Specification5

COP-4722

Assignment-5

The purpose of this assignment is to familiarize with the data mining concepts.

Consider a shopping database relational schema with one table as

Purchase(Customer, ItemBought)

The data for this table will be sent through email (to yourusername@fiu.edu).
The value on the first line of the data is the minimum support value.
Part I: Perform data mining manually on this given set of data values and output

- Frequent itemsets at ALL possible levels starting from L-1 that satisfy the specified minimum support value
- Only for the frequent itemset L-2,
  - enumerate all possible association rules
  - compute the confidence value and interest value for each association rule

A datafile for this assignment has been emailed to your fiu email address. If the email is not in your inbox, please check the spam folder.
The value on the first line of the data file is the minimum support value.
Note that support, confidence and interest are floating point values between 0.0 and 1.0
The minimum support value for asg5 data also follows this value range.

Part II: Validate the above results through an an embedded SQL program (asg5.java)
Create an empty table Purchase in your SQL Server:
using http://users.cis.fiu.edu/~prabakar/resource/SQL/SQL2014/createNewMarket.sql
Write an embedded SQL program (asg5.java) in Java that would construct the data set in the database server:

- prompt the user for a data file name
- delete all rows of the table
- read the data file and insert a row for each input line of purchase data

and perform data mining on this table and output all possible levels of frequent itemsets that satisfy the user specified minimum support value.

Upload Part I (manual answer) and Part II (asg5.java) together as in a zip file through the class website.
Submit a print out of these two parts along with the following signed declaration: