

Gesetz zur Änderung des Bundesdatenschutzgesetzes

Ein Modell zur Scoring-Beauskunftung

Der Deutsche Bundestag und der Bundesrat haben 2009 das „Gesetz zur Änderung des Bundesdatenschutzgesetzes“ (BDSG-Novelle I) betreffend Scoring und Auskunfteien beschlossen, das zum 1. April 2010 in Kraft getreten ist. Ziel der neuen Regelungen ist es unter anderem die Transparenz beim Einsatz von Scoring-Verfahren zu erhöhen. Unternehmen und Auskunfteien sind somit seit Anfang April 2010 dazu verpflichtet, ihren Kunden auf Anfrage deren Scoring schriftlich und verständlich offen zu legen.

Wir stellen in diesem Beitrag den rechtlichen Rahmen dieser Gesetzesnovelle vor und entwickeln daraufhin ein Modell, welches eine Offenlegung für Scoringssysteme aller Art ermöglicht. Das entwickelte Konzept garantiert, dass der Kunde ausreichend in die Stärken und Schwächen seiner Bonität Einsicht nehmen kann, ohne gleichzeitig methodisch die Grundlagen multivariater Merkmalsanalysen oder Regressionsmodelle verstehen zu müssen. In einem Beispiel aus dem Bereich des Konsumentencredit-Scoring wird das Konzept ausgeführt und zeigt eine technische Lösung, so wie sie von Banken jeder Größe implementiert werden kann.

Das rechtliche Problem und seine Auswirkung auf die Bankpraxis

Der neugefasste zweite Absatz in § 34 BDSG schafft einen Auskunftsanspruch des Kunden gegenüber der Bank bei Einsatz von Scoring-Verfahren im Sinne des § 28b BDSG. Bereits in der Vergangenheit hatte der Kunde der Bank gegenüber nach § 34 Abs. 1 BDSG einen Auskunftsanspruch über alle den Betroffenen gespei-

cherten personenbezogenen Daten. Seit dem 1. April 2010 können die Bankkunden bei Einsatz des Scorings von der für die Entscheidung verantwortlichen Stelle über folgende Aspekte Auskunft verlangen:

1. die innerhalb der letzten sechs Monate vor dem Zugang des Auskunftsverlangens erhobenen oder erstmalig gespeicherten Wahrscheinlichkeitswerte,

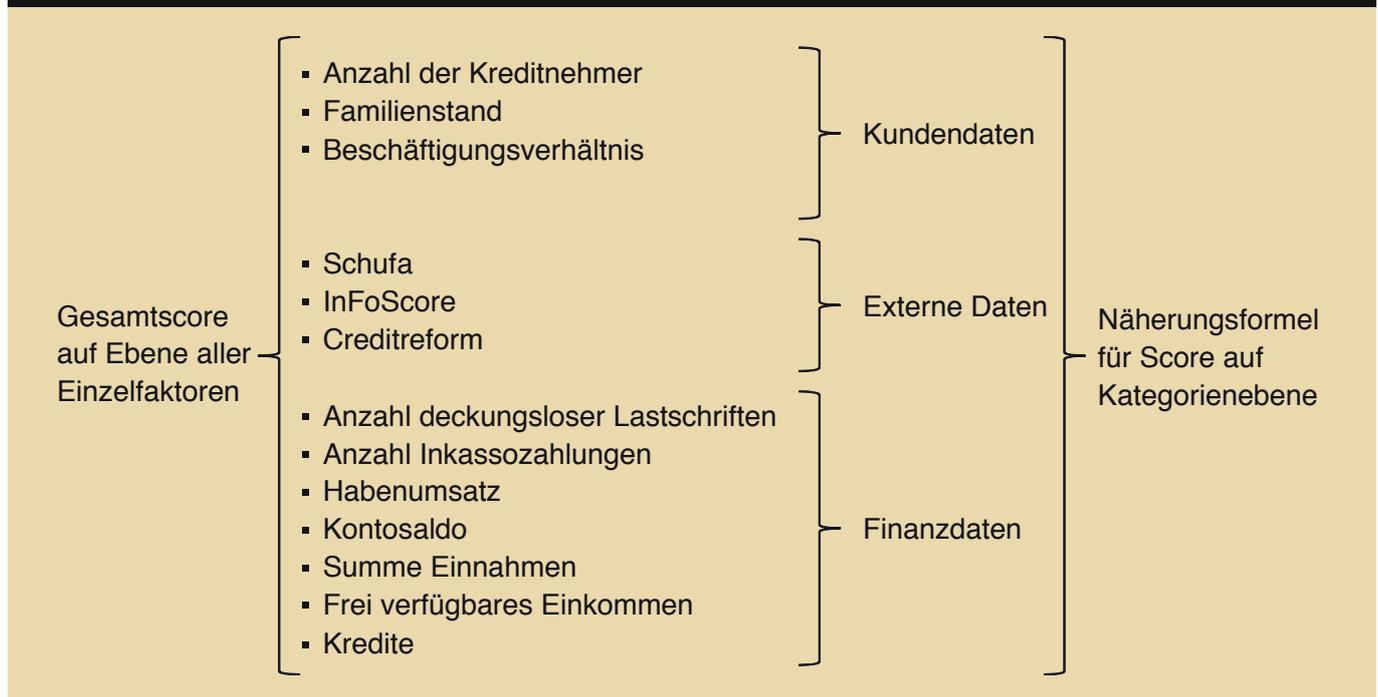
2. die zur Berechnung der Wahrscheinlichkeitswerte genutzten Datenarten und
3. das Zustandekommen und die Bedeutung der Wahrscheinlichkeitswerte.

