Anderson MacGyver.

Digitalization & AI – What's it all about?

Bijeenkomst Vereniging van Arbeidsrecht

May 2019

About the speaker

- John Philip Ottersbach
  - German
  - Diploma thesis and Ph.D. in high energy physics working for the ATLAS experiment at CERN
  - Joined Anderson MacGyver in the early start-up phase to help organization with their transition into a data-driven, digital world
    - Digital Strategy
    - Digital Organization
    - Digital Solutions

Introduction

An international research facility

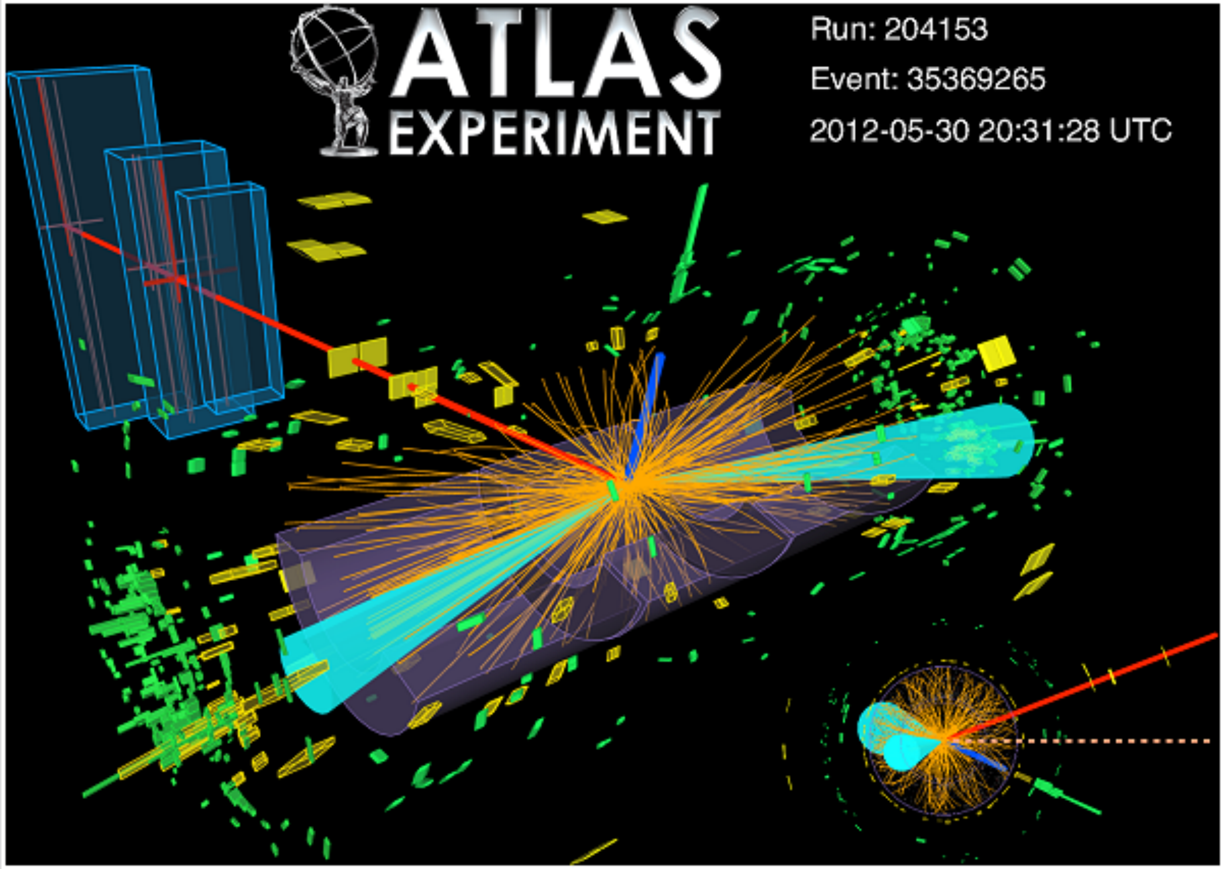
- CERN (Conseil Européen pour la Recherche Nucleaire):
  - Facility for fundamental research on matter in the Geneva area
  - Some 15.000 researches from 608 institutes/universities in 113 nations contribute
  - Objective: “Seek and find answer to the origin and nature of our universe.”

