„Transparente“ Algorithmen in Staat und Privatwirtschaft?

22.9.2016

Prof. Dr. Katharina A. Zweig

TECHNISCHE UNIVERSITÄT
KAISERSLAUTERN

Kurze Anmerkung

„Im Bereich der Computer bedeutet der Begriff „transparent“ nicht so sehr „durchsichtig“ sondern eher „unsichtbar“ oder „unbemerkt“. Transparente Computerprogramme oder –prozeduren sind typischerweise solche, die der Nutzer nicht bemerkt. Es ist eine wünschenswerte Eigenschaft in Situtionen, in denen technisch nicht bewanderte Nutzer eher verwirrt wären, wenn sie die dahinterliegende Technik bemerkten oder gar mit ihr interagieren müssten.“

„In computers, transparent means something a little different than its general meaning of *having the quality of being easily seen through* , coming closer to meaning *invisible* or *undetectable* . Computer programs and procedures that are said to be transparent are typically those that the user is - or could be - unaware of. Transparency is considered to be especially desirable in situations where users that are not particularly technically inclined would tend to be confused by seeing or having to interact directly with programming components.”

http://whatis.techtarget.com/definition/transparent

Das kleine ABC der Informatik

Wann gefährden

Algorithmen,

Big Data und

Cünstliche Intelligenz

unsere Demokratie?

Wie sagt man die Rückfallrate eines Verbrechers voraus?

Predictive Policing

Vorhersagen, wann und wo Straftaten wahrscheinlich sind.

Predictive Policing

Ein **Algorithmus**
hat mir geflüstert,
dass Du **fast** ein Krimineller bist.
Dann komm mal mit!

Aber auch: Vorhersagen,
ob ein Individuum
straffällig werden könnte!

Beispiel USA:

1) Oregon
2) Andere Bundesstaaten

Big Data

- Big Data Methoden nutzen, z.B.:
  - Alter der ersten Verhaftung
  - Alter des Delinquenten (der Delinquentin!)
  - Finanzielle Lage
  - Kriminelle Verwandte
  - Geschlecht
  - Art und Anzahl der Vorstrafen
  - Zeitpunkt der letzten kriminellen Akte
  - ….
  - Aber nicht: die (in den USA eindeutig zugeordnete) ‚race‘.

Algorithmus

- Die Algorithmendesignerinnen und -designer müssen nun entscheiden, welche der Daten vermutlich mit „Rückfallwahrscheinlichkeit“ korrelieren.
- Dies sollte am besten in einer einzigen Zahl münden, so dass man direkt sortieren kann.
- Beispiel Formel:

$$\begin{aligned} & 3 * \text{ bisherige Verhaftungen} \\ & - 2 * \text{ Anzahl Tage seit letzter Verhaftung} \\ & + 3 * (\text{Wenn Mann, dann 1, sonst 0}) \\ & + 2,5 * (\text{ Wenn Raubüberfall, dann 1, sonst 0}) + \ldots \end{aligned}$$

Allgemein

$$
\begin{aligned}
& w_1 * \text{ bisherige Verhaftungen} \\
- & w_2 * \text{ Anzahl Tage seit letzter Verhaftung} \\
+ & w_3 * \text{ (Wenn Mann, dann 1, sonst 0)} \\
+ & 2_4 * \text{ (Wenn Raubüberfall, dann 1, sonst 0)} + \ldots
\end{aligned}
$$

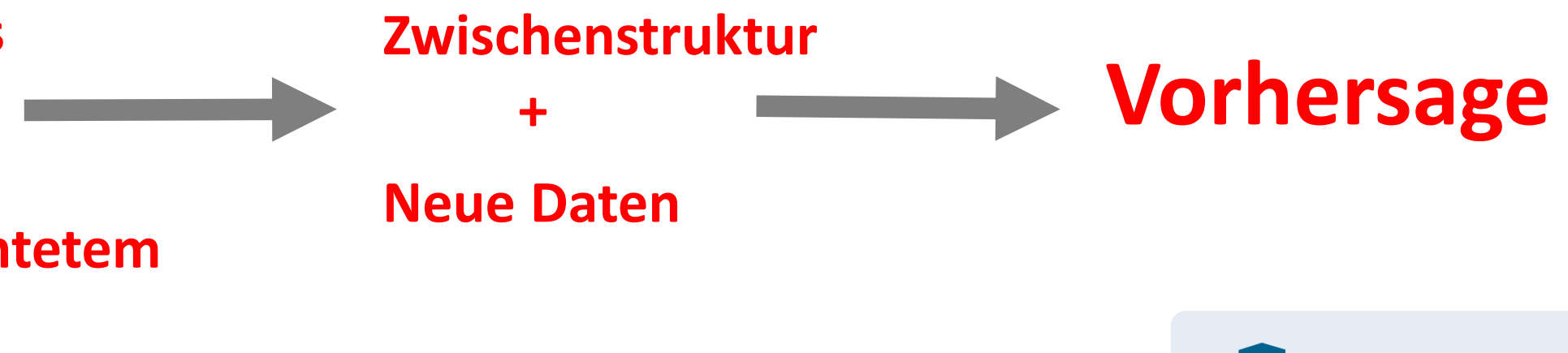
- Wer bestimmt die Gewichte so, dass möglichst die einen hohen Wert bekommen, die rückfällig geworden sind?
- Dazu bedarf es Algorithmen der künstlichen Intelligenz.

Künstliche Intelligenz

Lernende Algorithmen

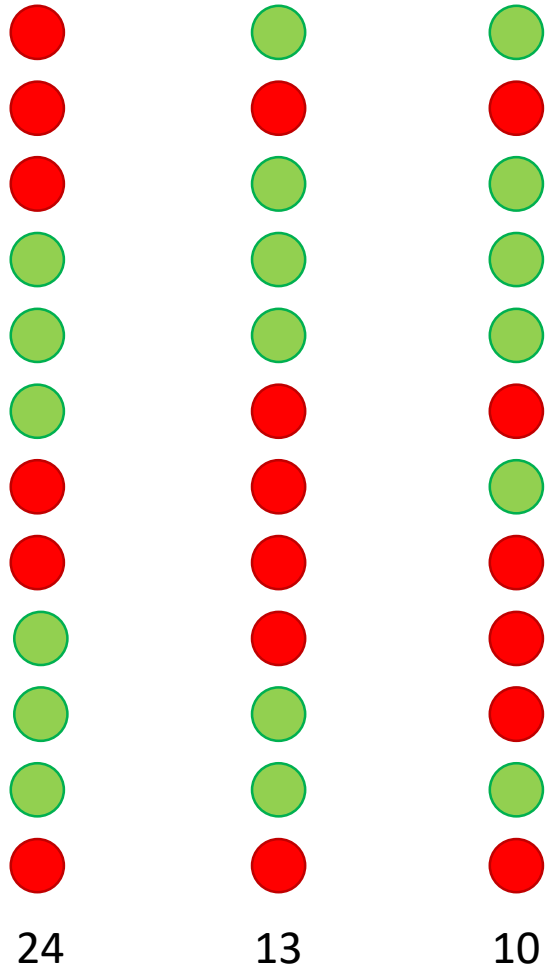
Künstliche Intelligenz

- **Problem:** gegeben eine Menge von bekannten Daten, finde Muster, die auf neuen Daten vorhersagen, wie sich etwas oder jemand verhalten wird.
- Algorithmus baut – basierend auf bekannten Daten – eine Zwischenstruktur auf, die dann Vorhersagen für neue Daten generiert.
- Der Algorithmus wird „auf den Daten trainiert“.



