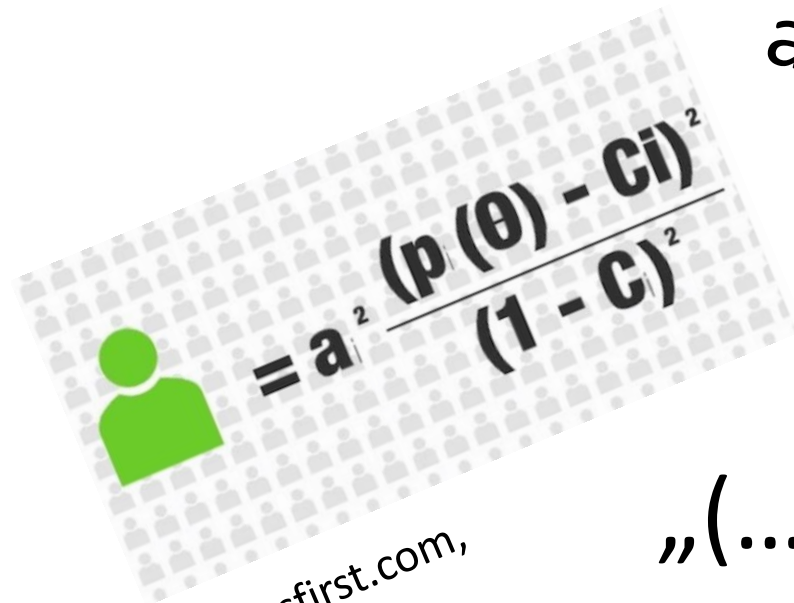


Interdisciplinary aspects of designing and implementing algorithmic decision making in societal processes

Prof. Dr. Katharina Zweig

TU Kaiserslautern
Head of the Algorithm Accountability Lab
Co-Founder of AlgorithmWatch

„Employment assessment software“


$$= a^2 \frac{(p(\theta) - ci)^2}{(1 - c)^2}$$

Assessfirst.com,
16.11.2017

„(...) with the availability of good data, the predictive possibilities are virtually unlimited (...)“

<https://www.inostix.com/predict-hiring-success/>
16.11.2017

Let's take the emotion out of the process and replace it with a data-driven approach...“

iNostix (by Deloitte),
16.11.2017

@netwerkerin
Prof. KA Zweig
TU Kaiserslautern

Will this criminal recidivate?



People so irrational, so sad!

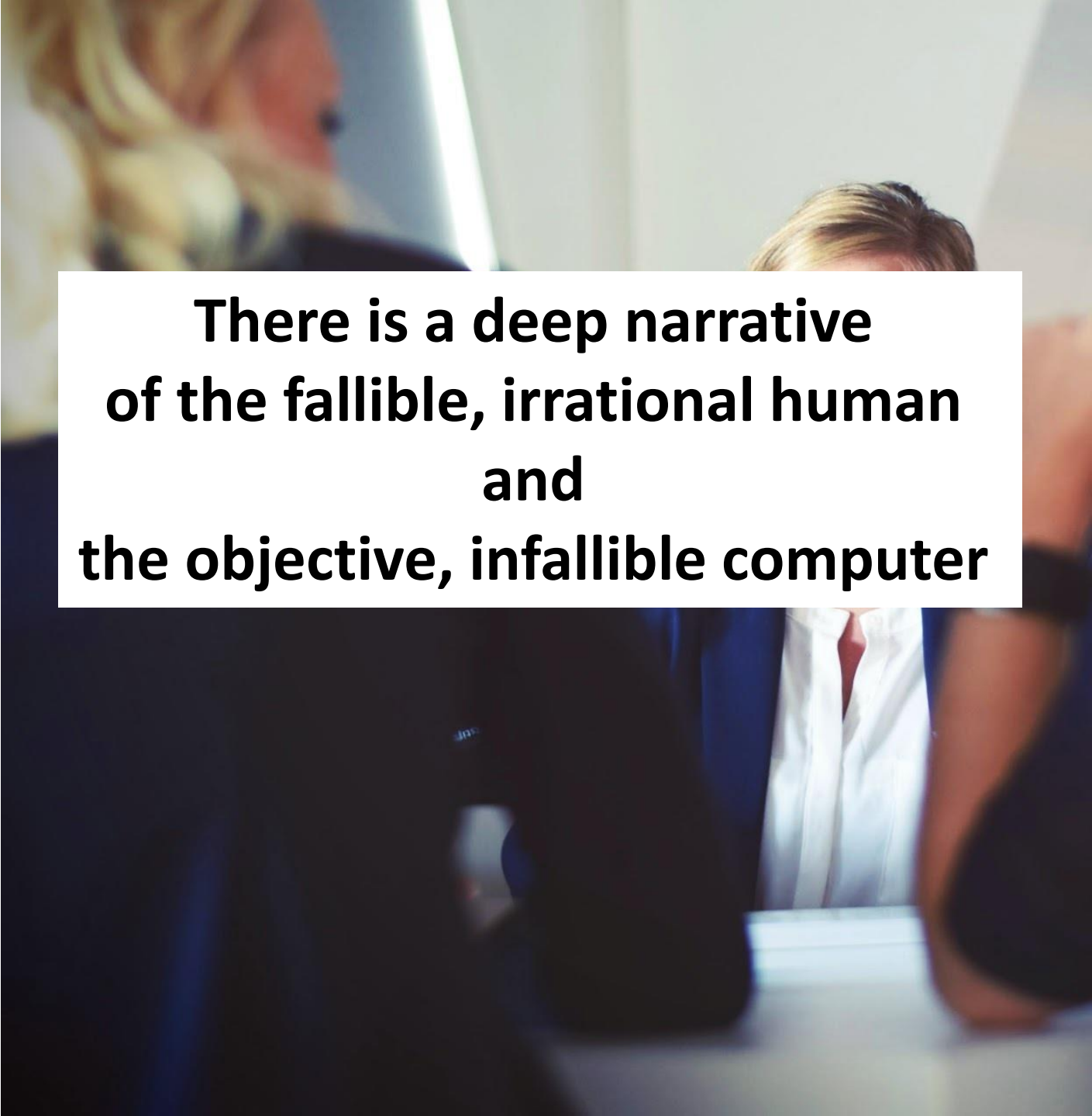
- Judges decide regularly over early release proposals.
- Study showed that judges take less risk the longer the time after the last break¹.
- Enormous number of research like this.
- Conclusion: Humans are biased and irrational.

