

Applied Analytical Data Science

- Übung 1 -

Dr. Jörg-Uwe Kietz,
Übung zur Vorlesung
Mittwoch, 14-16 Uhr Vorlesung,
16-18 Uhr Übung (alle 2 Wochen)
<http://www.kietz.ch/AADS/>

Vorbereitung

- 1) Download & Installation von RapidMiner (RM)
- 2) Installation der Weka-Extension von RM
- 3) Sich mit RM vertraut machen (Handbuch lesen)
Handbuch google: "rapidminer handbuch PDF"
- 4) Daten: <http://www.kietz.ch/AADS/AADSUebung.zip>
als Repository in RM anlegen

Business Verständniss

- Business Aufgabe: cup98doc.txt
- DM Aufgabe: IDEA_CeDR_0202.pdf, Chapter 6, page 23ff
- Beides lesen und Business-Unterstanding gemäss Vorlesung/CRISP-DM erarbeiten

Insbesondere:

- Was sind die Business Ziele unserer Aufgabe?
- Was ist das Business Erfolgskriterium?
- Was ist die daraus resultierende Data Mining Aufgabe?
- Was ist das Data Mining Erfolgskriterium?

Daten Verständnis

- cup98dic.txt beschreibt das Datenschema unserer Daten
- Initiales Data Understanding gemäss Vorlesung/CRISP-DM erarbeiten

Insbesondere:

- Sind sie ausreichend für die Aufgabe?
- Wie ist die Daten Qualität?
- Was müssen wir tun um die Datenqualität zu verbessern?
- Gibt es ein Schlüsselattribute?
- Welche Typen und Rollen sollten die Attribute in RM haben?

Vortrag jeder Gruppe für die nächste Übung

- Was sind die Business Ziele unserer Aufgabe?
- Was ist das Business Erfolgskriterium?
- Was ist die daraus resultierende Data Mining Aufgabe?
- Was ist das Data Mining Erfolgskriterium?
- Was für Daten haben wir?
- Sind sie ausreichend für die Aufgabe?
- Wie ist die Daten Qualität?
- Was müssen wir tun um die Datenqualität zu verbessern?

Abgabe: Vortrag (.pdf eine Folie quer pro Seite) jeder Gruppe per email am Montag abend vor der übernächsten Übung, jede Folie sollte Gruppen Nr. und Namen enthalten