributs *Berufsstatus*) zurückgegriffen oder mittels eigener Stichprobenuntersuchungen eine Schätzung vorgenommen werden. Die Messung erfolgt somit anhand obiger Metrik in Verbindung mit den Schätzern für die Verfallsparameter und den Metadaten bezüglich des Zeitpunktes der Datenerfassung.

Der nächste Abschnitt skizziert die praktische Anwendung der Metriken im Rahmen eines Fallbeispiels.

Praktischer Nutzen der Datenqualitätsmetriken

Im Weiteren wird anhand eines Fallbeispiels der praktische Nutzen der entwickelten DQ-Metriken für das Berichtswesen verdeutlicht, wobei mit dem Kundenwert eine Kenngröße im Vertriebsbereich und damit das Anwendungsgebiet des Kundenwertcontrollings fokussiert wird (Meyer/Shaffu 2007; Mengen/Mettler 2008). Dabei steht die Frage im Vordergrund, wie neben der Ermittlung der eigentlichen Kenngröße die Metriken für die Vollständigkeit und Aktualität verwendet werden können, um zu untersuchen, wie belastbar diese Kenngröße ist.

So ist es unmittelbar eingängig, dass ein hoher Kundenwert mit Vorsicht zu genießen ist, wenn die Inputdaten zu seiner Berechnung unvollständig oder mit hoher Wahrscheinlichkeit bereits veraltet sind. Die verwendeten Zahlen und Daten im Fallbeispiel wurden aus Vertraulichkeitsgründen verändert und anonymisiert, wobei die grundsätzlichen Ergebnisse erhalten blieben.

Das betrachtete Unternehmen versteht sich als unabhängiger Finanzdienstleister (FDL) mit dem Anspruch, seine Kunden umfassend und über einen Großteil des Lebenszyklus hinweg zu betreuen. Der Kundenwert dient dabei als Ausgangspunkt vieler Entscheidungen und hat damit eine zentrale Bedeutung.

Die Berechnung des Kundenwerts erfolgt beim FDL, der sich primär auf Akademiker spezialisiert hat, neben dem **Alter** vor allem auch unter Berücksichtigung des derzeitigen **Berufsstatus** (z. B. Student) sowie der **Fachrichtung** (z. B. Medizin) des Kunden (für detaillierte Ausführungen vgl. Heidemann/Klier 2008). Dies sind Attribute in einer Datenbank, die im Laufe der Kundenbeziehung erfasst und gespeichert werden. Allerdings sind die Kundendaten von sehr unterschiedlicher Aktualität und Vollständigkeit. Dies hat wiederum erheblichen Einfluss auf die Berechnung der Kundenwerte bzw. deren Belastbarkeit.

Anwendung der Metrik für das Datenqualitätsmerkmal Vollständigkeit

Ist in der Datenbank bspw. das Attribut *Fachrichtung* bei einem Kunden nicht mit einem Wert belegt, d. h. die Datenwerte des Kunden sind nicht vollständig, so wird im Rahmen der Kundenbewertung automatisiert auf einen je nach Anteil der einzelnen Fachrichtungen an der Zielgruppe gewichteten Durchschnittswert zurückgegriffen. Dies hat Implikationen auf den ermittelten Kundenwert. Tabelle 1 illustriert dies anhand von drei Beispielkunden:

Kunde	Alter	...	Berufsstatus	Fachrichtung	Kundenwert
1	27	...	Student	Medizin	4.500 €
2	27	...	Student	Architektur	2.000 €
3	27	...	Student		3.200 €

Tabelle 1: Ermittelter Kundenwert für beispielhafte Kunden

Bei der Kundenakquisition konzentriert sich der FDL insbesondere auf Studenten, um diese direkt beim Berufseintritt mit den relevanten Produkten versorgen zu können. Für die Kunden 1 und 2 wird aufgrund der Modellergebnisse aus (Heidemann/Klier 2008) ein Kundenwert von 4.500 € bzw. 2.000 € ausgewiesen. Dabei ist ersichtlich, dass der Kundenwert – bspw.

wegen den Unterschieden beim prognostizierten Einkommensverlauf und des damit einhergehenden unterschiedlichen Bedarfs an Finanzdienstleistungen – stark von der Fachrichtung des Kunden abhängt (hier Medizin vs. Architektur). Will man nunmehr für den Kunden 3, dessen Fachrichtung nicht bekannt ist, den Kundenwert berechnen, so wird beim FDL über alle Fachrichtungen ein gewichteter Durchschnittswert ermittelt (Basis: alle Kunden mit dem *Berufstatus* Student und dem *Alter* 27). Dieser liegt bei 3.200 € d. h. die fehlende Vollständigkeit der Kundendaten wird durch einen Schätzwert „ausgeglichen“. Dies kann deswegen problematisch sein, da ein Kundenwert von 3.200 € die Unvollständigkeit der Inputdatenwerte nicht transparent macht und somit ohne Weiteres aus dessen Verwendung falsche Entscheidungen resultieren können.

