### **DGI-Praxistage 2016**

**Iens Damm\*** 

# Der Blick in die Glaskugel bleibt Zukunftsmusik

Die 6. DGI-Praxistage boten einen Einblick in die spannende Welt der "Predictive Analytics"

DOI 10.1515/iwp-2017-0011

In den vergangenen Jahren sind Begriffe wie "Big Data" oder auch "Predictive Analytics" in Medien und Gesellschaft omnipräsent geworden. Dazu sind entsprechende Methoden und Lösungen zu einer Vielzahl von Unternehmen in der Praxis durchgedrungen. Nichtsdestotrotz scheint weiterhin eine gewisse Unsicherheit zu existieren, wie ebendiese Begriffe genau einzuordnen und zu bewerten sind und wie sie die alltäglichen Arbeitsfelder in den verschiedenen Unternehmen beinflussen.

Unter anderem aus diesen Gründen wurde bei den DGI-Praxistagen am 10. und 11. November 2016 die Ausgangsfrage gestellt:"Predictive Analytics – Blick in die Glaskugel oder glasklare Prognose?". Die Tagung begann mit drei Workshops. Zwei stellten konkrete Anwendungen vor: Expert Systems Deutschland GmbH präsentierte mit Cogito Studio Werkzeuge zur Erfassung und Auswertung von Daten in Textform, während Tableau Germany GmbH mit Tableau die Visualisierung von Daten für die Teilnehmer praktisch erfahrbar machte. Der dritte Workshop richtete sich an Studenten und Studentinnen und hatte das Thema "Wie Online Marketing politische Öffentlichkeit beeinflusst!" (s. Bericht S. 67)

Die Vortragsreihe am Donnerstag eröffnete schließlich Dr. Sven Schmeier, Mitarbeiter am Deutschen Forschungszentrum für Künstliche Intelligenz. Er beschrieb den Begriff Predictive Analytics als ein Zusammenspiel von zwei Elementen: Auf der einen Seite Daten, die entweder in strukturierter Form, als Zahlenreihen oder -tabellen, oder in unstrukturierter Form, als Textbeiträge oder Videodateien, verfügbar seien. Auf der anderen Seite eine gewisse Art der Intelligenz, etwa ein Computerprogramm. Beide Elemente zusammengeführt, erlauben eine Prognose, beispielsweise zum Wetter oder zum Ausgang eines bestimmten Pferderennens.

\*Kontaktperson: Jens Damm, Hochschule Darmstadt, Fachbereich Media, Max-Planck-Straße 2, 64807 Dieburg, E-Mail: jens.p.e.damm@stud.h-da.de Dabei sei das Konzept solcher Vorhersagen keineswegs eine neue Erfindung. Im 19. Jahrhundert erforschte der Londoner Arzt Dr. John Snow die Cholera-Epidemie und versuchte falsche Theorien über deren Ausbreitung zu widerlegen. Da er den Erreger in Flüssigkeiten vermutete, dies aber durch die zu dieser Zeit vorhandenen Messinstrumenten nicht nachweisen konnte, zeichnete er auf einer Karte die amtlich registrierten Todesfälle ein. Dadurch war es ihm möglich ein Epizentrum zu identifizieren, das sich um eine bestimmte Wasserpumpe anordnete. Snow verhinderte deren weitere Nutzung durch das brachiale Abbrechen des Schwengels. Dies stoppte die Epidemie tatsächlich. Er hatte also bekannte Fakten zu Rate gezogen, diese visualisiert und schließlich ein gewisses Muster identifiziert.

Herr Dr. Schmeier spannte den Bogen von der Geburtsstunde der künstlichen Intelligenz im Jahr 1956 über die Weiterentwicklung zu eigenständig lernenden Computern, wie die IBM-Supercomputer, die in Dame, Schach und letztendlich 2011 im Spiel Jeopardy triumphierten bis hin zum heutigen Entwicklungsstand. Im "Internet des Menschen" werden eben diese durch die Nutzung von Social Media, Dienstleistungen oder der Aufzeichnung ihres Surfverhaltens selbst zur Ware. Im Internet der Dinge werden Smartphones, Fahrzeuge und die eigenen vier Wände vernetzt und dadurch riesige Datenmengen generiert.

