Predictive Analytics

Paul Baumgarten

28.02.2019

Inhaltsverzeichnis

1	Einleitung	1
2	künstliche Intelligenz und Hype Zyklen	1
3	Der dritte Hype - Was ist neu?	3
3.1	Risikobewertung	3
3.2	Betrügerkennung	3
3.3	Frühwarnsysteme	4
3.3.1	Katastrophen	4
3.3.2	Krankheiten	5
3.4	Marketing	5
3.5	Gesichtererkennung	6
3.6	Autonomes Fahren	6
3.7	Diskriminierung	6
4	Diskussion	8
5	Fazit	8

Zusammenfassung

Diese Arbeit beschäftigt sich im groben mit künstlicher Intelligenz als Themenfeld, Anwendungsfällen von Predictive Analytics, Hypes und Winter. Es gibt einen Überblick über Subthemen und gewährt einen Einblick in Ethik und Software. Doch am Ende sitzt der Fehler immer vor der Maschine.

1 Einleitung

Predictive Analytics ist ein weit gefasster Begriff, der eine Auswahl von statistischen und analytischen Verfahren beschreibt, die Modelle repräsentieren und somit Aussagen über zukünftige Events oder Verhalten des Modells ermöglichen. [?]hitepaper [Nyce, Charles (2007), Predictive Analytics White Paper (PDF), American Institute for Chartered Property Casualty Underwriters/Insurance Institute of America, p. 1] Predictive Analytics ist daher dem Themenbereich künstliche Intelligenz zuzuordnen. Das Ziel ist es strukturierte oder unstrukturierte Daten also Informationen zu analysieren und somit zu Wissen zu konvertieren. [?]oe f hair Ein Beispiel dafür wäre eine Warenhauskette, die Kundendaten sammelt. Informationen wären dabei vor Allem zunächst die nicht identifizierenden Daten also Alter, Geschlecht, eventuelle Kleidergrößen und Ähnliches. Dem Gegenüber steht das zugehörige Kauverhalten. Dies sind zunächst Informationen, die so in der Art gesammelt werden könnten. Solange diese Informationen nur vorhanden sind, aus ihnen aber nichts marktstrategisches abgeleitet wird oder werden kann, handelt es sich wirklich nur um Informationen. Erst die Relationen zwischen den einzelnen Informationen größerer Datenmengen enthalten wirklich nutzbares Wissen. Um diese Relationen in großem Maße erstellen zu können, bedarf es großer Rechenleistung und entsprechender Algorithmen. Die Algorithmen sind schon länger vorhanden, die Rechenleistung ist erst im letzten Jahrzehnt so stark gestiegen, sodass mit Hilfe der statistischen Verfahren große Datenmengen verarbeitet werden können und damit die echten Probleme und nicht nur kleine Modelle lösbar geworden sind. rechenpower Dadurch sind die Kosten zur Wissensproduktion unter dessen Wert des gefallen. Es war also an der Zeit für einen neuen künstliche Intelligenz Hype

2 künstliche Intelligenz und Hype Zyklen

Im Gegensatz zu dem Hype um die meisten Innovationen lässt sich der um künstliche Intelligenz nicht komplett durch Gartners Hype Zyklus beschreiben. Zwar gab es zu den Anfängen der künstliche Intelligenz zunächst einen Gipfel der überzogenen Erwartungen gefolgt vom Tal der Enttäuschungen jedoch kam daraufhin der künstliche Intelligenz-Winter anstelle eines Plateaus der Produktivität. Diese Wintererscheinungen kamen nach dem nächsten künstliche Intelligenz-Hype wieder. Maßgeblich für diese verantwortlich sind so genannte „Setbacks“ also Events, die zum starken Rückgang des Interesses für künstliche Intelligenz geführt haben. Der erste Setback wurde durch den „Lighthill report“, einem Paper mit dem Titel „Artificial Intelligence: A Gernerall Survey“ von James Lighthill 1973 ausgelöst. Er resümierte, der erreichte Fortschritt der künstliche Intelligenz liege weit hinter den Erwartungen zurück. Der Report thematisierte das Versagen der künstliche Intelligenz in realistischen Modellen, also solchen, die Probleme der echten Welt zu lösen versuchten. Stattdessen wären nur kleine Modelle berechenbar und damit die Lösung nicht skalierbar. Das Hauptproblem war die kombinatorische Explosion, die bei vieldimensionalen Fragestellungen

