Wie gut ist Ihr Scoring Modell?

Prof. Dr. Thorsten Hock

Viele Unternehmen beurteilen die Attraktivität ihrer Kunden mit Hilfe von Scoring Modellen. Diese schließen auf Basis historischer Informationen auf zukünftige Kundenereignisse wie Zahlungsverzug, Zahlungsausfall oder Abwanderung. Das System aggregiert verschiedene Daten zu einem Score, der Hinweise auf die Wahrscheinlichkeit eben dieser Ereignisse liefert. Der große Vorteil dieser Modelle liegt in Ihrer Eigenschaft, die vorliegenden Informationen einheitlich, konsistent und insbesondere rational zu verarbeiten und zu bündeln. Mit den Fortschritten der Datenverarbeitung gewinnen Scoring- Ansätze immer mehr an Bedeutung. Insbesondere im Massengeschäft liegt ihr Charme in den hohen Effizienzgewinnen.
Dieser Beitrag beschäftigt sich mit der Frage, wie sich die Leistung von Scoring-Ansätzen – der Fokus liegt dabei auf Modellen, die die Bonität eines Kunden messen sollen (Creditscoring) – beurteilen lässt. Zunächst werden hierzu Anforderungen diskutiert, die man allgemein an Scoringsysteme stellen sollte. Als wesentlich zeigen sich hier die Prognosegüte, die Trennschärfe und die Stabilität. Für jede einzelne Anforderung werden nachvollziehbare Messkonzepte vorgestellt und anhand eines einheitlichen Beispiels erläutert.

ALLGEMEINE ANFORDERUNGEN AN SCORING MODELLE. Ein Credit-Score soll die Wahrscheinlichkeit prognostizieren, dass ein Kunde seinen Zahlungsverpflichtungen gegenüber dem Kreditgeber nicht vollumfänglich nachkommt. Optimale Scoring Modelle erfüllen dabei unter anderem die folgenden Anforderungen. [I]

- Prognosegüte: Einem Scoring-System kann dann eine hinreichende Prognosegüte attestiert werden, wenn gute Scores mit einer tiefen Ausfallwahrscheinlichkeit einhergehen und umgekehrt.

- Trennschärfe: Darunter versteht man die Fähigkeit eines Scoring-Modells, frühzeitig zwischen ausgefallenen Kreditnehmern und nicht ausgefallenen Kreditnehmern zu unterscheiden. Eine entscheidende Bedeutung kommt hier der so genannten Cut-Off-Schwelle zu: Demjenigen Score-Wert, ab dem sich die Entscheidung im Hinblick auf die Kreditvergabe ändert.

- Stabilität: Ein Scoring-Modell wird dann als stabil bezeichnet, wenn sich der Zusammenhang zwischen dem Score eines Kunden und seiner empirischen Ausfallwahrscheinlichkeit über die Zeit nicht ändert.

Zur Veranschaulichung der Berechnungen wird in den folgenden Ausführungen ein Beispiel-Modell verwendet, das den 1.000 Bestandskunden einen Score zwischen 1,00 (= geringstes Ausfallrisiko) und 5,00 (=höchstes Ausfallrisiko) zuordnet. Insgesamt 50 Kunden fallen hiervon am Periodenende aus (Abbildung 1).

ABBILDUNG 1
Exemplarische Belegung eines Scoring-Modells

Scorewerte (1)	Anzahl Schuldner (2)	Anzahl Ausfälle (3)	Anzahl Nicht-Ausfälle (4)
5.00	12	5	7
4.75	22	8	14
4.50	25	8	17
4.25	37	6	31
4.00	54	6	48
3.75	98	6	92
3.50	90	2	88
3.25	87	1	86
3.00	101	2	99
2.75	99	0	99
2.50	89	0	89
2.25	67	2	65
2.00	67	2	65
1.75	54	0	54
1.50	44	1	43
1.25	33	1	32
1.00	21	0	21
Summe	1000	50	950