Accelerating science

- Large Hadron Collider:
  - 27 km circumference proton accelerator
  - Center of mass energy: 14 TeV → important to produce heavy, i.e. rare particles ($E=\gamma mc^2$)
  - 4 dedicated collision points, experiments

Challenges and reward

- Finding the Higgs particle after an intensive search
- Analyzing years of collision data - 40 million events per second → 60 Terabyte of data per second
- Nobel prize 2013



From science to society and economy

- Developed at/in cooperation with CERN:
  - world wide web (Tim Berners-Lee)
  - grid computing
  - close to infinite hardware and software to deal with enormous amounts of data
- From science to economy = from fact to fiction (marketing)

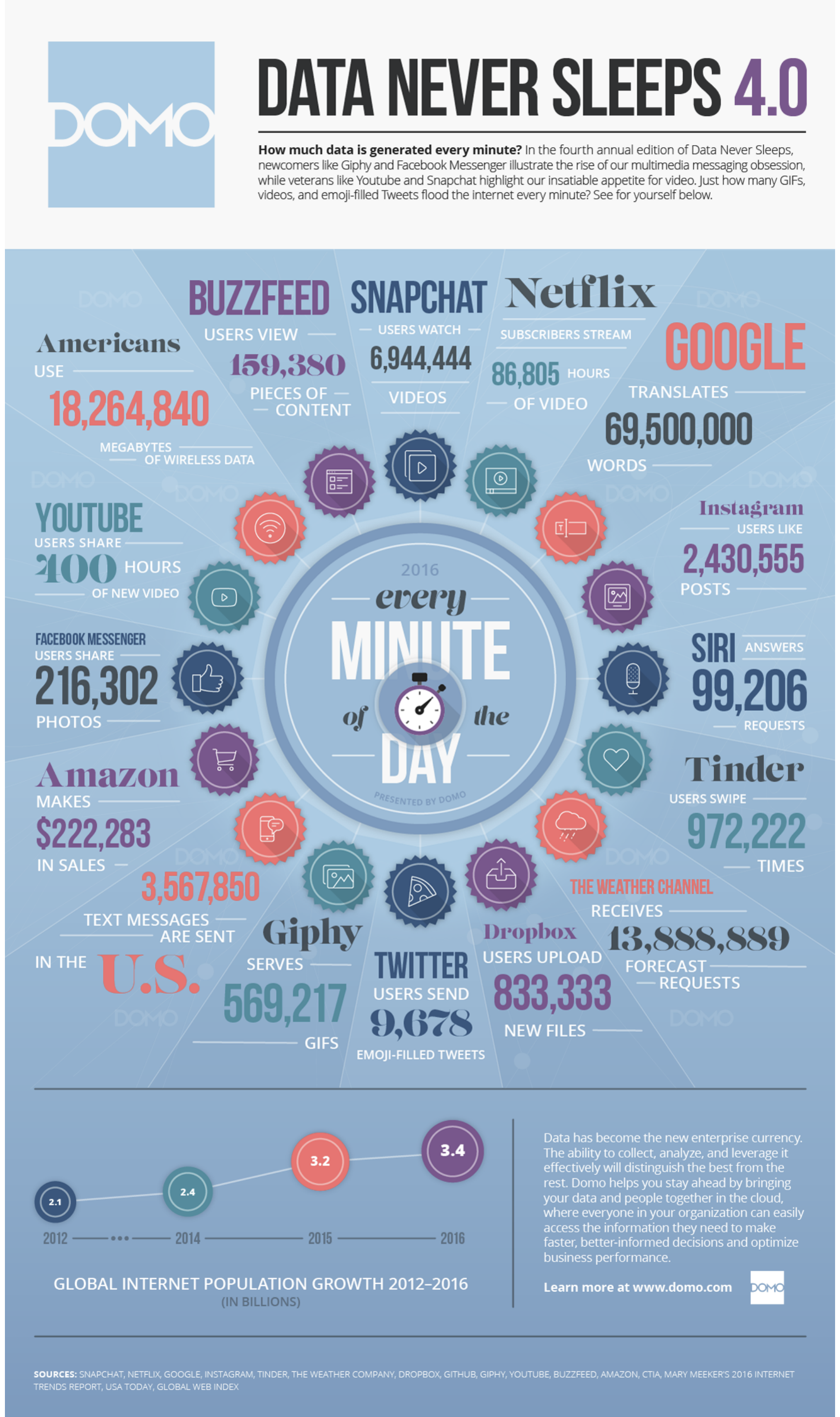
The hype

Marketing vs. reality

The world of data

Data as the new oil

- The amount of data is growing exponentially at a rate of about 40% p.a. => the world's data volume is doubling every 2 years
- > 55% of the global population has internet access
