

A Brief History of “Scoring” from a German Perspective

by Sarah Sommer and Gert G. Wagner

The main part of this paper is based on the section “Die Geschichte des Scorings” in the report “Verbrauchergerechtes Scoring” (Consumer-friendly scoring) by the German Advisory Council for Consumer Affairs (SVRV 2018). The section was drafted by Sarah Sommer and Gert G. Wagner

Scores are numerical ratings used to predict or steer people’s behavior. These numerical ratings are usually calculated via algorithmic processes based on a broad range of data (see, e. g., AlgorithmWatch 2019).

From a historical perspective, the desire to accurately assess the characteristics, behavior and preferences of people in order to predict future developments is nothing new. Even in the analogue world, particular characteristics and behavioral patterns of individuals were (and still are) associated with certain consequences, with people being assigned numerical values in specific contexts. As a result of the digital revolution, along with the complex algorithms and large databases associated with it, the topic of “scoring” has simply acquired added relevance. Many people believe that another dimension is added by the development of artificial intelligence (AI) and self-learning algorithms. However, the tradition of rating performances and abilities using numerical values or standardized terms of evaluation is well established.

Grades in school

Today, China uses “social credit” scores to rate all of its citizens (Kostka 2018) and to punish individuals for unwanted behavior, for example, –by blocking them from traveling (Kuo 2019). This is not a completely new phenomena: In Germany, for example, schools grades have been awarded since the 16th century (Lintorf 2012) and not just for learning performance: social behavior was also graded. Nowadays, the assessment of an individual’s performance and academic achievements is still associated with various consequences, such as skipping a year in school, gaining access to secondary education, or obtaining a school-leaving qualification (e.g. the

so-called *Abitur* in Germany). In Germany, the *Abitur* grade is one of the main admission criteria for getting into university.

A pupil's overall grade is calculated based on their grades in individual subjects, with some grades (i.e. the main subjects) being weighted more heavily than others. During the admission process, pupils with the highest final grades are then accepted (usually in combination with the time spent on the waiting list) until there are no more spaces available (the "Numerus Clausus" system based on the *Abitur* grade). In this context, it becomes particularly apparent that the final school grade is not merely considered an evaluation of past achievements, but that it also holds a degree of predictive power with regard to future performance. A good *Abitur* grade is supposed to indicate that good performance can be expected at university, too, and that a degree will likely be completed successfully.

Scores in sports

Another area of life that is traditionally based on measuring the performance of people is competitive sport. However, it is not only performance that is measured. Many disciplines measure the people themselves, dividing them into different weight categories (e.g. boxing and weightlifting), giving them handicaps (golf), or creating seeding lists (e.g. tennis), all with the aim of maximizing the excitement and fairness of the competition. Ultimately, athletes are constantly characterized by their scores. Boxers are labelled by scoring their bodies, e. g. as "Heavy Weight Champion". Others are scored by their performance. Take Armin Hary, for example, known as the first man to run 100 meters in 10.0 seconds.¹ In Germany he was "scored" and labelled as the "Ten-Point-Zero-Seconds-Man" (Hary 1961).

Credit scoring

In business relationships, where contracts are frequently entered into with previously unknown business partners and a certain amount of trust is required, risk

¹ Even though, according to more accurate electronic equipment than the official manual stopwatch, it actually took him around 10.2 seconds with a barely-legal tailwind – a fact that also illustrates measurement issues inherent to scores of any kind.

minimization plays a particularly important role. As a result, scoring has been highly significant in this field for several decades. In response to the demand for information regarding the reliability and solvency of business clients, the first “credit scorers” emerged in the 19th century. In Europe, these included the companies Wys Muller (founded in 1861), Schimmelpfeng (1872), and Creditreform (1879). These companies collected financially relevant information on individuals and firms, then sold it to other businesses and banks. Since then, these companies have been an essential part of a well-functioning credit system.

The first attempts to create scores, which quantitatively calculate and numerically display a person’s risk of default, took place in the 1940s. The rudimentary scoring systems that were prevalent before then (those used by mail-order firms, for example) were based on a list of criteria. Before credit was granted, various preconditions were checked for compliance with this list and the results were then added together (Thomas, Crook & Edelman 2017).

In a research project in 1941, mathematician Davis Durand was the first person to use discriminant analysis in order to determine the risk of credit default (Durand 1941). He analyzed data on loans that had already been granted to find out which factors determined whether repayment would be problem-free or whether there would be difficulties. Based on this information, he developed a “credit score.” The first company to commercially develop statistical models for the granting of credit was Fair, Isaac and Company (now known as FICO) in California. The company has sold scoring-related products to financial institutions, retailers and mail-order companies since the 1950s (Dixon & Gellmann 2014).