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



Would computers make better decisions?

- The first states are testing out Algorithmic Decision Making or Algorithmic Decision Support Systems (ADM systems)¹.
- Possible discriminating features can be hidden from them.
- They are objective and almost failure free.
- (objective here := „constantly the same decision when the same information is given“)

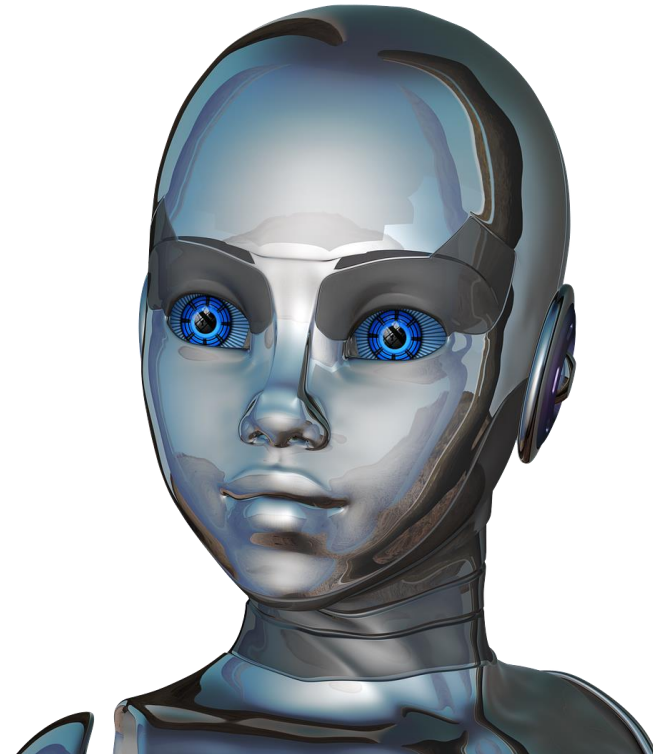


**There is a deep narrative
of the fallible, irrational human
and
the objective, infallible computer**

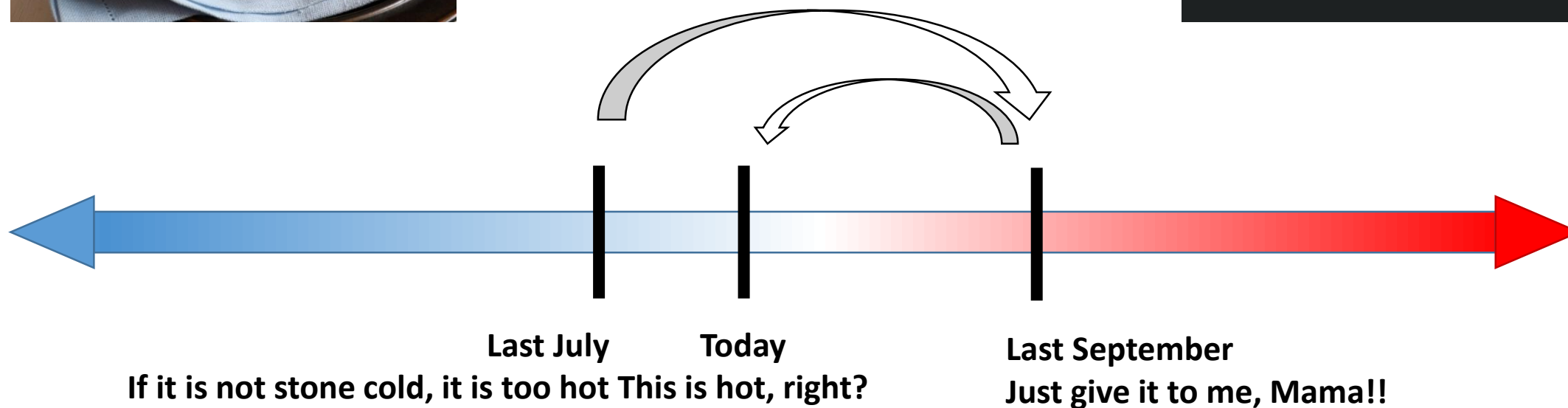
Can Computers learn?

Definition of learning used here

- Recognize a situation and show the behavior taught to you in these situations.
- Recognize that a new situation belongs to some category of situation and choose the most suitable behavior from a list of choices.