- Every minute: >31 million Facebook messages, 400h YouTube movies, etc.
- Technology finds one's way into more and more domains of our life's (example: wearable's like the Apple Watch) => "Internet of Things"
- We create (un)consciously more and more data
- However, it needs to be "refined" to add value

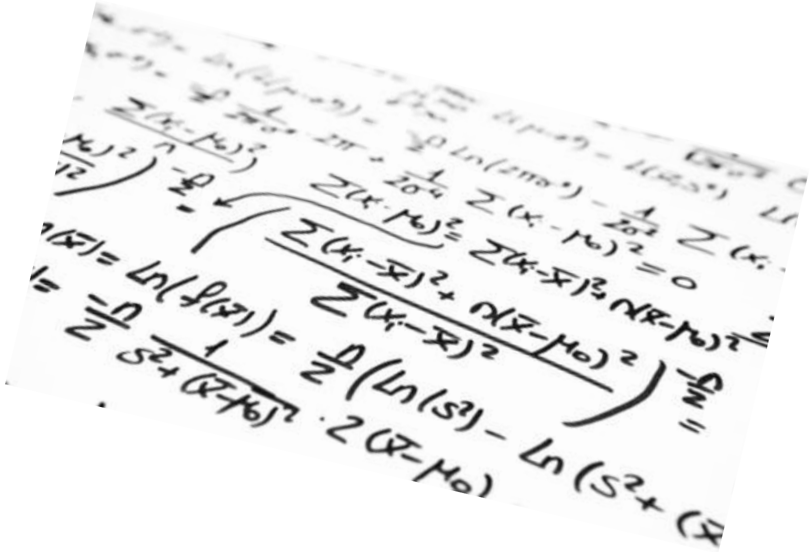


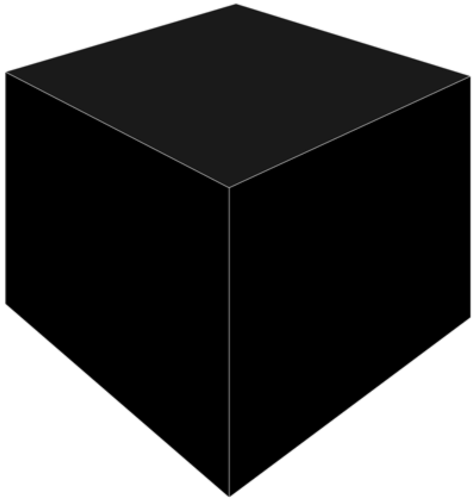
What data is good for …

Refining the new oil

The science of collecting, organizing and drawing conclusions from data:

- Classical statistics:
  - A domain of mathematics, partially centuries old, well proven and tested
  - Charts/graphs, parameter estimation, hypothesis testing, regression, maximum likelihood, …
  - Results are easily verifiable
- Data mining:
  - A relatively recent development (made possible through the advance of computing power -> Moore’s law)
  - Algorithms for predictions, classifications, segmentation (e.g. neural networks, boosted decision tree’s, support vector machines, …)
  - Results are often difficult to verify => black boxes

