PROJECT EVALUATION CRITERIA

Data Mining

Introduction

The present document summarizes the evaluation guidelines for the Data Mining main practical components. Please read it carefully to understand how your group work will be evaluated.

I.   Predictive Models

Build a predictive model for a dependent variable representing Customers that answer positively to a direct marketing campaign.

Deliverables

The following material must be submitted through Moodle platform within deadlines defined in the syllabus:

a) A written report, in pdf, that summarizes the analytical process and business conclusions.

 b) An excel template (provided by teachers) filled with information pertaining to the results of each group. This template must be sent alongside the pdf report. The name of the template delivered must be GROUP_XX_PM.xlsx

c) On the day of the oral presentation the group must bring the Enterprise Miner project in a Pen Drive should any questions arise that need to be checked within the project.

Guidelines for Project Evaluation

Standard criteria (18 points)

This component aims to evaluate the quality of the group work around what has been presented and discussed in practical classes:

a) Provided list of customers to be contacted. This list will be compared with the teacher's dataset containing the correct answers and used to calculate the real campaign Profit (in the project students can only estimate the profit)

b) The quality and synthesis of the written report.

c) Number of variables present in the model used to score the customers. The fewer the better. Each variable above 10 will "eat" 50€ from the profit in a).

d) Quality of proposed transformations

e) Total Number and % of correctly identified outliers (list provided by students) when compared with the teachers dataset containing the right answers.

f) Correctness of the metadata;

Advanced criteria (remaining 2 points)

To stimulate curiosity and self-learning skills, this component awards a bonus to groups willing to investigate themes discussed in theoretical classes but not covered/ presented in practical classes. This translates to the use of new nodes in the predictive modelling diagram from the list below (although they may not be part of the final best solution):

- Impute Node
- Variable Selection Node
- Principal Component Node
- MBR Node
- Ensemble node

If groups choose to use these bonus nodes, an additional appendix must be provided in the pdf report. This appendix will contain a theoretical presentation of the algorithm(s) used and a justification of the configuration options of the node(s). If a group chooses to go for the advanced criteria, this component will be a MANDATORY discussion topic in the oral defenses.

Discount criteria

a) Delay in the delivery of the project (1 point for each day of delay)

b) Excel template with considerable layout problems (1-3 points).