CS6243 HW1

Due Wed, 02/01/2012, before class starts.

1. (20 points) Identify and characterize a data set from your own experience (and different from the ones covered in lectures) that you think would benefit from data mining. Answer the following questions about the data:

a) What is the data about? (What are the instances and attributes?)
b) What benefit you hope to get from mining this data set?
c) What type of data mining (classification, clustering, etc.) you think would be relevant and briefly discuss why.
d) Briefly discuss if there are any data quality issues.

2. (20 points) Classify the following attributes as nominal, ordinal, interval, or ratio. Some cases may have more than one interpretation, so briefly indicate your reasoning if you think there may be some ambiguity.

(a) Time in terms of AM or PM.
(b) Time in terms of number of minutes.
(c) Bronze, Silver, and Gold medals as awarded at the Olympics.
(d) Day of the week.
(e) Ability to pass light in terms of the following values: opaque, translucent, transparent.
(f) Ethnicity
(g) Distance from the center of campus.
(h) US telephone area code
(i) US postal zip code
(j) Coat check number. (When you attend an event, you can often give your coat to someone who, in turn, gives you a number that you can use to claim your coat when you leave.)