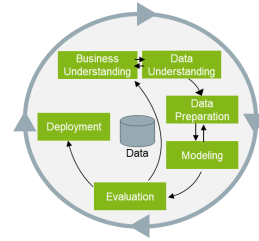


Three Ways to make your Industrial Data Science Projects a Success

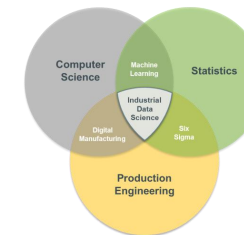
Prof. Dr.-Ing. Jochen Deuse
IDS 2019

- **Defining the Process**



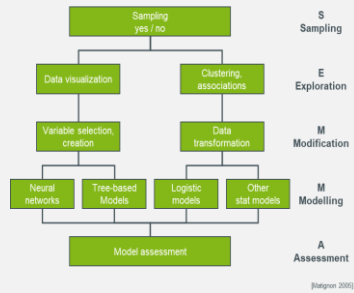
- **Dealing with Data Immaturity**

- **Combining Domain Knowledge with Data Science**

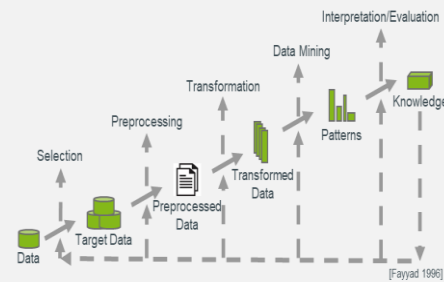


- **Conclusion**

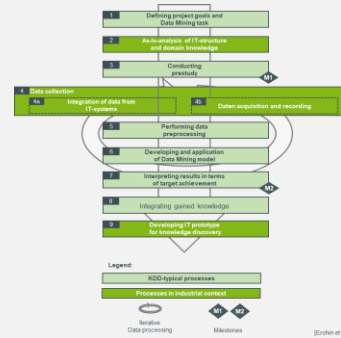
SEMMA of SAS



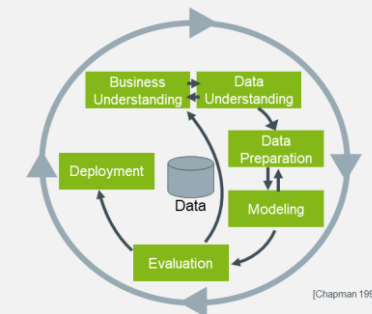
Knowledge Discovery in Databases (KDD)



Knowledge Discovery in Industrial Databases (KDID)



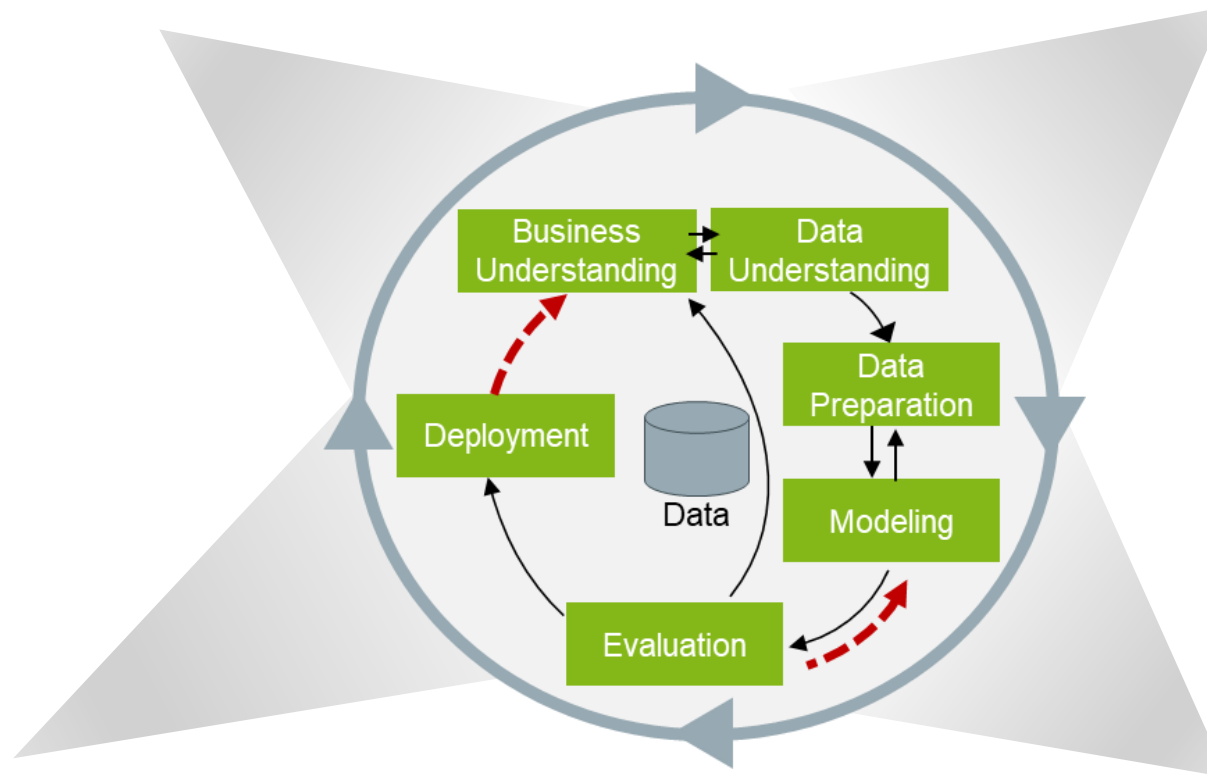
Cross-Industry Standard Process for Data Mining (CRISP-DM)



So why do we follow CRISP-DM?

- From a domain expert's perspective, the process is very intuitive

- It resembles a PDCA-respectively a DMAIC-circle



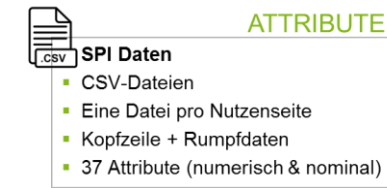
- It provides a well defined project structure

- It can easily be adapted across different industries

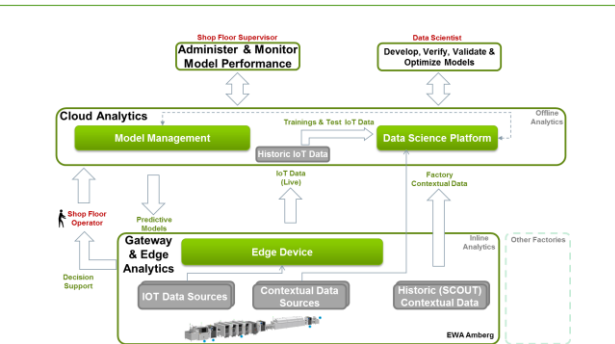
CRISP-DM provides a well defined Project Structure