„Lernen“ von Gewichten

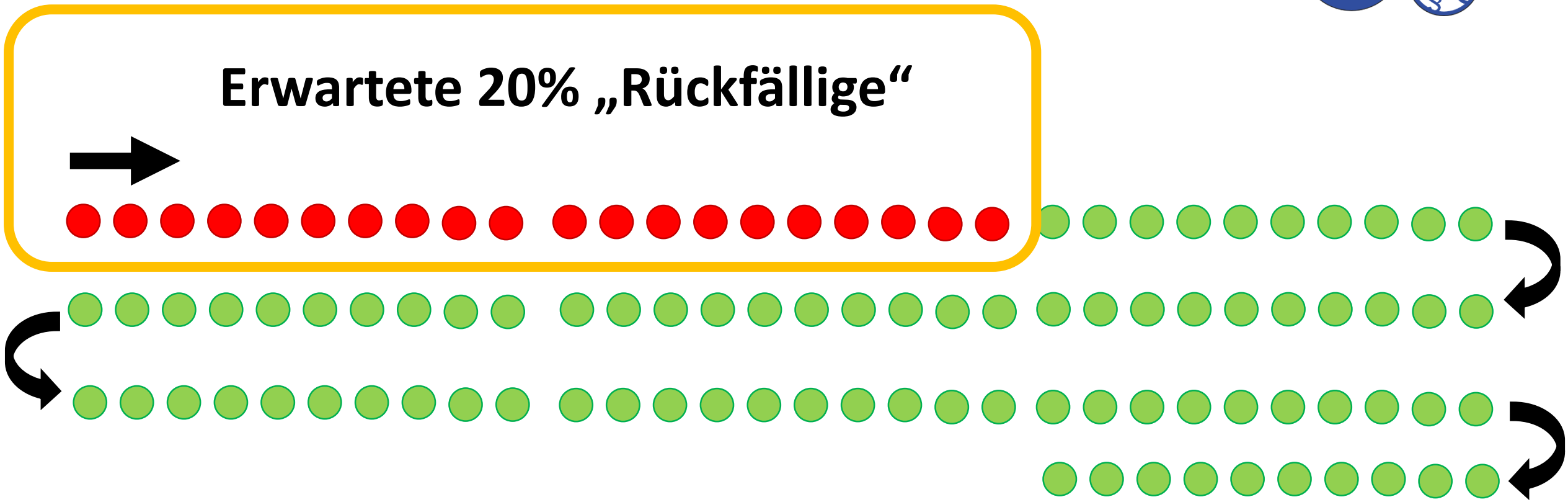
- Algorithmus probiert Gewichte
- Bewertet jeweils, wie viele bekannte Rückfällige möglichst weit oben stehen – für „alte“ Daten.
- Die Gewichtung, die das maximiert, wird für weitere Daten genommen.
- Kann im Wesentlichen für alles verwendet werden:
  - News Feed bei Facebook
  - Suchmaschinen
  - Produktempfehlung



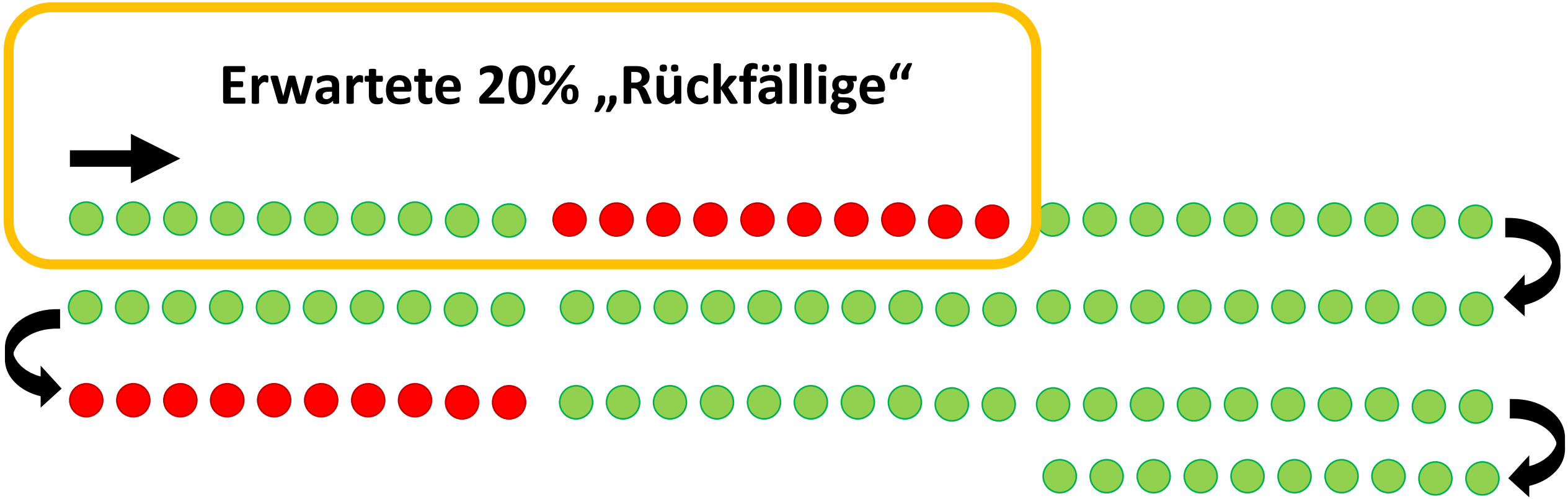
Oregon Recidivism Rate Algorithm

- Das oben genannte Qualitätsmaß dieses Algorithmus: 72 von 100 Paaren werden korrekt sortiert.
- Der in Oregon benutzte Algorithmus hat also, gegeben einen „Rückfaller“ und einen „Nichtrückfaller“, eine Chance von ca. 1:3 den Rückfaller höher zu gewichten als den Nichtrückfaller.
- Nur 28% aller so gemachten Prognosen sind falsch!
  - Das klingt doch ganz gut, oder?
- So werden aber keine Urteile gefällt!
- Problem: die Klassen sind ungleich verteilt!
  - 1000 Delinquenten
  - Ca. 200 werden rückfällig

Optimale Sortierung



Mögliche Sortierung eines Algorithmus mit dieser „Güte“ (ca. 70/100 Paaren)



Problem: Unbalancierte Klassen

- Bei optimaler Sortierung: die ersten 200 rot – keine Fehlentscheidung.
- Jetzt: nur die Hälfte!
- Damit 50% Fehlentscheidungen

Rückfallvorhersagealgorithmus ist rassistisch (Propublica)

- In einer Studie von Propublica (anderer Algorithmus) war die Quote noch schlechter:
  - Nur 20% der (vorhergesagten) Gewalttäter begingen eine Straftat
  - Bei allen möglichen Straftaten war die Vorhersage etwas besser als ein Münzwurf.
  - Bei schwarzen Mitbürgern war die Vorhersage immer zu pessimistisch;
  - Bei weißen zu optimistisch.
- Northpoint Software ist eine Firma, der Algorithmus ist unbekannt.
- Rasse ist an sich keine Variable des Algorithmus…

https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

Zweig‘sche Regel

Algorithmen der künstlichen Intelligenz werden da eingesetzt, wo es **keine einfachen Regeln** gibt.

