

---

Übungen für die Vorlesung

# Webmining

– Blatt 8: Supervised Learning –


---

Datum: 06/2017

Name(n) &amp; Matr.-Nr.:



**Aufgabe1: 'Web content classification'**

Laden Sie die entsprechenden Daten aus dem moodle Kurs (aktuelle Vorlesung)

- Wir verwenden hier als Quelle: <https://archive.ics.uci.edu/ml/datasets/Farm+Ads>
- Ueberlegen Sie wie Sie die Datensätze vorverarbeiten müssten 
- als Vorschlag sind die Daten etwas vorverarbeitet im moodle Kurs hinterlegt
- erstellen Sie ein SVM Klassifikationsmodell Daten (passend kodiert) mit Rapidminer (oder Matlab, ...)
- Evaluieren Sie Ihren Klassifikator

**Aufgabe2: 'SPAM Detektion'**

Laden Sie entweder die Daten aus dem elearning System oder erheben Sie eigene (mailbox) Spam/Nicht-SPAM Daten

- Ueberlegen Sie sich, wie sie die Daten weiter vorverarbeiten könnten 
- Welche Probleme gibt es neben der Datenkodierung? 
- erstellen Sie eine Prozessierung der Daten, so dass sie einen einfachen Spam detector erhalten
- modifizieren Sie Ihre pipeline, so dass die/ein Parameter ihres Klassifikators mit optimiert werden
- nutzen Sie eine 5-fache Crossvalidierung - was ist zu beachten?

**Abgabe:**

-