Digital gebrandmarkt

Wie Konsumentendaten gesammelt, gehandelt und genutzt werden

Algorithmen werten in großem Stil hinterrücks gesammelte Daten aus und prognostizieren Kreditwürdigkeit, Verhalten oder Gesundheit von Bürgern. Wo die eigenen Daten nicht reichen, kaufen Unternehmen fremde zu. Die Gefahr wächst, dass wir uns durch Big Data und Mathematik steuern und diskriminieren lassen.

Von Markus Morgenroth

Menschen sind dort besonders gut auszuspähen, wo sie sich sicher und unbeobachtet fühlen. Angenommen, Sie haben Probleme in der Beziehung und wollen einen Paartherapeuten konsultieren. Davon erzählen Sie nicht einmal den engsten Freunden. Aber würden Sie bei der Wahl und der Bezahlung dieses Therapeuten darüber nachdenken, ob damit Ihre Kreditwürdigkeit auf dem Spiel stehen könnte? Sicher nicht.

Ein US-amerikanisches Kreditkarten-Unternehmen dachte da einige Schritte weiter. Es war auf eine Korrelation zwischen Beziehungsproblemen und potenziellen finanziellen Belastungen in der Zukunft gestoßen: Wer sich scheiden lässt, gerät eher in finanzielle Schieflage. Das Unternehmen kürzte daraufhin Ehepartnern, die über ihre Kreditkarte eine Paartherapie buchten, die Kreditlinie, wie der Harvard-Juraprofessor Frank Pasquale nachwies [1].

Das Vorgehen wäre in Deutschland sicher nicht rechtmäßig (siehe Artikel auf S. 76). Dennoch zeigt das Beispiel, wie Datenanalysten aus einfachen Korrelationen Modelle, Vorhersagen und Anweisungen bauen. Meist geschieht das hinter dem Rücken der Betroffenen.

Unternehmen aus aller Welt sammeln und bewerten pausenlos persönliche Daten und verwenden sie dazu, zu analysieren, zu durchleuchten und digitale Stempel aufzudrücken. Jeder Zahlvorgang, jede Suchanfrage, jedes Posting in sozialen Netzen wird registriert; Smartphone und Fitnessarmband pumpen Standort und Vitalwerte in die Hersteller-Cloud und auch an vielen völlig unerwarteten Stellen hinterlässt jeder von uns auswertbare Datenspuren.

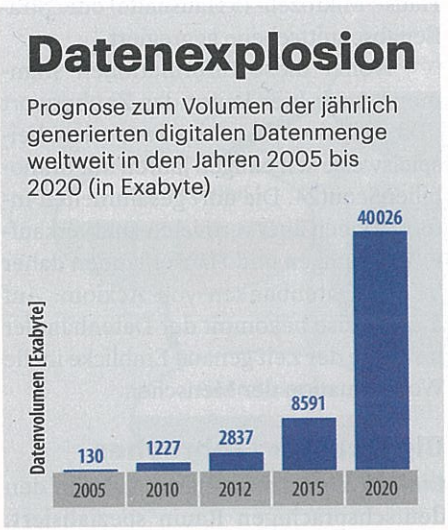
Wer meint, doch „nichts zu verbergen" zu haben, verkennt die Tragweite der Informationen, die sich aus Daten destillieren lassen: Unternehmen nutzen sie in großem Maßstab dazu, um auf Charaktereigenschaften, Leistungsfähigkeit, Intelligenz, Gemütsverfassung, Bildungsniveau, Krankheitswahrscheinlichkeiten, psychopathische Veranlagung, Kreditwürdigkeit und viele andere Eigenschaften von Menschen zu schließen – und dies oft ohne deren Wissen. Sind Sie ein wertvoller Kunde, ein loyaler Arbeitnehmer, ein vorbildliches Krankenversicherungsmitglied – oder eben nicht?

Anreicherung

Die Menge der von einzelnen Personen hinterlassenen Daten wächst exponentiell. 90 Prozent aller durch Internet-Nutzung erzeugten Daten sind in den letzten zwei Jahren entstanden. Der Festplattenhersteller Seagate prognostiziert, dass die jährlich produzierte Menge von 3,5 Zettabyte im Jahr 2013 auf 40 Zettabyte im Jahr 2020 ansteigen wird. Ein Zettabyte sind eine Milliarde Terabyte.