Sie suchen **Muster** in hoch-verrauschten Datensätzen.

Die Muster sind daher grundsätzlich **statistischer Natur.**

Versuchen fast immer, eine **kleine Gruppe** von Menschen zu identifizieren (Problem der **Unbalanciertheit**)

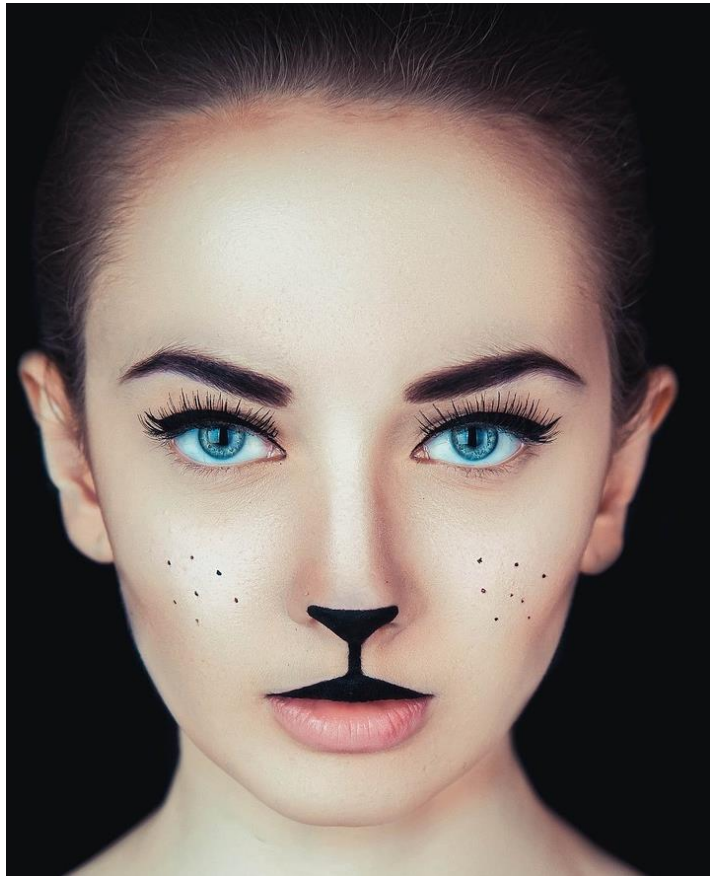
Wenn es **einfache Regeln zur Entscheidungsfindung gäbe, wären sie uns schon bekannt.**

Statistische Vorhersagen über Menschen

Was bedeutet das eigentlich?

Zu 70% ein Krimineller….

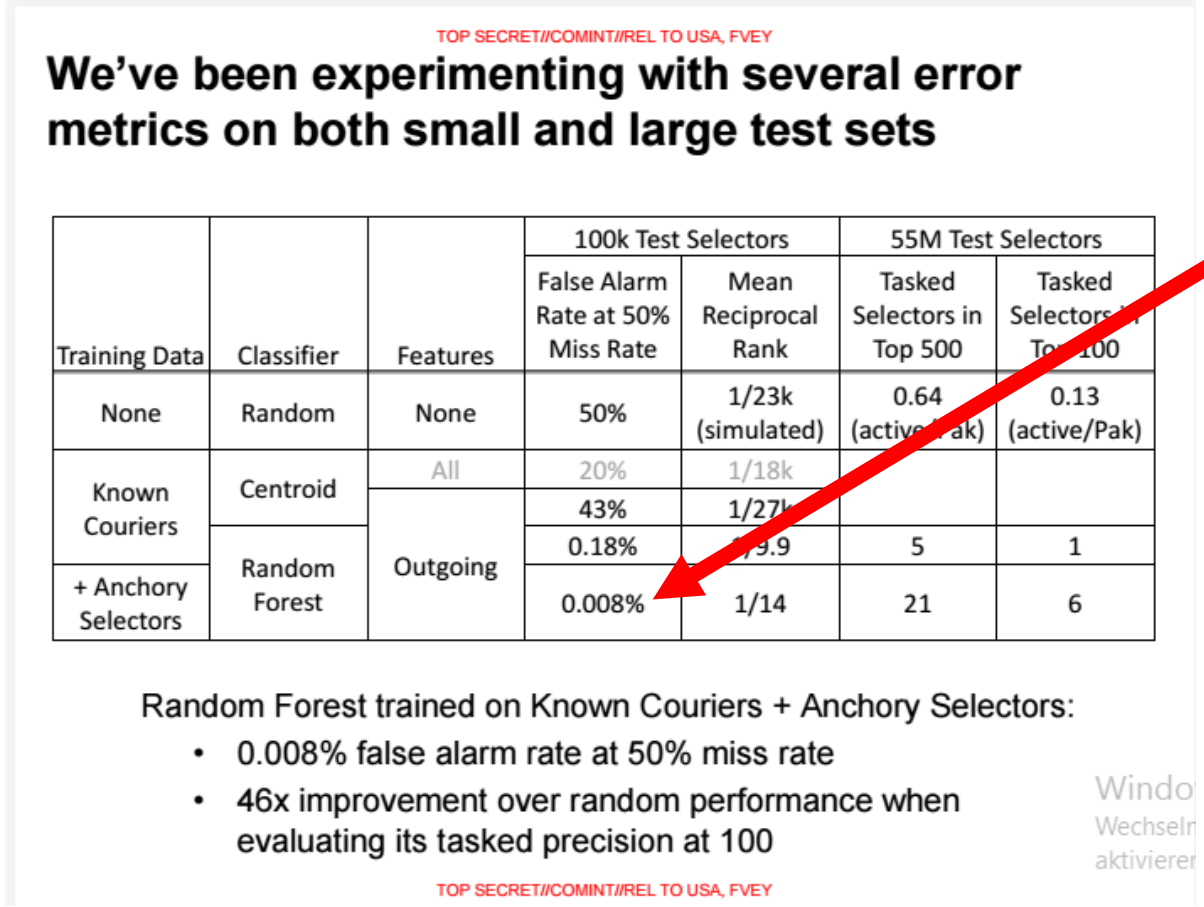
- Wenn dieser Mensch eine Katze wäre und 7 Leben hätte, würde er in 5 davon wieder rückfällig werden…
- Nein!
- **Algorithmische Sippenhaftung**
  - Von 100 Personen, die „genau so sind wie dieser Mensch“, werden 70 wieder rückfällig;
  - Mitgefangen, mitgehangen;
  - In einer dem Delinquenten (der Delinquentin) völlig unbekannten, algorithmisch bestimmten „Sippe“.