Die Auskunft ist einzelfallbezogen und in allgemein verständlicher Form nachvollziehbar zu erstellen.

Der Auskunftsanspruch erstreckt sich grundsätzlich auf alle innerhalb der letz-

Kategorisierung von Einzelmerkmalen

► Abb. 01



SICHERHEIT SCHAFFT VERTRAUEN. UND KARRIERECHANCEN.



www.bundesdruckerei.de

► „Identität schützen“: Mit diesem Ziel entwickelt die Bundesdruckerei zukunftsorientierte Identifikationssysteme und Dokumenten-Lösungen, die weltweit geschätzt und eingesetzt werden.

Teilen Sie unseren Erfolg und tragen Sie zur dynamischen Entwicklung unseres mittelständischen Unternehmens bei. Als innovative und engagierte Persönlichkeit bieten wir Ihnen langfristige Perspektiven und ausgezeichnete Aufstiegschancen. Dazu kommen eine ansprechende Vergütung, gute Sozialleistungen sowie flexible Arbeitszeitregelungen – und nicht zu vergessen: unser Standort mitten im Herzen Berlins! ◀

Manager Risk-Controlling (m/w)

Kennziffer: 1119

Ihre Aufgaben:

- Implementierung und Durchführung eines unabhängigen Risk-Controllings für internationale IT-Großprojekte und IT-Ausschreibungen
- Überwachung und Beratung des projektinternen Risikomanagements
- Festlegung des Vorgehens und verantwortliche Umsetzung
- Ermittlung und Bewertung von Projektrisiken aller Art (technische, kommerzielle, rechtliche und fiskalische Risiken)
- Abstimmung geeigneter Maßnahmen mit dem projektinternen Risikomanagement
- Überwachung der Abarbeitung durch das projektinterne Risikomanagement
- Mitarbeit bei der ständigen Verbesserung der Arbeitsprozesse des Risk-Controllings
- Erstellung von Reports und Dokumentationen
- Reporting gegenüber der Bereichsleitung und ggf. der Geschäftsführung

Ihr Profil:

- abgeschlossenes technisches Hochschulstudium oder vergleichbare Qualifikation
- Erfahrung im Bereich Risikomanagement oder im Qualitätsmanagement für Großprojekte
- hohe Kompetenz in der Analyse komplexer IT-Architekturen und technischer Abläufe
- Kenntnisse im Bereich IT-Sicherheit, in komplexen IT-Systemen u./o. Reise- und Identitätsdokumenten
- fundiertes Know-how im adäquaten Einsatz moderner Verfahren des Risk-Controllings
- verhandlungssicheres Englisch in Wort und Schrift
- wirtschaftliches Denken und Handeln
- Fähigkeit, komplexe Sachverhalte verständlich zu dokumentieren und managementgerecht zu präsentieren und zu vertreten
- selbstständiger Arbeitsstil und Teamgeist

Wir freuen uns auf Ihre Bewerbung.

Haben Sie Interesse? Dann bewerben Sie sich bei uns. Bitte nennen Sie in Ihrer Bewerbung Ihre Gehaltsvorstellung und den frühestmöglichen Eintrittstermin. Ihre Unterlagen senden Sie uns bitte unter Angabe der Kennziffer

**bevorzugt per E-Mail:
bewerbungen@bdr.de**

Bundesdruckerei GmbH
Jana Winkelmann
Oranienstraße 91
10969 Berlin

Tel. (0 30) 25 98 - 23 92
Fax (0 30) 25 98 - 23 16

**Unser Team freut sich, Sie zu
begrüßen!**



ten sechs Monate erhobenen bzw. gespeicherten entscheidungsrelevanten Scorewerte. Während bei einem abgelehnten Kreditantragsteller oftmals nur ein Antragsscorewert vorliegen dürfte, könnten bei Bestandskunden neben eventuellen Antragsscorewerten beispielsweise alle erstellten Bestandscores mitzuteilen sein. Der Umfang der Auskunft kann somit gerade im Bestandskundenbereich durchaus komplex sein.

Bei der Auskunftserteilung reicht es jedoch nicht aus, dem Kunden nur den „nackten“ Wahrscheinlichkeitswert mitzuteilen. Mit dieser Zahl könnte der Kunde in der Regel nichts anzufangen wissen. Vielmehr ist dieser Wert so in Relation zu setzen, dass der Kunde dessen Aussagegehalt verstehen kann. Dazu gehört, dass dem Betroffenen mitgeteilt wird, auf welches bestimmte künftige Verhalten sich der Scorewert bezieht. Zusätzlich ist die Skala der möglichen Scorewerte anzugeben.

Da mit der Beauskunftung des ermittelten Scorewerts letztlich erreicht werden soll, dass der Betroffene den Aussagegehalt seines persönlichen Scorewerts versteht, muss der Kunde darüber hinaus in die Lage versetzt werden, einzuschätzen, ob der zu seiner Person errechnete Scorewert ein guter, mittlerer oder schlechter Scorewert ist. Ohne diese Erläuterung wäre die Kenntnis des Scorewerts für den Kunden nicht aussagekräftig.

Zusätzlich sind dem Kunden nach § 34 Abs. 2 Satz 1 Nr. 2 BDSG die einzelnen Datenarten, die in die Scorewert-Berechnung einfließen, zu nennen. Bei der Bildung der Datenarten ist ein möglichst guter Mittelweg zwischen Aussagekraft für den Kunden und Wahrung von Geschäftsgeheimnissen zu finden. Dabei kann die Zahl der Datenarten je nach Scoring-Verfahren variieren. Werden beispielsweise nur vier Datenpunkte zum Scoring genutzt, könnte bereits die Nennung von zwei Datenarten ausreichen, bei mehr als zehn Datenpunkten wären zwei Datenarten zu aussageschwach, dagegen wäre hier die Nennung von vier Datenarten ausreichend informativ.

