Utrecht University

On the Role of AI/ML/Data in ChemEng. Education

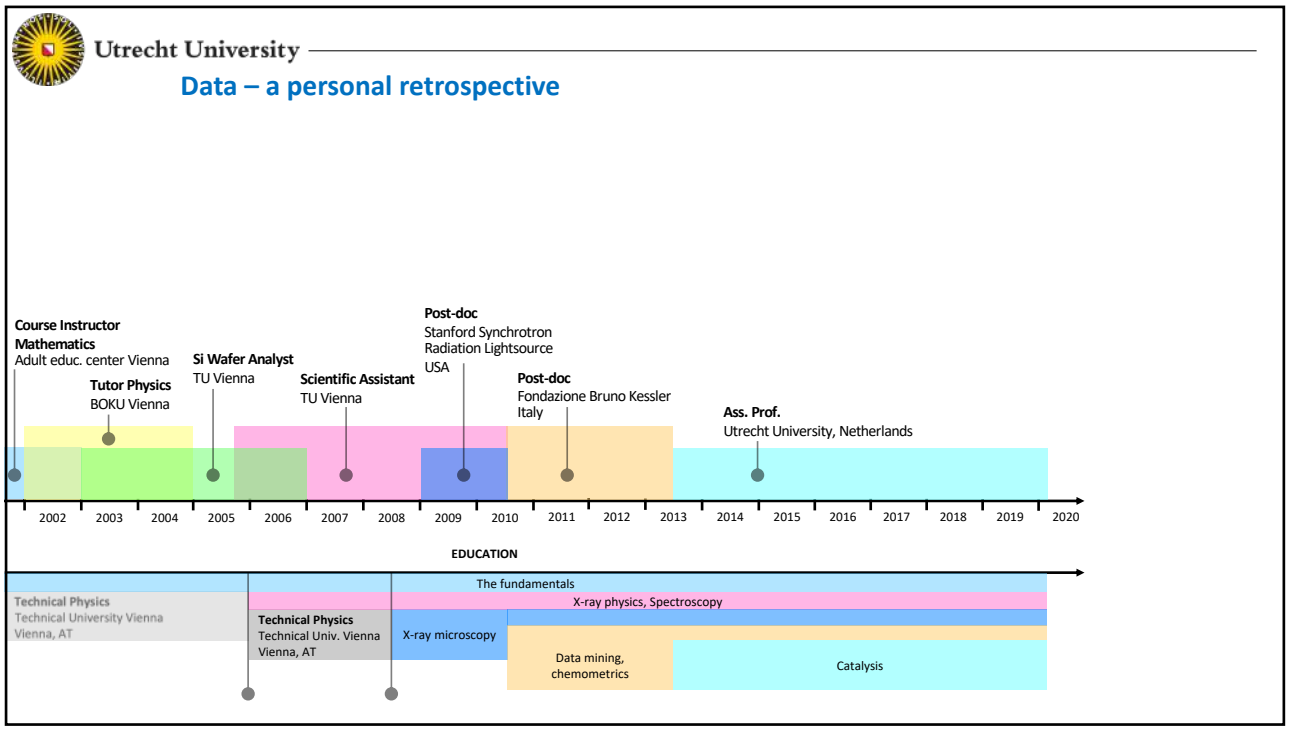
Some personal views

2

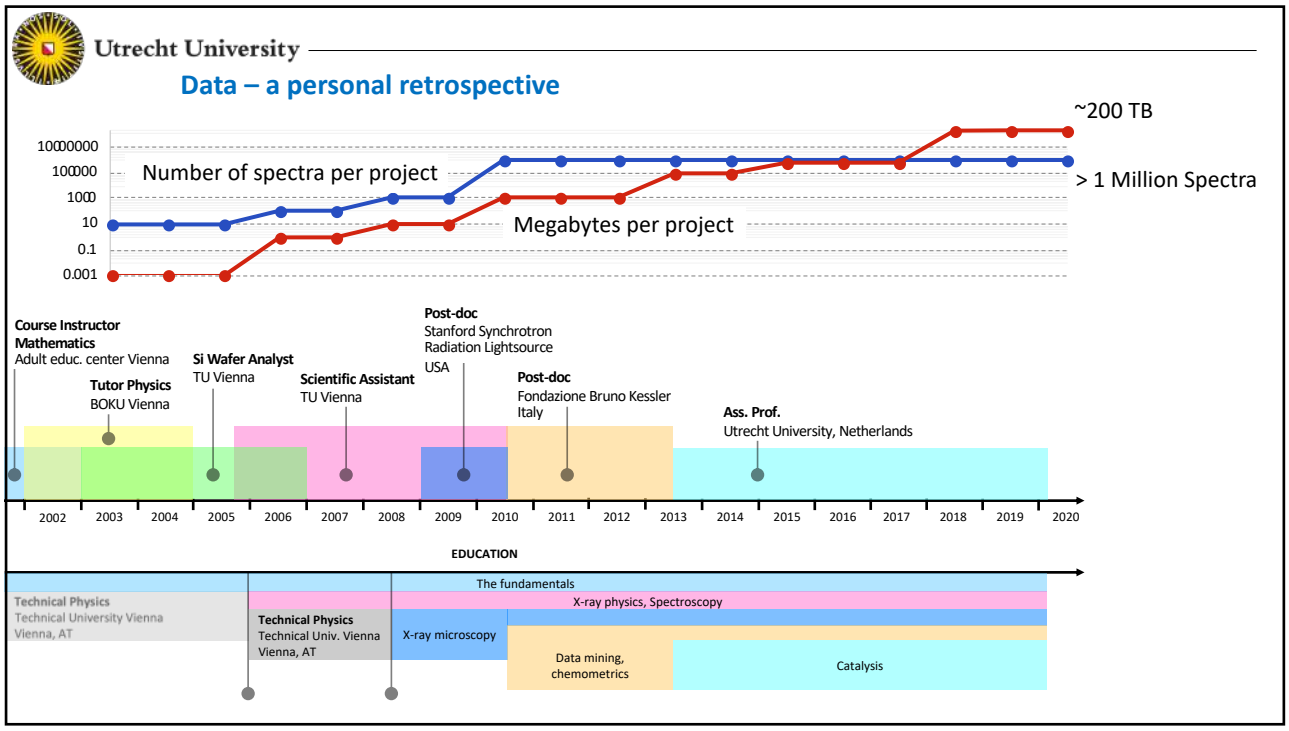
Utrecht University

Why would we need to teach about AI/ML/Data in Chemical Engineering Education?

3


Utrecht University
Data – a personal retrospective
Course Instructor
Mathematics
Adult educ. center Vienna
Tutor Physics
BOKU Vienna
Si Wafer Analyst
TU Vienna
Scientific Assistant
TU Vienna
Post-doc
Stanford Synchrotron
Radiation Lightsource
USA
Post-doc
Fondazione Bruno Kessler
Italy
Ass. Prof.
Utrecht University, Netherlands
2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014 2015 2016 2017 2018 2019 2020
EDUCATION
Technical Physics
Technical University Vienna
Vienna, AT
Technical Physics
Technical Univ. Vienna
Vienna, AT
The fundamentals
X-ray physics, Spectroscopy
X-ray microscopy
Data mining,
chemometrics
Catalysis

4


Utrecht University
Data – a personal retrospective
~200 TB
> 1 Million Spectra
10000000
100000
1000
10
0.1
0.001
Number of spectra per project
Megabytes per project