PROGNOSEGÜTE. Grundsätzlich hängt die Evaluation der Prognosegüte eines Scoring-Systems stark von deren Definition ab. Eine Möglichkeit könnte dabei lauten: Die Prognosegüte ist umso größer, je größer die Differenz zwischen dem Score-Durchschnitt (arithmetischer Mittelwert) der ausgefallenen Unternehmen und dem Score-Durchschnitt der nicht-ausgefallenen Unternehmen. Diese Kennzahl wird als Trenndistanz [II] bezeichnet. Je größer der Wert der Kennzahl, desto besser ist das Scoring,
Der Durchschnitts-Score der Ausfälle beträgt in unserem Beispiel 3,97, derjenige der Nicht-Ausfälle [III] beläuft sich auf 2,87. Für die Trenndistanz resultiert entsprechend ein Wert von 1,10. Ab welcher Trenndistanz kann ein Scoring-Modell letztendlich als gut eingeschätzt werden? Grundsätzlich existieren hierfür keine Richtwerte, weil die Bandbreiten der Werte stark von der Verteilung der errechneten Scores abhängen. Deshalb empfiehlt sich insbesondere eine Analyse der Trenndistanz über die Zeit. Zudem kann mit Hilfe eines statistischen Tests überprüft werden, ob sich die beiden Mittelwerte signifikant unterscheiden.

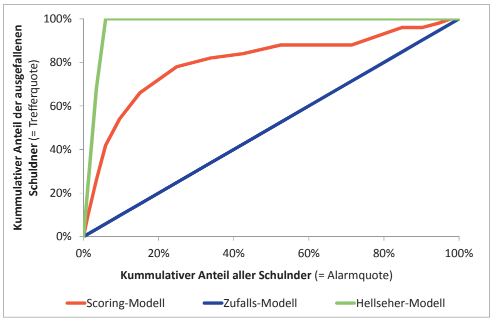
ABBILDUNG 2
Daten zur Bestimmung des Gini-Koeffizienten

			Schuldner		Ausfälle			
Scorewerte (1)	Anzahl Schuldner (2)	Anzahl Ausfälle (3)	Anteil pro Klasse (4)	kummuliert (5)	Anteil pro Klasse Scoringmodell (6)	kummuliert **Scoring-** Modell (7)	kummuliert **Zufalls-** Modell (8)	kummuliert **Hellseher-** Modell (9)
5.00	12	5	1.2%	1.2%	10.0%	10.0%	1.2%	24.0%
4.75	22	8	2.2%	3.4%	16.0%	26.0%	3.4%	68.0%
4.50	25	8	2.5%	5.9%	16.0%	42.0%	5.9%	100.0%
4.25	37	6	3.7%	9.6%	12.0%	54.0%	9.6%	100.0%
4.00	54	6	5.4%	15.0%	12.0%	66.0%	15.0%	100.0%
3.75	98	6	9.8%	24.8%	12.0%	78.0%	24.8%	100.0%
3.50	90	2	9.0%	33.8%	4.0%	82.0%	33.8%	100.0%
3.25	87	1	8.7%	42.5%	2.0%	84.0%	42.5%	100.0%
3.00	101	2	10.1%	52.6%	4.0%	88.0%	52.6%	100.0%
2.75	99	0	9.9%	62.5%	0.0%	88.0%	62.5%	100.0%
2.50	89	0	8.9%	71.4%	0.0%	88.0%	71.4%	100.0%
2.25	67	2	6.7%	78.1%	4.0%	92.0%	78.1%	100.0%
2.00	67	2	6.7%	84.8%	4.0%	96.0%	84.8%	100.0%
1.75	54	0	5.4%	90.2%	0.0%	96.0%	90.2%	100.0%
1.50	44	1	4.4%	94.6%	2.0%	98.0%	94.6%	100.0%
1.25	33	1	3.3%	97.9%	2.0%	100.0%	97.9%	100.0%
1.00	21	0	2.1%	100.0%	0.0%	100.0%	100.0%	100.0%
Summe	1000	50						

Eine weitere Definition lautet: Die Prognosegüte ist umso höher, je besser schwache Bonitäts-Scores mit hohen Ausfallwahrscheinlichkeiten korrespondieren und umgekehrt. Wie kann die Prognosegüte nach dieser Definition abgebildet werden? Die hierbei am weitesten verbreitete Kennzahl ist der Gini-Koeffizient.
Um die grundlegende Intuition zu vermitteln, unterstellen wir zunächst einen Hellseher. Dieser hat ein System entwickelt, das perfekt die Ausfälle von den Nicht-Ausfällen im Vorfeld unterscheiden kann. In Abbildung 3 wird die Performance des Hellsehers grafisch illustriert (grüne Kurve). Dieses Modell weist den ausgefallenen Schuldnern ex ante die schlechteste Score-Klasse zu. Die grüne Kurve verläuft zunächst linear nach oben. Ab dem Punkt 5% auf der X-Achse (insgesamt 5% Ausfälle in der Grundgesamt) und 100% auf der Y-Achse (alle Ausfälle wurden vom System erkannt) verläuft die Hellseher-Kurve dann horizontal. [IV]

