Sind Computer die besseren Richter?

Algorithmische Entscheidungs-systeme vor Gericht

Prof. Dr. Katharina Anna Zweig

TU Kaiserslautern
Leiterin des Algorithm Accountability Labs
Mitgründerin und wissenschaftliche Beraterin von AlgorithmWatch

TECHNISCHE UNIVERSITÄT
KAISERSLAUTERN

Menschen – so irrational!

- Richter müssen vorzeitige Haftentlassungsanträge begutachten.
- Studie: je weiter von der letzten Pause weg, desto weniger risikoreiche Entscheidungen[1].
- Eine Vielzahl solcher Studien scheint zu beweisen:

1 Danziger, S.; Levav, J. & Avnaim-Pesso, L.: “Extraneous factors in judicial decisions”, Proceedings of the National Academy of the Sciences, 2011 , 108 , 6889-6892

Menschen – so irrational!

- Richter müssen vorzeitige Haftentlassungsanträge begutac
- Studie: letzten weniger Entsche
- Eine Vielzahl solcher Studien scheint zu beweisen:

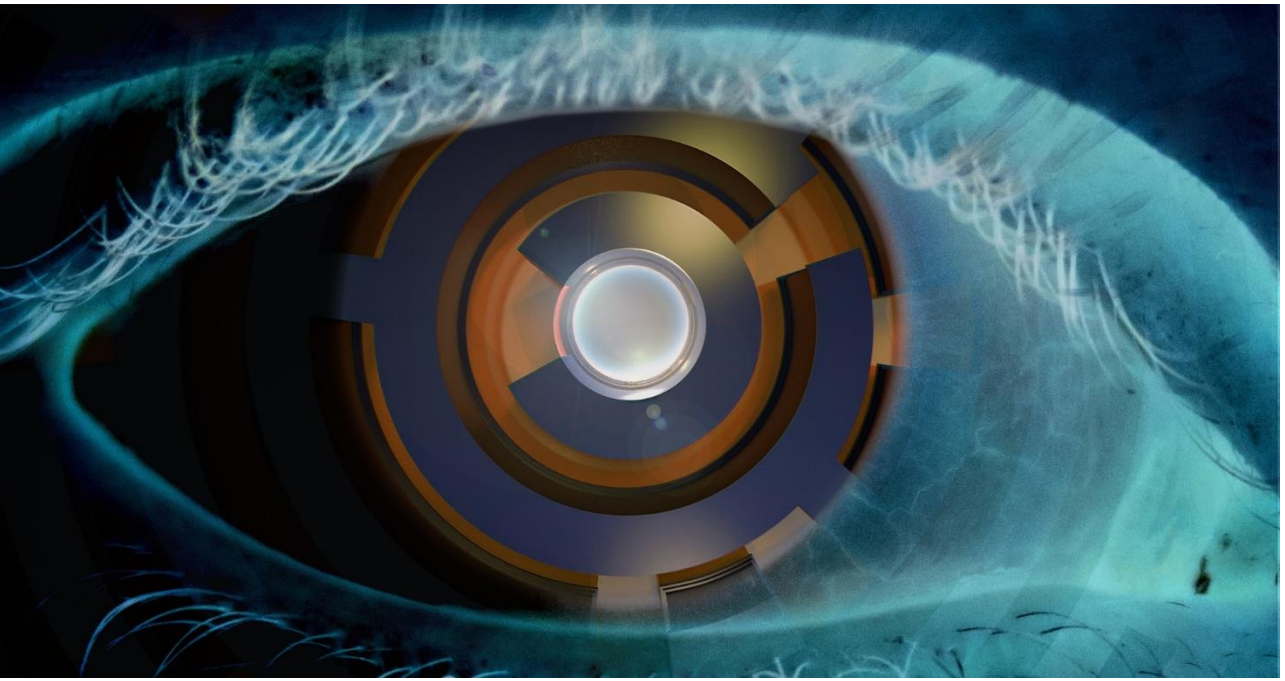
Menschen sind irrational und vorurteilsbeladen.

1 Danziger, S.; Levav, J. & Avnaim-Pesso, L.: “Extraneous factors in judicial decisions”, Proceedings of the National Academy of the Sciences, 2011 , 108 , 6889-6892

Problemfall USA

- Zweithöchste Inhaftierungssrate weltweit.
- 6x höhere Rate von Afroamerikanern und 2x höhere Rate von Latinos als von Weißen.

American Civil Liberties Union



- Amerikanische Bürgerrechtsunion (seit 1920) fordert:
- Algorithmische Entscheidungssysteme sollten überall im Prozess eingesetzt werden!

SMART REFORM IS POSSIBLE - States Reducing Incarceration Rates and Costs While Protecting Communities, Report from August 2011, p. 9

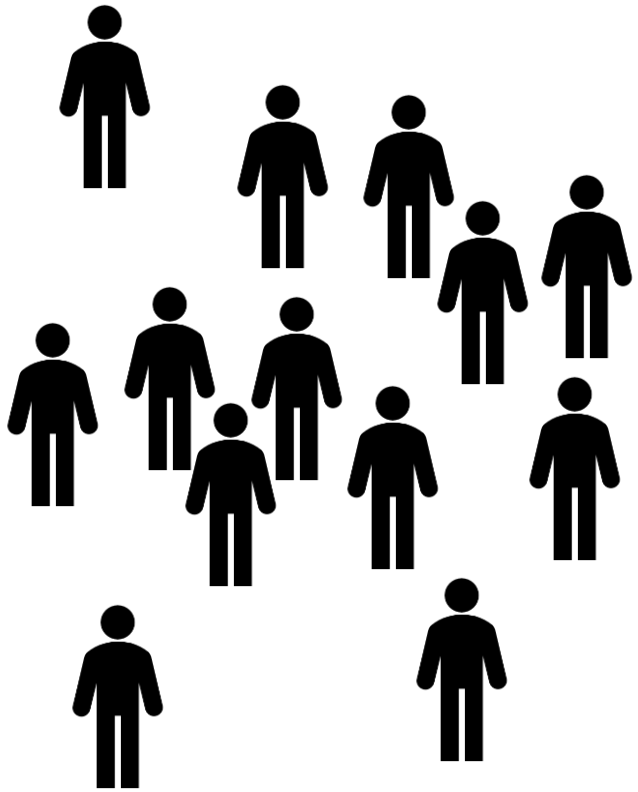
Könnten Computer das besser?

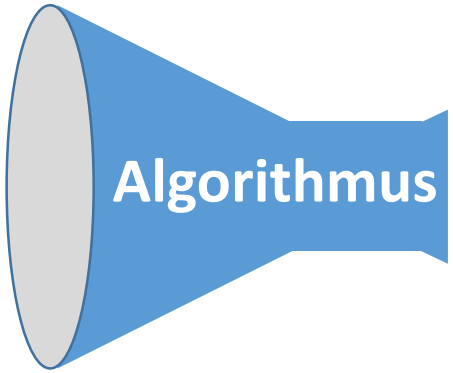
- Die ersten Länder testen *algorithmische Entscheidungssysteme* für gesellschaftlich relevante Probleme.
- Idee: Eigenschaften, nach denen nicht diskriminiert werden darf, könnten vor ihnen besser verborgen werden.