Course Instructor
Mathematics
Adult educ. center Vienna
Tutor Physics
BOKU Vienna
Si Wafer Analyst
TU Vienna
Scientific Assistant
TU Vienna
Post-doc
Stanford Synchrotron
Radiation Lightsource
USA
Post-doc
Fondazione Bruno Kessler
Italy
Ass. Prof.
Utrecht University, Netherlands
2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014 2015 2016 2017 2018 2019 2020
EDUCATION
Technical Physics
Technical University Vienna
Vienna, AT
Technical Physics
Technical Univ. Vienna
Vienna, AT
The fundamentals
X-ray physics, Spectroscopy
X-ray microscopy
Data mining,
chemometrics
Catalysis

5

Utrecht University

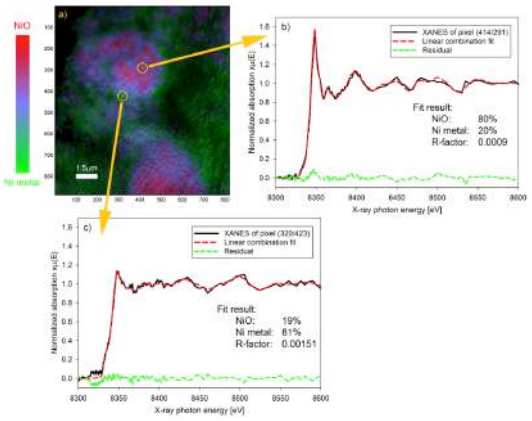
Data – a showcase

Chemical imaging is the analytical capability

-) to create a visual image of the distribution of the

-) components of a sample by

-) simultaneously collecting spectral and spatial (or time) information



Chemical imaging and the fundamental questions in science:

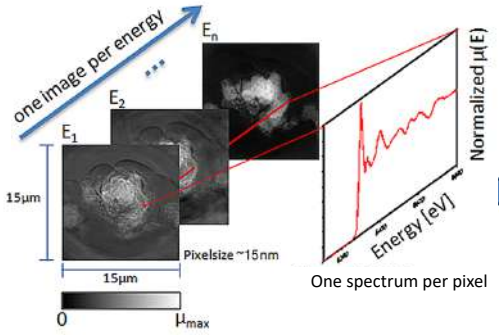
1. **What is there and where is it?**
2. How does it work?
3. How did it come to be this way?