auftreten[1]. Die Rechner von damals waren den großen Datenmengen bzw. Rechenaufgaben nicht gewachsen. Die künstliche Intelligenz hat zwar Anfang bis Mitte der 70er viele Fördergelder verloren, jedoch gab es immer noch Projekte und Forschung. So führten diese zu den sogenannten Expertensystemen, die später als erste Form erfolgreicher künstlicher Intelligenz identifiziert wurden (QUELLE). Sie arbeiteten mit Wenn-Dann-Regeln und Fakten um Entscheidungen anstelle eines Menschen zu treffen. (QUELLE EXPERT SYS) Diese Systeme haben einigen Firmen viele Millionen Dollar gespart und sorgten für einen neuen künstliche Intelligenz Hype. (AUSFÜHRBAR) Expertensysteme waren für ihre Anwendung spezialisierte Maschinen, samt Software und Hardware. Die wohl am besten bekannte war die Lisp Maschine und ihre zugehörige Sprache Lisp. In den späten 80ern wurden dann die Arbeitsrechner entwickelt, auf denen Lisp auch lief. Damit war die auf Lisp spezialisierte Hardware obsolet geworden und dessen Markt sowie der der Konkurrenzprodukte kollabierte. Mit dem Tod der Expertensysteme wurde auch wieder Gelder für die künstliche Intelligenz gestrichen. Bis Mitte der 2000er war der Ruf der künstliche Intelligenz bei Investoren noch schlecht.[2] Trotz der Setbacks und der Winter der künstliche Intelligenz ist das Feld nie wirklich verschwunden. künstliche Intelligenz wurde auch in den Wintern teils unter anderen Namen wie, Informatik, maschinelles Lernen, Analysen, Wissen-Basierte-Systeme usw. weitergeführt. Es war einfacher mit diesen Titeln Fördergelder zu bekommen. [3] Hinter diesen Themen stehen zum Teil heute erfolgreiche Technologien, die die Wenigsten der künstliche Intelligenz zuschreiben. Suchmaschinen, Übersetzungsprogramme, Logistikprogramme, Medizinische Programme zur Diagnoseunterstützung. Auch heute noch benutzen Automatikgetriebe eingebaute Fuzzylogik-Controller um den Gang auszusuchen oder für ein Doppelschaltgetriebe um zu (QUELLE)entscheiden, wann wie geschaltet werden soll. Einige der aufgezählten Technologien würde man möglicherweise im Jahr 2019 nicht mehr als Teil der künstlichen Intelligenz sehen. Dies liegt vermutlich an der unklaren Definition von künstlicher Intelligenz im Allgemeinen oder sogar an der umstrittenen Diskussion um Intelligenz als Konstrukt im Ganzen. Doch was ist aus der Geschichte zu lernen?

1. künstliche Intelligenz ist möglicherweise immer mit überzogenen Erwartungen verbunden
2. künstliche Intelligenz stirbt nicht, sondern wird im „Winter“ unter anderen Namen weiterbeforscht
3. Der Begriff wandelt sich, sobald man sich an die bisher mit ihm verbundenen Technologien gewöhnt hat.

Im Jahr 2019 befinden wir uns wieder in einem künstliche Intelligenz-Hype. Es gibt Verschwörungstheoretiker und Untergangsfantasien, in der uns die Maschine aus unserer Welt wirft. Es gibt auch solche, die sagen, die künstliche Intelligenz wird uns in kurzer Zeit in Allem überlegen sein. Was auch immer passieren mag, Vorhersagen haben im Feld der künstliche Intelligenz schon öfter enttäuscht als in den meisten anderen Gebieten. Aus diesem Grund wird hier exploriert, was in diesem

Hype neu ist, was die künstliche Intelligenz in dieser Phase auszeichnet. Was für neue „Buzzwörter“ gibt es?

3 Der dritte Hype - Was ist neu?

Ein wesentlicher Aspekt des aktuellen Hypes wird durch Technologien hervorgerufen, dessen theoretische Grundlagen schon länger existieren, dessen Umsetzung aber bislang an der Rechenpower scheiterten. So sind Statistische Verfahren wie Regressionen oder neuronale Netze und sogar große oder viele Entscheidungsbäume erst in den letzten 15 Jahren auf die Probleme anwendbar geworden. Diese Verfahren werden in der Domäne „predictive Analytics“ zusammengefasst. Doch in welchem Bereich werden Elemente aus „predictive Analytics“ angewendet - wo ist der Fortschritt? Welche Probleme konnten gelöst werden, welche sind neu dazugekommen oder mehr in den Fokus gerückt?

