Funktionsweise und Fehlerquellen bei selbstlernenden Algorithmen

Prof. Dr. Katharina Zweig, TU Kaiserslautern

zweig@cs.uni-kl.de

Algorithmische Entscheidungssysteme

Verständnis nötig für Sie als:

- Arbeitnehmer, Arbeitgeberin, Entscheider, Bürgerin.

Werden eingesetzt werden:

- In HR zur autom. Leistungsbewertung
- Zur Unterstützung von (Verwaltungs-)Prozessen, z.B. Hartz IV, Gefährderidentifikation
- Fahrerloses Fahren
- Entscheiden über Sie als Bürgerin und Bürger:
  - Filmempfehlungen, Suchmaschinenergebnisse, News Feed-Ergebnisse
  - Kreditwürdigkeit
  - ….

Am Beispiel der Rückfälligkeits-vorhersage für Kriminelle

Wie sagt man die Rückfallrate eines Verbrechers voraus?

Predictive Policing

Vorhersagen, wann und wo Straftaten wahrscheinlich sind basierend auf bisherigen Fällen.

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:

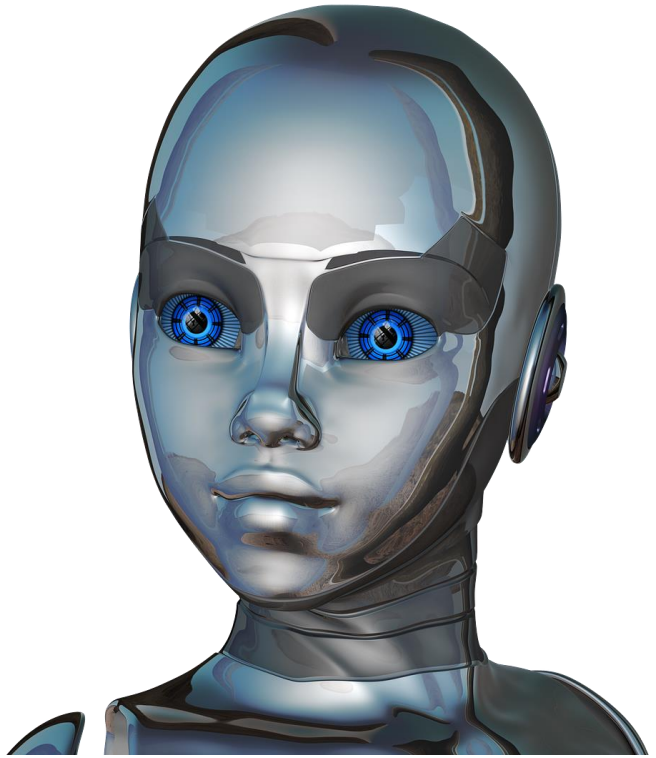
1) Oregon
2) Andere Bundesstaaten

Wie können Computer lernen?

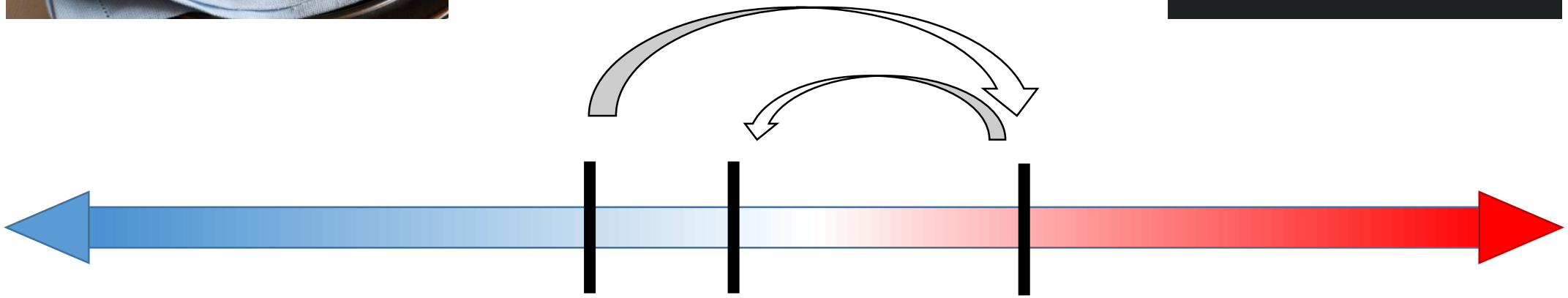
K-means clustering, Entscheidungsbäume und Regressionsansätze

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.