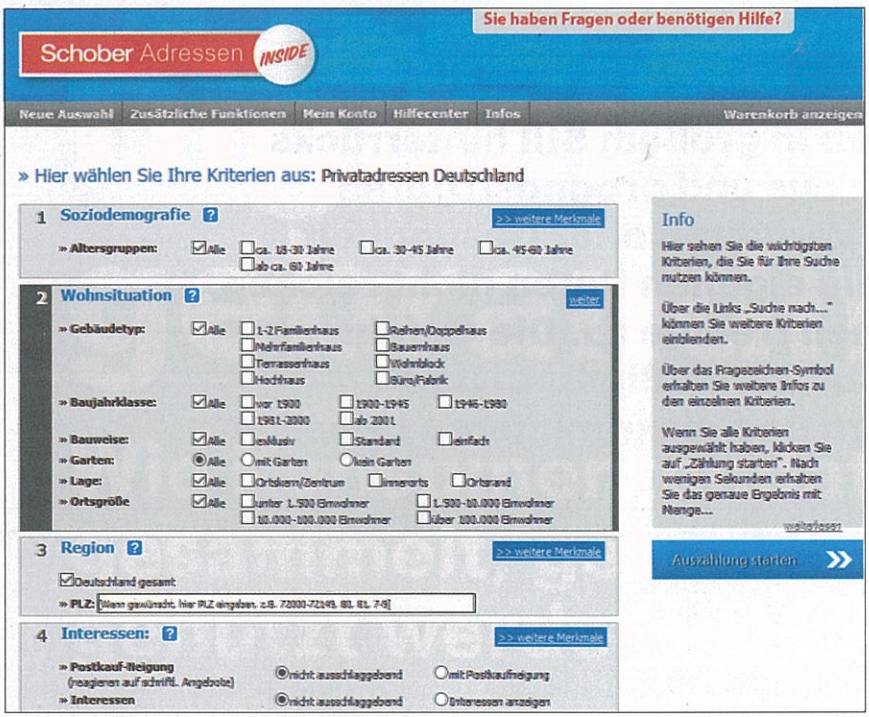
Derzeit sammeln Konzerne und staatliche Behörden wie die NSA alles, was möglich ist. Vieles wird erst einmal unstrukturiert weggespeichert, weil noch Auswertungsmethoden oder Rechenpower fehlen. Eines ist aber sicher: Nie war transparenter, was wir im letzten Sommer getan haben. Und im Sommer davor. Und dazwischen.

Big Data bedeutet nicht nur, viele Daten zu sammeln und auszuwerten, sondern insbesondere auch, verschiedenste Datenquellen zu fusionieren. Führt ein sogenannter Data Scientist eigene Daten mit Fremddaten zusammen, spricht er von einer „Anreicherung" – die Fusion führt zu einem Informationsgewinn: Ergänzt er etwa die E-Mail-Adresse einer Person um deren Postadresse oder gar Charaktereigenschaften, macht er die Daten für die Verwendung wertvoller. Deshalb erscheint es nur folgerichtig, dass um all unsere persönlichen Daten ein munteres Geschachere entstanden ist.



Big-Daten-Business

Viele dubiose, kleine Firmen bevölkern den Datenmarkt. Doch beherrscht wird er von großen, international agierenden Konzernen, wie zum Beispiel Acxiom, Datalogix, Rapleaf, Core Logic oder PeekYou. Acxiom, einer der Branchenriesen, erwirtschaftet weltweit mehr als eine Mil-



Beim Datenhändler Schober lässt sich die Zielgruppe für den Adressenkauf recht genau definieren.

liarde US-Dollar pro Jahr und verwaltet über 15.000 Datenbanken für seine über 7000 Kunden. Der Konzern verfügt über 700 Millionen aktive Konsumentenprofile, darunter mehr als 40 Millionen aus Deutschland.

Pro Haushalt listet Acxiom durchschnittlich 1500 Einzelangaben in seinen Datenbanken auf. In Deutschland teilt der Konzern die Bevölkerung – beruhend auf Alter, Familientyp und Sozialstatus – in 14 Hauptgruppen ein, etwa in „Alleinerziehend & statusarm", „Midlife-Single & gut situiert", „Goldener Ruhestand & aktiv". Nach verschiedenen Lifestyle-Merkmalen erfolgt eine weitere Kategorisierung in über 200 Untergruppen. Darin unterscheidet Acxiom dann etwa Raucher von Nichtrauchern oder nach Vorlieben in den Bereichen Sport, Freizeit, Technik, Telekommunikation und Tourismus.

Acxiom bietet anderen Datensammlern an, ihre bereits existierenden Kundenprofile mit Informationen aus den Acxiom-eigenen Datenbanken anzureichern. Auf diese Weise können äußerst umfangreiche Persönlichkeitsprofile entstehen. Das Unternehmen agiert sehr abgeschirmt. Alarmieren sollten aber bereits die Aussagen im Firmenprospekt: Man verfüge über ein „einzigartiges Spektrum an Markt- und Konsumentendaten" mit dem die Datenanreicherung und „präzise Qualifizierung nahezu jeder postalischen Anschrift in Deutschland" möglich sei.

Der Konzern ist einer von wenigen Partnern, von denen sich selbst der Datenkrake Facebook noch Informationsgewinn verspricht. Facebook kauft für verschiedene Länder, darunter auch Deutschland, Acxiom-Datensätze zu. Das Ziel der 2015 gestarteten Kooperation: Die Facebook-Werbekunden sollen ihre Zielgruppen noch präziser als ohnehin schon abschöpfen können. Im Ad-Management-Tool von Facebook ist nun der Hinweis zu finden: „Acxiom Mikrotyp beinhaltet umfangreichste Zielgruppendaten und entspricht dem BDSG. Alle Daten sind statistische Schätzwerte und je nach Sensitivität auf Haus-, Mikrozell- (5 Haushalte) oder Straßenabschnittsebene aggregiert."

