

# Big Data in der Personalauswahl

**FORSCHUNG.** Algorithmen sollen Persönlichkeitsprofile erkannt und damit die US-Wahl beeinflusst haben. Wie geht das – und was heißt das für die Personalarbeit?

Von **Heiko Weckmüller** und **Ricardo Büttner**

**M**itte Dezember 2016 sorgte ein Bericht zur Nutzung von Big Data im US-Wahlkampf für Furore. Hintergrund war ein Artikel der Schweizer Zeitschrift „Das Magazin“. Titel: „Ich habe nur gezeigt, dass es die Bombe gibt.“ Damit spielen die Autoren auf den Physiker Robert Oppenheimer an, der gemeinhin als „Vater der Atombombe“ gilt. Und doch ging es um einen anderen Wissenschaftler – nämlich um den Psychologen Michal Kosinski, der hier gewissermaßen als Vater jener Persönlichkeitsprofile vorgestellt wird, mit denen der US-Wahlkampf gezielt beeinflusst worden sein soll. Wir wollen nun einen nüchternen Blick auf die Arbeit von Kosinski werfen: Anhand seiner wissenschaftlichen Veröffentlichungen wollen wir zeigen, welche Aussagekraft Persönlichkeitsprofile haben, die auf Grundlage von Nutzerdaten aus den sozialen Netzwerken entstehen. Dabei geht es uns allerdings weniger um den US-Wahlkampf als vielmehr um die Frage, welchen Stellenwert Big Data künftig im Personalmanagement haben könnte und was in diesem Feld überhaupt realistisch ist.

## Forschung bei Facebook

Die Arbeiten von Michal Kosinski und seinen Kollegen beruhen auf einer einmaligen Erhebung von Persönlichkeitsprofilen bei Facebook-Nutzern. Für die Studie haben die Teilnehmer einen Fragebogen beantwortet und der wissenschaftlichen Verwendung ihrer Daten zugestimmt. Je nach Auswertung

stützt sich die Studie auf Angaben von bis zu 180.000 Facebook-Nutzern. Obwohl es sich nicht um eine Zufallsstichprobe handelt, dürfte die Repräsentativität trotzdem nur wenig eingeschränkt sein. Die verwendeten Fragen stammen aus dem sogenannten „International Personality Item Pool“. Damit basiert der Fragebogen auf dem wissenschaftlich etablierten Big-Five-Modell der Persönlichkeit. Durch die explizite Zustimmung der Nutzer war eine personenbezogene Verknüpfung mit den individuellen Facebook-Profilen möglich.

## Entzauberung der Algorithmen

Kosinski konzentriert sich auf die „Likes“: Er versucht, individuelle Variationen in den Persönlichkeitseigenschaften gemäß der Selbsteinschätzung in den Fragebögen mithilfe von Likes vorherzusagen. Dieser Schritt wird in den Medien oft mit dem nebulösen Begriff „Algorithmus“ umschrieben. Tatsächlich handelt es sich jedoch um statistische Zusammenhänge in Form von Korrelationskoeffizienten und linearen Mehrfachregressionen mit vergleichsweise kleinen Verfeinerungen. Kosinskis Datengrundlage für diese Auswertung beruht grob geschätzt auf etwa 70 Millionen Datenpunkten. Trotz dieser scheinbar großen Datenmenge und entgegen der Darstellung in den Medien spricht Kosinski nicht von Big Data. Das ist kaum verwunderlich, denn der Datenumfang seiner Untersuchung erfordert keine spezifische Software, die über die Funktionen handelsüblicher Statistikprogramme hinausgehen würde – eine

Voraussetzung, die zum Teil als Definitionsmerkmal von Big Data gilt.

## Zusammenhänge hinterfragen

Kosinski gibt seine Kernergebnisse meist als Korrelationskoeffizienten an. Der Korrelationskoeffizient beschreibt den statistischen Zusammenhang zwischen zwei Merkmalen im Bereich von -1 (vollständiger negativer Zusammenhang) bis +1 (vollständiger positiver Zusammenhang). In einer Studie von 2012 untersuchten Kosinski und seine Kollegen Yoram Bachrach, Thore Graepel, Pushmeet Kohli und David Stillwell zunächst Korrelationen bei abstrakten Metadaten – zum Beispiel, wie die Anzahl der Likes oder Gruppenzugehörigkeiten mit der Persönlichkeit zusammenhängt. Dabei stellten die Forscher nur sehr geringe Korrelationen mit Persönlichkeitseigenschaften fest, wobei diese spezielle Auswertung aufgrund der geringeren Verfügbarkeit von Metadaten ohnehin kaum repräsentativ ist. Gerade bei den personalwirtschaftlich zentralen Persönlichkeitseigenschaften Gewissenhaftigkeit und Verträglichkeit sind die Zusammenhänge besonders gering – diese Persönlichkeitseigenschaften sind etwa im Zusammenhang mit dem Berufserfolg und der Zusammensetzung von Teams relevant. Abstrakte Metadaten, wie sie Kosinski untersucht hat, eignen sich somit kaum zur validen Prognose von Persönlichkeitseigenschaften.