Probleme

- Aufmerksamkeitsökonomie der Richter und Richterinnen.
- „Best practice“ erfordert Nutzung der Software.
- Eine Nichtbeachtung der Empfehlung und gleichzeitige Fehleinschätzung wirkt viel schwerer als eine Beachtung der Empfehlung.
- Grundlegende Modellierung und Datenqualität kann schlecht sein.
- Der ins Gefängnis geschickte Delinquent **kann die Vorhersage prinzipiell nicht entkräften!**
  - Dies gilt auch für: Kreditvergaben, Bildungsangebote, Jobs, Personen, die von Drohnen erschossen werden oder als Terrorist eingesperrt werden, …

Terroristenidentifikation SKYNET



TOP SECRET//COMINT//REL TO USA, FVEY

We've been experimenting with several error metrics on both small and large test sets

Training Data	Classifier	Features	100k Test Selectors		55M Test Selectors	
			False Alarm Rate at 50% Miss Rate	Mean Reciprocal Rank	Tasked Selectors in Top 500	Tasked Selectors in Top 100
None	Random	None	50%	1/23k (simulated)	0.64 (active/Pak)	0.13 (active/Pak)
Known Couriers	Centroid	All	20%	1/18k		
		Outgoing	43%	1/27k		
	Random Forest		0.18%	1/9.9	5	1
+ Anchory Selectors			0.008%	1/14	21	6

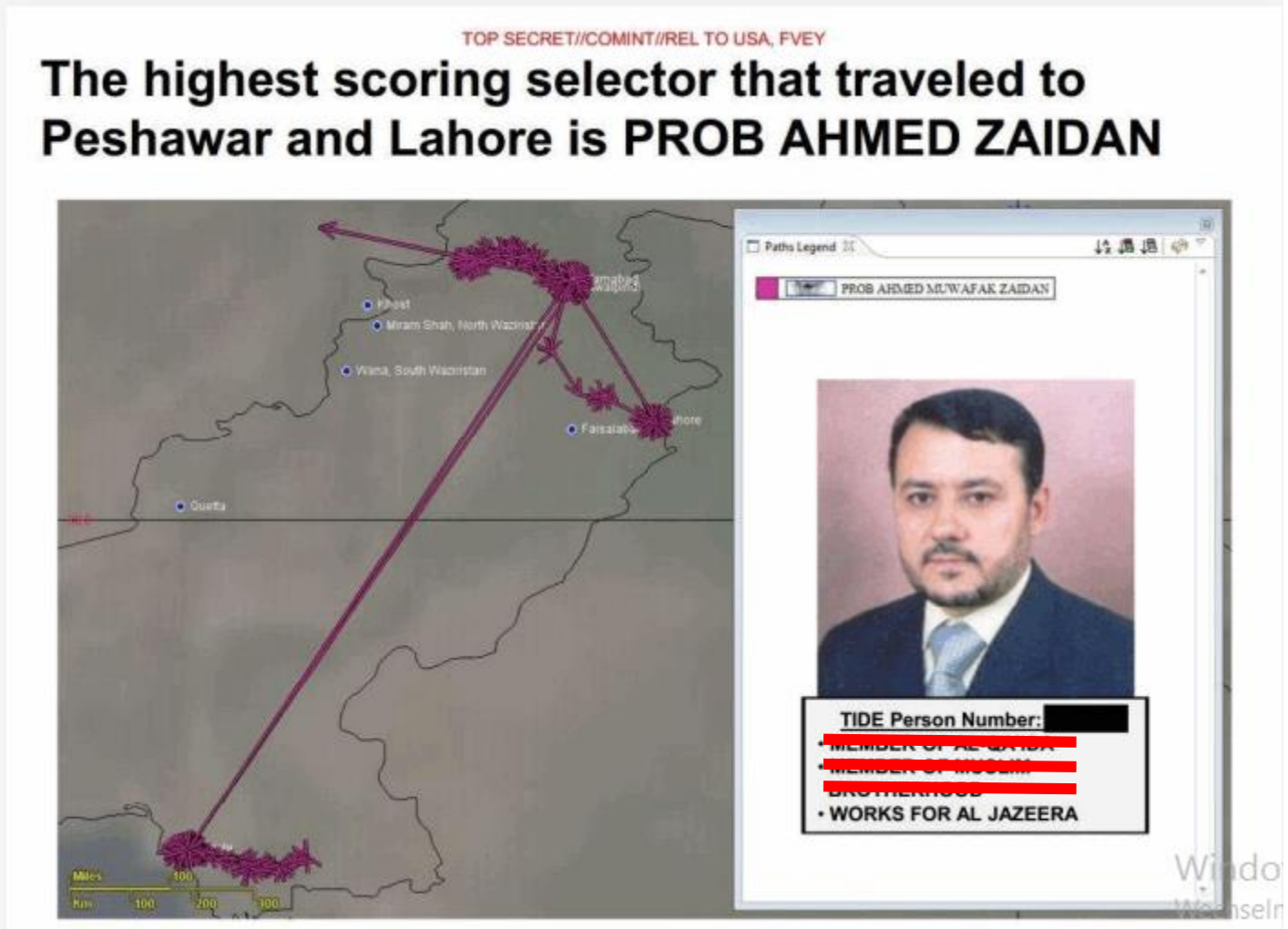
Random Forest trained on Known Couriers + Anchory Selectors:

- 0.008% false alarm rate at 50% miss rate
- 46x improvement over random performance when evaluating its tasked precision at 100

TOP SECRET//COMINT//REL TO USA, FVEY

https://theintercept.com/document/2015/05/08/skynet-courier/
https://theintercept.com/2015/05/08/u-s-government-designated-prominent-al-jazeera-journalist-al-qaeda-member-put-watch-list/

