



Algorithmen und ihr Einfluss auf Wahlen

–

Ließ ein Algorithmus Trump triumphieren?

Klausurtagung der CDU RLP
17.8.2017

Prof. Dr. Katharina A. Zweig
TU Kaiserslautern, Algorithm Watch

Als das Rennen noch offen war....



Vor der Wahl: „Google manipuliert für Hillary“



- SourceFed behauptete, negative Suchvervollständigungen würden von Google bei Mrs Clinton unterdrückt¹.
- Trielli et al. finden mehr positive Suchergebnisse für demokratische Kandidaten als für Republikaner².
- Psychologe Prof. Dr. Epstein sah massive Manipulation bei Google und betonte wiederholt, dass bis zu 20% der unentschlossenen Wähler durch Suchmaschinen manipuliert werden können³.

¹ Das Originalvideo ist nicht mehr erhältlich (!, 23.3.2017), hier ist eine gekürzte Fassung: https://scontent-frt3-1.xx.fbcdn.net/v/t42.1790-4/13418283_277976322591285_62218208_n.mp4?efg=eyJ2ZW5jb2RlX3RhZyI6InN2V9zZC9&oh=417e220b5ef8a001cd2a38641aaa783b&oe=58D3C44E

² <http://algorithmwatch.org/warum-die-google-suchergebnisse-in-den-usa-die-demokraten-bevorzugen/>

³ Zum Beispiel hier, zu demselben Thema: <https://sputniknews.com/us/201609121045214398-google-clinton-manipulation-election/>

Nach der Wahl:



- Ach nee.
- Doch nicht.
- Dann war's Facebook:
- Filterblasen und Echokammern werden algorithmisch erzeugt, so dass wir in einer postfaktischen Welt leben¹.



Das kleine ABC der Informatik



Das kleine ABC der Informatik

Gefährden

Algorithmen,

Big Data und

Computerintelligenz

unsere Demokratie?



A wie Algorithmus

Ein Algorithmus ist eine genaue Handlungsanweisung,
die ein allgemeines Problem löst

Woher kommt der Begriff „Algorithmus“



1. Verallgemeinerung von Logarithmus, als Beispiel für einen komplizierten, mathematischen Prozess
2. Von altgriechisch ‚logos‘, dem „Wort“, Wurzel auch zur „Logik“
3. Von einem arabischen Mathematiker mit Namen Al Chwarizmi.

$\log(x)$

λογος

Al Chwarizmi

Ein Algorithmus ist....



...eine für jede **erfahrene Programmiererin** ausreichend **detaillierte und systematische Handlungsanweisung** oder **Lösungsvorschrift**, so dass bei **korrekter Implementierung** der Computer für jede **korrekte Inputmenge** den **korrekten Output** berechnet – in endlicher Zeit.

Eine **Implementierung** ist die Übersetzung der für den Menschen verständlichen Handlungsanweisung in eine Programmiersprache.

Beispiel: Berechnung und Zuweisung von Sitzen im Landtag, Bundestag, etc.



- Unabhängig von dem genauen Ausgang der Wahl
- gibt es eine genaue Anweisung, nach der die Sitze den Parteien zugewiesen werden und diese die Sitze an ihre Abgeordneten verteilen.
- Küchenrezept keine so gute Analogie, aber okay.
 - Immer dieselben Zutaten und immer dasselbe Ergebnis (in der Theorie)



Ein Algorithmus ist....

- Handlungsanweisung:
Umrechnungsprozedur von
Stimmen zu Sitzen
- Klar definiert
- Für alle möglichen Ausgänge von
Wahlen definiert



...eine für jede **erfahrene Programmiererin** ausreichend **detaillierte und systematische Handlungsanweisung** oder **Lösungsvorschrift**, so dass bei **korrekter Implementierung** der Computer für jede **korrekte Inputmenge** den **korrekten Output** berechnet – in endlicher Zeit.

Eine **Implementierung** ist die Übersetzung der für den Menschen verständlichen Handlungsanweisung in eine Programmiersprache.



B wie Big Data und C wie Computerintelligenz

Big Data: sehr große Mengen an Daten, oft aus unterschiedlichen Quellen, unstrukturiert

Computerintelligenz: Algorithmen des maschinellen Lernens

Empfehlungssysteme



- Woher weiß Amazon, was ich lesen will?
- ...Google, was ich wissen will?
- ...Facebook, was mich unterhält?
- ...Twitter, was mich aufregt?
- ...Instagram, was mich freut?
- ...Pinterest, was ich basteln will?