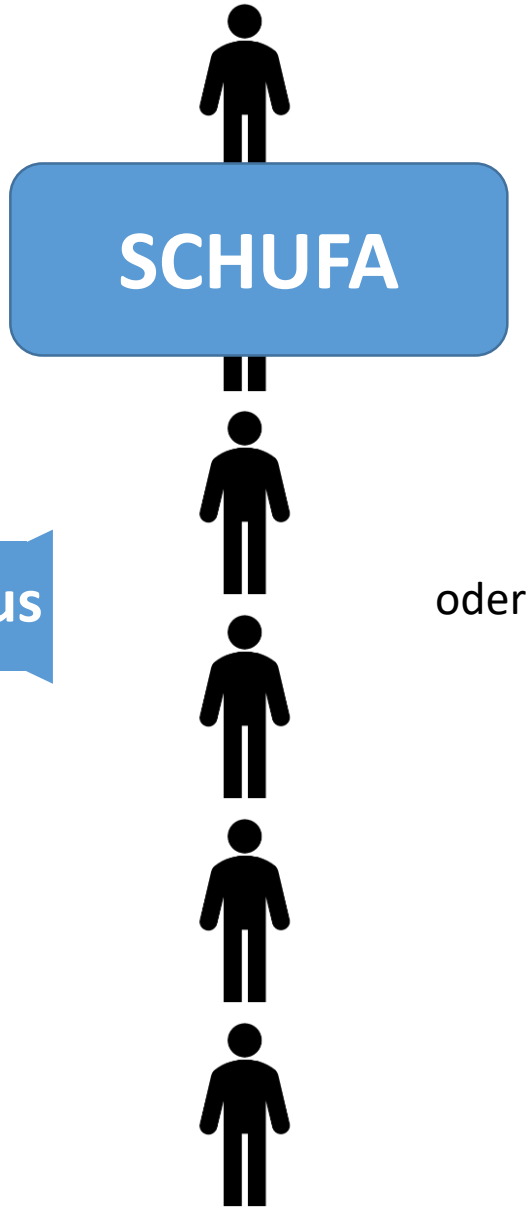
Könnten Computer das besser?

- *Algorithmische Entscheidungssysteme* sind objektiv und arbeiten nahezu fehlerfrei.
- objektiv = „reproduzierbar dieselbe Entscheidung bei derselben Eingabe von Daten“.