6

Utrecht University

Data – a showcase

Chemical imaging

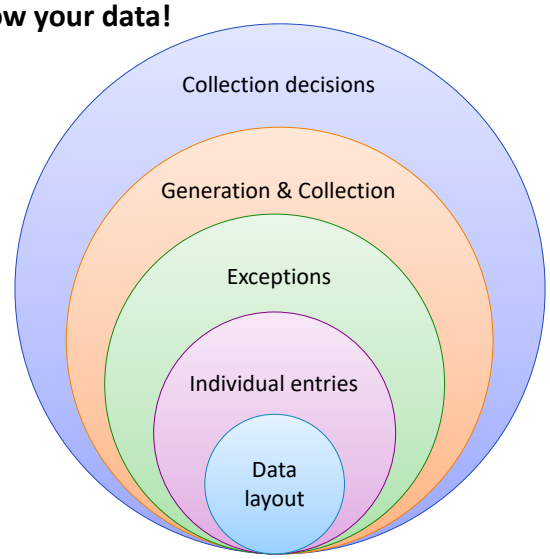


One spectrum per pixel

An image is a 1D data set, with the pixel position encoding the 2D information

⇒ for further data analysis this means we 'simply' have a list of spectra

Know your data!



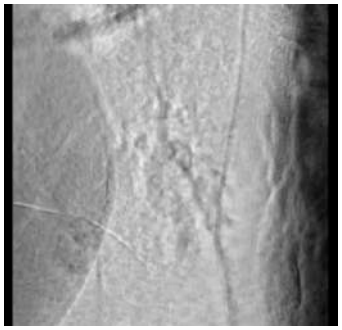
Levels of understanding data: the data knowledge onion

7

Utrecht University

Data – a showcase

146 images (905x927 pixels) recorded:
⇒ 146 x 838,935 pixels
⇒ 838,935 spectra
⇒ single scan: 20 minutes – 1 hour



Know your data!

Perform standard XANES data evaluation, i.e. a spectroscopist processing the XANES?

Spend 10 seconds per XANES spectrum => work for more than 97 days

⇒ need for a fast, efficient and reliable way to process the data

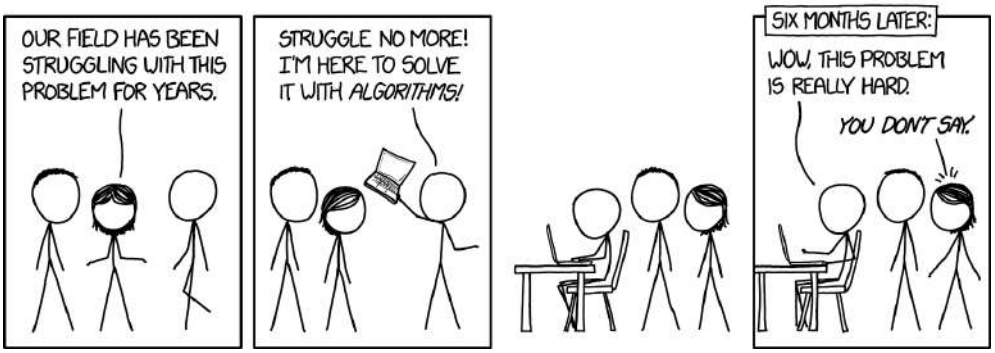
Development of software package to run TXM-XANES measurements and process datasets using the instrument: 'TXM-Wizard' [1] freely available for download at sourceforge.

[1] Y. Liu, F. Meirer et al., *J. Synchrotron Rad*. 19, 281–287 (2012).