3.1 Risikobewertung

Versicherungen sind besonders abhängig möglichst präzise Risiken abschätzen zu können. Sie wollen den Kunden ein Angebot machen, das möglichst attraktiv ist, sodass dieser nicht zur Konkurrenz geht. Darüber hinaus ist das Ziel eines Unternehmens so auch einer Versicherung, Gewinn zu erwirtschaften. Daher muss das Angebot auch für die Seite der Versicherung profitabel sein. Je genauer die Risikoabschätzung, desto geringer das Risiko für die Versicherung Verluste zu machen, oder Kunden an die Konkurrenz zu verlieren. Predictive Analytics soll hierbei helfen. In dieser Domäne könnte beispielweise ein Kreditpunktesystem helfen, das durch statistische Verfahren einen Score erstellt. Als Datengrundlage können das Einkommen, alte Kredite, aktuelle Kredite und familiäre Situationen dienen. Je höher der Score, desto eher kriegt der Kunde einen günstigen Kredit. Ähnlich würden eine Autovermietung bzw. Versicherung arbeiten. Durch verschiedene Parameter zum Beispiel Punkte in Flensburg, Fahrpraxis (Jahre und künstliche Intelligenzometer) Auto würde ein Fahrerscore berechnet werden können, anhand dessen das Angebot justiert würde. Durch Verfahren von predictive Analytics können auch diejenigen Parameter identifiziert werden, die maßgeblich dafür geeignet sind, als Prädiktoren benutzt zu werden, die also einen großen Einfluss auf den Score haben.

3.2 Betrugerkennung

Nach dem Global Payment Report 2015 ist die Kreditkarte das am meisten verbreitete Zahlungsmittel der Welt. Daher rufen die Kreditkartentransaktionen auch Betrüger auf den Plan. Im Nilson Report 2014 wurden die Schäden durch Kreditkartenbetrug auf 16 Milliarden Dollar beziffert. Für 2020 werden die Schäden auf 35 Milliarden Dollar geschätzt. Dabei gehen Betrüger in der Regel auf drei verschiedene Arten vor.[4]

1. Konventionelles Stehlen oder Fälschen der Karte
2. Onlinebetrug zum Beispiel als Fakehändler
3. Händler, die Daten weitergeben und am Betrug mitverdienen

Die Abwehr gegen solchen Betrug kann auf verschiedenen Ebenen Ansetzen. Für die Bank am zugänglichsten ist, sich die einzelnen Transaktionen anzusehen. Die Lerndaten sind erheblich bei so viel Betrug und die Transaktionen dem Betrug klar zuzuordnen. Eine Software die durch maschinelles Lernen, also statistischen Verfahren anhand vieler Parameter erkennt welche Transaktionen betrügerisch sind und welche nicht, liegt auf der Hand. Des Weiteren ist es denkbar Fakehändler mit maschinellem Lernen aufzuspüren. Fakehändler geben in der Regel keine Ware aus. Sie unterscheiden sich daher in jedem Fall von echten Händlern und sind so zumindest theoretisch von ihnen zu unterscheiden.

3.3 Frühwarnsysteme

Die der Domäne predictive Analytics zugrunde liegenden Techniken wurden vermutlich schon länger für die Früherkennung genutzt. Aufwändig programmierte Software, die Daten auswerten und Monitoring zulassen, dass schließlich Menschen dazu verleitet Warnungen rauszugeben, war lange Standard. Heute gibt es Software, in denen der Mensch nur noch die Parameter bestimmt, die für die Prädiktion relevant sein könnten. Die tatsächlichen Prädiktoren wählt dann das Programm aus, und überwacht auch die Datenlage. Das Implementieren von Frühwarnsystemen ist daher deutlich einfacher und günstiger geworden. Es sind auch kaum noch Menschen für die Überwachung notwendig.

