Bayesianisches Kredit-Scoring

Dr. Markus J. Rieder, data2impact

Abstract
Das Scoring von Kreditnehmern hinsichtlich ihres Ausfallrisikos ist ein Musterbeispiel für die praktische Tauglichkeit der Bayesianischen Statistik. Durch die einfache und klare Belegung der *a priori*-Verteilung und durch die intuitiv verständliche sowie direkt zu erhaltende *Likelihood*-Verteilung lässt sich eine robuste a-posteriori-Verteilung schätzen, die dazu dient, Kreditnehmern bei Antrag eine Ausfallwahrscheinlichkeit zuzuordnen. Wir schlagen ein Scoring-Verfahren vor, das die erwarteten Ausfälle explizit modelliert und mit der Verteilung von bekanntgewordenen Ausfällen verknüpft, jedem Kunden gemäß seiner Antragsdaten einen einfachen und konsistenten Schätzer seiner Ausfallwahrscheinlichkeit zuweist. Auf Basis von 39.000 Gewerbekunden mit je sechs Kundenmerkmalen wird die Vorhersagefähigkeit des Verfahrens demonstriert und der einer logistischen Regression gegenübergestellt. Es zeigt sich, dass das Bayesianische Verfahren bei vergleichbarer Trefferquote deutlich verständlicher und einfacher zu handhaben ist als die Methoden der klassischen Statistik. Somit ist das Bayesianische Kredit-Scoring eine Methode, die die traditionelle Gegensätzlichkeit der Expertenverfahren und statistischen Verfahren aufhebt und in ein gemeinsames Verfahren überführt.

1 Einleitung

Die klassischen Methoden zum Kredit-Scoring sind großteils ausgereizt und deren Vor- und Nachteile den Anwendern in der Praxis sattsam bekannt. Neben der Tatsache, dass die Punkte in den Scorekarten kaum zu interpretieren sind, wird die klassische Statistik nicht damit fertig, dass die Scorepunkte bei kleinen Datenmengen große Schwankungsbreiten aufweisen und durch unterbestimmte Regression die ganze Scorekarte instabil wird. Ausserdem ist es bei den klassischen Verfahren nur schwer möglich, das Wissen um die durchschnittliche Ausfallhäufigkeit einzubringen, und ein jedes Re-Scoring von im Zeitablauf sich verändernden Beständen ist mit größerem Aufwand verbunden als es nötig erscheint.
Von den Bayesianischen Methoden hingegen ist bekannt, dass sie abzubilden imstande sind, was intuitiv klar erscheint, und dabei auch für kleine Datenmengen stabile Schätzer liefern (für eine Einführung in die Bayesianische Statistik siehe etwa Berry, 1996). Das wird erreicht durch Einbeziehen von Information, die bereits vor Analyse der Daten vorhanden ist. Die Daten dienen somit als „Anreicherung“ dieser Vorinformation. Dieses Setup hat den Vorteil, dass auch qualitative Informationen zwanglos eingebracht werden können und erleichtert zum anderen das Neu-Analysieren von Daten, weil die Bayesianischen Verfahren sich aufgrund ihrer Updating-Relation für einen (zeit-)dynamischen Betrieb anbieten.
Aus dieser allgemeinen Einschätzung heraus wird evident, dass sich Bayesianische Verfahren zum Kredit-Scoring eignen. Die vorliegende Arbeit schlägt ein solches Verfahren vor und vergleicht es auf Grund von Antragsdaten aus dem Bereich selbständiger Gewerbekunden einer Retailbank mit den klassischen Verfahren.
Im zweiten Abschnitt werden kurz die wichtigsten klassischen Verfahren vorgestellt, der dritte Abschnitt zeigt, wie der Satz von Bayes zum Kredit-Scoring ausgebaut werden kann, worin seine Stärken bestehen, und dass die Bayesianische Statistik formal exakt auf eine vielmals vorgeschlagene Experten-Scorekarte führt. Im vierten Abschnitt wird das Verfahren einem Praxis-Test unterzogen, und im letzten Abschnitt zeigt die Zusammenfassung, dass die

vorgelegte Methode imstande ist, die sich traditionell gegenüberstehenden statistischen und expertenbefundenen Verfahren zu vereinen.

2 Klassische Verfahren des Kredit-Scorings

Die Methoden zum Kreditantrag-Scoring sind seit Anfang der 70er Jahre in der Theorie eingehend untersucht und in der Praxis umfangreich, vor allem zur Klassifizierung von Privatkunden, angewendet worden. Traditionell stehen sich dabei zwei Schulen gegenüber.

