



Der Fluch der Datensammler

Description

Unsere digitale Welt verleitet uns dazu, möglichst viele Daten zu speichern. Mehr Daten – so sagt man – geben mehr Sicherheit, mehr Informationen und eine bessere Möglichkeit, die Zukunft vorauszusagen. In hermetischen Systemen trifft dies zu, aber auch nicht immer.

Als die Steine Daten hatten



Die Geburt der Vorhersagen

Datenverarbeitung – im ureigenen Sinn – wurde von Menschen schon sehr lange vor der Erfindung von Computern betrieben. Sie war und ist ein wichtiger Teil der Überlebensstrategie und versucht zwei zentrale Fragen zu beantworten.

1. Welche Fehler wurden in der Vergangenheit gemacht?
2. Wie wird die Zukunft.

Als die Menschen noch gemeinsam auf die Jagd gingen, verlief diese sicher nicht immer erfolgreich. Dank der eigenen Sinne konnten, zumindest die Überlebenden, die Daten erfassen und mit dem Verstand verarbeiten, um neue Taktiken und Strategien zu generieren. Spätestens zu dieser Zeit

erfand der Mensch die ersten Mechanismen eines QS-Systems. Die Daten zeigten den Fehler, es wurden Maßnahmen definiert und getestet, bis das Ergebnis für „i. O.“ befunden wurde. Mangels Schrift blieb unseren Vorfahren das Ausfüllen von Formularen erspart.

Die Vorhersage der Zukunft war für Jäger und Sammler ebenfalls wichtig. Sie mussten wissen, zu welcher Jahreszeit die Tiere sich in der Umgebung aufhielten. Durch den Ackerbau erhielt dieser Aspekt eine immer größere Bedeutung. Die ersten Kalender erblickten das Licht der Welt und kurz darauf entdeckten ein paar Schamanen das Geschäftsmodell der Vorhersage. Mit möglichst komplexen, optisch und akustisch beeindruckenden Zeremonien verkündeten sie die Zukunft. Sie sagten gute und schlechte Jahre voraus. Das Konzept hat sich – mit zeitgemäßen Anpassungen – bis heute gehalten. Statt Schamanen spricht man heute von Wirtschaftsweisen und Experten, das Resultat ist aber nahezu identisch.

Der Traum der Prognose



Sagt die Börsendaten
voraus: Finanzexperte

Trotz Unmengen von Daten sind Prognosen schwierig – vor allem wenn sie die Zukunft betreffen. Und da wir aus den letzten fünfzigtausend Jahren nicht viel gelernt haben, sammeln wir noch mehr Daten um mit einer höheren Wahrscheinlichkeit bzw. Genauigkeit daneben zu liegen.

Der Traum geht so: Je mehr Daten wir erhalten, umso exakter können wir Tendenzen feststellen und mit entsprechenden Formeln und dank superschnellen Computern vorhersagen, was in den kommenden Tagen, Wochen oder gar Jahren passiert. In einfachen Systemen funktioniert das wunderbar. Es liegen fünf Äpfel im Korb. Alle zehn Minuten holt sich jemand einen Apfel daraus und nach einer Stunde greift jemand in einen leeren Korb. Das ist die Theorie.

Die Praxis geht so: Der Zweite greift in den Korb, stellt fest, dass der Apfel faulig ist und wirft alle Äpfel weg. Man kann bei dem Beispiel einwenden, dass einfach nicht genug Daten vorlagen und man bei einer längeren Beobachtung mehr Daten hätte und somit die Wahrscheinlichkeit für einen faulen Apfel

berechnen könnte. Man könnte auch Maßnahmen ergreifen um den faulen Apfel zu verhindern usw. Wir wissen aber aus dem Leben, dass es genug Systeme gibt, bei denen wir solche Daten haben und eine Vorhersage, sagen wir Mal, eher gewagt ist. Wenn Sie bspw. auf den Zug warten und fünf Minuten vor Abfahrt erfahren, dass er sich um eine Viertelstunde verspätet. Oder ausfällt.

Man sollte meinen, dass nach rund 180 Jahren Eisenbahn in Deutschland genügend Daten vorliegen würden, um Züge pünktlich von A nach B fahren zu lassen. In Japan stellt das trotz Erdbeben, Unwetter, Vulkanausbrüchen und anderen Schwierigkeiten weniger Probleme dar. Es hat den Anschein, als wären dort Bahn und Passagiere trotz genannter Unwägbarkeiten kalkulierbarer als bei uns.

Das Problem der Prognose



Eine Tür wird schon die richtige sein. Ene mene muh...

Eines der Kernprobleme liegt darin, dass selbst einfach scheinende Systeme sich bei genauer Betrachtung als sehr Komplex herausstellen. In solchen Fällen spielt die Fülle der Daten eine untergeordnete Rolle. Wenn sie nicht alle Einflussfaktoren erfassen (was schon quantenmechanisch nicht möglich ist), dann können sie nur unter fest definierten Bedingungen und einer bestimmten Zeitspanne eine Prognose wagen.