Top-“Kurier“ der Terroristen laut Algorithmus ist…



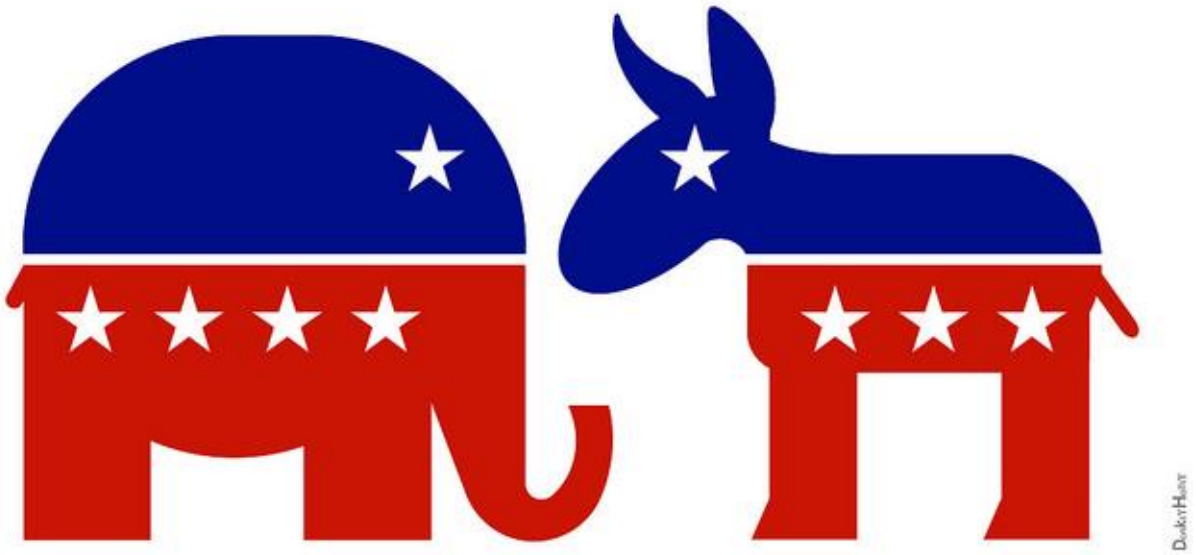
Spielkampsche Regel

Alle Algorithmen sind objektiv

Bis auf die von Menschen gemachten!

Können uns Algorithmen in unserer Meinung beeinflussen?

Bevorzugt Google Demokraten?



Studie von Trielli, Mussenden und Diakopolous[1]:

Unter 16 Präsidentschafts-kandidaten (USA) gab es bei Demokraten unter den ersten 10 Suchergebnissen 7 positive Berichte, bei Republikanern nur 5,9.

1 http://algorithmwatch.org/warum-die-google-suchergebnisse-in-den-usa-die-demokraten-bevorteile/

Sind wir beeinflussbar über Algorithmen?

- Suchergebnisreihenfolgen:
  - Manipulierte Suchreihenfolgen werden vom Nutzer nicht bemerkt und können die Tendenz eines unentschlossenen Wähler beeinflussen (Epstein & Robertson, 2015)
- Facebooks „Vote“ bzw. „Ich habe gewählt“-Button
  - Studie von Bond et al. über den Effekt auf das Wahlverhalten.
  - Effekt war klein, aber hochgerechnet ca. 60.000 mehr Wahlstimmen.

Epstein, R. & Robertson, R. E.: “The search engine manipulation effect (SEME) and its possible impact on the outcomes of elections”, Proceedings of the National Academy of Science, 2015, E4512-E4521

Bond, R. M.; Fariss, C. J.; Jones, J. J.; Kramer, A. D. I.; Marlow, C.; Settle, J. E. & Fowler, J. H.: “A 61-million-person experiment in social influence and political mobilization”, Nature, 2012, 489, 295-298

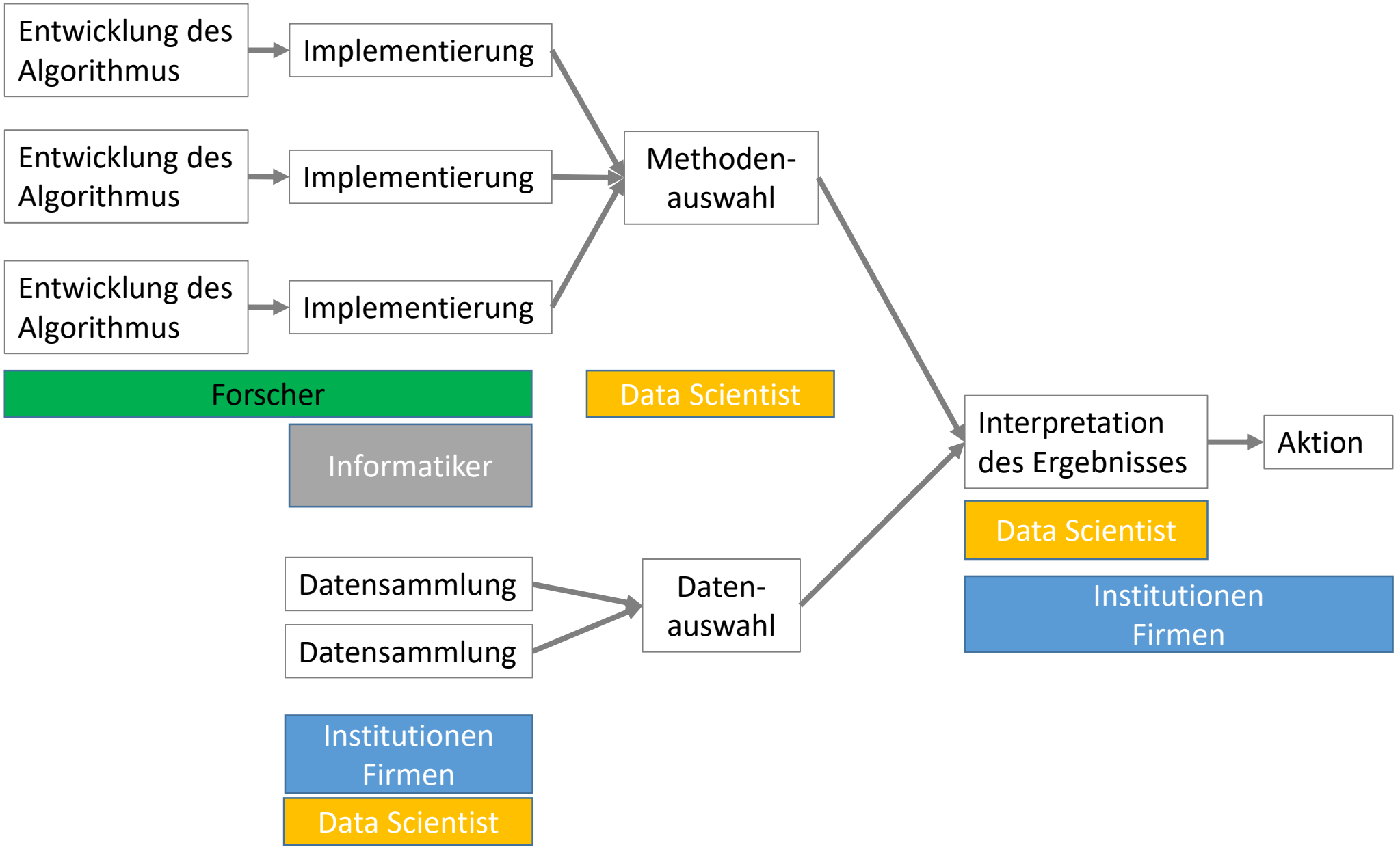
„Redirect Method“ by Google Jigsaw

- Könnten wir Suchmaschinen nicht au… Leute ‚umzudrehen‘, die anti-demokrati…
- Jigsaw sammelte Anti-ISIS-V… Tube,
- …identifizierte Suchw… ISIS-Interessierten stammen,
- …kreierte eine … pagne für ihren YouTube-Kanal mit dem gesammel…
- …und … ann an, wenn die oben genannten Stichworte kame…
- Sie erreichen mehr als 32.000 Interessierte, die sich ingesamt 500.000 Minuten Videomaterial ansahen.

