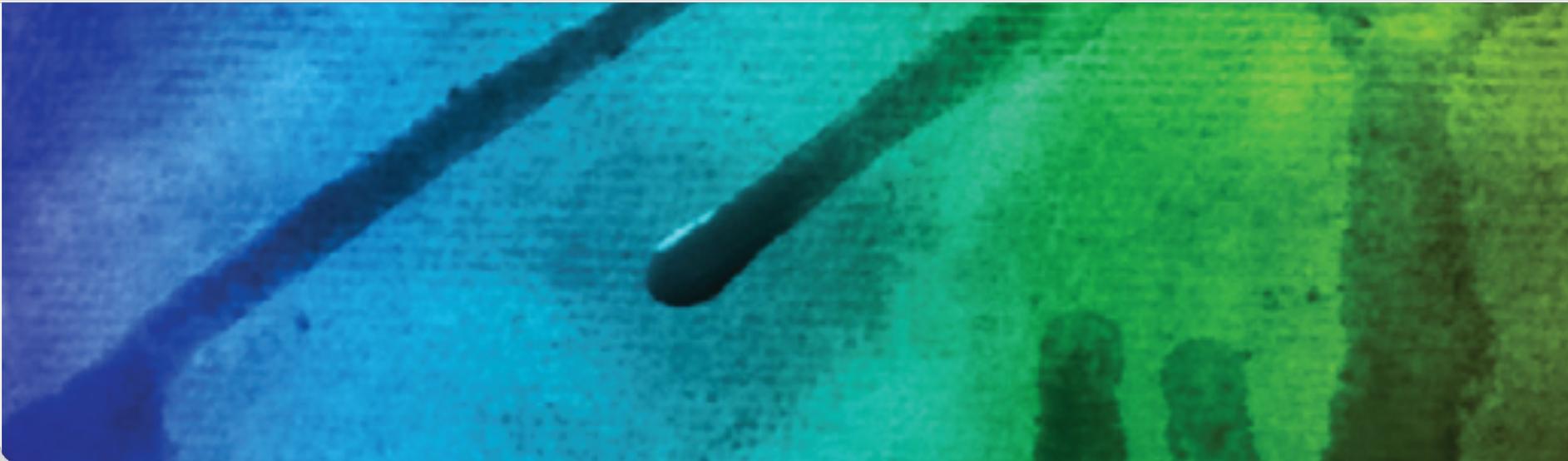


# Einführung in Machine Learning

Thematischer Einstieg und exemplarischer Projektablauf

parsQube GmbH – Markupforum 2018

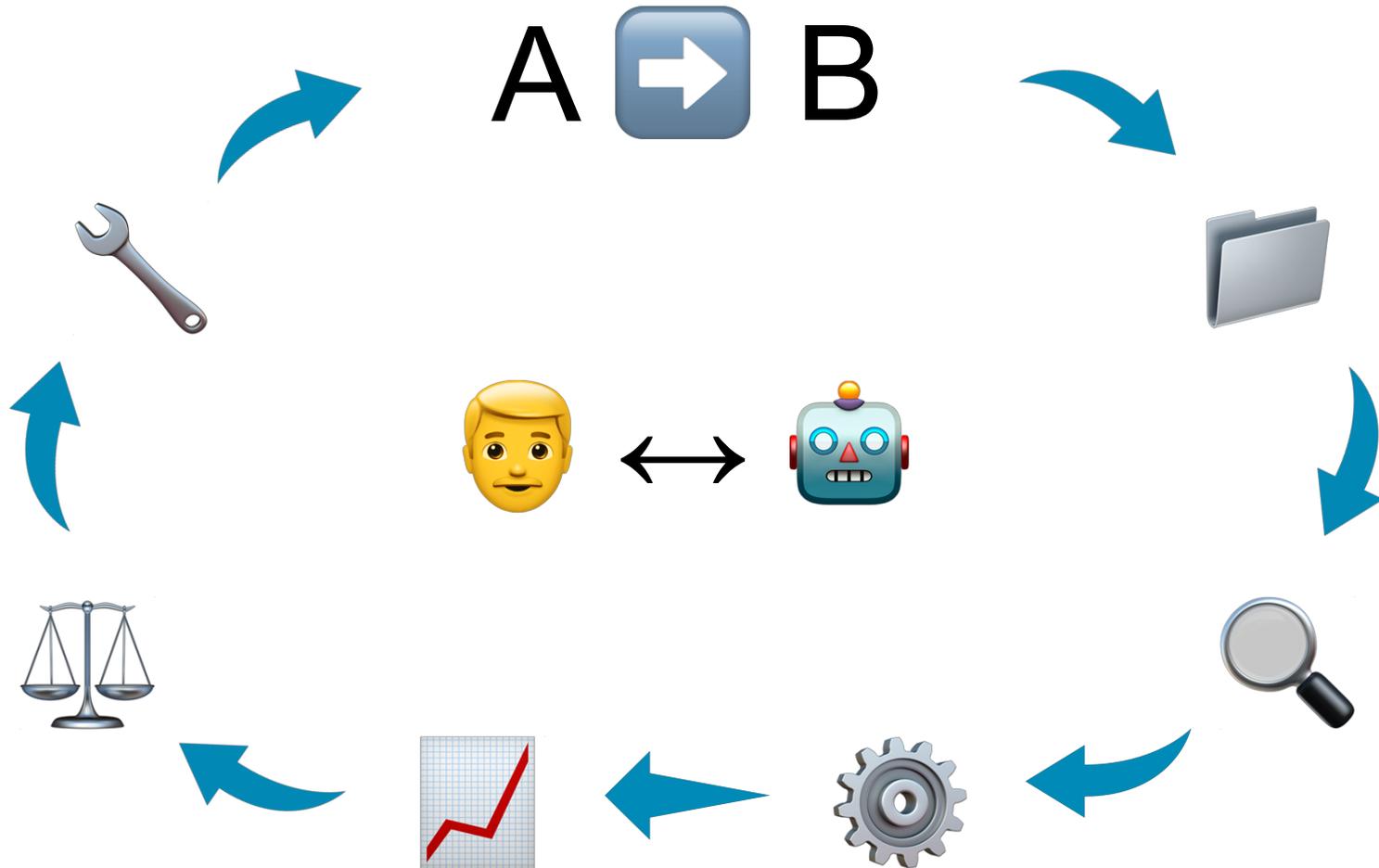
Simon Richebächer



# Agenda

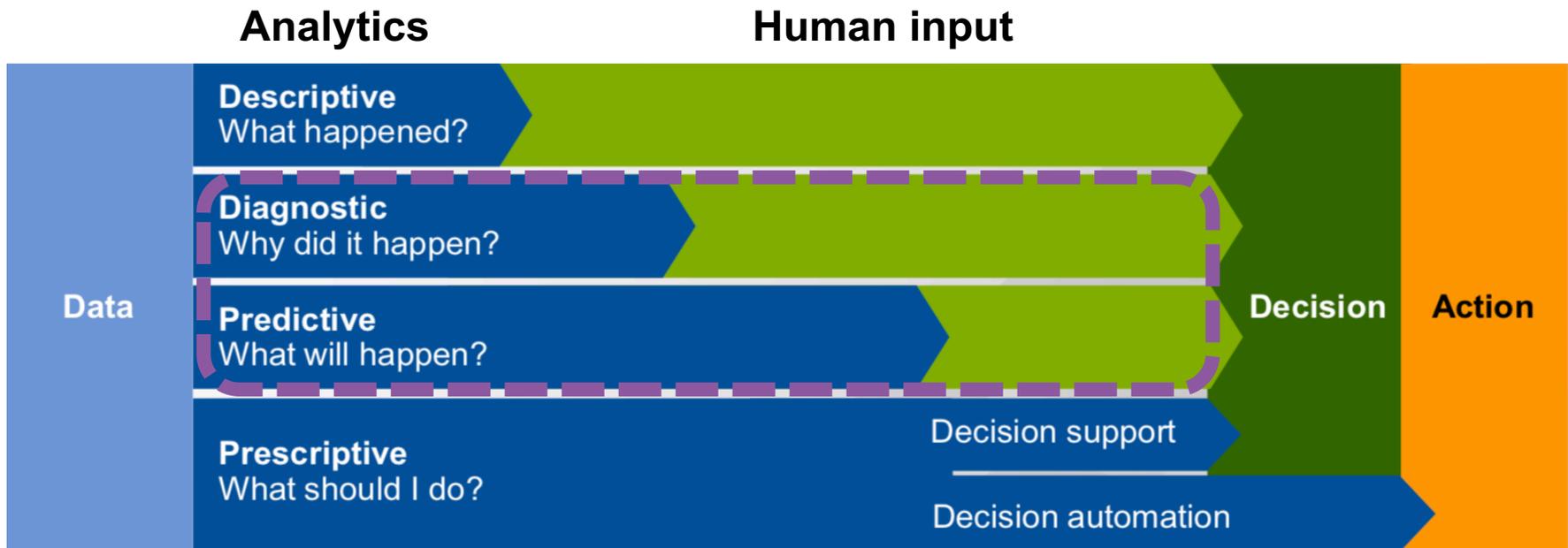
- Thematische Einführung Machine Learning
  - Einsatz und Bedeutung
  - Unterteilung und Arten
- Exemplarischer Projektablauf
  - Das Framework CRISP-DM
