ALGORITHM WATCH

Replacing HR by Computers - Can Software really evaluate People?

Prof. Dr. Katharina Zweig

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

TECHNISCHE UNIVERSITÄT KAISERSLAUTERN

Gender equality

This is what you get when you Google for images with the keyword „CEO"…

Ethnic Discrimination in Hiring and Employment

- CVs with a German sounding name receive 14% more call backs than Turkish sounding names[1].

- Women wearing a scarf /Hijab have a lower call back rate than those without[2].

1 Kaas, L. & Manger, C.: “Ethnic Discrimination in Germany's Labour Market: A Field Experiment”, German Economic Review, 2011 , 13 , 1-20
2 Ghumman, S. & Ryan, A. M.: “Not welcome here: Discrimination towards women who wear the Muslim headscarf , human relations, 2013 , 66(5) , 671-698

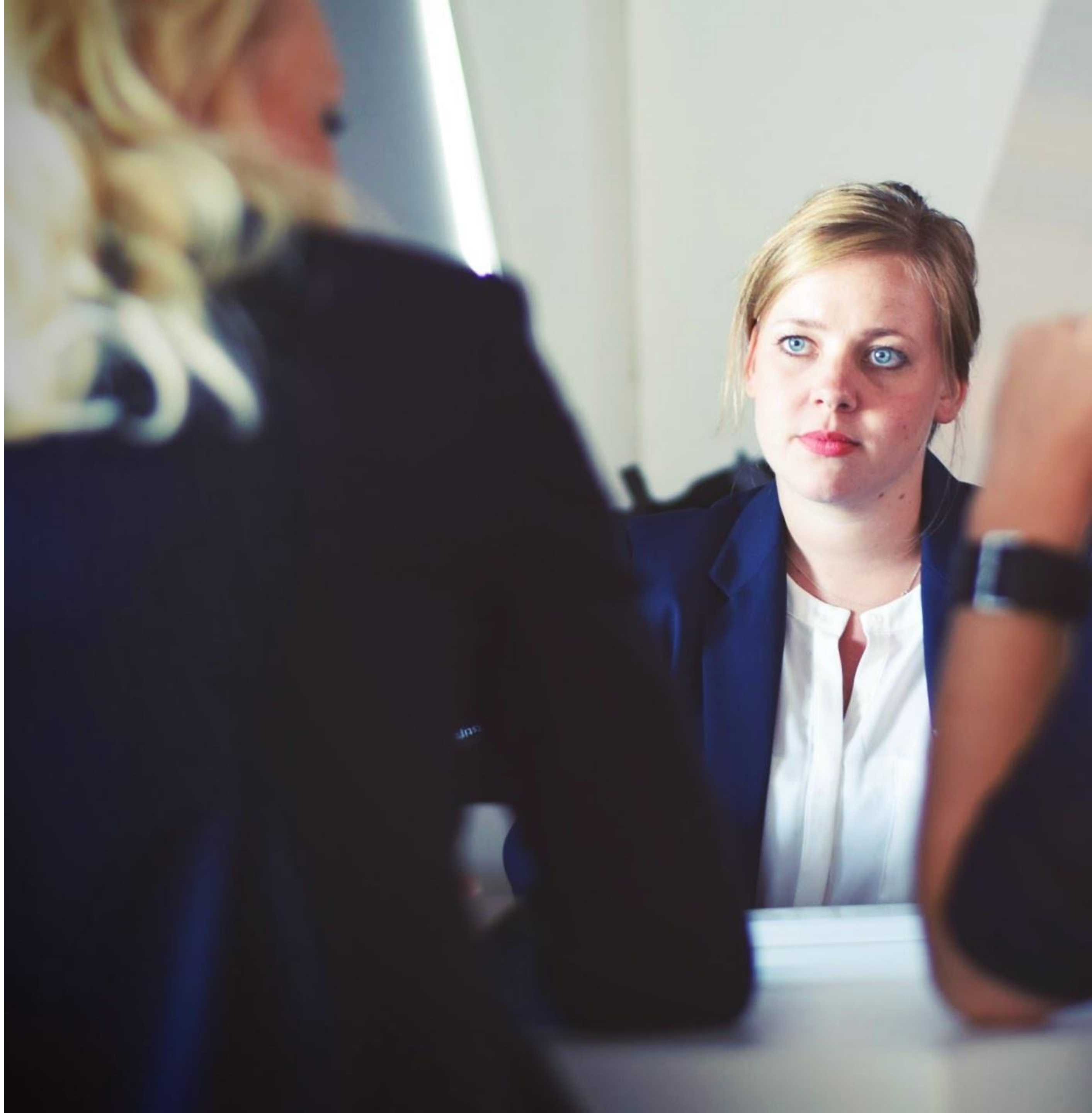
People
so irrational, so sad!

- Judges decide regularly over early release proposals.
- Study showed that judges take less risk the longer the time after the last break[1].
- 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 companies 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“)

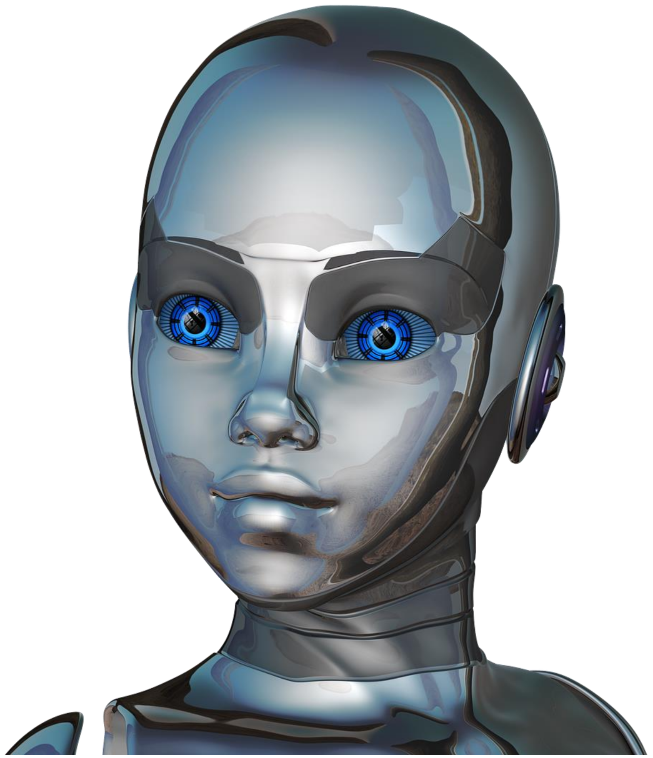


1 Claire Miller: “Can an Algorithm hire Better than a Human?”, The New York Times, June 25, 2015, https://www.nytimes.com/2015/06/26/upshot/can-an-algorithm-hire-better-than-a-human.html