Ist es das, was wir wollen?

Algorithmen in einer demokratischen Gesellschaft

Verkettete Verantwortlichkeiten



Wer überwacht die Auswir-kungen auf die Gesellschaft?

Medien?
Gesellschaft?
Politik?
Institutionen?
Firmen?
Recht?

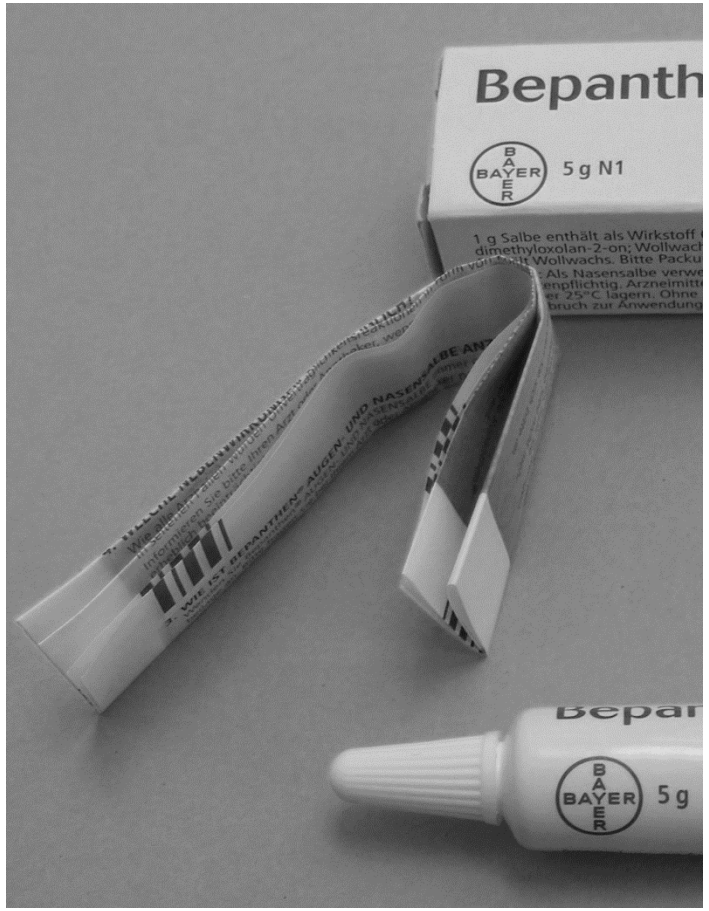
Quis custodiet ipsos algorithmos

Der „Automated Decision Making“-TÜV vulgo: „Algorithmen TÜV“

Notwendige Eigenschaften

- Unabhängige Prüfstelle mit Siegelvergabe
- Möglichst auch mit Forschungsauftrag
- Identifikation der **kleinstmöglichen Menge** an zu überprüfenden Algorithmen
  - Die meisten Algorithmen sind harmlos;
  - Produkthaftung ermöglicht, dass andere, z.B. Versicherungen, Interesse an korrekten Algorithmen haben;
  - Wettbewerb ermöglicht, dass andere ‚neutralere' Algorithmen anbieten.
  - Kein weiteres **Innovationshemmnis!**
- **Non-Profit**

Beipackzettel für Algorithmen



Welches Problem „kuriert“ der Algorithmus?

Was ist das Einsatzgebiet des Algorithmus, was seine Modellannahmen?

Welche „Nebenwirkungen“ hat der Algorithmus?

By User:Mattes (Own work) [Public domain], via Wikimedia Commons

Schlussformel

… zu Risiken und Nebenwirkungen der Digitalisierung befragen Sie bitte Ihren nächstgelegenen Data Scientist oder den deutschen Algorithmen TÜV.

Gründung von „Algorithm Watch“

Lorena Jaume-Palasí, Mitarbeiterin im iRights.Lab

Lorenz Matzat, Datenjournalist der 1. Stunde, Gründer von lokaler.de, Grimme-Preis-Träger

Matthias Spielkamp, Gründer von iRights.info, ebenfalls Grimme-Preis-Träger, Vorstandsmitglied von Reporter ohne Grenzen.

Prof. Dr. K.A. Zweig, Junior Fellow der Gesellschaft für Informatik, Digitaler Kopf 2014, TU Kaiserslautern