  - Beispielhafte Projektphasen
- Einblick in hilfreiche Tools
  - Kurze Live-Demo
- Zusammenfassung und Ausblick

# Thematische Einführung Machine Learning



# Thematische Einführung Machine Learning

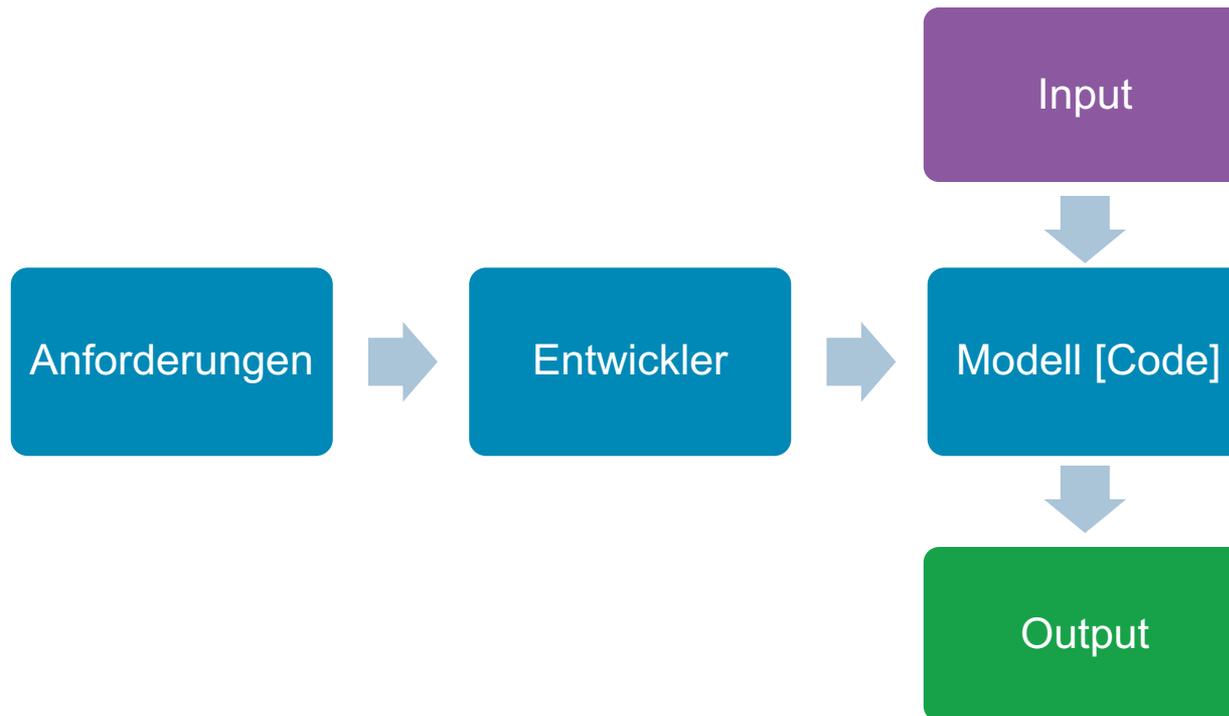
- Wo lässt sich Machine Learning (ML) einordnen?



Quelle: Gartner (September 2013)

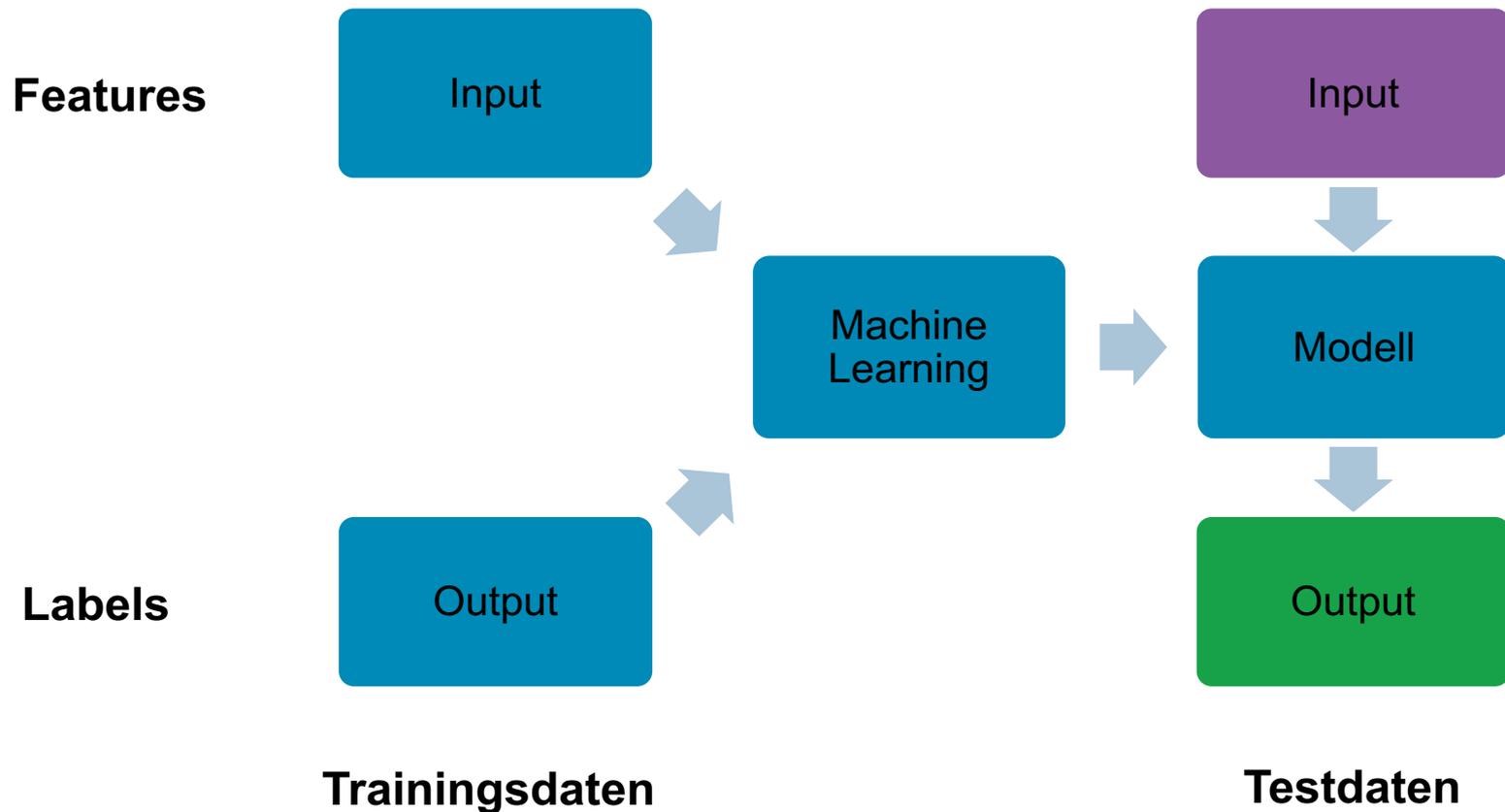
# Thematische Einführung Machine Learning