Unternehmen wie Toyota oder Google investierten Milliarden in die Erforschung künstlicher Intelligenz und vor allem sogenannter künstlicher neuronaler Netze, die sich am menschlichen Gehirn orientieren sollen. Diese Netze bieten mit einfachen mathematischen Modellen eine hohe Rechengeschwindigkeit. Dadurch seien viele Probleme gleichzeitig lös- und berechenbar, wie die Erkennung von Handschriften bei der Deutschen Post oder die Spracherkennung mit dem Smartphone. Warum und wie diese künstlichen Netze genau funktionierten, wisse niemand so recht – sie würden schließlich von den Computern eigenhändig kreiert.

Für die Zukunft erwartet Dr. Sven Schmeier das Vordringen der Digitalisierung in immer mehr Gesellschaftsbereiche, so beispielsweise in der Landwirtschaft, dem Gesundheitswesen oder der Arbeits- und Produktionswelt.

Belegt durch das sogenannte Mooresche Gesetz, verdoppele sich die Rechenleistung moderner Computer jährlich. Dies erweitere die Anwendungsmöglichkeiten von Predictive Analytics stark. Dazu erhöhe sich quantitativ wie auch qualitativ die Datengrundlage, weil neuronale Netze nicht mehr nur das ganze Wissen von Grund auf erlernen, sondern durch vorhandene Datensätze angereichert werden könnten.

Den Abschluss des ersten Tages der DGI-Praxistage bildete ein Kamingespräch verschiedener Branchenexperten. Dr. Thomas Keil vom SAS Institute begann die Vorstellungsrunde und schilderte einen Anwendungsfall, bei dem ein großer Lebensmittelkonzern durch Predictive Analytics vorhersagen konnte, welche der 10.000 Produkte zukünftig an einem Standort nachgefragt werden würden. Alexander Loth von Tableau Germany beschrieb ein "Predictive Maintenance Dashboard", das dem Endnutzer im Unternehmen durch Sensoren anzeige, wann ein bestimmter Aufzug gewartet werden müsse. Das Unternehmen Expert Systems mit Stefan Geißler hilft Schweizer Banken, auf ein verändertes rechtliches Umfeld zu reagieren und eine Vielzahl von Textdokumenten maschinell auszuwerten. Prof. Dr. Melanie Siegel untersuchte mit ihren Studenten der Hochschule Darmstadt, ob die Plattform Twitter das politische Stimmungsbild ähnlich wie traditionelle Meinungsumfragen abbilden könne<sup>1</sup>. Vervollständigt wurde die Runde durch Dr. Sven Schmeier, der mit seinem früheren Startup "XtraMind" den Kundenservice von Unternehmen zu verbessern half, indem eingehende E-Mails automatisch in Kategorien vorsortiert wurden.

Die Teilnehmer diskutierten unterschiedlichte Fragestellungen. Dr. Schmeier meinte, das Ziel von Predictive Analytics sei nicht, dass die Maschine den Menschen ersetze. Stattdessen stehe eine ergänzende Zusammenarbeit im Fokus, von der Letzterer durch eine gesteigerte Produktivität und Zufriedenheit profitieren könne. Frau Professor Siegel fügte hinzu, dass Computerprogramme nur die zugrunde liegende Datenbasis für eine fundierte Entscheidung des Menschen liefern könnten. Darüber hinaus wurde diskutiert, dass auch im Gesundheitswesen die Nutzung von Datenanalysen nicht grundsätzlich ausgeschlossen werden dürfe. Eine Überwachung der Essgewohnheiten für eine stark verlängerte Lebensdauer von transplantierten Organen oder die anonyme Auswertung von Behandlungsmethoden zur Bestimmung der Erfolgsquote könne im Alltag Menschenleben retten, ergänzte Dr. Keil. Anschließend bot sich bei einem Buffet, die Möglichkeit sich mit den Diskutanten und Zuhörern auszutauschen.