Die größte Bedeutung bekommt jedoch die Frage, wie der § 34 Abs. 2 Satz 1 Nr. 3 BDSG umgesetzt werden kann, wonach die Bank dem Kunden das Zustandekommen und die Bedeutung der Wahrscheinlichkeitswerte einzelfallbezogen und nachvollziehbar in allgemein verständlicher Form mitzuteilen hat.

Hinsichtlich des Zustandekommens des Scorewerts dürfte grundsätzlich der Hinweis ausreichen, dass anhand der Verknüpfung von individuellen Daten mit statistischen Erfahrungen eine möglichst zuverlässige Prognose erstellt wird. Hinsichtlich der Bedeutung des Scorewerts erfolgt eine Darstellung als „Ampellösung“, unterteilt nach Datenarten (siehe oben). Dazu wird dem Kunden die Relevanz je Datenart für das Gesamtergebnis mittels Ampel (Rot / Gelb / Grün) dargestellt.

Aufgrund der Anforderungen an die Verständlichkeit und Nachvollziehbarkeit der Auskunft sollte die Auskunftserteilung ausschließlich in Textform erfolgen.

Die Kommunikation hin zum Kunden

Risikofaktoren und Risikokategorien

Für die methodische Umsetzung der Scoringbeauskunftung ist zunächst wichtig, dass es nicht Sinn und Zweck sein kann, alle Details des Verfahrens offen zu legen. Zum einen können Scoringmodelle beliebig kompliziert – und damit für Privatpersonen kaum mehr verständlich kommunizierbar – sein, zum anderen wäre bei vollständiger Offenlegung auch jedem Missbrauch, aber auch der Gefährdung von Alleinstellungsmerkmalen gegenüber konkurrierenden Unternehmen Tür und Tor geöffnet.

Anstatt also die eingehenden Variablen in ihrer ganzen Breite und mit ihrem ge-

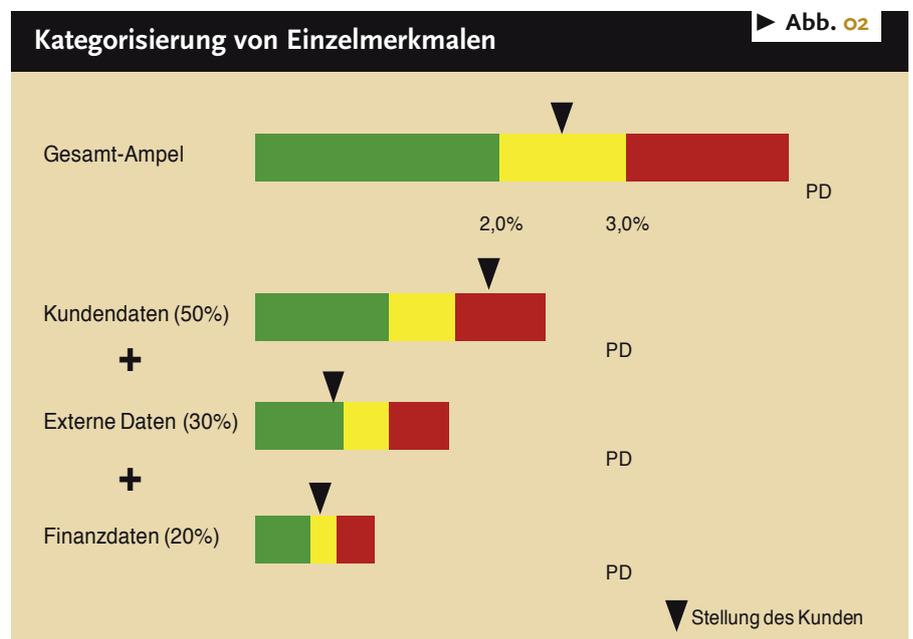
nauen Gewicht auszufächern, ist es angebracht, einzelne Risikokategorien in Form von aggregierten Datenarten darzustellen – wie etwa die Kategorie „Kundendaten“, „Finanzdaten“, oder „externe Daten“. Hinter jeder dieser Kategorien steckt eine definierte Anzahl an einzelnen Risikofaktoren, die in das Gesamtscoreing Eingang finden. Erste Aufgabe in der Scoring-Beauskunftung ist also, die einzelnen Risikofaktoren in Datenarten oder (Ober-)Kategorien zusammenzufassen (siehe ► **Abb. 01**). Das muss für jedes Scoringverfahren einzeln geschehen und kann nur per inhaltlicher Durchsicht erfolgen.

Die Ampellogik

Sind die Datenarten bzw. Risikokategorien mit ihren dahinter liegenden Faktoren definiert, geht es an die Frage: Wie ist ein Teilscoreing für eine Kategorie definiert, wenn man zunächst nur eine Formel für das Gesamtscoreing in Händen hält?

Der erste Schritt zur Erleuchtung dieses Problems ist die Erkenntnis, dass alle Scorings nur einen gemeinsamen Nenner haben: Die Ausfallwahrscheinlichkeit (probability of default, kurz PD). Am Ende einer jeden kalibrierten Bonitätsbeurteilung steht ein Mapping auf die PD – man ist daher gut beraten, diese PD als den gemeinsamen Maßstab aller Verfahren zu verwenden.

Weil aber die PD ein statistisches Maß ist und daher in der Kommunikation unverständlich wirkt, ist ein einfaches Ampelsystem anschaulicher und selbsterklä-



render. In einem solchen Ampelsystem kann sich der Kunde am Gesamtportfolio messen und gelangt darüber zu einer Einschätzung, wie es um seine Bonität innerhalb der verschiedenen Datenarten bestellt ist (siehe ► **Abb. 02**).

