

Interaktive Visuelle Datenanalyse I
—
Praktikum 3

Abgabetermin: 24.06.2025, 23:00 Uhr

Abnahme: ab dem 25.06.2025

Abgabeformat: zip für die Programme

Maximale Anzahl an Punkten: 22

**Logistische Regression, Evaluierung von Klassifizierern, Scatterplot, Bäume,
Interpretation der Ergebnisse**

Datensatz: Titanic.csv

Beachten Sie die Anforderungen aus der Einführung in das Praktikum!

1. Datenbereinigung (*1 Punkt*)

- Eingabe: Daten aus dem Datensatz
- Ausgabe: Bereinigte Daten
- Untersuchen Sie den Zweck der Daten sowie die Datenqualität.
- Identifizieren Sie mögliche Probleme in den Daten und beheben Sie die Fehler, falls möglich.
- Dokumentieren Sie Ihr Vorgehen.

2. Klassifizierung (*6 Punkte*)

- Eingabe: Bereinigte Daten aus dem Datensatz
- Erstellen Sie Trainings- und Testdatensätze mit den Methoden
 - 10-fold Cross-Validation
 - Bootstrapping 0.632
- Trainieren Sie folgende Klassifizierer mit diesen Trainings- und Testdatensätzen
 - Logistische Regression
 - Decision Tree
 - K-Nearest Neighbor mit $k = 3$

3. Evaluierung der Klassifikation (*4 Punkte*)

- Eingabe:
 - Trainings- und Testdaten der Klassifizierer
 - Ergebnisse der Klassifizierer
- Evaluieren Sie die Performance der Klassifizierer.
 - Berechnen Sie hierzu alle notwendigen statistischen Werte.

4. Visualisierung (*8 Punkte*)

- Eingabe:
 - Statistischen Ergebnisse der Evaluierung

-
- Stellen Sie folgende Ergebnisse visuell dar
 - Die statistischen Ergebnisse der Evaluierung
 - Die Confusion-Matrix
 - Decision Tree: Zusätzlich der generierte Baum

5. Analysieren/Interpretieren Sie die Ergebnisse. (*3 Punkte*)