Data Science 2

Vorlesung - NoCode

Prof. Dr. Christian Bockermann

Hochschule Bochum

Sommersemester 2022

Inhalt

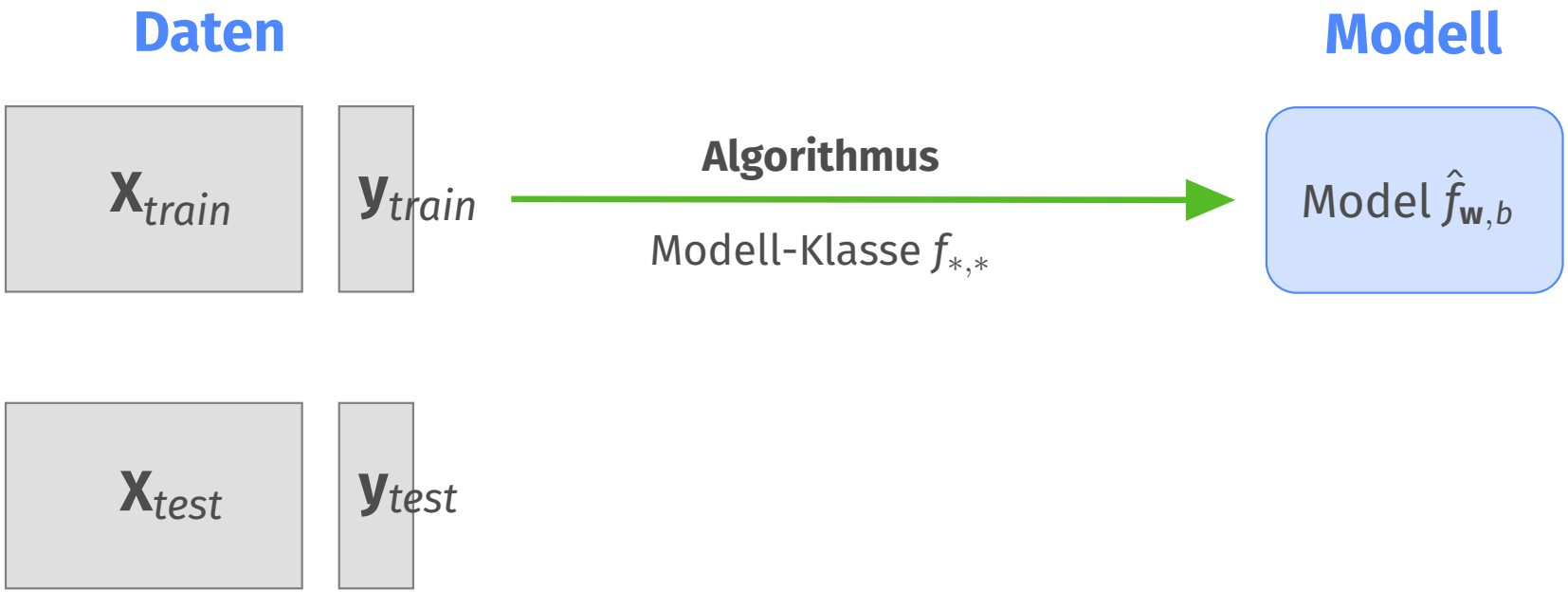
1 Datenanalyse mit Python

2 Weitere Software/Tools

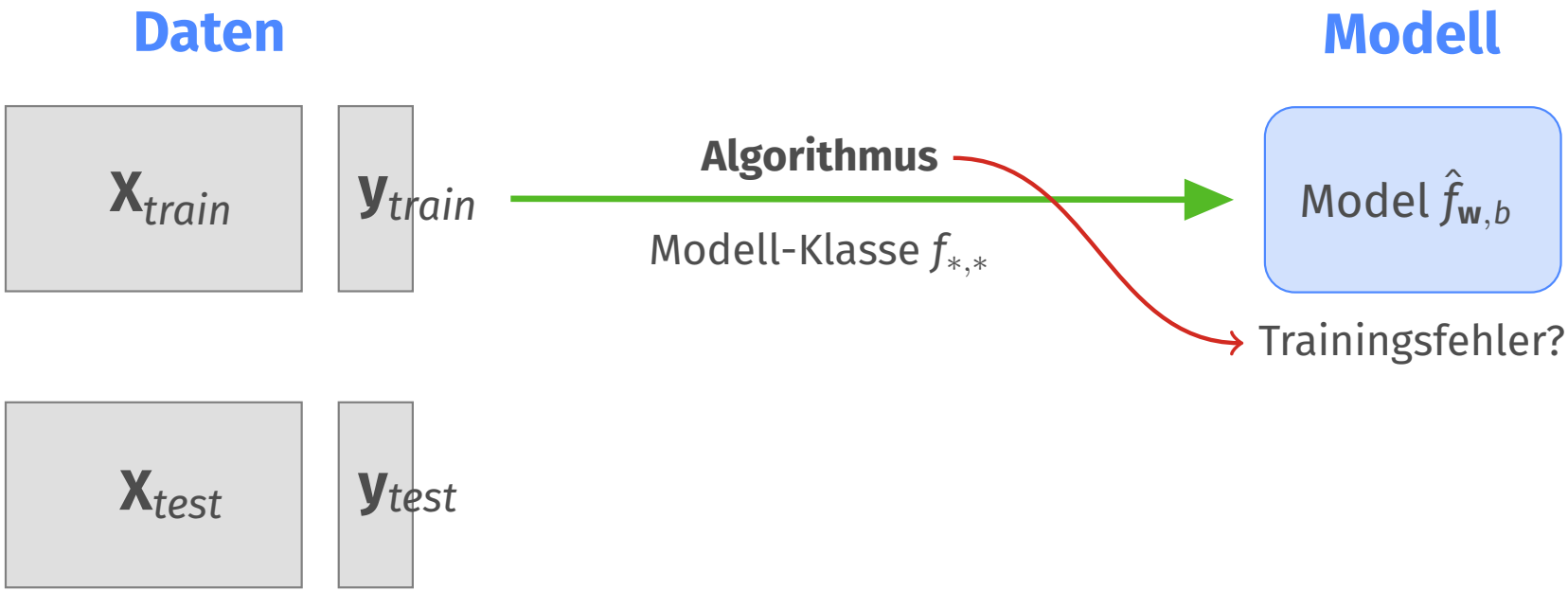
3 No-Code Ansätze

Datenanalyse mit Python

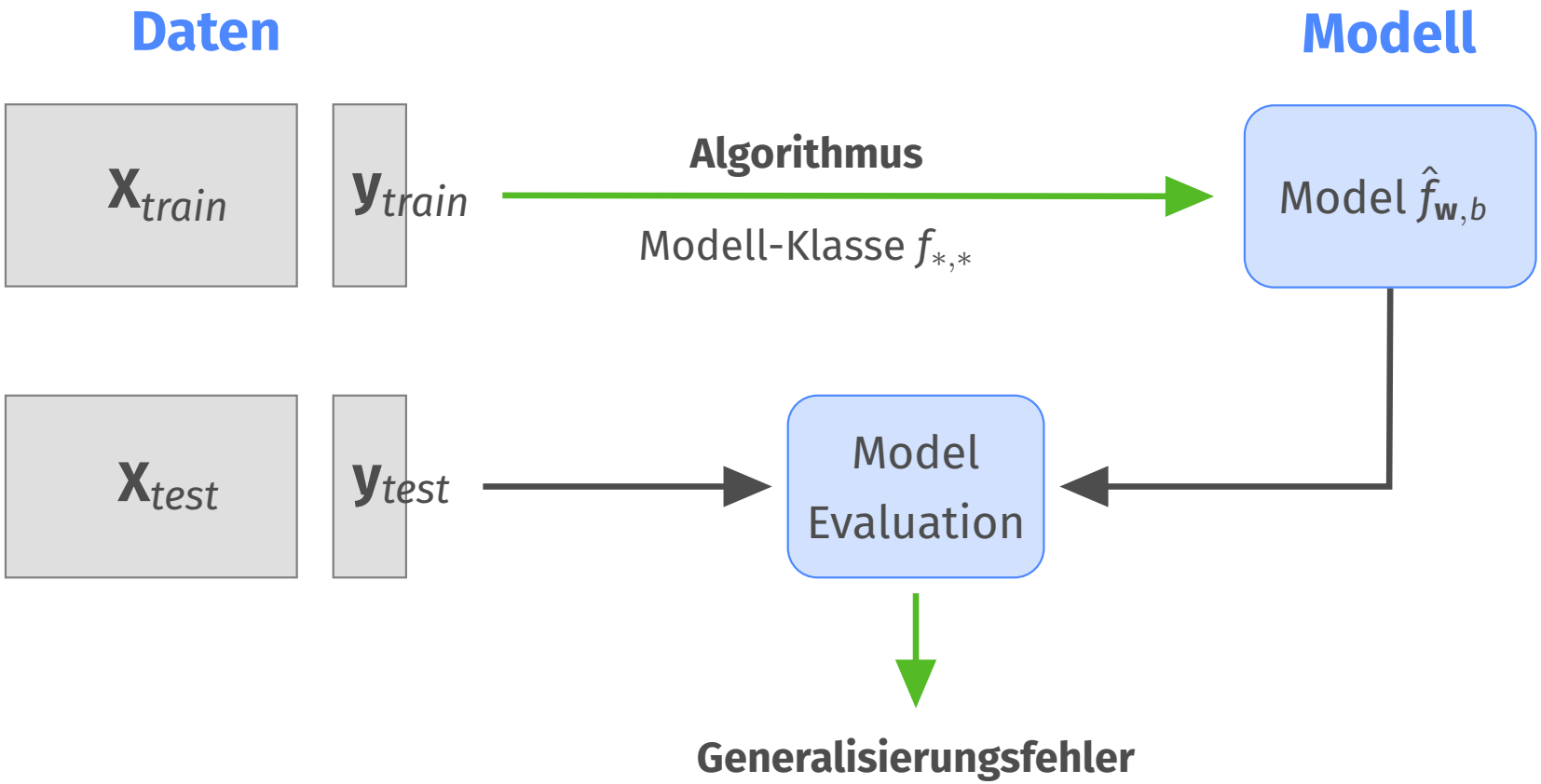
Vorgehen beim überwachten Lernen



Vorgehen beim überwachten Lernen



Vorgehen beim überwachten Lernen



```python
import pandas as pd

# read data from csv
df = pd.read_csv('daten.csv')
features = ['a1','a2','a3']

# Merkmale auswaehlen
X = df[features]
y = df['label']

# Daten aufteilen
X_tr,X_ts,y_tr,y_ts = train_test_split(X,y)

# Modell trainieren
m = DecisionTreeClassifier()
m.fit(X_tr, y_tr)
```

Programmiersprachen

- Julia, `http://julialang.org`
- Python mit Pandas, SciKit Learn `http://scikit-learn.org`
- R, `http://www.r-project.org`