2.1 Prinzipielle Ansätze

Ein erster Ansatz zum Kredit-Scoring sind die expertenbasierten Verfahren. Diese basieren auf der Überzeugung von Kreditanalysten, bestimmte Risikotreiber aus ihrem ökonomischen Verständnis des Bedienens einer Finanzbelastung ableiten zu können. Dieses Wissen kann in Expertenkarten umgesetzt und damit zur Beurteilung eines Kredites genutzt werden.
Neben dem in quantitative Form gegossenen Expertenwissen der Kreditentscheider ist es die klassische Statistik, die als prinzipielle Zugänge vor allem Klassifikationsbäume und Regressionen vorgeschlagen hat, um das Wissen über historische Ausfälle in die Kreditvergabe einbeziehen zu können.
Die viel beschworene Unvereinbarkeit expertenbasierter versus statistischer Verfahren ist auf die Gegensätzlichkeit entlang zweier Dimensionen zurückzuführen. Zum einen ist das die Frage, ob ein Kredit individuell oder im Umfeld aller anderen Kredite gesehen wird. Die Expertenverfahren favorisieren sehr stark eine Einzelgeschäftssicht, währenddessen die Statistik sich ganz klar aus einer Portfolio-Sicht ableitet. Zum anderen scheinen die Expertenverfahren den statistischen Verfahren entgegenzustehen, weil erstere sich auf Kausalitäten berufen, wo letztere nur die Diskriminanz zwischen historisch guten und historisch schlecht gewordenen Krediten optimieren.
Argumente in beiden Dimensionen fachen den Streit über einen besten Zugang zum Kredit-Scoring immer wieder an. Der hier entwickelte Bayesianische Zugang zeigt, dass die beiden traditionellen Zugänge sich nicht ausschließen, sondern elegant in eine gemeinsame Form gegossen werden können.

2.2 Expertenkarten

Abseits eines statistischen Zuganges zum Scoring sind es Kreditsachbearbeiter selbst, die ihr Wissen und ihre Erfahrung bei der Kreditvergabe einbringen können. Dabei wird anhand sogenannter „Scorekarten“ versucht, Kreditnehmer gemäß ihrer Eigenschaften mit Punkten zu bewerten, wobei die gesamte Punkteanzahl einem Maß für die Bonität des Schuldners gleichkommt. Die Merkmale der Scoretabellen bilden ab, was die Kreditsachbearbeiter zur Beurteilung der Rückzahlungsfähigkeit fachlich für richtig und wichtig halten, und die Gewichtung der Punkte spiegelt wider, wie sehr die einzelne Merkmalsausprägung als bonitätssteigernd eingeschätzt wird.
Der Vorteil von Expertenkarten besteht darin, dass die vergebenen Punkte einfach zu verstehen sind und die tatsächliche Einschätzung der Kreditentscheider abbilden, was in dem meisten Fällen für hohe Akzeptanz sorgt. Die Schwäche des Verfahrens besteht darin, dass die Experten, welche das Punktesystem installieren, kaum die Abhängigkeit der einzelnen Merkmale untereinander zu berücksichtigen vermögen. Die Multivariabilität ist somit eliminiert, was sich in einer niedrigeren Performance dieser Systeme äussert.

2.3 Klassifikationsbäume

Unter Klassifikations- und Regressionsbäumen subsummieren sich Methoden, die mit Hilfe der bekannten Merkmale versuchen, die Grundgesamtheit aller Kreditnehmer vorab genau so in Teile zu schneiden, dass diese sich in ihrem Ausfallverhalten möglichst stark voneinander unterscheiden. Durch eine Reihe von Schnitten ist das Ziel der Verfahren also, Zerlegungen zu erhalten, die möglichst wenig bzw. möglichst viele der Ausfälle enthalten.
Vorteil dieser Verfahren ist es etwa, dass ein und dasselbe Merkmal in mehreren Schnitten genutzt werden kann und daher Nichtlinearitäten aufgelöst werden können. Nachteil der Baumverfahren ist es, dass zu deren Kalibrierung sehr viele Daten (vor allem: sehr viele Ausfälle) zur Verfügung stehen müssen, da sonst die Statistiken nach nur wenigen Zerlegungen insignifikant werden. Im Falle zu weniger Zerlegungen wird das Verfahren auch zu „grobkörnig“, d.h, es können nicht ausreichend viele verschiedene Risikoklassen gebildet werden.

2.4 (Logistische) Regression

Die Regression, allen voran in der Form eines Logit-Modells, ist die am häufigsten anzutreffende statistische Methode im Kredit-Scoring. Dabei wird eine die Ausfallneigung kennzeichnende abhängige Variable als Funktion einer gewichteten Summe der beschreibenden Merkmale modelliert. Ist diese Funktion die logistische Verteilung, so bezeichnet man das Modell als logistische Regression. Bei der logistischen Regression ist es naheliegend, die abhängige Variable als Ausfallwahrscheinlichkeit zu interpretieren. Diese Interpretation hat sie mit den Klassifikationsbäumen und dem Bayesianischen Verfahren gemeinsam, was mit ein Grund für die Beliebtheit in der Praxis ist.
Vorteil eines stetigen Mappings auf die Ausfallwahrscheinlichkeit ist die theoretisch unendlich „feinkörnige“ Risikodifferenzierung. Die logistische Regression wird weiters geschätzt aufgrund ihrer Robustheit, die das Verfahren auch bei wenigen Daten anwendbar macht, und aufgrund der Verfügbarkeit der Rechenlogik in den statistischen Standard-Softwarepaketen. Nachteil der logistischen Regression sind die Schwierigkeiten, die mit der Interpretation der Regressionskoeffizienten verbunden sind. Die Gewichte der Scorekarten (Scorepunkte) sind in ihrer absoluten Höhe nur kaum zu interpretieren und schwierig zu kommunizieren, durch die Nichtlinearität der logistischen Verteilung entziehen sich auch die relativen Stellungen der Scorepunkte einer einfachen Deutung. In ihrer Anwendung auf das Kredit-Scoring ist die logistische Regression auch dort nicht ohne Vorskalierung anwendbar, wo Ausfälle nur sehr unterrepräsentiert auftreten.

