



Auftakt: DGI-Präsident Reinhard Karger kündigt den Keynote-Speaker Dr. Sven Schmeier an, der grundlegend in das Thema Datenanalyse einführte.

Vom Leben und Sterben in einer gewissen Wahrscheinlichkeit

Bericht über die 6. DGI-Praxistage „Predictive Analytics - Blick in die Glaskugel oder glasklare Prognose?“ Frankfurt am Main, 10./11.11. 2016

Vera Münch

Seit Menschengedenken wünscht sich der Mensch zu wissen, was morgen passiert. Aus Daten abgeleitete Vorhersagen lassen diesen Traum jetzt ein Stück näher rücken. Mit Big Data Analysemodellen und daran entlang entwickelten, selbstlernenden Softwaremaschinen wird in großen, manchmal über die ganze Welt verteilten Daten nach Mustern und Signalen gesucht, die von Maschinen interpretiert und zu Prognosen weiterverarbeitet werden. Was liefern diese sogenannten Predictive Analytics? Den Blick in die Glaskugel oder glasklare Prognosen? Die Deutsche Gesellschaft für Information und Wissen (DGI e.V.) hinterfragte bei den 6. DGI-Praxistagen unter dieser Überschrift die tatsächliche Prognosefähigkeit von Datenmustern, diskutierte Auswertungsmodelle, Softwarewerkzeuge und Anforderungen an die Datengrundlagen. Nach anderthalb informativen Tagen mit Fachleuten aus der Informationswirtschaft und der Informationswissenschaft stand fest: Predictive Analytics sind kein Blick in die Glaskugel der Wahrsagerin. Aber glasklare Prognosen kann es auch nicht geben, weil, so Dr. Thomas Keil von der SAS Institute GmbH, „immer unter Wahrscheinlichkeitsbedingungen gerechnet werden muss“ und „es immer wieder Ereignisse gibt, die sich komplett aus der Vorhersage herauskatapultieren“. Wenn die Modelle stimmen, sind ganz viele Prognosen aber schon verdammt gut. Die ersten Fachleute machen auch schon den nächsten Schritt in die Zukunft: Prescriptive Analytics - Handlungsempfehlungen, die aus Ergebnissen prognostischer Analysen automatisch generiert werden. Manche sagen dazu Data Science.



» „Muster entdecken und Voraussagen machen, das ist der Kern von Predictive Analytics“, erklärte Dr. Sven Schmeier vom Deutschen Forschungszentrum für Künstliche Intelligenz (DFKI) am Vorabend des Konferenztages in seiner Keynote „Predictive Analysis - gestern, heute, morgen“. Zwischen drei intensiven Workshops am Nachmittag und dem Kammingespräch am Abend vor dem Konferenztag führte Schmeier grundlegend in das Thema Datenanalyse ein. Er berichtete über die Geschichte der Auswertung von Daten, die, wie er an der Cholerabekämpfung von Dr. John Snow in London belegte, schon Mitte des 19. Jahrhunderts sehr gute Erfolge brachte. Die Forschung zum maschinellen Lernen, welches Softwaremaschinen in Daten Muster und Gesetzmäßigkeiten erkennen lässt und heute allen Big Data Analysen zugrunde liegt, nahm im Jahr 1949 ihren Anfang. Es gilt als das Geburtsjahr der sogenannten Künstlichen Intelligenz (KI, Artificial Intelligence AI). Wo KI heute steht und was Fachleute für ihre Zukunft vorhersagen, dazu gibt es im Web auf der (englischsprachigen) Seite KDnuggets einen interessanten Eintrag¹. Zehn Jahre nach dem Geburtsjahr von KI veröffentlichte IBM 1959 die erste Publikation zum Maschinellen Lernen, in der beschrieben wurde, wie das Unternehmen einem Computer das Damespiel beigebracht hat. Die Spielsoftware war zunächst nur mit Grundwissen ausgestattet. Sie lernte dann durch automatisches „Training“, welche Spielzüge und Spielverläufe zu welchen Ergebnissen führen. Schon nach kurzer Zeit hatte der Mensch keine Chance mehr, in diesem Spiel gegen den Computer zu gewinnen. Letzterer kann die Spielzüge einfach schneller und weiter vorausberechnen. Die Aktien von IBM stiegen mit den Erfolgen. Softwaremaschinen automatisch auf Daten zu trainieren ist für Analysen heute Standard.

In den fünf auf die Geburtsstunde der KI folgenden Jahrzehnten machten die Forschenden bei IBM aus ihrem legendären Computer Deep Blue einen maschinellen Schachspieler. Sie ließen ihn gegen den Großmeister Garri Kasparow antreten. Nach zahlreichen Niederlagen und Remis war es 1997 so weit: Deep Blue besiegte Kasparow. Vor fünf Jahren, 2011, gewann dann der IBM Watson in der Quizshow Jeopardy beim Beantworten von Quizfragen gegen die menschlichen Teilnehmer.

