Sozio-informatik

Algorithm Accountability

- why society needs insight into some algorithms

SAP INNOVATION@SST
7.12.2016

Prof. Dr. Katharina A. Zweig

TECHNISCHE UNIVERSITÄT
KAISERSLAUTERN

Introduction

Sozio-informatik

- Born 1976, in Hamburg, Germany
- Biochemist and Computer Scientist, working in the fields of „Network Analysis Literacy“ and „Algorithm Accountability“
- Professor for theoretical computer science at the TU Kaiserslautern
- Designed the nationally unique field of study called „Socioinformatics“
- Junior fellow of Germany‘s society of Computer Science; selected as one of Germany‘s „digital heads“ in 2014; member of the board on „Innovation and Technological Analysis“ of the federal ministry of research and education.

Can algorithms be worrisome?

Where were **you** on the 8th of November?

Von Gage Skidmore, CC BY-SA 3.0, https://commons.wikimedia.org/w/index.php?curid=51025646

Von Gage Skidmore, CC BY-SA 3.0, https://commons.wikimedia.org/w/index.php?curid=46897599

Before the election it was claimed „Google favors Hillary!“

- SourceFed published a viral video, saying that negative autocomplete statements were blocked when searching for Hillary
- Scientist Robert Epstein claimed to have evidence for massive manipulations at Google‘s search [1] and points to a study in which he claims to see evidence that up to 20% of all undecided voters can be biased by search engines [2].
- (The numbers are not backed up by his data)

[1] https://sputniknews.com/us/201609121045214398-google-clinton-manipulation-election/

[2] http://www.pnas.org/content/112/33/E4512.abstract

After the election

- Oh, well…
- Maybe it was not Google (or they were not very successfull)
- It must have been Facebook!
- With its algorithmically generated *filter bubbles* and *echo chambers,* we are just forced to live in a post-truth world…

Or was it...

Von MacedonianBoy - Own work GFDL, https://commons.wikimedia.org/w/index.php?curid=7842206

- BuzzFeed analyzed main partisan webpages on both sides and found out that, on average, conservative sources had a significantly higher percentage of fake news (or half true-half false news).
- Interestingly, a part of these „hyper partisan“ websites are located in ...
- ... **Macedonia!** [1]

[1] https://www.buzzfeed.com/craigsilverman/partisan-fb-pages-analysis?utm_term=.yhrZgyjK#.jckQXRPW

Motivation

„Most of the posts on these sites are aggregated, or completely plagiarized, from fringe and right-wing sites in the US. The Macedonians see a story elsewhere, write a sensationalized headline, and quickly post it to their site. Then they share it on Facebook to try and generate traffic. The more people who click through from Facebook, the more money they earn from ads on their website."

Craig Silverman und Lawrence Alexander: **„How Teens In The Balkans Are Duping Trump Supporters With Fake News"**, BuzzFeed Nov. 4th, downloaded on the 27th of November, 2016
https://www.buzzfeed.com/craigsilverman/how-macedonia-became-a-global-hub-for-pro-trump-misinfo?utm_term=.wvzZe7D5#.he3BElYV

So, algorithms might play a role in societal processes after all…

The ABC of Computer Science

Where and when can

Algorithms,

Big Data, and

Computer Intelligence

harm democracy?

A as in Algorithm

An algorithm is a problem solver!

Mathematical Problem

PUTPUT

INPUT

The part that defines the connection between input and output.

OUTPUT

Input: By User:Bluemoose - Own work, CC BY-SA 3.0
Putput: By Yann (talk) - Own work, GFDL
Output: CC BY-SA 3.0

Example: Navigation

Navigation

Input: Maps, length of streets, traffic jams, current location, goal, …

Given map and other input, compute the shortest path between current location and goal.

The problem itself does not **indicate** how to find the **solution**.

Output: optimal route

An algorithm is…

… a **sufficiently detailed** and **systematic sequence of instructions** describing how to find a **correct solution** for a **correct input** (in finite time), such that **any experienced programmer** can implement it on a **computer**.

Second example: Sorting

Problem: Sorting

Sorting 1: „Insertion sort“

- Start with one book, put it in the shelf.
- While there are more books,
  - take the next,
  - go along the shelf and put it in the correct position among the books already in the shelf.

- Will this produce the correct solution?
- Observation: all books in the shelf are already in the correct, relative order.
- Thus, when all books are in the shelf, they are correctly sorted.

Sorting 2: Bubble sort

- Put the books in the shelf in some random order.
- Go along the shelf. Whenever there are two books in the wrong order, swap them. Do this until you reach the end of the shelf and go back to the start of it.
- If there was at least one swap in the last run, repeat step 2 until finally no swap is necessary any more.
- If no swap was necessary anymore, this implies that the books are correctly sorted.

Generalizability - The power of algorithms!

- Given a set of objects or subjects…
- … and a sorting criterium that defines for each pair of things which goes left and which goes right…
- … any sorting algorithm can compute the correct solution (and it is the same solution for all of them).

- In that sense, different real-world problems can be **represented** by the same mathematical problem (e.g., sorting)
- The algorithm does not help us to **interpret** the meaning of its result, e.g.: these are the most relevant news, most important friends, or most popular products.

Summary: Algorithms

Algorithms…

- …can consist of only an equation,
- but in most cases they are a complex mix of instructions and computations
- that find one solution for a problem, based on the input.
- They are **frozen instructions,** devised by man, to tackle mathematical problems by computers.
- The result requires an **interpretation**!

