



HOCHSCHULE LANDSHUT
UNIVERSITY OF APPLIED SCIENCES

Machine Learning-Software für Controller/-innen*: ein Anwendungsbeispiel

Prof. Dr. Patrick Dieses
Professur für Controlling, Finanz- und Risikomanagement
Hochschule Landshut

Predictive Analytics Community
Online-Vortrag am 3. Dezember 2021

*Hinweis: Es wird im Folgenden das generische Maskulinum verwendet.



Prof. Dr. Patrick Dieses – Fakultät Betriebswirtschaft

Professur für Controlling, Risiko- und Finanzmanagement



- Seit 2016: Studiendekan der Fakultät Betriebswirtschaft an der HS Landshut
- Seit 2012: Inhaber der Professur für Controlling, Risiko- und Finanzmanagement
- Von 2003 bis 2012: Verschiedene Funktionen in der Industrie (E.ON, BASF) und Unternehmensberatung (McKinsey&Company) in den Bereichen Beteiligungs-, Finanz- sowie Risikocontrolling/-management



Bitte gerne kontaktieren!

Kontakt

E-Mail: patrick.dieses@haw-landshut.de

Homepage: <https://www.haw-landshut.de/hochschule/fakultaeten/betriebswirtschaft/prof-dr-patrick-dieses/profil.html>

„Predictive Planning and Forecasting wird durch die Reife der Technologie und die einfache Bereitstellung in der Cloud für immer mehr Unternehmen leistbar und relevant.“

BARC-Studie, 2018

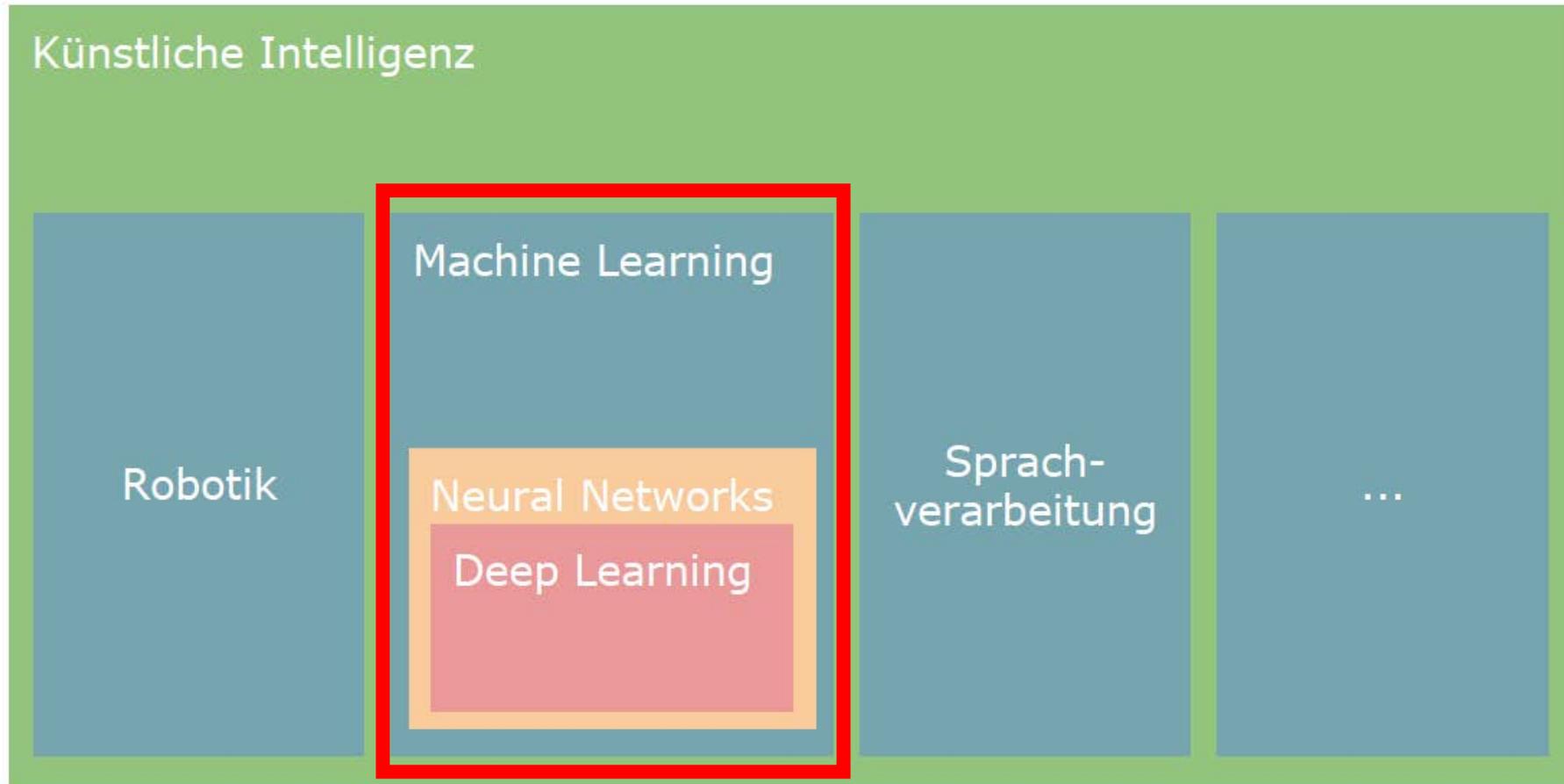
Wie schätzen Sie die aktuelle Reife Ihres Unternehmens im Bereich Predictive Planning and Forecasting ein? (n=308)



Quelle: BARC-Studie, 2018

1. Einordnung bisher vorgestellter Methoden in ML-Dimensionen
2. Vorstellung Data Analytics-Tool KNIME anhand eines Beispiels
3. Diskussion: Kompetenzanforderungen an Controller

Literatur zum Einstieg



Quelle: openHPI

Überwachtes Lernen
(Supervised Learning)

Unüberwachtes Lernen
(Unsupervised Learning)

Paradigmen

Halbüberwachtes Lernen
(Semisupervised Learning)

Bestärkendes Lernen
(Reinforcement Learning)

Überwachtes Lernen (Supervised Learning)

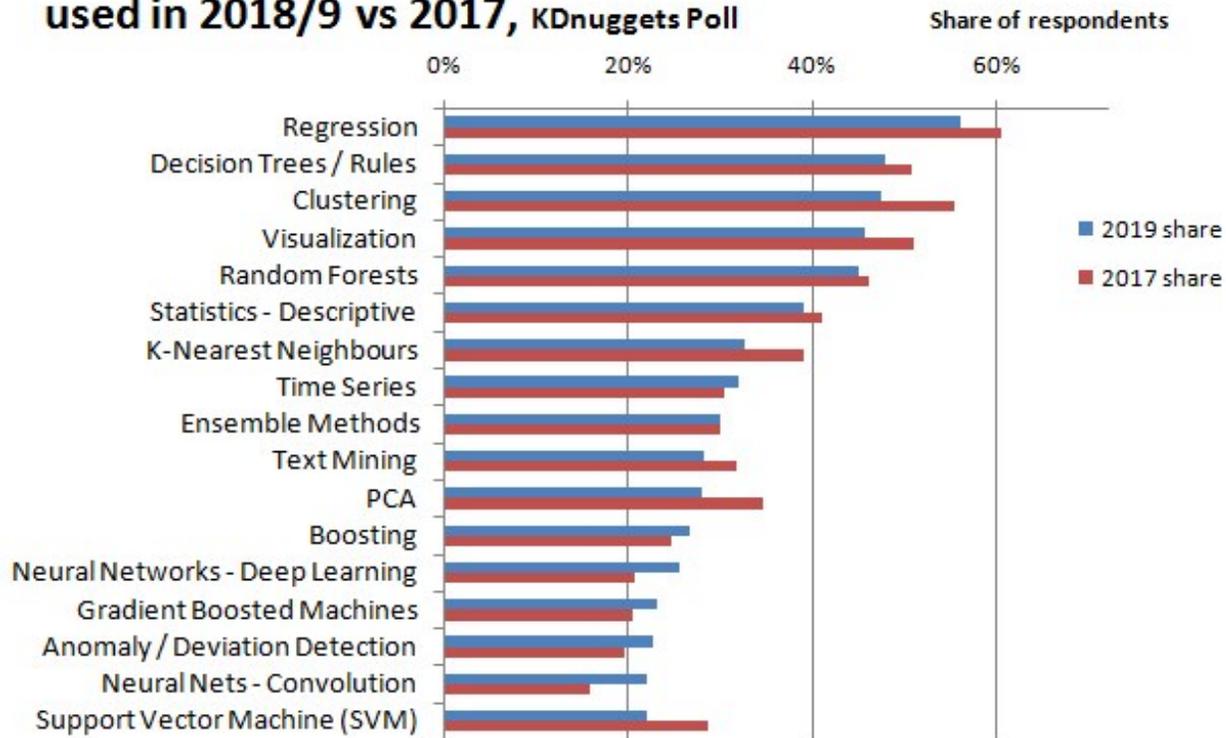
- Regressionsmodelle (univariate/multivariate, Autoregression → Zeitreihenanalysen)
- **Entscheidungsbäume**
- Random Forests
- Diskriminanzanalyse
- Logistische Regressionen (Logit)
- Naive Bayes Schätzer
- Neuronale Netze (→ Deep Learning)
- ...