#### Geschäfte mit Daten

Der zweite Tag der Veranstaltung wurde durch den Vortrag von Dr. Thomas Keil eröffnet. Gleich zu Beginn veranschaulichte eine Grafik die unglaublichen Datenmengen, die im Jahr 2016 pro Minute im Internet generiert wurden. Dieses exponentielle Wachstum werde auch durch die bis 2020 ca. 25 Milliarden vernetzten Internet-of-Things-Geräte weiter steigen. Strategien zur Vorhersage im Sinne von Predictive Analytics habe es vor 20 Jahren auch schon gegeben, heute würden diese aber aufgrund höherer Rechengeschwindigkeiten bei gleichzeitig geringerem Einsatz von Ressourcen zunehmend demokratisiert. Unternehmen, welche diese Chance nicht zu nutzen wüssten, werden nach Meinung von Dr. Keil den Anschluss verlieren. Für die Zukunft beschrieb der Referent den abzusehenden Einsatz und auch Verkauf bereits vorhandener Daten in bisher nicht bedachten Anwendungsfeldern. So könnten Telekommunikationsanbieter ihr Geschäftsmodell erweitern, indem von Sendemasten erhobene Wetterdaten, zum Beispiel zur Bestimmung von Unwettern aufgrund verringerter Signalstärken, an Versicherungen oder Wetterdienste weiterverkauft würden. Wie bereits in Einzelfällen üblich wäre es dann flächendeckend möglich. Tarife von Autoversicherungen mithilfe von Daten zum Fahrverhalten individuell auf den Halter zuzuschneiden. Auch könne die Ernährung der stark wachsenden Erdbevölkerung durch intelligente Analysemodelle in der Landwirtschaft gesichert werden, indem zum Beispiel Wetterdaten und andere Sensoren den perfekten Zeitpunkt der Ernte bestimmen würden.

Eine Live-Demo der sogenannten "Predictive Analytics Suite" zeigte Alexander Häußer von der Firma PwC in seinem darauffolgenden Vortrag. Diese mit der Programmiersprache R entwickelte Anwendung sei webbasiert und vereine die drei Säulen von Predictive Analytics in einem Programm. Diese würden die Datenanalyse, die Generierung von "Explanation Models" und das Erstellen von Vorhersagen umfassen. Als Beispiel wurden in der Systemvorführung die Absatzzahlen bestimmter Automarken vorhergesagt.

## Beispiele gesteigerter Effizienz

Nick Verstegen von der Firma StatSoft stellte bei seinem Vortrag fünf Kundenprojekte aus der Praxis vor. Für ein

<sup>1</sup> Vgl. hierzu den Beitrag von Melanie Siegel et al.: Automatische Erkennung von politischen Trends mit Twitter – brauchen wir Meinungsumfragen noch? auf Seite 67 in dieser Ausgabe.

Versicherungsunternehmen sei zum Beispiel eine klassische Werbekampagne über Briefpost optimiert worden. indem favorisierende Eigenschaften der Adressaten erst identifiziert und anschließend an Hand eines statistischen Modells die Zahl der verschickten Briefe halbiert wurde. Die Antwortrate erhöhte sich dadurch um 41 Prozent.

Bei einem großen Automobilzulieferer seien Produktionsprozesse automatisiert worden. Die Qualität eines millionenfach produzierten Bauteils musste vor dem Projekt mit großem Aufwand vollständig manuell überprüft werden. Durch den Einsatz von Predicitive Analytics konnte der personelle Einsatz um 90 Prozent verringert werden. Ein Beispiel aus dem Gesundheitsbereich konnte Verstegen auch vorstellen: So sollte das Risiko postoperativer Infektionen bei einem bestimmten Eingriff gesenkt werden. Durch die Verbindung der Daten aus der Patientenakte mit denen, die im Operationssaal entstanden, hätten noch während des Eingriffs bestimmte Gegenmaßnahmen vorgeschlagen werden können. Dies reduzierte zur Freude von Patienten und behandelnden Ärzten die postoperativen Infektionen um 74 Prozent.

Sogenannte unstrukturierte Daten waren das Thema von Joachim Lackner, Expert Sytsems. Diese können, wie bereits erwähnt, zum Beispiel in Text-, Audio- oder auch Videoform vorliegen und machen geschätzt bis zu 80 Prozent der allgemein vorhandenen Datenmenge aus. Die von dem Unternehmen entwickelte Software bringe der Maschine die menschliche Sprache bei - dies umfasse Zweideutigkeiten, Ironie, Wortvarianten oder Beziehungen innerhalb von Wörtern und Sätzen. Eine Pflege dieser semantischen Plattform sei sehr aufwendig, aber lohnens-

Ein mögliches Einsatzgebiet seien Anfragen über die Servicehotlines von Unternehmen. 70 Prozent dieser Anfragen seien standardisierbar, weshalb durch eine im System hinterlegte semantische Datenbank mögliche Fragen bereits ohne menschlichen Eingriff beantwortet werden könnten. Auch würden im Bereich des Schadenmanagements bei Versicherungen die Kosten um die Hälfte reduziert werden, wenn eine Software die Verwaltung der Dokumente automatisiere. So könne bereits bei Meldung eines Schadens eine Kategorie zugewiesen und geprüft werden, ob die Person überhaupt Versicherter des Unternehmens sei. Die Kommunikation und Auszahlung an den Kunden würde ebenfalls beschleunigt werden.