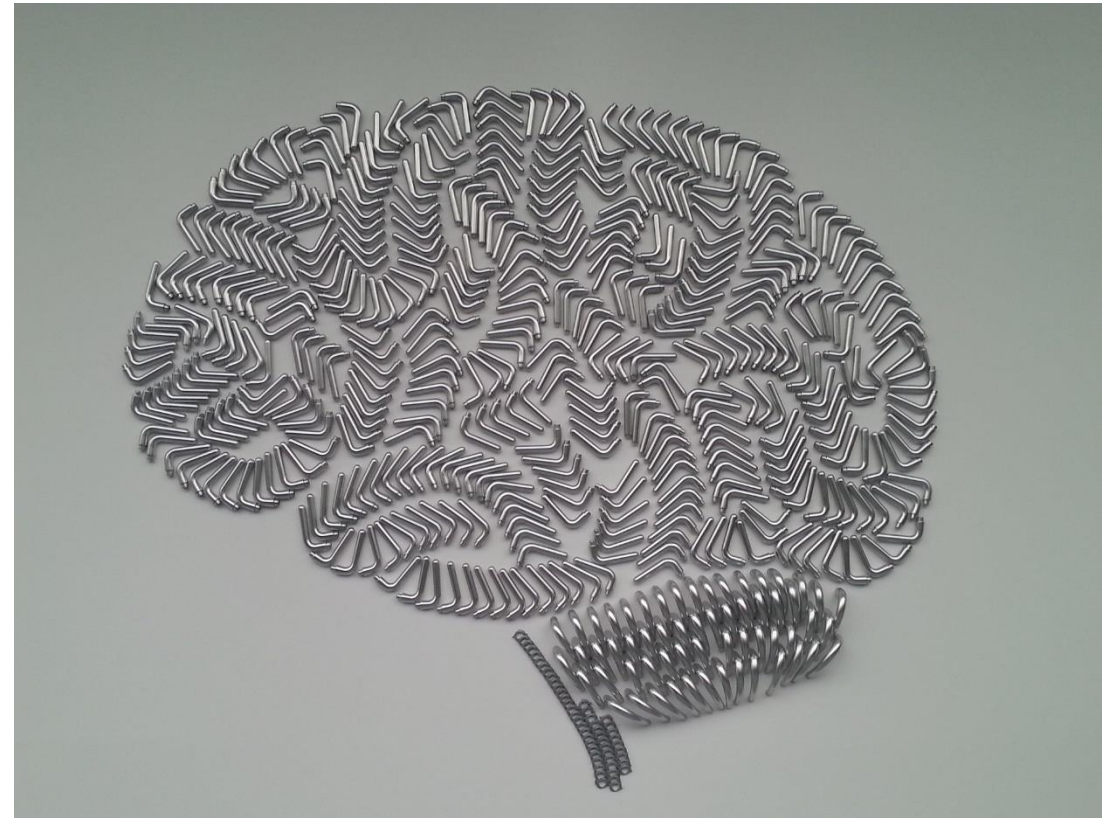


Sebastian learns „hot“ and „warm“



Sebastian learns...

- By **Feedback**: way more hot, way more cold than expected
- By **saving the rules in some structure**: in neurons and their connections.
- By **generalization of the learned rules**.

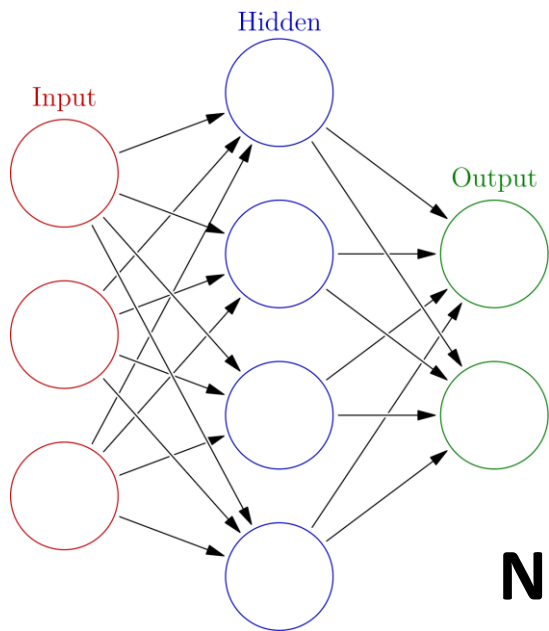


Computers learn...

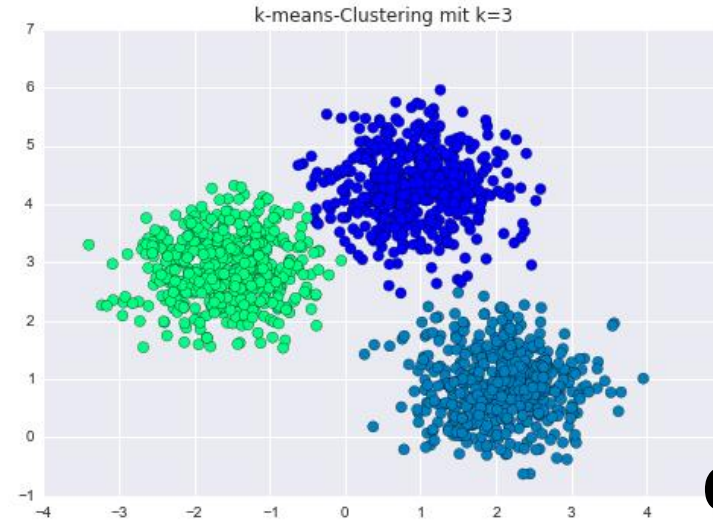
Computers also need a structure to learn and save the learned rules.

Optimally, the computers also get **feedback**.

They also learn **general rules** instead of being too specific.