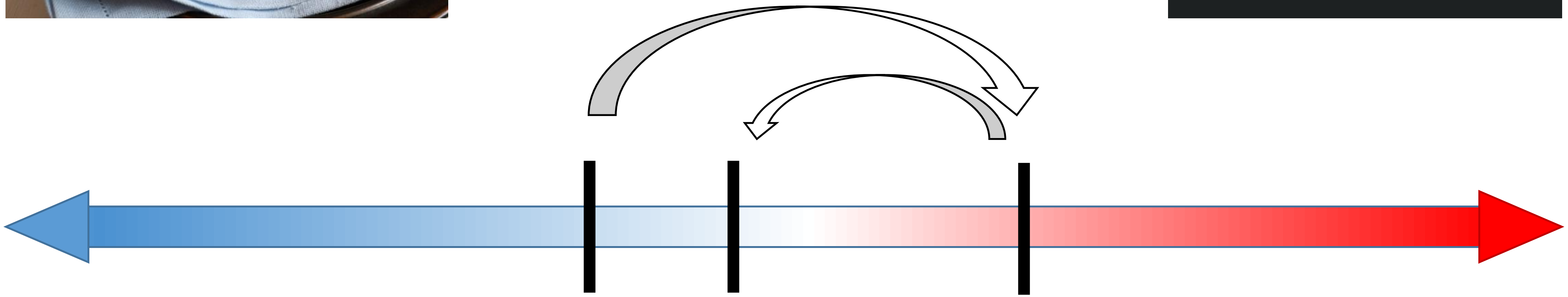
Can Computers learn?

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“



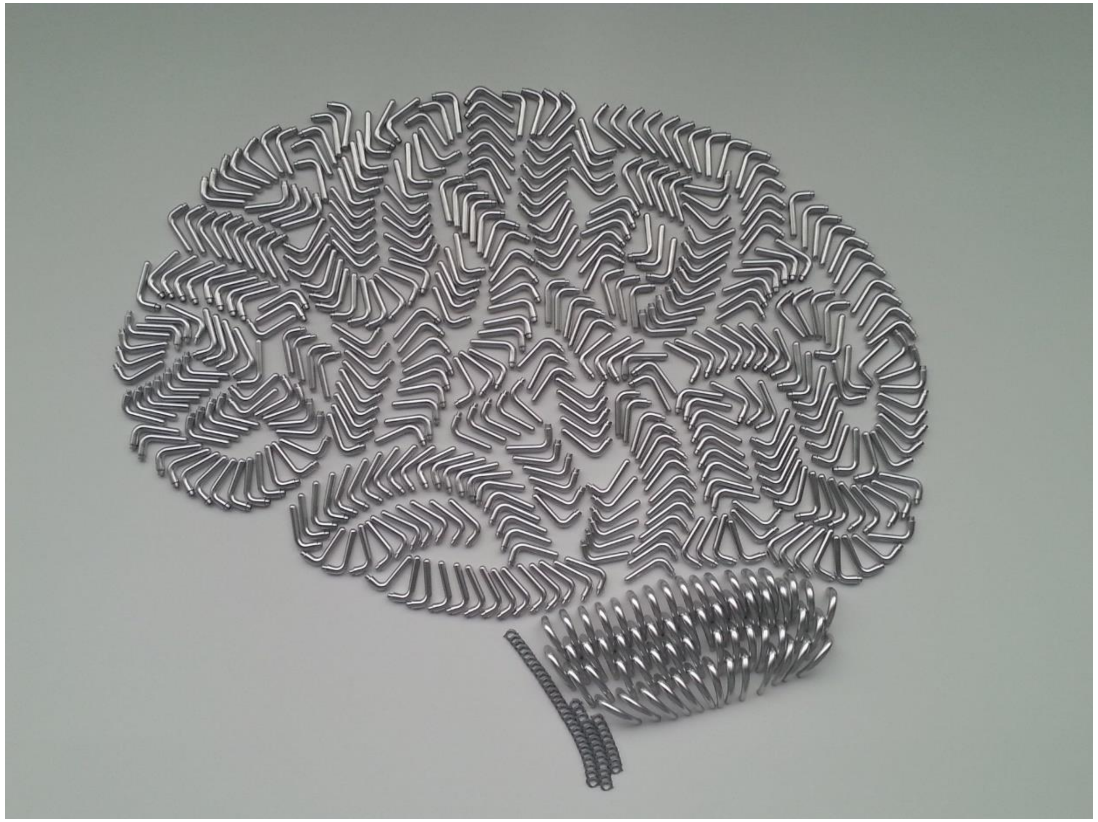
Last July
Everything steaming

Today
Eveything steaming unless somebody blew on it.

Last September
If it comes direct from the pot

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**.