3.3.1 Katastrophen

In einer Studie zur Entwicklung einer intelligenten Anwendung zum Schutz vor Katastrophen unter Benutzung von Mikrokontrollern und Sensoren, die das Wasserlevel messen, um das Überflutungsrisiko berechnen zu können, wurden Techniken von predictive Analytics zur Berechnung des Risikos benutzt. [5]

Auch hier wurde ein Score berechnet, der das Risiko für eine Überflutung angibt und auf einer Website anzeigt. Überschreitet der Wert eine kritische Grenze, wird eine SMS versendet. Durch solche Frühwarnsysteme können viele Leben gerettet werden. Die Datenmengen übersteigen dabei die menschliche Verarbeitungskompetenzen, sodass dieses Problem besonders gut und effizient durch Maschinen lösbar ist. Die Studie hat für ihr Frühwarnsystem Raspberry Pis benutzt. Wir sind inzwischen also in einer Zeit angekommen, wo wir ausreichend Rechenkapazität haben. Dafür sind andere Probleme vermehrt in den Fokus gerückt. Der "Information Bias", Ethische Diskussionen und die Fehlbarkeit der Menschen, die den Input für die Maschinen definieren.

3.3.2 Krankheiten

In den letzten 10 Jahren stieg die Inzidenz von akuten Nierenschädigungen pro Jahr um etwa 10% an [6], sodass sich diese als ernstzunehmende Belastungen für das Gesundheitssystem entwickelt haben[7]. Die geschätzten Kosten zur Behandlung dieser Erkrankung inklusive ihrer Komplikationen in Deutschland liegen bei 1,5 Mrd. Euro pro Jahr [8]. Hinzu kommt, dass es oftmals zu einer Unterversorgung von Patienten mit akuten Nierenschädigung kommt. Nur circa 5% der erkrankten Patienten erhielten ein nephrologisches Konsil, 8 % der Fälle wurden weiterhin ambulant betreut [7]. Mit Hilfe von Frühwarnsystemen konnte insbesondere in Großbritannien die Zeitspanne von der Diagnose bis zur Therapieeinleitung, eine höhere Versorgungsqualität und eine höhere Überlebensrate von intensiv betreuten Patienten erreicht werden [9]. Außerdem konnten nephrotoxische Medikamente durchschnittlich drei Tage früher abgesetzt werden und eine deutlichere Nierenfunktionserholung erreicht werden[10] [11]. Zusätzlich konnten auch auf Ärzteseite positive Veränderungen ausgemacht werden: In der Studie von XY (25) waren mehr als 80 % der befragten Ärzte zufrieden mit der Anwendung des Künstliche Intelligenz Frühwarnsystems. Dieses meldet auf Basis eines Algorithmus, der die Diagnosekriterien der künstliche Intelligenzndney Disease berücksichtigt, Patienten mit einem bestimmten Serumkreatininanstieg an den behandelnden Stationsarzt, um eine möglichst frühe Diagnose zu ermöglichen. Hierfür ist es notwendig zunächst den Kreatininausgangswert des Patienten zu bestimmen. So kann mit Hilfe eines Frühwarnsystems die Patientensicherheit erhöht werden sowie die Häufigkeit auftretender Komplikationen reduziert werden.

3.4 Marketing

Der wohl berühmteste Fall für Marketingstrategien, die durch predictive Analytics gestützt sind, ist der Fall der amerikanischen Supermarktkette Target. Es fing damit an, dass zwei Mitarbeiter von Target den neuen Statistiker der Supermarktkette Andrew Pole fragten, ob es möglich sei, in Erfahrung zu bringen, ob eine Kundin schwanger ist, selbst wenn diese das nicht mitteilen wollte. [12] Der Wunsch das in Erfahrung zu bringen kam daher, dass Target quasi alles verkauft, Kunden jedoch nur für ein paar Teile zu Target kamen. So ein festgefahrenes Kaufverhalten ließe sich jedoch nur selten ändern. Die Geburt eines Kindes wäre so eine Gelegenheit. Da in Amerika die Geburten in der Regel veröffentlicht werden, wissen zu diesem Zeitpunkt aber schon alle Supermärkte Bescheid. Daher ist es dann bereits zu spät der Erste zu sein, der gezielte Werbung platziert. Target wollte daher, den Kundinnen bereits im zweiten Trimester der Schwangerschaft Werbung schicken. Dazu war es Notwendig möglichst früh von der Schwangerschaft zu erfahren. Also haben sie Accounts für jeden Kunden angelegt, der mit der Kreditkarte bezahlt, einen Coupon genutzt, ein Formular ausgefüllt, eine Mail für eine Zurückname oder den Kundenservice kontaktiert hat. Zu den möglichst breit angelegten Daten wurde anschließend das Kaufverhalten gesammelt.