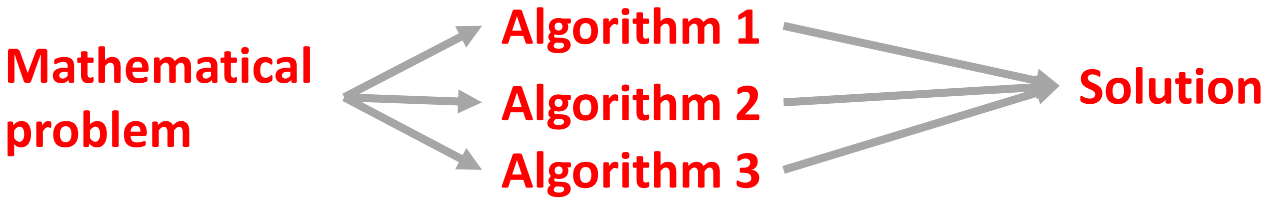
```
  var simple = typ
      forward = type.slice(
      ofType = what === "of-type";

  return first === 1 && last === 0 ?

      // Shortcut for :nth-*(n)
      function( elem ) {
          return !!elem.parentNode;
      } :

      function( elem, context, xml ) {
          var cache, outerCache, node, diff, nodeIndex, start,
              dir = simple !== forward ? "nextSibling" : "
                  previousSibling",
              parent = elem.parentNode,
              name = ofType && elem.nodeName.toLowerCase(),
              useCache = !xml && !ofType;

          if ( parent ) {
                                    (child|of-type)
```



Mathematical problem

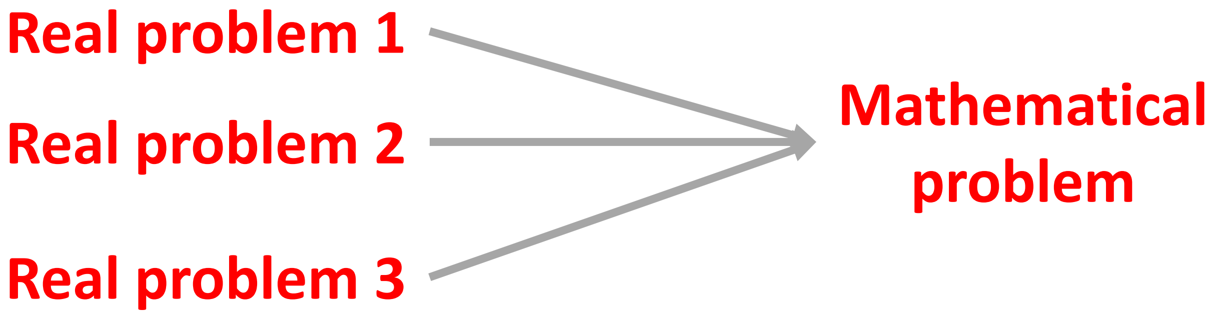
Algorithm 1

Algorithm 2

Algorithm 3

Solution

The matching between a real-world problem and a mathematical problem requires modeling decisions!



Real problem 1

Real problem 2

Real problem 3

Mathematical problem

„Harmless“ algorithms

- Most algorithms are harmless. They
  - search, save, filter, sort
  - do logistics and scheduling
  - or image recognition in industrial contexts
  - help manage
  - or help us create documents
  - …
- Mistakes are quite easily detected in these cases; humans are not directly affected by these algorithms.
- Similarly, algorithms with (yet) a small reach, e.g., by startups, are mainly „harmless“
- These do not need additional control or regulation in most cases
- Algorithms that have the potential to harm our fundamental rights or democrarcy itself are of a specific nature and most of them involve „big data“.

B as in Big Data

Let us see some example of a potentially harmful algorithm

Who should predict and take account of the likelihood of recidivism?

Now you‘re kidding, right? Isn‘t that outdated?

Predictive Policing

Predicts when and where criminal acts are most likely.

Predictive Policing

An **algorithm**
revealed,
that you‘re **almost** a criminal.
Let‘s go!

They can also predict, whether an individual is likely to become a criminal or to relapse into criminal behavior.

Used in the USA:

1) Oregon
2) Other states

Big Data

Big data describes:

- Very big data sets
- which are very often used outside of their original context
- which are likely erroneous
- but which can be used to find statistically valid correlations.

Big Data in Recidivism Prediction

- Big Data information could be:
  - Age of first arrest
  - Current age of culprit
  - Financial situation
  - Number of criminal relatives
  - Gender
  - Type and number of previous convictions
  - Time of last arrest
  - ….
  - But not the ethnicity of a person (neither in the US nor in Europe)

How can this be built into an algorithm?

- Algorithm designers decide principally which of these data most likely correlate with recidivism.
- They also decide on how these parameters are merged with each other to result – ideally – in a single number.
- Motivation: the higher the number, the higher the likelihood of recidivism.
- One way to merge them is to create a formula like this one:

$$\begin{aligned} & 3 * \text{ previous arrests} \\ - & 2 * \text{ number days since last arrest} \\ + & 3 * (\text{if male, then } 1, \text{ else } 0) \\ + & 2,5 * (\text{ if armed raid, then } 1, \text{ else } 0) + \ldots \end{aligned}$$

We are not **that** bold…

$$\begin{aligned} & w_1 * \text{ previous arrests} \\ & + w_2 * \text{ number days since last arrest} \\ & + w_3 * (\text{if male, then 1, else 0}) \\ & + w_4 * (\text{ if armed raid, then 1, else 0}) + \ldots \end{aligned}$$

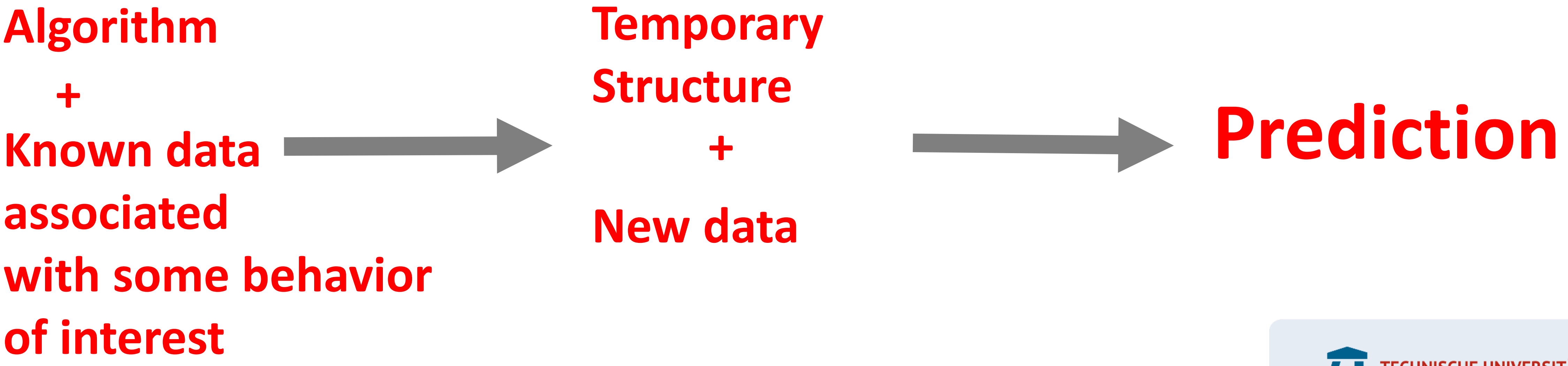
- Who determins the weights such that those who commit a crime in the next three years are those that get the highest overall numbers?
- For this, we need computer intelligence.