8

Utrecht University

Data – a personal retrospective



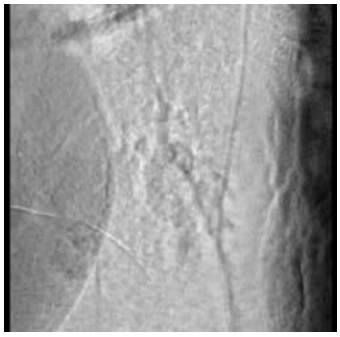
https://xkcd.com/1831/

9

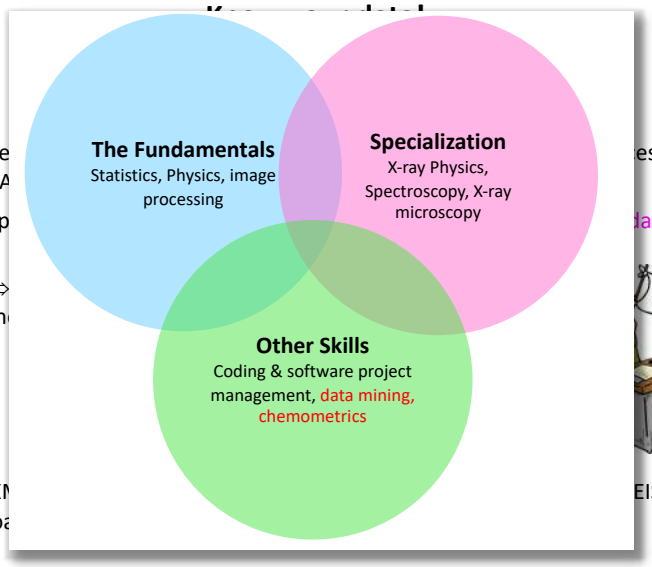
Utrecht University

Data – a showcase

146 images (905x927 pixels) recorded:
⇒ 146 x 838,935 pixels
⇒ 838,935 spectra
⇒ single scan: 20 minutes – 1 hour



Pe… essing the
XA…
Sp… days
⇒
th…



Development of software package to run TXM… EISS instrument:
'TXM-Wizard' [1] freely available for downloa…

[1] Y. Liu, F. Meirer et al., *J. Synchrotron Rad.* 19, 281–287 (2012).

10

Utrecht University

Data – a showcase

From Chemical imaging to Multivariate Analysis (MVA)

Object 1 (e.g. a voxel)	1	0.1	0.0	…
Object 2	2	0.2	0.3	…
.	3	0.5	0.2	…
.	4	1.3	0.0	…
.	…	…	…	…

Property 1 (e.g. position index) Property 2 . . .

e.g. one spectral axis

Every dimension ('a number') represents a property of the objects we study
⇒ we have more than one *dependent variable*
⇒ MVA:
the **statistical analysis** of data collected on more than one dependent variable

- Dimensionality reduction (PCA)
- Correlation analysis
- Clustering
- Classification
- Regression analysis
- Hypothesis testing
- Data mining
- …

If you know your math (lin. algebra, statistics) this is a piece of cake!

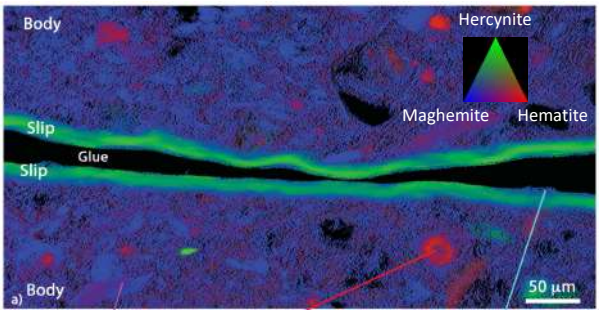
11

Utrecht University

Data – a showcase

From Chemical imaging to Multivariate Analysis (MVA)

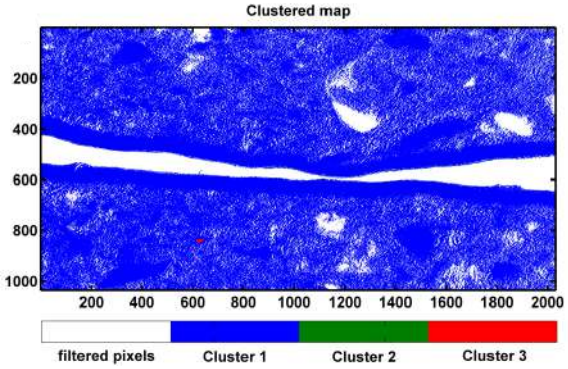
Example Clustering:
Finding the unexpected wodden needle in the haystack



F. Meirer et al., J. *Anal. At. Spectrom.* 28, 1870–1883 (2013).

Finding **~400 pixels** with a different (unexpected) spectrum within more than **2.1 million**

=> Detecting an unexpected phase that occupies less than 0.02% of the total volume of the sample.