In the decades that followed, mathematical progress coupled with innovation in the field of data processing ultimately enabled the development of credit scoring systems that were largely automatic. The combination of computers and algorithms, as well as the companies’ experience showing that scoring could significantly reduce the default rates on their loans and the risk of fraud, led to the credit agency scoring products we know today.

Calculation of insurance premiums

Another area in which these types of risk assessment are well-established is the insurance sector. In this field, scoring has predominantly been used to calculate both the amount being insured and the insurance premium for each individual customer. As early as the 1920s and 1930s, interest arose in Germany to establish a mathematical and statistical basis for calculating health insurance contributions. Using so-called “morbidity tables,” it was found that different medical costs could be expected depending on the sex, age, and occupation of the person being insured (Wagner-Braun 2002). Even today, in Germany, contributions are calculated individually when joining the private health insurance system. The same also applies to life insurance and occupational disability insurance. Consumers are classified (which is essentially calculating a score) based on a combination of individual factors (such as age or pre-existing conditions). This is then used to weigh up the consumer’s risk to the insurer and their premium is calculated accordingly.

One particularly complex example of premium calculation is found in the motor insurance industry, where tariffs are tailored to the individual customers based on a wide variety of different criteria. The main factors include the vehicle model type, the regional classification, and the deductible (“no-claims bracket”), as well as criteria such as the number of drivers, the age of the driver and vehicle, the number of miles being driven, and the parking location (Gesamtverband der Deutschen Versicherungswirtschaft e.V. 2016).

Another scoring system that German drivers are aware of is the Driver Fitness Assessment System, run by the Federal Motor Vehicle and Transport Authority which is based in the city of Flensburg. The system is known colloquially as “points in Flensburg”. Since 1974, this Authority has been noting “penalty points” for individual drivers in a register whenever regulatory offences and criminal offences are committed on the road. If a certain score (number of points) is reached, the Authority will revoke the license for a certain time and may order participation in driver fitness seminars before returning the license (Kraftfahrt-Bundesamt 2017).

Challenges

The various forms of scoring described above all have a long tradition and there is no doubt scoring existed in the analogue era. However, it is undeniable that the implementation of scoring has changed dramatically following the technological developments of the digital age. This year, for example, France started allocating its university places using a scoring algorithm called Parcoursup, which evaluates whether the admission criteria have been met and takes into account the place of residence and the preferences of the applicant (Joeres 2018). In online retail, a consumer's creditworthiness can be calculated automatically in just a matter of seconds so that appropriate payment options can then be offered. In the motor insurance sector, telematic tariffs now hold sway – continuously evaluating driving behavior and adjusting insurance premiums based on the resulting score.

Furthermore, algorithmic scoring is increasingly being used in many new areas and now evaluates consumers and consumer groups in the most diverse ways – with highly varied results (Dixon & Gellmann 2014, AlgorithmWatch 2019). There are scores that predict a household's purchasing power or its willingness to donate to charity (Equifax 2018; Blackbaud 2014), scores showing whether customers will migrate to other companies (Versium Analytics Inc. 2018), scores that aim to detect pregnancies (Duhig 2012), and scores that measure energy consumption behavior (Trove 2018). Dating services are also based on scores quantifying how well personal profiles match (Carr 2016).

A culture of evaluation and quantification is emerging (Mau 2017). From 'likes' on Facebook to the number of Twitter followers or stars on Airbnb – we are long past the point where only companies use algorithms to assess consumers and assign them numerical values. Scoring has truly become part of our daily lives.

Lessons learnt

The examples given above help to understand how scoring of people can work and under what circumstances scoring is accepted. The German Advisory Council for Consumer Affairs (SVRV 2018) concludes:

“The potential offered by scoring systems can be exploited to the full only when society’s various legitimate expectations are met. Only then will consumers accept and benefit from these systems. Consumer-friendly scoring is the state that will be achieved when these conditions are satisfied. This will include applying data protection in a way that minimizes the risk of mistaken identity, and providing simple and effective ways for individuals to appeal their score. Protected characteristics (such as gender) may not be used as a basis for unwarranted discrimination – directly or indirectly. When scores are calculated for predictive purposes, the quality of the criteria applied and the reliability of the predictions must be demonstrated. Moreover, predictive powers of this caliber should remain stable across a variety of socio-economic groups. Furthermore, the communicated objective of a score should not be misleading to the affected individuals. The predictions made must correspond to the objectives of the scoring system concerned, and should not be applied frivolously to areas other than those for which the score was calculated. Above all, however, scoring systems must be comprehensible to those who are scored.”