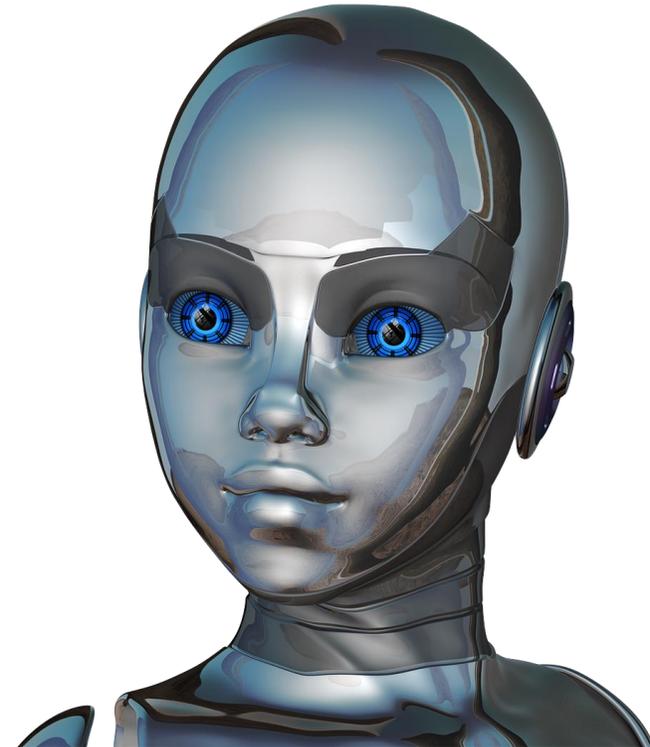




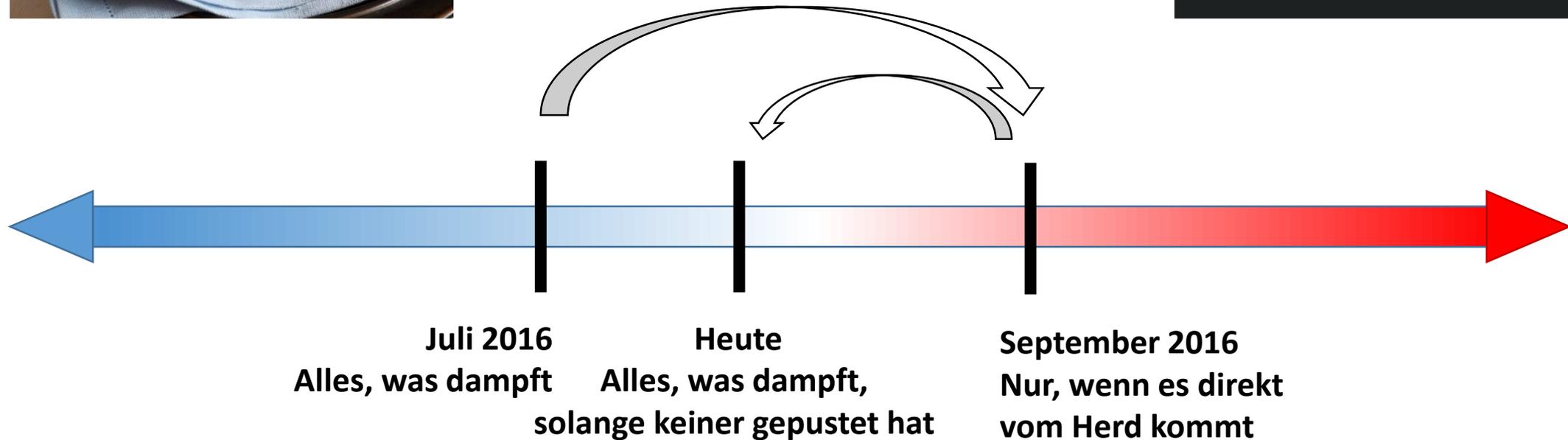
Können Computer lernen?

Was heißt Lernen?

- In derselben Situation ein vorher gezeigtes Verhalten wiederholen.
- In derselben Art von Situation das richtige Verhalten aus einer Reihe von Möglichkeiten auswählen.



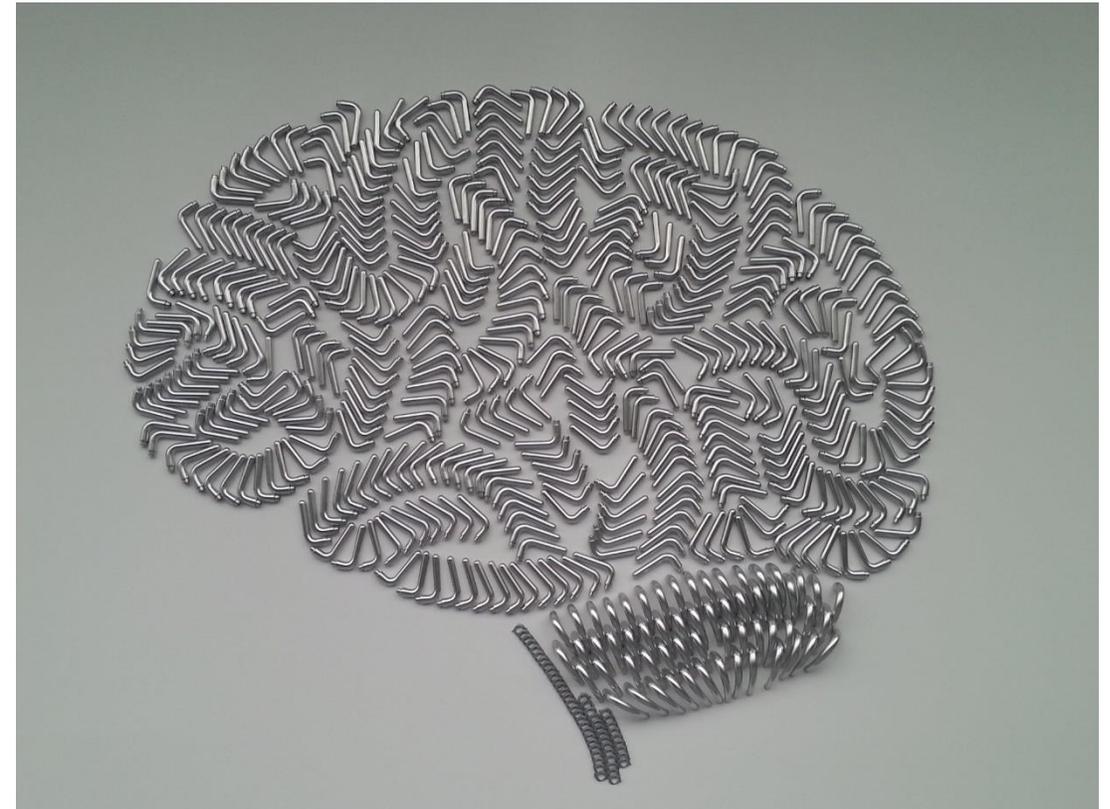
Sebastian lernt „heiss“ und „warm“



Sebastian lernt...