## Datenanalysten intern einsetzen

Thomas Gabriel, Mitgründer der Open Source Software KNIME, war der nächste Vortragende. Er vermutete, dass es in der Zukunft zu viele Daten geben werde, um diese an einem zentralen Ort speichern zu können. Die zu berücksichtigenden Daten würden außerdem viele verschiedene Fachgebiete umfassen. Dementsprechend sehe er die Zukunft einerseits in domänenübergreifenden Konzepten, also zum Beispiel der Kombination von Automobil- und Wetterdaten. Andererseits würden die Daten dort analysiert werden müssen, wo sie angefallen sind - ermöglicht durch eine parallele Ausführung von Rechenprozessen mit Tools wie zum Beispiel Apache Spark. Früher sei es üblich gewesen, dass Datenanalysten isoliert von der eigenen Organisation agierten. Heutzutage sei eine direkte Verzahnung mit den Fachgebieten unabdingbar, die Analysten müssten also direkt vor Ort eingesetzt werden. Die KNIME Software ermögliche durch eine interaktive Drag & Drop-Oberfläche und den optionalen Verzicht auf Programmiersprachen die Analyse für jedermann. KNIME beinhalte beispielsweise Elemente zur Datenvisualisierung als auch Text-Mining-Komponenten zur Auswertung unstrukturierter Daten. Die Anwendung könne zudem in Echtzeit auf einer Serverumgebung laufen.

### **Genios Smart Data Services**

Der Vormarsch von Predictive Analytics ermögliche auch klassischen Aggregatoren von Wirtschaftsinformationen wie Genios eine rasante Erweiterung der angebotenen Dienstleistungen, wie Ralf Hennemann beschrieb. Mehr als 400 Millionen lizensierter Dokumente sowie Inhalte von Webseiten, Foren oder Social Media bewertet, ordnet und komprimiert das Unternehmen für seine Kunden. Erste Predictive-Analytics-Projekte beinhalteten die Zusammenführung von Kundenstamm-Daten mit den Genios-Daten aus Presse, Fachpresse oder Handelsregister. Dadurch ließen sich laut Hennemann Rückschlüsse auf die Bewertung bestimmter Risiken durch den Unternehmenskunden ziehen. Dies wirke sich positiv auf die eigenen Vertriebsaktivitäten aus, weil Signale des Kunden für eine geplante Umsetzung bestimmter Projekte herausgefiltert werden könnten. Selbstverständlich habe Genios bei diesen neuartigen Dienstleistungen auch unterschiedlichste Herausforderungen bewältigen müssen. So sei es zu Beginn eines Projektes besonders schwierig, die konkreten Anforderungen des Kunden mit Hilfe der verschiedenen Teilnehmer zu identifizieren. Daran anschließend sei die Koordination und Entscheidungsfindung ebenfalls komplexer als im für Genios klassischen Research-Sektor, da verschiedenste Fachbereiche und Abteilungen zusammenkommen müssten. Auch sei die Integration zum Beispiel in vorhandene CRM-Systeme des Kunden sowie das Schaffen neuer Lizenzmodelle eine Herausforderung bei diesen Projekten. Auch aufgrund dieser Erfahrungen verfolge Genios mittlerweile einen kooperativen Ansatz, um den Kunden individuell zugeschnittene Lösungen anbieten zu können, so Hennemann weiter. Es werde regelmäßig mit der Fraunhofer-Gesellschaft, dem SAS Institute oder auch kleinen Start-ups kooperiert. Aktuell arbeite Genios beispielsweise an einer Projektdatenbank für eine Baufirma. Diese Anwendung registriere durch die Kombination mit Pressedaten, wie der Baufortschritt eines Projektes voranschreite. Dadurch könne der Kunde den richtigen Zeitpunkt für ein eigenes Angebot bestimmen, um die Chance auf einen Zuschlag zum Beispiel für Malerarbeiten zu erhöhen. Dies automatisiere einen vorher manuellen Vorgang und könne mit dem eigenen CRM-System kombiniert werden.

