

MACHINE LEARNING

DATA SCIENCE FOR BETTER DECISIONS

Bart Buelens, 17 Dec 2019, Brussels

BETTER DECISIONS

BETTER?



Data

Evidence based
Data access
Data architectures

BETTER?

Data

Evidence based
Data access
Data architectures

Methods

Algorithms
Uncertainty quantification
Trustworthiness

MACHINE LEARNING

DEFINITION & EXAMPLES

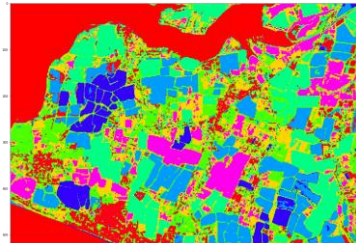
Learning of relation between inputs X -> outputs Y

- Supervised learning
Household characteristics -> spending

DEFINITION & EXAMPLES

Learning of relation between inputs X -> outputs Y

- Supervised learning
Household characteristics -> spending

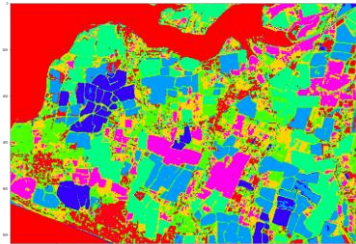


- Unsupervised learning
Satellite images -> vegetation types

DEFINITION & EXAMPLES

Learning of relation between inputs X -> outputs Y

- Supervised learning
Household characteristics -> spending



- Unsupervised learning
Satellite images -> vegetation types



- Reinforcement learning
Moves in game GO -> win the game

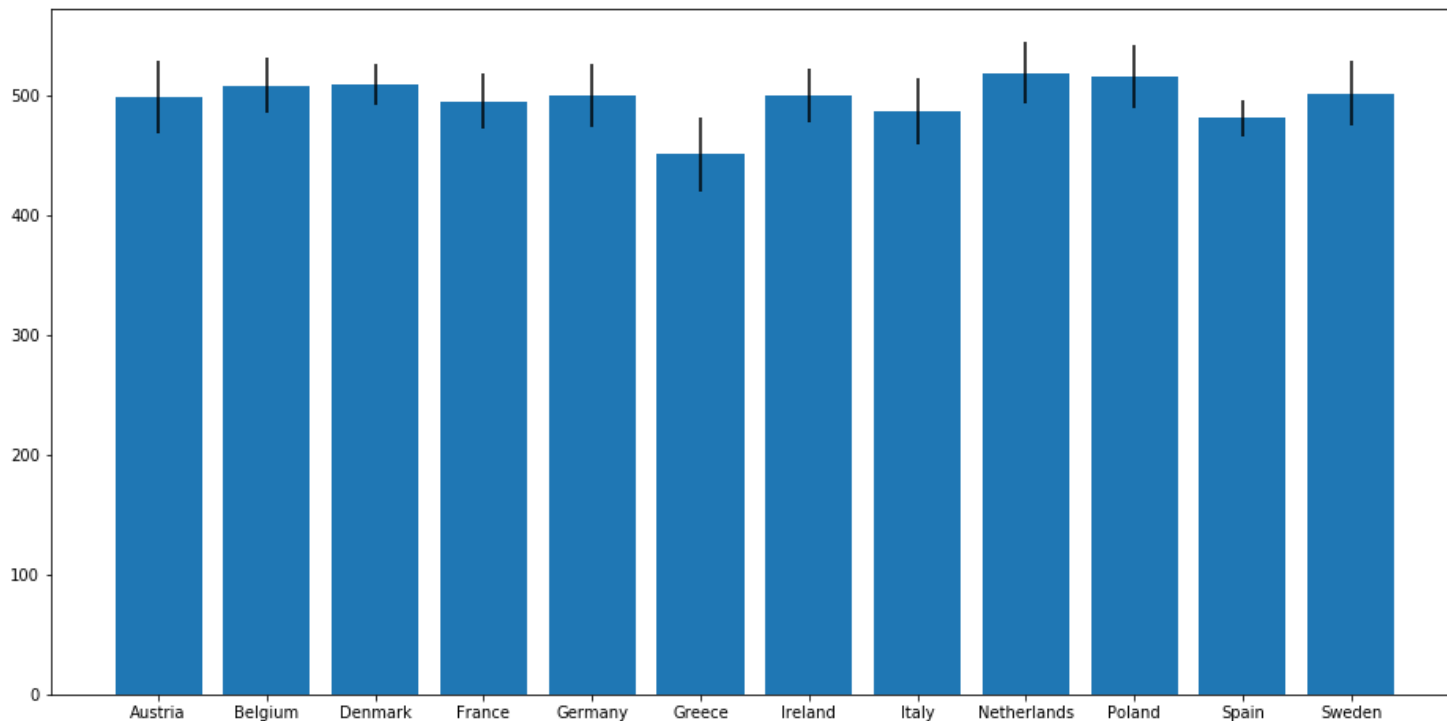
ERROR & UNCERTAINTY QUANTIFICATION (UQ)

