PROJECT DESCRIPTION AND ASSIGNMENT

Use the association rule mining module of the Weka system to mine association rules from the following dataset.
"churn.txt"

Mine, using the Weka system, association rules from the datasets above. Keep in mind that due to the representation of frequent itemsets in Weka, this system may run out of memory when mining datasets with as few as a dozen attributes. Run several experiments with your data and the system varying the parameters until you obtain a collection of association rules that represent your data well. The following are guidelines for your experiments:

- **Code**: Use the Weka system to mine the association rules as well as for preparing the data and presenting the results. Code by yourself any functionality that you need for manipulating the data and that is not offered in the Weka system.
- **Data**:
  - You can restrict your experiments to a subset of the dataset if Weka cannot handle the whole dataset. But remember that the more representative the association rules you mine from the data, the better.
  - Use the preprocessing techniques discussed in class to select, clean, and normalize the data.
  - Define concept hierarchies over the different attributes so that you can analyze your data at different levels of generality.
- **Experiments**: After you have cleaned and selected a subset of your data (if necessary), mine association rules using different parameter (confidence, support, etc.) settings. Analyze the resulting rules and repeat the experiment with other "view" of the data given by generalizing/specializing your data according to the concept hierarchies and/or by selecting different portions of the data.
- **Results**: Assume that you as the user/miner you want to obtain association rules for decision support, for understanding the data better, and/or for increasing your company's profit. Mine rules until you obtain a collection of rules that satisfies this objective.

REPORT

- **Implemented Model (by Weka)**.
- **Written Report**: Your report should contain the following sections with the corresponding discussions:
  1. **Code Description**: Describe the code that you used/wrote. Remember to acknowledge any sources of information/code you used.
  2. **Experiments**:
     - Describe what the objective of your analysis is. Is it to understand the data better? If so, what about the data you want to understand? Or is it for decision support? If so, what decisions you need to make based on the data? Or is it for classification/characterization/discrimination purposes? Explain.
     - For each experiment you ran describe:
       - Instances: What data did you use for the experiments?
       - Any pre-processing done to improve the quality of your results.
• Your system parameters.
• Any post-processing done to improve the quality of your results.
• Analysis of results of the experiment and their significance.

3. **Summary of Results**
   • What was the best collection of association rules that you obtained? Describe.
   • Discuss the strengths and the weaknesses of your project.

For any question, you can contact with:
Morteza Haghir Chehreghani { haghir@ce.sharif.edu }
Mahmmod Neshati { neshati@ce.sharif.edu }