Neural Network

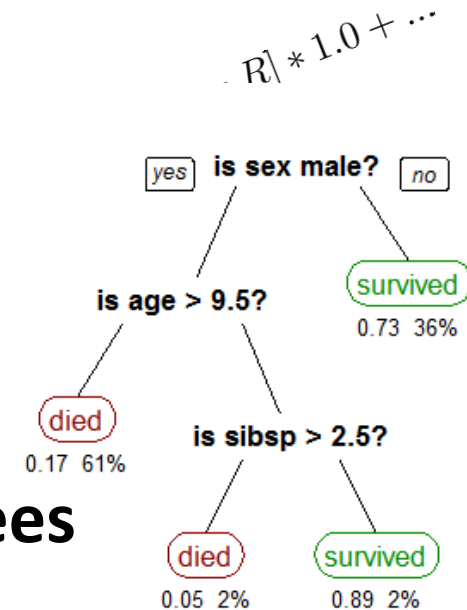


Clustering

Formula

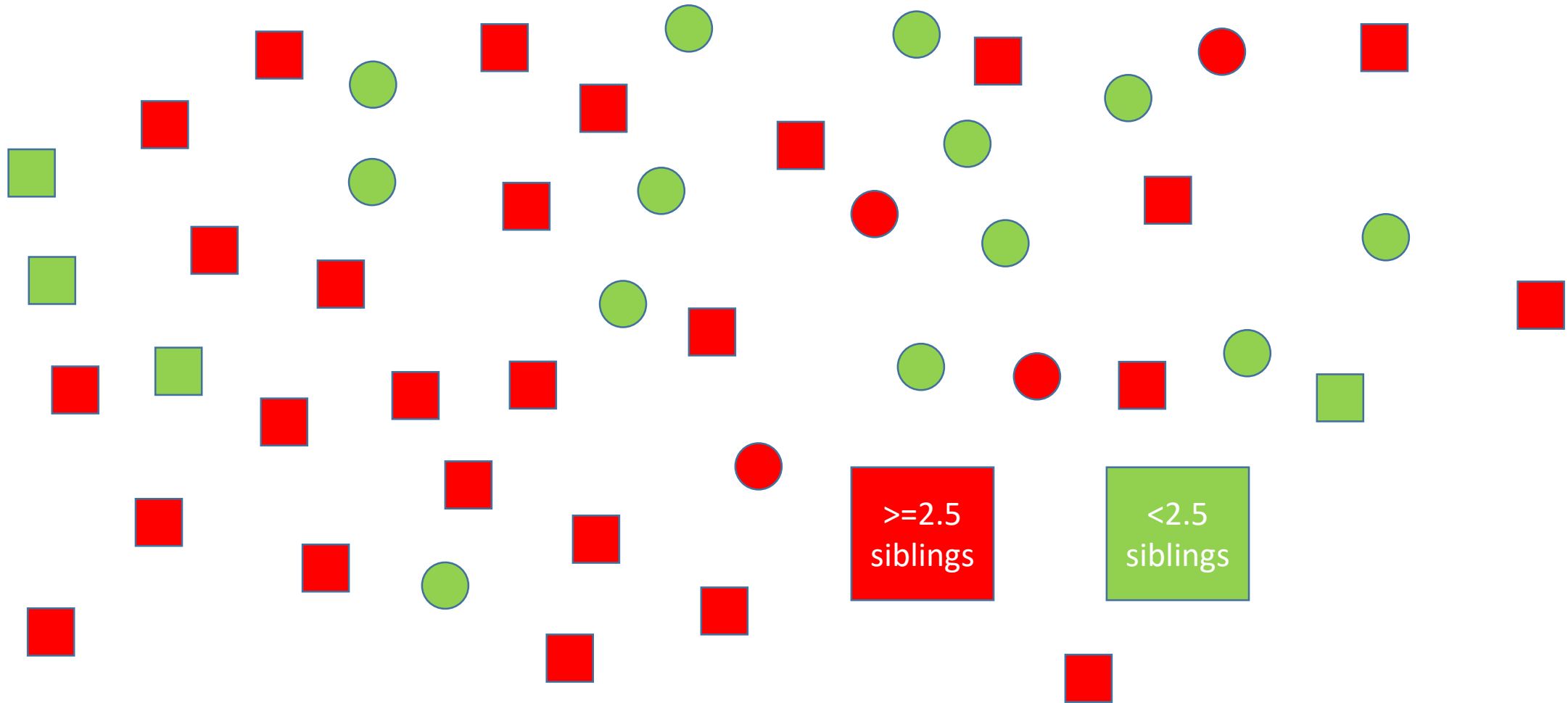
$$w_1 * \#V_h - w_2 * \#day_i V_h + w_3 * I[g = male]$$

Decision Trees



Learning with decision trees

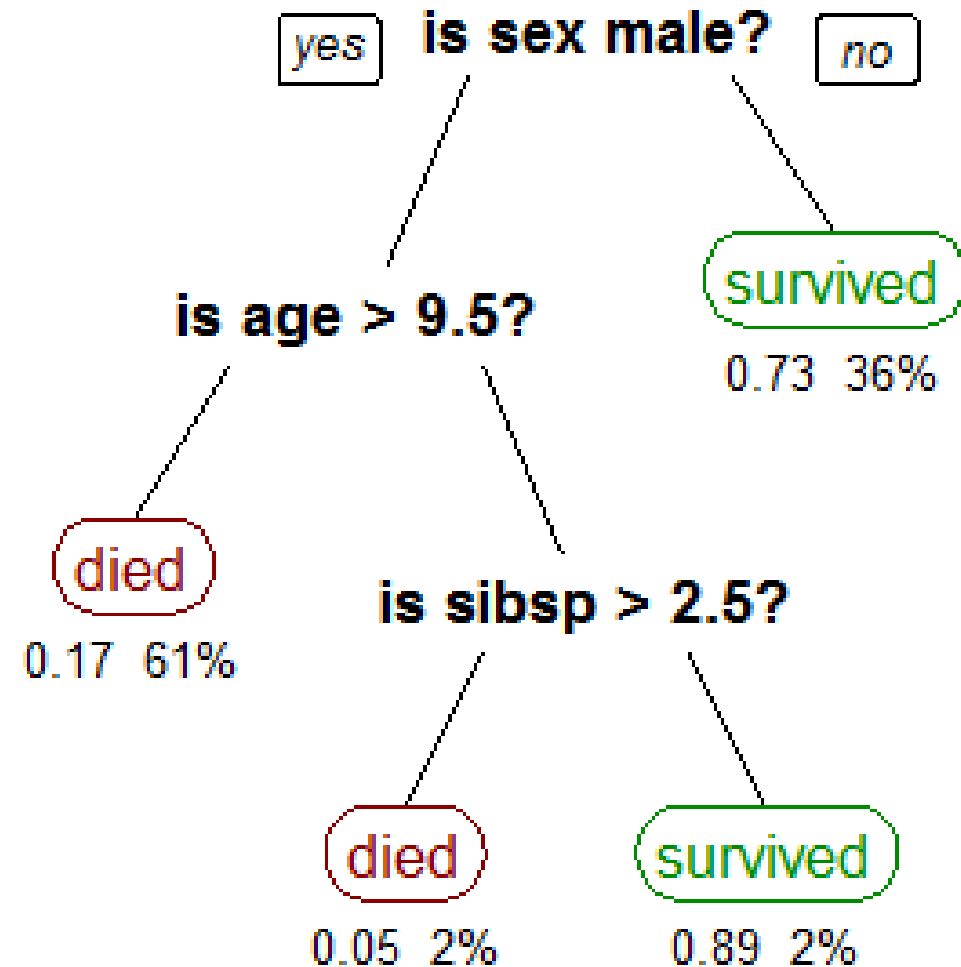
How to survive the Titanic



Learning with decision trees

- The tree shows the probability of survival of passengers of the Titanic.
- “sibsp” is number of family members.
- Numbers below the “leaves2” gives probability to survive and the fraction of people of this group with respect to all passengers.