Unüberwachtes Lernen (Unsupervised Learning)

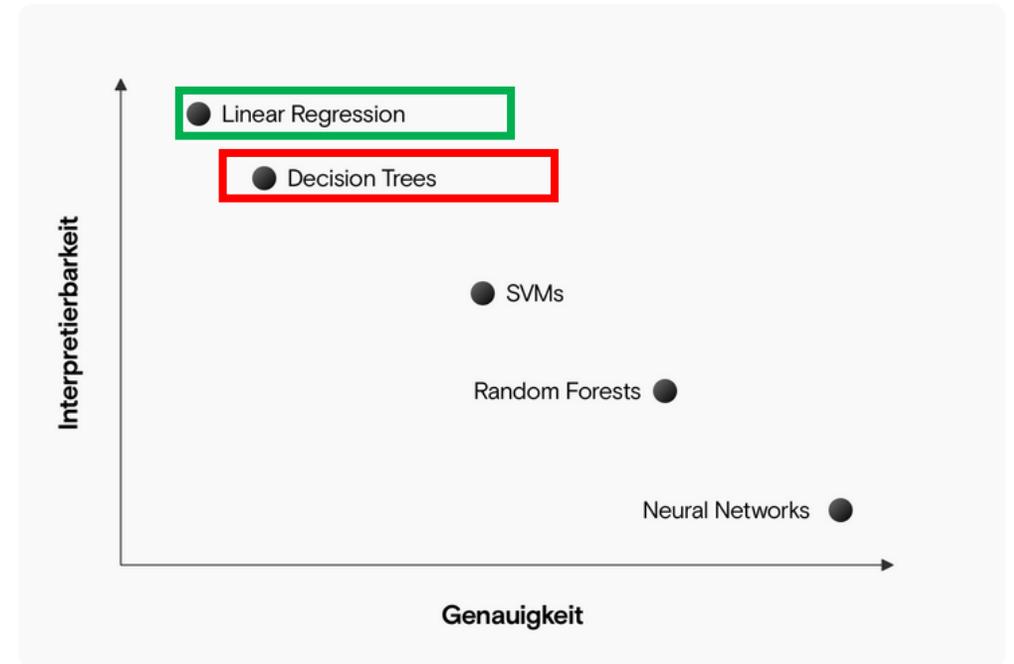
- Clusteranalysen
- Assoziationsmodelle
- Dimensionsreduktion
- ...

ML-Algorithmen: Verbreitung in der Anwendung

Top Data Science, Machine Learning Methods, Algorithms used in 2018/9 vs 2017, KDnuggets Poll



Quelle: www.kdnuggets.com



Genauere Modelle sind oft schwerer zu interpretieren.

Quelle: <https://www.datarevenue.com/de-blog/interpretierbares-machine-learning>

- Customer-Churn-Modelle
- Berechnung individueller CLV (Customer Lifetime Value)
- Personalfluktuatation
- Predictive Maintenance
- ...



Prescriptive Analytics

➔ Bessere (evidenzbasierte) Entscheidungen

Controlling!

1. Einordnung bisher vorgestellter Methoden in ML-Dimensionen
2. Vorstellung Data Analytics-Tool KNIME anhand eines Beispiels
3. Diskussion: Kompetenzanforderungen an Controller

Literatur zum Einstieg

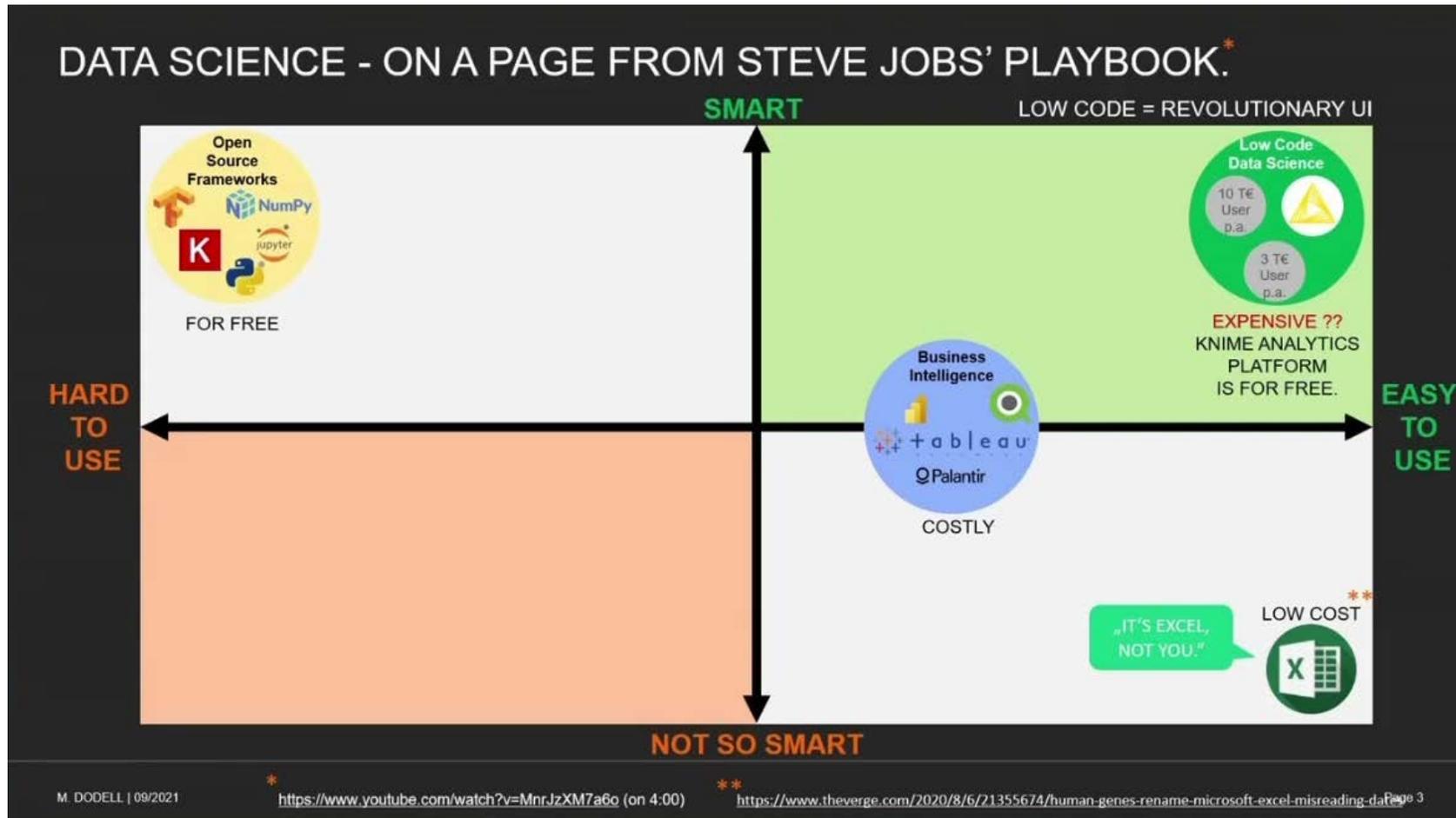
- Excel
- Spezielle Analytics-Software (KNIME, Rapid Miner,..)
- Planungssoftware mit offenen Schnittstellen (z.B. für Python, R)
- Planungssoftware mit eingebauten Analytics-Funktionen