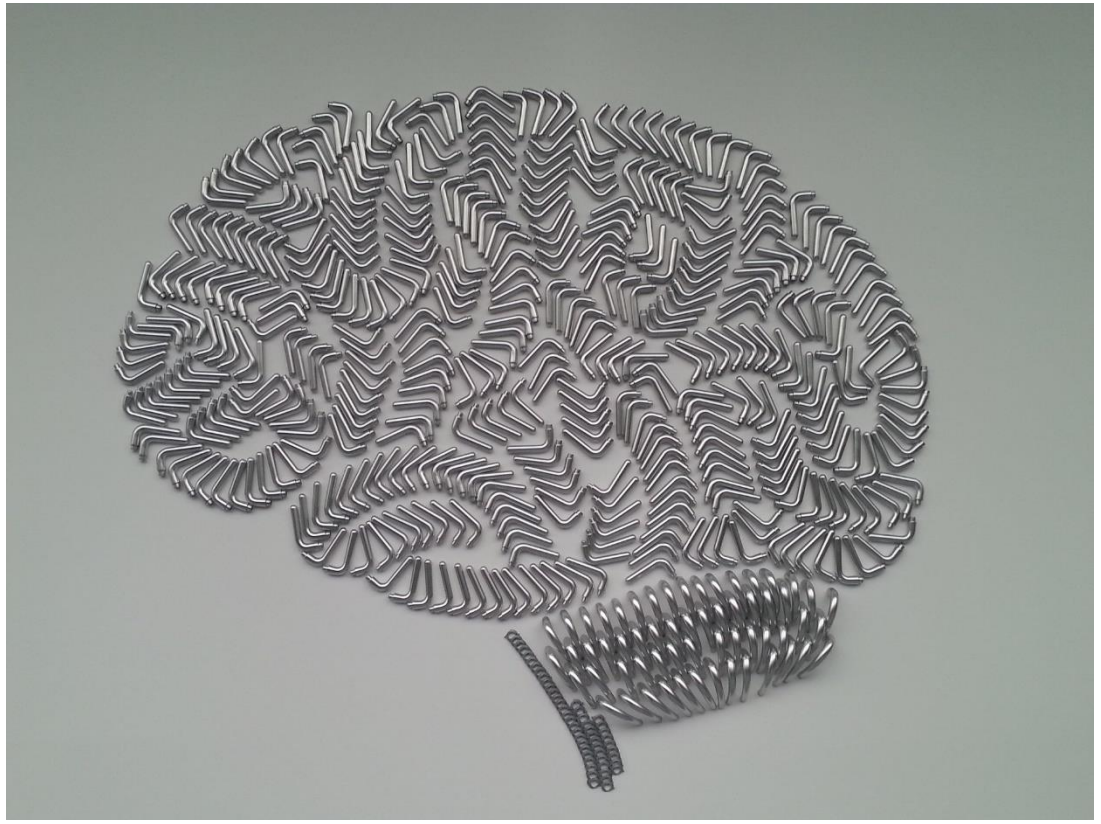


Sebastian lernt „heiss“ und „warm“



Sebastian lernt…

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

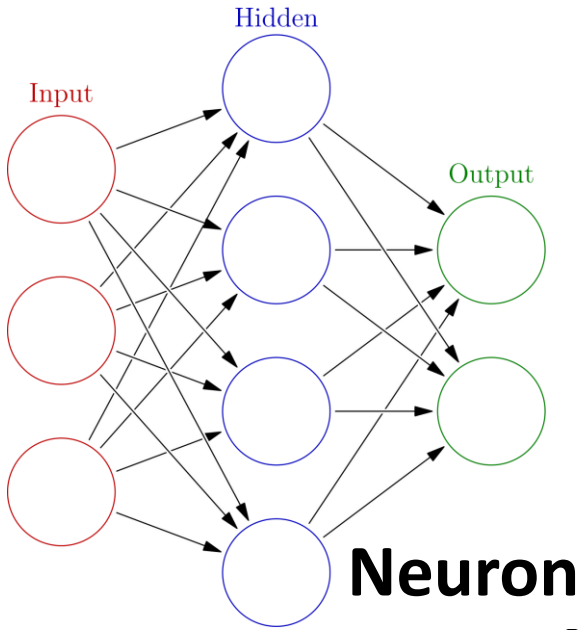


Computer lernen

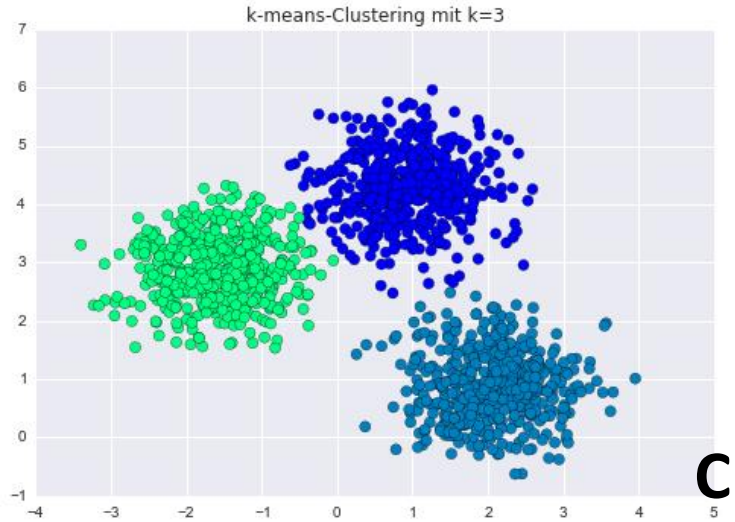
Damit ein Computer lernen kann, benötigt er ebenfalls eine **Struktur,** um Gelerntes abzuspeichern.

Optimal auch **Rückkopplung**.

Er lernt **generelle Regeln**.



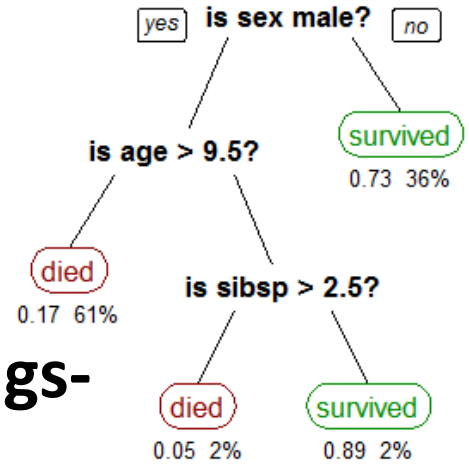
Neuronales Netz



Clustering

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

Formel



Entscheidungs-bäume

Lernen mit Formeln

Am Beispiel der **individuellen** Vorhersage von zukünftigem kriminellen Verhalten

Datengrundlagen

- Data Mining 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
  - Extra-Fragebogen
  - Aber bspw. nicht die (in den USA eindeutig zugec Zugehörigkeit.
- Wichtig: Beim Trainingsset ist bekannt, ob die Person rückfällig geworden ist oder nicht.