Ultimate goal: reduce **uncertainty**

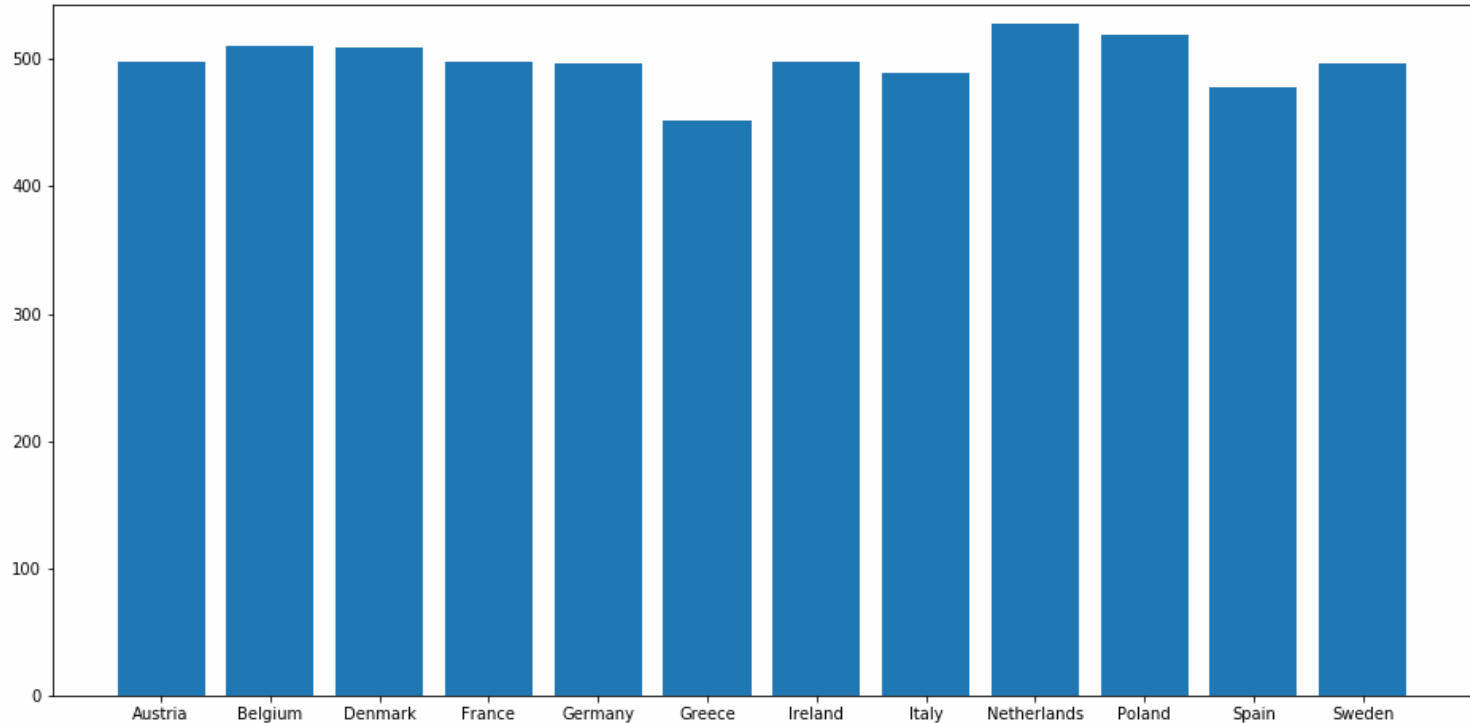
- Key component in predictive modeling and ML
- Insufficiently applied or not at all in predictive analytics in the past

- More & better data facilitates UQ
- More & different methods necessitate UQ

AVERAGES FOR AGE 15 YEARS PISA MATHEMATICS SCALE



AVERAGES FOR AGE 15 YEARS PISA MATHEMATICS SCALE



DATA

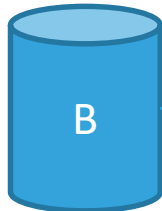
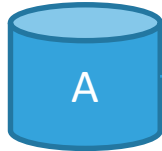
FEDERATED ARCHITECTURES

