Machine Learning Applications

TECHNISCHE UNIVERSITÄT DARMSTADT

Winter semester 2019/2020

Prof. Dr.-Ing. Uwe Klingauf

Organizational Information

Patrons of MLA @ TU Darmstadt…and a lot more!

Prof. Dr.-Ing. Uwe Klingauf
Institute of Flight Systems and Automatic Control

klingauf@fsr.tu-darmstadt.de

Prof. Dr.-Ing. Dipl.-Wirtsch.-Ing. Joachim Metternich
Institute of Production Management, Technology and Machine Tools

j.metternich@ptw.tu-darmstadt.de

Prof. Dr. Kristian Kersting
Computer Science Department and Centre for Cognitive Science

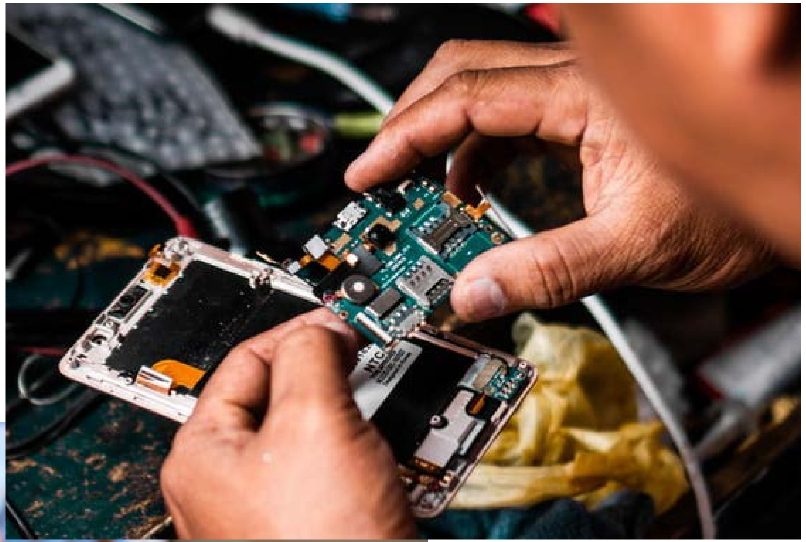
kersting@cs.tu-darmstadt.de

Prof. Dr.-Ing. Matthias Weigold
Institute of Production Management, Technology and Machine Tools

m.weigold@ptw.tu-darmstadt.de

Course elements

Lecture



Use Cases

Group Work
(„Hackathon“)

Written exam

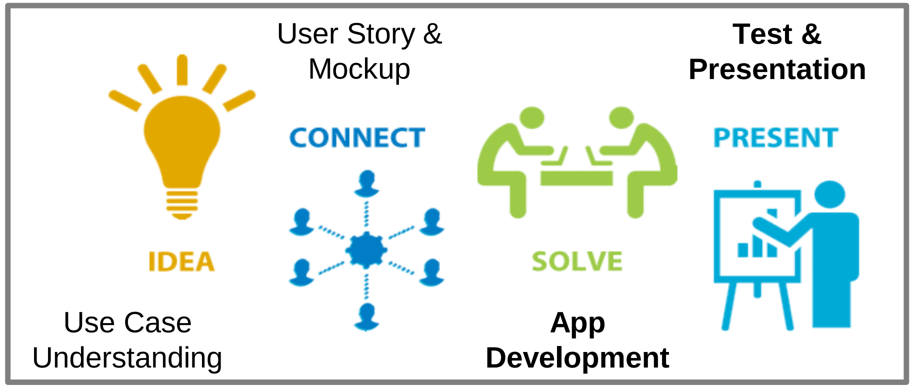
Picture source: freepik.com, unsplash.com

Structure and content of the lecture

VW	Date	Content	Lead
1	18th Oct 2019	Introduction, Motivation and Organizational Information	FSR
2	25th Oct 2019	Introduction to Machine Learning I	CS
3	1st Nov 2019	Introduction to Machine Learning II	CS
4	8th Nov 2019	Introduction to Machine Learning III	CS
5	15th Nov 2019	Introduction to and Application of Machine Learning	CS
6	22nd Nov 2019	Presentation of Industrial Partner and Introduction of Practical Exam (Hackathon)	FSR/EXT
7	29th Nov 2019	**Data Understanding:** Data Acquisition, Preprocessing and Feature Engineering	FSR
8	6th Dec 2019	**Diagnostics:** Feature Engineering, Regression, Health Assessment	FSR
9	13th Dec 2019	**Prognostics:** Prediction, Remaining Useful Life, Uncertainty	FSR
10	20th Dec 2019	**System of Systems:** System Level, Systemwide Predictions, Decision Support, IoT	FSR
11	17th Jan 2020	**Predictive Maintenance:** Band Saw Application, Neural Networks for Condition Monitoring and Prediction	PTW
12	24th Jan 2020	**Predictive Quality:** Prediction of Product Quality during Machining Process	PTW
13	31st Jan 2020	**Energy Forecasting:** Electric Load Forecasting by the Example of a Machine Tool	PTW
14	7th Feb 2020	**Operational Control:** Optimized Control of Cross-Linked Energy Systems by Means of Reinforcement Learning	PTW
15	14th Feb 2020	**Recap on Lecture for Written Exam**	FSR

Information on the exam

- Written exam on 17th February 2020
  - Covering the content of the lectures
- Group work from 2nd December 2019 till 2nd March 2020
  - Industry related task on data from a partner
  - Deliverable: Documentation, Presentation, Code
  - MATLAB or Python knowledge required



! Limited slots for the exam !

➔ Read "Hinweise zur Prüfung" document now available in Moodle!

Picture source: billardarchitectureinc.com

Contact for organizational matters

Katja Möller
Institute of Flight Systems and Automatic Control
Student Affairs

L1|01 590
+49 6151 16-21044
moeller@fsr.tu-darmstadt.de

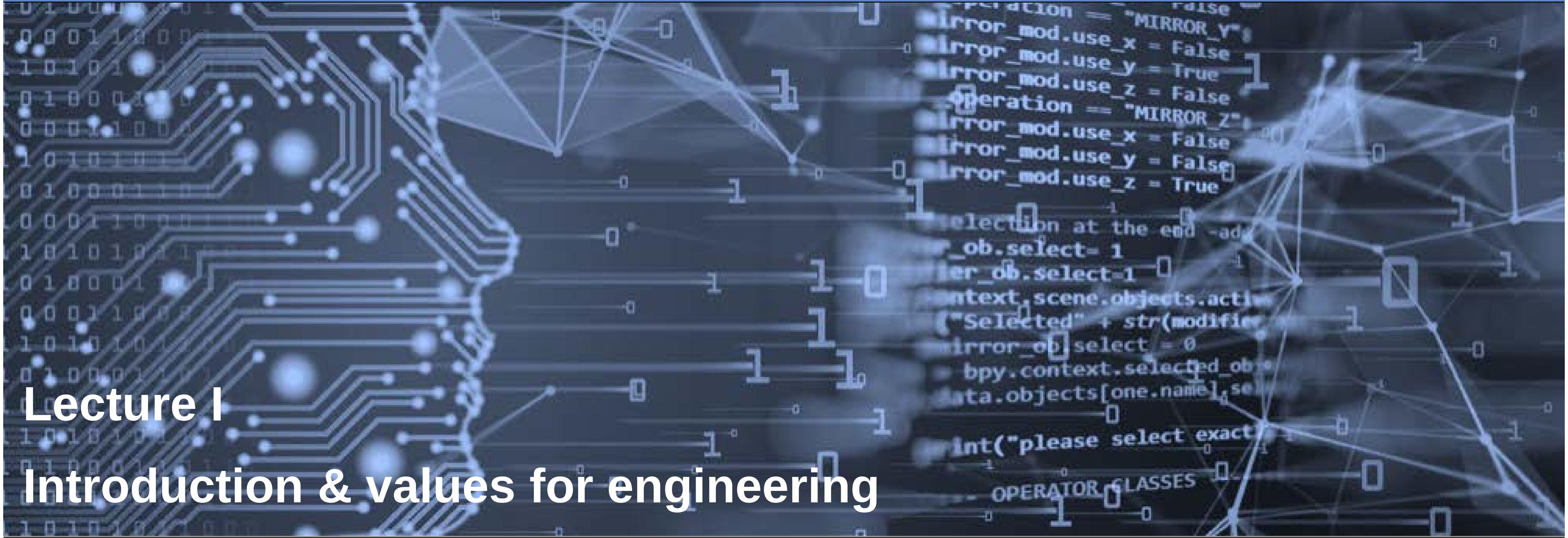
Simon Mehringskötter, M.Sc.
Institute of Flight Systems and Automatic Control
Research Associate

L1|01 573
+49 6151 16-21067
mehringskoetter@fsr.tu-darmstadt.de

Machine Learning Applications

Winter semester 2019/2020

Prof. Dr.-Ing. Uwe Klingauf



Lecture I

Introduction & values for engineering

What should you be able to take out of the lecture today?

- A simple overview and explanation what machine learning is
- Why you should deal with machine learning as an engineer
- Insights of suitable machine learning applications for engineering
- Barrier and limits to data-based developments (new business models, new digital products and services)
- Two process models for the approach to data science and data mining projects

Simple explanations, tasks and types

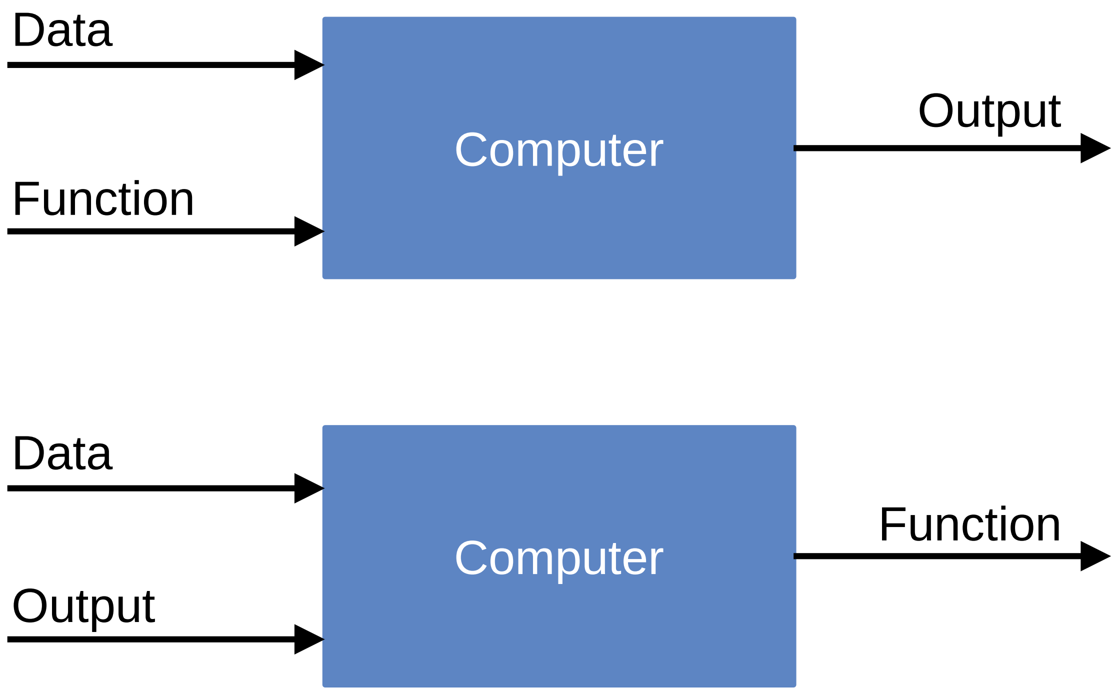
WHAT IS MACHINE LEARNING?