Algorithmische Entscheidungssysteme



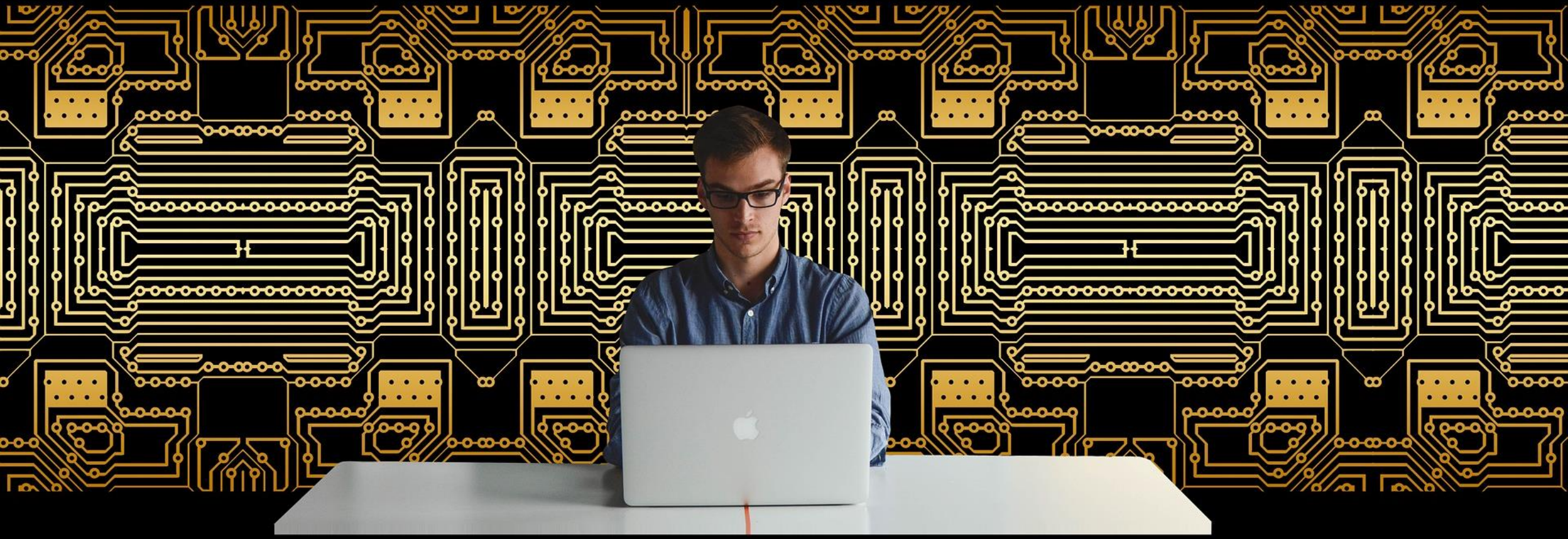




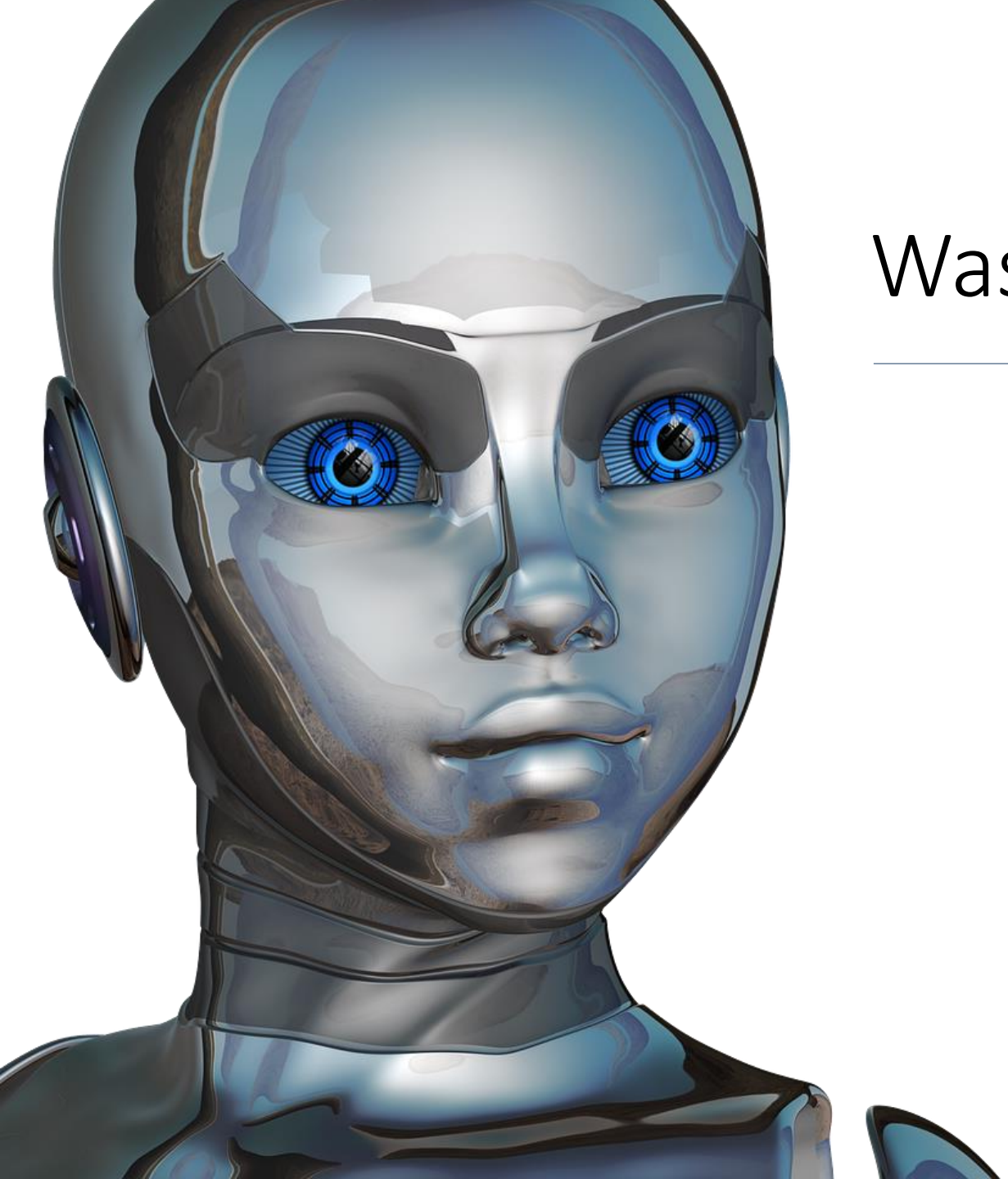
Scoring-Verfahren

oder

Klassifikation



Können Computer lernen?



Was heißt Lernen?

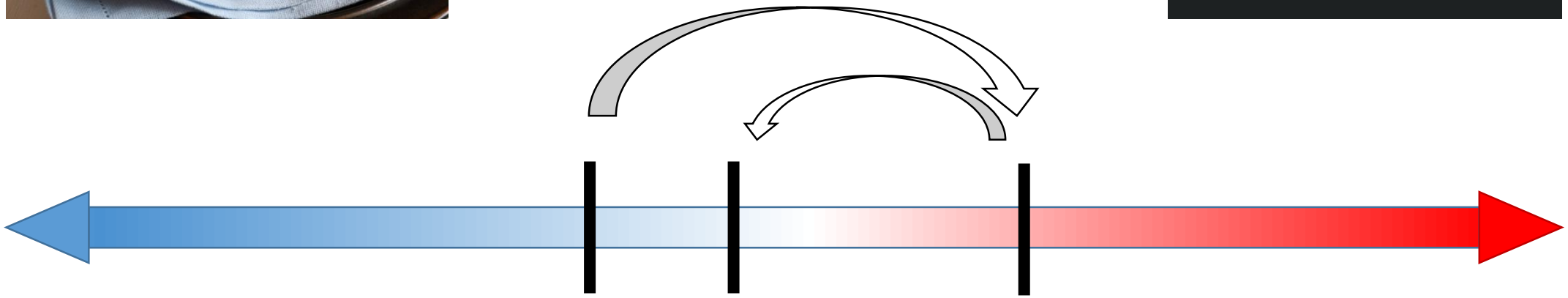
Einfach:

In derselben Situation ein vorher gezeigtes Verhalten wiederholen.

Generalisiert:

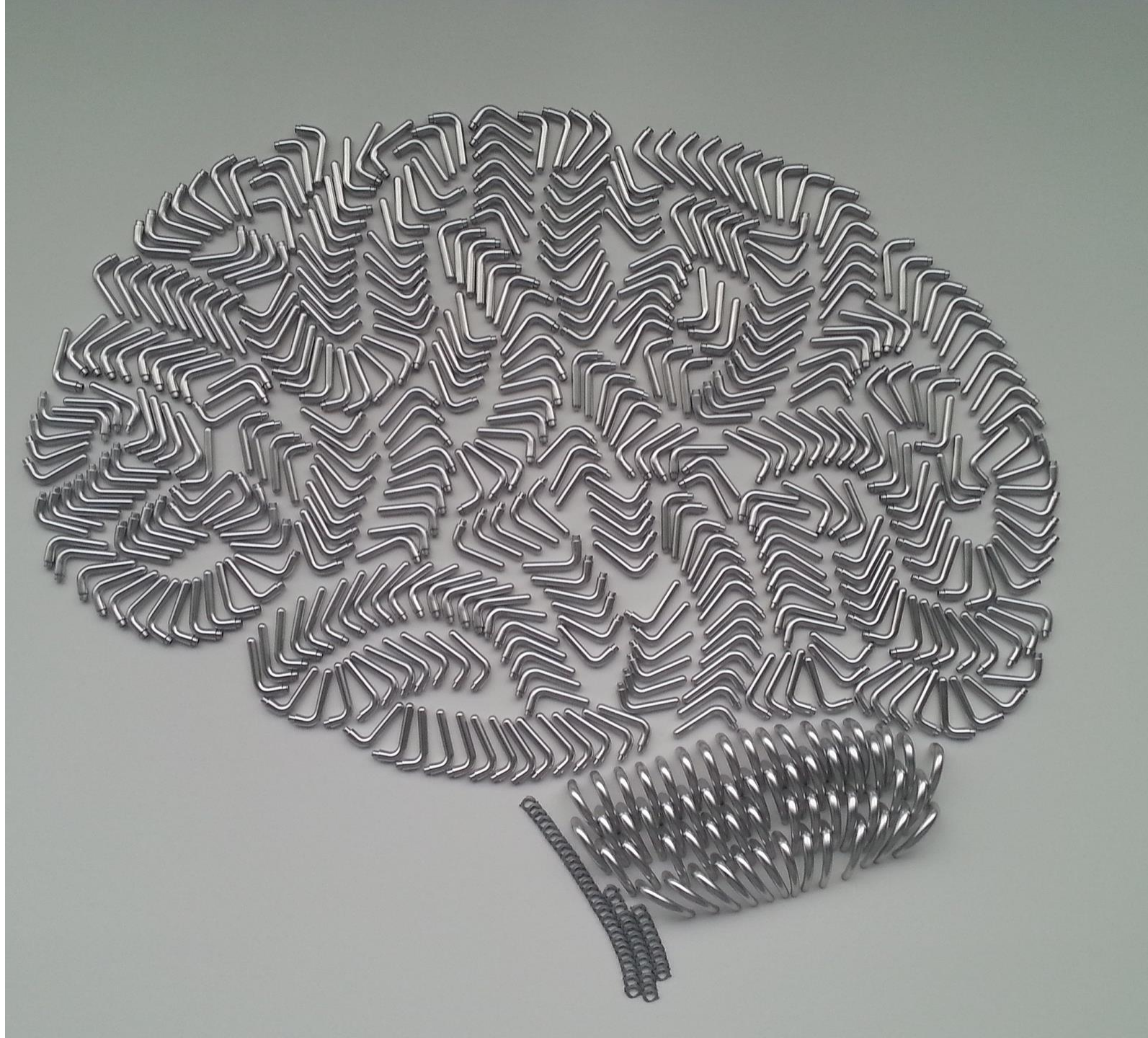
In derselben Art von Situation das richtige Verhalten aus einer Reihe von Möglichkeiten auswählen.

Sebastian lernt „heiss“ und „warm“



Sebastian lernt…

- Durch **Rückkopplung**: unerwartet heiß, unerwartet kalt
- Durch **Speicherung in einer Struktur:** in Neuronen und deren Verknüpfung.
- Durch viele **Datenpunkte.**
- Durch **Generalisierung des Gelernten**.

