KDD Cup 1998 Data

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.

Usage Notes

The KDD-CUP-98 data set and the accompanying documentation are now available for general use with the following restrictions:

1. The users of the data must notify Ismail Parsa (iparsa@epsilon.com) and Ken Howes (khowes@epsilon.com) in the event they produce results, visuals or tables, etc. from the data and send a note that includes a summary of the final result.
2. The authors of published and/or unpublished articles that use the KDD-Cup-98 data set must also notify the individuals listed above and send a copy of their published and/or unpublished work.
3. If you intend to use this data set for training or educational purposes, you must not reveal the name of the sponsor PVA (Paralyzed Veterans of America) to the trainees or students. You are allowed to say "a national veterans organization"...

For more information regarding the KDD-Cup (including the list of the participants and the results), please visit the KDD-Cup-98 web page at: http://www.epsilon.com/new. While there, scroll down to Data Mining Presentations where you will find the KDD-Cup-98 web page.

Ismail Parsa
Epsilon
50 Cambridge Street
Burlington MA 01803 USA
TEL: (781) 685-6734
FAX: (781) 685-0806

Information files:

- readme. This list, listing the files in the FTP server and their contents.
- instruct.txt . General instructions for the competition.
- cup98doc.txt. This file, an overview and pointer to more detailed information about the competition.
- cup98dic.txt. Data dictionary to accompany the analysis data set.
- cup98que.txt. KDD-CUP questionnaire. PARTICIPANTS ARE REQUIRED TO FILL-OUT THE QUESTIONNAIRE and turn in with the results.
- valtargt.readme. Describes the valtargt.txt file.

Data files:

- cup98lrn.zip PKZIP compressed raw LEARNING data set. (36.5M; 117.2M uncompressed)
- cup98val.zip PKZIP compressed raw VALIDATION data set. (36.8M; 117.9M uncompressed)
- cup98lrn.txt.Z UNIX COMPRESSed raw LEARNING data set. (36.6M; 117.2M uncompressed)
- cup98val.txt.Z UNIX COMPRESSed raw VALIDATION data set. (36.9M; 117.9M uncompressed)
- valtargt.txt. This file contains the target fields that were left out of the validation data set that was sent to the KDD CUP 98 participants. (1.1M)