3 Bayesianisches Schliessen

3.1 Satz von Bayes

Der Satz von Bayes eignet sich dazu, aus Daten Schätzungen über Zustände abzuleiten, ohne ein spezifisches Modell über das Verhältnis von Daten und Zuständen zu hinterlegen (Gelman et al., 1995). Mittels Bayesianischen Schließens kann eine nicht-parametrische Schätzgleichung aufgestellt werden, die die Wahrscheinlichkeitsverteilung der Zustände wiedergibt, wenn man die Daten kennt. Allgemein gilt

$$P(x \mid y) = P(y \mid x) * P(x) / P(y) \qquad \textit{(Gl. 1)}$$

In obiger Gleichung ist x der Zustand, y die darüber gemessenen Daten. P(x) ist die *a priori* Wahrscheinlichkeit, P(y | x) die sogenannte *Likelihood*. Erstere spiegelt das Wissen wider, das man bereits vor dem Informationseintrag der Daten besitzt, letztere gibt an, wie sich die Daten verteilen, wenn ein bestimmter Zustand auftritt.

3.2 Anwendung auf das Kredit-Scoring

Wenn wir den Satz von Bayes als Updating-Relation verstehen, der eine Wahrscheinlichkeitsverteilung liefert, sobald *a priori* Wissen mit Daten ergänzt wird, so ist eine Anwendung auf das Kredit-Scoring zwanglos zu erhalten. Sei x das Merkmal „Ausfall“ mit den Auspägungen „ja“ und „nein“, y der Satz der Kreditnehmer-Antragsdaten. Nach Bayes gilt demnach:

*P(Ausfall ïAntragsdaten) = P(Antragsdaten ï Ausfall) * P(Ausfall) / P(Antragsdaten) (Gl. 2)*

Die Interpretation, die sich aus diesem Satz anbietet, ist folgende: Einem gegebenen Kreditnehmer wird *a priori* unterstellt, dass er eine mittlere Ausfallwahrscheinlichkeit P(Ausfall) besitzt. Ausserdem weiß das Kreditinstitut aus der Historie, wie sich die Antragsdaten bei schlechten Kunden verhalten zu den Antragsdaten bei guten Kunden. Über die Antragsdaten, die er dem Kreditinstitut zur Verfügung stellt, trägt der Kunde also zusätzliche Information über seine Ausfallneigung ein. Diese Information, P(Antragsdaten | Ausfall) / P(Antragsdaten), wird genutzt, um seine *a priori* Ausfallwahrscheinlichkeit zu „adjustieren“, und *a posteriori* P(Ausfall |Antragsdaten) zu ermitteln. Der Satz von Bayes sagt also, dass die Ausfallwahrscheinlichkeit proportional ist zum einen zur *a priori* Ausfallwahrscheinlichkeit, zum anderen zum Verhältnis der im nachhinein schlecht gewordenen Antragsverteilung zur gesamten Antragsverteilung. Die Leistung des Bayesianischen Satzes besteht demnach darin, die Relation von Antrag zu Ausfall (die *Likelihood*) zu invertieren und konsistent mit dem *a priori* Wissen über Ausfall zu verknüpfen. Gleichung (2) gibt somit an, wie wahrscheinlich ein Ausfall ist, wenn bestimmte Antragsdaten vorliegen.
Die Zweistufigkeit des Verfahrens, nämlich vor Sichtung der Antragsdaten den allgemeinen Schätzer *a priori*-Verteilung zu haben, und diesen dann mittels *Likelihood* zu verfeinern, macht die Stärke des Verfahrens bei kleinen Datenmengen aus. Im Falle sehr stark gestreuter Antragsdaten üben dieselbigen über die Likelihood wenig Einfluss auf die Schätzung der Ausfallwahrscheinlichkeit aus. Im Extremfall nicht vorliegender Antragsdaten wird immer noch ein stabiler Schätzer gewonnen – die *a priori*-Verteilung.