➤ **Effektivität?**

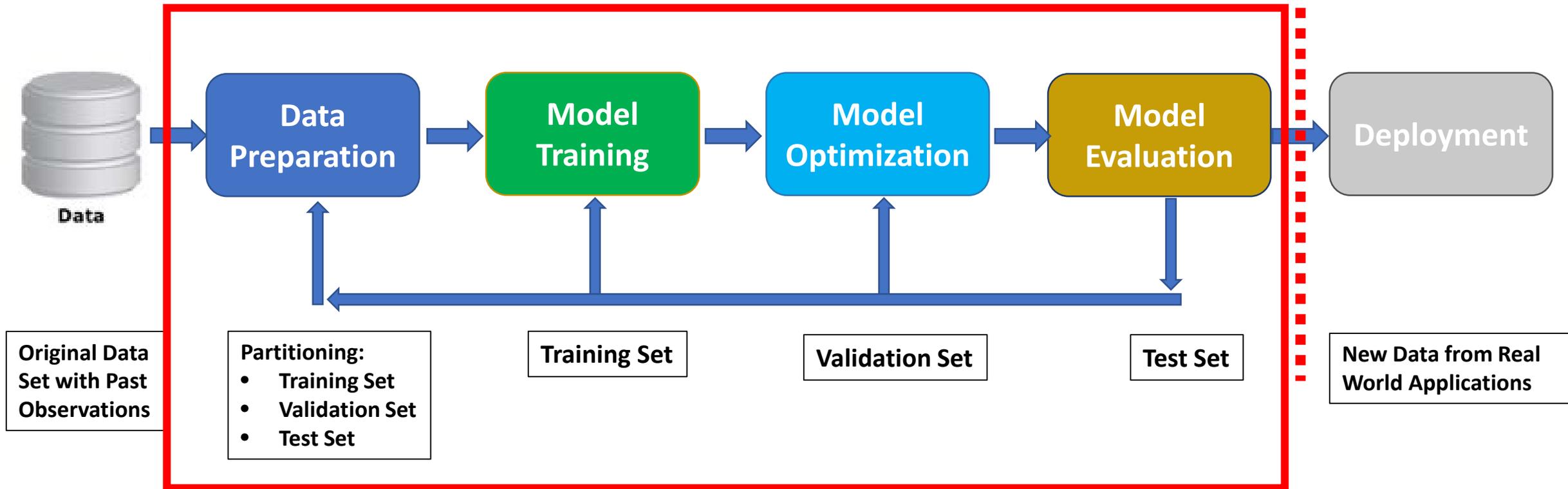
➤ **Effizienz?**

Data Science Analytics Software – Evaluation Matrix



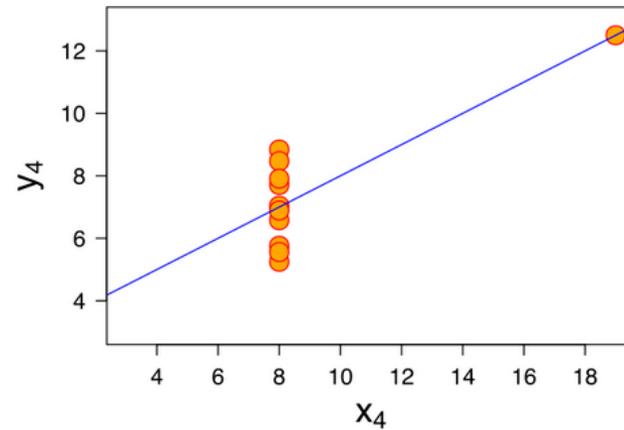
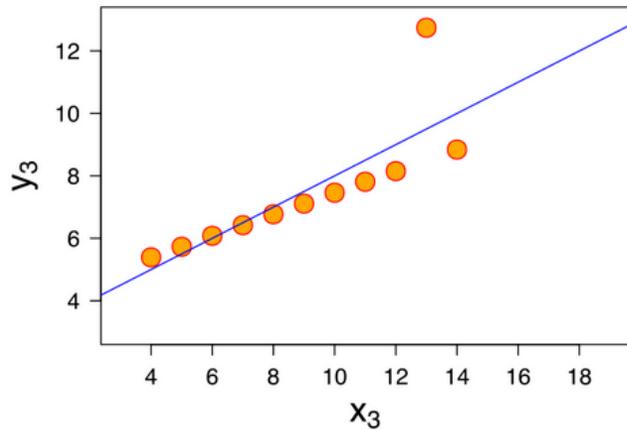
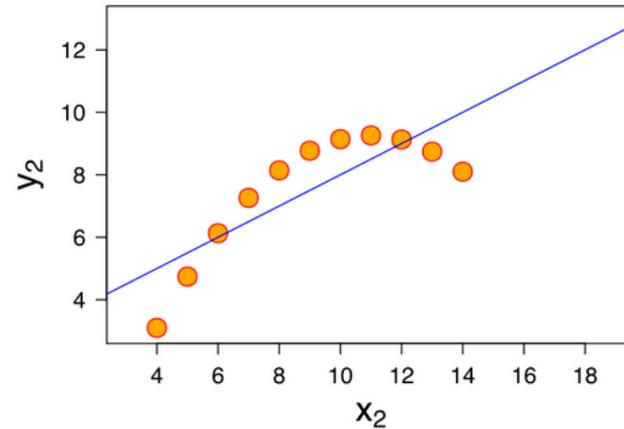
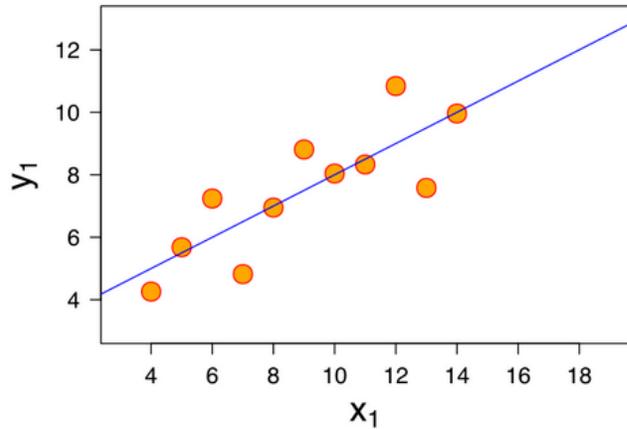
Source: https://www.knime.com/fall-data-talks-bringing-business-and-data-science-together-presentations?utm_source=pardot&utm_medium=email&utm_term=23&utm_content=newsletter&utm_campaign=KNIME-Inside (on 2:00)

Wesentliche Schritte



Quelle: KNIME AG

Wichtig: Check mit Datenvisualisierung!



| Eigenschaft | Wert |
|--|--|
| Mittelwert von x in jedem Fall | 9 (exakt) |
| Varianz von x in jedem Fall | 11 (exakt) |
| Mittelwert von y in jedem Fall | 7,50 (auf 2 Stellen) |
| Varianz von y in jedem Fall | 4,122 oder 4,127 (auf 3 Stellen) |
| Korrelation zwischen x und y in jedem Fall | 0,816 (auf 3 Stellen) |
| Lineare Regression in jedem Fall | $y = 3,00 + 0,500x$ (auf 2 bzw. 3 Stellen) |

Quelle: <https://de.wikipedia.org/wiki/Anscombe-Quartett>

- Screenshots der wichtigsten Schritte als BACKUP für Durchsicht im Nachgang
- Kurze Einführung in die grafische Benutzeroberfläche von KNIME
- Hinweis: Ergebnisse weichen von den Werten der Vorführung ab, da beim Durchführen des Knoten „Partitioning“ Datensatz neu (zufallsbasiert und geschichtet) in Trainings- und Testdaten aufgeteilt wird.

BACKUP
Screenshots

KNIME - Benutzeroberfläche

BACKUP
Screenshots



The screenshot displays the KNIME Analytics Platform interface. The main workspace shows a workflow with the following nodes: CSV Reader (Daten einlesen), Row Filter (Zeilen filtern), Rule Engine (Klassierung Einkommen), Color Manager (farblich markieren), Partitioning (Top: train set (75%), Bottom: test set (25%), Stratified sampling), Decision Tree Learner (Model trainieren), Decision Tree Predictor (Vorhersage), Scorer (JavaScript) (Model Evaluation I), and ROC Curve (Model Evaluation II). There are also three unconnected nodes: Data Explorer (Datencheck), Scatter Plot (Daten-visualisierung), and another Decision Tree Predictor (Vorhersage).

The left sidebar contains the KNIME Explorer, showing a project tree with folders like 'LOCAL (Local Workspace)' and 'Workflows', and a list of projects including 'KNIME_project'. Below it is the Node Repository with categories like IO, Manipulation, Views, Analytics, DB, etc.

The right sidebar shows the 'Description' panel for the current project, 'KNIME_proje...', with fields for Title, Description, Tags, Links, Creation Date (2021-11-5), and Author (dieses).

The bottom status bar includes 'Outline', 'Console', and 'Node Monitor' tabs, with the 'KNIME Console' currently active.