Um Entscheidern nicht nur den automatisiert berechneten Kundenwert zur Verfügung zu stellen – der ermittelte Wert für Kunde 3 ist aufgrund der mangelhaften Datengrundlage weniger belastbar –, sondern zugleich eine Qualitätsaussage zur Datenbasis mitzugeben, wird die oben dargestellte DQ-Metrik für Vollständigkeit herangezogen. Hierzu wurden einmalig alle Inputdaten, die zur Berechnung des Kundenwertes verwendet werden, hinsichtlich ihrer relativen Wichtigkeit in Bezug auf die Kundenwertermittlung gewichtet. Durch einen Vergleich, wie stark die berechneten Kundenwerte in Abhängigkeit von den einzelnen Ausprägungen eines Datenattributs streuen, konnten hierbei die Gewichte g_i für die Datenattribute festgelegt werden. Darauf basierend kann nun automatisiert mit Hilfe der Metrik für jeden Kunden neben dem Kundenwert auch die Vollständigkeit der entsprechenden Inputdaten berechnet und ausgewiesen werden. Zieht man vereinfachend nur die Datenattribute *Geburtsdatum* (zur Ermittlung des Kundenalters), *Berufstatus* und *Fachrichtung* heran und gewichtet diese im Verhältnis 2:1:1, so ergeben sich für das oben genannte Beispiel folgende Ergebnisse:

Kunde	Alter	...	Berufstatus	Fachrichtung	Kundenwert	Metrikergebnis
1	27	...	Student	Medizin	4.500 €	1,00
2	27	...	Student	Architektur	2.000 €	1,00
3	27	...	Student		3.200 €	0,75

Tabelle 2: Kundenwert und Metrikergebnis für Vollständigkeit für beispielhafte Kunden

Die so ermittelten Metrikergebnisse werden beim FDL wiederum (in aggregierter Form) im Berichtswesen berücksichtigt. So kann sich das Management bspw. nicht nur jederzeit über den Wert einzelner Kundensegmente (Aggregation der oben dargestellten Kundenwerte) informieren. Vielmehr wird im Management-Cockpit zugleich der Wert der Qualitätsmetrik für

Vollständigkeit angezeigt, um den Entscheidern einen Eindruck zu vermitteln, wie belastbar die ermittelte Zahl tatsächlich ist (waren bspw. für die Berechnung des Werts eines Kundensegments, der sich zu 200 Mio. € ergab, 90% oder lediglich 10% der erforderlichen Inputdaten vorhanden?). Solche Hintergrundinformationen sind dabei durchaus relevant für das Management und leisten einen Beitrag dazu, Entscheidungen auf belastbaren Kenngrößen zu treffen. Zugleich kann auf dieser Basis über mögliche DQ-Maßnahmen entschieden werden, um die Datengrundlage zu verbessern und somit Entscheidungen besser zu fundieren.

Anwendung der Metrik für das Datenqualitätsmerkmal Aktualität

Auch die Anwendung der Metrik für Aktualität soll am Kundenwertbeispiel illustriert werden. Bei der Kundenbewertung unterstellt der FDL, dass jeder Kunde grundsätzlich verschiedene Phasen eines Lebenszyklus durchläuft. Insofern wird versucht, die Profitabilität und den Bedarf eines Kunden an individuellen Lebensereignissen (z. B. Abschluss des Studiums) festzumachen und so die in den jeweiligen Phasen resultierenden Cashflows zu quantifizieren. Im Rahmen eines Projekts wurden auf diese Weise bereits Cashflowverläufe für idealtypische Zielgruppenkunden – je nach Ausprägung ausgewählter kundenspezifischer Charakteristika (z. B. Fachrichtung) ermittelt, aus denen sich die zugehörigen Kundenwerte ableiten lassen (vgl. Heidemann/Klier 2008). D. h. die Kundenwerte (Barwerte) ab dem Zeitpunkt des Berufseinstiegs (dann werden die Kunden für den FDL hinsichtlich der Produktverkäufe erst interessant) liegen bereits vor.

Natürlich verfügt der FDL aber auch über eine Vielzahl von Kunden, die den Zeitpunkt des Berufseinstiegs noch nicht erreicht haben und noch studieren. Diese werden mit dem Berufsstatus „Student“ in der Datenbank erfasst, wobei auch der Name, die Fachrichtung und die Anzahl der Semester (zum Zeitpunkt der Erfassung) hinterlegt werden. Hier stellt sich nun die Frage, wie diese Kunden zu bewerten sind. Da der FDL für Studenten keine bzw. nur wenige Produkte im Angebot hat, findet innerhalb dieser Kundengruppe i. d. R. lediglich ein sehr unregelmäßiger Kontakt statt. Dies führt – da Produkte meist (noch) nicht absetzbar sind – nicht nur zu negativen (wenn auch betragsmäßig geringen) Cashflows während des Studiums, sondern auch dazu, dass der genaue Zeitpunkt des Berufseinstiegs, ab dem positive Cashflows zu erwarten sind, nicht bekannt ist. Trotzdem ist eine Kontaktaufnahme mit allen potenziell noch Studierenden, um deren aktuellen Berufsstatus (regelmäßig zu den Zeitpunkten der Kundenbewertung) abzuprüfen, schon allein aus Kosten- und Ressourcengründen unrealistisch. Insofern ist es notwendig, die Kundenbewertung ohne diese aufwändige Prüfung durchzuführen, d. h. sie soll weitestgehend automatisiert erfolgen. Im Weiteren werden zunächst

die beiden bisherigen Verfahren des FDL zur Kundenbewertung dargestellt, um in die Problematik einzuführen. Danach wird der Einsatz der Metrik für Aktualität dargestellt.