Woher diese Informationen stammen? Auch hier lautet das Zauberwort „Datenfusion": Acxiom kooperiert beispielsweise seit einigen Jahren mit ImmobilienScout24. Die dort gesammelten Informationen über vermietete und verkaufte Wohnungen und Häuser landen daher in den Datenbanken von Acxiom. Auf diese Weise bekommt der Datenhändler im Laufe der Zeit genaue Einblicke in die Wohnsituation der Menschen.

Big Data für Einbrecher

Einige Daten-Broker haben sich auf den deutschsprachigen Raum spezialisiert. Die Datenbanken der zur Bertelsmann-Tochter Avarto gehörenden Firma AZ Direct enthalten durchschnittlich 600 Attribute für über 70 Millionen Menschen und 40 Millionen Haushalte in Deutschland – und somit über so gut wie jeden Einwohner des Landes.

Schober, ein weiterer Datenhändler aus Deutschland, hat nach eigenen Angaben zwar nur „30 Millionen Privat-Adressen mit jeweils über 300 Zusatzmerkmalen" in seiner Datenbank. Das Besondere aber ist, dass jeder zumindest einen sehr kleinen Ausschnitt dieser Daten kaufen kann.

Im Schober-Webshop [2] lässt sich eine gewünschte Zielgruppe auswählen, um die dazugehörigen Privatadressen zu erwerben. Selektiert wird nach Geschlecht, Alter, Haushaltsgröße, Kaufkraftprognose, Postleitzahlengebieten und Interessen. Auch die gewünschte Wohnsituation kann man auswählen, einschließlich bestimmter Gebäudetypen, Baujahren oder zugehörigem Garten.

Der Webshop stellt ab 24 Cent Stückpreis die zur Auswahl passenden Adressen bereit. Wer dort nach alleinlebenden, älteren Menschen sucht, die luxusaffin sind und in exklusiven Einfamilienhäusern wohnen, kann sich schnell und günstig eine Liste von vielversprechenden Adressen für Einbrüche besorgen.

Datenpool Gesundheitswesen

Einige Datenhändler fokussieren sich auf den lukrativen Gesundheitsbereich. Dominiert wird dieser Markt vom US-amerikanischen Konzern IMS Health, der auch in Deutschland 300 Mitarbeiter beschäftigt. Mehr als 16.000 Kunden nehmen die Beratungsdienste und Big-Data-Analysefähigkeiten des Konzerns in Anspruch, darunter Pharma-Unternehmen, Biotech-Firmen, Kostenträger, Ärzte und Krankenhäuser.

Das Unternehmen sammelt schon seit über 60 Jahren Gesundheitsdaten. Heute nutzt IMS Health nahezu 100.000 Datenquellen und aggregiert so Informationen über verschriebene Medikamente, eingereichte Krankenversicherungsansprüche, elektronische Krankenakten, Umfrageergebnisse sowie Profil- und Kontaktinformationen von Patienten. Auch die sozialen Netzwerke durchsucht

der Konzern nach verwertbaren Informationen über Patienten.

Wie skrupellos er auch in Europa Daten absaugt, wurde 2013 öffentlich: IMS Health erhielt den österreichischen Big Brother Award. Grund war ein Angebot, mit dem sich die Arztsoftware-Firma CompuGroup an österreichische Ärztinnen und Ärzte gewandt hatte. In Kooperation mit IMS Health wollte man sich Zugriff auf anonymisierte Patientendaten der Arztpraxen erkaufen. Mehrere Hundert Ärztinnen und Ärzte sollen das Angebot angenommen und für die Datenlieferungen 432 Euro pro Jahr erhalten haben.

Daten wie Geschlecht, Geburtsjahr, Krankenscheinart, Diagnose, Medikamente, Dosierung, Therapie und Laborwerte sollten anonymisiert an IMS Health geliefert werden. Doch Datenschützer waren sich einig, dass sich mit den betroffenen Daten Rückschlüsse auf einzelne Personen ziehen lassen – nachgewiesen wurde das allerdings nie.

Auch in Deutschland greift IMS Health Gesundheitsdaten ab und war in Kritik geraten, weil es Apothekenrechenzentren die von ihnen erfassten Rezept- und Patientendaten abkauft. Zwar ist der Handel mit Patientendaten grundsätzlich nicht verboten. Die Voraussetzung dafür ist, dass sie zuvor ausreichend anonymisiert werden. Wie in Österreich bezweifeln das Datenschutzbehörden auch hierzulande. IMS Health behauptet stets, dass der Aufwand einer Deanonymisierung in der Regel zu hoch sei.

Anonym oder Pseudonym

Genau dieser Punkt ist entscheidend: In Datenschutzerklärungen zu Online-Diensten liest man meist, dass persönliche Daten ausschließlich in anonymisierter Form an Drittfirmen weitergegeben werden. Also alles halb so wild? Nein, denn die Unternehmen sprechen häufig fälschlicherweise von anonymisierten Daten. In Wahrheit pseudonymisieren sie lediglich – sie tauschen Attribute eines Datensatzes, die eindeutig auf eine Person schließen lassen, durch Pseudonyme aus.

