Data Mining

Midterm Project

Dr. Jason T.L. Wang, Professor

Department of Computer Science

New Jersey Institute of Technology

Submission Rules

- Embed your last name and first name in your project file name. For example, if your name is John Smith, your file name should read: smith_john_midtermproj.doc. Only doc or pdf file is accepted.

  No tar/zip/rar is allowed.
- Your project will automatically lose **10** points if the above submission rules are violated.
- This is a single person project.
- Submit your project file in Moodle under Midterm Project Submission Site before the due time. The project file in Moodle is considered as the final version.
- No late project is accepted. A project is late if it is not submitted in Moodle before the due time. Zero points will be given to the late project.
- In addition, email an identical copy of your project file to TA with the same file name before the due time. If your project file is not emailed to TA before the due time, it is also considered late.

  Keep a copy of your email for verification when needed.

Midterm Project

Create 10 items usually seen in Amazon, K-mart, or any other supermarkets (e.g. diapers, clothes, etc.).

(1) Create a database of 20 transactions each containing some of these items. The information can be stored in a file, or a DBMS (e.g. ORACLE).

(2) Repeat (1) by creating 4 additional, different databases each containing 20 transactions.

Using the Apriori algorithm, generate and print out all the association rules and the input transactions for each of the 5 transactional databases you created (support and confidence should be user-specified parameters, so the output should show different support and confidence values with respect to different databases).

Platforms are open

- Programming language is open; any one of the following is allowed: C, C++, C#, Java, R, Matlab, Perl, Python, Php, visual studio, PL/SQL, etc. Use any programming language of your choice (specify the programming language you use in the project).
- Operating system is open; any one of the following is allowed: Windows, Solaris Unix, Linux, Mac OS, etc.
- Hardware is open; any one of the following is allowed: PC, Laptop, Sun Sparc, etc.

Project Grading

- Your project file will be examined by me and TA. TA will email the grade of the project to you.
- Note: There is a limit on the file size in Moodle and in NJIT's email box. So, keep your project file small to avoid any problem that may occur when submitting the file in Moodle and emailing it to TA. The project file should contain the source code and documentation including **screenshots**. The screenshots are used to demonstrate the running situation of your program, particularly how the program executes and produces output based on different input data and user-specified parameter values.