3.3 *A priori*-Verteilung

Die Schätzung der *a priori*-Verteilung, P(Ausfall), ist eine Punktschätzung für die Zustände „Ausfall=ja“ und „Ausfall=nein“. Sie leitet sich von der langjährigen Erfahrung ab, mit wie vielen Ausfällen in einem bestimmten Kreditsegment im Mittel zu rechnen ist. Für das hier durchgeführte Beispiel beträgt die *a priori*-Wahrscheinlichkeit für „Ausfall=ja“ (die Ausfallwahrscheinlichkeit) 2%, die Wahrscheinlichkeit für „Ausfall=nein“ (die Überlebenswahrscheinlichkeit) 98%. Die Schätzwerte mögen sich von der buchhaltungstechnischen Größe Einzelwertberichtigung ableiten oder aus der Controlling-Größe Standardrisikokosten berechnet werden. Die Schätzung, wie viele Ausfälle im nächsten Jahr auftreten werden, ist aber auch durch Kreditsachbearbeitern leistbar – es ist keineswegs schlechter, auf deren Wissen zurückzugreifen und die mittlere Ausfallrate als *a priori*-Wahrscheinlichkeit zu benutzen. Vielfach sind in den Kreditinstituten auch Historien über

Ausfälle vorhanden, die ein gutes Bild darüber abgeben, mit wie vielen Ausfällen in Zukunft zu rechnen ist.
Definiert man die *a priori*-Wahrscheinlichkeit als Parameter, so lässt sich später auch die Sensitivität gegenüber dieser Annahme prüfen oder die Annahme ganz eliminieren, indem man für einen gegebenen Kreditantrag angibt, wie weit die Wahrscheinlichkeit für „Ausfall=ja" von der Wahrscheinlichkeit für „Ausfall=nein" abweicht (sog. odds-ratio).

3.4 *Likelihood*-Verteilung

Die *Likelihood*-Verteilung P(Antragsdaten | Ausfall=ja) gibt an, wie wahrscheinlich die gegebenen Antragsdaten sind, wenn ein Ausfall vorliegt. Im Gegensatz dazu gibt P(Antragsdaten | Ausfall= nein) an, wie wahrscheinlich die gegebenen Antragsdaten sind, wenn kein Ausfall vorliegt. Beide Verteilungen lassen sich aus den historischen Kreditdaten ableiten oder gemäß einer Expertenschätzung eintragen. Die *Likelihood* hält also quantitativ fest, was ein Kreditsachbearbeiter *ex post* feststellen würde, wenn er heute einen ursprünglichen Antrag sieht, von dem heute klar ist, dass er einem Ausfall zugehört. Wenn z.Bsp. ein Kreditsachbearbeiter einen Ausfall mit den Worten „ja klar, schon wieder jemand unter 30 Jahren" quittiert, ist das in einer erhöhten Wahrscheinlichkeit für das Merkmal Alter in der Ausprägung <30 wiederzufinden, gemessen daran, wie häufig die Ausprägung <30 überhaupt auftritt.
Technisch ist das Erstellen der Wahrscheinlichkeitsverteilung für Antragsdaten auf Basis bereits realisierter Ausfälle sehr einfach. Aus einer Historie von genügend vielen Ausfällen lässt sich ersehen, wie sich die Verteilung über die Antragsdaten realisiert. Mit P(Antragsdaten | Ausfall=ja) wird also abgeleitet, wie die Verteilung über die Merkmalsausprägungen ist, wenn es sich um einen Ausfall handelt. Demgegenüber zeigt P(Antragsdaten | Ausfall=nein), wie die Verteilung der Antragsdaten aussieht, wenn es sich um einen gesunden Kreditnehmer handelt. Über die Relation der totalen Wahrscheinlichkeit,

$$P(\text{Antragsdaten}) = P(\text{Antragsdaten} \mid \text{Ausfall=ja}) * P(\text{Ausfall=ja}) + P(\text{Antragsdaten} \mid \text{Ausfall=nein}) * P(\text{Ausfall=nein}) \qquad (Gl.\ 3)$$

erstellt sich aus obigen beiden *Likelihoods* und den beiden *a priori*-Verteilungen zwanglos die multivariate Verteilung der Antragsdaten P(Antragsdaten). Diese gibt die Wahrscheinlichkeitsverteilung der Antragsdaten aller Kunden, guter wie schlechter, an.
Wenn, wie in den meisten Scorekarten der Fall, die beschreibenden Merkmale nominal skaliert sind (diskrete Wahrscheinlichkeitsverteilung), so ist jeder Kombination eines Antrages eine Wahrscheinlichkeit zuzuordnen. Im Falle eines metrisch skalierten Merkmales müsste die Verteilung durch eine kontinuierliche Wahrscheinlichkeitsverteilung, und damit parametrisch, beschrieben werden. In dieser Version wären die Scorepunkte nicht auf Ausprägungen bezogen, sondern würden sich auf ein ganzes Merkmal anwenden lassen.
Die Herausforderung besteht darin, die Multivariabilität darzustellen. Nachdem sich Antragsdaten typischerweise aus mehreren Merkmalen zusammensetzen, stellt sich bei ungenügend vielen Daten die Aufgabe, die Abhängigkeit der Einzelmerkmale untereinander abzubilden. Im Falle unabhängiger Merkmale würde sich für die Matrix P(Antragsdaten(M_1, M_2, ..., M_m)) = $P(M_1)*P(M_2)*....*P(M_m)$ ergeben. Die Annahme unabhängiger Einzelmerkmale ist, wie auch aus Untersuchungen in der Praxis bekannt, nicht sehr realistisch – die Korrelationen der Merkmale untereinander sind teils beträchtlich. Um Abhängigkeiten zwischen den Merkmalen zu quantifizieren, obwohl die Matrix P(Antragsdaten| Ausfall=ja) meist spärlich besetzt ist, müsste eine vollständige multivariate Verteilungsannahme getroffen werden. Dies mag nicht trivial erscheinen, es sei an dieser Stelle allerdings darauf hingewiesen, dass bereits die (relativ einfach anmutenden) Expertenkarten gleich schwierig (und damit: ebenso wenig) echt multivariat erstellt werden müssten.

