Wie kommt die Ethik in den Rechner?

Prof. Dr. K.A. Zweig
TU Kaiserslautern
Algorithm Accountability Lab
@nettwerkerin

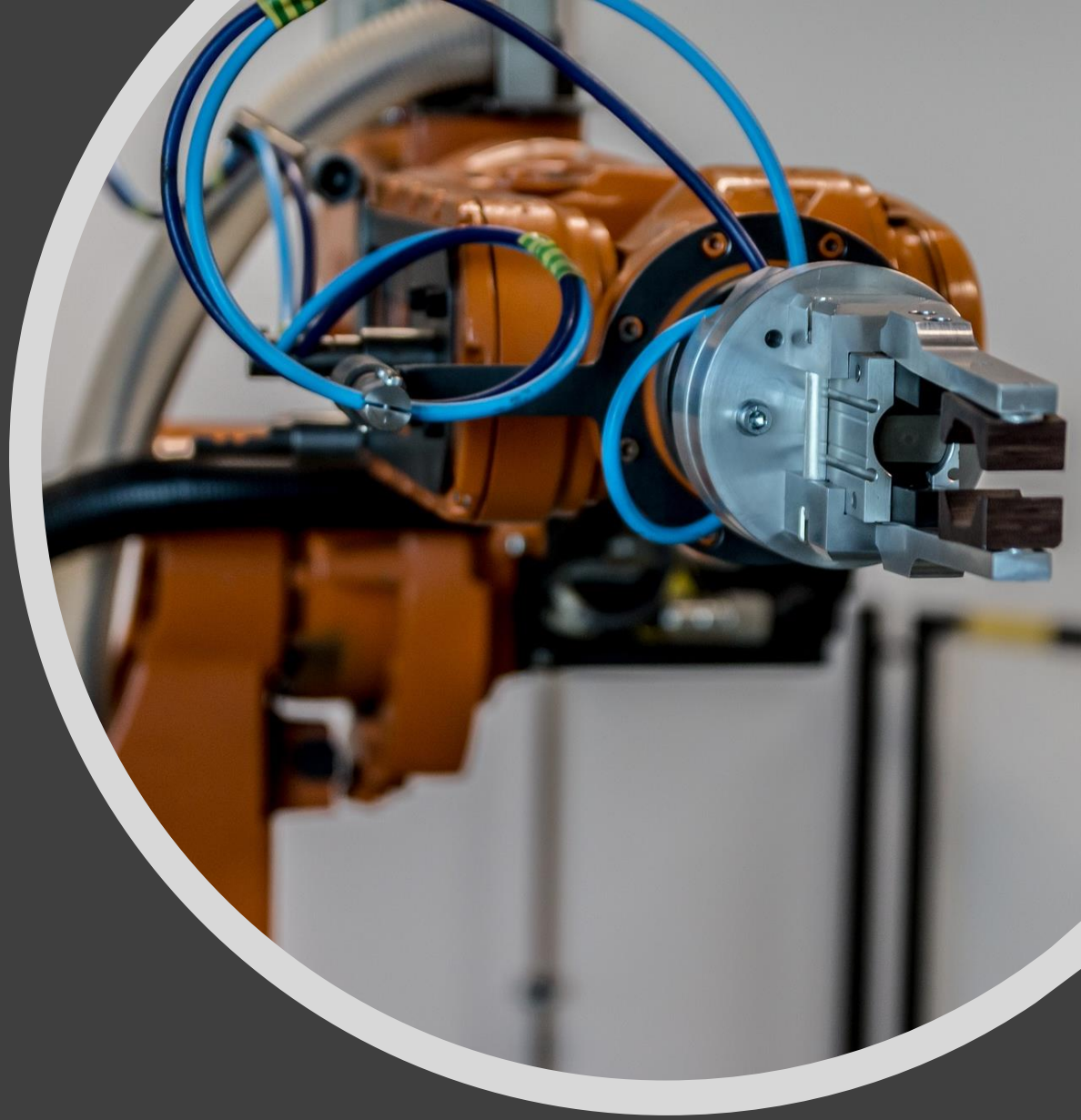
Maschinelles Lernen

Software, die aus Daten der Vergangenheit Entscheidungsregeln ableitet für zukünftige Daten.

Die Software trifft dann mit den gelernten Regeln Entscheidungen über neue Situationen.

Wann muss das auf technischer Ebene kontrolliert und reguliert werden?

Kontrolle von algorithmischen Entscheidungssystemen



Maschinelles Lernen muss um so stärker kontrolliert und reguliert werden, je höher das durch die Softare mögliche individuelle und gesamtgesellschaftliche Schadenspotenzial ist.

I.A. sind Entscheidungen über Objekte, z.B. im Produktionsprozess, nicht kritisch und bedürfen keiner technischen Kontrolle und Regulierung.

Wie „lernt“ das System von Daten?

DIY:

Sie sind heute meine „Support Vector Machine“

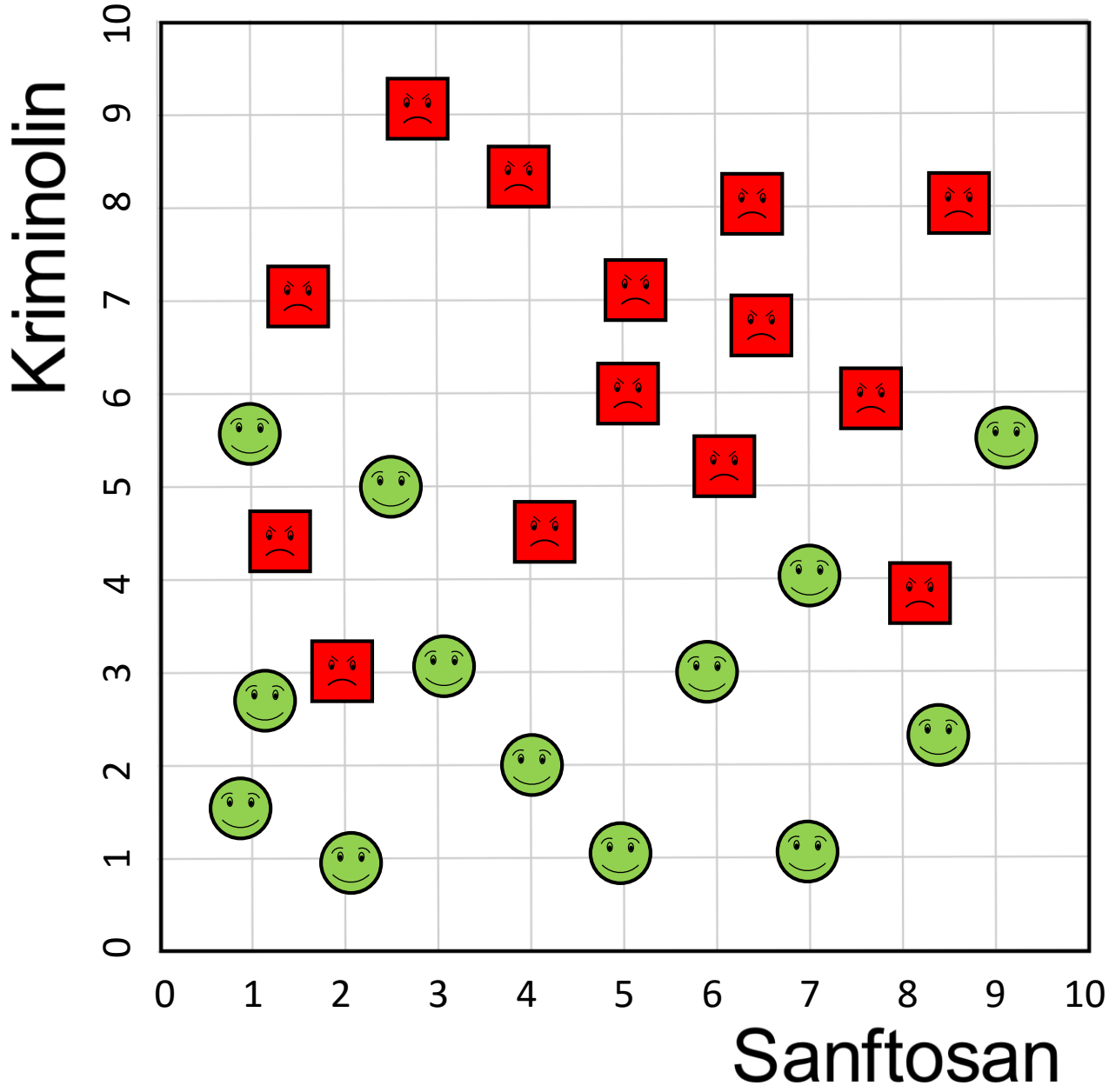
Bösartige Kriminelle

Unschuldige Bürger

Legen Sie den Holzspieß so zwischen die Smileys, dass die roten möglichst gut von den grünen getrennt sind. Kleben Sie ihn fest.

Gratulation: Sie haben eine Support Vector Machine trainiert!

Der dient nun als Entscheidungsregel, ob eine Person als kriminell gilt oder unschuldig zu sein scheint.

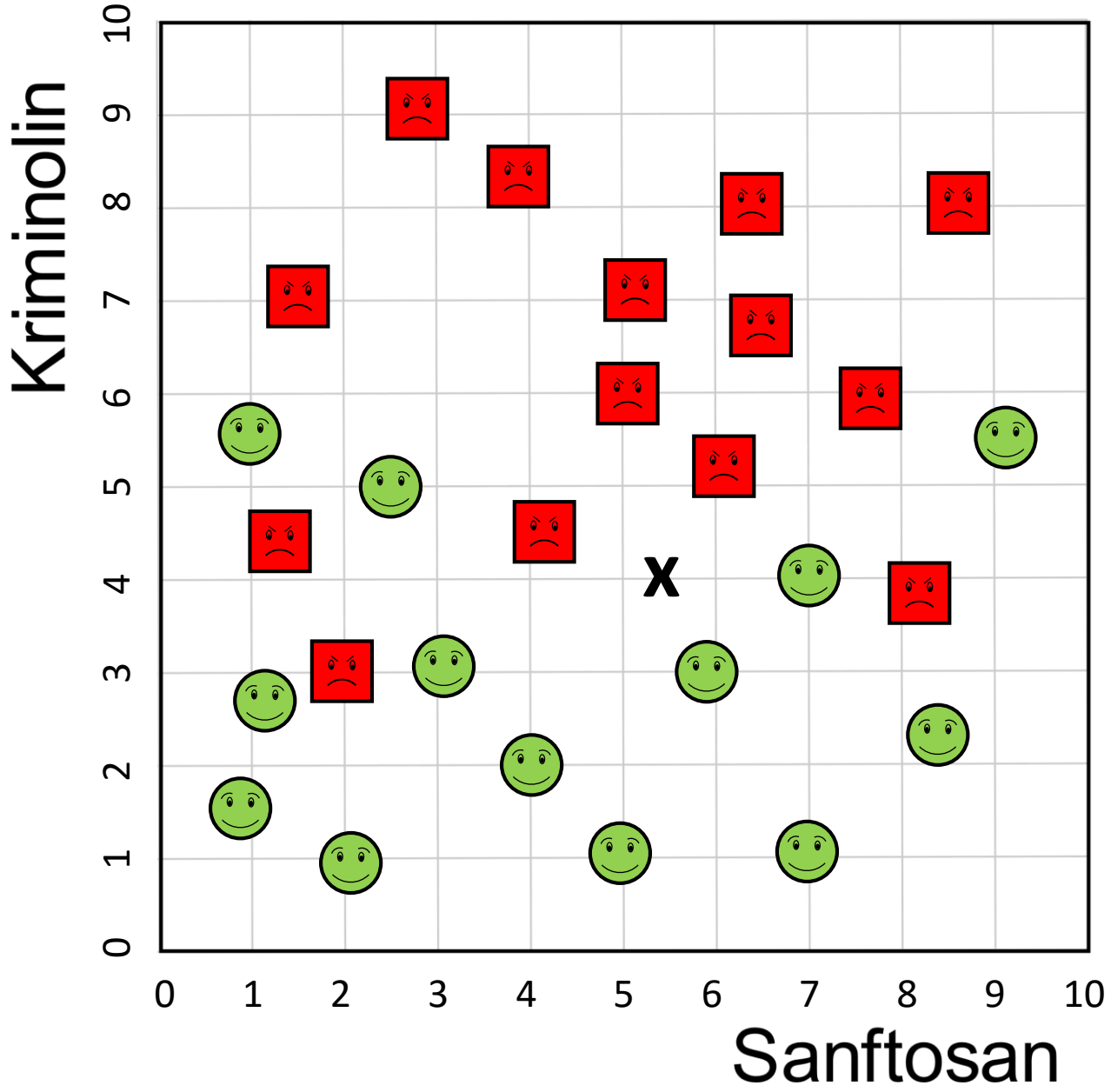


Bösartige Kriminelle

Unschuldige Bürger

Bewerten Sie Frau Müller:

5.5 Sanftosan
4.0 Kriminolin

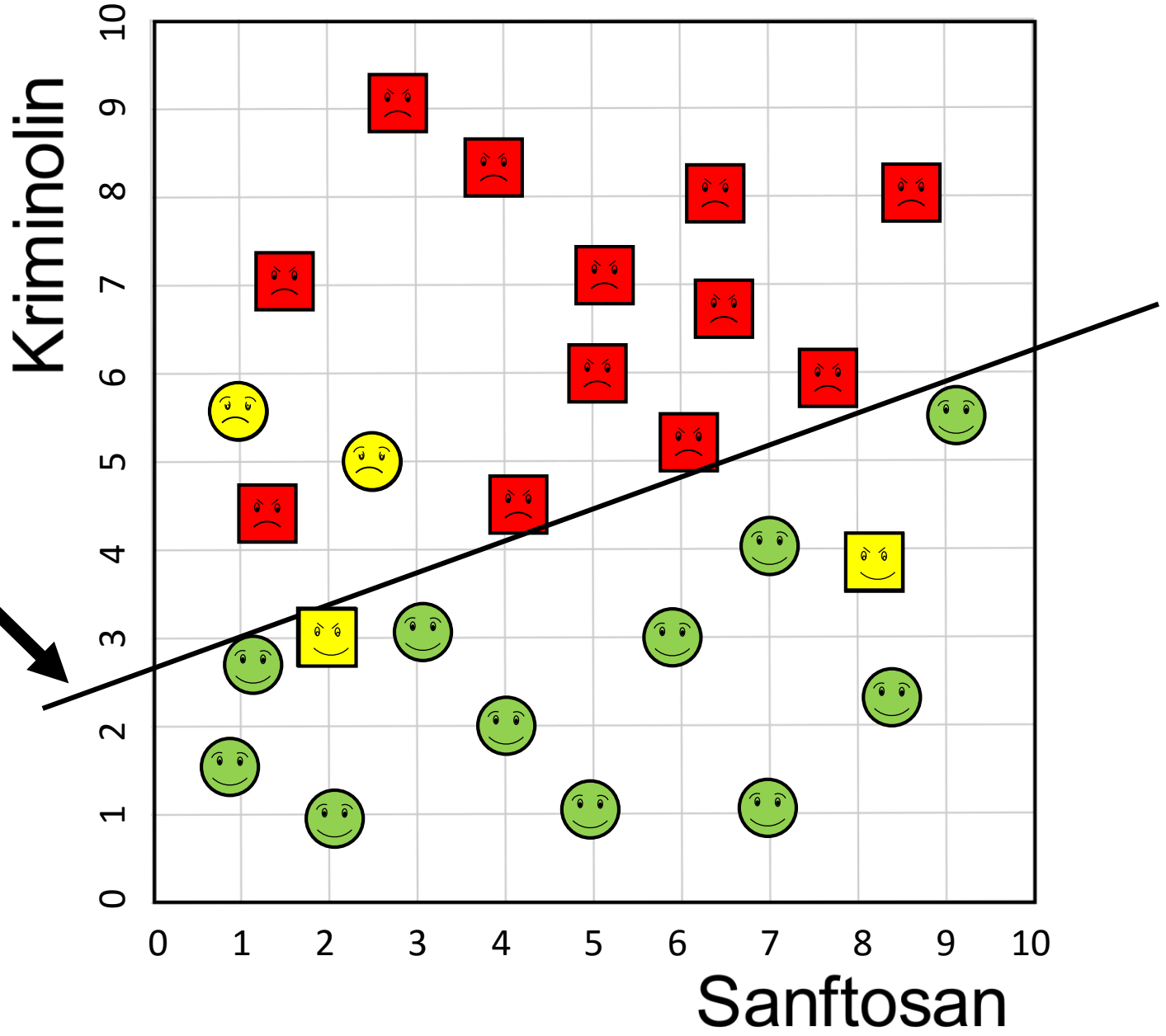


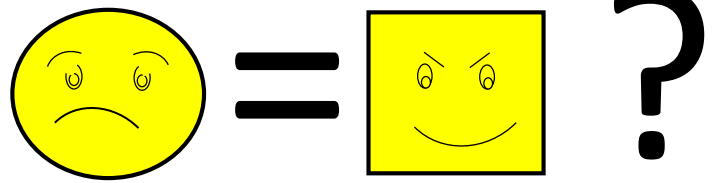
Eine der möglichen Trennlinien

Alle möglichen Trennlinien erzeugen Fehler:

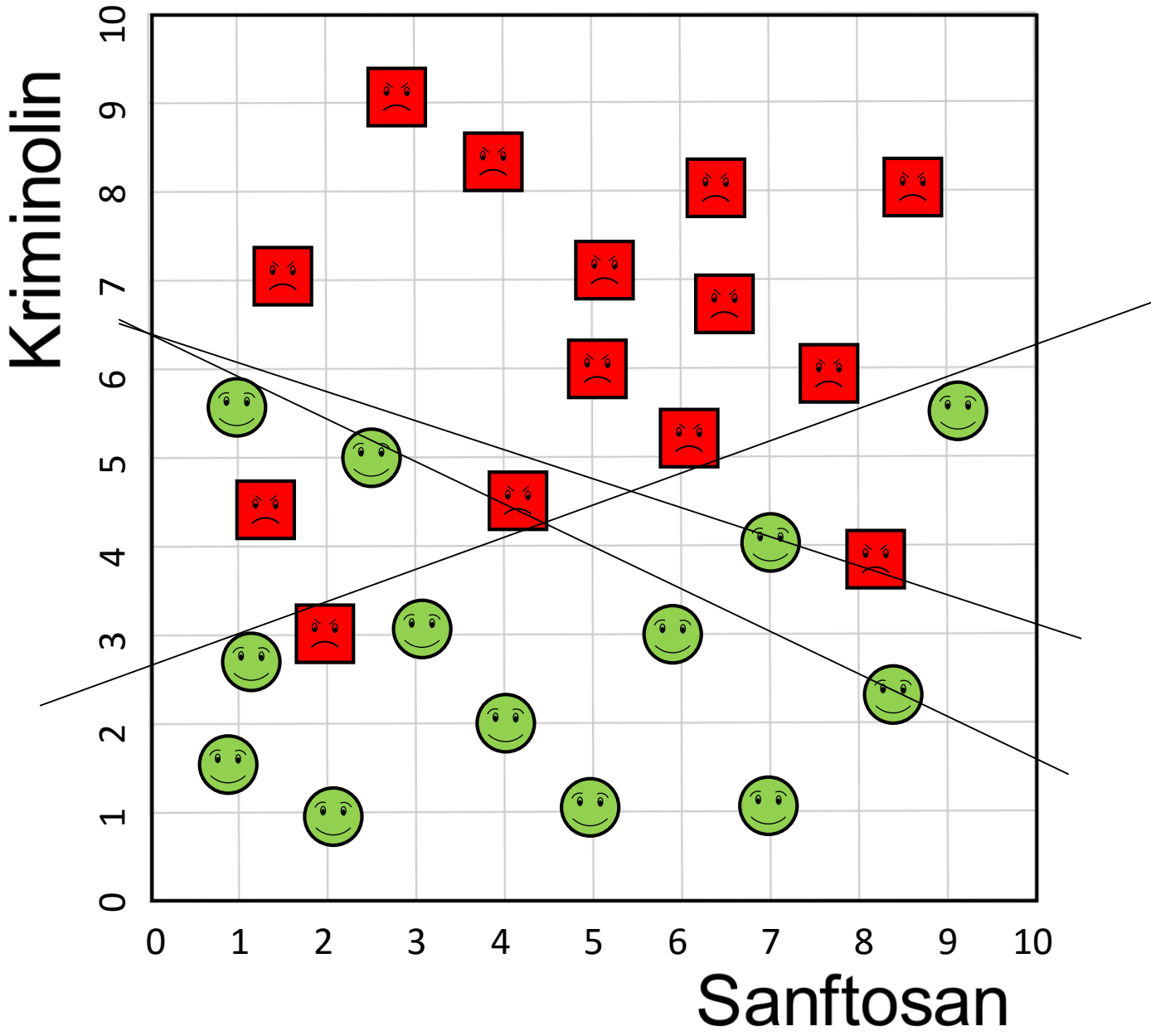
Bösartige Kriminelle, die unentdeckt bleiben

Unschuldige Bürger, die für kriminell gehalten werden





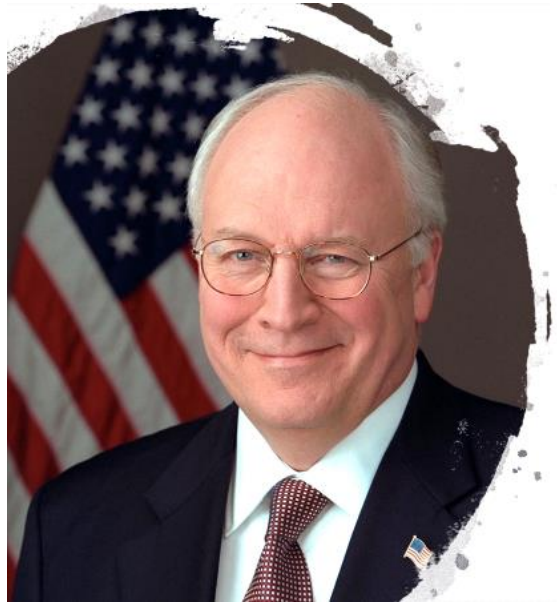
Wenn beide Fehler als gleich schlimm gelten, gibt es mehrere optimale Trennlinien mit möglichst wenigen Fehlern.




Sind beide Arten
von Fehler
gleich zu bewerten?

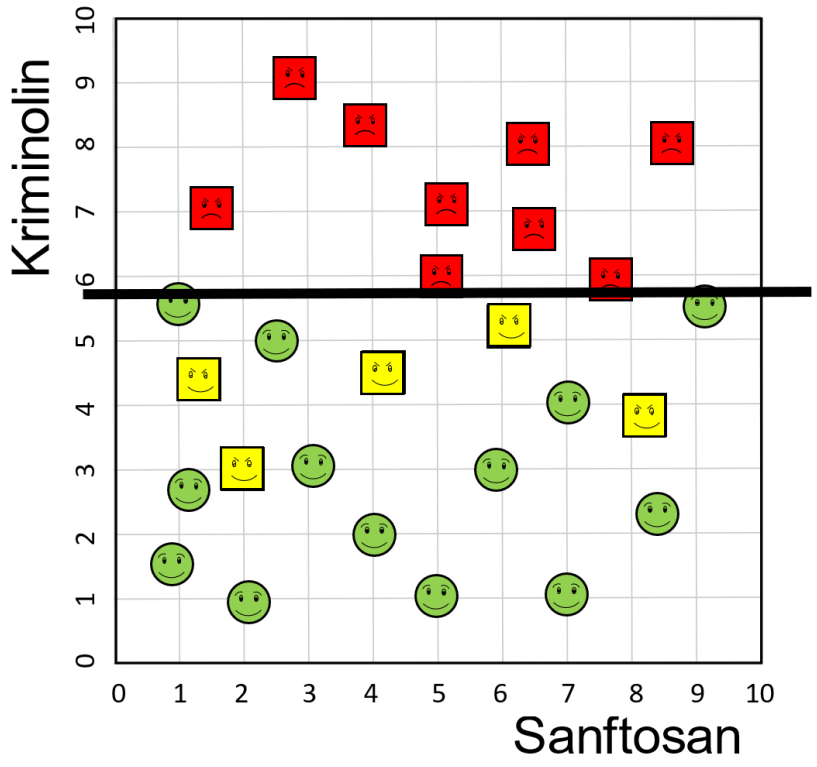
„It is better that ten guilty persons escape than that **one** innocent suffer.“