Regressionsansätze

- 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.
- Je höher die Zahl, desto höher die Rückfallwahrscheinlichkeit.
- Beispiel Formel:

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

Allgemein

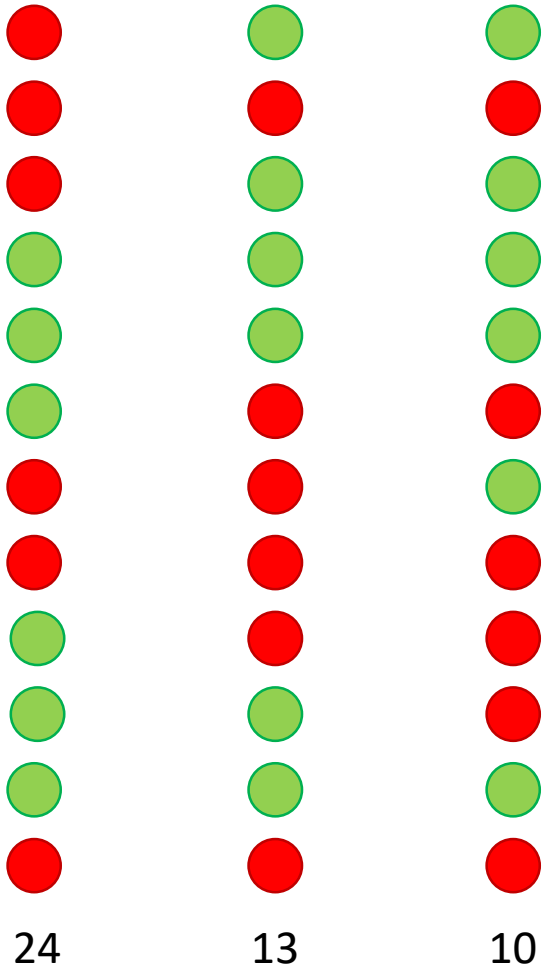
$$
\begin{aligned}
& w_1 * \text{ bisherige Verhaftungen} \\
- & w_2 * \text{ Anzahl Tage seit letzter Verhaftung} \\
+ & w_3 * \text{ (Wenn Mann, dann 1, sonst 0)} \\
+ & w_4 * \text{ (Wenn Raubüberfall, dann 1, sonst 0)} + \ldots
\end{aligned}
$$

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
- Bewertet jeweils, wie viele erwiesenermaßen Rückfällige möglichst weit oben stehen.
- Die Gewichtung, die das maximiert, wird für weitere Daten genommen.



Rote Kugeln symbolisieren rückfällige, grüne resozialisierte Personen.

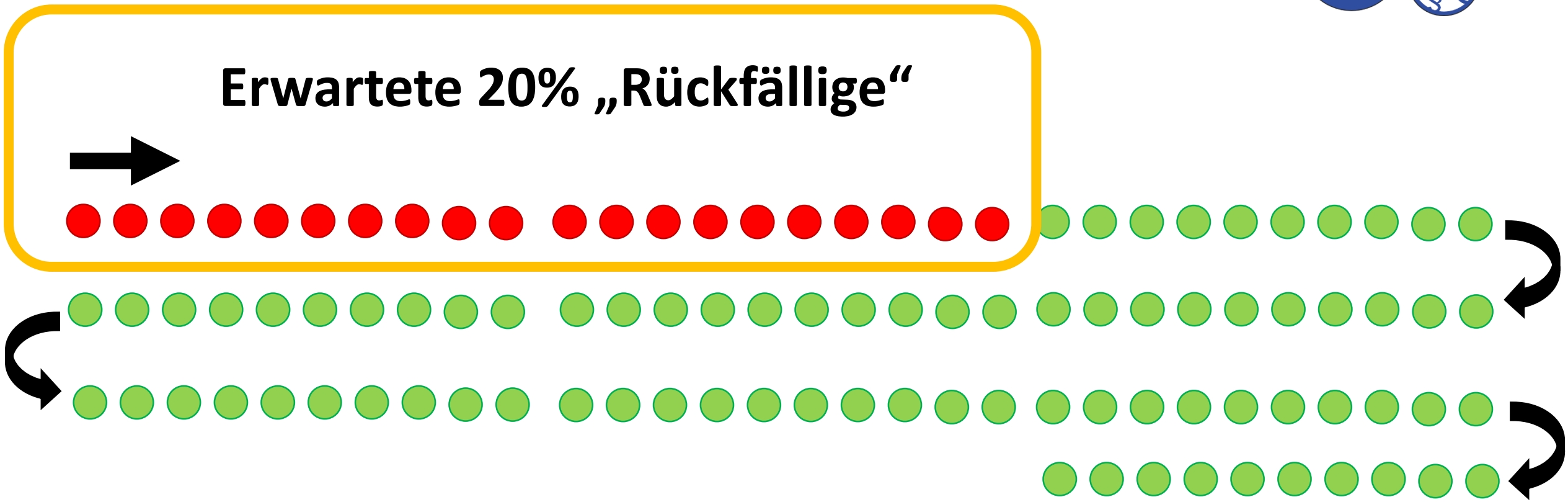
Optimale Sortierung: Alle roten oben, alle grünen darunter.