Auf Basis der Problemstellung lag es für den FDL zunächst nahe, den Wert eines Kunden mit dem Berufsstatus „Student“ wie folgt zu berechnen: Wenn zum Zeitpunkt des Erst- oder Folgekontakts die Semesteranzahl des Kunden (neu) gespeichert wird, so lässt sich – durch Vergleich mit der durchschnittlichen Studiendauer – die Reststudiendauer schätzen. Die durchschnittliche Studiendauer kann dabei aus öffentlichen Quellen anhand der durchschnittlichen Studiendauer von Absolventen bzw. Studienabbrechern (vgl. Statistisches Bundesamt 2007; Heublein/Schmelzer/Sommer 2008; Hackl/Sedlacek 2001) berechnet werden. So wird hier davon ausgegangen, dass die 64% aller Studienanfänger, die ihr Studium erfolgreich abschließen, durchschnittlich 15 Semester studieren, wohingegen die restlichen 36% nach durchschnittlich 5 bis 6 Semestern ihr Studium abbrechen. Insgesamt ergibt sich so eine durchschnittliche Studiendauer von ca. 12 Semestern ($0,64 \cdot 15,4 \text{ Semester} + 0,36 \cdot 5,5 \text{ Semester}$).

Für die so ermittelte Reststudiendauer, wurde bei der Kundenbewertung ein negativer Cashflow angesetzt¹. Erst bei Erreichen des Zeitpunkts der durchschnittlichen Studiendauer – d. h. zum geschätzten Berufseinstieg – wurde der gegebene Kundenwert laut Lebenszykluskonzept für die jeweilige Zielgruppe – als Durchschnitt über Absolventen und Abbrecher – veranschlagt. Ein Beispiel illustriert dieses einfache Vorgehen:

Für einen Kunden mit Fachrichtung Wirtschaftswissenschaften wurde im April 2005 erfasst, dass er im zweiten Semester studiert. Zum Zeitpunkt der Kundenbewertung im Mai 2008 wurde demzufolge davon ausgegangen, dass nunmehr acht Semester des Studiums absolviert sind. Setzt man dazu die durchschnittliche Studiendauer von zwölf Semestern in Beziehung, so ergibt sich eine geschätzte Reststudiendauer von vier Semestern. Für diesen Zeitraum wird ein negativer Cashflow (hier -150 € pro Semester) angesetzt. Zum Zeitpunkt des Berufseinstiegs sind für einen Absolventen der Wirtschaftswissenschaften laut Lebenszykluskonzept ein Barwert von 4.000 € für einen Abbrecher 2.000 € zu veranschlagen (hieraus lässt sich ein gewichteter durchschnittlicher Kundenwert von $0,64 \cdot 4.000 \text{ €} + 0,36 \cdot 2.000 \text{ €} = 3.280 \text{ €}$ berechnen). Tabelle 3 gibt eine Übersicht:

Zeit seit Studienbeginn	Geschätzte Reststudiendauer	Ø Cashflow als Student	Ø Kundenwert zum Berufseinstieg	Berechneter Kundenwert
8 Semester	4 Semester	-150 €/Semester	3.280 €	1.888 €

Tabelle 3: Kundenwertberechnung am Beispiel (ursprüngliches Vorgehen)

Wie das Beispiel verdeutlicht, wird bei einem Kalkulationszinssatz von 5% pro Semester mit 1.888 € ein – im Vergleich zum Kundenwert beim Berufseinstieg – relativ geringer Kundenwert ausgewiesen. Dies liegt u. a. daran, dass bei der Berechnung angenommen wurde, dass der Kunde mit Sicherheit noch genau vier Semester negative Cashflows generiert und erst danach in den Beruf einsteigt. Diese Annahme ist sehr vereinfachend, abstrahiert von der Qualität der Inputdaten und kann zu verzerrten, wenig belastbaren Kundenwerten führen.

Deswegen führte der FDL zunächst ein modifiziertes Verfahren ein, welches eine wahrscheinlichkeitsorientierte Betrachtung beinhaltet. So sollte die Annahme einer festen Studiendauer vermieden und die darauf basierende Berechnung unter Sicherheit verbessert werden. Das Verfahren sah vor, dass der Berufsstatus „Student“ eines zum Bewertungszeitpunkt gespeicherten Kunden jeweils in der Folgeperiode (Semester) unverändert bleiben oder sich ändern kann, falls der Kunde sein Studium beendet (und Absolvent oder Abbrecher wird). Diesen möglichen Übergängen im Berufsstatus wurden Wahrscheinlichkeiten zugewiesen, die zwar unter Berücksichtigung der geschätzten Reststudiendauer ermittelt wurden. Jedoch konnte durch das Verfahren, das an homogene Markov-Ketten angelehnt war (vgl. z. B. Pfeifer/Carraway 2000; Rudolf-Sipötz 2001), auch berücksichtigt werden, dass der Zeitpunkt des Berufseinstiegs ggf. schon vor bzw. erst nach dem Ablauf der durchschnittlichen Studiendauer erfolgt. Abbildung 2 veranschaulicht dieses Konzept:

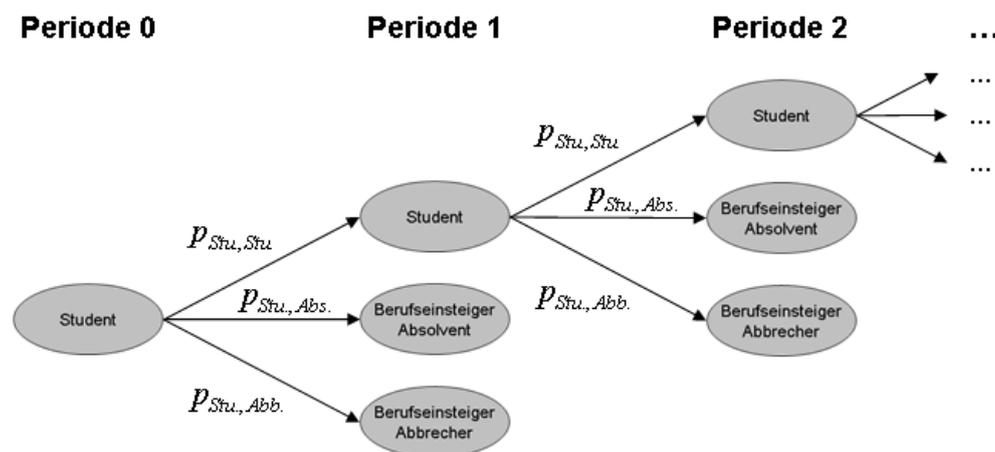


Abbildung 2: Übergänge vom Berufsstatus Student im modifizierten Verfahren

Die Wahrscheinlichkeit $p_{Stu, Stu}$ wurde dabei kundenindividuell so gewählt, dass die erwartete Semesteranzahl, in welcher der Kunde weiterhin im Berufsstatus „Student“ verbleibt, gleich der geschätzten Reststudiendauer ist. Im Mittel über viele Kunden wurde demnach zwar wieder die durchschnittliche Studiendauer (geschätzter Berufseinstieg) getroffen. Jedoch konnte

nun auch der (sehr realistische) Fall eintreten, dass ein Kunde in Bezug zur durchschnittlichen Studiendauer früher oder später sein Studium beendet (Streuung).

Legt man den obigen Beispielkunden zugrunde, so ergibt sich die Übergangswahrscheinlichkeit $p_{Stu.,Stu.}$ zu 61%. Hochgerechnet bedeutet dies, dass 61 von 100 Kunden, die sich im achten Semester befinden, im nächsten Semester immer noch studieren, wohingegen 39 ihr Studium beenden. Diese 39 Studenten sind wiederum anhand der obigen Anteile in 25 Absolventen und 14 Abbrecher zu unterscheiden. Auf Basis dieser Überlegungen lässt sich für den Beispielkunden nunmehr ein Kundenwert von 2.550 € berechnen (vgl. Abbildung 3).

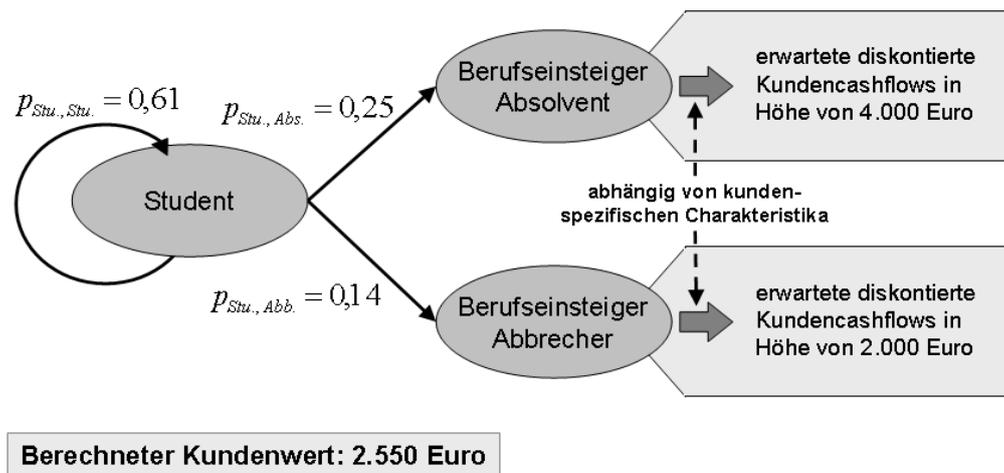


Abbildung 3: Kundenwertberechnung mit dem modifizierten Verfahren am Beispiel

Auch wenn dieses Verfahren nunmehr ein potenziell früheres oder späteres Ende des Studiums – im Vergleich zur durchschnittlichen Studiendauer – berücksichtigt, hängt die Frage wie belastbar der Kundenwert ist, natürlich wesentlich von der Qualität der Inputdaten ab. Konkret am Beispiel bezieht sich dies insbesondere auf die Annahme, dass sich der Kunde zum Bewertungszeitpunkt im Mai 2008 überhaupt im achten Semester befindet und faktisch noch Student ist. Diese Annahme liegt beiden bisherigen Berechnungen zugrunde und ist deshalb angreifbar, da der Fall ignoriert wird, dass die in der Datenbank gespeicherten Kundendaten als Input für die Kundenwertberechnung ggf. bereits veraltet sind und der Kunde im Mai 2008 überhaupt nicht mehr studiert. Gerade dies hatte sich ex post auch bei einigen Kontakten mit Kunden, die mit dem Berufsstatus „Student“ gespeichert waren, herausgestellt. Deshalb wurde untersucht, wie die Metrik für Aktualität eingesetzt werden kann, um dem DQ-Problem, dass der Berufsstatus „Student“ bei vielen Kunden nicht mehr den aktuellen Gegebenheiten entspricht, zu begegnen. Die grundsätzliche Idee lässt sich dabei wie folgt festhalten:

Da das Metrikergebnis für Aktualität die Wahrscheinlichkeit widerspiegelt, mit der ein Kunde zum Bewertungszeitpunkt noch Student ist, bietet sich das Metrikergebnis natürlich an, um die Übergangswahrscheinlichkeiten $p_{Stu.,Stu.}$ zu bestimmen. Um hierbei die restriktive Annahme, dass der Kunden zum Bewertungszeitpunkt mit Sicherheit noch Student ist, zu vermeiden, wurde im Beispiel in einem ersten Schritt ermittelt, wie wahrscheinlich es zum Bewertungszeitpunkt im Mai 2008 ist, dass der betrachtete Kunde (dessen Status „Student“ bereits im April 2005 erfasst wurde) überhaupt noch studiert ($Q_{Akt.}(w,A,0)=Q_{Akt.}(w,A)$). Für die folgenden Perioden (Semester) wurden dann die Übergangswahrscheinlichkeiten ebenfalls auf Basis der Metrik berechnet. Da es sich hierbei jeweils um bedingte Wahrscheinlichkeiten handelt, musste das Metrikergebnis jeweils mit einer um eins erhöhten Semesteranzahl erneut berechnet und für jede Periode t mit $t \geq 1$ der Quotient $p_{Stu.,Stu.}(t) = \frac{Q_{Akt.}(w,A,t)}{Q_{Akt.}(w,A,t-1)}$ gebildet werden, um die neuen, periodenspezifischen Übergangswahrscheinlichkeiten $p_{Stu.,Stu.}(t)$ zu erhalten². Die Berücksichtigung der Metrik hat dabei folgende Vorteile:

- I.) Bereits zum Bewertungszeitpunkt (im Beispiel im Mai 2008) kann eine Wahrscheinlichkeit dafür ermittelt werden, dass ein Kunde überhaupt noch Student ist. Dies verringert das Risiko einer Fehleinschätzung beim Kundenwert.
- II.) Die Übergangswahrscheinlichkeiten in den Folgeperioden müssen nicht mehr als konstant angenommen werden, sondern können periodenspezifisch bestimmt werden. Dies ist durchaus relevant im Hinblick auf die Belastbarkeit des Kundenwerts, da sich die Wahrscheinlichkeiten, dass Studenten direkt nach dem 5. bzw. 9. Semester das Studium beenden, auch in der Realität deutlich unterscheiden.
- III.) Die Qualität des Inputdatums (hier Berufsstatus „Student“) im Sinne seiner Aktualität kann kundenindividuell (bspw. unter Berücksichtigung der Fachrichtung) ermittelt werden. Dies ermöglicht eine differenzierte Berechnung des Kundenwerts, der somit weiter fundiert wird.

Abbildung 4 veranschaulicht das Verfahren unter Berücksichtigung der Metrikergebnisse $Q_{Akt.}(w,A,t)$:

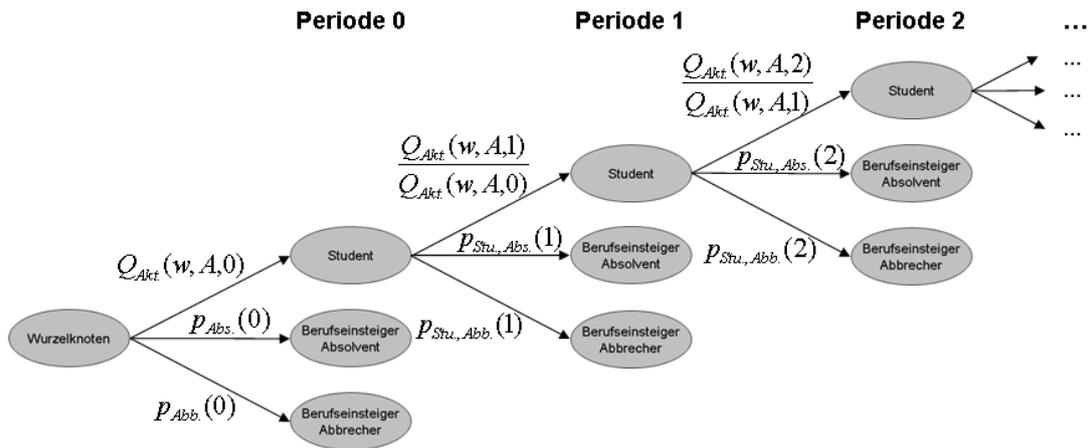


Abbildung 4: Übergangswahrscheinlichkeiten vom Berufsstatus Student im neuen Verfahren