Schmeier führte mit diesen Beispielen vor Augen, wozu Mustererkennung, Vorausberechnung und Prognose in Kombination mit starker Rechenleistung fähig sind und damit auch, welches Potential in Analysen und Predictive Analytics steckt. Er zeigte, wie die großen Internet-Unternehmen dieses Potential nutzen. Schon heute verfolgen (tracken) die Vorreiter der globalen Digitalwirtschaft Bewegungs- und Handlungsdaten von Nutzern im Internet in ganz großem Stil. Die erfassten, aufgezeichneten Daten analysieren sie und bauen darauf geschäftliche Handlungen auf; z.B. gezielte Produkt- und Dienstleistungsangebote, Anzeigenschaltungen, Trendanalysen, neue Geschäftsmodelle und ganze Geschäftsbereiche, den Verkauf von Information, gezielte Investitionsplanung und vieles mehr.

Silicon Valley Größen schließen KI-Forschungspartnerschaft

Wie umfassend die noch zu bewältigenden Entwicklungsaufgaben sind und wie groß die Hoffnungen, das Potential von Big Data Analytics gewinnbringend für die Wirtschaft und vielleicht auch die Menschheit zu nutzen, kann man an einer neuen Partnerschaft ablesen, über die Schmeier informierte: Google, Facebook, Amazon, IBM und Microsoft haben sich Mitte 2016 zu einer Kooperation für eine gemeinsame KI-Forschung zusammengeschlossen. Apple und der erfolgreiche Unternehmer Elon Musk, die zwei noch

verbleibenden Silicon Valley-Größen, beteiligen sich nicht an dieser „Partnership on AI“ was, wie Sven Schmeier berichtete, die Zeitung Guardian² zu der Vermutung verleitet, es gäbe wohl ethische Bedenken. Elon Musk selbst investiert nach Aussage des Vortragenden eine Milliarde US-Dollar in AI-Forschung, was, so glaubt Schmeier, dann „wohl eher in die ethische Forschung fließen dürfte“.

Als zu erwartende „qualitative Sprünge“ im Bereich KI und Analytics sagte Sven Schmeier abschließend für die nächsten Jahre voraus:

- Wissen wird aus Datenbanken in neuronale Netze transferiert werden
- Entscheidungen werden erklärbar (er),
- Mehr Daten werden Unsupervised Learning für verschiedene Aufgaben ermöglichen, z.B. Lernen durch Beobachtung

Big Data Analyse als Selbstbedienungsangebot

Im anschließenden Kammingespräch zeichneten Alexander Loth, Sales Consultant, Tableau Software, Dr. Thomas Keil, SAS Institute GmbH, Prof. Dr. Melanie Siegel, Hochschule Darmstadt, Dr. Sven Schmeier und Dr. Stefan Geißler, Expert System Deutschland GmbH dem Auditorium auf, welche Vorteile und Chancen aus Daten analysierte Prognosen für Wirtschaft und Gesellschaft bergen, überlegten aber auch gemeinsam, wo Analytics vielleicht besser nicht angewandt werden sollten. DGI-Präsident Reinhard Karger, Unternehmenssprecher des DFKI, führte durch das Gespräch. Alexander Loth, der schon am Nachmittag in einem fesselnden Workshop erläutert hatte, wie das Not-for-Profit-Unternehmen Tableau Software mit seinem an der amerikanischen Stanford University entwi-

¹ <http://www.kdnuggets.com/2016/11/13-forecasts-on-artificial-intelligence.html>

² <https://www.theguardian.com/technology/2016/sep/28/google-facebook-amazon-ibm-microsoft-partnership-on-ai-tech-firms>



Beim Kaminesgespräch ohne Kamin diskutierten die Analytics-Experten die Chancen und die Stolperfallen maschineller Big Data-Auswertungen für Vorhersagen und Empfehlungen. (V.l.n.r.) Alexander Loth, Tableau Software, DGI-Präsident Reinhard Karger, Thomas Keil, SAS, Prof. Dr. Melanie Siegel, Hochschule Darmstadt, Dr. Schmeier, DFKI, Stefan Geißler, Expert System.

ckelten Softwarewerkzeug Big Data Analysen für alle Angestellten in Unternehmen, aber auch für alle anderen Interessenten nutzbar machen will, wiederholte die Vision: Ziel von Tableau ist, allen Menschen die Chance zu geben, große Datenmengen aus ihrer eigenen Perspektive betrachten, analysieren und interpretieren zu können. Das Selbstbedienungsangebot für Big Data Analysen kann man in der Basisversion kostenlos nutzen. Der Zugang erfolgt über die Webseite von Tableau³. In seinem Blog alexloth.com schreibt der Vertreter von Tableau Berichte zu aktuellen Datenanalyseprojekten. Jüngstes Beispiel ist eine Untersuchung zum Wartungsbedarf von Fahrstühlen auf Bahnhöfen (Predictive Maintenance)⁴, für das Bahnkunden-Tweets ausgewertet wurden.