C as in Computer Intelligence

Learning Algorithm

Computer intelligence

- **Problem:** given a set of known data concerning human behavior of interest, find patterns in it which predict how another human (or the same human later) will behave.
- Algorithm builds – based on known data – some intermediary structure which then generates the predictions.



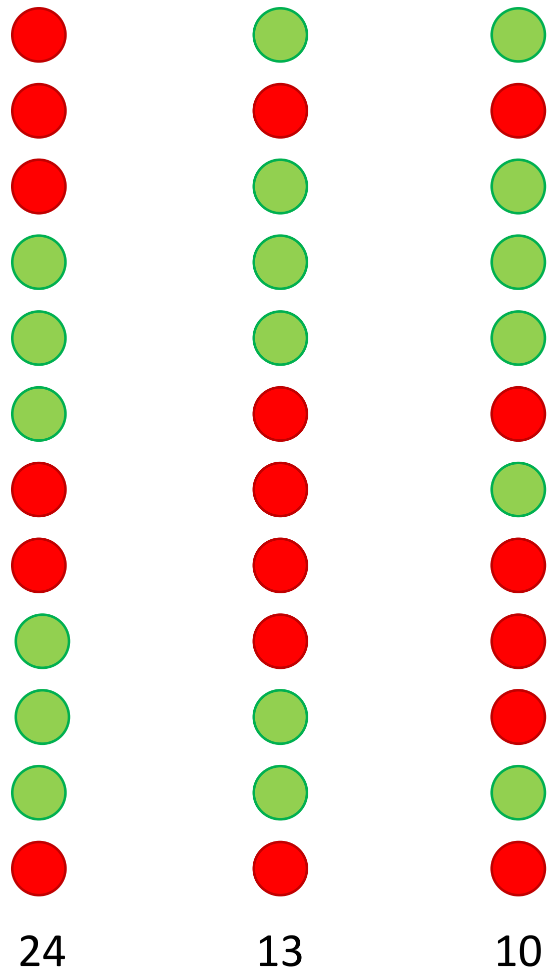
Algorithm
+
Known data associated with some behavior of interest

Temporary Structure
+
New data

Prediction

„Learning“ weights

- Algorithm basically tries out weight combinations.
- For each, it evaluates how many known recidivious criminals are sorted on top of the list.
- The weight combination maximizing this is then taken for further predictions.
- This basic principle can be used for almost anything:
  - News Feed at Facebook
  - Search engines
  - Product recommendations



Red circles = recidivious criminals;

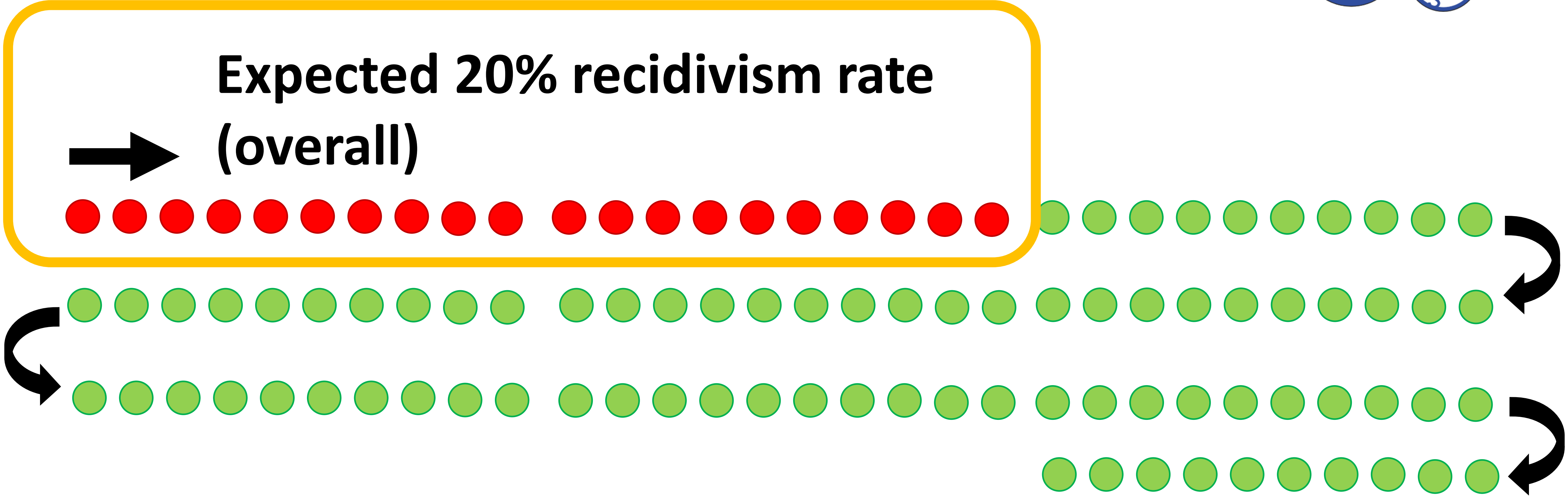
Optimal sorting: All red circles on top

Quality measure: Number of pairs of red and green circles, where red one is above gree one.