3.5 Einbringung qualitativer Faktoren

Ein weiterer Vorteil des Bayesianischen Ansatzes ist die Art und Weise, wie qualitative Informationen über Kreditnehmer in die Schätzung der Ausfallwahrscheinlichkeit eingetragen werden können. Nachdem der Satz von Bayes nicht davon abhängt, wie die zustandsbeschreibenden Parameter der Wahrscheinlichkeitsverteilungen skaliert sind, ist es ohne weiteres möglich, auch „weiche“ Faktoren zu modellieren. Konkret heisst das, man kann die auf den „harten“ Antragsdaten errechnete *a posteriori*-Schätzung P(Ausfall |Antragsdaten) erweitern um Wissen, das sich in Verteilungen der Art P(Ausfall |Qualitatives Urteil) abbildet. Eine Zusammenführung der weichen, qualitativen Faktoren aus der Einschätzung des Kreditsachbearbeiters und der harten, quantitativen Faktoren aus der Statistik der Antragsdaten zu einem Gesamturteil kann, weil Unabhängigkeit der Antragsdaten-Verteilung von der Verteilung des qualtitativen Urteils gegeben ist, gemäß der Formel P(Ausfall |Antragsdaten, Qualitatives Urteil) = P(Ausfall |Antragsdaten)*P(Ausfall |Qualitatives Urteil) erfolgen.

3.6 Dynamische Schätzungen

Aus unterschiedlichen Gründen (Eigenkapital-Berechnungen, Portfolio-Analysen, Stress-Tests, etc.) ist es wichtig, Scorings des gesamten Bestandes regelmäßig neu zu errechnen. Mit den klassischen Methoden wird hierbei jeder Kreditnehmer durch die Scoring-Gleichung mit seiner Ausfallwahrscheinlichkeit belegt. Mit dem Satz von Bayes ist es nun allerdings vorstellbar, einen Prozess zu installieren, der diese Verteilungen im Sinne einer Updating-Relation dynamisch schätzen kann. Dazu wird eine Bayes-Gleichung für die Verteilung der Ausfallwahrscheinlichkeit des gesamten Portfolios angeschrieben, die sich durch aktuelles Updaten der *a priori* Verteilung ergibt. Als *a priori* kann man dabei die letztmalig erhaltene *a posteriori*-Verteilung einstellen, wodurch sich ein dynamisches Re-Scoring installieren lässt (siehe dazu Rieder, 2003). Ein allfälliges Bestands-Scoring ist dadurch deutlich einfacher zu erhalten als ein mit den klassischen Verfahren geschätztes.

4 Vergleich der Methodenperformance

Zur Bewertung des hier vorgeschlagenen Verfahrens wird die Bayesianische Methode einer klassischen logistischen Regression gegenübergestellt, weil sich die Ergebnisse dieser auch auf der Skala Ausfallwahrscheinlichkeit messen lassen und die Regression daher dasjenige der oben vorgestellten klassischen Verfahren ist, für das eine direkte Vergleichbarkeit möglich ist.

Die Güte der beiden Ansätze seien entlang von quantitativen und qualitativen Dimensionen gemessen. Zum einen sind das Maße der Trennschärfe, zum anderen Beurteilungen der Interpretierbarkeit und praktischen Handhabbarkeit des jeweiligen Modells.

4.1 Beschreibung der Entwicklungsstichprobe

Als reales Beispiel wurden Kreditnehmer-Antragsdaten aus dem Segment der Gewerbekunden und dem Produkt Dispositionskredit gewählt. Die Entwicklungsstichprobe setzt sich aus 28.000 guten und 940 schlechten Schuldnern zusammen. Die beim Antrag bekannten Merkmale waren der Schufa-Auskunftei-Index, die Dauer der Kundenbeziehung, die Rechtsform des Unternehmens, die Branche, das Alter des Kreditnehmers und die Postleitregion des Gewerbestandortes.

