Business Intelligence Lab

Written exam 6/6/2013

Deliver exercises 1,2 within 2 h from start time
Deliver exercises 3,4 within 4 h from start time

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

Exercise 1 (10 pts). A customer is called *alive* in a month if he buyes something in that month or in a later month. Develop a SSIS package reading `sales_fact_1998` from the *foodmart* database, with the purpose of populating a table `*your_account*_alive` in the *lbi* database containing a row with `customer_id` and `the_month` for each customer `customer_id` alive in the month `the_month` of year 1998. The usage of 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 2 (6 pts). Solve Exercise 1 by developing a Java program Alive.java using JDBC. The usage of any type of SQL queries is now permitted.

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

Exercise 3 (4 pts). Write a MDX query to answer the following question on the Sales cube of the `ruggieri_foodmart` OLAP database:

- For each month, how many stores did sell at least one unit of the brand "Amigo"?

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

Exercise 4 (12 pts). In the problem of *ordinal classification*, the class attribute takes values on an ordinal scale from 1 to N. The prediction error is a function of the distance between the predicted class and the actual class value. Let us set as prediction error the Mean Absolute Error (MAE):

$$MAE = \frac{1}{M} \sum_{i,j=1}^{N} m_{ij}|i-j|$$

where m_{ij} is the number of cases in the test set with predicted class i and actual class j, and M is the total number of cases in the test set. Define a data mining process to solve an ordinal classification problem, and apply the proposed process to the dataset provided by the teacher. Discuss the results obtained.

What to deliver: either a Weka knowledge flow `.kfml` file or 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, with your name, surname, student ID, and computer IP address (http://www.whatismyip.com).

Results and oral exam. Results will be published on-line by tomorrow morning. Oral exams will start tomorrow afternoon at 14:30 at teacher office.