ABBILDUNG 3

GAP-Kurven verschiedener Modelle



Neben dem Hellseher existiert auch ein Glückspieler, der den Kunden rein zufällig Scores zuweist. Die blaue Kurve in der Grafik liegt auf der Winkelhalbierenden und repräsentiert genau dieses Vorgehen, denn statistisch gesehen erwischt der Glückspieler beispielsweise bei 20% der Schuldner insgesamt auch 20% der Kreditausfälle.
Unser Scoring-Modell wird in Abbildung 3 von der roten Kurve beschrieben und liegt erwartungsgemäß zwischen diesen beiden Extremlösungen.[V] Je näher nun die rote Kurve an der grünen Kurve liegt, desto besser kann das Scoring-Modell ex ante zwischen Ausfällen und Nicht-Ausfällen unterscheiden. Der Gini-Koeffizient versucht genau diesen Sachverhalt zu quantifizieren. Dazu setzt er die Fläche zwischen der roten und der blauen Kurve ins Verhältnis zur Fläche zwischen der grünen und der blauen Kurve. Je besser die Prognose des Scoring-Modells, desto näher liegt die rote Kurve an der grünen – und das beschriebene Verhältnis geht gegen 100%. Entsprechend haben Scoring-Systeme, die kaum besser als der Zufall sind, Werte für den Quotienten nahe 0%. Wie können wir die Flächeninhalte nun einfach berechnen?

Abbildung 4 zeigt die Ergebnisse. Um das Vorgehen bei der Berechnung des Flächeninhaltes zu verdeutlichen, soll das Scoring-Modell dienen. Die Fläche unter der schlechtesten Ratingklasse – Scorewert von 5,00 – kann mit einem Dreieck beschrieben werden, dass die 3 Eckkoordinaten (0%; 0%), (1,2%; 0%) und (1,2%; 10,0%) aufweist.[VI] Dieses hat einen Flächeninhalt von (1,2%*10,0%)/2=0,0006. Im Gegensatz dazu bildet die Fläche unter dem Scorewert von 4,75 ein Fünfeck. Dieses setzt sich aus dem Viereck mit der horizontalen Kantenlänge von 2,2% und der vertikalen Katenlänge von 10% zusammen und hat einen Flächeninhalt von 0,0022. Hinzu kommt das Dreieck mit den Koordinaten (1,2%; 10,0%), (3,4%; 10,0%) und (3,2%; 26,0%). Dessen Flächeninhalt liegt bei (2,2%*16,0%)/2=0,0018, was insgesamt die 0,0040 aus Abbildung 4 ergibt.

ABBILDUNG 4

Berechnung der Flächeninhalte

	Flächeninhalt unter		
Scorewerte (1)	Scoring-Modell (2)	Zufalls-Modell (3)	Hellseher-Modell (4)
5.00	0.0006	0.0001	0.0014
4.75	0.0040	0.0005	0.0101
4.50	0.0085	0.0012	0.0210
4.25	0.0178	0.0029	0.0370
4.00	0.0324	0.0066	0.0540
3.75	0.0706	0.0195	0.0980
3.50	0.0720	0.0264	0.0900
3.25	0.0722	0.0332	0.0870
3.00	0.0869	0.0480	0.1010
2.75	0.0871	0.0570	0.0990
2.50	0.0783	0.0596	0.0890
2.25	0.0603	0.0501	0.0670
2.00	0.0630	0.0546	0.0670
1.75	0.0518	0.0473	0.0540
1.50	0.0427	0.0407	0.0440
1.25	0.0327	0.0318	0.0330
1.00	0.0210	0.0208	0.0210
Summe	0.8018	0.5000	0.9736