Der Statistiker Pole lies einige Test laufen, um Muster in den großen Datenmengen erkennen zu

können und entdeckte, dass Frauen am Anfang des zweiten Schwangerschaftstrimesters besonders viel geruchsneutrale Seife kaufen. Auch festgestellt wurde, dass in den ersten 20 Wochen Nahrungsergänzungsmittel wie Zink, Magnesium und Kalzium nachgefragter waren.[12] Auf Basis solcher Erkenntnisse erstellte er dann einen Schwangerschaftsscore, der bis zur Geburt skalierte. So konnte er sogar einen ungefähren Geburtstermin feststellen. [12]

3.5 Gesichtererkennung

Im Jahr 2017 kam eine der ersten Kameras mit 4k Auflösung und integrierter Gesichtererkennung. Bis zu 10.000 Gesichter kann sie speichern und damit ohne Kommunikation mit einem Server erkennen. Das Gerät passt in eine Hand und kommt aus China von der Firma Dahua. Gesichtererkennung ist ein umstrittenes Thema. Erst im Dezember letzten Jahres(2018) eskaliert ein Streit zwischen dem Datenschutzbeauftragten der Stadt Hamburg, Johannes Caspar und Innensenator Andy Grote bezüglich der neuen Software zur Gesichtserkennung mit Hilfe derer G20-Randalierer erkannt werden sollen. „Caspar hält das Programm für bedenklich, da darin ohne Rechtsgrundlage die Gesichtsabdrücke Tausender Bürgerinnen und Bürger gespeichert seien.“[13] Aber nicht nur der Staat setzt Gesichtserkennungssoftware ein. Auch google und Facebook benutzen diese. Facebooks Gesichtserkennungssoftware soll besser sein, als ein Mensch. Schon vor der Einführung der Gesichtserkennung war es schwierig die Kontrolle über seine eigene Privatsphäre zu behalten, wenn man Mitglied bei Facebook war. Bekannte luden Fotos hoch und verlinkten einen eventuell, sodass jeder, der Zugriff auf die Fotos von Facebookfreunden hat, was unter Umständen jeder Mensch sein kann, Zugriff auf Fotos, auf denen man selbst abgebildet ist, hat. Natürlich kann man Verlinkungseinstellung ändern oder Freunde bitten, Verlinkungen zu entfernen. Aber einmal mehr zeigt Facebook damit, dass man seine Privatsphäre aktiv schützen muss, wenn man das möchte und Facebook beigetreten ist.

3.6 Autonomes Fahren

Ein Prestigeprojekt der künstlichen Intelligenz ist das autonome Fahren. Bereits 2011 wurde definiert, was nötig ist, um autonomes Fahren in echter Umgebung realisieren zu können. [14]

In Kalifornien durften selbstfahrende Autos ohne Lenkrad und Pedalen ab April 2018 fahren.[15] Kaum ein Bereich, in dem künstliche Intelligenz genutzt wird, ist so im Fokus der öffentlichen Diskussionen wie das autonome Fahren. Auch in anderen Domänen laufen ethische Diskussionen ab. Jedoch nirgends sind diese so präsent wie bei selbstfahrenden Automobilen.

3.7 Diskriminierung

Im Jahr 2013 wurde ein Mann in Wisconsin, USA verhaftet, weil er ein Auto gefahren ist, dass erst kürzlich in eine Schiesserei verwickelt war. Der Mann wurde verurteilt, da er versucht hatte

einem Polizisten zu entkommen, und ein Fahrzeug ohne Erlaubnis des Halters gefahren ist. Keines dieser Verbrechen ist im Staat Wisconsin zwangsweise mit Freiheitsentzug zu bestrafen. Bei der Urteilsverkündung verwies der Richter auf das hohe Rückfallrisiko des Verurteilten, das durch ein Computerprogramm namens Compas berechnet wurde. [16]