(https://en.wikipedia.org/wiki/Decision_tree_learning, 28th of June, 2015)



Predictions + Mistakes

- With learned rules we can predict important properties of new data points:
- Given a new data point, the location within the data structure shows its class.
- If there are only two classes, we can make two kinds of mistakes:
 - False positives
 - False negatives



Learning with formulas

How to predict the recidivism probability of criminals



https://de.wikipedia.org/wiki/Datei:Bundesarchiv_B_145_Bild-F083310-0001_Karlsruhe_Bundesverfassungsgericht.jpg

Von Bundesarchiv, B 145 Bild-F083310-0001 / Schaack, Lothar / CC-BY-SA 3.0, CC BY-SA 3.0 de, <https://commons.wikimedia.org/w/index.php?curid=5473096>

Algorithm

- Algorithm designers decide which of the data to use – which are likely to correlate with recidivism?
- The wanted formula should result in a single number, such that delinquents can be directly sorted by this number.
- The higher the number, the higher the recidivism rate.
- Example formula:

$$\begin{aligned} & 3 * \text{past convictions} \\ & - 2 * \text{number days since last arrest} \\ & + 3 * (\text{if male, then 1, 0 otherwise}) \\ & + 2,5 * (\text{if violent behavior, then 1, 0 otherwise}) + \dots \end{aligned}$$

Allgemein

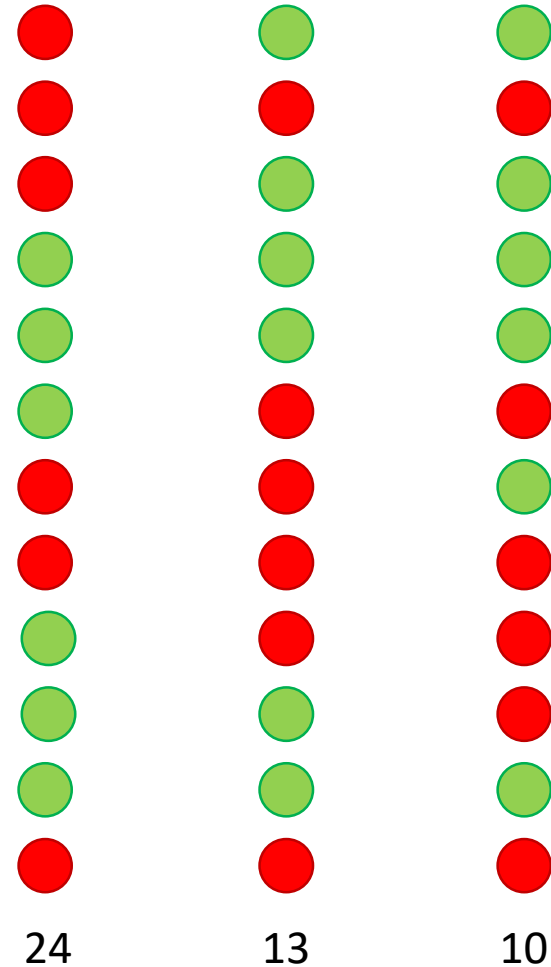
$$\begin{aligned} & w_1 * \text{past convictions} \\ + & w_2 * \text{number days since last arrest} \\ + & w_3 * (\text{if male, then 1, 0 otherwise}) \\ + & w_4 * (\text{if violent behavior, then 1, 0 otherwise}) + \dots \end{aligned}$$

The computer now determines the weights and gets a feedback, how well the predictions do on a data set with known behavior (past data).

Quality of an algorithm

„Learning“ of weights

- Algorithm just tries combinations of weights
- Evaluates how many recidivists are up front (high values)
- The weighting which maximizes the number is then set.



Red balls symbolise recidivating criminals, green ones resocialized persons.

Optimal sorting: all reds up front, all greens below.

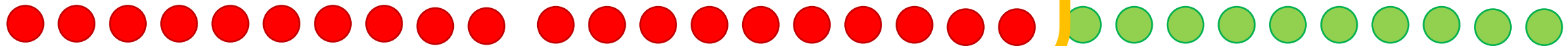
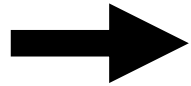
Measure of Quality: pairs of which the red ball is above the green one.

Oregon Recidivism Rate Algorithm

- For a concrete algorithm, this quality is 72 out of 100 pairs.
- Thus, given one future-recidivist and one future-resocialised person, the algorithm will give a higher value to the first with a chance of 1:3.
- Only about 25% of all predictions are expectedly wrong
- But this is not how predictions are made!
- Another problem: the classes are not balanced.
 - 10,000 Delinquenten
 - Approx. 2,000 will recidivate

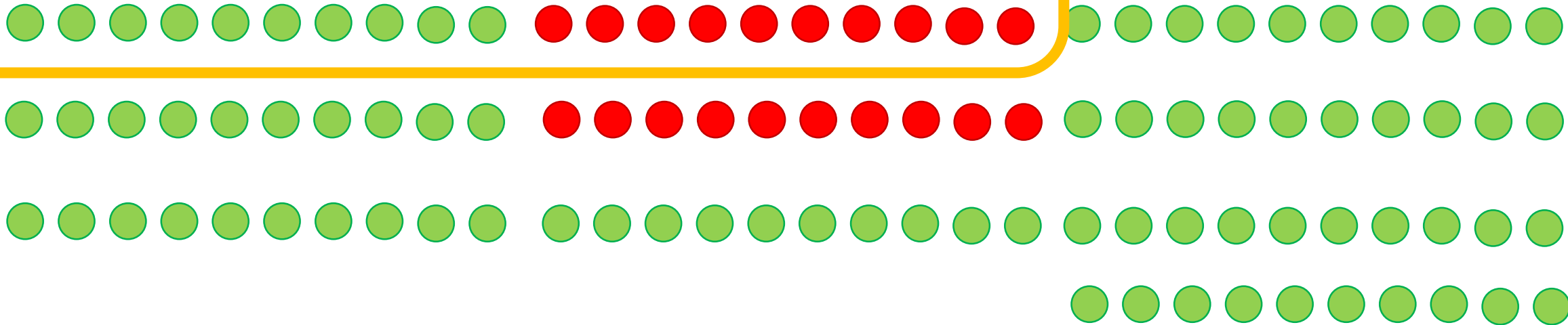
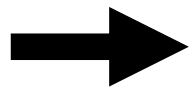
Optimal sorting