- As opposed to monolithic data silos
- Privacy preserving data sharing (talk by Chang Sun)
- For example eHealth.be

data keepers

=

data owners



TTP

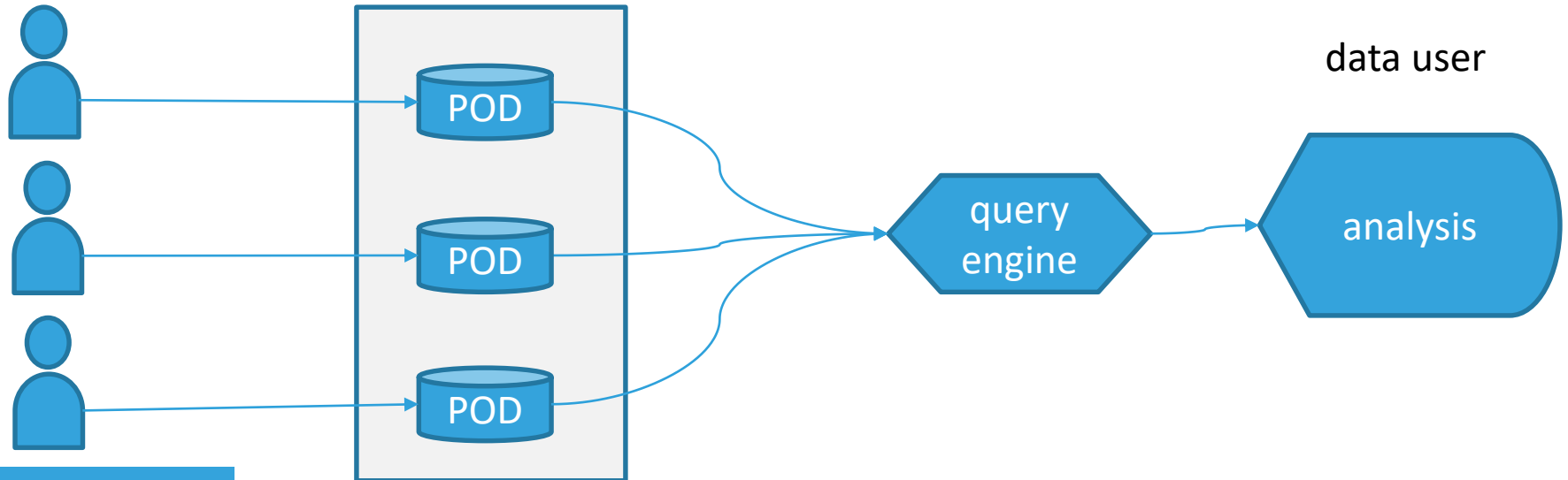
data user

analysis

DECENTRALIZED ARCHITECTURE

data owners

data keepers



DECENTRALIZED ARCHITECTURE

Solid

<https://solid.mit.edu/>

- VITO: pilot solid server
Medical questionnaire
- AI Flanders,
Distributed Data Intelligence,
with U Hasselt, VUB, UGent





Author's name and latest profile picture
stored in author's personal data pod

Work-related opinion about an article
stored in data pod of author's company

Discussed article title and photo
stored in news website's data pod

Likes on this post
each one in different individuals' data pods

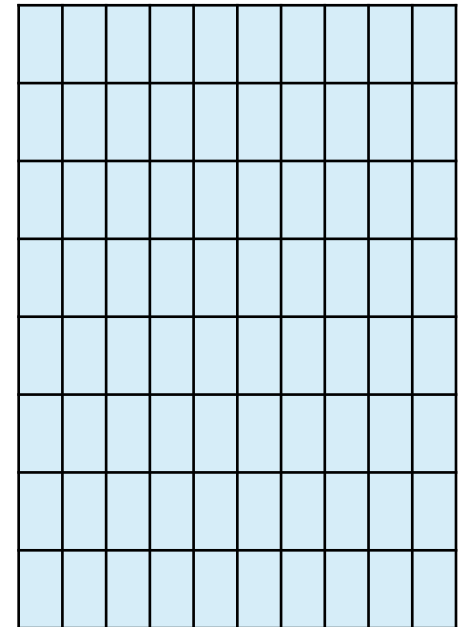
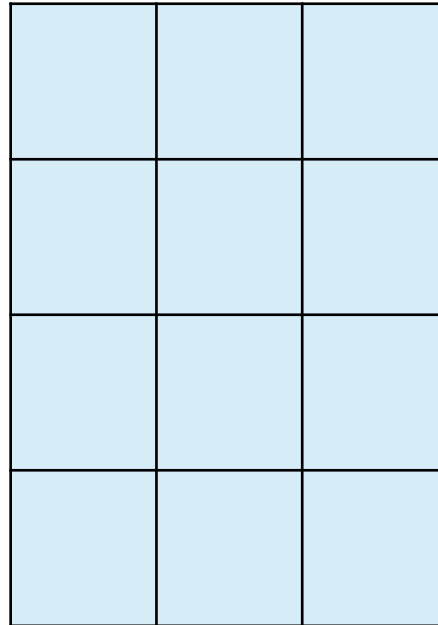
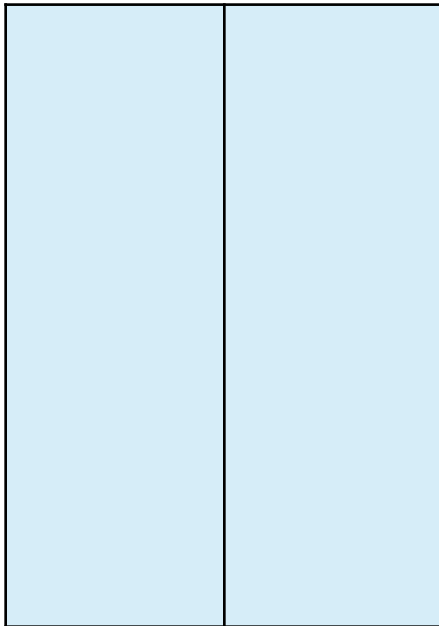
Comments on this post
each one in different individuals' data pods