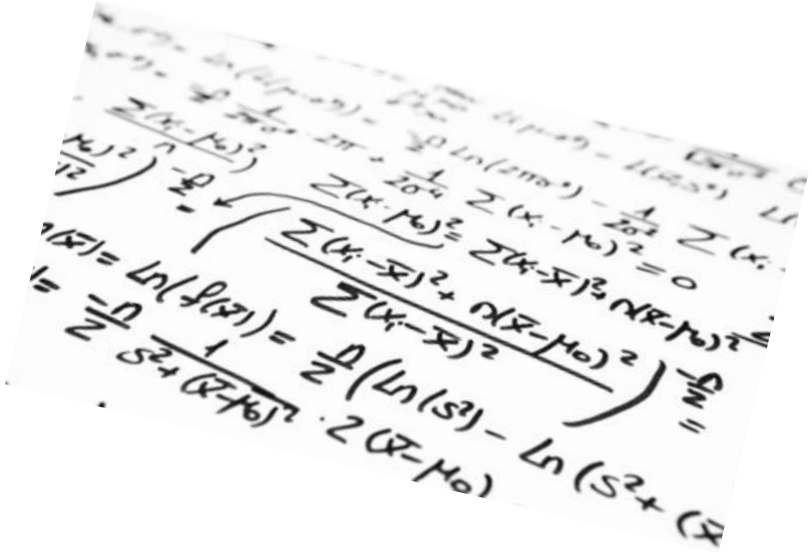


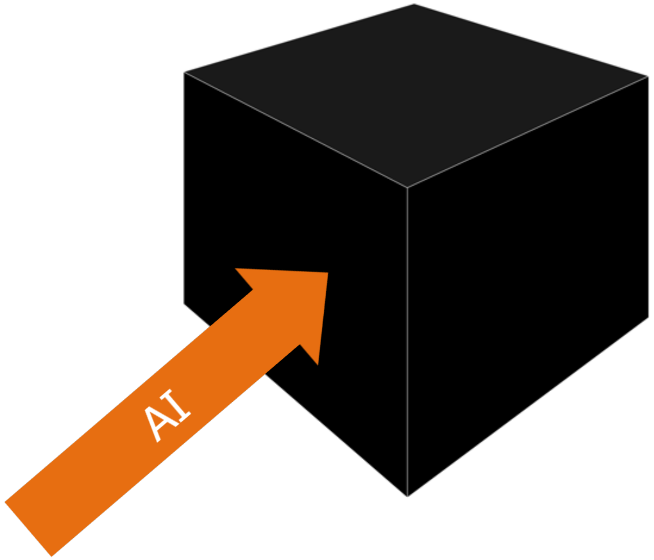


Refining the new oil

The science of collecting, organizing and drawing conclusions from data:

- Classical statistics:
  - A domain of mathematics, partially centuries old, well proven and tested
  - Charts/graphs, parameter estimation, hypothesis testing, regression, maximum likelihood, …
  - Results are easily verifiable
- Data mining:
  - A relatively recent development (made possible through the advance of computing power -> Moore’s law)
  - Algorithms for predictions, classifications, segmentation (e.g. neural networks, boosted decision tree’s, support vector machines, …)
  - Results are often difficult to verify => black boxes





The beginnings

- Artificial Intelligence is not a well and uniquely defined term
- Dynamic definitions; AI effect → as machines become more capable, tasks that are considered to require "intelligence" are removed from the definition of AI
- In the past decade, term captured by marketeers
- Research started in the mid 1950's
- Boost through nowadays available computing capacities

"The ability of a digital computer or computer-controlled robot to perform tasks commonly associated with intelligent beings."

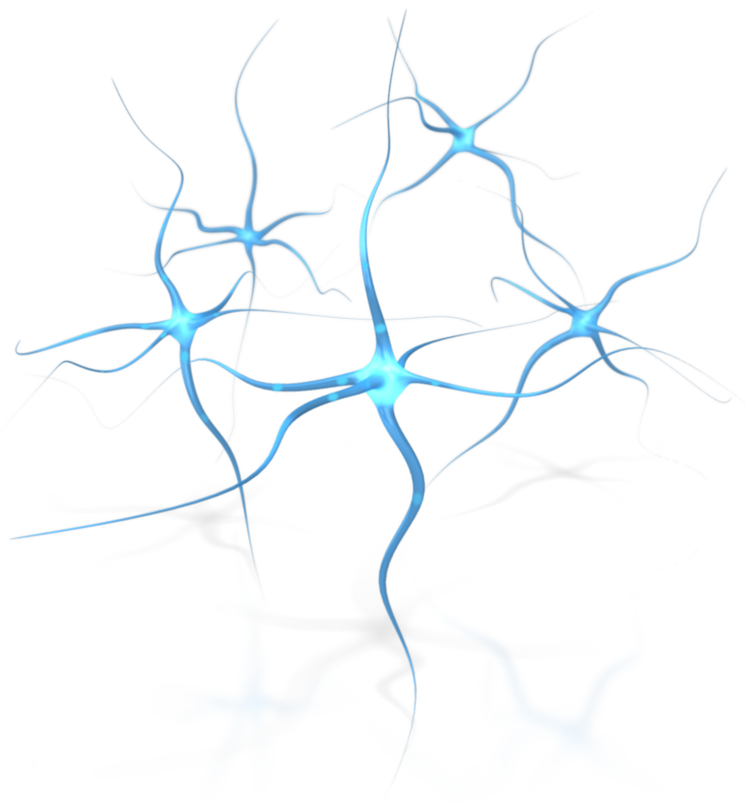
Encyclopedia Britannica

"The capability of a machine to imitate intelligent human behavior."