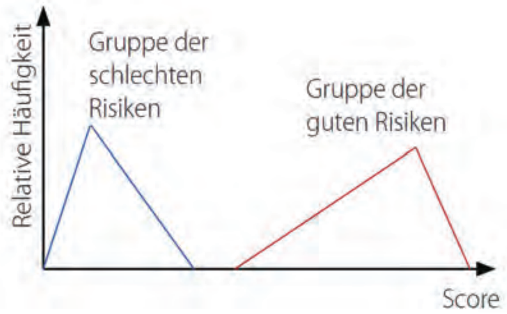
Mit diesem Vorgehen können auf einfache Art und Weise die Flächeninhalte der 3 Modelle bestimmt werden.[VII] Der Gini-Koeffizient bildet sich nun folgendermaßen:

$$\text{Gini} - \text{Koeffizient} = \frac{\text{Fläche Scoringmodell} - \text{Fläche Zufallsmodell}}{\text{Fläche Hellsehermodell} - \text{Fläche Zufallsmodell}} = \frac{0{,}8018 - 0{,}5}{0{,}9736 - 0{,}5} = 63{,}7\%$$

Sein Wertebereich liegt wie oben angedeutet zwischen 0% und 100%. Ab welchem Gini-Koeffizienten ein Ratingsystem nun gute Qualität aufweist, kann nur in der relativen Betrachtung beantwortet werden. Im B2B-Bereich liegen sehr gute Credit-Scoring-Modelle im Bereich von 55-70%.[VIII] Nach Erfahrungen von Munch (2009, S. 212) sind im Geschäftsfeld B2C Scoring-Modelle Gini-Koeffizienten zwischen 45% und 50% realistisch. Ein zentraler Nachteil der Beurteilung über diese Kennzahl ist, dass wichtige Informationen wie bspw. die Ausfallwahrscheinlichkeit eines Kunden bzw. eines Portfolios nicht Gegenstand der Evaluation sind.

TRENNSCHÄRFE. Die für den Entscheidungsträger in der Praxis entscheidende Frage lautet: Soll dem Kunden mit einem Score von XY ein Kredit gewährt werden oder nicht? Oder konkreter: Was ist der exakte Schwellenwert des Scoring-Systems, ab dem einem Kunden der Kredit verweigert werden sollte? Im Idealfall trennt ein Schwellenwert die Kunden ex ante in die Gruppe der guten Risiken und die Gruppe der schlechten Risiken. Abbildung 5 zeigt exemplarisch den Zusammenhang:

ABBILDUNG 5
Scoring bei optimaler Trennschärfe
(Quelle: Schröder et. al. (2014), S. 34)



In einem perfekten Scoring-Modell wären die Scores der schlechten Risiken ausnahmslos höher als diejenigen der guten Risiken. Für ein solches System könnten sehr einfach optimale Schwellenwerte gefunden werden. Diese sollten tiefer als die besten Scores der schlechten Risiken und höher als die schlechtesten Scores der guten Risiken sein. Leider weichen die Scoring-Modelle in der Praxis von diesen optimalen Bedingungen ab.

ABBILDUNG 6
Scoring bei realistischer Trennschärfe
(Quelle: Schröder et. al. (2014), S. 34)

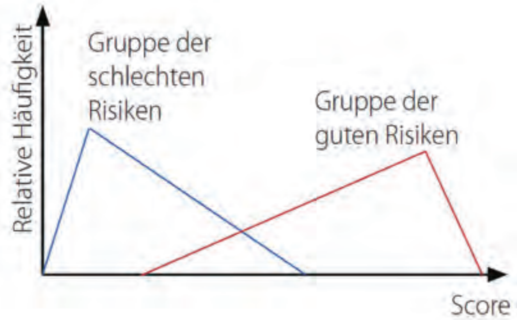


Abbildung 6 zeigt ein realistisches Bild, denn in der Praxis sind Fehlentscheidungen bei der Kreditvergabe unvermeidlich: Kunden mit verhältnismäßig guten Scores fallen aus und Kunden mit verhältnismäßig schlechten Scores kommen ihren Kreditverpflichtungen nach. Für die Wahl des Schwellenwertes (sogenannter Cut-Off) gilt entsprechend:

- Wird der Schwellenwert sehr hoch angesetzt (werden also auch Kreditnehmer mit schwächeren Ratings akzeptiert), dann fallen viele Kunden aus, die als kreditwürdig klassifiziert werden. Wir sprechen hier vom sogenannten Fehler erster Art (Alpha Fehler). Die Kosten für diesen Fehler sind umso höher, je größer die Kreditbeträge und je tiefer die Umsatzmargen des Unternehmens sind.
- Wird dagegen der Schwellenwert sehr tief angesetzt, dann erhalten viele Kunden keinen Kredit, obwohl sie in der Lage wären ihren Verpflichtungen vollumfänglich nachzukommen. Wir sprechen hier vom sogenannten Fehler zweiter Art (Beta-Fehler). Hier fallen Kosten in Form entgangener Gewinne (so genannte Opportunitätskosten) und verlorener Marktanteile an, weil auch ex post gute Risiken keine Kredite erhalten würden.

ABBILDUNG 7
Belegung und Gütemaße der Trennschärfe

	Belegung pro Klasse			Gütemaße		
Schwellenwerte (1)	Kunden (2)	Ausfälle (3)	Nicht-Ausfälle (4)	Alpha-Fehler (5)	Beta-Fehler (6)	Klassifikationsfehler (7)
				1.000	0.000	0.750
5.00	12	5	7	0.900	0.007	0.677
4.75	22	8	14	0.740	0.022	0.561
4.50	25	8	17	0.580	0.040	0.445
4.25	37	6	31	0.460	0.073	0.363
4.00	54	6	48	0.340	0.123	0.286
3.75	98	6	92	0.220	0.220	0.220
3.50	90	2	88	0.180	0.313	0.213
3.25	87	1	86	0.160	0.403	0.221
3.00	101	2	99	0.120	0.507	0.217
2.75	99	0	99	0.120	0.612	0.243
2.50	89	0	89	0.120	0.705	0.266
2.25	67	2	65	0.080	0.774	0.253
2.00	67	2	65	0.040	0.842	0.241
1.75	54	0	54	0.040	0.899	0.255
1.50	44	1	43	0.020	0.944	0.251
1.25	33	1	32	0.000	0.978	0.244
1.00	21	0	21	0.000	1.000	0.250
Summe	1000	50	950			

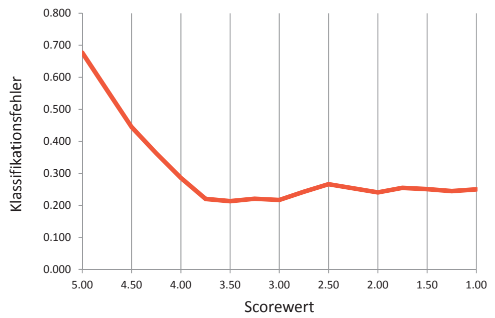
Für unser Beispiel errechnen sich die Fehlentscheidungen in Abhängigkeit vom Schwellenwert wie folgt (siehe Abbildung 7): Erhält jeder interessierte Kunde unabhängig vom Score einen Kredit, dann ist der Fehler 1. Art maximal – also bei 100%. In diesem Fall wird keiner der 50 Ausfälle im Beispielportfolio verhindert. Der Beta-Fehler liegt in diesem Fall dagegen bei Null. Zunächst wird die Schwelle bei einem Scorewert von 5,00 festgelegt. Entsprechend erhalten nur Kunden mit einem besseren Rating einen Kredit. In diesem Fall kommen genau 45 Kunden ihren Verpflichtungen nicht nach. Allerdings können jetzt 5 Ausfälle verhindert werden. Der Alpha-Fehler sinkt demnach auf 90,0% (Ausfälle bei Schwellenwert/Ausfälle gesamt = 45/50 = 0,90). Die neue Strategie führt aber auch dazu, dass insgesamt 12 Kunden ein Kredit verweigert wird. Entsprechend steigt der Beta-Fehler. Er liegt jetzt bei 0,7% (Nicht-Ausfälle bei Schwellenwert/Nicht-Ausfälle gesamt = 7/950 = 0,007). Analog errechnet sich bei einer gewählten Schwelle von 4,75 ein Alpha Fehler von 74,0% (= 37/50) und ein Beta-Fehler von 2,2%(= 21/950).