What is machine learning?

First, lets gain simple insights.

Traditional programming and data analysis

Machine Learning

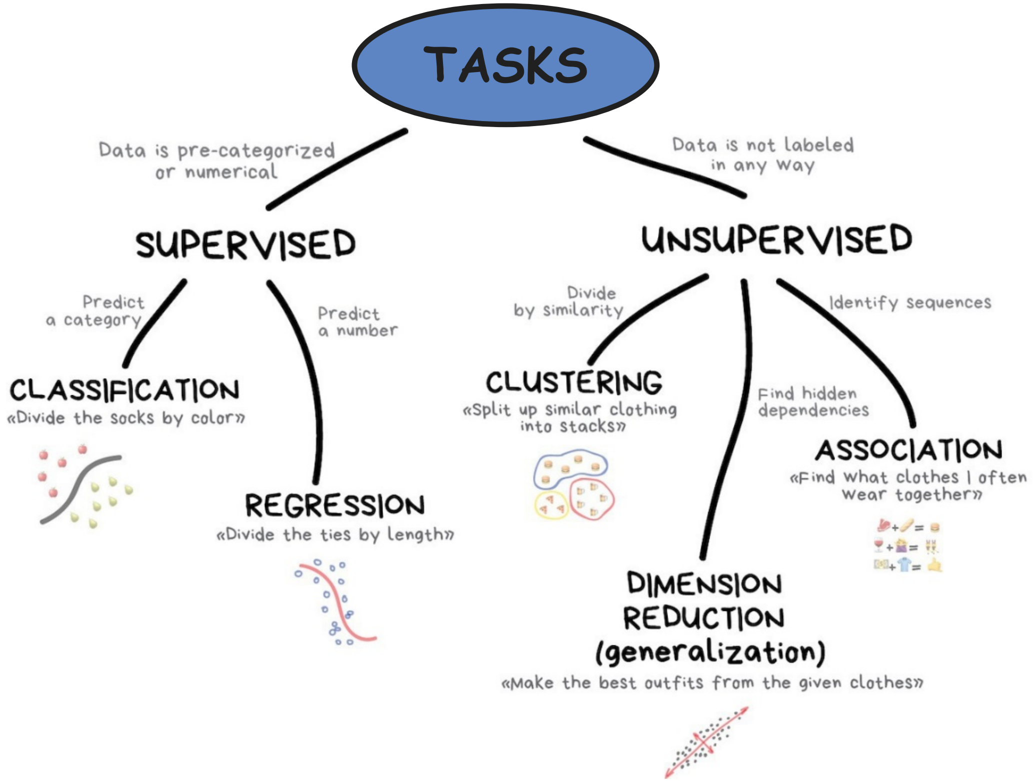


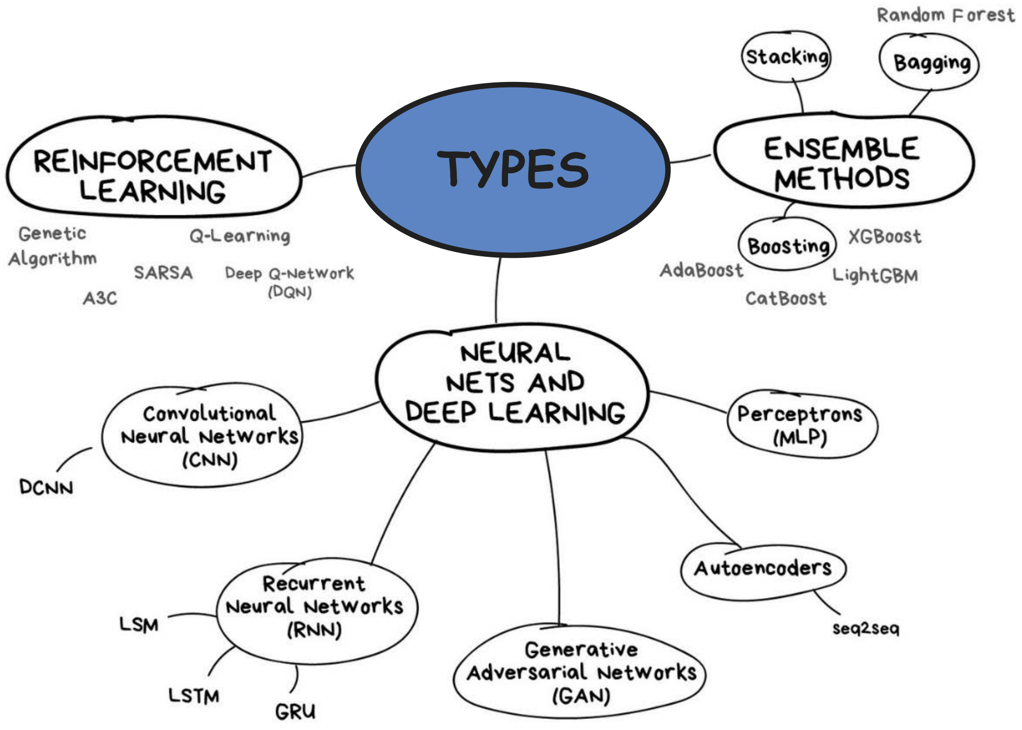
Let's look at a brief explanation from University of Oxford: https://www.youtube.com/watch?v=f_uwKZIAeM0

What is machine learning? A part of artificial intelligence!

It generalizes the experience so that a task performance can be improved.

Definition by Tom M. Mitchel (1997): "A computer program is said to learn from experience *E* with respect to some class of tasks *T* and performance measure *P* if its performance at tasks in *T*, as measured by *P*, improves with experience *E*.



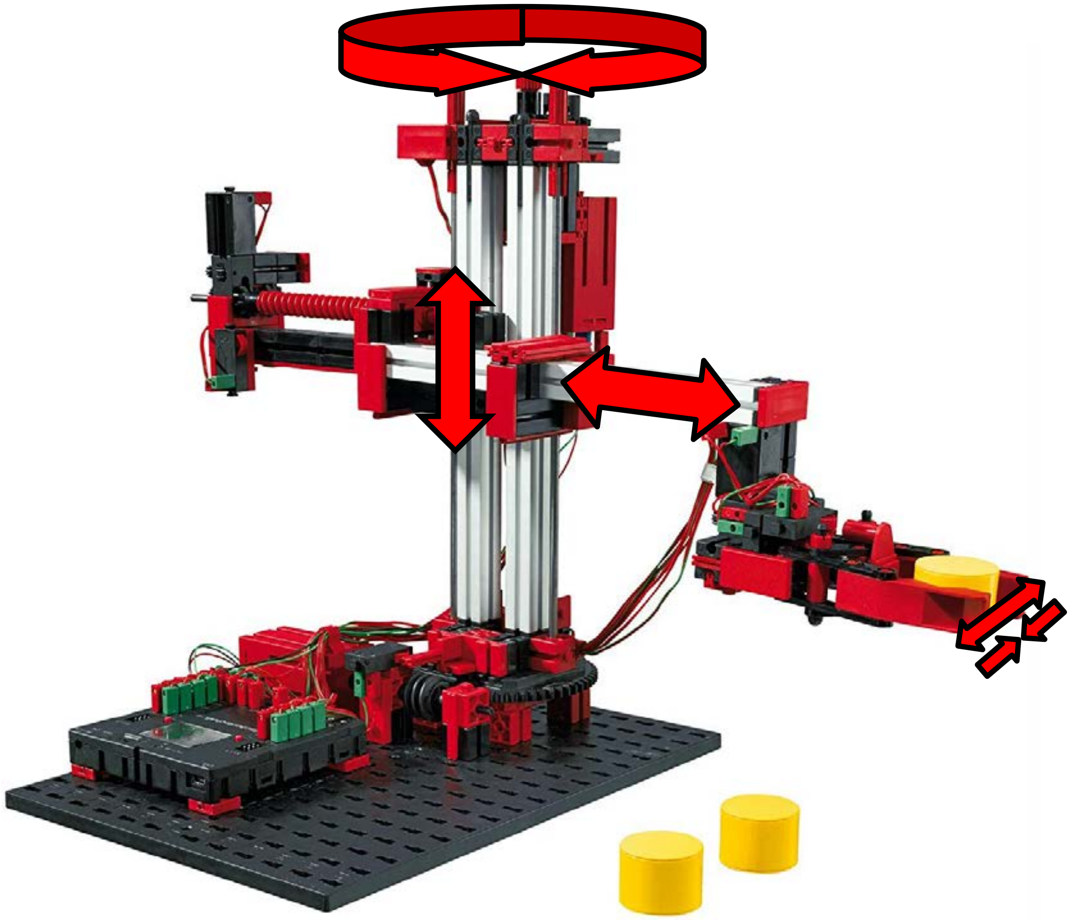


Picture source: in accordance with http://vas3k.com/blog/machine_learning/

Making automation more intelligent

EXAMPLE: AUTOMATIC HANDLING DEVICE

Setup of Automatic Handling Device



- 4-DoF robotic arm that moves boxes
- No integrated sensors → no process information

→ **Does the robotic arm move a heavy or a light container?**

- Custom retrofit of robotic arm
- Low-cost sensor and microcontroller (~20 €)
- Raw data send via Wi-Fi