- Die „manuelle“ Entwicklung eines Modells:



# Thematische Einführung Machine Learning

- Mit dem Einsatz von Machine Learning:



# Thematische Einführung Machine Learning

- Verlust der Fähigkeit zur Generalisierung durch:

## Overfitting



## Underfitting



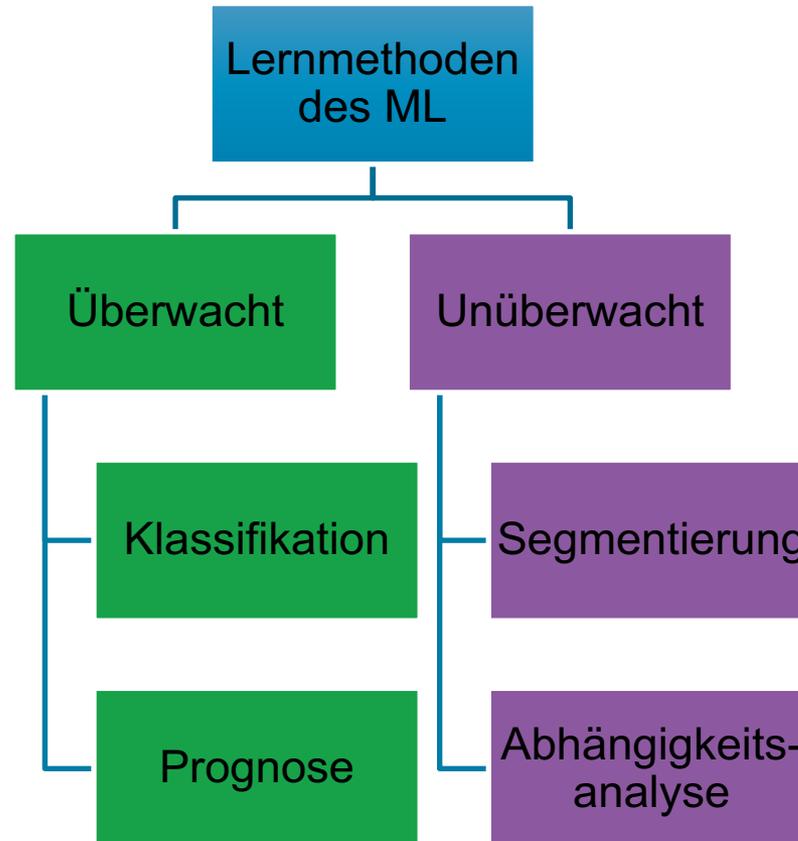
# Thematische Einführung Machine Learning

- Weitere Hürden im ML:



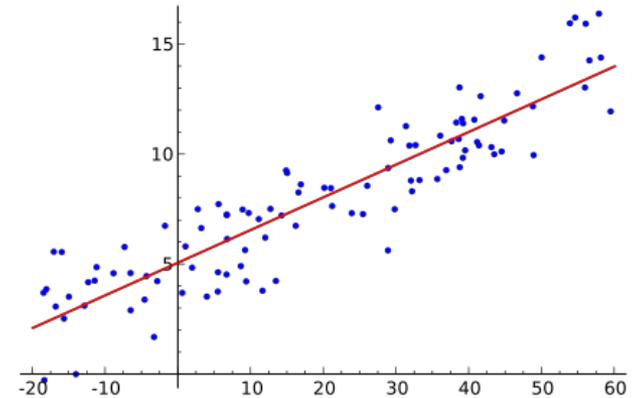
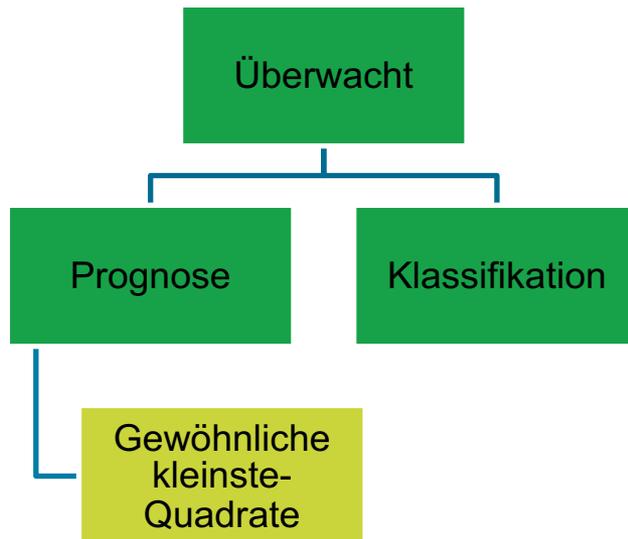
# Thematische Einführung Machine Learning

- Arten des ML: Unterteilung nach Lernmethodik



# Thematische Einführung Machine Learning

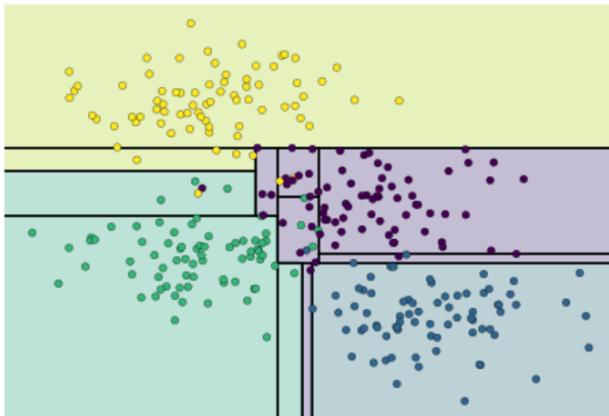
- Überwachte ML-Verfahren:



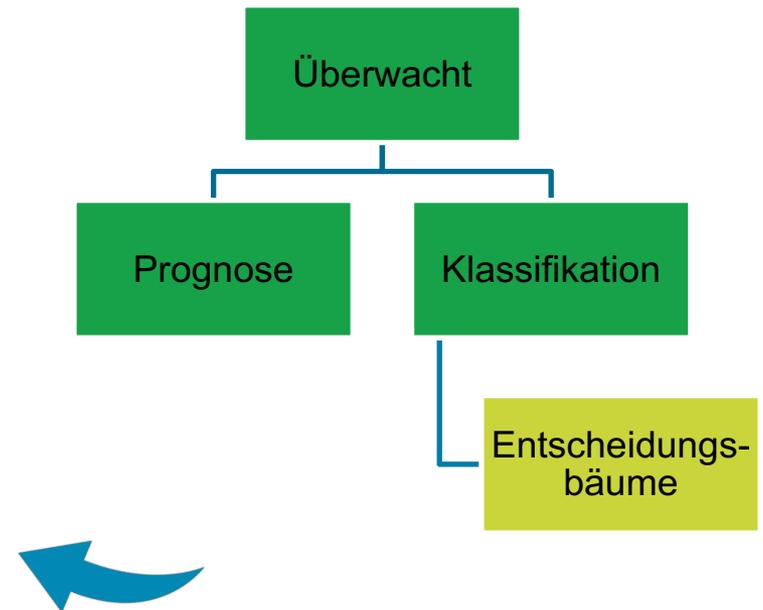
Quelle: VanderPlas, J. (2016)

# Thematische Einführung Machine Learning

- Überwachte ML-Verfahren:

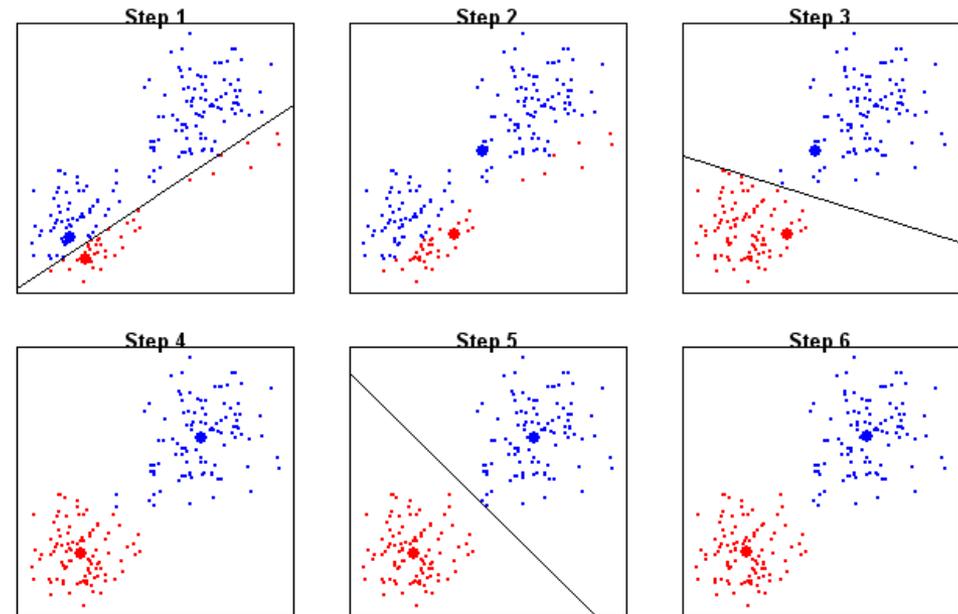
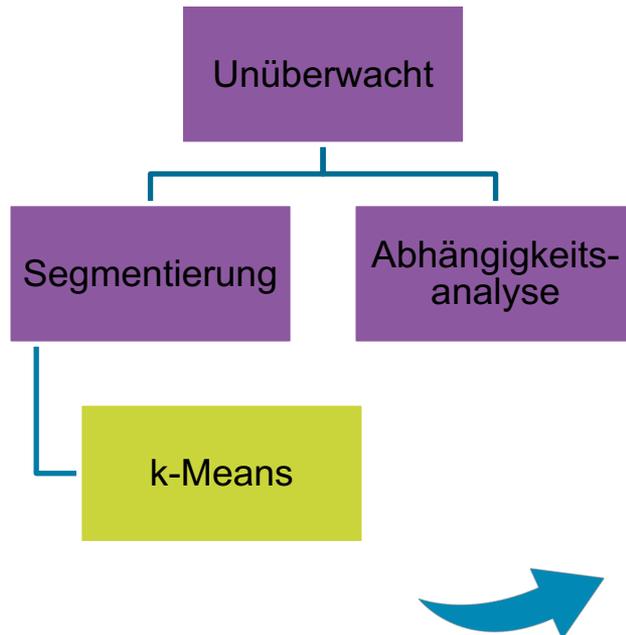


Quelle: VanderPlas, J. (2016)



# Thematische Einführung Machine Learning

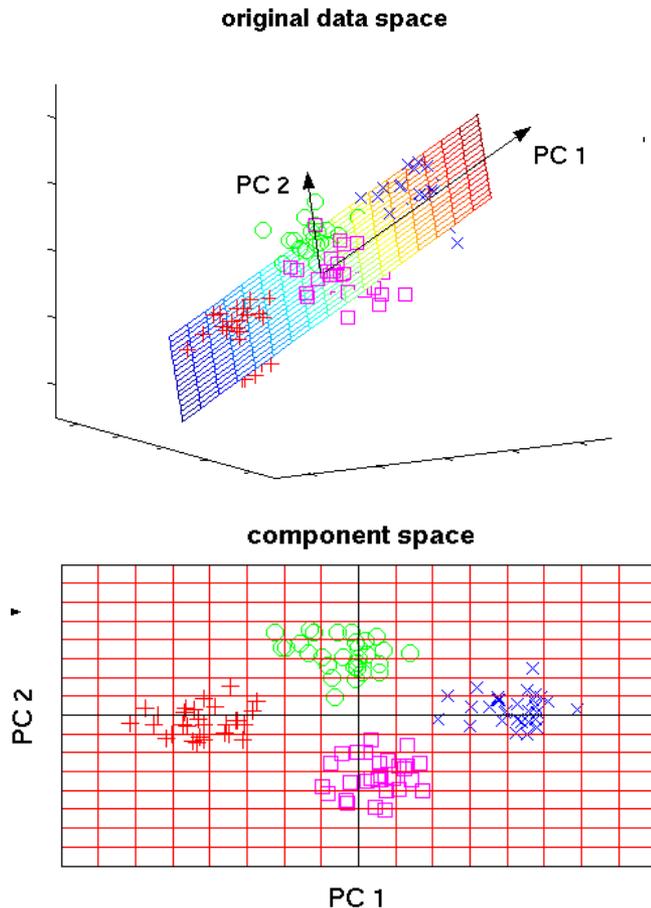
- Unüberwachte ML-Verfahren:



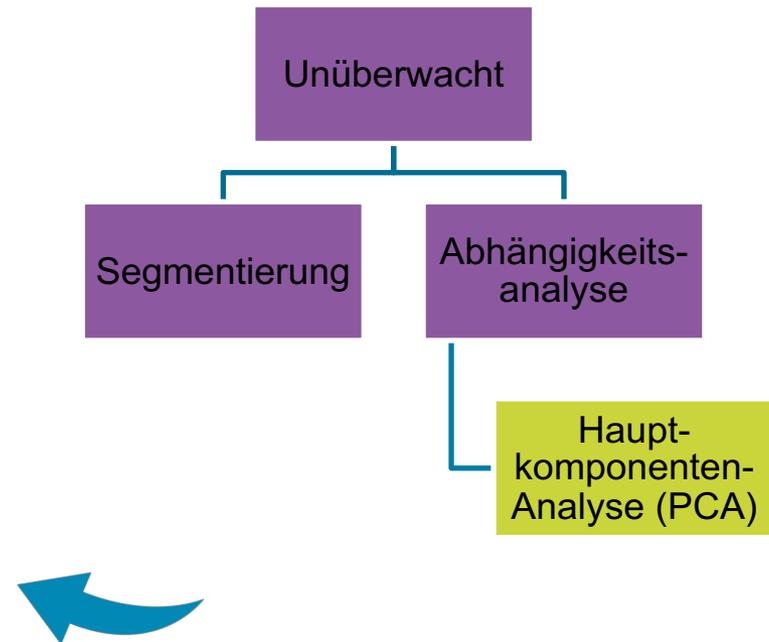
Quelle: medium.com (2017)

# Thematische Einführung Machine Learning

- Unüberwachte ML-Verfahren:

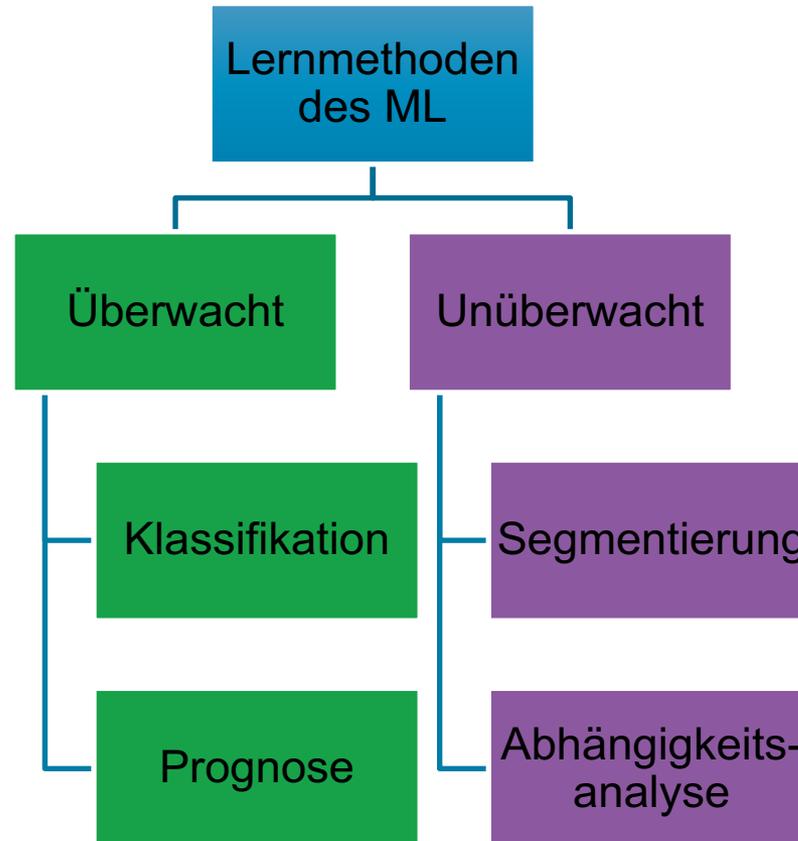


Quelle: hackernoon.com (2018)



# Thematische Einführung Machine Learning

- Arten des ML: Unterteilung nach Lernmethodik



# Exemplarischer Projektablauf

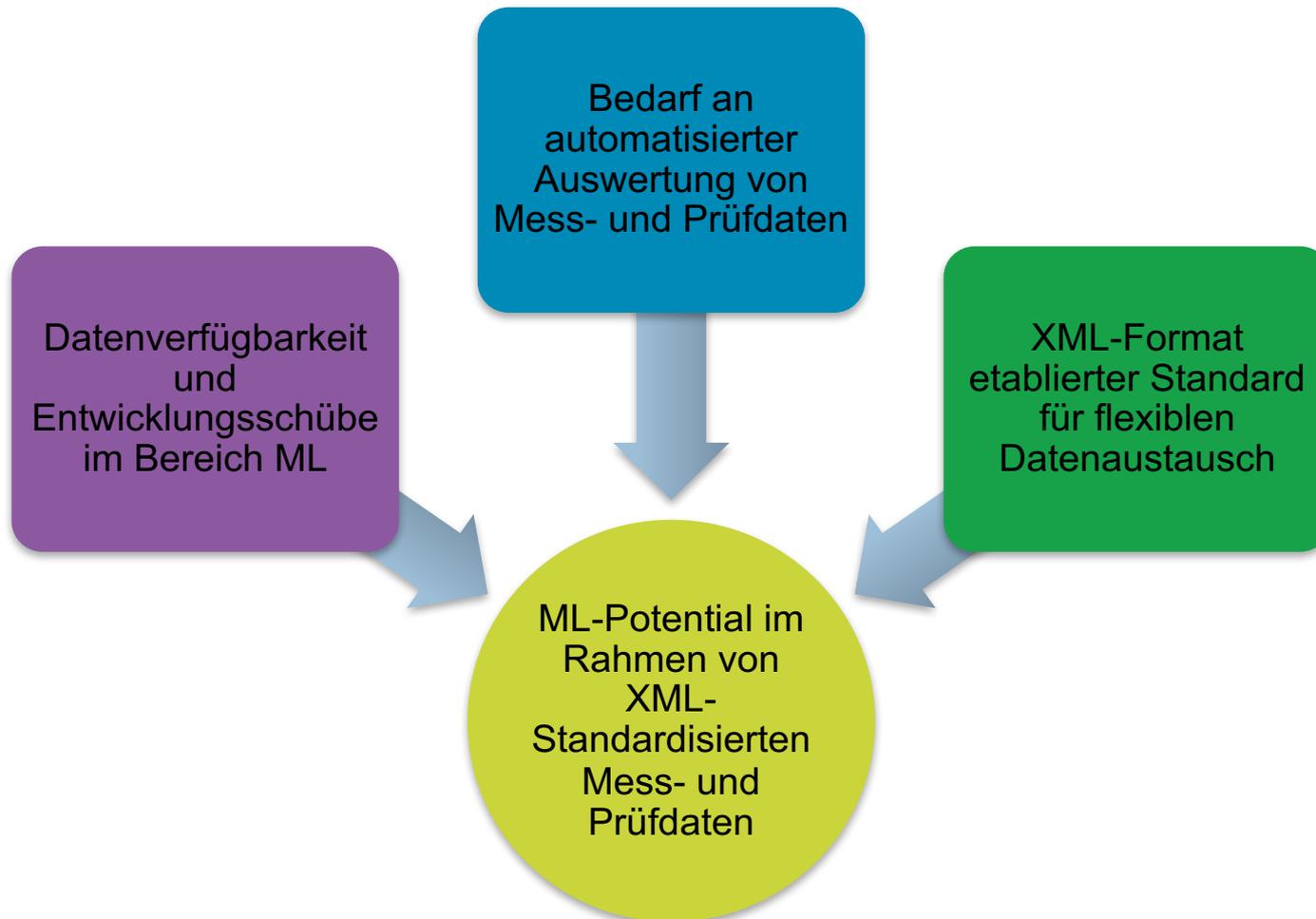
- Produktionsstätte ist Teil der Transportsektion von Thales
- Spezialisiert auf Leit- und Sicherungstechnik für den Schienennah- und –fernverkehr