Programmiersprache R für Statistik Aufgaben

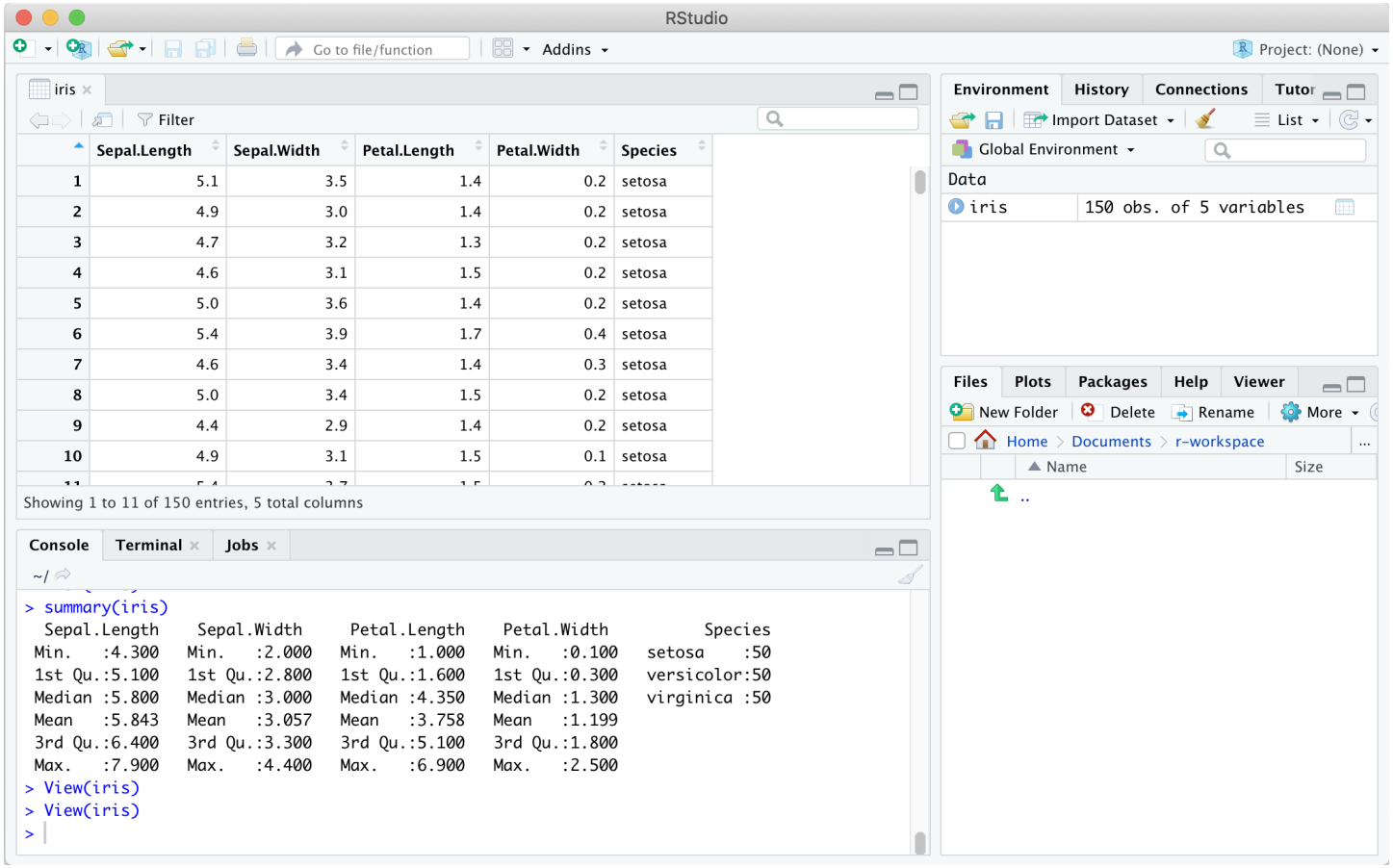


Abbildung: RStudio Umgebung für die Sprache R.

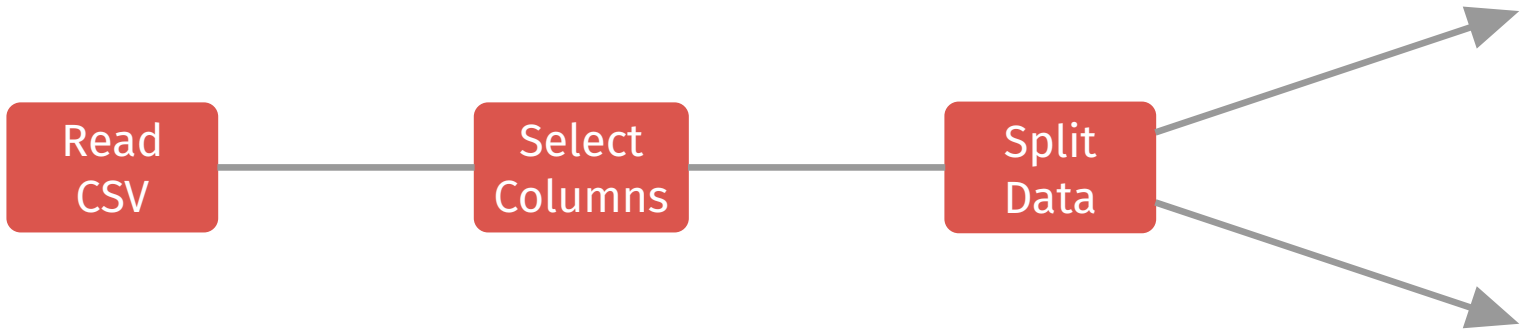
No-Code Ansätze

Trend: *No Code* Tools