→ **Automatic data-driven classification in MATLAB**

ESP 32
240 MHz
520 KB RAM
12 bit ADC
GPIO pins
Wi-Fi, BT

MPU-9250
3-axis Gyro
3-axis Accelerometer
3-axis Magnetometer

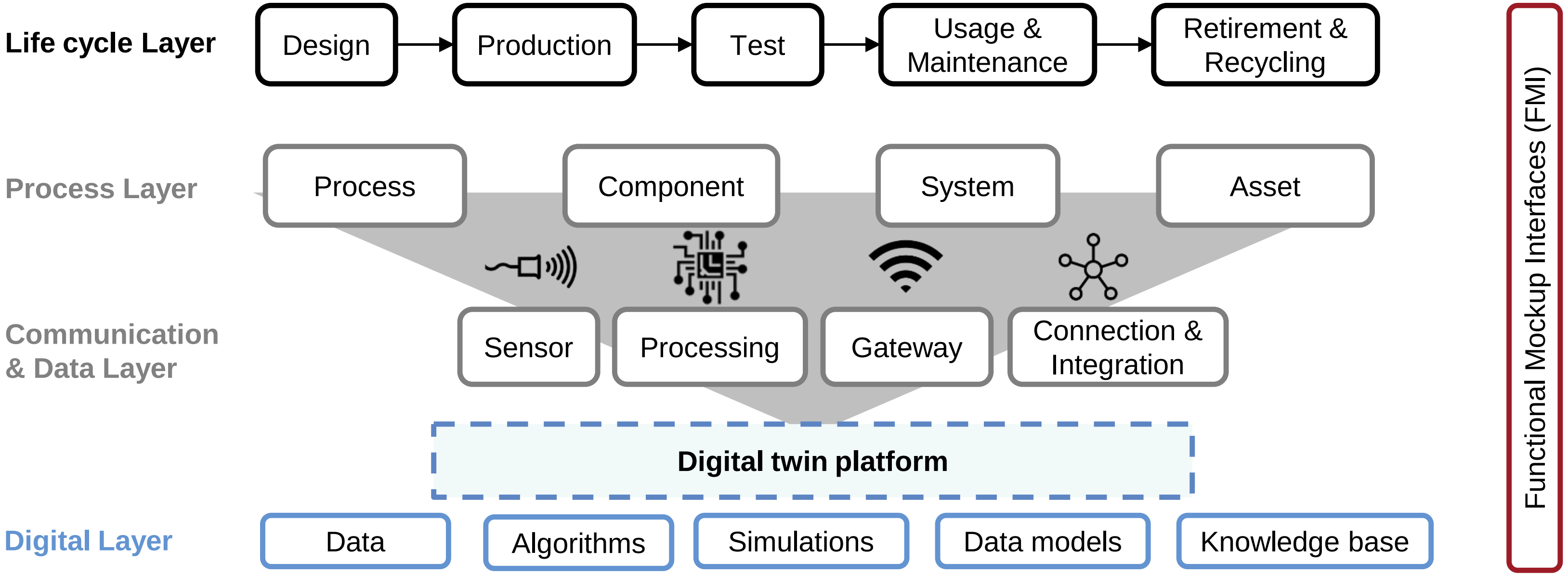
Dealing with resources: data is the new oil!

WHY IS IT IMPORTANT?

An example: Digital twins gain in importance

Definition and components of digital twins

- "Digital Twins are digital replications of living as well as nonliving entities that enable data to be seamlessly transmitted between the physical and virtual worlds." *El Saddik, University of Ottawa*
- „The concept is based on modeling assets with all their geometrical data, kinematic functionality and logical behavior using digital tools." *Dr. Sauer, Fraunhofer IOSB*



Digital twin is one of the key factors for industry 4.0

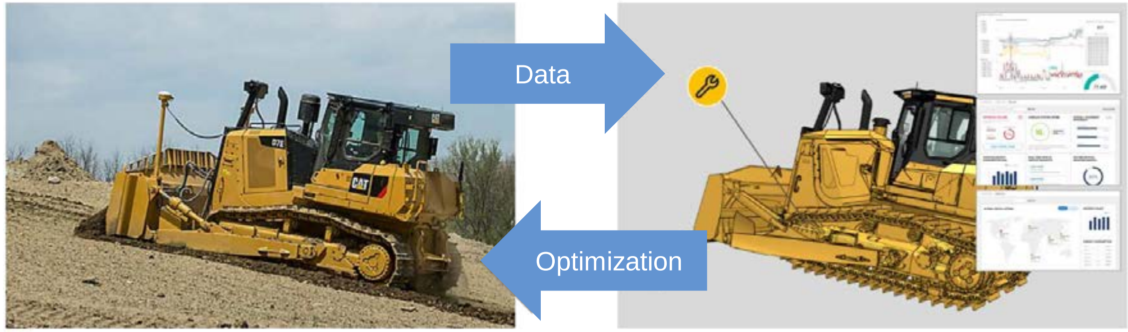
Answer to time-to-market and design-to-cost constraints

- Improve simulation and validation capabilities:

1. Higher accuracy in modelling, faster creation and solve of models
2. Assessments between simulation and physical system in operation
3. Cope with uncertainty of data, improvement of measuring systems on the physical system

- Providing decision support and alerts to users
- Discovering new application opportunities and revenue streams

Physical asset operation in the field   Virtual asset in digital space



Picture source: cat.com   Source: Microsoft: *The promise of a digital twin strategy* (2017)

Tesla Model S

Picture source: Tesla

CityAirbus - Urban Air Mobility

Picture Source: Airbus

Business model change: new digital products & services

Perfomance based contracting – a win-win situation!?

TECHNISCHE UNIVERSITÄT DARMSTADT

TotalCare – „Power by the hour“ business model by Rolls Royce

Advantages for the customer:

- Reliability of the engines is rewarded
- Financial risks are reduced and operating costs can be planned
- Engine availability is increased
- Improvement measures are carried out automatically

Advantages for the supplier:

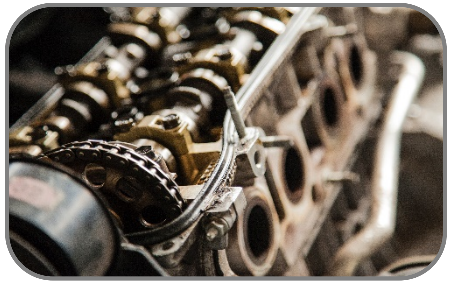
- The company can bypass aircraft manufacturers and establish direct business relationships with airlines
- Increase in sales
- The revenue stream has become more steady and is less susceptible to cyclical fluctuations

Source: http://www.zephram.de/blog/geschaeftsmodellinnovation/beispiel-servitization/

Why is it the “hot topic“ for engineering?

Job descriptions of engineers are changing, companies want all-rounders

Technical domain



Engineers

Data experts

Mediator:
Programming / Computer Science

Related job offers:

- ASAP: Ingenieur/Techniker | Datenanalyse und Testautomatisierung (m/w/d)
- accenture: Data Scientist/Consultant (all genders) Industrial and Operational Analytics
- DB: Ingenieur als Referent Fahrzeuganalysen Data Analytics (w/m/d)
- itk ENGINEERING: Data Engineer
- BMW GROUP: Software Quality Engineer - Autonomous Driving (f/m/x)
- BAYER: Validation Expert in DataScience/Industry 4.0 (m/w/d)

Foremost profile requests:
mechanical / industrial engineers

Source: stepstone.de

Digitalisation: An opportunity or a risk?

BARRIERS AND RISKS OF DIGITISATION

Challenges that come with machine learning

Data Policy, Actors and Stakeholders

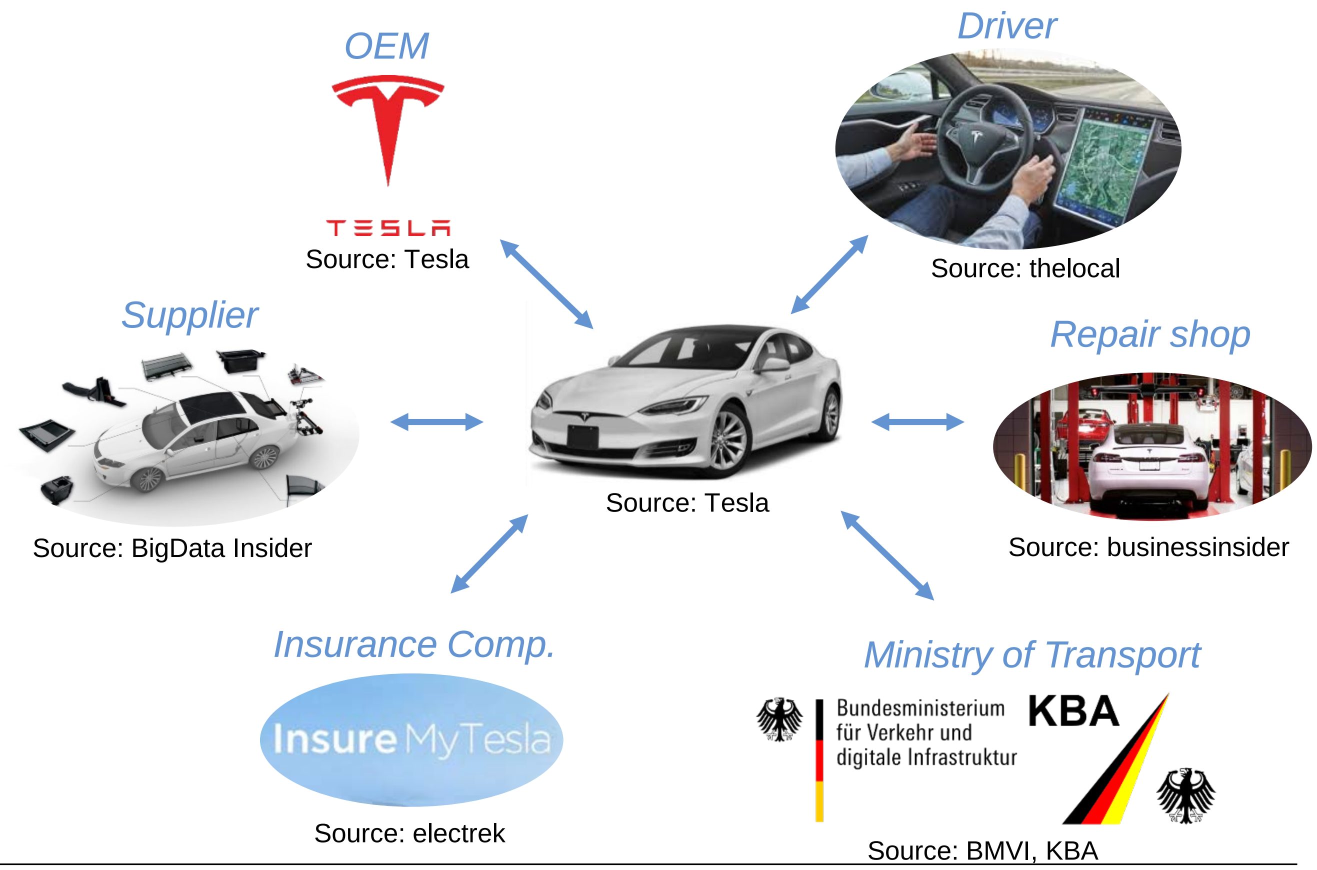
TECHNISCHE UNIVERSITÄT DARMSTADT

Key questions:

- Who owns the data?
- What data is appropriate to leverage for business use?
- What data can a company share?
- With whom can the company share data?