THALES

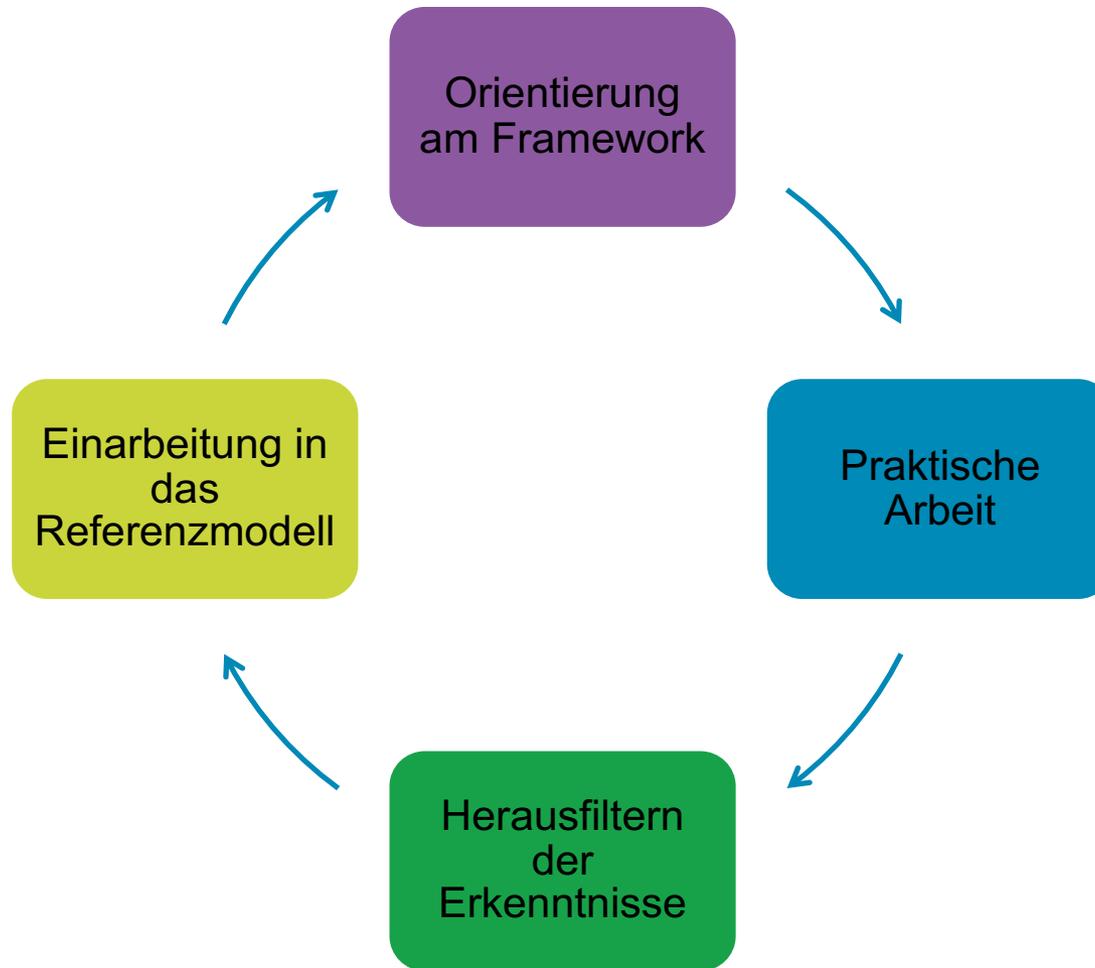


- parsQube setzt eine integrierte XML-Datenbank auf
- Erfasst die in der Produktion anfallenden Messwerte und Sensordaten
- (Grafische) Auswertung innerhalb eines Web Service zentral zugänglich

# Exemplarischer Projektablauf



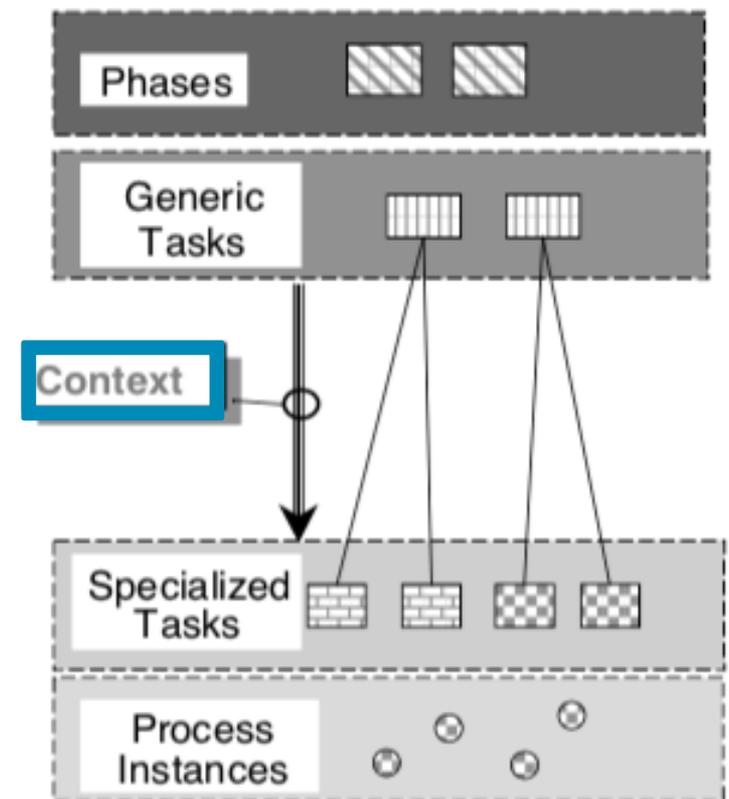
# Exemplarischer Projektablauf



# Exemplarischer Projektablauf

- Eine methodische Stütze für aktuelle und kommende Projekte ist hilfreich
- Benötigt: Etabliertes Framework für standardisierten Projektablauf
- **C**Ross Industry **S**tandard **P**rocess for **D**ata **M**ining (CRISP-DM)

## Reference Model



Quelle: Wirth & Hipp (2000)



# Exemplarischer Projektablauf

- Datengrundlage: Abteilung zur Funktionsprüfung von Leiterplatten
- Vielfältige Tests innerhalb automatisierter Testsysteme