Thomas Keil von SAS brachte in seinem Statement kurz und prägnant auf den Punkt, was Optimierung durch Predictive Analytics in der Detailfrage bedeutet: z.B., „dass die Bank mit Hilfe von SAS Software weiß, wann sie den Geldautomaten am Fußballstadion mit mehr Bargeld bestücken muss und wann

er eher weniger beansprucht wird“. Stefan Geißler berichtete aus seinem Tagesgeschäft von menschlichen Tragödien und großem Zeitgewinn durch maschinelle Analyse. Monatlang hätten sich in einer Schweizer Bank hoch qualifizierte, von der Arbeit frustrierte Uni-Absolventen durch Kundenakten quälen müssen, um herauszufinden, welche Steuerpflichtangaben nach neuen amerikanischen Steuergesetzen den dortigen Behörden zu melden waren, bevor sein Unternehmen Temis (jetzt Expert System) ins Boot geholt wurde. Nachdem das Modell zur Datenanalyse aufgesetzt war, schaffte das Analysewerkzeug an einem Wochenende das Äquivalent von sechs Monaten menschlicher Recherche. An einem Wochenende. Melanie Siegel brachte Personal Analytics ins Spiel, die z.B. in der Medizin sehr hilfreich sein können und beispielsweise für Kopfschmerztagebücher schon genutzt werden. Sie warnte aber auch vor Verfälschungen, die durch falsch interpretierte Aussagen oder absichtliche Irreführung entstehen können. Die Professorin für Informationswissenschaft untersucht mit ihren Studierenden an der Hochschule für Angewandte Wissenschaften in Darmstadt solche Fragen.

„Automatische Antworten machen andere Fehler als der Mensch“

Reinhard Karger fragte die Runde, für welche Gegenstandsbereiche sich Predictive Analytics nicht oder eher weniger eigenen, worauf Sven Schmeier aus einem Projekt zur automatischen Auswertung und Beantwortung von eMail-Anfragen an ein Unternehmen berichtete, bei dem man von zu hoch gesteckten Zielen hätte ablassen müssen. Zunächst hätte man die Vision gehabt, alles automatisieren zu können. Aber man hätte gelernt „automatische Antworten machen andere Fehler als der Mensch“. So sei etwa der Satzaufbau in Anfragen für Maschinen oft sehr schwer interpretierbar, beispielsweise, wenn ein Kunde schreibe „ich bestelle die Zeitung für meine Großmutter ab“. Alexander Loth verwies darauf, in Sprache ausgedrückte Gefühle seien noch nicht sehr gut erfassbar. Sentimentanalyse würde zwar an manchen Stellen bereits mit einigem Erfolg gemacht, aber z.B. Sarkasmus und Ironie würden in deutschen Tweets die Auswertung sehr schwer machen. Sven Schmeier erzählte von einem Selbstversuch mit 10 CDs, die er bei Amazon auf einen Schlag gekauft hat, um den Empfehlungsalgorithmus aus-

3 <http://www.tableau.com/de-de>

4 <http://alexloth.com/2016/10/30/predictive-maintenance-hilft-ihnen-wartungsmasnahmen-effizient-zu-gestalten/>

zuprobieren. Ab der 4. Empfehlung, was andere Kunden gekauft haben und ihn demnach auch interessieren müsste, hätten ihm die Vorschläge nicht mehr gefallen. Konsumbasierte Empfehlungen, so der Big Data Auswertungsexperte, seien „over-designed. Der Mensch ist doch eher ein Einzelfall“. Der Musik-Streamingdienst Spotify hingegen mache es mit dem Mix der Woche recht gut. Warum Spotify Empfehlungen so gut kann, verstand man am darauffolgenden Tag, als Thomas Keil in seinem Eröffnungsvortrag dazu Zahlen und Fakten nannte.

Was ist bei Trump, Brexit und der Bankenkrise passiert?

Natürlich beschäftigten die Prognosen zu Trump und Brexit als aus der Norm abweichende Ereignisse auch die im Gästehaus der Goethe-Universität versammelten Expertinnen und Experten. Wie es denn dazu hätte kommen können, dass die Meinungsforschung so versagte, wollte Reinhard Karger von der Gesprächsrunde wissen. Einer der Hauptgründe für die falschen Wahlprognosen zu Trump sei zunächst einmal die Tatsache, dass „Data Mining keine Strategien erkennt“, antwortete Sven Schmeier. Der Moderator ergänzte, das Trump-Team hätte auf Strategie und Intuition gesetzt. Zur fehlenden Strategiebeobachtung kommt bei den Phänomenen Trump und Brexit, was Alexander Loth angesprochen hatte und was nun die Mehrheitsmeinung auf dem Podium dazu war: Gefühle, unterschwellige Empfindungen und der Wahrheitsgehalt von Antworten auf Befragungen in den Analysen sind trotz guter statistischer Wahrscheinlichkeitsmodelle noch nicht erfassbar. Hier käme die sozialpsychologische Forschung zur Ergänzung der rein mathematischen Modelle ins Spiel, so Thomas Keil. Als Mensch hätte man vor der Bankenkrise die Signale gesehen. Aus den Datenmodellen seien sie trotz des durchdigitalisierten Bankgeschäfts nicht vorhersagbar gewesen.

Am Ende des Gedankenaustausches resümierte Reinhard Karger, das große Potential von Predictive Analytics sei unbestreitbar, doch könne man noch nicht sagen, das Problem Predictive Analytics sei gelöst.

