

## Projekt „DMC – Data Mining Cup 2008 unter Beteiligung von Studierenden der FH Köln“

### Ausgangspunkt

- Data Mining beschäftigt sich mit der automatischen Analyse von Daten mit dem Ziel, lernende Modelle für Klassifikations- oder Prognoseaufgaben zu bauen.
- Der [Data Mining Cup \(DMC\)](#) wird jedes Jahr als Challenge für Studenten angeboten und hat mittlerweile eine recht große Bekanntheit in der Data Mining Community erlangt. Es geht darum, eine praxisrelevante Aufgabe des Data Mining in einem Zeitraum von ca. 6 Wochen möglichst gut zu lösen.
- Durch das WPF „DMC – Data Mining praktisch“ (Dozenten: T. Bartz-Beielstein, W. Konen) wurden im vergangenen Semester Grundlagen gelegt, s. [Unterlagen in ILIAS](#). Eine interessierte Projektgruppe kann sich auf Basis dieser Unterlagen einarbeiten (oder besteht aus Teilnehmern, die das WPF besucht haben).
- Eckdaten des Wettbewerbs
  - Registrierung: ab 10. März 2008
  - Start des Wettbewerbs: 1. April 2008
  - Ende des Wettbewerbs: 15. Mai 2008
- Eine Durchführung ist sowohl im Rahmen eines QQ2-Projektes oder einer WPF-Veranstaltung möglich (evtl. auch „Projekt“ im Rahmen des Studienverlaufplans INF, hierzu jedoch gesonderte Absprache erforderlich).

### Ziele

1. Erstmalige Teilnahme von Studierenden der FH Köln am DMC
2. Evaluierung und Optimierung verschiedener State-of-the-Art-Verfahren in Anwendung auf die Wettbewerbsaufgabe (z.B. RandomForest, Design of Experiments, SVM u.a.m.)

### Schritte

- (falls notwendig:) Einarbeitung in verschiedene Tools für DM (z.B. R, Matlab, andere Tools)
- gute Projektplanung zur optimalen Aufteilung der Workload auf verschiedene Teilnehmer
- in der „heißen“ Projektphase ca. wöchentliche Projekttreffen
- Dokumentation des Projektweges

### Voraussetzungen

- Kenntnis Grundlagen DM + Tools DM (s.o.) oder Bereitschaft zur Einarbeitung
- genügend Zeit in der „heißen“ Projektphase

### Links / Literatur

- Data Mining Cup (DMC): <http://www.data-mining-cup.de/>
- WPF „DMC – Data Mining praktisch“: <http://ilias.fh-koeln.de>
- [Witten&Frank01] Ian H. Witten, Eibe Frank: **Data Mining**, Hanser (2001). S. 238-245. Das Buch zum [WEKA-System](#). Allgemeine Einführung in Data Mining, Schwerpunkt Deci-Trees + Klassifikation. Wenige Formeln, dafür textuell gute Beschreibung. Sehr gut verständliche Erklärung der Grundzüge von Bagging & Boosting.
- [Witten&Frank05, 2. Auflage]: mehr zu WEKA, einige Neuerungen (> 2 Ex. bestellt für Lib GM)

Bitte bei Interesse umgehend mit den untenstehenden Dozenten Kontakt aufnehmen!

gez. Wolfgang Konen, Thomas Bartz-Beielstein