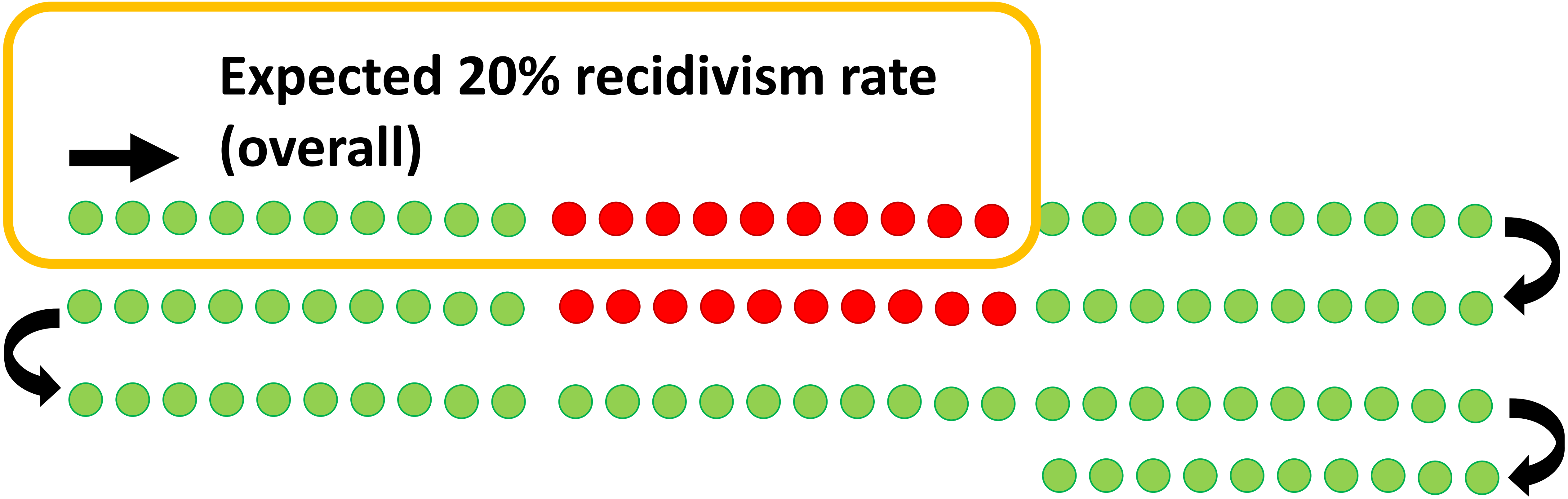
Oregon Recidivism Rate Algorithm

- The above mentioned quality measure is: 72 of 100 pairs are correctly sorted.
- The algorithm used in Oregon will thus, given on recidivious and one non-recidivious person, give a higher number to the recidivist in 3 out of 4 cases.
- Only about a quarter predictions are wrong – doesn‘t that sound good?
- Unfortunately, that‘s not the way a judges makes his sentence.
- Problem: the two classes do not have the same size (imbalanced)
  - Of 1000 culprits
  - About 2000 will relapse into criminal behavior

Optimal Sorting



Possible sorting with the above quality (75/100 Paaren)



Expected 20% recidivism rate (overall)

The recidivism rate algorithm COMPAS is racist (Propublica)

- In a study by Propublica (COMPAS algorithm) the result was even worse:
- Only 20% of the (predicted) violent criminals actually relapsed
    - Considering all kinds of criminal acts, the prediction was slightly better than flipping a coin.
    - The prediction was too pessimistic concerning black criminals;
    - And too optimistic concerning white criminals.
- Northpoint Software designed the algorithm, almost nothing is known about it

https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

Zweig‘s rules

Artificial intelligence looks for patterns in very noisy data.

The patterns are of a **statistical nature**, they show correlations, not causation.

Artificial intelligence almost always tries to identify a very small group of persons (Problem of **imbalance**).

It is used where **there are no simple rules**

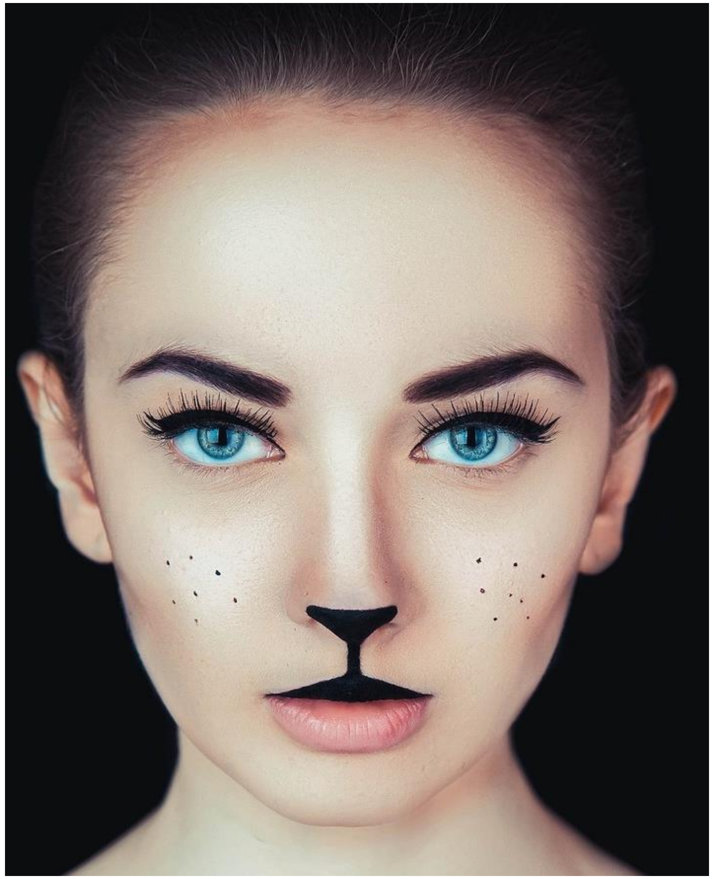
If there were simple rules, we would know them already.

Statistical predictions about the behavior of humans

What does it actually mean?

With 70% you‘re a Criminal….

