

## Big Data – Die Verfahren hinter dem Begriff

Der Sammelbegriff Big Data bezeichnet die verschiedenen Wege, wie die Datenmassen erhoben, verarbeitet und ausgewertet werden. Daher bezieht er sich auch auf eine Vielzahl technologischer Anwendungen. Die nachfolgenden Beispiele verdeutlichen, wie einerseits die Datenerhebung und andererseits die Datenauswertung erfolgen kann.

### Tracking

Beim Aufruf einer Webseite wird im verwendeten Browser ein sog. Cookie gespeichert. Es erfasst Nutzungsdaten, auch Trackingdaten genannt. Diese sind z. B. für die Wirtschaft interessant. Viele Unternehmen setzen solche Big-Data-Erhebungen ein, um ihre Produkte oder Dienstleistungen zu optimieren oder ihre Umsätze zu erhöhen.

Wenn ein Anbieter von Nachrichten durch Trackingdienste z. B. weiß, welche Beiträge nur teilweise gelesen oder welche Artikel häufig nacheinander gelesen werden, kann er Texte kürzen oder sich künftig auf bestimmte Themen konzentrieren.

Auch die zahlreichen Nutzerinnen und Nutzer sozialer Netzwerke wie z. B. Instagram erzeugen riesige Datenmengen, die durch die jeweilige Software gespeichert und analysiert werden. Mithilfe dieser Daten bieten soziale Netzwerke zielgruppenspezifische Werbung an, die sie an Unternehmen verkaufen. Die Werbekunden können wählen, wie alt die Personen sein sollen, die ihre Anzeige sehen, woher sie kommen, ob sie weiblich oder männlich sind oder welche Interessen sie haben. Die Betreiber sozialer Netzwerke verdienen Geld mit den erhobenen und verarbeiteten Daten. Zum Schutz vor solchen Anwendungen können Nutzerinnen und Nutzer Browsererweiterungen installieren, die das Tracking zumindest einschränken. Auch über die Browsereinstellungen selbst kann der Einsatz von Cookies begrenzt werden. Eine komplette Unterbindung führt jedoch meist dazu, dass Webseiten nicht mehr ohne Einschränkungen genutzt werden können.

Fingerprinting ist eine neuere Form des Trackings, die ohne Cookies auskommt. Hierbei ist es möglich, Personen über verschiedene Browser hinweg eindeutig zu identifizieren. Dabei werden z. B. Informationen zum Betriebssystem, zu verwendeten Schriftarten oder zur Bildschirmauflösung erfasst, um ein eindeutiges Profil erstellen zu können. Es existieren erste Browsererweiterungen, um auch das Tracking über Fingerprinting weitgehend zu unterbinden, wie z. B. der „CanvasBlocker“ für Firefox.

### Data-Mining

Mithilfe von Big-Data-Verfahren ist es möglich, Zusammenhänge und Muster zwischen einer Vielzahl von Daten zu ermitteln. Dadurch entstehen Modelle, mit denen man Vorhersagen tätigen kann. Wetterprognosen stellen beispielsweise eine solche Vorhersage dar. Deshalb kommen bei Big-Data-Auswertungen statistische Verfahren zum Einsatz.

Die Vorgehensweise, statistische Methoden systematisch auf umfassende Datenbestände anzuwenden, um Informationen aus ihnen herauszufiltern, wird auch als Data-Mining bezeichnet. Dabei kommen verschiedene Analysearten zum Einsatz, so z. B. die drei folgenden:

### Clustering

Bei dieser Methode sollen in großen Datenmassen Ähnlichkeiten zwischen einzelnen Fällen (z. B. Personen) ermittelt werden, um Gruppen zu bilden. Diese heißen dann Cluster. Zur Clusteranalyse können viele verschiedene Merkmale herangezogen werden. Beispielsweise entwickeln Parteien im Wahlkampf verschiedene Cluster von Wählertypen anhand von Alter, Geschlecht und Bil-

ungsgrad, um ein genaueres Bild von diesen zu erhalten. Auf Grundlage dieser ist es möglich, die Wählerinnen und Wähler gezielter anzusprechen.

### **Assoziationsanalyse**

Mithilfe dieses Verfahrens können Abhängigkeiten von Merkmalen und Regeln aufgestellt werden. Häufig geht es dabei um Entscheidungen: Wählt jemand A aus, wird er oder sie mit einer gewissen Wahrscheinlichkeit in einem nächsten Schritt auch B auswählen. Wenn man beispielsweise ein Buch in einen Online-Warenkorb legt und dann weitere Buchtitel vorgeschlagen bekommt, die einem ebenfalls gefallen könnten, hat eine Assoziationsanalyse stattgefunden. Vor einer solchen automatisierten Empfehlung wurden Einkäufe anderer Kundinnen und Kunden auf Abhängigkeiten analysiert und auf diesen Einkauf übertragen.

### **Regressionsanalyse**

Mit dieser Methode werden aus den vorhandenen Daten Zusammenhänge ermittelt. Sie ähnelt der Assoziationsanalyse und hilft herauszufinden, ob Daten abhängig oder unabhängig voneinander sind. Eine einfache Regression betrachtet nur zwei Werte wie z. B. a) die Häufigkeit von Fehlern, die in einer zweistündigen Klassenarbeit gemacht werden und b) die vergangene Zeit. Auf Basis eines Regressionsmodells können dann Vorhersagen getätigt werden, z. B. dass mit jedem zehn Minuten, die die Klassenarbeit dauert, die Fehlerhäufigkeit um einen Fehler steigt.