

Das Muster im Datenchaos

8. April 2013

Angela Barandun, Tages Anzeiger

Dank der richtigen Formel weiss ein Supermarkt, welche Kundinnen schwanger sind oder wie viel Grillfleisch er am nächsten Wochenende verkaufen wird. Solche Datenanalysen sind praktisch und gefährlich.

In der Welt des Detailhandels ist es so etwas wie ein Sechser im Lotto: eine Liste mit Zehntausenden von Namen von Frauen, die ziemlich sicher schwanger sind - inklusive ihres ungefähren Geburtstermins. Nie sind Menschen so leicht beeinflussbar, wie wenn sie zum ersten Mal Eltern werden. Das macht sie zum idealen Ziel für Werbung.

Zusammengestellt hat die Liste der Statistiker Andrew Pole im Auftrag der US-Supermarktkette Target. Pole hat das Einkaufsverhalten Tausender Kundinnen analysiert und ein Muster gefunden, das werdende Mütter identifiziert. Basis ist laut der «New York Times» eine Kombination aus 25 Produkten, die auch Anhaltspunkte über den ungefähren Geburtstermin liefern. Greift eine Target-Shopperin zu Vitaminpräparaten und wechselt auf parfümfreie Körperlotion, ist ihr Bauch noch nicht zu sehen. Deckt sie sich hingegen mit grösseren Mengen Wattebäuschen und Desinfektionsmittel ein, steht sie kurz vor der Geburt.

Unvorstellbare Menge an Daten

Möglich sind solche Berechnungen aufgrund einer Entwicklung namens «Big Data». Der Begriff bezeichnet einerseits den dramatischen Anstieg an Daten - produziert von smarten Handys, Maschinen, Stromnetzen, Autos oder Gebäuden; von Sensoren, Videokameras oder Kredit- und Kundenkarten. 2012 waren es laut der Marktforschungsfirma IDC rund 2,8 Zettabyte - eine unvorstellbare Menge. 2020 sollen es vierzehnmal mehr sein. Gleichzeitig sind die Kosten für Speicherplatz dramatisch gesunken - von 300'000 Dollar für ein Gigabyte 1980 auf rund 10 Cent.

Andererseits kann man Daten heute mit Mitteln analysieren, die bisher nicht zur Verfügung standen. Es geht einfacher, billiger - und vor allem schneller. Erst dadurch wird es möglich, im Datenchaos nach Mustern zu suchen.

Dr. Watson, Ihre Diagnose?

Das Paradebeispiel dafür ist Watson - der Supercomputer von IBM, der 2011 die US-Quizshow «Jeopardy» gewann. Für einen normalen Computer wäre das unmöglich gewesen, weil man dafür Sprache verstehen muss. Watson ruft aber nicht nur Begriffe aus Datenbanken ab, er versteht die Bedeutung eines Satzes, stellt Vermutungen an und überprüft sie. Diese Fähigkeiten sollen unter anderem das Gesundheitswesen revolutionieren.

Derzeit laufen zwei Pilotprojekte: Erstens durchforstet Watson Datenbanken mit bekannten Wirkstoffen nach neuen Anwendungsmöglichkeiten. Im Testlauf stiess er auf über ein Dutzend

Substanzen, deren Wirkung auf Malaria nun getestet wird. Zweitens soll Watson Ärzten bei der Diagnose ihrer Patienten helfen. Die Erwartungen sind gross, wann sich damit Geld verdienen lässt, ist offen.

Ankunftszeit am Flughafen

Dabei ist der Treiber der meisten Projekte ganz klar finanzieller Natur - wie ein Beispiel der «Harvard Business Review» zeigt. Bei einer US-Airline schätzten die Piloten ihre Ankunftszeit jeweils kurz vor dem Landeanflug, damit die Fluggesellschaft die Bodencrew koordinieren konnte. Die Trefferquote war jedoch miserabel: Bei jedem zehnten Flug lagen die Piloten über 10 Minuten daneben. Bei jedem dritten waren es über 5 Minuten. Die Wartezeiten beim Bodenpersonal kosteten jedes Jahr Millionen.

