

Andreas Oehler

# Grundsätze ordnungsgemäßer Bewertung durch Scoring

**Die digitale Durchleuchtung und Bewertung des Verhaltens von Konsumenten und Bürgern nach einem Punktesystem (Scoring) ist in repressiven Staaten wie China schon weit gediehen. Aber auch in Deutschland werden mit diesem System Bürger als Verbraucher in der Finanzwirtschaft oder als Patienten im Bereich Gesundheitsversorgung und Versicherungen bewertet. Die Problematik wird in der Öffentlichkeit wenig diskutiert. Dies überrascht insofern, als nahezu jede und jeder betroffen ist und zukünftig noch stärker als bisher im Rahmen der voranschreitenden Digitalisierung eingebunden sein dürfte.**

Denkt man an Scoring in Deutschland, so fallen vielen Bürgern die Auskunfteien zur Bonitätseinschätzung ein, weil viele damit bereits bei der Kontoeröffnung, der Kreditkarte oder bei Verbraucherkrediten, aber auch bei einem Telekom-/Internet-Vertrag und beim Online-Shopping zu tun hatten. Bekannt sind Schufa oder Creditreform, aber auch arvato infoscore von Bertelsmann. Scoring ist dann nichts anderes als eine Einschätzung des zukünftigen Verhaltens einer Person abzugeben – auf der Basis der bei diesen speziellen Informationsverarbeitern zu einer bestimmten Person vorliegenden Daten im Abgleich mit dem Datenbestand insgesamt oder Teilen daraus. Im Falle eines beantragten Kredits oder eines Handyvertrags interessiert also die Einschätzung, inwieweit mit einer pünktlichen Rückzahlung zu rechnen ist (Kreditwürdigkeit). Davor kann gegebenenfalls auch eine Einschätzung der reinen Kreditfähigkeit liegen, also inwieweit eine Person überhaupt – egal zu welchen Konditionen – einen Kredit erhalten soll.

## Was ist Scoring?

Allgemeiner geht es beim Scoring darum, für jede (angefragte) Person eine Verhaltenseinschätzung abzugeben, inwieweit die vertraglichen Verpflichtungen erfüllt werden. Eine solche Voraussage des Verhaltens betrifft aber in unserer digitalisierten Welt nicht mehr nur Kredite und Konten.

Längst gilt dies für Warenlieferungen ebenso wie für Internet- und Telekom-Verträge. Auch andere Dienstleister wie Versicherungen werten selbst aus oder lassen dies extern tun, um aus einem Datenbestand auf das Verhalten der potenziellen Vertragspartner zu schließen. Hier können Gesundheitsdaten aus Versicherungsanträgen grundsätzlich ebenso einfließen wie aus Wearables, wie z.B. Smart Watches oder Fitnessarmbändern, generierte Daten. Auch zur Erfassung des Fahrverhaltens in Verbindung mit Kfz-Versicherungsverträgen gibt es bereits Scoring. Dieses Vorgehen ist im ökonomischen Sinne keineswegs überraschend, versucht doch ein Vertragspartner aufgrund asymmetrisch geringerer Information mit Screening eine Verringerung der Informations-, Gestaltungs- und Betroffenheitsasymmetrien sowie der daraus resultierenden Risiken zu erreichen. Umgekehrt geben potenzielle Kunden Daten im Wege des Signalling ab, um bestimmte Verträge und Konditionen zu erhalten, z.B. können potenzielle Mieter ihre Bonitätsauskunft zur Wohnungsbesichtigung gleich mitbringen.

Die Thematik der vorvertraglichen Reduzierung von Asymmetrien durch Scoring ist also nicht wirklich neu. Sehr beachtenswert allerdings ist die im Zuge der Digitalisierung exponentiell wachsende Datenverfügbarkeit samt selbstlernender Auswertungstechnologien. Aufgrund der Durchdringung aller wesentlichen Lebensbereiche ist die Abhängigkeit der Nachfrager von der „Vermessung“ ihrer Person und ihres Verhaltens bis hin zu einer Anpassung von Verhaltensweisen an bekannte oder vermutete Scoring-Systeme exorbitant gestiegen. Dies ist deutlich weitreichender als der im Volksmund kolportierte Umzug zwecks Adressenänderung für eine bessere Bonitätseinschätzung.