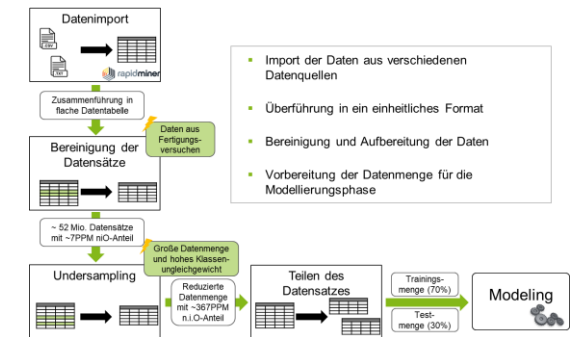
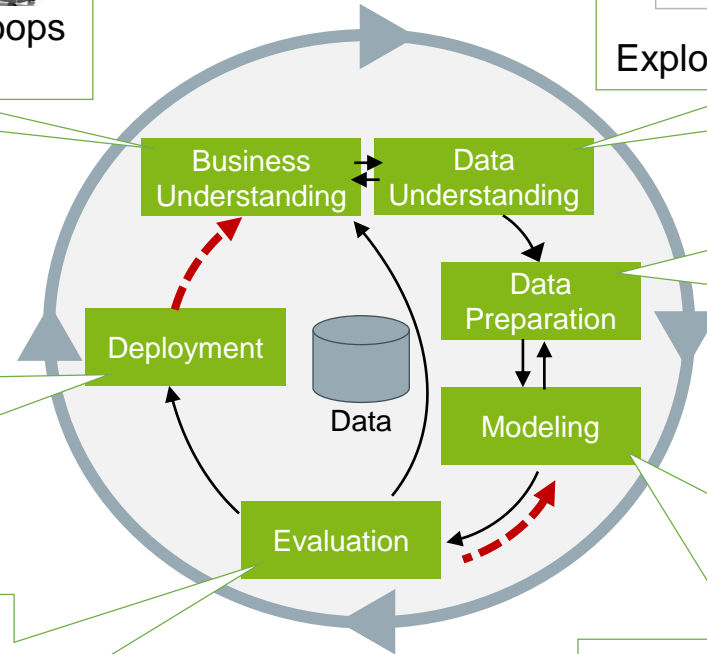
SMD-Value Stream: Shortening quality control loops and reducing the need for X-Ray inspection



Exploring process and inspection data



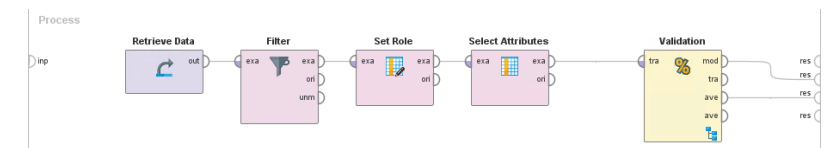
Deploying based on IoT-Architecture



Aggregating and cleansing of data

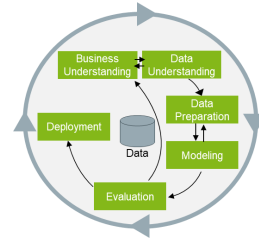
		Wahres Ergebnis			
Prognose	n.i.O. (Positive)	107	110.664	Precision 0,10 %	Pseudofehlerrate 99,90 %
	I.O. (Negative)	3	189.329	False Omission Rate 0,00 %	Negative Predictive Value 100 %
Sensitivity / Recall		False Positive Rate		Accuracy	
97,27 %		36,89 %		63,12 %	
Schlupfrate 2,73 %		Specificity			
		63,11 %			

Optimising slack rate and pseudo faults



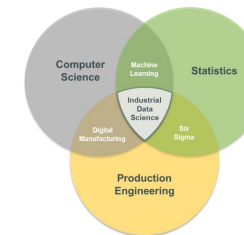
Selecting and configuring suitable prediction models

- Defining the Process



- Dealing with Data Immaturity

- Combining Domain Knowledge with Data Science



- Conclusion

- Data Acquisition – How is data collected along the value stream?
- Sample Size – Are there enough representatives of each class and are they evenly distributed?
- Reference Level – Is the data available in a high and uniform granularity?
- Consistency – Does the relevant data set contain logical contradictions?
- Traceability – Can label and feature value characteristics be joined unambiguously?
- ...

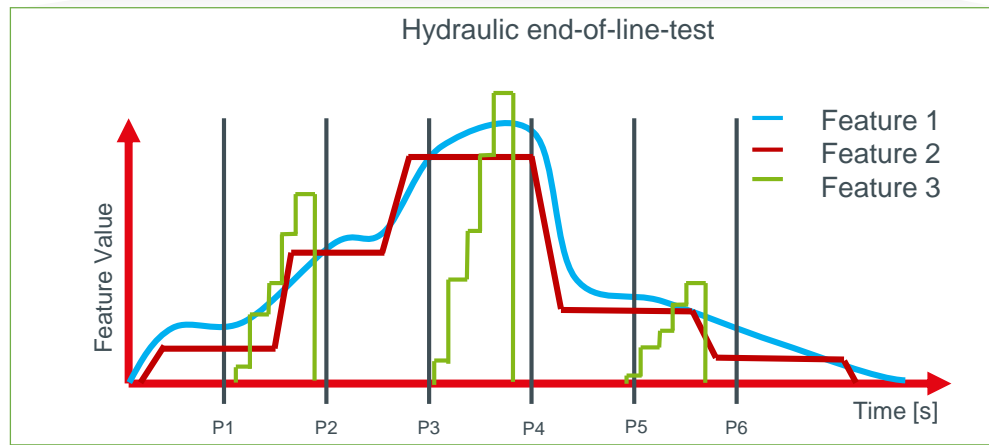
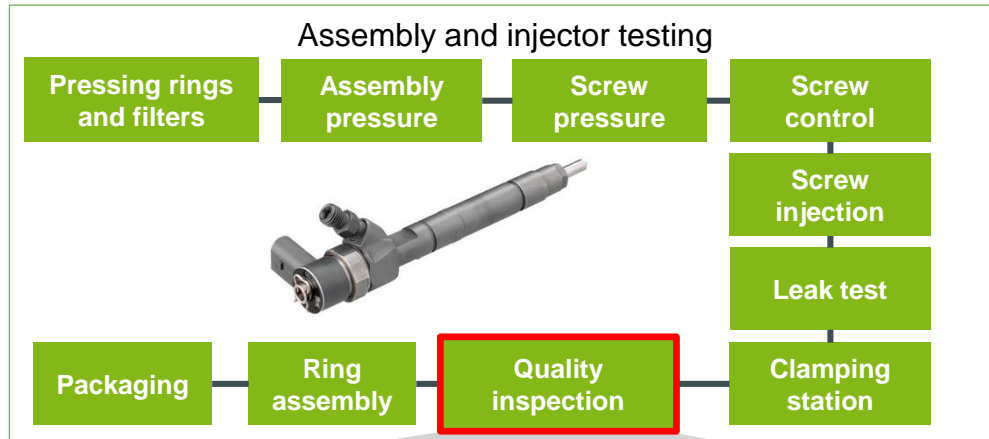
We have specified ten Criteria and four Levels of Maturity each

