

Aleksandra Sowa

Big Data Mining – oder: Antwort auf nie gestellte Fragen

»Big data is like teenage sex: everyone talks about it, nobody really knows how to do it, everyone thinks everyone else is doing it, so everyone claims they are doing it...«
(Dan Ariely)

Von der »Verdrängung des Theoretischen durch das Faktische« spricht Roberto Simanowski in seinem Buch *Data Love*. Und meint damit Big Data Mining, wie »die computergesteuerte Analyse großer Datensammlungen auf bestimmte Gesetzmäßigkeiten und unbekannte Zusammenhänge hin« heute genannt wird. Mit *Big Data Mining* wird die Datensammelwut der Internetkonzerne und der Behörden kanalisiert. Und treibt diese zugleich an.

Internetkonzerne wie Google und Facebook stehen für die Vorteile, die »eine resolute und effektive Datenakkumulation und -analyse dem Kunden verspricht«. Jared Cohen und Eric Schmidt reden deswegen in *Die Vernetzung der Welt* vom Datenstrom als Geschenk für Behörden und Unternehmen, »mit dem sie auf die Bedürfnisse ihrer Bürger und Kunden eingehen, spezifische demographische Gruppen erreichen und mithilfe neuer Methoden künftige Entwicklungen prognostizieren können«. Ob als der berühmte Grippeindex von Google oder als Kundenfürsorge: Ziel der Datensammlung und -analyse sei immer der Mensch. Und zwar: »die Verbesserung seiner wirtschaftlichen, sozialen, gesundheitlichen Situation«. Das betonen Internetkonzerne und Behörden gleichermaßen.

Erstmalig erfolgt die Datenverarbeitung jedoch ohne Theoriebildung. Datenströme werden auf Korrelationen ausgewertet ohne darauf zu achten, dass ein ursächlicher Zusammenhang besteht, sobald zwei Tatsachen zueinander in Beziehung gestellt werden. Nicht schon wieder Wissenschaft, sagen an dieser Stelle viele. Doch bei Massendatenauswertungen ist es ge-

nau die Wissenschaft, die Analyse, das Aufstellen von Vermutungen, Hypothesen, Annahmen etc. – all das, was ein repräsentatives Ergebnis von einem Haufen nutzloser Zahlen unterscheidet. Simanowski spricht im Kontext von Big Data sogar von einer gewissen Theoriemüdigkeit. Die neue Qualität wird durch mehr Quantität erreicht. Datenverarbeitung ohne Theoriebildung entfernt sich damit davon, was (traditionelle) Wissenschaft ausmacht. Big Data Mining führt nicht unbedingt zu mehr Wissen.

Oder eventuell doch, wenn man das »Wissen« betrachtet, das sich im Netz gegenwärtig im situativen Wissenstypus von Wikipedia spiegelt: »Der kumulative Besitz von Wissen [*having knowledge*] weicht dem virtuosen Jonglieren mit Wissen [*doing knowledge*]«.

Was damit gemeint ist, zeigt das Storch- und-Baby-Beispiel. Rudolf Flesch zitiert in *Besser schreiben, sprechen, denken* einen Statistiker, der entdeckte, dass es eine Korrelation von 0.9 zwischen der Anzahl von Storchennestern und den Geburtenraten in Stockholm während einer bestimmten Anzahl von Jahren gab. Der strengen Wenn-dann-Logik des Big Data Mining folgend müsste daraus zwangsläufig die Voraussage resultieren, dass die Geburtenrate in irgendeinen Zusammenhang mit der Anzahl von Storchennestern gebracht werden kann. Die Big-Data-Logik ist die Logik eines Dreijährigen.

»Big Angst« vor den Massendatenauswertungen liegt nicht nur in der potenziellen Beeinflussbarkeit der Statistik. Die »Angstworte der Zukunft« heißen tatsächlich *predictive analytics* und *algorithmic regulation*, also weniger Auswertung der

Datenströme zur Bestätigung bestimmter Hypothesen oder Vermutungen, sondern ihre Verwendung als Prognoseinstrument nach Art des Films *Minority Report*. Eine Mustersuche kann zur Identifizierung von potenziellen Tätern führen – oder eben nur zur Erkennung von Abweichlern vom Durchschnittswert.