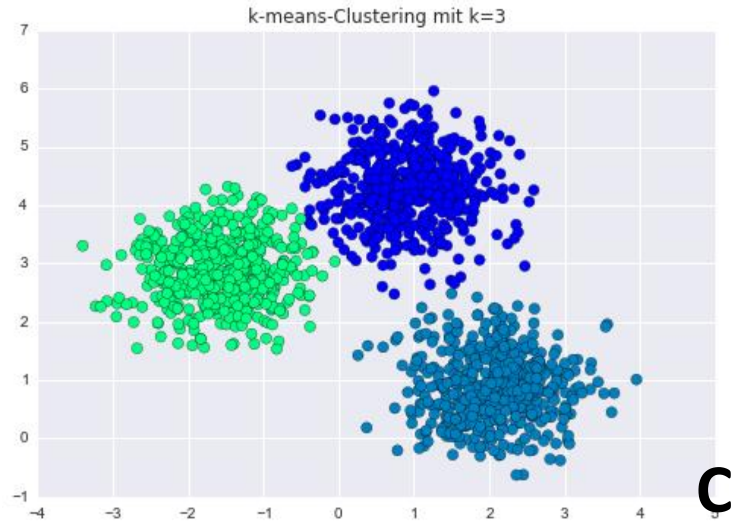


Computers learn…

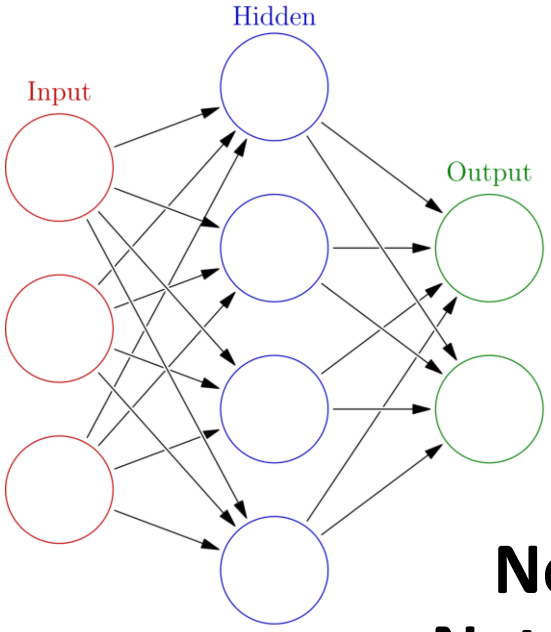
Computers also need a structure to learn and save the learned rules.

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

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



Clustering

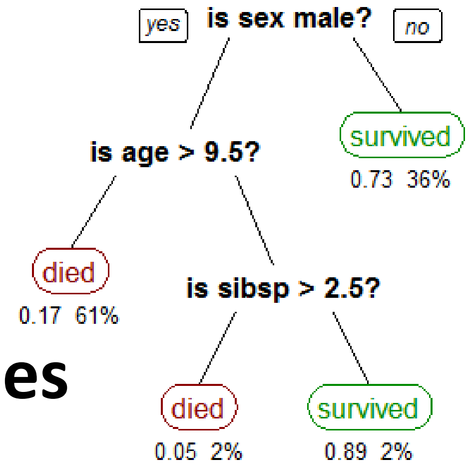


Neural Network

Formula

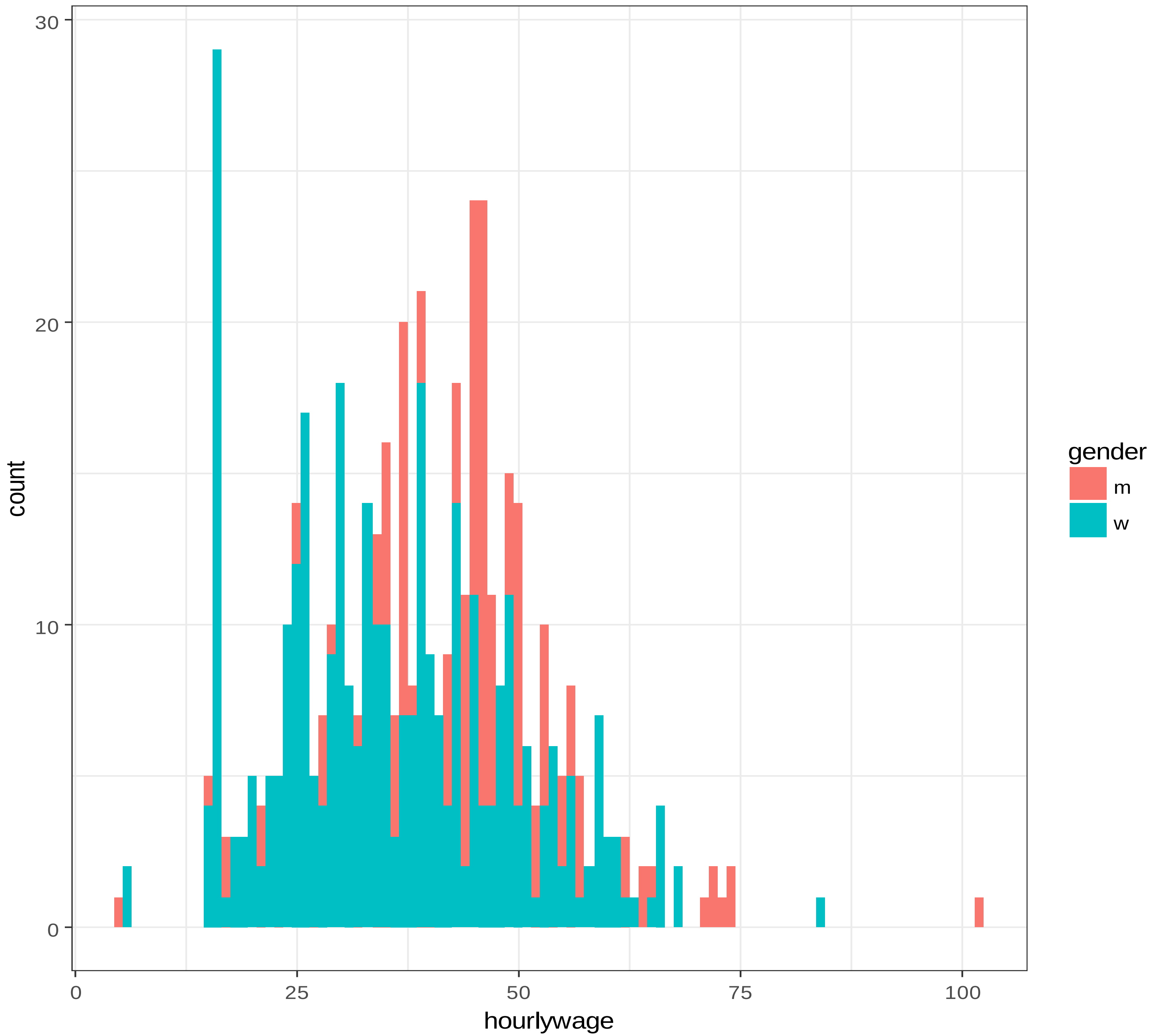
$w_1 * \#Vh - w_2 * \#day_I Vh + w_3 * I[g = male$