- Durch **Rückkopplung**: unerwartet heiß, unerwartet kalt
- Durch **Speicherung in einer Struktur**: in Neuronen und deren Verknüpfung.
- Durch **sehr viele Erlebnisse („Datenpunkte“)**
- Durch **Generalisierung des Gelernten**.

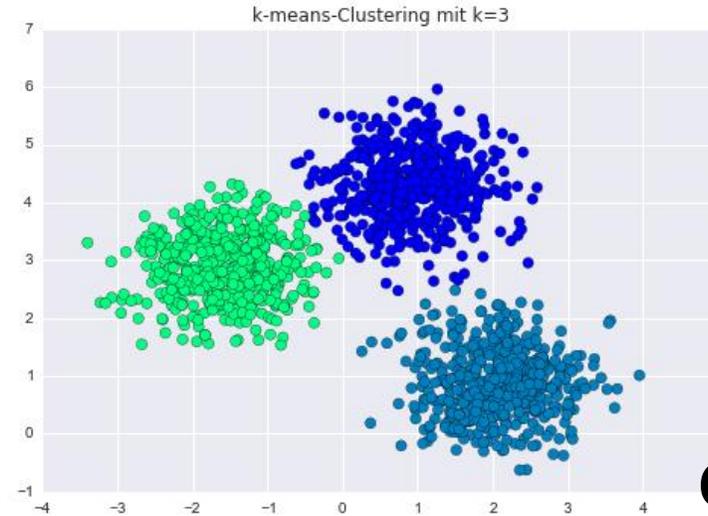
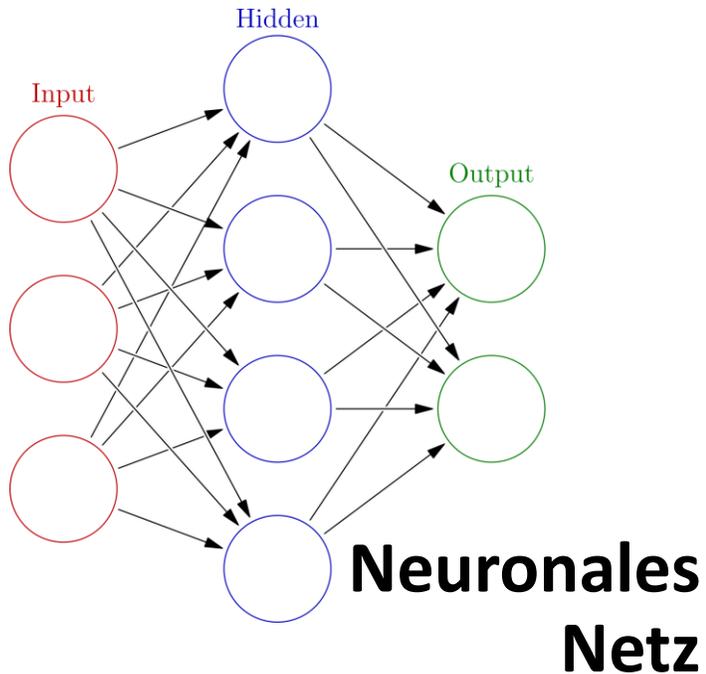


Computer lernen

Damit ein Computer lernen kann, benötigt er ebenfalls eine **Struktur**, um Gelerntes abzuspeichern.

Optimal auch **Rückkopplung**.

Er lernt **generelle Regeln**.



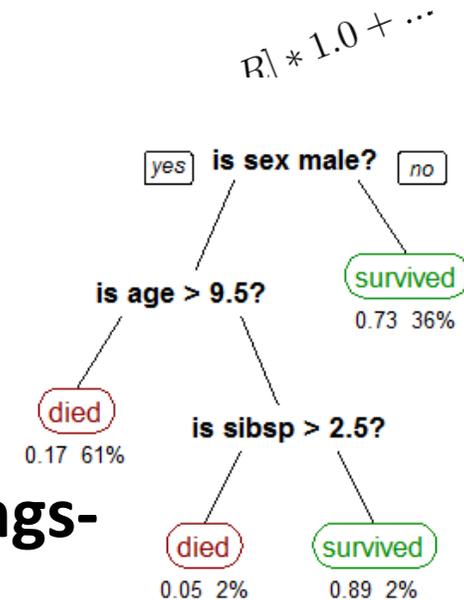
Clustering



Formel

$$w_1 * \#V_h - w_2 * \#day_i V_h + w_3 * I[g = male]$$

Entscheidungs- bäume





Lernen durch Clustern

Was ist ein Rrrr, was ein Hiha und was ein Ts?



