

MATH3346 – Data Mining Course 2006

Assignment for Module 1

Imagine you are being hired as a new data mining specialist by the *Australian Bureau of Statistics* (ABS). The ABS is collecting large amounts of data from various sources, and is conducting the next nationwide census in a few weeks.

For more information please see: <http://www.abs.gov.au>

For this assignment, you should write a report addressing the following six questions. Your report should be around 2-3 pages long (absolute maximum 4 pages!)

1. Describe what kind of data the ABS is collecting (types, sources, formats, volume, etc).
2. How would you design a data warehouse for the ABS?
3. What kind of data preprocessing and data integration issues will you have to consider when designing such a data warehouse?
4. What kind of questions do you think the ABS would like to have addressed by data mining? And what kind of data mining projects would you perform? What would the outcomes be?
5. What challenges will you most likely face as a data miner at the ABS?
6. Are there any legal/regulatory requirements that might limit the use of data mining in the ABS? If so, describe potential solutions to overcome such limitations.

Submission

- If you do include material from some other documents (e.g. Web pages), then it should be clearly identified as such, identifiable as quoted material by layout or quotation marks, and proper attribution made. You may include URL links to external documents.
- **Deadline:** Friday 11 August 2006, 5 pm.
- Please e-mail your report **as PDF file** to Peter:

`peter.christen@anu.edu.au`
- Late penalties will apply unless you have grounds for an extension (illness, family reason, etc.). Please contact me and provide written evidence in support of your case.
- This assignment is worth 10% of your final course mark.