Grün bedeutet hier, dass die Ausfallwahrscheinlichkeit niedrig ist und damit die Bonität als gut eingeschätzt wird. Gelb bedeutet, dass eine erhöhte (mittlere) Ausfallwahrscheinlichkeit vorliegt, der mit einer weiteren Prüfung entgegen gekommen werden kann. Rot (schlecht) kennzeichnet diejenige Einschätzung der Ausfallwahrscheinlichkeit, die für die Bank als zu riskant gilt, um hier Geschäft zu machen.

Ampelsysteme lassen sich nicht nur auf Gesamt-PD-Ebene, sondern auch auf Datenarten-Ebene darstellen und sind somit eine einfache Variante, wie man eine komplexe Bonitätseinschätzung grafisch transparent transportieren kann.

Konsistente Ampelfarben

Nachdem es vor der Novellierung des BDSG keine Notwendigkeit gab, Ampeln auf Teil-Kategorien zu definieren, kommt es in der Praxis durchaus vor, dass sich Scoringssysteme nicht ganz einfach in Teil-Kategorien zerlegen lassen. In einem solchen Falle muss das Modell sicherstellen, dass sich die Teilampeln nicht widersprechen. Es ist also zu verhindern, dass ein Fall, der insgesamt als „Rot“ ausgewiesen wurde, in denn Teilampeln auf „Grün“ und „Gelb“ steht. Eine solche Konstellation wäre kaum zu vermitteln und ist auch bei gründlicher Logik des Rechenapparates nicht zu erwarten. Aber gerade weil die Scoringmodelle nie von Datenarten bzw. Teil-Kategorien auf eine Gesamtnote hoch entwickelt wurden, sondern von vornherein nur auf eine Gesamtnote modelliert sind, ist ein solch inkonsistenter Fall in einem nichtlinearen Modell theoretisch vorstellbar.

Am Ende ist die wichtigste Anforderung an unser Rechenwerk also, dass dort, wo sich mehrere Kategorien gleicher Farbe treffen, auch diese Farbe ausgeworfen wird. In einer mehrdimensionalen Darstellung vermittelt das untenstehende Bild diese Vorstellung (siehe ► **Abb. 03**).

Die Achsen des Würfels oben geben die Farben der Teilampeln an, und die Würfelstücke darin zeigen, welche Gesamtampel-Farbe für diese Kombinationen zulässig wäre. Dort, wo also Kunden-, Finanzierungsdaten sowie externe Daten ein „Rot“ ausweisen, ist also auch auf Gesamtebene nur ein „Rot“ vorstellbar. Sollte sich in einem Massentest über das Portfolio erweisen, dass mehrfach Kombinationen von Teilampeln ausgewiesen werden, die nicht in das obige Würfelbild passen, wäre eine solche Scoring-Beauskunftung als inkonsistent zurückzuweisen.

Die Umkehrung dieser Aufgabe, nämlich von einem Gesamtampel „Rot“ auf die Farben in den gewünschten Teil-Ka-