Die vorhergesagte Ausfallwahrscheinlichkeit wurde in Klassen gemäß der Standard&Poor's Skala geschnitten. Hiernach bietet sich das folgende Bild der Verteilung Guter und Schlechter über die Ratingklassen:

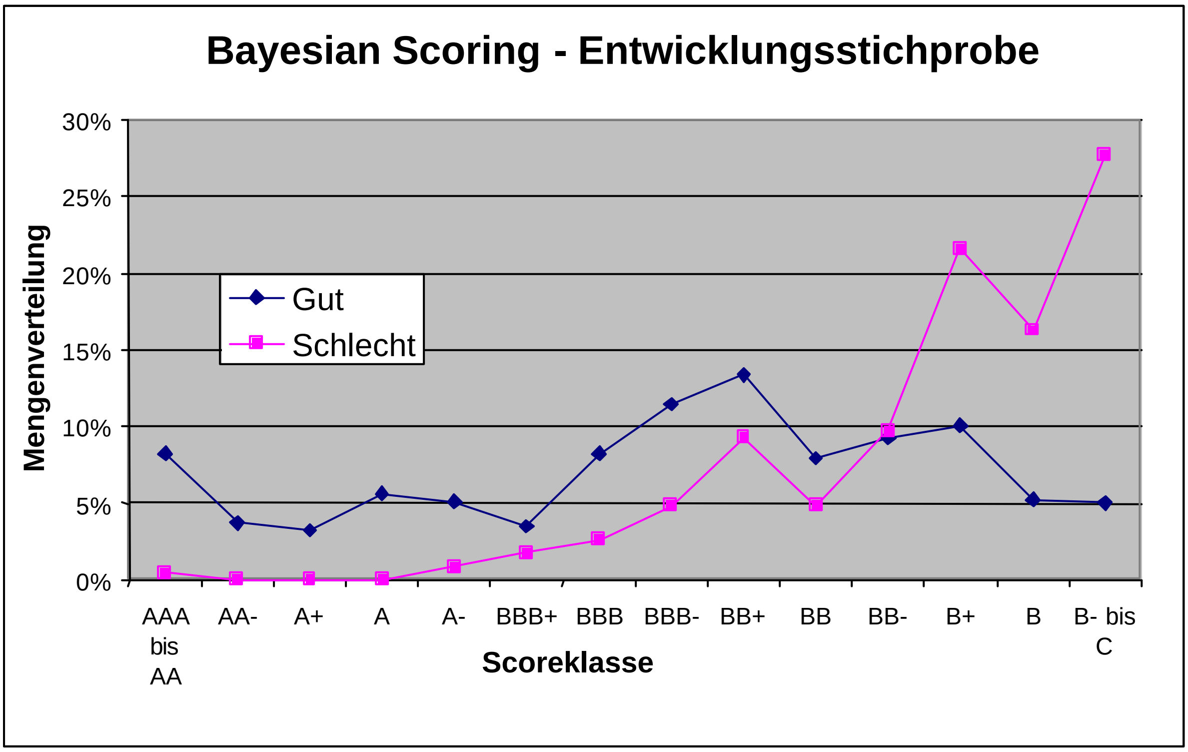

Abbildung 1: Scoreverteilung der Entwicklungsstichprobe (nach Bayesianischem Scoring).

An Abbildung 1 lässt sich eine bei Kreditausfällen häufig auszumachende Beobachtung ablesen. Die Verteilung der guten Fälle ist deutlich flacher als die Verteilung der schlechten Fälle. Mit anderen Worten, die Guten sind deutlich weniger differenzierbar als die Schlechten. Die Schlechten erweisen sich demgegenüber als viel einfacher zu separieren als die Guten. Beim Optimieren der Trennschärfe muss daher beachtet werden, dass der Alpha-Fehler mit steigenden Ratingklassen viel stärker ansteigt als der Beta-Fehler absinkt, welchselbiger über die Scoreklassen hinweg relativ gleichmäßig abnimmt.

4.2 *Out of sample*-Performance

Für die Feststellung der Güte der Scorekarte untersuchen wir Alpha- und Beta-Fehler hinsichtlich des idealen Cutoffs, sowie die Gütemaße Trefferquote und Gini-Koeffizient (Tabelle 1). Diese Maße wurden an einer Validierungsstichprobe (sog. *out of the sample* Probe), also für Kreditnehmer, die nicht Bestandteil der Entwicklungsstichprobe waren, auf einem Portfolio von 9500 Guten und 320 Schlechten errechnet.

Gütemaß	**Logistisch**	**Bayesianisch**
Alpha-Fehler (als Gut klassifizierte Schlechte / Schlechte)	32,3%	24,7%
Beta-Fehler (als Schlecht klassifizierte Gute / Gute)	20,7%	29,5%
Trefferquote (richtig klassifizierte / Alle)	73,2%	72,9%
Gini-Koeffizient	61,3%	60,0%

Tabelle 1: Gütemaße im Vergleich.