$R] * 1.0 + \ldots$



Decision Trees

Learning with probabilities



Wages in Seattle

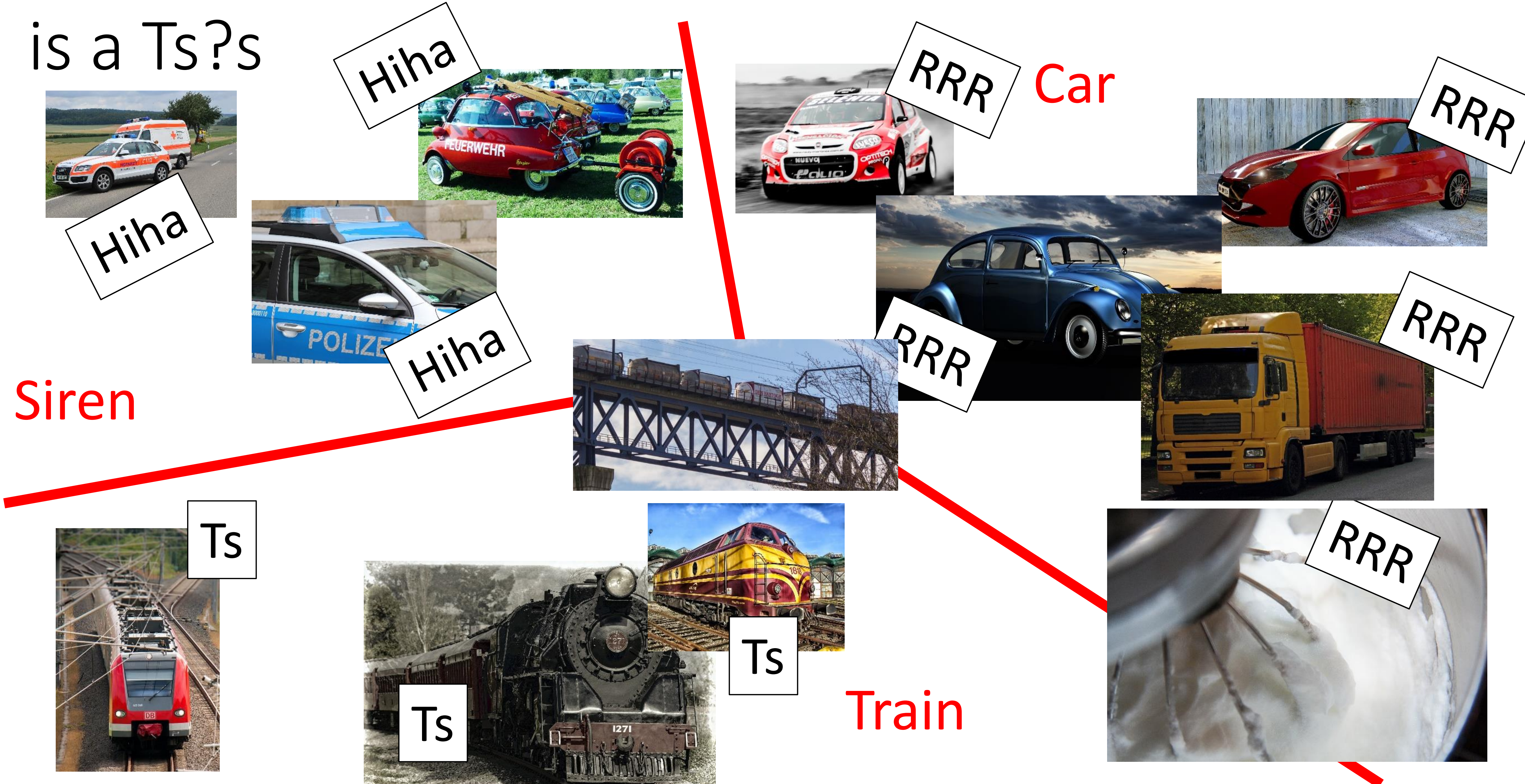
A new person is hired. The person earns less than $25 per hour.

Is the person male or female?

