Machine Learning

Lecture 11 – Changeover Prediction

Prof. Dr.-Ing. Bastian Engelmann

E-mail: bastian.engelmann@fhws.de

FH·W-S

Tecnológico de Monterrey

The course objectives

#	Date	Topic	Lecturer
1	17.03.2021	Introduction	All
2	24.03.2021	Data visualization and preprocessing (Titanic)	Prof. Schmitt
3	31.03.2021	Getting to know Machine Learning	Prof. Schmitt
4	07.04.2021	Linear / Multiple Regression	Prof. Ceballos-Cancino
5	14.04.2021	Logistic Regression	Prof. Engelmann
6	21.04.2021	Decision Trees	Prof. Ceballos-Cancino
7	28.04.2021	Neural Networks	Prof. Engelmann
8	05.05.2021	Time Series Forecasting	Prof. Batres
9	12.05.2021	Time Series Forecasting	Prof. Ceballos-Cancino
10	19.05.2021	Quality Assurance Legostones	Prof. Schmitt
11	**26.05.2021**	**Changeover Prediction**	**Prof. Engelmann**
12	02.06.2021	Poster Session	All

We meet via ZOOM – the link is permanently the same and posted in the eLearning announcement

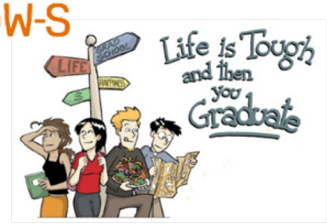
It´s a Machine Learning Tour

1. Cultural / virtual onboarding

2. The data tour starts

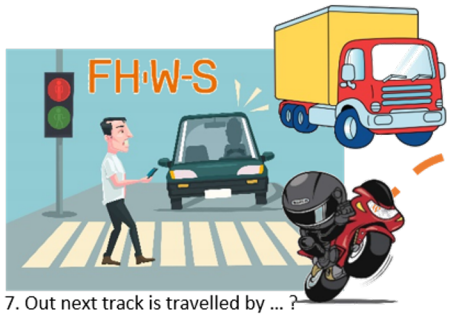
3. The propablity to survive

4. To be ill on tour is never a good idea



5. Will you pass the portfolio examination ?

6. Oh no, an illness again



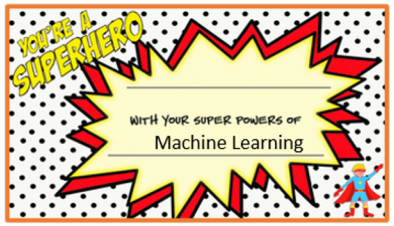
7. Out next track is travelled by ... ?

9. The time of our cruise is running out !

8. Time is ticking away !

10. We have to build bridges not walls !

11. Let´s go to the machine room



12. International Machine Learning Expert !

Agenda

Agenda

The lecture is structured in the following way:

- Introduction
  - OBerA project
  - The changeover task
- Data acquisition
  - Infrastructure
  - Sensor types
  - Data structure
- Let‘s get practical
  - CRISP-DM structure
  - Information on approach
  - Discussion

Source: colourbox educ. license

OBerA Project

PETER BREHM

EMUGE FRANKEN

HEISAB

HEITEC engineering solutions

KRITZNER Metalltechnik GmbH

PABST FIRMENGRUPPE

SIEMENS

Introduction to the research project „OBerA“

OBerA means: „Optimization of processes and machine tools through provision, analysis and target/actual comparison of production data“

- The project was established to support metalworking companies from the SME sector in the digitalization of their system landscape
- The consortium consists out of five production-oriented companies, Siemens as technology partner, the University of Applied Sciences FHWS and Heisab / Heitec for SAP integration
- The intention was to cover a broad spectrum of different companies standing exemplarily for the region of northern Bavaria
- A major goal of the project OBerA is to establish real-time data transparency in order to facilitate deviation management to increase productivity

OBerA Project

Companies in OBerA

Company	Turnover (Mio. €)	Number of Employees	Industrial Sector	Production Characteristic
Brehm	10–50	100	Medical implants	Single part production, series production
Franken	10–50	200	Tools for drilling, milling, and thread cutting	Small series and series production
Kritzner	10–50	100	Automation and automotive	Single part production, small series production
Pabst	6–10	60	Automotive and aerospace	Single part production, small series production
Siemens	>600	3000	Electronic motors	Series production

Source: Engelmann, B., Schmitt, S., Miller, E., Bräutigam, V., & Schmitt, J. Advances in Machine Learning Detecting Changeover Processes in Cyber Physical Production Systems, 2020

OBerA Project

SME

- SME are small and medium-sized enterprises
- Definition of EU for SME:
  Medium-sized companies have less than 250 employees and annual sales below 50 million euros

They often struggle with Industry 4.0:

- Old machines with no connectivity
- No infrastructure
- Not enough knowledge/resources

Source: colourbox educ. license

OBerA Project

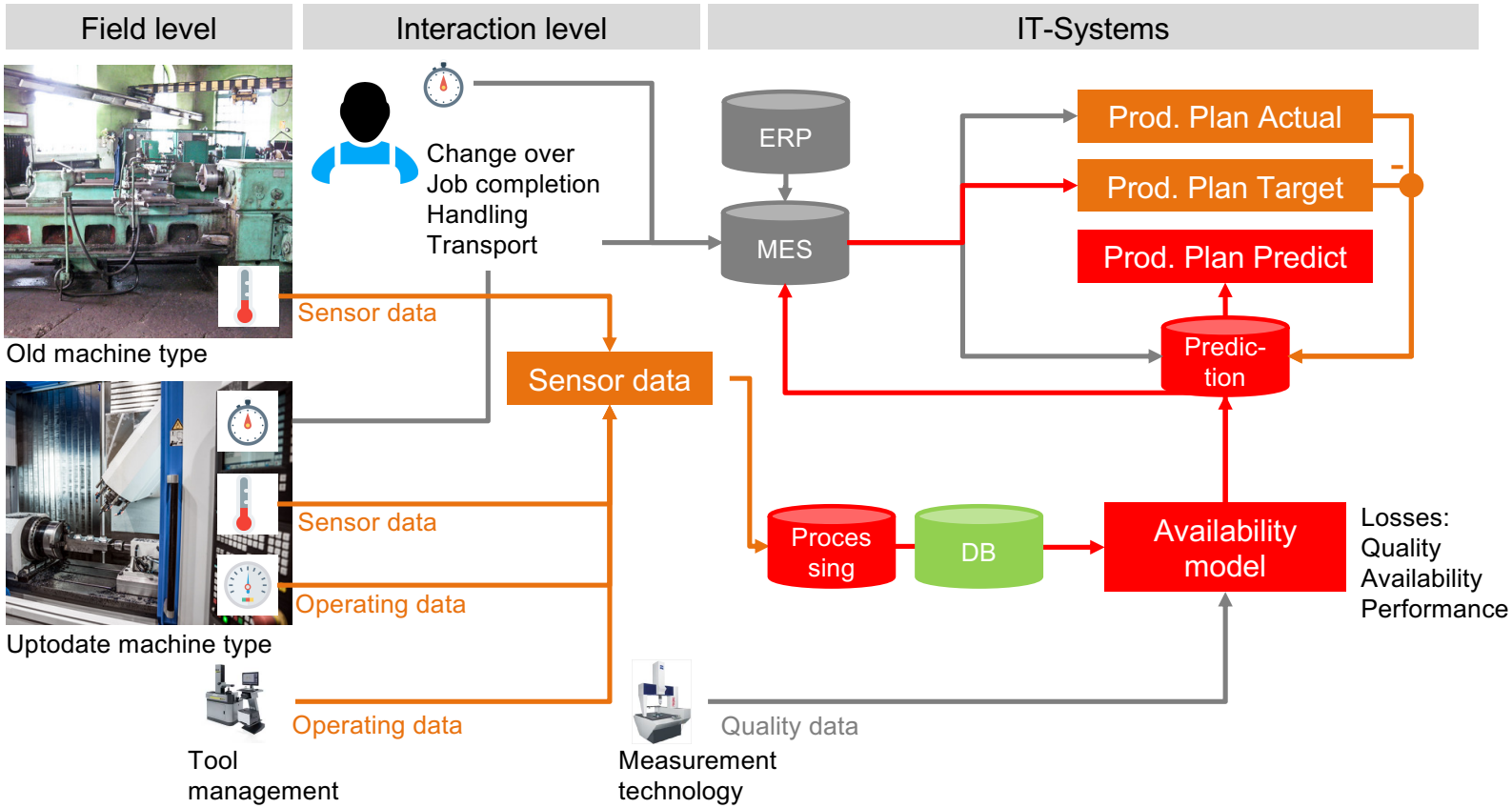
Volker and Manfred Pabst

Company Pabst

- Medium sized family business since 1979 founded by Roland Pabst
- Consists out of Pabst Components Manufacturing and Pabst Automotive
- Production of a wide range of machining components for the automotive and mechanical engineering industries as well as for the aerospace sector
- Managed by the Pabst brothers Manfred (CFO) und Volker (CTO)