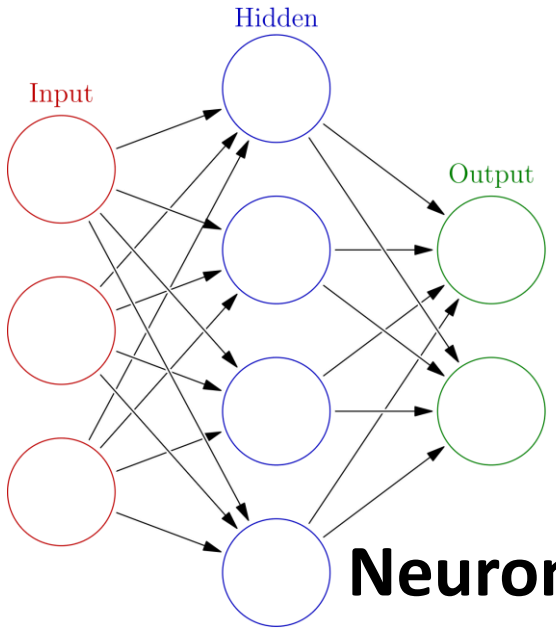


Computer lernen

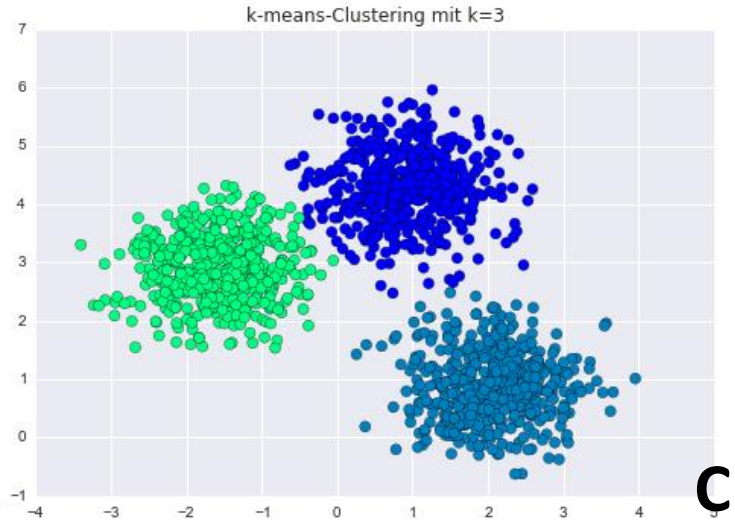
Damit ein Computer lernen kann, benötigt er ebenfalls eine **Struktur,** um Gelerntes abzuspeichern.

Optimal auch **Rückkopplung**.

Er lernt **generelle Regeln**.