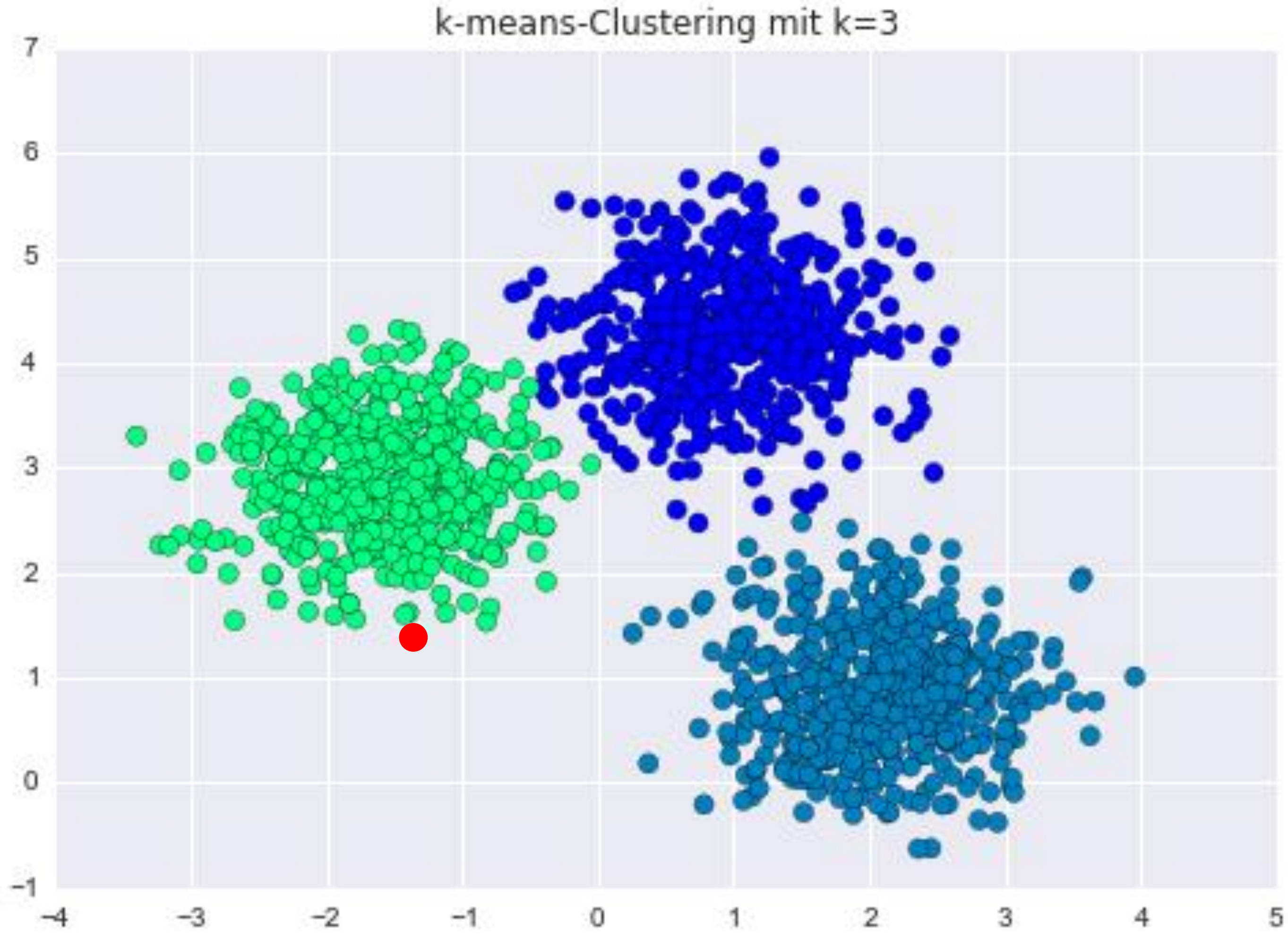
And what if the person earns more than $60/hour?

Learning by clustering of similar cases

What is a Rrrr, what is a Hiha and what is a Ts?s



K-Means-Clustering



Predictions + Mistakes

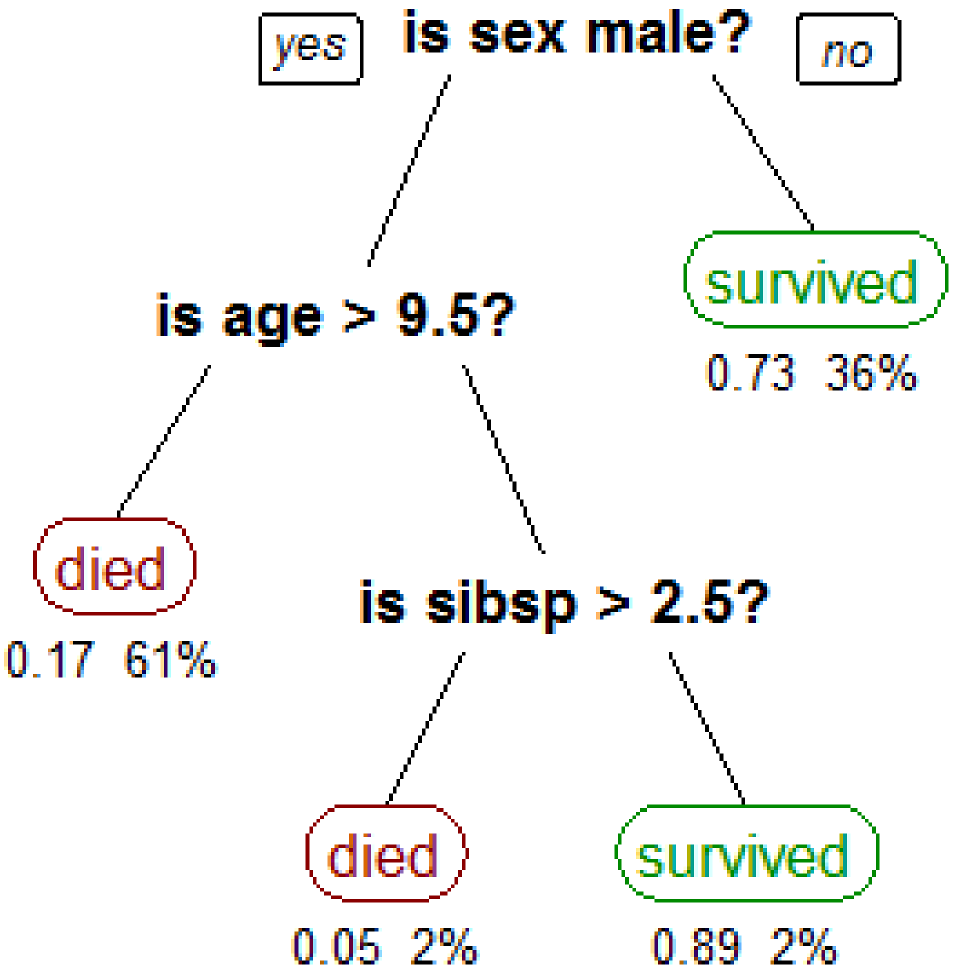
- With learned rules we can predict important properties of new data points:
- Given a new data point, the location within the data structure shows its class.
- If there are only two classes, we can make two kinds of mistakes:
  - False positives
  - False negatives

Learning with decision trees

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)



Learning with formulas

How to predict the recidivism probability of criminals

Predictive Policing

Waiting for you
since 10 minutes!

Algorithms that predict when and where criminal activities are most likely.

Predictive Policing

An **Algorithm** said
that youre **almost** a criminal!
You‘re arrested!

Also possible: predictions about whether a person is about to conduct a crime.

E.g. in Oregon and other states of the USA.

