Big Data and Scoring in the Financial Sector

Stefanie Eschholz and Jonathan Djabbarpour

Abstract Scoring is an assessment procedure, especially for the purpose of credit assessment. Big data did not "create" that kind of procedure but influences the calculation of probability forecasts by opening up additional data sources and by providing enhanced possibilities of analyzing data. Scoring is negatively connoted. While being connected to risks, it opens up opportunities for companies as well as for the data subject. Since 2009, scoring is regulated by the German Federal Data Protection act, which entitles the data subject to get information free of charge once a year. Currently, a draft amendment concerning scoring is discussed in Parliament.

1 Introduction

The catchphrases *scoring* and *big data* are frequently used by the media. Often, it is not clear what these phrases are supposed to mean. They are not always used with the same meaning, and they are sometimes used undifferentiated.

Therefore, the question arises what scoring actually is. Scoring describes a procedure, which assesses a person to compare him or her with others.[1] Those assessment procedures originate from banking: before a credit is given, a bank customer's credit default risk is assessed (so-called credit scoring).[2] For this purpose, a scale is determined. Depending on the position on that scale, the bank customer is assessed either as a "good" and therefore creditworthy customer or as a "bad" one. A "good" customer will be offered a credit with good conditions by the bank while a "bad" customer is not offered any credit at all or only one with bad conditions, for example higher interests or additional collateral are requested.

---

[1]BGH, NJW 2014, 1235 (1235 et seq.).
[2]BT-Drucks. 16/10529, p 9; Jandt, K&R 2015, 6 (6).

S. Eschholz (✉) · J. Djabbarpour
Institute for Information, Telecommunication and Media Law (ITM),
University of Münster, Münster, Germany
e-mail: stefanie.eschholz@uni-muenster.de

Furthermore, the question arises what big data is all about. Data is called "big" if it is characterized by the "three Vs": Volume, Velocity, Variety.[3] Additional characteristics such as Veracity are included in some definitions.[4] Big data is about analyzing masses of data.[5] Significant for big data is the quick and easy calculation of probability forecasts and correlations, which enables new insights and the deduction of (behavioral) patterns.[6]

2 Scoring Procedure

Usually, businesses that score do not publish any details or only few details about factors influencing the score and their weighting. One reason for this is that they consider this information as a business secret. Another reason is that fully transparent procedures entail the risk of manipulation.[7]

Generally, the scoring factors are gathered systematically to calculate one or more scores out of them by means of statistical methods. For instance, the Schufa, Germany's most noted credit agency,[8] calculates a basic score as well as sector-specific scores and collection scores. While the basic score reflects the customer's general creditworthiness, the sector-specific scores are supplemented with specifics of each sector, for example of the telecommunications sector. Collection scores indicate the probability of successfully collected debts. Different factors in different weightings are included in the calculation of the score to meet the different requests as good as possible.[9]

A single factor itself does not necessarily have a positive or negative influence on the score, but the factor can have such influence in context with or in dependency with other factors. For instance, one regularly paid mobile phone contract can

---

[3]Laney 2001, 3D Data Management: Controlling Data Volume, Velocity and Variety, http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf.

[4]E.g. Markl, in: Hoeren, Big Data und Recht, p 4 and Fraunhofer-Society, http://www.iml.fraunhofer.de/de/themengebiete/software_engineering/big-data.html.

[5]Markl, in: Hoeren 2014, Big Data und Recht, p 4; Jandt, K&R 2015, p 6.

[6]Jandt, K&R 2015, p 6.

[7]SCHUFA Holding Ahttps://www.schufa.de/de/ueber-uns/daten-scoring/scoring/transparente-scoreverfahren/.

[8]Credit agencies are private-law companies, not government agencies. They collect and file commercially personal data about companies and persons concerning their creditworthiness. They receive such data from other companies (e.g., banks, telecommunication companies, mail order companies, energy suppliers and collection companies), publicly available registers (e.g., concerning insolvency) or other public sources (e.g., internet, newspaper). They pass information to business partners for value. Besides Schufa, there are also other credit agencies like Infoscore, Deltavista and Bürgel. Sources: Ehmann 2014, section 29 m. n. 83, 84; LDI NRW, 2012; https://www.schufa.de/de/.

[9]https://www.schufa.de/de/unternehmenskunden/leistungen/bonitaet/.

have a positive influence, whereas many mobile phone contracts can have a negative influence. Furthermore, non-existing or not known factors can have an influence as well. Under certain circumstances, a customer without any record can be considered less creditworthy than a customer who regularly exceeds his credit line, but always repays his or her debts.[10]

The values used for scoring do not necessarily reflect reality. For instance, other factors are the number of people in the household or how long the household already exists. For a credit agency, a household does not exist until it gets to know about its existence. The scores are calculated with this value even if the household exists much longer. It is quite the same when it comes to the number of people in the household because sometimes outdated or simply wrong data is used here.[11]

In the past, the Schufa score could deteriorate when a customer asked different banks for offers even if he or she did not accept any of them. In the meantime, the Schufa has introduced the factor "request for conditions" that does not have any influence on the actual score. One has to obtain and prove your Schufa credit record to assure that the requests were used correctly and that no wrong negative factors influenced the score.[12]

3   Scoring in the Big Data Era

The extent and scope of scoring were increased considerably in recent years by new technologies for gathering and analyzing data—"big" data.[13] Scoring procedures infuse more and more areas of life and therefore they are the basis for decisions leading to a contract and its conditions.[14]

Admittedly, scoring is no specific manifestation of big data. The Schufa started in the 1920s, computerized its database already in the 1970s and began to develop credit scores in the 1990s.

However, big data opens up additional data sources. For instance, the Schufa considered using social media data from networks like Facebook, Twitter and Xing in 2012.[15] For this purpose, the Hasso-Plattner-Institute (HPI) of the University of Potsdam should start research on how information from social media could be used for credit scoring.[16] Because of the public reaction, the research never took place.

---

[10]Mansmann, c't 10/2014, p 80 et seq.

[11]Schulzki-Haddouti, c't 21/2014, p 39.

[12]Mansmann, c't 10/2014, p 80.

[13]Jandt, K&R 2015, p 6; ULD & GP Forschergruppe 2014, Scoring nach der Datenschutz-Novelle 2009 und neue Entwicklungen, 16. p 55 et seq., p 125.

[14]Jandt, K&R 2015, p 6 et seq.; Schulzki-Haddouti, c't 21/2014, p 38.

[15]Rieger, Kredit auf Daten, FAZ.net, http://www.faz.net/aktuell/feuilleton/schufa-facebook-kredit-auf-daten-11779657.html; Schmucker, DVP 8/2013, p 321 et seq.

[16]Rieger 2012, Kredit auf Daten, FAZ.de, http://www.faz.net/aktuell/feuilleton/schufa-facebook-kredit-auf-daten-11779657.html.

First, the HPI refrained from conducting the project SCHUFALab@HPI and finally the Schufa abandoned its plans.[17] Now, the Schufa does not use social media data at all according to its homepage. But other companies use social media data to assess credit default risks.[18] The company Kreditech, for example, uses big data (including social media data) to offer alternative financial services that are transacted fast and completely online and the company provides its service 24/7—but not in Germany.[19] Often, those alternative financial services to traditional bank credits are used especially by people who were assessed as risky potential customers by banks and therefore did not obtain any credit or did not obtain a low-interest credit.[20] The only option for people with a bad credit assessment who need a credit is to agree to a credit at the cost of their privacy.[21] It becomes apparent that personal data has an economic value that many customers are not aware of.

4   Risks and Chances

Striking headlines in the media[22] and statements made by politicians have shown the risks related to scoring.[23] The central points of criticism are the lack of transparency concerning the data used and concerning the procedures, the quality and correctness of data, the length of the retention period as well as the actual and legal possibilities to correct the data influencing the score.[24] Due to long retention periods mistakes made in the past influence the data subject's present and future.[25] Scores are derived from companies' experiences with their customers by generalization. Therefore, a person could get a score, which does not meet his or her current, individual circumstances.[26]

---

[17]Schmucker, DVP 8/2013, p 322.

[18]Morozov, Bonität übers Handy, FAZ.de, http://www.faz.net/aktuell/feuilleton/silicon-demokratie/kolumne-silicon- demokratie-bonitaet-uebers-handy-12060602.html.

[19]Kreditech Holding SSL GmbH, https://www.kreditech.com/what-we-do/.

[20]Morozov 2013, Bonität übers Handy, FAZ.net, http://www.faz.net/aktuell/feuilleton/silicon-demokratie/kolumne-silicon-demokratie-bonitaet-uebers-handy-12060602.html.

[21]Ibid.

[22]For examples see http://www.heise.de/: "Scoring zur Bonitätsprüfung schwer fehlerbehaftet", "Zügelloses Scoring—Kaum Kontrolle über die Bewertung der Kreditwürdigkeit", "Studie: Scoring 'oft unverständlich', 'Aussagekraft fragwürdig'".

[23]Steinebach et al., Begleitpapier Bürgerdialog—Chancen durch Big Data und die Frage des Privatsphäreschutzes, p 33.

[24]Jandt, K&R 2015, p 7 et seq.; Steinebach et al., Begleitpapier Bürgerdialog—Chancen durch Big Data und die Frage des Privatsphäreschutzes, p 32 et seq.; see BT-Drucks. 16/10529 and BT-Drucks. 18/4864.

[25]Jandt, K&R 2015, p 7 et seq.; Steinebach et al., Begleitpapier Bürgerdialog—Chancen durch Big Data und die Frage des Privatsphäreschutzes, p 9, 32, 39.

[26]BT-Drucks. 16/10529, p 17; Jandt, K&R 2015, p 6 et seq.; Steinebach et al., Begleitpapier Bürgerdialog—Chancen durch Big Data und die Frage des Privatsphäreschutzes, p 32 et seq.

It cannot be denied that the individual can suffer disadvantages based on scoring procedures. However, it has to be kept in mind that scoring has advantages as well, and not only for companies. One side of the coin is that companies are protected against payment defaults; the other side of the coin is the consumer's protection against over-indebtedness.[27] Banks would not have any indication which customer is able to cover repayment without the assessment by a score.[28] They would charge risk premiums und would grant less credits to make up for the risk of payment defaults. The result would be higher credit costs for all customers.[29] There would not be the opportunity to get an attractive credit offer due to a positive risk assessment any more.[30]

A reasonable risk assessment also contributes to macroeconomic stability. Based on scoring, credits are granted in accordance with the customer's economic performance and therefore crises like the Subprime-crisis in 2007, which resulted in a global financial crisis,[31] can be prevented.[32]

It also needs to be taken into consideration that scoring objectifies forecasts: decisions are based on an algorithm instead of a bank employee's subjective judgment, and therefore unconscious discrimination could be avoided.[33]

5 Legal Situation

In 2009, scoring was regulated by the German federal data protection act (BDSG) for the first time. Although the legislator wanted to regulate credit scoring, neither the law itself nor its explanatory memorandum is restricted to procedures to assess

---

[27]ULD & GP Forschergruppe 2014, Scoring nach der Datenschutz-Novelle 2009 und neue Entwicklungen, p 32; SCHUFA Holding AG, https://www.schufa.de/de/ueber-uns/daten-scoring/scoring/scoring/.

[28]ULD & GP Forschergruppe, Scoring nach der Datenschutz-Novelle 2009 und neue Entwicklungen, p 92; SCHUFA Holding AG, https://www.schufa.de/de/ueber-uns/daten-scoring/scoring/scoring/.

[29]SCHUFA Holding AG, https://www.schufa.de/de/ueber-uns/daten-scoring/scoring/scoring/.

[30]Steinebach et al., Begleitpapier Bürgerdialog—Chancen durch Big Data und die Frage des Privatsphäreschutzes, p 39; ULD & GP Forschergruppe, Scoring nach der Datenschutz-Novelle 2009 und neue Entwicklungen, p 92.

[31]Subprime mortgages are those that are given to people with a poor credit history. Credit defaults in the USA were increasing and resulted in a global financial crisis, since US mortgage credits were refinanced in the international capital markets. Sources with further information to the subprime-crisis: Budzinski and Michler, Gabler Wirtschaftslexikon; Steinebach et al. 2015.

[32]Steinebach et al., Begleitpapier Bürgerdialog—Chancen durch Big Data und die Frage des Privatsphäreschutzes, p 39.

[33]Steinebach et al., Begleitpapier Bürgerdialog—Chancen durch Big Data und die Frage des Privatsphäreschutzes, p 39; ULD & GP Forschergruppe, Scoring nach der Datenschutz-Novelle 2009 und neue Entwicklungen, p 92; SCHUFA Holding AG, https://www.schufa.de/de/ueber-uns/daten-scoring/scoring/scoring/.

credit default risks.[34] The law describes scoring as a procedure that is characterized by a means-end relation: the aim is to calculate how probable a certain, future behavior of the data subject is; as means mathematical-statistical methods are employed.[35]

Simultaneously, the scored data subject was entitled to get information free of charge once a year (section 34 para. 2, 4, 8 BDSG). According to the study "Scoring nach der Datenschutz-Novelle 2009", only one out of three consumers exercised their right, probably because of the fact that not every consumer knows his or her right to information. The right to information enables the data subject to exercise his or her right to correction, deletion and blocking of data (section 35 BDSG).[36] Besides, general civil law rules for damage claims and injunctive reliefs because of privacy violation have to be kept in mind as well as the specific damage claim of data protection law (section 7 BDSG).[37]

The Federal Court of Justice of Germany stated its position on scoring in two decisions. In the first decision, the court rejected an injunctive relief concerning a negative credit assessment because the freedom of expression protects the assessment of credit default risks as long as it is based on true fact.[38] In the second decision, the court confirmed the data subject's right to get to know, which personal data is filed about him or her and has influenced the score.[39] But, the algorithm with which the score is calculated is protected as business secret so that businesses do not have to inform the data subject about the weighting of single factors or the definition of comparison groups. The Federal Court of Justice of Germany argued that the credit agencies' competitiveness depends on the secrecy of the algorithm calculating the score. The right to information does not include the right to re-calculate and check the calculation of the score. It remains to be seen which position the Federal Constitutional Court will state deciding about the constitutional complaint brought against the second decision of the Federal Court of Justice of Germany.[40]

In May 2015, the parliamentary party BÜNDNIS 90/DIE GRÜNEN proposed a draft amendment[41] concerning scoring, which is still in the legislative process.[42] The draft aims at extending the data subject's right to information and access

---

[34]BT-Drucks. 16/10529, p 1, 9, 15 et seq.

[35]BT-Drucks. 16/10529, p 9; Ehmann, in: Simitis, BDSG, section 28b Ref. 22 et seq.

[36]BT-Drucks. 16/10529, p 17.

[37]ULD & GP Forschergruppe, Scoring nach der Datenschutz-Novelle 2009 und neue Entwicklungen, p 139.

[38]BGH, MMR 2011, p 409 et seq.

[39]BGH, NJW 2014, p 1235 et seq.; detailed ULD & GP Forschergruppe, Scoring nach der Datenschutz-Novelle 2009 und neue Entwicklungen, p 45 et seq.

[40]BVerfG, 1 BvR 756/14.

[41]BT-Drucks. 18/4864.

[42]For further information see http://dipbt.bundestag.de/extrakt/ba/WP18/669/66907.html.

against credit agencies and companies concerning his or her score. Following regulations shall be put in place:

- Ex ante disclosure of scoring procedures
- Right of access concerning single data sets, weighting of single factors, assignment to comparison groups and retention periods[43]
- For credit assessment, it shall be prohibited to use data that is not relevant to the data subject's creditworthiness or that is likely to discriminate
- Credit agencies shall be obligated to actively inform the data subject
- Supervisory authority shall control compliance with data protection legislation.

The legislative proposal points out, that scoring procedures need to become more transparent. It cites the study "Scoring nach der Datenschutznovelle 2009" to substantiate its demand. In the study, a lack of transparency in the procedures is criticized, stating that it deprives the data subject of the basis for effective legal protection. The study also states that the quality of the data influencing the score is not guaranteed. Moreover, the authors of the study doubt that the scientific integrity of the scoring procedures can be guaranteed. At present, there are no legally prescribed criteria for the measurement of the scientific integrity of the mathematical-statistical procedure.

This could be a reason why supervisory authorities practically do not control scoring procedures.[44] Although supervisory authorities are already under the current legal situation empowered to control whether the calculation is based on a scientific approved mathematical-statistical procedure (section 38 BDSG), they lack capacity to control by now.[45] Under these circumstances, it is not clear how the plans of BÜNDNIS 90/DIE GRÜNEN could be implemented. Besides, it can be doubted if it is actually possible to control when big data technologies and self-learning algorithms will be used more often in the future.[46]

6 Prospect

In the future, the economic usage of data and the data subject's interests must be balanced adequately in the scoring procedure as well as concerning any other manifestation of big data.[47] That is the only possible way to guarantee that decisions based on algorithms are reliable and legal.[48]

---

[43]Contrary to BGH, Decision of 28 January 2014—VI ZR 156/13.

[44]BT-Drucks. 18/4864, p 1; Steinebach et al., Begleitpapier Bürgerdialog—Chancen durch Big Data und die Frage des Privatsphäreschutzes, p 33, 55; ULD & GP Forschergruppe, Scoring nach der Datenschutz-Novelle 2009 und neue Entwicklungen, p 133.