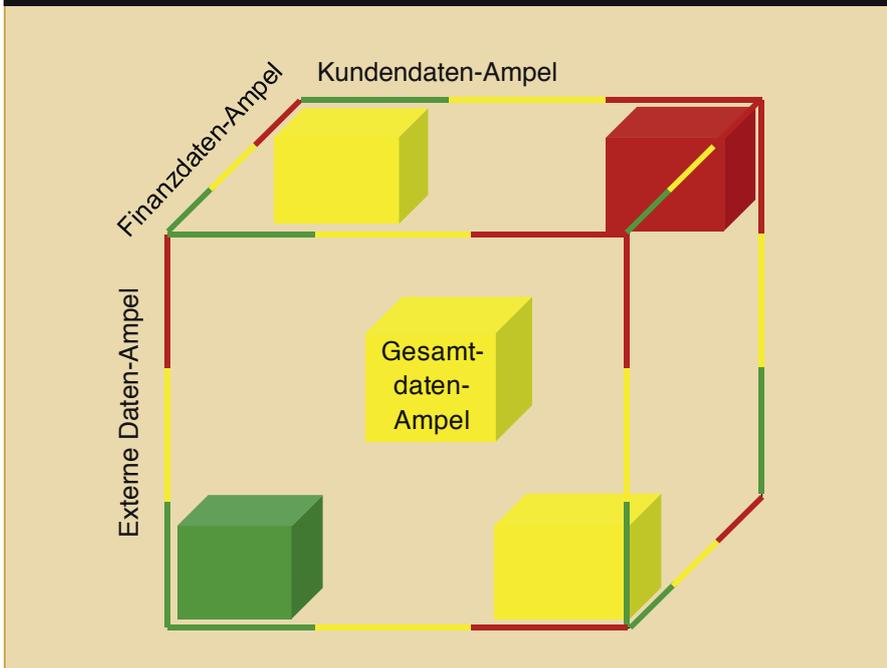
TICKER +++ TICKER +++ TICKER+++ TICKER +++ TICKER

+++ **Deutsche halten Griechenlandkrise für hausgemacht:** Laut einer Umfrage des Forschungsinstituts Emnid im Auftrag von N24 liegt die Hauptschuld an der Finanzkrise nach Ansicht von 59 Prozent der Deutschen bei den Griechen selbst. 13 Prozent halten ein Versagen der Banken für ursächlich, elf Prozent sehen die Schuld bei den europäischen Politikern. Dahingegen halten nur neun Prozent Spekulanten für die Schuldigen an der Griechenlandkrise. Mit Blick auf die Diskussion rund um die Hilfskredite wünschen sich 62 Prozent eine Beteiligung der Banken an der Rettung des Landes. Pessimistisch zeigen sich die Deutschen im Hinblick auf die Rückzahlung der Gelder: So sind 76 Prozent nicht davon überzeugt, dass es den Griechen gelingen wird, ihre Schulden zu tilgen. +++ **Fast 26 Mio. Deutsche nutzen Online-Banking:** Laut einer Umfrage des IT-Branchenverbands BITKOM erledigen fast 26 Mio. (und damit rund zwei Millionen mehr als im Vorjahr) Deutsche ihre Bankgeschäfte im Internet. Somit nutzen rund 41 Prozent aller Bundesbürger im Alter von 16 bis 74 Jahren Online-Banking, wodurch Deutschland im europaweiten Vergleich im Mittelfeld liegt. Spitzenreiter beim Online-Banking ist Norwegen mit 77 Prozent der Bevölkerung. An zweiter Stelle liegen die Niederlande mit 73 Prozent, gefolgt von Finnland, Island und Schweden mit jeweils gut zwei Dritteln. Insgesamt tätigt rund ein Drittel aller Bürger in der EU Bankgeschäfte via Internet. Gegen das Online-Banking werden vor allem Sicherheitsbedenken angeführt. So verzichten 28 Prozent der Internet-Nutzer aus diesem Grund auf den Internet-Zugang zu ihrer Bank. Zwei Prozent aller Internet-Nutzer sind nach eigenen Angaben bereits Opfer eines Betrugs beim Online-Banking geworden. +++ **BVI stellt Vorschläge zur Reform Offener Immobilienfonds vor:** Die im BVI Bundesverband Investment und Asset Management zusammengeschlossenen Anbieter von Offenen Immobilienfonds haben umfangreiche Vorschläge zur Reform ih-

rer Produkte präsentiert. Unter anderem sollte für alle Neu-Anleger eine Haltefrist von zwölf Monaten eingeführt werden. Nach deren Ablauf sollten Fondsanteile jedoch wie bisher täglich veräußerbar sein. Ein- und Auszahlpläne sollten weiterhin uneingeschränkt möglich bleiben, wobei durch ein entsprechendes Liquiditätsmanagement sichergestellt werden muss, dass Auszahlpläne auch bei einer etwaigen Rücknahmeaussetzung weiter bedient werden. Zudem plädiert der BVI für eine halbjährliche (statt wie bisher jährliche) Bewertung der Immobilien durch unabhängige Sachverständigenausschüsse. +++ **Zahlungsmoral deutscher Unternehmen steigt im April weiter an:** Die Zahlungsmoral deutscher Unternehmen hat sich einer Studie des Informationsdienstleisters D&B Deutschland zufolge auch im April weiter verbessert. So stieg der im Auftrag der Finanz-Nachrichtenagentur dpa-AFX erstellte D&B-Zahlungsindex im vergangenen Monat zum zehnten Mal in Folge auf einen Wert von 81,32 Punkten nach 80,79 Punkten im März. Höher stand der Indikator zuletzt vor mehr als zwei Jahren. Der Index reflektiert den Anteil der Unternehmen, die ihre Rechnungen pünktlich bezahlen. Von den rund 4,5 Millionen aktiven Unternehmen in Deutschland zahlten im April fast 130.000 mehr ihre Rechnungen innerhalb der vereinbarten Frist als noch vor einem Jahr. +++ **HRE mit höchstem Personalaufwand:** Laut einer Analyse des Karrieportals eFinacialCarreers weist der durchschnittliche Personalaufwand (im Wesentlichen also Gehälter, Personalnebenkosten und andere Komponenten wie etwa Abfindungen) innerhalb der deutschen Bankenlandschaft starke Unterschiede auf: Während Sparkassen und Genossenschaftsbanken mit knapp 47.000 bzw. gut 51.000 Euro pro Mitarbeiter das untere Ende markieren, fallen beim Branchenprimus Deutsche Bank bereits fast 147.000 Euro an. Den Spitzenwert markiert allerdings die HypoRealEstate mit über 152.000 Euro. +++

Kategorisierung von Einzelmerkmalen

▶ Abb. 03



tegorien zu schließen, wird im nächsten Kapitel angegangen.

Das mathematische Problem und seine Lösung

Es gibt keine wohldefinierten Teil-PDs

Wie oben ausgeführt, gibt es im allgemeinen Fall keine Bonitätseinschätzung auf den einzelnen Datenarten/Kategorien. Das Dilemma besteht nun also darin, ein Scoring, das beliebig viele Eingangsmerkmale zu einer PD zusammenführt, in bestimmte Anteile aufzuteilen. Nachdem aber jede Zahl in unendlich vielen Kombinationen aus anderen Zahlen zusammengesetzt werden kann, ist das Problem „wie werden die Ampelfarben des Gesamtscores auf Teilampeln heruntergebrochen?“ nicht eindeutig lösbar. Praktisch kann es also durchaus passieren, dass inkonsistente Fälle, so wie sie im Abschnitt oben dargestellt wurden, auftreten. In der Summe, also der finalen PD, sind die Ergebnisse der Teil-Kategorien nicht mehr auszumachen.

Dennoch verlangt die Novelle des BDSG von den auskunftsgewährenden Stellen, Teilbeurteilungen innerhalb von Kategorien abzugeben. Die PD, die ein Scoringverfahren liefert, muss also zunächst in ihre Treiber zerlegt werden. Nachdem die Datenarten/Kategorien aus den einzelnen Merkmalen heraus methodisch frei de-

finierbar (kombinierbar) sind, ist es am klügsten, gleich eine Zerlegung der PD in alle Einzelfaktoren vorzunehmen. Damit ist sichergestellt, dass die Kategorien beliebig aus den eingehenden Merkmalen zusammengestellt werden können.