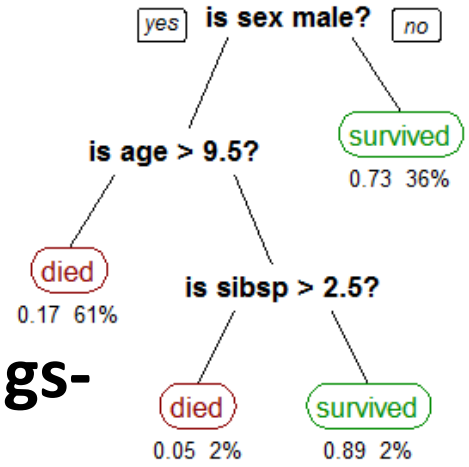
Neuronales Netz



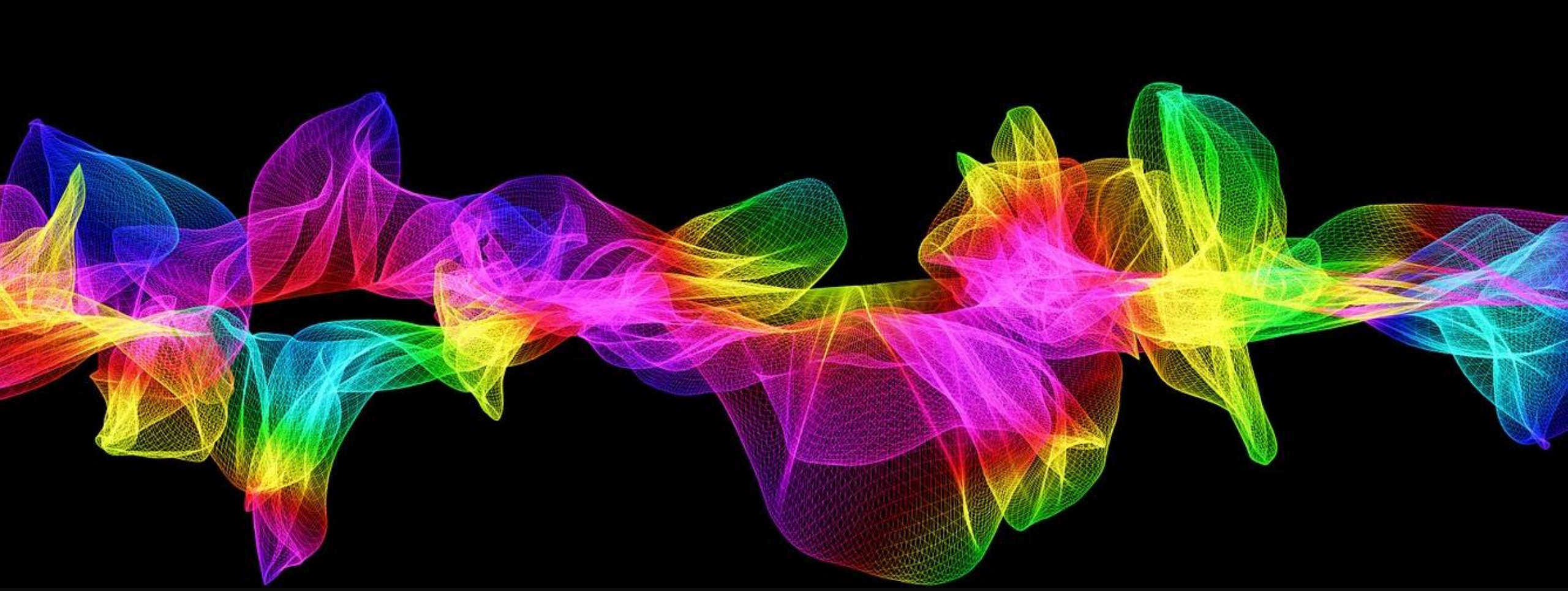
Clustering

$$w_1 * \#Vh - w_2 * \#dayiVh + w_3 * I[g = male] + B[\ * 1.0 + \cdots$$

Formel



Entscheidungs-bäume

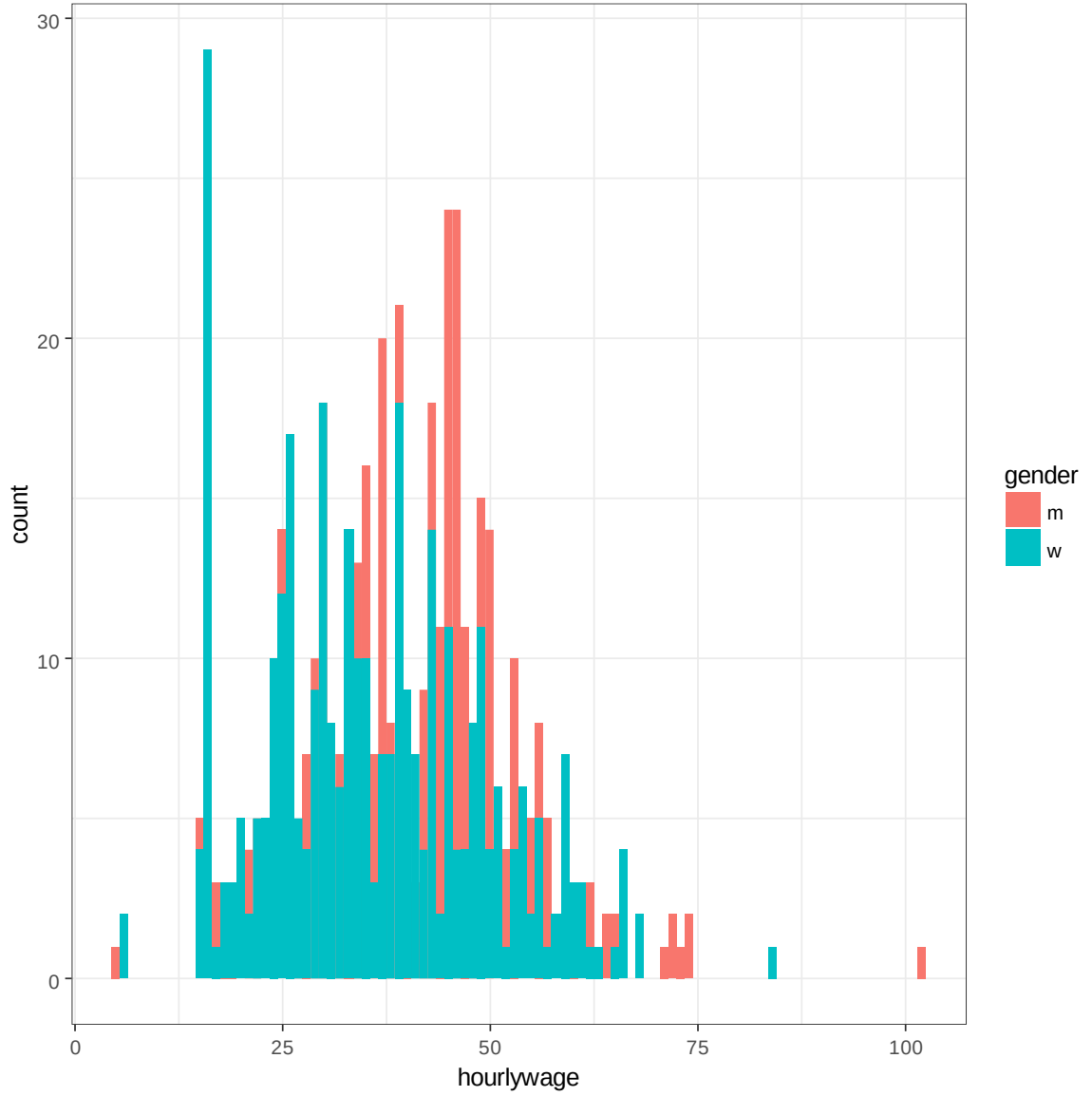


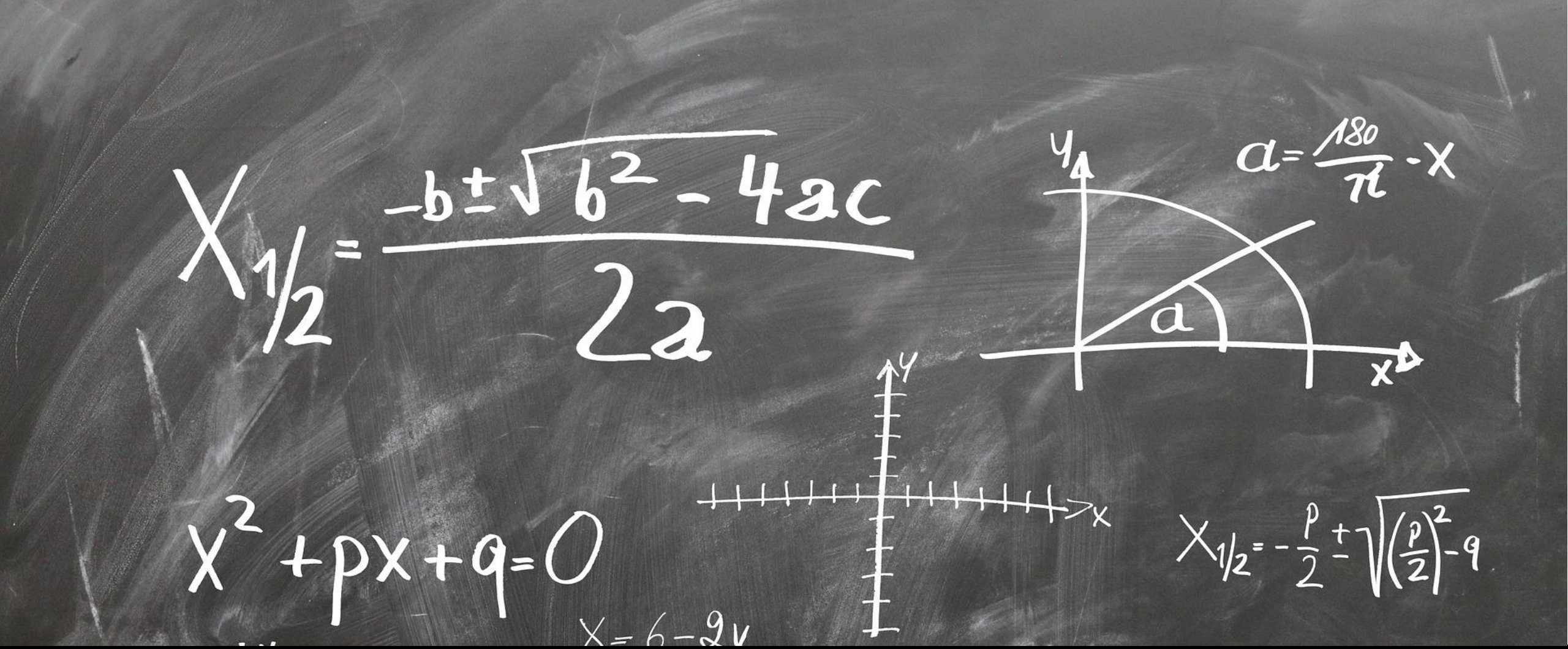
“Lernen” mit Korrelationen

Gehälter in Seattle

Sie bekommen Daten von einer Person – diese verdient weniger als $25 pro Stunde.

Basierend auf den Daten, ist die Person weiblich oder männlich?





Lernen mit Formeln

Am Beispiel der **Bewertung einer Bewerbung**

Datengrundlagen

- Data Mining Methoden nutzen verschiedene Informationen
- Am wichtigsten:

  Wurde der/die Kriminelle rückfällig?