Mittlerweile werden die Ankunftszeiten automatisch berechnet - von einem System, das öffentliche Daten über das Wetter oder Flugpläne mit Daten von Radarstationen nahe den Flughäfen kombiniert. Es erfasst die aktuelle Lage und berücksichtigt, wie lange es bei früheren Landeanflügen unter gleichen Bedingungen gedauert hat, bis das Flugzeug am Boden war. Seither tendieren die Abweichungen gegen null.

Es ist - Zufall!

Der britische Detailhändler Tesco setzt auf ähnliche Methoden. Er hat sein Bestellsystem an einen Wetterdienst gekoppelt, etwa um vorherzusagen, wie viel Grillfleisch er brauchen wird. Das System ist extrem komplex: Der Effekt, den das Wetter auf das Einkaufsverhalten hat, ist nicht nur von der Temperatur abhängig, sondern auch von Wochentag, Monat, Wetter der letzten Tage oder der Lage der Filiale. Ein sonniger Montag im April mit Temperaturen um die 17 Grad wird in der Stadt für einen höheren Sandwichabsatz sorgen, weil die Leute mittags draussen essen. Am Wochenende spürt der gleiche Laden davon kaum etwas. Analog wird der Absatz von Grillfleisch am ersten warmen Wochenende der Saison stärker ansteigen als am dritten oder vierten, weil der erste Ansturm bereits vorbei ist.

Banken setzen vor allem bei der Betrugsbekämpfung auf Datenanalysen. Kreditkartentransaktionen werden längst nicht mehr nur anhand logischer Kriterien geprüft - etwa, ob die Karte gleichzeitig auf zwei verschiedenen Kontinenten eingesetzt wird. Die Banken verfügen über Systeme, die registrieren, wenn ein Kunde, der kein Auto besitzt, plötzlich an der Tankstelle für Benzin bezahlt. Mithilfe von historischen Daten kann man sogar Soziogramme erstellen. Kauft sich eine verheiratete Frau plötzlich teuren Schmuck, steht sehr wahrscheinlich eine Trennung bevor.

Die Frage ist entscheidend

Nur: Was nützt dieses Wissen der Bank? Das ist eine der wichtigsten Fragen bei der Suche nach Mustern. Nur wenn man weiss, wonach man sucht und was das Resultat - auch finanziell - bringt, kann die Rechnung aufgehen. Mindestens genauso wichtig ist die Frage, ob das Ergebnis Sinn macht. Eine Studie etwa ging der Frage nach, was besonders langlebige Autos auszeichnet. Das Resultat: die Farbe. Gelbe Autos wurden statistisch mit Abstand am ältesten. Wieso? Keine Ahnung. Das Ergebnis macht keinen Sinn. Es ist - Zufall!

Kritiker sagen, dass wir heute ein beinahe religiöses Vertrauen in wissenschaftliche Methoden wie die Statistik setzen - und dabei den Kopf ausschalten. Wie gefährlich das ist, zeigt das Beispiel einer US-Studie. Sie kam zum Schluss, dass Fettleibigkeit ansteckend ist. Der Grund: Die Wahrscheinlichkeit, fettleibig zu werden, steigt, sobald es im sozialen Umfeld viele

fettleibige Menschen gibt. Dabei ist Übergewicht als Phänomen in den USA auf dem Vormarsch - unabhängig vom Umfeld.

Statistik kennt keine Moral

Nicht weniger beunruhigend sind die Prognosen von Statistik-Wunderkind Nate Silver zu den letzten US-Präsidentschaftswahlen in der «New York Times». Seine Analysen fanden international riesiges Echo. Nach den Wahlen zeigte sich: Er hat die Resultate beinahe punktgenau berechnet. Nur: «Hat Silver die Zukunft vorhergesagt - oder sie mit seinen Prognosen beeinflusst?», fragte etwa die Amerikanerin Valerie Casey kürzlich am GDI-Trendtag zum Thema Big Data.