ALGORITHM WATCH

Kontaktdaten

Prof. Dr. Katharina A. Zweig

TU Kaiserslautern

Gottlieb-Daimler-Str. 48

67663 Kaiserslautern

zweig@cs.uni-kl.de

Algorithmwatch.org

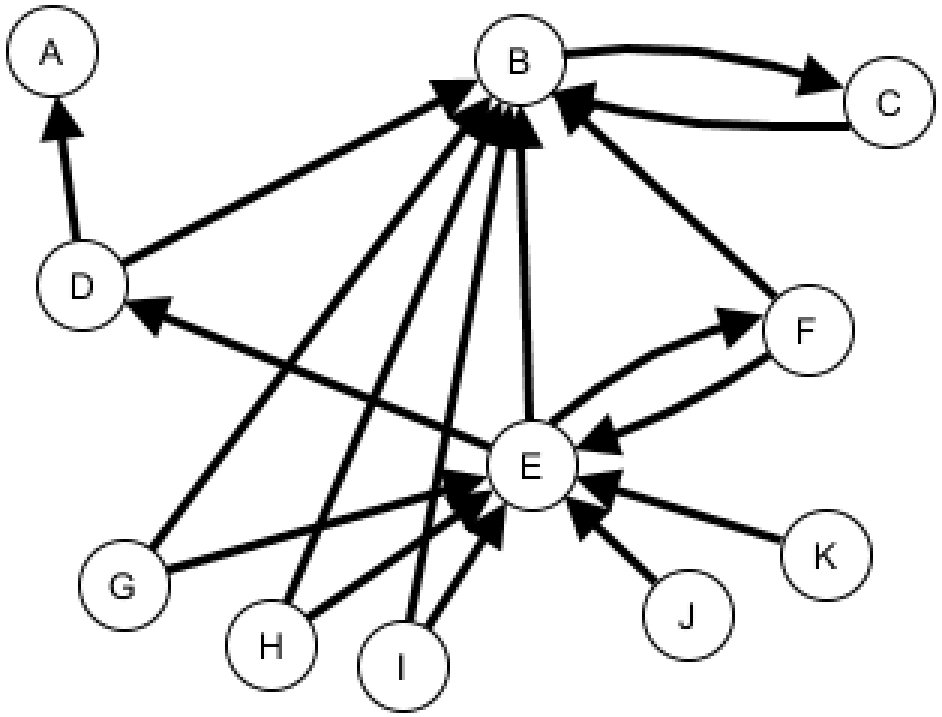
Komplexe Algorithmen

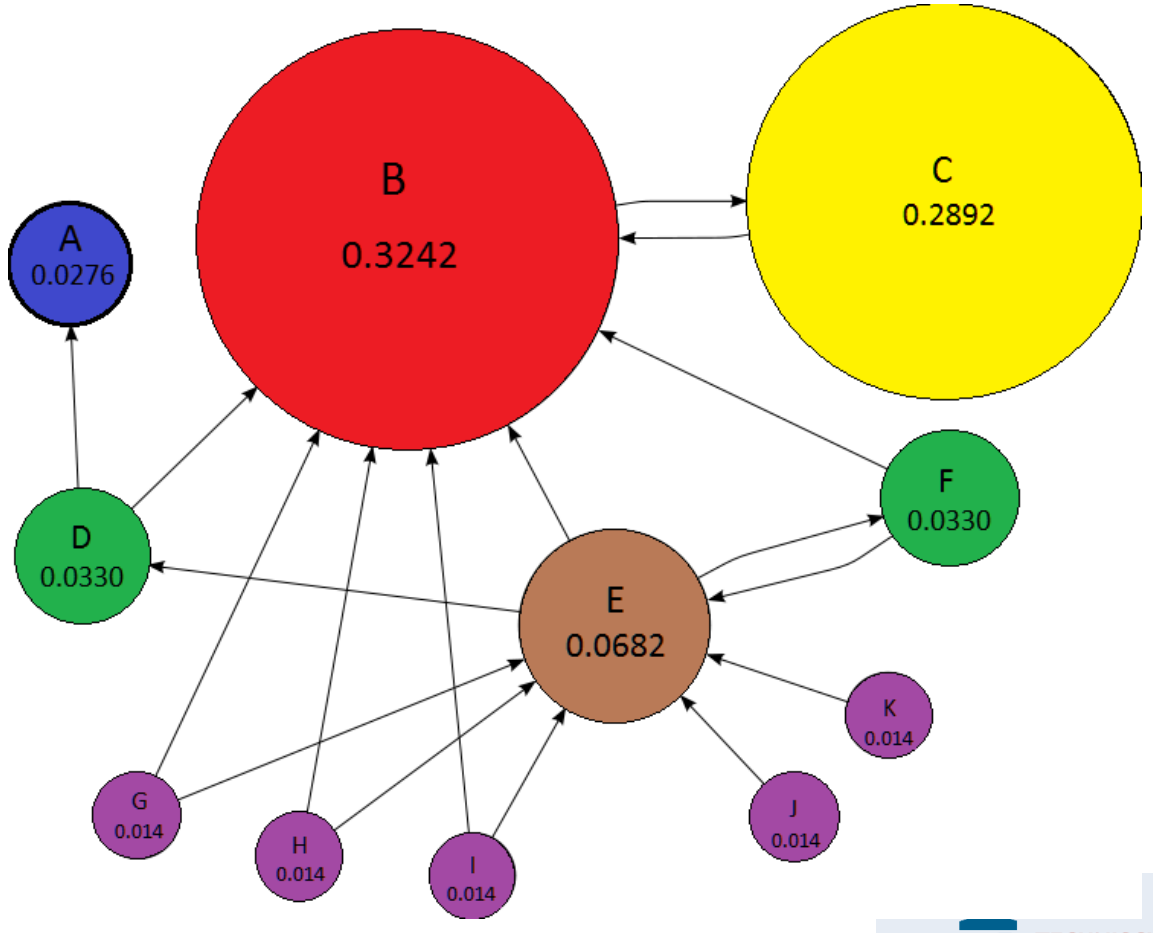
Können Suchmaschinenalgorithmen **objektiv** und **neutral** sein?

Suchmaschinen 101

1. Filtern aus allen ihnen bekannten Webseiten diejenigen, deren Text mit den angegebenen Suchbegriffen zusammenhängen.
2. Sortieren diese anhand:
   - Der Vernetzungsstruktur der Seiten untereinander
   - Dem Clickverhalten anderer Nutzer und Nutzerinnen bezüglich derselben Suche
   - Bei Personalisierung: auch nach dem eigenen, bisherigen Suchverhalten

PageRank





Idee hinter dem Algorithmus

https://commons.wikimedia.org/wiki/File:San_Diego_Comic-Con_2014_-_Silver_Surfer_%2814584916398%29.jpg, under CC:by-sa, by William Tung

Ein Modell menschlichen Verhaltens: der Random Surfer

- Ein Surfer klickt auf eine Webseite
- Folgt einem der Links auf der Webseite zufällig
- Von Zeit zu Zeit springt er auf eine völlig neue Webseite
  - Modelliert externes Wissen (z.B. Werbung, bekannte Seiten)

Modellierungsannahme

- Gibt nur dann **relevante und objektive** Ergebnisse, wenn Webseiten
  - Links auf ähnliche Seiten wie ihre eigene setzen,
  - Links auf relevante, meinungsangebende Seiten setzen, und
  - ihre Links **unabhängig** voneinander setzen.
- Unter dieser Bedingung ist der Algorithmus **neutra**l und gibt das kollektive Wissen der Welt nutzbringend weiter.
- Die Veröffentlichung des Algorithmus führte prompt zu Manipulationen seitens der Webseitenbetreiber.
  - Zu große Offenheit der Algorithmen ist manchmal **schädlich**.