Regressionsansätze

- Die Algorithmendesignerinnen entscheiden, welche der Daten vermutlich mit „Rückfallwahrscheinlichkeit“ korrelieren.
- Gewünschter Output: eine einzige Zahl, nach der Menschen sortiert werden.
- Verabredung: Je höher die Zahl, desto höher die Rückfallwahrscheinlichkeit.
- Beispiel Formel:

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

Allgemein

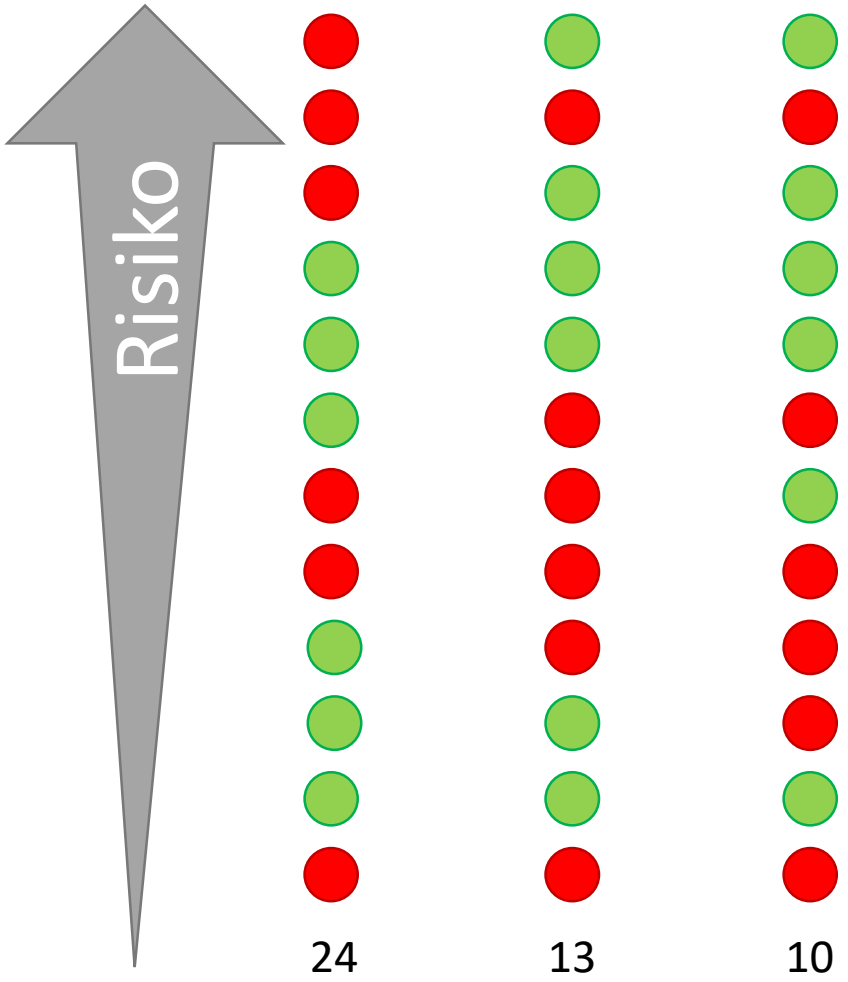
$$
\begin{aligned}
& w_1 * \text{ bisherige Verhaftungen} \\
- & w_2 * \text{ Anzahl Tage seit letzter Verhaftung} \\
+ & w_3 * \text{ (Wenn Mann, dann 1, sonst 0)} \\
+ & w_4 * \text{ (Wenn Raubüberfall, dann 1, sonst 0)} + ...
\end{aligned}
$$

Der Computer bestimmt die Gewichte und bekommt ein Feedback (Rückkopplung), inwieweit die damit resultierende Bewertung tatsächlich mit dem (beobachteten) Verhalten übereinstimmt.

Qualität eines Algorithmus

„Lernen“ von Gewichten

- Algorithmus probiert Gewichte und berechnet Risiko für alle Personen im Datenset.
- Bewertet jeweils, wie viele erwiesenermaßen Rückfällige möglichst weit oben stehen.
- Die Gewichtung, die das maximiert, wird für weitere Daten genommen.



Grüne Kugeln symbolisieren resozialisierte, rote rückfällige Kriminelle.

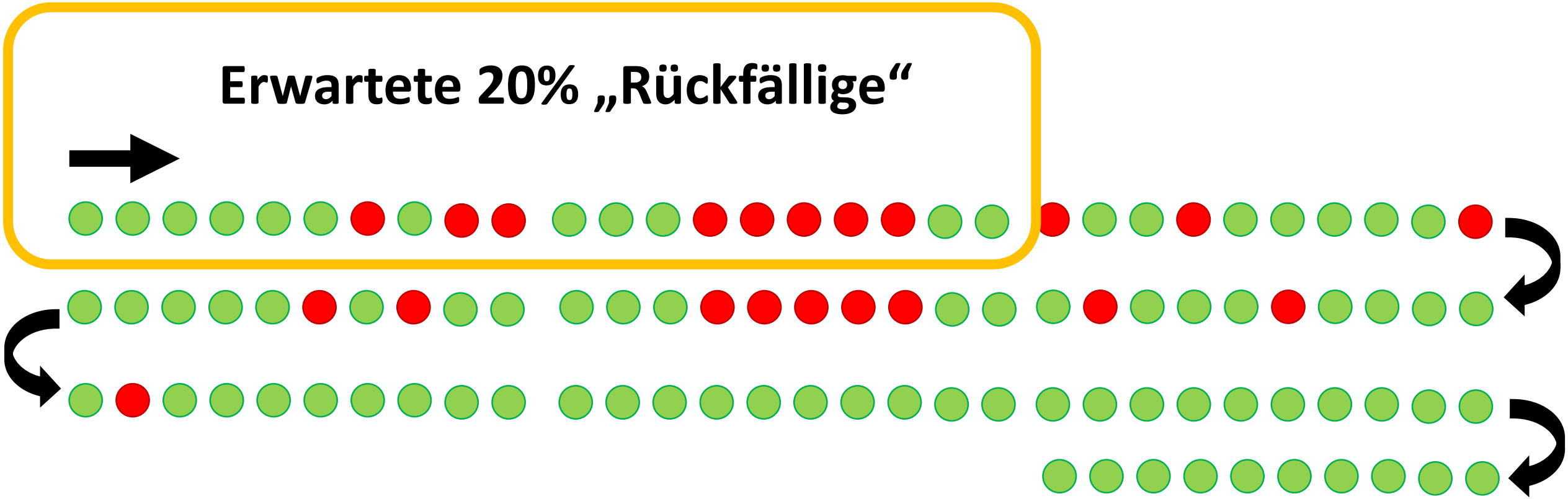
Optimale Sortierung: Alle roten oben, alle grünen darunter.

Qualitätsmaß: Paare von rot und grün, bei denen die rote Kugel über der grünen einsortiert ist.