Criteria	Maturity level			
	1	2	3	4
Data collection	manual entry	electronical, must be triggered manually	data acquisition is carried out automatically in most cases	fully automated data collection
Completeness of data collection	unilateral and incomplete recording of relevant characteristics	recording of the essential characteristics	recording of a large part of the relevant characteristics	recording of all relevant, (un)influenceable characteristics
Sample size	no historic data	small sample per object group	large sample per object group, but unbalanced data	large sample with large number per object group and class
Data sources	paperbased records	decentralised data storage with simple software (e.g. Excel)	different data management systems with central data storage	comprehensive Data Warehouse
Data format	formats that are difficult to process (e.g. scans, photos)	formats with limited processability (e.g. PDF)	different, directly processable formats (e.g. CSV, XML)	comprehensive standard format
Data structure	unstructured text or images	semi-structured data (e.g. XML, JSON)	structured, mixed-scaled data	structured, metrically scaled data and standardized codes
Feature type	only set points	highly aggregated actual values	aggregated actual values or raw data with low sampling rate	raw data in real time
Reference level	value characteristics at the highest reference level	value characteristics at the upper reference level	value characteristics at the next higher level	value characteristics at individual element level
Consistency of data	no consistency/integrity	massive amount of logical differences	few logical differences	full integrity/consistency
Traceability	no ID/ time stamp	different ID/ timestamp	comprehensive ID/ time stamp	comprehensive ID/ timestamp on same reference level

Non uniform Reference Levels prohibit Supervised Learning

Criteria	Maturity level			
	1	2	3	4
Data collection	manual entry	electronical, must be triggered manually	data acquisition is carried out automatically in most cases	fully automated data collection
Completeness of data collection	unilateral and incomplete recording of relevant characteristics	recording of the essential characteristics	recording of a large part of the relevant characteristics	recording of all relevant, (un)influenceable characteristics
Sample size	no historic data	small sample per object group	large sample per object group, but unbalanced data	large sample with large number per object group and class
Data sources	paperbased records	decentralised data storage with simple software (e.g. Excel)	different data management systems with central data storage	comprehensive Data Warehouse
Data format	formats that are difficult to process (e.g. scans, photos)	formats with limited processability (e.g. PDF)	different, directly processable formats (e.g. CSV, XML)	comprehensive standard format
Data structure	unstructured text or images	semi-structured data (e.g. XML, JSON)	structured, mixed-scaled data	structured, metrically scaled data and standardized codes
Feature type	only set points	highly aggregated actual values	aggregated actual values or raw data with low sampling rate	raw data in real time
Reference level	value characteristics at the highest reference level	value characteristics at the upper reference level	value characteristics at the next higher level	value characteristics at individual element level
Consistency of data	no consistency/integrity	massive amount of logical differences	few logical differences	full integrity/consistency
Traceability	no ID/ time stamp	different ID/ timestamp	comprehensive ID/ time stamp	comprehensive ID/ timestamp on same reference level

■ Diesel Injector Nozzle Manufacturing Value Stream

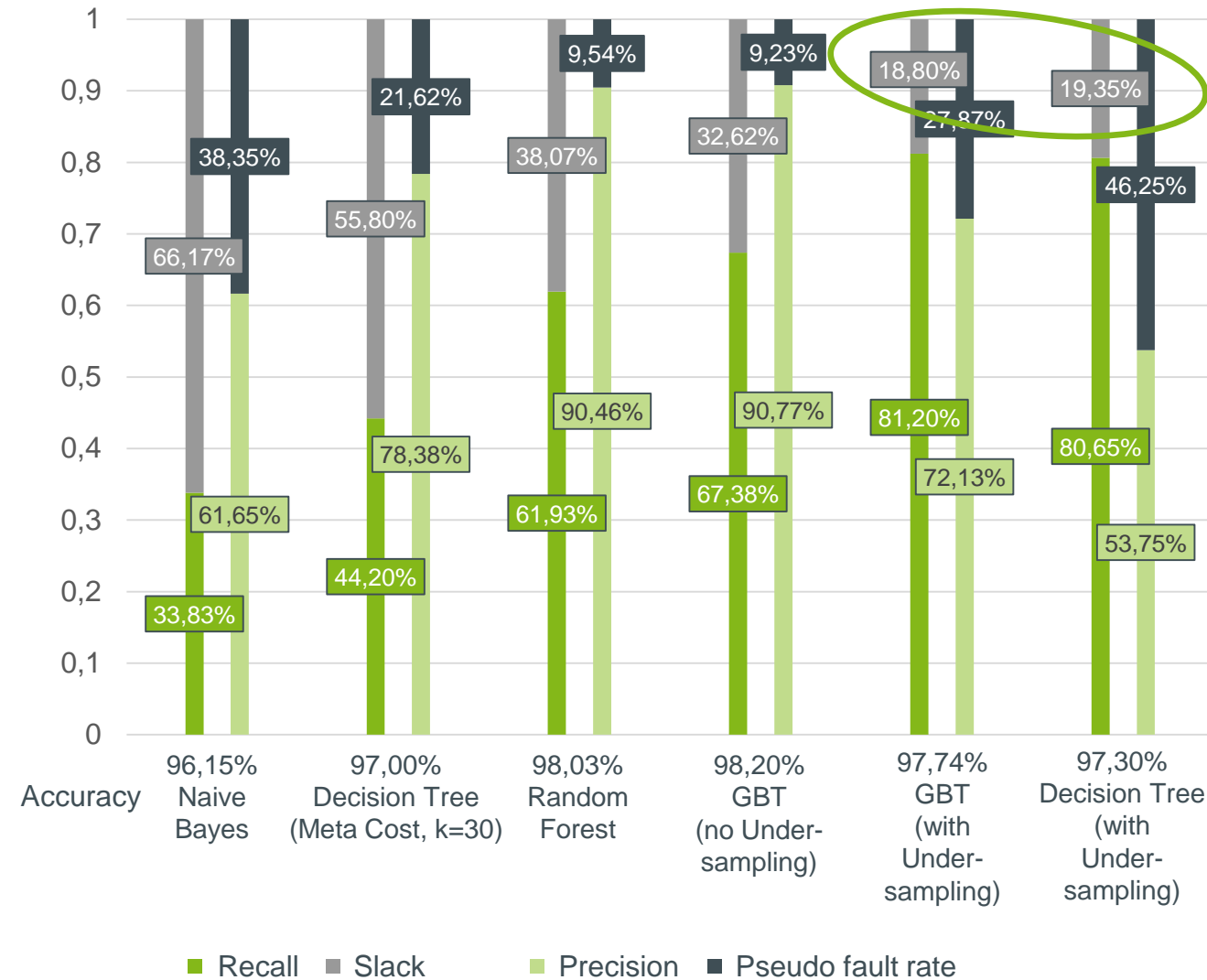
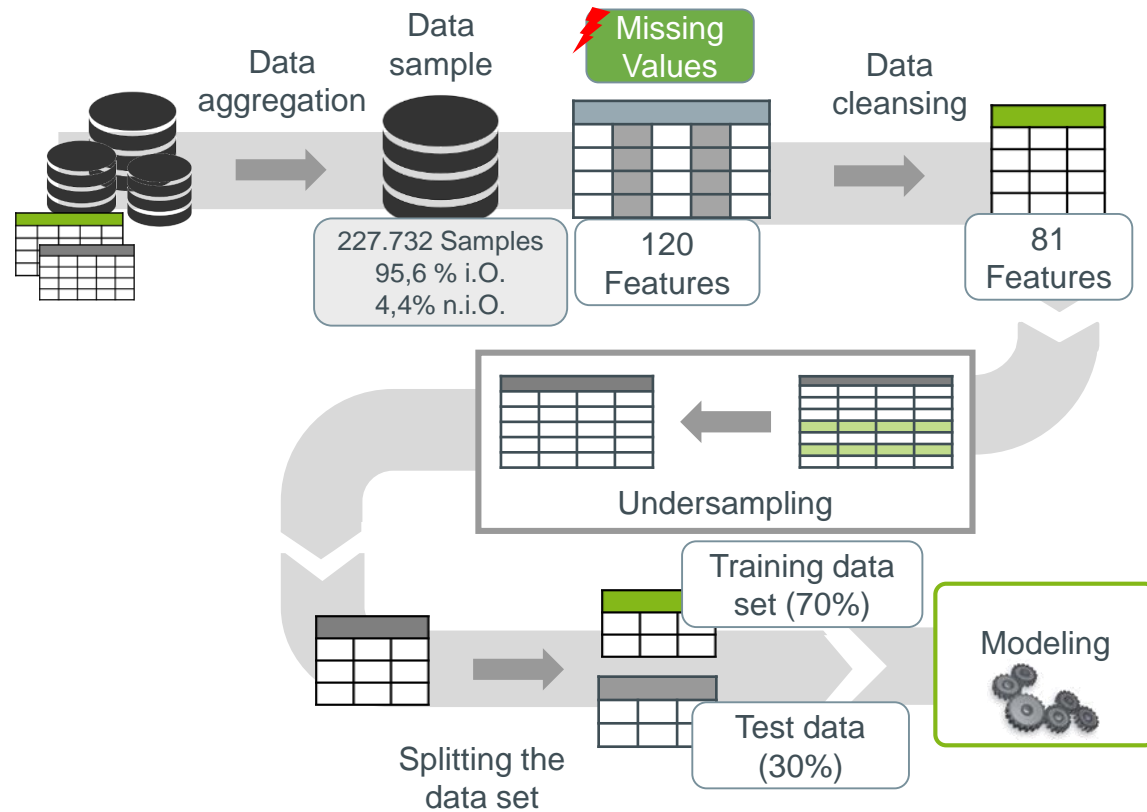


