# CS 634 Data Mining
# Midterm Project

**Yasser Abduallah, Instructor**
**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 or <name>.gz, <name>.zip, <name>.tar. Make sure to include the source code of your project as well as document any additional packages required to run your program.
- Your project will automatically lose **10** points if the above submission rules are violated.

- This is a single student project.

- Submit your project file in Canvas under Midterm Project Submission Site before the due time. The project file in Canvas is considered as the final version.

- No late project is accepted. A project is late if it is not submitted in Canvas before the due time. Zero points will be given to the late project.

# Midterm Project

Midterm Project due date is Sunday March 15th, 2020 at 11:59 PM

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. NJIT ORACLE or MySQL).

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. The support and confidence *should* be user-specified parameters, so the output should show different support and confidence values with respect to different databases.

**Important Notes:**

1) The purpose of this project to help you understand the algorithm, therefore, it must be "*your own*" implementation of the algorithm. If you use any existing package for the algorithm from Python, R, Matlab, or Guava, etc…, you will lose points.
   For example: Some software has a one-liner function for this algorithm, do not use them.
2) *Do not share or copy code from your peers or other resources. Your task is to implement the algorithm from scratch*

# 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, Linux, Mac OS, Ubuntu etc.

- ✓ **Hardware is open:**
  Any one of the following is allowed:
  PC, Laptop, Mac etc.

# Project Grading

- The grades will be posted on Canvas when they are completed.

- Note: There is a limit on the file size in Canvas and in NJIT's email box. So, keep your project file small to avoid any problem that may occur when submitting the file in Canvas.

- The project file must 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.

- Implementation, complexity of the code, code style, clarity of the report, and more are taking into consideration.

- Copying and sharing code with peers is prohibited and will result in 0 point for all parties that are involved.