Expectedly 20% recidivists



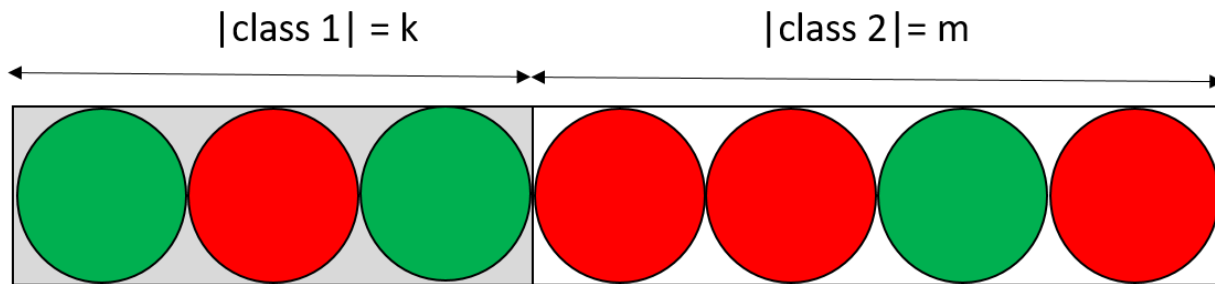
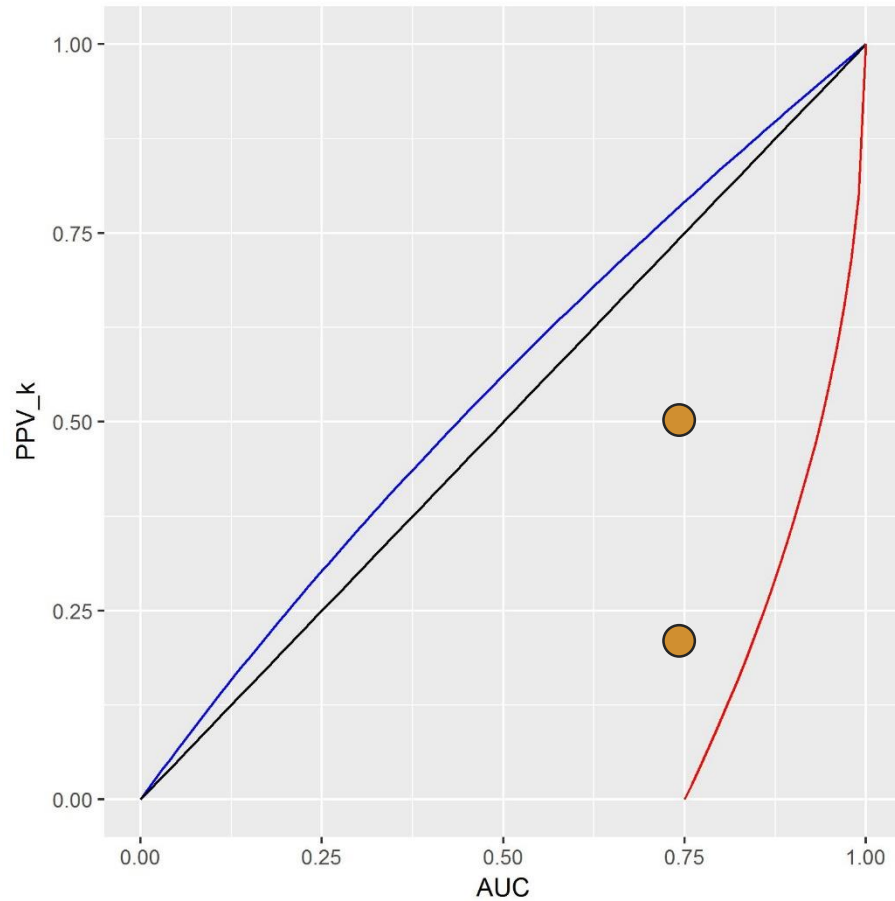
One possible sorting with this „quality“
(75/100 Pairs)

Expectedly 20% recidivists



Quality measures

- Diagram shows how much the „classic“ quality measure deviates from the more understandable measure



That is like selling this car

„You have to buy this gem of a car!
TÜV? Who needs TÜV anyway!
And just see the beautiful tires.
That is quality you just don't see
anymore!”



Propublica says: this algorithm is rassistic!

- In a study by Propublica this result was confirmed¹:
 - Only 20% of the predicted recidivists committed another (severe) crime.
 - If all kinds of crime are involved, the prediction is a bit better than throwing a coin.
 - Concerning afroamericans, the result was always too pessimistic;
 - Concerning whites, it was too optimistic.
- Northpoint Software is a company, the algorithm is not known.

¹ Angwin, J.; Larson, J.; Mattu, S. & Kirchner, L.: "Machine Bias - There's software used across the country to predict future criminals. And it's biased against blacks.", ProPublica, 2016 <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>

Rules of machine learning

Algorithm of machine learning are used where **there are no simple rules** but big data sets. **When there were simple rules, we likely already knew them.**

They search for patterns in **highly noisy data.**

The patterns are thus of a **statistical nature.**

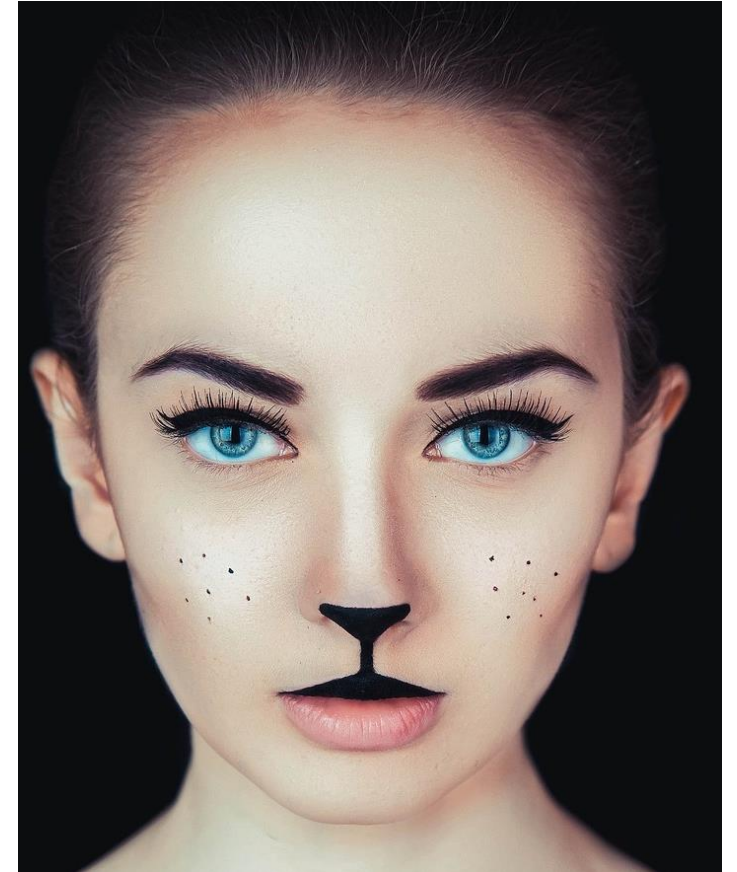
They almost always try to identify a small group of persons (problem of **imbalance**)