Die Tabelle zeigt, dass die beiden Verfahren sehr ähnlich performen, auch wenn die logistische Regression im summarischen (das heisst: vom Cutoff unabhängigen) Maß Gini-Koeffizient etwas besser abschneidet. Grund hierfür ist die Tatsache, dass wir für das Bayesianische Scoring im hier vorliegenden *proof-of-concept* keine Korrelation der Merkmale untereinander eingeführt haben. In Testläufen, wo das gemacht wurde, hat sich die Bayesianische Methodik nicht nur als ebenbürtig, sondern als performanter herausgestellt.
Die Performance nach der Trefferquote, welche sich aus Alpha- und Beta-Fehler zusammensetzt, ist vom Cutoff abhängig, und zeigt am idealen Cutoff (dem Cutoff mit der geringsten Summe aus beiden Fehlerarten) innerhalb der statistischen Konfidenzintervallen auf dem Signifikanzniveau von 95 % die gleiche Güte wie das klassische Verfahren.
Die oben angegebenen Performance-Maße für das Bayesianische Scoring sind unter der Maßgabe zu verstehen, dass die *Likelihood* ohne Abhängigkeit der Merkmale untereinander angesetzt wurde. Die errechneten Gütemaße sind daher Untergrenzen, die übertroffen werden, sobald die Korrelationen zwischen den Merkmalen berücksichtigt werden. Die Ergebnisse für die logistische Regression hingegen sind optimiert auf die Trennung guter und schlechter Kreditnehmer, unter Berücksichtigung der vollen Korrelationsmatrix der eingehenden Merkmale. Die Gütemaße der logistischen Regression sind daher die mit diesem Verfahren am höchsten zu erreichenden, was bei den direkten Vergleichen zum Bayesianischen Scoring im Auge zu behalten ist.

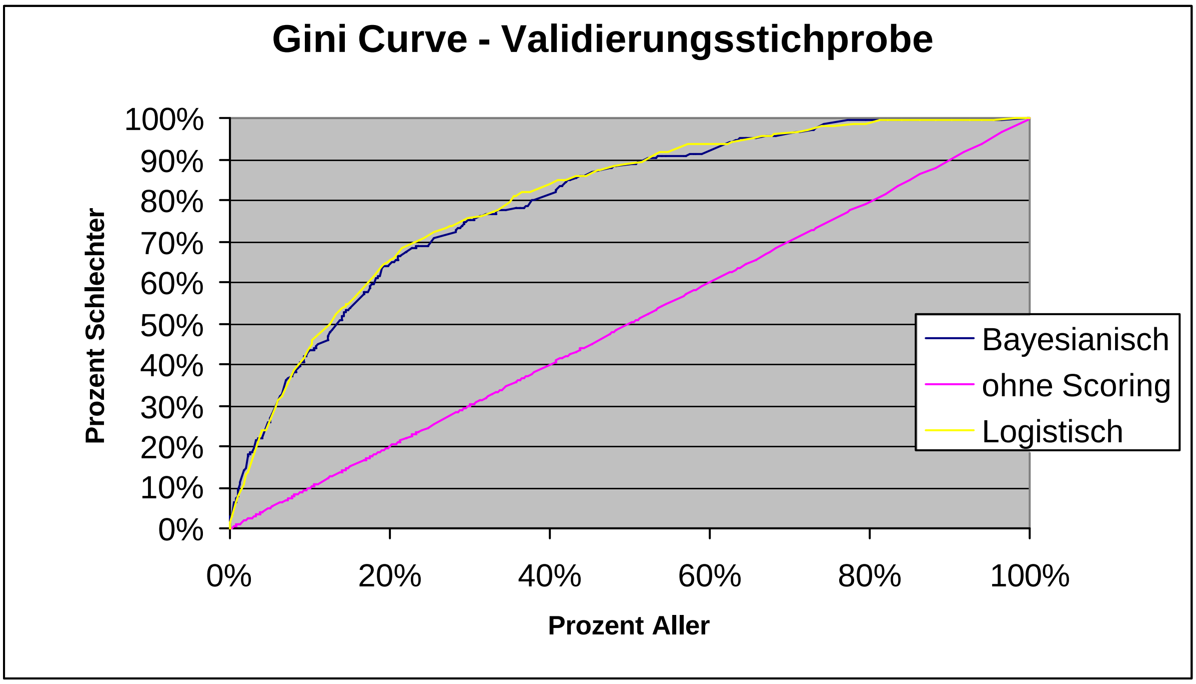

Abbildung 2: Gini-Curve der beiden Scorekarten “Logistisch” und “Bayesianisch”.