KNIME – Daten einlesen

BACKUP
Screenshots



| | A | B | C | D | E | F |
|----|--|---|---|---|---|---|
| 1 | User ID,Gender,Age,EstimatedSalary,Purchased | | | | | |
| 2 | 15624510,Male, ,19000,0 | | | | | |
| 3 | 15810944,Male,35,20000, | | | | | |
| 4 | 15668575,Female,26,43000,0 | | | | | |
| 5 | 15603246,Female,27,57000,0 | | | | | |
| 6 | 15804002,Male,19,76000,0 | | | | | |
| 7 | 15728773,Male,27,58000,0 | | | | | |
| 8 | 15598044,Female,27,84000,0 | | | | | |
| 9 | 15694829,Female,32,150000,1 | | | | | |
| 10 | 15600575,Male,25,33000,0 | | | | | |
| 11 | 15727311,Female,35,65000,0 | | | | | |
| 12 | 15570769,Female,26,80000,0 | | | | | |
| 13 | 15606274,Female,26,52000,0 | | | | | |
| 14 | 15746139,Male,20,86000,0 | | | | | |
| 15 | 15704987,Male,32,18000,0 | | | | | |
| 16 | 15628972,Male,18,82000,0 | | | | | |
| 17 | 15697686,Male,29,80000,0 | | | | | |
| 18 | 15733883,Male,47,25000,1 | | | | | |
| 19 | 15617482,Male,45,26000,1 | | | | | |

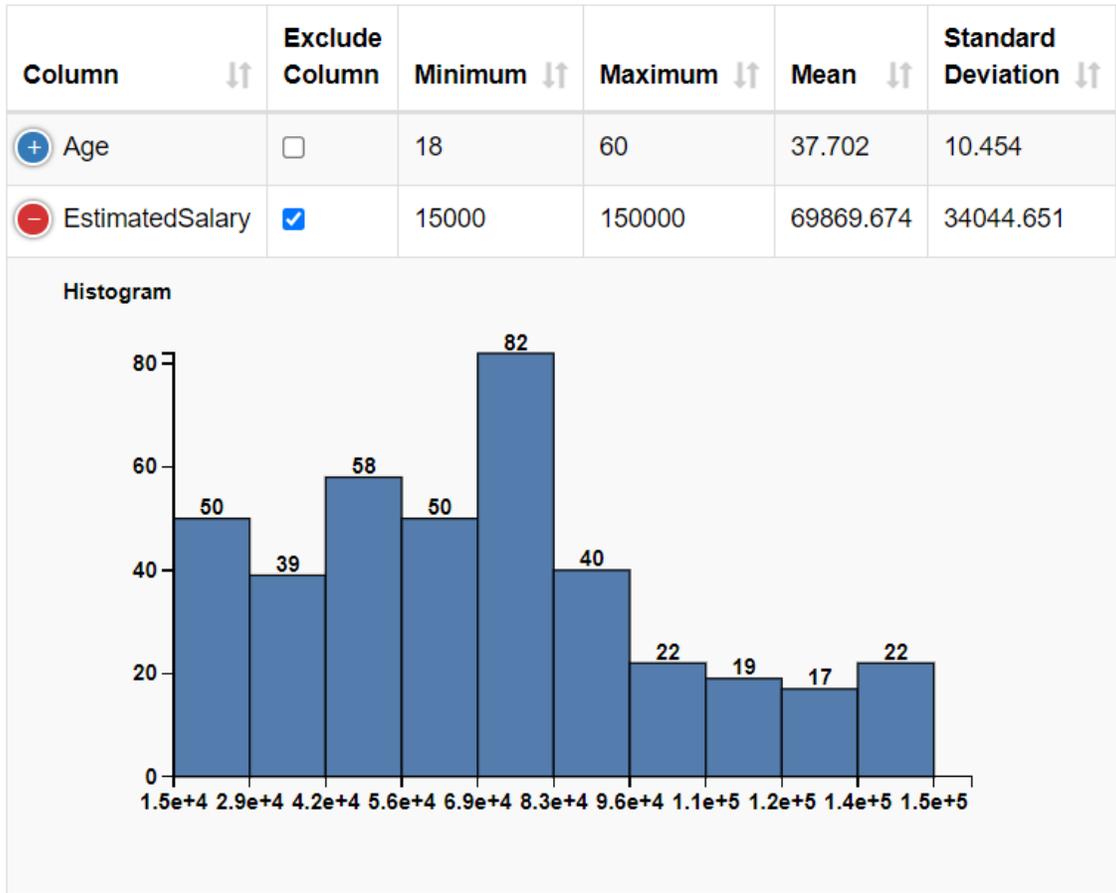
File Table - 0:1 - CSV Reader (Daten einlesen)

File Edit Hilite Navigation View

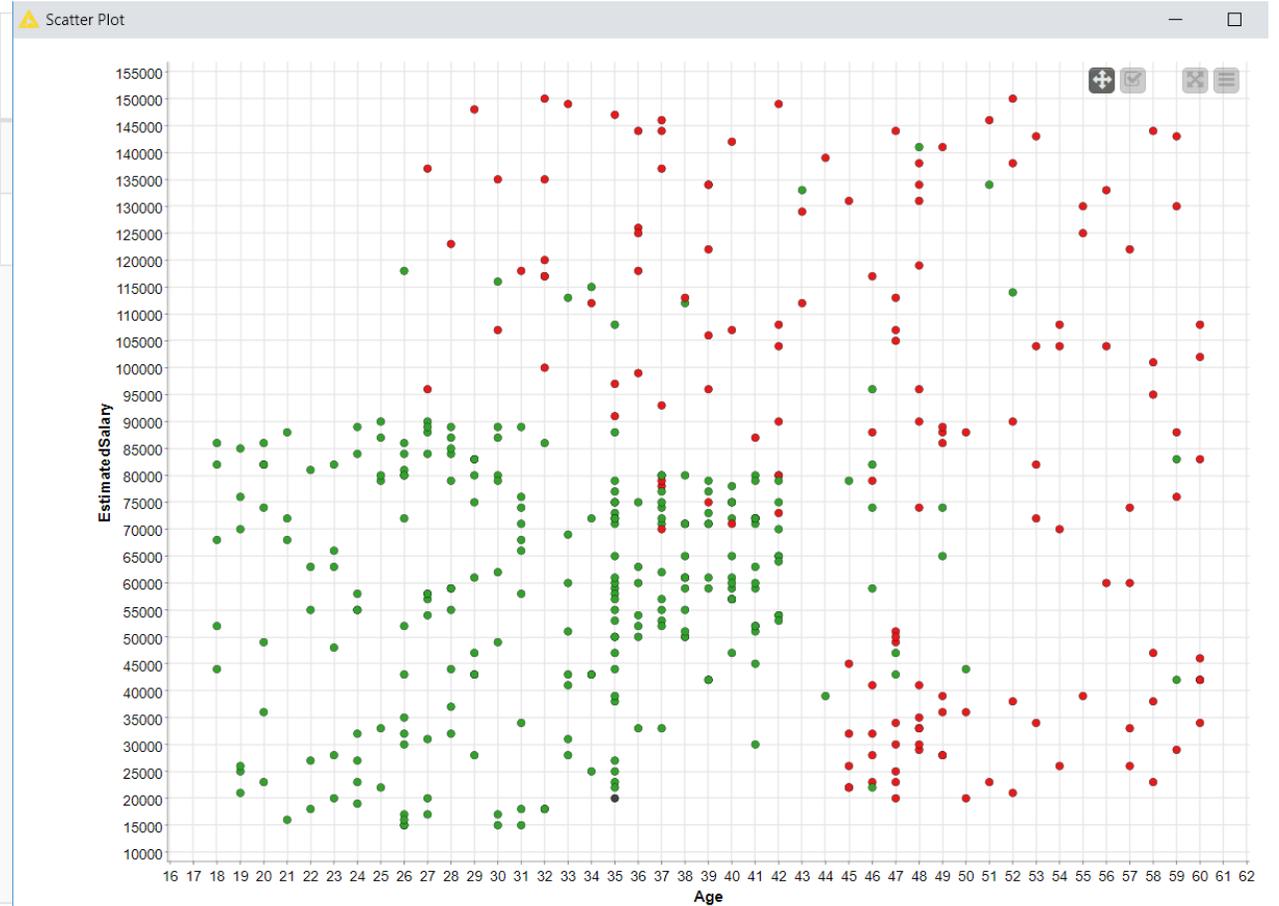
Table "default" - Rows: 400 Spec - Columns: 5 Properties Flow Variables

| Row ID | User ID | Gender | Age | Estimat... | Purcha... |
|--------|----------|--------|-----|------------|-----------|
| Row0 | 15624510 | Male | ? | 19000 | 0 |
| Row1 | 15810944 | Male | 35 | 20000 | ? |
| Row2 | 15668575 | Female | 26 | 43000 | 0 |
| Row3 | 15603246 | Female | 27 | 57000 | 0 |
| Row4 | 15804002 | Male | 19 | 76000 | 0 |
| Row5 | 15728773 | Male | 27 | 58000 | 0 |
| Row6 | 15598044 | Female | 27 | 84000 | 0 |
| Row7 | 15694829 | Female | 32 | 150000 | 1 |
| Row8 | 15600575 | Male | 25 | 33000 | 0 |
| Row9 | 15727311 | Female | 35 | 65000 | 0 |
| Row10 | 15570769 | Female | 26 | 80000 | 0 |
| Row11 | 15606274 | Female | 26 | 52000 | 0 |
| Row12 | 15746139 | Male | 20 | 86000 | 0 |
| Row13 | 15704987 | Male | 32 | 18000 | 0 |
| Row14 | 15628972 | Male | 18 | 82000 | 0 |
| Row15 | 15697686 | Male | 29 | 80000 | 0 |
| Row16 | 15733883 | Male | 47 | 25000 | 1 |
| Row17 | 15617482 | Male | 45 | 26000 | 1 |
| Row18 | 15704583 | Male | 46 | 28000 | 1 |
| Row19 | 15621083 | Female | 48 | 29000 | 1 |
| Row20 | 15649487 | Male | 45 | 22000 | 1 |
| Row21 | 15736760 | Female | 47 | 49000 | 1 |