Kundenverhaltensdaten füllen bei Spotify 40 PetaByte-Datenbank

Am nächsten Morgen präsentierte Thomas Keil in seinem Eröffnungsvortrag zum zweiten DGI-Praxistag unter dem Titel „Predictive Analytics - Vorhersagen mit Big Data“ einen allgemeinen Überblick über grundlegende Konzepte wie Modellbildung und Big Data. Der Marketingmanager für Business Analytics bei der SAS GmbH und Spezialist für Big Data unterfütterte seine Ausführungen mit interessanten Zahlen und Beispielen aus der Praxis und gab auch eine Antwort auf die Frage, warum Predictive Analytics als Arbeitswerkzeug für die Wirtschaft derzeit weltweit so große Aufmerksamkeit erfahren: „Heute kommt es nicht mehr darauf an, noch schneller zu wissen, was ich gestern verkauft habe, sondern, was muss ich als Händler morgen zu welchem Preis ins Regal stellen, damit ich meine Marge erhöhe. Darin steckt die Frage: was wird als nächstes passieren und wie kann ich optimal darauf reagieren“. Aus Daten abgeleitete Prognosen könnten darauf Antworten geben. Viele Unternehmen würden das heute auch einfach tun. Vor zehn oder zwanzig Jahren hätten das die ganz Großen auch schon gemacht. Dafür seien aber Spezialisten notwendig gewesen, die entsprechende Modelle entwickeln konnten. „Heute, da gebe ich Herrn Loth von Tableau recht, kann das jeder tun. Wer es nicht macht, ist tatsächlich abgehängt. Das ist Fakt“. Spotify scheint derzeit Weltmeister in der Kundendatenanalyse zu sein. Thomas Keil erzählte, die Nutzerdatenbank des Unternehmens sei 40 Petabyte groß. „Da ist kein MB und kein Kilobyte Musik drin. Das sind nur Trackingdaten zum Nutzerverhalten gespeichert. Welcher Nutzer hat

sich wann eingeloggt, welches Lied gehört...“ Spotify kenne die Vorlieben von allen Nutzern auf einer ganz detaillierten Datenbasis und werte diese aus. Das größere Team in dem Unternehmen sei nicht etwa das Rechte-Einkaufsteam, auch nicht die Technik, die Streamingverfahren entwickelt und bereitstellt, sondern „die Analysten, die herausfinden, welche Nutzer was im System machen und tun können“.

Verkaufsvorhersagen für eine Milliarde Produkte am Tag

Am Beispiel von Nestlé verdeutlichte der SAS Marketingmanager die Dimensionen, um die es bei der Optimierung im Warenverkauf geht. Nestlé verkauft pro Tag eine Milliarde Produkte unter 100.000 Marken. Dem Konzern stellt sich ununterbrochen die Frage: Welches Produkt verkaufe ich wann in welchem Markt in welcher Stückzahl, damit die Logistikkette entsprechend aufgebaut werden kann. Die Antwort geben Predictive Analytics, die aus Vergangenheitsdaten, Kalenderdaten, Produktionsdaten und weiteren Parametern errechnen, wann z.B. die Bio-Joghurts im Zwischenlager sein müssen, damit sie unter Berücksichtigung der Feiertage frisch mit noch 20 Tagen Haltbarkeit im Supermarkt in der Kleinstadt XY verfügbar sind. Die Analysen empfehlen tatsächlich auch noch den gerade geeigneten Verkaufspreis. „Je genauer die Modelle, desto besser werden die Abverkäufe“, so Keil.

Der unwahrscheinliche Tod des Truthahns

Warum die Ableitung von Vorhersagen aus Vergangenheitsdaten trotzdem nicht immer funktioniert, auch wenn die Modelle die Realität mathematisch einwandfrei abbilden und die Daten exakt analysiert werden, beschrieb Keil an einem Beispiel, das man nicht so schnell wieder vergisst. Der Truthahn würde jeden Tag seines Lebens spüren, dass er von seinem Bauern geliebt und ge-

häschtelt wird. Er wird täglich gefüttert, bekommt ordentlich zu fressen und das Futter wird im Verlauf der Zeit sogar immer besser. Bis er dann am Ende seiner Laufbahn geschlachtet wird. „Dieses Ereignis ist aus den in seinem Leben erhobenen Vergangenheitsdaten nicht vorhersehbar“. Bei Prognosen müsse immer unter Wahrscheinlichkeiten gerechnet werden. Keil ist trotzdem ein hundertprozentiger Verfechter von Predictive Analytics. „Niemand kann die Zukunft vorhersagen, egal wie viele Daten aus der Vergangenheit ich habe. Wir können aber ‚vorher sagen‘, was zukünftig passieren kann - und das immer besser“. Diese Fähigkeit, aus der Vergangenheit zu lernen, Hypo-

thetisch erzeugen kann, wirft ein weiteres Licht auf das Potential der Technologie. Keil berichtete von einem Telekommunikationsunternehmen, das neue Sendemasten bauen wollte, aber nicht wußte, wo anfangen. Daraufhin bestückte das Unternehmen vier Mobiltelefone mit jeweils einer SIM-Karte eines Mitbewerbers und eines mit der eigenen, machte das Ganze für fünf Sets und bestückte fünf Transporter damit. Dazu noch ein bißchen Technik und die Autos fahren los. Sie zeichneten im ganzen Land die Sendeleistung der verschiedenen Anbieter an den verschiedenen Orten auf. Danach konnte die Entscheidung, wo man mit den Investitionen am besten beginnt, auf

