Business Intelligence Lab

Written exam 6/11/2013

Deliver results within 4 h from start time

Notice: use your own SQL Server credentials (the lbi account is disabled)

Exercise 1 (8 pts). For a given customer c, the frequency of purchases in weekends (FPW) is the number of distinct weekend days (Saturdays or Sundays) in which the customer made a purchase. Write a Java program Frequency.java using JDBC that reads sales from the `sales_fact_1998` table from the *foodmart* database, and then it outputs the pairs c, d where c is a `customer_id` and d is its FPW. The usage of the GROUP BY clause in SQL queries to perform computation at server side is not permitted. The usage of the WHERE and ORDER BY clauses is permitted.

What to deliver: Frequency.java, myJDBCdef.props (with only the parameters needed for a test of the program)

Exercise 2 (8 pts). Develop a SSIS package reading `sales_fact_1998` from the *foodmart* database, with the purpose of writing on a text file all pairs c, d where c is a `customer_id` and d is its FPW. The usage of GROUP BY / WHERE / ORDER BY clauses in SQL queries to perform computation at server side is not permitted. All the work must be done by the SSIS package.

What to deliver: BIDS/SSDT solution.

Exercise 3 (8 pts). Write a MDX query to output the following report on the Sales cube of the `ruggieri_foodmart` OLAP database:

- for each customer, its FPW value

What to deliver: MDX query and a brief comment about it, a PowerPoint file with the screenshot of the MDX query result.

Exercise 4 (8 pts). Consider the J48 decision tree classifiers in Weka. Show experimentally on the `census.arff` dataset whether the following statement is true or false:

- for a fixed test set of 1000 instances, the larger is the training set the more accurate is the classifier.

What to deliver: a PowerPoint file with screenshots of Weka explorer, description of the steps of the analysis.

How to deliver: send an e-mail with a single <your surname>.zip file attached to ruggieri@di.unipi.it, including your name, surname, student ID, and computer IP address (http://www.whatismyip.com).

Results and oral exam. Results will be published on-line by today evening. Oral exams will start tomorrow at 9:00 at teacher office.