Gemäß Bundesdatenschutzgesetz (BDSG) ist ein Scoring-Ergebnis definiert als „ein Wahrscheinlichkeitswert für ein bestimmtes zukünftiges Verhalten des Betroffenen [...] zum Zweck der Entscheidung über die Begründung, Durch-

**Prof. Dr. Andreas Oehler** ist Lehrstuhlinhaber für Betriebswirtschaftslehre, insbesondere Finanzwirtschaft, sowie Direktor der Forschungsstelle Verbraucherfinanzen & Verbraucherbildung an der Universität Bamberg.

führung oder Beendigung eines Vertragsverhältnisses“<sup>1</sup>. Bereits an dieser Formulierung, z.B. weil die Nutzung von Anschriften nicht generell verboten ist, noch mehr aber hinsichtlich der ergänzenden Regulierung durch die Datenschutzgrundverordnung (DSGVO), die ab dem 25. Mai 2018 gelten soll, gibt es heftige Kritik.<sup>2</sup>

Aus ökonomischer, insbesondere aus informationsökonomischer Perspektive stellt eine zielgenaue und effiziente Datenanalyse mit Scoring in einem Umfeld asymmetrisch verteilter Informationen und asymmetrischer vertraglicher Gestaltungsmöglichkeiten und Betroffenheiten nichts anderes dar als eine intern oder extern vergebene Dienstleistung zur Gewährleistung der Wertschöpfungskette. Mit einer zunehmenden Anonymisierung von (potenziellen) Vertragsbeziehungen durch die fortschreitende Digitalisierung steigt der Analysebedarf rasant. Mit dem Scoring wirtschaftlich vergleichbar sind Tätigkeiten anderer Informationsintermediäre, z.B. (vergleichende) Tests, Vergleichsportale, Makler oder Ratingagenturen.<sup>3</sup>

### Warum sind Mindeststandards erforderlich?

Die ständig wachsende wirtschaftliche und auch gesellschaftliche Bedeutung reicht längst weit über den Anwendungsbereich des Risikomanagements mit Bonitätsanalyse und Kreditverträgen hinaus, insbesondere im Lebensbereich Gesundheit und Vorsorge, aber auch hinsichtlich des Zugangs zu Sachgütern und Dienstleistungen, z.B. Bezahlverfahren oder Kauf- und Buchungsmöglichkeiten. Entsprechend erscheint es notwendig, einen Mindeststandard für Scoring-Verfahren und ihre Pflege einzufordern. Auf einer Meta-Informationsebene liegt im Scoring für alle Beteiligten, auch für Nachfrager, die Chance, die Bewertung selbst zu nutzen oder Dritten zur Nutzung anzubieten, ohne aufwendige Recherchen selbst durchführen zu müssen, soweit dies überhaupt möglich ist. Damit diese genutzt werden kann, bedarf es einiger weniger Grundsätze oder

Mindestanforderungen für eine „good scoring“-Praxis, um eine valide Mindestqualität der Scores zu gewährleisten. Ein solcher Mindeststandard ließe sich als „Grundsätze ordnungsgemäßen Scorings“ (GoS) oder mit einer Reminiszenz an das anglo-amerikanische Mutterland des Scorings/Ratings als „Generally Accepted Scoring Principles“ (GASP) kennzeichnen. Erst auf diese Weise könnten sich Verbraucher wie auch Anbieter valide orientieren.