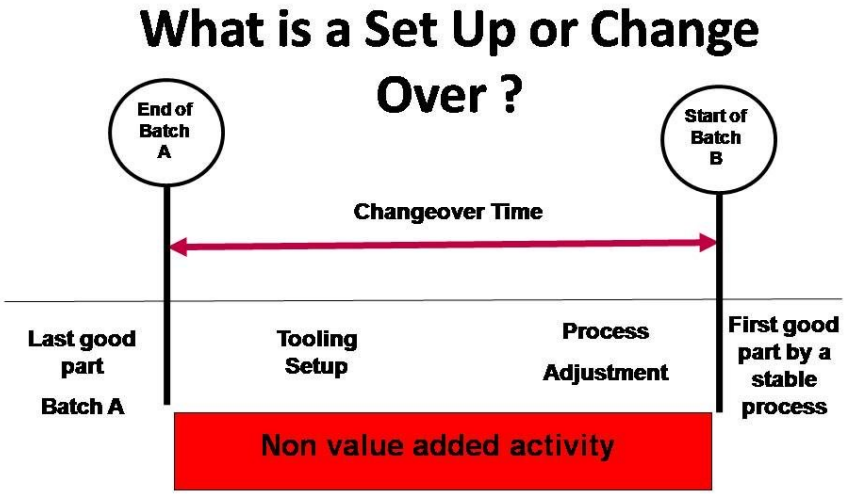
Source: http://www.pabst-gmbh.de/firmengruppe/organisation.html

Research approach of OBerA



What is „Changeover“?

- Preparing the necessary resources to perform a manufacturing task
- Refers to the time elapsed between producing the last good part of the first lot and the first good part of the next lot
- Efficient changeover process is an important economical and organisational element in production



“Quick Change Over” is a method to reduce the time spent going from producing one batch of material to the next through the use of a team.

Source: Ahmad, R., & Soberi, M. S. F. (2018). Changeover process improvement based on modified SMED method and other process improvement tools application
Image source: https://i.pinimg.com/originals/26/04/f7/2604f7a212dc3a6cca4d1d37ae7b08f8.jpg

OBerA Project

FHWS is preparing the BAZ1 milling machine

BAZ1:
- DECKEL MAHO DMG 100 U duoblock
- 5-axis milling center
- Built in 2005
- Control:
  Heidenhain Mill Plus IT / Version 530

Task Changeover@Pabst

- The change over process is done for the first item of each batch
- It includes all manufacturing steps that are needed to finish the piece (tool management, load CNC program, test milling etc.)
- Everything is done manually by the worker but only for the first part
- After the first part is finished, all the other parts go in to automated production using the settings of the change over from the first part
- The change over process often varies in time and remains untransparent for management

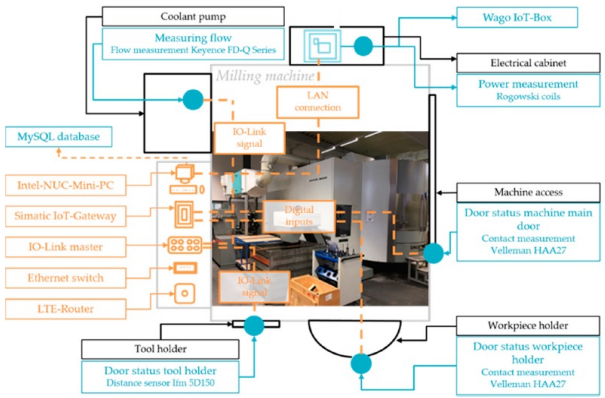
⇨ **Can it changeover be detected automatically from sensor data?**

OBerA Project

The milling process

OBerA Project

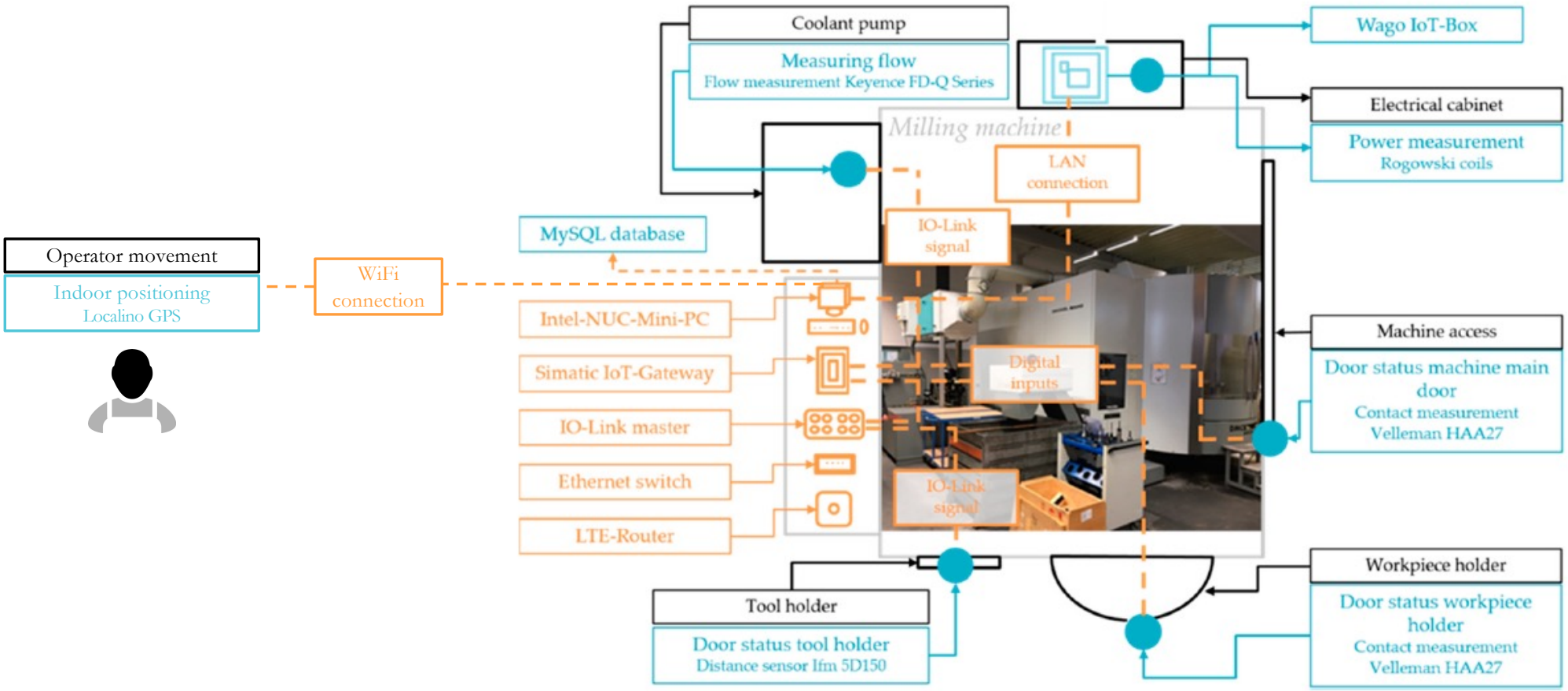
Data acquisition setup for brownfields



- Connecting new machines is quite easy because the bring state-of-the-art interfaces with them (“greenfield scenario”)
- Older machines are very difficult to upgrade to state-of-the-art interfaces
- In SME often older machines are kept forever (“brownfield scenario”)
- The approach in the project is to use a big variety of sensors, interfaces to learn also how to realize data acquisition in a brownfield scenario