Dabei ersetzen sie beispielsweise den Namen einer Person in verschiedenen Datensätzen aber mit demselben Pseudonym – der Zusammenhang bleibt erhalten, und man kann mit den Daten arbeiten. Oft enthalten diese pseudonymisierten Datensätze immer noch genügend Informationen, um sie eindeutig einer Person zuzuordnen. Genau dies soll bei den von IMS Health gekauften Datensätzen nach Aussage des damaligen Landesbeauftragten für den Datenschutz in Schleswig-Holstein Thilo Weichert der Fall gewesen sein.

Allein die Kombination aus Geschlecht, Geburtsdatum und der Postleitzahl des Wohnorts ist bei 87 Prozent der US-amerikanischen Bevölkerung laut einer Untersuchung einzigartig. Solange also diese drei Attribute noch in einem Datensatz stecken, braucht es meist weder den Namen, die E-Mail-Adresse, die Postadresse oder andere Attribute, um einen Datensatz eindeutig einer einzelnen Person zuzuordnen.

Heute nutzen Unternehmen meist Telefonnummern, E-Mail-Adresse oder Geräte-IDs zur Identifizierung einer Person. Diese Attribute werden gehasht. Damit soll dem Datenschutz Genüge getan sein, denn die Datensätze gelten nun als anonymisiert – weil die Hash-Funktion nur in eine Richtung funktioniert, ist es nicht mehr möglich, den ursprünglichen Wert wiederherzustellen.

Was aber, wenn andere Unternehmen ihre Datensätze mit derselben Hash-Methode anonymisieren? Eine E-Mail-Adresse würde in allen Datenbanken durch denselben Hashwert ersetzt. Wenn ein Unternehmen also anonymisierte Daten verkauft, dann kann der Empfänger diese Daten leicht mit den eigenen abgleichen. Auf diese Weise werden die Datensätze miteinander verbunden, über mehrere Datenbanken und Unternehmen hinweg. Daher fordern viele Datenschützer, gehashte E-Mail-Adressen, Telefonnummern und Geräte-IDs als personenbezogene Daten zu klassifizieren.

Big Five

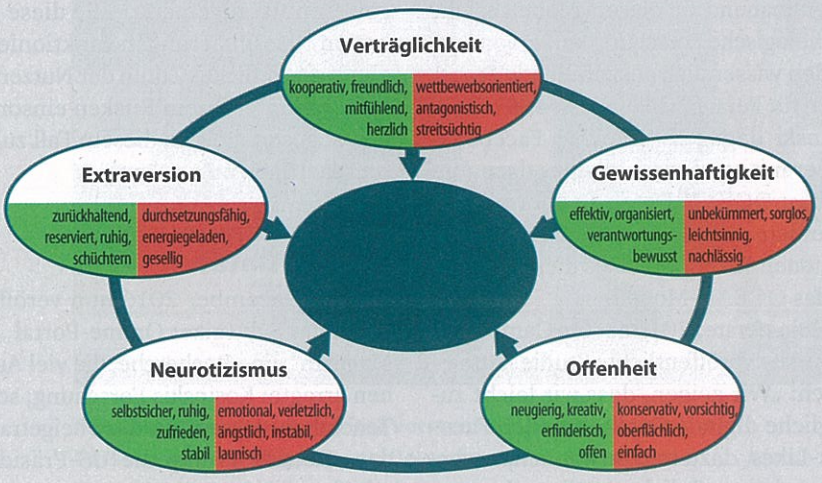
Über derlei Bestrebungen kann der Big-Data-Gigant Facebook nur lächeln. Ohne Argwohn beschenken mehr als eine Milliarde Menschen weltweit den US-Konzern mit personenbezogenen, höchst privaten Daten – teilweise, ohne es zu bemerken [3]. Facebook nutzt die daraus extrahierten Informationen, damit seine Werbekunden möglichst exakt ihre Zielgruppe zum genau richtigen Zeitpunkt erwischen.

Diese „Microtargeting“ genannte Methode funktioniert bei Facebook besser als irgendwo sonst. Der Konzern ist nicht nur ein riesiger Datensauger – er versteht es auch, die Daten intelligent zu verknüpfen und verwertbar zu machen. Deshalb lohnt sich Werbung auf der sozialen Plattform, und deshalb verdient Facebook Milliarden damit.

Facebook beschränkt sich bislang auf dieses Geschäftsfeld. Was sich mit dem ständig wachsenden Datenberg sonst

Das OCEAN-Modell

Jede Persönlichkeit lässt sich in fünf Teildimensionen „vermessen“. Diese lassen sich über Fragebögen ermitteln – oder via Big-Data-Analyse.



Ein Artikel über Big-Data-Analysen zur Manipulation der US-Wähler verbreitete sich in sozialen Netzwerken rasant.

noch anstellen ließe, zeigen immer wieder Experimente. Insbesondere die Forschungen von Michal Kosinski sorgen seit Jahren für Aufsehen und auch für Unruhe bei Facebook. Der gerade mal 34-jährige Kosinski ist Professor für Verhaltenspsychologie an der Stanford University.

