Business Intelligence Lab

Written exam 15/1/2014

Deliver results within 4 h from start time

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

Exercise 1 (8 pts). Consider a string, in input as a parameter, of the form "A_1, ..., A_n". Write a Java program Rollup.java that writes to a CSV file the result of the SQL query "SELECT A_1, ..., A_n, COUNT(*) FROM table GROUP BY ROLLUP(A_1, ..., A_n)", where "table" is the name of a table in input as another parameter. The usage of the GROUP BY clause in SQL queries to perform computation at server side is permitted, but the use of the ROLLUP keyword is not.

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

Exercise 2 (8 pts). Develop a SSIS package reading `customer` from the *foodmart* database, with the purpose of writing on a text file c, i where c is a `customer_id` and i ranges from 0 to the number `num_children_at_home` × `num_cars_owned` of that customer. 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 city, the sales and the margins of the product with the highest sales and of the product with the highest margin.

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). In the problem of *bi-classification*, there are two class attributes (e.g., sex and income) to be predicted by a classifier for a given tuple. Discuss how to solve the problem of bi-classification using standard decision trees. Provide an example, as accurate as possible, of your solution using the J48 decision tree classifiers in Weka and the `census` dataset (either in .arff format or in relational format).

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

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 teacher office.