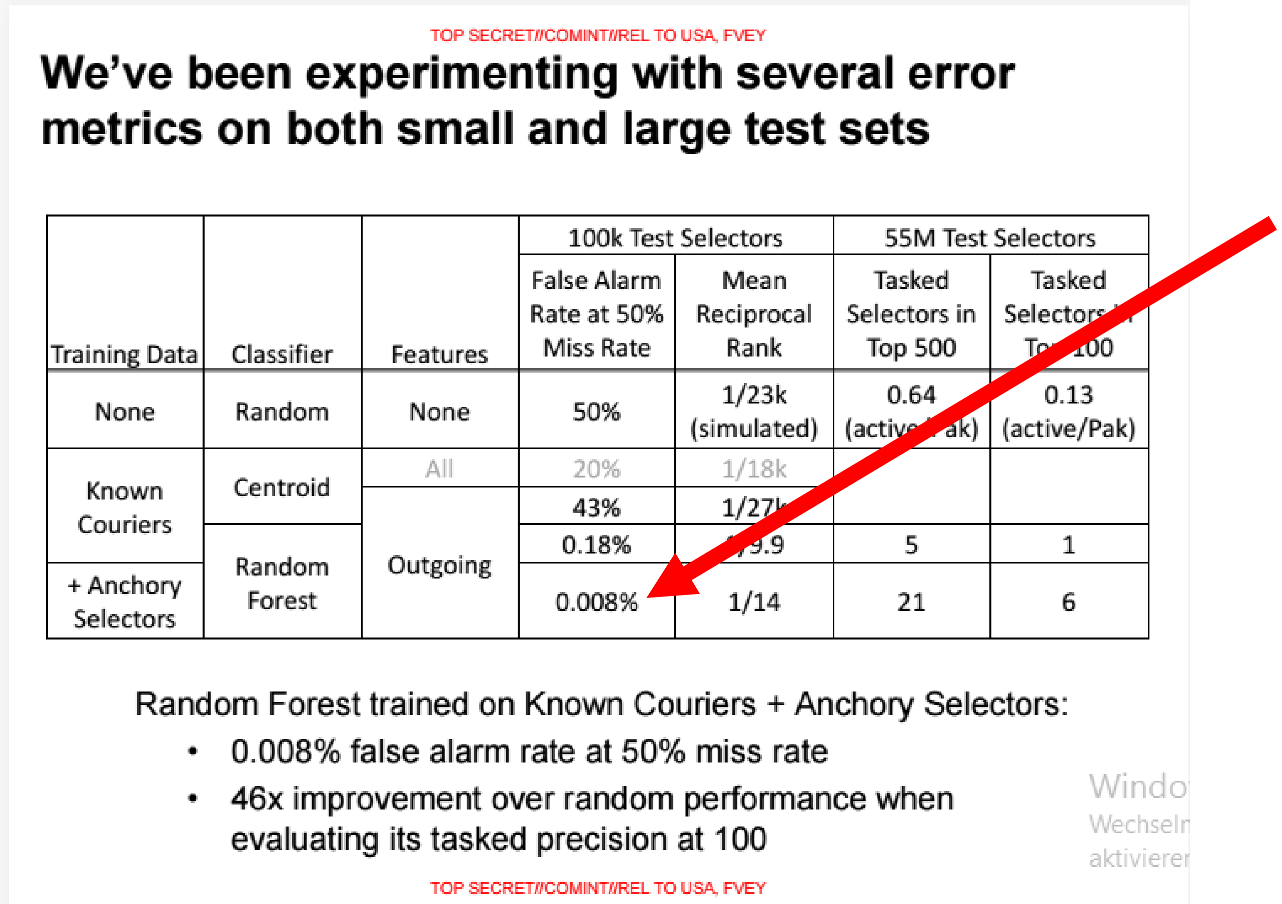
- If this person was a cat with 7 lives, he would commit a crime again in 5 of them…
- No, such an interpretation is absurd!
- **Algorithmic guilt by association**
  - Of 100 persons like you, 70% will commit a crime



Problems

- **Economy of attention** of judges.
- „Best practice“ requires usage of the software
- **Asymmetry of the decision:** The error in disregarding a correct recommendation by the algorithm is much higher than following a wrong recommendation by the algorithm.
- Basic modeling decision and data quality can be very bad.
- The person which is sent to prison can very principally not prove the algorithm wrong!
  - This is true also for: credit worthiness, education offers, job offers, people killed by drones, …

Terrorist identification SKYNET



TOP SECRET//COMINT//REL TO USA, FVEY

We've been experimenting with several error metrics on both small and large test sets

			100k Test Selectors		55M Test Selectors	
Training Data	Classifier	Features	False Alarm Rate at 50% Miss Rate	Mean Reciprocal Rank	Tasked Selectors in Top 500	Tasked Selectors in Top 100
None	Random	None	50%	1/23k (simulated)	0.64 (active/Pak)	0.13 (active/Pak)
Known Couriers	Centroid	All	20%	1/18k		
		Outgoing	43%	1/27k		
	Random Forest		0.18%	1/9.9	5	1
+ Anchory Selectors			0.008%	1/14	21	6

Random Forest trained on Known Couriers + Anchory Selectors:

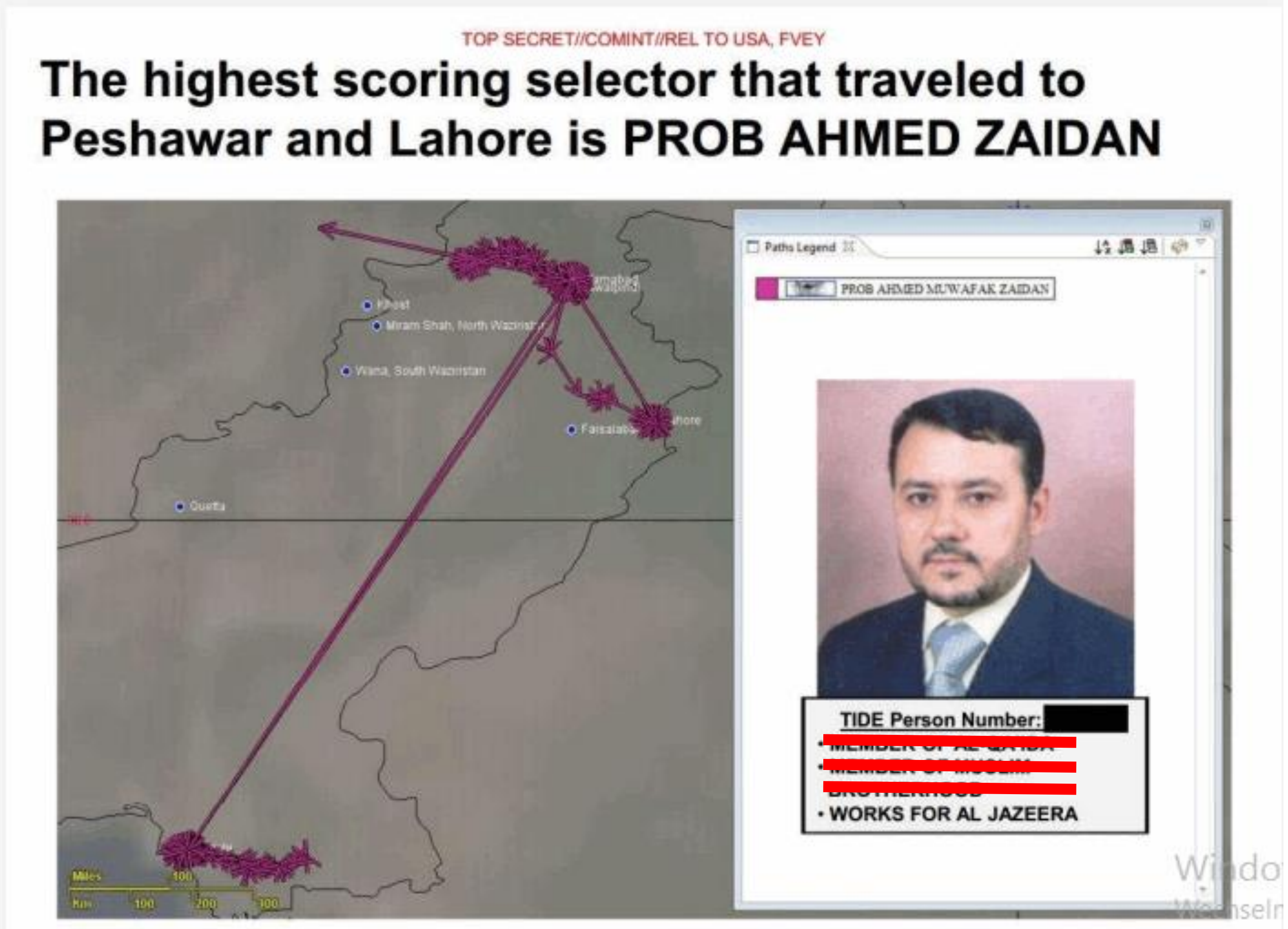
- 0.008% false alarm rate at 50% miss rate
- 46x improvement over random performance when evaluating its tasked precision at 100