Statistical predictions of human behavior

What does that actually mean?

You're 70% recidivist....

- If persons were cats, it would mean 5 of them they'd recidivate, and 2 they would'nt.
- However, humans are not cats.
- Algorithm Decision Making relies on statistically legitimized prejudice.
- **Algorithmic clan liability**
 - Of 100 persons „like you“, 70 recidivate;
 - People are categorized into an algorithmically determined clan totally unknown to them.



Is that a problem?

- Attention economy of judges.
- „Best practice“ requires usage of the software.
- A mistaken ignorance of the machine’s decision is more severe for the judge than following a wrong decision of the machine.
- Basic modelling and data quality can be very bad as well.
- The criminal sent to prison can – by definition – **not prove the prediction wrong!**
 - The same is true for: credits, education, jobs, persons killed by drones or arrested as terrorists, ...

Spielkamp's Rule



**All algorithms are objective -
besides those designed by humans!**

Algorithms in a democratic society

Data Scientist

Researcher

Development of analytic method

Implementation

Development of analytic method

Implementation

Development of analytic method

Implementation

Method selection

Trained Decision System

Interpretation of result

Decision of action

Feedback

Person or Institution

Operationalization

Person or Institution

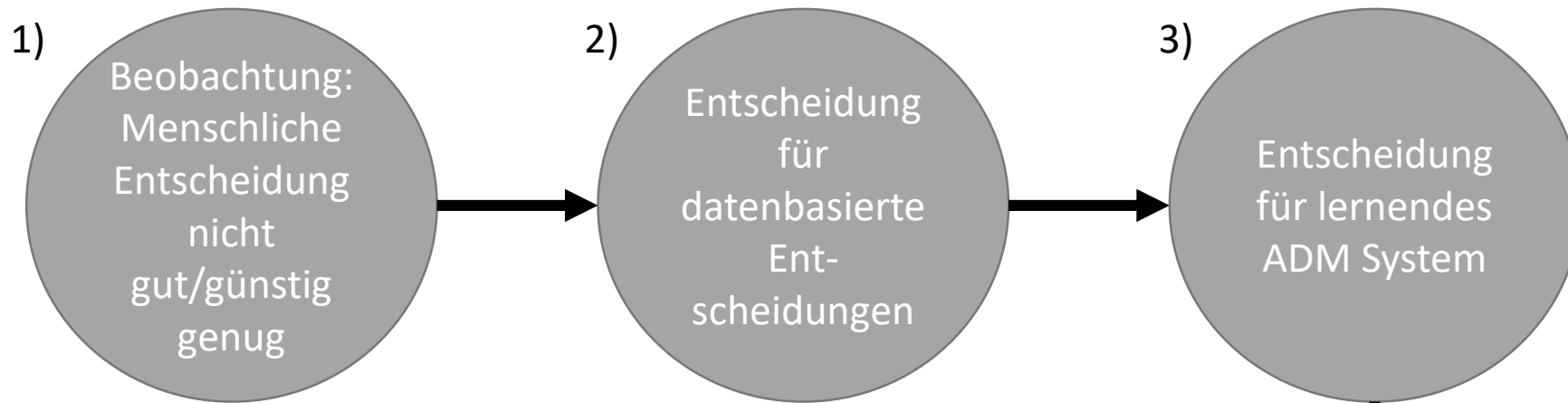
Data collection