Aus wirtschaftswissenschaftlicher Sicht handelt es sich beim Scoring grundsätzlich um einen Teilschritt im Rahmen einer Risikoabschätzung mittels Rating. Den Ausgangspunkt stellt die zielgerechte Extraktion von Einzelkriterien dar, die – letztlich aus Vergangenheits- und Gegenwartsinformationen – als bedeutsame und trennscharfe Merkmale (Indikatoren) beurteilt werden. Diesen werden mit mathematisch-statistischen Verfahren, häufig auf der Basis einer logistischen Regression, wie sie in künstlichen neuronalen Netzen eingesetzt sind, oder einer multivariaten Diskriminanzanalyse, Gewichte beigemessen (Schätzung, keine Rechnung). Auf Grundlage der Einzelkriterien mit ihren Gewichten werden jeweils Punkte („Scores“) vergeben, die zu einem Gesamt-Score zusammengeführt werden, der geschätzten Verhaltensausprägung oder, (Bonitäts-)„Note“. Beim Rating wird in einem weiteren Schritt eine Zuordnung zu einer (Bonitäts-)Klasse (Klasse von Ausfallwahrscheinlichkeiten) vorgenommen, die meist durch Symbole aus Buchstaben, Ziffern und Sonderzeichen gekennzeichnet werden. Solche quantitativ ausgerichteten und präskriptiv eingesetzten Scoring-Systeme führen zu einer (gewollten) Standardisierung in der Anwendung mit wenig „persönlichen“ Spielräumen.

Ausgehend von dieser Definition und der Eingangsanalyse ist es nur folgerichtig, dass in einer digitalisierten Informationsgesellschaft Scoring in immer mehr Anwendungsfeldern anzutreffen ist und mittlerweile infrastrukturelle Bedeutung gewonnen hat. Die systematische und automatisierte Einschätzung zu Verhaltensausprägungen von (potenziellen) Vertragspartnern betrifft längst alle existenziellen Lebensbereiche, also zum Beispiel Gesundheit, Finanzen, Vorsorge, Mobilität, Wohnen und Energie. Die angewendeten modellhaften Annahmen der eingesetzten Verfahren und deren eigendynamische, sich selbst fortentwickelnde Algorithmen schaffen dabei auch eine parallele Verhaltenswelt. Das stellt eine Entwicklung dar, die bereits aus dem Einsatz von Risikomanagement-Tools in der Wirtschaftspraxis und Finanzaufsicht bekannt ist. Analog dazu sind nicht nur die heute häufiger erörterten politischen und gesellschaftlichen, sondern insbesondere auch die systemischen Risiken anzusprechen. Die den Modellstrukturen inhärenten typisierten und datengetrieben erzeugten sowie dynamisierten Verhaltensmuster verändern selbst nicht nur die angebotsseitigen Verhaltensoptionen (z.B. Zulassung

1 Bundesdatenschutzgesetz (BDSG) §28b.

2 Vgl. z.B. Vorstand der Verbraucherzentrale Bundesverband, Klaus Müller beklagt sie als „massiven rechtlichen Rückschritt“, in K. Ludwig: Ärger um Schufa-Einträge, in: Süddeutsche Zeitung vom 9.5.2016, S. 19; oder A. Roßnagel: Studie: EU-Datenschutz-Grundverordnung verfehlt alle Ziele – Kasseler Juristen entwirren Rechtslage, Universität Kassel, Pressemitteilung vom 27.9.2016, <https://www.uni-kassel.de/uni/nc/universitaet/nachrichten/article/studie-eu-datenschutz-grundverordnung-verfehlt-alle-ziele-kasseler-juristen-entwirren-rechtslag.html> (21.8.2017).

3 Vgl. A. Oehler: Chancen der selbstbestimmten Datennutzung?!, in: Wirtschaftsdienst, 96. Jg. (2016) H. 11, S. 830-832, <https://archiv.wirtschaftsdienst.eu/jahr/2016/11/chancen-der-selbstbestimmten-datennutzung/> (18.8.2017); A. Oehler: Testen der Tester?, in: Wirtschaftsdienst, 94. Jg. (2014) H. 6, S. 444-447, <https://archiv.wirtschaftsdienst.eu/jahr/2014/6/testen-der-tester-grundsaezte-ordnungsgemaessen-testens/> (18.8.2017); vgl. auch A. Oehler, S. Wendt: Good Consumer Information: the Information Paradigm at its (Dead) End?, in: Journal of Consumer Policy, 40. Jg. (2017), S. 179-191.

zu medizinischen Operationen nach Scores aus Fitness-Apps oder Wearables, Zugang zu Krediten, Zahlung einer Reha-Maßnahme, Nutzung von Bezahlverfahren, Zugang zu Versicherungskonditionen nach Scores aus dem Fahrverhalten). Auch Nachfrager richten ihr Verhalten an den „geforderten“ Profilen und Mustern aus, um Leistungen überhaupt oder zu günstigen Konditionen erhalten zu können.