Windo
Wechseln
aktivieren

TOP SECRET//COMINT//REL TO USA, FVEY

https://theintercept.com/document/2015/05/08/skynet-courier/
https://theintercept.com/2015/05/08/u-s-government-designated-prominent-al-jazeera-journalist-al-qaeda-member-put-watch-list/

Top-“Courier“ according to the algorithm is…



TOP SECRET//COMINT//REL TO USA, FVEY

The highest scoring selector that traveled to Peshawar and Lahore is PROB AHMED ZAIDAN

TIDE Person Number: [redacted]

- ~~MEMBER OF AL-QA'IDA~~
- ~~MEMBER OF MUSLIM BROTHERHOOD~~
- WORKS FOR AL JAZEERA

Relevance of algorithms

Is there a neutral recommendation in search engines, news apps, or social networks?

- Most companies state they would sort news only by relevance.
- But: all recommendation algorithms filter, learn, and sort.
- They are based on modeling decisions, select certain variables and not others, and learn only from a small part of the data.
- All of these steps can be done well or less well. However, none of the steps is neutral in the sense of ‚objective‘.

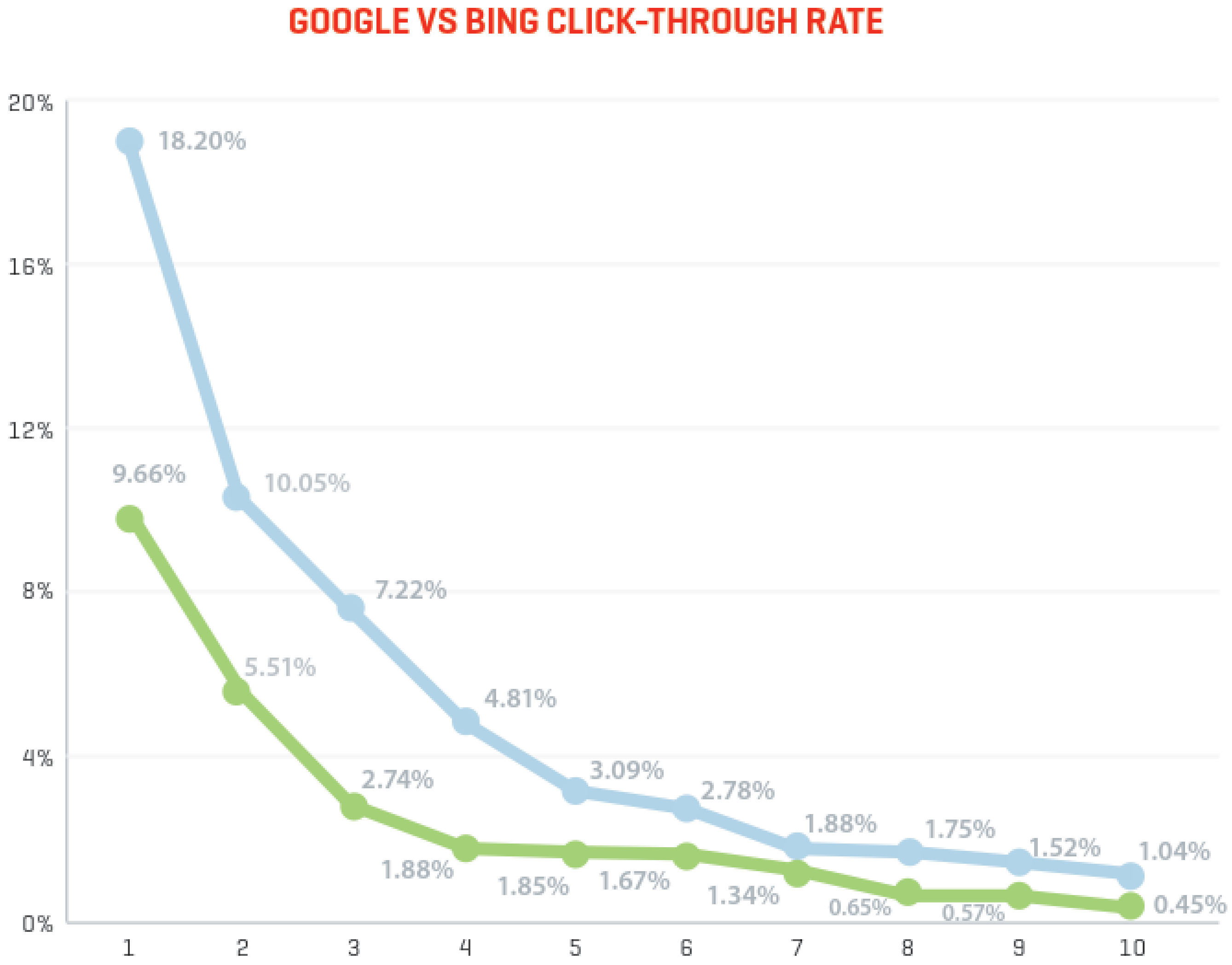
Spielkamp‘s Rule

All algorithms are objective…

…as long as they are not designed by humans!