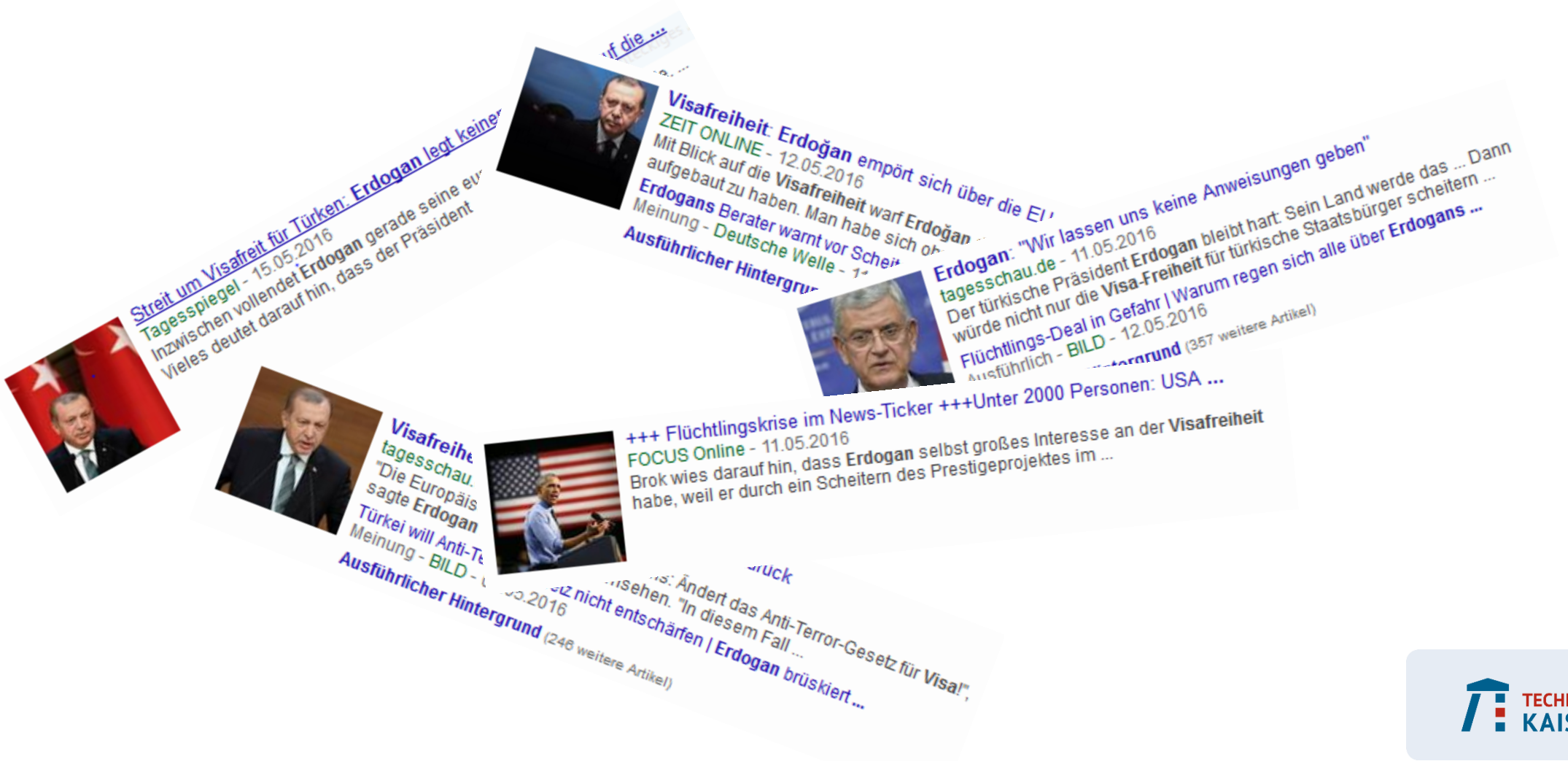
Relevanz – ein weites Feld

"A squirrel dying in front of your house may be MORE RELEVANT TO YOUR INTERESTS right now than people dying in Africa."

Mark Zuckerberg, CEO facebook,

nach David Kirkpatrick: „the facebook EFFECT“, Simon & Schuster New York, New York, USA, 2010, S. 181

Was ist die relevanteste Nachricht zur Anfrage: „Erdogan Visafreiheit“?



TECHNISCHE UNIVERSITÄT KAISERSLAUTERN

Big Data

- Wie kann Relevanz modelliert und „quantifiziert“ werden?
- Big Data Methoden nutzen, z.B.:
  - Sprache der Anfrage, Niveau der Anfrage, Wörter, Wortkombination
  - Tageszeit und geographische Informationen, Gerätetyp
  - Ihre bisherigen Suchanfragen und Ihr persönliches Klickverhalten
  - Welche Seiten wurden angeklickt, wie lange betrachtet, kam die Nutzerin wieder zurück zu den Ergebnissen?
  - Metadaten der Nachrichten/Medien: wann erstellt, durch wen, wo publiziert, Verschlagwortung, Wahl der Wörter
  - Verhalten anderer Nutzer, „liken“ auf sozialen Netzwerken, Interaktion mit Beiträgen

Big Data

Ganz allgemein:

- Große Datenmengen
- Außerhalb ihres spezifischen Zwecks genutzt
- Daher im Einzelnen vermutlich fehlerbehaftet
- Dank großer Masse und wenig individualisiertem Verhalten statistisch nutzbar

Frage + Big Data = mathematisches Problem?

Von der Schwierigkeit der Modellierung

Big Data + Frage

- Die Algorithmendesignerinnen und -designer müssen nun entscheiden, welche der Daten vermutlich mit „Relevanz“ korrelieren.
- Dies sollte am besten in einer einzigen Zahl pro Medium/Nachricht/Webseite münden, so dass man direkt sortieren kann.
- Beispiel Formel:

$$
\begin{aligned}
& 3 * \text{ bisherige Zugriffe} \\
- \ & \text{Anzahl Tage seit Publikation} \\
+ \ & \text{Beliebtheitsquotient des Publikationsortes} \\
+ \ & \text{Beliebtheitsquotient des Verfassers} + \ldots
\end{aligned}
$$

Allgemein

$$
\begin{aligned}
& w_1 * \text{ bisherige Zugriffe} \\
+\, & w_2 * \text{ Anzahl Tage seit Publikation} \\
+\, & w_3 * \text{ Beliebtheitsquotient des Publikationsortes} \\
+\, & w_4 * \text{ Beliebtheitsquotient des Verfassers} + \ldots
\end{aligned}
$$

- Wer bestimmt diese Gewichte, so dass insgesamt die „relevantesten“ (also die, die im Nachhinein am öftesten angeklickt werden) am weitesten nach oben sortiert werden?
- Dazu bedarf es Algorithmen der künstlichen Intelligenz.

Gnothi seauton! (Ganz ohne Computer)

- Kenne Dich selbst!
- Medienkompetenz heißt für mich zuallererst:
  - Welcher Konsum ist noch normal?
  - Wie wird man süchtig?
  - Wonach wird man süchtig?
  - Was heißt Manipulation?
- Journalistisches Verständnis?
  - Was heißt unabhängige Recherche?
  - Was ist eine gute Quelle, was ist ein Impressum?
- Kenntnis der Bürgerrechte und der Demokratie!
- Kenntnis von Massenphänomenen, z.B. Revolutionen, Shit-Storms

Chilon von Sparta:
„Erkenne Dich selbst!“

Computer! (Ganz ohne Computer)

- Vom Computer berechnete Lösungen suggerieren Objektivität.
- Wichtig:
  - Modellierung als subjektive Phase des Algorithmendesigns begreifen.
  - Unterschiede zwischen Mensch und Computer verstehen:
    - Riesiger zeitlicher Überblick
    - Statistiken
    - Ignorant gegenüber sozialen Konventionen und Kontexten
  - Personalisierung verstehen – intellektuelles Fast Food vs. Redaktionen

- Emergente Phänomene begreifen
  - Interaktion zwischen IT und Mensch

Algorithmen! (Ganz ohne Computer)

- Aufgaben aus dem Informatik-Biber und Informatik-Olympiade
  - Logik
  - Diskrete Mathematik
  - Erster Algorithmenentwurf
- Computer Science Unplugged (csunplugged.org), z.B.
  - Die Klasse als Simulation für einen Computer
  - Nachdenken über Algorithmen
- Was fehlt (Forschungsbedarf):
  - Ethische Dimensionen von Algorithmen
  - Didaktisch aufbereitete Algorithmen der KI