Die Eingriffstiefe und -breite moderner Scoring-Verfahren in alle wesentlichen Lebensbereiche rechtfertigt die Forderung nach einem Mindeststandard, der regelmäßig aufsichtlich mit Beweislastumkehr zu kontrollieren ist, analog zur Finanzaufsicht oder der Lebensmittelkontrolle. Gleichzeitig ist das Geschäftsmodell des Scorings selbst zu schützen, damit in den real unvollkommenen Märkten die genannten Vorteile ökonomischer Effizienz für Anbieter und Nachfrager gewahrt bleiben. In diesem Sinne sind die nachstehenden Grundsätze ordnungsgemäßen Scorings (GoS) zu verstehen. Die Mindestanforderungen sind geeignet, vorhandene Systeme zu überprüfen, auch durch eine staatliche Aufsicht, aber ebenso, um bei der Entwicklung neuer Strukturen zu helfen.<sup>4</sup>

### Grundsatz 1: Identifizierung und Transparenz

Grundsatzfragen: Ist klar erkennbar, wer das Scoring-Urteil vergibt? Ist das dahinter stehende Geschäftsmodell oder Unternehmen klar zu erkennen? Wer finanziert das Scoring, in welcher wirtschaftlichen, insbesondere finanziellen Beziehung stehen Scorer und Anbieter oder Score-Nutzer?

Theoretisch kann man inzwischen etwas leichter Auskunft über seine Daten und Scores erhalten (Auskunftsrecht als Holschuld). Wer aber weiß denn genau, wo überall eigene Verhaltenseinschätzungen vorliegen, wer sie wann und wie verwertet und verkauft? Ganz abgesehen davon, dass viele kaum wissen, mit welchen Folgen ihre Daten verwertet und verbreitet werden. In der Realität herrscht eher die sehr einseitige Praxis zulasten vieler vor, einerseits Daten abgeben zu müssen, um nicht wirtschaftliche Nachteile zu erleiden (z.B. Verweigerung von Verträgen), und andererseits faktisch nicht darüber informiert zu werden, mit welchen Folgewirkungen welche Daten in welcher Qualität verwertet und verkauft werden. Die Umsetzung des gesetzlich vorgeschriebenen Auskunftsrechts durch Auskunftfeien beur-

teilt die Stiftung Warentest jedenfalls kritisch.<sup>5</sup> Folgerichtig wäre, dass Bürger nicht erst mühsam klären müssen, ob inkorrekte oder falsche Daten vorliegen oder Daten fehlen. Vielmehr müssen die Scorer dazu verpflichtet sein, selbst aktiv im Dialog mit Bürgern dafür zu sorgen, dass alle Daten stimmen und sie vollständig über die Nutzung und Weitergabe aufgeklärt worden sind (Bringschuld der Scorer und Datennutzer mit Beweislastumkehr). Zusätzlich haben die Scorer und Datennutzer für aktuelle und bereinigte Daten zu sorgen, indem sie z.B. geeignete Verträge zwischen Agentur/Auskunftei und weiteren Datennutzern/-verwertern schließen. Aufgrund der hohen Bedeutung des Scorings für alle Bürger ist die bestmögliche Datenpflege ein sehr hohes Gut.

Notwendige Bedingungen guter Bürgerinformation zum Scoring sind Transparenz (jederzeitige Verfügbarkeit aller relevanten Informationen), Verständlichkeit (insbesondere müssen die relevanten Informationen zu verstehen sein, ohne sich hierfür zunächst Fachwissen aneignen zu müssen) und Vergleichbarkeit (Informationen zu Scores sollten nach Form und Inhalt vergleichbar zur Verfügung gestellt werden, um mögliche Auskünfte alternativer Anbieter gegenüberstellen zu können). Hinreichende Bedingungen guter Bürgerinformation zum Scoring sind Klarheit (die Darstellung der Information darf keinen Raum für alternative Interpretationsmöglichkeiten lassen), Geeignetheit für die Person und die persönliche Situation und Verifizierbarkeit.