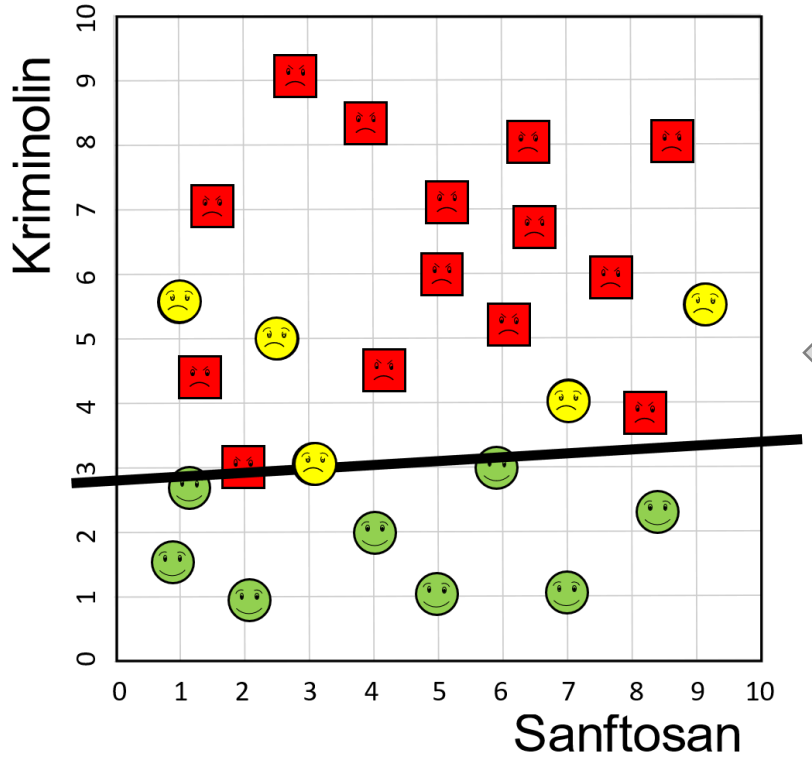
William Blackstone,
Rechtsphilosoph, 1760



"I am more concerned with bad guys who got out and released than I am with a few that, in fact, were innocent."

Dick Cheney, ehemaliger
Vizepräsident der USA,





1. Beobachtung

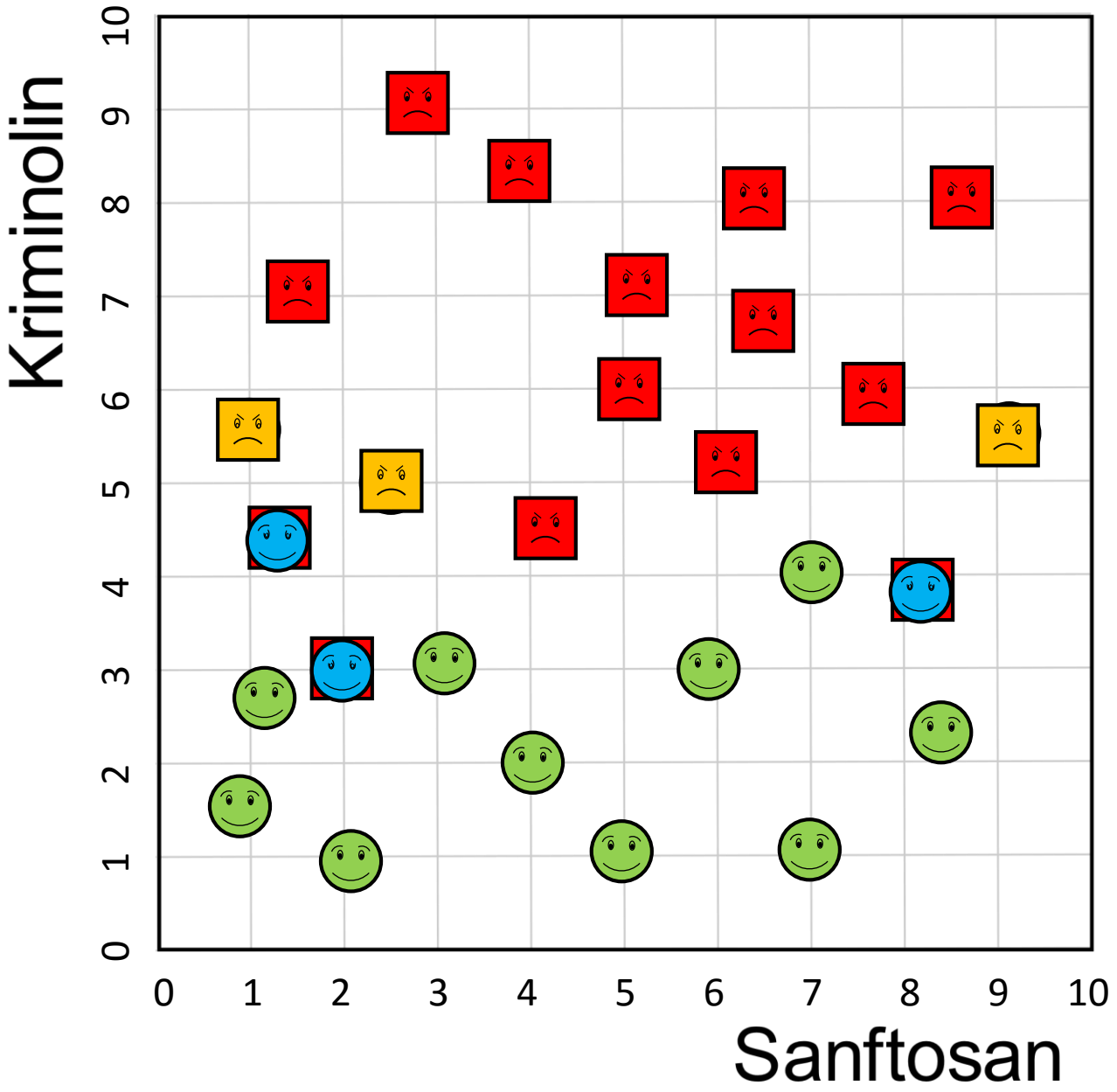
Was durch eine künstliche Intelligenz
optimiert werden soll,
ist eine gesellschaftliche Entscheidung!

Datenqualität

Noch nicht entdeckte Finanzbetrüger

Unschuldig im Gefängnis

Falsche Datenpunkt-zuordnungen haben Einfluss auf das Training der Support Vector Machine und damit auf die nachfolgenden Entscheidungen.