KNIME – Daten anschauen

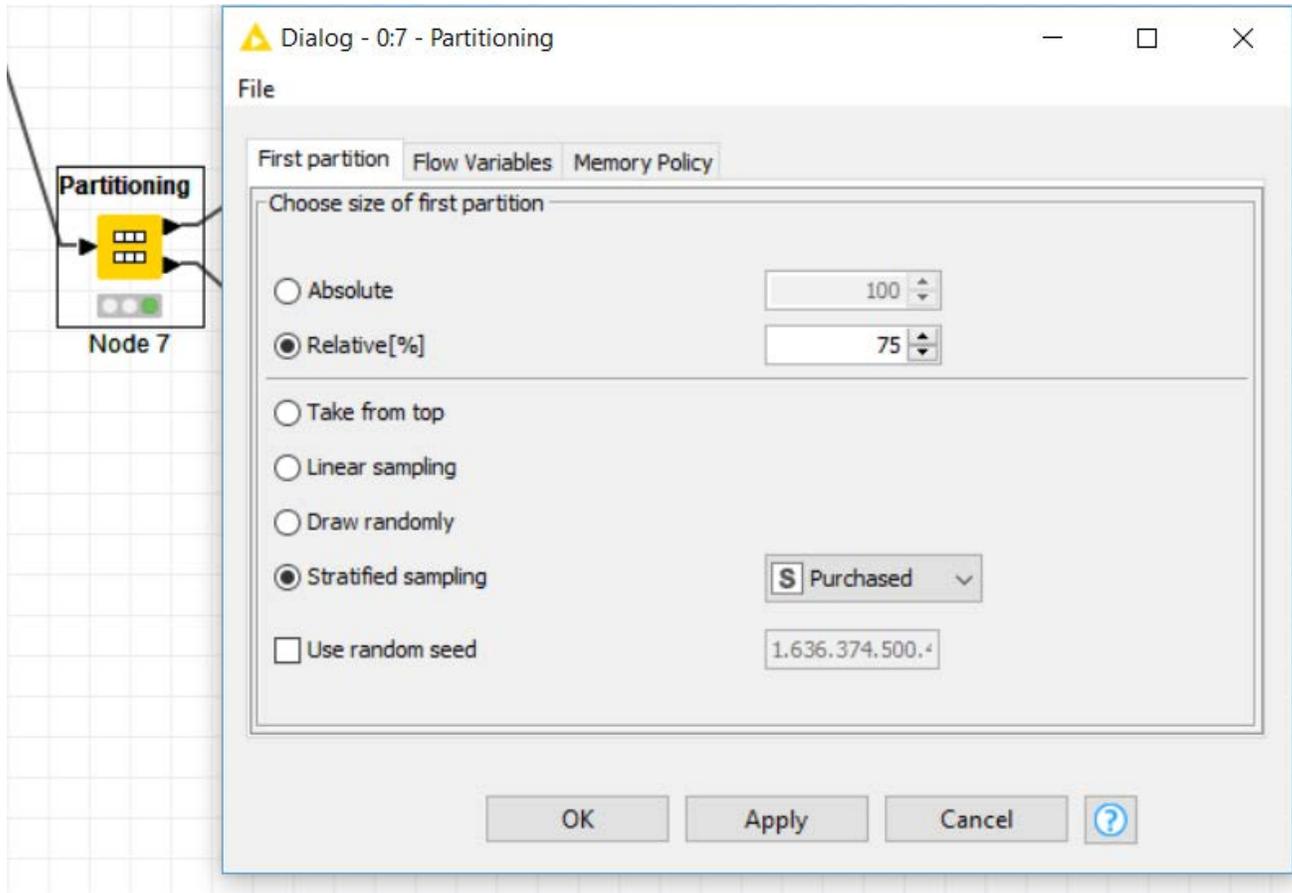


Showing 1 to 2 of 2 entries



KNIME – Datensatz teilen (Trainingsdatensatz und Testdatensatz)

BACKUP
Screenshots

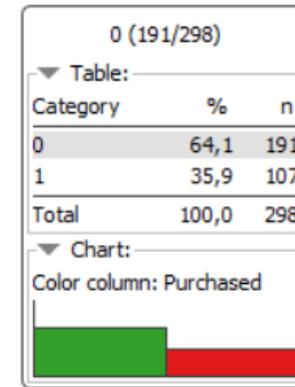
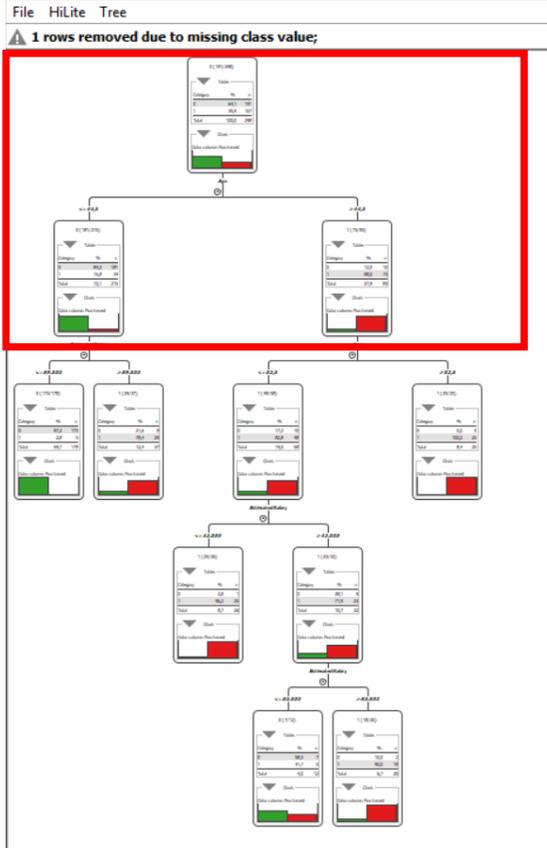


KNIME – Entscheidungsbaum basierend auf Trainingsdaten (1)

BACKUP
Screenshots



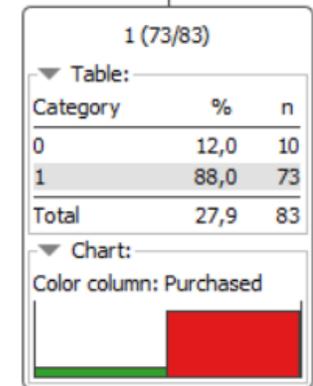
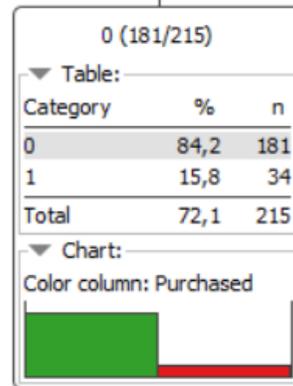
Decision Tree View - 0:8 - Decision Tree Learner (Top: train set (75%))