- RapidMiner, `http://rapidminer.com`
- Knime, `http://www.knime.com`
- WEKA, MOA, `http://www.cs.waikato.ac.nz/ml/weka`
- Talend (Data Processing)

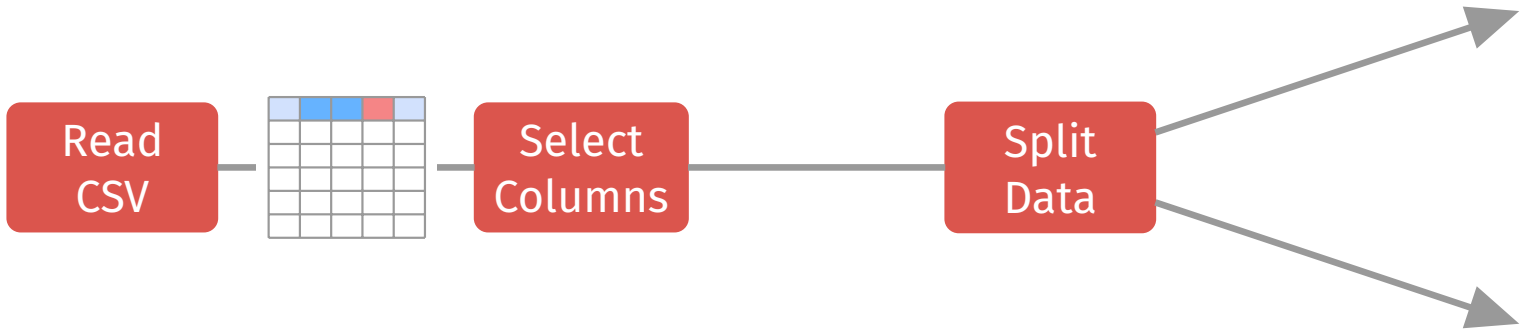
Werkzeuge um Prozesse mit graphischen Elementen zu entwerfen:

- Symbole für ausführbare Operationen
- Verbindungen zu Übergabe von Ergebnis-Objekten
- Einfache Start/Stopp Funktionen, Anhalten von Prozessen
- Möglichst ohne Programmierung auskommen



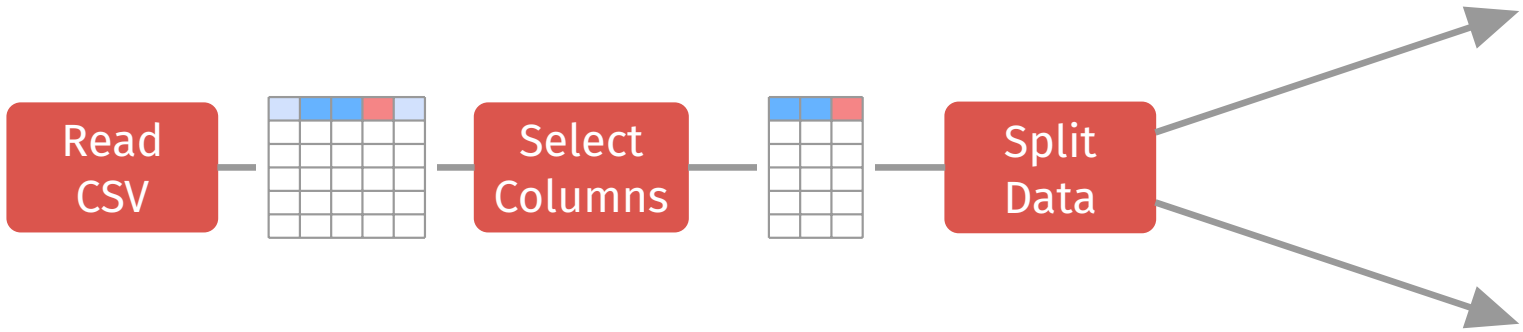
Werkzeuge um Prozesse mit graphischen Elementen zu entwerfen:

- Symbole für ausführbare Operationen
- Verbindungen zu Übergabe von Ergebnis-Objekten
- Einfache Start/Stopp Funktionen, Anhalten von Prozessen
- Möglichst ohne Programmierung auskommen



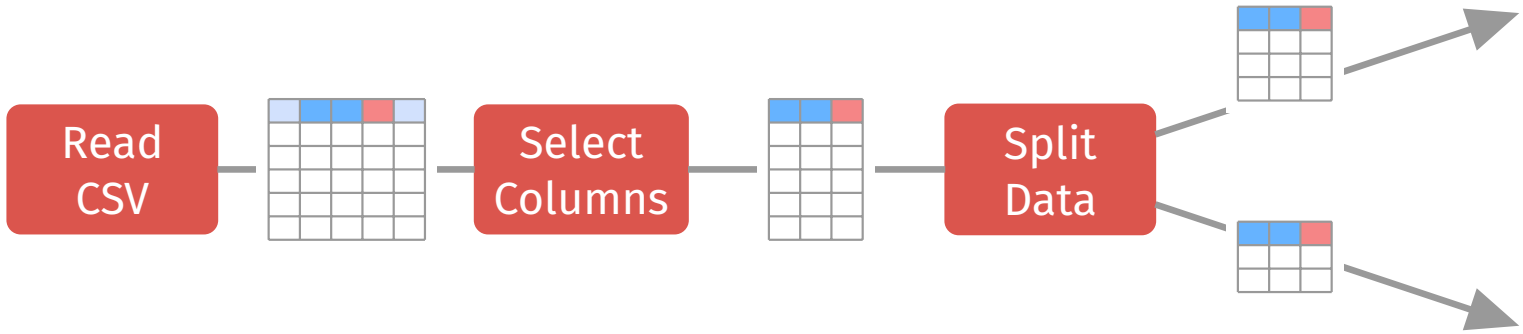
Werkzeuge um Prozesse mit graphischen Elementen zu entwerfen:

- Symbole für ausführbare Operationen
- Verbindungen zu Übergabe von Ergebnis-Objekten
- Einfache Start/Stopp Funktionen, Anhalten von Prozessen
- Möglichst ohne Programmierung auskommen



Werkzeuge um Prozesse mit graphischen Elementen zu entwerfen:

- Symbole für ausführbare Operationen
- Verbindungen zu Übergabe von Ergebnis-Objekten
- Einfache Start/Stopp Funktionen, Anhalten von Prozessen
- Möglichst ohne Programmierung auskommen