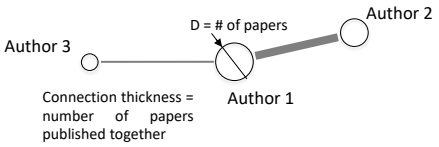


12

Utrecht University

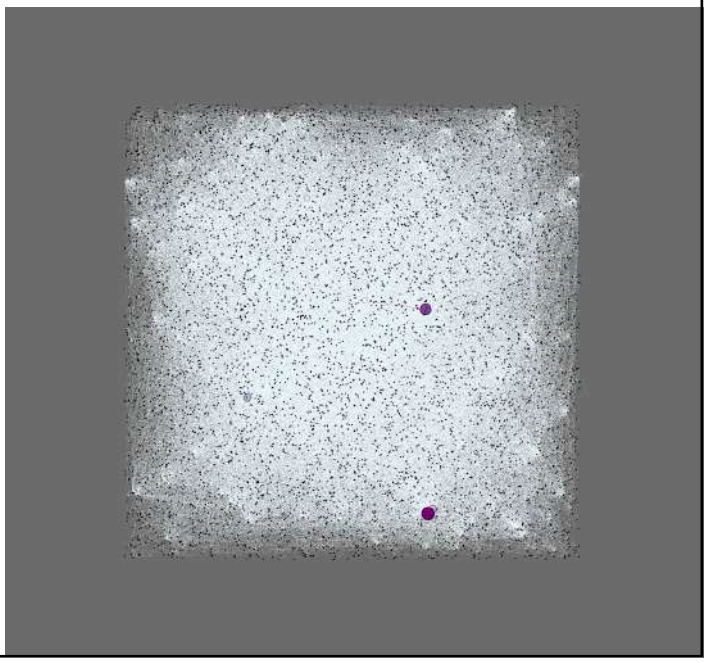
MVA - OK, but we want to use it…

A hobby:
Clustering in bibliometric studies



68 'key papers' generate a social network based upon 7081 papers with 6191 authors and 34720 connections between authors via citations

Data visualization!



Layout:
ForceAtlas2, Jacomy M., Venturini T., Heymann S., Bastian M., PLoS ONE 9(6): e98679 (2014).

13

Utrecht University

Unsupervised learning

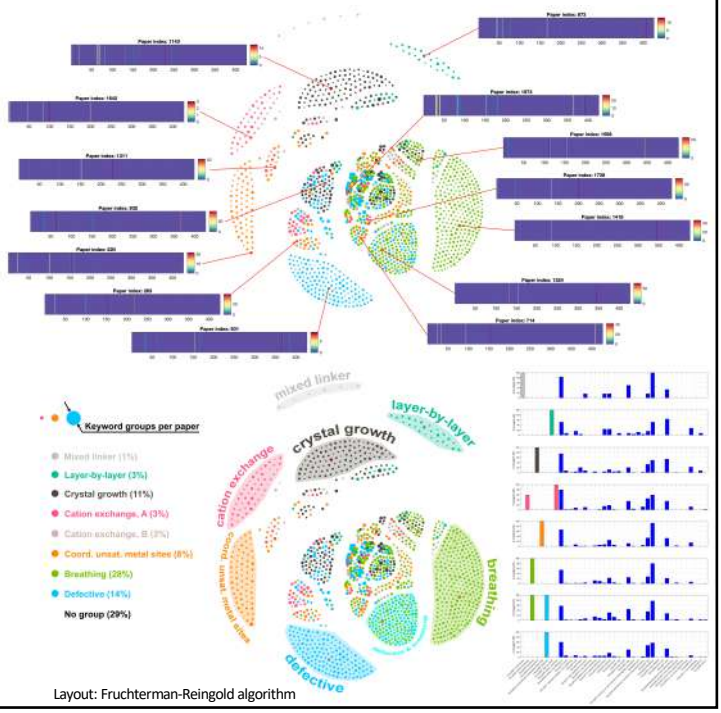
Example:
Bibliometric studies – a full text search correlating analysis techniques and material properties mentioned in the full text of each paper.

2016 papers were scanned for the occurrence of 426 keywords pooled into 65 groups

Groups 1-8 indicated phenomena studied

The 'properties' of these groups were then analyzed, e.g. what analytical methods had been used.

This is considered the first of three classes of ML: unsupervised learning



Miguel Rivera-Torrente et al., in preparation 2020

Layout: Fruchterman-Reingold algorithm

14

Utrecht University

… almost there

So…?

Why (the other 2 classes of) machine learning?

Consider using machine learning when you have a complex task or problem involving a large amount of data and lots of variables, but no existing formula or equation. For example, machine learning is a good option if you need to handle situations like these:

S. Raschka, Python Machine Learning, 1st ed. Packt Publishing (2015)

15

Utrecht University

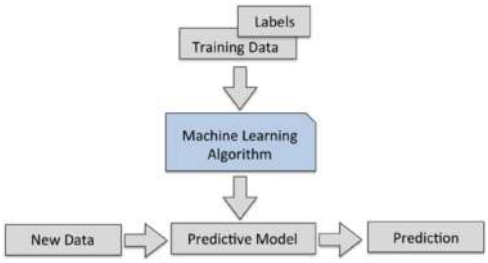
Machine learning 1-2-3

S. Raschka, Python Machine Learning, 1st ed. Packt Publishing (2015)

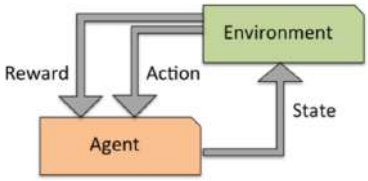
The other two classes: supervised and reinforced learning

Supervised learning: classification for 'predicting (*)' class labels
Handwritten character recognition:
Training data: handwritten example of each character of the alphabet
New data: user provides handwritten character
Prediction: which letter of the alphabet it is

Question: Can this handwriting algorithm predict a handwritten number?