Doch obwohl wir das wissen, sammeln wir möglichst viele Informationen. Ganz nach dem Witz: „Warum leckt sich der Hund die Eier? Weil er es kann!“ Dabei macht man uns vor, dass es für die Sammelwut gute Gründe gäbe. Schauen wir uns die Klassiker an:

1. Mehr Daten bedeutet mehr Sicherheit. So kann Terror effektiver bekämpft werden oder...
2. ...wenn das nicht gelingt, können die Verbrechen schneller und besser aufgeklärt werden. NSU lässt grüßen.
3. Durch die gesammelten Daten können Produkte / Benutzererfahrungen etc. verbessert werden.

Dabei wissen wir alle, dass es weder um Sicherheit, noch Aufklärung oder der Verbesserung von Produkten geht. Es geht ausschließlich um Kontrolle, Überwachung, Einschüchterung und letztlich um Geld. Und um noch den letzten Tropfen aus der Zitrone zu pressen, wird mit viel Aufwand alles an Daten abgegriffen, was man über das Internet erhalten kann. Da wundert es einen auch nicht mehr, wenn große Konzerne die Passwörter ihrer Kunden nicht einmal verschlüsseln sondern unverschlüsselt von A nach B senden und speichern.

Data2Money converter



Daten sammeln bedeutet Überwachung.

Ein kleines Beispiel, was mir kürzlich widerfuhr. Ich sprach mit einem Kollegen darüber, dass ich mir einen Raspberry Pi gekauft hatte. Er wusste zwar, was es ist, interessierte sich aber bisher nicht für das Thema. Wenige Minuten später bekam er auf seinem Computer genau zu dem Thema Werbung angezeigt. Bei jeder Gelegenheit. Hat er danach gesucht oder einschlägige Seiten besucht? Nein. Sein Handy hörte mit. Das reichte schon, um anschließend Werbung einzublenden.

Mein Kollege und ich verfügen über ausreichend technische Kenntnisse um solche Zusammenhänge zu erkennen. Wir sind sensibel genug und verstehen hinreichend die Mechanismen im Hintergrund. Obwohl wir mit Technik aufwachsen und von moderner Technologie umgeben sind, verstehen aber besonders in Deutschland immer weniger Menschen, wie diese Technologie funktioniert und welche Möglichkeiten sie bietet. Das beginnt schon bei einer Frau mittleren Alters, die sicher 20 Jahre im Internet unterwegs ist und man ihr dennoch erklären muss, dass man mit F5 die Seite im Browser erneut aufrufen bzw. aktualisieren kann und endet irgendwo bei pubertären Pickelgesichtern, die zwar ihren Smartphone, aber keinen PC bedienen können. Von Textverarbeitung, Tabellenkalkulation, ordentliche Mails schreiben oder dem Innenleben eines Computers ganz zu schweigen.

Auch Spiele sammeln Daten

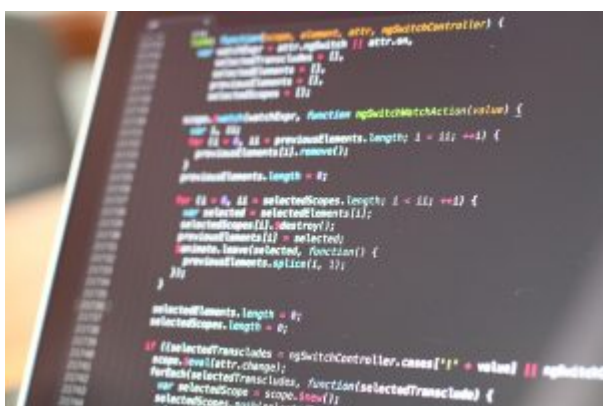


Geschafft! Alle Daten wurden erfolgreich an den Entwickler übertragen.

Die Sammelwut ist den meisten Menschen unklar oder wird als gegeben hingenommen. Selbst in Computerspielen weiß der Entwickler mittlerweile, wie oft in Zeitspanne X ein Icon angeklickt wurde. Wozu diese Information gut sein soll, weiß er wahrscheinlich selbst nicht. Aber da Daten sammeln so einfach ist und Internetverbindungen immer schneller werden, wird es gemacht. In Alpha- oder Betaphasen eines Spiels kann man dies noch logisch nachvollziehen, auch wenn es manchmal schwer fällt, wenn man die Resultate sieht. Am fertigen Produkt ist vieles, was Tracking angeht, höchst fragwürdig. Dabei wäre es für Spieler und Entwickler sinnvoll, vorher zu eruieren, welche Daten relevant sind und getrackt werden. Wird dies dem Spieler transparent erklärt, könnte man vielleicht sogar Vertrauen gewinnen.

Doch man muss nicht einmal ein Spiel starten, um Daten Preis zu geben. In den meisten Fällen reicht schon das Betriebssystem aus und wer so richtig Freude daran hat, Trackingopfer zu werden, startet den Browser seiner Wahl. Wenn nicht schon der Browser oder ein zugehöriges PlugIn die Daten erfasst und um den Globus sendet, dann spätestens die erste Webseite, die man besucht. Google Analytics und Co lauern um alles zu erfahren, was man halbwegs legal erfahren darf. Mittlerweile strafen einen Analysetools für Webentwickler dafür ab, wenn man Google Analytics nicht verwendet!