## Interesse und Persönlichkeit

Vielversprechender ist die inhaltliche Auswertung der Likes, also die Frage,

## SOCIAL MEDIA UND PERSÖNLICHKEITSANALYSE

Indikatorgruppe	Indikator	Offenheit für neue Erfahrungen	Gewissenhaftigkeit	Extraversion	Verträglichkeit	Neurotizismus
Statische Profilinformationen	Interessenanzahl	+				
	Gruppenanzahl					+
	Anzahl Profilbilder			+	+	-
Dynamische Profilinformationen	Häufigkeit Profilbildänderungen	+		+		+
	Häufigkeit neuer Bilder	+		+	+	
Nutzungsintensität	Loginhäufigkeit			+	-	
	Nutzungsdauer		-	+		+
	Besuchshäufigkeit anderer Profile		-	+	+	
Kommunikation	Anzahl gesendeter Nachrichten			+	+	-
	Anzahl eigener Kommentare			+	+	+
	Anzahl fremder Kommentare			+	+	
	Anzahl Fauxpas		-		-	
	Anzahl Blogbeiträge	+				
	Anzahl emotionaler Inhalte					+
Netzwerk	Kontaktanzahl	+	+	+	+	-
	Weitere Zentralitätsmaße	+			+	

Die Tabelle zeigt positive (+) und negative (-) Zusammenhänge zwischen Social-Media-Nutzungsattributen und den Big-Five-Persönlichkeitsmerkmalen.

IN ANL. AN GUETTNER 2016

welche und nicht wie viele Themen jemand unterstützt. Dieser Frage ging Kosinski zusammen mit seinen Kollegen Wu Youyou und David Stillwell in einer Studie von 2015 nach. Ergebnis: Bei Berücksichtigung aller Likes beträgt die durchschnittliche korrigierte Korrelation 0,56 – ein Wert, der auch vor dem Hintergrund von Vergleichsstudien als hoch zu bewerten ist. Hierbei weisen einzelne Themenseiten besonders hohe Korrelationen auf. Zum Beispiel „Mitt Romney“ für Gewissenhaftigkeit und „Hello Kitty“ für Neurotizismus, also für die emotionale Labilität des Charakters.

### Der Computer als Menschenkenner

Ein alternativer Vergleich zu den Meta-Studien, die Kosinski zur Einordnung seiner Messwerte heranzieht, wäre die Vorhersagekraft von Personalauswahlverfahren für den späteren Berufserfolg. Hier ergeben sich ähnlich hohe Korrela-

tionswerte von über 0,5 für Intelligenztests und strukturierte Interviews. Wissenschaftlich untersucht ist übrigens auch, wie verlässlich Personalmanager subjektiv Persönlichkeitseigenschaften auf Basis der Sichtung eines Facebook-profils einschätzen können: Die Prognosegüte in diesem Zusammenhang liegt nahe null. Daher sollten Personaler wohl eher den sogenannten Algorithmen als ihrer Intuition vertrauen. Damit wird es tatsächlich möglich, durch inhaltliche Analyse des Facebook-Verhaltens gute Rückschlüsse auf das Persönlichkeitsprofil zu ziehen. Diese Einsicht führt zur grundlegenden Frage, welche Bedeutung Persönlichkeitsprofile für die Personalarbeit haben.

### Persönlichkeit in der Personalarbeit

Die Prognosegüte der Persönlichkeitseigenschaften für das Personalmanagement ist gut untersucht und weitgehend

metaanalytisch hinterlegt. Es zeigen sich stabile, aber nur schwache bis mittlere Zusammenhänge. Zum Beispiel zwischen dem Persönlichkeitsmerkmal „Extraversion“ und transformationalem Führungsverhalten (Korrelation 0,24) oder zwischen der Arbeitszufriedenheit sowie Neurotizismus (0,29) und Gewissenhaftigkeit (0,26). Die Korrelationswerte für Zusammenhänge zwischen Persönlichkeitsmerkmalen und der Performance liegen berufsspezifisch im Bereich kleiner oder gleich 0,2.

Bei der Personalauswahl und -beurteilung stellt Persönlichkeit insofern nur ein Merkmal neben anderen dar. So ist beispielsweise der Zusammenhang zwischen allgemeiner Intelligenz und dem Berufserfolg deutlich stärker ausgeprägt. Die Forschungslage schlägt sich auch in der Praxis nieder, denn Personalmanager sind in der Regel skeptisch, tendenziell sogar überskeptisch, wenn

es um den Einsatz von Persönlichkeits-tests in der Personalauswahl geht.