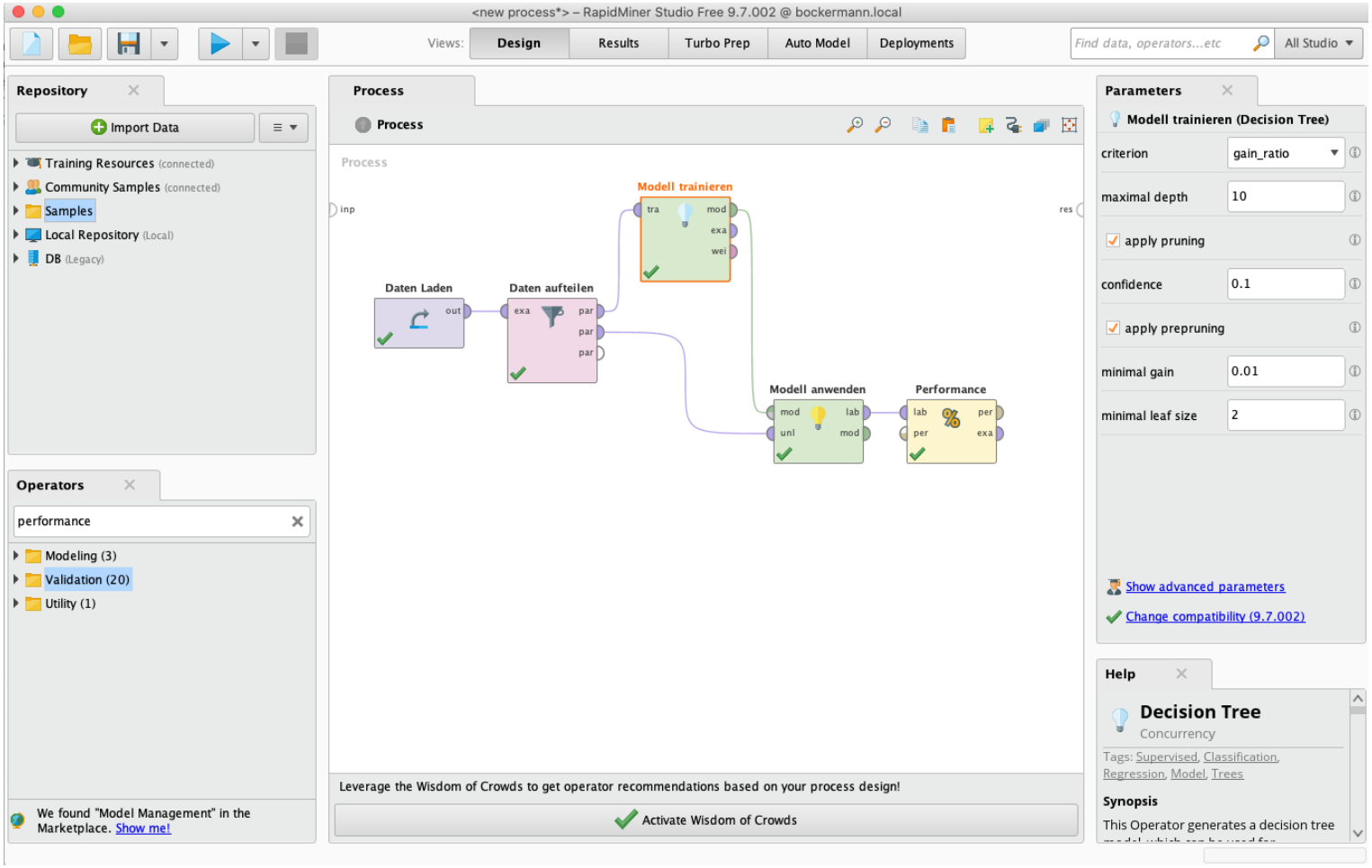


Abbildung: Die graphische Schnittstelle von RapidMiner.