Data basis

- Uses Big Data, e.g.:
    - Age of first arrest
    - Age of the criminal
    - Financial situation
    - Criminal family members
    - Gender
    - Type and number of convictions
    - Timepoint of the last criminal activity
    - Surveys
    - But not the ethnicity

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:

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

Allgemein

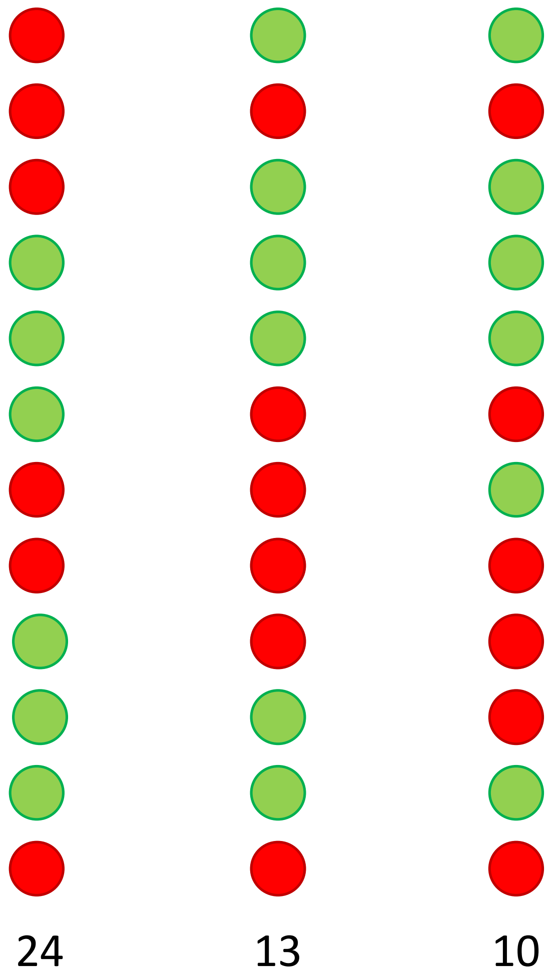
$$
\begin{aligned}
& w_1 * \text{ past convictions} \\
+ & w_2 * \text{ number days since last arrest} \\
+ & w_3 * (\text{if male, then 1, 0 otherwise}) \\
+ & w_4 * (\text{if violent behavior, then 1, 0 otherwise}) + ...
\end{aligned}
$$

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.

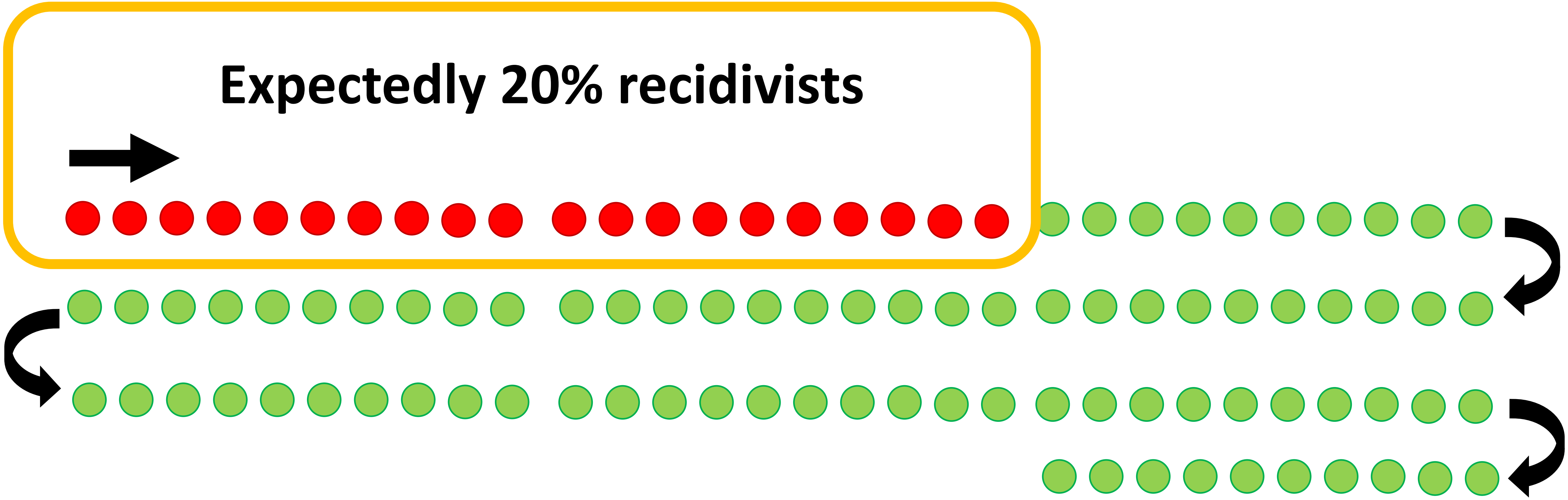
Optimal sorting: all reds up front, all greens below.