Age

$\leq 44,5$

$> 44,5$

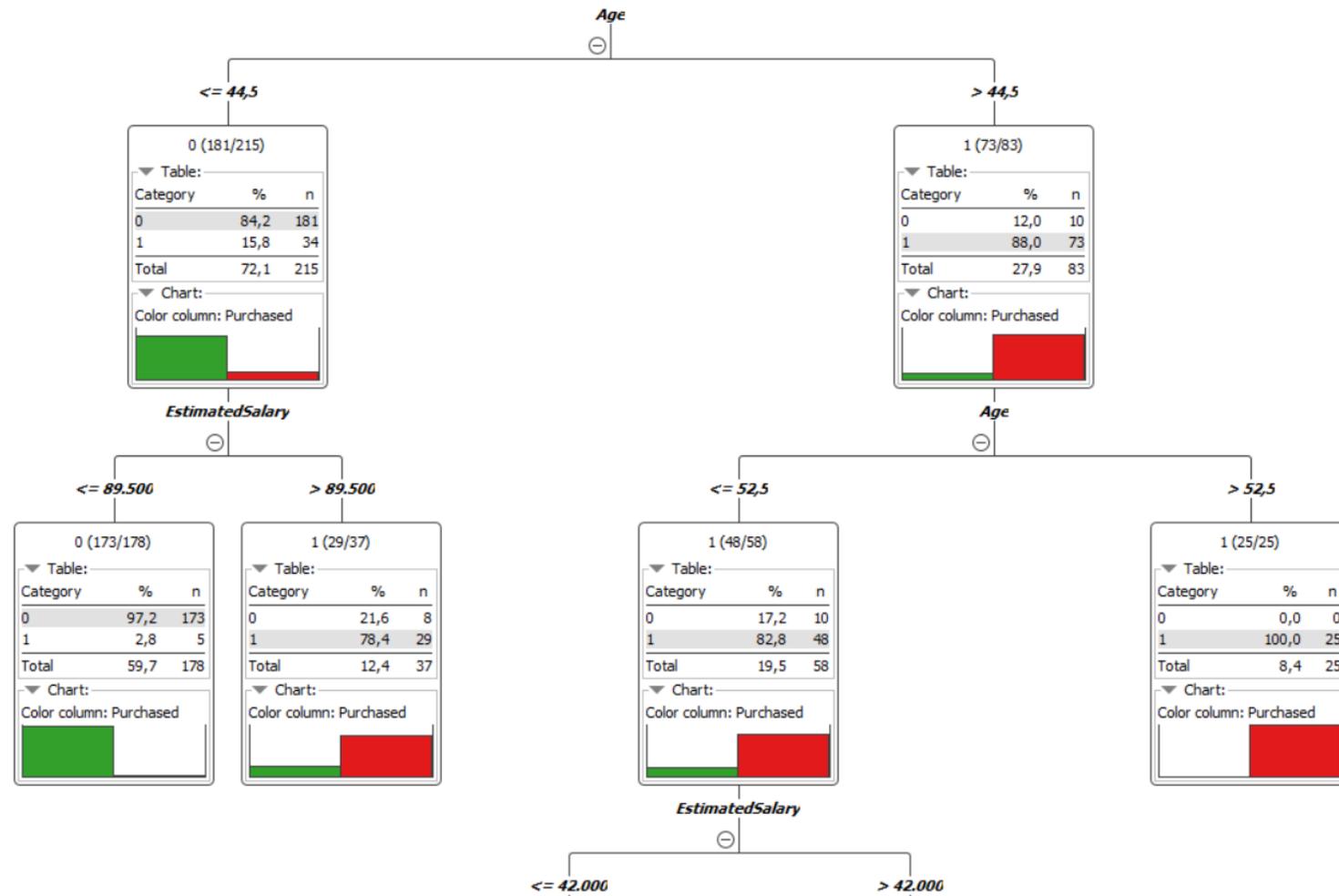
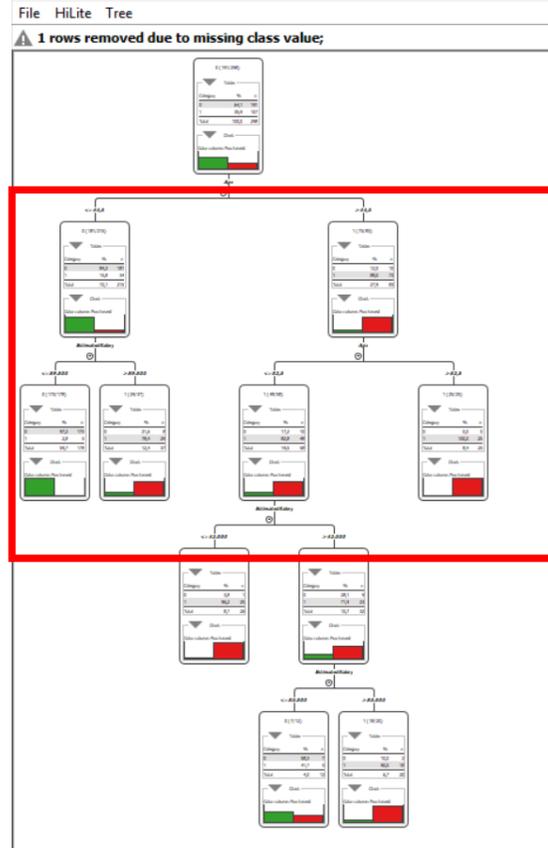


KNIME – Entscheidungsbaum basierend auf Trainingsdaten (2)

BACKUP
Screenshots



Decision Tree View - 0:8 - Decision Tree Learner (Top: train set (75%))



KNIME – Vorhersagen für Testdaten

BACKUP
Screenshots



Classified Data - 0:9 - Decision Tree Predictor

File Edit Hilite Navigation View

Table "default" - Rows: 100 Spec - Columns: 9 Properties Flow Variables

| Row ID | S User ID | S Gender | I Age | I EstimatedSalary | S Salary Group | D P (Purchased=0) | D P (Purchased=1) | S Prediction (Purchased) | S Purchased |
|---------|-----------|----------|-------|-------------------|----------------|-------------------|-------------------|--------------------------|-------------|
| Row 151 | 15576219 | Male | 41 | 45000 | Low | 0.972 | 0.028 | 0 | 0 |
| Row 160 | 15744279 | Male | 32 | 100000 | High | 0.216 | 0.784 | 1 | 1 |
| Row 161 | 15670619 | Male | 25 | 90000 | Medium | 0.216 | 0.784 | 1 | 0 |
| Row 165 | 15578738 | Female | 18 | 86000 | Medium | 0.972 | 0.028 | 0 | 0 |
| Row 168 | 15789815 | Male | 29 | 148000 | High | 0.216 | 0.784 | 1 | 1 |
| Row 170 | 15587013 | Male | 21 | 88000 | Medium | 0.972 | 0.028 | 0 | 0 |
| Row 172 | 15794661 | Female | 26 | 118000 | High | 0.216 | 0.784 | 1 | 0 |
| Row 174 | 15644296 | Female | 34 | 72000 | Medium | 0.972 | 0.028 | 0 | 0 |
| Row 175 | 15614420 | Female | 23 | 28000 | Low | 0.972 | 0.028 | 0 | 0 |
| Row 178 | 15584114 | Male | 24 | 23000 | Low | 0.972 | 0.028 | 0 | 0 |
| Row 183 | 15603319 | Male | 33 | 43000 | Low | 0.972 | 0.028 | 0 | 0 |
| Row 184 | 15759066 | Female | 33 | 60000 | Medium | 0.972 | 0.028 | 0 | 0 |
| Row 186 | 15724402 | Female | 20 | 82000 | Medium | 0.972 | 0.028 | 0 | 0 |
| Row 187 | 15571059 | Female | 33 | 41000 | Low | 0.972 | 0.028 | 0 | 0 |
| Row 189 | 15715160 | Male | 28 | 32000 | Low | 0.972 | 0.028 | 0 | 0 |
| Row 190 | 15730448 | Male | 24 | 84000 | Medium | 0.972 | 0.028 | 0 | 0 |
| Row 195 | 15667742 | Male | 34 | 43000 | Low | 0.972 | 0.028 | 0 | 0 |
| Row 196 | 15738448 | Female | 30 | 79000 | Medium | 0.972 | 0.028 | 0 | 0 |
| Row 197 | 15680243 | Female | 20 | 36000 | Low | 0.972 | 0.028 | 0 | 0 |
| Row 198 | 15745083 | Male | 26 | 80000 | Medium | 0.972 | 0.028 | 0 | 0 |
| Row 203 | 15809347 | Female | 41 | 71000 | Medium | 0.972 | 0.028 | 0 | 0 |

KNIME – Modellgüte (1)