Wie lässt sich nunmehr im Fallbeispiel die Metrik zur Abschätzung der Wahrscheinlichkeit $p_{Stu., Stu.}$ konkret entwickeln? Der Berufsstatus "Student" kann – wie dargestellt – seine Gültigkeit auf zwei Arten verlieren. Entweder das Studium wird abgeschlossen oder abgebrochen. Für beide Fälle musste die „Verfallsrate“ von Werten des Inputdatums (hier "Student") ermittelt werden, d. h. welcher Anteil an Datenwerten wird durchschnittlich innerhalb der jeweiligen Semester inaktuell. Diese „Verfallsraten“ für die verschiedenen Semester lassen sich anhand öffentlich zugänglicher Quellen ermitteln (vgl. Statistisches Bundesamt 2007; Heublein/Schmelzer/Sommer 2008; Hackl/Sedlacek 2001).

Betrachtet man bspw. den Fall der Studienabbrecher, so kann die relative Häufigkeitsverteilung von (Hackl/Sedlacek 2001) herangezogen werden (vgl. Abbildung 5). Diese beinhaltet den Prozentsatz an Abbrechern (bezogen auf alle Abbrecher) im jeweiligen Semester. So beenden bspw. ca. 18% aller Abbrecher bereits im ersten Semester ihr Studium. Dabei zeigte sich bei der Analyse, dass die Abbrecherquote in Bezug zu den im jeweiligen Semester noch Studierenden approximativ konstant ist (im Gegensatz dazu, verringert sich natürlich im Laufe der Semester die absolute Anzahl der Abbrecher). Insofern konnte eine konstante relative Verfallsrate und eine Exponentialverteilung angenommen werden.

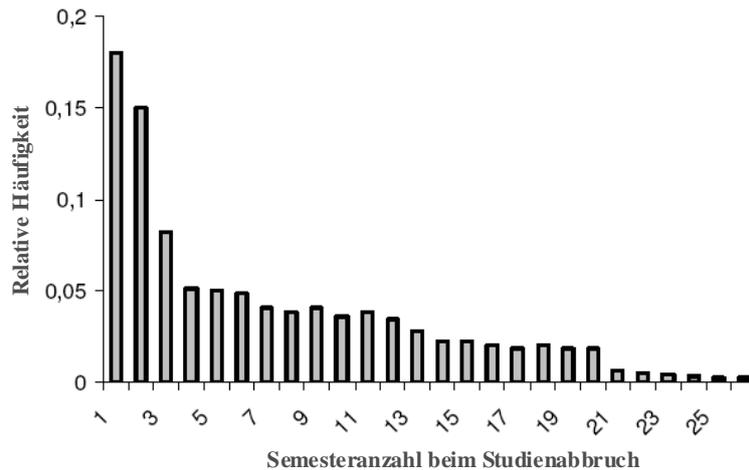


Abbildung 5: Relative Häufigkeitsverteilung bei Studienabbrechern (Hackl/Sedlacek 2001)

Auf Basis der Daten wurde eine Exponentialverteilung mit einer Verfallsrate von 0,18 ermittelt, was bedeutet, dass Studienabbrecher im Durchschnitt nach ca. 5,5 Semestern ihr Studium beenden. Die Verteilung $P_{\text{Abbrecher}}(x) = 1 - \exp(-0,18 \cdot x)$ repräsentiert dabei die Wahrscheinlichkeit, dass ein Student sein Studium nach x Semestern bereits abgebrochen hat. Analog konnte auch die Verteilung $P_{\text{Absolventen}}(x)$ ermittelt werden, welche die kumulierte Wahrscheinlichkeit widerspiegelt, dass ein Absolvent sein Studium nach x Semestern beendet hat. Allerdings war es hierbei notwendig, eine andere Verteilungsannahme zu treffen (zu alternativen Verteilungsannahmen vgl. Heinrich/Kaiser/Klier 2007a). Auf Basis beider Verteilungen ließ sich die Metrik für Aktualität als $Q_{\text{Akt.}}(w, A) = 1 - 0,64 \cdot P_{\text{Absolventen}}(x) - 0,36 \cdot P_{\text{Abbrecher}}(x)$ definieren und automatisiert sowie kundenindividuell das Metrikergebnis berechnen. Letztgenanntes stellt dabei zugleich die Wahrscheinlichkeit dafür dar, dass der Berufsstatus „Student“ für einen Kunden noch nicht veraltet ist und damit zum Bewertungszeitpunkt noch den aktuellen Gegebenheiten entspricht.

Wendet man das neue Verfahren im Fallbeispiel an, bedeutet dies, dass der Kunde zum Bewertungszeitpunkt nur noch mit einer Wahrscheinlichkeit von 67% Student ist, mit 5% bereits erfolgreich abgeschlossen und mit 28% abgebrochen hat. Bei den bisherigen Vorgehen wurden diesbezüglich keine Wahrscheinlichkeiten berechnet. Stattdessen wurde angenommen, dass der Kunde im Mai 2008 noch mit Sicherheit Student ist. Als weiterer Vorteil sind die Übergangswahrscheinlichkeiten (realistischerweise) im Zeitverlauf nun nicht mehr konstant und können kundenindividuell variieren. Insofern konnten die Ergebnisse der Kundenwertberechnung fundiert werden. Im Beispiel ergab sich ein Kundenwert von 2.045 €, der sich durch Berücksichtigung der Punkte I.), II. und III.) reduziert hat.