Hiha

Hiha



Hiha

Blaulicht



RRR

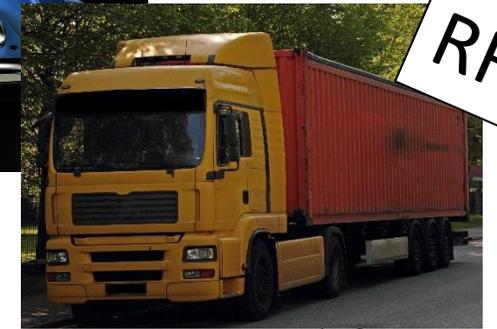
Auto



RRR



RRR



RRR



Ts



Ts



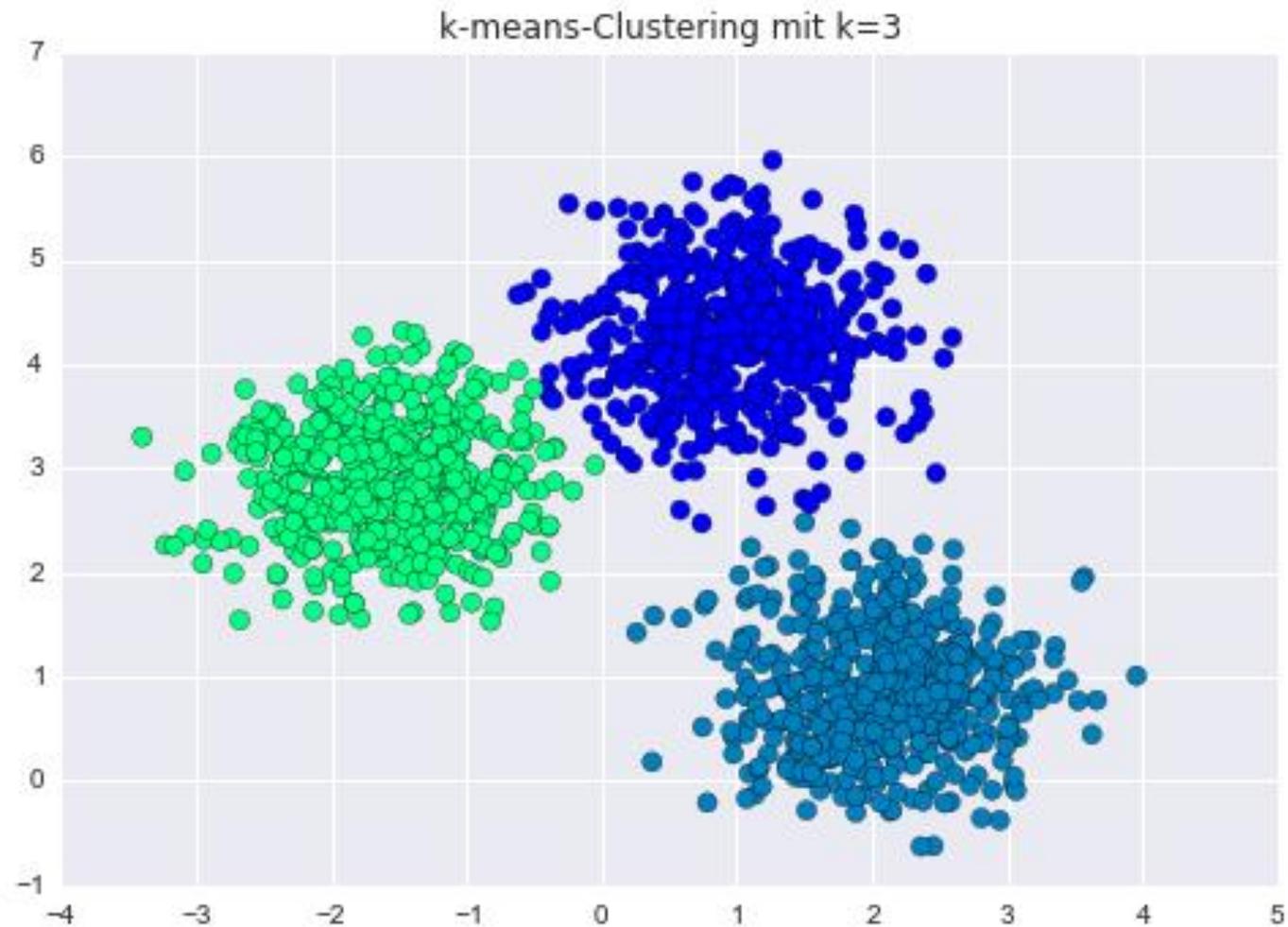
Ts

Zug



RRR

K-Means-Clustering



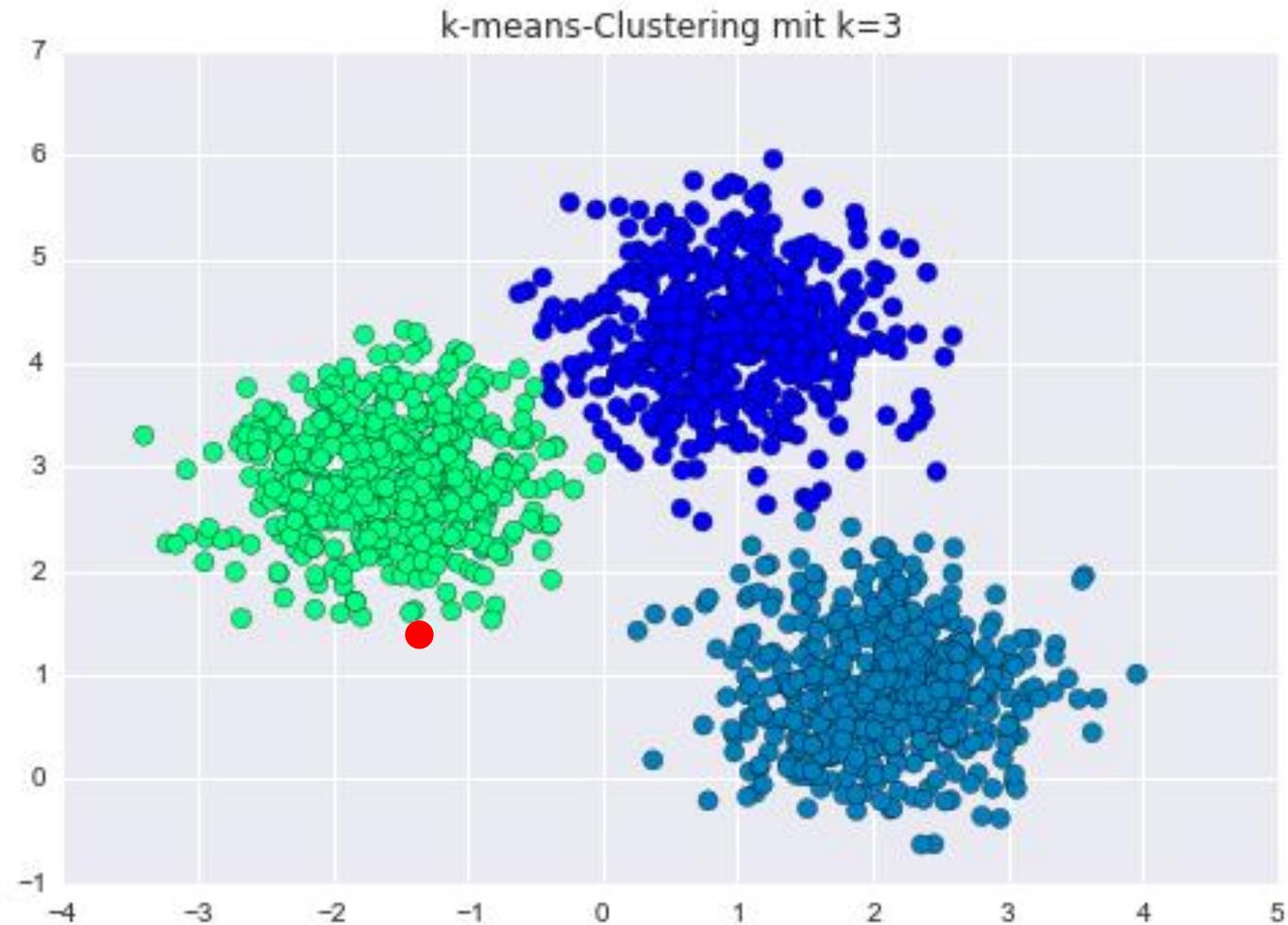
Computerintelligenz



- **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“.



K-Means-Clustering





Und wie funktioniert Google?

Weiß natürlich keine(r) so ganz genau...

Suchmaschinen 101

1. Filtern aus allen ihnen bekannten Webseiten diejenigen, deren Text mit den angegebenen Suchbegriffen zusammenhängen.
2. Sie bewerten diese Seiten anhand der Vernetzungsstruktur der Seiten untereinander
3. Es wurde vermutet, dass sie auch „lernen“, was ein Nutzer gerne mag
 1. Durch bisheriges Verhalten
 2. Durch Verhalten ähnlicher Nutzer