Für seine Studien greift er auf Psychometrie zurück, namentlich auf das in der Psychologie gängige OCEAN-Modell zur Vermessung von Persönlichkeit. Dieses auch „Big Five" genannte Modell geht davon aus, dass jede menschliche Persönlichkeit in fünf Dimensionen beschrieben werden kann, nämlich in Aufgeschlossenheit, Gewissenhaftigkeit, Extraversion (Geselligkeit), Verträglichkeit und Neurotizismus (seelische Verletzlichkeit).

Aufbauend auf diesem Modell hat die psychologische Forschung kurze, von Probanden wissentlich auszufüllende Fragebogen für Persönlichkeitstests entwickelt. Kosinski dagegen verlockte Facebook-Nutzer mit spielerischen Apps dazu, unbewusst OCEAN-Daten zu übermitteln. Außerdem projizierte er anhand von Korrelationen Facebook-Likes dieser Nutzer auf das OCEAN-Modell.

Seine erste, 2013 an der Cambridge University veröffentlichte Studie hatte es in sich: „Wir zeigen, dass wir leicht zugängliche digitale Daten, nämlich Facebook-Likes, dazu nutzen können, zuverlässig sehr persönliche Attribute herauszubekommen." Die Datenanalyse habe mit 88 Prozent Zuverlässigkeit homosexuelle von heterosexuellen Männer unterschieden, die Hautfarbe sei in 95 Prozent korrekt ermittelt worden, zu 85 Prozent habe er die politische Einstellung erkannt [4]. Diese Ergebnisse erhielt Kosinski, in dem er Korrelationen der Big-Five-Einordnungen zu Kontrollgruppen zog, deren Attribute bekannt sind.

Laut Aussage der beteiligten Forscher beruhen die Vorhersagen nur zu einem verschwindend geringen Anteil auf offensichtlichen Verbindungen, etwa wenn ein Raucher eine Zigarettenmarke geliked hat. Tatsächlich seien Korrelationen zwischen den Persönlichkeitseinschätzungen und Vergleichsgruppen entscheidender gewesen. Wohlgemerkt: Alle diese konkreten Klassifizierungen funktionierten ohne wissentliches Zutun der Nutzer – sie wurden hinter ihrem Rücken einsortiert und abgestempelt, in diesem Fall zu wissenschaftlichen Zwecken.

Wahlmanipulation mit Big Data?

Anfang Dezember 2016 nun veröffentlichte das Schweizer Online-Portal „Das Magazin" eine Recherche, die viel Aufsehen erregte: Kosinskis Forschung, so der Tenor, habe wesentlich dazu beigetragen, dass Donald Trump die US-Präsidentschaftswahl im November gewonnen hat. Es gebe ein für Trump arbeitendes Big-Data-Unternehmen namens Cambridge Analytica, das beruhend auf seinen Ergebnissen Wähler klassifiziert und anschließend manipuliert habe. „Ich habe nur gezeigt, dass es die Bombe gibt", zitieren die beiden Autoren des Artikels Kosinski direkt in der Headline [5].

Der Text verbreitete sich wie ein Lauffeuer in den sozialen Medien. Viele erkannten eine Art Sündenfall: Erstmalig sei ein Wahlergebnis über Facebook manipuliert worden. Andere wiederum bezweifeln, dass es möglich ist, mit zielgerichteter Werbung oder manipulativen Nachrichten die politische Stimmung in einem ganzen Staat zu drehen.

Die Autoren selbst ruderten daraufhin etwas zurück. Auch Kosinski relativierte in Interviews den Einfluss seiner Forschungen und den von Cambridge Analytica. Das britische Unternehmen selbst bestätigt zwar, über OCEAN-Persönlichkeitsprofilen von 230 Millionen US-Bürgern mit jeweils mindestens 5000 Datenpunkten zu verfügen. Allerdings widersprach ein Sprecher gegenüber dem Wired-Magazin jüngst in einem wichtigen Punkt der Recherche: „Wir verwenden keine Facebook-Daten."

Man habe „die Wählerdatenbank der Republikaner, politische Daten sowie online und offline erwerbliche Daten" herangezogen. Cambridge Analytica kauft seine Daten unter anderem bei dem erwähnten Konzern Axciom ein. Außerdem seien selbst durchgeführte Umfragen eingeflossen. Beruhend auf den ermittelten Profilen wurden zielgerichtet „diverse Video-Ads, Native-, Display- und Search-Advertising sowie Facebook, Twitter und Snapchat-Nachrichten" eingesetzt, teilte Cambridge Analytica mit.

Die Wahrheit dürfte wohl irgendwo in der Mitte liegen: Cambridge Analytica hebt über die Verknüpfung von Big Data mit OCEAN-Profilen das Targeting auf einen neuen Level der Manipulation. Aber es bleibt Microtargeting, wie es das Obama-Team auch schon 2012 eingesetzt hatte. Dass Haushalte und eventuell sogar Einzelpersonen so gezielt analysiert werden, wäre nach deutschen Recht nicht denkbar.

Eine düstere Prognose, nach der hierzulande ähnliche Methoden zur Wählermanipulation eingesetzt werden könnten, dürfte sich ohnehin nicht bewahrheiten:

Deutsche Wahlkampfstrategen verfügen über wesentlich weniger Daten als die US-Kollegen.

