

Praktische Aufgabe

Samstag, 28. Mai 2016 15:32

- Datensatz auswählen (UCI, <https://archive.ics.uci.edu/ml/datasets.html>)
 - o <https://archive.ics.uci.edu/ml/datasets/Bike+Sharing+Dataset>
 - o <https://archive.ics.uci.edu/ml/datasets/STUDENT+ALCOHOL+CONSUMPTION>
- Gemäß CRISP-DM Vorgehen
 - o Business Understanding
 - o Data Understanding
 - Einfache deskriptive (explorative) Analyse
 - Einfache Visualisierung
 - Überprüfung der Datenqualität
 - o Data Preparation
 - Missing Values erserten
 - Nutzlose Attribute löschen
 - Korrelierte Attribute löschen
 - Ausreißer identifizieren und entfernen (Masterstudenten)
 - Stichprobe ziehen (falls erforderlich)
 - o Modelling
 - Mindestens 3 Algorithmen verwenden und Ergebnisse wenn möglich interpretieren
 - Modelle evaluieren und das beste Modell auswählen
 - Mit dem besten Modell Klassifikation (Prognose) machen
 - Künstlich Erzeugen durch löschen der Zielwerte