Tracking der Webentwickler



Per Tracking lassen sich viele nutzlose Daten vom Seitenbesucher erfassen.

Es gibt tatsächlich Informationen, die für einen Webentwickler interessant sind. Seitenbesucher, Seitenzugriffe, Absprungrate, nationale Herkunft, ggf. Suchbegriffe, durch die man gefunden wurde. Browser und ihre Versionen können ebenso hilfreich sein wie das Wissen um die Betriebssysteme. Damit kann man zumindest ein bisschen was anfangen, aber die Tracking-Tools erfassen viele weitere Daten, die >90% der Webentwickler kaum oder nur am Anfang der Entwicklung nutzen. Irgendwann haben sich die Optimierungen erschöpft, die Daten werden aber weiterhin sinn- und grenzenlos gesammelt. Selbst wenn der Betreiber der Webseite das Tracking einstellt: Der Provider erfasst munter weiter alle Daten, die er bekommen kann.

Über den Status quo könnte man noch lachen oder ggf. einfach nur den Kopf schütteln. Schließlich kann es einem unbescholtenen Bürger egal sein, ob er im Internet getrackt wird oder ob Amazon weiß, über welche käuflich zu erwerbenden Produkte man sich unterhielt. Werbung kann man (mehr oder weniger gut) filtern oder schlicht ignorieren. Doch der Erfahrung nach gilt hier der Satz: Alles was man sagt kann und wird gegen einen verwendet.

Märchenland



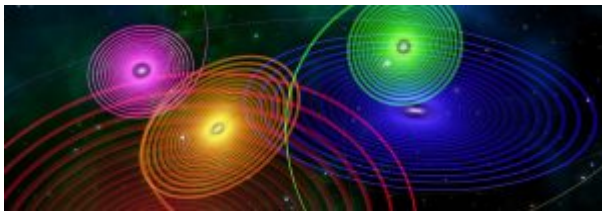
Die bis heute wissenschaftlich beste Methode, um Wetter und Klima vorauszusagen.

Aber zurück zum Märchen der Vorhersagbarkeit. Je mehr Daten man hat, umso besser ist die Prognose für die Zukunft. Innerhalb eines engen Rahmens oder eines Newtonschen Systems mag das stimmen, aber hier braucht man die Daten vor allem um allgemeine Gleichungen abzuleiten oder ein neuronales System zu trainieren, wobei Letzteres vorwiegend auf Versuch und Irrtum basiert. Ein Spiel wie Schach mag uns komplex erscheinen, für einen Computer ist es aber perfekt um zu berechnen, was bei einem entsprechenden Zug 20 oder gar 30 Züge später passiert. Hier treffen enger Rahmen und ein einfaches System gleichermaßen zusammen und die Prognosen sind korrekt. Außerhalb der Testumgebung sind aber viele andere Einflüsse möglich. Beispielsweise das ein Spieler einen ungünstigen Zug macht. Oder einen Schlaganfall erleidet.

Wie komplex die Realität werden kann, sieht man an den Wettervorhersagen. Wir wollen das Klima in 20 oder 30 Jahren vorhersagen können, wissen aber nicht einmal am Montag genau, wie das Wetter am Freitag wird. Dabei ist das Klima die Summe aller Wetterdaten über einen Zeitraum von Jahrzehnten.

Man spricht von Wahrscheinlichkeiten und lässt dabei bewusst eine Zahl x offen, die alle Parameter beinhalten, die man nicht vorhersagen kann oder an die kein Mensch denkt. So kann man auch eine Wahrscheinlichkeit dafür berechnen, dass 2022 der HSV Deutscher Fußballmeister wird. Solche Berechnungen sind aber ebenso sinnfrei wie die Berechnung des Ausgangs einer Darmspiegelung in zehn Jahren.

Die Wurzel allen Übels



Egal was man voraussagt: Irgendwo im Multiversum wird es schon stimmen.

Teilweise liegt das Kernproblem in der Mathematik selbst. Man kann im Prinzip alles berechnen, egal ob es das gibt oder nicht. Im Zweifelsfall setzt man ein paar Variablen ein, deren Größen man schätzt und schon kommt ein n -Dimensionaler Raum in einem Spiegeluniversum heraus, in dem alle auf einem Bein hüpfen müssen.

Obwohl das allgemein bekannt ist, wird der Unsinn natürlich weiter getrieben. Auf den Vermerk hin, dass das alles Quatsch sei, antworten Statistiker gerne damit, dass einfach nicht genug Daten vorliegen, um ein Muster zu erkennen. In Einzelfällen mag das stimmen, aber was bringt einem [eine Statistik](#), die besagt, dass 12% der Frauen zwischen 31 und 34 Brustkrebs bekommen, wenn sie Susanne heißen?

Statt blind Zahlen zu sammeln um darauf zu hoffen, etwas wirklich Sinnvolles aus ihnen abzuleiten, sollte die Sinnfrage zuerst gestellt werden. Noch bevor man überhaupt eine Datenbank anlegt.

Date Created

7. Oktober 2019

Author

sven