Data mining, Spring 2010. Group Work Assigment 1

Data mining plan for Supermarket Chain

1   The assignment

An imaginary supermarket chain X has over 500 retail outlets and over 1 million members in its loyalty card scheme, providing over 300 million sales transactions per year. Your assignment is to outline a data mining plan for the supermarkets loyalty card scheme that helps it to improve its operations.

2   Content of the plan

Outline the main components of the plan:

1. Business goals: Think about ways to improve the supermarkets shareholder values via data mining the loyalty card data (for example, factors affecting shareholder value: sales, profits, customer satisfaction, public image; you can pick one or more factor to improve)

2. Data mining process:

    (a) Data & quality: What data (about customers, stores, purchases) is available? In what format? What data quality issues need to be taken into account? How much space it takes to store the data?
    Make educated guesses or rough estimates.

    (b) Tools: Think about data mining methods to be used. Do you go for predictive (classification, regression) or descriptive (clustering, association analysis) approach?

    (c) Data Preprocessing: what kind of preprocessing might be needed?

    (d) Presentation: Think about how to present the results (Visualization, Summaries)

3. Privacy: what kind of privacy issues are associated with the data mining results? How are they dealt with?

3 Guidelines

- Get organized in the group. Time is short so you need to divide the workload between the members, but still work as a group.

- As the supermarket is imaginary, many points here are left open. Fill in the blanks by making estimates/guesses and projecting from real world supermarkets and your own experiences. As long as your plan is internally consistent it is ok.

- You can use internet freely to get information to help with the plan (e.g. web pages of real supermarket chains can be useful, look for pages targeted to investors/media). Don't get distracted too much as time is short.

- In the debrief session, present you plan on slides. Focus on content and getting your message through (Fanciness of presentation is not so important). The appropriate length of the presentation is 5–10 minutes.

- After the session, return the slides to Taru. Please remember to list the names of the group members on the slides.