

Universität Kassel
Fachbereich Mathematik/Informatik
Fachgebiet Wissensverarbeitung
Hertie-Stiftungslehrstuhl
Wilhelmshöher Allee 73
34121 Kassel

Email: hotho@cs.uni-kassel.de
Tel.: ++49 561 804-6252

Dr. Andreas Hotho

02.05.05

Knowledge Discovery

Übungsblatt 2

Sommersemester 2005

Infos zu den Rechnerübungen:

- Für die Rechnerübungen ist ein leistungsfähiger Rechner erforderlich.
- Das Fachgebiet kann Zugänge zu unseren Rechnern zur Verfügung stellen. Bitte bei Herrn Hotho melden.
- Die Rechnerübungen finden in Gruppen zu 2 bis 3 Personen statt.
- Melden Sie sich bitte im Projekt: <https://gforge.cs.uni-kassel.de/> an, und registrieren sich hier beim Projekt: KDD_Übung
- Der Informationsaustausch zur KDD Praxisübung wird über die Mailinglisten des gforge Systems erfolgen.
- **Bonus:** einen 0.3 Notensprung (also z.B. von 1.3 auf 1.0)
- **Was heißt Erfolg?**
Bis zum 11.07.2005 gibt jede Gruppe **eine** schriftliche Ausarbeitung von **maximal** 10 Seiten ab. Diese dokumentiert die Arbeit und die Ergebnisse der Gruppe, wobei der Schwerpunkt auf den durchgeführten Arbeiten liegt und erst sekundär auf den Ergebnissen.
- **Tool: Weka** (<http://www.cs.waikato.ac.nz/ml/weka/>), alternative Tools, wie Demoversion von Kommerziellen Anbietern oder auch Tools wie R können auch gern eingesetzt werden.
- **Was ist zu tun?**
Sie lösen die unten beschriebene Aufgabe!

Ein Teil der Übung am 12.7.05 wird dazu genutzt, um Ihre Erfahrungen mit dem Weka Tool vorzustellen und zu diskutieren. Dabei sollen aufgefallene Merkmale und Probleme des Tools (Benutzerfreundlichkeit,) diskutiert werden und mit möglichen alternativen Tools verglichen werden. Außerdem erhält jede Gruppe die Möglichkeit kurz ihre Lösungen sowie den Lösungsweg präsentieren (online :-).

Aufgabe

Lösen Sie die Aufgabe des KDD Cup 98 und beschreiben sie ihr Vorgehen.
<http://kdd.ics.uci.edu/databases/kddcup98/kddcup98.html>

Abstract:

This is the data set used for The Second International Knowledge Discovery and Data Mining Tools Competition, which was held in conjunction with KDD-98 The Fourth International Conference on Knowledge Discovery and Data Mining. The competition task is a regression problem where the goal is to estimate the return from a direct mailing in order to maximize donation profits.