ALGORITHMS

PRECISION STATISTICS

From stratification to personalization



PRECISION STATISTICS

- N units with p features, p large
- E.g. dichotomous (binary) features, $2^{24} = 16.777.216$

PRECISION STATISTICS

- N units with p features, p large
- E.g. dichotomous (binary) features, $2^{24} = 16.777.216$

Male / female	Netflix subscription	Saw 'breaking bad'	...
Young / old	Activity tracker	Sport once/week	...
Owning / renting	Been abroad	Been to UK	...
Owns a car	Took a train	Train to Bxl	...
High / low income	Is vegetarian	Bought meat	...
Works / not	Has PV panels	Produces > 3 MWh	...
City / rural	Been to a concert	Been to Werchter	...

PRECISION STATISTICS

- Precision farming



<https://watchitgrow.be/>

- Recommender systems (search engines, online shops, ...)
- Personalized advertising, marketing
- Personalized medicine
- Official statistics (Joep Burger)

PRECISION STATISTICS

‘Revival’ of ML methods: because they can handle *large p* appropriately

- penalty methods,
- random forests,
- neural networks,
- ...

ADDED VALUE OF ML (FOR OFFICIAL STATISTICS)

United Nations: **UNECE machine learning project**

Added value: increased relevance, better quality, reduced costs

Base decisions on results from ML-algorithms?

- VITO pilot: Energy Balance
- Early estimates ahead of full data availability



TRUSTWORTHY ML & AI

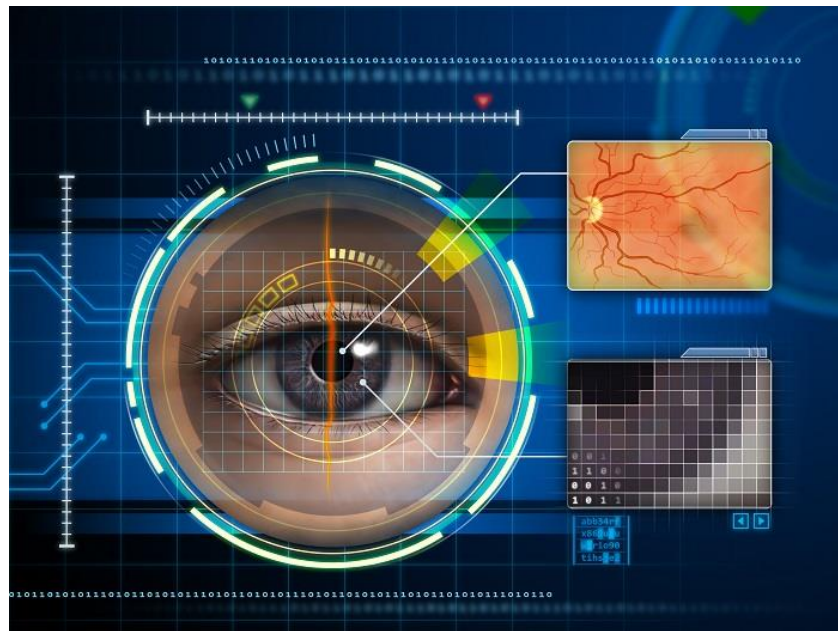
Ethics guidelines for trustworthy AI

Lawful – Ethical – Robust

Human oversight, resilience to attacks, accuracy, reliability, reproducibility, respect for data privacy, traceability, explainability, fair, non-discriminatory, accountability, ...

<https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai>

Assessment of guidelines @ VITO
Deep Learning on retinal images
<https://mona.health/>



CONCLUSION

Coming up next

Methods

Joep Burger

Machine learning in official statistics: two case studies

Data

Chang Sun

Analyzing personal data from multiple sources
in a privacy-preserving manner

THANK YOU

Bart Buelens
bart.buelens@vito.be