Ließ ein Algorithmus Trump triumphieren?

Der Fakten-Check: Google und Facebook

Disclaimer: Weder ich selbst noch meine Forschungsgruppe noch Algorithm Watch wird oder wurde in irgendeiner Form von diesen Firmen finanziell unterstützt.

Der Faktencheck: Google | *Suchvervollständigung*



- Im Vergleich zu anderen Suchmaschinen schwierig, ‚negative‘ Suchvervollständigung zu bekommen¹.
- Das gilt aber für **beide** Seiten (Clinton und Trump).
- Insofern **keine** einseitige Verzerrung erkennbar.
- Vermutlicher Grund: europäische Gerichtsfälle sorgen für Durchsetzung von Persönlichkeitsrechten
 - Fall Wulff

clinton ist|

clinton ist **ein reptil**

clinton ist **schlimmer als trump**

clinton ist **tot**

clinton **istanbul bilgi**

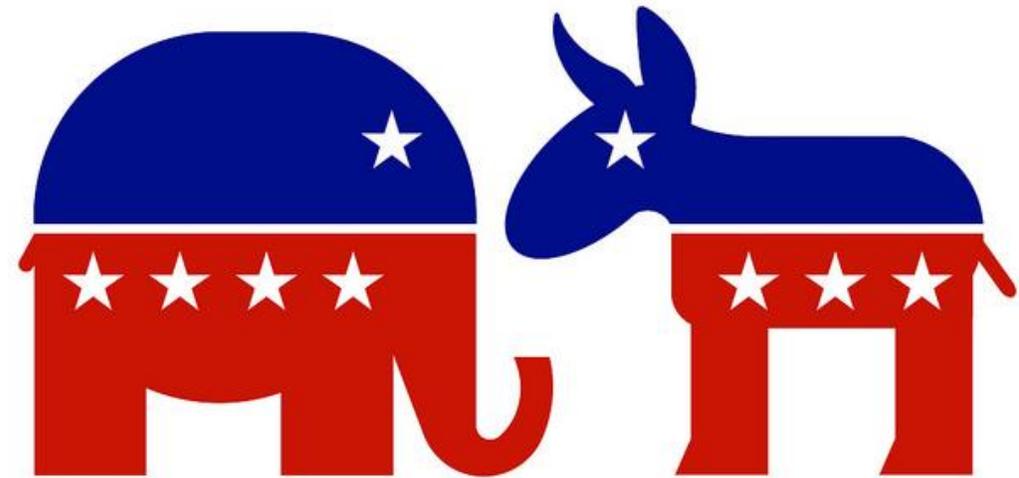
Zum Start der Suche Eingabetaste drücken