[45]Schulzki-Haddouti, c't 21/2014, p 38.

[46]Jandt, K&R 2015, p 7.

[47]Bitter/Buchmüller/Uecker, in: Hoeren, Big Data und Recht, p 71.

[48]Ibid.

References

Budzinski O, Michler AF (2017) Subprime-Krise. In: Gabler Wirtschaftslexikon, Springer Gabler Verlag (ed). http://wirtschaftslexikon.gabler.de/Archiv/72525/subprime-krise-v9.html. Accessed 4 Apr 2017

Ehmann E (2014) In: Simitis S (ed) Bundesdatenschutzgesetz, vol 8. Baden-Baden, Nomos

Hoeren T (ed) (2014) Big data und Recht. C. H. Beck, Munich

Jandt S (2015) Big data und die Zukunft des Scoring. Kommunikation und Recht 18(1):6–8

Landesbeauftragte für Datenschutz und Informationsfreiheit Nordrhein-Westfalen (LDI NRW) (2012) Auskunfteien—Fragen und Antworten (FAQ). https://www.ldi.nrw.de/mainmenu_Datenschutz/submenu_Datenschutzrecht/Inhalt/Auskunfteien/Inhalt/Auskunfteien/Auskunfteien_-_Haeufig_gestellte_Frage.pdf. Accessed 4 Apr 2017

Laney D (2001) 3D data management: controlling data volume, velocity and variety. http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf. Accessed 4 Apr 2017

Mansmann U (2014) Gut Gemeint. Wie die Schufa Verbraucher bewertet. c't (10):80–81. http://www.heise.de/ct/ausgabe/2014-10-Wie-die-Schufa-Verbraucher-bewertet-2172377.html. Accessed 4 Apr 2017

Morozov E (2013) Bonität übers Handy. http://www.faz.net/aktuell/feuilleton/silicon-demokratie/kolumne-silicon-demokratie-bonitaet-uebers-handy-12060602.html#Drucken. Accessed 4 Apr 2017

Rieger F (2012) Kredit auf Daten. http://www.faz.net/aktuell/feuilleton/schufa-facebook-kredit-auf-daten-11779657.html. Accessed 4 Apr 2017

Schmucker J (2013) Facebook kommunal—Kann das deutsche (Datenschutz-) Recht mit dem Wunsch nach Kommunikations- und Informationsfreiheitfreiheit in Einklang gebracht werden? Deutsche Verwaltungspraxis 64(8):319–324. http://www.dvp-digital.de/fileadmin/pdf/Zeitschriftenausgaben/DVP_Zeitschrift_2013-08.pdf. Accessed 4 Apr 2017

Schulzki-Haddouti C (2014) Zügelloses Scoring. Kaum Kontrolle über Bewertung der Kreditwürdigkeit. c't (21):38–39. http://www.heise.de/ct/ausgabe/2014-21-Kaum-Kontrolle-ueber-Bewertung-der-Kreditwuerdigkeit-2393099.html. Accessed 4 Apr 2017

Steinebach M, Winter C, Halvani O, Schäfer M, Yannikos Y (Fraunhofer-Institut für Sichere Informationstechnologie SIT) (2015) Begleitpapier Bürgerdialog. Chancen durch Big Data und die Frage des Privatsphäreschutzes. https://www.sit.fraunhofer.de/fileadmin/dokumente/studien_und_technical_reports/Big-Data-Studie2015_FraunhoferSIT.pdf. Accessed 4 Apr 2017

Unabhängiges Landeszentrum für Datenschutz Schleswig-Holstein (ULD), GP Forschungsgruppe (2014) Scoring nach der Datenschutz-Novelle 2009 und neue Entwicklungen. Abschlussbericht. http://www.bmi.bund.de/SharedDocs/Downloads/DE/Nachrichten/Kurzmeldungen/studie-scoring.pdf?__blob=publicationFile. Accessed 4 Apr 2017

Author Biographies

Stefanie Eschholz Dipl.-Jur., research associate at the Institute for Information, Telecommunication and Media Law (ITM) at the University of Münster until 2016. She holds a law degree from Münster.

Jonathan Djabbarpour B.Sc., research assistant at the Institute for Information, Telecommunication and Media Law (ITM) at the University of Münster until 2016. He currently completes his master studies in business informatics.