

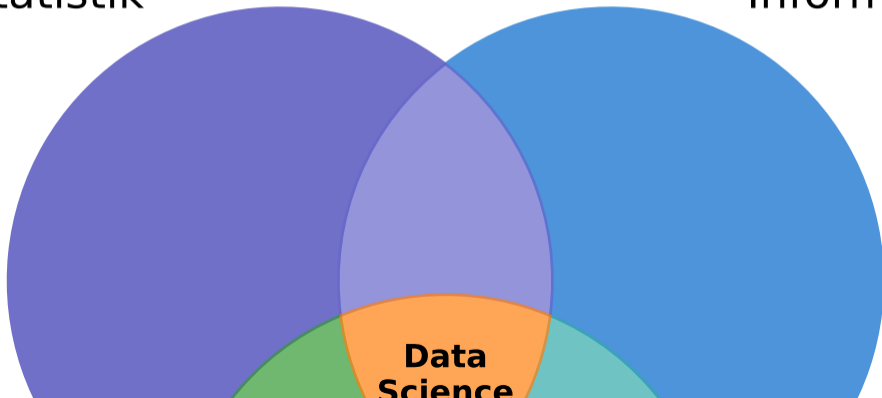
# Block 1: Data Science und Strategie

## Topic 1: Data Science Grundlagen

# Data Science als Schnittmenge

Statistik

Informatik



## Statistik

- Wahrscheinlichkeitstheorie
- Inferenz
- Hypothesentests
- Modellierung

## Informatik

- Programmierung
- Datenbanken
- Algorithmen
- Visualisierung

## Domänenwissen

- Fachexpertise
- Kontextverständnis
- Problemdefinition
- Interpretation

Nur die Kombination aller drei Bereiche ermöglicht erfolgreiche Data Science Projekte.

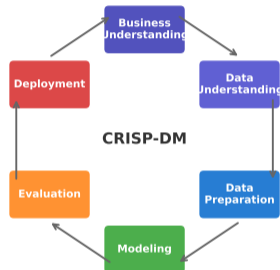
---

Alle drei Säulen sind für erfolgreiche Projekte erforderlich

## CRISP-DM: Cross-Industry Standard Process for Data Mining

### Sechs Phasen:

1. Business Understanding
2. Data Understanding
3. Data Preparation
4. Modeling
5. Evaluation
6. Deployment



Der Prozess ist **iterativ** und erlaubt Rücksprünge zu früheren Phasen.

CRISP-DM ist der Standard-Prozess für Data Science Projekte

1. **Business Understanding**  
Ziele definieren, Anforderungen klären, Projekterfolg messbar machen
2. **Data Understanding**  
Daten sammeln, Qualität prüfen, explorative Analyse durchführen
3. **Data Preparation**  
Daten bereinigen, transformieren, Features erstellen  
*Zeitaufwand: 60–80% des gesamten Projekts!*
4. **Modeling**  
Modelle auswählen, trainieren, Parameter optimieren
5. **Evaluation**  
Modellqualität bewerten, Business-Ziele überprüfen
6. **Deployment**  
Modell produktiv einsetzen, Monitoring etablieren

---

Data Preparation benötigt 60–80% der Projektzeit