Suche am 29.3.2017, 9:06

¹ Zum Beispiel Rhea Drysdale auf ihrem Blog: <https://medium.com/@rhea/hillary-clintons-search-results-manipulated-by-sourced-not-google-3dd9a5c68ca1#.v4c9tidnj>

Der Faktencheck: Google II

- Trielli et al. fanden 7 positive Nachrichten für Demokraten und nur 5.9 für Republikaner.
- Die Studie von Trielli et al.¹ beruht auf der normalen Google-Suche (**nicht** Nachrichten).
- Hier erwartet und findet man hauptsächlich selbstkuratierte Seiten.
- Es gab nur 3 demokr. und 13 republ. Kandidaten.
- Nimmt man den Outlier Bernie Sanders heraus, gibt es keinen Unterschied mehr (5.9 zu 6).



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

Der Faktencheck: Google III



- Epstein und Robertson warnen vor dem „Suchmaschinen-Manipulationseffekt“
 - 20% oder gar bis zu 80% unentschiedene Wähler könnten durch manipulierende Reihenfolgen „umgedreht“ werden.
- Mathematisch fehlerhaft².
 - Die ‚bis zu 80%‘ kommen aus einem nicht korrigierten ‚Multiple Testing‘.
- Die Daten zeigen nicht einen 20%-Effekt sondern 2-4%.
 - Das kann in den USA immer noch wahlentscheidend sein.
 - In Deutschland eher nicht.



1 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

2 Katharina A. Zweig: “Watching the watchers: Epstein and Robertson’s ‘Search Engine Manipulation Effect’”, Fallstudie von Algorithm Watch, publiziert am 7.4.2017, <https://algorithmwatch.org/watching-the-watchers-epstein-and-robertsons-search-engine-manipulation-effect/>

Datenspende Projekt BTW 2017



- Plugin
- 16 Suchanfragen, automatisch alle 4 Stunden
 - Auf Google und Google News
- Wir speichern die Top Stories
- Und die Links auf der ersten Suchergebnisseite.
- Zudem:
 - Ungefährer Standort
 - Status: Login als Google-Nutzer

The screenshot shows a Google search for 'Angela Merkel'. The search bar contains the name and the Google logo. Below the search bar, there are tabs for 'Alle', 'News', 'Bilder', 'Videos', 'Maps', 'Mehr', 'Einstellungen', and 'Tools'. The search results show approximately 58,000,000 results in 0.59 seconds. The 'Schlagzeilen' (Headlines) section features three articles: 'Civey-Umfrage zur Bundestagswahl: Die Hauptstadt mag Merkel' from Tagesspiegel, 'Angela Merkel - die Kanzlerin, in Bernstein gefangen' from Spiegel Online, and 'Merkels Deutschland: Zweigeteilt, dreigeteilt? Wie Stiefmutter Land und Leute zerlegt' from Contra Magazin. Below the headlines is a link to 'Mehr zu Angela Merkel'. The 'Angela Merkel' knowledge panel is visible, including her official website, a Wikipedia link, and a Facebook link. To the right, there is a gallery of images of Angela Merkel and a section titled 'Wird auch oft gesucht' (Also often searched) with thumbnails for Joachim Sauer, Wladimir Putin, Donald Trump, Martin Schulz, and Helmut Kohl.

Sind Sie vertraut mit der Filterblasentheorie?



- 1) Ja
- 2) Nein!



Erste Analyse



- Nicht viel Raum für Personalisierung
- Bei Analyse der Daten eines Suchtermins:
 - 8-9 gemeinsame Suchergebnisse bei Politikern
 - 5-6 gemeinsame Suchergebnisse bei Parteien
 - Der Rest weist viel Regionalisierung auf.

1	1	http://www.cdu-baden-baden.de/	
1	1	http://www.cdu-bad-godesberg.de/	
1	1	http://www.cdu-bad-honnef.de/	
1	1	http://www.cdu-bad-honnef.de/der-vorstand-cdu-	
1	1	http://www.cdu-bad-honnef.de/stadtverband-bad	
1	1	http://www.cdu-bad-kreuznach.de/	
1	1	http://www.cdu-bad-kreuznach.de/termine/	
1	1	http://www.cdu-bad-laer.de/	
1	1	http://www.cdu-bad-laer.de/Aktuelles.html	
1	1	http://www.cdu-badnenndorf.de/	
1	1	http://www.cdu-badnenndorf.de/2_24_Vereinigung	
1	1	http://www.cdu-bad-oldesloe.de/	
1	1	http://www.cdu-bad-oldesloe.de/fraktion/	
1	1	http://www.cdu-bad-oldesloe.de/kontakt/	
1	1	http://www.cdu-bad-oldesloe.de/presse/	
1	1	http://www.cdu-bad-oldesloe.de/vorstand/	

Der Faktencheck: Facebook I

Filterblasen und Echokammern



- Pro: Studie von Hagen, In der Au und Wieland zeigt, dass Polarisierung bei Personen ohne Abitur mit Intensität der Nutzung sozialer Medien steigt¹.
- Contra: Neue Studie des Hans-Bredow-Instituts (Schmidt et al., 2017) zeigt, dass Mediennutzung in Deutschland sehr divers ist – Einfluss der sozialen Medien auf Meinungsbildung eher gering².