➔ Each stakeholder has own interests on data

Tesla‘s full self-driving concept:



Challenges that come with machine learning

Who owns the data? An example of aircraft data ownership in Germany

Pilot

Source: AirlineGeeks

Repair shop

Lufthansa Technik

Source: Lufthansa Technik

Airlines

Source: which.co.uk

OEM

AIRBUS

Source: Airbus

- Airlines own the agregated operating and repair data
- Airlines have an interest in preventing disclosure of airline-specific data that could harm their competitive position

- OEMs own the design and test data
- Operational and maintenance data is usually the most valuable for predictive maintenance

Source: https://www.mro-network.com/technology/aircraft-data-ownership-debate-ongoing

Challenges that come with machine learning

What are the risks?

- **Bias** e.g. developer tendencies
- **Ethic** principles
- Wrong **correlations**
- Contaminated or insufficient reference **data**
- Feedback Loops
- „Dual-use“ – different intensions

  e.g. Google‘s Project Maven

Picture source: Suas News

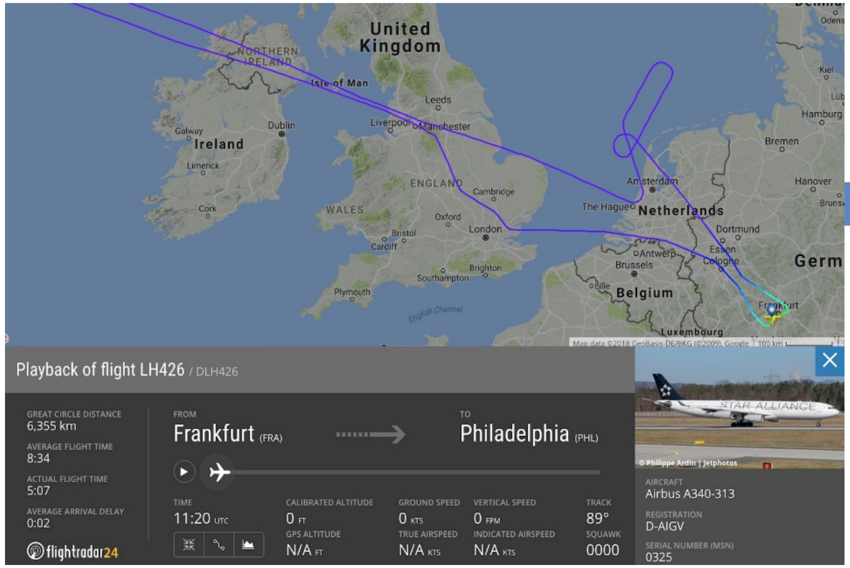
Another example

WHEN DOES THAT MAKE SENSE?

Example from aviation: Downtime costs money!

Unplanned outages represent significant financial risks for airlines.

TECHNISCHE UNIVERSITÄT DARMSTADT

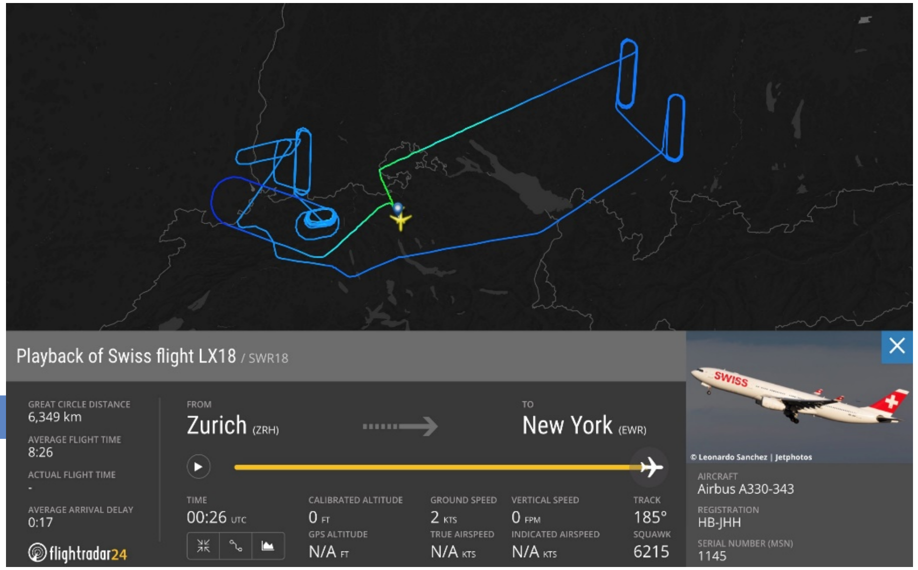


Impacts of technical issues

Lufthansa flight returns to Frankfurt but held briefly over the North Sea to further reduce landing weight

Swiss LX18 burning fuel in various holding patterns to reduce landing weight prior to return to Zurich

Source: https://www.flightradar24.com/blog



The loss account

Major airline
> 1 million flights per year

Approx. 250 flight diversions per year
due to technical malfunctions

Between 50K EUR and 500K EUR costs per incident

On average about 200K EUR costs per incident

Expenses: **20%** share of operational costs for maintenace, overhaul & repair
Daily business: **88%** planned events, **12%** unplanned events

Data-based PHM strategies are gaining popularity for replicating complex systems.

TECHNISCHE UNIVERSITÄT DARMSTADT

Physical / Analytical Models
(white-box)

Feature

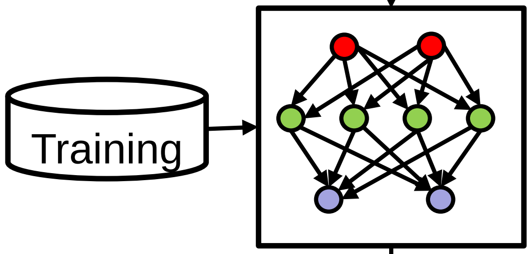
$$\frac{dN}{dA} = C\Delta K^m$$

Remaining Useful Lifetime (RUL)

\+ Traceability
\+ Stability verification
\- Deep domain knowledge necessary
\- High effort / high model complexity

Data-based / Iteratively learned models
(grey/black-box)

Feature

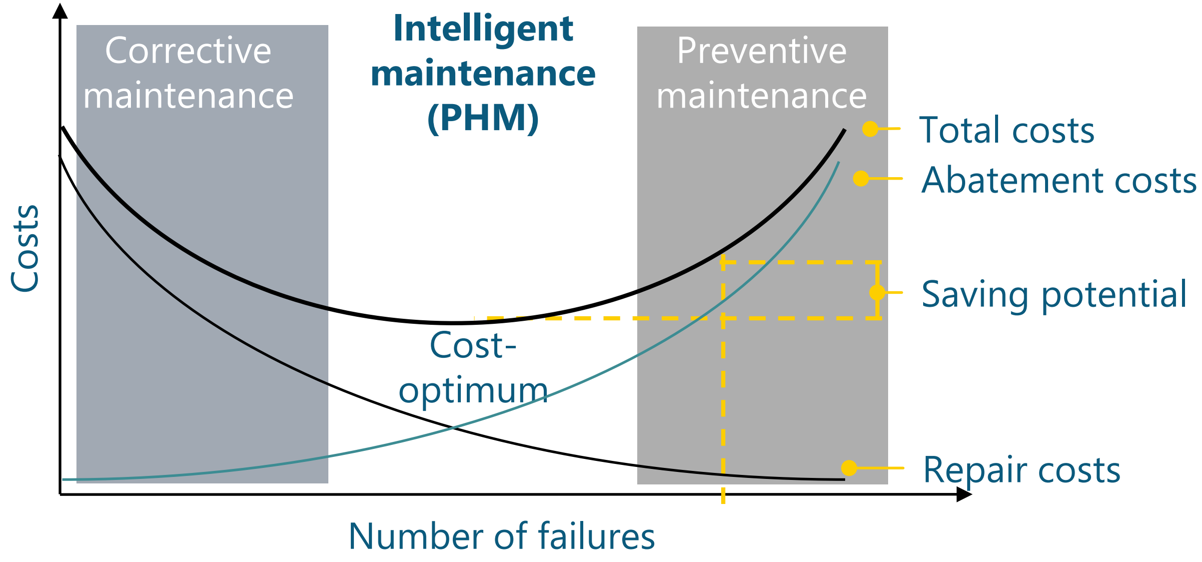


Remaining Useful Lifetime (RUL)

\+ Data availability increases through digitalisation
\+ Self-learning algorithms
\- Requires meaningful training data