		True result			
		NOK 3.018 (4,42 %)	OK 65.302 (95,58 %)		
Forecast	NOK 1.656 (2,42 %)	1.021	635 Pseudo faults	Precision 61,65 %	Pseudo fault rate 38,35 %
	OK 66.664 (97,58 %)	1.997 Slack	64.667	False omission rate 3,00 %	Negative predictive value 97,00 %
		Sensitivity / Recall 33,83 %	False Positive Rate 0,97 %	Accuracy 96,15 %	
		Slack rate 66,17 %	Specificity 99,03 %		

Unbalanced Label Proportions result in high Recall Rates

Criteria	Maturity level			
	1	2	3	4
Data collection	manual entry	electronical, must be triggered manually	data acquisition is carried out automatically in most cases	fully automated data collection
Completeness of data collection	unilateral and incomplete recording of relevant characteristics	recording of the essential characteristics	recording of a large part of the relevant characteristics	recording of all relevant, (un)influenceable characteristics
Sample size	no historic data	small sample per object group	large sample per object group, but unbalanced data	large sample with large number per object group and class
Data sources	paperbased records	decentralised data storage with simple software (e.g. Excel)	different data management systems with central data storage	comprehensive Data Warehouse
Data format	formats that are difficult to process (e.g. scans, photos)	formats with limited processability (e.g. PDF)	different, directly processable formats (e.g. CSV, XML)	comprehensive standard format
Data structure	unstructured text or images	semi-structured data (e.g. XML, JSON)	structured, mixed-scaled data	structured, metrically scaled data and standardized codes
Feature type	only set points	highly aggregated actual values	aggregated actual values or raw data with low sampling rate	raw data in real time
Reference level	value characteristics at the highest reference level	value characteristics at the upper reference level	value characteristics at the next higher level	value characteristics at individual element level
Consistency of data	no consistency/integrity	massive amount of logical differences	few logical differences	full integrity/consistency
Traceability	no ID/ time stamp	different ID/ timestamp	comprehensive ID/ time stamp	comprehensive ID/ timestamp on same reference level

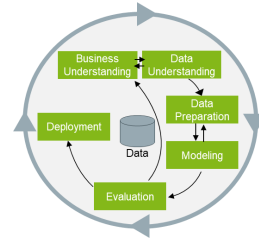
Undersampling reduces the Effect of unbalanced Label Proportions



A Lack of Traceability prohibits Supervised Learning

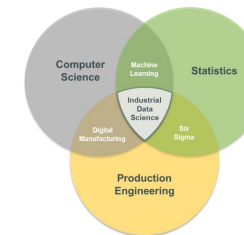
Criteria	Maturity level			
	1	2	3	4
Data collection	manual entry	electronical, must be triggered manually	data acquisition is carried out automatically in most cases	fully automated data collection
Completeness of data collection	unilateral and incomplete recording of relevant characteristics	recording of the essential characteristics	recording of a large part of the relevant characteristics	recording of all relevant, (un)influenceable characteristics
Sample size	no historic data	small sample per object group	large sample per object group, but unbalanced data	large sample with large number per object group and class
Data sources	paperbased records	decentralised data storage with simple software (e.g. Excel)	different data management systems with central data storage	comprehensive Data Warehouse
Data format	formats that are difficult to process (e.g. scans, photos)	formats with limited processability (e.g. PDF)	different, directly processable formats (e.g. CSV, XML)	comprehensive standard format
Data structure	unstructured text or images	semi-structured data (e.g. XML, JSON)	structured, mixed-scaled data	structured, metrically scaled data and standardized codes
Feature type	only set points	highly aggregated actual values	aggregated actual values or raw data with low sampling rate	raw data in real time
Reference level	value characteristics at the highest reference level	value characteristics at the upper reference level	value characteristics at the next higher level	value characteristics at individual element level
Consistency of data	no consistency/integrity	massive amount of logical differences	few logical differences	full integrity/consistency
Traceability	no ID/ time stamp	different ID/ timestamp	comprehensive ID/ time stamp	comprehensive ID/ timestamp on same reference level