¹ Lutz Hagen, Anne-Marie In der Au und Mareike Wieland: „Polarisierung im Social Web und der intervenierende Effekt von Bildung“, erscheint im März 2017 in der Sonderausgabe „Algorithmen, Kommunikation und Gesellschaft“ von kommunikation@gesellschaft (www.kommunikation-gesellschaft.de)

² Jan-Hinrik Schmidt et al.: „Relevanz einzelner Medienangebote und digitaler Dienste für die Meinungsbildung“, Arbeitspapier des Hans-Bredow-Instituts Nr. 40, 2017 <http://www.hans-bredow-institut.de/de/node/6330>

Der Faktencheck: Facebook II

Aktivierung durch Wahlbuttons



- 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.
- Bisher technisch von außen nicht untersuchbar:
 - Werden diese Wahlbuttons zufällig ausgerollt oder nur an Anhänger einer Partei?



Heute findet die Landtagswahl in Rheinland-Pfalz statt!

Sag deinen Freunden, dass du wählen gehst. Weitere Informationen gibt es hier.

[Ich bin Wähler!](#) [Mehr Informationen](#)

Manipulation durch Firmen



- Technisch möglich.
- Die Infrastruktur wird teilweise nahezu monopolistisch genutzt und könnte jederzeit zur Manipulation eingesetzt werden.
- Aber:
 - Es gibt keine belastbaren Anzeichen dafür.
 - Es fehlt ein sinnvolles Motiv.
 - Die Kosten durch ein ‚Leak‘ wären enorm.
- Persönliches Fazit: Momentan unwahrscheinlich
- Trotzdem (stetige, repräsentative) Überwachung notwendig





Weitere Verdächtige

Emergente Phänomene und Manipulation von außen

Oder waren es ...



- BuzzFeed schreibt, dass ein deutlicher Anteil der Nachrichten auf hyperkonservativen Webseiten falsch oder eine Mischung aus falschen und korrekten Nachrichten ist.
- Ein Teil dieser „Hyperpartisan“-Webseiten stammt aus....
- ...Mazedonien!

Motivation



„Die meisten Texte auf diesen Seiten sind zusammengeschrieben oder einfach abgeschrieben von rechtslastigen Webseiten der USA. Die Mazedonier sahen irgendwo eine Story, versetzten sie mit einer aufgeheizten Überschrift, und veröffentlichten es dann schnell auf ihrer Webseite. Dann teilen sie die Nachricht auf Facebook und veranlassen damit Leute, auf ihre Seite zu kommen. Je mehr Leute sich verführen lassen, desto mehr Geld verdienen sie durch Werbung auf ihrer Webseite.“

Craig Silverman und Lawrence Alexander: „**How Teens In The Balkans Are Duping Trump Supporters With Fake News**“, BuzzFeed Nov. 4th, downloaded on the 27th of November, 2016
https://www.buzzfeed.com/craigsilverman/how-macedonia-became-a-global-hub-for-pro-trump-misinfo?utm_term=.wvzZe7D5#.he3BEIYV

Weitere, unabhängige Quelle von Wired: Samantha Subramanian: “The Mazedonian Teens who mastered Fake News”, published on 15th of February, 2017 on WIRED,
<https://www.wired.com/2017/02/veles-macedonia-fake-news/>

Eigentlich sinnvolle Algorithmen ergänzen sich hier schlecht



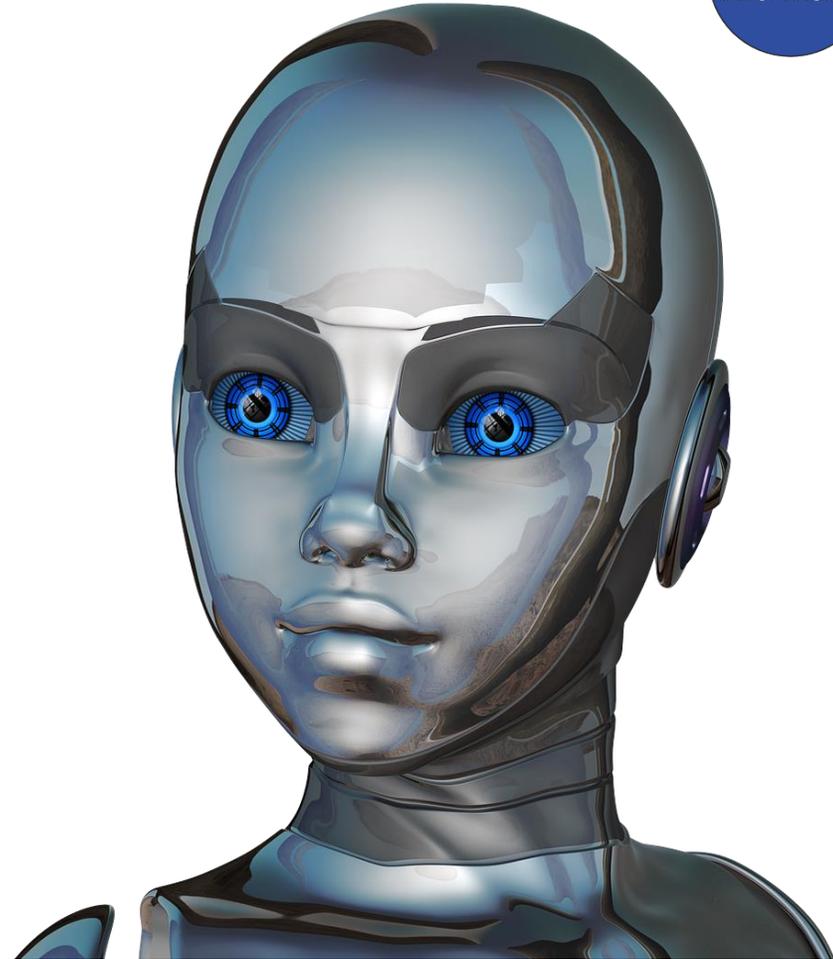
- Wir sprechen auf Emotionales mehr an als auf Faktisches.
- Suchmaschinen und News Feeds und andere Empfehlungssysteme servieren uns, was wir eh schon mögen.
- Werbeanzeigen werden (fast) unabhängig vom Inhalt der Webseiten verteilt – nur nach Demographie der Nutzerinnen und Nutzer.
- Rezept:
 - Möglichst emotionale (gerne falsche) Inhalte, die möglichst viel Traffic erzeugen.



