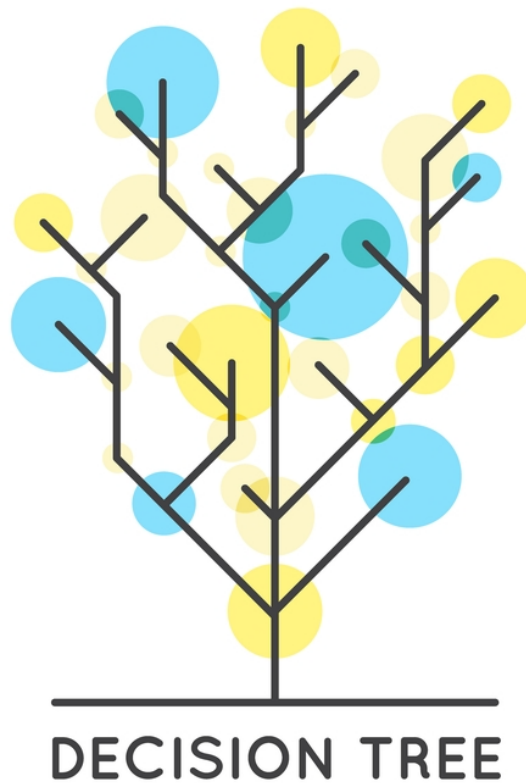


## **GDS2020\_Bericht\_Flavio\_Müller**

### **Entscheidungsbäume**



Flavio Müller

Windisch, 11.09.2020

## Inhaltsverzeichnis

1	Executive Summary	3
2	Einsatzgebiet	3
3	Funktionsweise des Algorithmus	3
3.1	Daten Teilen	3
3.2	Den Baum wachsen lassen	3
3.3	Den Baum zurückschneiden (Pruning)	3
3.4	Validating	3
3.5	Testing	3
4	Entwicklung für die Anwendung	3
5	Vor und Nachteile des Algorithmus	4
5.1	Vorteile	4
5.2	Nachteile	4
5.3	Vergleich mit anderen ML-Algorithmen	4
5.4	Verbesserungen	4
6	Anwendungsgebiete	5
6.1	Optimale Anwendungsgebiete	5
6.2	Nicht optimale Anwendungsgebiete	5
7	Quellenverzeichnis	5

# **1 Executive Summary**

## **2 Einsatzgebiet**

- Supervised Learning
- Klassifikation
- Regression

## **3 Funktionsweise des Algorithmus**

### **3.1 Daten Teilen**

- Train
- Validate
- Test

### **3.2 Den Baum wachsen lassen**

- Verschiedene Algorithmen
- Target Variable
- Error Algorithmen
- Evtl. einen Algorithmus genau analysieren

### **3.3 Den Baum zurückschneiden (Pruning)**

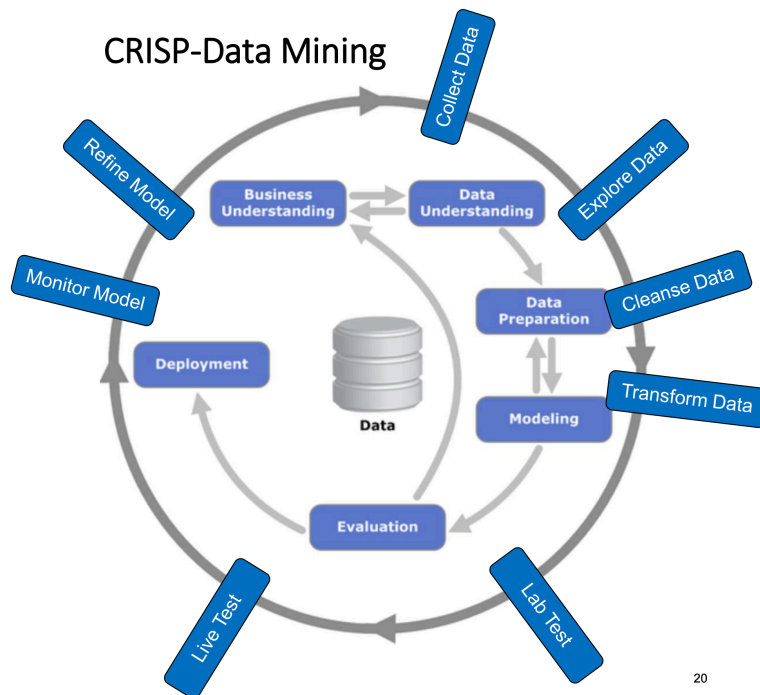
- Pre Pruning
- Post Pruning

### **3.4 Validating**

### **3.5 Testing**

## **4 Entwicklung für die Anwendung**

Schritte aus dem Crisp Data Mining Prozess



20

## 5 Vor und Nachteile des Algorithmus

### 5.1 Vorteile

- Leicht nachvollziehbar
- Einfach in der Anwendung
- Prädiktion sind auf klare Regeln zurückzuführen
- Keine Normalisierung der Daten notwendig

### 5.2 Nachteile

- Durch Schlichtheit, wenig differenziert
- Schlecht bei der Klassifizierung von Komplexen Daten z.B. Bilder
- Anfällig für Bias und Variance

### 5.3 Vergleich mit anderen ML-Algorithmen

- SVM
- KNN
- DNN

### 5.4 Verbesserungen

- Bagged Trees
- Boosted Trees
- Random Forest

## **6 Anwendungsgebiete**

### **6.1 Optimale Anwendungsgebiete**

- Versicherungswesen -> Einfachheit
- E-Commerce
- Banking
- Meist als Randomforest oder Boosted Trees
- Generell «einfache», für den Menschen verständliche Daten

### **6.2 Nicht optimale Anwendungsgebiete**

- Sehr komplexe Daten (Bilder, Audio usw.)
- Daten welche Menschen schlecht zuordnen können (Pixelwerte von Bildern)

## **7 Quellenverzeichnis**