Die Zusammenschau zeigt: Legt man einen sehr ambitionierten Schwellenwert fest, dann minimiert man zwar die Ausfälle (= Fehler erster Art), verhindert auf der anderen Seite aber eine lukrative Kreditbeziehung mit im Nachhinein solventen Kunden (= Beta-Fehler). Wie kann vor diesem Hintergrund ein optimaler Schwellenwert ermittelt werden? Die Beantwortung dieser Frage beginnt zunächst mit der Berechnung der durchschnittlichen Kosten für ein nicht-abgelehntes, schlechtes Risiko (= Kosten pro «Fehler erster Art –Kunde») und den durchschnittlichen Kosten für einen «Fehler zweiter Art – Kunden».[IX] Das Verhältnis dieser beiden Kostenarten ermöglicht es, einen gewichteten Klassifikationsfehler (Spalte 7 in Abbildung 7) zu berechnen. Sind beispielsweise bei einem ausgewählten Schwellenwert die Kosten für den Fehler erster Art um den Faktor drei größer als die Kosten für den Fehler zweiter Art, dann gilt:

$$\text{Klassifikationsfehler} = 0{,}75 * \text{Fehler erster Art} + 0{,}25 * \text{Fehler zweiter Art}$$

ABBILDUNG 8

Gewichteter Klassifikationsfehler in Abhängigkeit vom Schwellenwert



Im Beispielbestand sinkt bei sukzessiver Verschärfung der Kreditvergabe zunächst der Klassifikationsfehler. (Abbildung 8). Erst ab einem Scorewert von 3,75 bleibt der Klassifikationsfehler stabil. Folglich sollten die Entscheidungsträger hier die Schwelle ansetzen. Kunden mit einem Score von unter 3,75 sollten Kredite erhalten, Kunden mit 3,75 und schlechter sollten abgelehnt werden.

STABILITÄT.[X] Eine letzte Anforderung an ein adäquates Scoring-System ist dessen Stabilität. Sie zeichnet sich durch eine nachhaltige Modellierung des Zusammenhanges zwischen den Risikomerkmalen und der Bonität aus. Scheinabhängigkeiten und Überspezifikationen sollen vermieden werden. Denn hieraus resultieren meistens instabile Modelle, deren Prognosegüte sich im Zeitablauf stark verschlechtert.[XI] Erweitern wir nun unser Beispiel, in dem wir die ursprüngliche Evaluierungsperiode (Periode 1) um eine weitere (Periode 2) ergänzen.
Stabilität kann mit so genannten Stabilitätsindizes (StI) gemessen werden. Hierbei wird für jede Ratingklasse – gemessen am Score-Wert – der Kundenanteil in Periode 1 mit demjenigen in Periode 2 verglichen. Die Formel lautet:

$$\mathrm{StI} = \sum \left(\mathrm{Anteil}_{\text{Klasse i,Periode 1}} - \mathrm{Anteil}_{\text{Klasse i,Periode 2}}\right) \cdot \log \frac{\mathrm{Anteil}_{\text{Klasse i,Periode 1}}}{\mathrm{Anteil}_{\text{Klasse i,Periode 2}}}$$

So errechnet sich der StI für den Score-Wert 5,00 des gesamten Kundenportfolios (Abbildung 9, Spalte 6) als:

$$\mathrm{StI(Score\ 5{,}00)} = \sum \left(\frac{12}{1000} - \frac{4}{500}\right) \cdot \log \left(\frac{\frac{12}{1000}}{\frac{4}{500}}\right) = 0{,}0007$$

Die Kalkulationen liefern für den gesamten Bestand einen StI von 0,083 und bei ausschließlicher Berücksichtigung der Ausfälle einen StI von 0,212. Die Interpretation erfolgt über eine Gegenüberstellung mit Erfahrungswerten. Anderson (2007, S. 194) schlägt folgende Einteilung vor: für StI < 0,10 gibt es keine Änderungen im Kundenbestand, bei 0,10 < StI < 0,25 treten kleinere Diskrepanzen auf und ein StI > 0,25 weißt auf signifikante Veränderungen im Kundenbestand hin. Vor diesem Hintergrund kann in unserem Beispiel der gesamte Kundenbestand und die Nicht-Ausfälle als stabil angesehen werden, wohingegen die Teilmenge der ausgefallenen Kunden kleine Veränderungen in Periode 2 verglichen mit Periode 1 aufweist.