Use Case: Prognosis of Bearing Damages

Predictive health management saves costs & increases availability



Data acquisition

↓

Data processing

↓

Condition determination

↓

Condition assessment

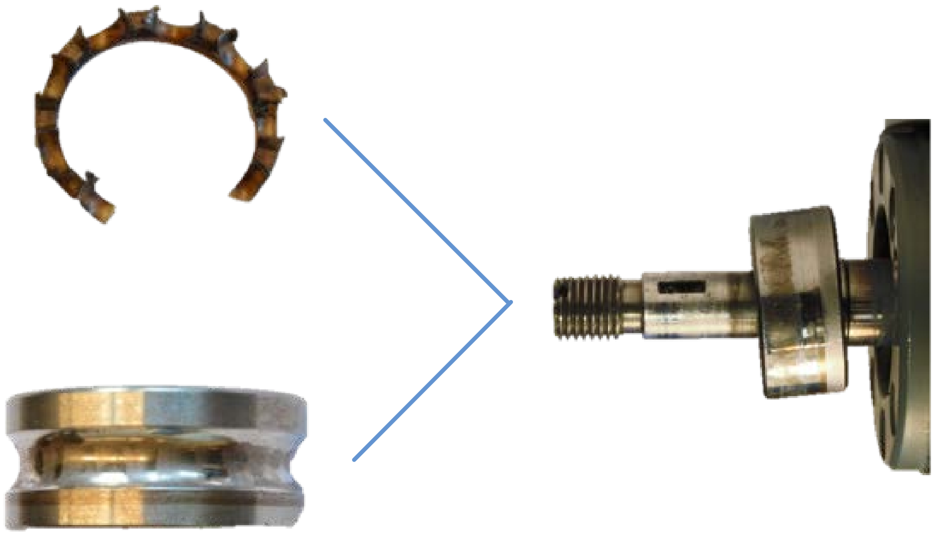
↓

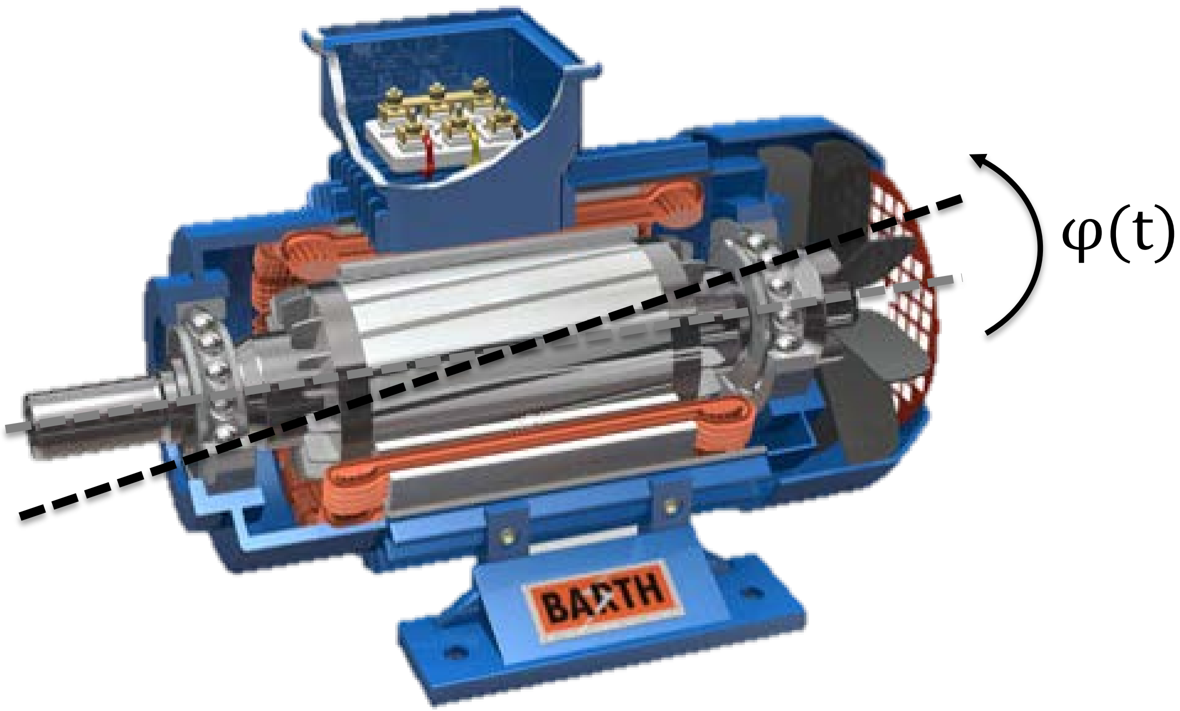
Condition prognosis

Bearing damage is the most common cause of failure with asynchronous motors, accounting for approx. 40%.

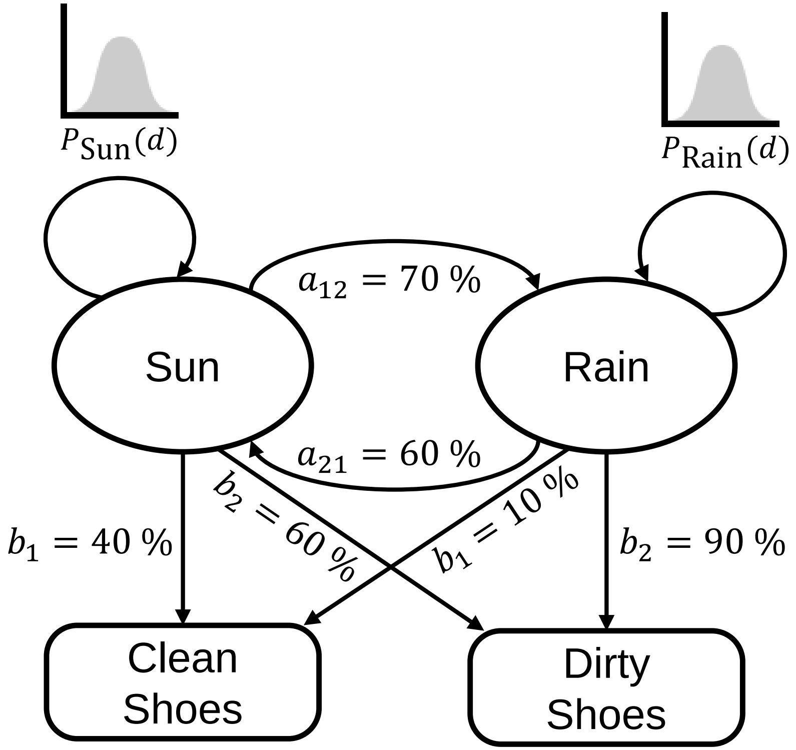
TECHNISCHE UNIVERSITÄT DARMSTADT

Damage → Shaft → Deviation → Rotor/Stator → Variable air gap → Stator current





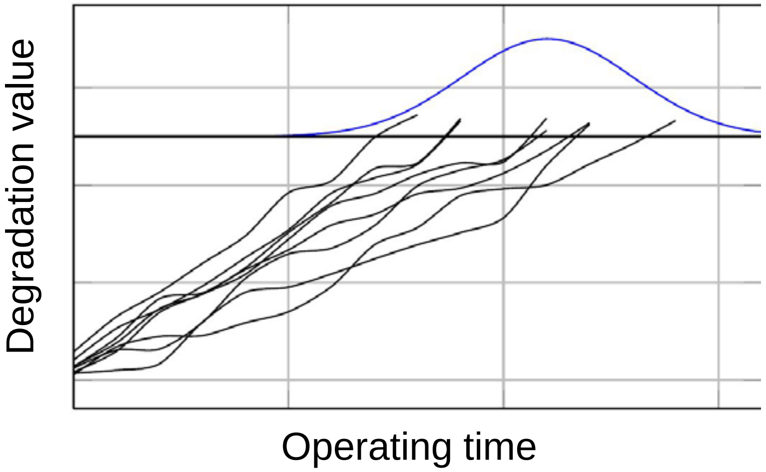
Methodology: Hidden-semi-Markov-Models (HsMM)

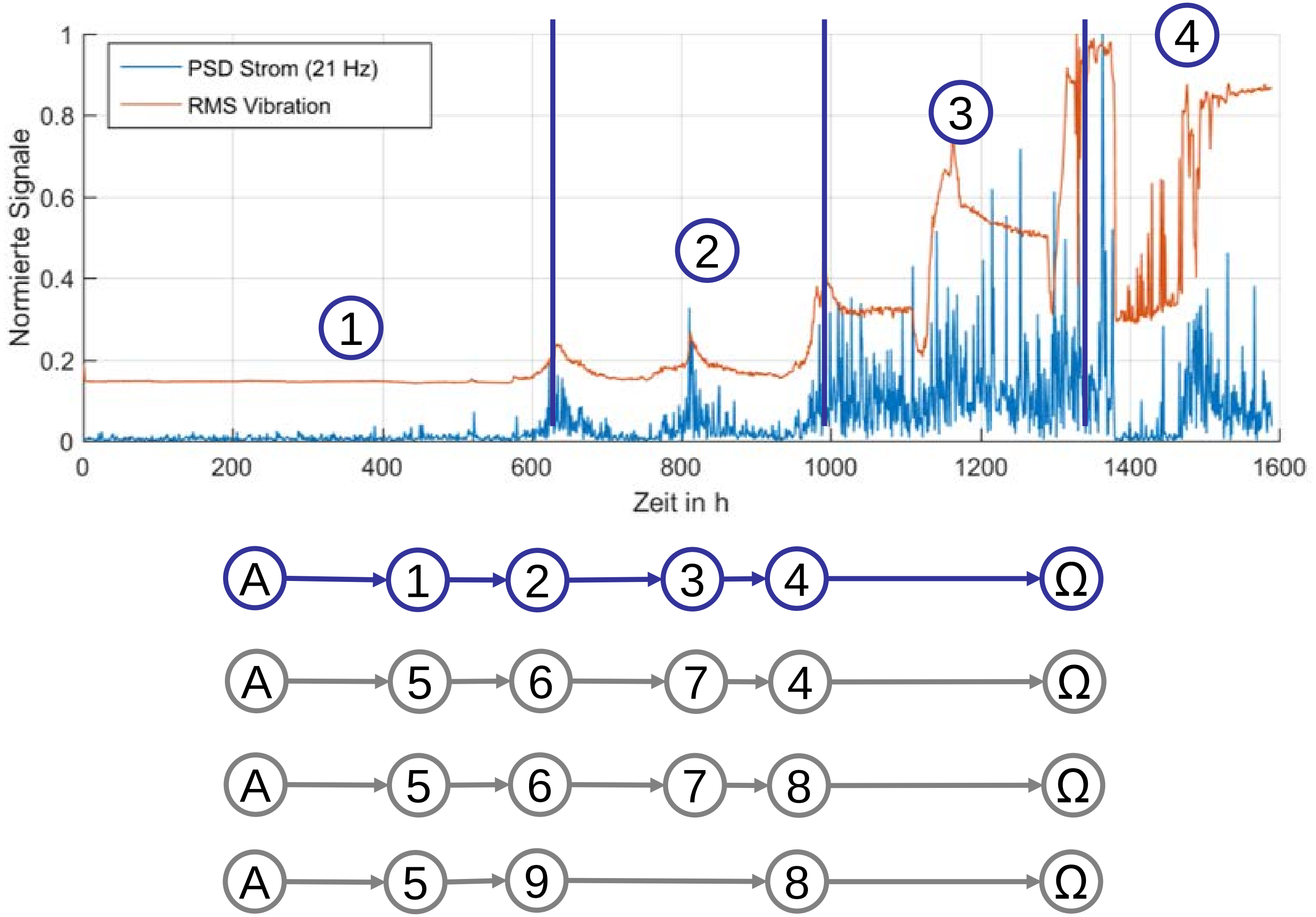


a_{ij}: state transition probabilities
b_n: observation probability
$P_i(d)$: state stay duration probability

- **MM:** fixed step time probability to change state (e.g. P = 70% that state *sun* transits to state *rain*)
- **HMM:** states are hidden – observation has probability measure for different states (e.g. guess from seeing dirty shoes that it's raining)
- **HsMM:** fixed step time is replaced by probability distribution to stay in a state (e.g. $P_{Sun}(d)$ distribution of time to stay in state *sun*)

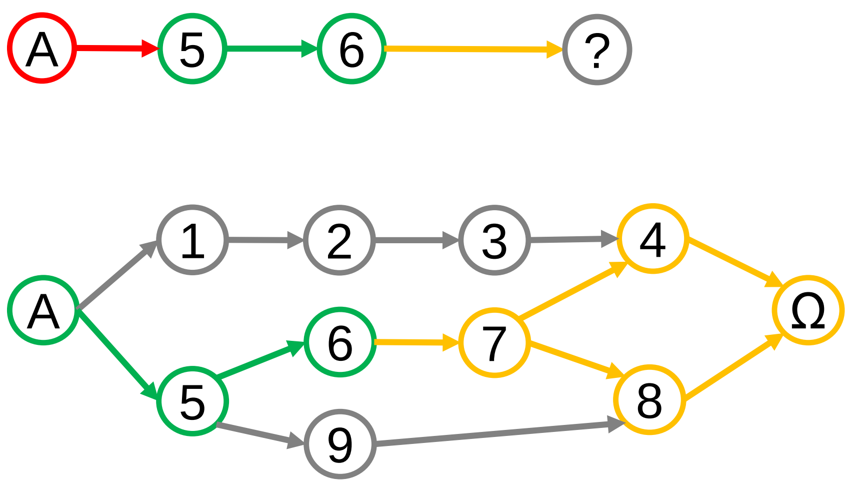
Data-based damage models are learned from the sensor data of the asynchronous motor.