Merriam-Webster

Neural networks

- First publication on the basic perceptron algorithm (Rossenblatt, 1957) => breakthrough in the 1990's with development of *multilayer perceptrons* and *backpropagation*
- Multilayer perceptrons have input, often hidden, and output layer
- Each connection has a weight (a real number)
- Each node performs a weighted sum of its inputs and thresholds the result (often using a sigmoid function)
- The output is determined using a linear combination of weighted input from the hidden layer
- Perceptron convergence theorem: Converges when cycled repeatedly through the training data, provided the problem is linearly separable!

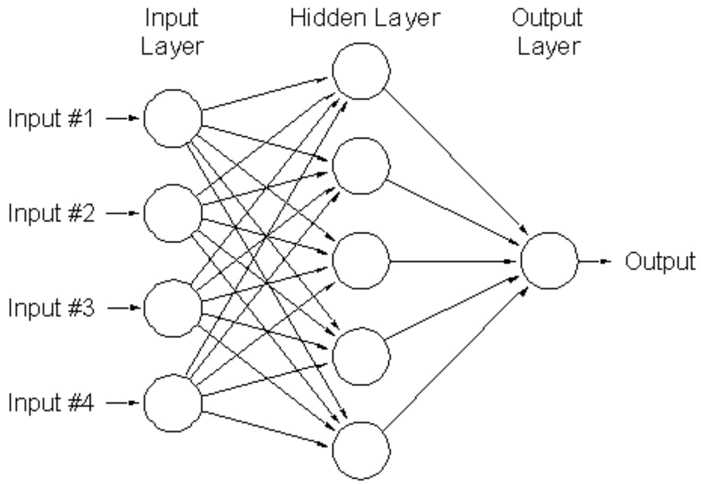


Neural networks

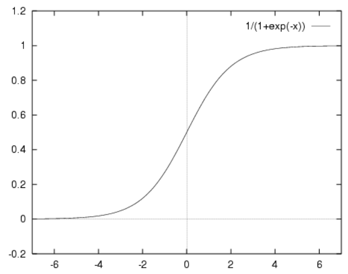
- First publication on the basic perceptron algorithm (Rossenblatt, 1957) => breakthrough in the 1990's with development of *multilayer perceptrons* and *backpropagation*
- Multilayer perceptrons have input, often hidden, and output layer
- Each connection has a weight (a real number)
- Each node performs a weighted sum of its inputs and thresholds the result (often using a sigmoid function)
- The output is determined using a linear combination of weighted input from the hidden layer
- Perceptron convergence theorem: Converges when cycled repeatedly through the training data, provided the problem is linearly separable!

attributes

class



$$x = w_0 + w_1 a_1 + w_2 a_2 + \ldots + w_n a_n = \sum_{i=0}^{n} w_i a_i$$



The case

A challenge for a company in the logistics sector:

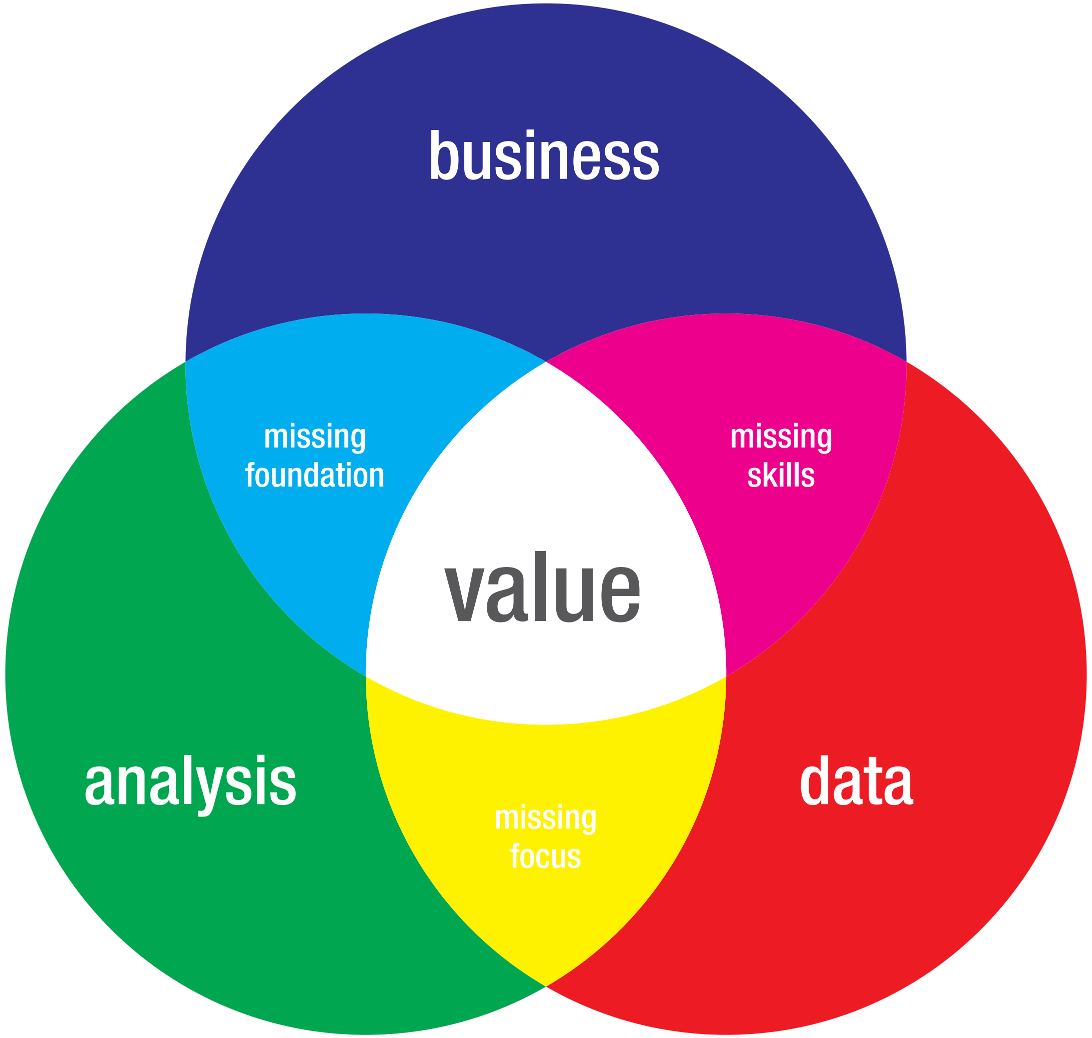
- Around 5000 independent parties can deliver goods daily and unannounced
- Rented cars and deployed staff for internal processing/distribution of the goods are expensive => good, reliable planning is crucial

Approach

- Can we predict the quantity of delivered goods based on historic data?
- The *class* is here the quantity of delivered goods (in number of cars) => regression problem
- Challenges and approach:
  - Find *attributes* that determine/have a large impact on the *class*
  - If we can find *attributes* that have an impact on future instances of the *class*, we can make predictions!
  - Find an algorithm that handles this specific regression problem well