2. Beobachtung

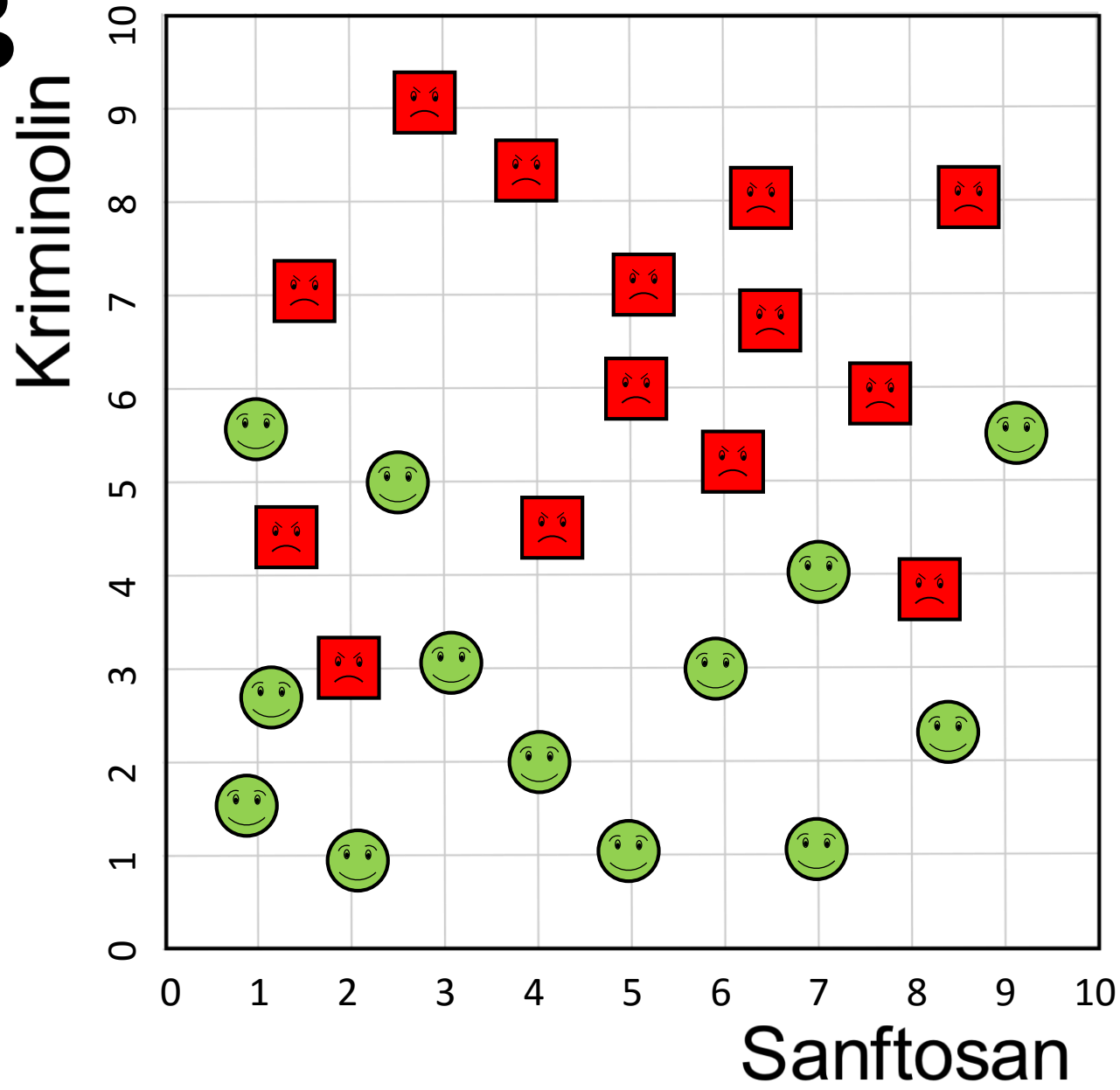
Wie gut die Maschine lernt, ist direkt abhängig von der Qualität der Daten.