- Defining the Process



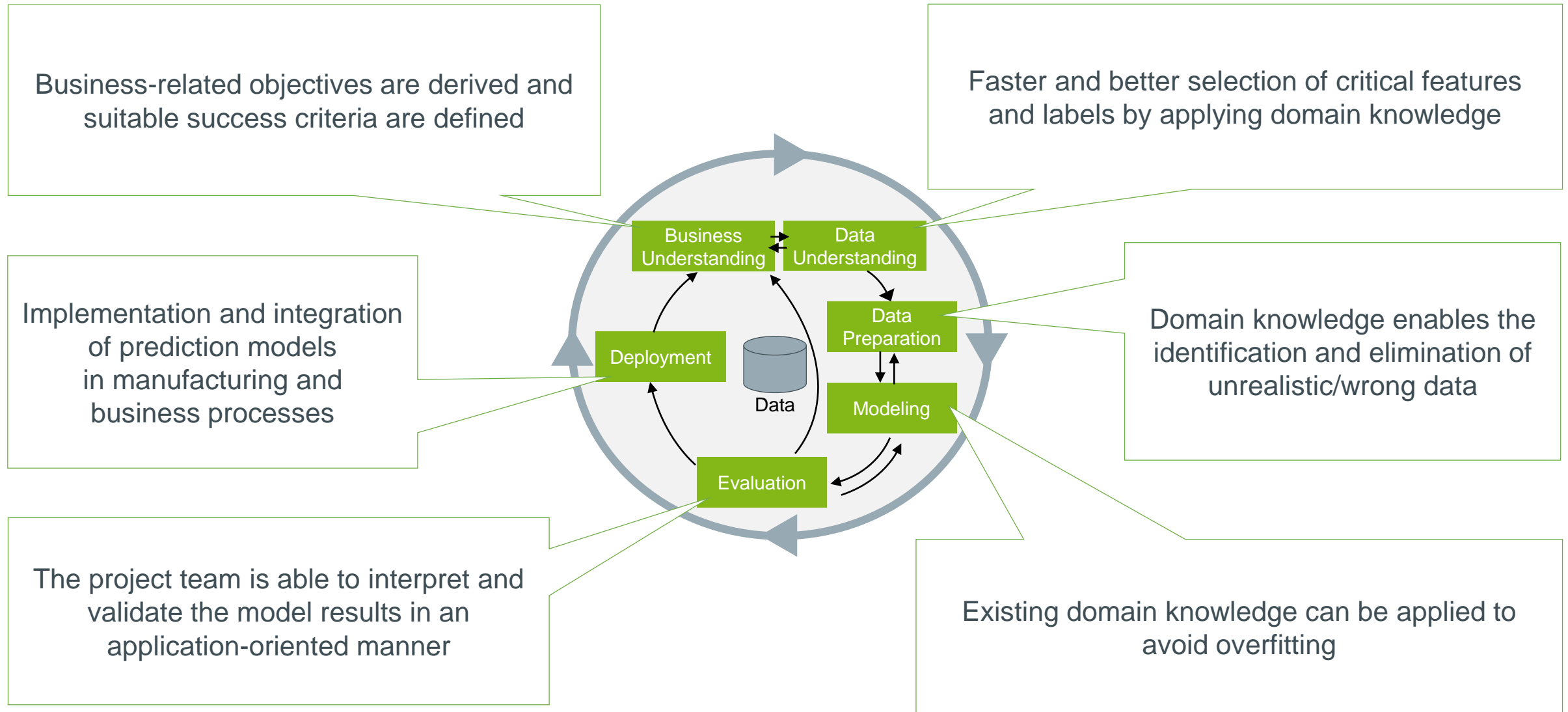
- Dealing with Data Immaturity

- Combining Domain Knowledge with Data Science



- Conclusion

Domain Knowledge is required in every Stage of CRISP-DM



SFB 876 - B3: Data Mining on Sensor Data of Automated Processes

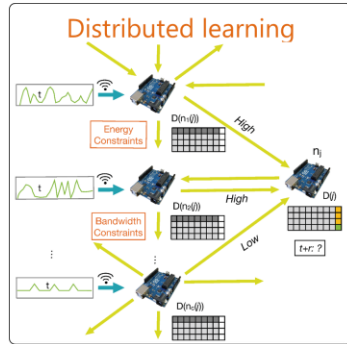


Prof. Dr. Katharina Morik,
Artificial Intelligence



Prof. Dr.-Ing. Jochen Deuse,
Production Systems

Distributed data analysis

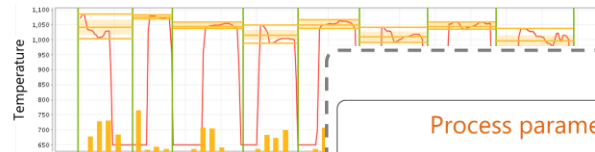


Learning from Label Proportions (LLP)

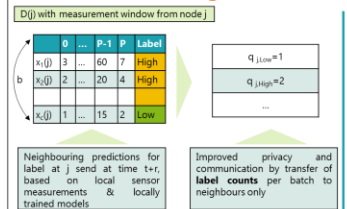
Step 1: Cluster Samples

Sensor data processing for quality prediction

Aggregation and feature extraction from time series

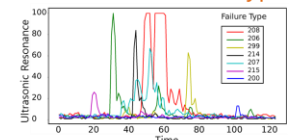


Training of Local Models



Distributed Learning of Local Models

Prediction of failure types

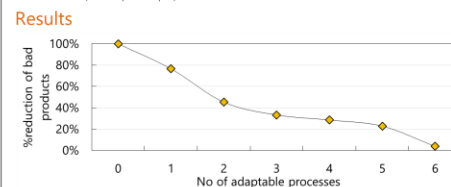
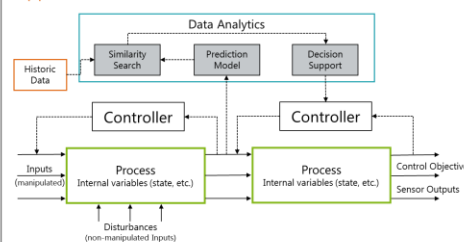


Ultrasonic spectra for quality testing

Method	SVM RBF	Random Forest
Raw time series (TS)	50,10%	50,11%
Extracted features	77,04%	77,22%
Meta-features + raw TS	70,16%	81,41%

Process parameter adaption

Application framework



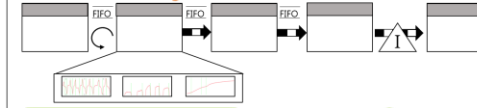
► Real-time process parameter adaption utilising quality prediction models and similarity-search algorithms

[Schmitt, Deuse, IEEM 2018]

Production process control strategies

Adaptive process control

Process Monitoring



Quality Prediction



Process Control



Decision Support



Decision

- Continue processing
- Adapt process parameters
- Reassign order
- Reject product