ABBILDUNG 9

Stabilitätsanalyse des Scoring-Modells

Schwellen-werte (1)	Periode 1 Kunden (2)	Periode 1 Ausfälle (3)	Periode 2 Kunden (4)	Periode 2 Ausfälle (5)	Stabilitätsindex (StI) StI Gesamt (6)	StI Ausfälle (7)	StI Nicht-Ausfälle (8)	Verschiebungsindex (ShI) ShI Gesamt (9)	ShI Ausfälle (10)	ShI Nicht-Ausfälle (11)
5.00	12	5	4	1	0.0007	0.0239	0.0001	0.0200	0.3000	0.0053
4.75	22	8	6	2	0.0026	0.0241	0.0015	0.0475	0.3800	0.0300
4.50	25	8	12	3	0.0000	0.0050	0.0000	0.0045	0.1800	-0.0047
4.25	37	6	21	4	0.0003	0.0050	0.0001	-0.0213	-0.1700	-0.0134
4.00	54	6	13	2	0.0089	0.0070	0.0093	0.1120	0.1600	0.1095
3.75	98	6	18	2	0.0270	0.0070	0.0290	0.2325	0.1500	0.2368
3.50	90	2	37	3	0.0014	0.0382	0.0024	0.0560	-0.2800	0.0737
3.25	87	1	41	3	0.0001	0.0778	0.0006	0.0163	-0.3250	0.0342
3.00	101	2	35	0	0.0049	0.0000	0.0046	0.0930	0.1200	0.0916
2.75	99	0	52	0	0.0001	0.0000	0.0001	-0.0138	0.0000	-0.0145
2.50	89	0	42	0	0.0001	0.0000	0.0001	0.0125	0.0000	0.0132
2.25	67	2	45	0	0.0029	0.0000	0.0037	-0.0518	0.0900	-0.0592
2.00	67	2	61	2	0.0143	0.0120	0.0144	-0.1100	-0.0800	-0.1116
1.75	54	0	52	1	0.0142	0.0000	0.0140	-0.0875	-0.0700	-0.0884
1.50	44	1	23	1	0.0000	0.0060	0.0000	-0.0030	-0.0300	-0.0016
1.25	33	1	18	1	0.0001	0.0060	0.0001	-0.0037	-0.0250	-0.0026
1.00	21	0	20	0	0.0053	0.0000	0.0056	-0.0190	0.0000	-0.0200
Summe	1000	50	500	25	0.0831	0.2121	0.0855	0.2843	0.4000	0.2782

Was sind die Gründe für die Veränderung der Stabilität? Eine Antwort hierauf können sogenannte Verschiebungsindizes (ShI) geben. Dieses Instrument der Veränderungsanalyse ermittelt die Verschiebungsrichtung und die Größe des Einflusses auf den Score. Der ShI kann folgendermaßen berechnet werden:

$$\mathrm{ShI} = \sum \left(\mathrm{Anteil}_{\text{i,Periode 1}} - \mathrm{Anteil}_{\text{i,Periode2}}\right) \cdot \mathrm{Scorewert}_{\mathrm{i}}$$

In unserem Beispiel errechnet sich entsprechend der erste Summand für den gesamten Kundenbestand (Abbildung 9, Spalte 9) als:

$$\mathrm{ShI\ (Score\ 5{,}00)} = \left(\frac{12}{1000} - \frac{4}{500}\right) \cdot 5{,}00 = 0{,}0200$$

Der ShI-Indexwert von +0,2843 für den gesamten Kundenbestand gibt an, dass die Scores des Kundenbestands aus Periode 2 um durchschnittlich ein 0,2843 Score-Punkte höher – also schlechter – ausfallen als in der Periode 1.[XII] StI und ShI liefern in Verbindung mit den bereits diskutierten Maßzahlen wie Gini-Koeffizient und Trenndistanz für beide Perioden wertvolle Inputfaktoren zur Qualitätsbeurteilung des Scoring-Systems.