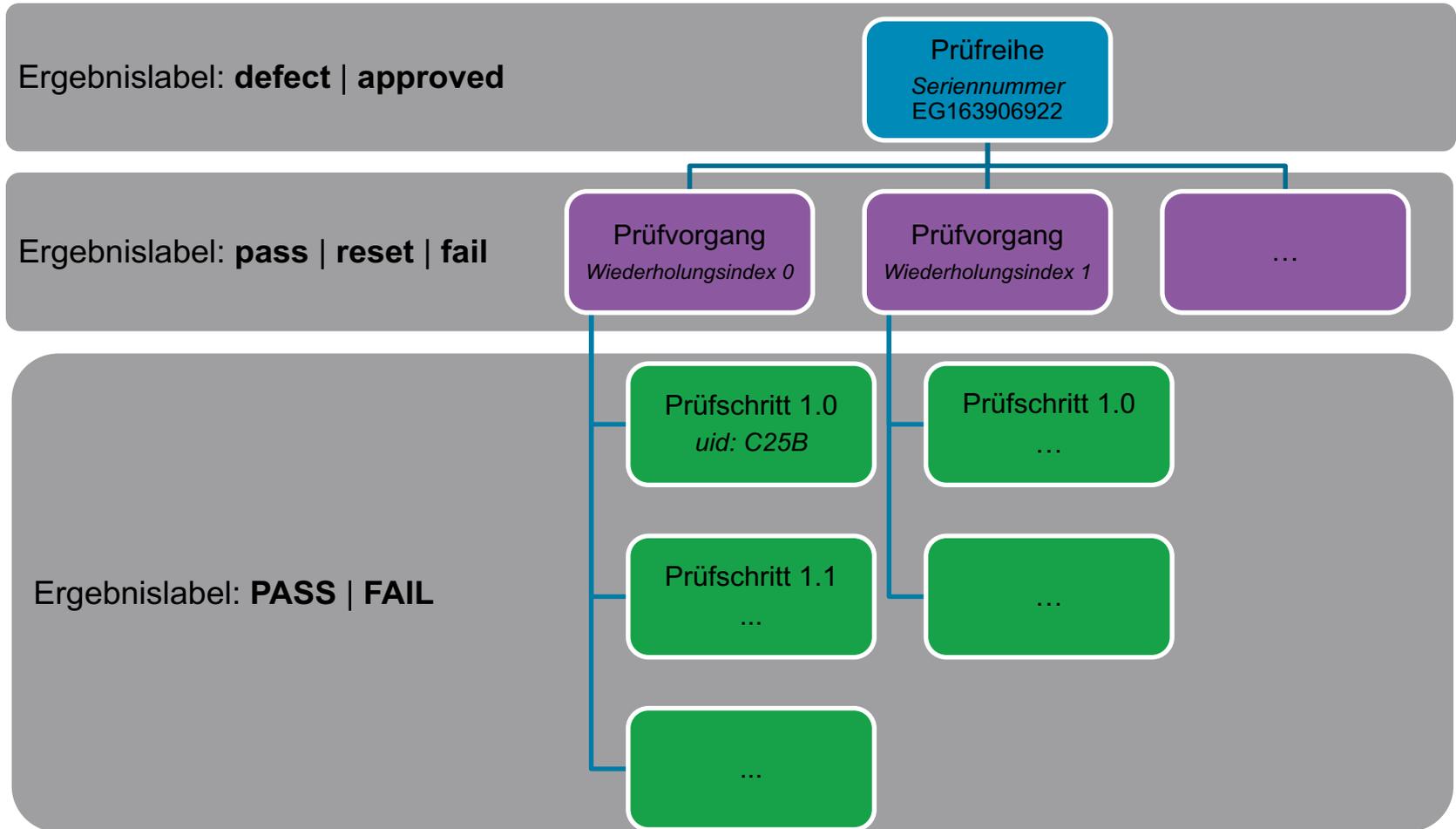


Quelle: wikipedia.org (2018)



Quelle: na.gardien.com (2018)

# Exemplarischer Projektablauf



# Exemplarischer Projektablauf

- XML-Logdatei: Metadaten

7HA02205ABAA\_EG180701227\_20180327\_104219.xml ×

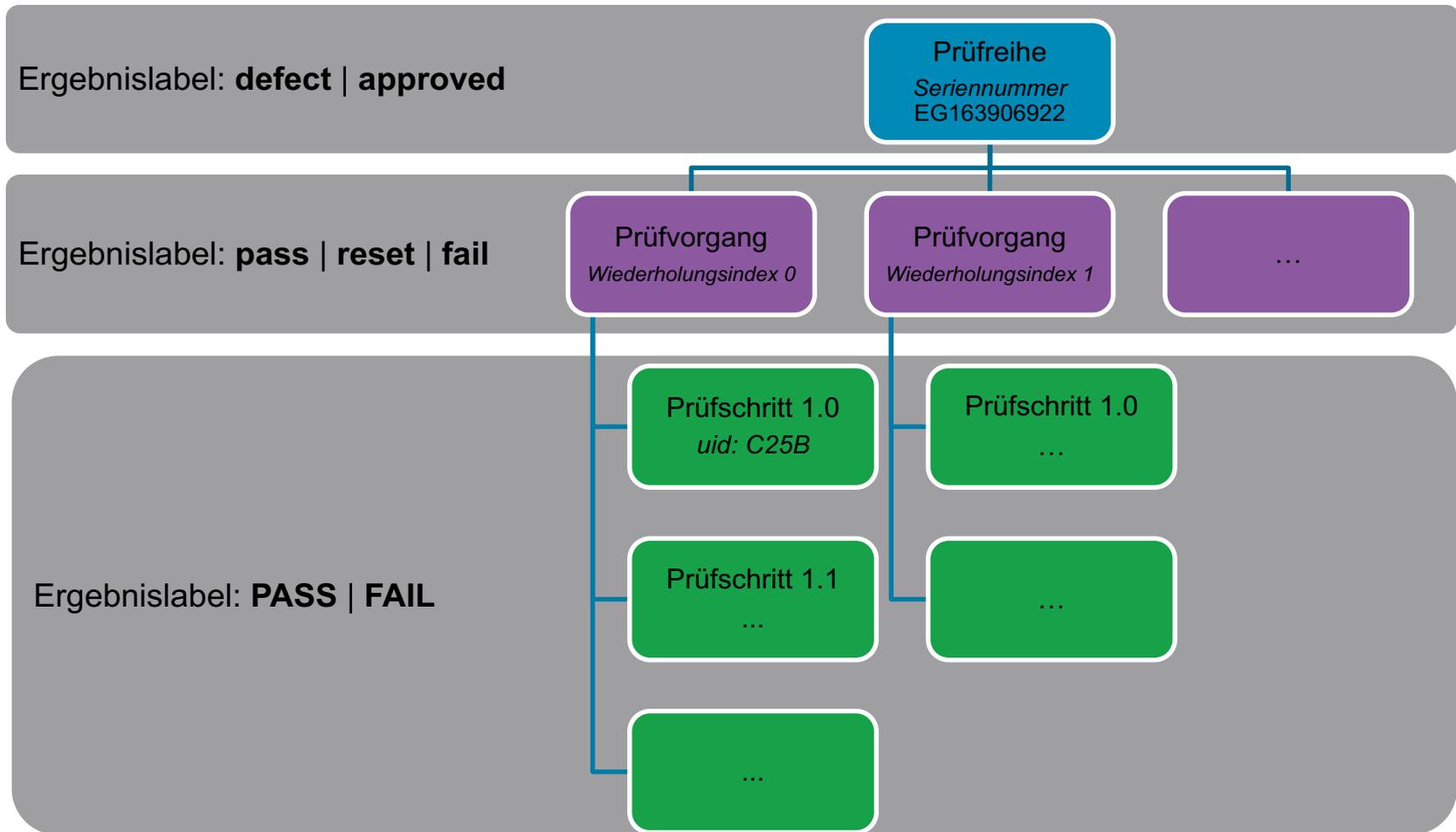
```
1  <?xml version="1.0" encoding="UTF-8"?>
2  <measurementData>
3      <test logVersion="1.3">
4          <metaData>
5              <rhodos version="1.1"/>
6              <testProgram partNumber="7HA02205ABAA" version="V2.0">
7                  <name lang="de">EAK-ZP30K</name>
8                  <configuration conformalCoating="0" processStep="860" repair="0" repairShop="0" runIn="
9              </testProgram>
10             <testStart year="2018" month="3" day="27" hour="10" minute="42" second="19" msec="452"/>
11             <boardData>
12                 <typeLabel specification="83552135" jobNumber="A100262875" serialNumber="EG180701227"
13             </boardData>
14             <testSystem>Tsvp1</testSystem>
15             <environmentData>
16                 <measurement lowerLimit="15" upperLimit="50" module="hardware.thales.M1998256239" id="
17                 <measurement lowerLimit="18.25" upperLimit="27.75" module="hardware.thales.M1998256239
18             </environmentData>
```