Verdeckte Datenquellen

Psychometriker Kosinski hat sich derweil neuen Datenquellen für seine Analysen zugewendet. Insbesondere das Smartphone hat es ihm angetan. Die Geräte seien „gewaltige psychologische Fragebögen, die wir konstant bewusst und unbewusst ausfüllen".

Tatsächlich haben Studien gezeigt, dass die Art und Weise, wie eine Person ein Smartphone nutzt, Aussagen über die Persönlichkeit zulässt. Wie häufig werden Mail- Kalender-, Office-, Chat- oder Spiele-Apps benutzt? Wie viele Anrufe gehen in einem bestimmten Zeitraum ein und wie viele gehen ab? Wie lange dauern die Anrufe? Wie viele SMS und Chat-Nachrichten werden empfangen und versendet und wie viele verschiedene Kontakte sind beteiligt? Welche Durchschnittslänge haben die verwendeten Wörter?

Versuche verschiedener Forschergruppen zeigen immer wieder, dass mittlerweile Algorithmen mithilfe von beiläufig erzeugten Daten wie Facebook-Likes oder der Smartphone-Nutzung die Persönlichkeitsmerkmale einer Person besser oder zumindest ebenso gut einschätzen wie Freunde oder Menschen aus dem erweiterten Familienkreis. Nur enge Familienmitglieder übertreffen die Algorithmen statistisch gesehen noch.

Selbst die Gemütsverfassung einer Person erkennen Algorithmen mittlerweile sehr zuverlässig. Seit längerer Zeit funktioniert das schon sehr gut bei der Auswertung der Stimme oder dem Gesichtsausdruck. Recht neu ist der Ansatz, Tastaturanschläge zu analysieren.

Dabei wurde die Zeitdauer zwischen dem Drücken und Loslassen einzelner Tasten gemessen und außerdem aufgezeichnet, wie oft die Entfernen-Taste und Tasten für Satz- und Sonderzeichen genutzt wurden. Die Ergebnisse waren von beeindruckender Genauigkeit. Der Algorithmus hat Personen, die ihren Gemütszustand selbst als traurig beschrieben, mit 88-prozentiger Wahrscheinlichkeit korrekt erkannt. Auch Personen, die sich müde, nervös, zuversichtlich oder unsicher fühlten, wurden mit über 80-prozentiger Wahrscheinlichkeit identifiziert.

Ähnliche Ergebnisse gibt es auch bei der Auswertung von Mausbewegungen. Dass solche Auswertungen technisch möglich sind, heißt natürlich nicht, dass sie bereits flächendeckend durchgeführt werden. Aber sie zeigen, welch enormes Potenzial in den Daten steckt. Und wie gefährlich diese einem werden, wenn sie in die Hände der Falschen geraten.

Gaming-Analysen

Auch mit Spielen lassen sich die Charaktereigenschaften einer Person sehr gut bestimmen. Gerade in Hinblick auf die Eignung für einen bestimmten Beruf stellen sie eine Art modernes Assessment-Center dar. Begonnen hat die Entwicklung vor einigen Jahren mit Spielen wie dem Wasabi Waiter, einem von Neurowissenschaftlern, Psychologen und Datenanalysten erstellten Programm, das es erlaubt, gute und vielversprechende Job-Bewerber von den weniger guten zu trennen.

Schauplatz des Spiels ist eine Sushi-Bar. Der Kandidat muss im Spiel dafür sorgen, dass leere Teller abgeräumt, neue Gäste begrüßt und die bestellten Mahlzeiten schnell zubereitet werden. Dabei muss er ständig unter Zeitdruck Entscheidungen treffen und Aufgaben priorisieren. Jede Entscheidung, jeder Klick, jede Mausbewegung wird aufgezeichnet und ausgewertet. Am Ende berechnen Algorithmen die Performance und die Eignung für ein bestimmtes Stellenprofil.

Mittlerweile haben auch einige traditionelle Computerspielehersteller diesen Trend entdeckt, wie wir hinter vorgehaltener Hand erfahren haben. Sie untersuchen Möglichkeiten, Daten, die beim Spielen anfallen, auszuwerten und zu kommerzialisieren. Eigentlich eine logische Konsequenz, denn egal ob man auf dem Smartphone daddelt oder am Computer spielt, der Hersteller kann in der Regel die erzeugten Daten leicht einer Person zuordnen.

Je nach Spiel muss man Geschicklichkeit beweisen, unter Stress Entscheidungen treffen, sich längere Zeit konzentrieren, mit Ressourcen haushalten, den Überblick behalten, strategisch denken – alles Eigenschaften und Verhaltensweisen, die auch im Berufsleben wichtig sind. Wer bei Spielen gut abschneidet, der wird auch im Job in dieser Hinsicht eine gute Figur machen, so die Annahme. Bislang ist allerdings noch kein Unternehmen bekannt, das Ingame-Daten in größerem Stil veräußert.