renden hat die Professorin im Wintersemester 2015/16 Twitter-Daten (Tweets) zu Politikerinnen und Politikern ausgewertet und die abgeleiteten Erkenntnisse mit dem ZDF-Politbarometer verglichen. Die Idee dahinter war, dass die Plattform Twitter vielfach für politische Diskussionen genutzt wird und Tweets sich durch das vorgegebene Format von 140 Zeichen und die häufige Verwendung von Hashtags für die Zuordnung zu einem Thema besonders für eine Sentimentanalyse „zu eignen scheinen“, wie sie sich ausdrückte. Für die automatische Analyse erstellte die Projektgruppe zunächst Wortlisten, lemmatisierte die Begriffe und baute Mechanismen zur Erkennung von Verstärkern und Negationen (z.B. durch „sehr“ oder „nicht“) in das Modell für den Twitter-Crawler und die Sentimentanalyse ein. Beides wurde im Rahmen der Forschungsarbeit implementiert. Über mehrere Wochen sammelte man damit Tweets zu zwei Politikerinnen und sieben Politikern, analysierte sie und visualisierte die Ergebnisse in Diagrammen. Zu den meisten Politikern waren die Meinungsäußerungen auf Twitter insgesamt vorwiegend negativ. Nur bei einem einzigen Politiker überstiegen die positiven Äußerungen auf Twitter die negativen. Der Vergleich mit dem ZDF-Politbarometer zeigte, das sich dieses nicht durch eine automatische Twitter-Analyse ersetzen lässt, es jedoch interessante Ergebnisse der automatischen Analyse gab. In den meisten Fällen konnten Meinungswechsel im Politbarometer (Anstieg oder Abfall der Beliebtheitskala) auch in den Twitterdaten nachgewiesen werden. Politische Ereignisse lassen sich zudem an der Visualisierung der Twitter-Auswertung direkt ablesen. Das Meinungsbild zu den beteiligten Politikern verändert sich direkt.

Als Grenzen des Ansatzes benannte Melanie Siegel, dass a) die automatische Klassifizierung notwendig, aber komplex ist, b) es gefälschte, angeblich unabhängige Meinungsäu-



Thomas Keil: „Aus der Vergangenheit zu lernen, Hypothesen zu bilden und dieses Wissen für künftige Entscheidungen zu nutzen ist seit eh und je die Grundlage menschlicher Evolution und zivilisatorischen Fortschritts.“

thesen zu bilden und dieses Wissen für zukünftige Entscheidungen zu nutzen, sei seit eh und je die Grundlage menschlicher Evolution und zivilisatorischen Fortschritts. Ob Wetter, Umsätze, Kundenverhalten, Kreditausfallrisiken, Kündigungen, Verkehr, Maschinenausfälle, Logistik, Produktionsstrecken, Netzauslastung, Therapien... der Phantasie zu Einsatzgebieten von Predictive Analytics seien kaum Grenzen gesetzt.

Sendemasten-Investitionsplanung mit fünf mal fünf Mobiltelefonen

Dass man „Vergangenheitsdaten“ zur Lösung einer Aufgabe auch ak-

fundiertem Wissen relativ leicht aufgesetzt werden.

Sentimentanalyse. Der Versuch, in Daten ausgedrückte Stimmungen zu erfassen

Bei den DGI-Praxistagen ist der Name Programm. Bei sechs der sieben weiteren Vorträgen des Tages standen aktuelle Entwicklungen und Problemstellungen aus dem Arbeitsalltag von Unternehmen sowie Erfahrungen im Sinne von Best Practice im Mittelpunkt. Der siebte Vortrag von Melanie Siegel beschäftigte sich mit einer wissenschaftlichen Untersuchung zur automatischen Erkennung von Trends. Mit einer Gruppe von Studie-

berungen gibt, c) das Social Bots, also Twitterautomaten, bei den Veröffentlichungen auf Twitter eingesetzt werden. Als letzten Punkt erwähnte sie unter d) dass man auch schwer bewerten könne, wie aussagekräftig ein Vergleich mit einer Meinungsforschung ist, die in der Krise steckt. Die Forschung zu Twitter-Analysen auf deutschsprachigen Tweets läuft nach Aussage der Informationswissenschaftlerin im Augenblick sehr stark an.

Ganz viel Wissen aus dem Arbeitsalltag des Analytics- und Forecast-Geschäfts

„Je mehr Muster, je mehr Konstanten in einer Zeitreihe sind, desto einfacher wird der Forecast“, erklärte Alexander Häußer vom Beratungshaus PricewaterhouseCoopers (PWC) in seinem Vortrag über prognostische Analysen für die Finanzwelt. Dort ist das Interesse an den neuen Verfahren sehr groß. Banken beschäftigen seit Jahrzehnten ganze Heerschaaren von Controllern und Analysten, die Daten und Kennzahlen intellektuell in Vorhersagen umwandeln. PWC berät die Finanzwelt zu Predictive Analytics und hat dazu mit der Predictive Analytics Suite (PAS) ein eigenes Softwaresystem entwickelt, das von PWC Digital Services in verschiedenen Vertriebsformen angeboten wird, unter anderem als Cloud-Service. „Vom Kunden selbst ist für den Umgang damit aber ein gewisse Expertise gefordert.“ Häußer stellte verschiedene Funktionen der Analyse und der interaktiven Ergebnisvisualisierung vor. Besonders interessiert ist die Finanzwelt an dem in der Suite bereits vorgesehene nächste Schritt nach den Predictive Analytics: den Prescriptive Analytics. Das sind Handlungsempfehlungen, die aus den Ergebnissen der prognostischen Analysen automatisch generiert werden. Dafür werden, so Alexander Häußer, Big Data, statistische Analysen, maschinelles Lernen und Unternehmensdaten verknüpft und ausgewertet. Aus dieser Daten-

grundlage werden generierte Empfehlungen für Maßnahmen zur Optimierung der Geschäftsgänge hergeleitet. Auf deutsch: Die Software schlägt den Verantwortlichen vor, was sie tun sollen.