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

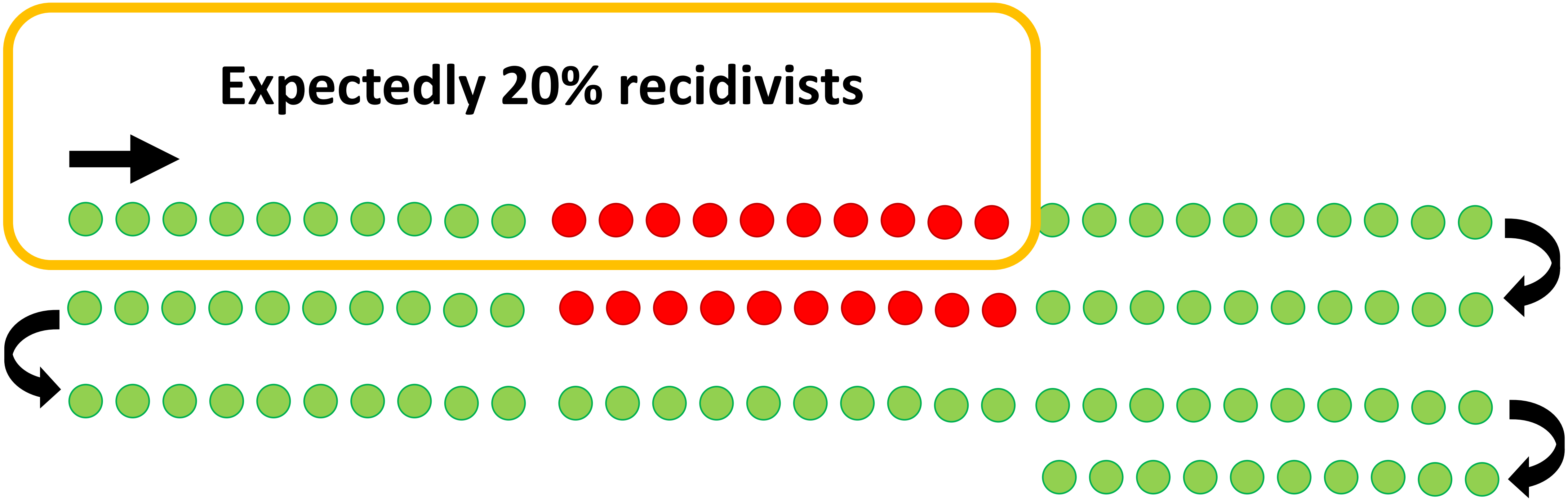
Oregon Recidivism Rate Algorithm

- 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.
  - 10000 Delinquenten
  - Ca. 2000 werden rückfällig

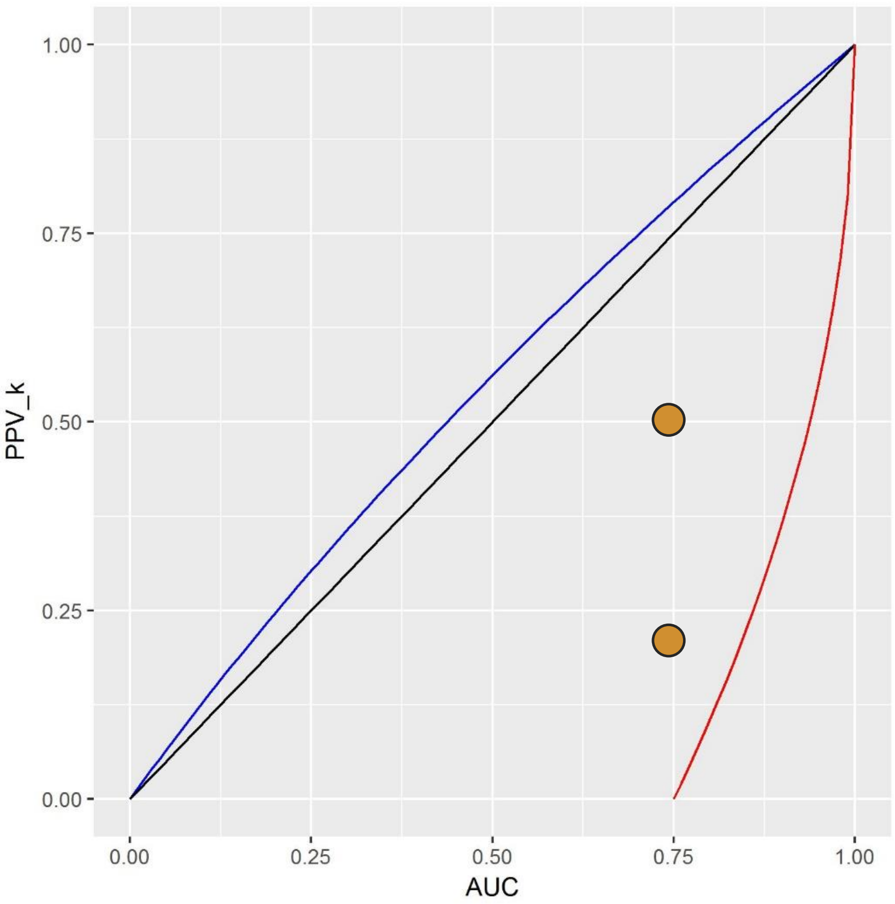
Optimal sorting

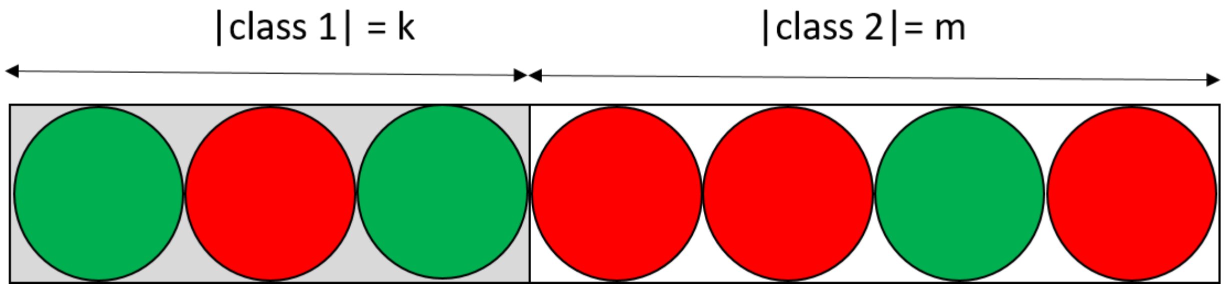


One possible sorting with this „quality“ (75/100 Pairs)



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 confirmed[1]:
  - Only 20% of the predicted recidivists commited 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.