4.3 Verständlichkeit und Interpretation der Modellgewichte

Die Einführung von Kredit-Scoringsystemen scheitert oftmals daran, dass es Akzeptanzschwierigkeiten der angebotenen Lösungen gibt. So werden zum Teil die Kartengewichte nicht verstanden, und Diskussion um deren genaue Bedeutung und fachliches Nachjustieren nach den statistischen Optimierungsläufen sind in der Praxis an der Tagesordnung. Die Bayesianische Methodik liefert als einzige Methode genau das, was Kreditsachbearbeiter ansetzen würden. Sie weiß einerseits um die mittlere zu erwartende Ausfallrate bereits vor Sichtung der Antragsdaten (*a priori*-Wahrscheinlichkeit), und agiert

andererseits mit dem Verhältnis von bestimmten Ausprägungen bei guten und schlechten Kreditnehmern.
Die Ergebnisse der vorgestellten Verfahren können allesamt als Scorekarten dargestellt werden. Die Interpretierbarkeit der abgeleiteten Gewichte jedoch ist jedes mal eine gänzlich verschiedene. Für alle Modelle gilt, dass es zur Risikodifferenzierung ausreicht, auf die relativen Verhältnisse der Gewichte zu achten. Die absoluten Zahlen führen jedoch in allen drei Methoden zu anderen Interpretationen.

- Expertenkarten geben ihren Gewichten meist gar keine Interpretation mit. Für die Expertenkarten genügt es, auf Differenzen und/oder Relationen zwischen verschiedenen Ausprägungen zu achten. Der summierte Gesamtscore dient lediglich dazu, Kreditnehmer in Klassen zu teilen, hat aber meist per se keine eigene Bedeutung.
- Klassifikationsbäume geben die Schnittpunkte der Merkmale an und teilen jeder der durch aufeinander folgende Schnitte erhaltenen Gruppen eine Ausfallwahrscheinlichkeit zu. Die logistische Regression und das Bayesianische Scoring geben ebenso Ausfallwahrscheinlichkeiten an. Die Gewichte der Scorekarten nach logistischer Regression sind zunächst nur indirekt über eine *Maximum-Likelihood* Methode zugänglich und haben dann per se auch nur eine kaum zu kommunizierende, direkte Aussagekraft. Die Interpretationsschwierigkeiten bei den Baumverfahren und der logistischen Regression sind zum einen bedingt durch die Komplexität der statistischen Modelle, zum anderen durch die Unstetigkeit (Bäume) bzw. Nichtlinearität (Regression) der angenommenen Funktionen.
- Die Gewichte der nach Bayesianischem Scoring errechneten Scorekarte hingegen sind sehr einfach zu erhalten und zudem intuitiv eingänglich. Wenn ein Kreditsachbearbeiter z. Bsp. weiß, dass bei guten Kreditnehmern eine bestimmte Altersgruppe doppelt so häufig vorkommt wie eine andere, bei schlechten sich dieses Verhältnis allerdings umkehrt, so ist dieses Wissen direkt in einem Gewicht abgebildet. Das Gewichtsverhältnis sagt somit aus, um wie viel wahrscheinlicher eine bestimmte Ausprägung bei Guten ist als bei Schlechten.

Neben der Herleitung ist auch die Anwendung der Bayesianischen Methodik sehr einfach – die Formel für die Ausfallwahrscheinlichkeit nach Bayes setzt sich aus Divisionen und Additionen zusammen. Nicht nur die technische Implementierbarkeit, auch Akzeptanz und Nachvollziehbarkeit bei den Kreditsachbearbeitern und Risiko-Managern ist der entscheidende Vorteil dieses Verfahrens.

5 Zusammenfassung und Schlussfolgerungen

Die traditionellen, der klassischen Statistik verpflichteten Methoden für das Kreditantrag-Scoring werden von Kreditsachbearbeitern regelmäßig in Frage gestellt, weil sie wenig intuitiv, schwer kommunizierbar und nur mäßig nachvollziehbar sind. Ohne eine fundierte Statistik-Ausbildung lassen sich die Gewichte aus der logistischen Regression oder die Schnitte aus Klassifikationsbäumen kaum verstehen und stoßen daher in der Praxis auf Unmut und Ablehnung.
Die hier vorgeschlagene Methodik entspricht exakt dem Verfahren, das zu einer Scorekarte führen würde, die die Expertenmeinung der Kreditsachbearbeiter widerspiegelt. Dass diese Methodik direkt aus der Bayesianischen Statistik abgeleitet werden kann und somit statistisch fundiert ist, ist ein Argument für deren Weiterenwicklung und Anwendung. Neben der einfachen Interpretation ist es auch die praktikable Handhabbarkeit bei der Berechnung von Ausfallwahrscheinlichkeiten, die simple Methodik zum Einbringen von qualitativen

Informationen, die Möglichkeit des einfachen Updatens des Bestands-Scorings und die hohe Trefferquote, die einem Bayesianischen Modell den Vorzug geben lassen.

Referenzen

Berry, D.A., *"Statistics - A Bayesian Perspective"*, Duxbury Press, 1996.

Gelman, A., J.B. Carlin, H.S. Stern, and D.B. Rubin, *"Bayesian Data Analysis"*, Chapman & Hall, 1995.

Rieder, M.J., *"A Kalman filter approach to dynamic portfolio scoring"*, in Arbeit, 2003.