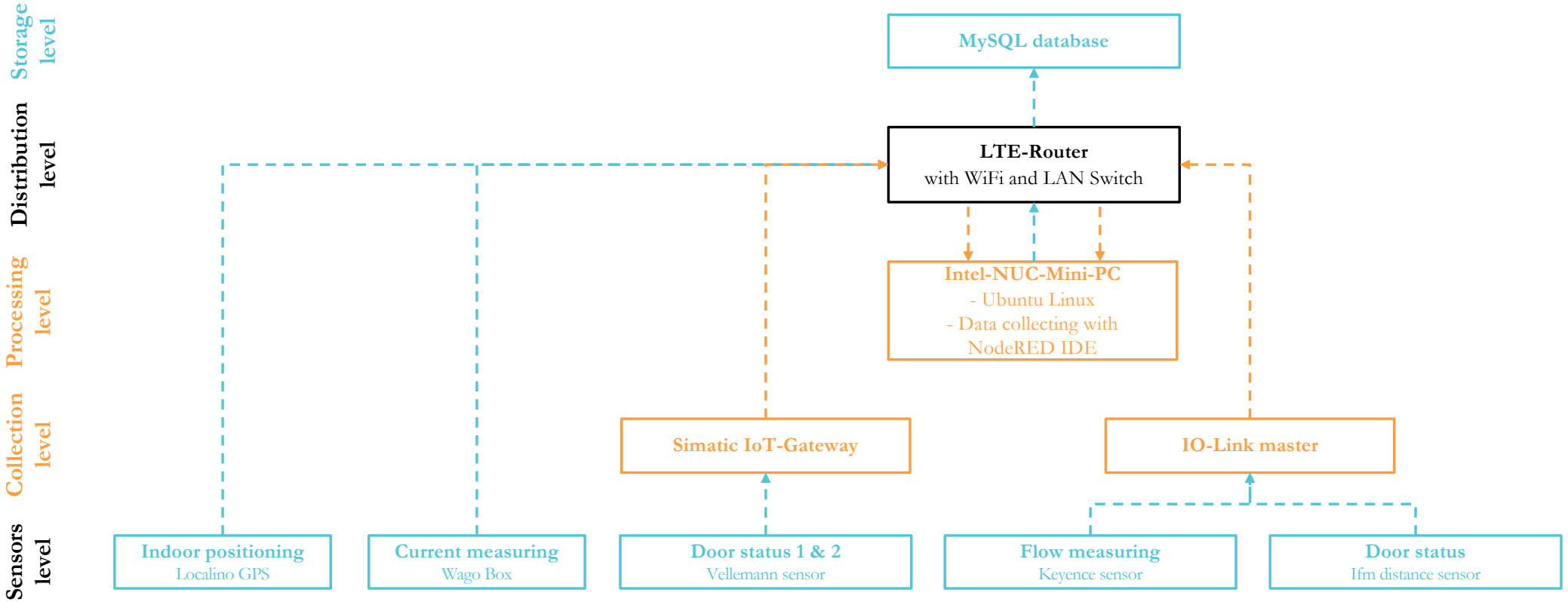
OBerA Project

Data acquisition setup



OBerA Project

Data acquisition setup

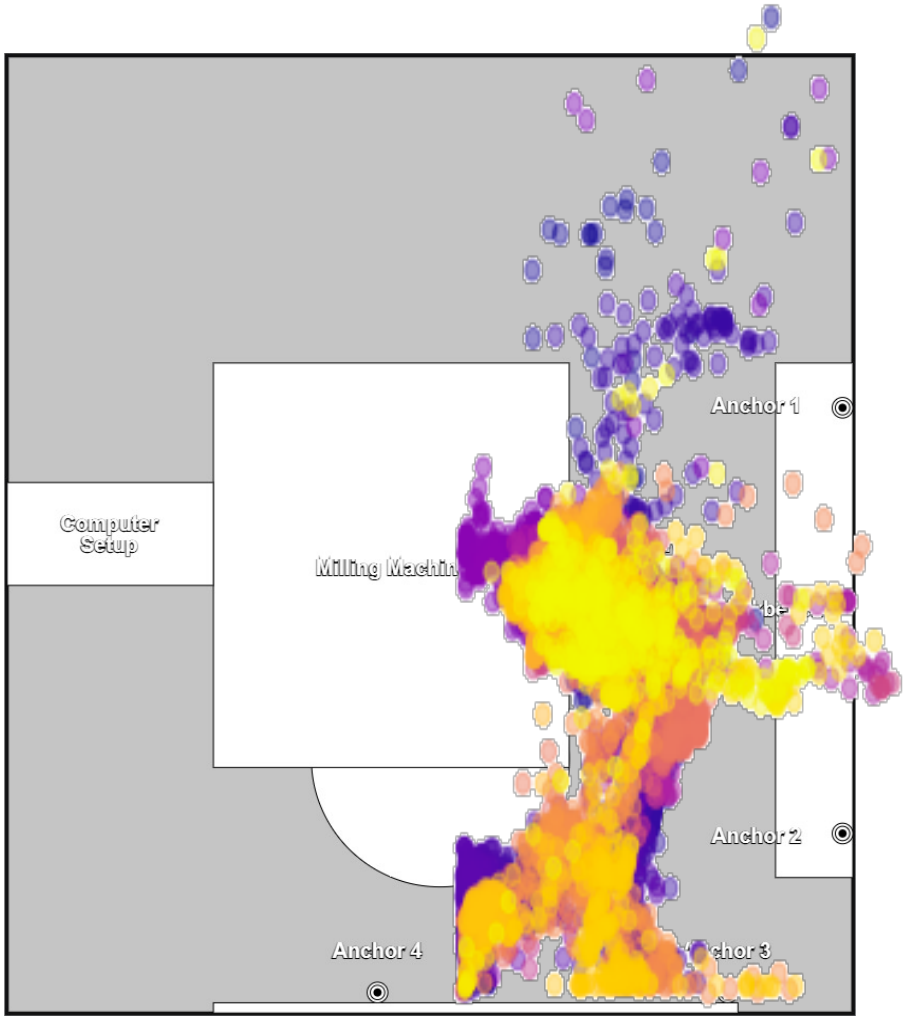


OBerA Project

Sensors: Localino Indoor GPS

OBerA Project

Sensors:
Indoor
positioning

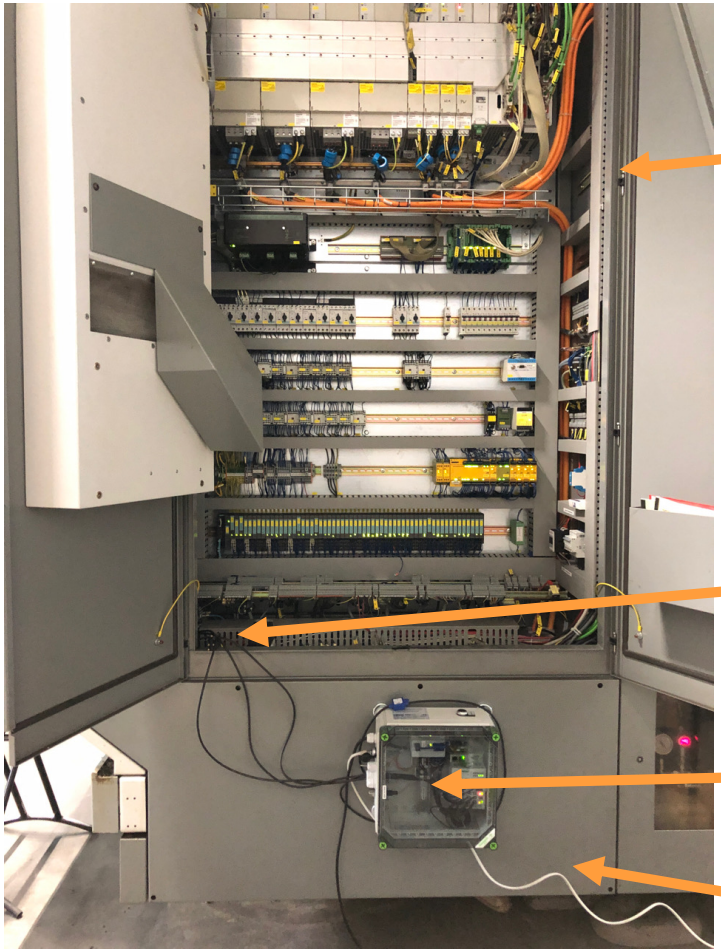


OBerA Project

Sensor: Rogowski coils - Measuring current in cables

OBerA Project

Sensor: Rogowski coils - Measuring current in cables

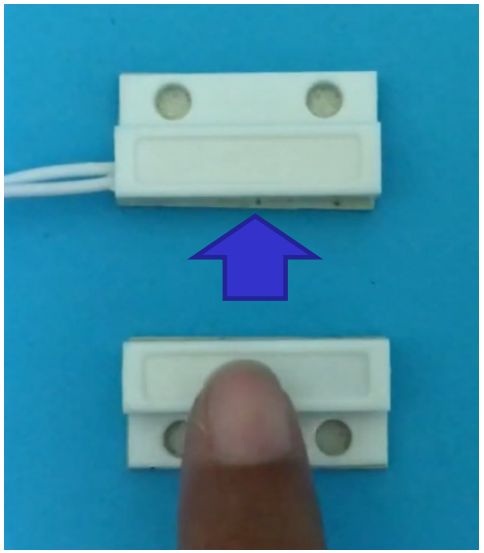