Von einer Darstellung des Scorings auf Ebene von Einzelfaktoren ist dennoch abzuraten, weil die Einzelfaktoren viel zu granular sind, als dass sich hierfür sinnvolle Ergebnisse erzielen ließen. Nur durch die Zusammenfassung in Kategorien werden die statistischen Schwankungen hinreichend gemittelt, so dass sich stabile Aussagen ergeben. Ergebnisse für Einzelfaktoren sind üblicherweise viel zu fehlerbehaftet.

Die Zerlegung der Ausfallwahrscheinlichkeit

Der intuitivste Zugang, eine Größe in Teile zu zerlegen, ist es, diese Größe als Summe ihrer Teile darzustellen. Mit x als dem Vektor der Eingangsmerkmale suchen wir also eine Zerlegung der PD in Summanden, so dass $P(D | x) = \sum_i P(D | x_i)$ gilt. Nutzt man an dieser Stelle die Technik einer Taylor-Reihen-Zerlegung, so wird klar, dass die Funktionen $P(D | x_i)$ – also die „Teilscores“ – nichts anderes darstellen als die partiellen Ableitungen der PD nach den einzelnen Merkmalen, multipliziert mit dem individuellen Abstand des betrachteten Falles von einem Referenzpunkt x_m .

Als Referenzpunkt, um den herum die Taylor-Entwicklung angesetzt wird, empfiehlt sich hier der Mittelwert-Kunde, also ein Kunde, der in allen Merkmalen den Erwartungswert über ein bestimmtes Portfolio aufweist. Somit ist sichergestellt, dass jeder Teilscore einfach interpretiert werden kann, weil er den absoluten Beitrag zur PD misst, den das betrachtete Merkmal im Vergleich zur mittleren PD liefert.

Bricht man die Taylor-Entwicklung beim zweiten Grade ab, ergibt sich die Ausfallwahrscheinlichkeit $P(D | x)$ zu

$$P(D | x) \approx P(D | x_m) + \sum_i \partial P(D | x_i) / \partial x_i \cdot (x_i - x_m) + \sum_i \partial^2 P(D | x_i) / \partial x_i^2 \cdot (x_i - x_m)^2 + \dots$$

Mit obiger Formel können wir jede PD annähern, indem wir von der mittleren PD, $P(D | x_m) = E[P(D | x)]$, ausgehend noch das erste und das zweite Taylor-Glied eines jeden Merkmals dazu addieren.

Anschaulich betrachtet wird die PD eines Falles also approximiert, indem man zunächst einen mittleren Kunden als Referenzfall ins Rennen schickt, diesen Referenzfall dann entlang jedes Merkmals verschiebt, den Effekt (Sensitivität mal Abstand) misst, und die einzelnen Beiträge für jedes Merkmal aufsummiert. Die Einzelbeiträge sind dennoch nicht für bare Münze zu nehmen, in der Praxis ergeben sich erst durch die Aggregation innerhalb von Kategorien sinnvolle Werte, die auf einer Skala mit Farbeinteilung aufgetragen werden können.

Farben der Teilampeln

Funktioniert die Approximation der PD zwar für die „Vorwärtsrechnung“, also von einem Antrag x zu einer $P(D | x)$, so bleibt noch die Frage bestehen, wie man von einer bestimmten Cut-Off-Ausfallwahrscheinlichkeit PD_C „zurück“ schließen kann auf Cut-Offs, die jeden Teilscore kategorisieren. Für den Ausweis von Ampelfarben für die Teilscores muss schließlich ein jeder Teilscore seine eigenen Cutoffs $k PD_C$ erhalten, um die Ampellogik, die zunächst nur auf dem Gesamtscore definiert war, auf die Kategorie-Teile herunter zu brechen.

Um einer Lösung näher zu kommen, ist zunächst festzuhalten, dass wir für jeden Teilscore k die gleiche Skalenteilung verwenden wollen. Wir wollen also nicht eine Kategorie mit einem unverhältnis-

TOP-Seminare für Risikomanagement-Profis

- **Bankenmanagement in der Krise** - Fachkonferenz für Sie als Vorstand oder Führungskraft
 - ▶ 10. bis 11. Juni 2010 in Bonn
- **Financial Engineering** - Kombination von Derivaten und Krediten
 - ▶ 10. bis 11. Juni 2010 in Berlin
- **Vermögen und Verbindlichkeiten außerhalb der Bilanz** (HBG, IFRS)
 - ▶ 14. Juni 2010 in Bonn
- **Liquiditätskosten als eine Komponente in der Preiskalkulation**
 - ▶ 17. Juni 2010 in Bonn
- **Meldewesenüberblick** - Gesetzliche Grundlagen
 - ▶ 13 bis 14. Juli 2010 in Berlin



Detaillierte Informationen zu den Seminaren erhalten Sie in unserem Webshop unter www.bank-verlag-shop.de.

Risikofaktoren in Kategorien mit ihrem mittleren Einfluss, dazu Sensitivitäten und die sich ergebende Ampelfarbe

Kategorie	Einfluss	Risikofaktor	Sensitivität	Abstand vom Mittelwert	Farbe Teilampel	Gesamtampel
Kundendaten	50%	Alter				
		Beruf				
		Familienstand				
		Einkommen				
Externe Daten	30%	Schufa-Auskunft				
		Infoscore-Auskunft				
		Creditreform				
Finanzdaten	20%	Kreditlaufzeit				
		Anzahl Antragsteller				
		Verwendungszweck				

mäßig großen Rotbereich, eine andere mit einem kleinen Rotbereich. Die Ampel der Teilscores sollte genau die gleichen „Abstände“ haben wie die Ampel des Gesamtscores.