Abschließend bleibt festzuhalten, dass das Fallbeispiel bewusst einfach gehalten wurde, um den praktischen Mehrwert der Metriken herauszuarbeiten. So sind insbesondere im Fall der Metrik für Aktualität durchaus komplexere Sachverhalte abbildbar, wie bspw. die Berücksichtigung weiterer Zustände neben Student, Studienabsolvent oder -abbrecher. Hier ist z. B. daran zu denken, dass für Kunden, deren Datenerfassung schon länger zurückliegt, auch Zustände nach der unmittelbaren Berufseinstiegsphase (bspw. in einer Unternehmensberatung nicht mehr als Consulting-Analyst sondern als Junior Consultant) möglich sind. Darüber hinaus kann die Metrik dazu genutzt werden, den Zeitpunkt des Berufseinstiegs kundenindividuell genauer zu spezifizieren, um den Kunden zielorientiert kontaktieren zu können. Eine Orientierung an der mittleren Studiendauer – wie bisher – führt gerade dazu, sehr gute und deshalb tendenziell schneller abschließende Studenten, welche oftmals die attraktiveren Kunden darstellen, zu spät zu kontaktieren. Nichtsdestotrotz sollte auch das einfache Fallbeispiel illustrieren können, dass die Ermittlung der Qualität von Inputdaten im Kundenwertcontrolling nicht nur hilft, die Frage besser zu beantworten, wie belastbar eine berechnete Kenngröße tatsächlich ist (wie im Fall des DQ-Merkmals Vollständigkeit). Vielmehr kann die Messung der DQ auch nützlich sein, um Kenngrößen selbst in ihrer Berechnung zu fundieren.

Zusammenfassung

Im Beitrag wurde die Fragestellung aufgegriffen, wie die Messung der DQ in Datenbanken erfolgen und welcher Mehrwert im Berichtswesen dadurch generiert werden kann. Ziel war es demnach nicht nur, Metriken für die Qualitätsmerkmale Vollständigkeit und Aktualität vorzustellen, die eine reproduzierbare und nachvollziehbare Messung ermöglichen. Vielmehr sollte auch – anhand des Fallbeispiels – verdeutlicht werden, welcher praktische Nutzen mit ihrer Anwendung einhergeht. So können Effekte auf die DQ, wie z. B. zeitlicher Verfall oder die Durchführung von Qualitätsmaßnahmen, untersucht und damit ex ante Planungswerte mit ex post Messwerten verglichen werden. Dies sowie die Eignung der Metriken konnte bereits in Zusammenarbeit mit Unternehmen gezeigt werden (vgl. auch Heinrich/Kaiser/Klier 2007a, Heinrich/Kaiser/Klier 2007b).

Im Kern können dadurch im Berichtswesen bereits bekannte Kenngrößen wie der Kundenwert mit einer Qualitätskennzahl untermauert werden. Dies führt zu einer verbesserten Entscheidungsgrundlage. Wird bspw. der Kundenwert dazu verwendet, die Intensität der Kundenbetreuung im Vertrieb zu steuern, so kann es zu völlig falschen Unternehmensentscheidungen kommen, falls die Berechnungen auf veralteten oder unvollständigen Daten basieren. Gleiches gilt in anderen Anwendungskontexten. Insofern ist immer zu hinterfragen, von

welcher Qualität die Inputdaten bei der Ermittlung einer Kenngröße sind. Die vorgestellten Metriken bieten hierfür Unterstützung und erlauben dabei zugleich, potenzielle Maßnahmen zur Verbesserung der DQ im Controlling hinsichtlich ihrer ökonomischen Vorteilhaftigkeit zu überprüfen (vgl. auch Heinrich/Kaiser/Klier 2007a).

Anmerkungen

¹ Ist die vergangene Zeit seit Studienbeginn größer als die durchschnittliche Studiendauer und liegen keine weiteren Daten vor, so wird die geschätzte Reststudiendauer auf ein Semester gesetzt.

² Die Übergangswahrscheinlichkeiten $p_{Stu.,Abs.}(t)$ sowie $p_{Stu.,Abb.}(t)$ können ebenfalls unter Berücksichtigung der Metrikergebnisse ermittelt werden. Auf eine konkrete Darstellung wird hier aus Platzgründen verzichtet.

Literaturhinweise

Al-Hakim, L.: Information quality factors affecting innovation process, in: International Journal of Information Quality, 1. Jg. (2007), Heft 2, S. 162-176.

Ballou, D. P./Tayi, G. K.: Enhancing Data Quality in Data Warehouse Environments, in: Communications of the ACM, 42. Jg. (1999), Heft 1, S. 73-78.

Ballou, D. P./Wang, R./Pazer, H./Tayi, G. K.: Modeling Information Manufacturing Systems to Determine Information Product Quality, in: Management Science, 44. Jg. (1998), Heft 4, S. 462-484.

Bobrowski, M./Marre, M./Yankelevich, D.: A homogeneous framework to measure data quality, in: Proceedings of the 4th International Conference on Information Quality (ICIQ), Boston 1999, S. 115-124.