Confusion Matrix

Scorer View

Confusion Matrix

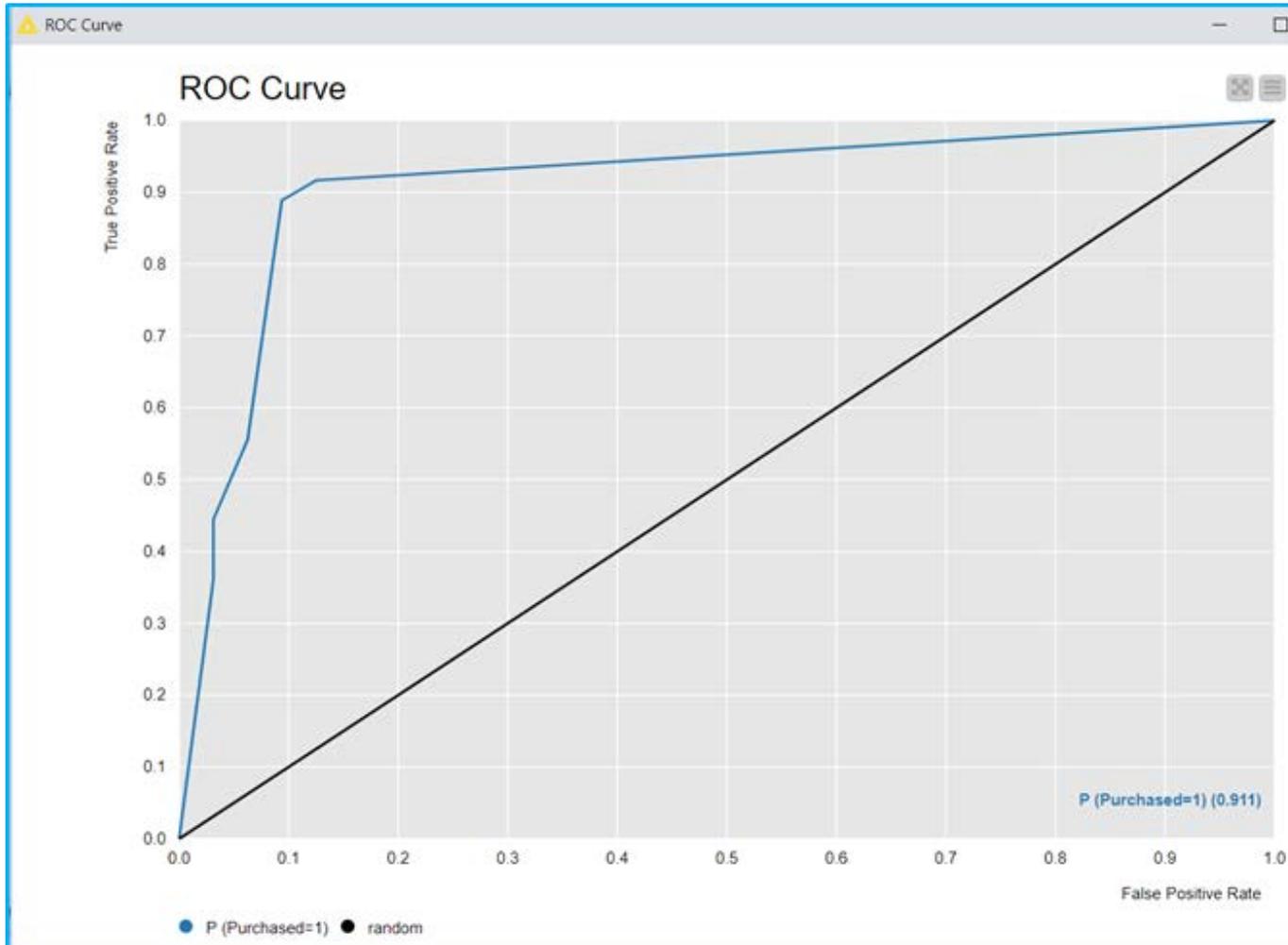
| | 0 (Predicted) | 1 (Predicted) | |
|------------|---------------|---------------|--------|
| 0 (Actual) | 58 | 6 | 90.63% |
| 1 (Actual) | 4 | 32 | 88.89% |
| | 93.55% | 84.21% | |

Class Statistics

| Class | Recall | Precision | Sensitivity | Specificity | F-measure |
|-------|--------|-----------|-------------|-------------|-----------|
| 0 | 90.63% | 93.55% | 90.63% | 88.89% | 92.06% |
| 1 | 88.89% | 84.21% | 88.89% | 90.63% | 86.49% |

Overall Statistics

| Overall Accuracy | Overall Error | Cohen's kappa (κ) | Correctly Classified | Incorrectly Classified |
|------------------|---------------|----------------------------|----------------------|------------------------|
| 90.00% | 10.00% | 0.786 | 90 | 10 |



ROC: Receiver Operating Characteristics-Kurve

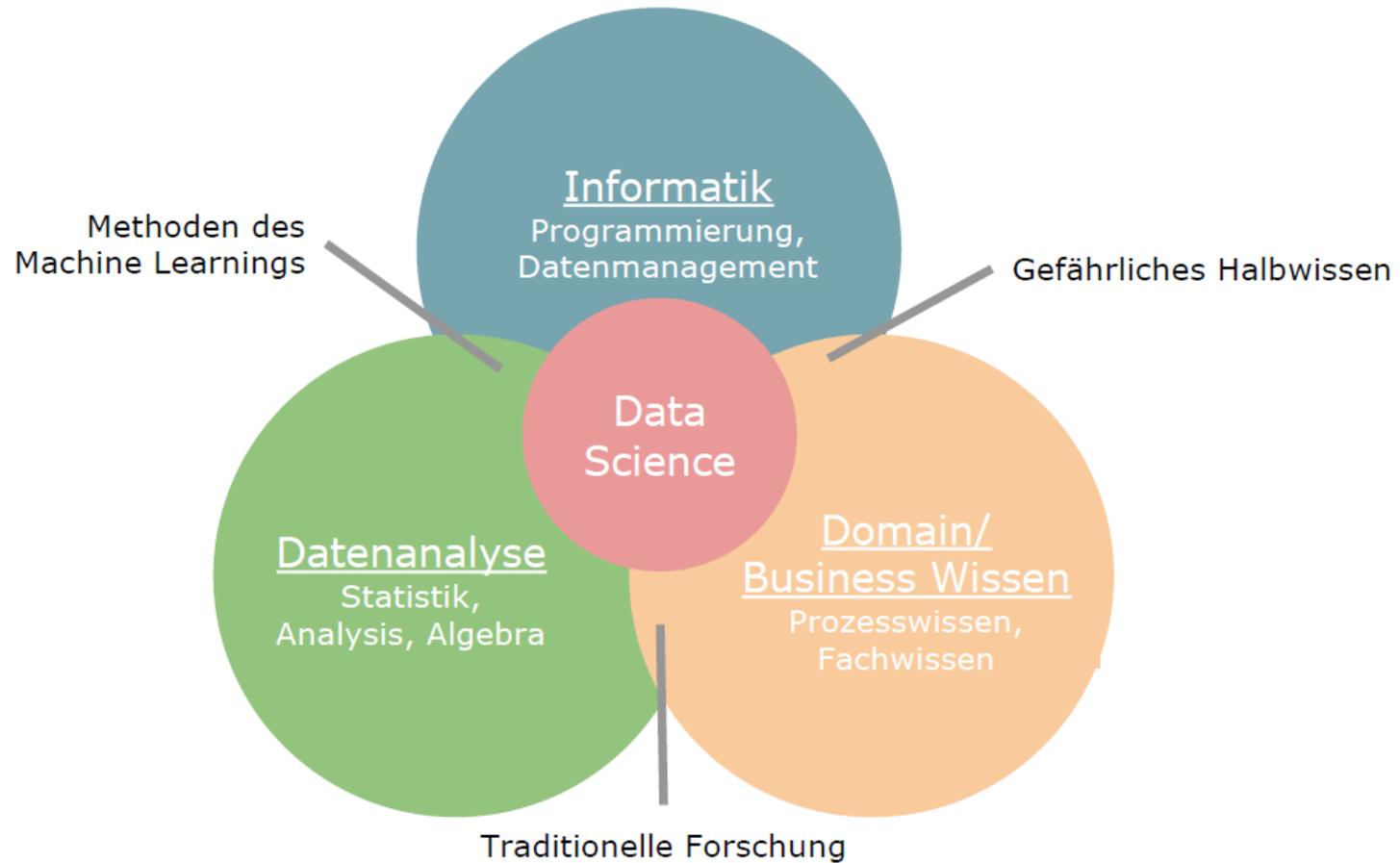
AUC: Area under the Curve (zwischen 0,5 und 1;
Referenzwert < 0,7)

True Positive Rate (TPR) =
 $\text{True Positives} / (\text{True positives} + \text{False negatives})$

False Positive Rate (TPR) =
 $\text{False Positives} / (\text{False positives} + \text{True negatives})$

1. Einordnung bisher vorgestellter Methoden in ML-Dimensionen
2. Vorstellung Data Analytics-Tool KNIME anhand eines Beispiels
3. Diskussion: Kompetenzanforderungen an Controller