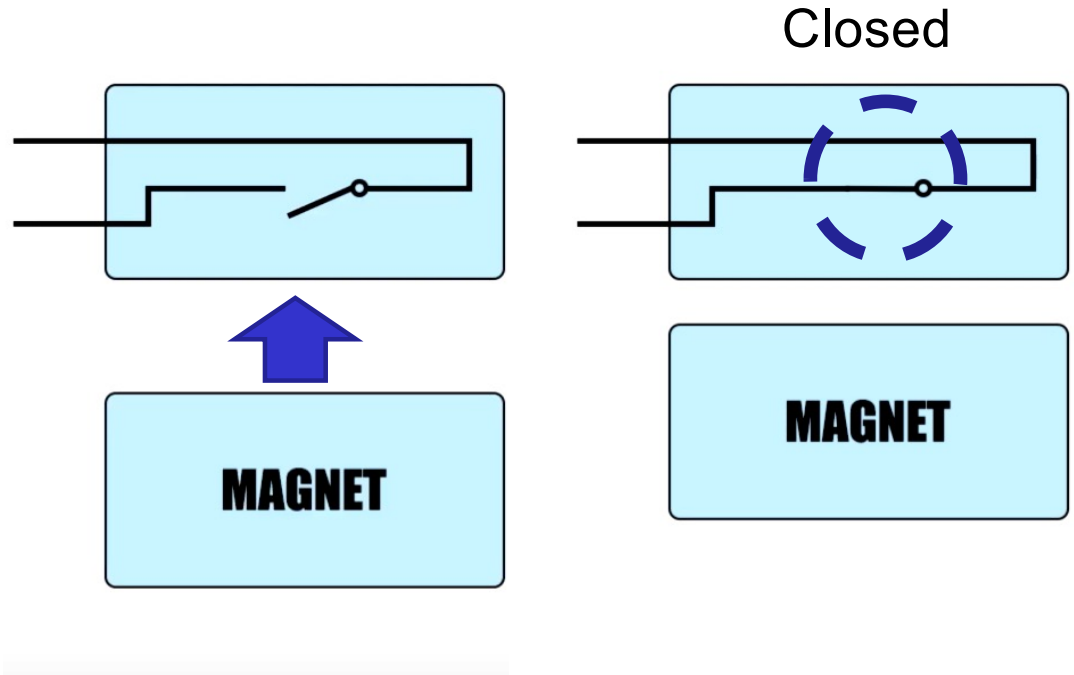
OBerA Project

Sensor: Binary door status

Reed switch

Magnet



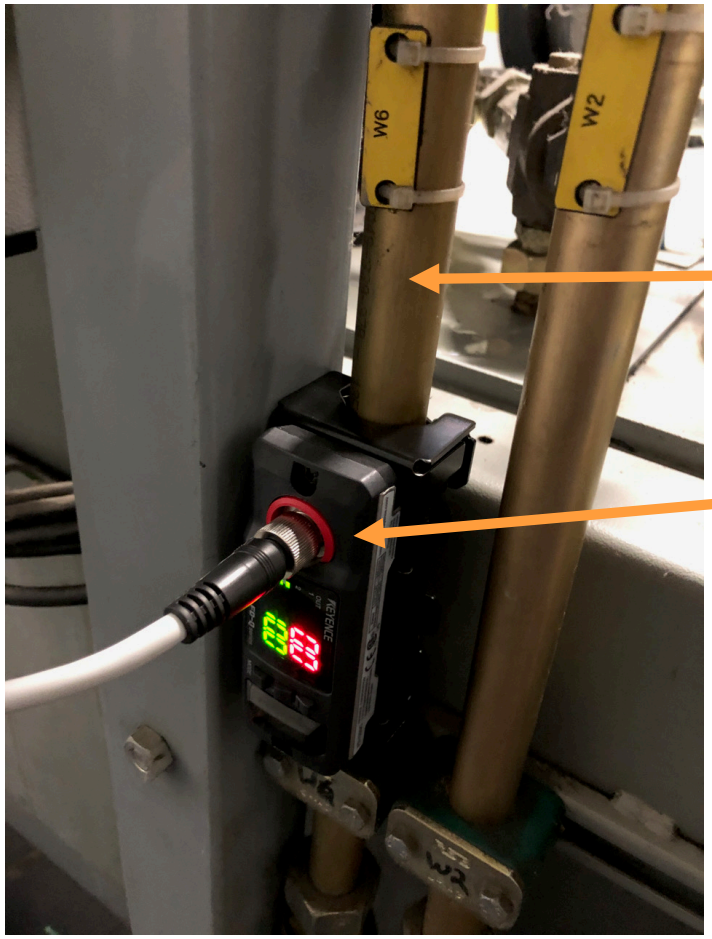


V1 - Main machine door	Vellemann 1
V2 - Changeover chamber door	Vellemann 2

Source: https://www.youtube.com/watch?v=7HR3jgthb3s

OBerA Project

Sensor: Measuring flow with ultra sonic pulse response

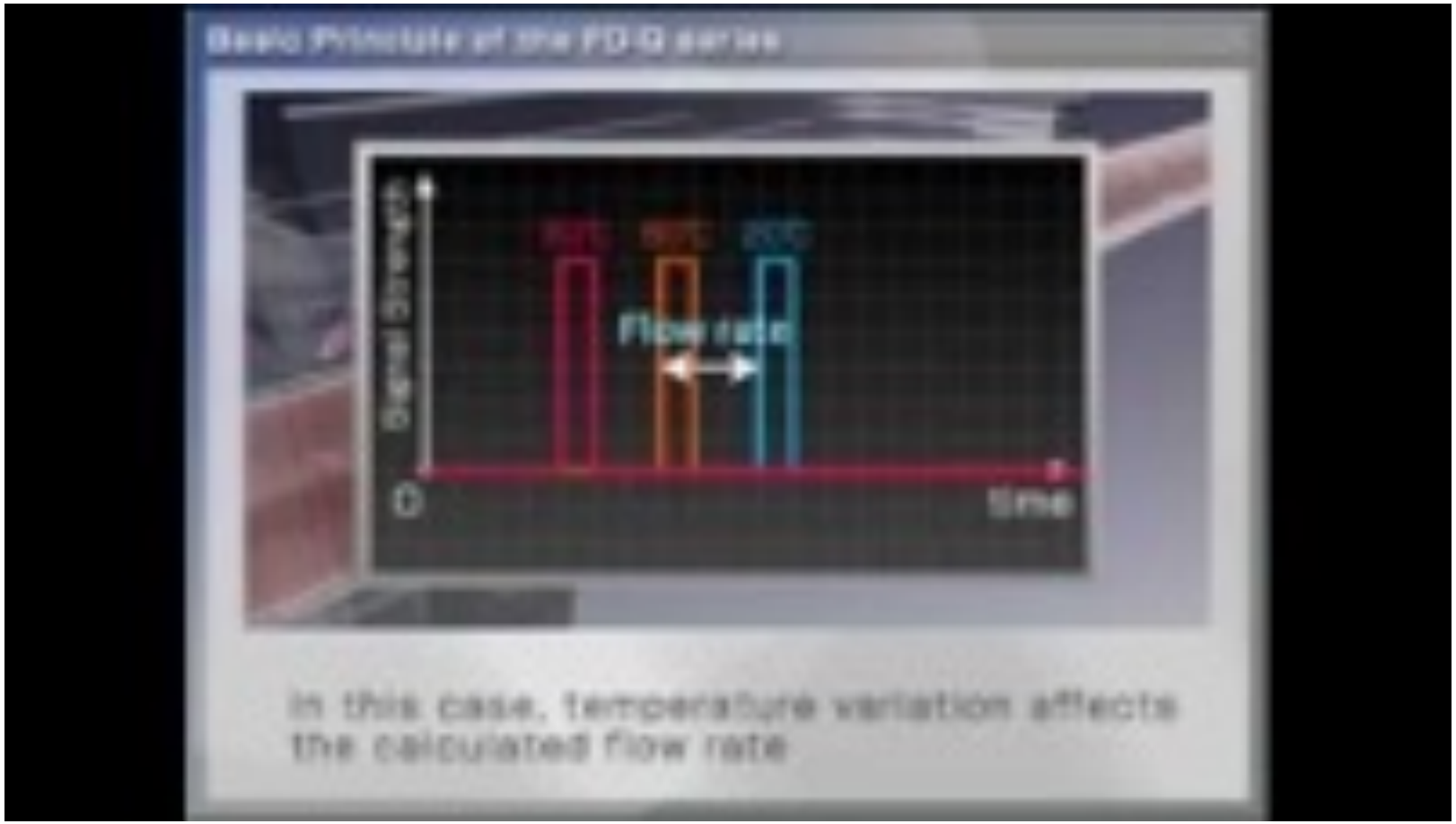


Main coolant pipe

Keyence FD-Q sensor

OBerA Project

Sensor: Measuring coolant flow



OBerA Project