800.000 Euro Kosteneinsparung durch Produktionsfehler-Vorhersage

Nicolas Versteegen, Trainer, Berater und Entwickler bei der StatSoft (Europe) GmbH, stellte in seinem Vortrag fünf Kundenprojekte vor, anhand derer er zeigte, welche vielfältigen Formen Predictive Analytics Projekte in der Praxis annehmen. Jedes Unternehmen stelle unterschiedliche Anforderungen bzgl. Transparenz, Automatisierbarkeit und Leistungsfähigkeit der Systeme und jedes Unternehmen bringe auch unterschiedliche Voraussetzungen mit, um die Analytics-Prozesse später im Betriebsablauf umzusetzen. Beides müsse individuell berücksichtigt werden. Nicolas Versteegen erklärte, dass deshalb unterschiedliche Methoden und Daten aus verschiedenen Quellen zum Einsatz kommen, auf die er kurz einging. Das Spektrum der vorgestellten Anwendungen reichte von der Verbesserung von Direct Mailings über die Konzeption eines Modells, das vorhersagt, wann in der Stahlproduktion voraussichtlich Zunderfehler auftreten werden, bis zur personalisierten Medizin. Bei großen Mailings konnten durch die Analyse und Optimierung bei dem betroffenen Kunden „mehrere 100.000 Euro an Druck- und Portokosten eingespart werden, ohne dass der Erfolg wesentlich geschmälert wird“. Die Fehlerprognose im Stahlwerk schlägt laut Versteegen mit 80% Reduktion der bisherigen Fehlerkosten zu Buche, was nach seiner Aussage etwa 800.000 Euro pro Jahr entspricht. Im medizinischen Bereich konnte das Risiko postoperativer Infektionen durch Gefährdungsvorhersage direkt im Operationssaal um 74% gesenkt werden. 400.000 US-Dollar Einsparungen und ganz viele

Patienten, denen zusätzliches Leid erspart blieb.

Kombinierte Vorhersagen aus strukturierten Zahlen und unstrukturiertem Text

Die meisten Predictive Analytics-Verfahren beziehen sich heute auf strukturiertes Zahlenmaterial und vernachlässigen dabei einen ganz großen Teil verfügbarer Informationen, zum Beispiel alle Studien und Berichte aus den letzten Jahren, erklärte Joachim Lackner in seinem Vortrag „Predictive Analytics unter Einbeziehung unstrukturierter Daten“. Er ist bei Expert System Deutschland verantwortlich für den Vertriebsbereich Financial Services. Das weltweit operierende Unternehmen bietet Semantic Intelligence Verfahren zur Dokumentenanalyse. Dabei werden aus dem Kontext der Dokumente durch automatisiertes Text Mining und Text Analytics relevante Information gefiltert und Vorhersagemodellen zugeführt, die in der Lage sind, hybride Datenbasen aus Zahlen und Text zu bedienen. Lackner stellte dem Auditorium die Vorgehensweise vor, ging auf Herausforderungen ein, die sich aus benutzten Wörtern und dem Satzbau ergeben und erklärte, wie diese durch Semantik gelöst werden können. „Semantik erschließt die DNA der Sprache“, beschrieb er plakativ. Mit den Verfahren von Expert System können Prozesse sehr stark optimiert werden, wie Lackners Kollege Stefan Geißler schon beim Kammingespräch am Vorabend am Beispiel einer Schweizer Bank und der amerikanischen Steuergesetzgebung dargestellt hatte. Gleichzeitig, so Lackner, trüge Semantic Intelligence auch noch zur Verbesserung der Vorhersagegüte bei.

Bigger Data, Faster Data, Crossed Data: Analyse wird zu Data Science

Thomas Gabriel, Chief Operating Officer (COO) für den Verkauf und den Geschäftsbetrieb von KNIME.com, begann seinen Vortrag „Advanced

Analytics Anwendungen in der Praxis mit KNIME Analytics Plattform“ mit der historischen Entwicklung der für Analysen verfügbaren Daten, Datenarten und -mengen. Er schloss die Erläuterung mit einer Vorhersage ab: Durch die allgemeine Entwicklung und das Internet der Dinge sei a) mit noch größeren Datenmengen - „Bigger Data“ - zu rechnen, zu vielen Daten, um sie in einem Datawarehouse speichern zu können, b) zudem würden die Datenflüsse schneller - „Faster Data“, und zwar zu schnell, um sie wirklich aufschließen zu können und c) rechnet er mit „ernsthaft he-



Der Online-Datenbankhost GBI-Genios hat Jahrzehnte Erfahrung in der Auswertung riesiger Datenmengen. Marketing- und Vertriebsleiter Ralf Hennemann kündigte in Frankfurt die neue Dienstleistung "Smart Data" an, mit der das Unternehmen in noch spezifischerer Weise als bisher aus den Informationspools Mehrwert ziehen will; auch aus unstrukturierten Textdaten wie z.B. Presseartikeln.