Literatur zum Einstieg



Controllerprofil: zukünftig vs. heute

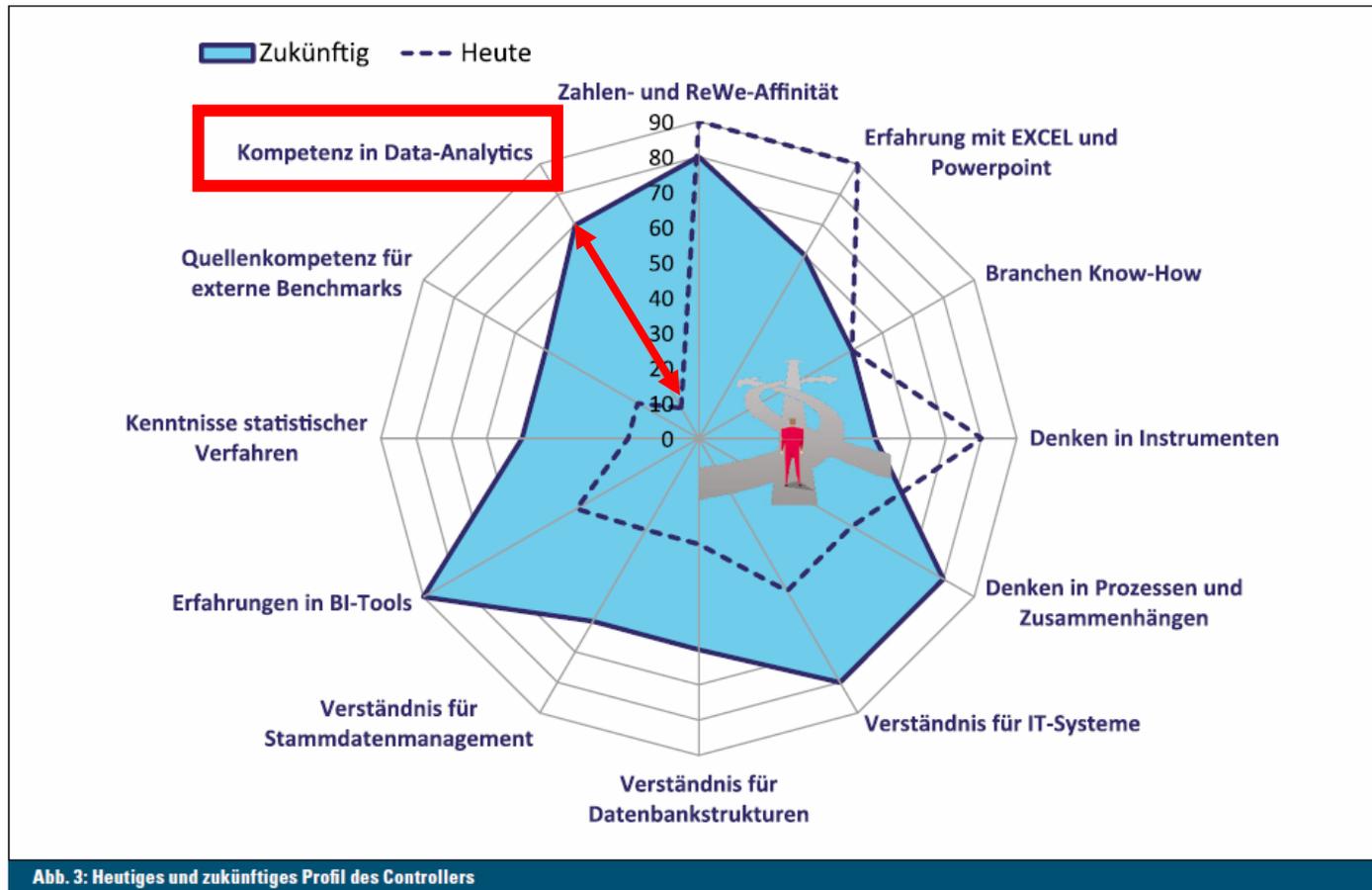
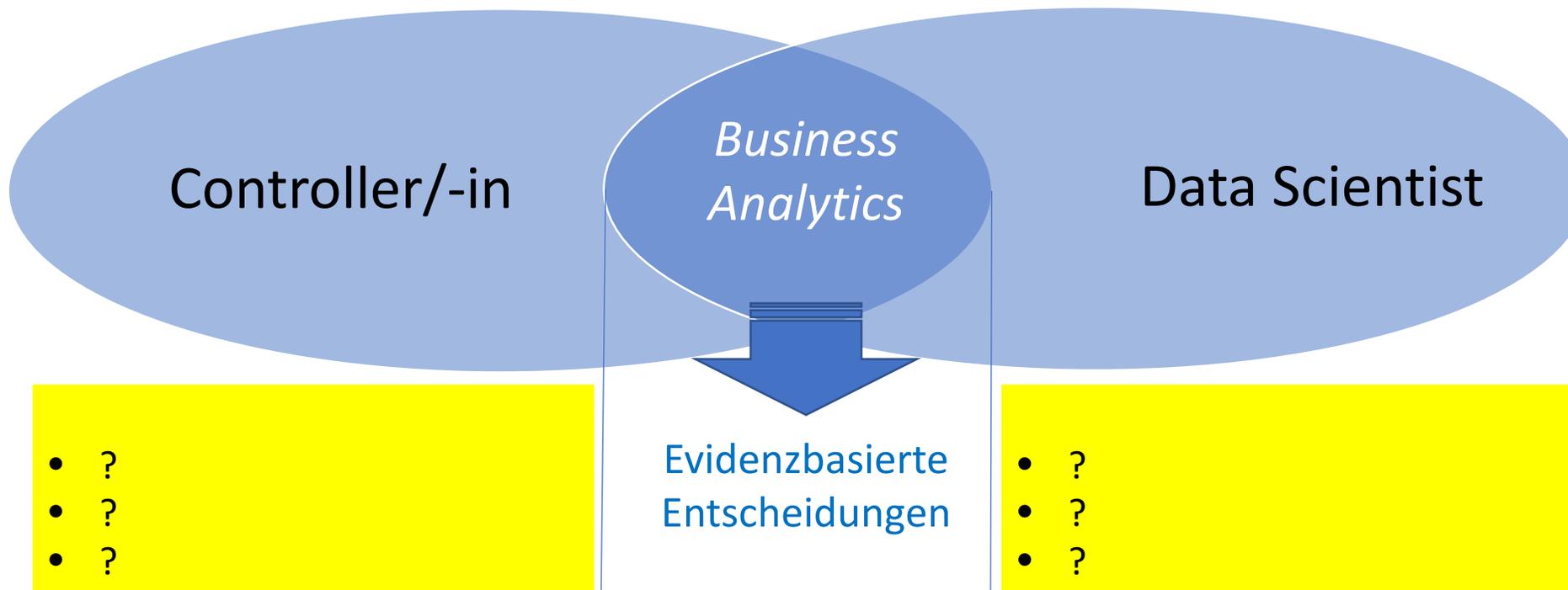


Abb. 3: Heutiges und zukünftiges Profil des Controllers

Quelle: Künstliche Intelligenz im Controlling von Günter Lubos, CONTROLLER Magazin, Heft 1 / 2020, S. 45-50

Versuch eines Zielbildes des Zusammenspiels Verantwortlichkeiten: Controller/-in vs. Data Scientist

Business Analytics: „Das Ziel von Business Analytics ist es, **betriebswirtschaftliche Probleme** im gesamten **Managementzyklus** von Planung, Steuerung und Kontrolle **evidenzbasiert** zu lösen.“*



* Quelle: Seiter, Misch (2019): Business Analytics. Wie sie Daten für die Steuerung von Unternehmen nutzen, 2. Aufl., München: Vahlen, S. 2

1. Einordnung bisher vorgestellter Methoden in ML-Dimensionen
2. Vorstellung Data Analytics-Tool KNIME anhand eines Beispiels
3. Diskussion: Kompetenzanforderungen an Controller

Literatur zum Einstieg

Albrecht, M./Schlüter, T. (2020): Erfolgsmodell Data Analytics. Use Case – Analysestrategien – Wettbewerbsvorsprünge, Berlin: Erich Schmidt Verlag.

Haneke, U./Trahasch, S./Zimmer, M./Felden, C. (Hrsg.) (2021): Data Science. Grundlagen; Architekturen und Anwendungen, Heidelberg: dpunkt.verlag

Nützliche Links:

<https://www.knime.com/>

<https://www.kaggle.com/>

<https://www.kdnuggets.com/>

BACKUP



These 1:

Der Controller kann nicht die Rolle des Data Scientists übernehmen.

These 2:

**Der Controller muss die Analysemöglichkeiten
der wichtigsten ML-Verfahren kennen.**

These 3:

**Das Controlling muss Meta-Wissen (inhaltlich, nicht technisch) über die
Datenbanken-Landschaft im Unternehmen haben.**