Human Perception influenced by Algorithms

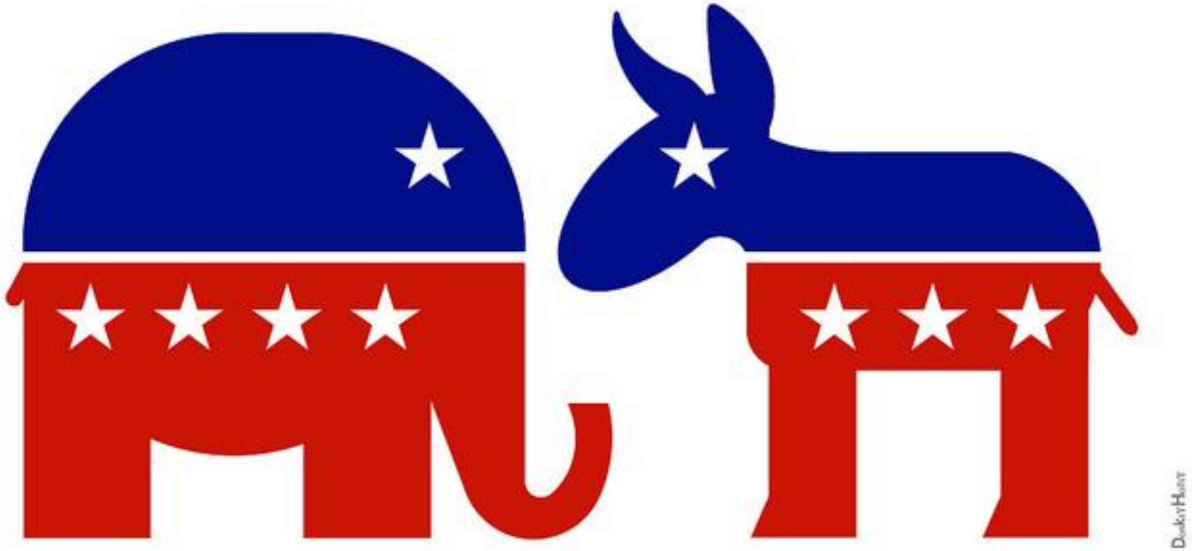
How important is being first?



Paul Davison at Digital Relevance™: „A Tale of Two Studies: Establishing Google & Bing Click-Through Rates“, Study by Digital Relevance™ using client data from Jan-June 2011, available from http://connect.relevance.com/a-tale-of-two-studies-establishing-google-bing-click_through-rates or research@relevance.com; published 2013.

Digitalrelevance, 8900 Keystone Crossing, Suite 100, Indianapolis, IN 46240

Does Google really favor democrats?



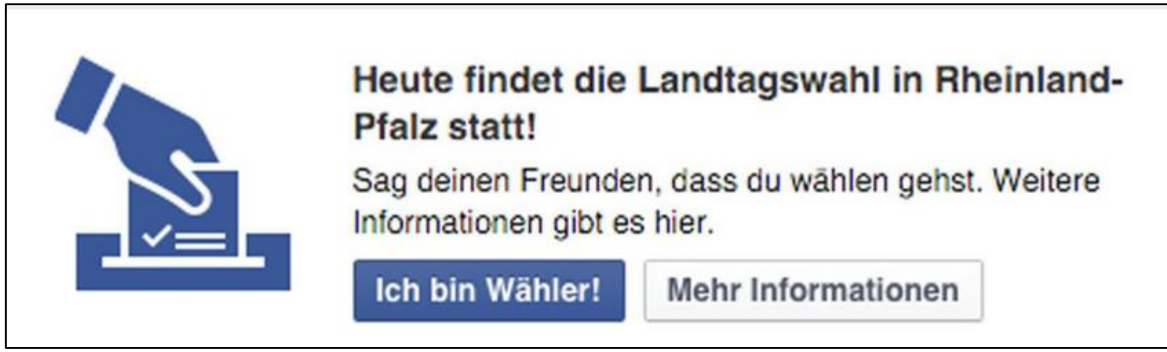
Studie by Trielli, Mussenden and Diakopolous[1]:

Among 16 candidates (USA) there were 7 positive results among the first 10 search engine results for democratic candidates, and only 5.9 among conservative candidates.

Is this difference significant?

[1] http://algorithmwatch.org/warum-die-google-suchergebnisse-in-den-usa-die-demokraten-bevorteile/

Can we be influenced?



Heute findet die Landtagswahl in Rheinland-Pfalz statt!

Sag deinen Freunden, dass du wählen gehst. Weitere Informationen gibt es hier.

- Order of search engine results:
  - Manipulated order is not recognized by users and might change the behavior of undecided voters (Epstein & Robertson, 2015)
- Facebook‘s „Vote“-Button
  - First small studie by Bond et al.
  - Effect (at last election in 2012) was small, but might have been around. 60.000 more votes overall.

Epstein, R. & Robertson, R. E.: “The search engine manipulation effect (SEME) and its possible impact on the outcomes of elections”, Proceedings of the National Academy of Science, 2015, E4512-E4521

Bond, R. M.; Fariss, C. J.; Jones, J. J.; Kramer, A. D. I.; Marlow, C.; Settle, J. E. & Fowler, J. H.: “A 61-million-person experiment in social influence and political mobilization”, Nature, 2012, 489, 295-298

Zuccherosconi

If Zuckerberg wanted to go for presidency –

who would be able to win against him by political topics alone?