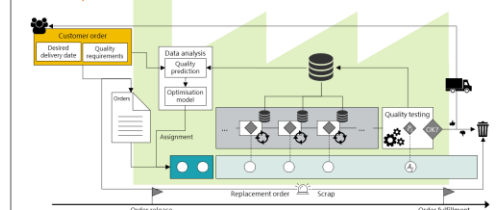
Constrained optimisation model
 $\min f(x)$ s.t. $g(x) \leq 0, h(x) = 0$

► Integrated adaptive process control framework for optimal decision support

[Wiegand, Stolpe, Deuse, Morik, AT 2016]

Customer-requirement-oriented process control

Concept



Realisation

(PCO-AP)

$$\begin{aligned} \min & \sum_{i \in I} \sum_{j \in J} c_{ij} x_{ij} \\ \text{s.t.} & \sum_{j \in J} x_{ij} = 1 \quad \text{f.a. } i \in I \\ & \sum_{i \in I} x_{ij} = 1 \quad \text{f.a. } j \in J \\ & x_{ij} \geq 0 \quad \text{f.a. } i \in I, j \in J \end{aligned}$$

► Assignment of intermediate products to customer orders based on quality predictions

[Schmitt, Wiegand, Deuse, ZWF 2018] [Schmitt, Hahn, Deuse, TecRep 2018]





Citizen data scientists



Domain experts



rapidminer
Data scientists

Prediction of...



...malt processability



...lautering duration



...yeast processing yield





TECHNISCHE
UNIVERSITÄT
DARMSTADT

Prof. Kristian Kersting:

- Deep Learning
- Tree-based procedures
- Ensemble strategies



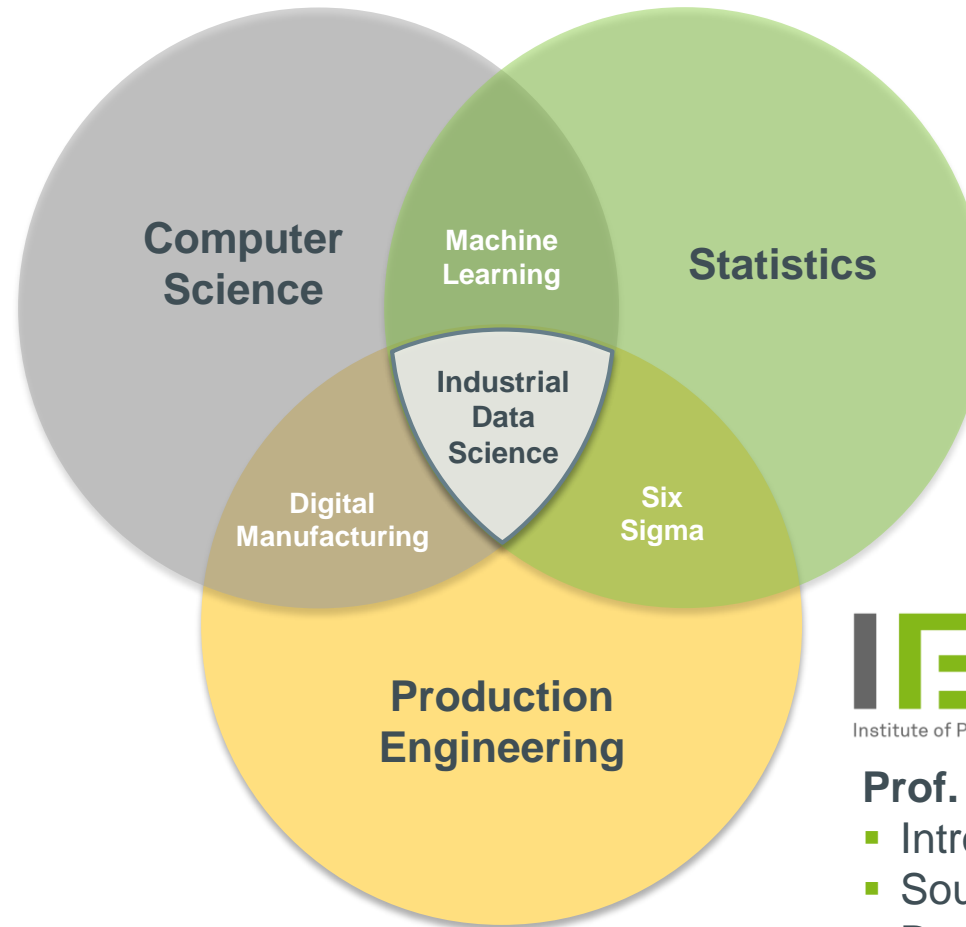
Prof. Claus Weihs:

- Association Analysis
- Data Transformations
- Concepts for Model Selection



Prof. Jens Teubner:

- Basics of Data Management
- Database Systems
- Data Warehouses



Prof. Jochen Deuse:

- Introduction to Industrial Data Science
- Sources of industrial data
- Data analysis in industrial environments