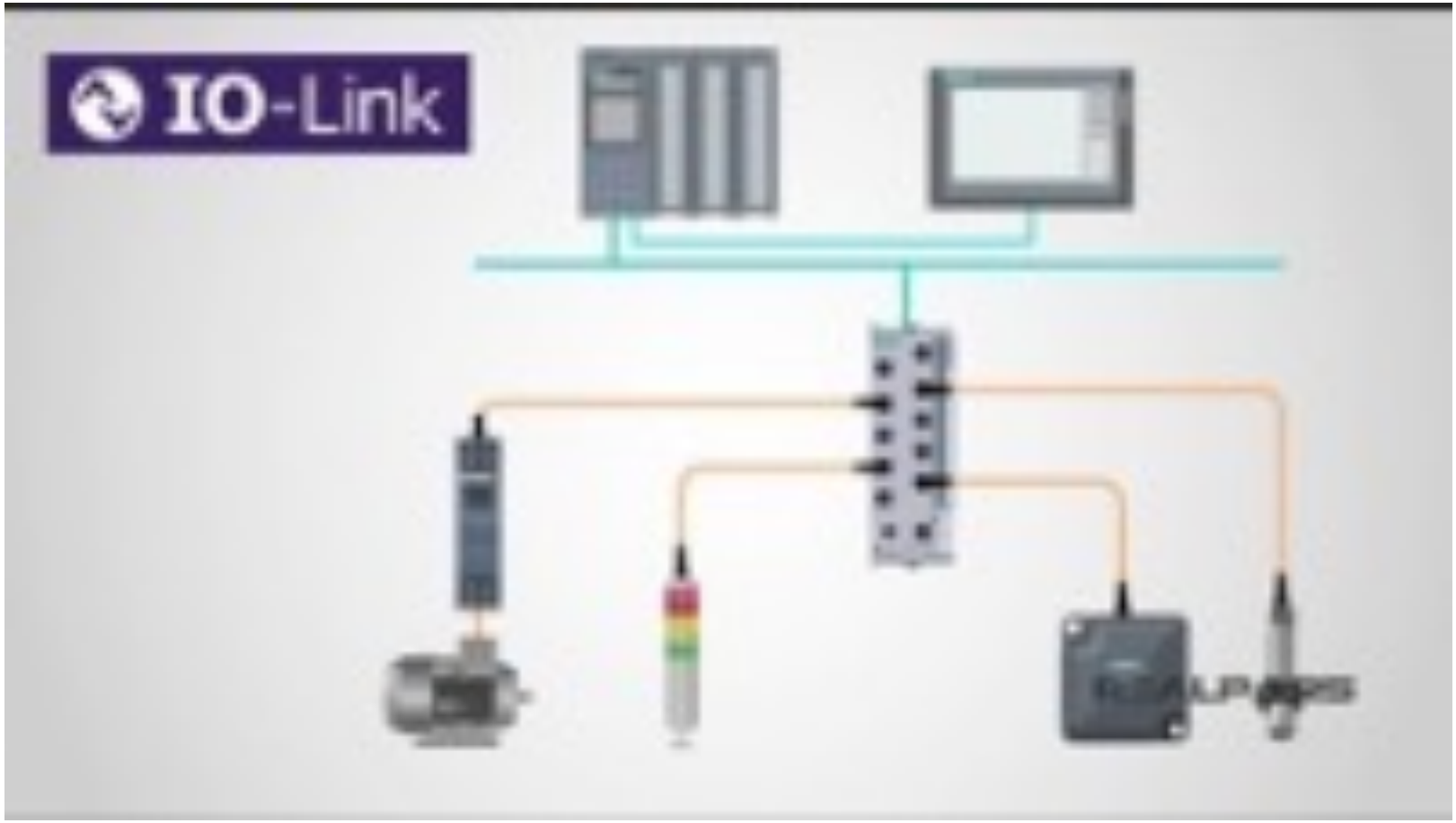
Sensor: Distance door status with Ifm O5D150

- Optical distance sensor measures distance to the door handle of the tool management cabin
- Very high range due to time-of-flight measurement
- Reliable background suppression and color-independent detection
- With display and pushbuttons for precise switching point adjustment
- Laser protection class 1 for the requirements of the automotive industry
- Reliable detection of glossy surfaces