Reinforced learning: the goal is to develop a system (*agent*) that improves its performance based on interactions with the *environment*.
Example: A chess engine:
the agent decides upon a series of moves depending on the state of the board (the environment), and the reward can be defined as *win* or *lose* at the end of the game.



(*) the word 'predicting' can be misleading; sometimes is means predicting future outcome (e.g. probability of sale), sometimes is means classifying after the event

16

Utrecht University

Artificial intelligence

Larry Tesler's Theorem (ca. 1970):

"Intelligence is whatever machines haven't done yet."

commonly quoted as (coined by D. Hofstadter [1]):

"AI is whatever hasn't been done yet."

Aim:
emulate the human behaviour

Source: http://xkcd.com/329/

[1] Hofstadter, Douglas (1980), Gödel, Escher, Bach: an Eternal Golden Braid, page 601

17

Utrecht University

Data informed decision making

intelligent Decision Support Systems (iDSS)

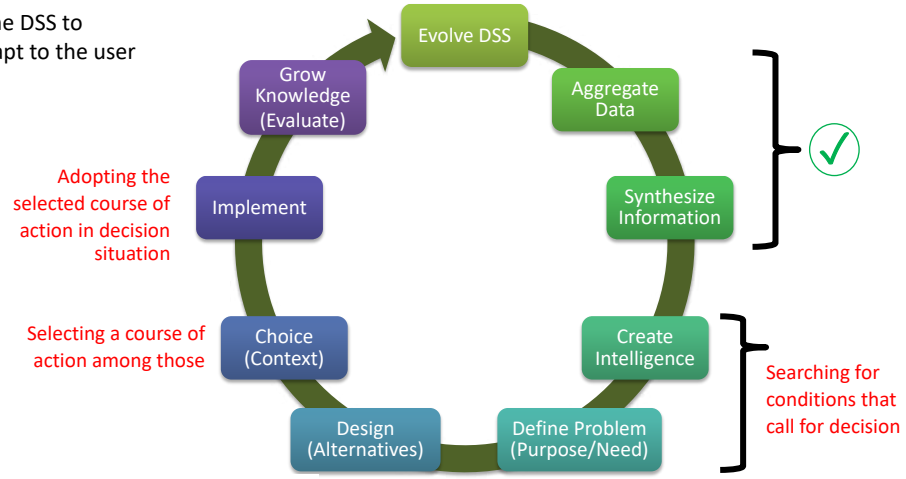
Here machine learning allows the DSS to obtain new knowledge or to adapt to the user or changing environment.

It is **unlikely** that ML will be used in the actual decision making <-> bias

Problems:
-) requires vast knowledge of the field… -> "knowledge acquisition problem"

-) Overfitting and over-generalization

Aim: emulate the human decision-making process

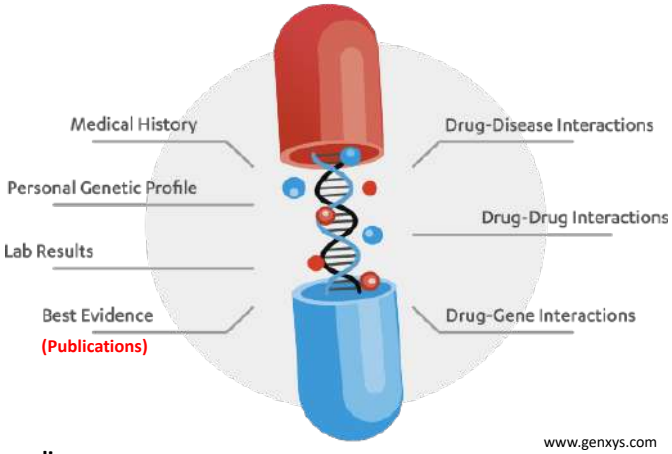


18

Utrecht University

Data informed decision making

Example:
Clinical Decision Support Systems (CDSS)