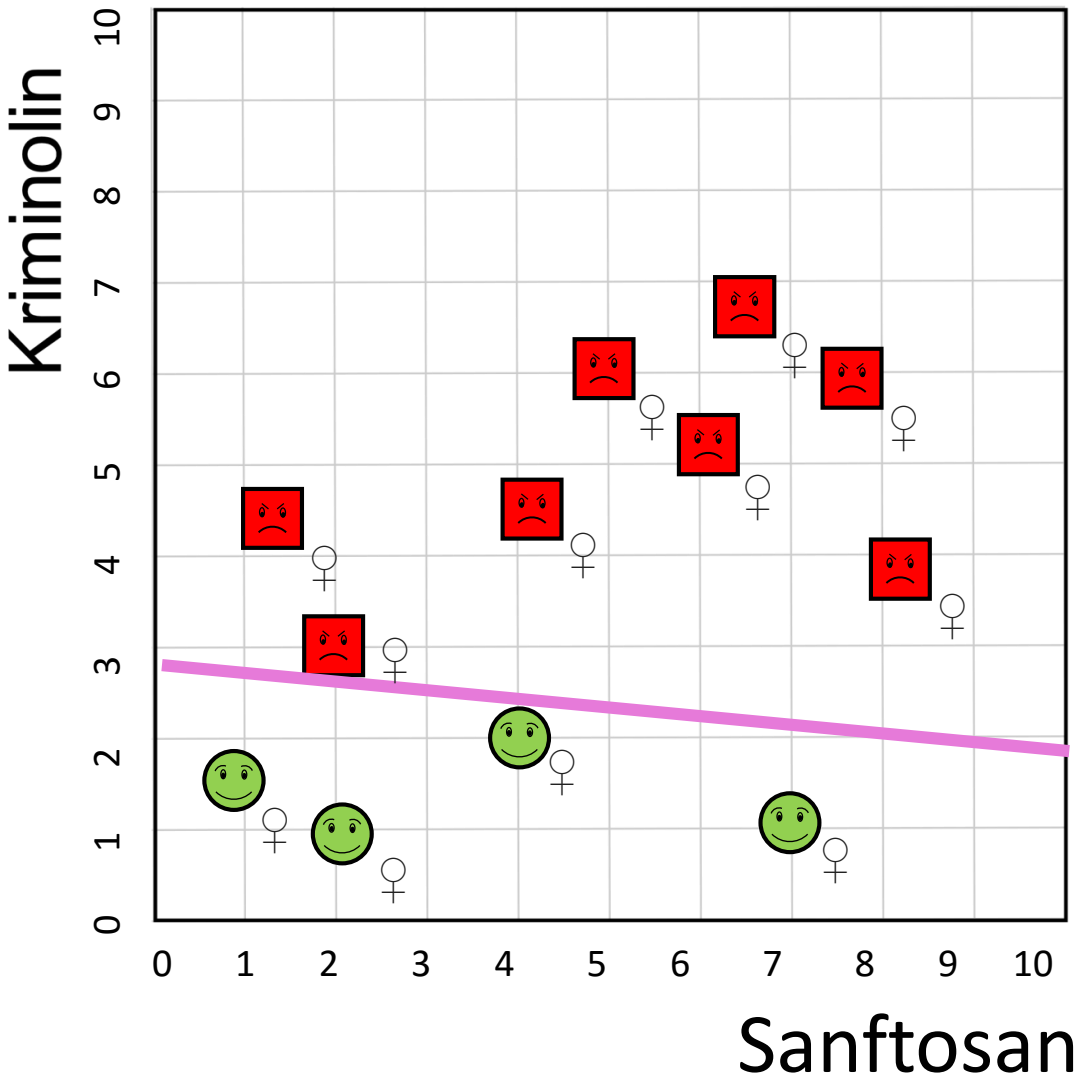
Diskriminierung

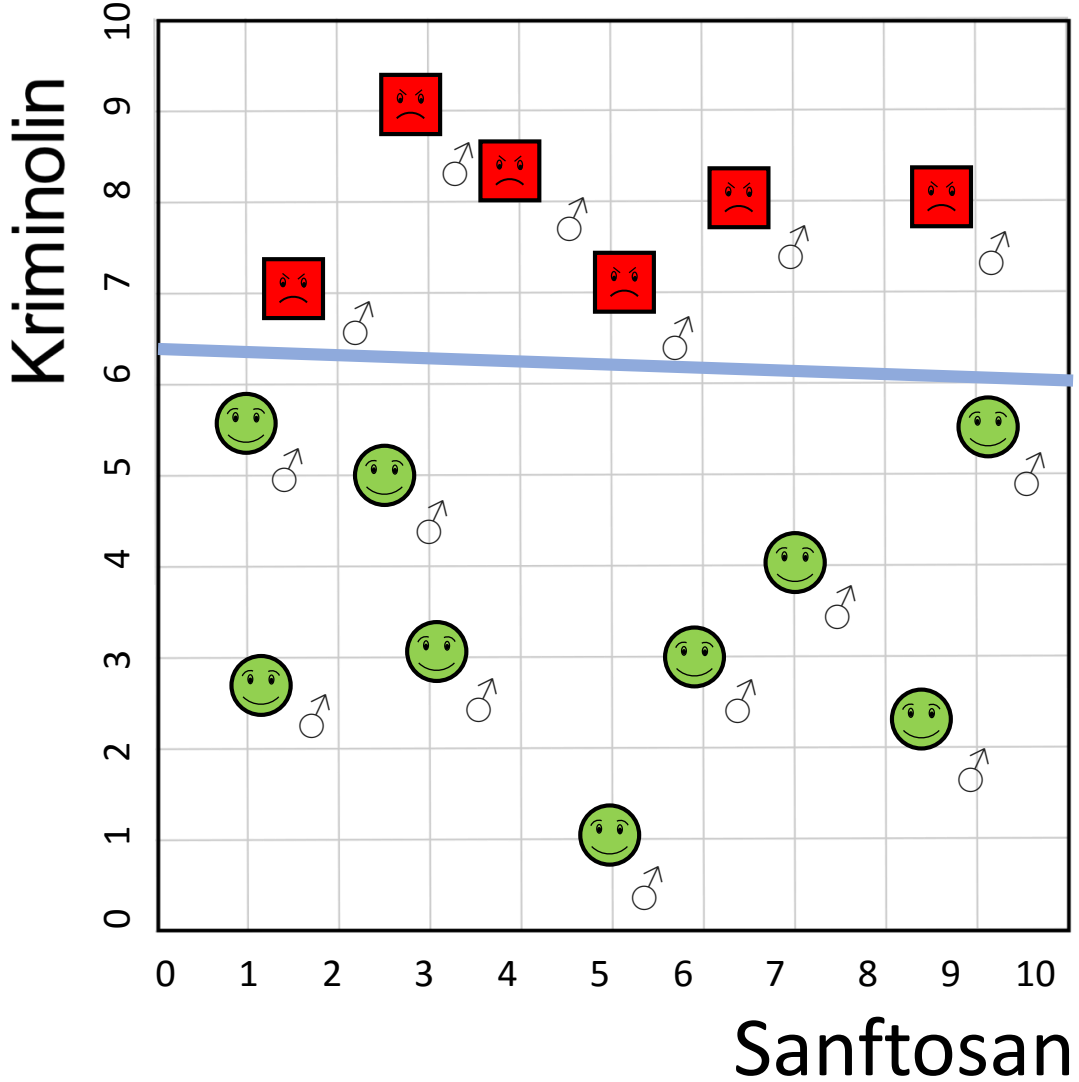
Kann ein Computer diskriminieren, wenn maschinelles Lernen verwendet wird?

Ja!

Auf der nächsten Seite trennen wir die Datenpunkte auf der rechten Seite auf in männliche und weibliche Personen und trainieren für jede Teilmenge jeweils eine Support Vector Machine.