Problematisch dabei ist, dass niemand außer dem Hersteller wirklich weiß, wie der Algorithmus arbeitet. Nur der finale Score am Ende wird bekannt. Der Verurteilte ging in Revision, da er nicht von einem Computer derart hart verurteilt werden wollte und verlor auch diesen Prozess. Seltsam dabei ist, dass ein Computerprogramm maßgeblich bei der Rechtsprechung beteiligt ist, obwohl weder Richter noch Staatsanwälte wissen, wie der Computer arbeitet. Bisher waren alle anderen Sachverständigen, Psychologen oder Experten anderer Themen deutlich durchsichtiger. Außerdem steht es einem Richter frei, den Score zu benutzen oder nicht. Dadurch erweitert sich das Strafmaß und es ist dem Richter einfacher anhand von moralisch problematischen Parametern zu entscheiden, zum Beispiel Geschlecht und Rasse.[16]

Dabei wird das letzte Argument vor Allem für die Verwendung eines Computerprogramms verwendet. Dem Programm schreibt man eine Vorurteilsfreiheit zu, eine Neutralität. In manchen Staaten ist eine Verwendung der Programme aus diesem Grund sogar Pflicht. Aber die Verantwortung über das Urteil auf einen Computer zu übertragen muss nicht dazu beitragen, dass Vorurteile verschwinden. Sie werden unter Umständen nur verschoben. Programme wie COMPAS entscheiden auf Basis der Daten, die ihnen gegeben werden. Das heißt, COMPAS wurde mit Daten gespeist, die bereits mit Vorurteilen belastet waren. Der Algorithmus bekam historische Urteile und Rückfalleinschätzungen von Richtern, also Menschen. Die Vorurteile und Muster die diese Richter angewandt haben, reproduziert COMPAS. Zusätzlich arbeiten diese Programme oft mit einem Feedback loop also einer Schleife, die dessen Ergebnisse zurückspeist, um den Algorithmus zu verbessern. Diese Technik ist an anderer Stelle schon dramatisch schiefgegangen. Googles Automatische Bildbeschreibung hat Menschen afrikanischer Herkunft als Gorillas betitelt, Microsofts Twitter Bot wurde schnell ein rassistischer Idiot.[16]

Die Fähigkeit eines Richters zu spüren, ob ein Angeklagter höhere Chancen hat rückfällig zu werden und daher länger ins Gefängnis sollte oder durch die längere Abwesenheit vom Umfeld, Familie und Job wahrscheinlicher rückfällig würde, erreichen diese Programme in der Regel nicht.[16] Und selbst wenn, bekäme es niemand mit, da die Öffentlichkeit keine Kontrolle, ja nicht mal Informationen über die genauen Arbeitsweisen hat (welche Features sind entscheidend?).

Es ist weder zielführend noch richtig das Problem im Konzept oder der Anwendung solcher Programme im Ganzen zu sehen. Diese Programme haben viel Potenzial. Nur muss man bei der Verwendung drauf achten, neutrale und nicht durch Menschen vor-beeinflusste Daten zu benutzen. Nähme man bei diesem Beispiel nur die Daten der tatsächlichen Rückfälle und nicht die Scores der Richter von früher, so hätte man schon mal einen Vorurteilsfreieren Algorithmus. Da aber insbesondere in den USA deutlich mehr Afroamerikaner prozentual zu ihrem Anteil in der Bevölkerung straffällig

werden, würde sich das vermutlich auch im Algorithmus widerspiegeln. Auch da könnte ein maßgebliches Feature die Hautfarbe sein. Selbst wenn man diese als Parameter manuell ausschließen würde, fände der Algorithmus Mediatorvariablen, über die die Hautfarbe dennoch einfließen könnte. Es geht schließlich nicht um Kausalitäten, sondern um Korrelationen.

Das was die Software damit allerdings vor Allem bestätigen würde, ist dass die Chancengleichheit eben definitiv nicht gegeben ist. Es reicht also nicht, sich mit dem Prozess, inwiefern die Hautfarbe im Rechtssystem bei Verurteilungen eine Rolle spielt, zu beschäftigen, man muss auch darüber nachdenken, wie man echte Chancengleichheit etablieren kann, nur dann würde der Algorithmus auch zeigen, dass die Hautfarbe einen Menschen eben nicht kriminell macht.