Bild von Berlusconi: European People's Party - https://www.flickr.com/photos/eppofficial/22194872598/, CC BY 2.0, https://commons.wikimedia.org/w/index.php?curid=46340059

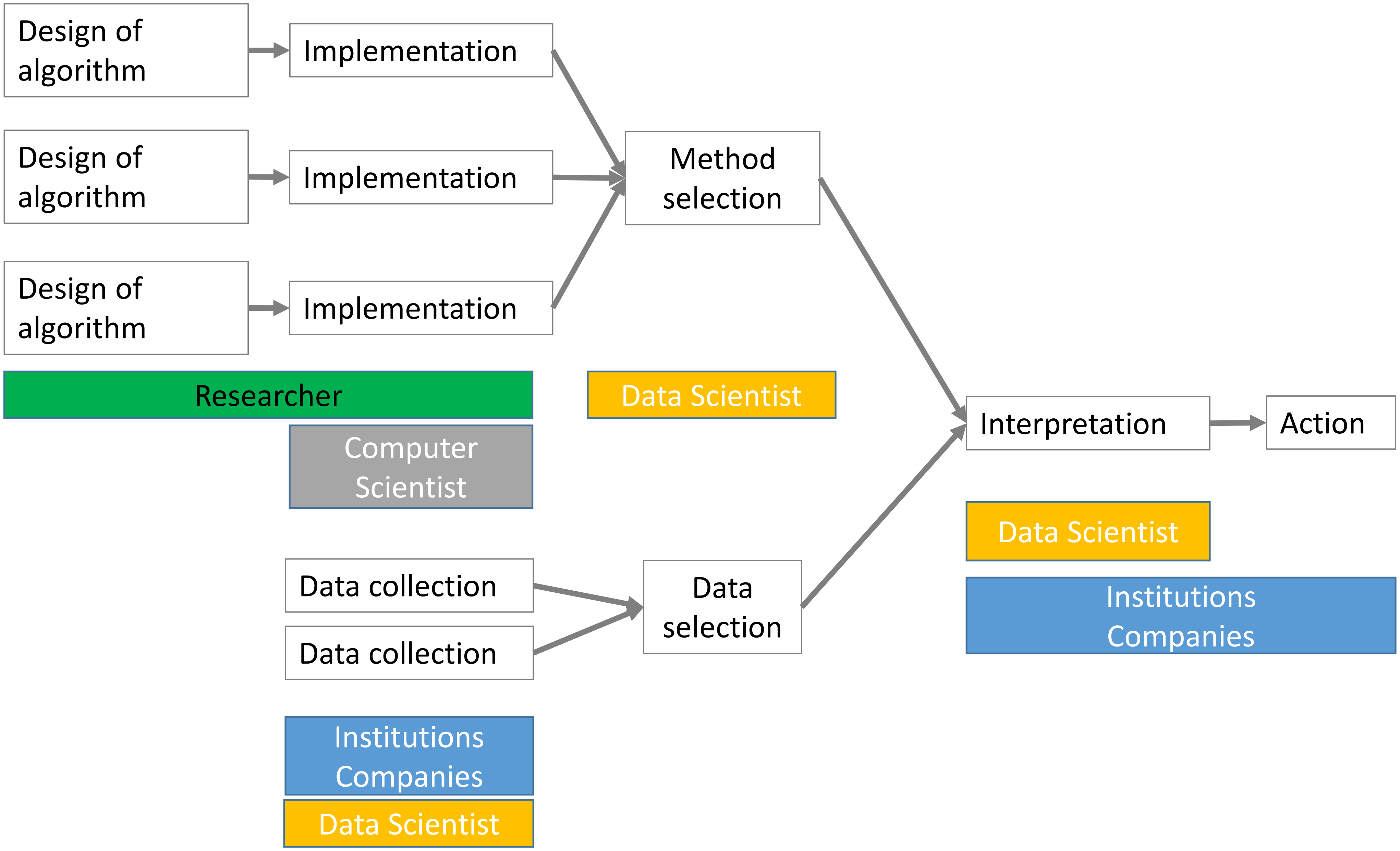
Von Guillaume Paumier - Eigenes Werk, CC BY 3.0, https://commons.wikimedia.org/w/index.php?curid=21461459

Algorithms in a democracy

Algorithms and Equality

- Not all will be in the focus of algorithms in the same way. The more individual you are (the richer, educated, more embedded), the less you will be in the focus. (Cathy O'Neil, Weapons of Math Destruction, 2016)
- There is the risk of building smaller and smaller groups of people by algorithms, e.g., for insurances
- Those that are once in the „wrong" group, cannot always prove that they are in the wrong group. True for credit worthiness, job applications, arrests, being accepted to university, etc.

Who is responsible?



Design of algorithm → Implementation

Design of algorithm → Implementation

Design of algorithm → Implementation

Implementation → Method selection

Researcher

Computer Scientist

Data Scientist

Data collection

Data collection

Data collection → Data selection

Institutions Companies

Data Scientist

Method selection → Interpretation

Data selection → Interpretation

Interpretation → Action

Data Scientist

Institutions Companies

Sozio-informatik

Who holds algorithms accountable?

Media?
Society?
Government?
NGOs?
Companies?
Law?

Quis custodiet ipsos algorithmos

Do we need an „Automated Decision Making“-Safety Commission?

(aka Algorithm TÜV or Algorithm MOT)

„Algorithm Watch“

Lorena Jaume-Palasí, Law Philosophy

Lorenz Matzat, Data Journalist

Matthias Spielkamp, Journalist

Prof. Dr. K.A. Zweig, TU Kaiserslautern

ALGORITHM WATCH

http://algorithmwatch.org/
(mainly in German)

Necessary properties of an algorithmic safety commission

- Independency, democratic legitimation
- Possibly with research budget
- Identification of the smallest possible set of algorithms to test, since
  - Most algorithms are harmless;
  - Products themselves are already well regulated and tested;
  - Algorithms in a competition might not need further regulation as well.
  - We cannot afford further impediments of innovation**!**
- **Non-Profit**

Algorithm information leaflet

Which problem is actually solved by this algorithm?

What is the domain of this algorithm, are there modeling assumptions that might not always be true?

Which side effects does this algorithm have?

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

With this, …

… for all kinds of risks and possible side effects of the digitization, ask your local data scientist!

Contact

Prof. Dr. Katharina A. Zweig
TU Kaiserslautern
Gottlieb-Daimler-Str. 48
67663 Kaiserslautern
zweig@cs.uni-kl.de

Find the initiative at:

algorithmwatch.org