### Social Media: Darstellung des Selbst oder doch Selbstdarstellung?

Neben Facebook-Likes sind auch andere Zusammenhänge zwischen der Social-Media-Nutzung und den Persönlichkeitseigenschaften bekannt: Bereits vor Kosinski fanden andere Wissenschaftler schwache Zusammenhänge etwa zwischen der Zahl der Kontakte, Nachrichten oder veröffentlichten Bilder und den Big-Five-Persönlichkeitsmerkmalen. Die ersten Studien hierzu stammen aus dem Jahr 2008. Vergleichbare Ergebnisse liegen inzwischen für

---

**Derzeit wird an maschinellen Verfahren gearbeitet, mit denen sich Persönlichkeitstendenzen kostengünstig und beinahe in Echtzeit ermitteln lassen.**

---

verschiedene Social-Media-Plattformen vor. All diese Untersuchungen haben jedoch dasselbe methodische Problem: Die als zusammenhängend erkannten Daten-Persönlichkeitspaare – etwa die Zahl von Kontakten und Extraversion – wurden durch die Studienteilnehmer selbst generiert. Die Social-Media-Nutzer haben ihre Persönlichkeit nämlich in der Regel anhand eines Persönlichkeitsfragebogens selbst eingeschätzt. Das Problem dabei ist, dass auf den Social-Media-Plattformen oft „Impression Management“, also Selbstoptimierung, betrieben wird. Deshalb ist es naheliegend, dass sich etwa eine Person, die gerne als eher extrovertiert wahrgenommen werden möchte, sich nicht nur et-

was extrovertierter einschätzt, sondern auch mehr Kontakte sammelt. Bei den hier vorgestellten Korrelationsanalysen ist entsprechend anzunehmen, dass die Datengrundlage durch Faktoren wie soziale Erwünschtheit, Impression Management und Kohärenzerwartungen zumindest leicht verzerrt wurde.

### Persönlichkeitstendenzen mithilfe von Maschinen erkennen

In der Summe können die schwachen Zusammenhänge zwischen verschiedenen Social-Media-Attributen und Persönlichkeitsmerkmalen trotzdem genutzt werden, um Tendenzen der Nutzerpersönlichkeit zu erkennen. Hier kommen zunehmend prädiktive maschinelle Lernverfahren zum Einsatz, die auch sehr komplexe Zusammenhänge abbilden können. Das Schlagwort für diese Verfahren lautet künstliche Intelligenz: Beispielsweise konnten Forscher zeigen, dass maschinell trainierte sogenannte Entscheidungsbäume deutlich bessere Ergebnisse liefern als lineare Regressionsmodelle. Derzeit wird daran gearbeitet, durch sogenannte „Deep-learning-Ansätze“ weitere Verbesserungen zu erreichen. Die wesentlichen Vorteile solcher Verfahren sind die geringen Einsatzkosten und die Geschwindigkeit: Zumindest grobe Persönlichkeitstendenzen lassen sich somit kostengünstig und beinahe in Echtzeit ermitteln, was beispielsweise für E-Recruiting Systeme hochinteressant ist.

### Die Datenlage bleibt ungewiss

Trotz der Offenheit des Internets ist absehbar, dass die Betreiber von Social-Media-Plattformen die benötigten Daten nicht kostenlos zur Verfügung stellen werden. In Sachen Anfragefrequenz, Anfragezeitraum und Nutzerbreite wird es bei den kostenlosen Datenschnittstellen aller großen Betreiber Einschränkungen geben. Zudem müssen die Nutzer häufig der Auswertung applikationsspezifisch zustimmen, bevor überhaupt Daten abrufbar sind. Hinzu kommt, dass sich einige prädiktive Social-Media-Attribute

ständig ändern: Als „Persönlichkeitsindikator“ weist beispielsweise das Interesse an der Late-Night-Show „Colbert Report“ eine hohe Prognosegenauigkeit auf; die Show ist inzwischen aber eingestellt. Ebenso dürfte die Bedeutung von „Mitt Romney“ in den sozialen Netzwerken wohl künftig abnehmen. Zudem sind die Ergebnisse in vielen Fällen länderspezifisch. Eine Übertragung auf Deutschland ist damit oft nur mit Analogieschlüssen alla „Colbert ist gleich Harald Schmidt“ möglich, womit der Zusammenhang jedoch fraglich wird. Daher müssten in Deutschland zunächst eigene Studien durchgeführt werden.

### Big Data als Orientierungshilfe

Analysen des Social-Media-Verhaltens lassen nur grobe Rückschlüsse auf die Persönlichkeit zu. Sie eignen sich damit zwar nicht für individuelle Aussagen, ermöglichen aber eine überblicksartige Segmentierung anhand von fünf Persönlichkeitseigenschaften. Da Big-Data-Analysen stets Wahrscheinlichkeitseinschätzungen sind, ist eine exakte Vorhersage nicht möglich und wird auch nie möglich sein. Trotzdem sollten sich Personalmanager jedenfalls die Grundlagen aneignen: Sie sollten Wahrscheinlichkeiten als solche anerkennen und sich mit deren Logik beschäftigen – dazu ist Übung erforderlich. Notwendig bleibt das jedoch allemal, denn aus wissenschaftlicher Sicht stehen die Chancen für den Algorithmus schon heute gut: Seine Prognosen sind verlässlicher als die des Personalers, der einen Kandidaten anhand seines Facebook-Profiles einschätzt. ■



**PROF. DR. HEIKO WECK-MÜLLER** lehrt Human Resource Management an der FOM Hochschule Bonn.



**PROF. DR. RICARDO BÜTTNER** lehrt Wirtschaftsinformatik an der Hochschule Aalen.