Prozesse werden als Graph mit vordefinierten Operator-Bausteinen gebaut

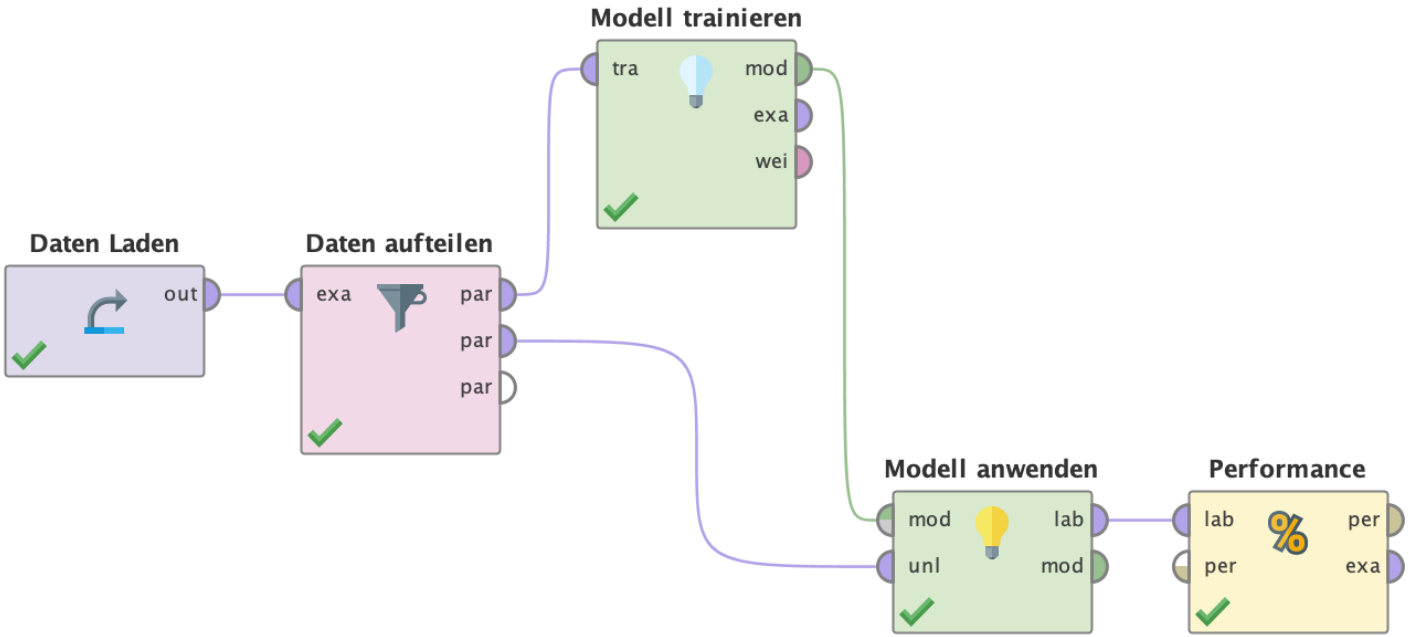


Abbildung: Ein Prozeß als Graph in RapidMiner.

RapidMiner wurde als OpenSource Tool am Lehrstuhl für künstliche Intelligenz der TU Dortmund entwickelt

- Prozess-Definition für ETL, Modellierung und Auswertung
- Einfaches Inspizieren / Exploration von Daten
- Enterprise Version für Unternehmen verfügbar
- Marktplatz mit Vielzahl von Erweiterungen
- *Wisdom of the crowds* Ansatz für schnellen Start