terogenen Daten“ - „Crossed Data“ - gekreuzt aus verschiedenen Domänen. Zu den Methoden für die Auswertung, die er ebenfalls in ihrer historischen Entwicklung vorstellte, erwartet Thomas Gabriel als nächstes erweiterte Algorithmen für Data Mining; mit Deep Learning würden gerade neuronale Netze wieder hochkommen und „Biosociative Knowledge Discovery“. Damit ist gemeint, dass viele Methoden entwickelt werden, die dann nach Bedarf verknüpfbar sind und verknüpft werden. Das

Buch⁵ dazu ist unter dem selben Titel 2012 bei Springer erschienen. Der Analysespezialist mit Promotion in Fuzzy Logik und Regelbasierten Systemen ordnet die von Gartner eingeteilten vier Spielarten der Datenanalyse - descriptive, diagnostic, predictive, prescriptive (die auf den DGI-Praxistagen von mehreren Vortragenden als kommender Service vorhergesagt wurden) - in zwei Blöcke: 1) Business Intelligence / Analytics und 2) Advanced Analytics / Data Science. Zu 1) zählt er die deskriptiven und diagnostischen Analysen, mit denen ermittelt wurde und wird, was geschehen ist oder gerade geschieht warum es wahrscheinlich geschah oder geschieht. Zu 2) prognostische Analysen und „Prescriptive Analytics“, die handlungsempfehlenden Analysen. Thomas Gabriel erklärte, dabei ginge es darum, von der Maschine zu erfahren, „Was können wir tun, damit das eintritt, was die Vorhersagen vorhersagen“. Auch er sieht am Horizont schon „Analytics for the Masses...“. Dateneinsicht würde für Jedermann zu jeder Zeit möglich werden. Analyseplattformen wie KNIME, Tableau und andere würden dies ermöglichen. Alexander Loth lieferte im Anschluss noch einmal die wichtigsten Informationen zu Tableau. Ausführlich vorgestellt hatte er die Analyseplattform für jedermann bereits im Workshop am Vortrag.

GBI-Genios bringt Genios Smart Data Services

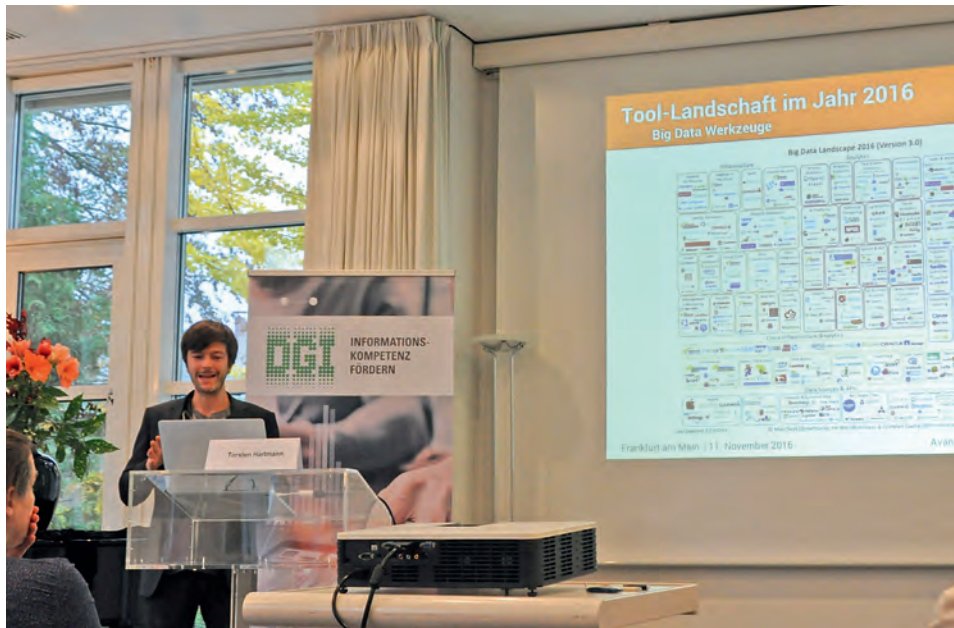
Als Pionier der Online-Informationsbranche hat GBI-Genios einen riesigen Datenpool aus lizenzpflichtigen Presseinhalten. Ideale Voraussetzungen für den Anbieter, der sich nun von der klassischen Recherche in Richtung Predictive Analytics bewegt. „Als Datenbankhost gehen wir ständig mit großen Datenmengen um und sind bestrebt, durch das Verknüpfen unterschiedlicher Inhalte einen Mehr-

wert für unsere Kunden zu schaffen. Künftig wollen wir das unter dem Titel ‚Smart Data‘ in noch spezifischerer Weise tun“, verkündete der Marketing- und Vertriebsleiter von GBI-Genios, Ralf Hennemann in Frankfurt. Das Unternehmen verbindet unter dem Label Smart Data seine Kompetenz in der Datenaufbereitung, -anreicherung und Content-Erschließung aus dem klassischen Geschäft mit der Auswertung von frei verfügbaren Daten (Big Data, Social Web, Internet). GBI-Genios will auf diese Weise aus Texten z.B. Stimmungswerte in der Presse und in Social Media erfassen. „Diese Art der Schaffung von Mehrwert erfordert das Erkennen von Signalen“, so Ralf Hennemann in seinem Vortrag ‚Identifikation schwacher Signale durch Auswertung externer Daten‘. „Was sind die Signale, damit man auf so etwas wie ein Investitionsvorhaben schließen kann?“, umriss er ein Anwendungsgebiet. Die Identifikation der richtigen Signalgeber und die Verknüpfung der Signale würden beispielsweise Vorhersagen von der ersten Idee für ein Bauvorhaben, entlang des Baufortschritts bis hin zur Fertigstellung möglich machen. Für Handwerker, die Installations- oder Malerarbeiten anbieten beispielsweise, seien dies wertvolle Informationen. Was sich aus dieser Verbindung der klassischen Online-Informationswirtschaft mit der neuen Predictive-Softwarewelt entwickelt, darauf darf man gespannt sein. Beide können sicherlich ganz viel voneinander lernen.