References

- AlgorithmWatch (2019). *Automating Society: Taking Stock of Automated Decision-Making in the EU – A report by AlgorithmWatch in cooperation with Bertelsmann Stiftung, supported by the Open Society Foundations*. Berlin:
AlgorithmWatch. www.algorithmwatch.org/automating-society, last accessed on March 4, 2019.
- Blackbaud (2014). *Target Analytics ProspectPoint*.
https://www.blackbaud.com/files/resources/downloads/01.14.ANLY_%20ProspectPoint.datasheet.pdf, last accessed on May 14, 2018.
- Carr, A. (2016). *I found out my secret internal tinder rating and now I wish I hadn't*.
<https://www.fastcompany.com/3054871/whats-your-tinder-score-inside-the-apps-internal-ranking-system>, last accessed on September 21, 2018.
- Dixon, P. & Gellmann, R. (2014). *The scoring of america: how secret consumer scores threaten your privacy and your future*. <http://www.worldprivacyforum.org/wp->

- [content/uploads/2014/04/WPF_Scoring_of_America_April2014_fs.pdf](#), last accessed on June 7, 2018.
- Duhigg, C. (2012). *How companies learn your secrets*.
https://www.nytimes.com/2012/02/19/magazine/shopping-habits.html?_r=1&pagewanted=all, last accessed on May 14, 2018.
- Durand, D. (1941). *Risk elements in consumer instalment financing*. New York: National Bureau of Economic Research.
- Equifax (2018). *Discretionary Spending Index*.
<https://www.equifax.com/business/discretionary-spending-index/>, last accessed on May 14, 2018.
- Gesamtverband der Deutschen Versicherungswirtschaft e.V. (2016). *So setzt sich der Versicherungsbeitrag für einen Pkw zusammen*.
<https://www.gdv.de/de/themen/news/so-setzt-sich-der-versicherungsbeitrag-fuer-einen-pkw-zusammen-11804>, last accessed on June 8, 2018.
- Hary, Armin (1961). 10,0. München: Copress Verlag.
- Joeres, A. (2018). *Parcoursup – das außerirdische Universitätsauswahlssystem der französischen Regierung*. <https://algorithmenethik.de/2018/05/30/parcoursup-das-ausserirdische-universitaetsauswahlssystem-der-franzoesischen-regierung/>, last accessed on June 15, 2018.
- Kostka, G. (2018). *China's Social Credit Systems and Public Opinion: Explaining High Levels of Approval*. <https://ssrn.com/abstract=3215138>, last accessed on September 27, 2018.
- Krafftahrt-Bundesamt (2017). *Rund um den Punkt - kurz gefasst*.
https://www.kba.de/SharedDocs/Publikationen/DE/Presse/rund_um_den_Punkt_kurz_gefasst_faltblatt_pdf.pdf?__blob=publicationFile&v=6, last accessed on Oktober 1st, 2018.
- Kuo, Lily (2019). *China bans 23m from buying travel tickets as part of 'social credit' system*. The Guardian, March 1, 2019 ,
<https://www.theguardian.com/world/2019/mar/01/china-bans-23m-discredited-citizens-from-buying-travel-tickets-social-credit-system>, last accessed on March 4, 2019.
- Lintorf, K. (2012). Zur individuellen und gesellschaftlichen Bedeutung von Schulnoten. In K. Lintorf, *Wie vorhersagbar sind Grundschulnoten? Prädiktionskraft individueller und kontextspezifischer Merkmale* (S. 19-35). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Mau, S. (2017). *Das metrische Wir. Über die Quantifizierung des Sozialen*. Berlin: Suhrkamp Verlag.

- SVRV (Sachverständigenrat für Verbraucherfragen: Advisory Council for Consumer Affairs) (2018), *Verbrauchergerechtes Scoring*, Berlin: http://www.svr-verbraucherfragen.de/wp-content/uploads/SVRV_Verbrauchergerechtes_Scoring.pdf.
- Thomas, L., Crook, J. & Edelman, D. (2017). *Credit scoring and its applications, second edition*. Philadelphia: Society for Industrial and Applied Mathematics.
- Trove (2018). *Predictive data science sharpens hourly load forecasting*. <http://trovedata.com/results/case-study/predictive-data-science-sharpens-hourly-load-forecasting>, last accessed on June 25, 2018.
- Versium Analytics Inc. (2018). *Custom predictive scores solve business challenges*. <https://versium.com/predictive-scores/>, last accessed on May 14, 2018.
- Wagner-Braun, M. (2002). *Zur Bedeutung berufsständischer Krankenkassen innerhalb der privaten Krankenversicherung in Deutschland bis zum Zweiten Weltkrieg*. Stuttgart: Franz-Steiner Verlag.