

3. Praktikumsaufgabe IVDA

a) Lesen Sie den Datensatz „titanic.csv“ ein und 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.

b) Trainieren Sie die folgende Klassifizierer mit den Daten, erstellen Sie hierfür Trainings- und Testdatensätze mit den Methoden Crossvalidation mit 10 folds und Bootstrapping 0.632:

- Logistische Regression
- Decision Tree
- K-nearest neighbor mit $k=3$

c) Evaluieren Sie die Performance der Klassifizierer. Berechnen Sie hierfür die notwendigen statistischen Werte und stellen Sie diese visuell für die Klassifizierer dar. Visualisieren Sie außerdem die confusion Matrix. Beim Verfahren Decision Tree soll zudem der generierte Baum visuell dargestellt werden.

Die Abgabe der Praktikumsaufgabe ist bis zum 3. Januar 2024.