### Grundsatz 2: Verifizierung

Grundsatzfragen: Sind die Kriterien und Prüfmethode öffentlich zugänglich und durch Dritte nachprüfbar? Wird offengelegt, wie geprüft wird und welche Standards für welche Untersuchung zugrundegelegt werden?

Zumindest gegenüber den Gremien mit Aufsichtsfunktion und weiteren Sachverständigen müssen die wesentlichen Merkmale eines Scoring-Verfahrens samt der zugehörigen Vorüberlegungen und Auswahlprozesse frühzeitig offengelegt werden, nicht erst nach Durchführung und Weitergabe. Aufgrund der erheblichen Bedeutung von Scores für die Entscheidungen von Bürgern und Anbietern sollte auf diesen Grundsatz der Verifizierung hohes Gewicht gelegt werden, um rechtzeitig fehlleitende Informationen aus Datenanalysen zu vermeiden. Dadurch kann auch das Bewusstsein für die Bedeutung von Scoring-Informationen auf Märkten geschärft werden. Möglich erscheint eine frühzeitige Klärung, inwieweit Scoring-Urteile aus Indizien und Einschätzungen oder entlang wissenschaftlicher Standards abgeleitet und nachweisbar sind. Hierzu müssen Einzelkriterien, die in einem Scoring verwendet werden sollen (nicht unbedingt der

4 Vgl. A. Oehler: Labeling & Co – Wegweiser, Irreführung oder Wettbewerbsanreiz?, in: EMF Spectrum 3, 2013, S. 31-34. Bezogen auf die Verwendung von Vertrauenslabels hat die Verbraucherkommission Baden-Württemberg bereits 2011 einen Forderungskatalog erstellt, der im Folgenden ebenfalls berücksichtigt wird. Vgl. Verbraucherkommission Baden-Württemberg: Vom Labelmissbrauch zu Vertrauenslabels, Stellungnahme, Stuttgart 2011.

5 Stiftung Warentest: Durchblick schwierig, in: Finanztest, 2/2016, S. 12-17.

Algorithmus oder die Gewichtungsfaktoren), und eine Kontrolle der Analyse- und Schätzqualität offengelegt und eine regelmäßige Qualitätsprüfung der Einzelkriterien (Gewichtveränderungen, Änderung der Bedeutung) nachgewiesen werden, ebenso eine regelmäßige Qualitätsprüfung der Scores selbst (einzelne Klassen oder Summe von Kriterien). Die Verifizierung hat sowohl die interne Validität der Schätzungen (Zuverlässigkeit, Reliabilität) als auch die externe Validität nachzuweisen (Gültigkeit, Generalisierbarkeit).

### Grundsatz 3: Relevanz und Nützlichkeit

Grundsatzfragen: Sind die Aussagen zur Verhaltenseinschätzung („Note“) maßgeblich für die wesentlichen Eigenschaften einer Person und Situation? Wird der Fokus nur auf einzelne Teilaspekte gelegt, z.B. Adressdaten, oder werden alle wesentlichen Eigenschaften gleichermaßen berücksichtigt? Gehen die Scoring-Kriterien über staatlich festgelegte Standards hinaus oder überprüfen sie im Wesentlichen nur deren Einhaltung (z.B. Adressdaten mit einem weiteren Kriterium grundsätzlich zulässig)?

Die genannten Kriterien beziehen sich zunächst darauf, sicherzustellen, dass tatsächlich auch das Angekündigte eingeschätzt wird: Ist im Score auch „drin“, was „drauf“ steht? So relevant für manche vielleicht die Wohnadresse sein mag, Kern ist zunächst die Verhaltenseinschätzung zur Einhaltung von Verträgen und gegebenenfalls Vertragsstörungen. Unklare Scorings behindern die Funktionalität der Informationen.