### Wortvektoren statt Wissensnetz

Den Abschluss der interessanten Vortragsreihe der diesjährigen Praxistage gestaltete Tobias Hartmann von Avantgarde Labs. Man helfe vor allem mittelständischen Unternehmen dabei, die Unmengen an Tools und Technologien von Predictive Analytics für unstrukturierte Daten auf das Notwendigste zu reduzieren. Dazu zähle zum Beispiel eine Produktsuche für Online-Shops. Die Kunden seien die umfassenden Suchmöglichkeiten der bekannten Anbieter gewohnt, die aber nur mit großem Aufwand nachzubilden seien. Die Generierung dynamischer Vorschläge von verwandten Produkten oder die Korrektur falscher Suchanfragen sei sehr komplex. Für Personaldienstleister und auch unternehmensinterne Personalabteilungen sei eine Lösung für das sogenannte Kompetenzmatching erarbeitet worden, so Hartmann weiter. Hierbei würden Lebensläufe der Bewerber automatisch mit passenden Jobangeboten kombiniert, indem auf historische Daten über erfolgte Einstellungen zugegriffen werde. Auch das Wissensmanagement einer Engineering-Firma konnte das Unternehmen optimieren. Abertausende von Serviceberichten wurden in Problemklassen kategorisiert und mit einer Suchanwendung kombiniert. So würden Techniker zeitnah eine Lösung für ein Problem identifizieren können, das bereits an anderer Stelle aufgetreten und gelöst worden war.

Tobias Hartmann beendete seinen Vortrag mit einer Live-Demo. Sogenannte Wortvektoren wären hochaktuell und würden nur durch den Einsatz statistischer Methoden, also ohne das manuelle Erstellen eines semantischen Wissensnetzes, die Beziehungen zwischen Begriffen in einem semantischen Raum visualisieren. So konnten die Zu-

schauer nachvollziehen, dass der Algorithmus die Beziehung der Wörter "Berlin" und "Spree" auf den Begriff "Prag" übertragen und so das Ergebnis "Moldau" liefern konnte. Als wichtige Erfolgsfaktoren der Projektarbeit von Avantgarde Labs nannte Hartmann noch, dass man mit kleinen Anwendungsfällen beginnt, die einen direktem Nutzen versprechen und dass man von Anfang an die verschiedenen Blickfelder aller Fachabteilungen integrieren sollte.

Nach den vielen verschiedenen Eindrücken, die während der DGI-Praxistage gesammelt wurden, diskutierten die Teilnehmer zum Schluss erneut die Frage, ob Predictive Analytics der "Blick in die Glaskugel" oder "eine glasklare Prognose" sei. Prognosen habe es schon immer gegeben, wie das Beispiel des Arztes John Snow zeige. Heutzutage gebe es aber durch die stark gesteigerte Rechenleistung und die Verfügbarkeit einfacher Visualisierungen Predictive Analytics für jedermann. Es bleibe jedoch bei einer "milchigen Glaskugel", die noch längst keine glasklaren Prognosen abliefern könne. Wichtig sei jetzt, voneinander zu lernen und disziplinübergreifend zu arbeiten, so das Fazit der Anwesenden.

Seit dem Jahr 2011 gibt es die DGI-Praxistage, in deren Mittelpunkt aktuelle Entwicklungen und Problemstellungen im Arbeitsalltag und Erfahrungsberichte im Sinne von Best Practice stehen. Fachvorträge und Diskussionsrunden ermöglichen einen intensiven Erfahrungs- und Wissensaustausch für alle Beteiligten.

Für weitere Informationen zu den einzelnen Vorträgen empfiehlt sich ein Besuch der Website der DGI, auf der einige der Vortragsfolien abzurufen sind. Außerdem sei darauf hingewiesen, dass sowohl Tableau Deutschland als auch das SAS Institute auf ihren Internetauftritten kostenlose Versionen ihrer Anwendungen anbieten.

**Deskriptoren:** Tagung, Datenanalyse, Prognose, Informationsanalyse, DGI-Praxistage 2016



Jens Damm Hochschule Darmstadt Fachbereich Media Max-Planck-Straße 2 64807 Dieburg jens.p.e.damm@stud.h-da.de

Jens Damm studiert Onlinekommunikation an der Hochschule Darmstadt. Er arbeitet außerdem als Werkstudent im Big-Data-Team der Merck-Gruppe und interessiert sich für das "Internet der Dinge" (IoT).