Intransparentes Tracking

Mit besonders großem Unbehagen verfolgen Datenschützer die Entwicklung bei Fitness-Apps und Wearables (siehe auch S. 86). Wie umfangreich der Handel

Mit dem Spiel Wasabi Waiter vermessen US-amerikanische Personaler die Fähigkeit von Bewerbern, mit Stress und Entscheidungsdruck umzugehen.

Die Oral-B Genius pumpt Zahnputzverhalten und Bildanalysen in die Cloud.

Procter & Gamble ermöglicht App-Entwicklern, via Bluetooth oder Rest-API auf Zahnputzdaten zuzugreifen – mit Erlaubnis des Nutzers.

mit den daraus gewonnenen Gesundheitsdaten ist, zeigte eine Untersuchung bereits im Jahr 2014: Die untersuchten Gesundheits-, Wellness- und Fitness-Apps haben Daten der Anwender an insgesamt 70 Drittfirmen weitergegeben. Darunter oft auch den Namen oder die E-Mail-Adresse.

Im Mai 2016 hat die norwegische Verbraucherschutzbehörde ein Verfahren gegen die Betreiber der beliebten Fitness-App Runkeeper angestrengt, weil sie illegal persönliche Daten der Anwender an eine Drittfirma weiterleitete. Auch verschiedene Hersteller von Fitnessarmbändern (Fitbit, Jawbone, Garmin und Mio), stehen in der Kritik, mehr Daten zu sammeln, als für den bestimmungsgemäßen Gebrauch des Geräts nötig wäre.

Im Dezember 2016 schlug hierzulande sogar die Bundesdatenschutzbeauftragte Andrea Voßhoff Alarm: „Viele Anbieter missachten oft gesetzliche Anforderungen", konstatierte sie als Ergebnis einer Stichprobe ihrer Behörde sowie nach Untersuchungen von Landesdatenschutzbehörden. „Nutzerinnen und Nutzer werden nicht oder nur mangelhaft darüber informiert, welche ihrer sensiblen Gesundheitsdaten von wem und zu welchem Zweck gespeichert werden. Gesammelte Daten können oftmals nicht gelöscht werden." Und: „Oft werden die durch die Geräte erhobenen Gesundheitsdaten durch externe Dritte verarbeitet. Durch die unklaren Regelungen zur Datenverarbeitung entgleiten diese Daten dabei der Kontrolle durch die Nutzer."

Zahnputz-Cloud

Dennoch nimmt die vom Nutzer tolerierte Datensammelei immer absurdere Formen an: Elektrische Zahnbürsten analysieren mithilfe eingebauter Sensoren das Zahnputzverhalten und senden die Daten per Bluetooth ans Handy, das eine Bewertung des Putzerfolgs präsentiert. Vorreiter Procter & Gamble (P&G) bringt mit seiner neuen Oral-B-Genius-Serie sogar die Kamera in den intimen Hygienebereich: Per Handy-Kamera überwacht sie, ob auch alle Zähne gleich lang geputzt werden [6].

Die detaillierten und personalisierten Angaben zum alltäglichen Zahnputzverhalten landen in der Cloud. P&G bietet App-Herstellern ein SDK für Apps an, die darauf zugreifen können – mit Zustimmung des Nutzers. Sogar ein Web-API gibt es, wie der Konzern stolz berichtet. Darüber lassen sich in Echtzeit Daten wie Zahnputzhäufigkeit, Dauer und Andruckstärke ziehen.

In den USA und einigen anderen Ländern gibt es bereits Versicherungskonzerne, die Tarife für vergünstigte Zahnzusatzversicherungen anbieten – speziell für Kunden, die einwilligen, ihr Zahnputzverhalten mit einer smarten Zahnbürste überwachen zu lassen.

In Deutschland haben erste Versicherungen damit begonnen, Fitnesstracker oder Smartwatches zu subventionieren. Schon seit über einem Jahr bietet etwa die Techniker Krankenkasse bis zu 250 Euro Zuschuss beim Kauf einer Apple Watch an, wenn der Versicherte bereit ist, an sieben Maßnahmen teilzunehmen, beispielsweise einer Vorsorgeuntersuchung und zwei Gesundheitskursen. Noch betonen die Versicherungen allerdings, nicht auf die erhobenen Daten der Smartwatch zugreifen zu wollen.

In Anbetracht des immer größer werdenden Kostendrucks im Gesundheitssektor ist es allerdings durchaus denkbar, dass sich dieses Geschäftsmodell in den nächsten Jahren ändert. Darauf deutet auch hin, dass manche Versicherungskonzerne schon einen Schritt weiter sind. Die Generali-Gruppe beispielsweise startete Mitte 2016 ein Programm, um Daten zur Bewegung und zum Lebensstil der Versicherten

zu erfassen. Zukünftig sollen dann etwa Supermärkte an die Versicherung melden dürfen, welche Lebensmittel ein Versicherter eingekauft hat. Noch sind solche Tarife freiwillig und jeder entscheidet selbst, inwieweit er sich gläsern macht.