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

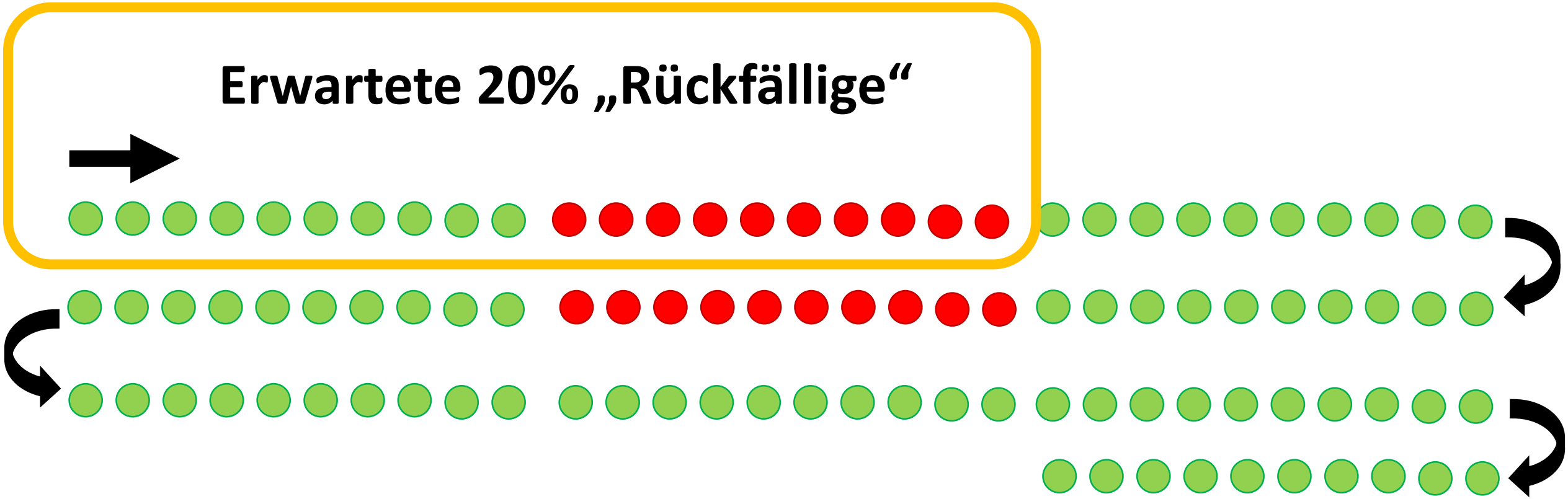
Oregon Recidivism Rate Algorithm

- 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 25% aller so gemachten Prognosen sind falsch!
  - Das klingt doch ganz gut, oder?
- 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%.

Optimale Sortierung



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



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!”