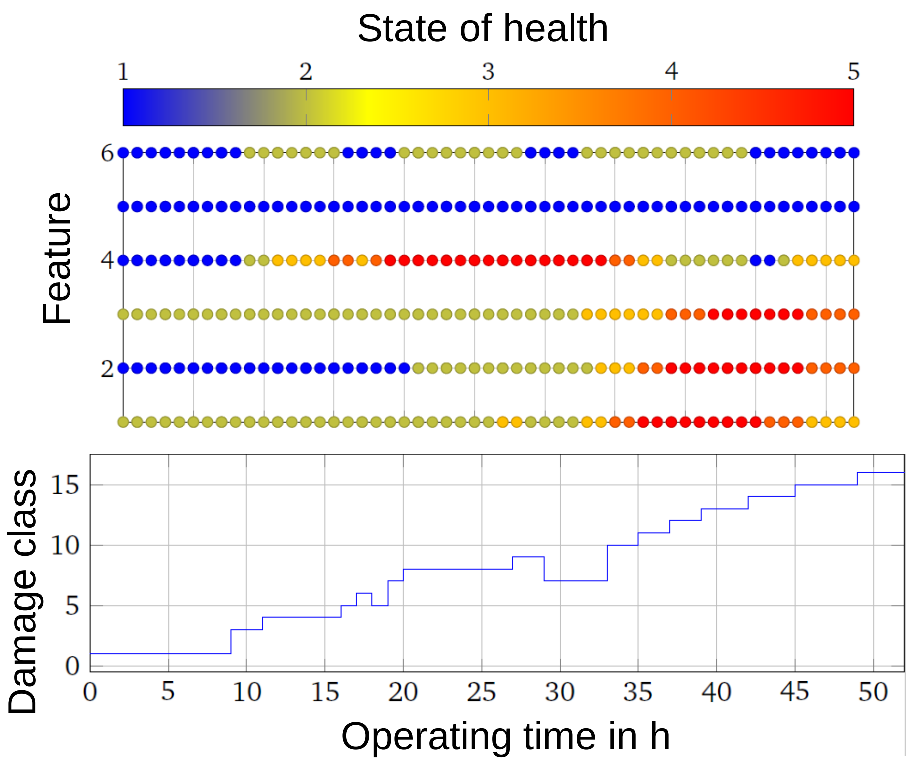


Markov models can be used to diagnose and predict the health status of the bearing.

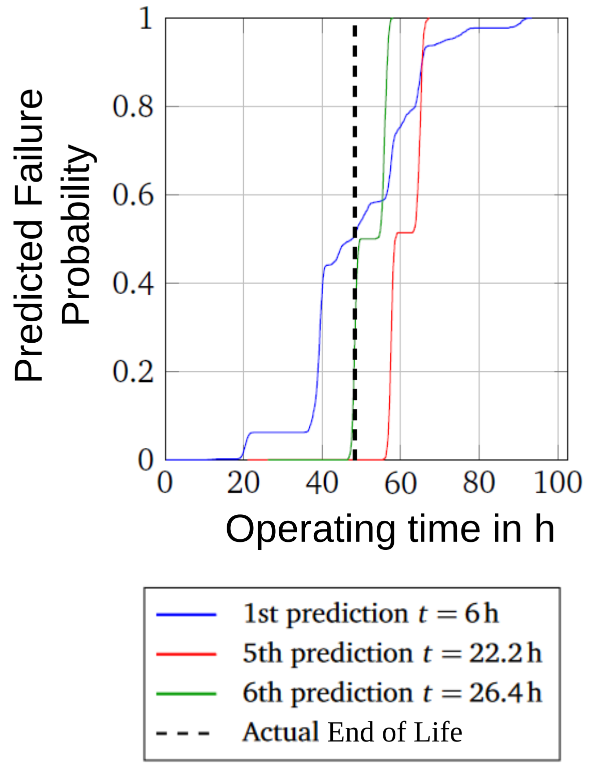
Modelling



Fault diagnosis



Fault prognosis



Field of application of machine learning

Machine learning can solve problems, but not all problems require learning!

TECHNISCHE UNIVERSITÄT DARMSTADT

Usefull field of application

- Speeding up the process of data modeling for data science
- Situations where there is a lot of ground truth data, but very little obvious correlation
- Problems where there are no human experts, so data cannot be labelled or categorised
- Problems with high complexity, rapidly changing rules and changing data
- Problems where there are human experts and it can be programmed, but where it is not cost effective to implement

Currently not applicable

- Solving unique problems for a particular business use case
- Cleaning the data in first place so that it is valuable in a machine learning workflow

↳ Area of application of machine learning must be **meaningfully** weighed up!

ARTIFICIAL INTELLIGENCE CANNOT DO EVERYTHING

Algorithms and iterative learning frameworks are universially applicable!





Pearson correlation coefficient

TECHNISCHE UNIVERSITÄT DARMSTADT

“Pearson’s correlation coefficient is a statistical measure of the strength of a linear relationship between paired data.”

$$r_{xy} = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2} \cdot \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

→ L = {-1; 1}

1 is total positive linear correlation
0 is no linear correlation
−1 is total negative linear correlation

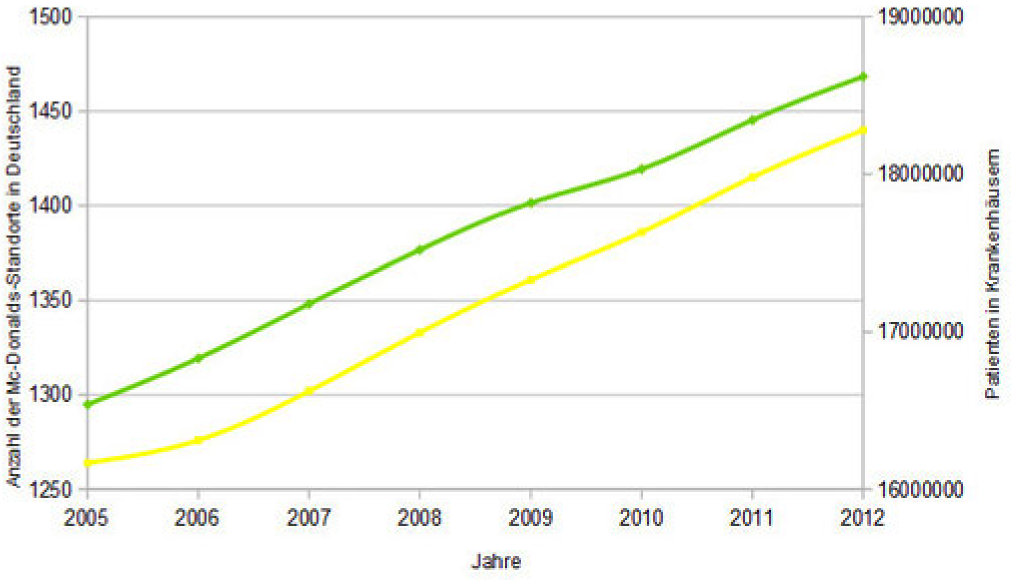
Spurious correlations

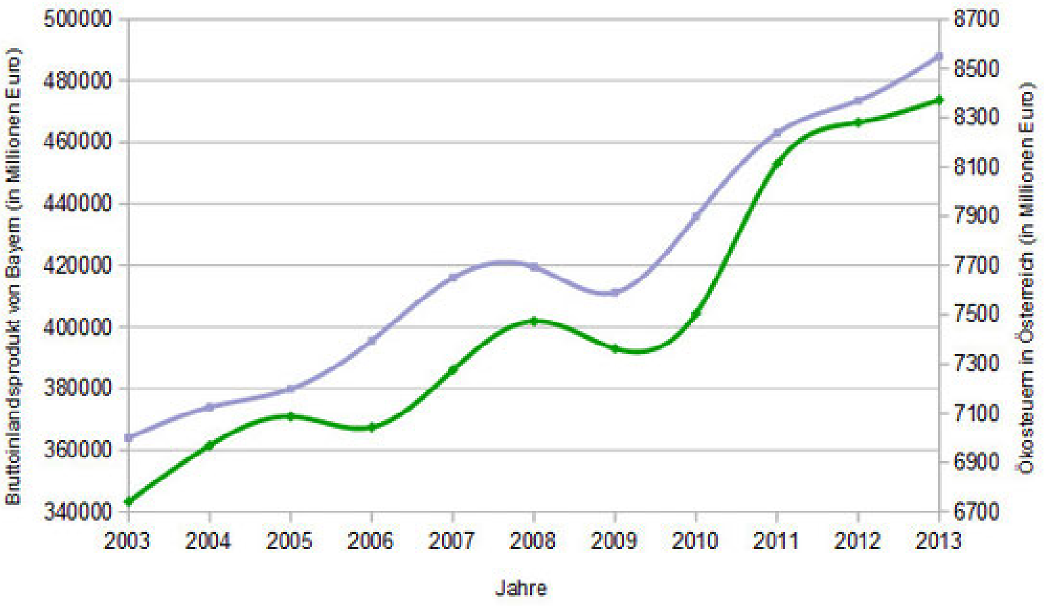
If a causal relation is not observed (1/2)

TECHNISCHE UNIVERSITÄT DARMSTADT

- Number of McDonalds sites in Germany and patients in hospitals in Germany
- Correlation: 0.9954

- Gross Domestic Product of Bavaria and Ecotaxes in Austria
- Correlation: 0.9817