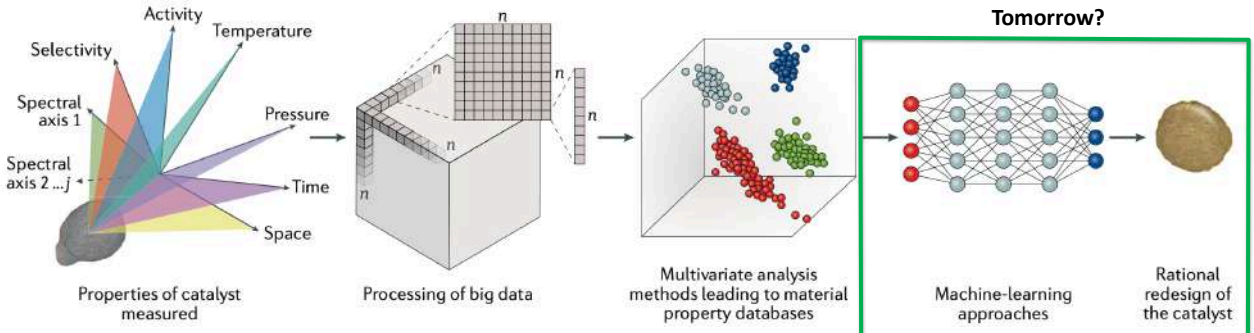
"Decision support systems are used **to help to analyse, diagnose and prescribe** based on patient data and current evidence. Typically, **a CDSS makes suggestions for the clinician** to look through, and the clinician is expected to select useful information from the results."

19

Utrecht University

Data informed decision making

In catalysis?



high-throughput analysis to establish material property databases
=> to allow predictive approaches via machine learning that allows the prediction of material properties at the inception of the design
=> the rational design of tomorrow's catalysts

F. Meirer, B.M. Weckhuysen, *Nat. Rev. Mat.* 3, 324 (2018).

20

Utrecht University

Why would we need to teach about AI/ML/Data in Chemical Engineering Education?

How do we teach this?

21

Utrecht University

It is already being taught… by the experts

AI Education

AI TUDelft

At the moment, Artificial Intelligence education is mainly embedded within the MSc programme Computer Science (in particular the Data Science & Technology track).

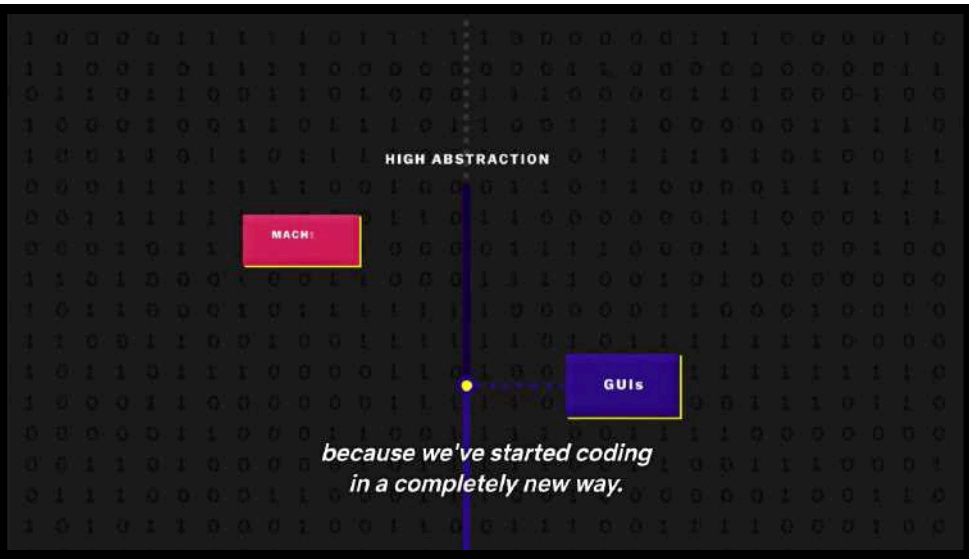
At the Bachelor level, several AI-related courses are offered in the BSc programme Computer Science & Engineering.

Finally, current TU Delft BSc students can also consider the Minor Robotics.

22

Utrecht University

CODING



201907_EMEA Culture Explained, Category Documentary, Netflix Premiere date: May 23rd 2018 9:00 AM CEST
Permission for educational screenings granted

23

Utrecht University

CODING

Source: https://xkcd.com/1838/

"… [ML] is a whole different kind of abstraction and way of doing things [in coding] …"

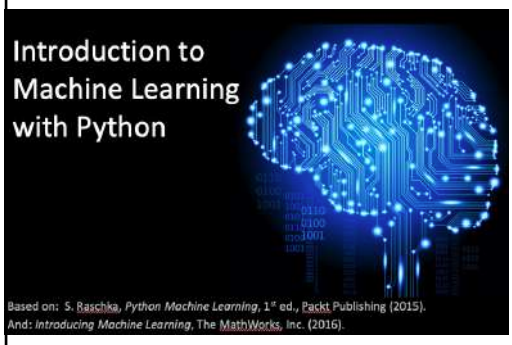
"… we should absolutely expand who gets to be in that design room [..] they will ask better questions … "