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.**

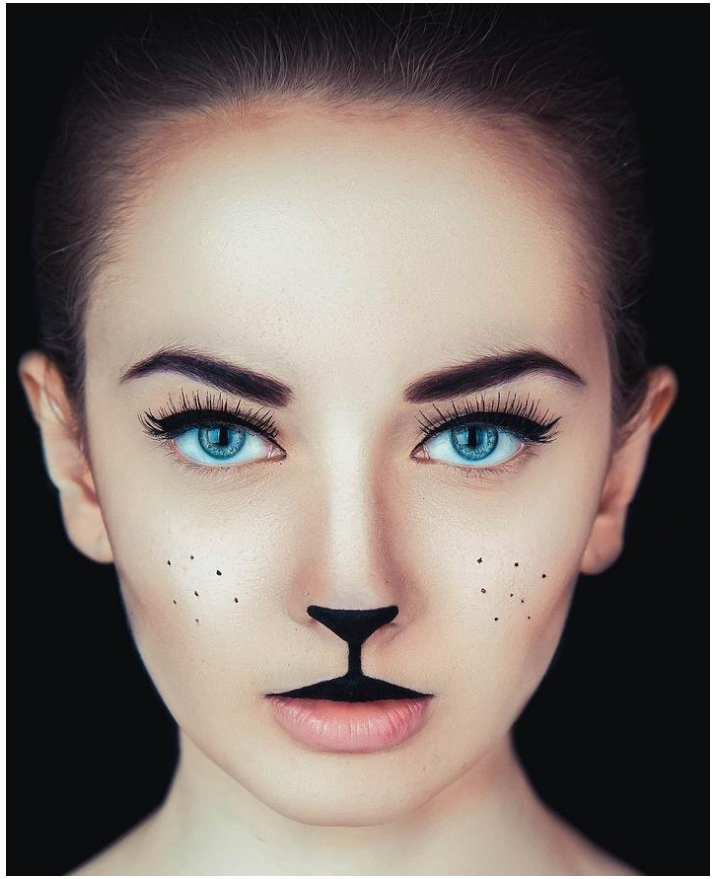
Versuchen fast immer, eine **kleine Gruppe** von Menschen zu identifizieren (Problem der **Unbalanciertheit**)

Wenn es **einfache Regeln zur Entscheidungsfindung gäbe, kennten wir sie schon.**

Statistische Vorhersagen über Menschen

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“.