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**)

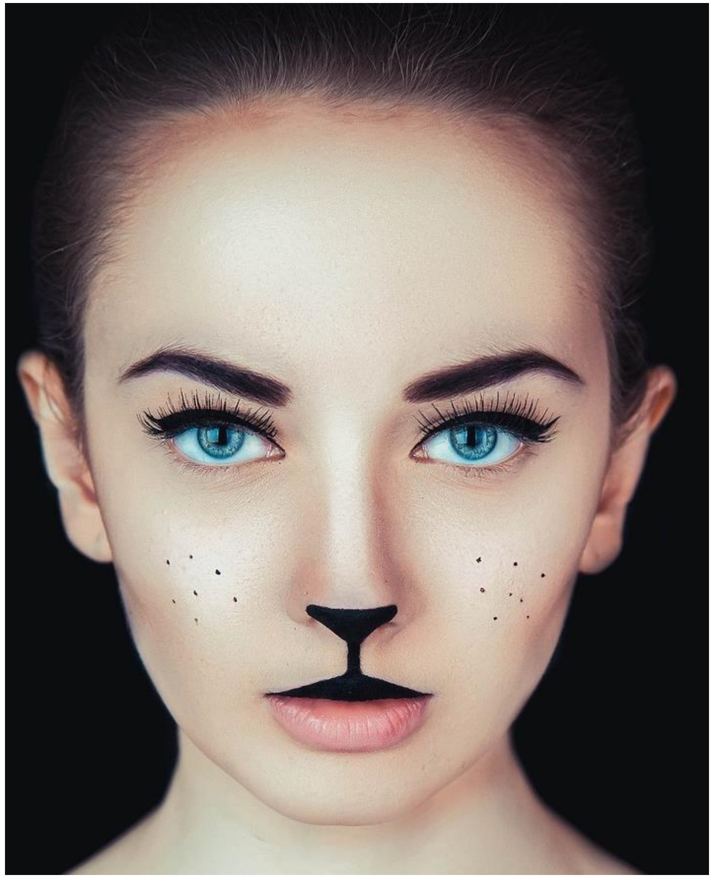
When there were simple rules, we likely already knew them.

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.



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

Is that a problem?

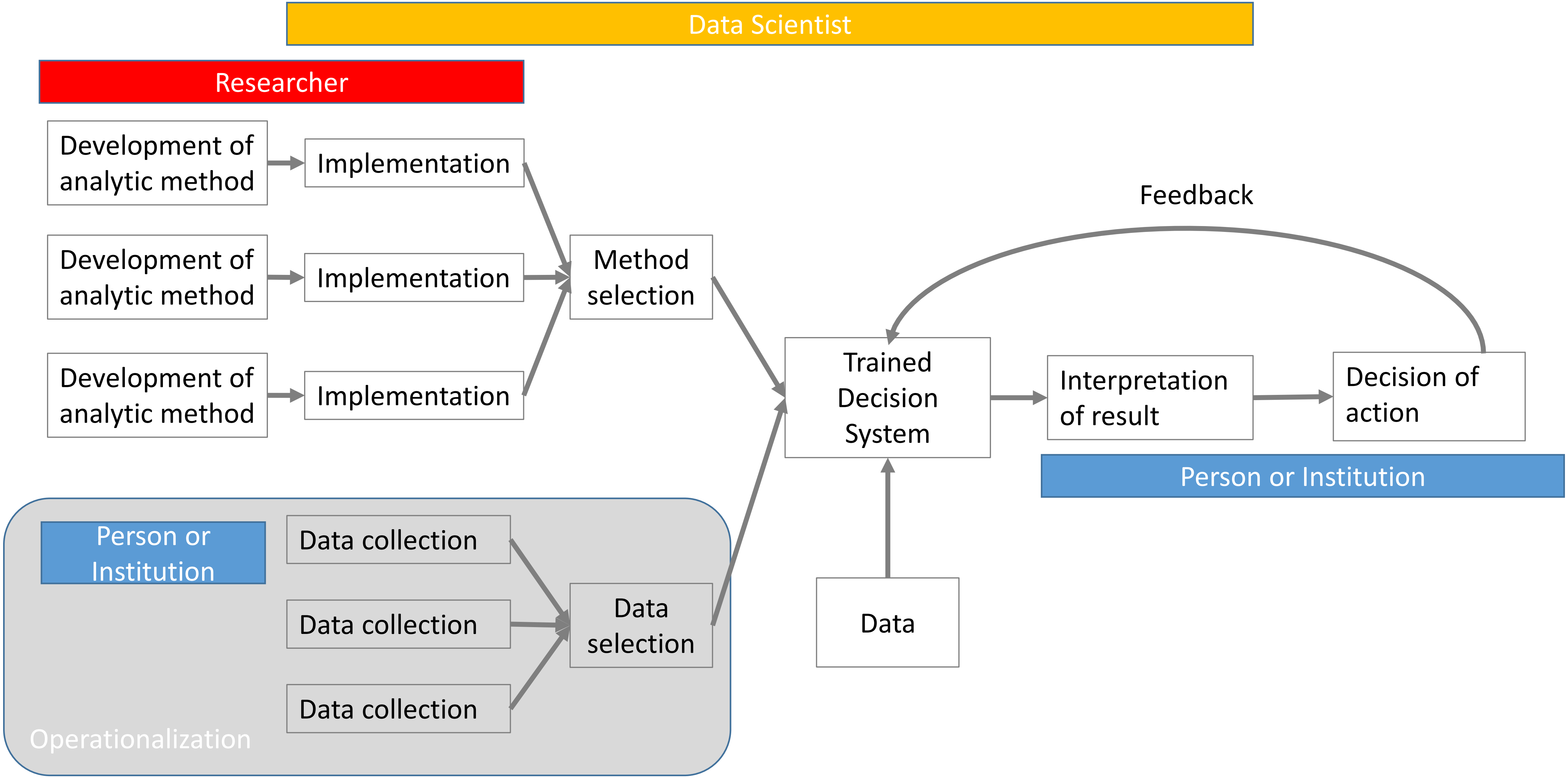
- Attention economy of judges.
- „Best practice“ requires usage of the software.
- A mistaken ignorance of the machine‘s decision is more severe for the judge than following a wrong decision of the machine.
- Basic modelling and data quality can be very bad as well.
- The criminal sent to prison can – 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
Feedback
Trained Decision System
Interpretation of result
Decision of action
Person or Institution
Person or Institution
Data collection
Data collection
Data collection
Data selection
Data
Operationalization

Quis custodiet ipsos algorithmos

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

Gründung von „Algorithm Watch“

Lorena Jaume-Palasí, 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

ALGORITHM WATCH

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.

Conclusion

Jump in

but don‘t drown

Prof. Dr. Katharina Zweig

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

zweig@cs.uni-kl.de