Auf diesem Wege kann »das Gesellschaftliche auf Mathematik reduziert« und können alternative Positionen blockiert werden. Eli Pariser, US-amerikanischer Jurist und Aktivist, klagte beispielsweise, dass ihm von Facebook nicht die Updates seiner konservativen Freunde angezeigt werden, da der Facebook-Algorithmus erkennt, dass er sich stärker für die Updates seiner linksorientierten Freunde interessiert. Die Personalisierung des Informationsangebots mag im Hinblick auf Werbung oder Einkäufe effizient und dem Internetnutzer willkommen sein, jedoch im Politischen, wo die Konfrontation verschiedener Standpunkte »das Lebenselixir der Demokratie« darstellt, mutiert das Internet eventuell zum »gemütlichen Heim der Autopropaganda, die das Subjekt permanent bestätigt und das, was es (noch) nicht ist, unablässig ausblendet«. Die Logik des Algorithmus siegt über die Logik des Subjekts, gelegentlich auch die Meinung der Gegenseite zur Kenntnis nehmen zu wollen.

In der Zeit vor Big Data waren sich Wissenschaftler einig darüber, dass es bei wissenschaftlicher Erforschung eines Problems zwei Gruppen gibt: Leute, die Daten sammeln, und solche, die Annahmen erstellen. Und man war sich einig, dass die, die Annahmen erstellen, wichtiger sind. »Vornehmer ausgedrückt«, schrieb Flesch, »bedeutet das, dass die Betonung mehr auf Deduktion als auf Induktion liegt und dass die aristotelische Methode ... höher einge-

schätzt wird als die von Bacon.« Genau die Theoriebildung fehlt aber bei Big Data Mining. Das *Das* (etwas ist so) mache das *Warum* (deshalb ist es so) überflüssig. Auf diese Weise verändert Big Data Mining unser Verhältnis zur Welt und zum Wissen, behauptet Simanowski. Schon heute würden die politischen Blogs zu 90 % auf »ähnlich denkende« Websites verlinken, stellt Simanowski fest.

Um den Vorwurf der »Feudalisierung« der Erkenntnisse aus dem Big Data Mining zu relativieren, antwortet die Wissenschaft mit einer »doppelten Demokratisierung« der Erkenntnis. Einerseits wird der Zahl eine Objektivität, ja Unbeeinflussbarkeit und Unabhängigkeit, unterstellt. Andererseits werden die Zahlen im Rahmen einer Self-Tracking-Bewegung von den Nutzern (und Untersuchungsobjekten zugleich) erzeugt, weitergegeben und/oder verkauft – »als empirische Soziologie von unten«. Bisher sei die Datenfreigabe vorrangig freiwilliger Natur, betonen Cohen und Schmidt. Die Nutzer veröffentlichen bestimmte Inhalte wie Vorlieben, Entscheidungen, Absichten oder Angewohnheiten aus ihren individuellen sozialen oder wirtschaftlichen Motiven. Das ist jedoch nicht unbedingt die Regel.

»Wenn neben der liberalen Marktwirtschaft kein anderes Gesellschaftsmodell mehr als Alternative zur Verfügung steht, überzeugt der Befund vom falschen Leben kaum mehr – oder eben nur in dem Sinne, wie Winston Churchill von der Demokratie sprach: als schlechteste aller Staatsformen, ausgenommen alle anderen«, schreibt Roberto Simanowski. Auch für die Demokratie kann es dann dank Big Data bald so weit sein, dass die neue Qualität durch mehr Quantität erreicht wird. Kurz: »Masse statt Klasse«.



Aleksandra Sowa

leitete zusammen mit dem deutschen Kryptologen Hans Dobbertin das Horst-Görtz-Institut für Sicherheit in der Informationstechnik. Sie ist Autorin zahlreicher Fachpublikationen und in einem großen Telekommunikationskonzern tätig.