Social Bots



- Alle Algorithmen reagieren auf Interaktion mit den Nachrichten.
- Interaktion erhöht i.A. das Ranking der Nachricht.
- Computerprogramme agieren wie Menschen („social bots“) und manipulieren damit die Algorithmen.
- Das kann auch aus anderen Ländern heraus passieren.





Austrocknen des Sumpfes

- Es werden Seiten finanziert, die viele Klicks haben, sich aber nicht unbedingt durch Qualität auszeichnen.
- Hier muss ein neuer Anreiz gesetzt werden.
- Z.B. Bewusstmachung, wer durch Werbung finanziert wird¹.
- Es gibt Berichte, dass sich auch Extremisten über durch Google und YouTube vermittelte Werbung teilweise finanzieren².
 - Es beginnt sich Protest zu firmieren, insbesondere in Großbritannien³.
- Aber: Firmen nicht zu Zensurbehörden machen!

1 <http://www.spiegel.de/netzwelt/web/breitbart-kritik-an-deutscher-werbung-auf-hetz-seite-a-1124526.html>

2 <http://www.usatoday.com/story/tech/news/2017/03/20/trouble-brewing-google-youtube-europe-advertising-british-media-outlets/99410760/>

3 <http://www.usatoday.com/story/tech/news/2017/03/20/trouble-brewing-google-youtube-europe-advertising-british-media-outlets/99410760/>

Nicht zuletzt: Die Parteien selbst

- *Search Engine Optimization* gehört zur Kunst des Parteimarketings.
- Zum ersten Mal ist auch wirklich feingranulare, personalisierte Werbung möglich.
 - Gibt es das Microtargeting aufgrund von Persönlichkeitsanalysen aus Facebook wirklich?¹
 - Microtargeting an sich ist auf jeden Fall Standard – basierend auf Datamining.



¹ <http://www.zeit.de/digital/internet/2016-12/us-wahl-donald-trump-facebook-big-data-cambridge-analytica>

Anteil selbstkuratierte Webseiten



Die Ergebnisse finden Sie in unserem
Zwischenbericht für das Datenspendeprojekt ab ca.
Ende August 2017 auf dem Blog von [algorithmwatch](http://algorithmwatch.org):

Algorithmwatch.org



Können uns Algorithmen
manipulieren?

Was tun?



Das Analoge nicht vergessen!

- In Deutschland ist der Medienmix bisher gut und das Vertrauen in Öffentlich-Rechtliche relativ hoch¹.
- Das ist nicht gottgegeben:
 - Erste Spaltungstendenzen ebenfalls erkennbar¹.
- Vertrauen in journalistisch arbeitende Medien muss gelehrt werden und erhalten bleiben!
 - Medienkompetenz und journalistisches Denken lehren.
 - Tageszeitungen, Nachrichten und Quellenkritik als festen Bestandteil der (schulischen) Ausbildung!



Zwischenfazit



- Momentan ist die Lage in Deutschland vermutlich nicht besorgniserregend
 - Weitere Grundlagenforschung ist wichtig
- Es ist aber damit zu rechnen, dass der Anteil der Nutzer, die Nachrichten über soziale Netzwerke beziehen, weiter steigt.
- Die aufgebaute Infrastruktur und die Monopolstellung der wichtigen Intermediäre ermöglicht interne und externe Manipulation der Algorithmen und damit Einflussnahme auf Nutzer.
- Vielleicht gibt es ein Recht einer Gesellschaft auf Intransparenz von Algorithmen?
 - Hier beißt sich natürlich die Katze in den Schwanz.
- Auf jeden Fall braucht die Gesellschaft Kontrolle über einige Algorithmen und algorithmische Entscheidungssysteme.



Algorithmen in einer demokratischen Gesellschaft

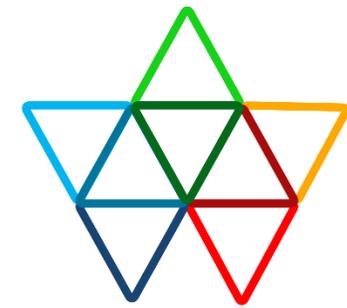
Algorithmische Entscheidungssysteme



- Empfehlungssysteme (wie der NewsFeed) oder Suchmaschinenalgorithmen gehören zu den **algorithmischen Entscheidungssystemen**.
- Weitere algorithmische Entscheidungssysteme gefährden gesellschaftliche Teilhabe noch deutlicher:
 - Kreditwürdigkeit, autom. Leistungsbewertung, Bewerbungsortierer, Predictive Policing, Rückfälligkeitsvorhersagealgorithmen, Terroristenidentifikation



Gründung von „Algorithm Watch“



ALGORITHM
WATCH



Lorena Jaume-Palasi, 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

Schlussformel



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