4 Diskussion

Ein großer Punkt der in öffentlichen Diskussionen selten Beachtung findet ist, dass wer Datenmüll in Programme einspeist auch immer Müll zurückbekommt. Eben ausgeführte Themen wie der Information Bias oder Gesichtserkennung, die nur mit weißen Männern Mitte 30 trainiert wurden und daher andere nicht erkennen, autonome Fahrzeuge dessen Teststrecken sich stark vom echten Verkehr unterscheiden zeigen den Effekt der durch schlecht ausgewählte Daten auftritt. Selbst wenn technisch alles richtig gemacht wurde, so kann man auch über die Anwendung bestimmter Technologie reden. Wie weit dürfen Firmen mit ihren Marketingstrategien gehen? Wie ist das mit Daten anderer die wir teilen? Wie können Kinder geschützt werden ohne dass sie ahnungslos und technologiefremd auf die Selbstverantwortung losgelassen würden? Auch kriminelle werden predictive Analytics oder andere KI Domänen benutzen, wenn es ihnen Hilft an Geld zu kommen. Wie wird ein Auto entscheiden, wen es im Zweifel überfährt? Wer entscheidet, wie das Auto sich entscheiden soll und wer sollte es lieber tun? Es gibt noch viele ungeklärten Fragen, doch die Technologien halten nicht still.

5 Fazit

Viele Anwendungsgebiete gibt es schon für predictive Analytics und einige werden vermutlich noch kommen. Man kann hoffen, dass der Hype nicht so groß wird, dass sich wieder ein Winter anschließt. Ein natürlicher Gegenspieler von Technologie ist die Skepsis. Außer den überbegeisterten KI-Fans, gibt es auch noch die Untergangverkünder, die komplett gegen alles mit KI sind und nicht von dem "Teufelszeug" hören wollen. Sicherlich bringen Technologien Gefahren mit sich. Es ist daher an uns den Einsatz und die Grenzen aktiv mitzugestalten, uns an Dinge zu gewöhnen, sie zu integrieren. Fortschritt lässt sich nicht aufhalten, wenn wir nicht auf den Zug springen, fahren die USA und China uns möglicherweise davon. Die Domäne predictive Analytics hat großes Potenzial, ist letztendlich aber auch nur eine Sammlung von Tools. Und wie Ron Weinstein schon sagte:

“A fool with a tool is still a fool“

“Ein Idiot mit einem Werkzeug ist immer noch ein Idiot“

Literatur

- [1] James Lighthill. Artificial intelligence: A general survey. 1973.
- [2] Alex Castro. Are u talking to me. 2008.
- [3] John Markoff. Behind artificial intelligence, a squadron of bright real people. *New York Times*, 2005.
- [4] Saravanan Sagadevan Ong Shu Yee and Nurul Hashimah Ahamed Hassain Malim. Credit card fraud detection using machine learning as data mining technique. 2017.
- [5] JonathanM. Caballero MichaelM. Orozco*. Smart disaster prediction application using flood risk analytics towards sustainable climate action. 2018.
- [6] statistisches Bundesamt. 2012.
- [7] akute nierenschädigung.
- [8] Matthews B Kerr M, Bedford M. The economic impact of acute kidney injury in england. *Nephrol Dial Transplant* 29, page 1362–1368, 2014.
- [9] Neal EB McCoy AB, Cox ZL. A real-time pharmacy surveillance and clinical decision support to reduce adverse drug events in acute kidney injury: a randomized, controlled trial. *Am J Kidney Dis* 56, page 221–238, 2012.
- [10] Moiz A. Balasubramanian G, Al-Aly Z. Early nephrologist involvement in hospital-acquired acute kidney injury: a pilot study. *Am J Kidney Dis* 57, page 228–234, 2011.
- [11] Steurbaut K. Colpaert K, Hoste EA. Impact of real-time electronic alerting of acute kidney injury on therapeutic intervention and progression of rife class. *rit Care Med* 40, page 1164–1170, 2012.
- [12] Charles Duhigg. How companies learn your secrets. *New York Times*, 2012.
- [13] Franziska Coesfeld. G20: Streit um Gesichtserkennung eskaliert. 2018.
- [14] Jan Becker Jennifer Dolson David Held Soeren Kamm el J. Zico Kolter Dirk Langer Oliver Pink Vaughan Pratt Michael Sokolsky Ganymed Stanek David Stavens Alex Teichman Moritz Werling Jesse Levinson, Jake Askeland and Sebastian Thrun. Towards fully autonomous driving: Systems and algorithms. 2011.
- [15] Reuters Zeit Online, AFP. Autonomes fahren: Kalifornien lässt selbstfahrende Autos ohne Lenker zu. 2018.
- [16] Ellora Thadaney Israni. *New York Times*, 2017.