KNIME ist ebenfalls ein graphisches Tool für Prozess-Design

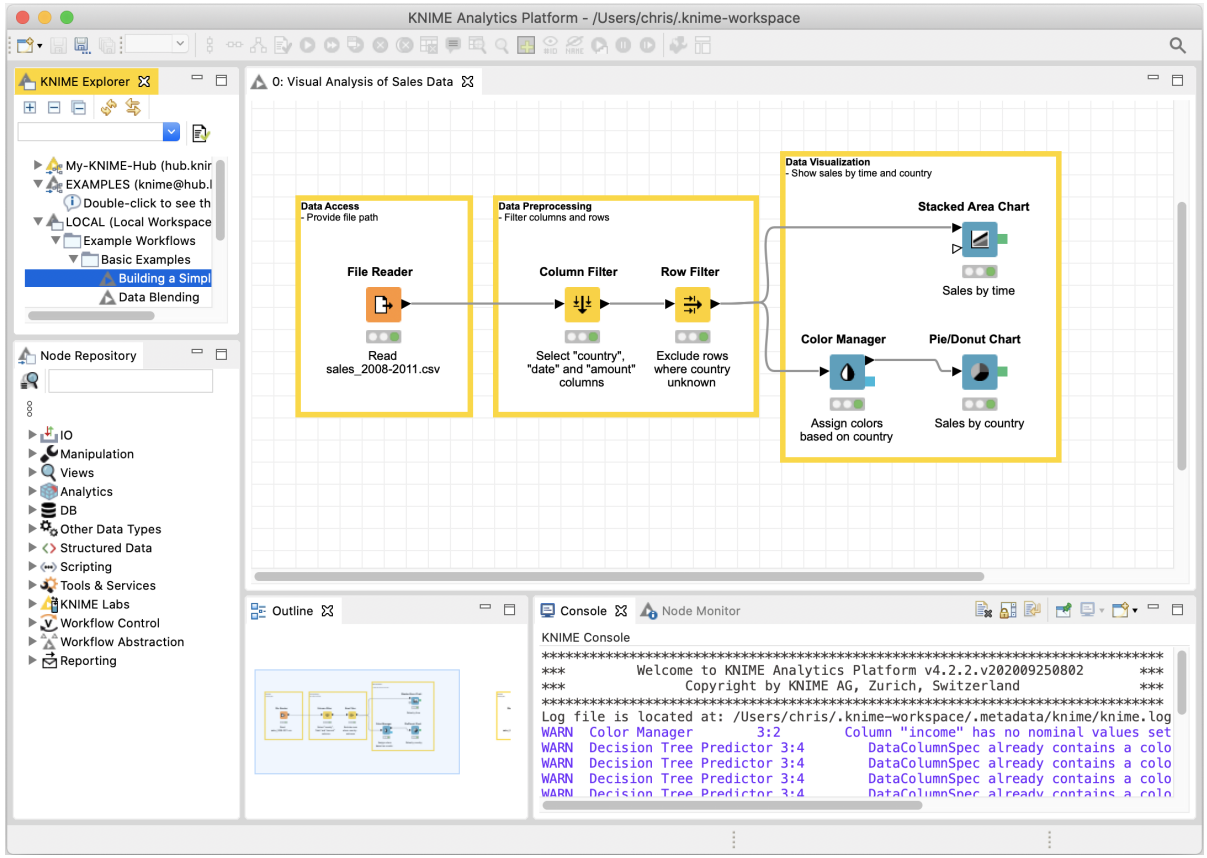


Abbildung: Die graphische Schnittstelle von KNIME.

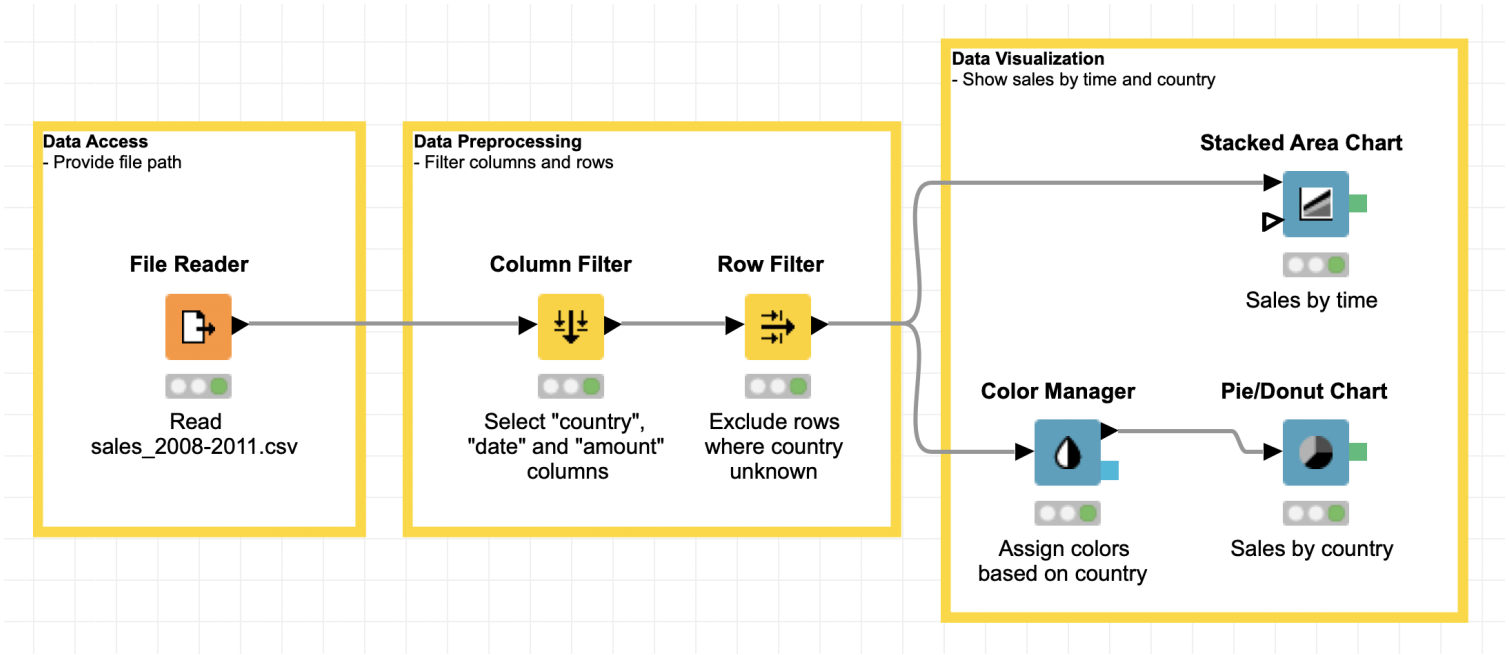


Abbildung: Ein Prozess zur Visualisierung mit KNIME.

Demo Rapidminer