⇨ Used here to test IO-Link capabilities

OBerA Project

Collection: IO-Link



OBerA Project

Collection: Simatic IoT-Gateway

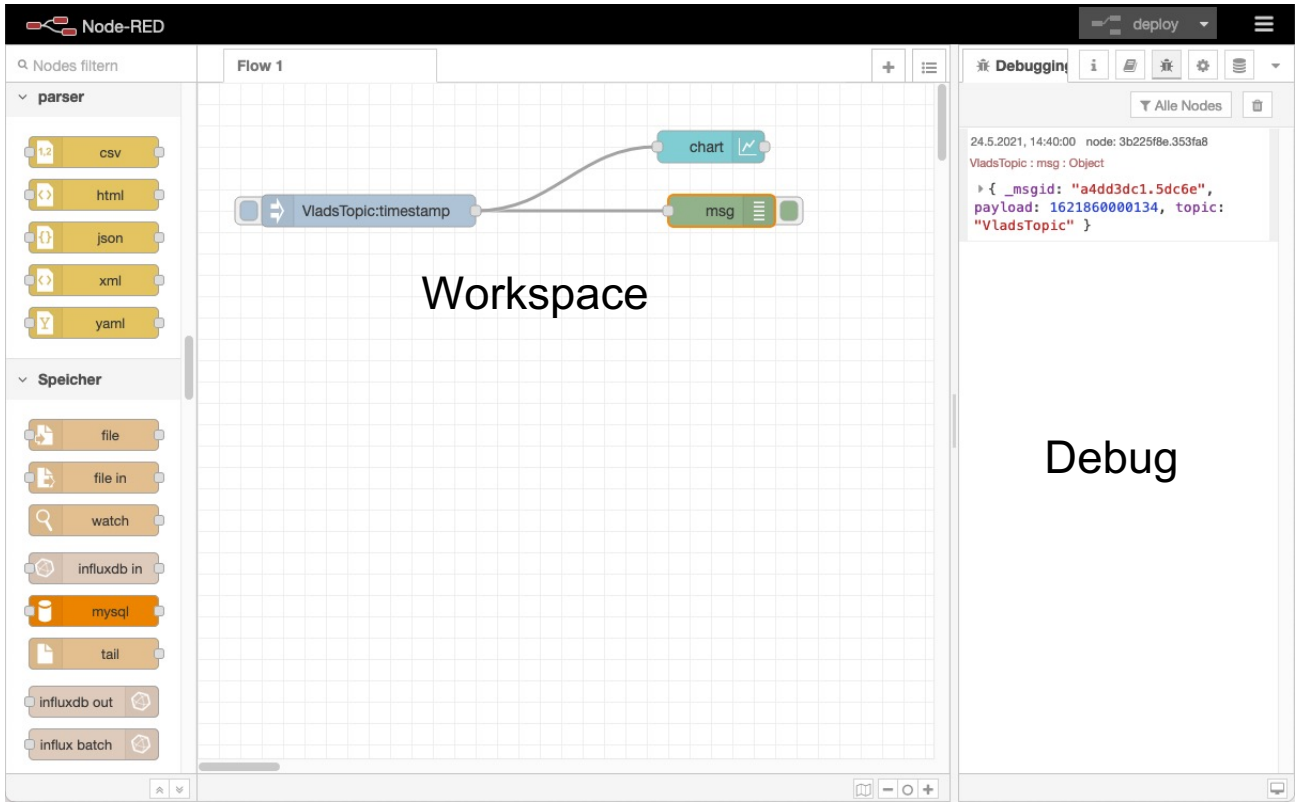
Data processing with NodeRED on the Intel NUC

Node-RED
Flow-based programming for the Internet of Things

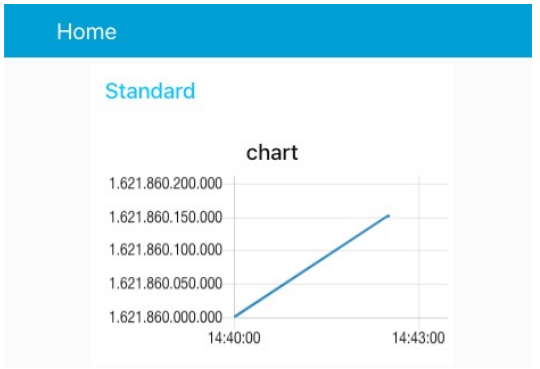
- "Node-RED is a programming tool for wiring together hardware devices, APIs and online services in new and interesting ways."
- Node-RED can be run from a web browser,
  - either via a local installation: http://localhost:1880/ or http://127.0.0.1:1880/
  - ... or in vLab!
- Node-RED is based on Node.js, which is an extension of the "Javascript" programming language.

OBerA Project

NodeRED in vLab



Node palette



Easy Dashboarding

OBerA Project

Stages of changeover of the milling process

	DATETIME	Distance	Power	Liter	V1	V2	X	Y	Phase	Phase_compressed	Production
3990	2021-03-31 08:33:25	20.0625	-241.060	0	1	1	4.02959	1.173720	13	2	0
5133	2021-03-31 08:58:43	20.0625	674.090	0	1	1	4.09991	1.094430	13	2	0
7887	2021-03-31 10:07:11	20.0625	-1342.350	64	1	1	3.90334	0.951932	21	5	1

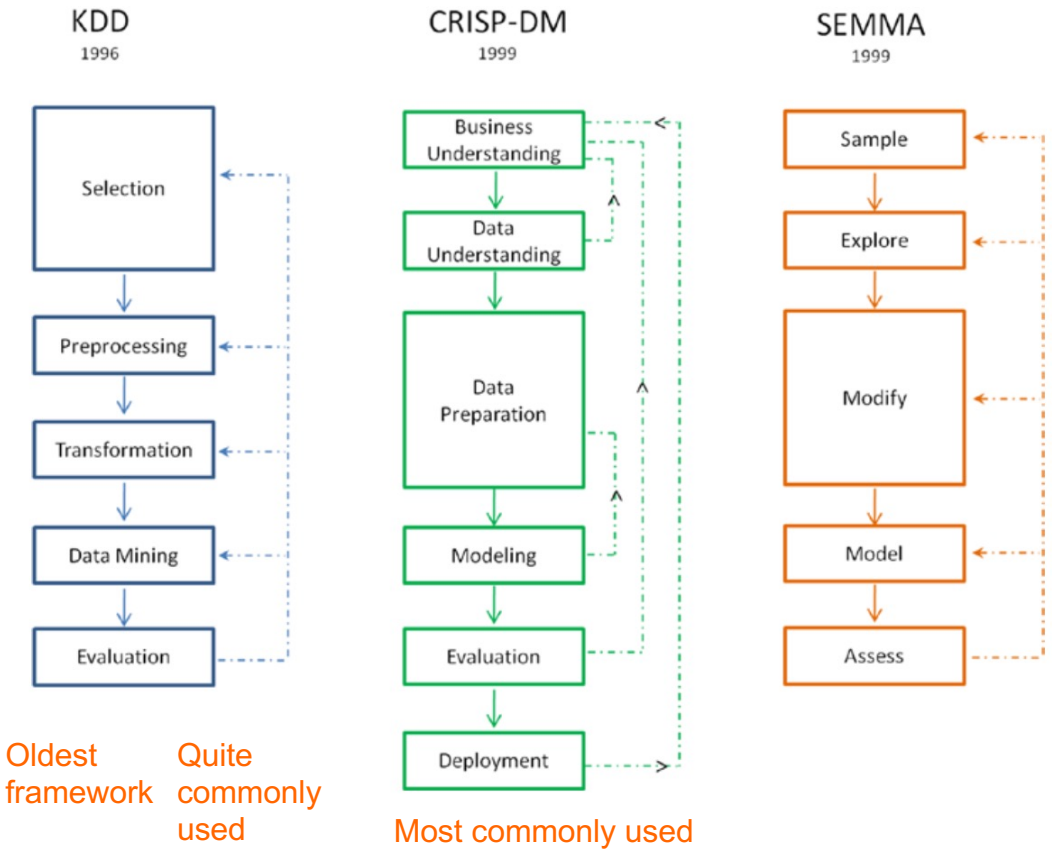
- Economically interesting is just the status “Change over / Production / Idle”
- For building a model, having more stages can be better suited for algorithmic classification
- Higher effort for more stages occurs for the supervision time of change over and then the labelling of the data for supervised learning
- Your CSV will have labels for 2 (“Production”), 5 (“Phase_compressed”) and 15 phases (“Phase”) - see table above