Semantische Magie mit Wortvektoren in der Datenwertschöpfungskette

Mit Thorsten Hartmann, Geschäftsführer der Avantgarde Labs GmbH in Dresden, klang der Vortragsreigen der 6. DGI-Praxistage mit einer fulminanten, tiefen Vermittlung neuesten Wissens zu Predictive Analytics mit Textdaten höchst unterhaltsam aus. „Die Tool-Landschaft für Big Data Analysen ist nicht mehr zu überblicken“ tröstete er gleich zu Beginn all

⁵ http://link.springer.com/chapter/10.1007/978-3-642-31830-6_2



Thorsten Hartmann, Geschäftsführer der Avantgarde Labs in Dresden, verkörpert die Generation der Digital Natives – und so selbstverständlich geht er auch an neue Methoden zur Erschließung von Daten zur Gewinnung von Information und Wissen heran. Er nennt es „semantische Magie“. Auf der Folie der Versuch, das unübersichtliche Angebot an Werkzeugen für die Datenanalyse irgendwie darzustellen. Die Grafik ist im Web.

jene, die sich neu mit dem Thema beschäftigen und sich von der Fülle der Analysewerkzeuge erschlagen fühlen. Auch die Experten von den Avantgarde Labs, zu deren Kerngeschäft die Beobachtung dieses Marktes gehört, kennen nicht mehr alle. Die Folie „Big Data Landscape 2016“⁶ (© Matt Turck, Jim Hao & FirstMark Capital) die Hartmann dazu zeigte, machte beim Draufgucken schwindelig. Sie ist als pdf im Internet verfügbar.

Thorsten Hartmann bestätigte Ralf Hennemann in dessen Bewertung der Textanalyse „Text als Wissensträger rückt immer mehr in den Fokus intelligenter Datenanalyse“. Der Grund sei, dass die jüngsten Entwicklungen in der Computerlinguistik und der künstlichen Intelligenz jetzt die automatische Erschließung und Verknüpfung großer Textmengen möglich mache. „Intelligente Algorithmen extrahieren Fachterminologie aus Texten, erzeugen geeignete Metadaten für Dokumente und verknüpfen diese mit existierenden Dokumentationssprachen und Wis-sensdatenbanken“. Höherwertige intellektuelle Aufgaben wie Klassifikation, Inhaltsanalyse und Verschlagwortung würden zunehmend von

Software übernommen.

Als krönenden Abschluss demonstrierte Thorsten Hartmann mit einem fröhlichen Schmunzeln, was die neuen Methoden und Werkzeuge können: „Semantische Magie mit Wortvektoren“. Man könne, erklärte der Experte, in einem semantischen Raum Vektoralgebra machen, was den Vektoren eine gewisse Semantik geben würde. Angewandt auf Textanalyse könne die Suchmaschine als Antwort auf die Frage „Was ist zu Frau wie Onkel zu Tante?“ die Antwort „Mann“ liefern. Oder auf die Frage „Was ist zu Prag wie Spree zu Berlin?“ Moldau antworten. Man muss Thorsten Hartmann recht geben: Es hat etwas von Magie.

3 Milliarden Wörter und 3 Millionen dreidimensionale Wortvektoren für die Forschung

Google hat seinen Google News Korpus für Forschung und Entwicklung freigegeben. Damit stehen laut Hartmann 3 Milliarden Wörter und drei Millionen 300-dimensionale Wortvektoren zur Verfügung.

Als praktische Tipps gab Hartmann den Teilnehmenden der 6. DGI-Praxistage mit auf den Weg, bei Analytics-Projekten

- den Fokus auf das Geschäftsproblem zu legen, nicht auf die Softwarewerkzeuge;

- klein zu starten und Anwendungsfälle mit unmittelbarem Nutzen zu wählen
- ein multidisziplinäres Projektsetup zu wählen, das Management und Fachabteilungen einbezieht, und dann
- Schritt für Schritt vorzugehen.

Damit gingen anderthalb spannende Tage zu Ende. **I**

I Die DGI wird auf dem 106. Bibliothekartag „Menschen – Medien – Märkte“ vom 30.5. bis 02.06. 2017 in Frankfurt als Aussteller vertreten sein.

Die nächste Konferenz der Gesellschaft ist das DGI-Forum Wittenberg 2017. Es steht unter dem Motto: „Am Anfang war das Wort – wer aber hat heute das Sagen?“. Termin 7. bis 9. September 2017 in Wittenberg.



Vera Münch

ist freie Journalistin mit Schwerpunkt Fachinformation und Wissensvermittlung

vera-muench@kabelmail.de

⁶ <http://mattturck.com/wp-content/uploads/2016/03/Big-Data-Landscape-2016-v18-FINAL.png>