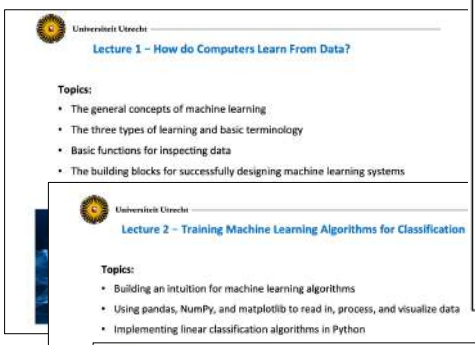
"… without people actively correcting for that [bias] historical data will lead us to repeat the mistakes of the past"

24

Utrecht University

My own attempt …




Universiteit Utrecht

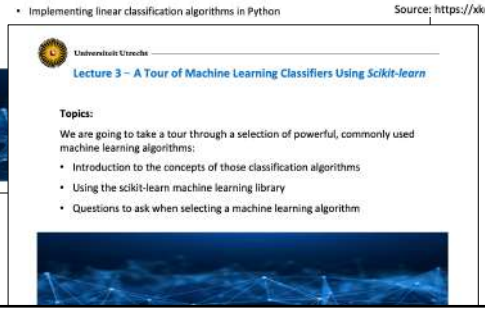
Lecture 1 – How do Computers Learn From Data?

Topics:
- The general concepts of machine learning
- The three types of learning and basic terminology
- Basic functions for inspecting data
- The building blocks for successfully designing machine learning systems

Universiteit Utrecht

Lecture 2 – Training Machine Learning Algorithms for Classification

Topics:
- Building an intuition for machine learning algorithms
- Using pandas, NumPy, and matplotlib to read in, process, and visualize data
- Implementing linear classification algorithms in Python


Universiteit Utrecht

Lecture 3 – A Tour of Machine Learning Classifiers Using *Scikit-learn*

Topics:

We are going to take a tour through a selection of powerful, commonly used machine learning algorithms:
- Introduction to the concepts of those classification algorithms
- Using the scikit-learn machine learning library
- Questions to ask when selecting a machine learning algorithm

Source: https://xkcd.com/1838/

A course for how to use Python, with a module on introducing the basics of ML

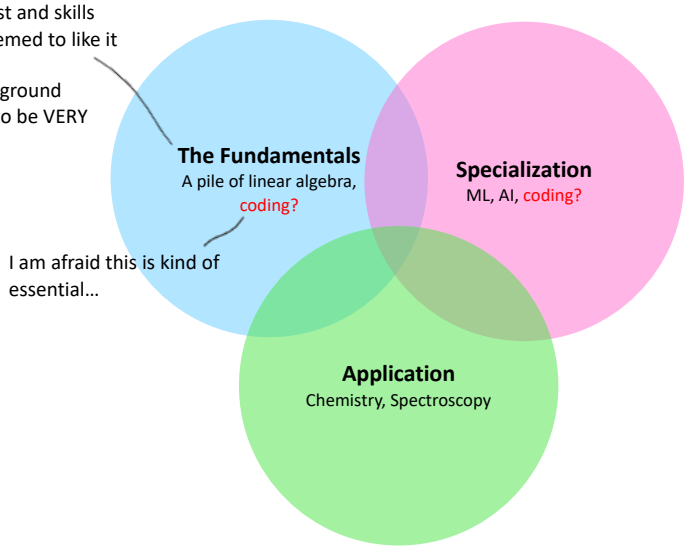
25

Utrecht University

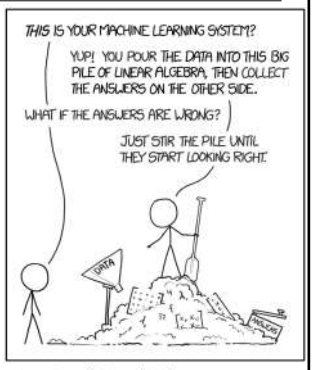
… also taught me something.

Students with interest and skills here did well and seemed to like it

The variation in background knowledge seemed to be VERY large



I am afraid this is kind of essential…



Source: https://xkcd.com/1838/

26

Utrecht University

So if we want the future generation to use AI/ML in chemical engineering…

Do we (chemists) all need to become Computer scientists?

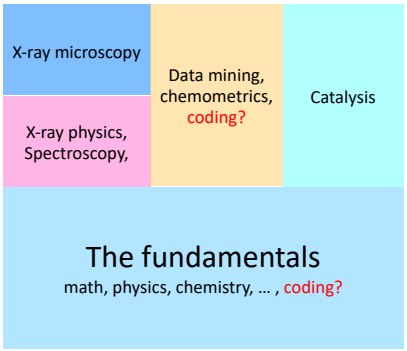
Would be nice, but certainly not.

Do we all need to learn how to code?

Probably yes …

Do we all need sufficient knowledge in linear algebra and statistics?

Certainly yes!



27