OBerA Project

Stages of changeover of the milling process

2+1 Stages	4+1 Stages	15+1 Stages		
Change over	Change over start	1/19	An/Abmelden des Rüstauftrages am Terminal	Logging on/off the setup job at the terminal
		2	Reinigen des Maschinentisches	Cleaning of the machine table
		3	Befördern der Vorrichtung von Maschinen- tisch zu Werkbank	Moving the fixture from machine table to workbench
		4	Befestigen von Bauteil an Vorrichtung	Fastening of component to fixture
		5	Befördern der Vorrichtung von Werkbank zu Maschinentisch	Transfer of fixture from workbench to machine table
		6	Befestigen von Vorrichtung an Maschinen- tisch	Attachment of fixture to machine table
		7+8	Laden des NC-Programmes / Durchführen der Werkzeugvoreinstellung	Loading the NC program / performing tool presetting
		9	Befüllen des Werkzeugmagazins	Filling the tool magazine
		10	Eingeben der aktuellen Werkzeugmaße	Entering the current tool dimensions
	Change over main stage	11	Schwenken der Maschinentische 1 und 2	Swiveling of machine tables 1 and 2
		12	Setzen des Nullpunkts	Setting the zero point
		13	Ablaufen und Optimieren des NC-Program- mes	Running and optimizing the NC program
	Change over stop	14	Reinigen des Bauteils & Abmontierten des Bauteils	Cleaning of the component & dismantling of the component
		16	Entgraten des Bauteils & Nachmessen des Bauteils	Deburring of the component & remeasurement of the component
		18	Laden des optimierten NC-Programms auf Server	Load the optimized NC program on the server
		1/19	An/Abmelden des Rüstauftrages am Terminal	Logging on/off the setup job at the terminal
Production	Production	Production		
(Idle)	(Idle)	(Idle)		

Let‘s get practical: Data mining frameworks



⇨ CRISP-DM used here, with more detailed phase descriptions (Mejbah, A. 2020)

Source: Swamynathan, M. 2017. Mastering Machine Learning with Python in Six Steps, Apress. S. 75

CRISP-DM: Business understanding

Business Understanding

- Definition of business objectives for the problem
- Success criteria for business and data mining efforts
- Budget allocation and resource planning
- Clear, well-defined Machine Learning and data mining methodologies to be followed, including high-level workflows from exploration to deployment
- Detailed project plan with all six phases of the CRISP-DM model defined with estimated timelines and risks

⇨ Done!

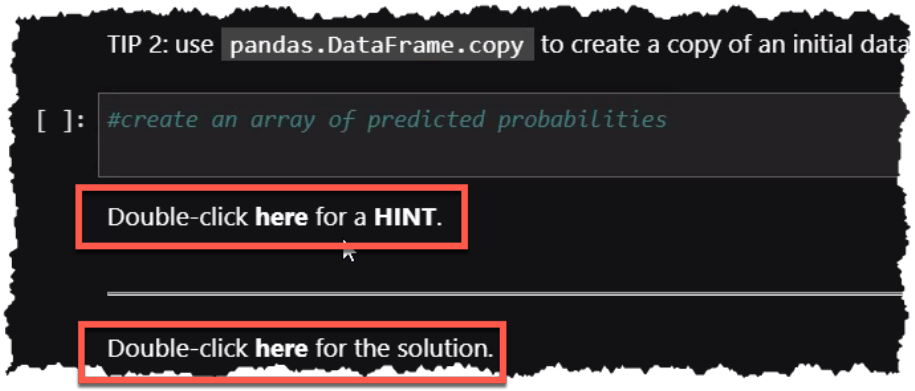
Source: Mejbah, Ahammad (2020). Machine Learning Successful Project Plan. Online: https://www.kaggle.com/getting-started/181153#1002081 (retrieved: 19.02.2021).

Let‘s get practical!

Information about the Jupyter notebook

- Vladyslav prepared a Jupyter notebook with all the necessary steps for you
- You find it along with the .csv data file online in eLearning
- It is intended that you solve single tasks from the following slides with your group in Zoom breakout rooms
- Please assign your team number to your Zoom name to facilitate group organisation in Zoom
- We will get you out of the Zoom rooms in the end to discuss the results in the audience
- If you need help then raise your hand in the Zoom room!
- There are hints in the notebook
- If you get stuck, then there are also solutions, but please first try to solve it on your own!

  Have fun!



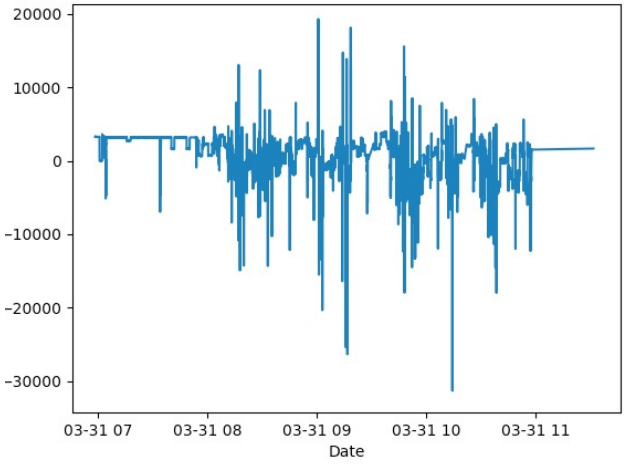
CRISP-DM: Data understanding

Data Understanding

- *Data Collection* - Sensor Data (Done!)
- *Data Description*
  Data sources (SQL, NoSQL, Big Data), record of origin (ROO), record of reference(ROR) (TASK 1.1)
  Data volume (size, number of records, total databases, tables) (TASK1.2)
  Data attributes and their description (variables, data types) (TASK1.3, TASK1.4)
  Relationship and mapping schemes (understand attribute representations)
  Basic descriptive statistics (mean, median, variance)
- *Exploratory Data Analysis* - descriptive statistics, plots, charts, and visualizations (TASK1.5)
  Explore, describe, and visualize data attributes (TASK1.6)
  Select data and attributes subsets that seem most important for the problem (Slide)
  Extensive analysis to find correlations and associations and test hypotheses (TASK1.7)
  Note missing data points if any (Slide)
- *Data Quality Analysis*
  Missing values (Slide)
  Inconsistent values (Slide)
  Wrong information due to data errors (manual/automated)
  Wrong metadata information
  } (TASK 1.8)

Source: Mejbah, Ahammad (2020). Machine Learning Successful Project Plan. Online: https://www.kaggle.com/getting-started/181153#1002081 (retrieved: 19.02.2021).

CRISP-DM: Data understanding



Remark on data selection

- All sensor data is recorded with a time stamp:

	DATETIME	Distance	Power	Liter	V1	V2	X	Y	Phase	Phase_compressed	Production
3990	2021-03-31 08:33:25	20.0625	-241.060	0	1	1	4.02959	1.173720	13	2	0
5133	2021-03-31 08:58:43	20.0625	674.090	0	1	1	4.09991	1.094430	13	2	0
7887	2021-03-31 10:07:11	20.0625	-1342.350	64	1	1	3.90334	0.951932	21	5	1

- For the model you must decide, if you would use classification or forecasting algorithms
- For binary classification you won’t need the time stamps
- Binary classification is suggested in the Jupyter notebook

⇨ Please feel free to use a classification or forecast approach!

CRISP-DM: Data understanding

```
Start: 2021-03-31 08:45:44 | Duration: 0:00:03
Start: 2021-03-31 08:45:48 | Duration: 0:00:02
Start: 2021-03-31 08:45:54 | Duration: 0:00:02
Start: 2021-03-31 08:45:59 | Duration: 0:00:04
Start: 2021-03-31 08:46:07 | Duration: 0:00:02
Start: 2021-03-31 08:46:13 | Duration: 0:00:02
Start: 2021-03-31 08:46:19 | Duration: 0:00:03
Start: 2021-03-31 08:46:26 | Duration: 0:00:02
Start: 2021-03-31 08:46:32 | Duration: 0:00:02
Start: 2021-03-31 08:46:38 | Duration: 0:00:02
Start: 2021-03-31 08:46:41 | Duration: 0:00:02
Start: 2021-03-31 08:46:44 | Duration: 0:00:02
Start: 2021-03-31 08:46:51 | Duration: 0:00:02
Start: 2021-03-31 08:46:53 | Duration: 0:00:02
Start: 2021-03-31 08:46:57 | Duration: 0:00:02
Start: 2021-03-31 08:47:02 | Duration: 0:00:03
Start: 2021-03-31 08:47:09 | Duration: 0:00:02
Start: 2021-03-31 08:47:16 | Duration: 0:00:02
...                        | ...
Start: 2021-03-31 10:58:05 | Duration: 0:33:48
```

Timestamp gaps

- There was a breakfast break when the Localino Indoor GPS still recorded (33min)
- The timestamps often show gaps > 2 seconds
- In the analysis it turned out that the Localino GPS data often has gaps
- The missing data from the GPS is sent in a batch, so that often for specific time stamps several (X, Y) positions from the sensor exist
- These incidents were filtered out as duplicates (strategy “take first value”)

⇨ From the originally recorded 24.000 data rows, only 10.000 are remaining int the CSV-file!