Die Idee und Zielsetzung einer Basisregulierung sollte sich aufgrund der Vor- und Nachteile regulierender Eingriffe auf die genannten Grundsätze (GoS/GASP) im Sinne eines Mindeststandards beschränken. Damit würden mögliche Nachteile einer weitergehenden Regulierung wie die Beschränkung des Scorings nach Anwendungsbereichen, Produkten oder Dienstleistungen samt damit verbundene bürokratischen Aufwand vermieden oder vermindert. Gleichzeitig blieben die Vorteile der Mindestregulierung erhalten. Hierzu gehören unter anderem die Gewährleistung der Orientierung für Bürger sowie der Anbieter, die Sicherstellung des Markteintritts neuer Wettbewerber/Scorer, die

Möglichkeit für alternative Scoring-Verfahren und -Anwendungen und eine grundsätzlich von Angebot und Nachfrage abhängige Auswahl der einzuschätzenden wirtschaftlichen Verhaltensweisen.

Die Einhaltung des rechtlich verbindlichen Mindeststandards der Grundsätze ordnungsgemäßen Scorings (GoS) benötigt kein aufwendiges Aufsichtswesen, insbesondere kann auf bestehende Regelungen in verwandten Themenfeldern rekuriert werden. Inhaltlich muss über die gesamte Scoring-Prozesskette ein „gelebtes“ Risikomanagement- und Qualitätssicherungssystem nachgewiesen werden. Die regelmäßige Überprüfung kann z.B. im Rahmen der Jahresabschlussprüfung durch die bestellten Wirtschaftsprüfer mit entsprechendem Prüfauftrag vorgenommen werden, so wie dies bereits für andere Themenbereiche üblich ist.<sup>6</sup> Zu diesem „Lebendigkeits“-Nachweis für das implementierte interne Risikomanagement mit Kontroll- und Fehlermanagementsystemen gehört auch die Zusammensetzung und Arbeit der eingesetzten Gremien und die notwendige Unabhängigkeit der beauftragten Personen, z.B. in Fachgremien, die das Scoring-Verfahren erörtern und begleiten, oder in Gremien, die für die Initiierung von Scoring-Vorhaben zuständig sind. Einbezogen werden sollten auch die installierten oder zu installierenden Aufsichtsorgane der Geschäftsführung der Scoring-Agenturen.

Das Zusammenwirken der Beteiligten birgt die Chance, das Vertrauen in Scoring-Urteile zu rechtfertigen und einem weiteren Vertrauensverlust durch wirkungslose Kontrollpyramiden entgegenzuwirken (Aufseher der Aufseher der Aufseher ...). Die vorgeschlagenen Grundsätze ordnungsgemäßen Scorings (GoS) sichern den wichtigen Informationscharakter guter Scorings und stärken die notwendige Reputation. Es liegt im Interesse der Bürger, der Anbieter und seriöser Scorer, dauerhaft sorgfältig recherchierte, überprüfbare, wirklich relevante und nachvollziehbare Scoring-Ergebnisse zu publizieren und nicht mit dem Fokus auf geschäftstüchtiger Erledigung mit ungenauen Informationen das wertvolle Gut verlässlicher Scoring-Informationen zu beschädigen.

<sup>6</sup> Vgl. A. Oehler: Testen der Tester?, a.a.O.

#### Title: *The Basics of Generally Accepted Scoring Principles*

**Abstract:** *Scoring and comparable methods that are used to rate people's future behaviour can provide useful information for suppliers and thereby reduce the asymmetrical information deficits and the corresponding risks (screening). Likewise, consumers can profit from scoring if they use the scores to reveal their creditworthiness (signalling). In a market economy where markets are not fully transparent, scoring agencies act as information intermediaries. They provide information on future behaviour at the metalevel so that not every supplier or consumer has to rate themselves. This is why the reliability of scores and scoring agencies is so fundamental. This article examines the principles or minimum standards for valid scoring and discusses how to introduce, regulate and test them.*

**JEL Classification:** D11, D18, K2