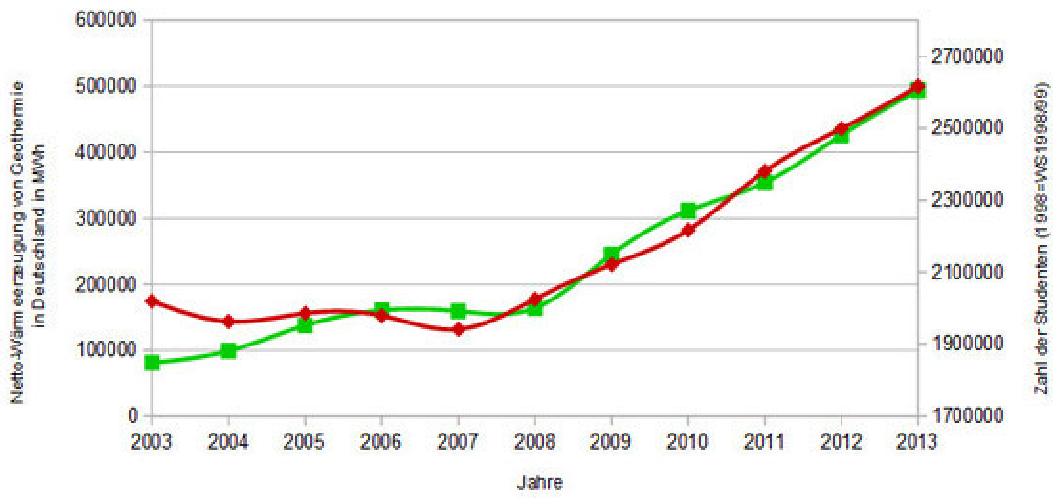
Source: in accordance to N. Zellmer, https://scheinkorrelation.jimdo.com/

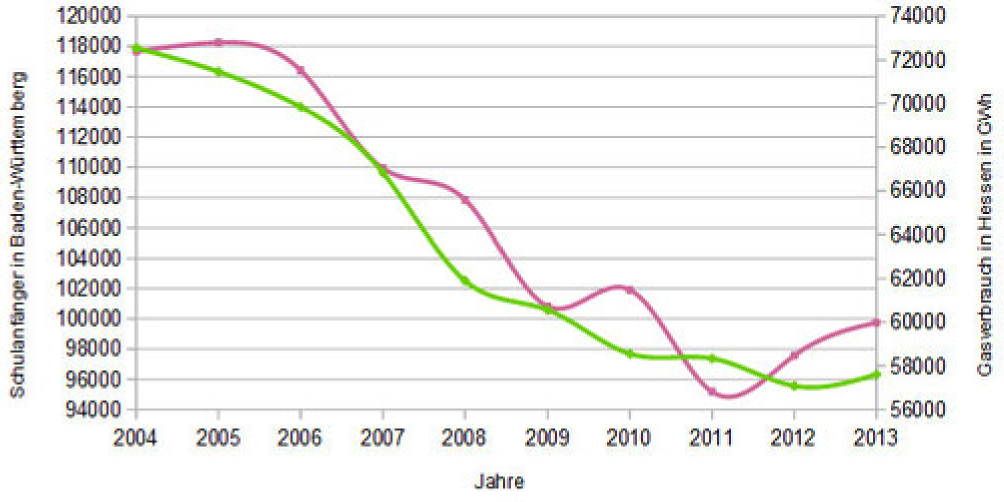
Spurious correlations

If a causal relation is not observed (2/2)

- Net heat generation from geothermal energy in Germany in MWh (green) and number of students (red)
- Correlation: 0.9682

- School beginners in Baden-Württemberg (green) and gas consumption in Hesse (pink)
- Correlation: 0.9665





Source: in accordance to N. Zellmer, https://scheinkorrelation.jimdo.com/

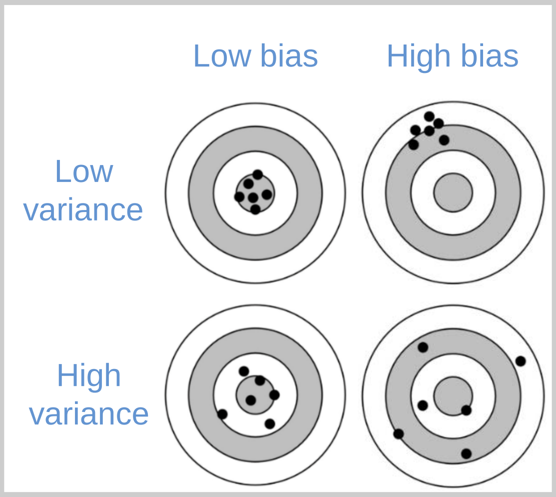
Bias-variance tradeoff

Decomposing errors in machine learning models

Reducible error:

- **Bias** error: occurring by erroneous assumptions in the underlying model
- **Variance** error: sensitivity to small fluctuations in the training set

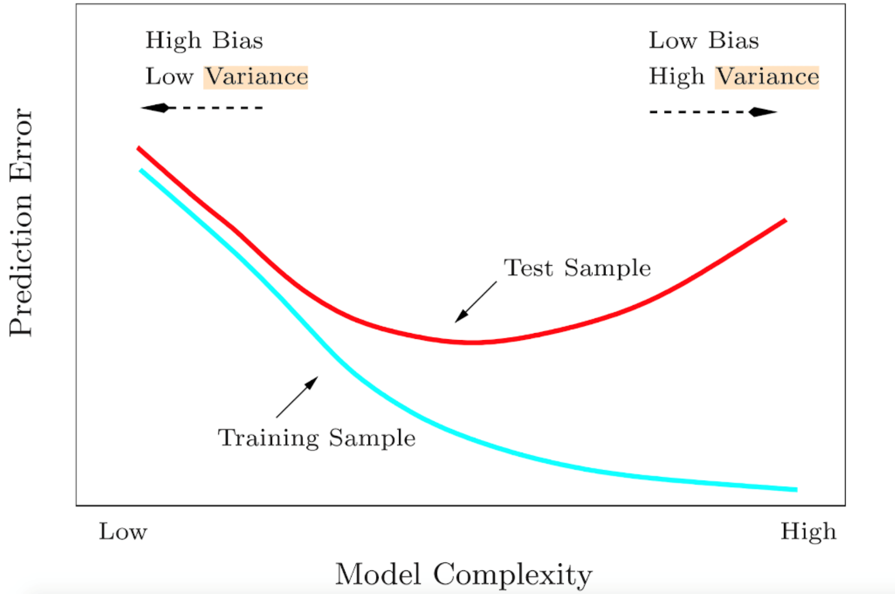
If you try to reduce one error, the other might go up!



Source: http://snoek.ddns.net/~oliver/mysite/the-bias-variance-tradeoff.html

Irreducible error:

Natural variability in a system caused by unknown/unpredictable factors



Source: https://techpolicyinstitute.org/wp-content/uploads/2017/12/Woloszko_Forecasting-GDP-growth-with-adaptive-trees-002.pdf

Bias-variance tradeoff

Overfitting and Underfitting and how to tackle these phenomenon.

- **Overfitting problem:**

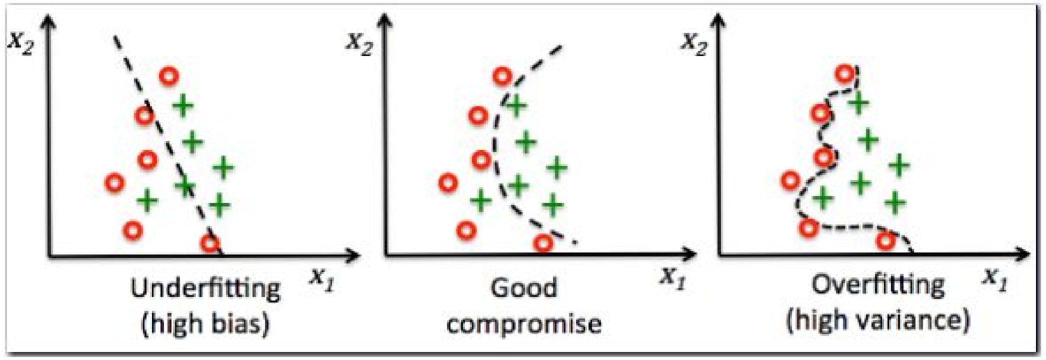
  Starting position high bias: Reducing the bias causes the variance to go up which leads to an overfitting problem

- **Underfitting problem:**

  Starting position high variance: Reducing the variance causes the bias to go up which leads to an underfitting problem

- How to tackle these phenomenon?

  Build a more complex model, Cross Validation, Dropout method, etc.



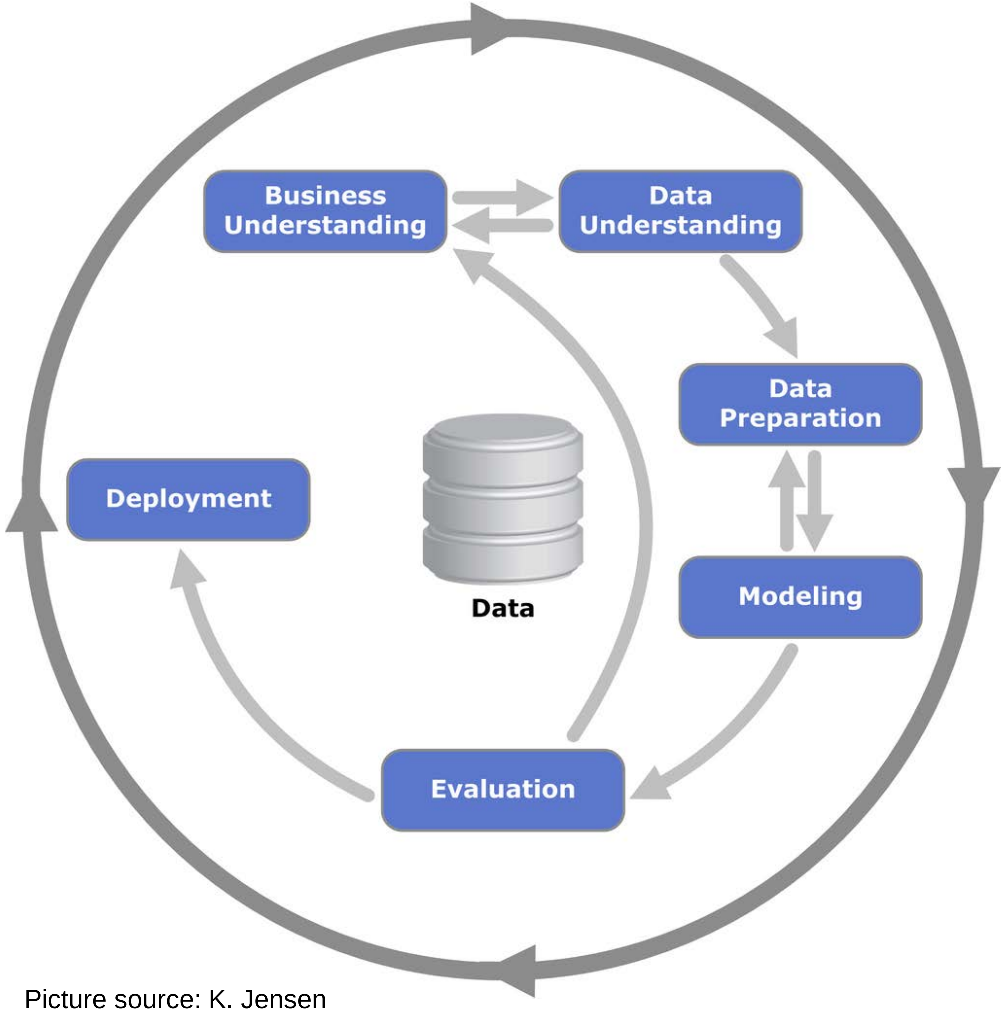
Source: https://www.sigs-datacom.de/trendletter/2019-11/2-ki-und-testen.html

Process models for data science and data mining

WHAT ARE SUITFUL APPROACHES?