CRISP-DM

Data Preparation

Clean, wrangle, curate and prepare the data

- *Data Integration*
  Same attributes: combining data
  Different attributes: keys
- *Data Wrangling*
  processing, cleaning, normalization, and formatting (Task 2.1)
  underlying errors and inconsistencies
  Handling missing values (remove rows, impute missing values) (Task 2.2)
  Handling data inconsistencies (delete rows, attributes, fix inconsistencies)
  Fixing incorrect metadata and annotations
  Handling ambiguous attribute values
  Curating and formatting data into necessary formats (CSV, Json, relational)
- *Attribute Generation and Selection*
  Feature extraction and engineering (Task 2.3)
  creating new attributes or variables (Slide)
  selecting a subset of features or attributes from the dataset

Source: Mejbah, Ahammad (2020). Machine Learning Successful Project Plan. Online: https://www.kaggle.com/getting-started/181153#1002081 (retrieved: 19.02.2021).

CRISP-DM

Creating new attributes or variables

- To improve the accuracy of ML models often variables are aggregated
- Examples:

Distance door status

Distance
20.0625
20.0625
20.0625
20.0625

With rule „> 20.0“ transform into categorical data „open“ & „close“

GPS Position

X	Y
4.02959	1.173720
4.09991	1.094430
3.90334	0.951932
3.37632	0.861243

Define squared checkboard with n^2 fields and use field number as categories

⇨ Not used in CSV

⇨ Feel free to aggregate existing variables in new ones!

CRISP-DM

Modelling

Success criteria, data mining objectives, and business objective

- *Selecting modelling techniques* (TASK3.1: You are free to use what ever you want, but there is a suggestion in the notebook)
  data mining tools, frameworks, techniques, and algorithms
  depending on data available, business goals, data mining goals, algorithm requirements, and constraints
- *Model Building*: Training Model (TASK3.1)
- *Model Evaluation and Tuning* (parameter optimization): model accuracy, (TASK3.2)
  precision, recall, F1 score, (in TASK3.2 & Slide)
  grid search and
  cross validation (TASK3.3)
- *Model Assessment*
  Reproducible and consistent results from models Scalability, robustness, and ease of deployment Future extensibility of the model

Source: Mejbah, Ahammad (2020). Machine Learning Successful Project Plan. Online: https://www.kaggle.com/getting-started/181153#1002081 (retrieved: 19.02.2021).

CRISP-DM: Models for imbalanced data

F1 score

- There are several metrics to assess the quality of a model like “Accuracy” or “F1 score”
- “Accuracy” can be used when categories / label are equally distributed while “F1 score” is a better metric when there are imbalanced classes
- “Imbalanced” means that one category / label occurs more often than the other (e.g. many more rows for “Change over” than for “Production”)
- F1 score reaches its best value at 1 and worst score at 0

⇨ Compute the F1 score, also known as balanced F-score or F-measure with **sklearn.metrics.f1_score**

Source:
Image: Colourbox educational license
https://scikit-learn.org/stable/modules/generated/sklearn.metrics.f1_score.html https://medium.com/analytics-vidhya/accuracy-vs-f1-score-6258237beca2

CRISP-DM

Evaluation

Detailed assessment and review of the final models

- Ranking final models based on the quality of results and their relevancy (Slide)
- Any assumptions or constraints that were invalidated by the models (Slide)
- Cost of deployment of the entire Machine Learning pipeline from data extraction and processing to modelling and predictions
- Any pain points in the whole process? What should be recommended? What should be avoided?
- Data sufficiency report based on results
- Final suggestions, feedback, and recommendations from solutions team and SMEs (Mindmap)
- Based on the report formed from these points, after a discussion, the team can decide whether they want to proceed to the next phase of model deployment or a full reiteration is needed, starting from business and data understanding to modelling

Source: Mejbah, Ahammad (2020). Machine Learning Successful Project Plan. Online: https://www.kaggle.com/getting-started/181153#1002081 (retrieved: 19.02.2021).

Model evaluation

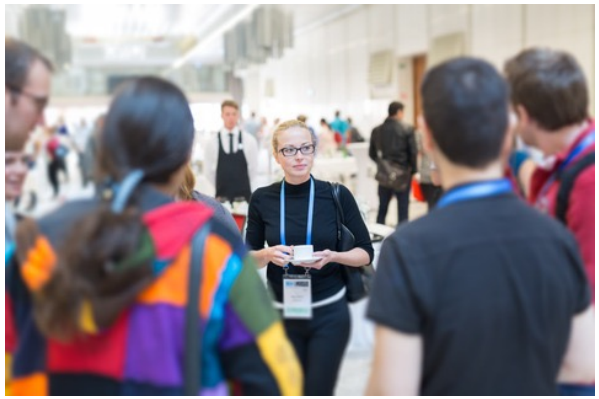
Our results so far

ML Approach	Important settings	Libraries	Accuracy	Remark
BalancedRandomForestClassifier	n_estimators=100	https://pypi.org/project/imbalanced-learn/	~ 88%	Predict two stages of change over
Decision Tree (Ensemble; RUS Boosted)	N/A	imbalanced-learn (https://imbalanced-learn.org/stable/)	~ 88%	Predict two stages of change over
Support Vector Machine (SVC)	kernel='rbf', C=100000	sklearn.svm	~ 82%	Predict five stages of change over
Neural network	Hidden Layer = 2 Activation function = „relu“	Tensorflow Keras	~ 83%	Predict two stages of change over

⇨ Can you beat our results? Reward: a crate of German beer!

Source: Mejbah, Ahammad (2020). Machine Learning Successful Project Plan. Online: https://www.kaggle.com/getting-started/181153#1002081 (retrieved: 19.02.2021).

CRISP-DM



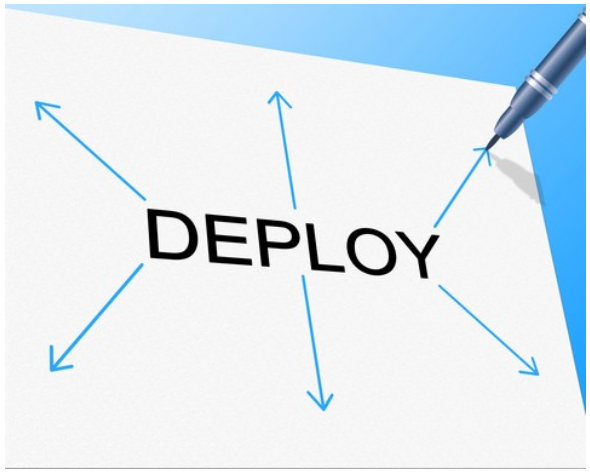
Discussion

Please share your suggestions, results, ideas etc. on the following mindmap on a Miro board:

https://miro.com/app/board/o9J_lC7tl-Y=/

Image source: Colourbox educational license

CRISP-DM: Deployment



Deployment

- Transition from development to production seamlessly
- Path-to-production methodology
- Servers, hardware, software
- Put in place for regular monitoring and maintenance of models to continuously evaluate their performance, check for results and their validity, and retire, replace, and update

Source: Mejbah, Ahammad (2020). Machine Learning Successful Project Plan. Online: https://www.kaggle.com/getting-started/181153#1002081 (retrieved: 19.02.2021).

Minute paper

Please take 10min to write a minute paper!

Come back into Zoom session when you‘re finished