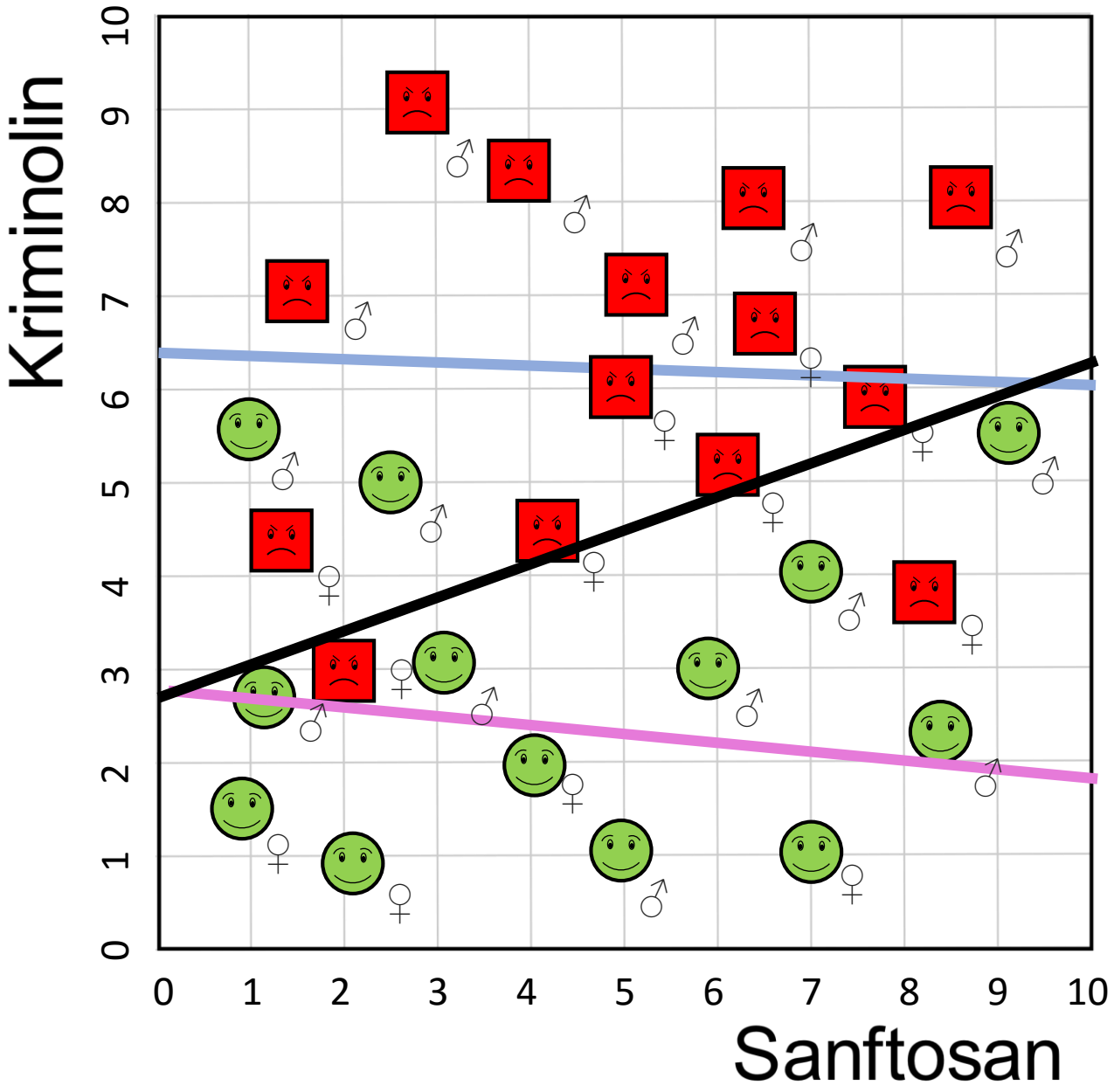
Ergebnis:

In diesem fiktiven Beispiel wird für jede Teilgruppe eine optimale Entscheidungsregel ohne Fehler gefunden.

Wirft man dagegen beide Gruppen zusammen, diskriminiert die trainierte Support Vector Machine

Männer:

Zwei weibliche Kriminelle gelten als unschuldig, zwei unschuldige Bürger als kriminell.



3. Beobachtung

Eine geschützte Information kann wichtig sein, um bessere Entscheidungen zu treffen. Diskriminierung wird nicht per se dadurch vermieden, dass die Information vorenthalten wird.

Ethische Entscheidungen im maschinellen Lernen

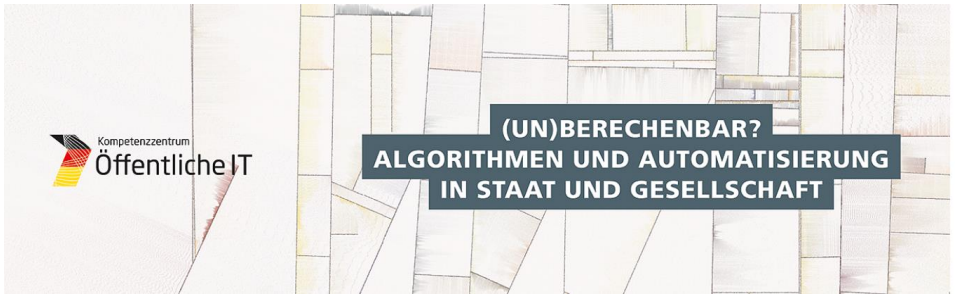
- Was genau „gelernt“ (optimiert) werden soll, ist eine gesellschaftliche Frage, wenn es um Entscheidungen über Menschen und gesellschaftliche Teilhabe geht.
- Ob Daten als Grundlage für eine soziale Fragestellung geeignet sind, muss dann auch die Gesellschaft entscheiden.
- Auch wahrhaftige Daten stellen immer nur einen Ausschnitt aus der Wirklichkeit dar – sie bedürfen der Einordnung und Interpretation.
- Die Frage nach Diskriminierung, ihrer Entdeckung und ihres Ausgleichs bedarf der gesellschaftlichen Diskussion – unabhängig davon, wer die Entscheidung trifft – Mensch oder Maschine.


Schlussfolgerung

Wie kommt die Ethik in den Rechner?

Über Sie,
über mich,
über uns.

Weitere Informationen



1. Studie für die Bertelsmann-Stiftung:
   Zweig, Fischer & Lischka: „Wo Maschinen irren können“ (Serie AlgoEthik, No. 4, 2018)
2. Zwei Kapitel im Sammelband (Un)Berechenbar? des Fraunhofer FOKUS, Kompetenzzentrum ÖFIT, 2018
   1. Zweig & Krafft: „Fairness und Qualität algorithmischer Entscheidungen“
   2. Krafft & Zweig: „Wie Gesellschaft algorithmischen Entscheidungen auf den Zahn fühlen kann“
3. Studie für die Konrad-Adenauer-Stiftung „Algorithmische Entscheidungen: Transparenz und Kontrolle“ (Zweig, erscheint 2019)
4. Studie vom Fraunhofer FOKUS, Kompetenzzentrum Öffentliche IT (ÖFIT): Opiela, Mohabbat Kar, Thapa & Weber: Exekutive KI 2030 – Vier Zukunftsszenarien für Künstliche Intelligenz in der öffentlichen Verwaltung, 2018)

Kontakt

Prof. Dr. Katharina A. Zweig
Algorithm Accountability Lab
Gottlieb-Daimler-Str. 48
67663 Kaiserslautern

aalab.informatik.uni-kl.de

zweig@cs.uni-kl.de
@nettwerkerin bei Twitter