Data mining life cycle: CRISP-DM

Cross-Industry Standard Process for Data Mining



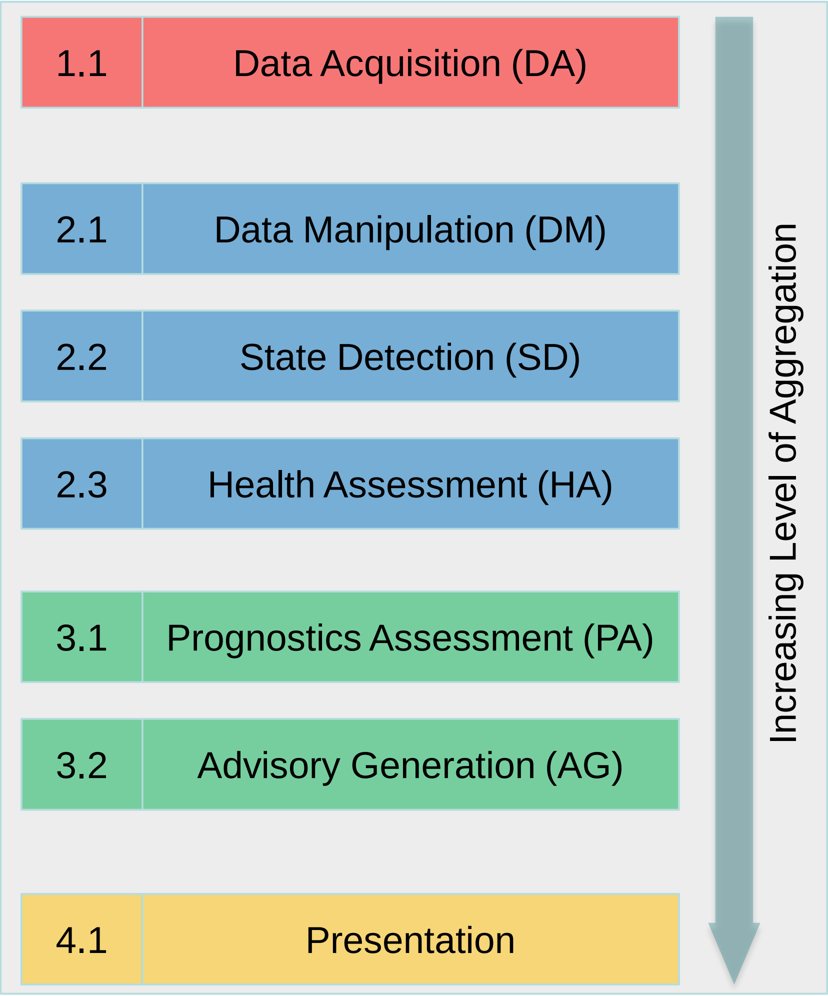
Picture source: K. Jensen

- Iterative open standard process model to gain knowledge from data related to business goals
- Feedback loops allow iteration of goals and models

1. Define project goals and business objectives
2. Understand the available data and there quality
3. Filter and select useful and relevant data
4. Create data-models that might meet the defined goals
5. Evaluate models' performance related to the goals
6. Set the best model into operation

Source: ftp://public.dhe.ibm.com/software/analytics/spss/documentation/modeler/18.0/en/ModelerCRISPDM.pdf

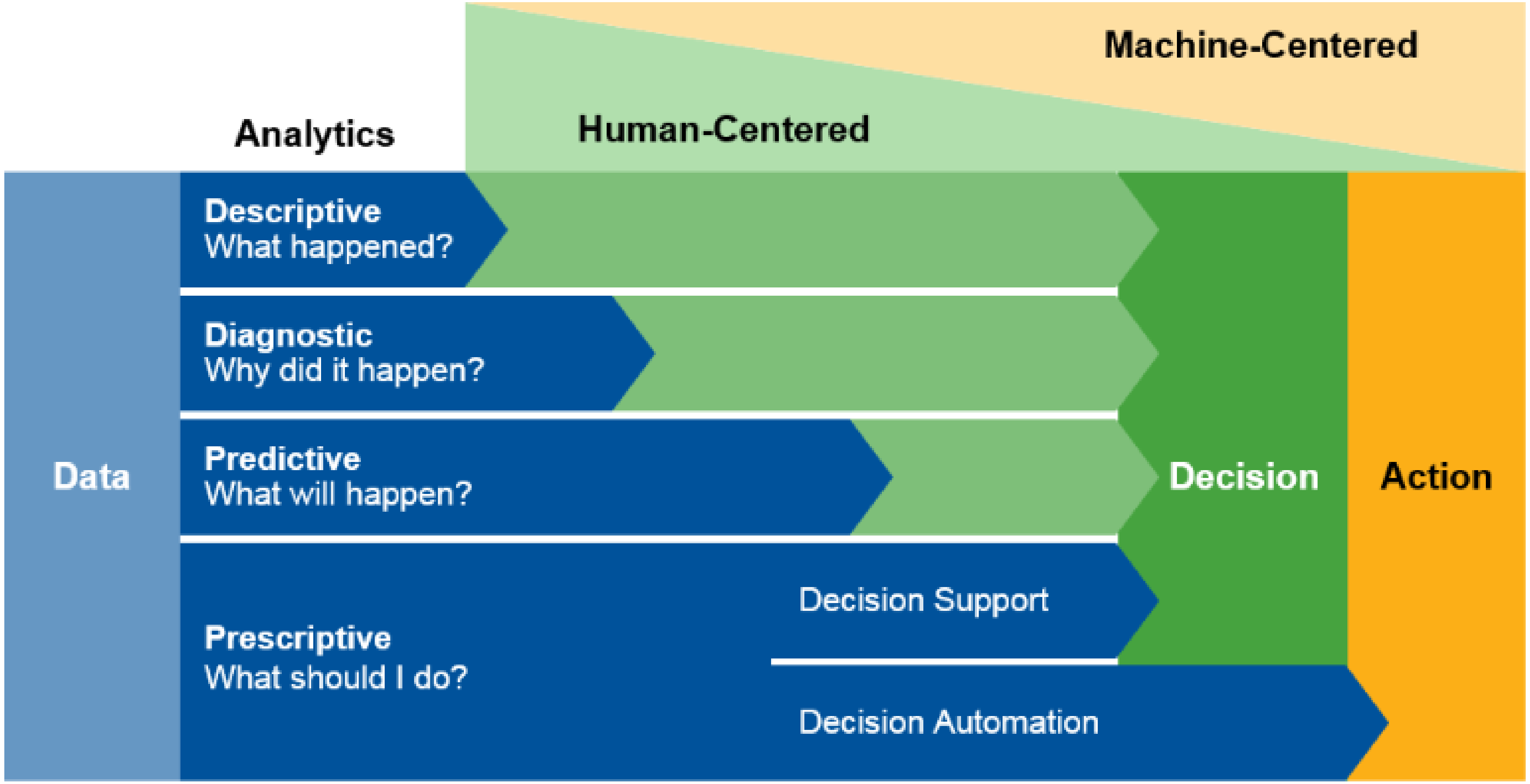
Open System Architecture for Condition-Based Maintenance (OSA-CBM)



- Developed by industry team and incorporated in ISO-13374-1
- OSA-CBM standard architecture for information processing
- OSA-CBM is a pure technical representation and does not value financial benefits
- Transformation of raw data into simple usable information to optimally plan maintenance operations

- 1.X sensors that measure physical parameters
- 2.X filter sensor data, create features, assess health
- 3.X predict health degradation, estimate measures/advisories
- 4.X inform the corresponding user (human-machine-interface)

For transformation of business intelligence, companies need to become aware of the four analytical capabilities



Source: Gartner Inc. [Publ.]: 2017 Planning Guide for Data and Analytics. Technical Professional Advice, G00311517 (2016)

What to take with you?

LEARNING OUTCOMES

Key Findings

TECHNISCHE UNIVERSITÄT DARMSTADT

- Machine learning changes the way to approach problems and opens new opportunities e.g. in automation, digital twins
- Machine learning is interesting for a broad range of industry tasks and can profit significantly from engineering knowledge
- Machine learning generalizes the experience so that a task performance can be improved
- Machine learning can solve problems, but not all problems require learning
- Machine learning models must be critically reviewed to avoid spurious results
- Machine learning rises new challenges regarding data security, data ownership, etc.

References

- Mitchel, T. M.: Machine Learning. McGraw-Hill (1997)
- Nam-Ho Kim, Dawn An, Joo-Ho Choi: Prognostics and Health Management of Engineering Systems. Springer (2017)
- Jörg Frochte: Maschinelles Lernen: Grundlagen und Algorithmen in Python. Hanser Fachbuchverlag (2018)
- A. El Saddik, "Digital Twins: The Convergence of Multimedia Technologies," in IEEE MultiMedia, vol. 25, no. 2, pp. 87-92, Apr.-Jun. 2018.
- Christoph Anger: Hidden semi-Markov Models for Predictive Maintenance of Rotating Elements; PhdThesis 2018; TU Darmstadt
- Christian Preusche: Clusterbasierte Zustandsbewertung von technischen Systemen zur Unterstützung der prädiktiven Instandhaltung. PhdThesis 2018; TU Darmstadt
- ISO-13374-1 Condition monitoring and diagnostics of machines — Data processing, communication and presentation — Part 1: General guidelines

Time for your questions and suggestions…