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

Sozio-informatische Gesamtbetrachtung

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, …

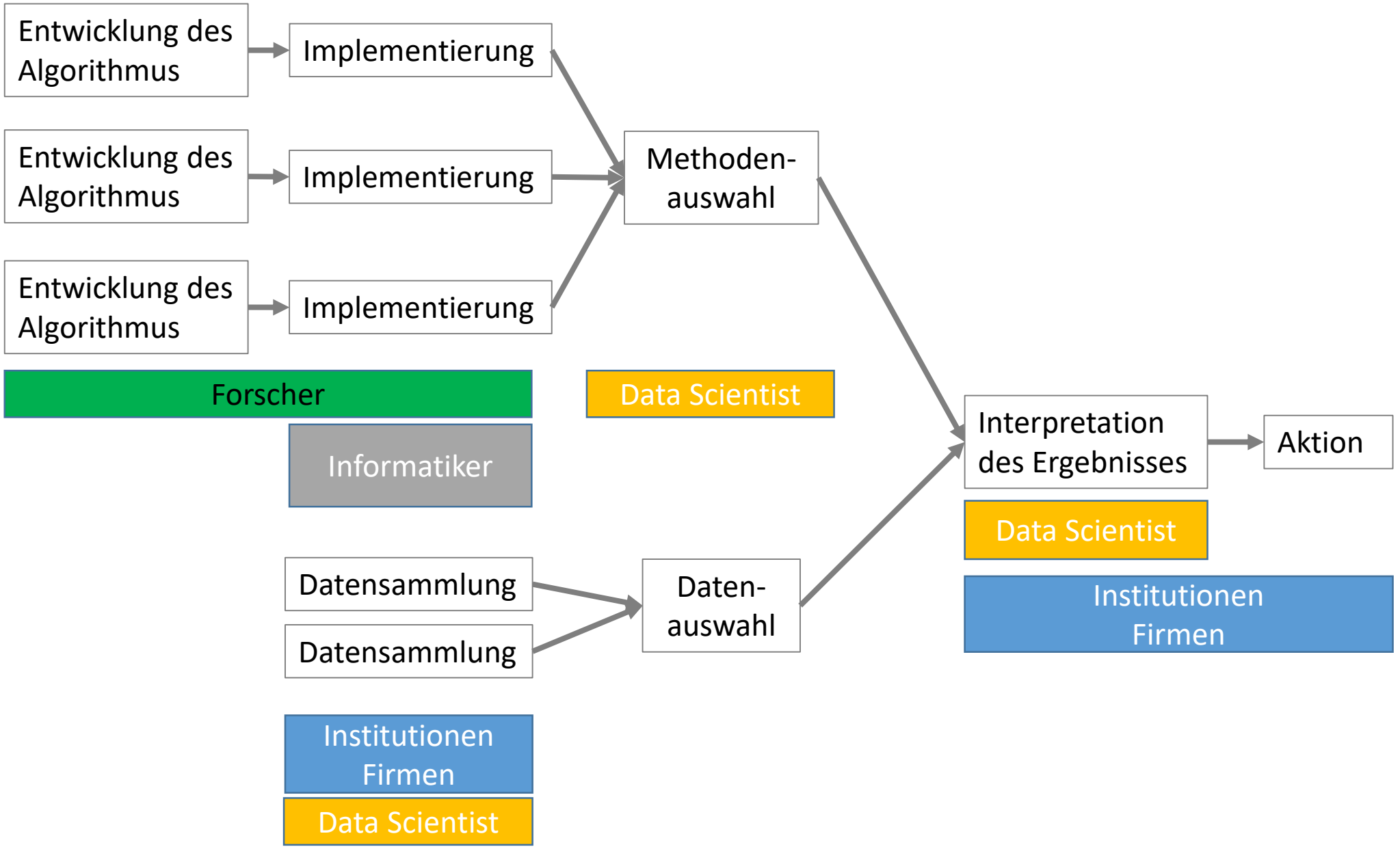
Generell

Gilt auch für:

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

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“(Kenneth Cukier und Viktor Mayer-Schönberger: „Big Data“)

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.

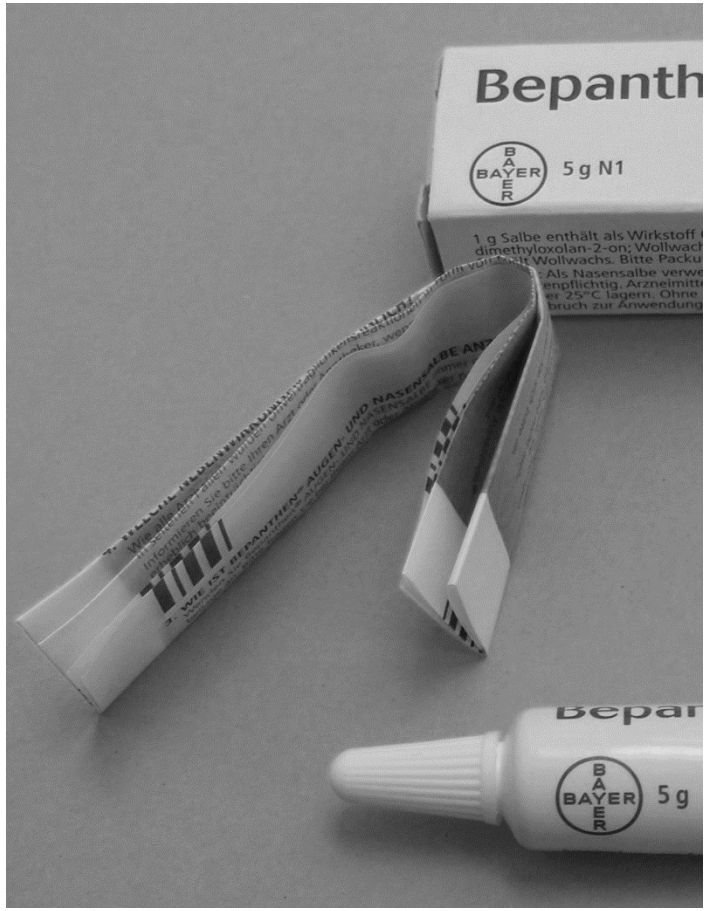
Prof. Dr. K.A. Zweig, Junior Fellow der Gesellschaft für Informatik, Digitaler Kopf 2014, TU Kaiserslautern

ALGORITHM WATCH

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.