Project coaching



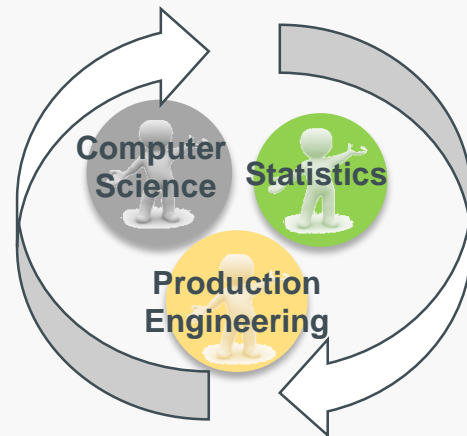
TECHNISCHE
UNIVERSITÄT
DARMSTADT

fi fakultät für
informatik

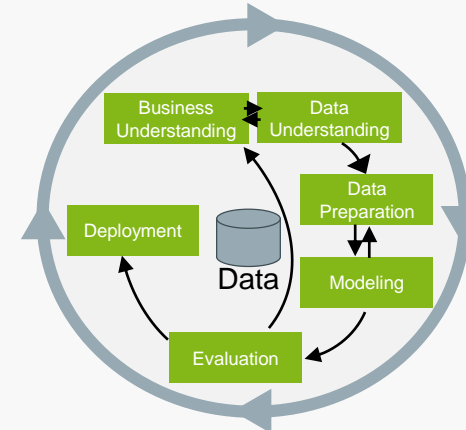
**fakultät
statistik**

IPS
Institute of Production Systems

Use case driven competence development



Interdisciplinary
project teams



Common
project structure

Industrial use cases

Weidmüller

Quality prediction
for injection molding

EVONIK
KRAFT FÜR NEUES

Yield prediction
for chemical processes

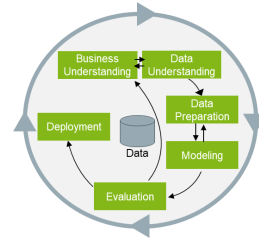
ebm papst

Quality prediction
for fan assembly

BMW GROUP

Quality prediction
for engine assembly

- Defining the Process

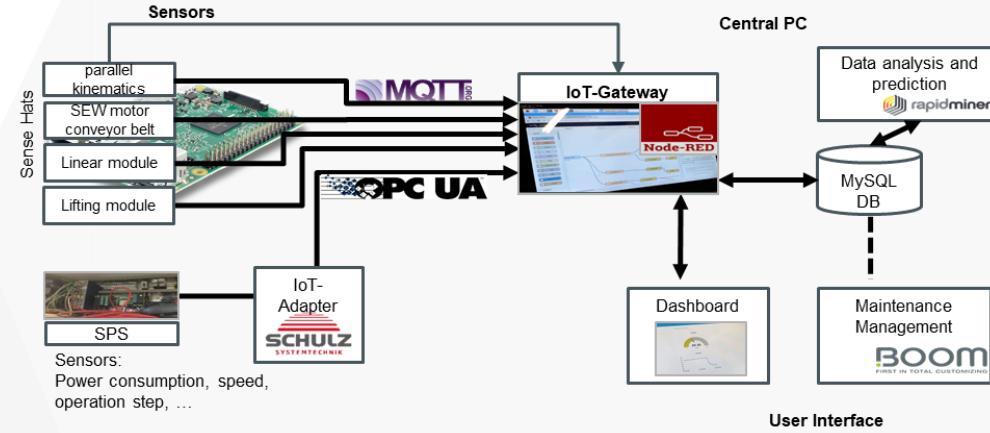


- Dealing with Data Immaturity

Production Data			
Time	Machine	Operator	Value
00:00	M1	O1	10
00:01	M1	O1	12
00:02	M1	O1	15
00:03	M1	O1	18
00:04	M1	O1	20
00:05	M1	O1	22
00:06	M1	O1	25
00:07	M1	O1	28
00:08	M1	O1	30
00:09	M1	O1	32
00:10	M1	O1	35
00:11	M1	O1	38
00:12	M1	O1	40
00:13	M1	O1	42
00:14	M1	O1	45
00:15	M1	O1	48
00:16	M1	O1	50
00:17	M1	O1	52
00:18	M1	O1	55
00:19	M1	O1	58
00:20	M1	O1	60
00:21	M1	O1	62
00:22	M1	O1	65
00:23	M1	O1	68
00:24	M1	O1	70
00:25	M1	O1	72
00:26	M1	O1	75
00:27	M1	O1	78
00:28	M1	O1	80
00:29	M1	O1	82
00:30	M1	O1	85
00:31	M1	O1	88
00:32	M1	O1	90
00:33	M1	O1	92
00:34	M1	O1	95
00:35	M1	O1	98
00:36	M1	O1	100
00:37	M1	O1	102
00:38	M1	O1	105
00:39	M1	O1	108
00:40	M1	O1	110
00:41	M1	O1	112
00:42	M1	O1	115
00:43	M1	O1	118
00:44	M1	O1	120
00:45	M1	O1	122
00:46	M1	O1	125
00:47	M1	O1	128
00:48	M1	O1	130
00:49	M1	O1	132
00:50	M1	O1	135
00:51	M1	O1	138
00:52	M1	O1	140
00:53	M1	O1	142
00:54	M1	O1	145
00:55	M1	O1	148
00:56	M1	O1	150
00:57	M1	O1	152
00:58	M1	O1	155
00:59	M1	O1	158
01:00	M1	O1	160
01:01	M1	O1	162
01:02	M1	O1	165
01:03	M1	O1	168
01:04	M1	O1	170
01:05	M1	O1	172
01:06	M1	O1	175
01:07	M1	O1	178
01:08	M1	O1	180
01:09	M1	O1	182
01:10	M1	O1	185
01:11	M1	O1	188
01:12	M1	O1	190
01:13	M1	O1	192
01:14	M1	O1	195
01:15	M1	O1	198
01:16	M1	O1	200
01:17	M1	O1	202
01:18	M1	O1	205
01:19	M1	O1	208
01:20	M1	O1	210
01:21	M1	O1	212
01:22	M1	O1	215
01:23	M1	O1	218
01:24	M1	O1	220
01:25	M1	O1	222
01:26	M1	O1	225
01:27	M1	O1	228
01:28	M1	O1	230
01:29	M1	O1	232
01:30	M1	O1	235
01:31	M1	O1	238
01:32	M1	O1	240
01:33	M1	O1	242
01:34	M1	O1	245
01:35	M1	O1	248
01:36	M1	O1	250
01:37	M1	O1	252
01:38	M1	O1	255
01:39	M1	O1	258
01:40	M1	O1	260
01:41	M1	O1	262
01:42	M1	O1	265
01:43	M1	O1	268
01:44	M1	O1	270
01:45	M1	O1	272
01:46	M1	O1	275
01:47	M1	O1	278
01:48	M1	O1	280
01:49	M1	O1	282
01:50	M1	O1	285
01:51	M1	O1	288
01:52	M1	O1	290
01:53	M1	O1	292
01:54	M1	O1	295
01:55	M1	O1	298
01:56	M1	O1	300
01:57	M1	O1	302
01:58	M1	O1	305
01:59	M1	O1	308
02:00	M1	O1	310
02:01	M1	O1	312
02:02	M1	O1	315
02:03	M1	O1	318
02:04	M1	O1	320
02:05	M1	O1	322
02:06	M1	O1	325
02:07	M1	O1	328
02:08	M1	O1	330
02:09	M1	O1	332
02:10	M1	O1	335
02:11	M1	O1	338
02:12	M1	O1	340
02:13	M1	O1	342
02:14	M1	O1	345
02:15	M1	O1	348
02:16	M1	O1	350
02:17	M1	O1	352
02:18	M1	O1	355
02:19	M1	O1	358
02:20	M1	O1	360
02:21	M1	O1	362
02:22	M1	O1	365
02:23	M1	O1	368
02:24	M1	O1	370
02:25	M1	O1	372
02:26	M1	O1	375
02:27	M1	O1	378
02:28	M1	O1	380
02:29	M1	O1	382
02:30	M1	O1	385
02:31	M1	O1	388
02:32	M1	O1	390
02:33	M1	O1	392
02:34	M1	O1	395
02:35	M1	O1	398
02:36	M1	O1	400
02:37	M1	O1	402
02:38	M1	O1	405
02:39	M1	O1	408
02:40	M1	O1	410
02:41	M1	O1	412
02:42	M1	O1	415
02:43	M1	O1	418
02:44	M1	O1	420
02:45	M1	O1	422
02:46	M1	O1	425
02:47	M1	O1	428
02:48	M1	O1	430
02:49	M1	O1	432
02:50	M1	O1	435
02:51	M1	O1	438
02:52	M1	O1	440
02:53	M1	O1	442
02:54	M1	O1	445
02:55	M1	O1	448
02:56	M1	O1	450
02:57	M1	O1	452
02:58	M1	O1	455
02:59	M1	O1	458
03:00	M1	O1	460
03:01	M1	O1	462
03:02	M1	O1	465
03:03	M1	O1	468
03:04	M1	O1	470
03:05	M1	O1	472
03:06	M1	O1	475
03:07	M1	O1	478
03:08	M1	O1	480
03:09	M1	O1	482
03:10	M1	O1	485
03:11	M1	O1	488
03:12	M1	O1	490
03:13	M1	O1	492
03:14	M1	O1	495
03:15	M1	O1	498
03:16	M1	O1	500
03:17	M1	O1	502
03:18	M1	O1	505
03:19	M1	O1	508
03:20	M1	O1	510
03:21	M1	O1	512
03:22	M1	O1	515
03:23	M1	O1	518
03:24	M1	O1	520
03:25	M1	O1	522
03:26	M1	O1	525
03:27	M1	O1	528
03:28	M1	O1	530
03:29	M1	O1	532
03:30	M1	O1	535
03:31	M1	O1	538
03:32	M1	O1	540
03:33	M1	O1	542
03:34	M1	O1	545
03:35	M1	O1	548
03:36	M1	O1	550
03:37	M1	O1	552
03:38	M1	O1	555
03:39	M1	O1	558
03:40	M1	O1	560
03:41	M1	O1	562
03:42	M1	O1	565
03:43	M1	O1	568
03:44	M1	O1	570
03:45	M1	O1	572
03:46	M1	O1	575
03:47	M1	O1	578
03:48	M1	O1	580
03:49	M1	O1	582
03:50	M1	O1	585
03:51	M1	O1	588
03:52	M1	O1	590
03:53	M1	O1	592
03:54	M1	O1	595
03:55	M1	O1	598
03:56	M1	O1	600
03:57	M1	O1	602
03:58	M1	O1	605
03:59	M1	O1	608
04:00	M1	O1	610
04:01	M1	O1	612
04:02	M1	O1	615
04:03	M1	O1	618
04:04	M1	O1	620
04:05	M1	O1	622
04:06	M1	O1	625
04:07	M1	O1	628
04:08	M1	O1	630
04:09	M1	O1	632
04:10	M1	O1	635
04:11	M1	O1	638
04:12	M1	O1	640
04:13	M1	O1	642
04:14	M1	O1	645
04:15	M1	O1	648
04:16	M1	O1	650
04:17	M1	O1	652
04:18	M1	O1	655
04:19	M1	O1	658
04:20	M1	O1	660
04:21	M1	O1	662
04:22	M1	O1	665
04:23	M1	O1	668
04:24	M1	O1	670
04:25	M1	O1	672
04:26	M1	O1	675
04:27	M1	O1	678
04:28	M1	O1	680
04:29	M1	O1	682
04:30	M1	O1	685
04:31	M1	O1	688
04:32	M1	O1	690
04:33	M1	O1	692
04:34	M1	O1	695
04:35	M1	O1	698
04:36	M1	O1	700
04:37	M1	O1	702
04:38	M1	O1	705
04:39	M1	O1	708
04:40	M1	O1	710
04:41	M1	O1	712
04:42	M1	O1	715
04:43	M1	O1	718
04:44	M1	O1	720
04:45	M1	O1	722
04:46	M1	O1	725
04:47	M1	O1	728
04:48	M1	O1	730
04:49	M1	O1	732
04:50	M1	O1	735
04:51	M1	O1	738
04:52	M1	O1	740
04:53	M1	O1	742
04:54	M1	O1	745
04:55	M1	O1	748
04:56	M1	O1	750
04:57	M1	O1	752
04:58	M1	O1	755
04:59	M1	O1	758
05:00	M1	O1	760
05:01	M1	O1	762
05:02	M1	O1	765
05:03	M1	O1	768
05:04	M1	O1	770
05:05	M1	O1	772
05:06	M1	O1	775
05:07	M1	O1	778
05:08	M1	O1	780
05:09	M1	O1	782
05:10	M1	O1	785
05:11	M1	O1	788
05:12	M1	O1	790
05:13	M1	O1	792
05:14	M1	O1	795
05:15	M1	O1	798
05:16	M1	O1	800
05:17	M1	O1	802
05:18	M1	O1	805
05:19	M1	O1	808
05:20	M1	O1	810
05:21	M1	O1	812
05:22	M1	O1	815
05:23	M1	O1	818
05:24	M1	O1	820
05:25	M1	O1	822
05:26	M1	O1	825
05:27	M1	O1	828
05:28	M1	O1	830
05:29	M1	O1	832
05:30	M1	O1	835
05:31	M1	O1	838
05:32	M1	O1	840
05:33	M1	O1	842
05:34	M1	O1	845
05:35	M1	O1	848
05:36	M1	O1	850
05:37	M1	O1	852
05:38	M1	O1	855
05:39	M1	O1	858
05:40	M1	O1	860
05:41	M1	O1	862
05:42	M1	O1	865
05:43	M1	O1	868
05:44	M1	O1	870
05:45	M1	O1	872
05:46	M1	O1	875
05:47	M1	O1	878
05:48	M1	O1	880
05:49	M1	O1	882
05:50	M1	O1	885
05:51	M1	O1	888
05:52	M1	O1	890
05:53	M1	O1	892
05:54	M1	O1	895
05:55	M1	O1	898
05:56	M1	O1	900
05:57	M1	O1	902
05:58	M1	O1	905
05:59	M1	O1	908
06:00	M1	O1	910
06:01	M1	O1	912
06:02	M1	O1	915
06:0			