English, L.: Improving Data Warehouse and Business Information Quality, 1. Aufl., New York 1999.

Eschenbach, R.: Controlling, 2. Aufl., Stuttgart 1996.

Even, A./Shankaranarayanan, G.: Utility-Driven Assessment of Data Quality, in: The DATA BASE for Advances in Information Systems, 38. Jg. (2007), Heft 2, S. 75-93.

Fisher, C. W./Chengalur-Smith, I./Ballou, D. P.: The Impact of Experience and Time on the Use of Data Quality Information in Decision Making, in: Information Systems Research, 14. Jg. (2003), Heft 2, S. 170-188.

Hackl, P./Sedlacek, G.: Analyse der Studiendauer am Beispiel der Wirtschaftsuniversität Wien, in: Dutter, R. (Hrsg.): Festschrift 50 Jahre Österreichische Statistische Gesellschaft, Wien 2001, S. 41-59.

Harris Interactive: Studie Information Workers Juni 2006 für USA, Großbritannien, Frankreich und Deutschland, 2006.

Heidemann, J./Klier, M.: Ein Ansatz zur Operationalisierung des Kundenwerts und seine praktische Anwendung am Beispiel eines Finanzdienstleisters, Diskussionspapier des Lehrstuhls WI-IF der Universität Augsburg 2008.

Helfert, M.: Eine empirische Untersuchung von Forschungsfragen beim Data Warehousing aus Sicht der Unternehmenspraxis. Arbeitsbericht BE HSG/CC DWS/05 des Instituts für Wirtschaftsinformatik der Universität St. Gallen, St. Gallen 2000.

Helfert, M./Heinrich, B.: Analyzing Data Quality Investments in CRM - A model-based approach, in: Proceedings of the 8th International Conference on Information Quality, Boston 2003, S. 80-95.

Heinrich, B./Kaiser, M./Klier, M.: DQ Metrics: A Novel Approach to Quantify Timeliness and its Application in CRM, in: Proceedings of the 12th International Conference on Information Quality (ICIQ), Boston 2007a.

Heinrich, B./Kaiser, M./Klier, M.: How to measure data quality? – a metric based approach, in: Proceedings of the 28th International Conference on Information Systems (ICIS), Montreal 2007b.

Heublein, U./Schmelzer, R./Sommer, D.: Die Entwicklung der Studienabbruchquote an den deutschen Hochschulen: Ergebnisse einer Berechnung des Studienabbruchs auf der Basis des Absolventenjahrgangs 2006, Hamburg 2008.

Horváth, P.: Controlling, 10. Aufl., München 2006.

-
- Jung, H.: Controlling, 2. Aufl., München 2007.
- Juran, J. M.: How to think about Quality, 1. Aufl., New York 1998.
- Lee, Y. W./Strong, D. M./Kahn, B. K./Wang, R. Y.: AIMQ: a methodology for information quality assessment, in: Information & Management, 40. Jg. (2002), Heft 2, S. 133-146.
- Mengen, A./Mettler, A.: Kundenwertermittlung - wie viel Vertrieb ist uns der Kunde wert? in: Zeitschrift für Controlling und Management, 52. Jg. (2008), Heft 1, S. 30-36
- Meyer, M./Shaffu, N.: Unternehmenswertorientiertes Kundenmanagement auf Basis des Customer Lifetime Value: Eine empirische Studie bei Mobilfunknutzern in Deutschland und Frankreich, in: Zeitschrift für Controlling und Management, 51. Jg. (2007), Heft 1, S. 54-62
- Pfeifer, P. E./Carraway, R. L.: Modelling Customer Relationships as Markov Chains, in: Journal of Interactive Marketing, 14 Jg. (2000), Heft 2, S. 43-55.
- Redman, T. C.: Data Quality for the Information Age, 1. Aufl., Boston 1996.
- Rohwedder, J. P./Kasten, G./Malzahn, D./Piro, A./Schmid, J.: Informationsqualität – Definitionen, Dimensionen und Begriffe, in: Hildebrand, K./Gebauer, M./Hinrichs, H./Mielke, M. (Hrsg.): Daten- und Informationsqualität – Auf dem Weg zur Information Excellence, Wiesbaden 2008, S. 25-46.
- Rudolf-Sipötz, E.: Kundenwert: Konzeption – Determinanten – Management, Dissertation, St. Gallen 2001.
- Samleben M./Stadlbauer, F./Hess, T.: Anwendungssystemintegration im Controlling: Aktueller Stand und wichtige Trends, in Zeitschrift für Controlling & Management, 50. Jg. (2006), Heft 2, S. 86-93.
- Statistisches Bundesamt: Bildung und Kultur 2006 – Prüfungen an Hochschulen (Fachserie 11 Reihe 4.2), Wiesbaden 2007.
- Teboul, J.: Managing Quality Dynamics, 1. Aufl., New York 1991.
- The Data Warehousing Institute: Taking Data Quality to the Enterprise through Data Governance, USA 2006.
- Wang, R. Y./Storey, V. C./Firth, C. P.: A Framework for analysis of data quality research, in: IEEE Transactions on Knowledge and Data Engineering, 7. Jg. (1995), Heft 4, S. 623-640.