Weiterhin ist intuitiv klar, dass die einzelnen Teilkategorien so etwas wie ein Gewicht tragen, also unterschiedlich starken Einfluss auf den Gesamtscore haben können. Wir würden also gerne die Vorstellung formalisieren, dass etwa die Kategorie „Kundendaten“ 50 Prozent beiträgt, während die Kategorie „externe Daten“ nur 30 Prozent, die Kategorie „Finanzdaten“ gar nur 20 Prozent beiträgt.

Gewichte der Kategorien

Will man die erste Anforderung, die nach gleichen Abständen der Ampelskalen und die zweite Anforderung, die nach Gewichten für jeden Teilscore, möglichst einfach realisieren, so bietet sich hierfür eigentlich nur eine Linearkombination der k Teilscore-Cut-Offs zum Gesamt-Cut-Off an. Man kann also schreiben

$$PD_C = \sum_k \kappa_k \cdot {}_k PD_C,$$

wobei die Faktoren κ_k damit nichts anderes sind als „prozentuale“ Beiträge zum Gesamtscore. Die Idee, wie die Faktoren κ_k (also der mittlere Einfluss

eines jeden Merkmals) zu finden sind, ist nun angelehnt an eine multivariate Regression: Auch hier kann man von einem mittleren Einfluss sprechen, wenn man danach sucht, welche unabhängige Variable die abhängige Variable in welchem Ausmaß treibt. Die richtige Repräsentation dieser intuitiv einfachen Idee ist die Aufteilung der Summe der Varianzen aller unabhängigen Variablen in ihre Beiträge. Mittelwert und Varianz erhält man, wenn man über den Gesamtbestand aggregiert:

$$\begin{aligned} E\{P(D|x) - P(D|x_m)\} &\cong \\ \sum_i \partial P(D|x_i) / \partial x_i \cdot E\{(x_i - x_m)\} \\ \text{var}[P(D|x) - P(D|x_m)] &\cong \\ \sum_i \partial P(D|x_i) / \partial x_i \cdot \text{var}\{(x_i - x_m)\} \end{aligned}$$

Die PD wird somit gedacht als über dem Gesamtbestand normalverteilt. Wenn die Teil-PDs ebenso normalverteilte Variablen sind und diese Teil-PDs in Summe die Gesamt-PD ergeben, so wird, sieht man von Korrelationen ab, die Varianz der Gesamt-PD gleich der Summe der Varianzen aller Teil-PDs sein. Teilt man also die Summanden der rechten Seite in der letzten Gleichung durch ihre Summe, werden die „Einflüsse“ der einzelnen Merkmale auf die PD offenbar.

Ein Beispiel: Konsumentencredit-Scoring

Um das obige Konzept greifbarer zu machen, sei ein fiktives Konsumentencredit-Scoring beschrieben, das auf zehn Einzelmerkmalen beruht. In sinnvolle Kategorien gruppiert, sind dies Merkmale aus Kundendaten, externen Daten, und Finanzdaten. Weil das Scoring per Methode der logistischen Regression erstellt wurde, ist klar, dass auf Kategorien-Ebene keine Definition eines Scores (oder einer PD) vorliegt.

Die Tabelle unten zeigt, welchen prozentualen Einfluss auf das Gesamtscore jede der drei Kategorien trägt, wenn man diesen Einfluss mittels Analyse der Varianz über das gesamte Portfolio (siehe Gleichung links) ermittelt. Außerdem kann über die Formel die Sensitivität eines jeden Einzelmerkmals errechnet werden. Diese Sensitivität wird, ebenso wie der Abstand eines betrachteten Beispielfalles grafisch dargestellt, weil eine Wiedergabe der realen Zahlen nicht instruktiv ist, solange die Zahlen von der jeweiligen Repräsentation der Variablen abhängen.

Im Ergebnis erhält man für einen Beispielskunden die in der vorletzten Spalte angeführten Ampelfarben pro Datenart/Kategorie, indem das Summenprodukt aus Sensitivität und Abstand mit der jewei-

ligen Cut-Off-PD verglichen wird (siehe ► **Tab. 01**). Die Detailspalten brauchen dem Kunden gegenüber vor dem Hintergrund der Allgemeinverständlichkeit und Transparenz nicht dargestellt zu werden. Damit kann eine gesetzeskonforme Beauskunftung für jeden Fall, der gescoret wird, in Form einer Teilampel und Gesamtampel erstellt werden.

Technische Umsetzung

Die technische Umsetzung der komplexen mathematischen Anforderung hat eine Vielzahl von Aufgaben zu erfüllen.

Zunächst muss die Umsetzung in einem extrem kurzen Zeitfenster, parallel zur Entwicklung des hier aufgezeigten Modells erfolgen. Das hatte zur Folge, dass die Umsetzung sehr flexibel erfolgen muss. Zusätzlich gilt es eine Umsetzung zu finden, die zukünftig kurzfristig auf Änderungsanforderungen reagieren kann und auch versionsfähig ist, um beispielsweise bei Cut-Off Anpassungen oder Veränderungen einzelner Bepunktungen von Variablen oder dem Austausch von Variablen ohne großen Aufwand in der Programmierung umgesetzt werden kann.