Oregon Recidivism Rate Algorithm

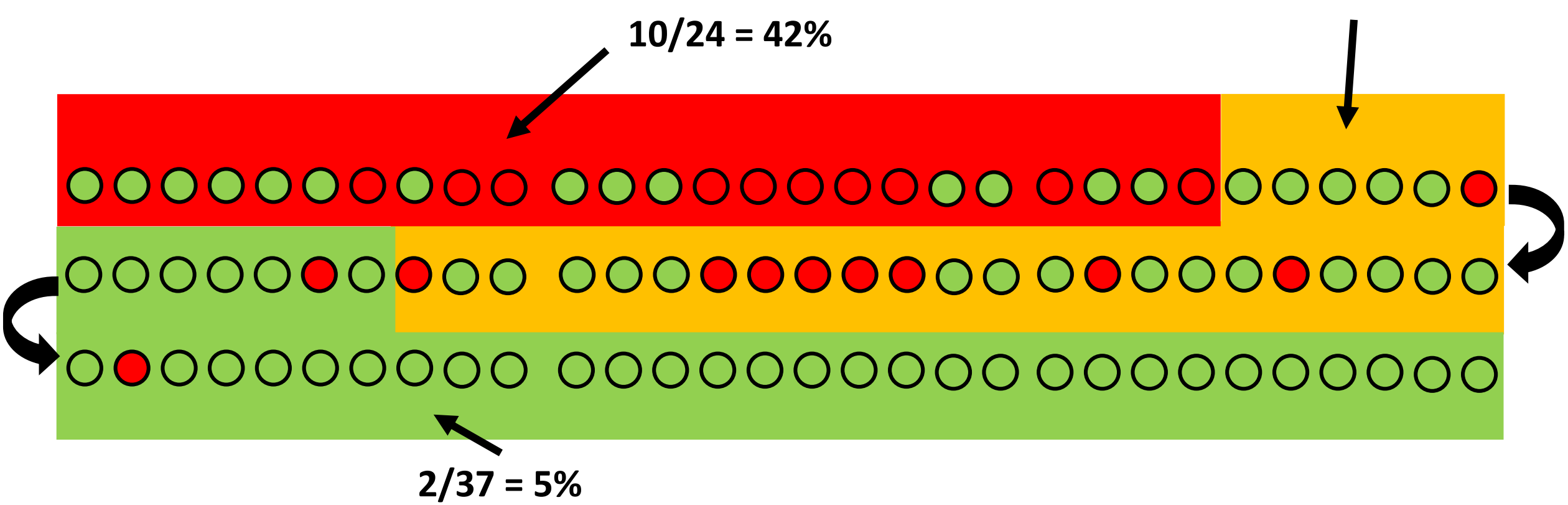
- 72 von 100 Paaren werden korrekt sortiert.
- So werden aber keine Urteile gefällt!
- Sondern: Reihe von Angeklagten, von denen diejenigen mit dem höchsten Rückfallrisiko benannt werden sollen.
- Rückfallquote bei jugendlichen Kriminellen liegt z.B. bei 20%.

Mögliche Sortierung eines Algorithmus mit dieser „Güte“ (75/100 Paaren)



Vom Scoring zur Klassifikation

- ACLU fordert: Es soll drei Klassen geben.
- Niedriges, mittleres, hohes Risiko.



Das ist wie…

„Kaufen Sie diesen wunderbaren Wagen. TÜV? Brauchen Sie nicht! Und sehen Sie nur, die unglaublich gut erhaltenen Sommerreifen. Das ist noch Qualität!”

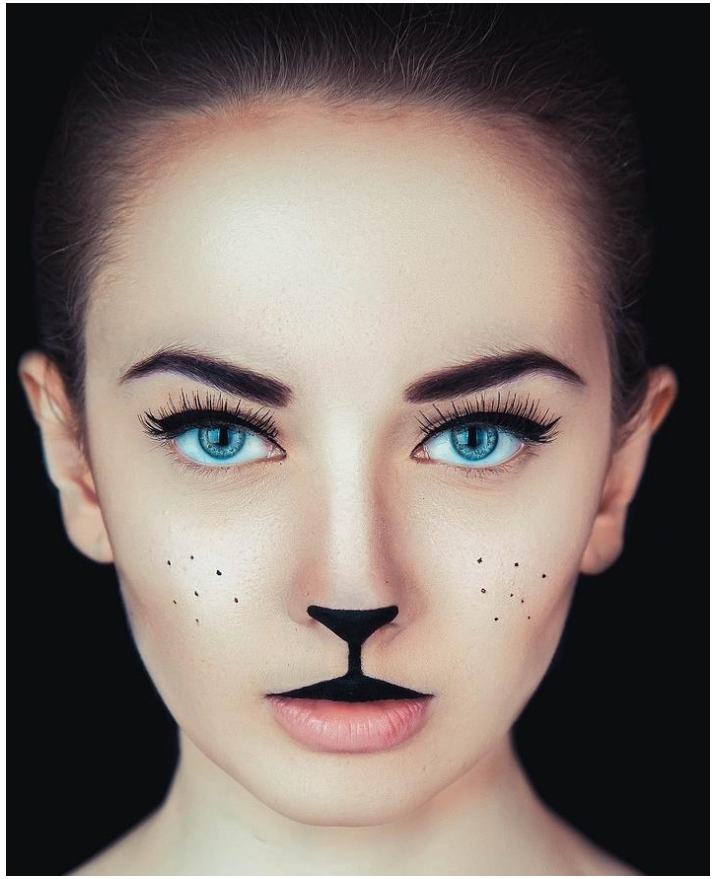
Probleme von algorithmischen Entscheidungssystemen (ADM Systemen)

1. **Wer entscheidet, wann ein ADM System „gut“ ist?**

Statistische Vorhersagen über Menschen

Zu 40% ein Krimineller….

- Wenn dieser Mensch eine Katze wäre und 7 Leben hätte, würde er in 3 davon wieder rückfällig werden…
- Nein!
- **Algorithmische Sippenhaftung**
  - Von 100 Personen, die „genau so sind wie dieser Mensch“, werden 40 wieder rückfällig;
  - Wir folgen einem **algorithmisch legitimierten Vorurteil**.



https://pixabay.com/de/m%C3%A4dchen-katze-k%C3%A4tzchen-mode-1048728/

Probleme von algorithmischen Entscheidungssystemen (ADM Systemen)

1. Wer entscheidet, wann ein ADM System „gut“ ist?
2. **ADM Systeme ergeben nur Wahrscheinlichkeiten, keine Wahrheiten.**

Können Algorithmen diskriminieren?

Gleichberechtigung

Wenn man auf Google nach „CEO“ sucht…

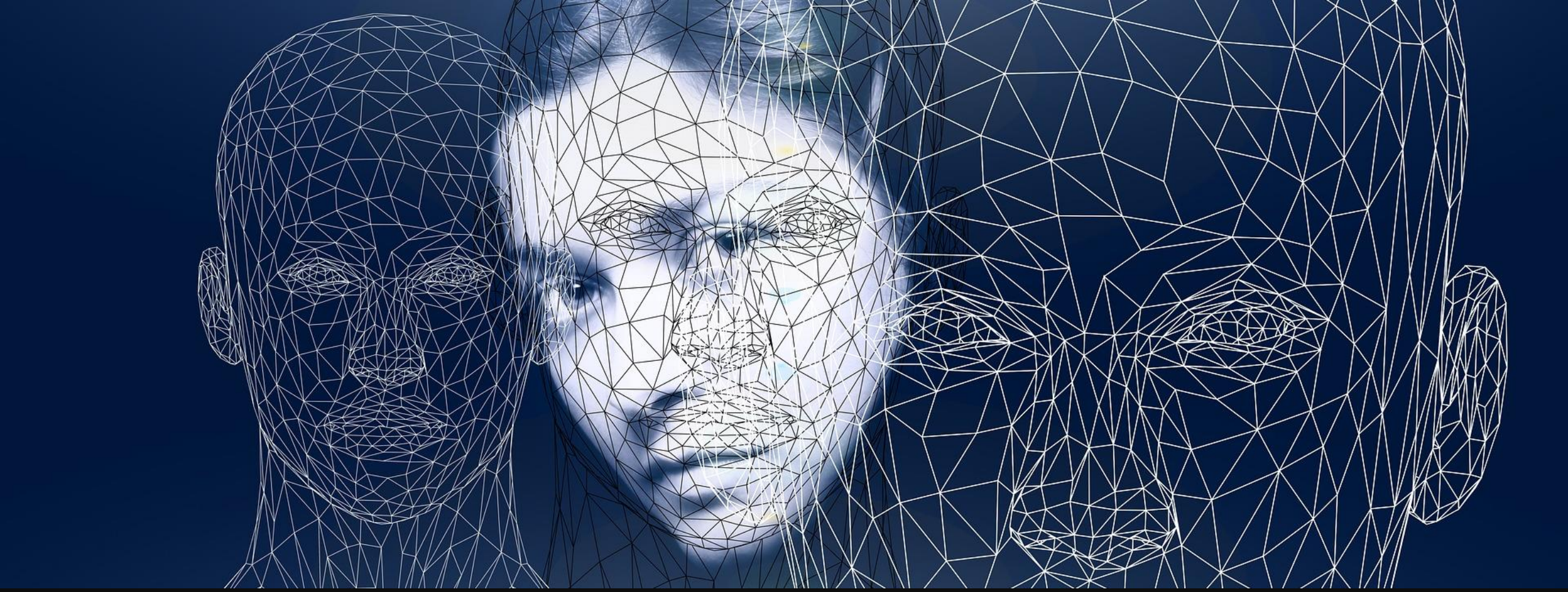
Und das, wenn ich auf Pixabay nach „Chef“ suche…