Different Approaches can overcome Data Maturity Challenges

Improving data maturity (e.g. retrofitting)



Criteria	Maturity Level			
	1	2	3	4
Data Collection	manual entry	electrical, must be triggered manually	data acquisition is carried out automatically in most cases	fully automated data collection
Completeness of data collection	unilateral and incomplete recording of relevant characteristics	recording of the essential characteristics	recording of a large part of the relevant characteristics, but unbalanced data	recording of all relevant, (un)influenceable characteristics
Sample Size	no historic data	small sample per object group	large sample per object group, but unbalanced data	large sample with large number per object group and class
Data Sources	paperbased records	decentralised data storage with simple software (e.g. Excel)	different data management systems with central data storage	comprehensive Data Warehouse
Data Format	formats that are difficult to process (e.g. scans, photos)	formats with limited processability (e.g. PDF)	different, directly processable formats (e.g. CSV, XML)	comprehensive standard format
Data Structure	unstructured text or images	Semi-structured data (e.g. XML, JSON)	structured, mixed-scaled data	structured, metrically scaled data and standardized codes
Feature Type	only set points	highly aggregated actual values	aggregated actual values or raw data with low sampling rate	raw data in real
Reference Level	attributes of the values at the highest reference level	attributes of the values at the upper reference level	attributes of the values at the next higher level	attributes of the values at individual element level
Consistency of data	No consistency/integrity	Massive amount of logical differences	Few logical differences	Full integrity/consistency
Traceability	No ID/ time stamp	Different ID/ timestamp	Comprehensive ID/ time stamp	Comprehensive ID/ timestamp on same reference level

Involving senior data scientists