SCHLUSSBEMERKUNGEN. Der vorliegende Beitrag möchte dem Leser einen Überblick vermitteln, wie die Qualität von Scoring-Systemen quantitativ evaluiert werden kann. Aber auch qualitative Aspekte spielen bei der Modellbeurteilung eine ebenso zentrale Rolle. Entscheidend ist hierbei die Datenqualität. Sie ist zum einen ausreichend erfüllt, wenn eine konsistente Erhebung der Daten gesichert ist. Zudem muss der Datenumfang auch groß genug sein, damit statistisch gesicherte Aussagen überhaupt möglich sind. Zentral ist dabei eine hinreichende Anzahl dokumentierter Ausfälle. Sie sind Voraussetzung dafür, dass die prognostischen Eigenschaften der Risikomerkmale aufgeschlüsselt und eindeutig nachgewiesen werden können. Nur dann führt die quantitative Herangehensweise zu optimalen Kredit- und Limit-Entscheidungen bei überschaubarem Aufwand.

AUTOR

Prof. Dr. Thorsten Hock lehrt an der Ostbayerisch Technischen Hochschule Amberg-Weiden und ist für den Fachbereich Finanzen und quantitative Methoden zuständig. Er berät Unternehmen bei der risikoorientierten Kundenbewertung.

Literaturverzeichnis

- Anderson (2007), «The Credit Scoring Toolkit: Theory and Practice for Retail Credit Risk Management and Decision Automation», Oxford University Press.
- Deutsche Bundesbank (2003), Validierungsansätze für interne Rating Systeme, Monatsbericht 9/2003, S. 61 bis 74.
- Khomski (2012), «Einige Validierungsaspekte von Scoring-Systemen», Fachbeitrag der 1 Plus i GmbH 1/12, S. 1 bis 7.
- Munch (2009), «Wie steht es um Ihr Scoring ? – Das Rating von Scoring-Systemen», Finanzierung – Leasing – Factoring 5/2009, S. 209 bis 213.
- Schröder, Lang, Lerbs und Radev (2014), «Ökonomische Bedeutung und Funktionsweise von Credit Scoring», in: Schröder, Traeger (Hrsg.), «Scoring im Fokus: Ökonomische Bedeutung und rechtliche Rahmenbedingungen im internationalen Vergleich», BIS Verlag der Carl von Ossietzky Universität Oldenburg, S. 8 bis 90.
- Schulte-Mattler, Daun und Manns (2004), «Trennschärfemaße zur Validierung von internen Rating-Systemen», RATINGaktuell 06/2004, S. 46 bis 52.

Fussnoten

[I] Die Unterscheidung zwischen Prognosegüte und Trennschärfe ist unter Berücksichtigung der einschlägigen Fachliteratur nicht unbedingt üblich. Die Unterscheidung findet ausschließlich zum Zwecke einer besseren Systematisierung der vorgestellten Maßzahlen statt.

[II] Vgl. bspw. Munsch (2009, S. 212).

[III] Berechnung (siehe Abbildung 1): Jeder Wert in (1) muss mit dem dazugehörigen Wert aus (3) multipliziert werden. Summiert man diese Produkte für alle Score-Klassen auf und teilt durch die Anzahl 50, dann resultiert der Durchschnittsscore 3,97.

[IV] Die dazugehörigen X-Werte des Hellsehermodells entsprechen der Spalte (5) und die Y-Werte der Spalte (9) aus Abbildung 2.

[V] Die dazugehörigen X-Werte des Scoring-Modells entsprechen der Spalte (5) und die Y-Werte der Spalte (7) aus Abbildung 2.

[VI] Siehe Abbildung 2, erste Zeile und Spalte (5) bzw. (7) für die Werte.

[VII] Der gesamte Flächeninhalt des Zufallsmodells errechnet sich viel einfacher mit (100%*100%)/2=0,5.

[VIII] Siehe bspw. Schulte-Mattler et. al. (2004, S. 50). Diese Einschätzung für sehr gute Scorings-Systeme für Geschäftskunden teilt Munsch (2009, S. 212).

[IX] Die Kosten für einen Fehler zweiter Art Kunden könnte approximativ berechnet werden als der gesamte Rohertrag aller Kunden geteilt durch die Anzahl der Nicht-Ausfälle.

[X] Der Abschnitt bezieht sich größtenteils auf Khomski (2012) und die dort diskutierte Literatur.

[XI] Vgl. Deutsche Bundesbank (2003, S. 64).

[XII] Ein negatives Vorzeichen des ShI würde eine Verbesserung in der Score-Belegung in Periode 2 implizieren.