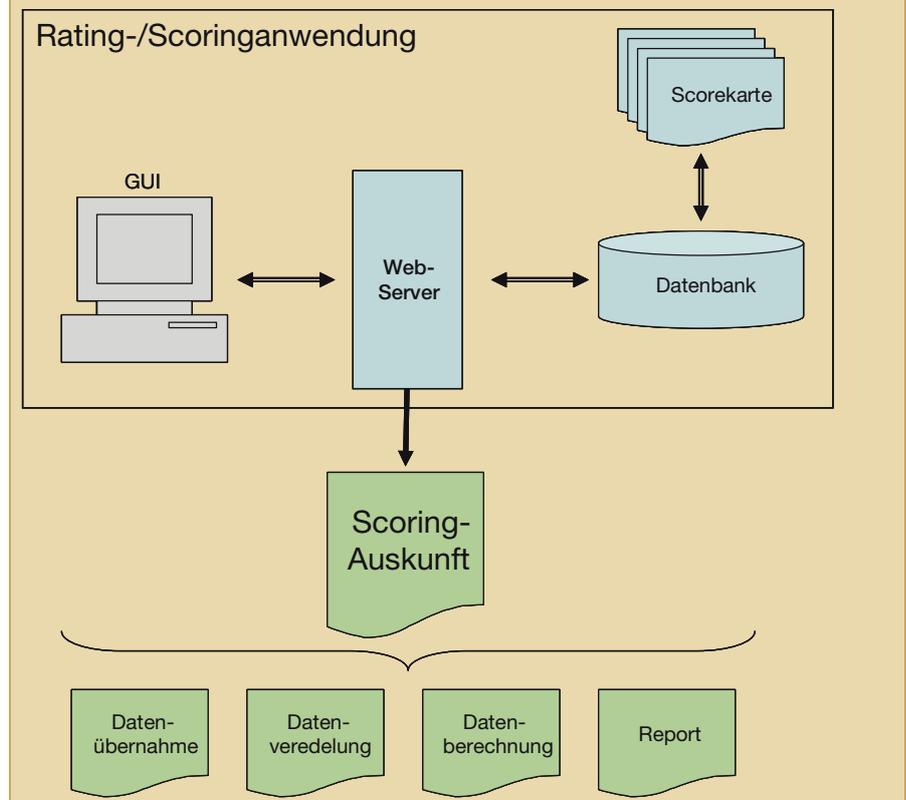
Eine weitere Anforderung besteht in der Option der manuellen Nachbearbeitung, Anpassung und Kontrolle der Scoringauskunft durch die bearbeitende Stelle. Diese manuelle Nachbearbeitung reicht von der Ergänzung im Scoring nicht erfasster Informationen (wie beispielsweise die Adresse) bis zur Übersetzung von Auskünften von Auskunfteien in eine für den Kunden lesbare Form.

Vor diesem Hintergrund ist eine Lösung konzipiert worden, die die im Scoring erfassten Daten vollständig an ein Scoring-Auskunftstool übergibt, das sich aus vier Komponenten zusammensetzt (siehe ► **Abb. 04**).

1. Datenübernahme: Standardisierte Übernahme der vollständigen Daten mit eindeutigen Identifiern je Datenfeld;
2. Datenveredelung: Normalisierung der Daten in einer neutralen Datenschicht;
3. Datenberechnung: Versionsabhängige Berechnung der Teilampeln und der Gesamtampel für die Scoringauskunft;
4. Report: Aggregation der Daten aus der Datenveredelung und Datenberechnung in den für den Kunden ausdrückbaren Report. □

Schematische Architekturdarstellung „Scoring-Auskunft“

► **Abb. 04**



Fazit

Das in diesem Artikel vorgestellte Modell legt die methodischen Grundlagen dar, wie ein beliebiges Scoringssystem mittels Ampelsystem entlang einzelner Datenarten/Kategorien an Kunden kommuniziert werden kann. Gemäß dem neugefassten zweite Absatz in § 34 BDSG haben Kunden einen Auskunftsanspruch gegenüber der Bank bei Einsatz von Scoring-Verfahren. Die Bank braucht für die Erfüllung des Gesetzesanspruchs eine Umsetzung des hier vorgestellten und entwickelten Konzeptes:

- Das Mapping der Einzelfaktoren auf Datenarten/Kategorien, die kommuniziert werden sollen;
- Die Sensitivität des Scorings auf jedes Einzelmerkmal;
- Die Stellung eines mittleren Referenzkundens, gemessen am gesamten Portfolio;
- Den Einfluss (die Varianz) jedes Einzelmerkmals, gemessen am gesamten Portfolio.

In Summe ist das Darstellen der einzelnen Informationsstücke ein nicht unerheblicher Aufwand für Banken, der komplettiert wird durch Anforderungen an die technische Umsetzung:

- auf externe und interne Anforderungen flexibel reagierende Umsetzung;
- versionsfähige Umsetzung für zukünftige Modellanpassungen und parallele Beauskunftung mehrerer Versionen von Scoringmodellen;
- Option zur manuellen Nachbearbeitung, Anpassung und Kontrolle der Scoringauskunft.

Mithin ist zu schließen, dass die Novelle des BDSG an Banken eine Liste an Herausforderungen stellt, die auf den ersten Blick wohl nicht absehbar waren. Die intuitiv einfache Vorstellung von Scoring-Beauskunftung ist jedenfalls weit übertroffen, denn eine geforderte Detailtreue in Kombination mit Kommunizierbarkeit generiert Komplexitäten, die vom Gesetzgeber in dieser Form kaum beabsichtigt sein konnten.

Quellenverzeichnis

BDSG-Novelle I im BGBl. Nr. 48 vom 31. Juli 2009, S.2254-2257

Autoren:

Dr. Markus J. Rieder, MBA, ist selbständiger Datenanalyst im Finanzbereich.

Olaf Schulze ist Project Manager bei der Deutschen Postbank AG.