Die Gefahr steigt allerdings, dass Versicherungsnehmer irgendwann nicht mehr die Wahl haben, weil die intransparente, flächendeckende Auswertung sämtlicher verfügbaren Daten zum Standard geworden ist. Dies könnte dazu führen, dass Menschen, die zu hohe Kosten generieren, systematisch von Versicherungsleistungen ausgeschlossen werden und so das Solidaritätsprinzip heutiger Krankenversicherungssysteme nicht mehr eingehalten wird: Wer sich dagegen entscheidet, seine Daten preiszugeben, der benachteiligt sich selbst, weil auch er günstige Tarife oder bestimmte Leistungen nicht erhält.

Sortieren Arbeitgeber zukünftig Bewerber mit überdurchschnittlich hohen Krankheitswahrscheinlichkeiten aus? Aus heutiger Sicht erscheint dieses Szenario fast dystopisch. Aber wenn man die derzeitige Entwicklung konsequent weiterdenkt, sind solche Szenarien längst nicht mehr auszuschließen.

Viel zu oft herrscht eine fast unendliche Technologiegläubigkeit vor: Was die Daten sagen, muss wahr sein, obwohl gerade Verhaltensprognosen auf Korrelation, Statistik und Wahrscheinlichkeit beruht. Die Big-Data-Algorithmen können immer nur so gut sein, wie die Menschen, die sie erschaffen haben.

Fehlertoleranz?

Und genauso, wie Software fast nie ohne Fehler ist, sind auch die Ergebnisse der Big-Data-Analysen selten komplett korrekt. Das fängt schon bei fehlerhafter Datenerhebung an. Allein in den USA entstehen durch inkonsistente, redundante und absichtlich verfälschte Daten laut Technologieberatungsfirma Artemis Ventures über drei Milliarden US-Dollar Schaden jährlich.

Auch die oft fehlende Aktualität der Daten wird immer mehr zum Problem: Wie stellt man sicher, dass eine nicht mehr zutreffende Information, beispielsweise ein negativer Eintrag über die Kreditwürdigkeit einer Person, wirklich aus allen Datenbanken verschwindet? Den Weg der Daten kann kaum jemand kontrollieren. Gespeichert bleibt, was gespeichert werden kann.

In den meisten Fällen bekommt man von fehlerhaften Daten ohnehin erst mal gar nichts mit, weil sie keine direkt spürbaren Auswirkungen haben. Erst wenn man vielleicht später bei einer Bewerbung auf einen neuen Job die Stelle nicht bekommt, eine neue Versicherung abschließt und höhere Raten zahlen muss oder von bestimmten Angeboten ausgeschlossen wird, kommt man ins Grübeln.

Egal ob in der Logistik, in der Medizin, Verkehrsplanung oder in der Klimaforschung, es gibt viele Beispiele, wo Big-Data-Auswertungen sinnvoll genutzt werden. Problematisch wird es meistens dann, wenn Daten über Menschen im Spiel sind, weil dann ungenaue Analysen, intransparente Algorithmen oder fehlerhafte Daten durchaus dramatische Auswirkungen auf das Leben eines Einzelnen haben können.

Welcher Mensch oder welcher Algorithmus die Analysen deutet, bleibt in aller Regel im Dunkeln. Was bedeutet es denn, wenn die Analyse für einen gesunden Menschen vorhersagt, dass er mit einer 60-prozentigen Wahrscheinlichkeit in den nächsten Jahren Bluthochdruck, Diabetes oder eine andere chronische Krankheit bekommen wird und damit höhere Kosten für die Krankenversicherung erzeugt? Was geschieht, wenn die Analyse ergibt, dass die Eignung eines Bewerbers für den neuen Job nur zu 55 Prozent gegeben ist, obwohl der Personalchef denkt, er habe einen perfekt geeigneten Kandidaten gefunden? Verlässt er sich auf seine Intuition oder glaubt er den vermeintlich objektiven Zahlen?

Dieses Verhältnis hat längst begonnen sich zu verschieben: Die Algorithmen werden in Zukunft in vielen Bereichen unser Leben bestimmen. Genauso, wie man sich vor 15 Jahren kaum vorstellen konnte, was sich heute alles mit einem Smartphone anstellen lässt, so unrealistisch mag es aus heutiger Sicht sein, was Unternehmen in 15 Jahren mit den Daten anstellen, die schon heute von uns gesammelt werden. *(hob@ct.de)* c't

Markus Morgenroth ist Informatiker und Autor des Buchs „Sie kennen dich! – Die wahre Macht der Datensammler“ (siehe c't 22/14, S. 192).

Im Programm „Generali Vitality“ erfasst der Versicherer Daten zu Fitness und Lebensstil seiner Kunden.

Literatur

[1] Frank Pasquale, The Black Box Society: The Secret Algorithms That Control Money and Information, Harvard University Press, 2015
[2] http://shop.schober.com
[3] Herbert Braun, Sammelleidenschaft, Wie und wo Facebook seine Daten zusammenträgt, c't 24/16, S. 76
[4] Michal Kosinski, David Stillwell, Thore Graepel, Private traits and attributes are predictable from digital records of human behavior, PNAS, vol. 110 no. 15, 2013
[5] https://www.dasmagazin.ch/2016/12/03/ich-habe-nur-gezeigt-dass-es-die-bombe-gibt/
[6] Positionsgesteuert putzen, c't 18/16, S. 54