Diskriminierung

- Google zeigt weiblichen Surfern schlechtere Jobs an.
- Rückfälligkeitsvorhersage-algorithmen sind rassistisch.
- Diskriminierungen in Trainingsdaten werden „mitgelernt“.
- Wenn Trainingsdaten zu wenig Daten über Minderheiten enthalten, werden deren Eigenschaften nicht „mitgelernt“.

Probleme von algorithmischen Entscheidungssystemen (ADM Systemen)

1. Wer entscheidet, wann ein ADM System „gut“ ist?
2. ADM Systeme ergeben nur Wahrscheinlichkeiten, keine Wahrheiten.
3. **ADM Systeme können diskriminieren.**



Sozio-informatische Gesamtbetrachtung

Probleme der Einbettung der ADM in den sozialen Prozess

- **Aufmerksamkeitsökonomie** von Entscheiderinnen und Entscheidern.
- **„Best practice“** erfordert Nutzung der Software.
- **Delegation von Verantwortung!**
- Manchmal kann ein(e) falsch-negativ Beurteilte(r) **die Vorhersage prinzipiell nicht entkräften!**
  - Z.B. Krimineller im Gefängnis

Probleme von algorithmischen Entscheidungssystemen (ADM Systemen)

1. Wer entscheidet, wann ein ADM System „gut“ ist?
2. ADM Systeme ergeben nur Wahrscheinlichkeiten, keine Wahrheiten.
3. ADM Systeme können diskriminieren.
4. **ADM Systeme können soziale Prozesse verändern.**

Probleme von algorithmischen Entscheidungssystemen (ADM Systemen)

1. **Wer entscheidet, wann ein ADM System „gut“ ist?**
2. **ADM Systeme ergeben nur Wahrscheinlichkeiten, keine Wahrheiten.**
3. **ADM Systeme können diskriminieren.**
4. **ADM Systeme können soziale Prozesse verändern.**

Algorithmen in einer demokratischen Gesellschaft

Generell

Prinzipiell können algorithmische Entscheidungssysteme für sehr viele, schwierige Fragestellungen in derselben Art gebaut werden:

- Automatische Leistungsbewertung
- Kreditvergabe
- Schulische und universitäre Ausbildungen, die durch algorithmische Entscheidungssysteme unterstützt werden
- Gefährder-, Terroristenidentifikation
- …

Quis custodiet ipsos algorithmos

Der „Automated Decision Making“-TÜV vulgo: „Algorithmen TÜV“(Kenneth Cukier und Viktor Mayer-Schönberger: „Big Data“)

Gründung von „Algorithm Watch“

Lorena Jaume-Palasí, Rechtsphilosophin

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

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

ALGORITHM WATCH

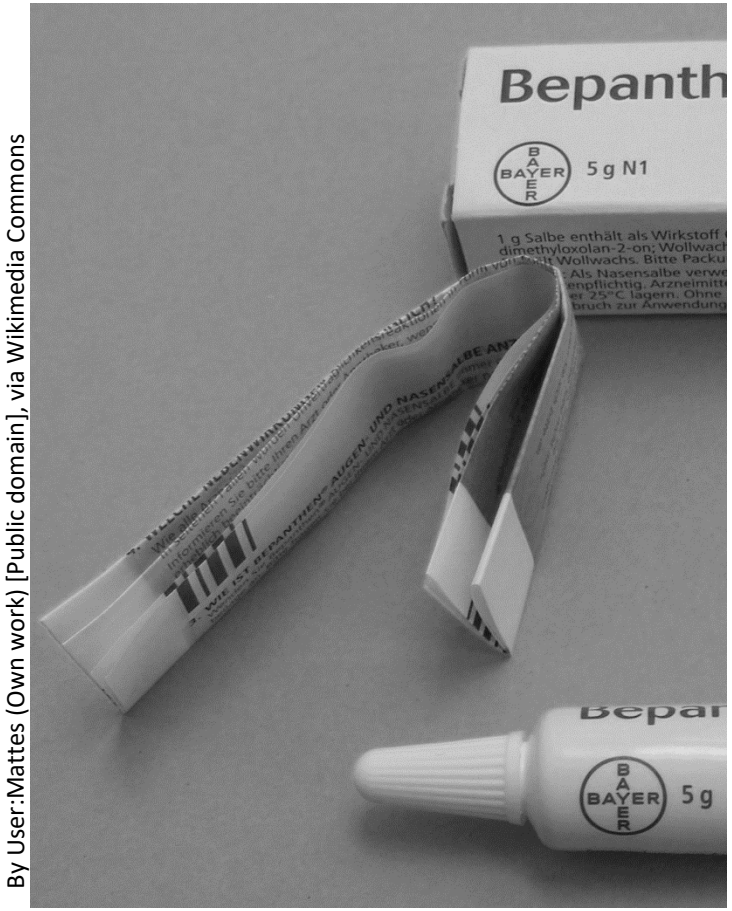
Gründung des Algorithm Accountability Labs an der TU Kaiserslautern

ALGORITHM ACCOUNTABILITY LAB

Wie könnte ein „Algorithmen-TÜV“ aussehen?

- 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



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

Welches Problem „kuriert“ der Algorithmus?

Was ist das Einsatzgebiet des Algorithmus, wo ist er geeignet?

Welche „Nebenwirkungen“ hat der Algorithmus durch seine Einbettung in einen sozialen Prozess?

Schlussformel

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

Weitere Literatur

Studie für die Bertelsmann-Stiftung (2018)
Katharina Zweig mit Sarah Fischer und Konrad Lischka
https://www.bertelsmann-stiftung.de/de/publikationen/publikation/did/wo-maschinen-irren-koennen/

Katharina Zweig (2018)
Auch Algorithmen können diskriminieren
https://merton-magazin.de/auch-algorithmen-koennen-diskriminieren?tags=Personalmanagement

Weitere Informationen

Broschüre der Bayerischen Landesmedienanstalt
Kostenlos zu beziehen von der BLM

Googlen nach „BLM Dein Algorithmus meine Meinung“

Prof. Dr. Katharina A. Zweig
zweig@cs.uni-kl.de
@nettwerkerin bei Twitter