# Exemplarischer Projektablauf

- XML-Logdatei: Mess- und Prüfdaten

```
19 </metaData>
20 <step stepNumber="1.0" rerun="false" uid="C25B" auxiliaryStep="false">
21   <stepName lang="de">Adapterhilfsspannung 12 V</stepName>
22   <measurement type="standard" time="10.42.20:304" result="PASS" count="1" rerun="false">
23     <element>
24       <targetValue dataType="number" unit="V" compareMethod="minMax">
25         <min>11.9</min>
26         <max>12.1</max>
27       </targetValue>
28       <measurementValue dataType="number" unit="V">12.017089325907683</measurementValue>
29     </element>
30   </measurement>
31 </step>
```

# Exemplarischer Projektablauf



# Exemplarischer Projektablauf

- Die Prüfschritte charakterisieren einen Prüfvorgang eindeutig: *Features*
- Ergebnis der Prüfreihe ist das *Label*

```
baseline_df.head(10)
```

		result	1.0	1.1	1.2	1.3	1.4	1.5.1	1.5.2	1.5.3	1.5.4	...
serNbr	loopindex											
EG153500559	0	reset	12.015524	0.094321	47.829467	0.071510	NaN	NaN	NaN	NaN	NaN	...
	1	reset	12.015524	0.089922	47.814212	0.071739	37.840891	24.616489	0.897911	5.070776	1.341215	...
	2	reset	12.020220	0.081463	47.821839	0.603992	37.859960	24.544025	0.882655	5.069251	1.343884	...
	3	reset	12.026482	0.085975	47.802770	0.072351	NaN	NaN	NaN	NaN	NaN	...
	4	reset	12.026482	0.087215	47.818026	0.072198	NaN	NaN	NaN	NaN	NaN	...
	5	reset	12.018655	0.090825	47.818026	0.128619	37.875216	24.626024	24.629838	5.070776	5.076115	...
	6	reset	12.021786	0.090261	47.810398	0.129078	37.856146	24.694674	24.692767	5.071158	5.076877	...
	7	pass	12.012393	0.096351	47.810398	0.129231	37.852332	24.748068	24.746162	5.070395	5.076496	...
EG153500605	0	reset	12.020220	0.093757	47.810398	0.096662	37.844705	24.664163	24.624117	5.081834	5.095180	...
	1	reset	12.023351	0.086200	47.829467	0.129001	37.867588	24.738534	24.734720	5.082216	5.095180	...

# Einblick in hilfreiche Tools

- Mit dieser vorverarbeiteten Datengrundlage lässt sich ML betreiben
- Einen kurzen live Einblick in die Entwicklungsumgebung Jupyter Notebook:



Quelle: jupyter.org (2018)

[localhost:8888/notebooks](http://localhost:8888/notebooks)



Quelle: python.org (2018)

- Weitere hilfreiche Open-Source Software:



Quelle: orange.biolab.si (2018)



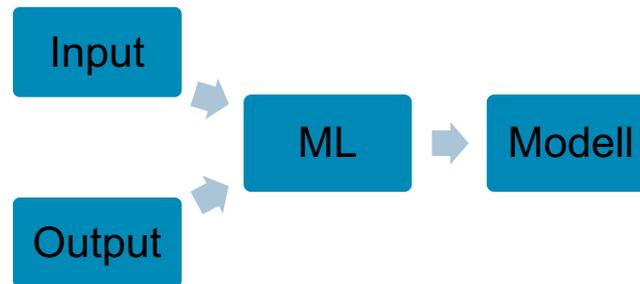
Quelle: anaconda.com (2018)

# Zusammenfassung und Anregungen

- ML findet Einsatz in Diagnostic und Predictive Analytics



- ML ermöglicht die „automatisierte“ Entwicklung von Modellen



- Es gibt Lernstrategien zur Erkennung bekannter und unbekannter Muster

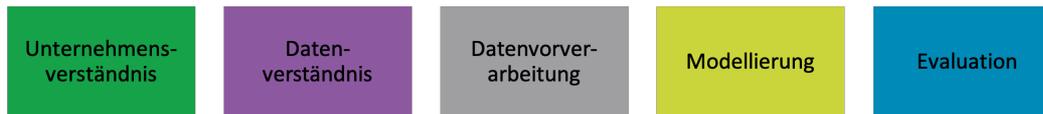


# Zusammenfassung und Anregungen

- ML ist im Zusammenhang mit XML ein gefragtes Werkzeug



- CRISP-DM bietet eine Orientierung für ML-Projekte



- Open-Source-Software wie Jupyter Notebook erleichtert den Einstieg und Umgang mit der Thematik



Danke für eure Aufmerksamkeit

# Literaturhinweise

- Duda, R. O., Hart, P. E. & Stork, D. G., 2001. *Pattern Classification*. New York: John Wiley & Sons Inc.
- Oettinger, M., 2017. *Data Science - Eine praxisorientierte Einführung im Umfeld von Machine Learning, künstlicher Intelligenz und Big Data*. 1 ed. Hamburg: tredition GmbH.
- Salem, M. V. M. et al., 2011. *Information Retrieval Technology*. Heidelberg: Springer-Verlag.
- VanderPlas, J., 2016. *Python Data Science Handbook*. 1 ed. s.l.:O'Reilly Media.
- Vermeulen, A. F., 2018. *Practical Data Science: A Guide to Building the Technology Stack for Turning Data Lakes into Business Assets*. 1 ed. Berkeley, CA: Apress.
- Wirth, R. & Hipp, J., 2000. CRISP-DM: Towards a Standard Process Model for Data Mining. *Proceedings of the Fourth International Conference on the Practical Application of Knowledge Discovery and Data Mining*, pp. 29-39.
- Gartner 2013. Business Intelligence and Analytics Trends. *Gartner Symposium IT XPO*, p.57.

# Bildquellen

- <https://hackernoon.com/a-laymans-introduction-to-principal-components-2fca55>
- [https://de.wikipedia.org/wiki/Leiterplatte\\_c19fa0](https://de.wikipedia.org/wiki/Leiterplatte_c19fa0)
- <https://medium.com/@dilekamadushan/introduction-to-k-means-clustering-7c0ebc997e00>
- <http://na.gardien.com/news/de/polytron-print-tests-with-gardiens-automatic-tester>
- <http://jupyter.org>
- <https://www.python.org>
- <https://www.anaconda.com>
- <https://orange.biolab.si>