Data collection

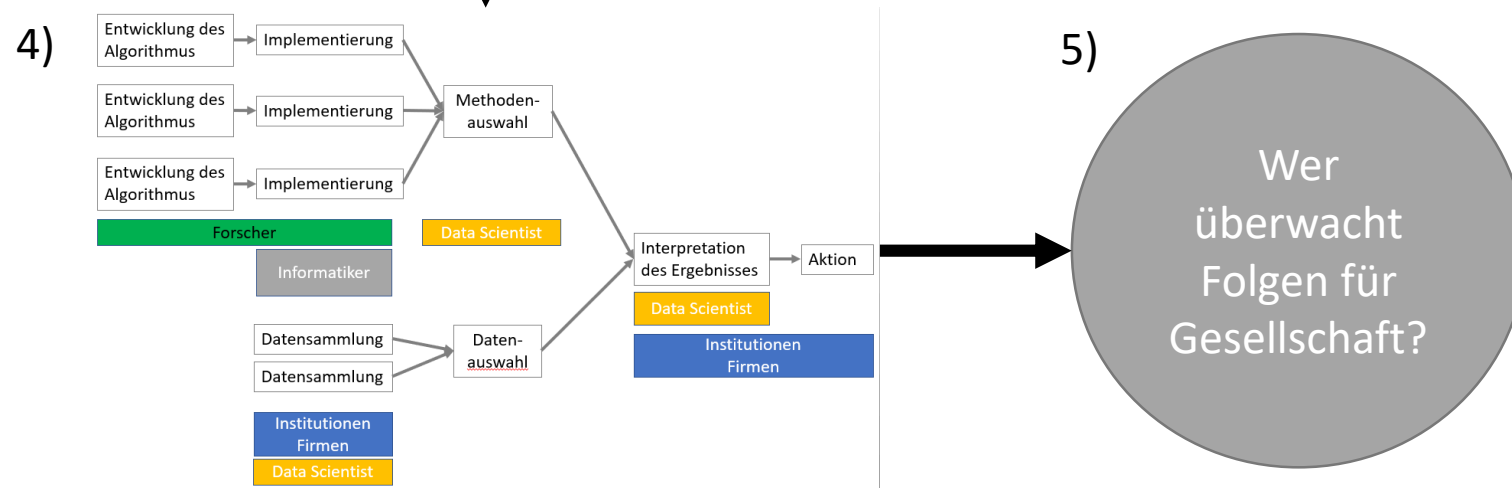
Data collection

Data selection

Data

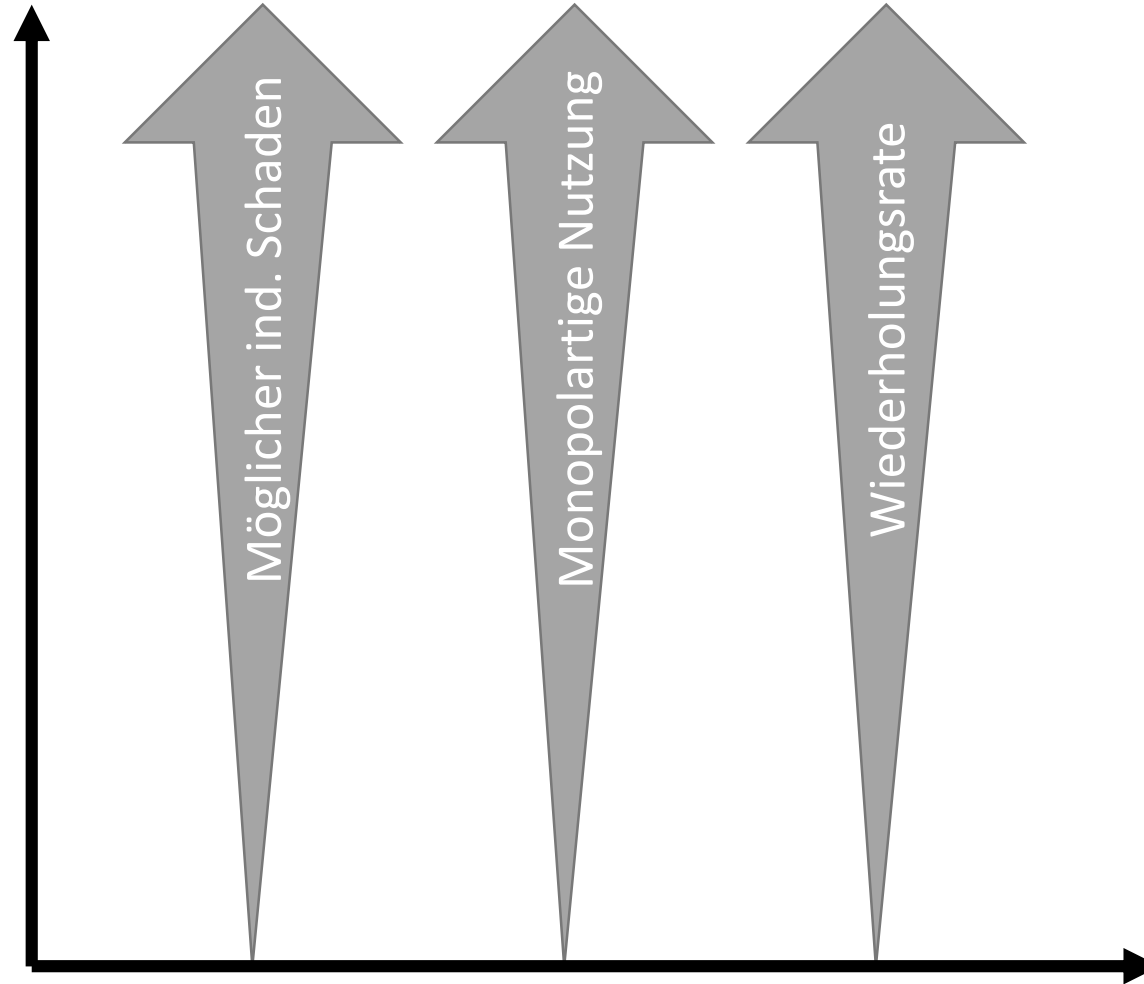


- Was ist gesellschaftliches oder ökonomisches Ziel?
- Was genau wird optimiert?
- Wie wird Datensatz zusammengestellt?
- Wie wird Qualität gemessen?
- Wer entscheidet dies?



Notwendigkeit von TA und Technikfolgenüberwachung

TA und
Technikfolgen-
überwachung
notwendig

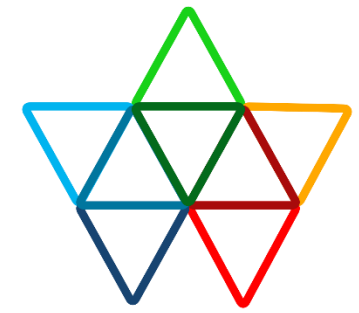


Im Nachhinein,
bei Verdachtsfall
ausreichend?

Quis custodiet ipsos algorithmos

„Automated Decision Making“-TÜV vulgo: „Algorithm TÜV“

Gründung von „Algorithm Watch“



ALGORITHM
WATCH



Lorena Jaume-Palasi, Law Philosopher



Lorenz Matzat, Data journalist, Grimme-Award



Matthias Spielkamp, founder of iRights.info, Grimme-Award



Prof. Dr. K.A. Zweig, Junior Fellow of the German Society of Computer Science (Gesellschaft für Informatik), Digitaler Kopf 2014, TU Kaiserslautern

Summary

- There are definitely chances in using algorithms to make decisions – also about humans
 - Reliable
 - Can be made more transparent
 - Could be less discriminating
- However, the ADMs we looked at so far do not seem to be of this sort.



Why Volkswagen

- We need to go beyond trans-disciplinary narrative sharing.
- We need high-level, understandable summaries of each others' insights.
- We need a place for meeting that everyone acknowledges.
- We need to do real interdisciplinary work over the largest disciplinary gaps.