A real-world example of applied AI

Finding the right attributes



A real-world example of applied AI

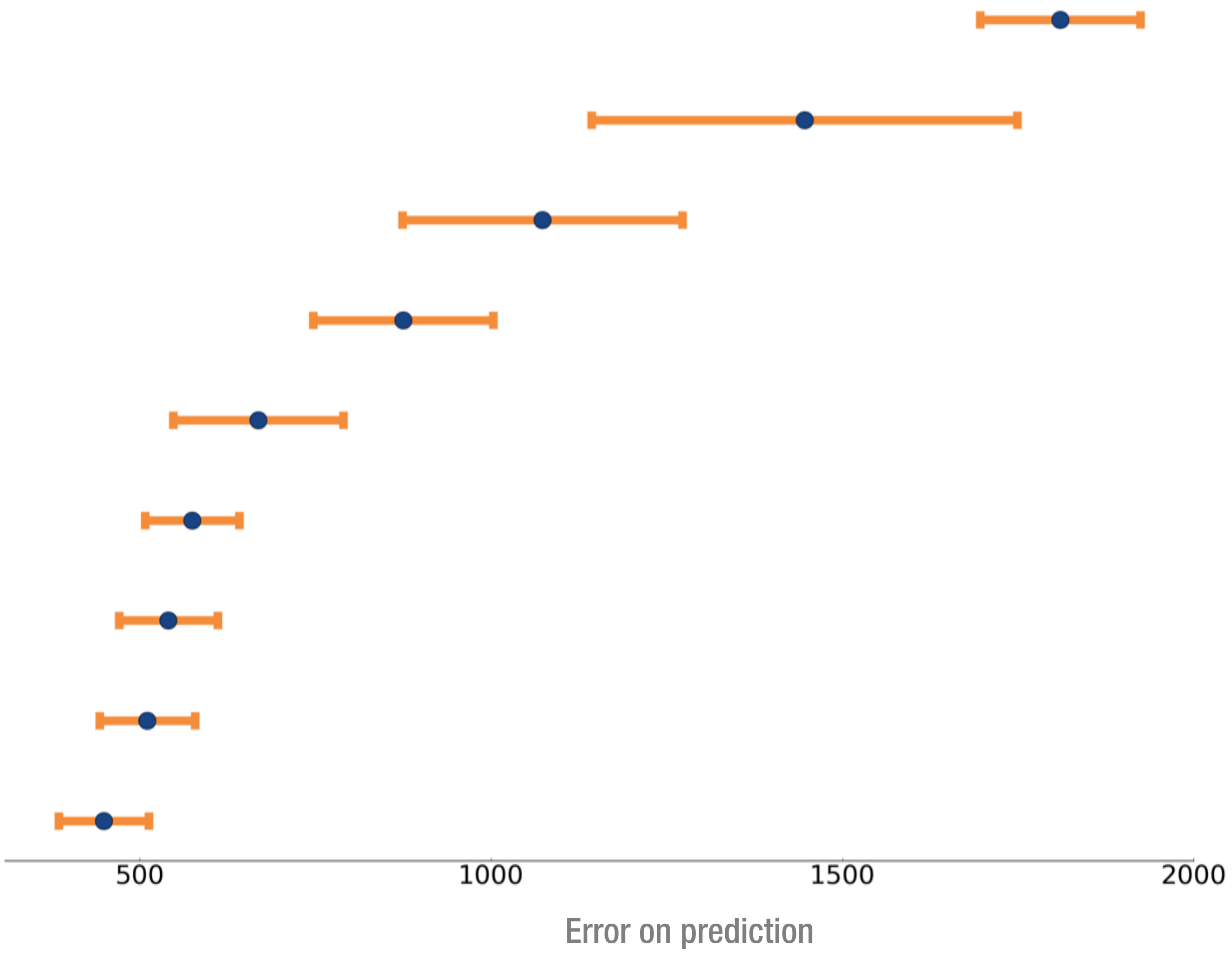
Finding the right attributes

Heuristic, iterative process:

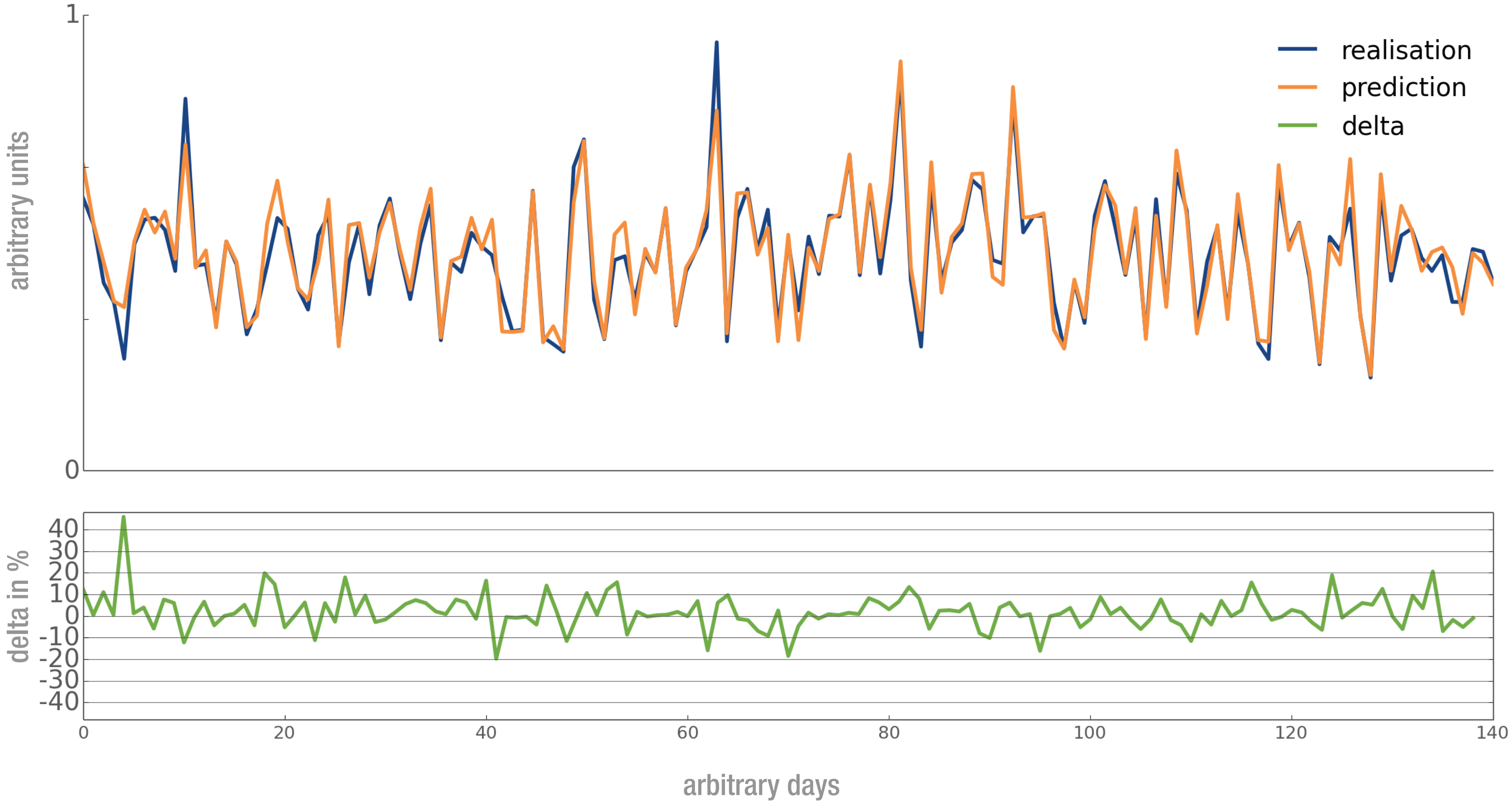
1. Think of effects that might have an influence on the *class*. What might a model of the *class* depend on (e.g. the weather) ?
2. Translate them into variables that can be computed using data. What data will be needed? How can we get access to still missing data (e.g.: use open weather data to get access to temperatures, precipitation, etc.) ?
3. Try different implementations of each variable and see what works best (e.g.: average air temperature in different altitudes, averages over different time periods, etc.)
4. Continue adding attributes until the optimal performance of the model is reached

Attribute selection

- Benchmark (ZeroR)
- Neural network
  - + Holidays
  - + Weather + pricing
  - + Trend
  - + Parameter optimisation
  - + Additional data
  - + Holiday optimisation
  - + Removing history



Results


1
0
arbitrary units
realisation
prediction
delta
delta in %
40
30
20
10
0
-10
-20
-30
-40
0
20
40
60
80
100
120
140
arbitrary days

Take aways

- What's in the black box?
- Artificial intelligence – painstakingly stupid
- The smarter you want to get, the more data (of the same type) you need
- Predictions – looking back in time
- Nevertheless, with the right amount of data and proper analysis you can get a very long way with AI (e.g. Siri, autonomous cars, real-time translation, etc.)
- It's booming and it will continue to do so! Focus areas:
  - specific computer chips for more computing power
  - lots of R&D to shed some light on that black box

Anderson MacGyver.

Anderson MacGyver B.V. | Kronenburgplantsoen 10, 3401 BP IJsselstein | www.andersonmacgyver.com | info@andersonmacgyver.com
https://www.linkedin.com/company/anderson-macgyver https://twitter.com/andersonmcgyver