Sogar Nate Silver ist kritisch. Eine seiner Kernaussagen lautet: «Big data needs big judgment.» Was so viel heisst wie: Ohne gesunden Menschenverstand sind statistische Ergebnisse nichts wert. Sie stützen sich auf die Vergangenheit - ihre Prognosefähigkeit muss daher regelmässig versagen. Exemplarisch dafür steht die Hilflosigkeit ökonomischer Modelle nach der Finanzkrise. Hinzu kommt ein weiterer Punkt: Die statistische Analyse kennt keine Moral. Das mussten auch Andrew Pole und Target erkennen. Kurz nachdem die Werber angefangen hatten, Schwangeren Werbung für Babykleider, Wiegen und Nuggi zu schicken, kam es zu einem Zwischenfall. Ein erzürnter Vater warf einem Filialleiter vor, seine schulpflichtige Tochter mit derartiger Werbung auf dumme Gedanken bringen zu wollen. Als sich der Filialleiter telefonisch erneut entschuldigen wollte, gestand ihm der Vater, er habe mit seiner Tochter gesprochen - sie sei tatsächlich schwanger.

Es sollte nicht der einzige Zwischenfall bleiben. Die werdenden Mütter reagierten äusserst sensibel darauf, dass der Supermarkt so etwas Privates wusste. Die Kampagne wurde aber nicht etwa gestoppt. Der Babywerbung wurde einfach so viel anderes beigemischt, dass die Schwangeren keinen Verdacht mehr schöpften - und die Gutscheine brav beim nächsten Einkauf einlösten.

Big Data in der Schweiz

2 Millionen Anrufe, 15 Millionen Verbindungen zu Handyantennen, jeden Tag, allein in Genf. Was das bedeutet – wohin die Massen strömen und wann die Stadt schläft –, zeigt das Kunstprojekt «Ville Vivant» anhand von Daten der Swisscom. Jede Linie steht dabei für einen Handynutzer, der sich von A nach B bewegt. Die Swisscom stellt solche Daten auch kommerziellen Anbietern zur Verfügung. Etwa dem Navi-Spezialisten Tomtom, der damit in Echtzeit Informationen zu Staus gewinnt – oder sie vorhersagen kann. Auch der Bund interessiert sich für diese Daten. Die Swisscom evaluiere Möglichkeiten, was man mit den Daten sonst noch machen kann, so Sprecher Josef Huber. Konkurrentin Telefonica etwa hat bereits eine Abteilung gegründet, die solche Daten vermarktet. Die Swisscom betont, sie verkaufe keine Personendaten an Dritte – was nur mit Einwilligung der Kunden möglich wäre –, sondern verwende anonymisierte Standortdaten.

Auch Postfinance wertet Kundendaten statistisch aus. Das sogenannte E-Cockpit ist Teil des E-Banking-Portals und schlüsselt für jeden Kunden Einkäufe mit Postcard in verschiedene Kategorien wie Essen, Wohnen oder Kleidung auf. Der Dienst ist freiwillig, wer mitmacht, akzeptiert die Auswertung seiner Daten. Negative Reaktionen gab es laut Sprecher Alex Josty bislang keine. Postfinance mache sich derzeit grundsätzliche Gedanken über die kommerzielle Nutzung von Kundendaten – natürlich unter Einhaltung des Datenschutzgesetzes. Migros und

Coop erfassen mit ihren Kundenkarten seit Jahren Einkäufe, bei Coop werden die Daten erst seit letztem Herbst ausgewertet. Migros setzt sie seit langem gezielt im Marketing ein. Kauft jemand häufig Tierfutter, wird er auf neue Produkte hingewiesen und erhält Rabatt-Bons. Auch ob man die Bons nutzt, wird erfasst. «Wir können auch feststellen, in welcher Region ein Kunde häufig einkauft», sagt Sprecherin Monika Weibel. Wetterdaten fliessen nicht ins Bestellsystem ein. «Allerdings ist das prüfenswert», so Weibel. Eine Liste aller schwangeren Kundinnen wäre für Migros und Coop zwar interessant. Allerdings sprechen laut Coop «moralisch-ethische Grundsätze» gegen eine solche Auswertung. Stattdessen setzen beide Detailhändler auf Clubs, bei denen die werdenden Eltern freiwillig Mitglied werden, um dadurch von Spezialaktionen zu profitieren.